CHEM1 VIRTUAL TEXTBOOK



Stephen Lower
Simon Fraser University

LibreTexts™

Simon Fraser University

Chem1 Virtual Textbook

Stephen Lower

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 08/19/2026

TABLE OF CONTENTS

Licensing

1: Fundamentals of Science and Chemistry
- 1.1: Introduction to Chemistry
- 1.2: Pseudoscience

2: Essential Background
- 2.1: Classification and Properties of Matter
- 2.2: Energy, Heat, and Temperature
- 2.3: The Measure of Matter
- 2.4: The Meaning of Measure

3: Measuring Matter
- 3.1: Units and Dimensions
- 3.2: The Meaning of Measure
- 3.3: Significant Figures and Rounding off
- 3.4: Reliability of a measurement
- 3.5: Drawing Conclusions from Data

4: The Basics of Chemistry
- 4.1: Atoms, Elements, and the Nucleus
- 4.2: Avogadro's Number and the Mole
- 4.3: Formulas and Their Meaning
- 4.4: Chemical Equations and Stoichiometry
- 4.5: Introduction to Chemical Nomenclature
- 4.6: Significant Figures and Rounding

5: Atoms and the Periodic Table
- 5.1: Primer on Quantum Theory
- 5.2: Quanta - A New View of the World
- 5.3: Light, Particles, and Waves
- 5.4: The Bohr Atom
- 5.5: The Quantum Atom
- 5.6: Atomic Electron Configurations
- 5.7: Periodic Properties of the Elements
- 5.8: Why Don't Electrons Fall into the Nucleus?

6: Properties of Gases
- 6.1: Observable Properties of Gas
- 6.2: Ideal Gas Model - The Basic Gas Laws
- 6.3: Dalton's Law
- 6.4: Kinetic Molecular Theory (Overview)

6.5: More on Kinetic Molecular Theory
6.6: Real Gases and Critical Phenomena

7: Solids and Liquids

7.1: Matter under the Microscope
7.2: Intermolecular Interactions
7.3: Hydrogen-Bonding and Water
7.4: Liquids and their Interfaces
7.5: Changes of State
7.6: Introduction to Crystals
7.7: Ionic and Ion-Derived Solids
7.8: Cubic Lattices and Close Packing
7.9: Polymers and Plastics
7.10: Colloids and their Uses
Index

8: Solutions

8.1: Solutions and their Concentrations
8.2: Thermodynamics of Solutions
8.2.2A: Solutions of Gaseous Solutes in Gaseous Solvents
8.2.2B: Solutions of Gaseous Solutes in Liquid Solvents
8.2.2C: Solutions of Liquid Solutes in Liquid Solvents
8.2.2D: Solutions of Solid Solutes in Liquid Solvents
8.3: Colligative Properties- Raoult's Law
8.4: Colligative Properties- Boiling Point Elevation and Freezing Point Depression
8.5: Colligative Properties - Osmotic Pressure
8.6: Reverse Osmosis
8.7: Colligative Properties and Entropy
8.8: Ideal vs. Real Solutions
8.9: Distillation
8.10: Ions and Electrolytes
8.10.9A: Electrolytes and Electrolytic Solutions
8.10.9B: The nature of ions in aqueous solution
8.10.9C: Weak and Strong Electrolytes
8.10.9D: Ionic migration
8.10.9E: Some applications of electrolytic conduction

9: Chemical Bonding and Molecular Structure

9.1: Three Views of Chemical Bonding
9.2: Molecules - Properties of Bonded Atoms
9.3: Models of Chemical Bonding
9.4: Polar Covalence

9.5: Molecular Geometry
9.6: The Hybrid Orbital Model
9.7: The Hybrid Orbital Model II
9.8: Molecular Orbital Theory
9.9: Bonding in Coordination Complexes
9.10: Bonding in Metals
9.11: Bonding in Semiconductors
9.12: The Shared-Electron Covalent Bond

10: Fundamentals of Acids and Bases

10.1: Introduction to Acids and Bases
10.2: Aqueous Solutions- pH and Titrations
10.3: Acid-base reactions à la Brønsted
10.4: Acid-Base Reactions
10.5: Lewis Acids and Bases
10.6: Types of Acids and Bases
10.7: Acid-Base Gallery

11: Chemical Equilibrium

11.1: Introduction to Chemical Equilibrium
11.2: Le Chatelier's Principle
11.3: Reaction Quotient
11.4: Equilibrium Expressions
11.5: Equilibrium Calculations
11.6: Phase Distribution Equilibria

12: Solubility Equilibria

13: Acid-Base Equilibria

13.1: Introduction to Acid/Base Equilibria
13.2: Strong Monoprotic Acids and Bases
13.3: Finding the pH of weak Acids, Bases, and Salts
13.4: Conjugate Pairs and Buffers
13.5: Acid/Base Titration
13.6: Applications of Acid-Base Equilibria
13.7: Exact Calculations and Approximations

14: Thermochemistry

14.1: Energy, Heat and Work
14.2: The First Law of Thermodynamics
14.3: Molecules as Energy Carriers and Converters
14.4: Thermochemistry and Calorimetry
14.5: Calorimetry
14.6: Applications of Thermochemistry

14.E: Thermochemistry (Exercises)

15: Thermodynamics of Chemical Equilibria

15.1: Energy Spreading Drives Spontaneous Change

15.2: Entropy Rules

15.3: The Second Law of Thermodynamics

15.4: Free Energy and the Gibbs Function

15.5: Thermodynamics of Mixing and Dilution

15.6: Free Energy and Equilibrium

15.7: Some Applications of Entropy and Free Energy

15.8: Quantum states, Microstates, and Energy spreading in Reactions

16: Electrochemistry

16.1: Chemistry and Electricity

16.2: Galvanic cells and Electrodes

16.3: Cell Potentials and Thermodynamics

16.4: The Nernst Equation

16.5: Applications of the Nernst Equation

16.6: Batteries and Fuel Cells

16.7: Timeline of Battery Development

16.8: Electrochemical Corrosion

16.9: Corrosion Gallery

16.10: Electrolytic Cells and Electrolysis

17: Chemical Kinetics and Dynamics

17.1: Rates of reactions and rate laws

17.2: Reaction Rates Typically Change with Time

17.3: Collision and activation- the Arrhenius law

17.4: Reaction Mechanisms

17.5: Kinetics of Reactions in Solution

17.6: Catalysts and Catalysis

17.7: Experimental methods of chemical kinetics

Index

Glossary

Detailed Licensing

This page titled Chem1 (Lower) is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

This page titled Licensing is shared under a license and was authored, remixed, and/or curated by LibreTexts.

Simon Fraser University

7: Solids and Liquids

Stephen Lower

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden born from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and was supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 08/19/2026

1: FUNDAMENTALS OF SCIENCE AND CHEMISTRY

1.1: Introduction to Chemistry

1.2: Pseudoscience

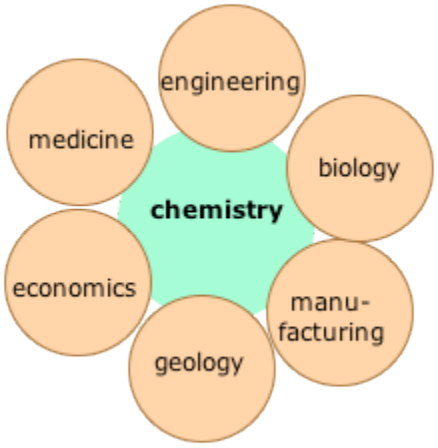
This page titled 1: Fundamentals of Science and Chemistry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

1.1: Introduction to Chemistry

Learning Objectives

- Distinguish beween *chemistry* and *physics*;
- Suggest ways in which the fields of engineering, economics, and geology relate to Chemistry;
- Define the following terms, and classify them as primarily *microscopic* or *macroscopic* concepts: element, atom, compound, molecule, formula, structure.
- The two underlying concepts that govern chemical change are *energetics* and *dynamics*. What aspects of chemical change does each of these areas describe?

Chemistry is a very universal and dynamically-changing subject to be confined to a fixed definition; it might be better to think of chemistry more as a *point of view* that places its major focus on the structure and properties of *substances*— particular kinds of matter— and especially on the changes they undergo.



The real importance of Chemistry is that it serves as the interface to practically all of the other sciences, as well as to many other areas of human endeavor. For this reason, Chemistry is often said (at least by chemists!) to be the "central science". Chemistry can be "central" in a much more personal way: with a solid background in Chemistry, you will find it far easier to migrate into other fields as your interests develop.

> Research or teaching not for you? Chemistry is so deeply ingrained into so many areas of business, government, and environmental management that some background in the subject can be useful (and able to give you a career edge as a team member having special skills) in fields as varied as product development, marketing, management, computer science, technical writing, and even law.

So just what is chemistry?

Do you remember the story about the group of blind men who encountered an elephant? Each one moved his hands over a different part of the elephant's body— the trunk, an ear, or a leg— and came up with an entirely different description of the beast. Chemistry can similarly be approached in different ways, each yielding a different, valid, and yet hopelessly incomplete view of the subject. Thus we can view chemistry from multiple standpoints ranging from the theoretical to the eminently practical:

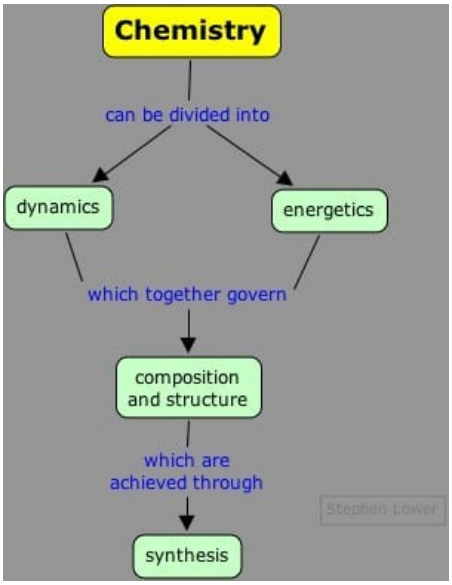
Mainly theoretical	Mainly practical
Why do particular combinations of atoms hold together, but not others?	What are the properties of a certain compound?
How can I predict the shape of a molecule?	How can I prepare a certain compound?

Mainly theoretical	Mainly practical
Why are some reactions slow, while others occur rapidly?	Does a certain reaction proceed to completion?
Is a certain reaction possible?	How can I determine the composition of an unknown substance?

Boiling it down to the basics

At the most fundamental level, chemistry can be organized along the lines shown here.

- *Dynamics* refers to the details of that rearrangements of atoms that occur during chemical change, and that strongly affect the rate at which change occurs.
- *Energetics* refers to the thermodynamics of chemical change, related to the uptake or release of heat. This aspect of chemistry controls the direction in which change occurs, and the mixture of substances that are produced as a result.
- *Composition and structure* define the substances that are produced because of a chemical change. *Structure* specifically refers to the relative arrangements of the atoms in space. The extent to which a given structure can persist is determined by energetics and dynamics.
- *Synthesis* refers to formation of new (and usually more complex) substances from simpler ones, but in the present context we use it in the more general sense to denote the operations required to bring about chemical change and to isolate the desired products.



This view of Chemistry is a rather a stringent one that is probably more appreciated by people who already know the subject than by those who are about to learn it, so we will use a somewhat expanded scheme to organize the fundamental concepts of chemical science. But if you need a single-sentence "definition" of Chemistry, this one wraps it up pretty well:

> Chemistry is the study of *substances*; their properties, structure, and the changes they undergo.

Micro-macro: the forest or the trees

Chemistry, like all the natural sciences, begins with the direct observation of nature — in this case, of matter. But when we look at matter in bulk, we see only the "forest", not the "trees" — the atoms and molecules of which matter is composed of — whose properties ultimately determine the nature and behavior of the matter we are looking at. This dichotomy between what we can and cannot directly see constitutes two contrasting views which run through all of chemistry, which we call *macroscopic* and *microscopic*.

In the context of Chemistry, "microscopic" implies the atomic or subatomic levels which cannot be seen directly (even with a microscope!) whereas "macroscopic" implies things that we can know by direct observations of physical properties such as mass,

volume, etc. The following table provides a conceptual overview of Chemical science according to the macroscopic/microscopic dichotomy we have discussed above. It is of course only one of the many ways of looking at the subject, but you may find it helpful in organizing the many facts and ideas that you will encounter in your study of Chemistry. We will organize the discussion in this lesson along similar lines.

realm	macroscopic view	microscopic view
composition	formulas, mixtures	structures of solids, molecules, and atoms
properties	intensive properties of bulk matter	particle sizes, masses and interactions
change (energetics)	energetics and equilibrium	statistics of energy distribution
change (dynamics)	kinetics (rates of reactions)	mechanisms

Chemical composition

Mixture or "pure substance" ?

In science it is necessary to know what we are talking about, so before we can begin to consider matter from a chemical point of view, we need to know its *composition*; whether it is a *single substance*, or a *mixture*? (We will get into the details of the definitions later, but for the moment you probably have a fair understanding of the distinction; think of a sample of salt (sodium chloride) as opposed to a *solution* of salt in water— a *mixture* of salt and water.)

Elements and compounds

It has been known for at least a thousand years that some substances can be broken down by heating or chemical treatment into "simpler" ones, but there is always a limit; we eventually get substances known as *elements* that cannot be reduced to any simpler forms by ordinary chemical or physical means. What is our criterion for "simpler"? The most observable (and therefore macroscopic) property is the weight.

The idea of a minimal unit of chemical identity that we call an *element* developed from experimental observations of the relative weights of substances involved in chemical reactions. For example, the compound mercuric oxide can be broken down by heating into two other substances:

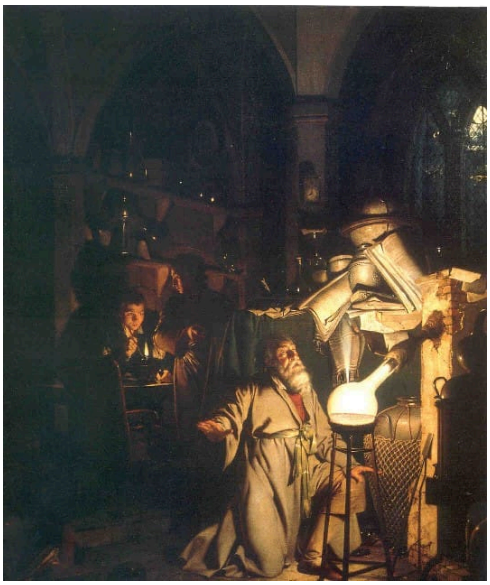
$$2\,\mathrm{HgO} \rightarrow 2\,\mathrm{Hg} + \mathrm{O_2} \tag{1.1.1}$$

... but the two products, metallic mercury and dioxygen, cannot be decomposed into simpler substances, so they must be elements.

Elements and atoms

The definition of an element given above is an *operational* one; a certain result (or in this case, a non-result!) of a procedure that might lead to the decomposition of a substance into lighter units will tentatively place that substance into one of the categories: element or compound. Because this operation is carried out on bulk matter, the concept of the element is also a *macroscopic* one. The *atom*, by contrast, is a *microscopic* concept which in modern chemistry relates the unique character of every chemical element to an actual physical particle.

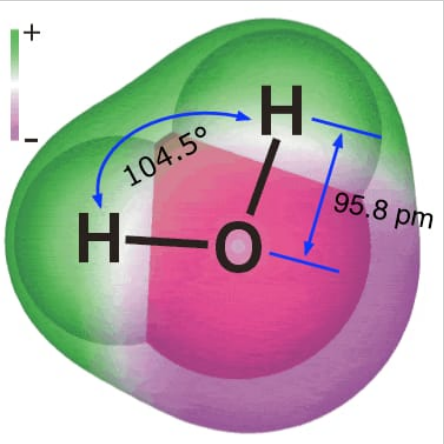
The idea of the atom as the smallest particle of matter had its origins in Greek philosophy around 400 BCE but was controversial from the start (both Plato and Aristotle maintained that matter was infinitely divisible.) It was not until 1803 that John Dalton proposed a rational atomic theory to explain the facts of chemical combination as they were then known, thus being the first to employ macroscopic evidence to illuminate the microscopic world. It wasn't until the 1900s that the atomic theory became universally accepted. In the 1920's it became possible to measure the sizes and masses of atoms, and in the 1970's techniques were developed that produced images of individual atoms.

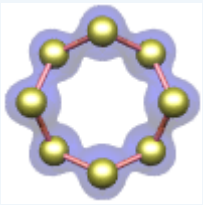


Formula and structure

The formula of a substance expresses the relative number of atoms of each element it contains. Because the formula can be determined by experiments on bulk matter, it is a macroscopic concept even though it is expressed in terms of atoms.

What the ordinary chemical formula does *not* tell us is the order in which the component atoms are connected, whether they are grouped into discrete units (*molecules*) or are two- or three dimensional extended structures, as is the case with solids such as ordinary salt. The microscopic aspect of composition is the *structure*, which gives detailed relative locations (in two or three dimensional space) of each atom within the minimum collection needed to define the structure of the substance.

Macroscopic	Microscopic	
Substances are defined at the macroscopic level by their **formulas** or **compositions**, and at the microscopic level by their **structures**.	The elements hydrogen and oxygen combine to form a compound whose composition is expressed by the formula H_2O.	The molecule of water has the structure shown here. 

Macroscopic	Microscopic	
Chemical substances that cannot be broken down into simpler ones are known as **elements**. The actual physical particles of which elements are composed are **atoms** or **molecules**.	Sulfur – the **element** in its orthorhombic crystalline form. 	**molecule** is an octagonal ring of sulfur **atoms**. The crystal shown at the left is composed of an ordered array of these molecules. (No, they don't actually move around like this, although they are in a constant state of vibrational motion.) 

Compounds and molecules

As we indicated above, a *compound* is a substance containing more than one element. Since the concept of an element is macroscopic and the distinction between elements and compounds was recognized long before the existence of physical atoms was accepted, the concept of a compound must also be a macroscopic one that makes no assumptions about the nature of the ultimate. Thus when carbon burns in the presence of oxygen, the product carbon dioxide can be shown by (macroscopic) weight measurements to contain both of the original elements:

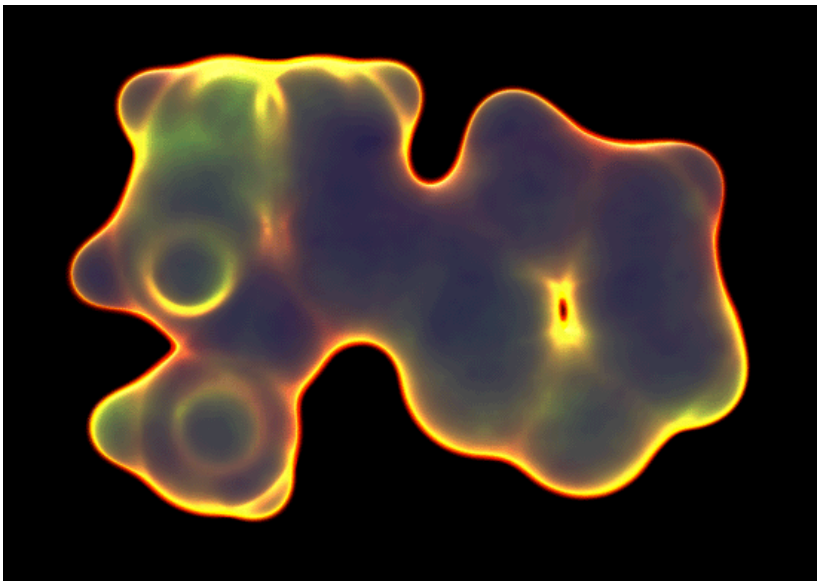
$$\mathrm{C} + \mathrm{O}_2 \longrightarrow \mathrm{CO2} \tag{1.1.2}$$

10.0 g + 26.7 g = 36.7 g

One of the important characteristics of a compound is that the proportions by weight of each element in a given compound are constant. For example, no matter what weight of carbon dioxide we have, the percentage of carbon it contains is (10.0 / 36.7) = 0.27, or 27%.

Molecules

A *molecule* is an assembly of atoms having a fixed composition, structure, and distinctive, measurable properties.



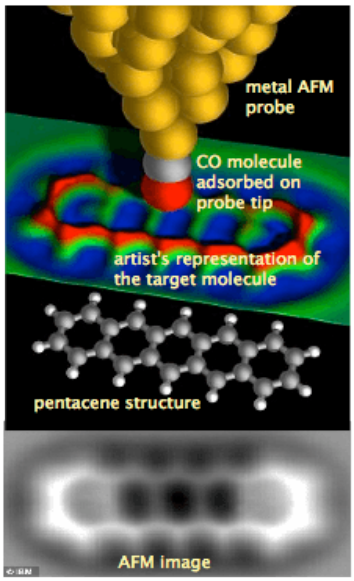
Computer-model of Nicotine molecule $C_{10}H_{14}N_2$ by Ronald Perry

"Molecule" refers to a kind of particle, and is therefore a microscopic concept. Even at the end of the 19th century, when compounds and their formulas had long been in use, some prominent chemists doubted that molecules (or atoms) were any more than a convenient model.

Molecules suddenly became real in 1905, when Albert Einstein showed that Brownian motion, the irregular microscopic movements of tiny pollen grains floating in water, could be directly attributed to collisions with molecule-sized particles.

In 2009, IBM scientists in Switzerland succeeded in imaging a real molecule, using a technique known as atomic force microscopy in which an atoms thin metallic probe is drawn ever-so-slightly above the surface of an immobilized pentacene molecule cooled to nearly absolute zero. In order to improve the image quality, a molecule of carbon monoxide was placed on the end of the probe.

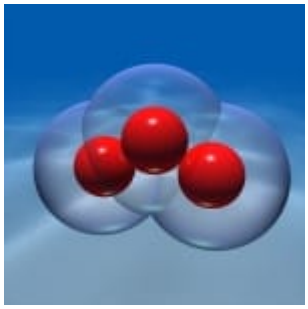
The image produced by the AFM probe is shown at the very bottom. What is actually being imaged is the surface of the electron clouds of the molecule, which consists of six hexagonal rings of carbon atoms with hydrogen on its periphery. The tiny bumps that correspond to these hydrogen atoms attest to the remarkable resolution of this experiment.



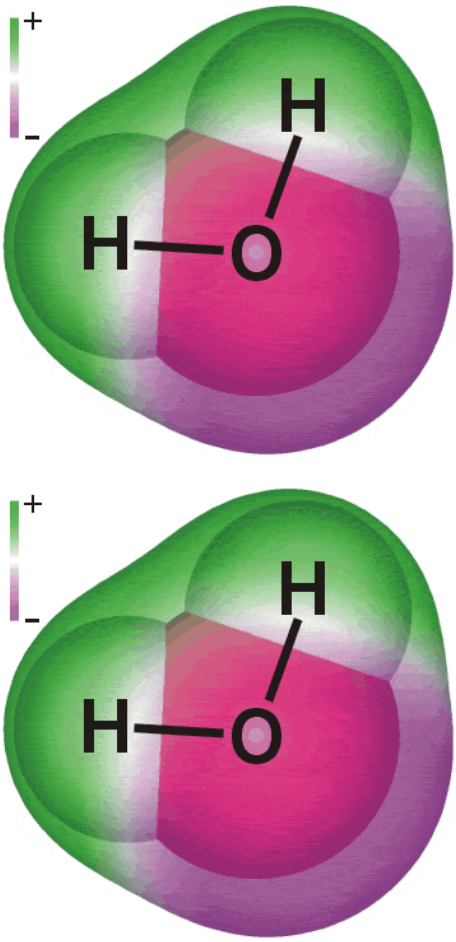
The atomic composition of a molecule is given by its *formula*. Thus the formulas CO, CH_4, and O_2 represent the molecules carbon monoxide, methane, and dioxygen. However, the fact that we can write a formula for a compound does not imply the existence of molecules having that composition. Gases and most liquids consist of molecules, but many solids exist as extended lattices of atoms or ions (electrically charged atoms or molecules.) For example, there is no such thing as a "molecule" of ordinary salt, NaCl (see below.)

Confused about the distinction between molecules and compounds?

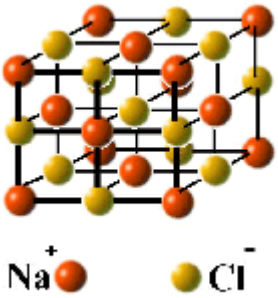
Maybe the following will help:



A molecule but not a compound - Ozone, O_3, is *not* a compound because it contains only a single element.



This well-known molecule is a compound because it contains more than one element.



Ordinary solid **salt is a compound but not a molecule**. It is built from interpenetrating lattices of sodium and chloride ions that extend indefinitely.

Structure and properties

Composition and structure lie at the core of Chemistry, but they encompass only a very small part of it. It is largely the *properties* of chemical substances that interest us; it is through these that we experience and find uses for substances, and much of chemistry as a science is devoted to understanding the relation between structure and properties. For some purposes it is convenient to distinguish between chemical properties and physical properties, but as with most human-constructed dichotomies, the distinction becomes more fuzzy as one looks more closely.

Chemical change

Chemical change is defined macroscopically as a process in which new substances are formed. On a microscopic basis it can be thought of as a re-arrangement of atoms. A given chemical change is commonly referred to as a chemical reaction and is described by a chemical equation that has the form

reactants → products

In elementary courses it is customary to distinguish between "chemical" and "physical" change, the latter usually relating to changes in physical state such as melting and vaporization. As with most human-created dichotomies, this begins to break down when examined closely. This is largely because of some ambiguity in what we regard as a distinct "substance".

Example 1: Chlorine

Elemental chlorine exists as the diatomic molecule $\mathrm{Cl_2}$ in the gas, liquid, and solid states; the major difference between them lies in the degree of organization. In the gas the molecules move about randomly, whereas in the solid they are constrained to locations in a 3-dimensional lattice. In the liquid, this tight organization is relaxed, allowing the molecules to slip and slide around each other.

Since the basic molecular units remain the same in all three states, the processes of melting, freezing, condensation and vaporization are usually regarded as *physical* rather than chemical changes.

Example 2: Sodium Chloride

Solid salt consists of an indefinitely extended 3-dimensional array of Na^+ and Cl^- *ions* (electrically charged atoms.)

When heated above 801°C, the solid melts to form a liquid consisting of these same ions. This liquid boils at 1430° to form a vapor made up of discrete molecules having the formula $\mathrm{Na_2Cl_2\,(aq)}$.. Because the ions in the solid, the hydrated ions in the solution, and the molecule $\mathrm{Na_2Cl_2}$ are really different chemical species, so the distinction between physical and chemical change becomes a bit fuzzy.

Energetics and Equilibrium

You have probably seen chemical reaction equations such as the "generic" one shown below:

$$\mathrm{A + B \longrightarrow C + D} \tag{1.1.3}$$

An equation of this kind does *not* imply that the reactants A and B will change entirely into the products C and D, although in many cases this will be what appears to happen. Most chemical reactions proceed to some intermediate point that yields a mixture of reactants and products.

Example 3

For example, if the two gases phosphorus trichloride and chlorine are mixed together at room temperature, they will combine until about half of them have changed into phosphorus pentachloride:

$$\mathrm{PCl_3 + Cl_2 \ \rightarrow PCl_5} \tag{1.1.4}$$

At other temperatures the extent of reaction will be smaller or greater. The result, in any case, will be an equilibrium mixture of reactants and products.

The most important question we can ask about any reaction is "what is the equilibrium composition"?

- If the answer is "all products and negligible quantities of reactants", then we say the reaction can takes place and that it "goes to completion ".
- If the answer is "negligible quantities of products", then we say the reaction cannot take place in the forward direction, but that the reverse reaction can occur.
- If the answer is "significant quantities of all components" (both reactants and products) are present in the equilibrium mixture, then we say the reaction is "reversible" or "incomplete".

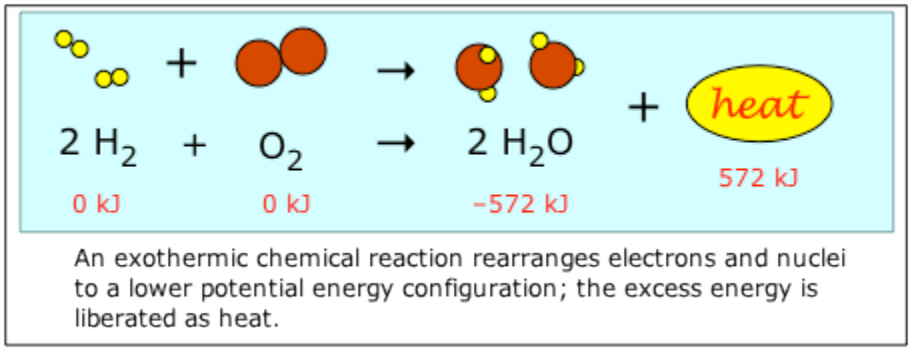
The aspect of "change" we are looking at here is a *property of a chemical reaction*, rather than of any one substance. But if you stop to think of the huge number of possible reactions between the more than 15 million known substances, you can see that it would be an impossible task to measure and record the equilibrium compositions of every possible combination.

One or two directly measurable properties of the individual reactants and products can be combined to give a number from which the equilibrium composition at any temperature can be easily calculated. There is no need to do an experiment!

This is very much a macroscopic view because the properties we need to directly concern ourselves with are those of the reactants and products. Similarly, the equilibrium composition — the measure of the extent to which a reaction takes place — is expressed in terms of the quantities of these substances.

Chemical Energetics

Virtually all chemical changes involve the uptake or release of energy, usually in the form of heat. It turns out that these energy changes, which are the province of *chemical thermodynamics*, serve as a powerful means of predicting whether or not a given reaction can proceed, and to what extent. Moreover, all we need in order to make this prediction is information about the energetic properties of the reactants and products; there is no need to study the reaction itself. Because these are bulk properties of matter, chemical thermodynamics is entirely macroscopic in its outlook.

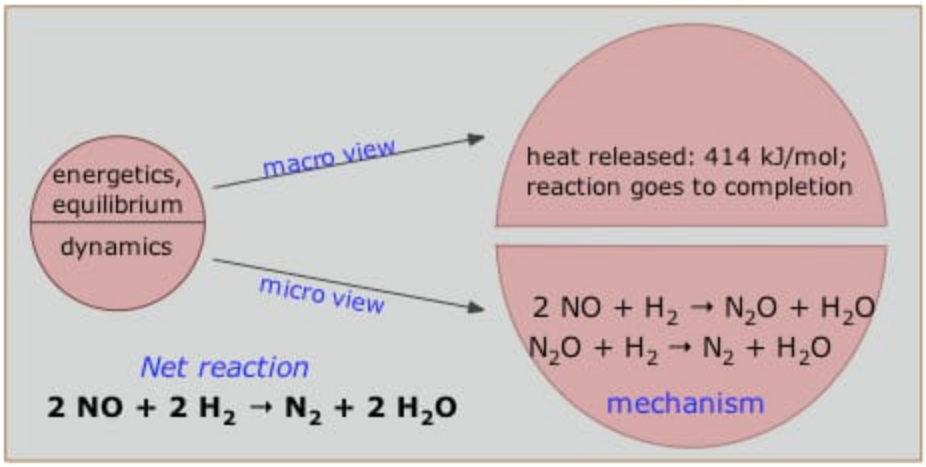


An exothermic chemical reaction rearranges electrons and nuclei to a lower potential energy configuration; the excess energy is liberated as heat.

Dynamics: Kinetics and Mechanism

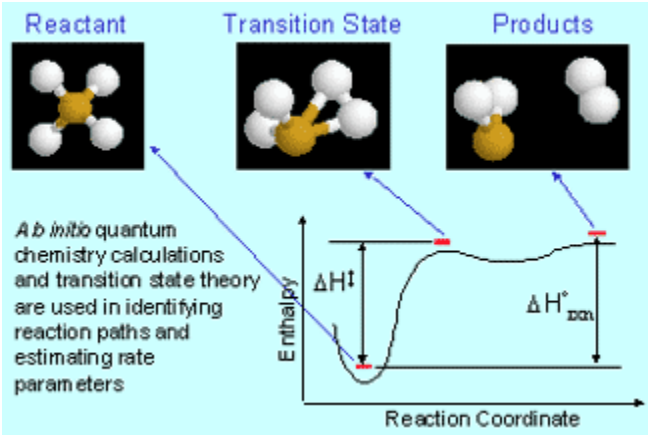
The energetics of chemical change that we discussed immediately above relate to the ***end result*** of chemical change: the composition of the final reaction mixture, and the quantity of heat liberated or absorbed. The *dynamics* of chemical change are concerned with ***how*** the reaction takes place:

- What has to happen to get the reaction started (which molecule gets bumped first, how hard, and from what direction?)
- Does the reaction take place in a single step, or are multiple steps and intermediate structures involved?



These details constitute what chemists call the *mechanism* of the reaction. For example, the reaction between nitric oxide and hydrogen (identified as the net reaction at the bottom left), is believed to take place in the two steps shown here. Notice that the nitrous oxide, N_2O, is formed in the first step and consumed in the second, so it does not appear in the net reaction equation. The N_2O is said to act as an *intermediate* in this reaction. Some intermediates are unstable species, often distorted or incomplete molecules that have no independent existence; these are known as *transition states*.

The microscopic side of dynamics looks at the mechanisms of chemical reactions. This refers to a "blow-by-blow" description of what happens when the atoms in the reacting species re-arrange themselves into the configurations they have in the products.



Mechanism represents the *microscopic* aspect of chemical change. Mechanisms, unlike energetics, cannot be predicted from information about the reactants and products; chemical theory has not yet advanced to the point where we can do much more than make educated guesses. To make matters even more complicated (or, to chemists, *interesting!*), the same reaction can often proceed via different mechanisms under different conditions.

Kinetics

Because we cannot directly watch the molecules as they react, the best we can usually do is to infer a reaction mechanism from experimental data, particularly that which relates to the rate of the reaction as it is influenced by the concentrations of the reactants. This entirely experimental area of chemical dynamics is known as *kinetics*.

Reaction rates, as they are called, vary immensely: some reactions are completed in microseconds, others may take years; many are so slow that their rates are essentially zero. To make things even more interesting, there is no relation between reaction rates and "tendency to react" as governed by the factors in the top half of the above diagram; the latter can be accurately predicted from energetic data on the substances (the properties we mentioned in the previous screen), but reaction rates must be determined by experiment.

Catalysts

Catalysts can make dramatic changes in rates of reactions, especially in those whose un-catalyzed rate is essentially zero. Consider, for example, this rate data on the decomposition of hydrogen peroxide. H_2O_2 is a by-product of respiration that is poisonous to living cells which have, as a consequence, evolved a highly efficient enzyme (a biological catalyst) that is able to destroy peroxide as quickly as it forms. Catalysts work by enabling a reaction to proceed by an alternative mechanism.

$$H_2O_2 \rightarrow 2\ H_2O + O_2$$

	relative rate
no catalyst	1
platimum catalyst	24,000
catalase enzyme	2,900,000,000

In some reactions, even ***light*** can act as a catalyst. For example, the gaseous elements hydrogen and chlorine can remain mixed together in the dark indefinitely without any sign of a reaction, but in the sunlight they combine explosively.

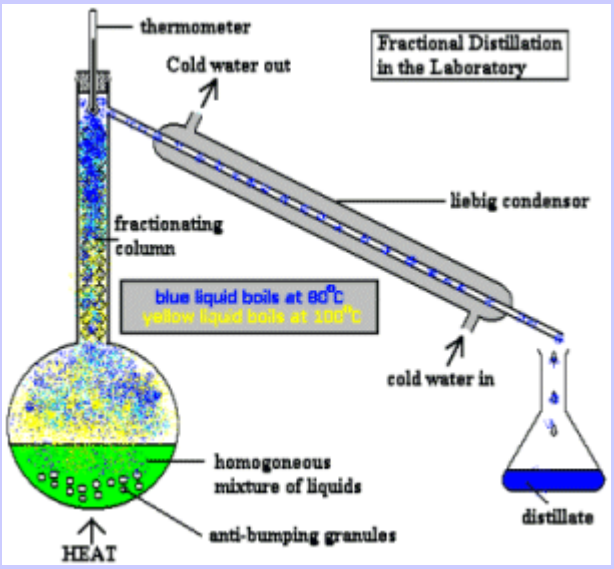
Currents of modern Chemistry

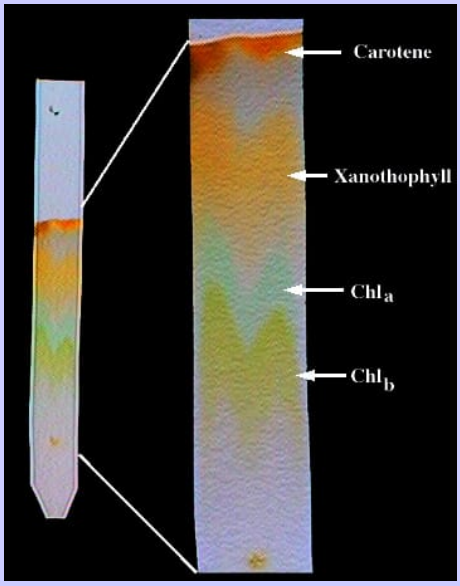
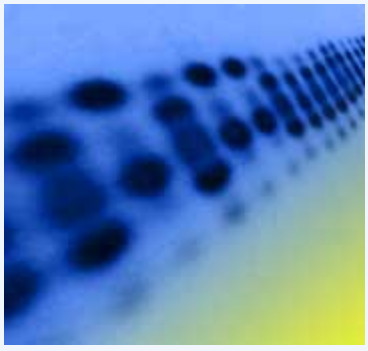
In the preceding section we looked at chemistry from a conceptual standpoint. If this can be considered a "macroscopic" view of chemistry, what is the "microscopic" view? It would likely be what chemists actually do. Because a thorough exploration of this would lead us into far more detail than we can accommodate here, we will mention only a few of the areas that have emerged as being especially important in modern chemistry.

Separation science

A surprisingly large part of chemistry has to do with isolating one component from a mixture. This may occur at any number of stages in a manufacturing process, including the very critical steps involved in removing toxic, odiferous, or otherwise undesirable by-products from a waste stream. But even in the research lab, a considerable amount of effort is often devoted to separating the desired substance from the many components of a reaction mixture, or in separating a component from a complex mixture (for example, a drug metabolite from a urine sample), prior to measuring the amount present.

Distillation - separation of liquids having different boiling points. This ancient technique (believed to have originated with Arabic alchemists in 3500 BCE), is still one of the most widely employed operations both in the laboratory and in industrial processes such as oil refining.



Solvent extraction - separation of substances based on their differing solubilities. A common laboratory tool for isolating substances from plants and chemical reaction mixtures. Practical uses include processing of radioactive wastes and decaffeination of coffee beans. The separatory funnel shown here is the simplest apparatus for liquid-liquid extraction; for solid-liquid extraction, the Soxhlet apparatus is commonly used.	
Chromatography - This extremely versatile method depends on the tendency of different kinds of molecules to adsorb (attach) to different surfaces as they travel along a "column" of the adsorbent material. Just as the progress of people walking through a shopping mall depends on how long they spend looking in the windows they pass, those molecules that adsorb more strongly to a material will emerge from the chromatography column more slowly than molecules that are not so strongly adsorbed.	 Paper chromatography of plant juice [link]
Gel electrophoresis - a powerful method for separating and "fingerprinting" macromolecules such as nucleic acids or proteins on the basis of physical properties such as size and electric charge.	

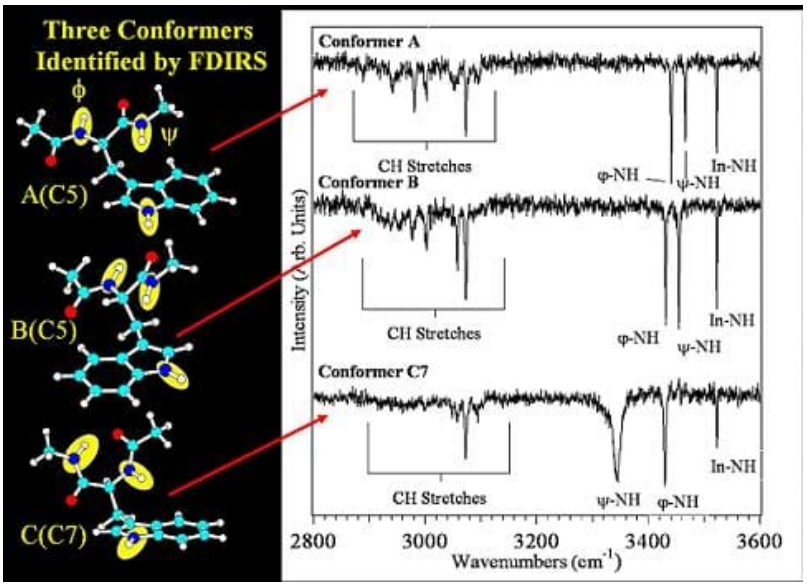
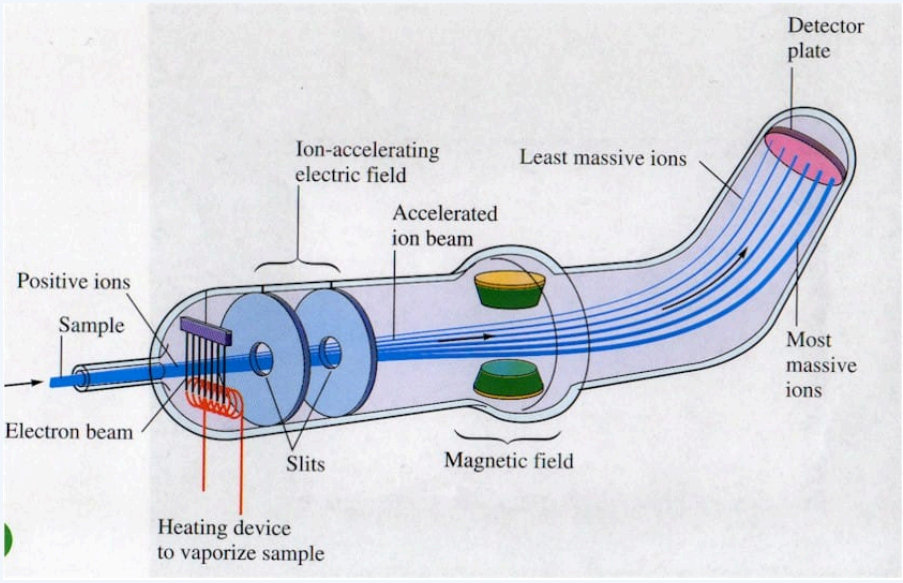
Identification and assay

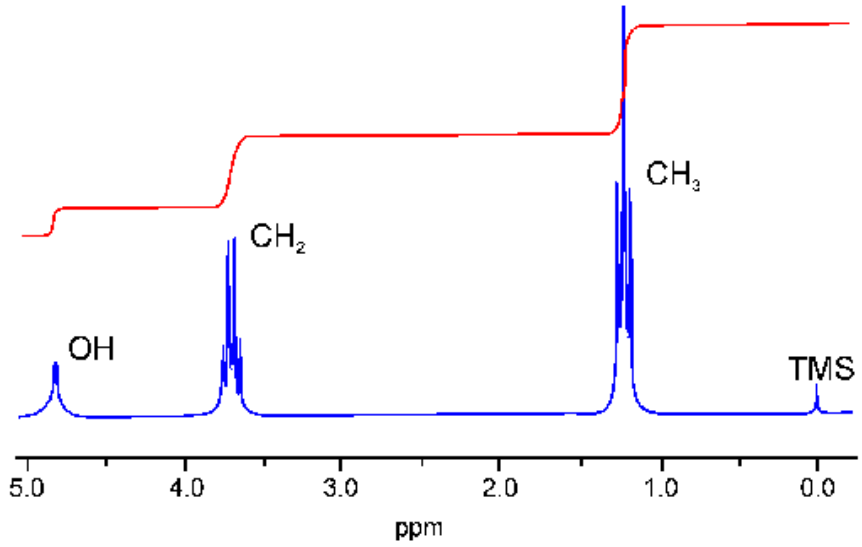
What do the following people have in common?

- A plant manager deciding on whether to accept a rail tank car of vinyl chloride for manufacture into plastic pipe

- An agricultural chemist who wants to know about the vitamin content of a new vegetable hybrid
- The manager of a city water-treatment plant who needs to make sure that the carbonate content of the water is maintained high enough to prevent corrosion, but low enough to prevent scale build-up

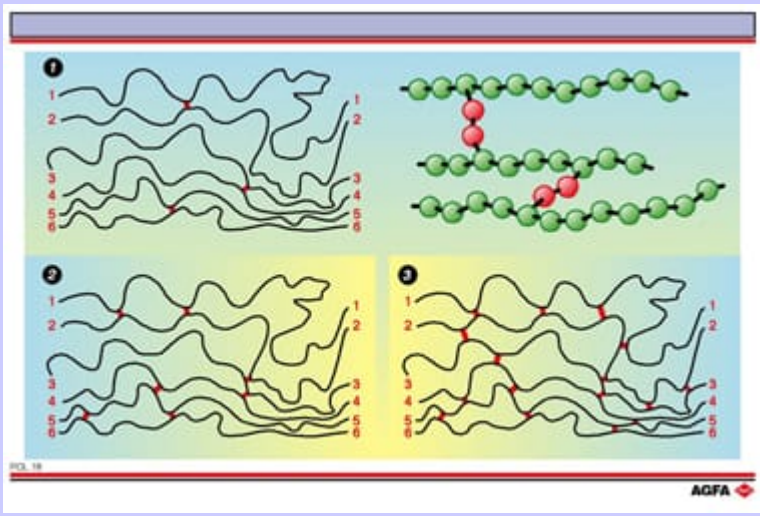
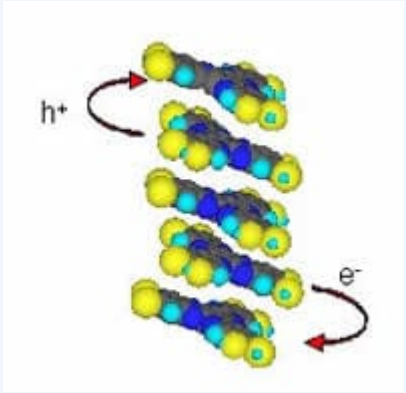
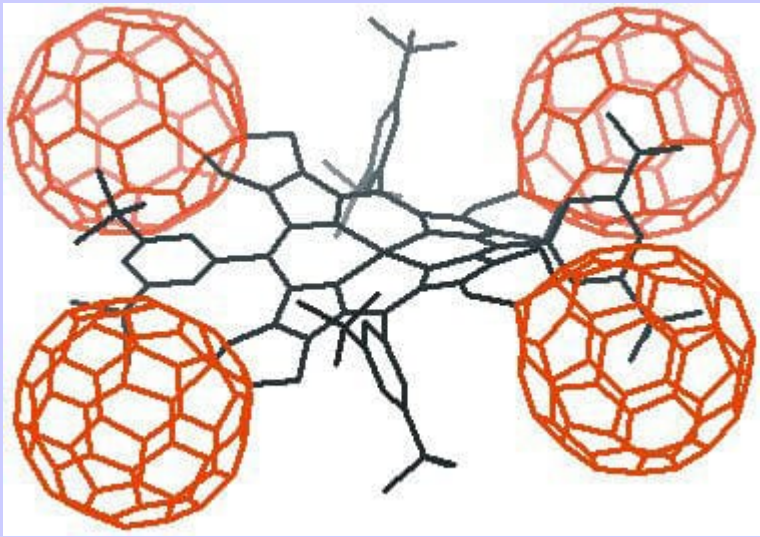
The answer is that all depend on *analytical techniques* — measurements of the nature or quantity ("assays") of some substance of interest, sometimes at very low concentrations. A large amount of research is devoted to finding more accurate and convenient means of making such measurements. Many of these involve sophisticated instruments; among the most widely used are the following:

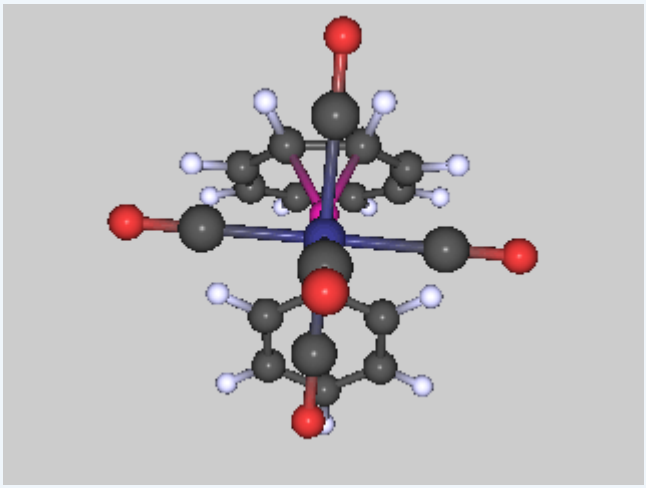
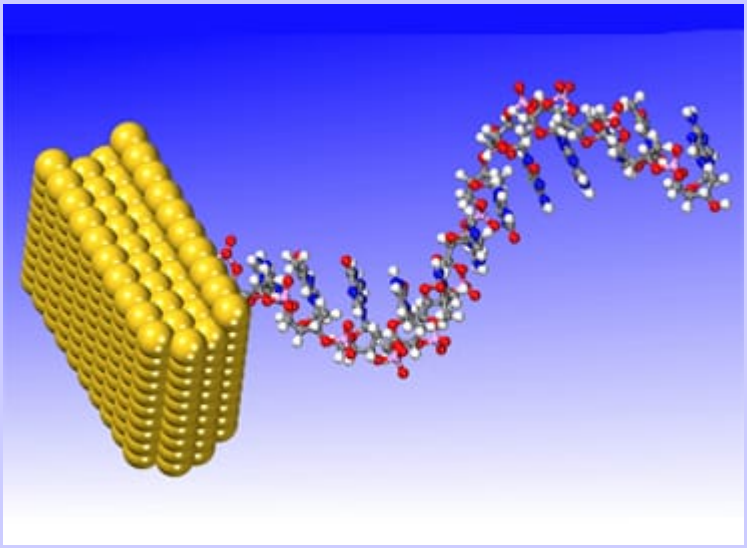
Spectrophotometers that examine the ways that light of various wavelengths is absorbed, emitted, or altered by atomic and molecular species.	
Mass spectrometers that break up molecules into fragments that can be characterized by electrical methods.	

Instruments (NMR spectrometers) that analyze the action of radio waves and magnetic fields on atomic nuclei in order to examine the nature of the chemical bonds attached to a particular kind of atom.	
"In the early 1900's a chemist could analyze about 200 samples per year for the major rock-forming elements. Today, using X-ray fluorescence spectrometry, two chemists can perform the same type of analysis on 7,000 samples per year."	

Materials, polymers, and nanotechnology

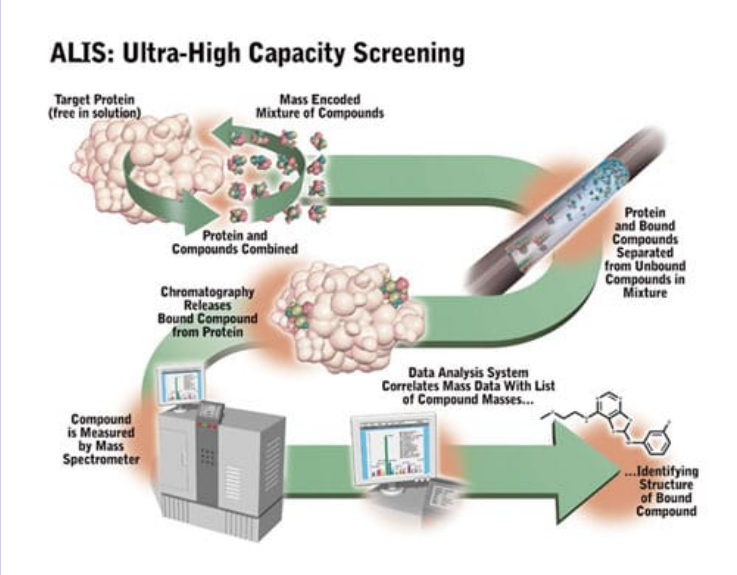
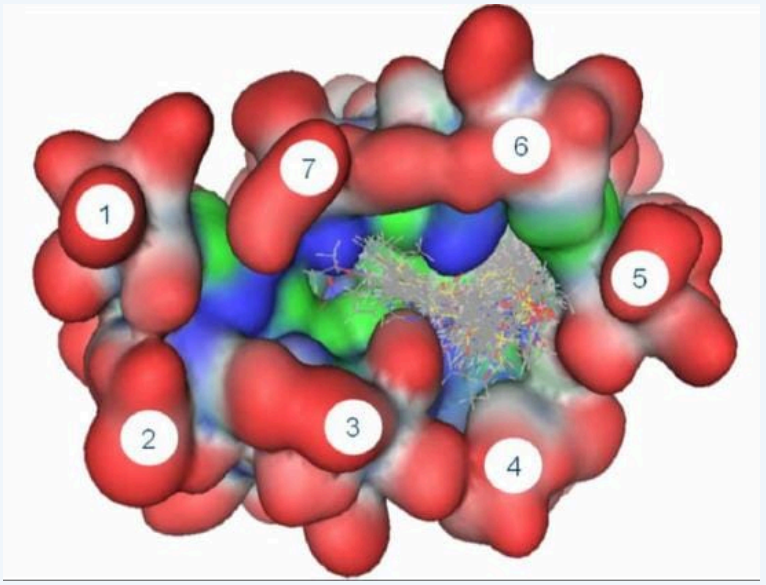
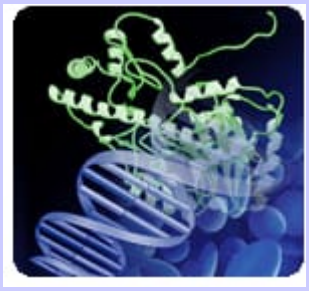
Materials science attempts to relate the physical properties and performance of engineering materials to their underlying chemical structure with the aim of developing improved materials for various applications.

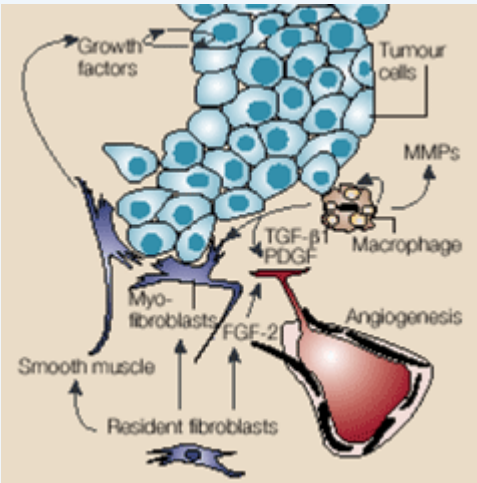
Polymer chemistry - developing polymeric ("plastic") materials for industrial uses. Connecting individual polymer molecules by cross-links (red) increases the strength of the material. Thus ordinary polyethylene is a fairly soft material with a low melting point, but the cross-linked form is more rigid and resistent to heat.	
Organic semiconductors offer a number of potential advantages over conventional metalloid-based devices.	
Fullerenes, nanotubes, and nanowires - Fullerenes were first identified in 1985 as products of experiments in which graphite was vaporized using a laser, work for which R. F. Curl, Jr., R. E. Smally, and H. W. Kroto shared the 1996 Nobel Prize in Chemistry. Fullerene research is expected to lead to new materials, lubricants, coatings, catalysts, electro-optical devices, and medical applications.	

Nanodevice chemistry — constructing molecular-scale assemblies for specific tasks such as computing, producing motions, etc.	
Biosensors and biochips - the surfaces of metals and semiconductors "decorated" with biopolymers can serve as extremely sensitive detectors of biological substances and infectious agents.	

Biochemistry and Molecular biology

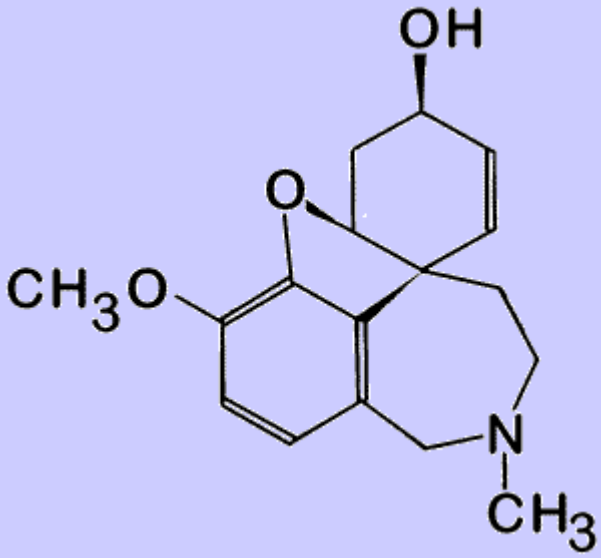
This field covers a wide range of studies ranging from fundamental studies on the chemistry of gene expression and enzyme-substrate interactions to drug design. Much of the activity in this area is directed to efforts in ***drug discovery.***

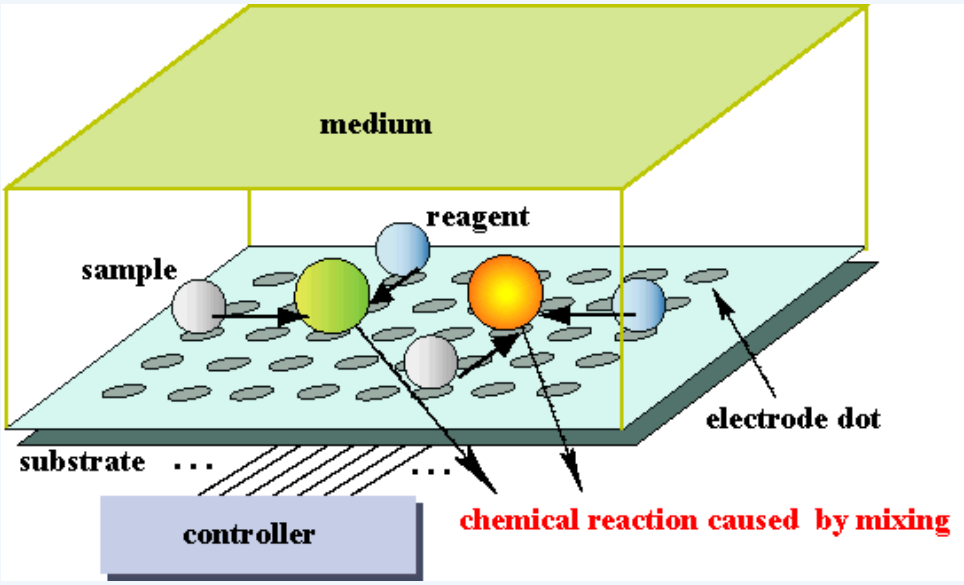
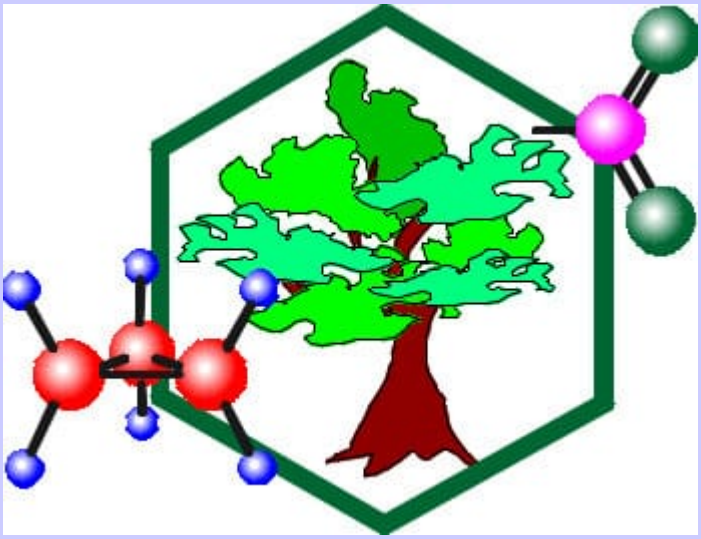
Drug screening began as a largely scattershot approach in which a pathogen or a cancer cell line was screened against hundreds or thousands of candidate substance in the hope of finding a few "leads" that might result in a useful therapy. This field is now highly automated and usually involves combinatorial chemistry (see below) combined with innovative separation and assay methods.	
Drug design looks at interactions between enzymes and possible inhibitors. Computer-modeling is an essential tool in this work.	
Proteomics - This huge field focuses on the relations between structure and function of proteins— of which there are about 400,000 different kinds in humans. Proteomics is related to genetics in that the DNA sequences in genes get decoded into proteins which eventually define and regulate a particular organism.	 [*Science* article image link]

Chemical genomics explores the chain of events in which signaling molecules regulate gene expression.	

Synthesis

In its most general sense, this word refers to any reaction that leads to the formation of a particular molecule. It is both one of the oldest areas of chemistry and one of the most actively pursued. Some of the major threads are

New-molecule synthesis - Chemists are always challenged to come up with molecules containing novel features such as new shapes or unusual types of bonds.	

Combinatorial chemistry refers to a group of largely-automated techniques for generating tiny quantities of huge numbers of different molecules ("libraries") and then picking out those having certain desired properties. Although it is a major drug discovery technique, it also has many other applications.	
Green chemistry - synthetic methods that focus on reducing or eliminating the use or release of toxic or non-biodegradable chemicals or byproducts.	
Process chemistry bridges the gap between chemical synthesis and chemical engineering by adapting synthetic routes to the efficient, safe, and environmentally-responsible methods for large-scale synthesis.	

Congratulations! You have just covered all of Chemistry, condensed into one quick and painless lesson— the world's shortest Chemistry course! Yes, we left out a lot of the details, the most important of which will take you a few months of happy discovery to pick up. But if you keep in mind the global hierarchy of composition/ structure, properties of substances, and change (equilibrium and dynamics) that we have developed in both macroscopic and microscopic views, you will find it much easier to assemble the details as you encounter them and to see where they fit into the bigger picture.

This page titled 1.1: Introduction to Chemistry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

1.2: Pseudoscience

A *pseudoscience* is a belief or process which masquerades as science in an attempt to claim a legitimacy which it would not otherwise be able to achieve on its own terms; it is often known as *fringe*-or *alternative* science. The most important of its defects is usually the lack of the carefully controlled and thoughtfully interpreted experiments which provide the foundation of the natural sciences and which contribute to their advancement.

Of course, the pursuit of scientific knowledge usually involves elements of intuition and guesswork; experiments do not always test a theory adequately, and experimental results can be incorrectly interpreted or even wrong. In legitimate science, however, these problems tend to be self-correcting, if not by the original researchers themselves, then through the critical scrutiny of the greater scientific community. Critical thinking is an essential element of science.

Other Types of Defective Science

There have been several well-documented instances in which the correction process referred to above was delayed until after the initial incorrect interpretation became widely publicized, resulting in what has been called ***pathological science***. The best known of these incidents are the "discoveries" of N-rays, of polywaters, and of cold fusion. All of these could have been averted if the researchers had not been so enthused with their results that they publicized them before they had received proper review by others. Human nature being what it is, there is always some danger of this happening; to discourage it, most of the prestigious scientific journals will refuse to accept reports of noteworthy work that has already been made public.

Another term, *junk science*, is often used to describe scientific theories or data which, while perhaps legitimate in themselves, are believed to be mistakenly used to support an opposing position. There is usually an element of political or ideological bias in the use of the term. Thus the arguments in favor of limiting the use of fossil fuels in order to reduce global warming are often characterized as junk science by those who do not wish to see such restrictions imposed, and who claim that other factors may well be the cause of global warming. A wide variety of **commercial advertising** (ranging from hype to outright fraud) would also fall into this category; at its most egregious it might better be described as *deceptive science*.

This description of **Ivory Soap** is a classic example of junk science from the 19th century. Not only is the term "pure" meaningless when applied to an undefined mixture such as bath soap, but the implication that its ability to float is evidence of this purity is deceptive. The low density is achieved by beating air bubbles into it, actually reducing the "purity" of the product and in a sense cheating the consumer.

Hoax science is another category that describes deliberately contrived sensationalist writings that have received wide publicity (and earned substantial royalties for their authors.) Immanual Velikovsky's *Worlds in Collision* (1950) is now probably the best known of these, followed by Erich von Däniken's *Chariots of the Gods?* (1968). Perhaps the most recent contender in this field is David Talbott, who with Wallace Thornhill wrote *The Electric Universe* and *Thunderbolts of the Gods*.

Fraudulent science and *Scientific Misconduct* refer to work that is intentionally fabricated or misrepresented for personal (recognition or career-advancement) or commercial (marketing or regulatory) reasons. Suppression of science for political reasons often occurred during the second Bush administration. The tobacco and pharmaceutical industries have been notoriously implicated in the latter category. The tobacco industry even published a phoney "scientific" journal containing articles written by hack authors that disputed warnings about smoking-induced cancer.

Charges of minor fudging go back to the days of Ptolemy, Gailileo, and Isaac Newton, but revelations of more contemporary frauds and their contamination of the scientific literature makes these far more problematic. Since about 1980, scientists at several major U.S. universities have been compelled to withdraw articles from prestigious journals. One of the most widely-publicized cases was that of the eminent Korean researcher who reported bogus stem cell results. Some rather troubling recent cases involve at least seventy articles describing bogus chemical structures reported by a group of scientists in China, and about an equal number of papers that had been forged or falsified by one or more scientists at a university in India.

Finally, there is just plain *bad science*, which would logically encompass all of the evils being discussed here, but is commonly

used to describe well-intentioned but incorrect, obsolete, incomplete, or over-simplified expositions of scientific ideas. An example would be the statement that electrons revolve in orbits around the atomic nucleus, a picture that was discredited in the 1920's, but is so much more vivid and easily grasped than the one that supplanted it that it shows no sign of dying out.

In ordinary conversation, the word "theory" connotes an opinion, a conjecture, or a supposition. But in science, the term has a much more limited meaning. A scientific theory is an attempt to explain some aspect of the natural world in terms of empirical evidence and observation. It commonly draws upon established principles and knowledge with the aim of extending them in a logical and consistent way that enables one to make useful predictions. All scientific theories are tentative and subject to being tested and modified. As theories become more mature, they grow into more organized bodies of knowledge that enable us to understand and predict a wider range of phenomena. Examples of such theories are quantum theory, Einstein's theories of relativity, and evolution.

Scientific theories fall into two categories:

1. Theories that have been shown to be incorrect, usually because they are not consistent with new observations;
2. All other theories

Hence, theories cannot be proven to be correct; there is always the possibility that further observations will disprove the theory. Furthremore, a theory that cannot be refuted or falsified is not a scientific theory.

For example, the theories that underlie astrology (the doctrine that the positions of the stars can influence one's life) are not falsifiable because they, and the predictions that follow from them, are so vaguely stated that the failure of these predictions can always be "explained away" by assuming that various other influences were not taken into account. It is similarly impossible to falsify so-called "creation science" or "intelligent design" because one can simply evoke the "then a miracle occurs" at any desired stage.

Recognizing Pseudoscience?

There is no single test that unambiguously distinguishes between science and pseudoscience, but as the two diverge more and more from one another, certain differences become apparent, and these tend to be remarkably consistent across all fields of interest. In examining the following table, it might be helpful to consider examples of astronomy *vs.* astrology, or of chemistry vs. alchemy, which at one time were single fields that gradually diverged into sciences and pseudosciences.

Many scientists' ordinary response to pseudoscientific claims is simply to laugh at them. But mythology has always been an important part of human culture, often by giving people the illusion of having some direct control over their lives. This can lead to their becoming advocates for various kinds of health quackery, to commercial scams, and to cult-like organizations such as scientology. Worst of all, they can pressure political and educational circles to adopt their ideologies.

Does the "Establishment" Actively Suppress new Ideas?

Anyone who has been around for long enough has encountered statements like these:

- An inventor's design for a device that uses water as a fuel has been bought up and suppressed by the oil companies.
- "Alternative health" techniques (homeopathy, chiropractic, chelation therapy— you name it!) are actively suppressed by the medical profession or the pharmaceutical industry in a desperate attempt to serve their selfish interests.
- Reports of unidentified flying objects (UFO's) are suppressed by the U.S. Government in an attempt to prevent panic and/or to maintain control over citizens.
- Editors of scientific journals and the reviewers they call on to assess the worth of submitted papers reject out-of-hand anything that comes from persons who are not members of the scientific "establishment" or which report results not consistent with presently-accepted science.

Claims of these kinds are frequently made and widely believed, especially by those who are inclined to see conspiracies around every corner. There is little if any evidence for any of these claims. The real reason that new devices or new theories get thrown aside is that the arguments or evidence adduced to support them is inadequate or not credible. The individuals who believe themselves to be unfairly thwarted by the scientific community are very often so isolated from it that they are unable to appreciate

its norms of clarity, rigor, and consistency with existing science.

A common refrain is that " they laughed at Galileo, at Thompson, and at Wegner," whose theores were eventually supported. Well, with Galileo, they did not exactly laugh; it was more a case of challenge to religious doctrine that forced him to recant his assertion that the Sun, and not the Earth, is at the center of the solar system. There have been innumerable cases in which the world was simply not ready to accept a new idea. This was especially common before the scientific method had been developed, and before the technology needed to apply it had become available.

When J.J. Thomson discovered evidence that the atom is not the ultimate fundamental particle and could be broken up into smaller units, even Thomson himself was reluctant to accept it, and he became a laughingstock for several years until more definitive evidence became available.

Alfred Wegener's theory of continental drift was bitterly attacked when it was first published in 1915, and it did not become generally accepted until about 50 years later. Others had made similar proposals based on the way the continents of Africa and South America could be fitted together, but Wegener was the first to make a careful study of fossil and geological similarities between the two continents. Nevertheless, the idea that continents could float around was too hard to accept at a time when nothing was known about the interior structure of the Earth, and the evidence he presented was rejected as inadequate.

On the other hand, the even-more-revolutionary concepts of special- and general relativity, and of quantum theory (which developed in several stages), achieved rapid acceptance when they were first presented, as did Louis Pasteur's germ theory of disease. In all of these cases the new theories provided credible explanations for what was previously unexplainable, and the tools for confirming them existed at the time, or in the case of general relativity, would soon become available.

This page titled 1.2: Pseudoscience is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

2: ESSENTIAL BACKGROUND

Most courses in Chemistry, especially those at the college/university level, assume that their students have had prior courses in general science, and often in physics, which provide them with an understanding of important concepts such as significant figures, units of measure, treatment measurement error, density and buoyancy. But if most of that has receded into the fuzzy past, the six sections of this unit will bring you up to speed. Neglect this stuff at your peril — it will come up in one way or another in any Chemistry course you take!

2.1: Classification and Properties of Matter

2.2: Energy, Heat, and Temperature

2.3: The Measure of Matter

2.4: The Meaning of Measure

This page titled 2: Essential Background is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Classification and Properties of Matter

Learning Objectives

- Give examples of *extensive* and *intensive* properties of a sample of matter. Which kind of property is more useful for describing a particular kind of matter?
- Explain what distinguishes *heterogeneous* matter from *homogeneous* matter.
- Describe the following separation processes: *distillation*, *crystallization*, *liquid-liquid extraction*, *chromatography*.
- To the somewhat limited extent to which it is meaningful, classify a given property as a *physical* or *chemical* property of matter.

Matter is "anything that has mass and occupies space", we were taught in school. True enough, but not very satisfying. A really complete answer is unfortunately beyond the scope of this course, but we will offer a hint of it in a later chapter on atomic structure. For the moment, let's put off trying to define matter and focus on the chemist's view: matter is what chemical substances are composed of. But what do we mean by chemical substances? How do we organize our view of matter and its properties? These very practical questions will be the subjects of this lesson.

Properties of Matter

The science of chemistry developed from observations made about the nature and behavior of different kinds of matter, which we refer to collectively as the *properties* of matter. The properties we refer to in this lesson are all *macroscopic* properties: those that can be observed in bulk matter. At the *microscopic* level, matter is of course characterized by its *structure*: the spatial arrangement of the individual atoms in a molecular unit or an extended solid. By observing a sample of matter and measuring its various properties, we gradually acquire enough information to *characterize* it; to distinguish it from other kinds of matter. This is the first step in the development of chemical science, in which interest is focused on specific kinds of matter and the transformations between them.

If you think about the various observable properties of matter, it will become apparent that these fall into two classes. Some properties, such as mass and volume, depend on the *quantity* of matter in the sample we are studying. Clearly, these properties, as important as they may be, cannot by themselves be used to characterize a kind of matter; to say that "water has a mass of 2 kg" is nonsense, although it may be quite true in a particular instance. Properties of this kind are called *extensive properties* of matter.

Suppose we take further measurements, and find that the same quantity of water whose mass is 2.0 kg also occupies a volume of 2.0 liters. We have measured two extensive properties (mass and volume) of the same sample of matter. This allows us to define a new quantity, the quotient m/V which defines another property of water which we call the *density*. Unlike the mass and the volume, which by themselves refer only to individual samples of water, the density (mass per unit volume) is a property of *all* samples of pure water at the same temperature. Density is an example of an *intensive property* of matter.

This definition of the density illustrates an important general rule: the ratio of two **extensive** properties is always an **intensive** property.

Intensive properties are extremely important, because every possible kind of matter possesses a unique set of intensive properties that distinguishes it from every other kind of matter. Some intensive properties can be determined by simple observations: color (absorption spectrum), melting point, density, solubility, acidic or alkaline nature, and density are common examples. Even more fundamental, but less directly observable, is chemical composition.

The more intensive properties we know, the more precisely we can characterize a sample of matter.

Intensive properties are extremely important, because every possible kind of matter possesses a unique set of intensive properties that distinguishes it from every other kind of matter. In other words, **intensive properties serve to characterize matter**. Many of the intensive properties depend on such variables as the temperature and pressure, but the ways in which these properties change with such variables can themselves be regarded as intensive properties.

Example 2.1.1

Classify each of the following as an extensive or intensive property.

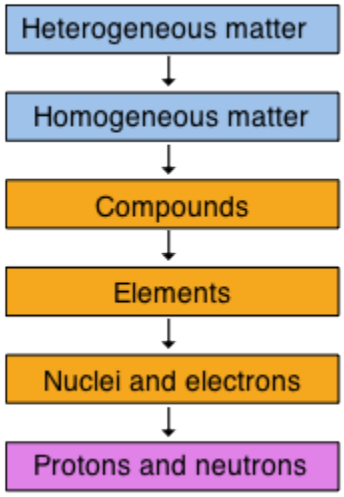
a. The volume of beer in a mug
b. The percentage of alcohol in the beer
c. The number of calories of energy you derive from eating a banana
d. The number of calories of energy made available to your body when you consume 10.0 g of sugar
e. The mass of iron present in your blood
f. The mass of iron present in 5 mL of your blood
g. The electrical resistance of a piece of 22-gauge copper wire.
h. The electrical resistance of a 1-km length of 22-gauge copper wire
i. The pressure of air in a bicycle tire

Answer a available at the end of this section.

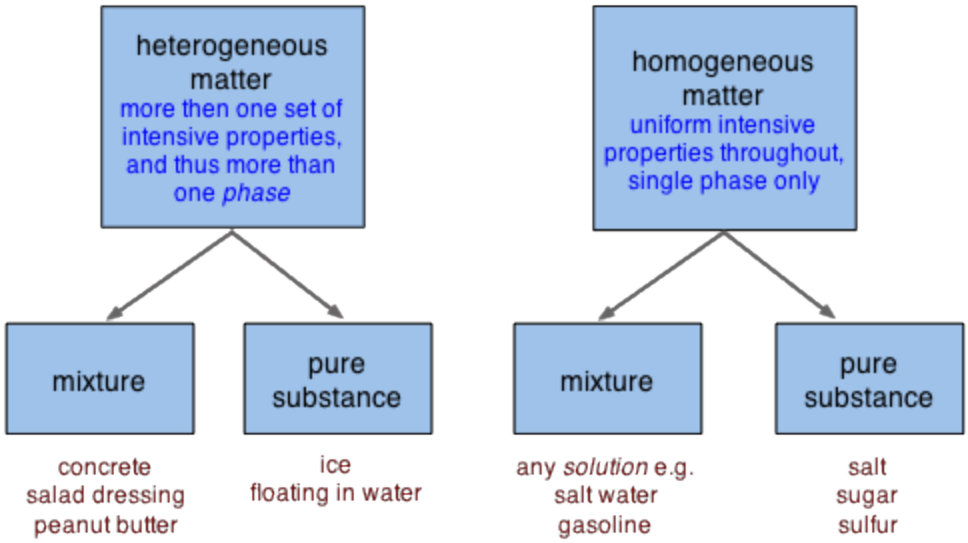
The last example shows that not everything is black or white! But we often encounter matter that is not uniform throughout, whose different parts exhibit different sets of intensive properties. This brings up another distinction that we address immediately below.

How to classify matter?

One useful way of organizing our understanding of matter is to think of a hierarchy that extends down from the most general and complex to the simplest and most fundamental. The orange-colored boxes represent the central realm of chemistry, which deals ultimately with specific chemical substances, but as a practical matter, chemical science extends both above and below this region.



Alternatively, when we are thinking about specific samples of matter, it may be more useful to re-cast our classification in two dimensions:



Notice, in the bottom line of boxes above, that "mixtures" and "pure substances" can fall into either the homogeneous or heterogeneous categories.

Homogeneous and heterogeneous: it's a matter of *phases*

Homogeneous matter (from the Greek *homo* = same) can be thought of as being uniform and continuous, whereas *heterogeneous matter* (*hetero* = different) implies non-uniformity and discontinuity. To take this further, we first need to define "uniformity" in a more precise way, and this takes us to the concept of *phases*.

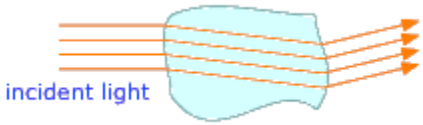
A *phase* is a region of matter that possesses *uniform intensive properties* throughout its volume. A volume of water, a chunk of ice, a grain of sand, a piece of copper— each of these constitutes a single phase, and by the above definition, is said to be homogeneous. A sample of matter can contain more than a single phase; a cool drink with ice floating in it consists of at least two phases, the liquid and the ice. If it is a carbonated beverage, you can probably see gas bubbles in it that make up a third phase.



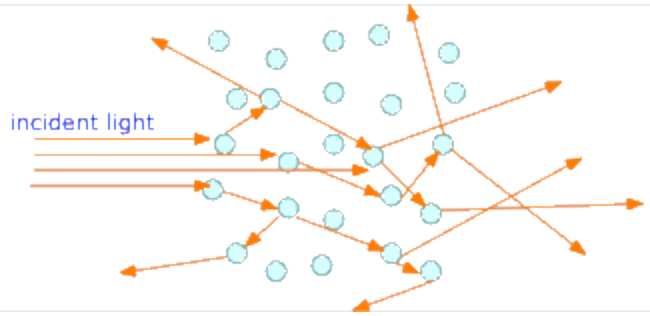
Figure 2.1.1: The phase boundaries in an oil-vinegar salad dressing are easy to see (left), but those in milk are so small that they are only revealed by scattered light (right).

Phase boundaries

Each phase in a multiphase system is separated from its neighbors by a *phase boundary*, a thin region in which the intensive properties change discontinuously. Have you ever wondered why you can easily see the ice floating in a glass of water although both the water and the ice are transparent? The answer is that when light crosses a phase boundary, its direction of travel is slightly bent, and a portion of the light gets reflected back; it is these reflected and distorted light rays emerging from that reveal the chunks of ice floating in the liquid.



If, instead of visible chunks of material, the second phase is broken into tiny particles, the light rays usually bounce off the surfaces of many of these particles in random directions before they emerge from the medium and are detected by the eye. This phenomenon, known as *scattering*, gives multiphase systems of this kind a cloudy appearance, rendering them translucent instead of transparent. Two very common examples are ordinary *fog*, in which water droplets are suspended in the air, and *milk*, which consists of butterfat globules suspended in an aqueous solution.



Getting back to our classification, we can say that Homogeneous matter consists of a single phase throughout its volume; heterogeneous matter contains two or more phases.

Dichotomies ("either-or" classifications) often tend to break down when closely examined, and the distinction between homogeneous and heterogeneous matter is a good example; this is really a matter of degree, since at the microscopic level all matter is made up of atoms or molecules separated by empty space! For most practical purposes, we consider matter as homogeneous when any discontinuities it contains are too small to affect its visual appearance.

How large must a molecule or an agglomeration of molecules be before it begins to exhibit properties of a being a separate phase? Such particles span the gap between the micro and macro worlds, and have been known as *colloids* since they began to be studied around 1900. But with the development of nanotechnology in the 1990s, this distinction has become even more fuzzy.

Pure Substances and Mixtures

The air around us, most of the liquids and solids we encounter, and all too much of the water we drink consists not of pure substances, but of *mixtures*. You probably have a general idea of what a mixture is, and how it differs from a pure substance; what is the scientific criterion for making this distinction?

To a chemist, a *pure substance* usually refers to a sample of matter that has a distinct set of properties that are common to all other samples of that substance. A good example would be ordinary salt, sodium chloride. No matter what its source (from a mine, evaporated from seawater, or made in the laboratory), all samples of this substance, *once they have been purified*, possess the same unique set of properties.

A *pure substance* is one whose intensive properties are the same in any purified sample of that same substance.

A *mixture*, in contrast, is composed of two or more substances, and it can exhibit a wide range of properties depending on the relative amounts of the components present in the mixture. For example, you can dissolve up to 357 g of salt in one litre of water at room temperature, making possible an infinite variety of "salt water" solutions. For each of these concentrations, properties such as the density, boiling and freezing points, and the vapor pressure of the resulting solution will be different.

Those of us who enjoy **peanut butter** would never willingly purchase a brand advertised as "impure". But a *Consumer Reports* article published some years ago showed a table listing the number of "mouse droppings" and "insect parts" (presumably from peanut storage silos) they found in samples of all the major brands. *Bon appetit*!

Finally, we all prefer to drink "pure" water, but we don't usually concern ourselves with the dissolved atmospheric gases and ions such as Ca^{2+} and HCO_3^- that are present in most drinking waters. But these harmless "impurities" are always present in those "pure" spring waters.

and *para*-water, respectively.

The bottom line: To a chemist, the term "pure" has meaning only in the context of a particular application or process.

Operational and conceptual classifications

Since chemistry is an experimental science, we need a set of experimental criteria for placing a given sample of matter in one of these categories. There is no single experiment that will always succeed unambiguously deciding this kind of question. However, there is one principle that will always work in theory, if not in practice. This is based on the fact that the various components of a mixture can, in principle, always be separated into pure substances.

Consider a heterogeneous mixture of salt water and sand. The sand can be separated from the salt water by the mechanical process of filtration. Similarly, the butterfat contained in milk may be separated from the water by a process known as *churning*, in which mechanical agitation forces the butterfat droplets to coalesce into the solid mass we know as butter. These examples illustrate the general principle that heterogeneous matter may be separated into homogeneous matter by mechanical means.

Turning this around, we have an *operational definition* of heterogeneous matter: If, by some *mechanical operation* we can separate a sample of matter into two or more other kinds of matter, then our original sample was heterogeneous. To find a similar operational definition for homogeneous mixtures, consider how we might separate the two components of a solution of salt water. The most obvious way would be to evaporate off the water, leaving the salt as a solid residue. Thus a homogeneous mixture can be separated into pure substances by undergoing appropriate *partial* changes of state— that is, by evaporation, freezing, etc.

Note the term *partial* in the above sentence; in the last example, we evaporate only the water, not the salt (which would be very difficult to do anyway!) The idea is that one component of the mixture is preferentially affected by the process we are carrying out. This principle will be emphasized in the following examples.

Separating homogeneous mixtures

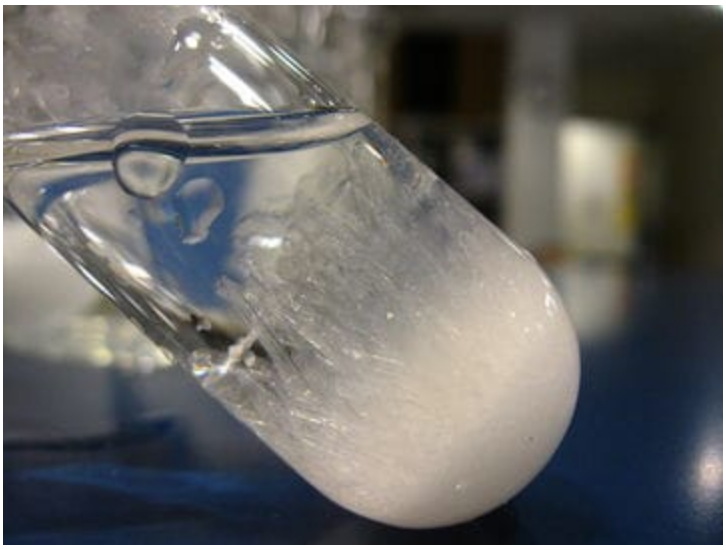
Some common methods of separating homogeneous mixtures into their components are outlined below.

Distillation

A mixture of two volatile liquids is partly boiled away; the first portions of the condensed vapor will be enriched in the component having the lower boiling point. Note that if *all* the liquid were boiled away, the distillate would be identical with the original liquid. But if, say, half of the liquid is distilled, the distillate would contain a larger fraction of the more volatile component. If the distillate is then re-distilled, it can be further enriched in the low-boiling liquid. By repeating this process many times (aided by the fractionating column above the boiling vessel), a high degree of separation can be achieved.

Fractional crystallization

A hot saturated solution containing two or more dissolved solids is allowed to cool slowly; the least-soluble material crystallizes out first, and can be separated by filtration. This process is widely employed both in the laboratory and, on a much larger scale, in industry.



Similarly, a molten mixture of several components, when slowly cooled, will first yield crystals of the material having the highest melting point. This process occurs on a huge scale in nature when molten magma from the earth's mantle rises into the lithosphere and cools underground — a process that can take up to a million years. This is how the common rock known as granite is formed. Eventually these rocks rise and become exposed on the earth's surface.



Figure 2.1.1: Granite is a major component of California's Sierra Nevada Mountains, where it forms the well-known peak Half Dome in Yosemite Valley (left). A magnified image of granite shows the various components that have crystallized out over time as the melt gradually cooled. Note that because the various components of granite are not removed as they crystallize out, they remain embedded in the solidified material.

Liquid-liquid Extraction

Two mutually-insoluble liquids, one containing two or more solutes (dissolved substances), are shaken together in a *separatory funnel*. Each solute will concentrate in the liquid in which it is more soluble. The two solutions are then separated by opening the stopcock at the bottom, allowing the more dense solution to drain out.



Solid-liquid Extraction

In working with natural products such as plant materials, a first step is often to extract soluble substances from the plant parts. This, and similar extractions of the soluble components of complex solids, is carried out in an apparatus known as a Soxhlet extractor.



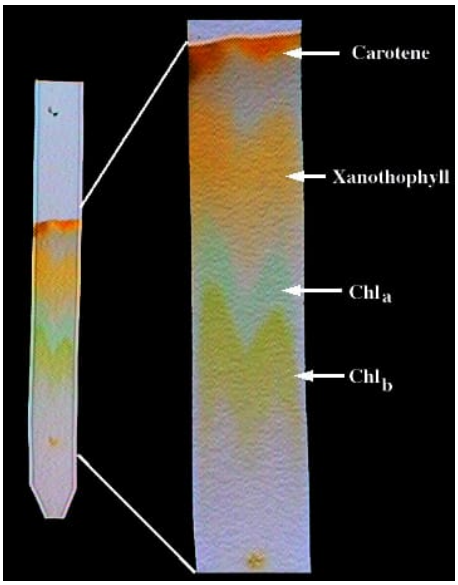
Figure 2.1.1: Soxlet extractor

The idea is to continuously percolate an appropriate hot solvent through the material, which is contained in a porous paper "thimble". Hot vapor from the boiling flask bypasses the extraction chamber through the arm at the left (labeled "vapor" in the illustration →) and into the condenser, from which it drips down into the extraction chamber, where a portion of the soluble material mixes with the solvent. When the condensate reaches the top of the chamber, it flows out through the siphon arm, emptying its contents into the boiling flask, which becomes increasingly concentrated in the extracted material.

The advantage of this arrangement is that the percolation-and-extraction process can be repeated indefinitely (usually hours to days) without much attention.

Chromatography

As a liquid or gaseous mixture flows along a column containing an adsorbant material, the more strongly-adsorbed components tend to move more slowly and emerge later than the less-strongly adsorbed components. In this example, an extract made from plant leaves is separated into its principle components: carotene, xanthophyll, and chlorophylls A and B.



Although chromatography originated in the mid-19th Century, it was not widely employed until the 1950's. Since that time, it has encompassed a huge variety of techniques and is no longer limited to colored substances. Chromatography is now one of the most widely-employed methods for the analysis and separation of complex mixtures of liquids and gases.

Physical and Chemical Properties

Since chemistry is partly the study of the transformations that matter can undergo, we can also assign to any substance a set of chemical properties that express the various changes of composition the substance is known to undergo. Chemical properties also include the conditions of temperature, etc., required to bring about the change, and the amount of energy released or absorbed as the change takes place.

The properties that we described above are traditionally known as *physical properties*, and are to be distinguished from *chemical properties* that usually refer to changes in composition that a substance can undergo. For example, we can state some of the more distinctive physical and chemical properties of the element **sodium**:

Physical properties (25 °C)	Chemical properties
• appearance: a soft, shiny metal • density: 0.97 g/cm^3 • melting point: 97.5 °C • boiling point: 960 °C	• forms an oxide Na_2O and a hydride NaH • burns in air to form sodium peroxide Na_2O_2 • reacts violently with water to release hydrogen gas • dissolves in liquid ammonia to form a deep blue solution

The more closely one looks at the distinction between physical and chemical properties, the more blurred this distinction becomes. For example, the high boiling point of water compared to that of methane, CH_4, is a consequence of the electrostatic attractions between O-H bonds in adjacent molecules, in contrast to those between C-H bonds; at this level, we are really getting into chemistry! So although you will likely be expected to "distinguish between" physical and chemical properties on an exam, don't take it too seriously — this turns out to be a rather dubious dichotomy, loved by teachers, but of limited usefulness!

Answers

Example 2.1.1

extensive; depends on size of the mug.
intensive; same for any same-sized sample.
extensive; depends on size and sugar content of the banana.
intensive; same for any 10 g portion of sugar.

extensive; depends on volume of blood in the body.
intensive; the same for any 5 mL sample.
extensive; depends on length of the wire.
intensive; same for any 1 km length of the same wire.
pressure itself is intensive, but is also dependent on the quantity of air in the tire.

This page titled 2.1: Classification and Properties of Matter is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Energy, Heat, and Temperature

Learning Objectives

- Explain the difference between *kinetic energy* and *potential energy*.
- Define *chemical energy* and *thermal energy*.
- Define *heat* and *work*, and describe an important limitation in their interconversion.
- Describe the physical meaning of temperature.
- Explain the meaning of a *temperature scale* and describe how a particular scale is defined.
- Convert a temperature expressed in Fahrenheit or Celsius to the other scale.
- Describe the *Kelvin temperature scale* and its special significance.
- Define *heat capacity* and *specific heat*, and explain how they can be measured.

All chemical changes are accompanied by the absorption or release of heat. The intimate connection between matter and energy has been a source of wonder and speculation from the most primitive times; it is no accident that fire was considered one of the four basic elements (along with earth, air, and water) as early as the fifth century BCE. This unit will cover only the very basic aspects of the subject, just enough to get you started; there is a far more complete set of lessons on **chemical energetics** elsewhere.

What is Energy?

Energy is one of the most fundamental and universal concepts of physical science, but one that is remarkably difficult to define in way that is meaningful to most people. This perhaps reflects the fact that energy is not a “thing” that exists by itself, but is rather an attribute of matter (and also of electromagnetic radiation) that can manifest itself in various ways. It can be observed and measured only indirectly through its effects on matter that acquires, loses, or possesses it. Energy can take many forms: mechanical, chemical, electrical, radiation (light), and thermal. You also know that energy is *conserved*; it can be passed from one object or place to another, but it can never simply disappear.

In the 17th Century, the great mathematician **Gottfried Leibniz** (1646-1716) suggested the distinction between *vis viva* ("live energy") and *vis mortua* ("dead energy"), which later became known as kinetic energy and potential energy. Except for radiant energy that is transmitted through an electromagnetic field, most practical forms of energy we encounter are of two kinds: kinetic and potential.

- Kinetic energy is associated with the **motion** of an object; a body with a mass, *m,* and moving at a velocity, *v,* possesses the kinetic energy $\frac{1}{2}mv^2$. This "v-squared" part is important; if you double your speed, you consume four times as much fuel (glucose for the runner, gasoline or electricity for your car.
- **Potential energy** is energy a body has by virtue of its location in a *force field*— a gravitational, electrical, or magnetic field. For example, if an object of mass *m* is raised off the floor to a height *h*, its potential energy increases by *mgh*, where *g* is a proportionality constant known as the *acceleration of gravity*. Similarly, the potential energy of a particle having an electric charge *q* depends on its location in an electrostatic field.

Kinetic and potential energy are freely interconvertible

Pick up a book and hold it above the table top; you have just increased its potential energy in the force field of the earth's gravity. Now let it drop. Its newly-acquired potential energy begins to re-appear as kinetic energy as it accelerates downward at a velocity increasing by 9.8 m/sec every second (9.8 m sec^{-2} or 32 ft sec^{-2}). At the instant it strikes the surface, the potential energy you gave supplied to the book has now been entirely converted into kinetic energy.

And what happens to that kinetic energy after the book stops moving? It is still there, but you can no longer see its effect; it has now become dispersed as *thermal kinetic energy* ("heat") into the molecules of the book, the table top, and, ultimately, into the surroundings, including the air.

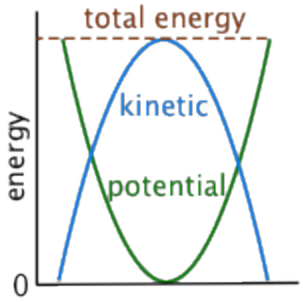


Figure: In absence of friction, the total energy remains constant

The more you think about it, the more examples of kinetic-potential conversion you will find in everyday life. In many other instances, however, the energy of an object can be seen to repeatedly alternate between potential and kinetic forms. Left alone, the process continues indefinitely until friction has dissipated the energy into the surroundings.

Energy's graveyard: Thermal energy

Energy is conserved: it can neither be created nor destroyed. But it can, and eventually always will, disappear from our view and into the microscopic world of individual molecular particles. All molecules are in a continual state of motion, and they therefore possess kinetic energy. But unlike the motion of a massive body such as a baseball or a car that is moving along a defined trajectory, the motions of individual atoms or molecules are random and chaotic, forever changing in magnitude and direction as they collide with each other or (as in the case of a gas) with the walls of the container.

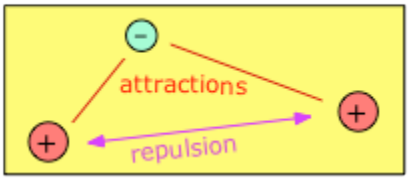
The sum total of all of this microscopic-scale randomized kinetic energy within a body is given a special name, *thermal energy*. Although we cannot directly *see* thermal energy in action, we can certainly *feel* it; as we will see further, it correlates directly with the *temperature* of an object.

The chemistry connection

Atoms and molecules are the principle actors of thermal energy, but they possess other kinds of energy as well that plays a major role in chemistry.

Bond energy

H_2^+ is energetically stable enough to exist as an identifiable entity, and thus fits the definition a molecule. But it is also extremely reactive, so it does not sit around for very long. It can only be observed when a high-voltage electrical discharge is passed through hydrogen gas; the blue glow one sees represents its demise as it picks up electrons and reverts to the far more stable dihydrogen molecule H_2.



Consider, for example, the *simplest possible molecule*. This is the *hydrogen molecule ion*, H_2^+, in which a single electron simultaneously attracts two protons. These protons, having identical charges, repel each other, but this is overcome by the electron-proton attractions, leading to a net decrease in potential energy when an electron combines with two protons. This potential energy decrease is sufficient to enable H_2^+ to exist as a discrete molecule which we can represent as $[H—H]^+$ in order to explicitly depict the *chemical bond* that joins the two atoms.

The *strength* of a chemical bond increases as the potential energy associated with its formation becomes more negative.

Chemical bonds also possess some *kinetic energy* that is associated with the "motion" of the electron as it spreads itself into

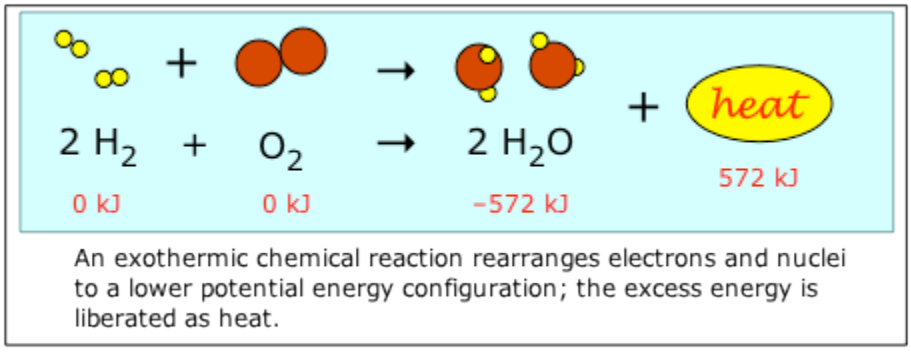
the extended space it occupies in what we call the "bond". This is a quantum effect that has no classical counterpart. The kinetic energy has only half the magnitude of the potential energy and works against it; the total bond energy is the sum of the two energies.

Chemical energy

The chemical bonds in the glucose molecules store the energy that fuels our bodies.

Molecules are vehicles both for **storing and transporting energy**, and the means of **converting it from one form to another** when the formation, breaking, or rearrangement of the chemical bonds within them is accompanied by the uptake or release of energy, most commonly in the form of *heat*.

Chemical energy refers to the potential and kinetic energy associated with the *chemical bonds* in a molecule. Consider what happens when hydrogen and oxygen combine to form water. The reactants H_2 and O_2 contain more bond energy than H_2O, so when they combine, the excess energy is liberated given off in the form of thermal energy, or "heat".



An exothermic chemical reaction rearranges electrons and nuclei to a lower potential energy configuration; the excess energy is liberated as heat.

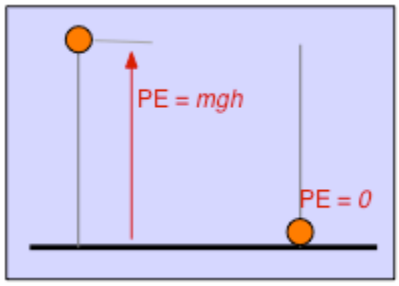
By convention, the energy content of the chemical elements in their natural state (H_2 and O_2 in this example) are defined as "zero". This makes calculations much easier, and gives most compounds negative "energies of formation". (see below)

Chemical energy manifests itself in many different ways:

- chemical → thermal → kinetic chemical → thermal → kinetic + radiant
- chemical → electrical → kinetic (nerve function, muscle movement)
- chemical → electrical

Energy scales are always arbitrary

You might at first think that a book sitting on the table has zero kinetic energy since it is not moving. In truth, however, the earth itself is moving; it is spinning on its axis, it is orbiting the sun, and the sun itself is moving away from the other stars in the general expansion of the universe. Since these motions are normally of no interest to us, we are free to adopt an arbitrary scale in which the velocity of the book is measured with respect to the table; on this so-called *laboratory coordinate system*, the kinetic energy of the book can be considered zero.



We do the same thing with potential energy. If we define the height of the table top as the zero of potential energy, then an object

having a mass m suspended at a height *h* above the table top will have a potential energy of *mgh*. Now let the object fall; as it accelerates in the earth's gravitational field, its potential energy changes into kinetic energy. An instant before it strikes the table top, this transformation is complete and the kinetic energy $\frac{1}{2}mv^2$ is identical with the original *mgh*. As the object comes to rest, its kinetic energy appears as heat (in both the object itself and in the table top) as the kinetic energy becomes randomized as thermal energy.

Energy units

Energy is measured in terms of its ability to perform work or to transfer heat. Mechanical work is done when a force *f* displaces an object by a distance *d*:

$$W = f \cdot d \tag{2.2.1}$$

The basic unit of energy is the *joule*. One joule is the amount of work done when a force of 1 newton acts over a distance of 1 m; thus 1 J = 1 N-m. One newton is the amount of force required to accelerate a 1 - kg mass by 1 meter per second in one second - 1 m sec^{-2}, so the basic dimensions of the joule are kg m^2 s^{-2}. The other two units of energy that are in wide use are the *calorie* and the *BTU* (British thermal unit). These are defined in terms of the heating effect on water. For the moment, we will confine our attention to joule and calorie.

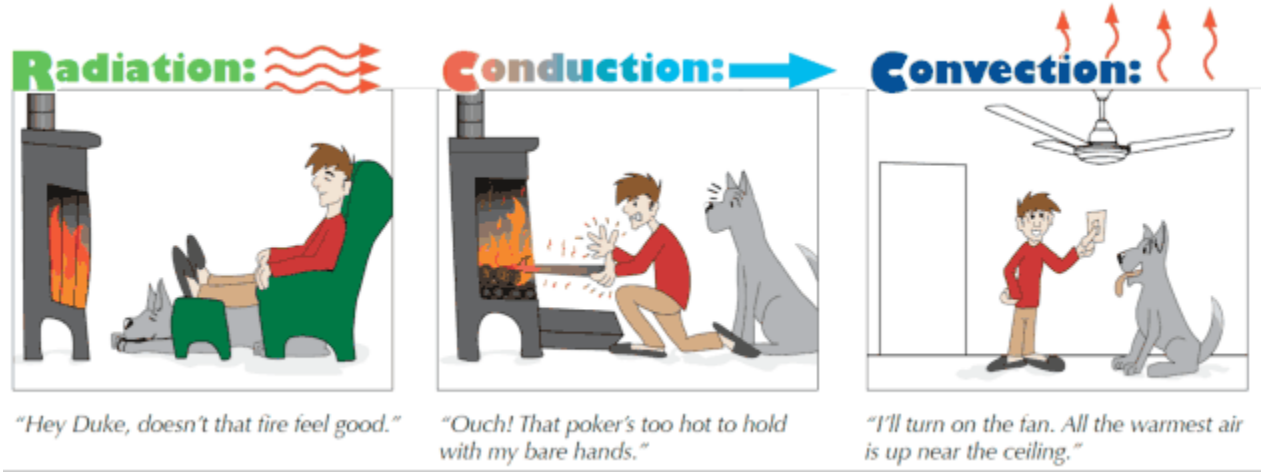
Heat and *work* are both measured in energy units, but they do not constitute energy itself. As we will explain below, they refer to *processes* by which energy is transferred to or from something— a block of metal, a motor, or a cup of water.

Heat

When a warmer body is brought into contact with a cooler body, thermal energy flows from the warmer one to the cooler until their two temperatures are identical. The warmer body loses a quantity of thermal energy ΔE, and the cooler body acquires the same amount of energy. We describe this process by saying that "ΔE joules of heat has passed from the warmer body to the cooler one." It is important, however, to understand that Heat is the transfer of energy due to a difference in temperature.

We often refer to a "flow" of heat, recalling the 18th-century notion that heat was an actual substance called “caloric” that could flow like a liquid. This is a misnomer; heat is a *process* and is not something that can be contained or stored in a body. It is important that you understand this, because the use of the term in our ordinary conversation ("the heat is terrible today") tends to make us forget this distinction.

There are basically three mechanisms by which heat can be transferred: conduction, radiation, and convection. The latter process occurs when the two different temperatures cause different parts of a fluid to have different densities.



Work

Work is the transfer of energy by any process other than heat.

Work, like energy, can take various forms: mechanical, electrical, gravitational, etc. All have in common the fact that they are the product of two factors, an *intensity term* and a *capacity term*. For example, the simplest form of mechanical work arises when an object moves a certain distance against an opposing force. Electrical work is done when a body having a certain charge moves through a potential difference.

type of work	intensity factor	capacity factor	formula
mechanical	force	change in distance	$f\Delta x$
gravitational	gravitational potential (a function of height)	mass	*mgh*
electrical	potential difference	quantity of charge	$Q\Delta V$

Performance of work involves a transformation of energy; thus when a book drops to the floor, gravitational work is done (a mass moves through a gravitational potential difference), and the potential energy the book had before it was dropped is converted into kinetic energy which is ultimately dispersed as thermal energy.

Mechanical work is the product of the force exerted on a body and the distance it is moved: 1 N-m = 1 J.

Heat and work are best thought of as processes by which energy is exchanged, rather than as energy itself. That is, heat "exists" only when it is flowing, work "exists" only when it is being done.

When two bodies are placed in thermal contact and energy flows from the warmer body to the cooler one,we call the process "heat". A transfer of energy to or from a system by any means other than heat is called "work".

So you can think of heat and work as just different ways of accomplishing the same thing: the transfer of energy from one place or object to another.

To make sure you understand this, suppose you are given two identical containers of water at 25°C. Into one container you place an electrical immersion heater until the water has absorbed 100 joules of heat. The second container you stir vigorously until 100 J of work has been performed on it. At the end, both samples of water will have been warmed to the same temperature and will contain the same increased quantity of thermal energy. There is no way you can tell which contains "more work" or "more heat".

An important limitation on energy conversion

A gas engine converts the chemical energy available in its fuel into thermal energy. Only a part of this is available to perform work; the remainder is dispersed into the surroundings through the exhaust. This limitation is the essence of the Second Law of Thermodynamics which we will get to much later in this course

Thermal energy is very special in one crucial way. All other forms of energy are *interconvertible*: mechanical energy can be completely converted to electrical energy, and the latter can be completely converted to thermal, as in the water-heating example described above. But although work can be completely converted into thermal energy, **complete conversion of thermal energy into work is impossible.** A device that *partially* accomplishes this conversion is known as a *heat engine*; a steam engine, a jet engine, and the internal combustion engine in a car are well-known examples.

Temperature and its meaning

We all have a general idea of what *temperature* means, and we commonly associate it with "heat", which, as we noted above, is a widely misunderstood word. Both relate to what we described above as *thermal energy*—the randomized kinetic energy associated with the various motions of matter at the atomic and molecular levels.

Heat, you will recall, is not something that is "contained within" a body, but is rather a *process* in which [thermal] energy enters or leaves a body as the result of a *temperature difference*.

So when you warm up your cup of tea by allowing it to absorb 1000 J of heat from the stove, you can say that the water has

acquired 1000 J of *energy* — but *not* of *heat*. If, instead, you "heat" your tea in a microwave oven, the water acquires its added energy by direct absorption of electromagnetic energy; because this process is not driven by a temperature difference, heat was not involved at all!!

Thermometry

We commonly measure temperature by means of a *thermometer* — a device that employs some material possessing a property that varies in direct proportion to the temperature. The most common of these properties are the density of a liquid, the thermal expansion of a metal, or the electrical resistance of a material.



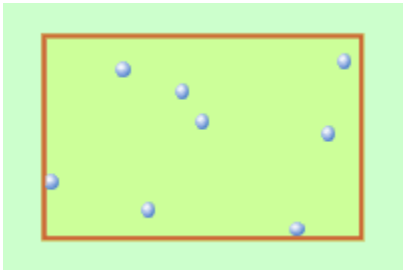
The ordinary thermometer we usually think of employs a reservoir of liquid whose thermal expansion (decrease in density) causes it to rise in a capillary tube. Metallic mercury has traditionally been used for this purpose, as has an alcohol (usually isopropyl) containing a red dye.

Mercury was the standard thermometric liquid of choice for more than 200 years, but its use for this purpose has been gradually phased out owing to its neurotoxicity. Although coal-burning, disposal of fluorescent lamps, incineration and battery disposal are major sources of mercury input to the environment, broken thermometers have long been known to release hundreds of tons of mercury. Once spilled, tiny drops of the liquid metal tend to lodge in floor depressions and cracks where they can emit vapor for years.

Temperature

Temperature is a measure of the *average kinetic energy* of the molecules within the water. You can think of temperature as an expression of the "intensity" with which the thermal energy in a body manifests itself in terms of chaotic, microscopic molecular motion.

- ***Heat*** is the *quantity* of thermal energy that enters or leaves a body.
- ***Temperature*** measures the *average translational kinetic energy* of the molecules in a body.



This animation depicts thermal translational motions of molecules in a gas. In liquids and solids, there is vary little empty space between molecules, and they mostly just bump against and jostle one another.

You will notice that we have sneaked the the word "*translational*" into this definition of temperature. *Translation* refers to a change in location: in this case, molecules moving around in random directions. This is the major form of thermal energy under ordinary conditions, but molecules can also undergo other kinds of motion, namely rotations and internal vibrations. These latter two forms

of thermal energy are not really "chaotic" and do not contribute to the temperature.

Energy is measured in *joules*, and temperature in *degrees*. This difference reflects the important distinction between energy and temperature:

- We can say that 100 g of hot water contains more energy (*not heat*!) than 100 g of cold water. And because energy is an extensive quantity, we know that a 10-g portion of this hot water contains only ten percent as much energy as the entire 100-g amount.
- Temperature, by contrast, is *not* a measure of quantity; being an intensive property, it is more of a "quality" that describes the "intensity" with which thermal energy manifests itself. So both the 100-g and 10-g portions of the hot water described above possess the same temperature.

Temperature scales

Temperature is measured by observing its effect on some temperature-dependent variable such as the volume of a liquid or the electrical resistance of a solid. In order to express a temperature numerically, we need to define a *scale* which is marked off in uniform increments which we call *degrees*. The nature of this scale — its zero point and the magnitude of a degree, are completely arbitrary.

Although rough means of estimating and comparing temperatures have been around since AD 170, the first mercury thermometer and temperature scale were introduced in Holland in 1714 by Gabriel Daniel Fahrenheit.



Fahrenheit established three fixed points on his thermometer. Zero degrees was the temperature of an ice, water, and salt mixture, which was about the coldest temperature that could be reproduced in a laboratory of the time. When he omitted salt from the slurry, he reached his second fixed point when the water-ice combination stabilized at "the thirty-second degree." His third fixed point was "found as the ninety-sixth degree, and the spirit expands to this degree when the thermometer is held in the mouth or under the armpit of a living man in good health." After Fahrenheit died in 1736, his thermometer was recalibrated using 212 degrees, the temperature at which water boils, as the upper fixed point. Normal human body temperature registered 98.6 rather than 96.

Belize and the U.S.A. are the only countries that still use the Fahrenheit scale!

In 1743, the Swedish astronomer Anders Celsius devised the aptly-named *centigrade* scale that places exactly 100 degrees between the two reference points defined by the freezing- and boiling points of water.

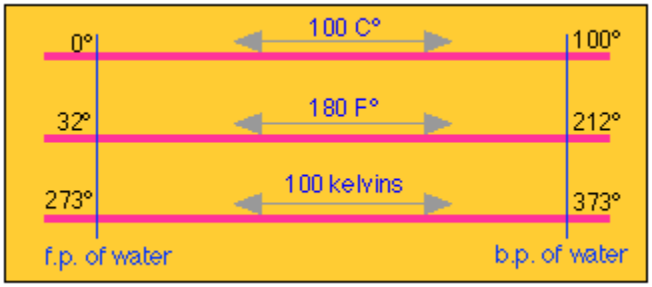
For reasons best known to Celsius, he assigned 100 degrees to the freezing point of water and 0 degrees to its boiling point, resulting in an inverted scale that nobody liked. After his death a year later, the scale was put the other way around. The revised centigrade scale was quickly adopted everywhere except in the English-speaking world, and became the metric unit of temperature. In 1948 it was officially renamed as the Celsius scale.

Temperature comparisons and conversions

When we say that the temperature is so many degrees, we must specify the particular scale on which we are expressing that temperature. A temperature scale has two defining characteristics, both of which can be chosen arbitrarily:

- The temperature that corresponds to 0° on the scale;
- The magnitude of the *unit increment* of temperature– that is, the size of the *degree*.

In order to express a temperature given on one scale in terms of another, it is necessary to take both of these factors into account.



Converting between Celsius and Fahrenheit is easy if you bear in mind that between the so-called ice- and steam points of water there are 180 Fahrenheit degrees, but only 100 Celsius degrees, making the F° 100/180 = 5/9 the magnitude of the C°.

Because the ice point is at 32 °F, the two scales are offset by this amount. If you remember this, there is no need to memorize a conversion formula; you can work it out whenever you need it. Note the distinction between "°C" (a temperature) and "C°" (a temperature *increment*).

Absolute temperature scales

Near the end of the 19th Century when the physical significance of temperature began to be understood, the need was felt for a temperature scale whose zero really means *zero* — that is, the complete absence of thermal motion. This gave rise to the *absolute temperature scale* whose zero point is –273.15 °C, but which retains the same degree magnitude as the Celsius scale. This was eventually renamed after Lord Kelvin (William Thompson) thus the Celsius degree became the *kelvin*. It is now common to express an increment such as five C° as "five kelvins"

In 1859 the Scottish engineer and physicist William J.M. Rankine proposed an absolute temperature scale based on the Fahrenheit degree. Absolute zero (0° Ra) corresponds to –459.67°F. The *Rankine scale* has been used extensively by those same American and British engineers who delight in expressing energies in units of BTUs and masses in pounds.

The importance of absolute temperature scales is that absolute temperatures can be entered directly in all the fundamental formulas of physics and chemistry in which temperature is a variable. Perhaps the most common example, known to all beginning students, is the ideal gas equation state.

$$PV = nRT \tag{2.2.2}$$

Heat capacity

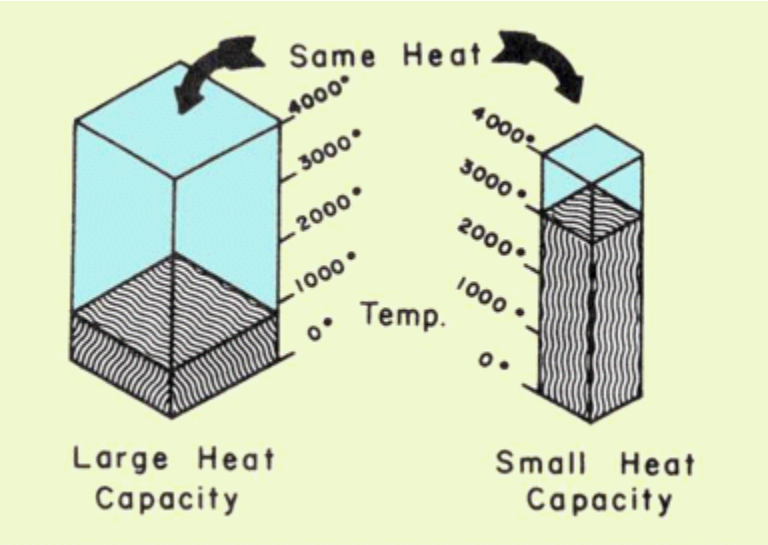
As a body loses or gains heat, its temperature changes in direct proportion to the amount of thermal energy *q* transferred:

$$q = C\Delta T \tag{2.2.3}$$

The proportionality constant *C* is known as the *heat capacity*

$$C = \frac{q}{\Delta T} \tag{2.2.4}$$

If ΔT is expressed in kelvins (degrees) and q in joules, the units of C are J K^{-1}. In other words, the heat capacity tells us how many joules of energy it takes to change the temperature of a body by 1 C°. The greater the value of C, the the smaller will be the effect of a given energy change on the temperature.



It should be clear that C is an *extensive* property— that is, it depends on the quantity of matter. Everyone knows that a much larger amount of energy is required to bring about a 10 C° change in the temperature of 1 L of water compared to 10 mL of water. For this reason, it is customary to express C in terms of unit quantity, such as per gram, in which case it becomes the *specific heat capacity*, commonly referred to as the "specific heat" and has the units J $K^{-1}g^{-1}$.

Thus if identical quantities of heat flow into two bodies having different heat capacities, the one having the smaller heat capacity will undergo the greater change in temperature. (You might find it helpful to think of heat capacity as a measure of a body's ability to *resist* a change of temperature when absorbing or losing heat.) Note: you are expected to know the units of specific heat. The advantage of doing so is that you need not learn a "formula" for solving specific heat problems.

Example 2.2.1

How many joules of heat must flow into 150 mL of water at 0 °C to raise its temperature to 25 °C?

Solution

The mass of the water is (150 mL) × (1.00 g mL^{-1}) = 150 g. The specific heat of water is 4.18 J K^{-1} g^{-1}. From the definition of specific heat, the quantity of energy

$$q = \Delta E \text{ is } (150 \text{ g})(25.0 \text{ K})(4.18 \text{ J K}^{-1} \text{ g}^{-1}) = 16700 \text{ J}.$$

How can I rationalize this procedure? It should be obvious that the greater the mass of water and the greater the temperature change, the more heat will be required, so these two quantities go in the numerator. Similarly, the energy required will vary inversely with the specific heat, which therefore goes in the denominator.

Table 2.2.1: Specific heat capacities of some common substances

Substance	C, J /g-K
Aluminum	0.900
Copper	0.386

Lead	0.128
Mercury	0.140
Zinc	0.387
Alcohol (ethanol)	2.4
Water	4.18
Ice (-10° C)	2.05
Gasoline (*n*-octane)	0.53
Glass	0.84
Carbon (graphite/diamond)	0.710 / .509
Sodium chloride	0.854
Rock (granite)	0.790
Air	1.01

Note especially the following:

- The molar heat capacities of the metallic elements are almost identical. This is the basis of the *Law of Dulong and Petit*, which served as an important tool for estimating the atomic weights of some elements.
- The intermolecular hydrogen bonding in **water** and alcohols results in anomalously high heat capacities for these liquids; the same is true for ice, compared to other solids.
- The values for graphite and diamond are consistent with the principle that solids that are more “ordered” tend to have larger heat capacities.

Example 2.2.1:

A piece of nickel weighing 2.40 g is heated to 200.0 °C, and is then dropped into 10.0 mL of water at 15.0 °C. The temperature of the metal falls and that of the water rises until thermal equilibrium is attained and both are at 18.0 °C. What is the specific heat of the metal?

Solution

The mass of the water is (10 mL) × (1.00 g mL^{-1}) = 10 g. The specific heat of water is 4.18 1 J K^{-1} g^{-1} and its temperature increased by 3.0 C°, indicating that it absorbed (10 g)(3 K)(4.18 J K^{-1} g^{-1}) = 125 J of energy. The metal sample lost this same quantity of energy, undergoing a temperature drop of 182 C° as the result. The specific heat capacity of the metal is:

$$(125\ J) / (2.40\ g)(182\ K) = 0.287\ J\ K^{-1}\ g^{-1}.$$

Notice that no "formula" is required here as long as you know the units of specific heat; you simply place the relevant quantities in the numerator or denominator to make the units come out correctly.

This page titled 2.2: Energy, Heat, and Temperature is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

2.3: The Measure of Matter

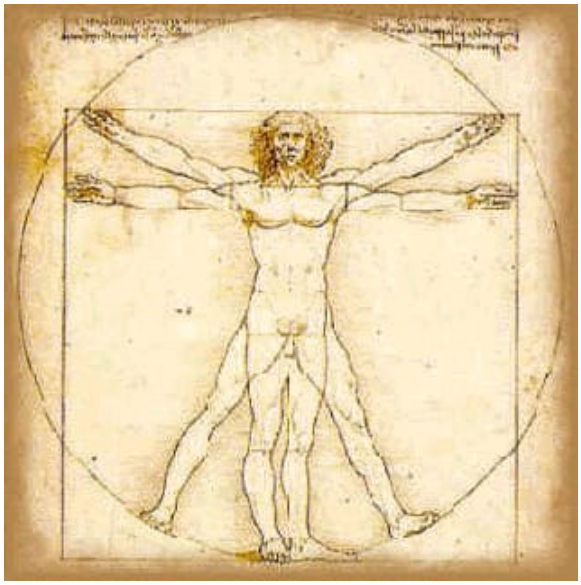
Learning Objectives

- Describe the names and abbreviations of the *SI base units* and the *SI decimal prefixes*.
- Define the *liter* and the *metric ton* in these units.
- Explain the meaning and use of *unit dimensions*; state the dimensions of *volume*.
- State the quantities that are needed to define a *temperature scale*, and show how these apply to the *Celsius, Kelvin,* and *Fahrenheit* temperature scales.
- Explain how a Torricellian *barometer* works.

The natural sciences begin with ***observation*****,** and this usually involves ***numerical measurements*** of quantities such as length, volume, density, and temperature. Most of these quantities have ***units*** of some kind associated with them, and these units must be retained when you use them in calculations. All measuring units can be defined in terms of a very small number of fundamental ones that, through "dimensional analysis", provide insight into their derivation and meaning, and must be understood when converting between different unit systems.

Units of Measure

Have you ever estimated a distance by “stepping it off”— that is, by counting the number of steps required to take you a certain distance? Or perhaps you have used the width of your hand, or the distance from your elbow to a fingertip to compare two dimensions. If so, you have engaged in what is probably the first kind of measurement ever undertaken by primitive mankind.



Leonardo da Vinci - Vitruvian Man

The results of a measurement are always expressed on some kind of a *scale* that is defined in terms of a particular kind of *unit*. The first scales of distance were likely related to the human body, either directly (the length of a limb) or indirectly (the distance a man could walk in a day). As civilization developed, a wide variety of measuring scales came into existence, many for the same quantity (such as length), but adapted to particular activities or trades. Eventually, it became apparent that in order for trade and commerce to be possible, these scales had to be defined in terms of standards that would allow measures to be verified, and, when expressed in different units (bushels and pecks, for example), to be correlated or converted.

Over the centuries, hundreds of measurement units and scales have developed in the many civilizations that achieved some literate means of recording them. Some, such as those used by the Aztecs, fell out of use and were largely forgotten as these

civilizations died out. Other units, such as the various systems of measurement that developed in England, achieved prominence through extension of the Empire and widespread trade; many of these were confined to specific trades or industries. The examples shown here are only some of those that have been used to measure length or distance. The history of measuring units provides a fascinating reflection on the history of industrial development.

Ångstrom	1E−6 cm
Astronomical unit	1.495E8 km
Bolt (U.S., cloth)	120 ft
Cable length	720 ft
Chain (engineer's)	100 ft
Chain (surveyor's)	0.1 furlong
Cubit	18 in
Ell	45 in
Foot (U.S.)	12 in
Foot (British)	0.4 pace
Furlong	1/8 mile
Hand	4 in
Inch (U.S.)	1/12 ft
Inch (British)	1/36 yard
League	3 miles
Light year	9.46E12 km
Mile (U.S., staute)	5280 ft
Mile (nautical)	1.853 km
Nail (British)	2.25 in
Pace (British)	30 in
Parsec	3.084E13 km
Point (printer's)	1/72 in
Rope (British)	20 ft
Span	9 in

The most influential event in the history of measurement was undoubtedly the French Revolution and the Age of Enlightenment that followed. This led directly to the metric system that attempted to do away with the confusing multiplicity of measurement scales by reducing them to a few fundamental ones that could be combined in order to express any kind of quantity. The metric system spread rapidly over much of the world, and eventually even to England and the rest of the U.K. when that country established closer economic ties with Europe in the latter part of the 20th Century. The United States is presently the only major country in which “metrication” has made little progress within its own society, probably because of its relative geographical isolation and its vibrant internal economy.

The SI Units

Science, being a truly international endeavor, adopted metric measurement very early on; engineering and related technologies have been slower to make this change, but are gradually doing so. Even the within the metric system, however, a variety of units were employed to measure the same fundamental quantity; for example, energy could be expressed within the metric system in units of ergs, electron-volts, joules, and two kinds of calories. This led, in the mid-1960s, to the adoption of a more basic set of units, the *Systeme Internationale* (*SI*) units that are now recognized as the standard for science and, increasingly, for technology of all kinds.

In principle, any physical quantity can be expressed in terms of only seven *base units*. Each base unit is defined by a standard which is described in the NIST Web site.

kelvin

The seven base units in the SI system

Observable	Base Unit	Abbreviation
length	meter	m
mass	kilogram	kg
time	second	s
temperature (absolute)	Kelvin	K
amount of substance	mole	mol

Observable	Base Unit	Abbreviation
electric current	ampere	A
luminous intensity	candela	cd

- The base unit of **mass** is unique in that a decimal prefix (see below) is built-in to it; that is, it is not the *gram*, as you might expect.
- The base unit of **time** is the only one that is not metric. Numerous attempts to make it so have never garnered any success; we are still stuck with the 24:60:60 system that we inherited from ancient times. (The ancient Egyptians of around 1500 BC invented the 12-hour day, and the 60:60 part is a remnant of the base-60 system that the Sumerians used for their astronomical calculations around 100 BCE.)
- Of special interest to Chemistry is the *mole*, the base unit for expressing the **quantity of matter**. The number is also known as Avogadro's number and is exactly $6.02214076 \times 10^{23}$ of anything.

Owing to the wide range of values that quantities can have, it has long been the practice to employ prefixes such as milli and mega to indicate decimal fractions and multiples of metric units. As part of the SI standard, this system has been extended and formalized.

deci

The SI decimal prefixes

prefix	abbreviation	multiplier	--	prefix	abbreviation	multiplier
peta	P	10^{15}		deca	d	10^{-1}
tera	T	10^{12}		centi	c	10^{-2}
giga	G	10^{9}		milli	m	10^{-3}
mega	M	10^{6}		micro	μ	10^{-6}
kilo	k	10^{3}		nano	n	10^{-9}
hecto	h	10^{2}		pico	p	10^{-12}
deca	da	10		femto	f	10^{-15}

There is a category of units that are "honorary" members of the SI in the sense that it is acceptable to use them along with the base units defined above. These include such mundane units as the hour, minute, and degree (of angle), etc., but the three shown here are of particular interest to chemistry, and you will need to know them.

liter (litre)	L	1 L = 1 dm^3 = 10^{-3} m^3
metric ton	t	1 t = 10^3 kg
united atomic mass unit	u	1 u = 1.66054×10^{-27} kg

SI-Derived Units and Dimensional Analysis

Most of the physical quantities we actually deal with in science and also in our daily lives, have units of their own: volume, pressure, energy and electrical resistance are only a few of hundreds of possible examples. It is important to understand, however, that all of these can be expressed in terms of the SI base units; they are consequently known as derived units. In fact, most physical quantities can be expressed in terms of one or more of the following five fundamental units:

mass	length	time	electric charge	temperature
M	L	T	Q	Θ (theta)

Dimensional analysis is an important tool in working with and converting units in calculations. Consider, for example, the unit of ***volume***, which we denote as V. To measure the volume of a rectangular box, we need to multiply the lengths as measured along the three coordinates:

$$V = x \cdot y \cdot z \tag{2.3.1}$$

We say, therefore, that volume has the dimensions of length-cubed:

$$dim.V = L^3 \tag{2.3.2}$$

Thus the units of volume will be m^3 (in the SI) or cm^3, ft^3 (English), etc. Moreover, any formula that calculates a volume must contain within it the L^3dimension; thus the volume of a sphere is 4/3 πr^3.

Units and their Ranges in Chemistry

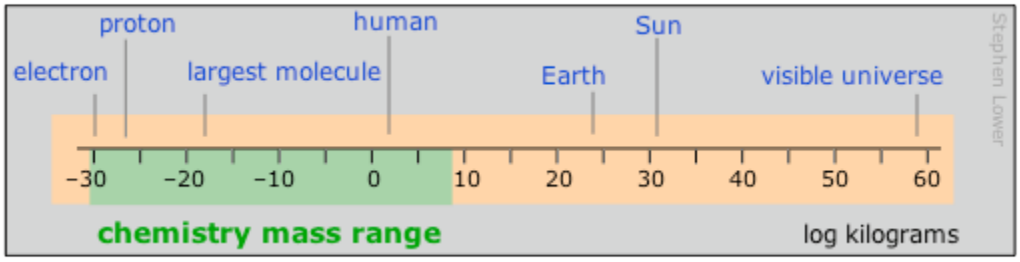
In this section, we will look at some of the quantities that are widely encountered in Chemistry, and at the units in which they are commonly expressed. In doing so, we will also consider the actual range of values these quantities can assume, both in nature in general, and also within the subset of nature that chemistry normally addresses. In looking over the various units of measure, it is interesting to note that their unit values are set close to those encountered in everyday human experience

Ranges of Mass and Weight in Chemistry

These two quantities are widely confused. Although they are often used synonymously in informal speech and writing, they have different dimensions: *weight* is the *force* exerted on a mass by the local gravitational field:

$$f = ma = mg \tag{2.3.3}$$

where *g* is the acceleration of gravity. While the nominal value of the latter quantity is 9.80 m s^{-2} at the Earth's surface, its exact value varies locally. Because it is a force, the proper SI unit of weight is the *newton*, but it is common practice (except in physics classes!) to use the terms "weight" and "mass" interchangeably, so the units *kilograms* and *grams* are acceptable in almost all ordinary laboratory contexts.



Please note that in this diagram and in those that follow, the numeric scale represents the *logarithm* of the number shown. For example, the mass of the electron is 10^{-30} kg.

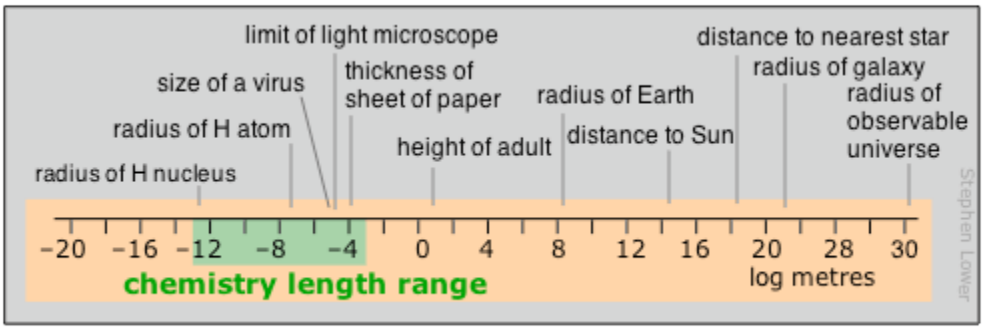
The range of masses spans 90 orders of magnitude, more than any other unit. The range that chemistry ordinarily deals with has greatly expanded since the days when a microgram was an almost inconceivably small amount of material to handle in the laboratory; this lower limit has now fallen to the atomic level with the development of tools for directly manipulating these particles. The upper level reflects the largest masses that are handled in industrial operations, but in the recently developed fields of geochemistry and environmental chemistry, the range can be extended indefinitely. Flows of elements between the various regions of the environment (atmosphere to oceans, for example) are often quoted in t

Range of Distances Encountered in Chemistry

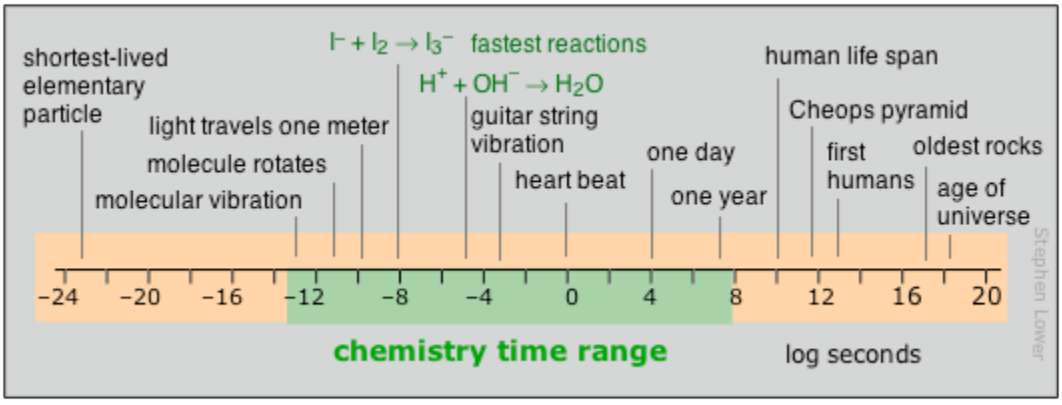
Chemists tend to work mostly in the moderately-small part of the distance range. Those who live in the lilliputian world of crystal- and molecular structures and atomic radii find the *picometer* a convenient currency, but one still sees the older non-SI unit called the *Ångstrom* used in this context;

$$1\,\text{Å} = 10^{-10}\,\text{m} = 100\,\text{pm}. \tag{2.3.4}$$

Nanotechnology, the rage of the present era, also resides in this realm. The largest polymeric molecules and colloids define the top end of the particulate range; beyond that, in the normal world of doing things in the lab, the *centimeter* and occasionally the *millimeter* commonly rule.



For humans, time moves by the heartbeat; beyond that, it is the motions of our planet that count out the hours, days, and years that eventually define our lifetimes. Beyond the few thousands of years of history behind us, those years-to-the-powers-of-tens that are the fare for such fields as evolutionary biology, geology, and cosmology, cease to convey any real meaning to us. Perhaps this is why so many people are not very inclined to accept the validity of these sciences.



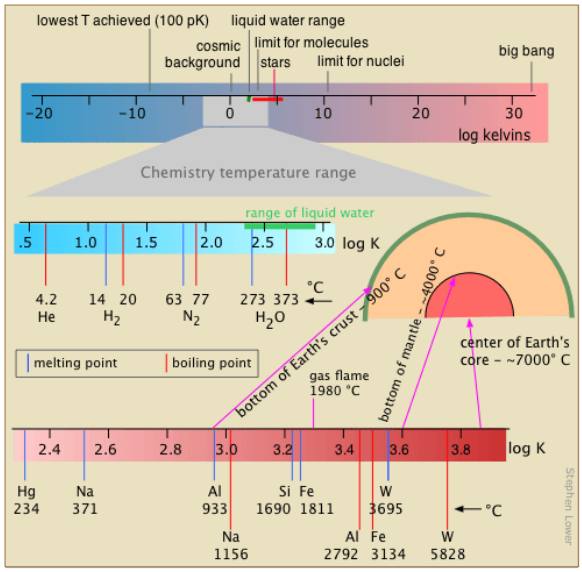
Most of what actually takes place in the chemist's test tube operates on a far shorter time scale, although there is no limit to how *slow* a reaction can be; the upper limits of those we can directly study in the lab are in part determined by how long a graduate student can wait around before moving on to gainful employment.

Looking at the microscopic world of atoms and molecules themselves, the time scale again shifts us into an unreal world where numbers tend to lose their meaning. You can gain some appreciation of the duration of a nanosecond by noting that this is about how long it takes a beam of light to travel between your two outstretched hands. In a sense, the material foundations of chemistry itself are defined by time: neither a new element nor a molecule can be recognized as such unless it lasts around sufficiently long enough to have its "picture" taken through measurement of its distinguishing properties.

Range of Temperatures in Chemistry

Temperature, the measure of thermal intensity, spans the narrowest range of any of the base units of the chemist's measure. The

reason for this is tied into temperature's meaning as an indicator of the intensity of thermal kinetic energy. Chemical change occurs when atoms are jostled into new arrangements, and the increasing weakness of these motions brings most chemistry to a halt as absolute zero is approached. At the upper end of the scale, thermal motions become sufficiently vigorous to shake molecules into atoms, and eventually, as in stars, strip off the electrons, leaving an essentially reaction-less gaseous fluid, or plasma, of bare nuclei (ions) and electrons.



We all know that temperature is expressed in degrees. What we frequently forget is that the degree is really an *increment* of temperature, a fixed fraction of the distance between two defined reference points on a *temperature scale*.

Range of Pressures in Chemistry

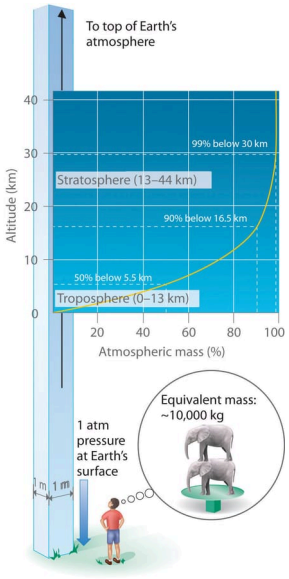
Pressure is the measure of the *force* exerted on a unit area of surface. Its SI units are therefore newtons per square meter, but we make such frequent use of pressure that a derived SI unit, the *pascal*, is commonly used:

$$1\,\mathrm{Pa} = 1\,\mathrm{N\,m^{-2}} \tag{2.3.5}$$

Pressure of the Atmosphere

The concept of pressure first developed in connection with studies relating to the atmosphere and vacuum that were first carried out in the 17th century. The molecules of a gas are in a state of constant thermal motion, moving in straight lines until experiencing a collision that exchanges momentum between pairs of molecules and sends them bouncing off in other directions.

This leads to a completely random distribution of the molecular velocities both in speed and direction— or it would in the absence of the Earth's gravitational field which exerts a tiny downward force on each molecule, giving motions in that direction a very slight advantage. In an ordinary container this effect is too small to be noticeable, but in a very tall column of air the effect adds up: the molecules in each vertical layer experience more downward-directed hits from those above it. The resulting force is quickly randomized, resulting in an increased pressure in that layer which is then propagated downward into the layers below.



At sea level, the total mass of the sea of air pressing down on each 1 m^2 of surface is about 1034 g, or 10340 kg m^{-2}. The force (weight) that the Earth's gravitational acceleration g exerts on this mass is:

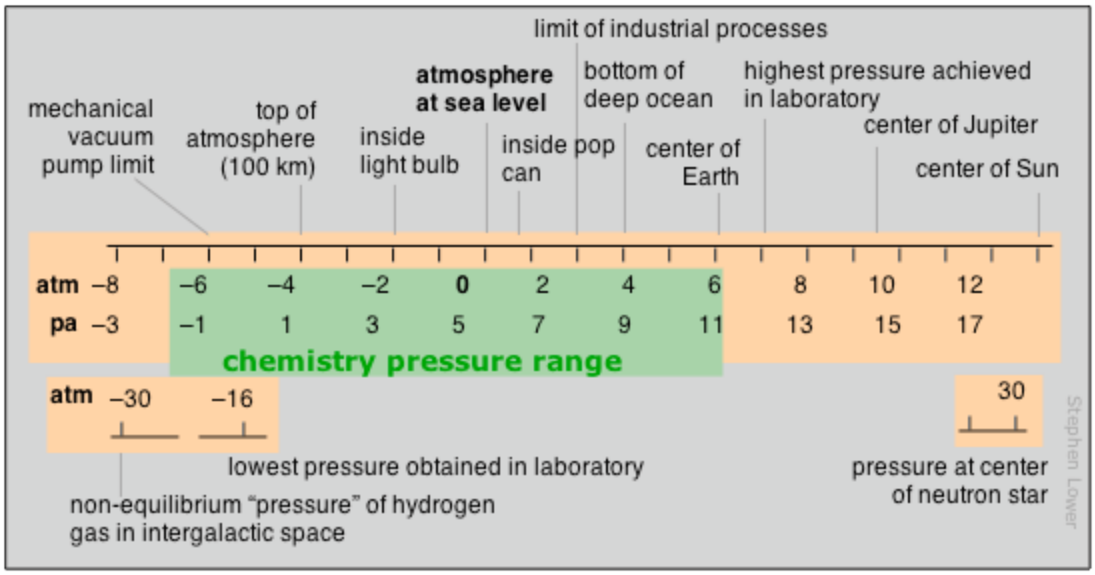
$$\begin{aligned} f &= ma \\ &= mg \\ &= (10340\ \text{kg})(9.81\ \text{m s}^{-2}) \\ &= 1.013 \times 10^5\ \text{kg m s}^{-2} \\ &= 1.013 \times 10^5\ \text{newtons} \end{aligned} \tag{2.3.6}$$

resulting in a pressure of

$$1.013 \times 10^5\,\text{N m}^{-2} = 1.013 \times 10^5\ \text{Pa}. \tag{2.3.7}$$

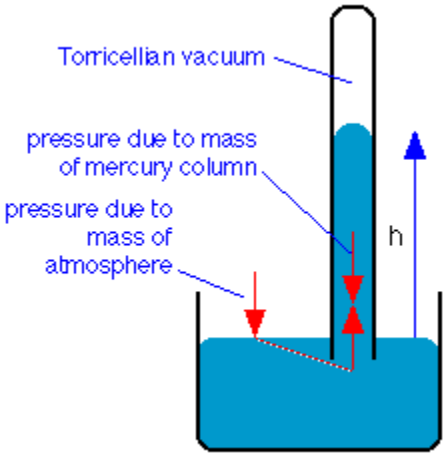
The actual pressure at sea level varies with atmospheric conditions, so it is customary to define standard atmospheric pressure as 1 atm = 1.013 × 10^5 Pa or 101 kPa.

Although the *standard atmosphere* (atm) is not an SI unit, it is still widely employed. In meteorology, the *bar*, exactly 1.000 × 10^5 = 0.967 atm, is often used.



The Barometer

In the early 17th century, the Italian physicist and mathematician Evangalisto Torricelli invented a device to measure atmospheric pressure. The Torricellian *barometer* consists of a vertical glass tube closed at the top and open at the bottom. It is filled with a liquid, traditionally mercury, and is then inverted, with its open end immersed in the container of the same liquid. The liquid level in the tube will fall under its own weight until the downward force is balanced by the vertical force transmitted hydrostatically to the column by the downward force of the atmosphere acting on the liquid surface in the open container. Torricelli was also the first to recognize that the space above the mercury constituted a vacuum, and is credited with being the first to create a vacuum.



One *standard atmosphere* will support a column of mercury that is 76 cm high, so the "millimeter of mercury", now more commonly known as the *torr*, has long been a common pressure unit in the sciences: 1 atm = 760 torr.

Contributors and Attributions

This page titled 2.3: The Measure of Matter is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

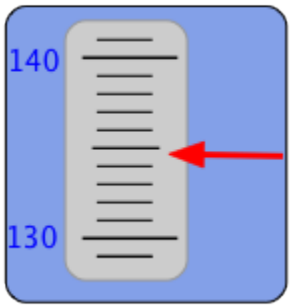
2.4: The Meaning of Measure

Learning Objectives

- Give an example of a measured numerical value, and explain what distinguishes it from a "pure" number.
- Give examples of random and systematic errors in measurements.
- Find the mean value of a series of similar measurements.
- State the principal factors that affect the difference between the mean value of a series of measurements, and the "true value" of the quantity being measured.
- Calculate the absolute and relative precisions of a given measurement, and explain why the latter is generally more useful.
- Distinguish between the accuracy and the precision of a measured value, and on the roles of random and systematic error.

In science, there are numbers and there are "numbers". What we ordinarily think of as a "number" and will refer to here as a *pure number* is just that: an expression of a precise value. The first of these you ever learned were the counting numbers, or *integers*; later on, you were introduced to the *decimal numbers*, and the *rational numbers*, which include numbers such as 1/3 and π (pi) that cannot be expressed as exact decimal values. The other kind of numeric quantity that we encounter in the natural sciences is a *measured value* of something– the length or weight of an object, the volume of a fluid, or perhaps the reading on an instrument. Although we express these values numerically, it would be a mistake to regard them as the kind of pure numbers described above.

Confusing? Suppose our instrument has an indicator such as you see here. The pointer moves up and down so as to display the measured value on this scale. What number would you write in your notebook when recording this measurement? Clearly, the value is somewhere between 130 and 140 on the scale, but the graduations enable us to be more exact and place the value between 134 and 135. The indicator points more closely to the latter value, and we can go one more step by estimating the value as perhaps 134.8, so this is the value you would report for this measurement.



Now here's the important thing to understand: although "134.8" is itself a *number*, the quantity we are measuring is almost certainly *not* 134.8 — at least, not exactly. The reason is obvious if you note that the instrument scale is such that we are barely able to distinguish between 134.7, 134.8, and 134.9. In reporting the value 134.8 we are effectively saying that the value is probably somewhere with the range 134.75 to 134.85. In other words, there is an uncertainty of ±0.05 unit in our measurement.

All measurements of quantities that can assume a continuous range of values (lengths, masses, volumes, etc.) consist of **two** parts: the reported value itself (never an exactly known number), and the uncertainty associated with the measurement. By "error", we do not mean just outright mistakes, such as incorrect use of an instrument or failure to read a scale properly; although such gross errors do sometimes happen, they usually yield results that are sufficiently unexpected to call attention to themselves.

Scale-reading error

When you measure a volume or weight, you observe a reading on a *scale* of some kind, such as the one illustrated above. Scales, by their very nature, are limited to fixed increments of value, indicated by the division marks. The actual quantities we are measuring, in contrast, can vary continuously, so there is an inherent limitation in how finely we can discriminate between two values that fall between the marked divisions of the measuring scale. Scale-reading error is often classified as *random error* (see

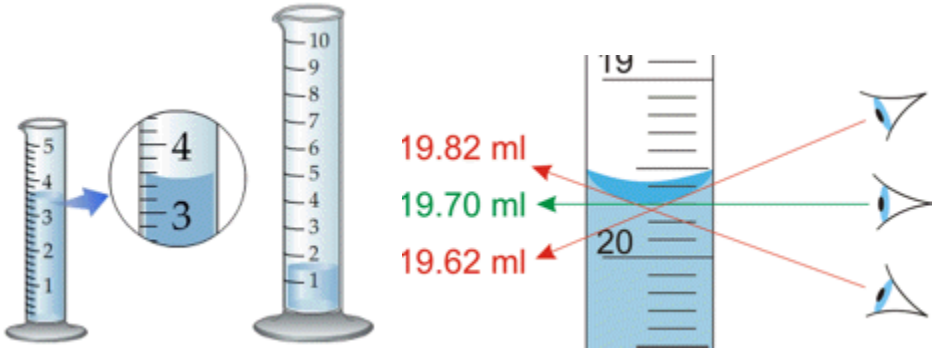
below), but it occurs so commonly that we treat it separately here.

The same problem remains if we substitute an instrument with a digital display; there will always be a point at which some value that lies between the two smallest divisions must arbitrarily toggle between two numbers on the readout display. This introduces an element of randomness into the value we observe, even if the "true" value remains unchanged. The more sensitive the measuring instrument, the less likely it is that two successive measurements of the same sample will yield identical results. In the example we discussed above, distinguishing between the values 134.8 and 134.9 may be too difficult to do in a consistent way, so two independent observers may record different values even when viewing the same reading.

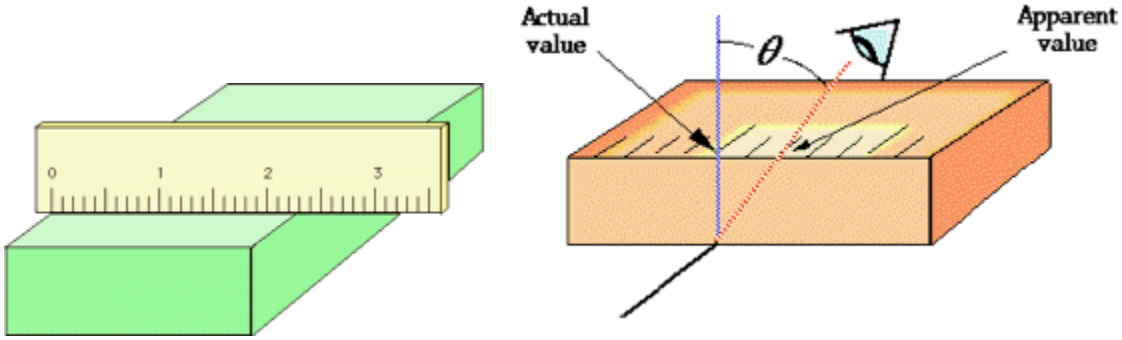
Parallax error

One form of scale-reading error that often afflicts beginners in the science laboratory is failure to properly align the eye with the part of the scale you are reading. This gives rise to *parallax error*. *Parallax* refers to the change in the apparent position of an object when viewed from different points.

The most notorious example encountered in the introductory chemistry laboratory is failure to read the volume of a liquid properly in a graduated cylinder or burette. Getting all of their students trained to make sure their eye is level with the bottom of the meniscus is the lab instructors' hope and despair.



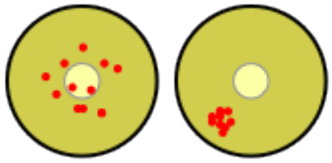
Proper use of a measuring device can help reduce the possibility of parallax error. For example, a length scale should be in direct contact with the object (*left*), not above it as on the right.



Analog meters (those having pointer needles) are most accurate when read at about 2/3 of the length of the scale. Analog-type meters, unlike those having digital readouts, are also subject to parallax error. Those intended for high-accuracy applications often have a mirrored arc along the scale in which a reflection of the pointer needle can be seen if the viewer is not properly aligned with the instrument.



- **Random (indeterminate) error:** Each measurement is also influenced by a myriad of minor events, such as building vibrations, electrical fluctuations, motions of the air, and friction in any moving parts of the instrument. These tiny influences constitute a kind of "noise" that also has a random character. Whether we are conscious of it or not, all measured values contain an element of random error.
- **Systematic error**: Suppose that you weigh yourself on a bathroom scale, not noticing that the dial reads “1.5 kg” even before you have placed your weight on it. Similarly, you might use an old ruler with a worn-down end to measure the length of a piece of wood. In both of these examples, all subsequent measurements, either of the same object or of different ones, will be off by a constant amount. Unlike random error, which is impossible to eliminate, these *systematic error* (also known as *determinate error*) is usually quite easy to avoid or compensate for, but only by a conscious effort in the conduct of the observation, usually by proper zeroing and calibration of the measuring instrument. However, once systematic error has found its way into the data, it is can be very hard to detect.



Accuracy and precision

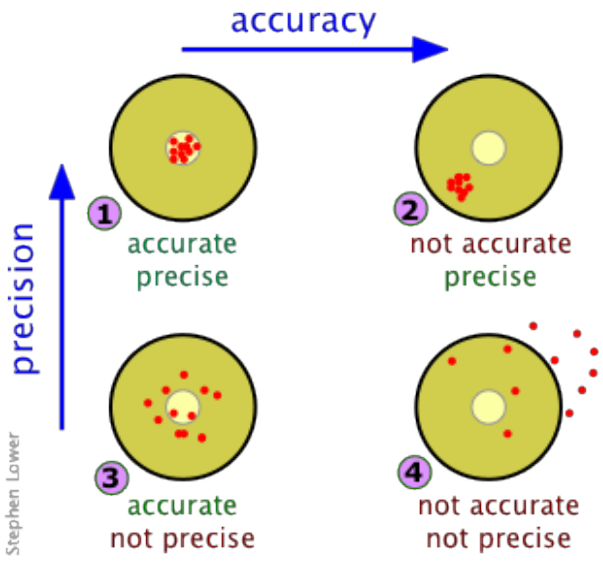
We tend to use these two terms interchangeably in our ordinary conversation, but in the context of scientific measurement, they have very different meanings:

- ***Accuracy*** refers to how closely the measured value of a quantity corresponds to its “true” value.
- ***Precision*** expresses the degree of reproducibility, or agreement between repeated measurements.

Accuracy, of course, is the goal we strive for in scientific measurements. Unfortunately, however, there is no obvious way of knowing how closely we have achieved it; the “true” value, whether it be of a well-defined quantity such as the mass of a particular object, or an average that pertains to a collection of objects, can never be known — and thus we can never recognize it if we are fortunate enough to find it.

Four Scenarios

A target on a dart board serves as a convenient analogy. The results of four sets of measurements (or four dart games) are illustrated below. Each set is made up of ten observations (or throws of darts.) Each red dot corresponds to the point at which a dart has hit the target — or alternatively, to the value of an individual observation. For measurements, assume the true value of the quantity being measured lies at the center of each target. Now consider the following four sets of results:



1 Right on! You win the dart game, and get an A grade on your measurement results.

2 Your results are beautifully replicable, but your measuring device may not have been calibrated properly or your observations suffer from a systematic error of some kind. Accuracy: F, Precision, A; overall grade C.

3 Extremely unlikely, and probably due to pure luck; the only reason for the accurate mean is that your misses mostly canceled out. Grade D.

Pretty sad; consider switching to music or politics — or have your eyes examined.

When we make real measurements, there is no dart board or target that enables one to immediately judge the quality of the result. If we make only a few observations, we may be unable distinguish between any of these scenarios.

The "true value" of a desired measurement can be quite elusive, and may not even be definable at all. This is a very common difficulty in both the social sciences (as in opinion surveys), in medicine (evaluating the efficacy of a drug or other treatment), and in all other natural sciences. The proper treatment of such problems is to make multiple observations of individual instances of what is being measured, and then use statistical methods to evaluate the results. In this introductory unit on measurement, we will defer discussion of concepts such as standard deviation and confidence intervals which become essential in courses at the second-year level and beyond. We will restrict our treatment here to the elementary considerations that are likely to be needed in a typical first-year course.

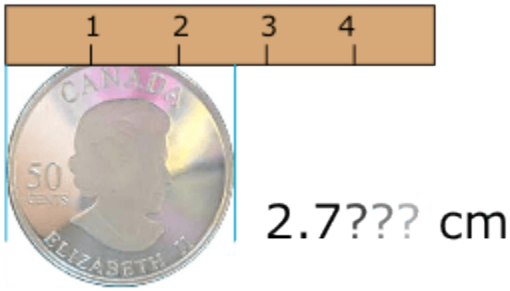
How many measurements do I need?

One measurement may be enough. If you wish to measure your height to the nearest centimeter or inch, or the volume of a liquid cooking ingredient to the nearest 1/8 “cup”, you don't ordinarily worry about random error. The error will still be present, but its magnitude will be such a small fraction of the value that it will not significantly affect whatever we are trying to achieve. Thus random error is not something we are concerned about in our daily lives. In the scientific laboratory, there are many contexts in which a single observation of a volume, mass, or instrument reading makes perfect sense; part of the "art" of science lies in making an informed judgment of how exact a given measurement must be. If we are measuring a directly observable quantity such as the weight of a solid or volume of a liquid, then a single measurement, carefully done and reported to a precision that is consistent with that of the measuring instrument, will usually be sufficient.

However more measurements are needed when there is no clearly-defined "true" value. A collection of objects (or of people) is known in statistics as a *population*. There is often a need to determine some quantity that describes a collection of objects. For

example, a pharmaceutical researcher will need to determine the time required for half of a standard dose of a certain drug to be eliminated by the body, or a manufacturer of light bulbs might want to know how many hours a certain type of light bulb will operate before it burns out. In these cases a value for any individual sample can be determined easily enough, but since no two samples (patients or light bulbs) are identical, we are compelled to repeat the same measurement on multiple objects. And naturally, we get a variety of results, usually referred to as *scatter*. Even for a single object, there may be no clearly defined "true" value.

Suppose that you wish to determine the diameter of a certain type of coin. You make one measurement and record the results. If you then make a similar measurement along a different cross-section of the coin, you will likely get a different result. The same thing will happen if you make successive measurements on other coins of the same kind.



Here we are faced with two kinds of problems. First, there is the inherent limitation of the measuring device: we can never reliably measure more finely than the marked divisions on the ruler. Secondly, we cannot assume that the coin is perfectly circular; careful inspection will likely reveal some distortion resulting from a slight imperfection in the manufacturing process. In these cases, it turns out that there is no single, true value of the quantity we are trying to measure.

Mean, median, and range of a series of observations

There are a variety of ways to express the average, or central tendency of a series of measurements, with mean (more precisely, arithmetic mean) being most commonly employed. Our ordinary use of the term "average" also refers to the mean. These concepts are usually all you need as a first step in the analysis of data you are likely to collect in a first-year chemistry laboratory course.

The mean and its meaning

In our ordinary speech, the term "average" is synonymous with "mean". In statistics, however, "average" is a more general term that can refer to median, mode, and range, as well as to mean. When we obtain more than one result for a given measurement (either made repeatedly on a single sample, or more commonly, on different samples of the same material), the simplest procedure is to report the *mean*, or *average* value. The *mean* is defined mathematically as the sum of the values, divided by the number of measurements:

$$x_m = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.4.1}$$

If you are not familiar with this notation, don't let it scare you! It's no different from the *average* that you are likely already familiar with. Take a moment to see how it expresses the previous sentence; if there are n measurements, each yielding a value xi , then we sum over all i and divide by n to get the mean value x_m. For example, if there are only two measurements, x_1 and x_1, then the mean is

$$x_m = \frac{x_1 + x_2}{2} \tag{2.4.2}$$

The general problem of determining the uncertainty of a calculated result turns out to be rather more complicated than you might think, and will not be treated here. There are, however, some very simple rules that are sufficient for most practical purposes.

Absolute and Relative Uncertainty

If you weigh out 74.1 mg of a solid sample on a laboratory balance that is accurate to within 0.1 milligram, then the actual weight

of the sample is likely to fall somewhere in the range of 74.0 to 74.2 mg; the absolute uncertainty in the weight you observe is 0.2 mg, or ±0.1 mg. If you use the same balance to weigh out 3.2914 g of another sample, the actual weight is between 3.2913 g and 3.2915 g, and the *absolute uncertainty* is still ±0.1 mg.

Although the absolute uncertainties in these two examples are identical, we would probably consider the second measurement to be more precise because the uncertainty is a smaller fraction of the measured value. The *relative uncertainties* of the two results would be

$$0.2 \div 74.1 = 0.0027 \text{ (about 3 parts in 1000 (PPT), or 0.3\%)}$$

$$0.0002 \div 3.2913 = 0.000084 \text{ (about 0.8 PPT , or 0.008 \%)}$$

Relative uncertainties are widely used to express the reliability of measurements, even those for a single observation, in which case the uncertainty is that of the measuring device. Relative uncertainties can be expressed as parts per hundred (*percent*), per thousand (PPT), per million, (PPM), and so on.

Questions

1. Addition and subtraction, both numbers have uncertainties

The simplest method is to just add the absolute uncertainties.

Example: (6.3 ± 0.05 cm) – (2.1 ± 0.05 cm) = 4.2 ± 0.10 cm
However, this tends to over-estimate the uncertainty by assuming the worst possible case in which the error in one of the quantities is at its maximum positive value, while that of the other quantity is at its maximum minimum value.

Statistical theory informs us that a more realistic value for the uncertainty of a sum or difference is to add the s*quares* of each absolute uncertainty, and then take the square root of this sum. Applying this to the above values, we have

$[(.05)^2 + (.05)^2]^{½} = 0.07$, so the result is 4.2 ± 0.07 cm.

2. Multiplication or division, both numbers have uncertainties.

Convert the absolute uncertainties into *relative* uncertainties, and add these. Or better, add their squares and take the square root of the sum.

Problem Example 3

Estimate the absolute error in the density calculated by dividing (12.7 ± .05 g) by (10.0 ± 0.02 mL).

Solution: Relative uncertainty of the mass: 0.05 / 12.7 = 0.0039 = 0.39%
Relative uncertainty of the volume: 0.02 / 10.0 = 0.002 = 0.2%
Relative uncertainty of the density: $[(.39)^2 + (0.2)^2]^{½} = 0.44$ %
Mass ÷ volume: (12.7 g) ÷ (10.0 mL) = 1.27 g mL^{-1}
Absolute uncertainty of the density: (± 0.044) x (1.27 g mL^{-1}) = ±0.06 g mL^{-1}

3. Multiplication or division by a pure number

Trivial case; multiply or divide the uncertainty by the pure number.

This page titled 2.4: The Meaning of Measure is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

3: MEASURING MATTER

The natural sciences begin with ***observation***, and this usually involves ***numerical measurements*** of quantities such as length, volume, density, and temperature. Most of these quantities have ***units*** of some kind associated with them, and these units must be retained when you use them in calculations. All measuring units can be defined in terms of a very small number of fundamental ones that, through "dimensional analysis", provide insight into their derivation and meaning, and must be understood when converting between different unit systems.

http://www.chem1.com/acad/webtext/ma...ure/index.html

3.1: Units and Dimensions

3.2: The Meaning of Measure

3.3: Significant Figures and Rounding off

3.4: Reliability of a measurement

3.5: Drawing Conclusions from Data

Contributors and Attributions

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook
- Thumbnail: Unsplash License (William Warby via Unsplash)

This page titled 3: Measuring Matter is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

3.1: Units and Dimensions

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the italicized terms in the context of this topic.

- Describe the names and abbreviations of the *SI base units* and the *SI decimal prefixes*.
- Define the *liter* and the *metric ton* in these units.
- Explain the meaning and use of *unit dimensions*; state the dimensions of *volume*.
- State the quantities that are needed to define a *temperature scale*, and show how these apply to the *Celsius, Kelvin,* and *Fahrenheit* temperature scales.
- Explain how a Torricellian *barometer* works.

Have you ever estimated a distance by "stepping it off"— that is, by counting the number of steps required to take you a certain distance? Or perhaps you have used the width of your hand, or the distance from your elbow to a fingertip to compare two dimensions. If so, you have engaged in what is probably the first kind of measurement ever undertaken by primitive mankind. The results of a measurement are always expressed on some kind of a *scale* that is defined in terms of a particular kind of *unit*. The first scales of distance were likely related to the human body, either directly (the length of a limb) or indirectly (the distance a man could walk in a day).

As civilization developed, a wide variety of measuring scales came into existence, many for the same quantity (such as length), but adapted to particular activities or trades. Eventually, it became apparent that in order for trade and commerce to be possible, these scales had to be defined in terms of standards that would allow measures to be verified, and, when expressed in different units (bushels and pecks, for example), to be correlated or converted.

Over the centuries, hundreds of measurement units and scales have developed in the many civilizations that achieved some literate means of recording them. Some, such as those used by the Aztecs, fell out of use and were largely forgotten as these civilizations died out. Other units, such as the various systems of measurement that developed in England, achieved prominence through extension of the Empire and widespread trade; many of these were confined to specific trades or industries. The examples shown here are only some of those that have been used to measure length or distance. The history of measuring units provides a fascinating reflection on the history of industrial development.

The most influential event in the history of measurement was undoubtedly the French Revolution and the Age of Rationality that followed. This led directly to the metric system that attempted to do away with the confusing multiplicity of measurement scales by reducing them to a few fundamental ones that could be combined in order to express any kind of quantity. The metric system spread rapidly over much of the world, and eventually even to England and the rest of the U.K. when that country established closer economic ties with Europe in the latter part of the 20th Century. The United States is presently the only major country in which "metrication" has made little progress within its own society, probably because of its relative geographical isolation and its vibrant internal economy.

Science, being a truly international endeavor, adopted metric measurement very early on; engineering and related technologies have been slower to make this change, but are gradually doing so. Even the within the metric system, however, a variety of units were employed to measure the same fundamental quantity; for example, energy could be expressed within the metric system in units of ergs, electron-volts, joules, and two kinds of calories. This led, in the mid-1960s, to the adoption of a more basic set of units, the *Systeme Internationale* (*SI*) units that are now recognized as the standard for science and, increasingly, for technology of all kinds.

The SI base Units

In principle, any physical quantity can be expressed in terms of only seven *base units*. Each base unit is defined by a standard which is described in the NIST Web site.

The SI base units

length	meter	m
mass	kilogram	kg
time	second	s
temperature (absolute)	kelvin	K
amount of substance	mole	mol
electric current	ampere	A
luminous intensity	candela	cd

A few special points about some of these units are worth noting:

- The base unit of **mass** is unique in that a decimal prefix (see below) is built-in to it; that is, it is not the *gram*, as you might expect.
- The base unit of **time** is the only one that is not metric. Numerous attempts to make it so have never garnered any success; we are still stuck with the 24:60:60 system that we inherited from ancient times. (The ancient Egyptians of around 1500 BC invented the 12-hour day, and the 60:60 part is a remnant of the base-60 system that the Sumerians used for their astronomical calculations around 100 BCE.)
- Of special interest to Chemistry is the *mole*, the base unit for expressing the **quantity of matter**. Although the number is not explicitly mentioned in the official definition, chemists define the mole as Avogadro's number (approximately 6.02×10^{23}) of anything.

The *SI* decimal prefixes

Owing to the wide range of values that quantities can have, it has long been the practice to employ prefixes such as milli and mega to indicate decimal fractions and multiples of metric units. As part of the SI standard, this system has been extended and formalized.

prefix	abbreviation	multiplier	prefix	abbreviation	multiplier
exa	E	10^{18}	deci	d	10^{-1}
peta	P	10^{15}	centi	c	10^{-2}
tera	T	10^{12}	milli	m	10^{-3}
giga	G	10^{9}	micro	μ	10^{-6}
mega	M	10^{6}	nano	n	10^{-9}
kilo	k	10^{3}	pico	p	10^{-12}
hecto	h	10^{2}	femto	f	10^{-15}
deca	da	10	atto	a	10^{-18}

For a more complete table, see the NIST page on SI prefixes

Non-SI Units

There is a category of units that are "honorary" members of the SI in the sense that it is acceptable to use them along with the base units defined above. These include such mundane units as the hour, minute, and degree (of angle), etc., but the three shown here are of particular interest to chemistry, and you will need to know them.

- **liter** (L)

$$1\,L = 1\,dm^3 = 10^{-3}m^3$$

- **metric ton** (t)

$$1\,t = 10^3 kg$$

- **atomic mass unit** (u)

$$1\,u = 1.66054 \times 10^{-27}\,kg$$

Most of the physical quantities we actually deal with in science and also in our daily lives, have units of their own: volume, pressure, energy and electrical resistance are only a few of hundreds of possible examples. It is important to understand, however, that all of these can be expressed in terms of the SI base units; they are consequently known as *derived units*.

In fact, most physical quantities can be expressed in terms of one or more of the following five fundamental units:

mass ***M***	length ***L***	time ***T***	electric charge **Q**	temperature **Θ** (*theta*)

Consider, for example, the unit of ***volume***, which we denote as V. To measure the volume of a rectangular box, we need to multiply the lengths as measured along the three coordinates:

$$V = x \cdot y \cdot z \tag{3.1.1}$$

We say, therefore, that volume has the dimensions of length-cubed:

$$dim.V = L^3 \tag{3.1.2}$$

Thus the units of volume will be m^3 (in the SI) or cm^3, ft^3 (English), etc. Moreover, any formula that calculates a volume must contain within it the L^3 dimension; thus the volume of a sphere is $4/3\ \pi r^3$.

Consider, for example, the unit of volume, which we denote as V. To measure the volume of a rectangular box, we need to multiply the lengths as measured along the three coordinates: V = x · y · z We say, therefore, that volume has the dimensions of length-cubed: dim.V = L^3 Thus the units of volume will be m^3 (in the SI) or cm^3, ft^3 (English), etc. Moreover, any formula that calculates a volume must contain within it the L^3 dimension; thus the volume of a sphere is $4/3\ \pi r^3$.

Example 3.1.1: Energy Units

Find the dimensions of energy.

Solution

When mechanical work is performed on a body, its energy increases by the amount of work done, so the two quantities are equivalent and we can concentrate on work. The latter is the product of the force applied to the object and the distance it is displaced. From Newton's law, force is the product of mass and acceleration, and the latter is the rate of change of velocity, typically expressed in meters per second per second. Combining these quantities and their dimensions yields the result shown here.

Units and Their Ranges in Chemistry

In this section, we will look at some of the quantities that are widely encountered in Chemistry, and at the units in which they are commonly expressed. In doing so, we will also consider the actual range of values these quantities can assume, both in nature in general, and also within the subset of nature that chemistry normally addresses. In looking over the various units of measure, it is interesting to note that their unit values are set close to those encountered in everyday human experience

Mass and Weight

These two quantities are widely confused. Although they are often used synonymously in informal speech and writing, they have

different dimensions: *weight* is the *force* exerted on a mass by the local gravational field:

$$f = ma = mg \tag{3.1.3}$$

where *g* is the acceleration of gravity. While the nominal value of the latter quantity is 9.80 m s^{-2} at the Earth's surface, its exact value varies locally. Because it is a force, the SI unit of weight is properly the *newton*, but it is common practice (except in physics classes!) to use the terms "weight" and "mass" interchangeably, so the units *kilograms* and *grams* are acceptable in almost all ordinary laboratory contexts.

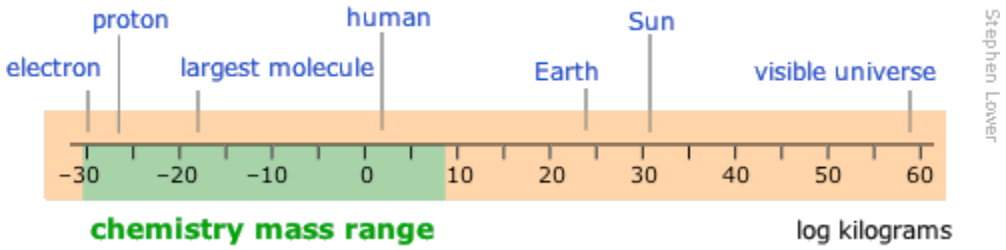
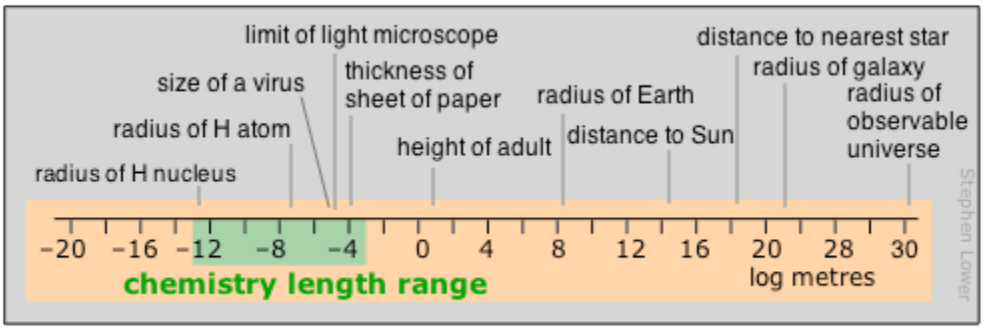


Figure 1: Please note that in this diagram and in those that follow, the numeric scale represents the logarithm of the number shown. For example, the mass of the electron is 10^{-30} kg.

The range of masses spans 90 orders of magnitude, more than any other unit. The range that chemistry ordinarily deals with has greatly expanded since the days when a microgram was an almost inconceivably small amount of material to handle in the laboratory; this lower limit has now fallen to the atomic level with the development of tools for directly manipulating these particles. The upper level reflects the largest masses that are handled in industrial operations, but in the recently developed fields of geochemistry and environmental chemistry, the range can be extended indefinitely. Flows of elements between the various regions of the environment (atmosphere to oceans, for example) are often quoted in teragrams.

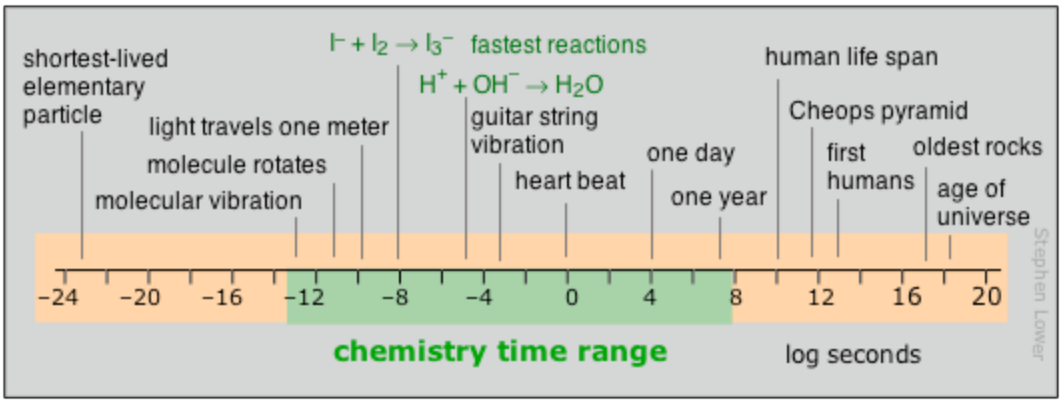
Length

Chemists tend to work mostly in the moderately-small part of the distance range. Those who live in the lilliputian world of crystal- and molecular structures and atomic radii find the *picometer* a convenient currency, but one still sees the older non-SI unit called the *Ångstrom* used in this context; 1Å = 10^{-10} m = 100pm. Nanotechnology, the rage of the present era, also resides in this realm. The largest polymeric molecules and colloids define the top end of the particulate range; beyond that, in the normal world of doing things in the lab, the *centimeter* and occasionally the *millimeter* commonly rule.



Time

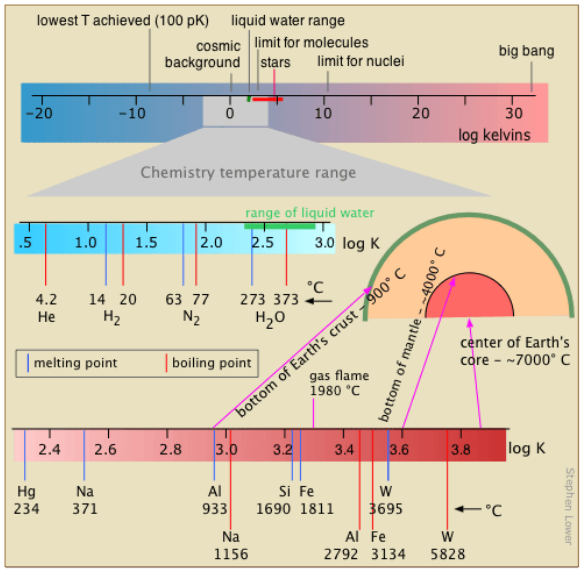
For humans, time moves by the heartbeat; beyond that, it is the motions of our planet that count out the hours, days, and years that eventually define our lifetimes. Beyond the few thousands of years of history behind us, those years-to-the-powers-of-tens that are the fare for such fields as evolutionary biology, geology, and cosmology, cease to convey any real meaning for us. Perhaps this is why so many people are not very inclined to accept their validity.



Most of what actually takes place in the chemist's test tube operates on a far shorter time scale, although there is no limit to how slow a reaction can be; the upper limits of those we can directly study in the lab are in part determined by how long a graduate student can wait around before moving on to gainful employment. Looking at the microscopic world of atoms and molecules themselves, the time scale again shifts us into an unreal world where numbers tend to lose their meaning. You can gain some appreciation of the duration of a nanosecond by noting that this is about how long it takes a beam of light to travel between your two outstretched hands. In a sense, the material foundations of chemistry itself are defined by time: neither a new element nor a molecule can be recognized as such unless it lasts around sufficiently long enough to have its "picture" taken through measurement of its distinguishing properties.

Temperature

Temperature, the measure of thermal intensity, spans the narrowest range of any of the base units of the chemist's measure. The reason for this is tied into temperature's meaning as a measure of the intensity of thermal kinetic energy. Chemical change occurs when atoms are jostled into new arrangements, and the weakness of these motions brings most chemistry to a halt as absolute zero is approached. At the upper end of the scale, thermal motions become sufficiently vigorous to shake molecules into atoms, and eventually, as in stars, strip off the electrons, leaving an essentially reaction-less gaseous fluid, or plasma, of bare nuclei (ions) and electrons.



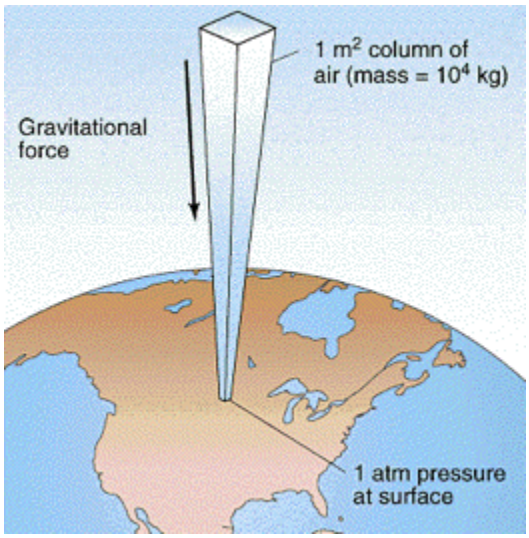
We all know that temperature is expressed in degrees. What we frequently forget is that the degree is really an *increment* of temperature, a fixed fraction of the distance between two defined reference points on a *temperature scale*.

Pressure

Pressure is the measure of the *force* exerted on a unit area of surface. Its SI units are therefore newtons per square meter, but we make such frequent use of pressure that a derived SI unit, the *pascal*, is commonly used:

$$1\ \text{Pa} = 1\ \text{N m}^{-2}$$

The concept of pressure first developed in connection with studies relating to the atmosphere and vacuum that were first carried out in the 17th century. The molecules of a gas are in a state of constant thermal motion, moving in straight lines until experiencing a collision that exchanges momentum between pairs of molecules and sends them bouncing off in other directions. This leads to a completely random distribution of the molecular velocities both in speed and direction— or it would in the absence of the Earth's gravitational field which exerts a tiny downward force on each molecule, giving motions in that direction a very slight advantage. In an ordinary container this effect is too small to be noticeable, but in a very tall column of air the effect adds up: the molecules in each vertical layer experience more downward-directed hits from those above it. The resulting force is quickly randomized, resulting in an increased pressure in that layer which is then propagated downward into the layers below.



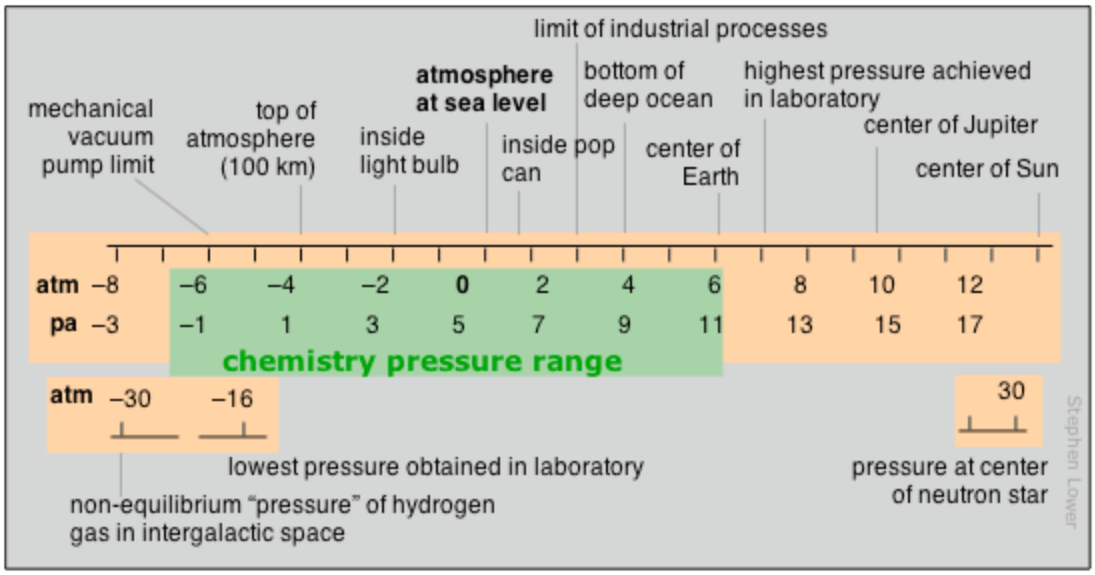
At sea level, the total mass of the sea of air pressing down on each 1-cm^2 of surface is about 1034 g, or 10340 kg m^{-2}. The force (weight) that the Earth's gravitional acceleration g exerts on this mass is

$$f = ma = mg = (10340\ \text{kg m}^{-2})(9.81\ \text{m s}^{-2}) = 1.013 \times 10^5\ \text{kg m s}^{-2} = 1.013 \times 10^5\ \text{newtons}$$

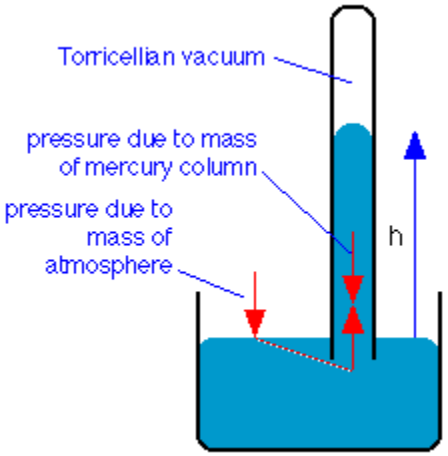
resulting in a pressure of

$$1.013 \times 10^5 N\, m^{-2} = 1.013 \times 10^5 pa. \tag{3.1.4}$$

The actual pressure at sea level varies with atmospheric conditions, so it is customary to define standard atmospheric pressure as 1 atm = 1.013 ×10^5 pa or 101 kpa. Although the standard atmosphere is not an SI unit, it is still widely employed. In meteorology, the *bar*, exactly 1.000 × 10^5 = 0.967 atm, is often used.



In the early 17th century, the Italian physicist and mathematician Evangalisto Torricelli invented a device to measure atmospheric pressure. The Torricellian *barometer* consists of a vertical glass tube closed at the top and open at the bottom. It is filled with a liquid, traditionally mercury, and is then inverted, with its open end immersed in the container of the same liquid. The liquid level in the tube will fall under its own weight until the downward force is balanced by the vertical force transmitted hydrostatically to the column by the downward force of the atmosphere acting on the liquid surface in the open container. Torricelli was also the first to recognize that the space above the mercury constituted a vacuum, and is credited with being the first to create a vacuum.



One *standard atmosphere* will support a column of mercury that is 76 cm high, so the "millimeter of mercury", now more commonly known as the *torr*, has long been a common pressure unit in the sciences: 1 atm = 760 torr.

This page titled 3.1: Units and Dimensions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

3.2: The Meaning of Measure

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Give an example of a measured numerical value, and explain what distinguishes it from a "pure" number.
- Give examples of random and systematic errors in measurements.
- Find the mean value of a series of similar measurements.
- State the principal factors that affect the difference between the mean value of a series of measurements, and the "true value" of the quantity being measured.
- Calculate the absolute and relative precisions of a given measurement, and explain why the latter is generally more useful.
- Distinguish between the accuracy and the precision of a measured value, and on the roles of random and systematic error.

The exact distance between the upper lip and the tip of the dorsal fin will forever be hidden in a fog of uncertainty. The angle at which we hold the calipers and the force with which we close them on the object will never be exactly reproducible. A more fundamental limitation occurs whenever we try to compare a continuously-varying quantity such as distance with the fixed intervals on a measuring scale; between 59 and 60 mils there is the same infinity of distances that exists between 59 and 60 miles!

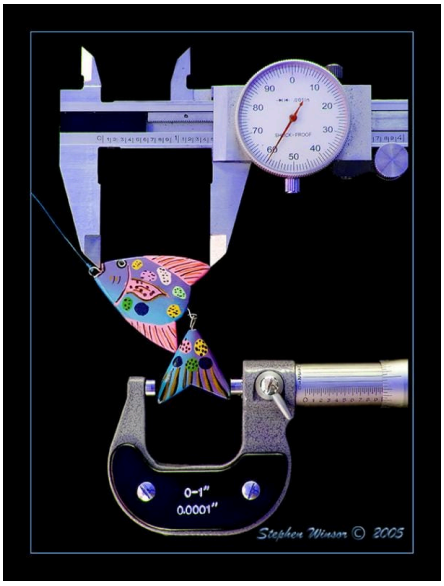


Image by Stephen Winsor; used with permission of the artist.

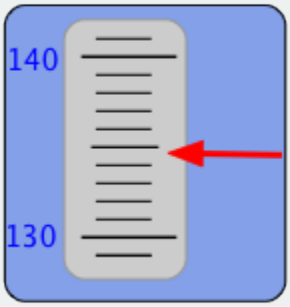
The "true value" of a measured quantity, if it exists at all, will always elude us; the best we can do is learn how to make meaningful use (and to avoid mis-use!) of the numbers we read off of our measuring devices.

Uncertainty is Certain!

In science, there are numbers and there are "numbers". What we ordinarily think of as a "number" and will refer to here as a *pure number* is just that: an expression of a precise value. The first of these you ever learned were the counting numbers, or *integers*; later on, you were introduced to the *decimal numbers*, and the *rational numbers*, which include numbers such as 1/3 and π (pi) that cannot be expressed as exact decimal values.

The other kind of numeric quantity that we encounter in the natural sciences is a *measured value* of something– the length or weight of an object, the volume of a fluid, or perhaps the reading on an instrument. Although we express these values numerically, it would be a mistake to regard them as the kind of pure numbers described above.

Confusing? Suppose our instrument has an indicator such as you see here. The pointer moves up and down so as to display the measured value on this scale. What number would you write in your notebook when recording this measurement? Clearly, the value is somewhere between 130 and 140 on the scale, but the graduations enable us to be more exact and place the value between 134 and 135. The indicator points more closely to the latter value, and we can go one more step by estimating the value as perhaps 134.8, so this is the value you would report for this measurement.



Now here's the important thing to understand: although "134.8" is itself a *number*, the quantity we are measuring is almost certainly *not* 134.8— at least, not exactly. The reason is obvious if you note that the instrument scale is such that we are barely able to distinguish between 134.7, 134.8, and 134.9. In reporting the value 134.8 we are effectively saying that the value is probably somewhere with the range 134.75 to 134.85. In other words, there is an uncertainty of ±0.05 unit in our measurement.

All measurements of quantities that can assume a continuous range of values (lengths, masses, volumes, etc.) consist of **two parts**: the reported value itself (never an exactly known number), and the uncertainty associated with the measurement. By "error", we do not mean just outright mistakes, such as incorrect use of an instrument or failure to read a scale properly; although such gross errors do sometimes happen, they usually yield results that are sufficiently unexpected to call attention to themselves.

Errors in Measurements

All measurements are subject to error which contributes to the uncertainty of the result. By "error", we do not mean just outright mistakes, such as incorrect use of an instrument or failure to read a scale properly; although such gross errors do sometimes happen, they usually yield results that are sufficiently unexpected to call attention to themselves.

When you measure a volume or weight, you observe a reading on a *scale* of some kind. Scales, by their very nature, are limited to fixed increments of value, indicated by the division marks. The actual quantities we are measuring, in contrast, can vary continuously, so there is an inherent limitation in how finely we can discriminate between two values that fall between the marked divisions of the measuring scale. The same problem remains if we substitute an instrument with a digital display; there will always be some point at which some value that lies between the two smallest divisions must arbitrarily toggle between two numbers on the readout display. This introduces an element of randomness into the value we observe, even if the "true" value remains unchanged.

The more sensitive the measuring instrument, the less likely it is that two successive measurements of the same sample will yield identical results. In the example we discussed above, distinguishing between the values 134.8 and 134.9 may be too difficult to do in a consistent way, so two independent observers may record different values even when viewing the same reading. Each measurement is also influenced by a myriad of minor events, such as building vibrations, electrical fluctuations, motions of the air, and friction in any moving parts of the instrument. These tiny influences constitute a kind of "noise" that also has a random character. Whether we are conscious of it or not, all measured values contain an element of random error.

Each measurement is also influenced by a myriad of minor events, such as building vibrations, electrical fluctuations, motions of the air, and friction in any moving parts of the instrument. These tiny influences constitute a kind of "noise" that also has a

random character. Whether we are conscious of it or not, all measured values contain an element of random error.

Suppose that you weigh yourself on a bathroom scale, not noticing that the dial reads “1.5 kg” even before you have placed your weight on it. Similarly, you might use an old ruler with a worn-down end to measure the length of a piece of wood. In both of these examples, all subsequent measurements, either of the same object or of different ones, will be off by a constant amount. Unlike random error, which is impossible to eliminate, these *systematic errors* are usually quite easy to avoid or compensate for, but only by a conscious effort in the conduct of the observation, usually by proper zeroing and calibration of the measuring instrument. However, once systematic error has found its way into the data, it is can be very hard to detect.

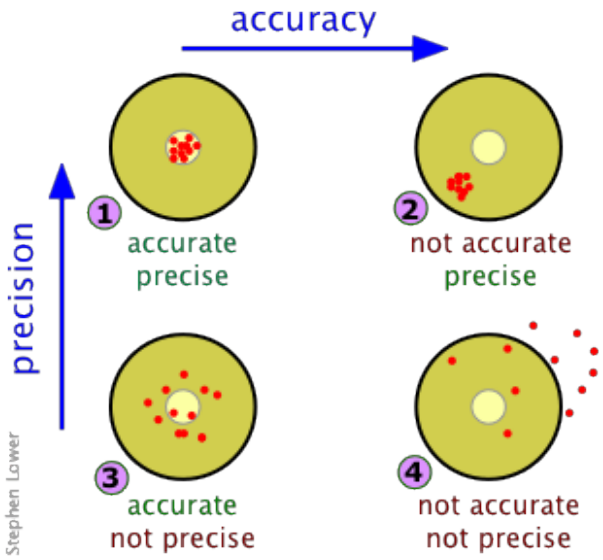
The Difference Between Accuracy and Precision

We tend to use these two terms interchangeably in our ordinary conversation, but in the context of scientific measurement, they have very different meanings:

- ***Accuracy*** refers to how closely the measured value of a quantity corresponds to its “true” value.
- ***Precision*** expresses the degree of reproducibility, or agreement between repeated measurements.

Accuracy, of course, is the goal we strive for in scientific measurements. Unfortunately, however, there is no obvious way of knowing how closely we have achieved it; the “true” value, whether it be of a well-defined quantity such as the mass of a particular object, or an average that pertains to a collection of objects, can never be known– and thus we can never recognize it if we are fortunate enough to find it.

A target on a dart board serves as a convenient analogy. The results of four sets of measurements (or four dart games) are illustrated below. Each set is made up of ten observations (or throws of darts.) Each red dot corresponds to the point at which a dart has hit the target — or alternatively, to the value of an individual observation. For measurements, assume the true value of the quantity being measured lies at the center of each target.



Now consider the following four sets of results:

1 - Right on! You win the dart game, and get an A grade on your measurement results.

2 - Your results are beautifully replicable, but your measuring device may not have been calibrated properly or your observations suffer from a systematic error of some kind. Accuracy: F, Precision, A; overall grade C.

3 - Extremely unlikely, and probably due to pure luck; the only reason for the accurate mean is that your misses mostly canceled out. Grade D.

4 - Pretty sad; consider switching to music or politics — or have your eyes examined.

Note carefully that when we make real measurements, there is no dart board or target that enables one to immediately judge the quality of the result. If we make only a few observations, we may be unable distinguish between any of these scenarios. Thus we cannot distinguish between the four scenarios illustrated above by simply examining the results of the two measurements. We can, however, judge the precision of the results, and then apply simple statistics to estimate how closely the mean value is likely to reflect the true value in the absence of systematic error.

More than one answer in Replicate Measurements

If you wish to measure your height to the nearest centimeter or inch, or the volume of a liquid cooking ingredient to the nearest “cup”, you can probably do so without having to worry about random error. The error will still be present, but its magnitude will be such a small fraction of the value that it will not be detected. Thus random error is not something we worry about too much in our daily lives.

If we are making scientific observations, however, we need to be more careful, particularly if we are trying to exploit the full sensitivity of our measuring instruments in order to achieve a result that is as reliable as possible. If we are measuring a directly observable quantity such as the weight or volume of an object, then a single measurement, carefully done and reported to a precision that is consistent with that of the measuring instrument, will usually be sufficient.

More commonly, however, we are called upon to find the value of some quantity whose determination depends on several other measured values, each of which is subject to its own sources of error. Consider a common laboratory experiment in which you must determine the percentage of acid in a sample of vinegar by observing the volume of sodium hydroxide solution required to neutralize a given volume of the vinegar. You carry out the experiment and obtain a value. Just to be on the safe side, you repeat the procedure on another identical sample from the same bottle of vinegar. If you have actually done this in the laboratory, you will know it is highly unlikely that the second trial will yield the same result as the first. In fact, if you run a number of replicate (that is, identical in every way) determinations, you will probably obtain a scatter of results.

> To understand why, consider all the individual measurements that go into each determination; the volume of the vinegar sample, your judgment of the point at which the vinegar is neutralized, and the volume of solution used to reach this point. And how accurately do you know the concentration of the sodium hydroxide solution, which was made up by dissolving a measured weight of the solid in water and then adding more water until the solution reaches some measured volume. Each of these many observations is subject to random error; because such errors are random, they can occasionally cancel out, but for most trials we will not be so lucky– hence the scatter in the results.

A similar difficulty arises when we need to determine some quantity that describes a collection of objects. For example, a pharmaceutical researcher will need to determine the time required for half of a standard dose of a certain drug to be eliminated by the body, or a manufacturer of light bulbs might want to know how many hours a certain type of light bulb will operate before it burns out. In these cases a value for any individual sample can be determined easily enough, but since no two samples (patients or light bulbs) are identical, we are compelled to repeat the same measurement on multiple samples, and once again, are faced with a scattering of results.

As a final example, suppose that you wish to determine the diameter of a certain type of coin. You make one measurement and record the results. If you then make a similar measurement along a different cross-section of the coin, you will likely get a different result. The same thing will happen if you make successive measurements on other coins of the same kind.

> Here we are faced with two kinds of problems. First, there is the inherent limitation of the measuring device: we can never reliably measure more finely than the marked divisions on the ruler. Secondly, we cannot assume that the coin is perfectly circular; careful inspection will likely reveal some distortion resulting from a slight imperfection in the manufacturing process. In these cases, it turns out that there is no single, true value of either quantity we are trying to measure.

Mean, Median, and Range of a Series of Observations

There are a variety of ways to express the average, or central tendency of a series of measurements, with mean (more precisely, arithmetic mean) being most commonly employed. Our ordinary use of the term "average" also refers to the mean. When we obtain more than one result for a given measurement (either made repeatedly on a single sample, or more commonly, on different

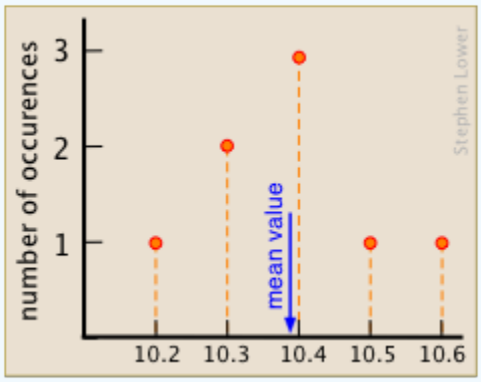
samples), the simplest procedure is to report the mean, or average value. The mean is defined mathematically as the sum of the values, divided by the number of measurements:

$$x_m = \frac{\sum_i x_i}{n} \tag{3.2.1}$$

If you are not familiar with this notation, don't let it scare you! Take a moment to see how it expresses the previous sentence; if there are n measurements, each yielding a value xI, then we sum over all i and divide by n to get the mean value x_m. For example, if there are only two measurements, x_1 and x_1, then the mean is $(x_1 + x_2)/2$.

Example 3.2.1

Calculate the mean value of the set of eight measurements illustrated here.



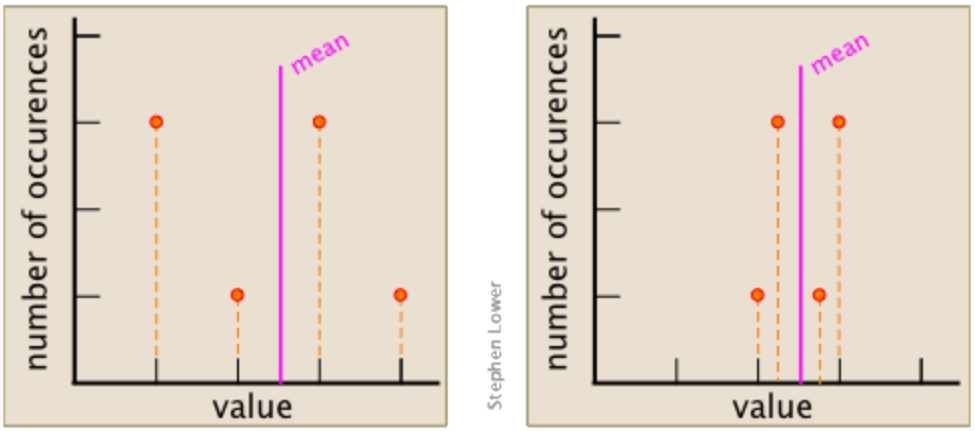
Solution

There are eight data points (10.4 was found in three trials, 10.5 in two), so $n = 8$. The mean is (via Equation 3.2.1):

$$\frac{10.2 + 10.3 + (3x10.4) + 10.5 + 10.5 + 10.8}{8} = 10.4.$$

Range

The range of a data set is the difference between its smallest and largest values. As such, its value reflects the precision of the result. For example, the following data sets have the same average, but the one having the smaller range is clearly more precise.



same means, different ranges

If you arrange the list of measured values in order of their magnitude, the median is the one that has as many values above it as below it.

*Example*s: for the data set [22 23 23 24 26 28] the mode would be 23.

For an odd number of values *n*, the median is the [(*n*+1)/2]th member of the set. Thus for [22 23 23 24 24 27], (*n*+1)/2 =3, so 23 is the median value.

Mode

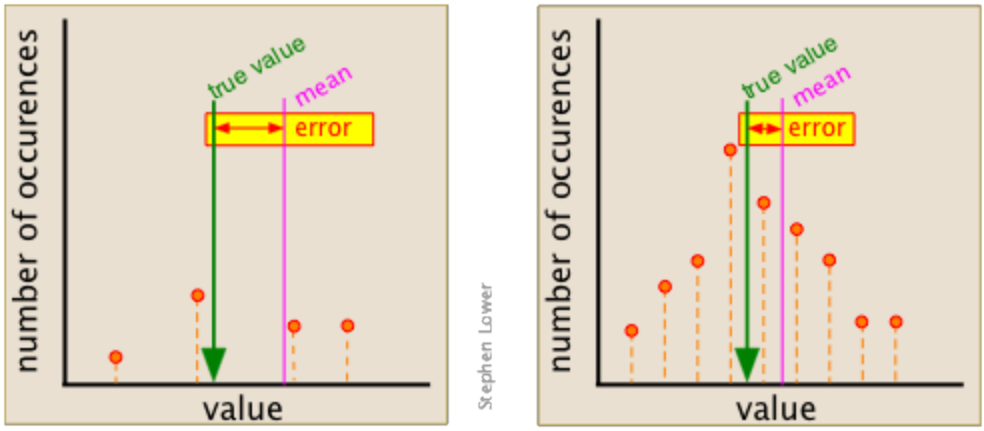
This refers to the value that is observed most frequently in a series of measurements. If two or more values tie for the highest frequency, then there can be multiple modes. Mode is most useful in describing larger data sets.

Example: for the data set [22 23 23 24 26 26] the modes are 23 and 24.

The more observations, the more reliable the mean value. If this is not immediately obvious, think about it this way. You would not want to predict the outcome of the next election on the basis of interviews with only two or three voters; you would want a sample of ten to twenty at a minimum, and if the election is an important national one, a fair sample would require hundreds to thousands of people distributed over the entire geographic area and representing a variety of socio-economic groups. Similarly, you would want to test a large number of light bulbs in order to estimate the mean lifetime of bulbs of that type.

> Statistical theory tells us that the more samples we have, the greater will be the chance that the mean of the results will correspond to the “true” value, which in this case would be the mean obtained if samples could be taken from the entire population (of people or of light bulbs.)

This point can be better appreciated by examining the two sets of data shown here. The set on the left consists of only three points (shown in orange), and gives a mean that is quite far removed from the "true" value, which is arbitrarily chosen for this example.

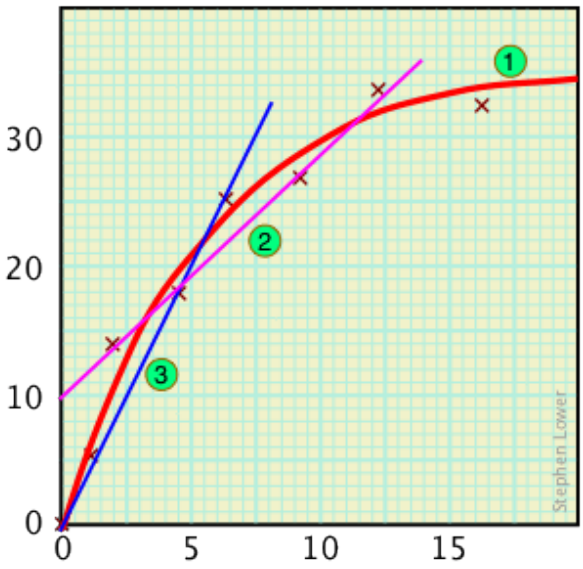


In the data set on the right, composed of nine measurements, the deviation of the mean from the true value is much smaller.

> Deviation of the mean from the "true value" becomes smaller when more measurements are made.

Plots and points

A similar problem arises when you try to fit a curve to a series of plotted points. Suppose, for example, that curve **1** (red) represents the true relationship between the quantities indicated on the *y*-axis (dependent variable) and those on the *x*-axis (independent variable). This curve is derived from the seven points indicated on the plot.



Contrast this curve with the false straight-line relationships that might be obtained if only four or three points had been recorded.

Absolute and Relative Uncertainty

If you weigh out 74.1 mg of a solid sample on a laboratory balance that is accurate to within 0.1 milligram, then the actual weight of the sample is likely to fall somewhere in the range of 74.0 to 74.2 mg; the *absolute uncertainty* in the weight you observe is 0.2 mg, or ±0.1 mg. If you use the same balance to weigh out 3.2914 g of another sample, the actual weight is between 3.2913 g and 3.2915 g, and the absolute uncertainty is still ±0.1 mg. Thus the absolute uncertainty is is unrelated to the magnitude of the observed value.

When expressing the uncertainty of a value given in scientific notation, the exponential part should include both the value itself and the uncertainty. An example of the proper form would be (3.19 ± 0.02) × 10^4 m.

Although the absolute uncertainties in these two examples are identical, we would probably consider the second measurement to be more precise because the uncertainty is a smaller fraction of the measured value. A quantity calculated in this way is known as the *relative uncertainty*.

Example 3.2.1

Calculate the relative uncertainties of the following absolute uncertainties:

a. 74.1 ± 0.1 mg,
b. 3.2914 ± 0.1 mg.

Solution

a. $$\frac{0.2\,mg}{74.1\,mg} = 0.0027 \text{ or } 0.003$$

(note that the quotient is dimensionless) this can be expressed as 0.3% (3 parts per hundred) or 3 parts per thousand.

b. $$\frac{0.0002\,g}{3.2913\,g} = 8.4 \times 10^{-5} \text{ or roughly } 8 \times 10^{-5}$$

, which we can express as $8 \times 10^{-3}\%$ (0.008 parts per hundred), or (8E-5 / 10) = 8E-6 = 8 PPM.

Relative uncertainties are widely used to express the reliability of measurements, even those for a single observation, in which case the uncertainty is that of the measuring device. Relative uncertainties can be expressed as parts per hundred (percent), per thousand (PPT), per million, (PPM), and so on.

Propagation of Error

We are often called upon to find the value of some quantity whose determination depends on several other measured values, each of which is subject to its own sources of error.

Consider a common laboratory experiment in which you must determine the percentage of acid in a sample of vinegar by observing the volume of sodium hydroxide solution required to neutralize a given volume of the vinegar. You carry out the experiment and obtain a value. Just to be on the safe side, you repeat the procedure on another identical sample from the same bottle of vinegar. If you have actually done this in the laboratory, you will know it is highly unlikely that the second trial will yield the same result as the first. In fact, if you run a number of replicate (that is, identical in every way) determinations, you will probably obtain a scatter of results.

To understand why, consider all the individual measurements that go into each determination; the volume of the vinegar sample, your judgement of the point at which the vinegar is neutralized, and the volume of solution used to reach this point. And how accurately do you know the concentration of the sodium hydroxide solution, which was made up by dissolving a measured weight of the solid in water and then adding more water until the solution reaches some measured volume. Each of these many observations is subject to random error; because such errors are random, they can occasionally cancel out, but for most trials we will not be so lucky — hence the scatter in the results.

Rules for estimating errors in calculated results

Suppose you measure the mass and volume of a sample, and are required to calculate its density by dividing one quantity by the other:

$$d = m / V.$$

Both components of this quotient have uncertainties associated with them, and you wish to attach an uncertainty to the calculated density. The general problem of determining the uncertainty of a calculated result turns out to be rather more complicated than

you might think, and will not be treated here. There are, however, some very simple rules that are sufficient for most practical purposes.

- **Addition and subtraction, both numbers have uncertainties**: The simplest method is to just add the absolute uncertainties.
- **Multiplication or division, both numbers have uncertainties**: Convert the absolute uncertainties into *relative* uncertainties, and add these. Or better, add their squares and take the square root of the sum.
- **Multiplication or division by a pure number:** Trivial case; multiply or divide the uncertainty by the pure number.

Example 3.2.2: Addition and Subtraction of Numbers with Uncertainties

$$(6.3 \pm 0.05\,cm) – (2.1 \pm 0.05\,cm) = 4.2 \pm 0.10\,cm$$

However, this tends to over-estimate the uncertainty by assuming the worst possible case in which the error in one of the quantities is at its maximum positive value, while that of the other quantity is at its maximum minimum value.

Statistical theory informs us that a more realistic value for the uncertainty of a sum or difference is to add the *squares* of each absolute uncertainty, and then take the square root of this sum. Applying this to the above values, we have

$$\sqrt{(0.05)^2 + (0.05)^2} = 0.07$$

so the result is 4.2 ± 0.07 cm.

Example 3.2.3

Estimate the absolute error in the density calculated by dividing (12.7 ± 0.05\, g) by (10.0 ± 0.02\, mL).

Solution

Relative uncertainty of the mass:

$$\frac{0.05}{12.7} = 0.0039 = 0.39\%$$

Relative uncertainty of the volume:

$$\frac{0.02}{10.0} = 0.002 = 0.2\%$$

Relative uncertainty of the density:

$$\sqrt{(0.39)^2 + (0.2)^2} = 0.44\%$$

Mass ÷ volume:

$$(12.7\,g) \div (10.0\,mL) = 1.27\,g\,mL^{-1}$$

Absolute uncertainty of the density:

$$(\pm 0.044) \times (1.27\,g\,mL^{-1}) = \pm 0.06\,g\,mL^{-1}$$

This page titled 3.2: The Meaning of Measure is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

3.3: Significant Figures and Rounding off

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Give an example of a measurement whose number of significant digits is clearly too great, and explain why.
- State the purpose of rounding off, and describe the information that must be known to do it properly.
- Round off a number to a specified number of significant digits.
- Explain how to round off a number whose second-most-significant digit is 9.
- Carry out a simple calculation that involves two or more observed quantities, and express the result in the appropriate number of significant figures.

The numerical values we deal with in science (and in many other aspects of life) represent measurements whose values are never known exactly. Our pocket-calculators or computers don't know this; they treat the numbers we punch into them as "pure" mathematical entities, with the result that the operations of arithmetic frequently yield answers that are physically ridiculous even though mathematically correct. The purpose of this unit is to help you understand why this happens, and show you what to do about it.

Digits: significant and otherwise

- "*The population of our city is 157,872.*"
- "*The number of registered voters as of Jan 1 was 27,833.*"

Consider the two statements shown above. Which of these would you be justified in dismissing immediately? Certainly not the second one, because it probably comes from a database which contains one record for each voter, so the number is found simply by counting the number of records.

The first statement cannot possibly be correct. Even if a city's population could be defined in a precise way (Permanent residents? Warm bodies?), how can we account for the minute-by minute changes that occur as people are born and die, or move in and move away?

What is the difference between the two population numbers stated above? The first one expresses a quantity that cannot be known exactly– that is, it carries with it a degree of uncertainty. It is quite possible that the last census yielded precisely 157,872 records, and that this might be the "population of the city" for legal purposes, but it is surely not the "true" population. To better reflect this fact, one might list the population (in an atlas, for example) as **157,900** or even **158,000**. These two quantities have been rounded off to four and three significant figures, respectively, and the have the following meanings:

- **157900** (the significant digits are underlined here) implies that the population is believed to be within the range of about 157850 to about 157950. In other words, the population is 157900±50. The "plus-or-minus 50" appended to this number means that we consider the absolute uncertainty of the population measurement to be 50 – (–50) = 100. We can also say that the relative uncertainty is 100/157900, which we can also express as 1 part in 1579, or 1/1579 = 0.000633, or about 0.06 percent.
- The value **158000** implies that the population is likely between about 157500 and 158500, or 158000±500. The absolute uncertainty of 1000 translates into a relative uncertainty of 1000/158000 or 1 part in 158, or about 0.6 percent.

Which of these two values we would report as "the population" will depend on the degree of confidence we have in the original census figure; if the census was completed last week, we might round to four significant digits, but if it was a year or so ago, rounding to three places might be a more prudent choice. In a case such as this, there is no really objective way of choosing between the two alternatives.

This illustrates an important point: the concept of *significant digits* has less to do with mathematics than with our confidence in a measurement. This confidence can often be expressed numerically (for example, the height of a liquid in a measuring tube can be

read to ±0.05 cm), but when it cannot, as in our population example, we must depend on our personal experience and judgment.

So, what is a significant digit? According to the usual definition, it is all the numerals in a measured quantity (counting from the left) whose values are considered as known exactly, plus one more whose value could be one more or one less:

- In "157900" (four significant digits), the leftmost three digits are known exactly, but the fourth digit, "9" could well be "8" if the "true value" is within the implied range of 157850 to 157950.
- In "*158*000" (three significant digits), the leftmost two digits are known exactly, while the third digit could be either "7" or "8" if the true value is within the implied range of 157500 to 158500.



Although rounding off always leads to the loss of numeric information, what we are getting rid of can be considered to be "numeric noise" that does not contribute to the quality of the measurement.

The purpose in rounding off is to avoid expressing a value to a greater degree of precision than is consistent with the uncertainty in the measurement.

Implied Uncertainty and Round-off error

If you know that a balance is accurate to within 0.1 mg, say, then the uncertainty in any measurement of mass carried out on this balance will be ±0.1 mg. Suppose, however, that you are simply told that an object has a length of 0.42 cm, with no indication of its precision. In this case, all you have to go on is the number of digits contained in the data. Thus the quantity "0.42 cm" is specified to 0.01 unit in 0 42, or one part in 42 . The *implied relative uncertainty* in this figure is 1/42, or about 2%. The precision of any numeric answer calculated from this value is therefore limited to about the same amount.

It is important to understand that the number of significant digits in a value provides only a rough indication of its precision, and that information is lost when rounding off occurs.

Suppose, for example, that we measure the weight of an object as 3.28 g on a balance believed to be accurate to within ±0.05 gram. The resulting value of 3.28±.05 gram tells us that the true weight of the object could be anywhere between 3.23 g and 3.33 g. The absolute uncertainty here is 0.1 g (±0.05 g), and the relative uncertainty is 1 part in 32.8, or about 3 percent.

How many significant digits should there be in the reported measurement? Since only the leftmost "3" in "3.28" is certain, you would probably elect to round the value to 3.3 g. So far, so good. But what is someone else supposed to make of this figure when they see it in your report? The value "3.3 g" suggests an *implied uncertainty* of 3.3±0.05 g, meaning that the true value is likely between 3.25 g and 3.35 g. This range is 0.02 g below that associated with the original measurement, and so rounding off has introduced a bias of this amount into the result. Since this is less than half of the ±0.05 g uncertainty in the weighing, it is not a very serious matter in itself. However, if several values that were rounded in this way are combined in a calculation, the rounding-off errors could become significant.

The standard rules for rounding off are well known. Before we set them out, let us agree on what to call the various components of a numeric value.

- The *most significant digit* is the leftmost digit (not counting any leading zeros which function only as placeholders and are never significant digits.)

- If you are rounding off to ***n*** significant digits, then the *least significant digit* is the ***n***th digit from the most significant digit.The least significant digit can be a zero.
- The first non-significant digit is the ***n***+1th digit.

- If the first non-significant digit is less than 5, then the least significant digit remains unchanged.
- If the first non-significant digit is greater than 5, the least significant digit is incremented by 1.
- If the first non-significant digit is 5, the least significant digit can either be incremented or left unchanged (*see below!*)
- All non-significant digits are removed.

Students are sometimes told to increment the least significant digit by 1 if it is odd, and to leave it unchanged if it is even. One wonders if this reflects some idea that even numbers are somehow “better” than odd ones! (The ancient superstition is just the opposite, that only the odd numbers are "lucky".)

In fact, you could do it equally the other way around, incrementing only the even numbers. If you are only rounding a single number, it doesn’t really matter what you do. However, when you are rounding a series of numbers that will be used in a calculation, if you treated each first-nonsignificant 5 in the same way, you would be over- or underestimating the value of the rounded number, thus accumulating round-off error. Since there are equal numbers of even and odd digits, incrementing only the one kind will keep this kind of error from building up.

You could do just as well, of course, by flipping a coin!

Table: Examples of rounding-off

number to round / no. of sig. digits	result	comment
34.216 / 3	34.2	First non-significant digit (1) is less than 5, so number is simply truncated.
2.252 / 2	2.2 or 2.3	First non-significant digit is 5, so least sig. digit can either remain unchanged or be incremented.
39.99 / 3	40.0	Crossing "decimal boundary", so all numbers change.
85,381 / 3	85,400	The two zeros are just placeholders
0.04597 / 3	0.0460	The two leading zeros are not significant digits.

Rounding Up The Nines

Suppose that an object is found to have a weight of 3.98 ± 0.05 g. This would place its true weight somewhere in the range of **3.93** g to **4.03** g. In judging how to round this number, you count the number of digits in “3.98” that are known exactly, and you find none! Since the “4” is the leftmost digit whose value is uncertain, this would imply that the result should be rounded to one significant figure and reported simply as 4 g. An alternative would be to bend the rule and round off to two significant digits, yielding 4.0 g. How can you decide what to do?

In a case such as this, you should look at the implied uncertainties in the two values, and compare them with the uncertainty associated with the original measurement.

rounded value	implied max	implied min	absolute uncertainty	relative uncertainty
3.98 g	3.985 g	3.975 g	±.005 g or 0.01 g	1 in 400, or 0.25%
4 g	4.5 g	3.5 g	±.5 g or 1 g	1 in 4, 25%
4.0 g	**4.05 g**	**3.95 g**	**±.05 g or 0.1 g**	**1 in 40, 2.5%**

Clearly, rounding off to two digits is the only reasonable course in this example.

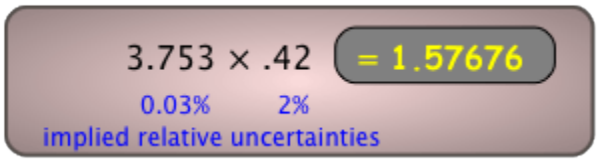
The same kind of thing could happen if the original measurement was 9.98 ± 0.05 g. Again, the true value is believed to be in the range of 10.03 g to 9.93 g. The fact that no digit is certain here is an artifact of decimal notation. The absolute uncertainty in the observed value is 0.1 g, so the value itself is known to about 1 part in 100, or 1%. Rounding this value to three digits yields 10.0 g with an implied uncertainty of ±.05 g, or 1 part in 100, consistent with the uncertainty in the observed value.

Observed values should be rounded off to the number of digits that most accurately conveys the uncertainty in the measurement.

- Usually, this means rounding off to the number of significant digits in in the quantity; that is, the number of digits (counting from the left) that are known exactly, plus one more.
- When this cannot be applied (as in the example above when addition of subtraction of the absolute uncertainty bridges a power of ten), then we round in such a way that the relative implied uncertainty in the result is as close as possible to that of the observed value.

Rounding off the Results of Calculations

When carrying out calculations that involve multiple steps, you should avoid doing any rounding until you obtain the final result. In science, we frequently need to carry out calculations on measured values. For example, you might use your pocket calculator to work out the area of a rectangle:



Your calculator is of course correct as far as the pure numbers go, but you would be wrong to write down 1.57676 cm^2 as the answer. Two possible options for rounding off the calculator answer are shown below:

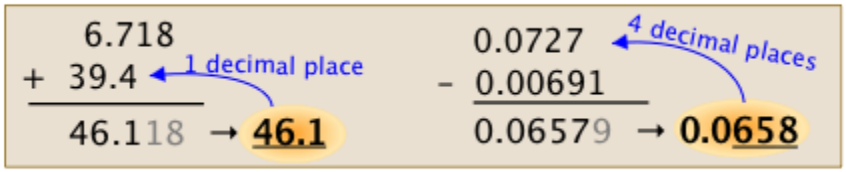
Rounded Value	Precision
1.58	1 part in 158, or 0.6%
1.6	1 part in 16, or 6 %

It is clear that neither option is entirely satisfactory; rounding to 3 significant digits leaves the answer too precisely specified, whereas following the rule and rounding to 2 digits has the effect of throwing away some precision. In this case, it could be argued that rounding to three digits is justified because the implied relative uncertainty in the answer, 0.6%, is more consistent with those of the two factors.

The above example is intended to point out that the rounding-off rules, although convenient to apply, do not always yield the most desirable result. When in doubt, it is better to rely on relative implied uncertainties.

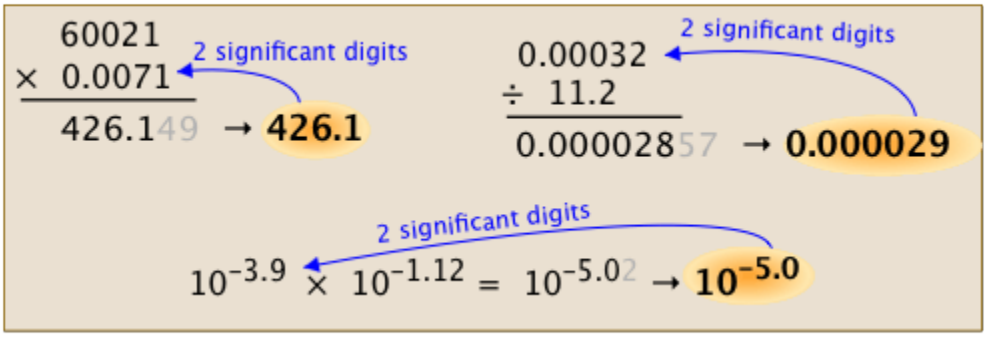
Addition and Subtraction

When adding or subtracting, we go by the number of *decimal places* rather than by the number of significant digits. Identify the quantity having the smallest number of decimal places, and use this number to set the number of decimal places in the answer.



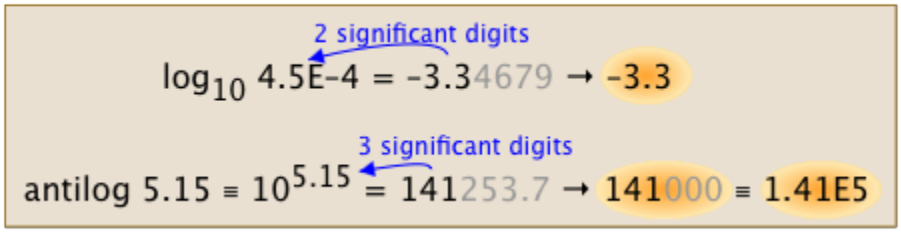
Multiplication and Division

The result must contain the same number of significant figures as in the value having the least number of significant figures.



Logarithms and Antilogarithms

Express the base-10 logarithm of a value using the same number of significant figures as is present in the *normalized form* of that value. Similarly, for antilogarithms (numbers expressed as powers of 10), use the same number of significant figures as are in that power.

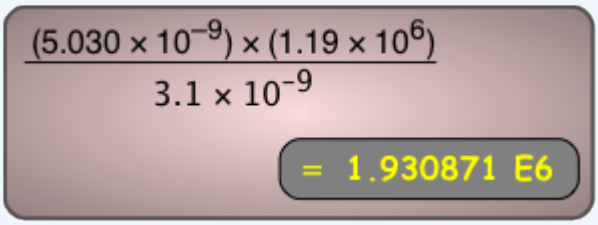
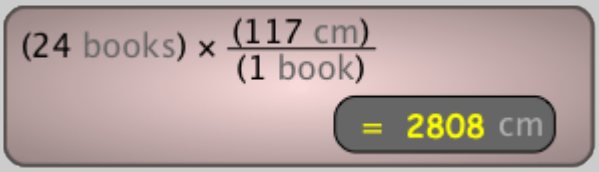
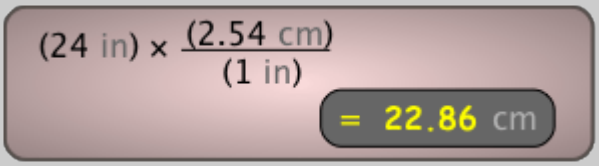


If a number is expressed in the form $a \times 10^b$ ("scientific notation") with the additional restriction that the coefficient a is no less than 1 and less than 10, the number is in its *normalized* form.

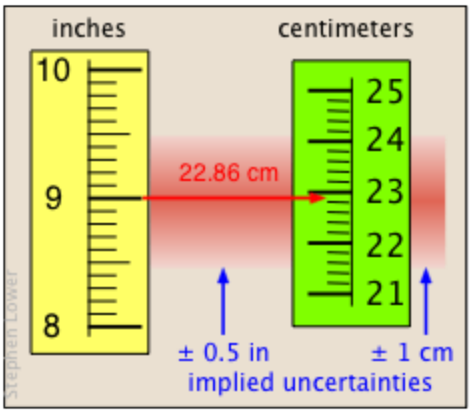
More Rounding Examples

The following examples will illustrate the most common problems you are likely to encounter in rounding off the results of calculations. They deserve your careful study!

calculator result	rounded	remarks
3.753 × .42 = 1.57676 0.03% 2% implied relative uncertainties	1.6	Rounding to two significant figures yields an implied uncertainty of 1/16 or 6%, three times greater than that in the least-preciseely known factor. This is a good illustration of how rounding can lead to the loss of information.

calculator result	rounded	remarks
	1.9E6	The "3.1" factor is specified to 1 part in 31, or 3%. In the answer 1.9, the value is expressed to 1 part in 19, or 5%. These precisions are comparable, so the rounding-off rule has given us a reasonable result.
A certain book has a thickness of 117 mm; find the height of a stack of 24 identical books: 	2810 mm	The “24” and the “1” are exact, so the only uncertain value is the thickness of each book, given to 3 significant digits. The trailing zero in the answer is only a placeholder.
	10.4	In addition or subtraction, look for the term having the smallest number of decimal places, and round off the answer to the same number of places.
	23 cm	[see below]

The last of the examples shown above represents the very common operation of converting one unit into another. There is a certain amount of ambiguity here; if we take "9 in" to mean a distance in the range 8.5 to 9.5 in, then the uncertainty is ±0.5 in, which is 1 part in 18, or about ± 6%. The relative uncertainty in the answer must be the same, since all the values are multiplied by the same factor, 2.54 cm/in. In this case we are justified in writing the answer to two significant digits, yielding an uncertainty of about ±4 cm; if we had used the answer "20 cm" (one significant digit), its implied uncertainty would be ±5 cm, or ±25%.



When the appropriate number of significant digits is in question, calculating the relative uncertainty can help you decide.

This page titled 3.3: Significant Figures and Rounding off is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

3.4: Reliability of a measurement

Learning Objectives

- Explain the distinction between the *mean* value of a series of measurements and the *population mean*.
- What quantity besides the mean value do we need in order to evaluate the quality of a series of measurements?
- Explain the meaning and significance of the dispersion of the mean, and state what factor controls it.
- Explain the distinction between *determinate* and *indeterminate* error.
- Describe the purpose process of using a *blank* and *control* value when making a series of measurements. What principal assumption must be made in doing this?

In this day of pervasive media, we are continually being bombarded with data of all kinds— public opinion polls, advertising hype, government reports and statements by politicians. Very frequently, the purveyors of this information are hoping to "sell" us on a product, an idea, or a way of thinking about someone or something, and in doing so, they are all too often willing to take advantage of the average person's inability to make informed judgments about the reliability of the data, especially when it is presented in a particular context (popularly known as "spin".) In Science, we do not have this option: we collect data and make measurements in order to get closer to whatever "truth" we are seeking, but it's not really "science" until others can have confidence in the reliability of our measurements.

Attributes of a measurement

The kinds of measurements we will deal with here are those in which a number of separate observations are made on individual samples taken from a larger *population*.

Population, when used in a statistical context, does not necessarily refer to people, but rather to the set of all members of the group of objects under consideration.

For example, you might wish to determine the amount of nicotine in a manufacturing run of one million cigarettes. Because no two cigarettes are likely to be exactly identical, and even if they were, random error would cause each analysis to yield a different result, the best you can do would be to test a representative sample of, say, twenty to one hundred cigarettes. You take the average (mean) of these values, and are then faced with the need to estimate how closely this sample mean is likely to approximate the population mean. The latter is the "true value" we can never know; what we can do, however, is make a reasonable estimate of the likelihood that the sample mean does not differ from the population mean by more than a certain amount.

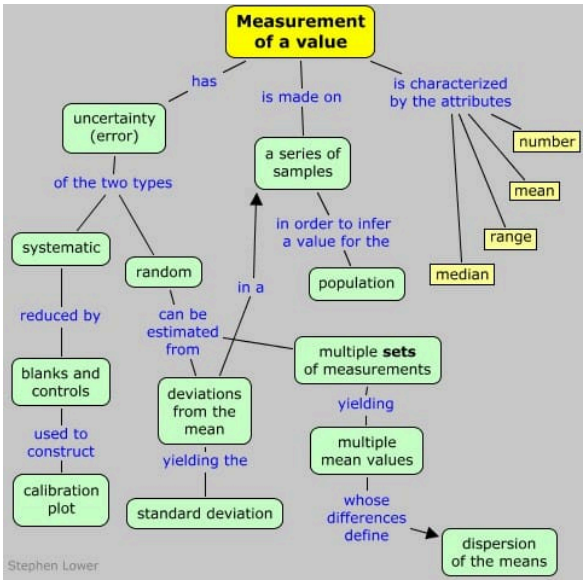
The attributes we can assign to an individual set of measurements of some quantity x within a population are listed below. It is important that you learn the meaning of these terms:

Number of measurements

This quantity is usually represented by n.

Mean

The *mean* value x_m (commonly known as the *average*), defined as



Don't let this notation scare you! It just means that you add up all the values and divide by the number of values.

Median

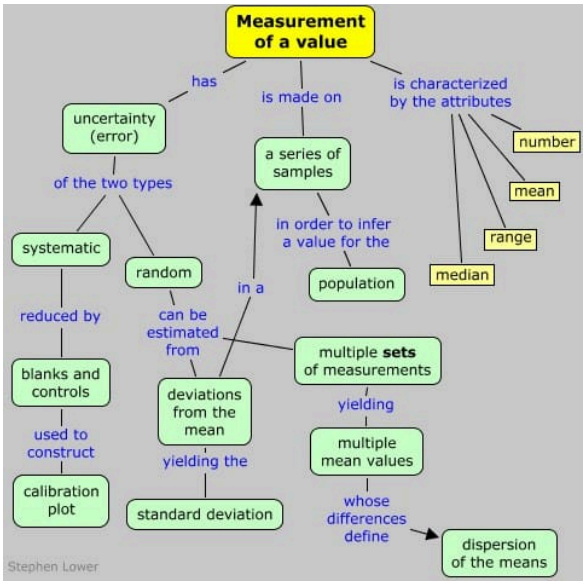
The *median* value, which we will not deal with in this brief presentation, is essentially the one in the middle of the list resulting from writing the individual values in order of increasing or decreasing magnitude.

Range

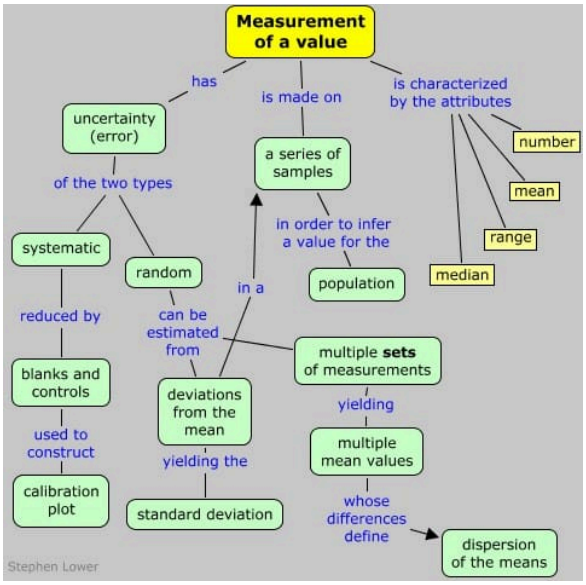
The *range* is the difference between the largest and smallest value in the set.

Problem example:

Find the mean value and range of the set of measurements depicted here.



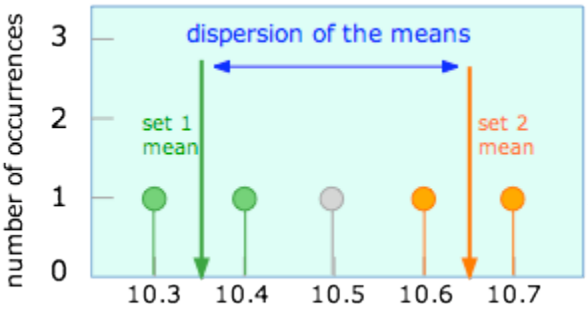
Solution: This set contains 8 measurements. The range is (10.7 – 10.3) = 0.4, and the mean value is



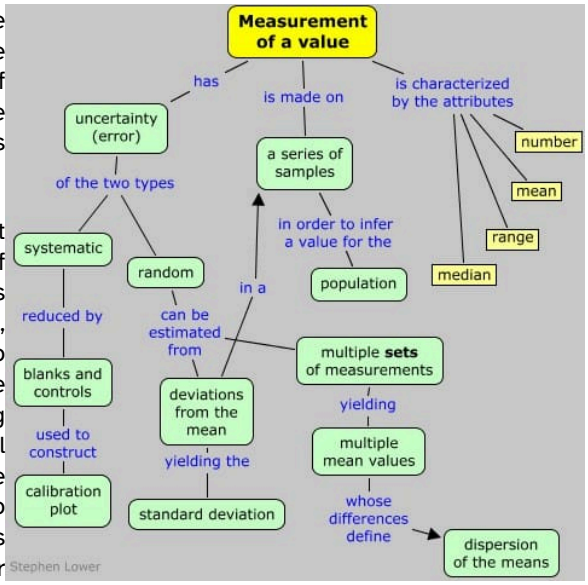
6 More than one answer: dispersion of the mean

"Dispersion" means "spread-outedness". If you make a few measurements and average them, you get a certain value for the mean. But if you make another set of measurements, the mean of these will likely be different. The greater the difference between the means, the greater is their *dispersion*.

Suppose that instead of taking the five measurements as in the above example, we had made only two observations which, by chance, yielded the values that are highlighted here. This would result in a sample mean of 10.45. Of course, any number of other pairs of values could equally well have been observed, including multiple occurances of any single value, such as 10.6.



Shown at the left are the results of two possible pairs of observations, each giving rise to its own sample mean. Assuming that all observations are subject only to random error, it is easy to see that successive pairs of experiments could yield many other sample means. The range of possible sample means is known as the dispersion of the mean.



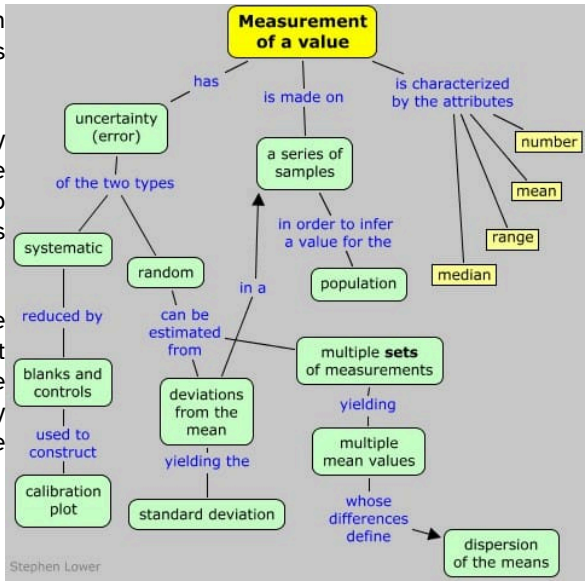
It is clear that both of the two sample means cannot correspond to the population mean, whose value we are really trying to discover. In fact, it is quite likely that neither sample mean is the “correct” one in this sense. It is a fundamental principle of statistics, however, that the more observations we make in order to obtain a sample mean, the smaller will be the dispersion of the sample means that result from repeated sets of the same number of observations. (This is important; please read the preceding sentence at least three times to make sure you understand it!)

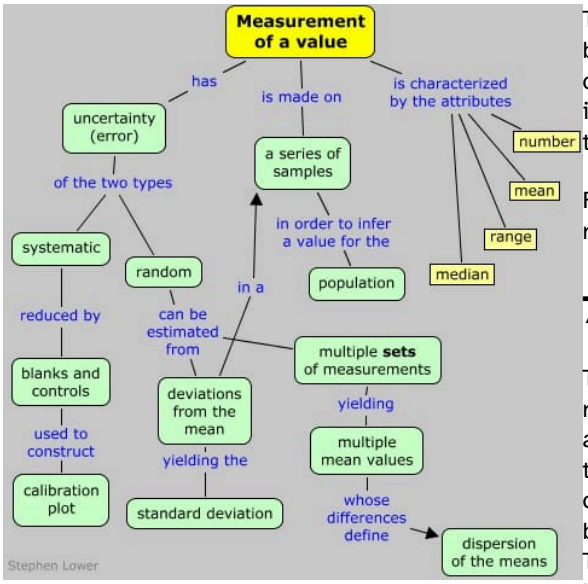
How the dispersion of the mean depends on the number of observations

The difference between the sample mean (blue) and the population mean (the "true value", green) is the *error* of the measurement. It is clear that this error diminishes as the number of observations is made larger.

What is stated above is just another way of saying what you probably already know: larger samples produce more reliable results. This is the same principle that tells us that flipping a coin 100 times will be more likely to yield a 50:50 ratio of heads to tails than will be found if only ten flips (observations) are made.

The reason for this inverse relation between the sample size and the dispersion of the mean is that if the factors giving rise to the different observed values are truly random, then the more samples we observe, the more likely will these errors cancel out. It turns out that if the errors are truly random, then as you plot the number of occurrences of each value, the results begin to trace out a very special kind of curve.



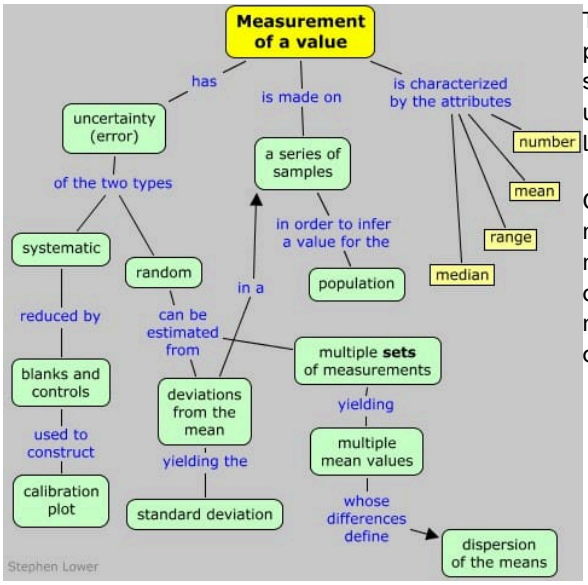


The significance of this is much greater than you might at first think, because the G*aussian curve* has special mathematical properties that we can exploit, through the methods of statistics, to obtain some very useful information about the reliability of our data. This will be the major topic of the next lesson in this set.

For now, however, we need to establish some important principles regarding measurement error.

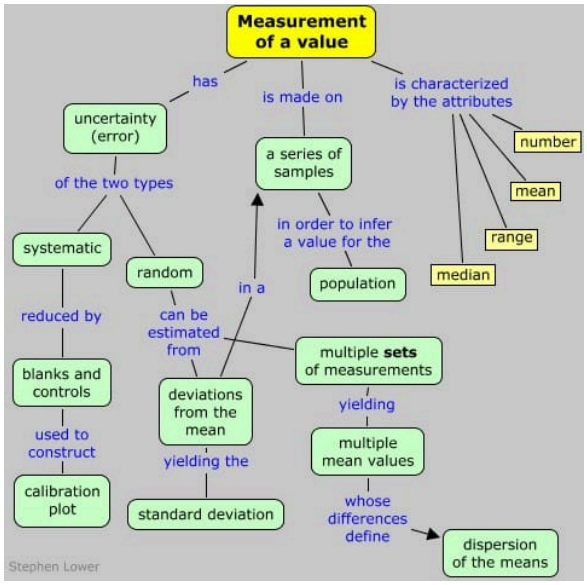
7 Systematic error

The scatter in measured results that we have been discussing arises from random variations in the myriad of events that affect the observed value, and over which the experimenter has no or only limited control. If we are trying to determine the properties of a collection of objects (nicotine content of cigarettes or lifetimes of lamp bulbs), then random variations between individual members of the population are an ever-present factor. This type of error is called random or *indeterminate error*, and it is the only kind we can deal with directly by means of statistics.

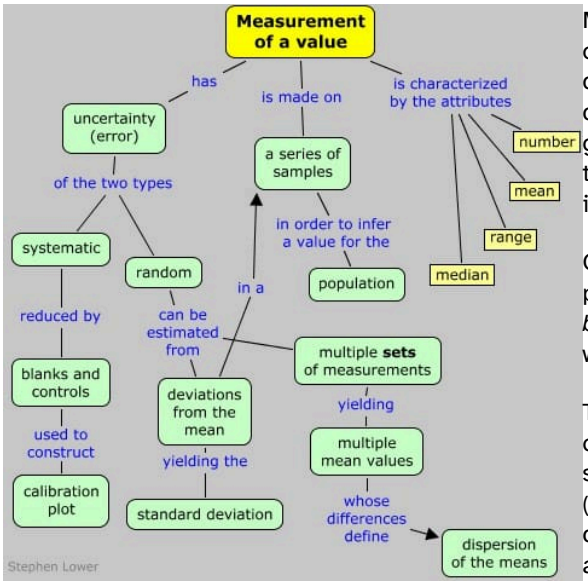


There is, however, another type of error that can afflict the measuring process. It is known as *systematic* or *determinate error*, and its effect is to shift an entire set of data points by a constant amount. Systematic error, unlike random error, is not apparent in the data itself, and must be explicitly looked for in the design of the experiment.

One common source of systematic error is failure to use a reliable measuring scale, or to misread a scale. For example, you might be measuring the length of an object with a ruler whose left end is worn, or you could misread the volume of liquid in a burette by looking at the top of the meniscus rather than at its bottom, or not having your eye level with the object being viewed against the scale, thus introducing *parallax* error.



8 Blanks and controls



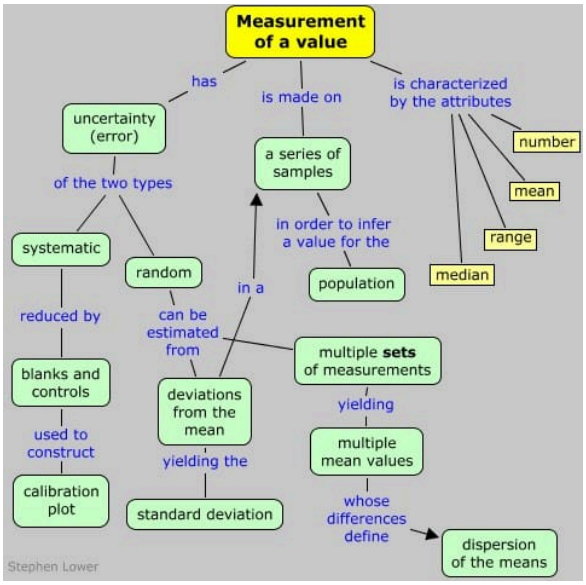
Many kinds of measurements are made by devices that produce a response of some kind (often an electric current) that is directly proportional to the quantity being measured. For example, you might determine the amount of dissolved iron in a solution by adding a reagent that reacts with the iron to give a red color, which you measure by observing the intensity of green light that passes through a fixed thickness of the solution. In a case such as this, it is common practice to make two additional kinds of measurements:

One measurement is done on a solution as similar to the unknowns as possible except that it contains no iron at all. This sample is called the *blank*. You adjust a control on the photometer to set its reading to zero when examining the blank.

The other measurement is made on a sample containing a known concentration of iron; this is usually called the *control*. You adjust the sensitivity of the photometer to produce a reading of some arbitrary value (50, say) with the control solution. Assuming the photometer reading is directly proportional to the concentration of iron in the sample (this might also have to be checked, in which case a *calibration curve* must be constructed), the photometer reading can then be converted into iron concentration by simple proportion.

9 The standard deviation

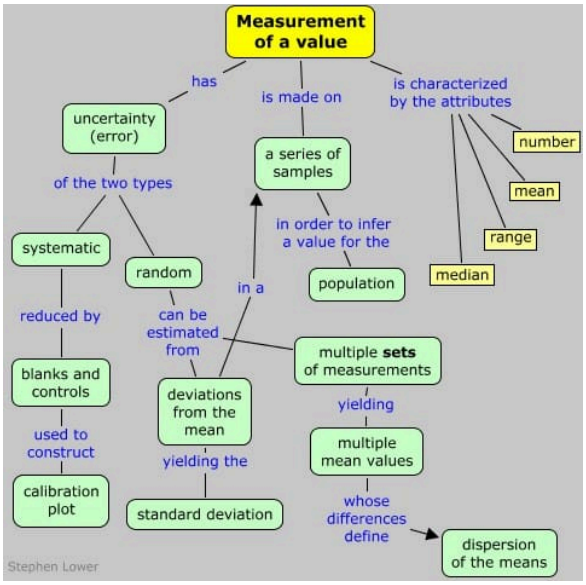
Consider the two pairs of observations depicted here:



Notice that the sample means happen to have the same value of “40” (pure luck!), but the difference in the precisions of the two measurements makes it obvious that the set shown on the right is more reliable. How can we express this fact in a succinct way? We might say that one experiment yields a value of 40 ±20, and the other 40 ±5. Although this information might be useful for some purposes, it is unable to provide an answer to such questions as "how likely would another independent set of measurements yield a mean value within a certain range of values?" The answer to this question is perhaps the most meaningful way of assessing the "quality" or reliability of experimental data, but obtaining such an answer requires that we employ some formal statistics.

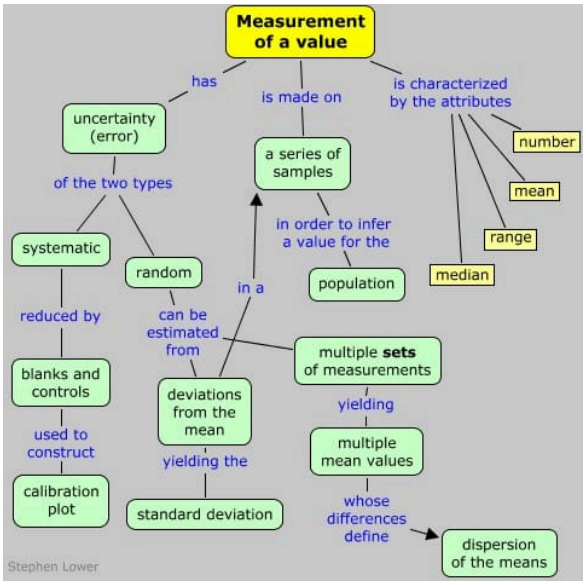
Deviations from the mean

We begin by looking at the differences between the sample mean and the individual data values used to compute the mean. These differences are known as *deviations from the mean*, $x_i - x_m$. These values are depicted below; note that the only difference from the plots above is placement of the mean value at 0 on the horizontal axis.

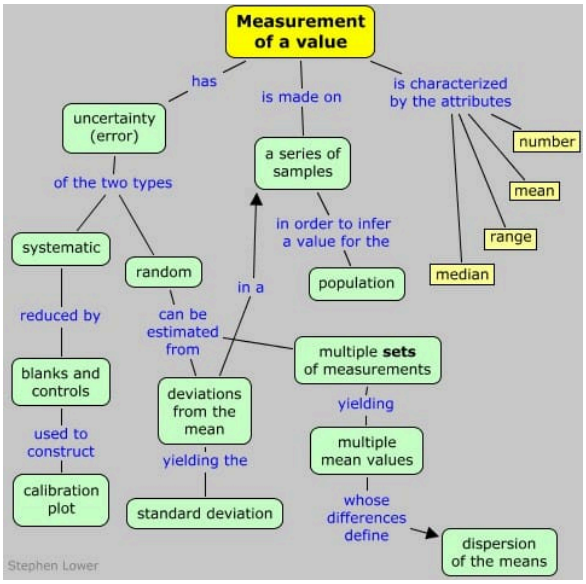


The variance and its square root

Next, we need to find the average of these deviations. Taking a simple average, however, will not distinguish between these two particular sets of data, because both deviations average out to zero. We therefore take the average of the *squares* of the deviations (squaring makes the signs of the deviations disappear so they cannot cancel out). Also, we compute the average by dividing by one *less* than the number of measurements, that is, by $n-1$ rather than by n. The result, usually denoted by $\mathbf{s^2}$, is known as the *variance*:



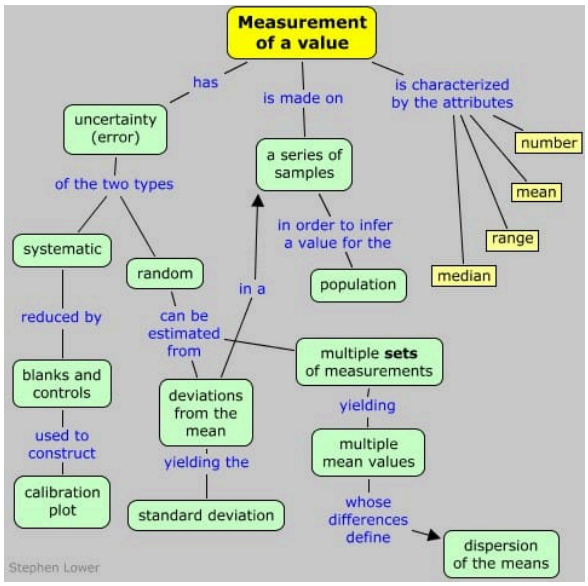
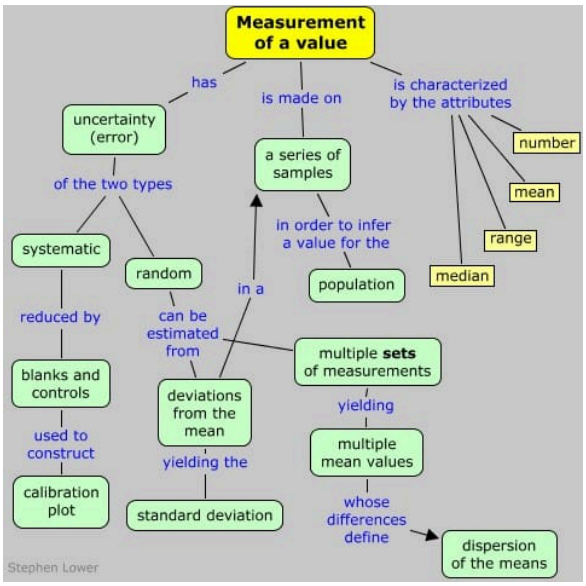
Finally, we take the square root of the variance to obtain the *standard deviation* s:



This is the most important formula in statistics; it is so widely used that most scientific calculators provide built-in means of calculating S from the raw data.

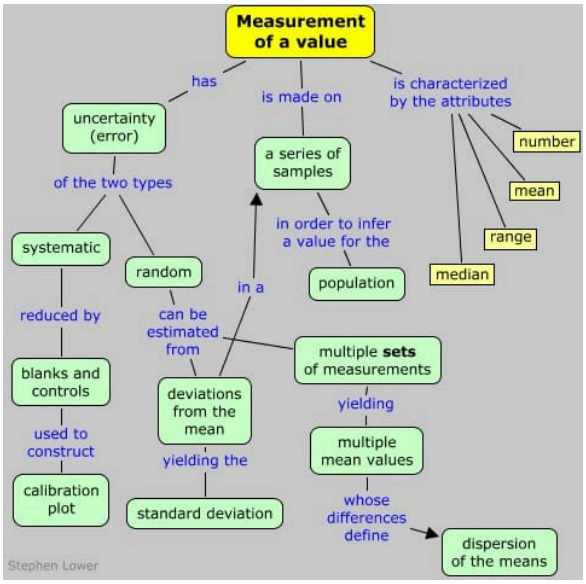
Problem example: Calculate the variance and standard deviation for each of the two data sets shown above.

Solution: Substitution into the two formulas yields the following results:

data values	20, 60	35,45
sample mean	40	40
variance S^2		
standard deviation	28	7.1

Comment: Notice how the contrasting values of S reflect the difference in the precisions of the two data sets— something that is entirely lost if only the two means are considered.

Now that we have developed the very important concept of standard deviation, we can employ it in the next section to answer practical questions about how to interpret the results of a measurement.



Back to Essential Background for General Chemistry

Back to chem1 (Stephen Lower/Simon Frasier University)

This page titled 3.4: Reliability of a measurement is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

3.5: Drawing Conclusions from Data

Learning Objectives

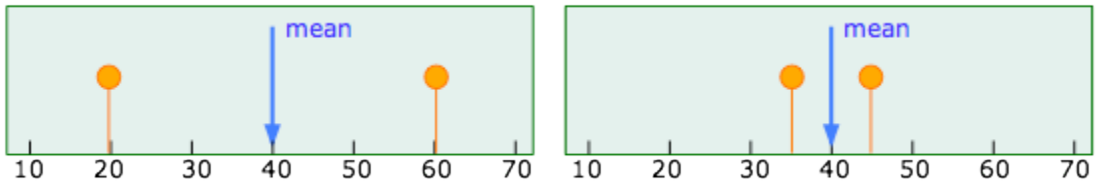
Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the italicized terms in the context of this topic.

- What is the *deviation from the population mean*, why can we not know its value, and why is it neverthless a fundamentally important quantity in statistics?
- Sketch out a *Gaussian curve*, and label the two axes, showing on the x-axis the deviation from the mean in terms of *standard deviations*. Shade in the area corresponding to the *95.4-percent confidence level*,
- State the meaning of a confidence interval and how it relates to the standard deviation on a plot of the Gaussian curve.
- State the meaning of a *confidence interval* and how it relates to the *standard deviation* on a plot of the *Gaussian curve*.
- What is the distinction between a *confidence interval* and the *confidence level*?
- Describe the circumstances when a *Student*'s *t* statistic is useful.
- Describe some of the major problems that can cause statistics to be erroneous or misleading.

OK, you have collected your data, so what does it mean? This question commonly arises when measurements made on different samples yield different values. How well do measurements of mercury concentrations in ten cans of tuna reflect the composition of the factory's entire output? Why can't you just use the average of these measurements? How much better would the results of 100 such tests be? This final lesson on measurement will examine these questions and introduce you to some of the methods of dealing with data. This stuff is important not only for scientists, but also for any intelligent citizen who wishes to independently evaluate the flood of numbers served up by advertisers, politicians, "experts", and yes— by other scientists.

The Standard Deviation

Each of these sets has the same mean value of 40, but the "quality" of the set shown on the right is greater because the data points are less scattered; the *precision* of the result is greater.



The quantitative measure of this precision is given by the *standard deviation*

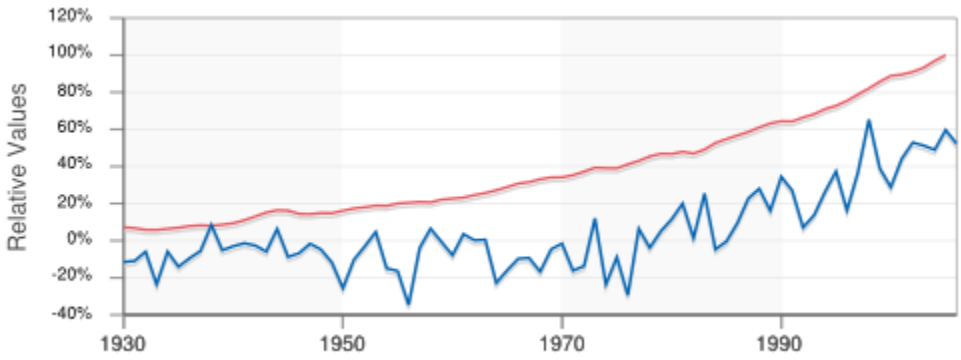


whose value works out to 28 and 7 for the two sets illustrated above. A data set containing only two values is far too small for a proper statistical analysis— you would not want to judge the average mercury content of canned tuna on the basis of only two samples, for instance. Suppose, then, for purposes of illustration, that we have accumulated many more data points but the standard deviations of the two sets remain at 28 and 7 as before. What conclusions can we draw about how close the mean value of 40 is likely to come to the "true value" (the *population mean* μ) in each case?

Although we cannot ordinarily know the value of μ, we can assign to each data point x_i a quantity $(x_i - x_m)$ which we call the *deviation from the [population] mean*, an index of how far each data point differs from the elusive “true value”. We now divide this deviation from the mean by the standard deviation of the entire data set:



If we plot the values of *z* that correspond to each data point, we obtain the following curves for the two data sets we are using as examples:



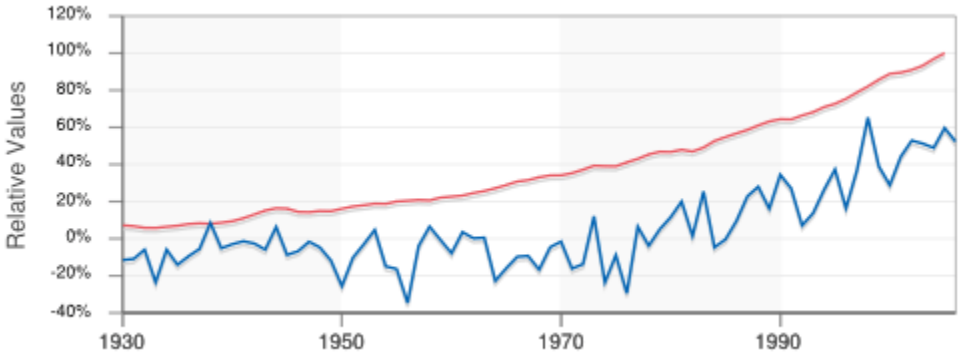
Bear in mind that we cannot actually plot these curves from our experimental data points because we don't know the value of the population mean μ (if we did, there would be no need to make the measurements in the first place!), and we are unlikely to have enough data points to obtain a smooth curve anyway.

We won't attempt to prove it here, but the mathematical properties of a Gaussian curve are such that its shape depends on the scale of units along the x-axis and on the standard deviation of the corresponding data set. In other words, if we know the standard deviation of a data set, we can construct a plot of ***z*** that shows how the measurements would be distributed

- if the number of observations is very large
- if the different values are due only to random error

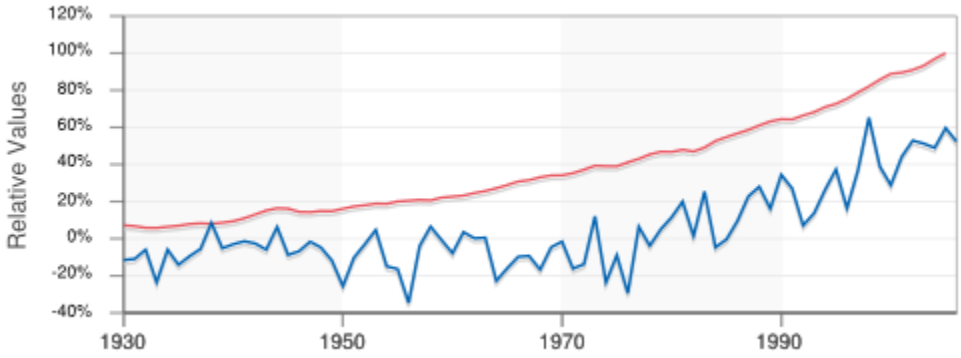
An important corollary to the second condition is that if the data points do **not** approximate the shape of this curve, then it is likely that the sample is not representative, or that some complicating factor is involved. The latter often happens when a teacher plots a set of student exam scores, and gets a curve having two peaks instead of one— representing perhaps the two sub-populations of students who devote their time to studying and partying.

This minor gem was devised by the statistician W.J. Youdan and appears in *The visual display of quantitative information*, an engaging book by Edward R. Tufte (Graphics Press, Cheshire CT, 1983).



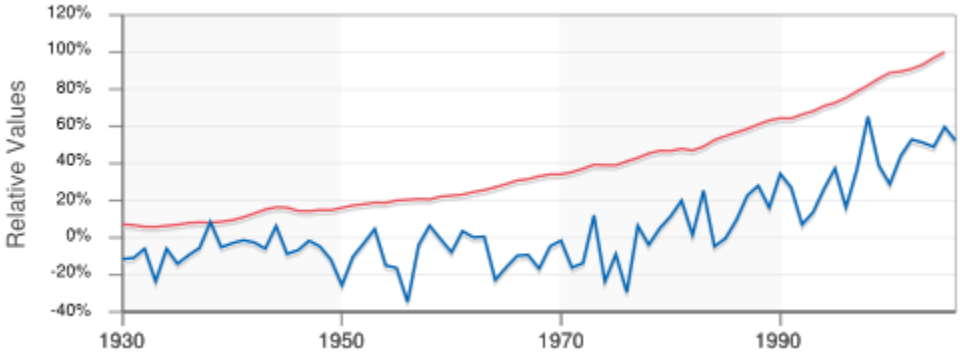
Confidence intervals

Clearly, the sharper and more narrow the standard error curve for a set of measurement, the more likely it will be that any single observed value approximates the true value we are trying to find. Because the shape of the curve is determined by S, we can make quantitative predictions about the reliability of our data from its standard deviation. In particular, if we plot *z* as a function of the number of standard deviations from the mean (rather than as the number of absolute deviations from the mean as was done above), the shape of the curve depends only on the value of S. That is, the dependence on the particular units of measurement is removed.



Moreover, it can be shown that if all measurement error is truly random, 68.3 percent (about two-thirds) of the data points will fall within one standard deviation of the population mean, while 95.4 percent of the observations will differ from the population mean by no more than two standard deviations. This is extremely important, because it allows us to express the reliability of a measurement quantitatively, in terms of*confidence intervals*.

You might occasionally see or hear a news report stating that the results of a certain public opinion poll are considered reliable to within, say, 5%, “nineteen times out of twenty”. This is just another way of saying that the confidence interval in the poll is 95%, the standard deviation is about 2.5% of the stated result, and that there is no more than a 5% chance that an identical poll carried out on another set of randomly-selected individuals from the same population would yield a different result. This is as close to “the truth” as we can get in scientific measurements.



Note carefully: **Confidence *interval*(CI) and confidence *level* (CL) are not the same!**

A given CI (denoted by the shaded range of 18-33 ppm in the diagram) is always defined in relation to some particular CL; specifying the first without the second is meaningless. If the CI illustrated here is at the 90% CL, then a CI for a higher CL would be wider, while that for a smaller CL would encompass a smaller range of values.

The units of CI are those of the *measurement* (e.g., ppm); CL itself is usually expressed in *percent*.

How the confidence level depends on the number of measurements

The more measurements we make, the more likely will their average value approximate the true value. The width of the confidence interval (expressed in the actual units of measurement) is directly proportional to the standard deviation **S** and to the value of ***z*** (both of these terms are defined above). The confidence interval of a **single measurement** in terms these quantities and of the observed sample mean is given by:

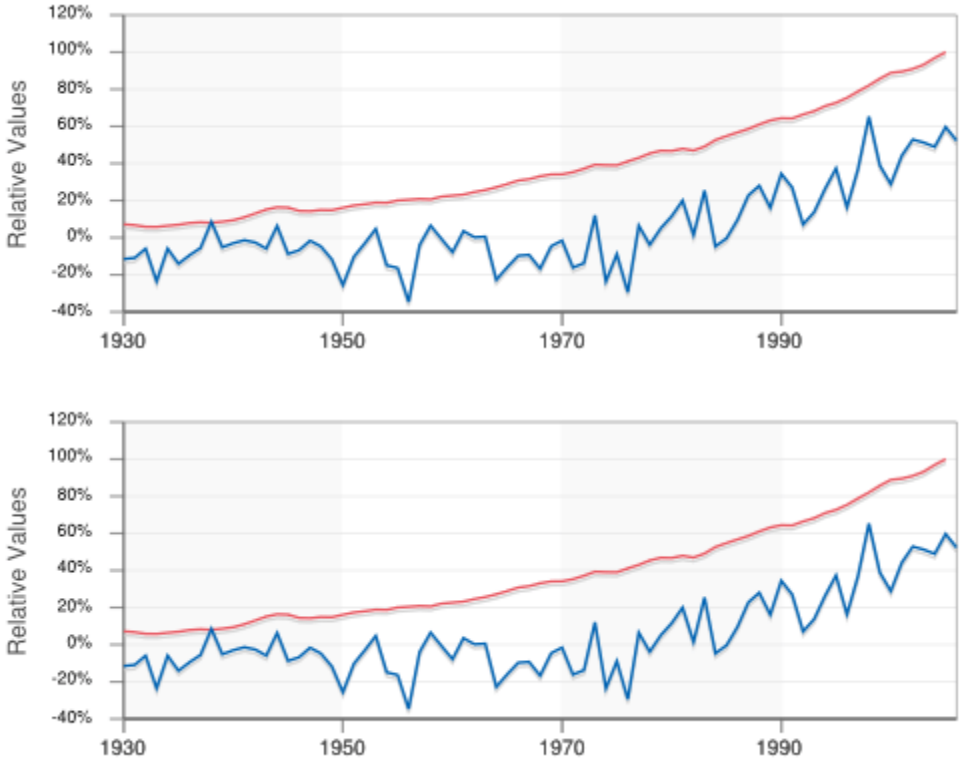
$CI = x_m + z\ S$

If *n* replicate measurements are made, the confidence interval becomes smaller:



This relation is often used “in reverse”, that is, to determine how many replicate measurements *n* must be carried out in order to obtain a value within a desired confidence interval.

As we pointed out above, any relation involving the quantity z (which the standard error curve is a plot of) is of limited use unless we have some idea of the value of the population mean μ. If we make a very large number of measurements (100 to 1000, for example), then we can expect that our observed sample mean approximates μ quite closely, so there is no difficulty.



The shaded area in each plot shows the fraction of measurements that fall within two standard deviations (2S) of the "true" value (that is, the population mean μ). It is evident that the width of the confidence interval diminishes as the number of measurements becomes greater. This is basically a result of the fact that relatively large random errors tend to be less common than smaller ones, and are therefore less likely to cancel out if only a small number of measurements is made.

Dealing with small data sets

OK, so larger data sets are better than small ones. But what if it is simply not practical to measure the mercury content of 10,000 cans of tuna? Or if you were carrying out a forensic examination of a tiny chip of paint, you might have only enough sample (or enough time) to do two or three replicate analyses. There are two common ways of dealing with such a difficulty.

One way of getting around this is to use *pooled data*; that is, to rely on similar prior determinations, carried out on other comparable samples, to arrive at a standard deviation that is representative of this particular type of determination. The other common way of dealing with small numbers of replicate measurements is to look up, in a table, a quantity ***t***, whose value depends on the number of measurements and on the desired confidence level. For example, for a confidence level of 95%, ***t*** would be 4.3 for three samples and 2.8 for five. The magnitude of the confidence interval is then given by

$$CI = \pm t\ S$$

This procedure is not black magic, but is based on a careful analysis of the way that the Gaussian curve becomes distorted as the number of samples diminishes. **Why was the *t*-test invented in a brewery?** And why does it have such a funny name?

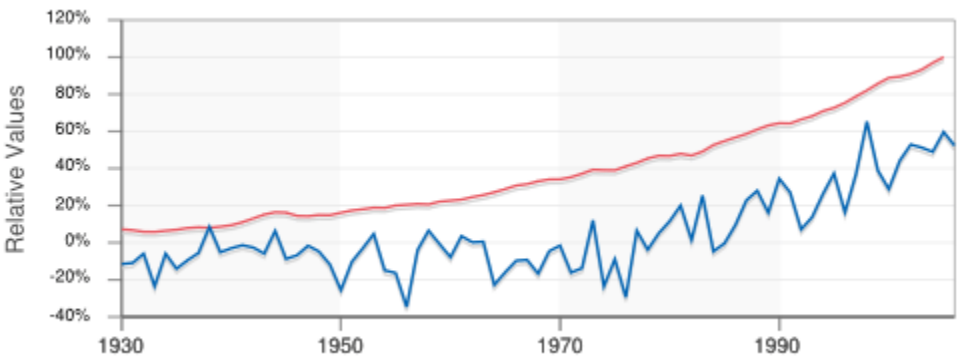
Using statistical tests to make decisions

Once we have obtained enough information on a given sample to evaluate parameters such as means and standard deviations, we are often faced with the necessity of comparing that sample (or the population it represents) with another sample or with some kind of a standard. The following sections paraphrase some of the typical questions that can be decided by statistical tests based on the quantities we have defined above. It is important to understand, however, that because we are treating the questions statistically, we can only answer them in terms of statistics— that is, to a given confidence level.

The usual approach is to begin by assuming that the answer to any of the questions given below is "no" (this is called the *null hypothesis*), and then use the appropriate statistical test to judge the validity of this hypothesis to the desired confidence level. Because our purpose here is to show you **what** can be done rather than **how** to do it, the following sections do not present formulas or example calculations, which are covered in most textbooks on analytical chemistry. You should concentrate here on trying to understand why questions of this kind are of importance.

"Should I throw this measurement out?"

That is, is it likely that something other than ordinary indeterminate error is responsible for this suspiciously different result? Anyone who collects data of almost any kind will occasionally be faced with this question. Very often, ordinary common sense will be sufficient, but if you need some help, two statistical tests, called the *Q*test and the *T* test, are widely employed for this purpose.



We won't describe them here, but both tests involve computing a quantity (*Q* or*T*) for a particular result by means of a simple formula, and then consulting a table to determine the likelihood that the value being questioned is a member of the population represented by the other values in the data set.

"Does this method yield reliable results?"

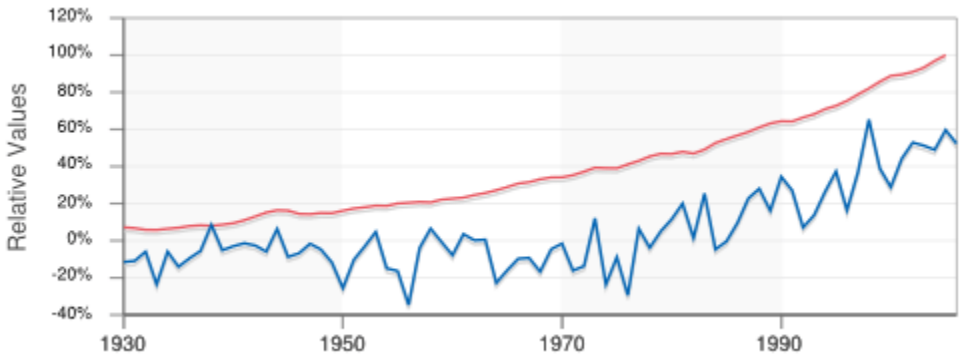
This must always be asked when trying a new method for the first time; it is essentially a matter of testing for determinate error. The answer can only be had by running the same procedure on a sample whose composition is known. The deviation of the mean

value of the “known” $x_m$from its true value μ is used to compute a *Student's t* for the desired confidence level. You then apply this value of t to the measurements on your unknown samples.

“Are these two samples identical?”

You wish to compare the means x_{m1} and x_{m2} from two sets of measurements in order to assess whether their difference could be due to indeterminate error. Suppose, for example, that you are comparing the percent of chromium in a sample of paint removed from a car's fender with a sample found on the clothing of a hit-and-run victim. You run replicate analyses on both samples, and obtain different mean values, but the confidence intervals overlap. What are the chances that the two samples are in fact identical, and that the difference in the means is due solely to indeterminate error?

A fairly simple formula, using Student’s *t*, the standard deviation, and the numbers of replicate measurements made on both samples, provides an answer to this question, but only to a specified confidence level. If this is a forensic investigation that you will be presenting in court, be prepared to have your testimony demolished by the opposing lawyer if the CL is less than 99%.



“What is the smallest quantity I can detect?”

This is just a variant of the preceding question. Estimation of the detection limit of a substance by a given method begins with a set of measurements on a *blank*, that is, a sample in which the substance of question is assumed to be absent, but is otherwise as similar as possible to the actual samples to be tested. We then ask if any difference between the mean of the blank measurements and of the sample replicates can be attributed to *indeterminate error* at a given *confidence level*.

For example, a question that arises at every world Olympics event, is what is the minimum level of a drug metabolite that can be detected in an athlete's urine? Many sensitive methods are subject to random errors that can lead to a non-zero result even in a sample known to be entirely free of what is being tested for. So how far from "zero" must the mean value of a test be in order to be certain that the drug was present in a particular sample? A similar question comes up very frequently in environmental pollution studies.

How to Lie with Statistics

How to lie with statistics is the title of an amusing book by Darrell Huff (Norton, 1954). Some of Irving Geiss’s illustrations for this book appear below. See also

Throwing away “wrong” answers.

It occasionally happens that a few data values are so greatly separated from the rest that they cannot reasonably be regarded as representative. If these “outliers” clearly fall outside the range of reasonable statistical error, they can usually be disregarded as likely due to instrumental malfunctions or external interferences such as mechanical jolts or electrical fluctuations.

Some care must be exercised when data is thrown away however; There have been a number of well-documented cases in which investigators who had certain anticipations about the outcome of their experiments were able to bring these expectations about by removing conflicting results from the data set on the grounds that these particular data “had to be wrong”

Beware of too-small samples

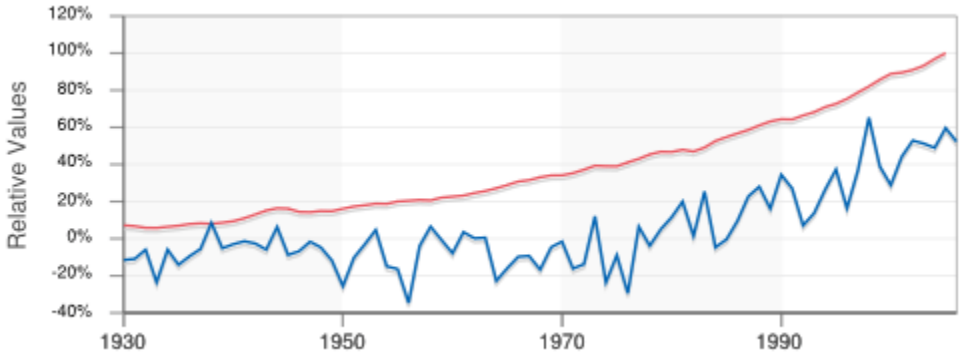
The probability of ten successive flips of a coin yielding 8 heads is given by



... indicating that it is not very likely, but can be expected to happen about eight times in a thousand runs. But there is no law of nature that says it cannot happen on your first run, so it would clearly be foolish to cry “Eureka” and stop the experiment after one— or even a few tries. Or to forget about the runs that did not turn up 8 heads!

Perils of dubious "correlations"

The fact that two sets of statistics show the same trend does not prove they are connected, even in cases where a logical correlation could be argued. Thus it has been suggested that according to the two plots below, "In relative terms, the global temperature seems to be tracking the average global GDP quite nicely over the last 70 years."



The difference between confidence levels of 90% and 95% may not seem like much, but getting it wrong can transform science into junk science — a not-unknown practice by special interests intent on manipulating science to influence public policy; see the excellent 2008 book by David Michaels "*Doubt is Their Product: How Industry's Assault on Science Threatens Your Health*".

This page titled 3.5: Drawing Conclusions from Data is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

4: THE BASICS OF CHEMISTRY

The chapters in this unit are absolutely essential for anyone embarking on the serious study of Chemistry. The material covered here will be needed in virtually every topic you will encounter in the remainder of your first-year course, as well as in subsequent Chemistry courses — so you might as well master it now!

4.1: Atoms, Elements, and the Nucleus

4.2: Avogadro's Number and the Mole

4.3: Formulas and Their Meaning

4.4: Chemical Equations and Stoichiometry

4.5: Introduction to Chemical Nomenclature

4.6: Significant Figures and Rounding

This page titled 4: The Basics of Chemistry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Atoms, Elements, and the Nucleus

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Give a chemical definition of *element,* and comment on the distinction between the terms *atom* and *element*.
- You should know the *names and symbols* of the more common elements, including those whose symbols are derived from their Latin names.
- Describe, in your own words, the *Laws of Chemical Change*: mass conservation, constant composition, and multiple proportions.
- Explain how these laws follow from *Dalton*'s *atomic theory*.
- Describe Rutherford's alpha-ray scattering experiment and how it led to the present model of the atom.
- Define *atomic number* and *mass number*, and explain the relation between them.
- Define *isotope* and *nuclide*, and write the symbol for a nuclide of a given element with a given number of neutrons.
- Explain the purpose of a *mass spectrometer* and its general principle of operation.
- Describe the *atomic weight scale*.
- Find the *molecular weight* or *formula weight* from a chemical formula.
- Define the *unified atomic mass unit*, and write out the mass numbers of the proton, neutron, and electron.

The parallel concepts of the *element* and the *atom* constitute the very foundations of chemical science. As such, the concept of the element is a *macroscopic* one that relates to the world that we can observe with our senses. The *atom* is the *microscopic* realization of this concept; that is, it is the *actual* ***physical particle*** that is unique to each chemical element. Their very small size has long prevented atoms from being observable by direct means, so their existence was not universally accepted until the late 19th Century. The fact that we still hear the mention of the "*atomic* ***theory*** of matter" should not imply that there is now any doubt about the existence of atoms. Few theories in the history of science have been as thoroughly validated and are as well understood.

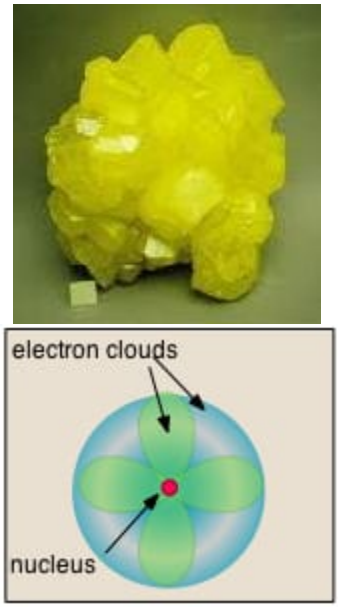


Figure 4.1.1: Sulfur the element (left) and Sulfur the atom (right)

Although the word *atom* usually refers to a specific *kind* of particle (an "atom of magnesium", for example), our everyday use of *element* tends to be more general, referring not only to a substance composed of a particular type of atom ("bromine is one of the few elements that are liquids at room temperature"), but also to atoms in a collective sense ("magnesium is one of the elements having two electrons in its outer shell").

The underlying concept of atoms as the basic building blocks of matter has been around for a long time. As early as 600 BCE, the Gujarati (Indian) philosopher Acharya Kanad wrote that "*Every object of creation is made of atoms which in turn connect with each other to form molecule*s". A couple of centuries later in 460 BCE, the Greek philosopher Democritus reasoned that if you keep breaking a piece of matter into smaller and smaller fragments, there will be some point at which the pieces cannot be made any smaller. He called these "basic matter particles"— in other words, atoms. But this was just *philosophy*; it would not become *science* until 1800 when John Dalton showed how the atomic concept followed naturally from the results of quantitative experiments based on weight measurements.

Elements

The *element* is the fundamental unit of chemical identity. The concept of the element is very ancient. It was developed in many different civilizations in an attempt to rationalize the variety of the world and to understand the nature of change, such as, change that occurs when a piece of wood rots, or is burnt to produce charcoal or ash. Most well known to us are the four elements "earth, air, fire and water" that were popularized by Greek philosophers (principally Empedocoles and Aristotle) in the period 500-400 BCE.

To these, Vedic (Hindu) philosophers of India added s*pace*, while the ancient Chinese concept of *Wu Xing* regarded earth, metal, wood, fire and water as fundamental. These basic elements were not generally considered to exist as the actual materials we know as earth, water, etc., but rather to represent the "principles" or *essence*s that the elements conveyed to the various kinds of matter we encounter in the world.

Eventually, practical experience (largely connected with the extraction of metals from ores) and the beginnings of scientific experimentation in the 18th Century led to our modern concept of the chemical element. An element is a s*ubstance*: the simplest form to which any other chemical substance can be reduced through appropriate thermal or chemical treatment. "Simplest", in the context of experimentation at the time, was defined in terms of weight; cinnabar (mercuric sulfide) can be broken down into two substances, mercury and sulfur, which themselves cannot be reduced to any lighter forms.



Figure 4.1.2: Antoine Lavoisier (1743-1794)

Although Lavoisier got many of these right, he did manage to include a few things that do not quite fit into our modern idea of what constitutes a chemical element. There are two such mistakes in the top section of the table that you should be able to identify even if your French is less than tip-top— can you find them?

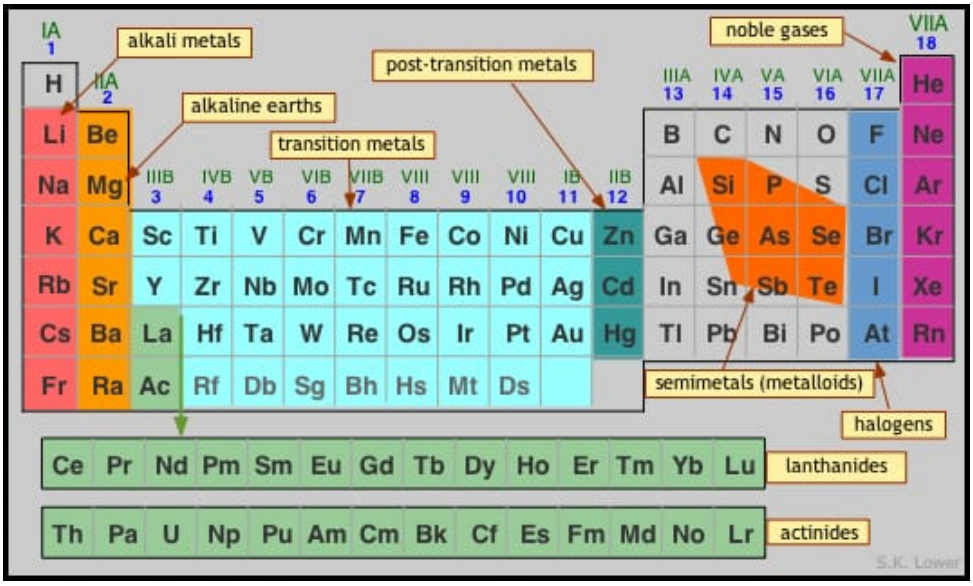
Lavoisier's other misassignment of the elements in the bottom section was not really his fault. Chalk, magnesia, barytes, alumina and silica are highly stable oxygen-containing compounds; the high temperatures required to break them down could not be achieved in Lavoisier's time (magnesia is what fire brick is made of). The proper classification of these substances was delayed until further experimentation revealed their true nature. Ten of the chemical elements have been known since ancient times and five more were discovered through the 17th Century.

- **How many elements are there?** Ninety-two elements have been found in nature. Around 25 more have been made

artificially, but all of these decay into lighter elements, with some of them disappearing in minutes or even seconds.

- **Where do the elements come from?** The present belief is that helium and a few other very light elements were formed within about three minutes of the "big bang", and that the next 23 elements (up through iron) are formed mostly by nuclear fusion processes within stars, in which lighter nuclei combine into successively heavier elements. Elements heavier than iron cannot be formed in this way, and are produced only during the catastrophic collapse of massive stars (supernovae explosions).
- **How do the elements vary in abundance?** Quite markedly, and very differently in different bodies in the cosmos. Most of the atoms in the universe still consist of hydrogen, with helium being a distant second. On Earth, oxygen, silicon, and aluminum are most abundant. These profiles serve as useful guides for constructing models for the formation of the earth and other planetary bodies.
  The system of element symbols we use today was established by the Swedish chemist Jons Jacob Berzelius in 1814. Prior to that time, graphical alchemical symbols were used, which were later modified and popularized by John Dalton. Fortunately for English speakers, the symbols of most of the elements serve as mnemonics for their names, but this is not true for the seven metals known from antiquity, whose symbols are derived from their Latin names. The other exception is *tungsten* (a name derived from Swedish), whose symbol W reflects the German name which is more widely used.
- **How are the elements organized?** Two general organizing principles developed in the 19th Century: one was based on the increasing relative weights (atomic weights) of the elements, yielding a list that begins this way:

H He Li Be B C N O F Ne Na Mg Al Si P S Cl Ar K Ca...



Atoms become real

Throughout most of history the idea that matter is composed of minute particles had languished as a philosophical abstraction known as atomism, and no clear relation between these "atoms" and the chemical "elements" had been established. This began to change in the early 1800's when the development of balances that permitted reasonably precise measurements of the weight changes associated with chemical reactions ushered in a new and fruitful era of experimental chemistry. This resulted in the recognition of several *laws of chemical change* that laid the groundwork for the atomic theory of matter.



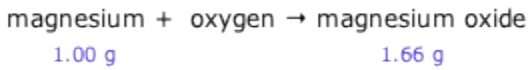
Figure 4.1.3: The word "atom" comes from the Greek a-tomos, meaning "un-cut-able"

Laws of Chemical Change

Recall that a "law", in the context of science, is just a relationship, discovered through experimentation, that is sufficiently well established to be regarded as beyond question for most practical purposes. Because it is the nature of scientists to question the "unquestionable", it occasionally happens that exceptions do arise, in which case the law must undergo appropriate modification.

Conservation of mass-energy is usually considered the most fundamental of law of nature. It is also a good example of a law that had to be modified; it was known simply as *Conservation of Mass* until Einstein showed that energy and mass are interchangeable. However, the older term is perfectly acceptable within the field of ordinary chemistry in which energy changes are too small to have a measurable effect on mass relations. Within the context of chemistry, conservation of mass can be thought of as "conservation of atoms". Chemical change just shuffles them around into new arrangements.

Mass conservation had special significance in understanding chemical changes involving gases, which were for some time not always regarded as real matter at all. (Owing to their very small densities, carrying out actual weight measurements on gases is quite difficult to do, and was far beyond the capabilities of the early experimenters.) Thus when magnesium metal is burned in air, the weight of the solid product always exceeds that of the original metal, implying that the process is one in which the metal combines with what might have been thought to be a "weightless" component of the air, which we now know to be oxygen.



More importantly, this experimental result tells us something very important about the mass of the oxygen atom relative to that of the magnesium atom.

The **Law of Definite Proportions,** also known as the *law of constant composition*, states that the proportion by weight of the element present in any pure substance is always the same. This enables us to generalize the relationship we illustrated above.

Example 4.1.1: Law of Definite Proportions

How many kilograms of metallic magnesium could theoretically be obtained by decomposing 0.400 kg of magnesium oxide into its elements?

Solution

The mass ratio of Mg to O in this compound is

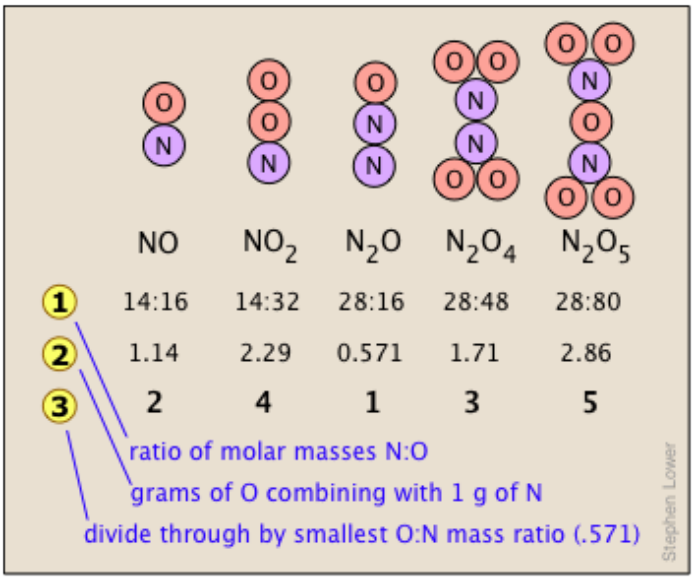
$$\frac{1}{1.66} = 0.602$$

so 0.400 kg of the oxide contains

$$(0.400\ kg) \times 0.602 = 0.241 \text{ kg of Mg.}$$

The fact that we are concerned with the *reverse* of the reaction cited above is irrelevant.

The **Law of Multiple Proportions** address the fact that Many combinations of elements can react to form more than one compound. In such cases, this law states that the weights of one element that combine with a fixed weight of another of these elements are integer multiples of one another. It's easy to say this, but please make sure that you understand how it works. Nitrogen forms a very large number of oxides, five of which are shown here.



- Line 1 shows the ratio of the relative weights of the two elements in each compound. These ratios were calculated by simply taking the molar mass of each element, and multiplying by the number of atoms of that element per mole of the compound. Thus for NO_2, we have (1 × 14) : (2 × 16) = 13:32. (These numbers were not known in the early days of Chemistry because atomic weights (i.e., molar masses) of most elements were not reliably known.)
- The numbers in Line 2 are just the mass ratios of O:N, found by dividing the corresponding ratios in line 1. But someone who depends solely on experiment would work these out by finding the mass of O that combines with unit mass (1 g) of nitrogen.
- Line is obtained by dividing the figures the previous line by the smallest O:N ratio in the line above, which is the one for N_2O. Note that just as the law of multiple proportions says, the weight of oxygen that combines with unit weight of nitrogen work out to small integers.
- Of course we just as easily could have illustrated the law by considering the mass of nitrogen that combines with one gram of oxygen; it works both ways!

Example 4.1.2: Law of Multiple Proportions

Nitrogen and hydrogen form many compounds, some of which involve other elements as well. The mass of hydrogen that combines with 1.00 g of nitrogen to form three of these compounds are: urea, 0.1428 g; ammonia, 0.0714 g; ammonium chloride, 0.2857 g. Show that this data is consistent with the Law of Multiple Proportions.

Solution

The "fixed weight" we are considering here is the nitrogen. Inspection of the numbers above shows that the ammonia contains the smallest weight ratio H:N = 0.0714, while the weight ratio of H:N in urea is twice this number, and that in ammonium chloride is four times 0.0714. Thus the H:N ratios are themselves stand in the ratio of 2:1:4, respectively, and the Law is confirmed.

Dalton's Interpretation Established Atomic Theory

The idea that matter is composed of tiny "atoms" of some kind had been around for at least 2000 years. Dalton's accomplishment was to identify atoms with actual chemical elements.



If Nobel prizes had existed in the early 1800's, the English school teacher/ meteorologist/ chemist John Dalton (1766-1844) would certainly have won one for showing how the experimental information available at that time, as embodied in the laws of chemical change that we have just described, are fully consistent with the hypothesis that atoms are the smallest units of chemical identity. These points of Dalton's atomic theory provided satisfactory explanations of all the laws of chemical change noted above:

Dalton's explanation of the Law of Conservation of Mass was that it is really a consequence of "*conservation of atoms*" which are presumed to be indestructible by chemical means. In chemical reactions, the atoms are simply rearranged, but never destroyed.

Dalton's Explanation of the law of constant composition was that if compounds are made up of definite numbers of atoms, each of which has its own characteristic mass, then the relative mass of each element in a compound must always be the same. Thus the elements **must** always be present in a pure sample of a compound in the same proportions by mass.

A given set of elements can usually form two or more compounds in which the numbers of atoms of some of the elements are different. Because these numbers must be integers (you can't have "half" an atom!), the mass of one element combined with a fixed mass of any other elements in any two such compounds can differ only by integer numbers. Thus, for the series of nitrogen-hydrogen compounds cited in the Problem Example above, we have the following relations:

Compound	Formula	weight ratio H:N	ratio to 0.0714
urea	$CO(NH_2)_2$	0.1428	2
ammonia	NH_3	0.0714	1
ammonium chloride	NH_4Cl	0.2857	4

Although Dalton's atomic theory was immediately found to be a useful tool for organizing chemical knowledge, it was some time before it became accepted as a true representation of the world. Thus, as late as 1887, one commentator observed

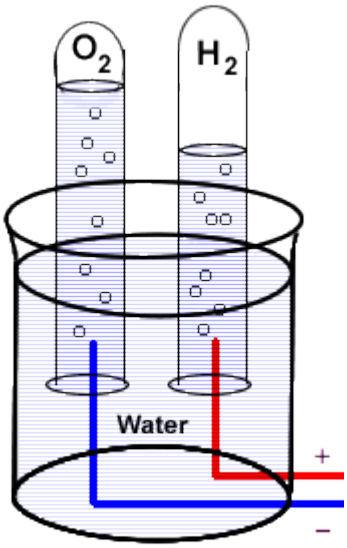
"Atoms are round bits of wood invented by Mr. Dalton."

These wooden balls have evolved into computer-generated images derived from the *atomic force microscope* (AFM), an exquisitely sensitive electromechanical device in which the distance between the tip of a submicroscopic wire probe and the surface directly below it is recorded as the probe moves along a surface to which atoms are adsorbed. The general principle of the AFM is quite simple, but its realization in an actual device can appear somewhat intimidating! This highly specialized atomic force microscope is one of several similar devices described at this Argonne National Laboratory.

Relative Masses

Dalton's atomic theory immediately led to the realization that although atoms are far too small to be studied directly, their *relative masses* can be estimated by observing the weights of elements that combine to form similar compounds. These weights are sometimes referred to as *combining weights*. There is one difficulty, however: we need to know the *formulas* of the compounds we are considering in order to make valid comparisons. For example, we can find the relative masses of two atoms X and Y that combine with oxygen only if we assume that the values of *n* in the two formulas XO_n and YO_n are the same. But the very relative masses we are trying to find must be known in order to determine these formulas.

The way to work around this was to focus on binary (two-element) compounds that were assumed to have simple atom ratios such as 1:1, 1:2, etc., and to hope that enough 1:1 compounds would be found to provide a starting point for comparing the various pairs of combining weights. Compounds of oxygen, known as oxides, played an especially important role here, partly because almost all of the chemical elements form compounds with oxygen, and most of them do have very simple formulas.



--

The first proof that water is composed of hydrogen and oxygen was because of the discovery, in 1800, that an electric current could decompose water into these elements. Notice the 2:1 volumes of the two gases displacing the water at the tops of the tubes.

Of these oxygen compounds, the one with hydrogen — ordinary *water* — had been extensively studied. Earlier experiments had given the composition of water is 87.4 percent oxygen and 12.6 percent hydrogen by weight. This means that if the formula of water is assumed [incorrectly] to be HO, then the mass ratio of the two kinds of atoms must be O:H = 87.4/12.6 = 6.9. Later work corrected this figure to 8, but the wrong assumption about the formula of water would remain to plague chemistry for almost fifty years until studies on gas volumes proved that water is H_2O.

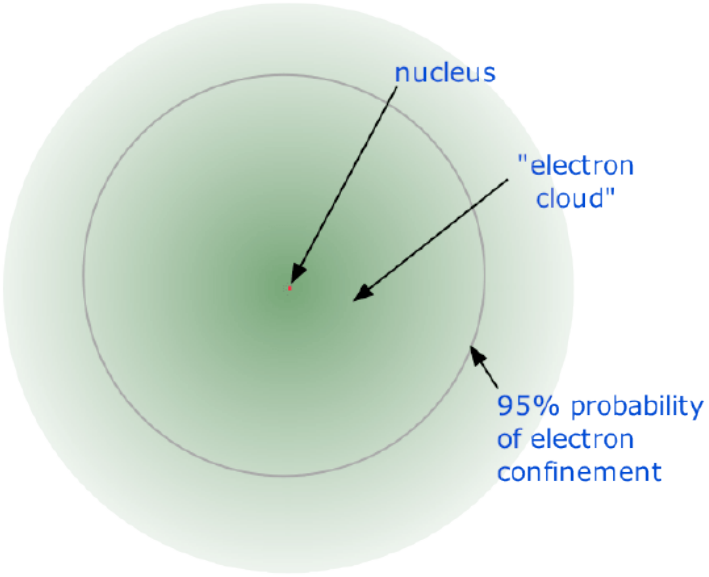
Dalton fully acknowledged the tentative nature of weight ratios based on assumed simple formulas such as HO for water, but was nevertheless able to compile in 1810 a list of the relative weights of the atoms of some of the elements he investigated by observing weight changes in chemical reactions.

hydrogen	nitrogen	carbon	oxygen	phosphorus	sulfur	iron	zinc	copper	lead
1	5	5.4	7	9	13	50	56	56	95

Because hydrogen is the lightest element, it was assigned a relative weight of unity. By assigning definite relative masses to atoms of the different elements, Dalton had given reality to the concept of the atom and established the link between atom and element. Once the correct chemical formulas of more compounds became known, more precise combining-weight studies eventually led to the relative weights of the atoms we know today as the *atomic weights*, which we discuss farther on.

The Nuclear atom

The precise physical nature of atoms finally emerged from a series of elegant experiments carried out between 1895 and 1915. The most notable of these achievements was Ernest Rutherford's famous 1911 alpha-ray scattering experiment, which established that



- Almost all of the *mass* of an atom is contained within a tiny (and therefore extremely dense)*nucleus* which carries a positive electric charge whose value identifies each element and is known as the *atomic number* of the element.
- Almost all of the *volume* of an atom consists of empty space in which electrons, the fundamental carriers of negative electric charge, reside. The extremely small mass of the electron (1/1840 the mass of the hydrogen nucleus) causes it to behave as a quantum particle, which means that its location at any moment cannot be specified; the best we can do is describe its behavior in terms of the probability of its manifesting itself at any point in space. It is common (but somewhat misleading) to describe the volume of space in which the electrons of an atom have a significant probability of being found as the *electron cloud*. The latter has no definite outer boundary, so neither does the atom. The radius of an atom must be defined arbitrarily, such as the boundary in which the electron can be found with 95% probability. Atomic radii are typically 30-300 pm.

Protons and Neutrons

The nucleus is itself composed of two kinds of particles. *Protons* are the carriers of positive electric charge in the nucleus; the proton charge is exactly the same as the electron charge, but of opposite sign. This means that in any [electrically neutral] atom, the number of protons in the nucleus (often referred to as the *nuclear charge*) is balanced by the same number of electrons outside the nucleus.

Because the electrons of an atom are in contact with the outside world, it is possible for one or more electrons to be lost, or some new ones to be added. The resulting electrically-charged atom is called an ion.

The other nuclear particle is the *neutron*. As its name implies, this particle carries no electrical charge. Its mass is almost the same as that of the proton. Most nuclei contain roughly equal numbers of neutrons and protons, so we can say that these two particles together account for almost all the mass of the atom.

Atomic Number (Z)

What single parameter uniquely characterizes the atom of a given element? It is not the atom's relative mass, as we will see in the section on isotopes below. It is, rather, the number of protons in the nucleus, which we call the *atomic number* and denote by

the symbol *Z*. Each proton carries an electric charge of +1, so the atomic number also specifies the electric charge of the nucleus. In the neutral atom, the *Z protons* within the nucleus are balanced by *Z electrons* outside it.



Atomic numbers were first worked out in 1913 by Henry Moseley, a young member of Rutherford's research group in Manchester.

Moseley searched for a measurable property of each element that increases linearly with atomic number. He found this in a class of X-rays emitted by an element when it is bombarded with electrons. The frequencies of these X-rays are unique to each element, and they increase uniformly in successive elements. Moseley found that the square roots of these frequencies give a straight line when plotted against Z; this enabled him to sort the elements in order of increasing atomic number.

You can think of the atomic number as a kind of serial number of an element, commencing at 1 for hydrogen and increasing by one for each successive element. The chemical name of the element and its symbol are uniquely tied to the atomic number; thus the symbol "Sr" stands for strontium, whose atoms all have *Z* = 38.

Mass number (A)

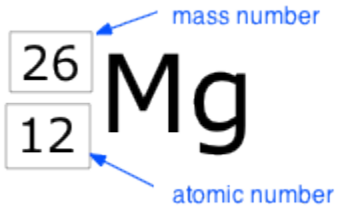
This is just the sum of the numbers of protons and neutrons in the nucleus. It is sometimes represented by the symbol *A*, so

$$A = Z + N \tag{4.1.1}$$

in which *Z* is the atomic number and *N* is the *neutron number*.

Nuclides and their Symbols

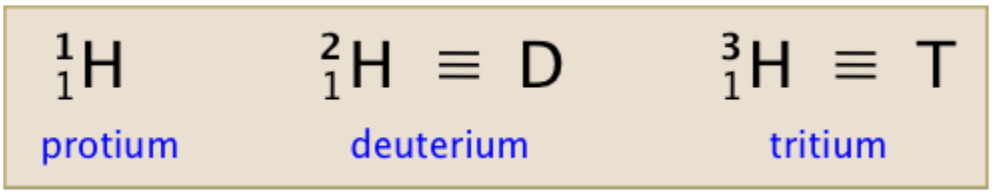
The term *nuclide* simply refers to any particular kind of nucleus. For example, a nucleus of atomic number 7 is a nuclide of nitrogen. Any nuclide is characterized by the pair of numbers (*Z* ,*A*). The element symbol depends on *Z* alone, so the symbol ^{26}Mg is used to specify the mass-26 nuclide of magnesium, whose name implies *Z*=12. A more explicit way of denoting a particular kind of nucleus is to add the atomic number as a subscript. Of course, this is somewhat redundant, since the symbol Mg **always** implies *Z*=12, but it is sometimes a convenience when discussing several nuclides.



Two nuclides having the **same** atomic number but different mass numbers are known as *isotopes*. Most elements occur in nature as mixtures of isotopes, but twenty-three of them (including beryllium and fluorine, shown in the table) are monoisotopic. For example, there are three *natural isotopes* of magnesium: ^{24}Mg (79% of all Mg atoms), ^{25}Mg (10%), and ^{26}Mg (11%); all three are present in all compounds of magnesium in about these same proportions.

	Z	mass numbers
H	1	1 2 3
He	2	3 4
Li	3	6 7
Be	4	9
B	5	10 11
C	6	12 13 14
N	7	14 15
O	8	16 17 18
F	9	19
Ne	10	20 21 22

Approximately 290 isotopes occur in nature. The two heavy isotopes of hydrogen are especially important— so much so that they have names and symbols of their own:



Deuterium accounts for only about 15 out of every one million atoms of hydrogen. Tritium, which is radioactive, is even less abundant. All the tritium on the earth is a by-product of the decay of other radioactive elements.

Atomic Weights

Atoms are of course far too small to be weighed directly; weight measurements can only be made on the massive (but unknown) numbers of atoms that are observed in chemical reactions. The early combining-weight experiments of Dalton and others established that hydrogen is the lightest of the atoms, but the crude nature of the measurements and uncertainties about the formulas of many compounds made it difficult to develop a reliable scale of the relative weights of atoms. Even the most exacting weight measurements we can make today are subject to experimental uncertainties that limit the precision to four significant figures at best.

Weighing atoms: Mass Spectrometry

An alternative way of examining the behavior of individual atomic particles became evident in 1912, when J.J. Thomson and F.W. Aston showed that a stream of gaseous neon atoms, broken up by means of an electrical discharge, yielded two kinds of subatomic particles having opposite electrical charges, as revealed by their deflections in externally-applied magnetic and electrostatic fields. (The deflections themselves could be observed by the spots the particles made when they impinged on a photographic plate.) This, combined with the finding made a year earlier by Wilhelm Wien that the degree of deflection of a particle in these fields is proportional to the ratio of its electric charge to its mass, opened the way to characterizing these otherwise invisible particles.

Neutral atoms, having no charge, cannot be accelerated along a path so as to form a beam, nor can they be deflected. They can, however, be made to acquire electric charges by directing an electron beam at them, and this was the basis of the first mass spectrometer developed by Thomson's former student F.W. Aston (1877-1945, 1922 Nobel Prize) in 1919. This enabled him to quickly identify 212 of the 287 naturally occurring isotopes.

The mass spectrometer has become one of the most widely used laboratory instruments. Mass spectrometry is now mostly used to identify molecules. Ionization usually breaks a molecule up into fragments having different charge-to-mass ratios, each molecule resulting in a unique "fingerprint" of particles whose origin can be deduced by a jigsaw puzzle-like reconstruction. For many years, "mass-spec" had been limited to small molecules, but with the development of novel ways of creating ions from molecules, it has now become a major tool for analyzing materials and large biomolecules, including proteins.

The scale of relative weights (the *atomic weight scale*) we now use is based on $^{12}_{6}C$, whose relative mass is defined as exactly

12. Atomic weights are the ratios of the weights of an element to the weight of an identical number of $^{12}_{6}C$ atoms. Being ratios, **atomic weights are dimensionless**.

From 1850 to 1961, the atomic weight scale was defined relative to oxygen = 16.

Example 4.1.3: A Zillion Atoms

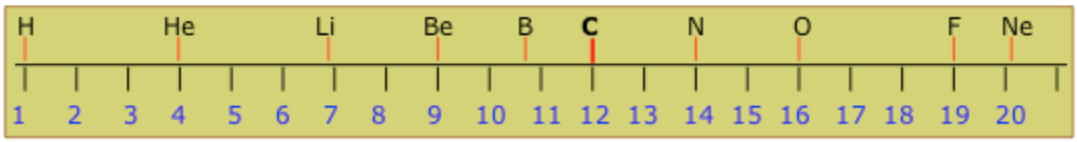
A certain number (call it "one zillion") of oxygen atoms weighs 1.200 g. What will be the weight of an equal number of lithium atoms?

Solution

From the atomic weight table, the mass ratio of Li/O = 6.94/16.00, so the weight of one zillion lithium atoms will be

$$(1.200\ g) \times \frac{6.94}{16.00} = 0.570\ g$$

You can visualize the atomic weight scale as a long line of numbers that runs from 1 to around 280. The beginning of the scale looks like this:



You will notice that the relative masses of the different elements (shown in the upper part) are not all integers. If the nuclei all differ by integral numbers of protons and neutrons that have virtually identical masses, we would expect the atomic weights to be integers. Some are very close to integers (the reason they are not exactly integral will be explained in the next section), but many are nowhere near integral. This puzzling observation eventually led to the concept of *isotopes*.

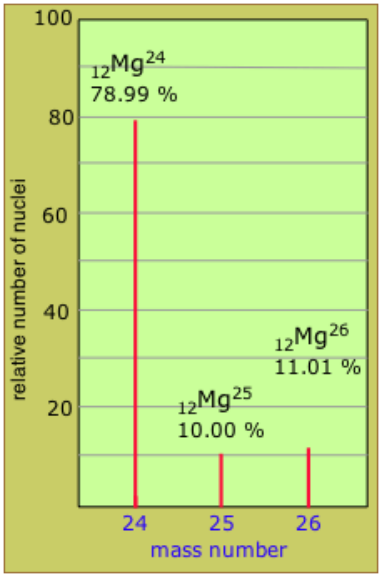


Figure 4.1.4: Aston's first experiments with his newly-invented mass spectrometer in 1919 immediately revealed that most of the elements exist in nature as mixtures of isotopes. For example, the mass spectrum of magnesium shows three isotopes of masses 24 through 26 having the natural abundances shown on the plot. Of the 92 natural elements, 23 have only a single isotope, and are said to be *monoisotopic*.

The atomic weights that are determined experimentally and listed in tables are *weighted averages* of these isotopic mixtures.

Example 4.1.4: Estimating Average Atomic Weight

Estimate the average atomic weight of magnesium from the isotopic abundance data shown in the above mass spectrometry plot.

Solution

We just take the weighted average of the mass numbers:

$$(0.7899 \times 24) + (0.1000 \times 25) + (0.1101 \times 26) = \mathbf{24.32}$$

Note: The measured atomic weight of Mg (24.305) is slightly smaller than this because atomic masses of nuclear components are not strictly additive, as will be explained further below.

When there are only two significantly abundant isotopes, you can estimate the relative abundances from the mass numbers and the average atomic weight. The following is a favorite exam problem:

Example 4.1.5: Relative Abundances

The average atomic weight of chlorine is 35.45 and the element has two stable isotopes $^{35}_{17}\mathrm{Cl}$ and $^{37}_{17}\mathrm{Cl}$. Estimate the relative abundances of these two isotopes.

Solution

Here you finally get to put your high-school algebra to work! If we let x represent the fraction of $^{35}\mathrm{Cl}$, then (1-x) gives the fraction of $^{37}\mathrm{Cl}$. The weighted average atomic weight is then

$$35x + 37(1\text{-}x) = 35.45$$

Solving for x gives $2x = 1.55$, $x = 0.775$, so the abundances are 77.5% Cl^{35} and 22.5% Cl^{37}.

Example 4.1.6: Mass Spectra

Elemental chlorine, Cl_2, is made up of the two isotopes mentioned in the previous example. How many peaks would you expect to observe in the mass spectrum of Cl_2?

Solution

The mass spectrometer will detect a peak for each possible combination of the two isotopes in dichlorine: ^{35}Cl-^{35}Cl, ^{35}Cl-^{37}Cl, and ^{37}Cl-^{37}Cl.

Tables of Atomic Weights

are updated every few years as better data becomes available.

	atomic number	atomic weight		atomic number	atomic weight
H	1	1.00794	C	6	12.0107
He	2	4.00262	N	7	14.0067
Li	3	6.941	O	8	15.9994
Be	4	9.0121821	F	9	18.998403
B	5	10.811	Ne	10	20.1797

One peculiarity you might notice is that the number of significant figures varies from element to element. It tends to be highest for monoisotopic elements, as you can see here for beryllium and fluorine. For some elements, the isotopic abundances vary slightly, depending on the source; this variance reduces the useful precision of a value.

Atomic weights, molecular weights and formula weights

Molecules are composed of atoms, so a *molecular weight* is just the sum of the atomic weights of the elements it contains.

Example 4.1.7: Molecular Weight of Sulfuric Acid

What is the molecular weight of sulfuric acid, H_2SO_4?

Solution

The atomic weights of hydrogen and of oxygen are 1.01 and 16.00, respectively (you should have these common values memorized.) From a table, you can find that the atomic weight of sulfur is 32.06. Adding everything up, we have

$$(2 \times 1.01) + 32.06 + (4 \times 16.00) = 98.08 \tag{4.1.2}$$

Because some solids are not made up of discrete molecules (sodium chloride, NaCl, and silica, SiO_2 are common examples), the term *formula weight* is often used in place of molecular weight. In general, the terms molecular weight and formula weight are interchangeable.

Isotopic Fractionation

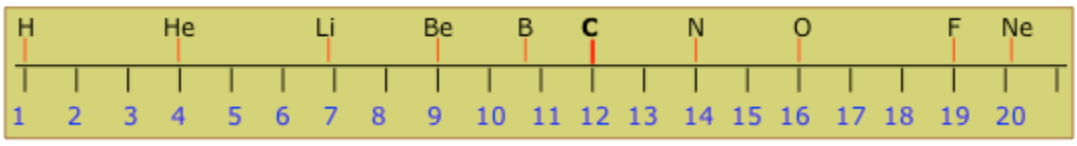
The isotopes of a given element are so similar in their chemical behavior that what small differences may exist can be considered negligible for most practical purposes. However, heavier isotopes do tend to react or evaporate slightly more slowly than lighter ones, so that given enough time, various geochemical processes can result in an enrichment of one isotope over the other, an effect known as geochemical isotopic fractionation.

What differences do exist are most evident in the lighter elements, and especially in hydrogen, whose three isotopes differ in mass by relatively large amounts. Thus "heavy water", D_2O (^2H_2O) is not decomposed by electrolysis quite as rapidly as is ^1H_2O, so it becomes enriched in the the un-decomposed portion of the water in an electrolysis apparatus. Its boiling point is 101.7°C and it freezes at 3.8°. Animals will die if they drink heavy water in place of ordinary water.

The minute differences between the behaviors of most isotopes constitute an invaluable tool for research in geochemistry. For example, the tiny fraction of water molecules containing O^{18} evaporates more slowly than the lighter (and far more abundant) H_2O^{16}. But the ratio of O^{18} to O^{16} in the water that evaporates depends on the temperature at which this process occurs. By observing this ratio in glacial ice cores and in marine carbonate deposits, it is possible to determine the average temperature of the earth at various times in the past.

Atomic masses

Here again is the beginning of the atomic weight scale that you saw above:



You understand by now that atomic weights are relative weights, based on a scale defined by $_6C^{12}$ = 12. But what is the *absolute* weight of an atom, expressed in grams or kilograms? In other words, what actual mass does each unit on the atomic weight scale represent?

The answer is $1.66053886 \times 10^{-27}$ kg. This quantity (whose value you do *not* need to memorize) is known as the *unified atomic mass unit*, denoted by the abbreviation *u* or *amu*. The unified atomic mass unit is defined as 1/12 of the mass of one atom of carbon-12. Fortunately, you do not need to memorize this value, because you can easily calculate its value from Avogadro's number,N_A, which you *are* expected to know:

$$1\,u = \frac{1}{N_A}\,g = \frac{1}{1000\,N_A}\,Kg \tag{4.1.3}$$

The unified atomic mass unit is defined as 1/12 of the mass of one atom of carbon-12.

Masses of the Subatomic Particles

Atoms are composed of protons, neutrons, and electrons, whose properties are shown below:

Table 4.1.1*: Masses of Subatomic Particles*

particle	mass, g	mass, u	charge	symbol
electron	$9.1093897 \times 10^{-28}$	$5.48579903 \times 10^{-4}$	1–	${}^{0}_{-1}\mathrm{e}$
proton	$1.6726231 \times 10^{-24}$	1.007276470	1+	${}^{0}_{1}\mathrm{H}^{+}$ or ${}^{0}_{1}\mathrm{p}$
neutron	$1.6749286 \times 10^{-24}$	1.008664904	0	${}^{0}_{1}\mathrm{n}$

Two important points should be noted from Table 4.1.1:

- The mass of the electron is negligible compared to that of the two nuclear particles;
- The proton and neutron have masses that are almost, but not exactly, identical.

Nuclear Masses

As we mentioned in one of the problem examples above, the mass of a nucleus is always slightly different from the masses of the *nucleons* (protons and neutrons) of which it is composed. The difference, known as the *mass defect*, is related to the energy associated with the formation of the nucleus through Einstein's famous formula $e = mc^2$. This is the one instance in chemistry in which conservation of *mass-energy*, rather than of mass alone, must be taken into account. But there is no need for you to be concerned with this in this part of the course.

For all practical purposes, until you come to the section of the course on nuclear chemistry, you can consider that the proton and neutron have masses of about 1 *u*, and that the mass of an atom (in *u*) is just the sum of the neutron and proton numbers.

This page titled 4.1: Atoms, Elements, and the Nucleus is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

4.2: Avogadro's Number and the Mole

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Define *Avogadro's number* and explain why it is important to know.
- Define the *mole*. Be able to calculate the number of moles in a given mass of a substance, or the mass corresponding to a given number of moles.
- Define *molecular weight*, *formula weight*, and *molar mass*; explain how the latter differs from the first two.
- Be able to find the number of atoms or molecules in a given weight of a substance.
- Find the *molar volume* of a solid or liquid, given its density and molar mass.
- Explain how the molar volume of a metallic solid can lead to an estimate of atomic diameter.

The chemical changes we observe always involve *discrete numbers of atoms* that rearrange themselves into new configurations. These numbers are HUGE— far too large in magnitude for us to count or even visualize, but they are still *numbers*, and we need to have a way to deal with them. We also need a bridge between these numbers, which we are unable to measure directly, and the weights of substances, which we do measure and observe. The *mole concept* provides this bridge, and is central to all of quantitative chemistry.

Counting Atoms: Avogadro's Number

Owing to their tiny size, atoms and molecules cannot be counted by direct observation. But much as we do when "counting" beans in a jar, we can estimate the number of particles in a sample of an element or compound if we have some idea of the volume occupied by each particle and the volume of the container. Once this has been done, we know the number of formula units (to use the most general term for any combination of atoms we wish to define) in any arbitrary weight of the substance. The number will of course depend both on the formula of the substance and on the weight of the sample. However, if we consider a weight of substance that is the same as its formula (molecular) weight expressed in grams, we have only one number to know: **Avogadro's number**.

Avogadro's number is known to ten significant digits:

$$N_A = 6.022141527 \times 10^{23}. \tag{4.2.1}$$

However, you only need to know it to three significant figures:

$$N_A \approx 6.02 \times 10^{23}. \tag{4.2.2}$$

So 6.02×10^{23} of *what*? Well, of anything you like: apples, stars in the sky, burritos. However, the only *practical* use for N_A is to have a more convenient way of expressing the huge numbers of the tiny particles such as atoms or molecules that we deal with in chemistry. Avogadro's number is a *collective number*, just like a dozen. Students can think of 6.02×10^{23} as the "chemist's dozen".

Before getting into the use of Avogadro's number in problems, take a moment to convince yourself of the reasoning embodied in the following examples.

Example 4.2.1: Mass ratio from atomic weights

The atomic weights of oxygen and carbon are 16.0 and 12.0 atomic mass units (u), respectively. How much heavier is the oxygen atom in relation to carbon?

Solution

Atomic weights represent the relative masses of different kinds of atoms. This means that the atom of oxygen has a mass that is

$$\frac{16\ \cancel{u}}{12\ \cancel{u}} = \frac{4}{3} \approx 1.33$$

as great as the mass of a carbon atom.

Example 4.2.2: Mass of a single atom

The absolute mass of a carbon atom is 12.0 unified atomic mass units (u). How many grams will a single oxygen atom weigh?

Solution

The absolute mass of a carbon atom is 12.0 u or

$$12\ \cancel{u} \times \frac{1.6605 \times 10^{-24}\, g}{1\ \cancel{u}} = 1.99 \times 10^{-23}\, g\ (\text{per carbon atom})$$

The mass of the oxygen atom will be 4/3 greater (from Example 4.2.1):

$$\left(\frac{4}{3}\right) 1.99 \times 10^{-23}\, g = 2.66 \times 10^{-23}\, g\ (\text{per oxygen atom})$$

Alternatively we can do the calculation directly like with carbon:

$$16\ \cancel{u} \times \frac{1.6605 \times 10^{-24}\, g}{1\ \cancel{u}} = 2.66 \times 10^{-23}\, g\ (\text{per oxygen atom})$$

Example 4.2.3: Relative masses from atomic weights

Suppose that we have N carbon atoms, where N is a number large enough to give us a pile of carbon atoms whose mass is 12.0 grams. How much would the same number, N, of oxygen atoms weigh?

Solution

We use the results from Example 4.2.1 again. The collection of N oxygen atoms would have a mass of

$$\frac{4}{3} \times 12\, g = 16.0\, g.$$

Exercise 4.2.1

What is the numerical value of N in Example 4.2.3?

Answer available at the end of this section.

- It is a *number*, just as is "dozen", and thus is *dimensionless*.
- It is a *huge* number, far greater in magnitude than we can visualize
- Its practical use is limited to counting tiny things like atoms, molecules, "formula units", electrons, or photons.
- The value of N_A can be known only to the precision that the number of atoms in a measurable weight of a substance can be estimated. Because large numbers of atoms cannot be counted directly, a variety of ingenious indirect measurements have been made involving such things as Brownian motion and X-ray scattering.

- The current value was determined by measuring the distances between the atoms of silicon in an ultrapure crystal of this element that was shaped into a perfect sphere. (The measurement was made by X-ray scattering.) When combined with the measured mass of this sphere, it yields Avogadro's number. However, there are two problems with this:
  - The silicon sphere is an artifact, rather than being something that occurs in nature, and thus may not be perfectly reproducible.
  - The standard of mass, the kilogram, is not precisely known, and its value appears to be changing. For these reasons, there are proposals to revise the definitions of both N_A and the kilogram.

Moles and their Uses

The *mole* (abbreviated mol) is the the SI measure of *quantity of a "chemical entity"*, which can be an atom, molecule, formula unit, electron or photon. One mole of anything is just Avogadro's number of that something. Or, if you think like a lawyer, you might prefer the official SI definition:

Definition: The Mole

The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12

Avogadro's number (Equation 4.2.2) like any pure number, is dimensionless. However, it also defines the mole, so we can also express N_A as 6.02×10^{23} mol^{-1}; in this form, it is properly known as ***Avogadro's constant***. This construction emphasizes the role of Avogadro's number as a *conversion factor* between number of moles and number of "entities".

Example 4.2.4: number of moles in N particles

How many moles of nickel atoms are there in 80 nickel atoms?

Solution

$$\frac{80\ atoms}{6.02 \times 10^{23}\ atoms\ mol^{-1}} = 1.33 \times 10^{-22} mol$$

Is this answer reasonable? Yes, because 80 is an extremely small fraction of N_A.

Molar Mass

The atomic weight, molecular weight, or formula weight of one mole of the fundamental units (atoms, molecules, or groups of atoms that correspond to the formula of a pure substance) is the ratio of its mass to 1/12 the mass of one mole of C^{12} atoms, and being a ratio, is dimensionless. But at the same time, this *molar mass* (as many now prefer to call it) is also the observable mass of one mole (N_A) of the substance, so we frequently emphasize this by stating it explicitly as so many grams (or kilograms) per mole: g mol^{-1}.

> It is important always to bear in mind that the mole is a *number* and not a *mass*. But each individual particle has a mass of its own, so a mole of any specific substance will always *correspond to* a certain mass of that substance.

Example 4.2.5: Boron content of borax

Borax is the common name of sodium tetraborate, $Na_2B_4O_7$.

a. how many moles of boron are present in 20.0 g of borax?
b. how many grams of boron are present in 20.0 g of borax?

Solution

The formula weight of $Na_2B_4O_7$ so the molecular weight is:

$$(2 \times 23.0) + (4 \times 10.8) + (7 \times 16.0) = 201.2$$

a. 20 g of borax contains (20.0 g) ÷ (201 g mol^{-1}) = 0.10 mol of borax, and thus **0.40 mol** of B.
b. 0.40 mol of boron has a mass of (0.40 mol) × (10.8 g mol^{-1}) = **4.3 g**.

Example 4.2.6: Magnesium in chlorophyll

The plant photosynthetic pigment chlorophyll contains 2.68 percent magnesium by weight. How many atoms of Mg will there be in 1.00 g of chlorophyll?

Solution

Each gram of chlorophyll contains 0.0268 g of Mg, atomic weight 24.3.

- Number of moles in this weight of Mg: (.0268 g) / (24.2 g mol^{-1}) = 0.00110 mol
- Number of atoms: (0.00110 mol) × (6.02E23 mol^{-1}) = 6.64×10^{20}

Is this answer reasonable? (Always be suspicious of huge-number answers!) Yes, because we would expect to have huge numbers of atoms in any observable quantity of a substance.

Molar Volume

This is the volume occupied by one mole of a pure substance. Molar volume depends on the density of a substance and, like density, varies with temperature owing to thermal expansion, and also with the pressure. For solids and liquids, these variables ordinarily have little practical effect, so the values quoted for 1 atm pressure and 25°C are generally useful over a fairly wide range of conditions. This is definitely **not** the case with gases, whose molar volumes must be calculated for a specific temperature and pressure.

Example 4.2.7: Molar Volume of a Liquid

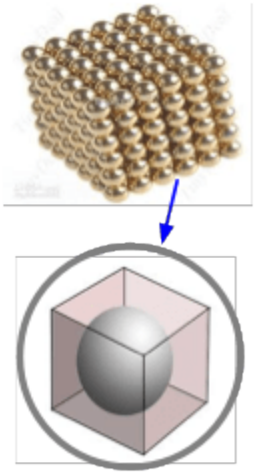
Methanol, CH_3OH, is a liquid having a density of 0.79 g per milliliter. Calculate the molar volume of methanol.

Solution

The molar volume will be the volume occupied by one molar mass (32 g) of the liquid. Expressing the density in liters instead of mL, we have

$$V_M = \frac{32\ g\ mol^{-1}}{790\ g\ L^{-1}} = 0.0405\ L\ mol^{-1}$$

The molar volume of a metallic element allows one to estimate the size of the atom. The idea is to mentally divide a piece of the metal into as many little cubic boxes as there are atoms, and then calculate the length of each box. Assuming that an atom sits in the center of each box and that each atom is in direct contact with its six neighbors (two along each dimension), this gives the diameter of the atom. The manner in which atoms pack together in actual metallic crystals is usually more complicated than this and it varies from metal to metal, so this calculation only provides an approximate value.



Example 4.2.8: Radius of a Strontium Atom

The density of metallic strontium is 2.60 g cm^{-3}. Use this value to estimate the radius of the atom of Sr, whose atomic weight is 87.6.

Solution

The molar volume of Sr is:

$$\frac{87.6\ g\ mol^{-1}}{2.60\ g\ cm^{-3}} = 33.7\ cm^3\ mol^{-1} \tag{4.2.3}$$

The volume of each "box" is"

$$\frac{33.7\ cm^3 mol^{-1}}{6.02 \times 10^{23}\ mol^{-1}} = 5.48 \times 10^{-23}\ cm^3 \tag{4.2.4}$$

The side length of each box will be the cube root of this value, $3.79 \times 10^{-8}\ \ cm$. The atomic radius will be half this value, or

$$1.9 \times 10^{-8}\ cm = 1.9 \times 10^{-10}\ m = 190pm \tag{4.2.5}$$

Note: Your calculator probably has no cube root button, but you *are* expected to be able to find cube roots; you can usually use the x^y button with y=0.333. You should also be able estimate the magnitude of this value for checking. The easiest way is to express the number so that the exponent is a multiple of 3. Take 54×10^{-24}, for example. Since 3^3=27 and 4^3 = 64, you know that the cube root of 55 will be between 3 and 4, so the cube root should be a bit less than 4 × 10^{-8}.

So how good is our atomic radius? Standard tables give the atomic radius of strontium is in the range 192-220 pm.

Answers

Exercise 4.2.1

Using the results of Examples 4.2.2 and 4.2.3.

$$N \times 1.99 \times 10^{-23}\ g\ \text{(per carbon atom)} = 12\ g$$

or

$$N = \frac{12\ \cancel{g}}{1.99 \times 10^{-23}\ \cancel{g}\ \text{(per carbon atom)}} = 6.03 \times 10^{23}\text{atoms}$$

There are a lot of atoms in 12 g of carbon.

This page titled 4.2: Avogadro's Number and the Mole is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

4.3: Formulas and Their Meaning

Learning Objectives

Make sure you thoroughly understand the following essential concepts that have been presented above.

- Explain why the s*ymbol* of an element often differs from the *formula* of the element..
- Define an *ion*, and explain the meaning of its formula.
- Find the *simplest ("empirical") formula* of a substance from a more complex molecular formula. Explain the meaning of the formula of an ionic solid such as NaCl.
- Define *molecular weight, formula weight*, and *molar mass*. Calculate any of these from any chemical formula.
- Given a chemical formula, express the *mole ratios* of any two elements, or the *mole fraction* of one of its elements.
- Find the *percentage composition* of a compound from its formula.
- Calculate the *mass ratio* of any two elements present in a compound from its formula.
- Find the empirical formula of a binary compound from the *mole ratio* of its two elements, expressed as a decimal number.
- Find the empirical formula of a binary compound from the *mass ratio* of its two elements.
- Find the empirical formula of a compound from its mass- or percentage composition.

At the heart of chemistry are *substances* — elements or compounds— which have a*definite composition* which is expressed by a *chemical formula*. In this unit you will learn how to write and interpret chemical formulas both in terms of moles and masses, and to go in the reverse direction, in which we use experimental information about the composition of a compound to work out a formula.

The *formula* of a compound specifies the number of each kind of atom present in one molecular unit of a compound. Since every unique chemical substance has a definite composition, every such substance must be describable by a chemical formula.

Example 4.3.1: Writing a Molecular Formula

The well-known alcohol *ethanol* is composed of molecules containing two atoms of carbon, five atoms of hydrogen, and one atom of oxygen. What is its molecular formula?

Solution

Just write the symbol of each element, following by a subscript indicating the number of atoms if more than one is present. Thus: **C_2H_5O**

Note that:

- The number of atoms of each element in a molecular formula is written as a subscript;
- When only a single atom of an element in a molecular formula is present, the subscript is omitted.
- In the case of organic (carbon-containing) compounds, it is customary to place the symbols of the elements C, H, (and if present,) O, N in this order in the formula.

Formulas of Elements and Ions

The *symbol* of an element is the one- or two-letter combination that represents the atom of a particular element, such as Au (gold) or O (oxygen). The symbol can be used as an *abbreviation* for an element name (it is easier to write "Mb" instead of "molybdenum"!) In more formal chemical use, an element symbol can also stand for one atom, or, depending on the context, for *one mole of atoms* of the element.

Some of the non-metallic elements exist in the form of molecules containing two or more atoms of the element. These molecules

are described by *formulas* such as N_2, S_6, and P_4. Some of these elements can form more than one kind of molecule; the best-known example of this is oxygen, which can exist as O_2 (the common form that makes up 21% of the molecules in air), and also as O_3, an unstable and highly reactive molecule known as *ozone*. The soccer-ball-shaped carbon molecules sometimes called *buckyballs* have the formula C_{60}.

Allotropes

Different molecular forms of the same element (such as O_2 and O_3) are called allotropes.

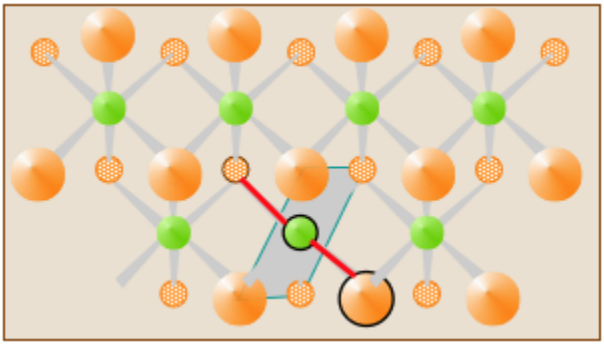
Ions are atoms or molecules that carry an electrical charge. These charges are represented as superscripts in the ionic formulas. Thus:

Cl^-	the chloride ion, with one negative charge per atom
S^{2-}	the sulfide ion carries two negative charges
$HCO_3{}^{2-}$	the carbonate ion— a molecular ion
$NH_4{}^+$	the ammonium ion

Note that the number of charges (in units of the electron charge) should always **precede** the positive or negative sign, but this number is omitted when the charge is ±1.

Formulas of Extended Solids

In solid $CdCl_2$, the Cl and Cd atoms are organized into sheets that extend indefinitely. Each atom is surrounded by six atoms of the opposite kind, so one can arbitrarily select any Cl–Cd–Cl as the "molecular unit". One such $CdCl_2$ unit is indicated by the two red-colored bonds in the diagram, but it does not constitute a discrete "molecule" of $CdCl_2$.



Many apparently "simple" solids exist only as ionic solids (such as NaCl) or as extended solids (such as $CuCl_2$) in which no discrete molecules can be identified. The formulas we write for these compounds simply express relative numbers of the different kinds of atoms in the compound in the smallest possible integer numbers. These are identical with the empirical or "simplest" formulas that we discuss further on.

Many minerals and most rocks contain varying ratios of certain elements and can only be precisely characterized at the structural level. Because these are usually not pure substances, the "formulas" conventionally used to describe them have limited meanings. For example the common rock olivine, which can be considered a solid solution of Mg_2SiO_4 and Fe_2SiO_4, can be represented by $(Mg,Fe)_2SiO_4$. This implies that the ratio of the metals to SiO_4 is constant, and that magnesium is usually present in greater amount than iron.

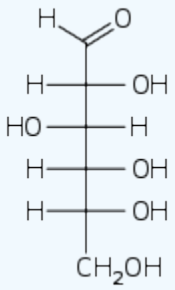
Empirical Formulas

Empirical formulas give the *relative* numbers of the different elements in a sample of a compound, expressed in the smallest possible integers. The term *empirical* refers to the fact that formulas of this kind are determined experimentally; such formulas

are also commonly referred to as *empirical formulas*.

Example 4.3.2: Empirical formula from molecular formula

Glucose (the "fuel" your body runs on) is composed of molecular units having the formula $C_6H_{12}O_6$. What is the empirical formula of glucose?



Solution

The glucose molecule contains twice as many atoms of hydrogen as carbons or oxygens, so we divide through by 6 to get **CH_2O**.

Note: this empirical formula, which applies to all 6-carbon sugars, indicates that these compounds are "composed" of carbon and water, which explains why sugars are known as *carbohydrates*.

Some solid compounds do not exist as discrete molecular units, but are built up as extended two- or three-dimensional lattices of atoms or ions. The compositions of such compounds are commonly described by their empirical formulas. In the very common case of *ionic solids*, such a formula also expresses the minimum numbers of positive and negative ions required to produce an electrically neutral unit, as in NaCl or $CuCl_2$.

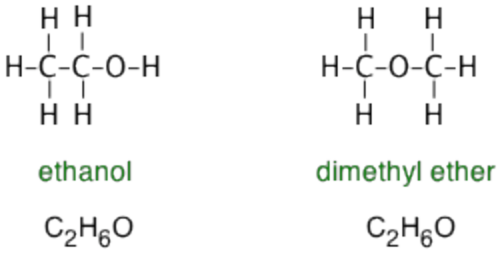
Example 4.3.3: Molecular formula from ionic charges

a. Write the formula of ferric bromide, given that the ferric (iron-III) ion is Fe^{3+} and the bromide ion carries a single negative charge.
b. Write the formula of bismuth sulfide, formed when the ions Bi^{3+} and S^{2-}combine.

Solution:

a. Three Br^- ions are required to balance the three positive charges of Fe^{3+}, hence the formula **$FeBr_3$**.
b. The only way to get equal numbers of opposite charges is to have six of each, so the formula will be **Bi_2S_3**.

The formulas we ordinarily write convey no information about the compound's *structure*— that is, the order in which the atoms are connected by chemical bonds or are arranged in three-dimensional space. This limitation is especially significant in organic compounds, in which hundreds if not thousands of different molecules may share the same empirical formula. For example, both ethanol and dimethyl ether both have the *empirical formula* C_2H_6O, however *the structural formulas* reveal the very different nature of these two molecules:



More Complex Formulas

It is often useful to write formulas in such as way as to convey at least some information about the structure of a compound. For example, the formula of the solid $(NH_4)_2CO_3$ is immediately identifiable as ammonium carbonate, and essentially a compound of ammonium and carbonate ions in a 2:1 ratio, whereas the *simplest* or *empirical* formula $N_2H_8CO_3$ obscures this information.

Similarly, the distinction between ethanol and dimethyl ether can be made by writing the formulas as C_2H_5OH and $CH_3–O–CH_3$, respectively. Although neither of these formulas specifies the structures precisely, anyone who has studied organic chemistry can work them out, and will immediately recognize the –OH (hydroxyl) group which is the defining characteristic of the large class of organic compounds known as *alcohols*. The –O– atom linking two carbons is similarly the defining feature of *ethers*.

Several related terms are used to express the mass of one mole of a substance.

- **Molecular weight** :This is analogous to atomic weight: it is the relative weight of one formula unit of the compound, based on the carbon-12 scale. The molecular weight is found by adding atomic weights of all the atoms present in the formula unit. Molecular weights, like atomic weights, are dimensionless; i.e., they have no units.
- **Formula weight** :The same thing as molecular weight. This term is sometimes used in connection with ionic solids and other substances in which discrete molecules do not exist.
- **Molar mass:** The mass (in grams, kilograms, or any other unit) of one mole of particles or formula units. When expressed in grams, the molar mass is numerically the same as the molecular weight, but it must be accompanied by the mass unit.

Example 4.3.4: Formula weight and molar mass

a. Calculate the formula weight of copper(II) chloride, $\mathrm{CuCl_2}$.
b. How would you express this same quantity as a *molar mass*?

Solution

a. The atomic weights of Cu and Cl are, respectively 63.55 and 35.45; the sum of each atomic weight, multiplied by the numbers of each kind of atom in the formula unit, yields:

$$63.55 + 2(25.35) = 134.45. \tag{4.3.1}$$

b. The masses of one mole of Cu and Cl atoms are, respectively, 63.55 g and 35.45 g; the mass of one mole of $CuCl_2$ units is:

$$(63.55g) + 2(25.35g) = 134.45g. \tag{4.3.2}$$

Interpreting formulas in terms of mole ratios and mole fractions

The information contained in formulas can be used to compare the compositions of related compounds as in the following example:

Example 4.3.5: mole ratio calculation

The ratio of hydrogen to carbon is often of interest in comparing different fuels. Calculate these ratios for methanol (CH_3OH) and ethanol (C_2H_5OH).

Solution

The H:C ratios for the two alcohols are 4:1 = 4.0 for methanol and 6:2 (3.0) for ethanol.

Alternatively, one sometimes uses mole fractions to express the same thing. The mole fraction of an element M in a compound is just the number of atoms of M divided by the total number of atoms in the formula unit.

Example 4.3.6: mole fraction and mole percent

Calculate the mole fraction and mole-percent of carbon in ethanol (C_2H_5OH).

Solution

The formula unit contains nine atoms, two of which are carbon. The mole fraction of carbon in the compound is 2/9 = .22. Thus 22 percent of the atoms in ethanol are carbon.

Interpreting formulas in terms of masses of the elements

Since the formula of a compound expresses the ratio of the numbers of its constituent atoms, a formula also conveys information about the relative masses of the elements it contains. But in order to make this connection, we need to know the relative masses of the different elements.

Example 4.3.7: mass of each element in a given mass of compound

Find the masses of carbon, hydrogen and oxygen in one mole of ethanol (C_2H_5OH).

Solution

Using the atomic weights (molar masses) of these three elements, we have

- carbon: (2 mol)(12.0 g mol^{-1}) = 24 g of C
- hydrogen: (6 mol)(1.01 g mol^{-1}) = 6 g of H
- oxygen: (1 mol)(16.0 g mol^{-1}) = 16 g of O

The *mass fraction* of an element in a compound is just the ratio of the mass of that element to the mass of the entire formula unit. Mass fractions are always between 0 and 1, but are frequently expressed as percent.

Example 4.3.8: Mass fraction & Mass percent of an element in a Compound

Find the mass fraction and mass percentage of oxygen in ethanol (C_2H_5OH)

Solution

Using the information developed in the preceding example, the molar mass of ethanol is (24 + 6 + 16)g mol^{-1} = 46 g mol^{-1}. Of this, 16 g is due to oxygen, so its mass fraction in the compound is (16 g)/(46 g) = 0.35 which corresponds to 35%.

Finding the **percentage composition** of a compound from its formula is a fundamental calculation that you must master; the technique is exactly as shown above. Finding a mass fraction is often the first step in solving related kinds of problems:

Example 4.3.9: Mass of an Element in a Given mass of Compound

How many tons of potassium are contained in 10 tons of KCl?

Solution

The mass fraction of K in KCl is 39.1/74.6=.524; 10 tons of KCl contains(39.1/74.6) × 10 tons of K, or **5.24 tons** of K. (Atomic weights: K = 39.1, Cl = 35.5.)

Note that there is no need to deal explicitly with moles, which would require converting tons to kg.

Example 4.3.10: Mass of compound containing given mass of an element

How many grams of KCl will contain 10 g of potassium?

Solution

The mass ratio of KCl/K is 74.6 ÷ 39.1; 10 g of potassium will be present in (74.6/39.1) × 10 grams of KCl, or **19 grams**.

Mass ratios of two elements in a compound can be found directly from the mole ratios that are expressed in formulas.

Example 4.3.11: Mass ratio of elements from formula

Molten magnesium chloride ($MgCl_2$) can be decomposed into its elements by passing an electric current through it. How many kg of chlorine will be released when 2.5 kg of magnesium is formed? (Mg = 24.3, Cl = 35.5)

Solution

The mass ratio of Cl/Mg is (35.5 ×2)/24.3, or 2.9; thus 2.9 kg of chlorine will be produced for every kg of Mg, or (2.9 × 2.5) = 7.2 kg of chlorine for 2.5 kg of Mg. (Note that is is not necessary to know the formula of elemental chlorine (Cl_2) in order to solve this problem.)

Empirical formulas from Experimental data

As was explained above, the *empirical formula* (*empirical formula*) is one in which the relative numbers of the various elements are expressed in the smallest possible whole numbers. Aluminum chloride, for example, exists in the form of structural units having the composition Al_2Cl_6; the empirical formula of this substance is $AlCl_3$. Some methods of analysis provide information about the relative numbers of the different kinds of atoms in a compound. The process of finding the formula of a compound from an analysis of its composition depends on your ability to recognize the decimal equivalents of common integer ratios such as 2:3, 3:2, 4:5, etc.

Example 4.3.12: Empirical formula from mole ratio

Analysis of an aluminum compound showed that 1.7 mol of Al is combined with 5.1 mol of chlorine. Write the empirical formula of this compound.

Solution

The formula $Al_{1.7}Cl_{5.1}$ expresses the relative numbers of moles of the two elements in the compound. It can be converted into the empirical formula by dividing both subscripts by the smaller one, yielding **$AlCl_3$** .

More commonly, an arbitrary mass of a compound is found to contain certain masses of its elements. These must be converted to moles in order to find the formula.

Example 4.3.13: Empirical formula from combustion masses

In a student lab experiment, it was found that 0.5684 g of magnesium burns in air to form 0.9426 g of magnesium oxide. Find the empirical formula of this compound. Atomic weights: Mg = 24.305, O=16.00.

Solution

Express this ratio as 0.375 g of C to 1.00 g of O.

- moles of carbon: (.375 g)/(12 g/mol) = 0.03125 mol C;
- moles of oxygen: (1.00 g)/(16 g/mol) = 0.0625 mol O
- mole ratio of C/O = 0.03125/0.0625 = 0.5;

this corresponds to the formula $C_{0.5}O$, which we express in integers as **CO_2**.

Example 4.3.14: Empirical formula from element masses

A 4.67-g sample of an aluminum compound was found to contain 0.945 g of Al and 3.72 g of Cl. Find the empirical formula of this compound. Atomic weights: Al = 27.0, Cl=35.45.

Solution

The sample contains (0.945 g)/(27.0 g mol^{-1}) = .035 mol of aluminum and (3.72 g)(35.45) = 0.105 mol of chlorine. The formula $Al_{.035}Cl_{.105}$ expresses the relative numbers of moles of the two elements in the compound. It can be converted into the empirical formula by dividing both subscripts by the smaller one, yielding **$AlCl_3$**.

The composition of a binary (two-element) compound is sometimes expressed as a mass ratio. The easiest approach here is to treat the numbers that express the ratio as masses, thus turning the problem into the kind described immediately above.

Example 4.3.15: Empirical Formula from element mass ratio

A compound composed of only carbon and oxygen contains these two elements in a mass ratio C:H of 0.375. Find the empirical formula.

Solution

Express this ratio as 0.375 g of C to 1.00 g of O.

- moles of carbon: (.375 g)/(12 g/mol) = .03125 mol C;
- moles of oxygen: (1.00 g)/(16 g/mol) = .0625 mol O
- mole ratio of C/O = .03125/.0625 = 0.5;

this corresponds to the formula $C_{0.5}O$, which we express in integers as **CO_2**.

The composition-by-mass of a compound is most commonly expressed as weight percent (grams per 100 grams of compound). The first step is again to convert these to relative numbers of moles of each element in a fixed mass of the compound. Although this fixed mass is completely arbitrary (there is nothing special about 100 grams!), the *ratios* of the mole amounts of the various

elements are not arbitrary: these ratios must be expressible as integers, since they represent ratios of integral numbers of atoms.

Example 4.3.16: Empirical Formula from mass-percent composition

Find the empirical formula of a compound having the following mass-percent composition. Atomic weights are given in parentheses: 36.4 % Mn (54.9), 21.2 % S (32.06), 42.4 % O (16.0)

Solution

100 g of this compound contains:

- Mn: (36.4 g) / (54.9 g mol^{-1}) = 0.663 mol
- S: (21.2 g) / (32.06 g mol^{-1}) = 0.660 mol
- O: (42.4 g) / (16.0 g mol^{-1}) = 2.65 mol

The formula $Mn_{.663}S_{.660}O_{2.65}$ expresses the relative numbers of moles of the three elements in the compound. It can be converted into the empirical formula by dividing all subscripts by the smallest one, yielding $Mn_{1.00}S_{1.00}O_{4.01}$ which we write as **$MnSO_4$**.

Note: because experimentally-determined masses are subject to small errors, it is usually necessary to neglect small deviations from integer values.

Example 4.3.17: Empirical Formula from mass-percent composition

Find the empirical formula of a compound having the following mass-percent composition. Atomic weights are given in parentheses: 27.6 % Mn (54.9), 24.2 % S (32.06), 48.2 % O (16.0).

Solution

A preliminary formula based on 100 g of this compound can be written as

or

$$Mn_{.503}S_{.754}O_{3.01}$$

Dividing through by the smallest subscript yields $Mn_1S_{1.5}O_6$. Inspection of this formula suggests that multiplying each subscript by 2 yields the all-integer formula **$Mn_2S_3O_{12}$** .

Notes on experimental methods

One of the most fundamental operations in chemistry consists of breaking down a compound into its elements (a process known as *analysis*) and then determining the empirical formula from the relative amounts of each kind of atom present in the compound. In only a very few cases is it practical to carry out such a process directly: thus heating mercury(II) sulfide results in its direct decomposition:

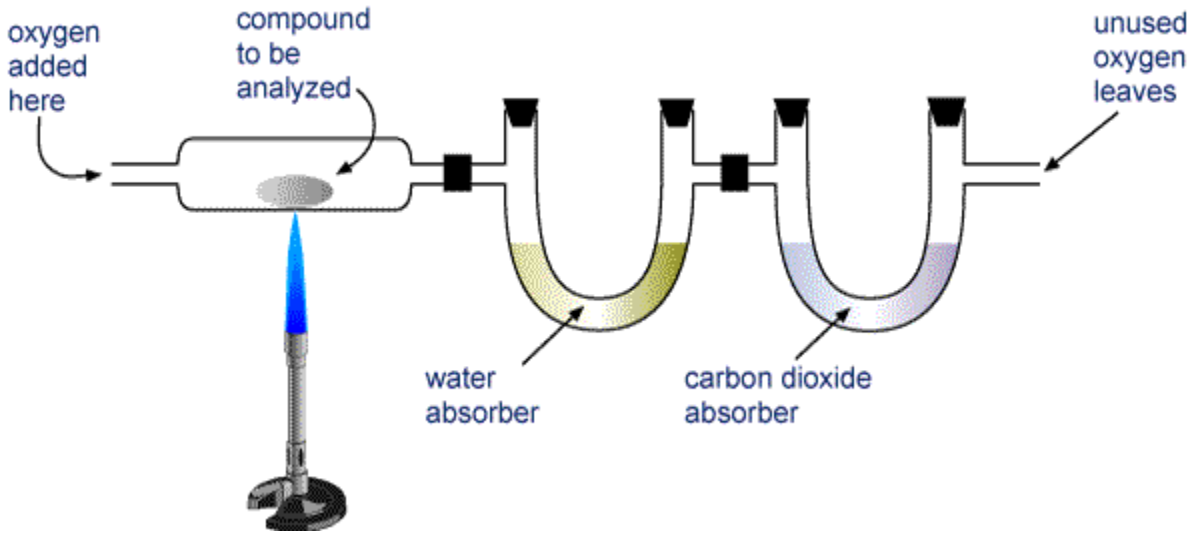
$$2\,HgS \longrightarrow 2\,Hg + O_2. \tag{4.3.3}$$

Similarly, electrolysis of water produces the gases H_2 and O_2 in a 2:1 volume ratio.

Most elemental analyses must be carried out indirectly, however. The most widely used of these methods has traditionally been the *combustion analysis* of organic compounds. An unknown hydrocarbon $C_aH_bO_c$ can be characterized by heating it in an oxygen stream so that it is completely decomposed into gaseous CO_2 and H_2O. These gases are passed through tubes containing substances which absorb each gas selectively. By careful weighing of each tube before and after the combustion process, the values of *a* and *b* for carbon and hydrogen, respectively, can be calculated. The subscript c for oxygen is found by subtracting the

calculated masses of carbon and hydrogen from that of the original sample.

Since the 1970s, it has been possible to carry out combustion analyses with automated equipment. This one can also determine nitrogen and sulfur:



Measurements of mass or weight have long been the principal tool for understanding chemical change in a quantitative way. Balances and weighing scales have been in use for commercial and pharmaceutical purposes since the beginning of recorded history, but these devices lacked the 0.001-g precision required for quantitative chemistry and elemental analysis carried out on the laboratory scale.

It was not until the mid-18th century that the Scottish chemist Joseph Black invented the *equal arm analytical balance*. The key feature of this invention was a lightweight, rigid beam supported on a knife-edged fulcrum; additional knife-edges supported the weighing pans. The knife-edges greatly reduced the friction that limited the sensitivity of previous designs; it is no coincidence that accurate measurements of combining weights and atomic weights began at about this time.

Analytical balances are enclosed in a glass case to avoid interference from air currents, and the calibrated weights are handled with forceps to prevent adsorption of moisture or oils from bare fingers.

Anyone who was enrolled in college-level general chemistry up through the 1960's will recall the training (and tedium) associated with these devices. These could read directly to 1 milligram and allow estimates to ±0.1 mg. Later technical refinements added magnetic damping of beam swinging, pan brakes, and built-in weight sets operated by knobs. The very best research-grade balances achieved precisions of 0.001 mg.

Beginning in the 1970's, **electronic balances** have come into wide use, with single-pan types being especially popular. A single-pan balance eliminates the need for comparing the weight of the sample with that of calibrated weights. Addition of a sample to the pan causes a displacement of a *load cell*which generates a compensating electromagnetic field of sufficient magnitude to raise the pan to its original position. The current required to accomplish this is sensed and converted into a weight measurement. The best research-grade electronic balances can read to 1 microgram, but 0.1-mg sensitivities are more common for student laboratory use.

This page titled 4.3: Formulas and Their Meaning is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

4.4: Chemical Equations and Stoichiometry

Learning Objectives

Make sure you thoroughly understand the following essential ideas

- Given the formulas of reactants and products, write a *balanced chemical equation* for the reaction.
- Given the relative solubilities, write a balanced *net ionic equation* for a reaction between aqueous solutions of two ionic compounds.
- Write appropriate chemical *conversion factors* to calculate the masses of all components of a chemical reaction when the mass of any single component is specified in any system of units.
  and find the masses of all components present when the reaction is complete.
- Describe the manner in which the concept of limiting reactant relates to combustion and human exercise.

A chemical equation expresses the net change in composition associated with a chemical reaction by showing the number of moles of reactants and products. But because each component has its own molar mass, equations also implicitly define the way in which the masses of products and reactants are related. In this unit we will concentrate on understanding and making use of these mass relations.

In a *chemical reaction*, one or more *reactants* are transformed into *products*:

$$\text{reactants} \rightarrow \text{products}$$

The purpose of a *chemical equation* is to express this relation in terms of the formulas of the actual reactants and products that define a particular chemical change. For example, the reaction of mercury with oxygen to produce mercuric oxide would be expressed by the equation

$$Hg + O_2 \rightarrow HgO_2$$

Sometimes, for convenience, it is desirable to indicate the physical state (gas, liquid or solid) of one or more of the species by appropriate abbreviations:

$$Hg(l) + O_2(g) \rightarrow HgO_2(s)$$

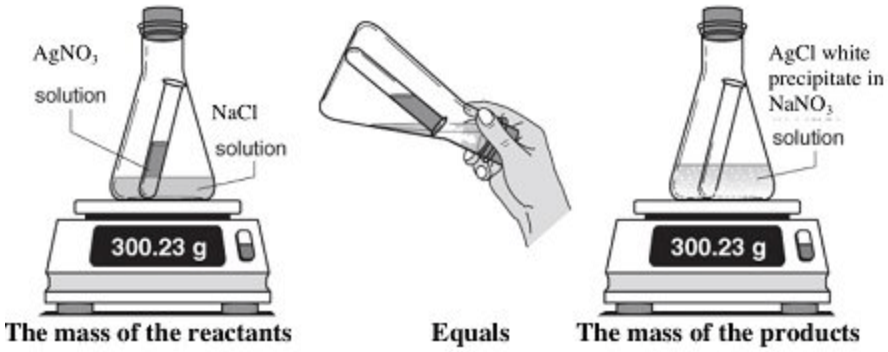
$$C(\text{graphite}) + O_2(g) \rightarrow CO_2(g)$$

$$C(\text{diamond}) + O_2(g) \rightarrow CO_2(g)$$

However, this is always optional.

How to read and write chemical equations

A chemical equation is a statement of a fact: it expresses the net change that occurs as the result of a chemical reaction. In doing so, is must be consistent with the law of conservation of mass:



In the context of an ordinary chemical reaction, conservation of mass means that atoms are neither created nor destroyed. This requirement is easily met by making sure that there are equal numbers of all atoms on both sides of the equation.

> When we balance an equation, we simply make it consistent with the observed fact that individual atoms are conserved in chemical changes. There is no set "recipe'' for balancing ordinary chemical equations; it is best to begin by carefully studying selected examples such as those given below.

Example 4.4.1: Combustion of Propane

Write a balanced equation for the combustion of propane C_3H_8 in oxygen O_2. The products are carbon dioxide CO_2 and water H_2O.

Solution

Begin by writing the unbalanced equation

$$C_3H_8 + O_2 \rightarrow CO_2 + H_2O$$

It is usually best to begin by balancing compounds containing the least abundant element, so we first balance the equation for carbon:

$$C_3H_8 + O_2 \rightarrow \mathbf{3}\ CO_2 + H_2O$$

In balancing the oxygen, we see that there is no way that an even number of O_2 molecules on the left can yield the uneven number of O atoms shown on the right. Don't worry about this now— just use the appropriate fractional coefficient:

$$C_3H_8 + \mathbf{3\ ½}\ O_2 \rightarrow 3\ CO_2 + H_2O$$

Finally, we balance the hydrogens by adding more waters on the right:

$$C_3H_8 + 7/2\ O_2 \rightarrow 3\ CO_2 + \mathbf{4}\ H_2O$$

Ah, but now the oxygens are off again — fixing this also allows us to get rid of the fraction on the left side:

$$C_3H_8 + \mathbf{5}\ O_2 \rightarrow 3\ CO_2 + 4\ H_2O$$

It often happens, however, that we do end up with a fractional coefficient, as in this variant of the above example.

Example 4.4.2: combustion of ethane

Write a balanced equation for the combustion of ethane C_2H_6 in oxygen O_2. The products are carbon dioxide CO_2 and water H_2O.

Solution

Begin by writing the unbalanced equation

$$C_2H_6 + O_2 \rightarrow CO_2 + H_2O$$

...then balance the carbon:

$$C_2H_6 + O_2 \rightarrow \mathbf{2}\ CO_2 + H_2O$$

Let's balance the hydrogen next:

$$C_2H_6 + O_2 \rightarrow 2\ CO_2 + \mathbf{3}\ H_2O$$

...but now we need a non-integral number of dioxygen molecules on the left:

$$C_2H_6 + 7/2\ O_2 \rightarrow 2\ CO_2 + 3\ H_2O$$

My preference is to simply leave it in this form; there is nothing wrong with 7/2 = 3 ½ moles of O_2, and little to be gained by multiplying every term by two— not unless your teacher is a real stickler for doing it "by the book", in which case you had better write

$$2\ C_2H_6 + 7\ O_2 \rightarrow 4\ CO_2 + 6\ H_2O$$

Net Ionic Equations

Ionic compounds are usually dissociated in aqueous solution; thus if we combine solutions of silver nitrate $AgNO_3$ and sodium chloride NaCl we are really combining four different species: the *cations* (positive ions) $Ag^+$and $Na^+$and the *anions* (negative ions) NO_3^- and Cl^-. It happens that when the ions Ag^+ and Cl^- are brought together, they will combine to form an insoluble precipitate of silver chloride. The net equation for this reaction is

$$Ag^+_{(aq)} + Cl^-_{(aq)} \rightarrow AgCl_{(s)} \tag{4.4.1}$$

Note that

- the ions NO_3^- and Cl^- are not directly involved in this reaction; the equation expresses only the *net change*, which is the removal of the silver and chloride ions from the solution to form an insoluble solid.
- the symbol (*aq*) signifies that the ions are in aqueous solution, and thus are *hydrated*, or attached to water molecules.
- the symbol (s) indicates that the substance AgCl exists as a solid. When a solid is formed in a reaction that takes place in solution, it is known as a *precipitate*. The formation of a precipitate is often indicated by underscoring.

From the above example involving silver chloride, it is clear that a meaningful net ionic equation can be written only if two ions combine to form an insoluble compound. To make this determination, it helps to know the *solubility rules*— which all students of chemistry were at one time required to memorize, but are nowadays usually obtained from tables such as the one shown below.

Anion (negative ion)	Cation (positive ion)	Soluble?
any anion	alkali metal ions (Li^+, Na^+, K^+, etc.)	yes
nitrate, NO_3^-	any cation	yes
acetate, CH_3COO^-	any cation except Ag^+	yes
halide ions Cl^-, Br^-, or I^-	Ag^+, Pb^{2+}, Hg_2^{2+}, Cu^{2+}	no
halide ions Cl^-, Br^-, or I^-	any other cation	yes
sulfate, SO_4^{2-}	Ca^{2+}, Sr^{2+}, Ba^{2+}, Ag^+, Pb^{2+}	no

sulfate, SO_4^{2-}	any other cation	yes
sulfide, S_2^-	alkali metal ions or NH_4^+	yes
sulfide, S_2^-	Be^{2+}, Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+}, Ra^{2+}	yes
sulfide, S_2^-	any other cation	no
hydroxide, OH^-	alkali metal ions or NH_4^+	yes
hydroxide, OH^-	Sr^{2+}, Ba^{2+}, Ra^{2+}	slightly
hydroxide, OH^-	any other cation	no
phosphate, PO_4^{3-}, carbonate CO_3^{2-}	alkali metal ions or NH_4^+	yes
phosphate, PO_4^{3-}, carbonate CO_3^{2-}	any other cation	no

Example 4.4.3: net ionic equations

Write net ionic equations for what happens when aqueous solutions of the following salts are combined:

a. $PbCl_2 + K_2SO_4$
b. $K_2CO_3 + Sr(NO_3)_2$
c. $AlCl_3 + CaSO_4$
d. $Na_3PO_4 + CaCl_2$

Use the solubility rules table(above) to find the insoluble combinations:

a. $Pb^{2+}(aq) + SO_4^{2-}(aq) \rightarrow PbSO_4(s)$
b. $Sr^{2+}(aq) + CO_3^{2-}(aq) \rightarrow SrCO_3(s)$
c. no net reaction
d. $3\ Ca^{2+}(aq) + 2\ PO_4^{3-}(aq) \rightarrow 3\ Ca_3(PO_4)_2(s)$

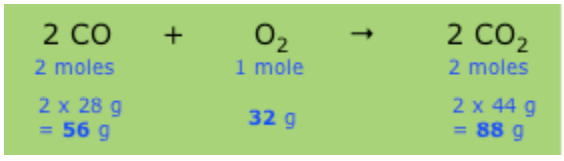
Note the need to balance the electric charges

Mass Relations in Chemical Equations

A balanced chemical equation expresses the relative number of moles of each component (product or reactant), but because each formula in the equation implies a definite mass of the substance (its molar mass), the equation also implies that certain weight relations exist between the components. For example, the equation describing the combustion of carbon monoxide to carbon dioxide

$$2CO + O_2 \rightarrow 2CO_2 \tag{4.4.2}$$

implies the following relations:



The relative masses shown in the bottom line establish the *stoichiometry* of the reaction, that is, the relations between the masses of the various components. Since these masses vary in direct proportion to one another, we can define what amounts to a conversion factor (sometimes referred to as a *chemical factor*) that relates the mass of any one component to that of any other

component.

Example **3.4.4: chemical factor and mass conversion**

Evaluate the chemical factor and the conversion factor that relates the mass of carbon dioxide to that of the CO consumed in the reaction.

Solution

From the above box, the mass ratio of CO_2 to CO in this reaction is 88/56 = **1.57**; this is the *chemical factor* for the conversion of CO into CO_2. The *conversion factor* is just 1.57/1 with the mass units explicitly stated:

$$\frac{1.57\ g\ CO_2}{1\ g\ CO} = 1 \tag{4.4.3}$$

Example 4.4.5: mass-mass calculations in various units

a. How many tons of CO_2 can be obtained from the combustion of 10 tons of CO?
b. How many kg of CO must be burnt to produce 20 kg of CO_2?

Solutions

a. (1.57 T CO_2 / 1 T CO) × (10 T CO) = **15.7 T CO_2**
b. Notice the answer to this one must refer to carbon monoxide, not CO_2, so we write the conversion factor in reverse:

(1 kg CO / 1.57 kg CO_2) × (20 kg CO_2) = (20/1.57)g CO = **12.7 kg CO**.

Is this answer reasonable? Yes, because the mass of CO must always be smaller than that of CO_2 in this reaction.

More mass-mass problems

Don't expect to pass Chemistry unless you can handle problems such as the ones below; they come up frequently in all kinds of contexts. If you feel the need for more guidance, see one of the video tutorials listed near the bottom of this page.

Example 4.4.6

The ore FeS_2 can be converted into the important industrial chemical *sulfuric acid* H_2SO_4 by a series of processes. Assuming that the conversion is complete, how many liters of sulfuric acid (density 1.86 kg L^{-1}) can be made from 50 kg of ore?

Solution

As with most problems, this breaks down into several simpler ones. We begin by working out the stoichiometry on the assumption that all the sulfur in the or ends up as H_2SO_4, allowing us to write

$$FeS_2 \rightarrow 2\ H_2SO_4$$

*is*balanced in respect to the two components of interest, and this is all we need here. The molar masses of the two components are 120.0 and 98 g mol^{-1}, respectively, so the equation can be interpreted in terms of masses as

[120 mass units] FeS_2 → [2 × 98 mass units] H_2SO_4

Thus 50 kg of ore will yield (50 kg) × (196/120) = 81.7 kg of product.
[*Check*: is this answer reasonable? Yes, because the factor (196/120) is close to (200/120) = 5/3, so the mass of product should be slightly smaller than twice the mass of ore consumed.]

From the density information we find that the volume of liquid H_2SO_4 is

$$(81.7\text{ kg}) \div (1.86\text{ kg L}^{-1}) = \mathbf{43.9\ L}$$

[*Check*: is this answer reasonable? Yes, because density tells us that the number of liters of acid will be slightly greater than half of its weight.]

Example 4.4.7

Barium chloride forms a crystalline hydrate, $BaCl_2 \cdot xH_2O$, in which *x* molecules of water are incorporated into the crystalline solid for every unit of $BaCl_2$. This water can be driven off by heat; if 1.10 g of the hydrated salt is heated and reweighed several times until no further loss of weight (i.e., loss of water) occurs, the final weight of the sample is 0.937 g. What is the value of *x* in the formula of the hydrate?

Solution

The first step is to find the number of moles of $BaCl_2$ (molecular weight 208.2) from the mass of the dehydrated sample.

$$(0.937\text{ g}) / (208.2\text{ g mol}^{-1}) = 0.00450\text{ mol}$$

Now find the moles of H_2O (molecular weight 18) lost when the sample was dried:

$$(1.10 - .937)\text{g} / (18\text{ g mol}^{-1}) = .00905\text{ mol}$$

Allowing for a reasonable amount of measurement error, it is apparent that the mole ratio of $BaCl_2$:H2O = 1:2. The formula of the hydrate is $\mathbf{BaCl_2 \cdot 2H_2O}$.

Limiting Reactants

Most chemical reactions that take place in the real world begin with more or less arbitrary amounts of the various reactants; we usually have to make a special effort if we want to ensure that stoichiometric amounts of the reactants are combined. This means that one or more reactant will usually be present in excess; there will be more present than can react, and some will remain after the reaction is over. At the same time, one reactant will be completely used up; we call this the *limiting reactant*because the amount of this substance present will control, or limit, the quantities of the other reactants that are consumed as well as the amounts of products produced.

Limiting reactant problems are handled in the same way as ordinary stoichiometry problems with one additional preliminary step: you must first determine which of the reactants is limiting— that is, which one will be completely used up. To start you off, consider the following very simple example

Example 4.4.8

For the hypothetical reaction **3 A + 4 B → [products]**, determine which reactant will be completely consumed when we combine

a. equimolar quantities of A and B;
b. 0.57 mol of A and 0.68 mol of B.

Solution

a) Simple inspection of the equation shows clearly that more moles of B are required, so this component will be consumed (and is thus the limiting reactant), leaving behind ¾ as many moles of A.

b) How many moles of B will react with .57 mol of A? The answer will be (4/3 × 0.57 mol). If this comes to less than 0.68 mol, then B will be the limiting reactant, and you must continue the problem on the basis of the amount of B present. If the limiting reactant is A, then all 0.57 mol of A will react, leaving some of the B in excess. Work it out!

Example 4.4.9

Sulfur and copper, when heated together, react to form copper(I) sulfide, Cu_2S. How many grams of Cu_2S can be made from 10 g of sulfur and 15 g of copper?

Solution

From the atomic weights of Cu (63.55) and S (32.06) we can interpret the he reaction 2 Cu + S → Cu_2S as

[2 × 63.55 = 127.1 mass units] **Cu** + [32.06 mass units] S
→ [159.2 mass units] **Cu_2S**

Thus 10 g of S will require

(10 g S) × (127.1 g Cu)/(32.06 g S) = 39.6 g Cu

...which is a lot more than what is available, so **copper is the limiting reactant** here.

[Check: is this answer reasonable? Yes, because the chemical factor (127/32) works out to about 4, indicating that sulfur reacts with about four times its weight of copper.]

The mass of copper sulfide formed will be determined by the mass of copper available:

(15 g Cu) × (159.2 g Cu_2S) / (127.1 g Cu) = **18.8 g Cu_2S**

[Check: is this answer reasonable? Yes, because the chemical factor (159.2/127.1) is just a bit greater than unity, indicating that the mass of the product will slightly exceed that of the copper consumed.]

The concept of limiting reactants touches us all in our everyday lives — and as we will show in the second example below, even in the maintenance of life itself!

Air-to-fuel ratios in combustion

Combustion is an exothermic process in which a fuel is combined with oxygen; *complete* combustion of a hydrocarbon fuel such as methane or gasoline yields carbon dioxide and water:

$CH_4 + 2\ O_2 \rightarrow CO_2 + 2\ H_2O(g)$

Example 4.4.10: Fuel-to-Oxygen mass ratio in combustion of methane

Calculate the mass ratio of CH_4 to O_2 required for complete combustion.

Solution

This is just the ratio of the molar mass of CH_4 (16 g) to that of two moles of dioxygen (2 x 32 g)

Thus (64 g) / (16 g) = 4/1 = **4.0**.

Complete combustion of each kg of methane consumes 4 kg of dioxygen, which is supplied by the air. In the classic Bunsen burner, this air is admitted through an adjustable opening near the base. When it is fully open, the flame is blue in color and achieves its maximum temperature, indicating that combustion is approximately stoichiometric. If the opening is gradually closed, the appearance of the flame changes as illustrated. Under these conditions, oxygen becomes the limiting reactant and combustion is incomplete.



Incomplete combustion is generally undesirable because it wastes fuel, produces less heat, and releases pollutants such as carbon soot. Energy-producing combustion processes should always operate in fuel-limited mode.

In ordinary combustion processes, the source of oxygen is air. Because only about 20 percent of the molecules in dry air consist of O_2, the volume of air that must be supplied is five times greater than what would be required for pure O_2. Calculation of the *air-to-fuel mass ratio* ("A/F ratio") employed by combustion engineers is complicated by the differing molar masses of dioxygen and air. For methane combustion, the A/F ratio works out to about 17.2. A/F ratios which exceed the stoichiometric values are said to be *lean*, while those in which air becomes the limiting component are characterized as *rich*. In order to ensure complete combustion, it is common practice to maintain a slightly lean mixture. The quantities of so-called *excess air* commonly admitted to burners vary from 5-10% for natural gas to up to 100% for certain grades of coal.

For internal combustion engines fueled by gasoline (roughly equivalent to C_7H_{14}), the stoichiometric A/F ratio is 15:1. However, practical considerations necessitate differing ratios at various stages of operation. Typical values vary from a rich ratio for starting or acceleration to slightly lean ratios for ordinary driving. These ratios are set by the carburetor, with additional control by the engine computer and exhaust-line oxygen sensor in modern vehicles, or by a manual *choke* in earlier ones.

Aerobic and anaerobic respiration

Our bodies require a continual supply of energy in order to maintain neural activity, synthesize proteins and other essential biochemical components, replace cells, and to power muscular action. The "fuel" — the carrier of chemical energy — glucose, a simple sugar which is released as needed from the starch-like polymer glycogen, the form in which the energy we derive from food is stored. Arterial blood carries dissolved glucose along with hemoglobin-bound dioxygen to individual cells which are the sites of glucose "combustion":

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O \tag{4.4.4}$$

The net reaction and the quantity of energy released are the same as if the glucose were burned in the open air, but within the cells the reaction proceeds in a series of tiny steps which capture most of this energy for the body's use, liberating only a small fraction of it as thermal energy (heat). Because this process utilizes oxygen from the air we breath, it is known as *aerobic respiration*. And as with any efficient combustion process, glucose is the limiting reactant here.

However, there are times when vigorous physical activity causes muscles to consume glucose at a rate that exceeds the capacity of the blood to deliver the required quantity of oxygen. Under these conditions, cellular respiration shifts to an alternative *anaerobic* mode:

$$C_6H_{12}O_6 \rightarrow 2CH_3CH(OH)COOH \tag{4.4.5}$$

As you can see from this equation, glucose is only partially broken down (into *lactic acid*), and thus only part of its chemical energy is captured by the body. There are numerous health benefits to aerobic exercise including increased ability of the body to maintain an aerobic condition. But if you are into short-distance running (sprinting) or being pursued by a tiger, the reduced efficiency of anaerobic exercise may be a small price to pay.

This page titled 4.4: Chemical Equations and Stoichiometry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

4.5: Introduction to Chemical Nomenclature

Learning Objectives

Different instructors set out widely varying requirements for chemical nomenclature. The following are probably the most commonly expected:

- You should know the name and symbols of at least the first twenty elements, as well as all of the halogen and noble gas groups (groups 17-18).
- Name any binary molecule, using the standard prefixes for 1-10.
- All of the commonly-encountered ions.
- Salts and other ion-derived compounds, including the acids listed here. In some courses you will not need to know the *-ous/-ic* names for salts of copper, iron, etc., but in others you will.
- Find out from your instructor which organic compounds you must be able to name.

Chemical nomenclature is far too big a topic to treat comprehensively, and it would be a useless diversion to attempt to do so in a beginning course; most chemistry students pick up chemical names and the rules governing them as they go along. But we can hardly talk about chemistry without mentioning *some* chemical substances, all of which do have names— and often, more than one! All we will try to do here is cover what you need to know to make sense of first-year chemistry. For those of you who plan to go on in chemistry, the really fun stuff comes later!

There are more than 100 million named chemical substances. Who thinks up the names for all these chemicals? Are we in danger of running out of new names? The answer to the last question is "no", for the simple reason that the vast majority of the names are not "thought up"; there are elaborate rules for assigning names to chemical substances on the basis of their structures. These are called *systematic names*; they may be a bit ponderous, but they uniquely identify a given substance. The rules for these names are defined by an international body. But in order to make indexing and identification easier, every known chemical substance has its own numeric "personal ID", known as a CAS registry number. For example, caffeine is uniquely identified by the registry number 58-08-2. About 15,000 new numbers are issued every day.

Common Names vs. Systematic Names

Many chemicals are so much a part of our life that we know them by their familiar names, just like our other friends. A given substance may have several common or *trivial names*; ordinary cane sugar, for example, is more formally known as "sucrose", but asking for it at the dinner table by that name will likely be a conversation-stopper, and I won't even venture to predict the outcome if you try using its systematic name in the same context:

"please pass the α-D-glucopyranosyl-(1,2)- β-D-fructofuranoside!"

But "sucrose" would be quite appropriate if you need to distinguish this particular sugar from the hundreds of other named sugars. The only place you would come across a systematic name like the rather unwieldy one mentioned here is when referring (in print or in a computer data base) to a sugar that has no common name.

Chemical substances have been a part the fabric of civilization and culture for thousands of years, and present-day chemistry retains a lot of this ancient baggage in the form of terms whose hidden cultural and historic connections add color and interest to the subject. Many common chemical names have reached us only after remarkably long journeys through time and place, as the following two examples illustrate.

Most people can associate the name ***ammonia*** (NH_3) with a gas having a pungent odor; the systematic name "nitrogen trihydride" (which is rarely used) will tell you its formula. What it will not tell you is that smoke from burning camel dung (the staple fuel of North Africa) condenses on cool surfaces to form a crystalline deposit. The ancient Romans first noticed this on the walls and ceiling of the temple that the Egyptians had built to the Sun-god Amun in Thebes, and they named the material sal ammoniac, meaning "salt of Amun". In 1774, Joseph Priestly (the discoverer of oxygen) found that heating sal ammoniac produced a gas with

a pungent odor, which a T. Bergman named "ammonia" eight years later.

Alcohol entered the English language in the 17th Century with the meaning of a "sublimated" substance, then became the "pure spirit" of anything, and only became associated with "spirit of wine" in 1753. Finally, in 1852, it become a part of chemical nomenclature that denoted a common class of organic compound. But it's still common practice to refer to the specific substance CH_3CH_2OH as "alcohol" rather then its systematic name *ethanol*.

Arabic alchemy has given us a number of chemical terms; for example,alcohol is believed to derive from Arabic or al-ghawl whose original meaning was a metallic powder used to darken women's eyelids (kohl).

The general practice among chemists is to use the more common chemical names whenever it is practical to do so, especially in spoken or informal written communication. For many of the very simplest compounds (including most of those you will encounter in a first-year course), the systematic and common names are the same, but where there is a difference and if the context permits it, the common name is usually preferred.

Many of the "common" names we refer to in this lesson are known and used mainly by the scientific community. Chemical substances that are employed in the home, the arts, or in industry have acquired traditional or "popular" names that are still in wide use. Many, like sal ammoniac mentioned above, have fascinating stories to tell.

$B_4O_7{\cdot}10H2O$

Table 4.5.1*:*

popular name	chemical name	formula
borax	sodium tetraborate decahydrate	
calomel	mercury(I) chloride	Hg_2Cl_2
milk of magnesia	magnesium hydroxide	$Mg(OH)_2$
muriatic acid	hydrochloric acid	$HCl(aq)$
oil of vitriol	sulfuric acid	H_2SO_4
saltpeter	sodium nitrate	$NaNO_3$
slaked lime	calcium hydroxide	$Ca(OH)_2$

Minerals: Minerals are solid materials that occur in the earth which are classified and named according to their compositions (which often vary over a continuous range) and the arrangement of the atoms in their crystal lattices. There are about 4000 named minerals. Many are named after places, people, or properties, and most frequently end with -ite.

Proprietary names: Chemistry is a major industry, so it is not surprising that many substances are sold under trademarked names. This is especially common in the pharmaceutical industry, which uses computers to churn out names that they hope will distinguish a new product from those of its competitors. Perhaps the most famous of these is Aspirin, whose name was coined by the German company Bayer in 1899. This trade name was seized by the U.S. government following World War I, and is no longer a protected trade mark in that country.

Names and symbols of the Elements

Naming of chemical substances begins with the names of the *elements*. The discoverer of an element has traditionally had the right to name it, and one can find some interesting human and cultural history in these names, many of which refer to the element's properties or to geographic locations. Only some of the more recently-discovered (and artificially produced) elements are named after people. Some elements were not really "discovered", but have been known since ancient times; many of these have symbols that are derived from the Latin names of the elements. There are nine elements whose Latin-derived symbols you are expected to know (Table 4.5.2).

Table 4.5.2

element name	symbol	Latin name
antimony	Sb	stibium
copper	Cu	cuprum
gold	Au	aurum
iron	Fe	ferrum
lead	Pb	plumbum
mercury	Hg	hydrargyrum
potassium	K	kalium
sodium	Na	natrium
tin	Sn	stannum

There is a lot of history and tradition in many of these names. For example, the Latin name for mercury, *hydrargyrum*, means "water silver", or quicksilver. The appellation "quack", as applied to an incompetent physician, is a corruption of the Flemish word for quicksilver, and derives from the use of mercury compounds in 17th century medicine. The name "mercury" is of alchemical origin and is of course derived from the name of the Greek god after whom the planet is named; the enigmatic properties of the element, at the same time metallic, fluid, and vaporizable, suggest the same messenger with the winged feet who circles through the heavens close to the sun.

Naming the binary molecules

The system used for naming chemical substances depends on the nature of the molecular units making up the compound. These are usually either ions or molecules; different rules apply to each. In this section, we discuss the simplest binary (two-atom) molecules.

It is often necessary to distinguish between compounds in which the same elements are present in different proportions; carbon monoxide CO and carbon dioxide CO_2 are familiar to everyone. Chemists, perhaps hoping it will legitimize them as scholars, employ Greek (of sometimes Latin) prefixes to designate numbers within names; you will encounter these frequently, and you should know them:

Table 4.5.3

1/2.	1	2	3	4	5	6	7	8	9	10
hemi	*mono*	*di*	*tri*	*tetra*	*penta*	*hexa*	*hepta*	*octa*	*nona*	*deca*

You will occasionally see names such as *di*hydrogen and *di*chlorine used to distinguish the common forms of these elements (H_2, Cl_2) from the atoms that have the same name when it is required for clarity.

Examples:

- N_2O_4 - dinitrogen tetroxide [note the missing *a* preceding the vowel]
- N_2O - dinitrogen oxide [more commonly, *nitrous oxide*]
- SF_6 - sulfur hexafluoride
- P_4S_3 - tetraphosphorus trisulfide [more commonly, *phosphorus sesquisulfide*]
- Na_2HPO_4 - disodium hydrogen phosphate
- H_2S - hydrogen sulfide [we skip both the *di* and *mono*]
- CO - carbon monoxide [*mono-* to distinguish it from the dioxide]
- $CaSO_4{\cdot}½H_2O$ - calcium sulfate hemihydrate [In this solid, two $CaSO_4$ units share one water of hydration between them; more commonly called Plaster of Paris]

It will be apparent from these examples that chemists are in the habit of taking a few liberties in applying the strict numeric

prefixes to the more commonly known substances.

These two-element compounds are usually quite easy to name because most of them follow the systematic rule of adding the suffix *-ide* to the root name of the second element, which is normally the more "negative" one. Several such examples are shown above. But as noted above, there are some important exceptions in which common or *trivial* names take precedence over systematic names:

- H_2O (*water*, not dihydrogen oxide)/
- H_2O_2 (*hydrogen peroxide*, not dihydrogen dioxide)
- H_2S (*hydrogen sulfide*, not dihydrogen sulfide)
- NH_3 (*ammonia*, not nitrogen trihydride)
- NO (*nitric oxide*, not nitrogen monoxide)
- N_2O (*nitrous oxide*, not dinitrogen oxide)
- CH_4 (*methane*, not carbon tetrahydride)

Naming the ions

An *ion* is an electrically charged atom or molecule— that is, one in which the number of electrons differs from the number of nuclear protons. Many simple compounds can be regarded, at least in a formal way, as being made up of a pair of ions having opposite charge signs. The positive ions, also known as *cations*, are mostly those of metallic elements which simply take the name of the element itself.

calcium	sodium	magnesium	cadmium	potassium
Ca^{2+}	Na^{+}	Mg^{2+}	Cd^{2+}	K^{+}

The only important non-metallic cations you need to know about are

hydrogen	hydronium	ammonium
H^{+}	H_3O^{+}	NH_4^{+}

(Later on, when you study acids and bases, you will learn that the first two represent the same chemical species.)

Some of the metallic ions are *multivalent*, meaning that they can exhibit more than one electric charge. For these there are systematic names that use Roman numerals, and the much older and less cumbersome common names that mostly employ the Latin names of the elements, using the endings *-ous* and *-ic* to denote the lower and higher charges, respectively (Table 4.5.4). (In cases where more than two charge values are possible, the systematic names are used.)

Table 4.5.4: Metalic names

Cu^{+}	Cu^{2+}	Fe^{2+}	Fe^{3+}	$^{*}Hg_2^{2+}$	Hg^{2+}	Sn^{2+}	Sn^{4+}
copper(I)	copper(II)	iron(II)	iron(III)	mercury(I)	mercury(II)	tin(II)	tin(IV)
cuprous	cupric	ferrous	ferric	mercurous	mercuric	stannous	stannic
* The mercurous ion is a unique *double cation* that is sometimes incorrectly represented as Hg^{+}.							

The non-metallic elements generally form *negative* ions (*anions*). The names of the monatomic anions all end with the *-ide* suffix:

Cl^{-}	S^{2-}	O^{2-}	C^{4-}	I^{-}	H^{-}
chloride	sulfide	oxide	carbide	iodide	hydride

There are a number of important polyatomic anions which, for naming purposes, can be divided into several categories. A few follow the pattern for the monatomic anions:

OH^-	CN^-	O_2^-
hydroxide	cyanide	peroxide

Oxyanions

The most common oxygen-containing anions (*oxyanions*) have names ending in *-ate*, but if a variant containing a small number of oxygen atoms exists, it takes the suffix *-ite*.

CO_3^{2-}	NO_3^-	NO_2^-	SO_4^{2-}	SO_3^{2-}	PO_4^{3-}
carbonate	nitrate	nitrite	sulfate	sulfite	phosphate

The above ions (with the exception of nitrate) can also combine with H+ to produce "acid" forms having smaller negative charges. For rather obscure historic reasons, some of them have common names that begin with *-bi*which, although officially discouraged, are still in wide use:

ion	systematic name	common name
HCO_3^-	hydrogen carbonate	bicarbonate
HSO_4^-	hydrogen sulfate	bisulfate
HSO_3^-	hydrogen sulfite	bisulfite

Chlorine, and to a smaller extent bromine and iodine, form a more extensive series of oxyanions that requires a somewhat more intricate naming convention:

ClO^-	ClO_2^-	ClO_3^-	ClO_4^-
hypochlorite	chlorite	chlorate	perchlorate

Ion-derived compounds

These compounds are formally derived from positive ions (*cations*) and negative ions (*anions*) in a ratio that gives an electrically neutral unit. Salts, of which ordinary "salt" (sodium chloride) is the most common example, are all solids under ordinary conditions. A small number of these (such as NaCl) do retain their component ions and are properly called "ionic solids". In many cases, however, the ions lose their electrically charged character and form largely-non-ionic solids such as $CuCl_2$. The term "ion-derived solids" encompasses both of these classes of compounds.

Most of the cations and anions described above can combine to form solid compounds that are usually known as *salts*. The one overriding requirement is that the resulting compound must be electrically neutral: thus the ions Ca^{2+} and Br^- combine only in a 1:2 ratio to form calcium bromide, $CaBr_2$. Because no other simplest formula is possible, there is no need to name it "calcium dibromide".

Since some metallic elements form cations having different positive charges, the names of ionic compounds derived from these elements must contain some indication of the cation charge. The older method uses the suffixes *-ous* and *-ic* to denote the lower and higher charges, respectively. In the cases of iron and copper, the Latin names of the elements are used: *ferrous*, *cupric*.

This system is still widely used, although it has been officially supplanted by the more precise, if slightly cumbersome Stock system in which one indicates the cationic charge (actually, the oxidation number) by means of Roman numerals following the symbol for the cation. In both systems, the name of the anion ends in *-ide*.

Table 4.5.5:

formula	systematic name	common name
CuCl	copper(I) chloride	cuprous chloride
$CuCl_2$	copper(II) chloride	cupric chloride
Hg_2Cl	mercury(I) chloride	mercurous chloride
HgO	mercury(II) oxide	mercuric oxide
FeS	iron(II) sulfide	ferrous sulfide
Fe_2S_3	iron(III) sulfide	ferric sulfide

Acids

Most acids can be regarded as a combination of a hydrogen ion H^+ with an anion; the name of the anion is reflected in the name of the acid. Notice, in the case of the oxyacids, how the anion suffixes *-ate* and *-ite* become *-ic* and *-ous*, respectively, in the acid name.

Table 4.5.6

anion	anion name	acid	acid name
Cl^-	chloride ion	HCl	hydrochloric acid
CO_3^{2-}	carbonate ion	H_2CO_3	carbonic acid
NO_2^-	nitrite ion	HNO_2	nitrous acid
NO_3^-	nitrate ion	HNO_3	nitric acid
SO_3^{2-}	sulfite ion	H_2SO_3	sulfurous acid
SO_4^{2-}	sulfate ion	H_2SO_4	sulfuric acid
CH_3COO^-	acetate ion	CH_3COOH	acetic acid

Organic compounds

Since organic (carbon) compounds constitute the vast majority of all known chemical substances, organic nomenclature is a huge subject in itself. We present here only the very basic part of it that you need to know in first-year chemistry— much more awaits those of you who are to experience the pleasures of an organic chemistry course later on. The simplest organic compounds are built of straight chains of carbon atoms which are named by means of prefixes that denote the number of carbons in the chain. Using the convention $C_n$to denote a straight chain of *n* atoms (don't even *ask* about branched chains!), the prefixes for chain lengths from 1 through 10 are given here:

C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
meth-	eth-	prop-	but-	pent-	hex-	hept-	oct-	non-	dec-

As you can see, chains from C_5 onward use Greek number prefixes, so you don't have a lot new to learn here. The simplest of these compounds are hydrocarbons having the general formula C_nH_{2n+2}. They are known generically as *alkanes*, and their names all combine the appropriate numerical prefix with the ending *-ane*:

CH_4	C_2H_6	C_3H_8	C_8H_{18}
C	C–C	C–C–C	C–C–C–C–C–C–C–C

CH_4	C_2H_6	C_3H_8	C_8H_{18}
methane	ethane	propane	octane

All carbon atoms must have four bonds attached to them; notice the common convention of not showing hydrogen atoms explicitly.

Functional groups

and higher chains, the substituent can be in more than one location, thus giving rise to numerous *isomers*.

Alcohols: the hydroxyl group

formula	common name	systematic name
CH_3OH	methyl alcohol	methanol
CH_3CH_2OH	ethyl alcohol	ethanol
$C_8H_{15}OH$	octyl alcohol	octanol

Acids: The carboxyl group

formula	common name	systematic name
HCOOH	formic acid	methanoic acid
CH_3COOH	acetic acid	ethanoic acid
C_4H_9COOH	butyric acid	butanoic acid

A few others...

class	example	name
amine	methylamine	CH_3NH_2
ketone	acetone (dimethylketone)	CH_3-CO-CH_3
ether	diethyl ether	C_2H_5-O-C_2H_5

This page titled 4.5: Introduction to Chemical Nomenclature is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

4.6: Significant Figures and Rounding

Learning Objectives

- Give an example of a measurement whose number of significant digits is clearly too great, and explain why.
- State the purpose of rounding off, and describe the information that must be known to do it properly.
- Round off a number to a specified number of significant digits.
- Explain how to round off a number whose second-most-significant digit is 9.
- Carry out a simple calculation that involves two or more observed quantities, and express the result in the appropriate number of significant figures.

The numerical values we deal with in science (and in many other aspects of life) represent measurements whose values are never known exactly. Our pocket-calculators or computers don't know this; they treat the numbers we punch into them as "pure" mathematical entities, with the result that the operations of arithmetic frequently yield answers that are physically ridiculous even though mathematically correct. The purpose of this unit is to help you understand why this happens, and to show you what to do about it.

Digits: Significant and otherwise

Consider the two statements shown below:

- **"The population of our city is 157,872."**
- **"The number of registered voters as of Jan 1 was 27,833.**

Which of these would you be justified in dismissing immediately? Certainly not the second one, because it probably comes from a database which contains one record for each voter, so the number is found simply by counting the number of records. The first statement cannot possibly be correct. Even if a city's population could be defined in a precise way (Permanent residents? Warm bodies?), how can we account for the minute-by minute changes that occur as people are born and die, or move in and move away?

What is the difference between the two population numbers stated above? The first one expresses a quantity that cannot be known exactly — that is, it carries with it a degree of uncertainty. It is quite possible that the last census yielded precisely 157,872 records, and that this might be the "population of the city" for legal purposes, but it is surely not the "true" population. To better reflect this fact, one might list the population (in an atlas, for example) as **157,900** or even **158,000**. These two quantities have been rounded off to four and three significant figures, respectively, and the have the following meanings:

- **157900** (the significant digits are underlined here) implies that the population is believed to be within the range of about 157850 to about 157950. In other words, the population is 157900±50. The "plus-or-minus 50" appended to this number means that we consider the absolute uncertainty of the population measurement to be 50 – (–50) = 100. We can also say that the relative uncertainty is 100/157900, which we can also express as 1 part in 1579, or 1/1579 = 0.000633, or about 0.06 percent.
- The value **158000** implies that the population is likely between about 157500 and 158500, or 158000±500. The absolute uncertainty of 1000 translates into a relative uncertainty of 1000/158000 or 1 part in 158, or about 0.6 percent.

Which of these two values we would report as "the population" will depend on the degree of confidence we have in the original census figure; if the census was completed last week, we might round to four significant digits, but if it was a year or so ago, rounding to three places might be a more prudent choice. In a case such as this, there is no really objective way of choosing between the two alternatives.

This illustrates an important point: the concept of *significant digits* has less to do with mathematics than with our confidence in a measurement. This confidence can often be expressed numerically (for example, the height of a liquid in a measuring tube can be read to ±0.05 cm), but when it cannot, as in our population example, we must depend on our personal experience and judgment.

So, what is a significant digit? According to the usual definition, it is all the numerals in a measured quantity (counting from the

left) whose values are considered as known exactly, plus one more whose value could be one more or one less:

- In “157900” (four significant digits), the left most three digits are known exactly, but the fourth digit, “9” could well be “8” if the “true value” is within the implied range of 157850 to 157950.
- In “*158*000” (three significant digits), the left most two digits are known exactly, while the third digit could be either “7” or “8” if the true value is within the implied range of 157500 to 158500.



Although rounding off always leads to the loss of numeric information, what we are getting rid of can be considered to be “numeric noise” that does not contribute to the quality of the measurement. The purpose in rounding off is to avoid expressing a value to a greater degree of precision than is consistent with the uncertainty in the measurement.

Implied Uncertainty

If you know that a balance is accurate to within 0.1 mg, say, then the uncertainty in any measurement of mass carried out on this balance will be ±0.1 mg. Suppose, however, that you are simply told that an object has a length of 0.42 cm, with no indication of its precision. In this case, all you have to go on is the number of digits contained in the data. Thus the quantity “0.42 cm” is specified to 0.01 unit in 0 42, or one part in 42 . The implied relative uncertainty in this figure is 1/42, or about 2%. The precision of any numeric answer calculated from this value is therefore limited to about the same amount.

Rounding Error

It is important to understand that the number of significant digits in a value provides only a rough indication of its precision, and that information is lost when rounding off occurs. Suppose, for example, that we measure the weight of an object as 3.28 g on a balance believed to be accurate to within ±0.05 gram. The resulting value of 3.28±.05 gram tells us that the true weight of the object could be anywhere between 3.23 g and 3.33 g. The absolute uncertainty here is 0.1 g (±0.05 g), and the relative uncertainty is 1 part in 32.8, or about 3 percent.

How many significant digits should there be in the reported measurement? Since only the left most “3” in “3.28” is certain, you would probably elect to round the value to 3.3 g. So far, so good. But what is someone else supposed to make of this figure when they see it in your report? The value “3.3 g” suggests an *implied uncertainty* of 3.3±0.05 g, meaning that the true value is likely between 3.25 g and 3.35 g. This range is 0.02 g below that associated with the original measurement, and so rounding off has introduced a bias of this amount into the result. Since this is less than half of the ±0.05 g uncertainty in the weighing, it is not a very serious matter in itself. However, if several values that were rounded in this way are combined in a calculation, the rounding-off errors could become significant.

Rules for Rounding

The standard rules for rounding off are well known. Before we set them out, let us agree on what to call the various components of a numeric value.

- The *most significant digit* is the left most digit (not counting any leading zeros which function only as placeholders and are

never significant digits.)
- If you are rounding off to ***n*** significant digits, then the *least significant digit* is the ***n***th digit from the most significant digit. The least significant digit can be a zero.
- The first non-significant digit is the ***n***+1th digit.

Rounding-off rules

- If the first non-significant digit is less than 5, then the least significant digit remains unchanged.
- If the first non-significant digit is greater than 5, the least significant digit is incremented by 1.
- If the first non-significant digit is 5, the least significant digit can either be incremented or left unchanged (*see below!*)
- All non-significant digits are removed.

Fantasies about fives

Students are sometimes told to increment the least significant digit by 1 if it is odd, and to leave it unchanged if it is even. One wonders if this reflects some idea that even numbers are somehow “better” than odd ones! (The ancient superstition is just the opposite, that only the odd numbers are "lucky".)

In fact, you could do it equally the other way around, incrementing only the even numbers. If you are only rounding a single number, it doesn't really matter what you do. However, when you are rounding a series of numbers that will be used in a calculation, if you treated each first nonsignificant 5 in the same way, you would be over- or understating the value of the rounded number, thus accumulating round-off error. Since there are equal numbers of even and odd digits, incrementing only the one kind will keep this kind of error from building up. You could do just as well, of course, by flipping a coin!

Table 4.6.1*: Examples of rounding-off*

number to round	number of significant digits	result	comment
34.216	3	34.2	First non-significant digit (1) is less than 5, so number is simply truncated.
2.252	2	2.2 or 2.3	First non-significant digit is 5, so least sig. digit can either remain unchanged or be incremented.
39.99	3	40.0	Crossing "decimal boundary", so all numbers change.
85,381	3	85,400	The two zeros are just placeholders
0.04597	3	0.0460	The two leading zeros are not significant digits.

Rounding up the Nines

Suppose that an object is found to have a weight of 3.98 ± 0.05 g. This would place its true weight somewhere in the range of **3.93** g to **4.03** g. In judging how to round this number, you count the number of digits in “3.98” that are known exactly, and you find none! Since the “4” is the left most digit whose value is uncertain, this would imply that the result should be rounded to one significant figure and reported simply as 4 g. An alternative would be to bend the rule and round off to two significant digits, yielding 4.0 g. How can you decide what to do? In a case such as this, you should look at the implied uncertainties in the two values, and compare them with the uncertainty associated with the original measurement.

Table 4.6.2

rounded value	implied max	implied min	absolute uncertainty	relative uncertainty
3.98 g	3.985 g	3.975 g	±.005 g or 0.01 g	1 in 400, or 0.25%
4 g	4.5 g	3.5 g	±.5 g or 1 g	1 in 4, 25%
4.0 g	**4.05 g**	**3.95 g**	**±.05 g or 0.1 g**	**1 in 40, 2.5%**

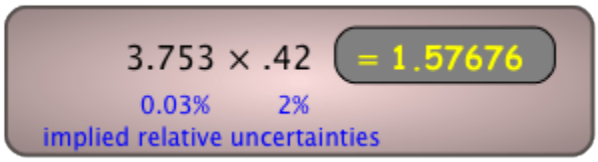
Clearly, rounding off to two digits is the only reasonable course in this example. Observed values should be rounded off to the

number of digits that most accurately conveys the uncertainty in the measurement.

- Usually, this means rounding off to the number of significant digits in in the quantity; that is, the number of digits (counting from the left) that are known exactly, plus one more.
- When this cannot be applied (as in the example above when addition of subtraction of the absolute uncertainty bridges a power of ten), then we round in such a way that the relative implied uncertainty in the result is as close as possible to that of the observed value.

Rounding the Results of Calculations

When carrying out calculations that involve multiple steps, you should avoid doing any rounding until you obtain the final result. Suppose you use your calculator to work out the area of a rectangle:



rounded value	relative implied uncertainty
1.58	1 part in 158, or 0.6%
1.6	1 part in 16, or 6 %

Note

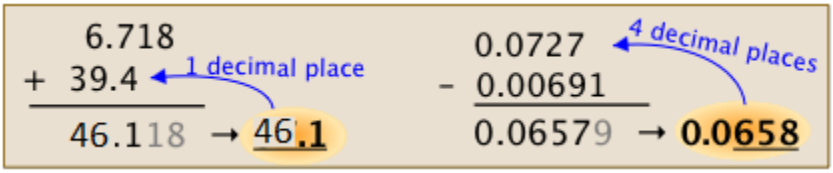
Your calculator is of course correct as far as the pure numbers go, but you would be wrong to write down "1.57676 cm^2" as the answer. Two possible options for rounding off the calculator answer are shown at the right.

It is clear that neither option is entirely satisfactory; rounding to 3 significant digits overstates the precision of the answer, whereas following the rule and rounding to the two digits in ".42" has the effect of throwing away some precision. In this case, it could be argued that rounding to three digits is justified because the implied relative uncertainty in the answer, 0.6%, is more consistent with those of the two factors.

> The "rules" for rounding off are generally useful, convenient guidelines, but they do not always yield the most desirable result. When in doubt, it is better to rely on relative implied uncertainties.

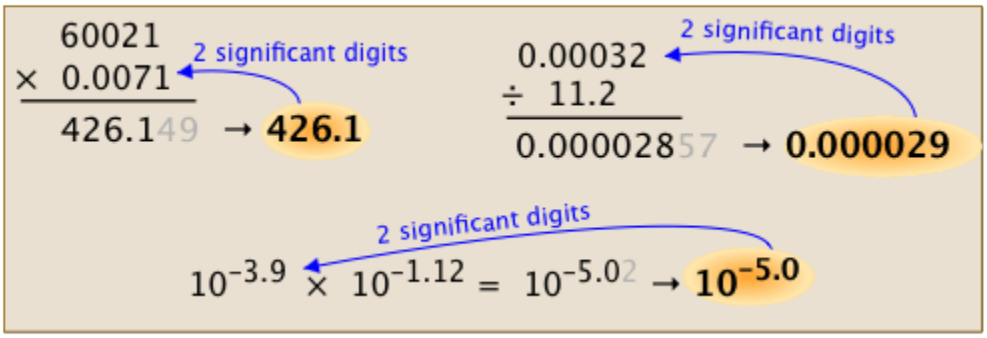
Addition and Subtraction

In operations involving significant figures, the answer is reported in such a way that it reflects the reliability of the **least precise** operation. An answer is no more precise that the least precise number used to get the answer. When adding or subtracting, we go by the number of *decimal places* (i.e., the number of digits on the right side of the decimal point) rather than by the number of significant digits. Identify the quantity having the smallest number of decimal places, and use this number to set the number of decimal places in the answer.



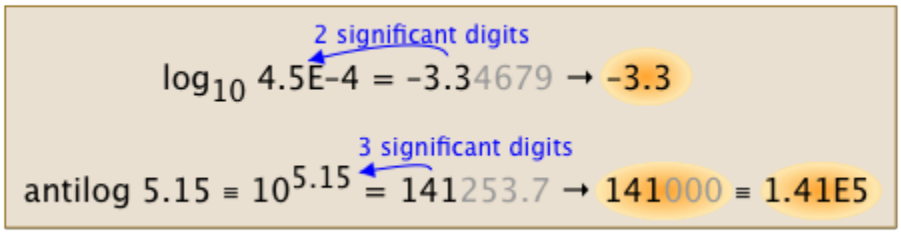
Multiplication and Division

The result must contain the same number of significant figures as in the value having the least number of significant figures.



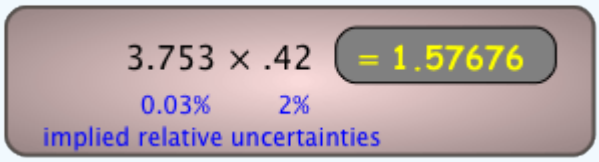
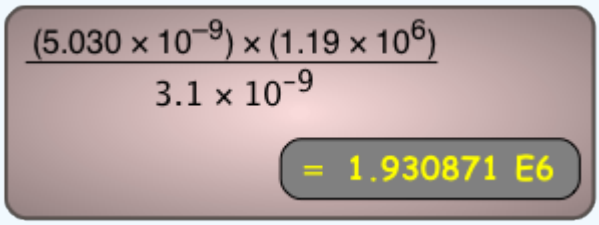
Logarithms and antilogarithms

If a number is expressed in the form $a \times 10^b$ ("scientific notation") with the additional restriction that the coefficient a is no less than 1 and less than 10, the number is in its *normalized* form. Express the base-10 logarithm of a value using the same number of significant figures as is present in the *normalized form* of that value. Similarly, for antilogarithms (numbers expressed as powers of 10), use the same number of significant figures as are in that power.



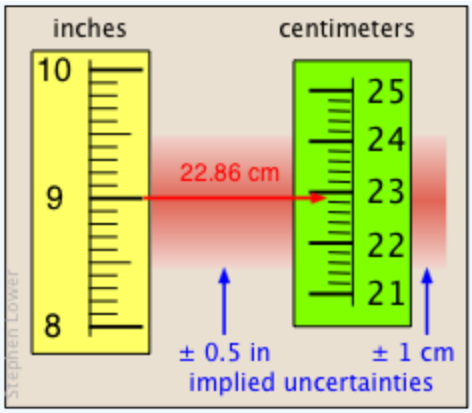
Examples 4.6.1

The following examples will illustrate the most common problems you are likely to encounter in rounding off the results of calculations. They deserve your careful study!

calculator result	rounded	remarks
	1.6	Rounding t two significant figures yields an implied uncertainty of 1/16 or 6%, three times greater than that ir the least-precisely known factor. This is a good illustration of how rounding can lead to the loss of informatior
	1.9E6	The "3.1" factor is specified t 1 part in 31 or 3%. In the answer 1.9, the value is expressed to 1 part in 19, or 5%. These precisions are comparable so the rounding-off rule has given us a reasonable result.

A certain book has a thickness of 117 mm; find the height of a stack of 24 identical books: $(24\ \text{books}) \times \frac{(117\ \text{cm})}{(1\ \text{book})} = 2808\ \text{cm}$	2810 mm	The “24” and the “1” are exact, so the only uncertain value is the thickness o each book, given to 3 significant digits. The trailing zer in the answer is only a placeholde
$7.01 + 0.0007 + 3.4 = 10.4107$	10.4	In addition or subtraction look for the term having the smalles number of decimal places, and round off the answer to the same number of places.
$(9\ \text{in}) \times \frac{(2.54\ \text{cm})}{(1\ \text{in})} = 22.86\ \text{cm}$		
	23 cm	see below

The last of the examples shown above represents the very common operation of converting one unit into another. There is a certain amount of ambiguity here; if we take "9 in" to mean a distance in the range 8.5 to 9.5 inches, then the implied uncertainty is ±0.5 in, which is 1 part in 18, or about ± 6%. The relative uncertainty in the answer must be the same, since all the values are multiplied by the same factor, 2.54 cm/in. In this case we are justified in writing the answer to two significant digits, yielding an uncertainty of about ±1 cm; if we had used the answer "20 cm" (one significant digit), its implied uncertainty would be ±5 cm, or ±25%.



When the appropriate number of significant digits is in question, calculating the relative uncertainty can help you decide.

This page titled 4.6: Significant Figures and Rounding is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5: ATOMS AND THE PERIODIC TABLE

Everything you need to know in a first-year college course about the principal concepts of quantum theory as applied to the atom, and how this determines the organization of the periodic table.

5.1: Primer on Quantum Theory

5.2: Quanta - A New View of the World

5.3: Light, Particles, and Waves

5.4: The Bohr Atom

5.5: The Quantum Atom

5.6: Atomic Electron Configurations

5.7: Periodic Properties of the Elements

5.8: Why Don't Electrons Fall into the Nucleus?

This page titled 5: Atoms and the Periodic Table is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Primer on Quantum Theory

Part 1: Particles and waves

Q1. What is a particle?

A **particle** is a discrete unit of matter having the attributes of **mass**, **momentum** (and thus **kinetic energy**) and optionally of **electric charge**.

Q2. What is a wave?

A wave is a periodic variation of some quantity as a function of location or time. For example, the wave motion of a vibrating guitar string is defined by the displacement of the string from its center as a function of distance along the string. A sound wave consists of variations in the pressure with location.

A wave is characterized by its **wavelength \(**λ\) (lambda) and **frequency** \(\nu**)** (nu), which are connected by the relation

\[\lmabda =\dfrac{u}{\nu}

in which *u* is the velocity of propagation of the disturbance in the medium.

Example

The velocity of sound in the air is 330 m s^{-1}. What is the wavelength of A440 on the piano keyboard?

Solution

$$\lambda = \frac{330\, mS^{-1}}{440\, s^{-1}} = 0.75\, m \tag{5.1.1}$$

Two other attributes of waves are the *amplitude* (the height of the wave crests with respect to the base line) and the *phase*, which measures the position of a crest with respect to some fixed point. The square of the amplitude gives the *intensity* of the wave: the energy transmitted per unit time).

A unique property of waves is their ability to combine constructively or destructively, depending on the relative phases of the combining waves.

Q3. What is light?

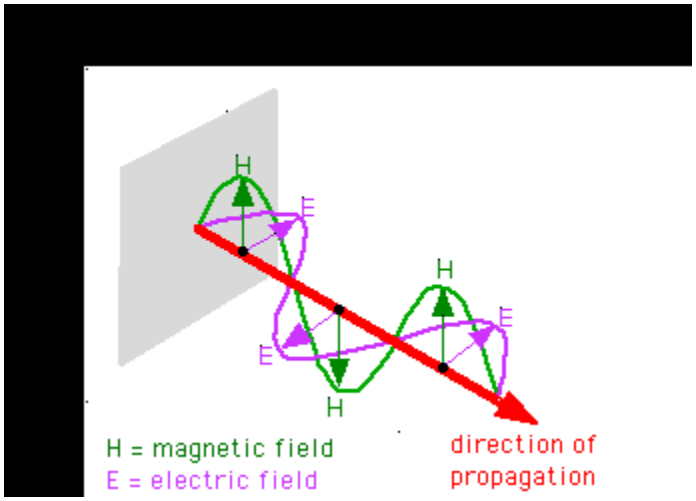
Phrasing the question in this way reflects the deterministic mode of Western thought which assumes that something cannot "be" two quite different things at the same time. The short response to this question is that all we know about light (or anything else, for that matter) are the results of experiments, and that some kinds of experiments show that light behaves like **particles**, and that other experiments reveal light to have the properties of **waves**.

Q4. What is the wave theory of light?

In the early 19th century, the English scientist Thomas Young carried out the famous two-slit experiment which demonstrated that a beam of light, when split into two beams and then recombined, will show interference effects that can only be explained by assuming that light is a wavelike disturbance. By 1820, Augustin Fresnel had put this theory on a sound mathematical basis, but the exact nature of the waves remained unclear until the 1860's when James Clerk Maxwell developed his **electromagnetic theory**.

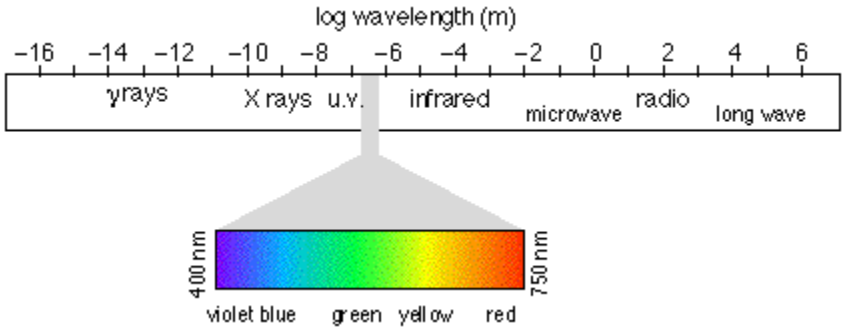
From the laws of electromagnetic induction that were discovered in the period 1820-1830 by Hans Christien Oersted and Michael Faraday, it was known that a moving electric charge gives rise to a magnetic field, and that a changing magnetic field can induce electric charges to move. Maxwell showed theoretically that when an electric charge is accelerated (by being made to oscillate within a piece of wire, for example), electrical energy will be lost, and an equivalent amount of energy is radiated into space, spreading out as a series of waves extending in all directions.

What is "waving" in electromagnetic radiation? According to Maxwell, it is the strengths of the electric and magnetic fields as they travel through space. The two fields are oriented at right angles to each other and to the direction of travel.



As the electric field changes, it induces a magnetic field, which then induces a new electric field, etc., allowing the wave to propagate itself through space

These waves consist of periodic variations in the electrostatic and electromagnetic field strengths. These variations occur at right angles to each other. Each electrostatic component of the wave induces a magnetic component, which then creates a new electrostatic component, so that the wave, once formed, continues to propagate through space, essentially feeding on itself. In one of the most brilliant mathematical developments in the history of science, Maxwell expounded a detailed theory, and even showed that these waves should travel at about 3E8 m s^{-1}, a value which experimental observations had shown corresponded to the speed of light. In 1887, the German physicist Heinrich Hertz demonstrated that an oscillating electric charge (in what was in essence the world's first radio transmitting antenna) actually does produce electromagnetic radiation just as Maxwell had predicted, and that these waves behave exactly like light.



It is now understood that *light* is electromagnetic radiation that falls within a range of wavelengths that can be perceived by the eye. The entire electromagnetic spectrum runs from radio waves at the long-wavelength end, through heat, light, X-rays, and to gamma radiation.

Part 2: Quantum theory of light

Q5. How did the quantum theory of light come about?

It did not arise from any attempt to explain the behavior of light itself; by 1890 it was generally accepted that the electromagnetic theory could explain all of the properties of light that were then known.

Certain aspects of the interaction between light and matter that were observed during the next decade proved rather troublesome, however. The relation between the temperature of an object and the peak wavelength emitted by it was established empirically by Wilhelm Wien in 1893. This put on a quantitative basis what everyone knows: the hotter the object, the "bluer" the light it emits.

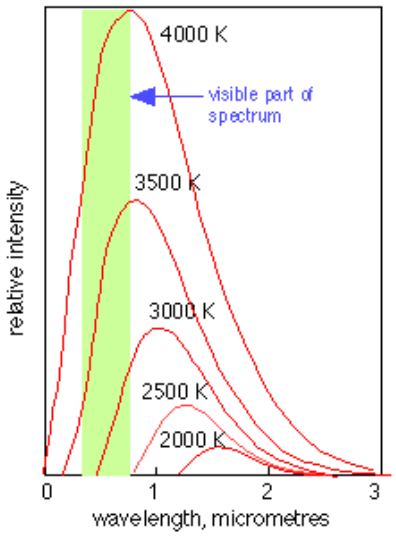
Q6. What is black body radiation?

All objects above the temperature of absolute zero emit electromagnetic radiation consisting of a broad range of wavelengths described by a distribution curve whose peak wavelength l at absolute temperature *T* for a "perfect radiator" known as a black body is given by Wien's law.

$$\lambda_{peak}(cm) = 0.0029\, m\, K \tag{5.1.2}$$

At ordinary temperatures this radiation is entirely in the infrared region of the spectrum, but as the temperature rises above about 1000K, more energy is emitted in the visible wavelength region and the object begins to glow, first with red light, and then shifting toward the blue as the temperature is increased.

This type of radiation has two important characteristics. First, the spectrum is a continuous one, meaning that all wavelengths are emitted, although with intensities that vary smoothly with wavelength. The other curious property of black body radiation is that it is independent of the composition of the object; all that is important is the temperature.



Q7. How did black body radiation lead to quantum physics?

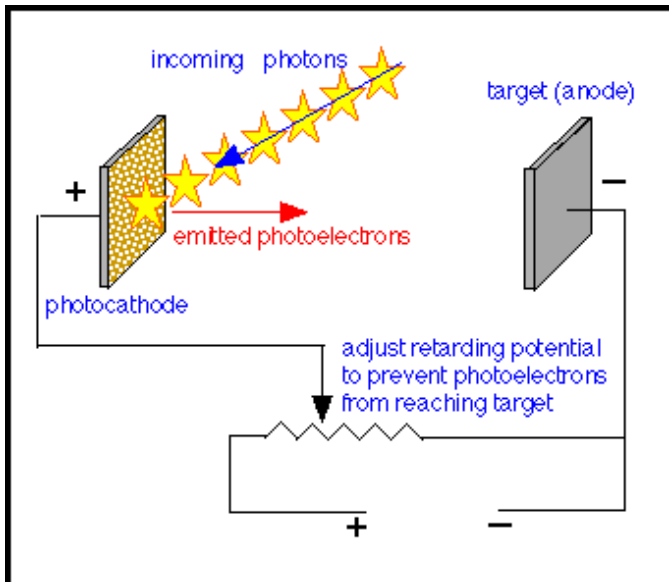
Black body radiation, like all electromagnetic radiation, must originate from oscillations of electric charges which in this case were assumed to be the electrons within the atoms of an object acting somewhat as miniature Hertzian oscillators. It was presumed that since all wavelengths seemed to be present in the continuous spectrum of a glowing body, these tiny oscillators could send or receive any portion of their total energy. However, all attempts to predict the actual shape of the emission spectrum of a glowing object on the basis of classical physical theory proved futile.



In 1900, the great German physicist Max Planck (who earlier in the same year had worked out an empirical formula giving the detailed shape of the black body emission spectrum) showed that the shape of the observed spectrum could be exactly predicted if the energies emitted or absorbed by each oscillator were restricted to integral values of hν, where ν ("nu") is the frequency and *h* is a constant 6.626E–34 J s which we now know as *Planck's Constant*. The allowable energies of each oscillator are quantized, but the emission spectrum of the body remains continuous because of differences in frequency among the uncountable numbers of oscillators it contains.This modification of classical theory, the first use of the quantum concept, was as unprecedented as it was simple, and it set the stage for the development of modern quantum physics.

Q8. What is the photoelectric effect?

Shortly after J.J. Thompson's experiments led to the identification of the elementary charged particles we now know as electrons, it was discovered that the illumination of a metallic surface by light can cause electrons to be emitted from the surface. This phenomenon, the photoelectric effect, is studied by illuminating one of two metal plates in an evacuated tube.



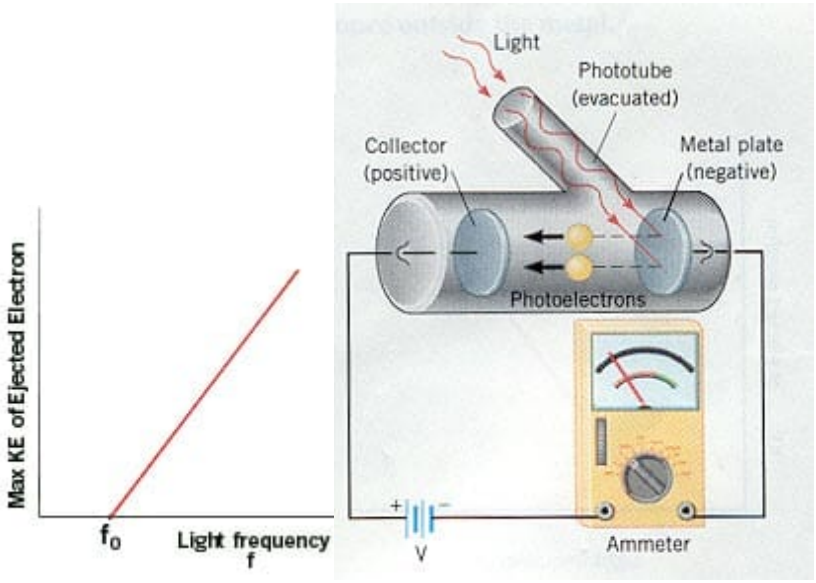
The kinetic energy of the photoelectrons causes them to move to the opposite electrode, thus completing the circuit and producing a measurable current. However, if an opposing potential (the *retarding potential*) is imposed between the two plates, the kinetic energy can be reduced to zero so that the electron current is stopped. By observing the value of the retarding potential V_r, the kinetic energy of the photoelectrons can be calculated from the electron charge *e*, its mass *m* and the frequency of the incident light:

$$E = \frac{n^2 h^2}{8 m_e a^2}$$

$$V_r e = \frac{m_e v_{max}}{2}$$

These two diagrams are taken from a web page by Joseph Alward of the University of the Pacific.

The plot at the right shows how the kinetic energy of the photoelectrons falls to zero at the critical wavelength corresponding to frequency f_0.



Q9. What peculiarity of the photoelectric effect led to the photon?

Although the number of electrons ejected from the metal surface per second depends on the intensity of the light, as expected, the kinetic energies of these electrons (as determined by measuring the retarding potential needed to stop them) does not, and this was definitely not expected. Just as a more intense physical disturbance will produce higher energy waves on the surface of the ocean, it was supposed that a more intense light beam would confer greater energy on the photoelectrons. But what was found, to everyone's surprise, is that the photoelectron energy is controlled by the wavelength of the light, and that there is a critical wavelength below which no photoelectrons are emitted at all.

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

Albert Einstein quickly saw that if the kinetic energy of the photoelectrons depends on the wavelength of the light, then so must its energy. Further, if Planck was correct in supposing that energy must be exchanged in packets restricted to certain values, then light must similarly be organized into energy packets. But a light ray consists of electric and magnetic fields that spread out in a uniform, continuous manner; how can a continuously-varying wave front exchange energy in discrete amounts? Einstein's answer was that the energy contained in each packet of the light must be concentrated into a tiny region of the wave front. This is tantamount to saying that light has the nature of a quantized particle whose energy is given by the product of Planck's constant and the frequency:

$$e = h\nu = \frac{hc}{\lambda}$$

Einstein's publication of this explanation in 1905 led to the rapid acceptance of Planck's idea of energy quantization, which had not previously attracted much support from the physics community of the time. It is interesting to note, however, that this did not make Planck happy at all. Planck, ever the conservative, had been reluctant to accept that his own quantized-energy hypothesis was much more than an artifice to explain black-body radiation; to extend it to light seemed an absurdity that would negate the well-established electromagnetic theory and would set science back to the time before Maxwell.

Q10. Where does relativity come in?

Einstein's special theory of relativity arose from his attempt to understand why the laws of physics that describe the current induced in a fixed conductor when a magnet moves past it are not formulated in the same way as the ones that describe the magnetic field produced by a moving conductor. The details of this development are not relevant to our immediate purpose, but some of the conclusions that this line of thinking led to very definitely are. Einstein showed that the velocity of light, unlike that of a material body, has the same value no matter what velocity the observer has. Further, the mass of any material object, which had previously been regarded as an absolute, is itself a function of the velocity of the body relative to that of the observer (hence "relativity"), the relation being given by

$$m = \frac{m_0}{\sqrt{1 - v^2 / c^2}}$$

in which m_o is the rest mass of the particle, v is its velocity with respect to the observer, and c is the velocity of light.

According to this formula, the mass of an object increases without limit as the velocity approaches that of light. Where does the increased mass come from? Einstein's answer was that the increased mass is that of the kinetic energy of the object; that is, energy itself has mass, so that mass and energy are equivalent according to the famous formula

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

The only particle that can move at the velocity of light is the photon itself, due to its zero rest mass.

Q11. Can the mass-less photon have momentum?

Although the photon has no rest mass, its energy, given by , confers upon it an effective mass of

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

and a momentum of

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

Q12. If waves can be particles, can particles be waves?

In 1924, the French physicist Louis de Broglie proposed (in his doctoral thesis) that just as light possesses particle-like properties, so should particles of matter exhibit a wave-like character. Within two years this hypothesis had been confirmed experimentally by observing the diffraction (a wave interference effect) produced by a beam of electrons as they were scattered by the row of atoms at the surface of a metal. $E = \frac{n^2 h^2}{8 m_e d^2}$

deBroglie showed that the wavelength of a particle is inversely proportional to its momentum:

$$\lambda = \frac{h}{mv}$$

Notice that the wavelength of a stationary particle is infinitely large, while that of a particle of large mass approaches zero. For most practical purposes, the only particle of interest to chemistry that is sufficiently small to exhibit wavelike behavior is the electron (mass 9.11E31 kg).

Q13. Exactly what is it that is "waving"?

We pointed out earlier that a wave is a change that varies with location in a periodic, repeating way. What kind of a change do the crests and hollows of a "matter wave" trace out? The answer is that the wave represents the value of a quantity whose square is a measure of the probability of finding the particle in that particular location. In other words, what is "waving" is the value of a **mathematical probability function**.

Q14. What is the uncertainty principle?

In 1927, Werner Heisenberg proposed that certain pairs of properties of a particle cannot simultaneously have exact values. In particular, the position and the momentum of a particle have associated with them uncertainties x and $E = \frac{n^2 h^2}{8 m_e d^2}$ $E = \frac{n^2 h^2}{8 m_e d^2}$ p given by $E = \frac{n^2 h^2}{8 m_e d^2}$

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

As with the de Broglie particle wavelength, this has practical consequences only for electrons and other particles of very small mass. It is very important to understand that these "uncertainties" are not merely limitations related to experimental error or observational technique, but instead they express an underlying fact that Nature does not allow a particle to possess definite values of position and momentum at the same time. This principle (which would be better described by the term "indeterminacy" than "uncertainty") has been thoroughly verified and has far-reaching practical consequences which extend to chemical bonding and molecular structure.

Q15. Is the uncertainty principle consistent with particle waves?

Yes; either one really implies the other. Consider the following two limiting cases: · A particle whose velocity is known to within a very small uncertainty will have a sharply-defined energy (because its kinetic energy is known) which can be represented by a probability wave having a single, sharply-defined frequency. A "monochromatic" wave of this kind must extend infinitely in space:

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

But if the peaks of the wave represent locations at which the particle is most likely to manifest itself, we are forced to the

conclusion that it can "be" virtually anywhere, since the number of such peaks is infinite! Now think of the opposite extreme: a particle whose location is closely known. Such a particle would be described by a short wave train having only a single peak, the smaller the uncertainty in position, the more narrow the peak.

To help you see how waveforms of different wavelength combine, two such combinations are shown below:

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

It is apparent that as more waves of different frequency are mixed, the regions in which they add constructively diminish in extent. The extreme case would be a wave train in which destructive interference occurs at all locations except one, resulting in a single pulse:

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

Is such a wave possible, and if so, what is its wavelength? Such a wave is possible, but only as the sum (interference) of other waves whose wavelengths are all slightly different. Each component wave possesses its own energy (momentum), and adds that value to the range of momenta carried by the particle, thus increasing the uncertainty δp. In the extreme case of a quantum particle whose location is known exactly, the probability wavelet would have zero width which could be achieved only by combining waves of all wavelengths-- an infinite number of wavelengths, and thus an infinite range of momentum dp and thus kinetic energy.

Q16. Are they particles or are they waves?

Suppose we direct a beam of photons (or electrons; the experiment works with both) toward a piece of metal having a narrow opening. On the other side there are two more openings, or slits. Finally the particles impinge on a photographic plate or some other recording device. Taking into account their wavelike character, we would expect the probability waves to produce an interference pattern of the kind that is well known for sound and light waves, and this is exactly what is observed; the plate records a series of alternating dark and light bands, thus demonstrating beyond doubt that electrons and light have the character of waves.

$E = \frac{n^2 h^2}{8 m_e d^2}$Now let us reduce the intensity of the light so that only one photon at a time passes through the apparatus (it is experimentally possible to count single photons, so this is a practical experiment). Each photon passes through the first slit, and then through one or the other of the second set of slits, eventually striking the photographic film where it creates a tiny dot. If we develop the film after a sufficient number of photons have passed through, we find the very same interference pattern we obtained previously.

There is something strange here. Each photon, acting as a particle, must pass through one or the other of the pair of slits, so we would expect to get only two groups of spots on the film, each opposite one of the two slits. Instead, it appears that the each particle, on passing through one slit, "knows" about the other, and adjusts its final trajectory so as to build up a wavelike interference pattern.

It gets even stranger: suppose that we set up a detector to determine which slit a photon is heading for, and then block off the other slit with a shutter. We find that the photon sails straight through the open slit and onto the film without trying to create any kind of an interference pattern. Apparently, any attempt to observe the photon as a discrete particle causes it to behave like one.

The only conclusion possible is that quantum particles have no well defined paths; each photon (or electron) seems to have an infinity of paths which thread their way through space, seeking out and collecting information about all possible routes, and then adjusting its behavior so that its final trajectory, when combined with that of others, produces the same overall effect that we would see from a train of waves of wavelength $\lambda = h/mv$.

Part 3: Electrons in atoms

Q17. What are line spectra?

We have already seen that a glowing body (or actually, any body whose temperature is above absolute zero) emits and absorbs radiation of all wavelength in a *continuous spectrum*. In striking contrast is the spectrum of light produced when certain substances are volatilized in a flame, or when an electric discharge is passed through a tube containing gaseous atoms of an element. The light emitted by such sources consists entirely of discrete wavelengths. This kind of emission is known as a discrete spectrum or *line spectrum* (the "lines" that appear on photographic images of the spectrum are really images of the slit through which the light passes before being dispersed by the prism in the spectrograph).

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

Every element has its own line spectrum which serves as a sensitive and useful tool for detecting the presence and relative abundance of the element, not only in terrestrial samples but also in stars. (As a matter of fact, the element helium was discovered in the sun, through its line spectrum, before it had been found on Earth.) In some elements, most of the energy in the visible part of the emission spectrum is concentrated into just a few lines, giving their light characteristic colors: yellow-orange for sodium, blue-green for mercury (these are commonly seen in street lights) and orange for neon.

Line spectra were well known early in the 19th century, and were widely used for the analysis of ores and metals. The German spectroscopist R.W. Bunsen, now famous for his gas burner, was then best known for discovering two new elements, rubidium and cesium, from the line spectrum he obtained from samples of mineral spring waters.

Q18. How are line spectra organized?

Until 1885, line spectra were little more than "fingerprints" of the elements; extremely useful in themselves, but incapable of revealing any more than the identify of the individual atoms from which they arise. In that year a Swiss school teacher named Johann Balmer published a formula that related the wavelengths of the four known lines in the emission spectrum of hydrogen in a simple way. Balmer's formula was not based on theory; it was probably a case of cut-and-try, but it worked: he was able to predict the wavelength of a fifth, yet-to-be discovered emission line of hydrogen, and as spectroscopic and astronomical techniques improved (the only way of observing highly excited hydrogen atoms at the time was to observe the solar spectrum during an eclipse), a total of 35 lines were discovered, all having wavelengths given by the formula which we write in the modern manner as

$$\frac{1}{\lambda} = R\left(\frac{1}{m^2} - \frac{1}{n^2}\right)$$

in which m = 2 and R is a constant (the *Rydberg constant*, after the Swedish spectroscopist) whose value is 1.09678E7 m^{-1}. The variable n is an integer whose values 1, 2, etc. give the wavelengths of the different lines.

It was soon discovered that by replacing m with integers other than 2, other series of hydrogen lines could be accounted for. These series, which span the wavelength region from the ultraviolet through infrared, are named after their discoverers.

name of series	when discovered	value of *m*
Lyman	1906-14	1
Balmer	1885	2

Paschen	1908	3
Brackett	1922	4
Pfund	1924	5

Attempts to adapt Balmer's formula to describe the spectra of atoms other than hydrogen generally failed, although certain lines of some of the spectra seemed to fit this same scheme, with the same value of *R*.

Q19. How large can n be?

There is no limit; values in the hundreds have been observed, although doing so is very difficult because of the increasingly close spacing of successive levels as n becomes large. Atoms excited to very high values of n are said to be in *Rydberg states*.

Q20. Why do line spectra become continuous at short wavelengths?

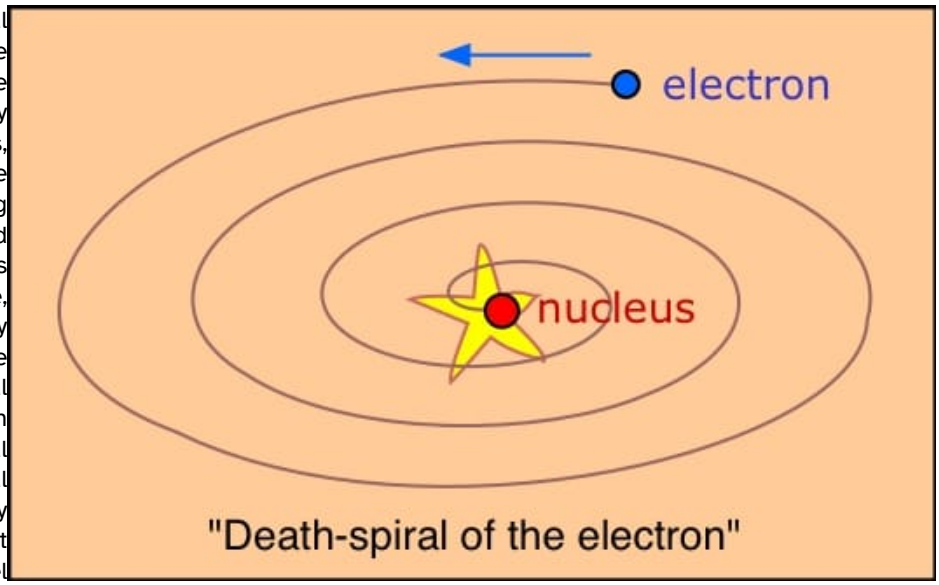
As n becomes larger, the spacing between neighboring levels diminishes and the discrete lines merge into a *continuum*. This can mean only one thing: the energy levels converge as n approaches infinity. This convergence limit corresponds to the energy required to completely remove the electron from the atom; it is the *ionization energy*.

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

At energies in excess of this, the electron is no longer bound to the rest of the atom, which is now of course a positive ion. But an unbound system is not quantized; the kinetic energy of the ion and electron can now have any value in excess of the ionization energy. When such an ion and electron pair recombine to form a new atom, the light emitted will have a wavelength that falls in the continuum region of the spectrum. Spectroscopic observation of the convergence limit is an important method of measuring the ionization energies of atoms.

Q21. What were the problems with the planetary model of the atom?

Rutherford's demonstration that the mass and the positive charge of the atom is mostly concentrated in a very tiny region called the nucleus, forced the question of just how the electrons are disposed outside the nucleus. By analogy with the solar system, a planetary model was suggested: if the electrons were orbiting the nucleus, there would be a centrifugal force that could oppose the electrostatic attraction and thus keep the electrons from falling into the nucleus. This of course is similar to the way in which the centrifugal force produced by an orbiting planet exactly balances the force due to its gravitational attraction to the sun.



The planetary model suffers from one fatal weakness: electrons, unlike planets, are electrically charged. An electric charge revolving in an orbit is continually undergoing a change of direction, that is, acceleration. It has been well known since the time of Hertz that an accelerating electric charge radiates energy. We would therefore expect all atoms to act as miniature radio stations. Even worse, conservation of energy requires that any energy that is radiated must be at the expense of the kinetic energy of the orbital motion of the electron. Thus the electron would slow down, reducing the centrifugal force and allowing the electron to spiral closer and closer to the nucleus, eventually falling into it. In short, no atom that operates according to the planetary model would last long enough for us to talk about it.

As if this were not enough, the planetary model was totally unable to explain any of the observed properties of atoms, including their line spectra.

Q22. How did Bohr's theory save the planetary model... for a while?

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

Niels Bohr was born in the same year (1885) that Balmer published his formula for the line spectrum of hydrogen. Beginning in 1913, the brilliant Danish physicist published a series of papers that would ultimately derive Balmer's formula from first principles.

Bohr's first task was to explain why the orbiting electron does not radiate energy as it moves around the nucleus. This energy loss, if it were to occur at all, would do so gradually and smoothly. But Planck had shown that black body radiation could only be explained if energy changes were limited to jumps instead of gradual changes. If this were a universal characteristic of energy- that is, if all energy changes were quantized, then very small changes in energy would be impossible, so that the electron would in effect be "locked in" to its orbit.

From this, Bohr went on to propose that there are certain stable orbits in which the electron can exist without radiating and thus without falling into a "death spiral". This supposition was a daring one at the time because it was inconsistent with classical physics, and the theory which would eventually lend it support would not come along until the work of de Broglie and Heisenberg more than ten years later.

Since Planck's quanta came in multiples of h, Bohr restricted his allowed orbits to those in which the product of the radius r and the momentum of the electron mv (which has the same units as h, J–s) are integral multiples of h:

$$2\pi rmv = nh \ (n = 1,2,3, \ldots)$$

Each orbit corresponds to a different energy, with the electron normally occupying the one having the lowest energy, which would be the innermost orbit of the hydrogen atom.

Taking the lead from Einstein's explanation of the photoelectric effect, Bohr assumed that each spectral line emitted by an atom that has been excited by absorption of energy from an electrical discharge or a flame represents a change in energy given by $\Delta E = hn = hc/\lambda$, the energy lost when the electron falls from a higher orbit (value of n) into a lower one.

Finally, as a crowning triumph, Bohr derived an expression giving the radius of the nth orbit for the electron in hydrogen as

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

Substitution of the observed values of the electron mass and electron charge into this equation yielded a value of 0.529E10 m for the radius of the first orbit, a value that corresponds to the radius of the hydrogen atom obtained experimentally from the kinetic theory of gases. Bohr was also able to derive a formula giving the value of the Rydberg constant, and thus in effect predict the entire emission spectrum of the hydrogen atom.

Q23. What were the main problems with Bohr's theory?

There were two kinds of difficulties. First, there was the practical limitation that it only works for atoms that have one electron--that is, for H, He^+, Li^{2+}, etc. The second problem was that Bohr was unable to provide any theoretical justification for his assumption that electrons in orbits described by the preceding equation would not lose energy by radiation. This reflects the fundamental underlying difficulty: because de Broglie's picture of matter waves would not come until a decade later, Bohr had to regard the electron as a classical particle traversing a definite orbital path.

Q24. How did the wave picture of the electron save Bohr's theory?

Once it became apparent that the electron must have a wavelike character, things began to fall into place. The possible states of an electron confined to a fixed space are in many ways analogous to the allowed states of a vibrating guitar string. These states are described as standing waves that must possess integral numbers of nodes. The states of vibration of the string are described by a series of integral numbers n = 1,2,... which we call the fundamental, first overtone, second overtone, etc. The energy of vibration is proportional to n^2. Each mode of vibration contains one more complete wave than the one below it.

In exactly the same way, the mathematical function that defines the probability of finding the electron at any given location within a confined space possesses n peaks and corresponds to states in which the energy is proportional to n^2.

The electron in a hydrogen atom is bound to the nucleus by its spherically symmetrical electrostatic charge, and should therefore exhibit a similar kind of wave behavior. This is most easily visualized in a two-dimensional cross section that corresponds to the conventional electron orbit. But if the particle picture is replaced by de Broglie's probability wave, this wave must follow a circular path, and- most important of all- its wavelength (and consequently its energy) is restricted to integral multiples n = 1,2,.. of the circumference $2\pi r = n\lambda$.

for otherwise the wave would collapse owing to self-interference. That is, the energy of the electron must be quantized; what Bohr had taken as a daring but arbitrary assumption was now seen as a fundamental requirement. Indeed the above equation can be derived very simply by combining Bohr's quantum condition $2\pi rmv = nh$ with the expression $mv = h/\lambda$ for the deBroglie wavelength of a particle.

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

Viewing the electron as a standing-wave pattern also explains its failure to lose energy by radiating. Classical theory predicts that an accelerating electric charge will act as a radio transmitter; an electron traveling around a circular wire would certainly act in this way, and so would one rotating in an orbit around the nucleus. In a standing wave, however, the charge is distributed over space in a regular and unchanging way; there is no motion of the charge itself, and thus no radiation.

Q25. What is an orbital?

Because the classical view of an electron as a localizable particle is now seen to be untenable, so is the concept of a definite trajectory, or "orbit". Instead, we now use the word *orbital* to describe the state of existence of an electron. An orbital is really no more than a mathematical function describing the standing wave that gives the probability of the electron manifesting itself at any given location in space. More commonly (and loosely) we use the word to describe the region of space in which an electron is likely to be found. Each kind of orbital is characterized by a set of quantum numbers *n, l*, and *m* These relate, respectively, to the average distance of the electron from the nucleus, to the shape of the orbital, and to its orientation in space.

Q26. If the electron cannot be localized, can it be moving?

In its lowest state in the hydrogen atom (in which l=0) the electron has zero angular momentum, so electrons in s orbitals are not in motion. In orbitals for which l>0 the electron does have an effective angular momentum, and since the electron also has a definite rest mass m_e = 9.11E31 kg, it must possess an effective velocity. Its value can be estimated from the Uncertainty Principle; if the volume in which the electron is confined is about 10^{-10} m, then the uncertainty in its momentum is at least $h/(10^{10})$ = 6.6E–24 kg m s^{-1}, which implies a velocity of around 10^{7} m s^{-1}, or almost one-tenth the velocity of light.

The stronger the electrostatic force of attraction by the nucleus, the faster the effective electron velocity. In fact, the innermost electrons of the heavier elements have effective velocities so high that relativistic effects set in; that is, the effective mass of the electron significantly exceeds its rest mass. This has direct chemical effects; it is the cause, for example, of the low melting point of metallic mercury and of the color of gold.

Q27. Why does the electron not fall into the nucleus?

The negatively-charged electron is attracted to the positive charge of the nucleus. What prevents it from falling in? This question can be answered in various ways at various levels. All start with the statement that the electron, being a quantum particle, has a dual character and cannot be treated solely by the laws of Newtonian mechanics.

We saw above that in its wavelike guise, the electron exists as a standing wave which must circle the nucleus at a sufficient distance to allow at least one wavelength to fit on its circumference. This means that the smaller the radius of the circle, the shorter must be the wavelength of the electron, and thus the higher the energy. Thus it ends up "costing" the electron energy if it gets too close to the nucleus. The normal orbital radius represents the balance between the electrostatic force trying to pull the electron in, and what we might call the "confinement energy" that opposes the electrostatic energy. This confinement energy can be related to both the particle and wave character of the electron.

If the electron as a particle were to approach the nucleus, the uncertainty in its position would become so small (owing to the very small volume of space close to the nucleus) that the momentum, and therefore the energy, would have to become very large. The electron would, in effect, be "kicked out" of the nuclear region by the confinement energy.

The standing-wave patterns of an electron in a box can be calculated quite easily. For a spherical enclosure of diameter d, the energy is given by

$$E = \frac{n^2 h^2}{8 m_e d^2}$$

in which n = 1,2,3. etc.

Q28. What is electron spin?

Each electron in an atom has associated with it a magnetic field whose direction is quantized; there are only two possible values that point in opposite directions. We usually refer to these as "up" and "down", but the actual directions are parallel and antiparallel to the local magnetic field associated with the orbital motion of the electron.

The term *spin* implies that this magnetic moment is produced by the electron charge as the electron rotates about its own axis.

Although this conveys a vivid mental picture of the source of the magnetism, the electron is not an extended body and its rotation is meaningless. Electron spin has no classical counterpart and no simple explanation; the magnetic moment is a consequence of relativistic shifts in local space and time due to the high effective velocity of the electron in the atom. This effect was predicted theoretically by P.A.M. Dirac in 1928.

This page titled 5.1: Primer on Quantum Theory is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.2: Quanta - A New View of the World

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- What was the ***caloric theory of heat*****,** and how did Rumford's experiments in boring cannon barrels lead to its overthrow?
- Define ***thermal radiation*** and the "scandal of the ultraviolet" and the role Max Planck played in introducing the ***quantum*** concept.
- What is the ***photoelectric effect***? Describe the crucial insight that led Einstein to the concept of the ***photon***.

What we call "classical" physics is based on our experience of what we perceive as the "real world". Even without knowing the details of Newton's laws of motion that describe the behavior of macroscopic bodies, we have all developed an intuitive understanding of this behavior; it is a part of everyone's personal view of the world. By extension, we tend to view atoms and molecules in much the same way, that is, simply as miniature versions of the macroscopic objects we know from everyday life. It turns out, however, that our everyday view of the macroscopic world is only a first approximation of the reality that becomes apparent at the atomic level. Many of those who first encounter this microscopic world of quantum weirdness find it so foreign to prior experience that their first reaction is to dismiss it as pure fantasy.

The fact is, however, that it is not only for real, but serves as the key that unlocks even some of the simplest aspects of modern Chemistry. Our goal in this lesson is to introduce you to this new reality, and to provide you with a conceptual understanding of it that will make Chemistry a more meaningful part of your own personal world.

The Limits of Classical Physics

Near the end of the nineteenth century, the enormous success of the recently developed kinetic molecular theory of gases had dispelled most doubts about the atomic nature of matter; the material world was seen to consist of particles that had distinct masses and sizes, and which moved in trajectories just as definite as those of billiard balls.

In the 1890s, however, certain phenomena began to be noticed that seemed to be inconsistent with this dichotomy of particles and waves. This prompted further questions and further experiments which led eventually to the realization that classical physics, while it appears to be "the truth", is by no means the whole truth. In particular, it cannot accurately describe the behavior of objects that are extremely small or fast-moving.

Chemistry began as an entirely empirical, experimental science, dealing with the classification and properties of substances and with their transformations in chemical reactions. As this large body of facts developed into a science (one of whose functions is always to explain and correlate known facts and to predict new ones), it has become necessary to focus increasingly on the nature and behavior of individual atoms and of their own constituent parts, especially the *electrons*. Owing to their extremely small masses, electrons behave as *quantum particles* which do not obey the rules of classical physics.

The purpose of this introductory unit is to summarize the major ideas of quantum theory that will be needed to treat atomic and molecular structure later on in the course.

Quantum theory can be presented simply as a set of assumptions which are developed through mathematical treatment. This is in fact the best route to take if one is to use quantum mechanics as a working tool. More than this, however, quantum theory brings with it a set of *concepts* that have far-reaching philosophical implications and which should be a part of the intellectual equipment of anyone who claims to have a general education in the sciences. A major objective of this chapter will be to introduce you to "the quantum way of thinking" and to show how this led to a profound break with the past, and a shift in our way of viewing the world that has no parallel in Western intellectual history.

Light

The development of our ideas about light and radiation was not quite as direct. In the 17th century, heat was regarded as a substance called *caloric* whose invisible atoms could flow from one object to another, thus explaining thermal conduction. This view of heat as a material fluid seemed to be confirmed by the observation that heat can pass through a vacuum, a phenomenon that we now call *radiant heat*. Isaac Newton, whose experiments with a prism in 1672 led to his famous textbook "*Optiks*", noted that light seemed to react with green plants to produce growth, and must therefore be a "substance" having atoms of its own. By 1800, the *corpuscular (particle) theory of light* was generally accepted.

And yet there were questions. Count Rumford's observation that the drill bits employed in boring cannons produced more frictional heat when they were worn and dull led to the overthrow of the caloric theory.

The caloric theory of heat assumed that small particles are able to contain more heat than large ones, so that when a material is sawn or drilled, some of its heat is released as the filings are produced. A dull drill produces few filings, and according to this theory, should produce little heat, but Rumford was able to show that the amount of heat produced is in fact independent of the state of the drill, and depends only on the amount of mechanical work done in turning it.

In 1812, Christiaan Huygens showed how a number of optical effects could be explained if light had a wavelike nature, and this led Fresnel to develop an elaborate wave theory of light. By 1818 the question of "particle or wave" had become so confused that the French Academy held a great debate intended to settle the matter once for all. The mathematician Poisson pointed out that Fresnel's wave theory had a ridiculous consequence: the shadow cast by a circular disk should have a bright spot of light at its center, where waves arriving in phase would reinforce each other. Fresnel performed the experiment and was entirely vindicated: if the light source is sufficiently point-like (an extended source such as the sun or an ordinary lamp will not work), this diffraction effect is indeed observed.

Heat

By this time it was known that radiant heat and "cold" could be focused and transmitted by mirrors, and in 1800 William Herschel discovered that radiant heat could be sensed in the dark region just beyond the red light refracted by a prism. Light and radiant heat, which had formerly been considered separate, were now recognized as one, although the question of precisely what was doing the "waving" was something of an embarrassment.

The quantum revolution

By 1890, physicists thought they had tidied up the world into the two realms of particulate matter and of wavelike radiant energy, which by then had been shown by James Clerk Maxwell to be forms of electromagnetic energy. No sooner had all this been accomplished, than the cracks began to appear; these quickly widened into chasms, and within twenty years the entire foundations of classical physics had disintegrated; it would not be until the 1920's that anyone with a serious interest in the nature of the microscopic world would find a steady place to stand.

Cathode rays

The atom was the first to go. It had been known for some time that when a high voltage is applied to two separated pieces of metal in an evacuated tube, "cathode rays" pass between them. These rays could be detected by their ability to cause certain materials to give off light, or *fluoresce*, and were believed to be another form of electromagnetic radiation. Then, in the 1890s, J.J. Thompson and Jean Perrin showed that cathode rays are composed of particles having a measurable mass (less than 1/1000 of that of the hydrogen atom), they carry a fixed negative electric charge, and that they come from atoms. This last conclusion went so strongly against the prevailing view of atoms as the ultimate, un-cuttable stuff of the world that Thompson only reluctantly accepted it, and having done so, quickly became the object of widespread ridicule.

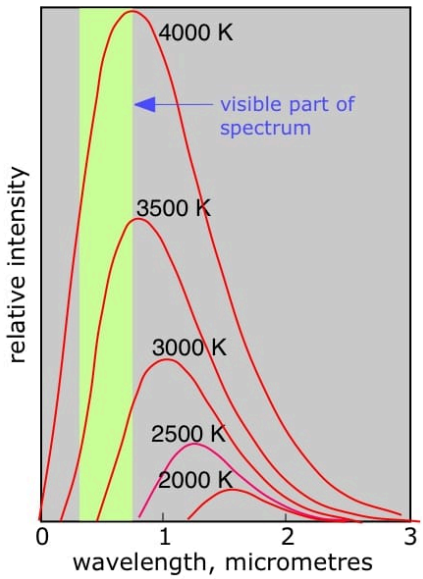
Radioactivity

But worse was soon to come; not only were atoms shown *not* to be the smallest units of matter, but the work of the Curies established that atoms are not even immutable; atoms of high atomic weight such as uranium and radium give off penetrating beams of radiation and in the process change into other elements, disintegrating through a series of stages until they turn into

lead. Among the various kinds of radiation that accompany radioactive disintegration are the very same cathode rays that had been produced artificially by Thompson, and which we now know as *electrons*.

Radiation is quantized

The wave theory of radiation was also running into difficulties. Any object at a temperature above absolute zero gives off radiant energy; if the object is moderately warm, we sense this as *radiant heat*. As the temperature is raised, a larger proportion of shorter-wavelength radiation is given off, so that at sufficiently high temperatures the object becomes luminous. The origin of this radiation was thought to lie in the thermally-induced oscillations of the atoms within the object, and on this basis the mathematical physicist James Rayleigh had worked out a formula that related the wavelengths given off to the temperature. Unfortunately, this formula did not work; it predicted that most of the radiation given off at any temperature would be of very short wavelength, which would place it in the ultraviolet region of the spectrum. What was most disconcerting is that no one could say *why* Rayleigh's formula did not work, based as it was on sound classical physics; this puzzle became known as the "scandal of the ultraviolet".



Thermal emission of radiation. These plots show something you already know: when a piece of metal or some other material becomes very hot, it begins to glow. Actually, all bodies warmer than absolute zero emit thermal radiation, but at ordinary temperatures the wavelengths are in the far-infrared range, well out of the visible region. At about 3000 K the wavelengths become short enough to reveal themselves to the eye: the body becomes “red hot”. At higher temperatures the radiation extends through more of the visible spectrum and the emission becomes yellow and then white. Notice that at all temperatures shown here, the bulk of the radiation is still in the infrared region; this is the reason that incandescent (tungsten filament) lamps are so inefficient; most of the energy goes into the production of heat.

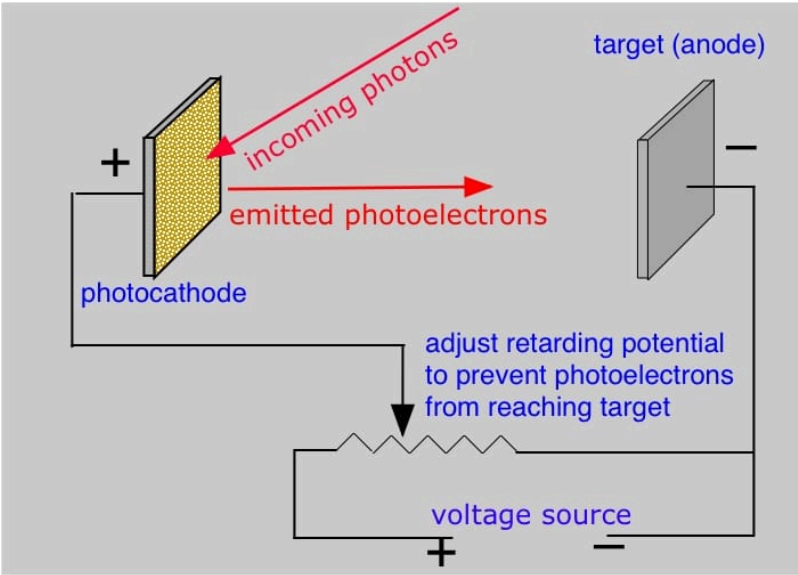
Quanta

In 1899 the German physicist Max Planck pointed out that one simple change in Rayleigh's argument would produce a formula that accurately describes the radiation spectrum of a perfect radiator, which is known as a "black body". Rayleigh assumed that such an object would absorb and emit amounts of radiation in amounts of any magnitude, ranging from minute to very large. This is just what one would expect on the basis of the similar theory of mechanical physics which had long been well established. Planck's change, for which he could offer no physical justification other than that it works, was to discard this assumption, and to require that the absorption or emission of radiation occur only in discrete chunks, or *quanta*. Max Planck had unlocked the door that would lead to the resurrection of the corpuscular theory of radiation. Only a few years later, Albert Einstein would kick the door open and walk through.

The photoelectric effect

By 1900 it was known that a beam of light, falling on a piece of metal, could cause electrons to be ejected from its surface. Evidently the energy associated with the light overcomes the binding energy of the electron in the metal; any energy the light

supplies in excess of this binding energy appears as kinetic energy of the emitted electron. What seemed peculiar, however, was that the energy of the ejected electrons did not depend on the intensity of the light as classical physics would predict. Instead, the energy of the photoelectrons (as they are called) varies with the color, or *wavelength* of the light; the higher the frequency (the shorter the wavelength), the greater the energy of the ejected electrons.



The photoelectric experiment. In order to be knocked out of the metal, an electron must be given a minimum amount of energy (the ionization energy.) Any energy in excess of this will show up as kinetic energy of the photoelectrons. The purpose of the retarding potential is to counteract this kinetic energy so that we can determine what photon wavelength (energy) corresponds to just the ionization energy.

In 1905, Albert Einstein, then an unknown clerk in the Swiss Patent Office, published a remarkable paper in which he showed that if light were regarded as a collection of individual particles, a number of phenomena, including the photoelectric effect, could be explained. Each particle of light, which we now know as a *photon* , has associated with it a distinct energy that is proportional to the frequency of the light, and which corresponds to Planck's energy quanta. The energy of the photon is given by

$$E = hu = \frac{hc}{\lambda} \tag{5.2.1}$$

in which *h* is Planck's constant, 6.63×10^{-34} J-s, ν (Greek *nu*) is the frequency, λ (*lamda*) is the wavelength, and *c* is the velocity of light, 3.00×10^{8} m s^{-1}. The photoelectric effect is only seen if the photon energy *e* exceeds the binding energy of the electron in the metal; it is clear from the above equation that as the wavelength increases, *e* decreases, and eventually no electrons will be released. Einstein had in effect revived the corpuscular theory of light, although it would not be until about 1915 that sufficient experimental evidence would be at hand to convince most of the scientific world— but not all of it: Max Planck, whose work had led directly to the revival of the particle theory of light, remained one of the strongest doubters.

The 1905 volume of *Annalen der Physik* is now an expensive collector's item, for in that year Einstein published three major papers, any one of which would have guaranteed him his place in posterity. The first, on the photoelectric effect, eventually won him the Nobel Prize. The second paper, on Brownian motion, amounted to the first direct confirmation of the atomic theory of matter. The third paper, his most famous, "On the electrodynamics of moving bodies", set forth the special theory of relativity.

The appearance of his general theory of relativity in 1919 would finally make Einstein into a reluctant public celebrity and scientific superstar. This theory explained gravity as a consequence of the curvature of space-time.

Matter and energy united

Energy

The concept of energy was slow to develop in science, partly because it was not adequately differentiated from the related quantities of force and motion. It was generally agreed that some agent of motion and change must exist; Descartes suggested,

for example, that God, when creating the world, had filled it with "vortices" whose motions never ceased, but which could be transferred to other objects and thus give them motion. Gradually the concepts of *vis viva* and *vis mortua* developed; these later became kinetic and potential energy. Later on, the cannon-boring experiments of Benjamin Thompson (Count Rumford) revealed the connections between heat and work. Finally, the invention of the steam engine forced the birth of the science of thermodynamics, whose founding law was that a quantity known as *energy* can be transferred from one object to another through the processes of heat and work, but that the energy itself is strictly conserved.

Relativity

If Einstein's first 1905 paper put him on the scientific map, the third one made him a scientific celebrity.

In effect, Einstein merely asked a simple question about Faraday's law of electromagnetic induction, which says that a moving electric charge (such as is produced by an electric current flowing in a conductor) will create a magnetic field. Similarly, a moving magnetic field will induce an electric current. In either case, something has to be moving. Why, Einstein asked, does this motion have to be relative to that of the room in which the experiment is performed— that is, relative to the Earth? A stationary charge creates no field, but we know that there is really no such thing as a stationary charge, since the Earth itself is in motion; what, then, do motion and velocity ultimately relate to?

The answer, Einstein suggested, is that the only constant and unchanging velocity in the universe is that of light. This being so, the beam emitted by the headlight of a moving vehicle, for example, can travel no faster than the light coming from a stationary one. This in turn suggested (through the Lorentz transformation - we are leaving out a few steps here!) that mass, as well as velocity (and thus also, time) are relative in that they depend entirely on the motion of the observer. Two observers, moving at different velocities relative to each other, will report different masses for the same object, and will age at different rates. Further, the faster an object moves with respect to an observer, the greater is its mass, and the harder it becomes to accelerate it to a still greater velocity. As the velocity of an object approaches the speed of light, its mass approaches infinity, making it impossible for an object to move as fast as light.

According to Einstein, the speed of light is really the *only* speed in the universe. If you are sitting still, you are moving through the *time* dimension at the speed of light. If you are flying in an airplane, your motion along the three cartesian dimensions subtracts from that along the fourth (time) coordinate, with the result that time, for you, passes more slowly.

Relativity comes into chemistry in two rather indirect ways: it is responsible for the magnetic moment ("spin") of the electron, and in high-atomic weight atoms in which the electrons have especially high effective velocities, their greater [relativistic] masses cause them to be bound more tightly to the nucleus— accounting, among other things, for the color of gold, and for the unusual physical and chemical properties of mercury.

Mass-energy

Where does the additional mass of a moving body come from? Simply from the kinetic energy of the object; this equivalence of mass and energy, expressed by the famous relation $e = mc^2$, is the most well known consequence of special relativity. The reason that photons alone can travel at the velocity of light is that these particles possess zero rest mass to start with. You can think of ordinary matter as "congealed energy", trapped by its possession of rest mass, whereas light is energy that has been liberated of its mass.

This page titled 5.2: Quanta - A New View of the World is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.3: Light, Particles, and Waves

Learning Objectives

Make sure you thoroughly understand the following essential ideas

- Cite two pieces of experimental evidence that demonstrate, respectively, the wave- and particle-like nature of light.
- Define the terms *amplitude*, *wavelength*, and *frequency* as they apply to wave phenomena.
- Give a qualitative description of *electromagnetic radiation* in terms of electrostatic and magnetic fields.
- Be able to name the principal regions of the *electromagnetic spectrum* (X-rays, infrared region, etc.) and specify their sequence in terms of either wavelength or energy per photon.
- Describe the difference between *line spectra* and *continuous spectra* in terms of both their appearance and their origins.
- What is meant by the *de Broglie wavelength* of a particle? How will the particle's mass and velocity affect the wavelength?
- State the consequences of the *Heisenberg uncertainty principle* in your own words.

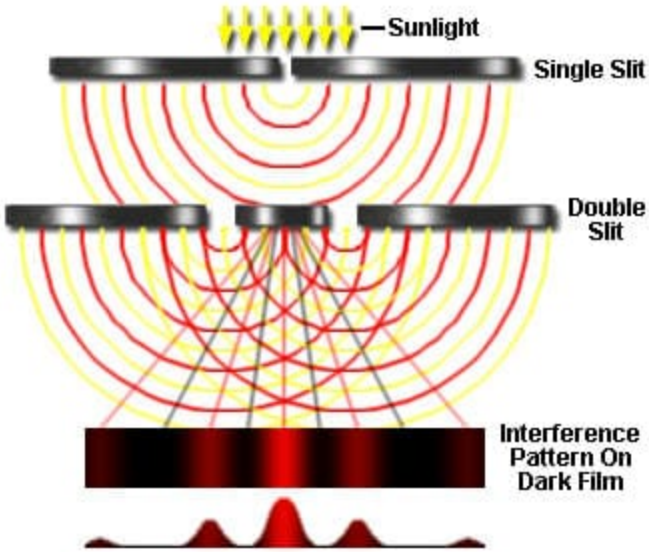
Our intuitive view of the "real world" is one in which objects have definite masses, sizes, locations and velocities. Once we get down to the atomic level, this simple view begins to break down. It becomes totally useless when we move down to the subatomic level and consider the lightest of all chemically-significant particles, the *electron*. The chemical properties of a particular kind of atom depend on the arrangement and behavior of the electrons which make up almost the entire volume of the atom. The electronic structure of an atom can only be determined indirectly by observing the manner in which atoms absorb and emit light. Light, as you already know, has wavelike properties, so we need to know something about waves in order to interpret these observations. But because the electrons are themselves *quantum particles* and therefore have wavelike properties of their own, we will find that an understanding of the behavior of electrons in atoms can only be gained through the language of waves.

The language of light

Atoms are far too small to see directly, even with the most powerful optical microscopes. But atoms do interact with and under some circumstances emit light in ways that reveal their internal structures in amazingly fine detail. It is through the "language of light" that we communicate with the world of the atom. This section will introduce you to the rudiments of this language.

Wave, particle, or what?

In the early 19th century, the English scientist Thomas Young carried out the famous double-slit experiment which demonstrated that a beam of light, when split into two beams and then recombined, will show interference effects that can only be explained by assuming that light is a wavelike disturbance. By 1820, Augustin Fresnel had put this theory on a sound mathematical basis, but the exact nature of the waves remained unclear until the 1860's when James Clerk Maxwell developed his electromagnetic theory.



But Einstein's 1905 explanation of the photoelectric effect showed that light also exhibits a particle-like nature. The *photon* is the smallest possible packet (*quantum*) of light; it has zero mass but a definite energy.

When light-wave interference experiments are conducted with extremely low intensities of light, the wave theory breaks down; instead of recording a smooth succession of interference patterns as shown above, an extremely sensitive detector sees individual pulses— that is, individual *photons*.

Suppose we conduct the double-slit interference experiment using a beam of light so weak that only one photon at a time passes through the apparatus (it is experimentally possible to count single photons, so this is a practical experiment.) Each photon passes through the first slit, and then through one or the other of the second set of slits, eventually striking the photographic film where it creates a tiny dot. If we develop the film after a sufficient number of photons have passed through, we find the very same interference pattern we obtained with higher-intensity light whose behavior was could be explained by wave interference.

There is something strange here. Each photon, acting as a particle, must pass through one or the other of the pair of slits, so we would expect to get only two groups of spots on the film, each opposite one of the two slits. Instead, it appears that the each particle, on passing through one slit, "knows" about the other, and adjusts its final trajectory so as to build up a wavelike interference pattern.

It gets even stranger: suppose that we set up a detector to determine which slit a photon is heading for, and then block off the other slit with a shutter. We find that the photon sails straight through the open slit and onto the film without trying to create any kind of an interference pattern. Apparently, any attempt to observe the photon as a discrete particle causes it to behave like one.

One well-known physicist (Landé) suggested that perhaps we should coin a new word, *wavicle*, to reflect this duality.

Later on, virtually the same experiment was repeated with electrons, thus showing that particles can have wavelike properties (as the French physicist Louis de Broglie predicted in 1923), just as what were conventionally thought to be electromagnetic waves possess particle-like properties.

Is it a particle or is it a wave?

For large bodies (most atoms, baseballs, cars) there is no question: the wave properties are insignificant, and the laws of classical mechanics can adequately describe their behaviors. But for particles as tiny as electrons (*quantum particles*), the situation is quite different: instead of moving along well defined paths, a quantum particle seems to have an infinity of paths which thread their way through space, seeking out and collecting information about all possible routes, and then adjusting its behavior so that its final trajectory, when combined with that of others, produces the same overall effect that we would see from a train of waves of wavelength = h/mv.



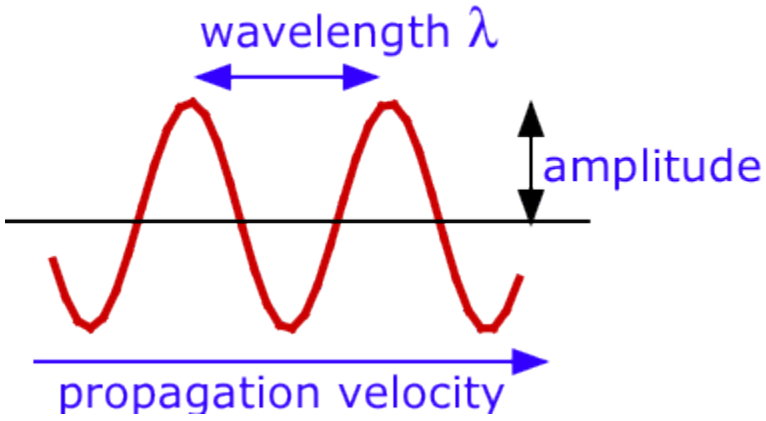
Figure 5.3.1: If cars behaved as quantum particles, this would present no problem at all!

Taking this idea of quantum indeterminacy to its most extreme, the physicist Erwin Schrödinger proposed a "thought experiment" in which the radioactive decay of an atom would initiate a chain of events that would lead to the death of a cat placed in a closed box. The atom has a 50% chance of decaying in an hour, meaning that its wave representation will contain both possibilities until an observation is made. The question, then, is will the cat be simultaneously in an alive-and-dead state until the box is opened? If so, this raises all kinds of interesting questions about the nature of being.

What you need to know about waves

We use the term "wave" to refer to a quantity which changes with time. Waves in which the changes occur in a repeating or periodic manner are of special importance and are widespread in nature; think of the motions of the ocean surface, the pressure variations in an organ pipe, or the vibrations of a plucked guitar string. What is interesting about all such repeating phenomena is that they can be described by the same mathematical equations.

Wave motion arises when a periodic disturbance of some kind is propagated through a medium; pressure variations through air, transverse motions along a guitar string, or variations in the intensities of the local electric and magnetic fields in space, which constitutes electromagnetic radiation. For each medium, there is a characteristic *velocity* at which the disturbance travels.



There are three measurable properties of wave motion: *amplitude*,*wavelength*, and *frequency*, the number of vibrations per second. The relation between the wavelength λ (Greek *lambda*) and frequency of a wave u (Greek *nu*) is determined by the propagation velocity *v*.

$$v = u\lambda \tag{5.3.1}$$

Example 5.3.1

What is the wavelength of the musical note A = 440 hz when it is propagated through air in which the velocity of sound is 343 m s^{-1}?

Solution

$$\lambda = \frac{v}{u} = \frac{343\ m\ s^{-1}}{440\ s^{-1}} = 0.80\ m \tag{5.3.2}$$

Light and electromagnetic radiation

Michael Faraday's discovery that electric currents could give rise to magnetic fields and *vice versa* raised the question of how these effects are transmitted through space. Around 1870, the Scottish physicist James Clerk Maxwell (1831-1879) showed that this electromagnetic radiation can be described as a train of perpendicular oscillating electric and magnetic fields.

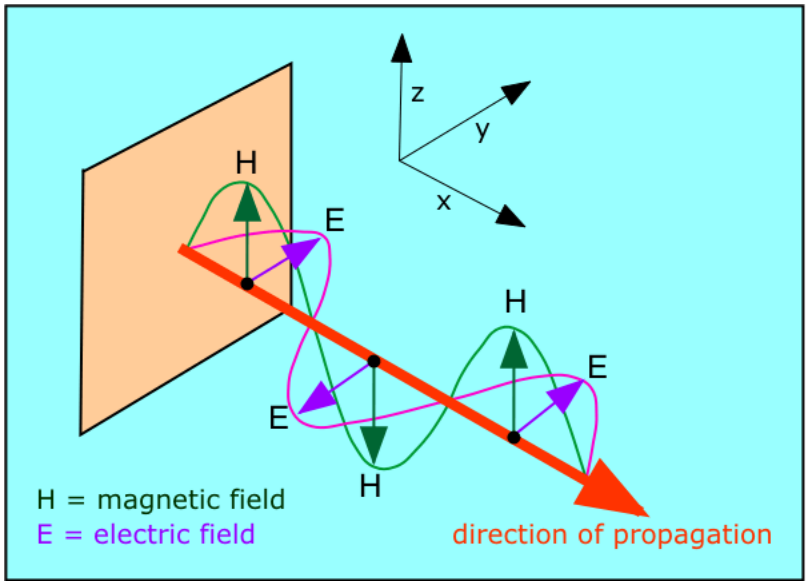
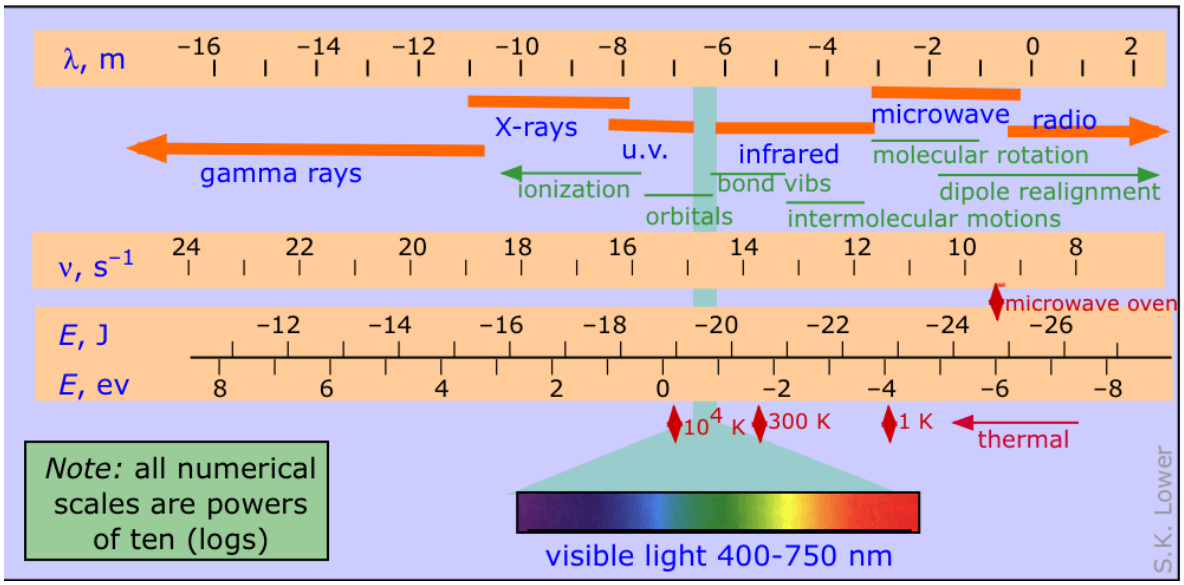


Figure 5.3.1: What is "waving" in electromagnetic radiation?: According to Maxwell, it is the strengths of the electric and magnetic fields as they travel through space. The two fields are oriented at right angles to each other and to the direction of travel. As the electric field changes, it induces a magnetic field, which then induces a new electric field, etc., allowing the wave to propagate itself through space.

Maxwell was able to calculate the speed at which electromagnetic disturbances are propagated, and found that this speed is the same as that of light. He therefore proposed that light is itself a form of electromagnetic radiation whose wavelength range forms only a very small part of the entire electromagnetic spectrum. Maxwell's work served to unify what were once thought to be entirely separate realms of wave motion.

The electromagnetic spectrum

The electromagnetic spectrum is conventionally divided into various parts as depicted in the diagram below, in which the four logarithmic scales correlate the wavelength of electromagnetic radiation with its frequency in herz (units of s^{-1}) and the energy per photon, expressed both in joules and electron-volts.



The other items shown on the diagram, from the top down, are:

- the names used to denote the various wavelength ranges of radiation (you should know their names and the order in which they appear)
- the principal effects of the radiation on atoms and molecules
- the peaks of thermal radiation emitted by black bodies at three different temperatures

Electromagnetic radiation and chemistry. It's worth noting that radiation in the ultraviolet range can have direct chemical effects by ionizing atoms and disrupting chemical bonds. Longer-wavelength radiation can interact with atoms and molecules in ways that provide a valuable means of identifying them and revealing particular structural features.

Energy units and magnitudes

It is useful to develop some feeling for the various magnitudes of energy that we must deal with. The basic SI unit of energy is the *Joule*; the appearance of this unit in *Planck*'s *constant h* allows us to express the energy equivalent of light in joules. For example, light of wavelength 500 nm, which appears blue-green to the human eye, would have a frequency of

$$\nu = \frac{c}{\lambda} = \frac{(3 \times 10^{8}\ \mathrm{m\,s^{-1}})}{(5 \times 10^{-7}\ \mathrm{m})} = (6 \times 10^{14}\ \mathrm{s^{-1}})$$

The quantum of energy carried by a single photon of this frequency is

$$\nu = \frac{c}{\lambda} = \frac{(3 \times 10^{8}\ \mathrm{m\,s^{-1}})}{(5 \times 10^{-7}\ \mathrm{m})} = (6 \times 10^{14}\ \mathrm{s^{-1}})$$

Another energy unit that is commonly employed in atomic physics is the *electron volt*; this is the kinetic energy that an electron acquires upon being accelerated across a 1-volt potential difference. The relationship 1 eV = 1.6022E–19 J gives an energy of 2.5 eV for the photons of blue-green light.

Two small flashlight batteries will produce about 2.5 volts, and thus could, in principle, give an electron about the same amount of kinetic energy that blue-green light can supply. Because the energy produced by a battery derives from a chemical reaction, this quantity of energy is representative of the magnitude of the energy changes that accompany chemical reactions.

In more familiar terms, one mole of 500-nm photons would have an energy equivalent of Avogadro's number times 4E–19 J, or

240 kJ per mole. This is comparable to the amount of energy required to break some chemical bonds. Many substances are able to undergo chemical reactions following light-induced disruption of their internal bonding; such molecules are said to be *photochemically active*.

Spectra: Interaction of light and matter

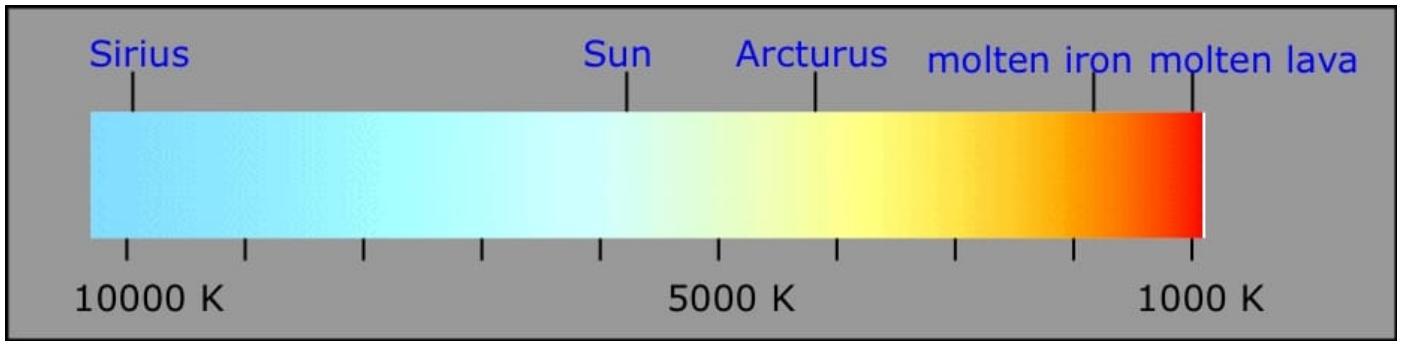
Continuous spectra

Any body whose temperature is above absolute zero emits radiation covering a broad range of wavelengths. At very low temperatures the predominant wavelengths are in the radio microwave region. As the temperature increases, the wavelengths decrease; at room temperature, most of the emission is in the infrared.

At still higher temperatures, objects begin to emit in the visible region, at first in the red, and then moving toward the blue as the temperature is raised. These *thermal emission spectra* are described as *continuous spectra*, since all wavelengths within the broad emission range are present.



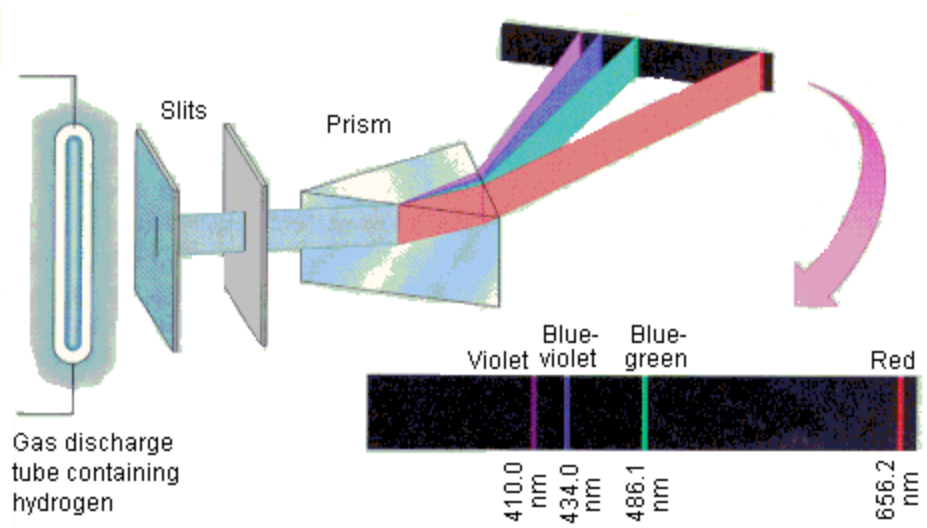
The source of thermal emission most familiar to us is the Sun. When sunlight is refracted by rain droplets into a rainbow or by a prism onto a viewing screen, we see the visible part of the spectrum.



Red hot, white hot, blue hot... your rough guide to temperatures of hot objects.

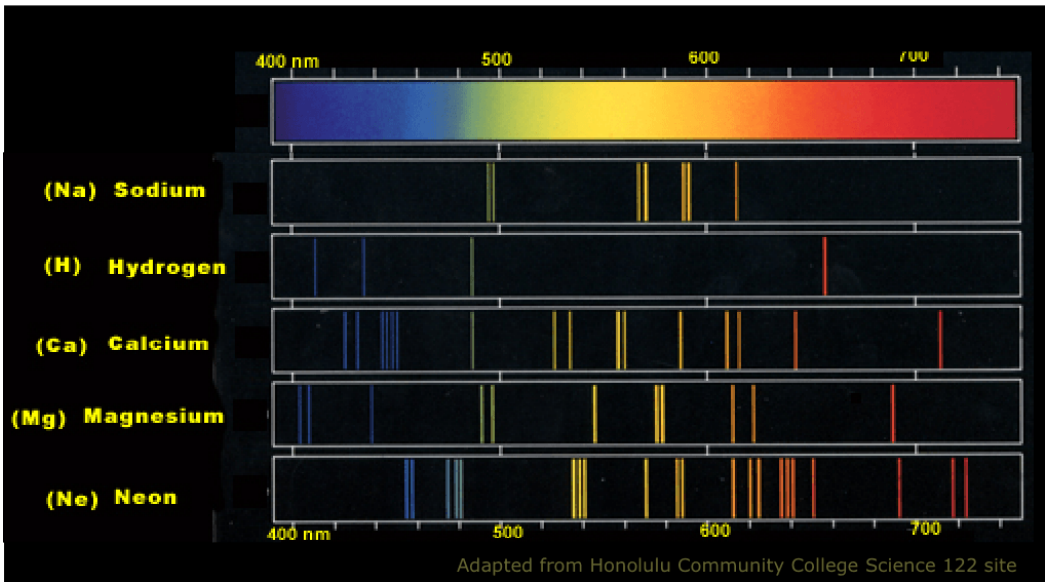
Line spectra

Heat a piece of iron up to near its melting point and it will emit a broad continuous spectrum that the eye perceives as orange-yellow. But if you zap the iron with an electric spark, some of the iron atoms will vaporize and have one or more of their electrons temporarily knocked out of them. As they cool down the electrons will re-combine with the iron ions, losing energy as the move in toward the nucleus and giving up this excess energy as light. The spectrum of this light is anything but continuous; it consists of a series of discrete wavelengths which we call *lines*.



A spectrum is most accurately expressed as a plot of intensity as a function of wavelength. Historically, the first spectra were obtained by passing the radiation from a source through a series or narrow slits to obtain a thin beam which was dispersed by a prism so that the different wavelengths are spread out onto a photographic film or viewing screen. The "lines" in *line spectra* are really images of the slit nearest the prism. Modern spectrophotometers more often employ diffraction gratings rather then prisms, and use electronic sensors connected to a computer.

Each chemical element has its own characteristic line spectrum which serves very much like a "fingerprint" capable of identifying a particular element in a complex mixture. Shown below is what you would see if you could look at several different atomic line spectra directly.



Atomic line spectra are extremely useful for identifying small quantities of different elements in a mixture.

- Companies that own large fleets of trucks and buses regularly submit their crankcase engine oil samples to spectrographic analysis. If they find high levels of certain elements (such as vanadium) that occur only in certain alloys, this can signal that certain parts of the engine are undergoing severe wear. This allows the mechanical staff to take corrective action before engine failure occurs.
- Several elements (Rb, Cs, Tl) were discovered by observing spectral lines that did not correspond to any of the then-known elements. Helium, which is present only in traces on Earth, was first discovered by observing the spectrum of the Sun.
- A more prosaic application of atomic spectra is determination of the elements present in stars.

If you live in a city, you probably see atomic line light sources every night! "Neon" signs are the most colorful and spectacular, but high-intensity street lighting is the most widespread source. A look at the emission spectrum (above) of sodium explains

the intense yellow color of these lamps. The spectrum of mercury (not shown) similarly has its strongest lines in the blue-green region.

Particles and waves

There is one more fundamental concept you need to know before we can get into the details of atoms and their spectra. If light has a particle nature, why should particles not possess wavelike characteristics? In 1923 a young French physicist, Louis de Broglie, published an argument showing that matter should indeed have a wavelike nature. The *de Broglie wavelength* of a body is inversely proportional to its momentum *mv*:

$$\lambda = \frac{h}{mv} \tag{5.3.3}$$

If you explore the magnitude of the quantities in this equation (recall that *h* is around 10^{-33} J s), it will be apparent that the wavelengths of all but the lightest bodies are insignificantly small fractions of their dimensions, so that the objects of our everyday world all have definite boundaries. Even individual atoms are sufficiently massive that their wave character is not observable in most kinds of experiments. Electrons, however, are another matter; the electron was in fact the first particle whose wavelike character was seen experimentally, following de Broglie's prediction. Its small mass (9.1E–31 kg) made it an obvious candidate, and velocities of around 100 km/s are easily obtained, yielding a value of λ in the above equation that well exceeds what we think of as the "radius" of the electron. At such velocities the electron behaves as if it is "spread out" to atomic dimensions; a beam of these electrons can be diffracted by the ordered rows of atoms in a crystal in much the same way as visible light is diffracted by the closely-spaced groves of a CD recording.

Electron diffraction has become an important tool for investigating the structures of molecules and of solid surfaces.

A more familiar exploitation of the wavelike properties of electrons is seen in the *electron microscope*, whose utility depends on the fact that the wavelength of the electrons is much less than that of visible light, thus allowing the electron beam to reveal detail on a correspondingly smaller scale.

The uncertainty principle

In 1927, the German physicist Werner Heisenberg pointed out that the wave nature of matter leads to a profound and far-reaching conclusion: no method of observation, however perfectly it is carried out, can reveal both the exact *location* and *momentum* (and thus the *velocity*) of a particle. This is the origin of the widely known concept that the very process of observation will change the value of the quantity being observed. The Heisenberg principle can be expressed mathematically by the inequality

$$\Delta x \Delta p \leq \frac{h}{2\pi} \tag{5.3.4}$$

in which the Δ (deltas) represent the uncertainties with which the location and momentum are known.

Suppose that you wish to measure the exact location of a particle that is at rest (zero momentum). To accomplish this, you must "see" the molecule by illuminating it with light or other radiation. But the light acts like a beam of photons, each of which possesses the momentum h/λ in which λ is the wavelength of the light. When a photon collides with the particle, it transfers some of its momentum to the particle, thus altering both its position and momentum.

Notice how the form of this expression predicts that if the location of an object is known exactly ($\Delta x = 0$), then the uncertainty in the momentum must be infinite, meaning that nothing at all about the velocity can be known. Similarly, if the velocity were specified exactly, then the location would be entirely uncertain and the particle could be anywhere. One interesting consequence of this principle is that even at a temperature of absolute zero, the molecules in a crystal must still possess a small amount of *zero point vibrational motion*, sufficient to limit the precision to which we can measure their locations in the crystal lattice.

An equivalent formulation of the uncertainty principle relates the uncertainties associated with a measurement of the energy of a system to the time Δt taken to make the measurement:

$$\Delta E \Delta t \leq \frac{h}{2\pi} \tag{5.3.5}$$

The "uncertainty" referred to here goes much deeper than merely limiting our ability to *observe* the quantity $\Delta x \Delta p$ to a greater precision than \(*h*/2\pi\). It means, rather, that this product has no exact value, nor, by extension, do *position* and *momentum* on a microscopic scale. A more appropriate term would be *indeterminacy*, which is closer to Heisenberg's original word *Ungenauigkeit*.

The revolutionary nature Heisenberg's uncertainty principle soon extended far beyond the arcane world of physics; its consequences quickly entered the realm of ideas and has inspired numerous creative works in the arts— few of which really have much to do with the Principle! A possible exception is Michael Frayn's widely acclaimed play (see below) that has brought a sense of Heisenberg's thinking to a wide segment of the public.

This page titled 5.3: Light, Particles, and Waves is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.4: The Bohr Atom

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Describe the Thompson, Rutherford, and early planetary models of the atom, and explain why the latter is not consistent with classical physics.
- State the major concepts that distinguished Bohr's model of the atom from the earlier planetary model.
- Give an example of a mechanical standing wave; state the meaning and importance of its boundary conditions.
- Sketch out a diagram showing how the concept of a standing wave applies to the description of the electron in a hydrogen atom.
- What is an atomic line emission spectrum? What is the significance of the continuum region of an emission spectrum? Sketch out a drawing showing the essentials of such a spectrum, including the ionization limit and the continuum.
- Describe the way in which Bohr's quantum numbers explain the observed spectrum of a typical atom.
- Explain the relation between the absorption and emission spectrum of an atom.

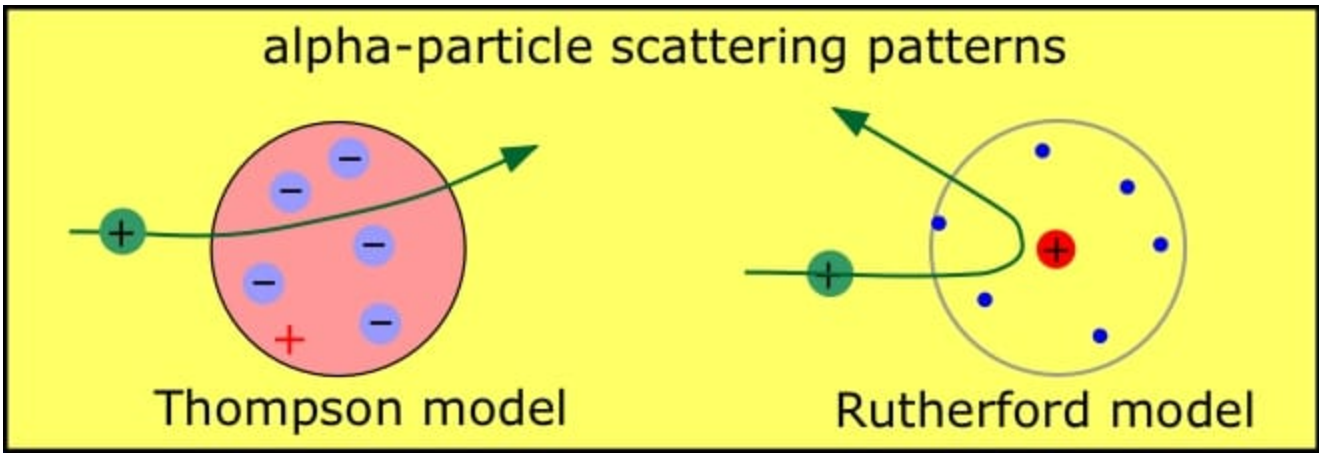
Our goal in this unit is to help you understand how the arrangement of the periodic table of the elements must follow as a necessary consequence of the fundamental laws of the quantum behavior of matter. The modern theory of the atom makes full use of the wave-particle duality of matter. In order to develop and present this theory in a comprehensive way, we would require a number of mathematical tools that lie beyond the scope of this course. We will therefore present the theory in a semi-qualitative manner, emphasizing its results and their applications, rather than its derivation.

Models of the atom

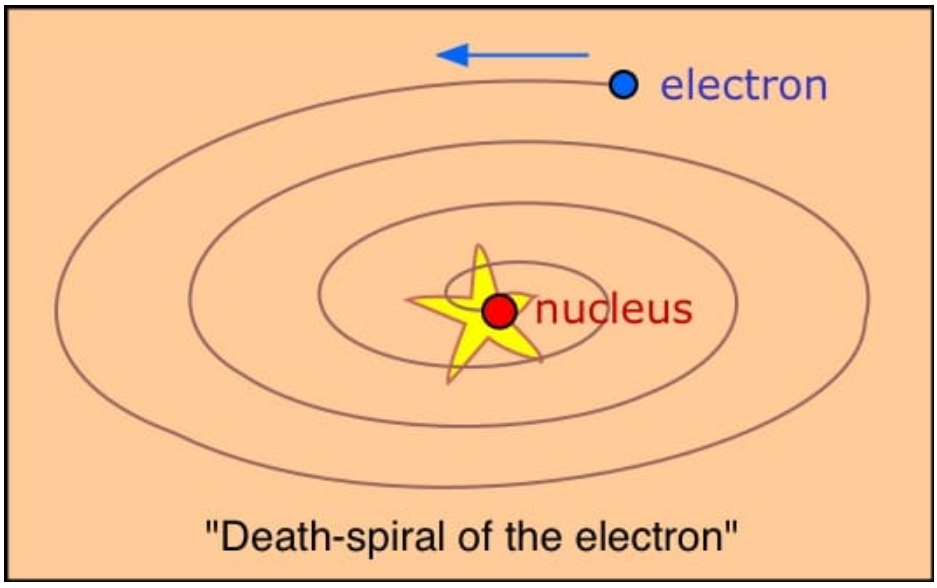
Models are widely employed in science to help understand things that cannot be viewed directly. The idea is to imagine a simplified system or process that might be expected to exhibit the basic properties or behavior of the real thing, and then to test this model against more complicated examples and modify it as necessary. Although one is always on shaky philosophical ground in trying to *equate* a model with reality, there comes a point when the difference between them becomes insignificant for most practical purposes.

The planetary model

The demonstration by Thompson in 1867 that all atoms contain units of negative electric charge led to the first science-based model of the atom which envisaged the electrons being spread out uniformly throughout the spherical volume of the atom. Ernest Rutherford, a New Zealander who started out as Thompson's student at Cambridge, distrusted this "plum pudding" model (as he called it) and soon put it to rest; Rutherford's famous alpha-ray bombardment experiment (carried out, in 1909, by his students Hans Geiger and Ernest Marsden) showed that nearly all the mass of the atom is concentrated in an extremely small (and thus extremely dense) body called the nucleus. This led him to suggest the *planetary model* of the atom, in which the electrons revolve in orbits around the nuclear "sun".



Even though the planetary model has long since been discredited, it seems to have found a permanent place in popular depictions of the atom, and certain aspects of it remain useful in describing and classifying atomic structure and behavior. The planetary model of the atom assumed that the electrostatic attraction between the central nucleus and the electron is exactly balanced by the centrifugal force created by the revolution of the electron in its orbit. If this balance were not present, the electron would either fall into the nucleus, or it would be flung out of the atom.

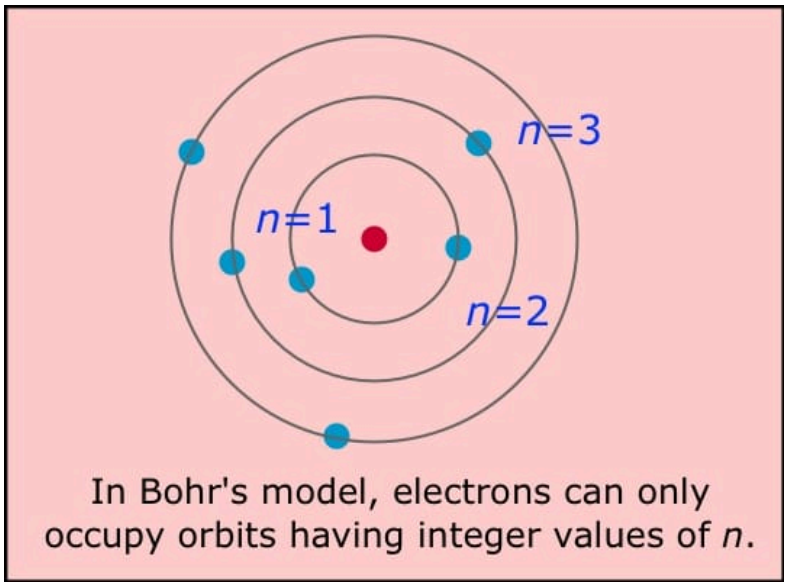


The difficulty with this picture is that it is inconsistent with a well established fact of classical electrodynamics which says that whenever an electric charge undergoes a change in velocity or direction (that is, *acceleration*, which must happen if the electron circles around the nucleus), it must continually radiate energy. If electrons actually followed such a trajectory, all atoms would act is miniature broadcasting stations. Moreover, the radiated energy would come from the kinetic energy of the orbiting electron; as this energy gets radiated away, there is less centrifugal force to oppose the attractive force due to the nucleus. The electron would quickly fall into the nucleus, following a trajectory that became known as the "death spiral of the electron". According to classical physics, no atom based on this model could exist for more than a brief fraction of a second.

Bohr's Model

Niels Bohr was a brilliant Danish physicist who came to dominate the world of atomic and nuclear physics during the first half of the twentieth century. Bohr suggested that the planetary model could be saved if one new assumption were made: certain "special states of motion" of the electron, corresponding to different orbital radii, would not result in radiation, and could therefore persist indefinitely without the electron falling into the nucleus. Specifically, Bohr postulated that the angular momentum of the electron, *mvr* (the mass and angular velocity of the electron and in an orbit of radius r) is restricted to values that are integral multiples of $h/2\pi$. The radius of one of these allowed Bohr orbits is given by

$$r = \frac{nh}{2\pi mu} \tag{5.4.1}$$



in which *h* is Planck's constant, *m* is the mass of the electron, *v* is the orbital velocity, and *n* can have only the integer values 1, 2, 3, etc. The most revolutionary aspect of this assumption was its use of the variable integer *n*; this was the first application of the concept of the *quantum number* to matter. The larger the value of *n*, the larger the radius of the electron orbit, and the greater the potential energy of the electron.

As the electron moves to orbits of increasing radius, it does so in opposition to the restoring force due to the positive nucleus, and its potential energy is thereby raised. This is entirely analogous to the increase in potential energy that occurs when any mechanical system moves against a restoring force— as, for example, when a rubber band is stretched or a weight is lifted.

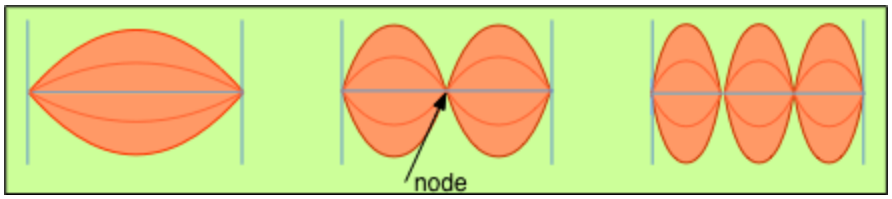
Thus what Bohr was saying, in effect, is that the atom can exist only in certain discrete energy states: the energy of the atom is *quantized*. Bohr noted that this quantization nicely explained the observed emission spectrum of the hydrogen atom. The electron is normally in its smallest allowed orbit, corresponding to $n = 1$; upon excitation in an electrical discharge or by ultraviolet light, the atom absorbs energy and the electron gets promoted to higher quantum levels. These higher excited states of the atom are unstable, so after a very short time (around 10^{-9} sec) the electron falls into lower orbits and finally into the innermost one, which corresponds to the atom's *ground state*. The energy lost on each jump is given off as a photon, and the frequency of this light provides a direct experimental measurement of the difference in the energies of the two states, according to the Planck-Einstein relationship $e = h\nu$.

Vibrations, standing waves and bound states

Bohr's theory worked; it completely explained the observed spectrum of the hydrogen atom, and this triumph would later win him a Nobel prize. The main weakness of the theory, as Bohr himself was the first to admit, is that it could offer no good explanation of *why* these special orbits immunized the electron from radiating its energy away. The only justification for the proposal, other than that it seems to work, comes from its analogy to certain aspects of the behavior of vibrating mechanical systems.

Spectrum of a guitar string

In order to produce a tone when plucked, a guitar string must be fixed at each end (that is, it must be a *bound system*) and must be under some tension. Only under these conditions will a transverse disturbance be countered by a restoring force (the string's tension) so as to set up a sustained vibration. Having the string tied down at both ends places a very important *boundary condition* on the motion: the only allowed modes of vibration are those whose wavelengths produce zero displacements at the bound ends of the string; if the string breaks or becomes unattached at one end, it becomes silent.



In its lowest-energy mode of vibration there is a single wave whose point of maximum displacement is placed at the center of the string. In musical terms, this corresponds to the fundamental note to which the string is tuned; in terms of the theory of vibrations, it corresponds to a "quantum number" of 1. Higher modes, known as *overtones* (and in music, as *octaves*), contain 2, 3, 4 and more points of maximum displacement (*antinodes*) spaced evenly along the string, separated by points of zero displacement (*nodes*). These correspond to successively higher quantum numbers and higher energies.

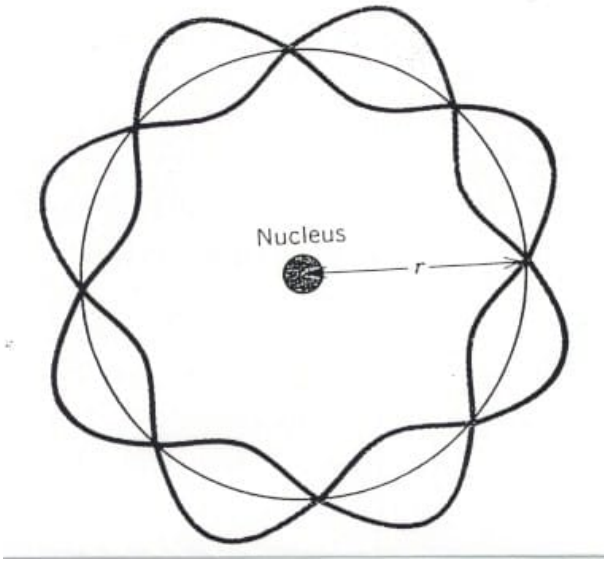
The vibrational states of the string are quantized in the sense that an integral number of antinodes must be present. Note again that this condition is imposed by the boundary condition that the ends of the string, being fixed in place, must be nodes. Because the locations of the nodes and antinodes do not change as the string vibrates, the vibrational patterns are known as *standing waves*.

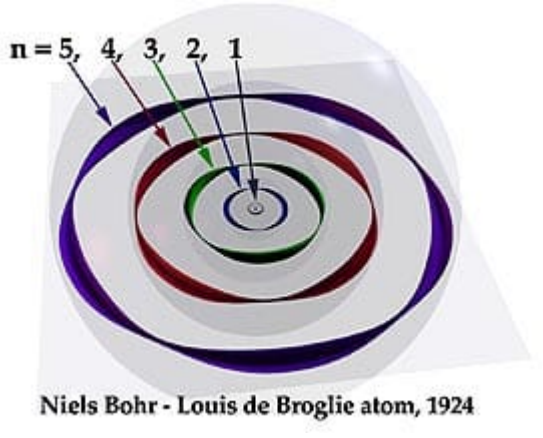
A similar kind of quantization occurs in other musical instruments; in each case the vibrations, whether of a stretched string, a column of air, or of a stretched membrane.

Standing waves live in places other than atoms and musical instruments: every time you turn on your microwave oven, a complex set of standing waves fills the interior. What is "waving" here is the alternating electrostatic field as a function of location; the wave patterns are determined by the dimensions of the oven interior and by the objects placed within it. But the part of a pizza that happens to be located at a node would not get very hot, so all microwave ovens provide a mechanical means of rotating either the food (on a circular platform) or the microwave beam (by means of a rotating deflector) so that all parts will pass through high-amplitude parts of the waves.

Standing waves in the hydrogen atom

The analogy with the atom can be seen by imagining a guitar string that has been closed into a circle. The circle is the electron orbit, and the boundary condition is that the waves must not interfere with themselves along the circle. This condition can only be met if the circumference of an orbit can exactly accommodate an integral number of wavelengths. Thus only certain discrete orbital radii and energies are allowed, as depicted in the two diagrams below.



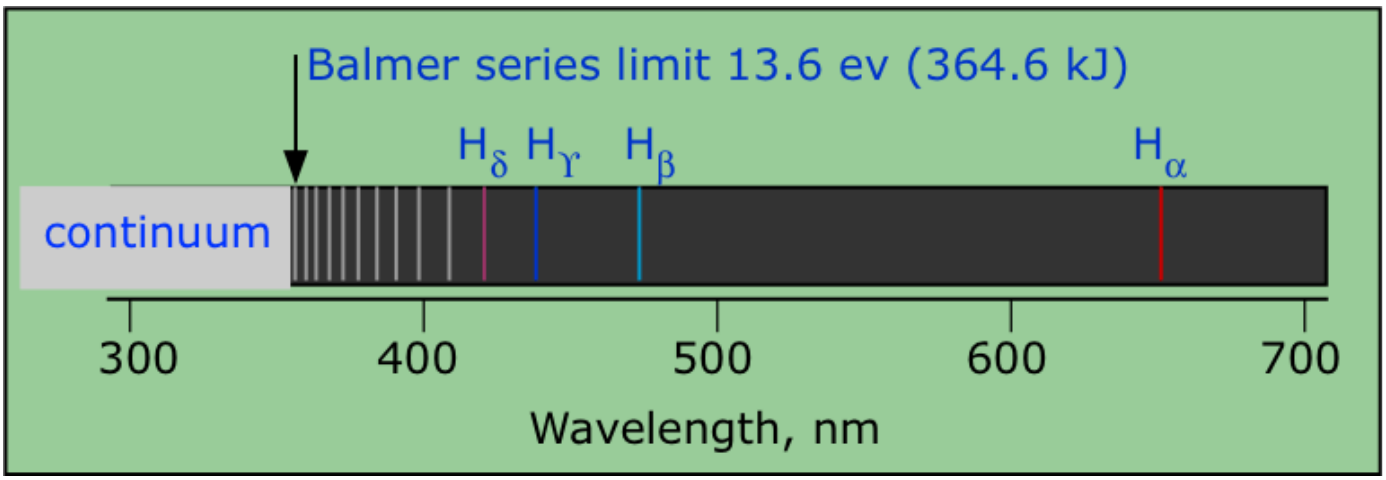


Unbound states

If a guitar string is plucked so harshly that it breaks, the restoring force and boundary conditions that restricted its motions to a few discrete harmonically related frequencies are suddenly absent; with no constraint on its movement, the string's mechanical energy is dissipated in a random way without musical effect. In the same way, if an atom absorbs so much energy that the electron is no longer bound to the nucleus, then the energy states of the atom are no longer quantized; instead of the line spectrum associated with discrete energy jumps, the spectrum degenerates into a continuum in which all possible electron energies are allowed. The energy at which the *ionization continuum* of an atom begins is easily observed spectroscopically, and serves as a simple method of experimentally measuring the energy with which the electron is bound to the atom.

Spectrum of the hydrogen atom

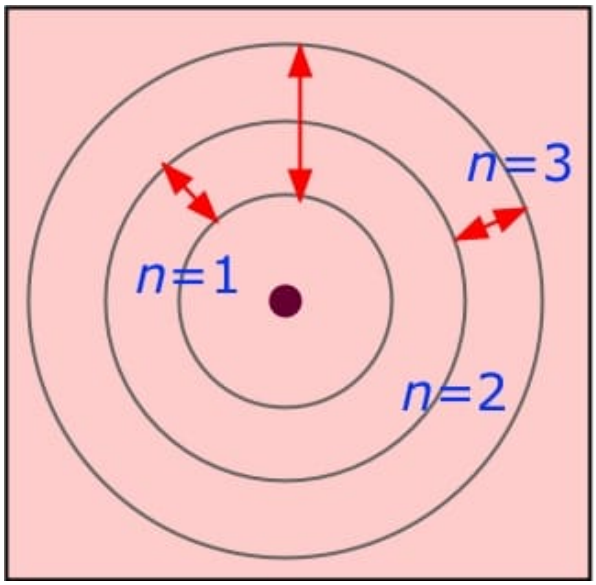
Hydrogen, the simplest atom, also has the simplest line spectrum (line spectra were briefly introduced in the previous chapter.) The hydrogen spectrum was the first to be observed (by Ånders Ångström in the 1860's). Johannn Balmer, a German high school teacher, discovered a simple mathematical formula that related the wavelengths of the various lines that are observable in the visible and near-uv parts of the spectrum. This set of lines is now known as the *Balmer Series*.



The four lines in the visible spectrum (designated by α through δ) were the first observed by Balmer. Notice how the lines crowd together as they approach the *ionization limit* in the near-ultraviolet part of the spectrum. Once the electron has left the atom, it is in an unbound state and its energy is no longer quantized. When such electrons return to the atom, they possess random amounts of kinetic energies over and above the binding energy. This reveals itself as the radiation at the short-wavelength end of the spectrum known as the *continuum* radiation. Other named sets of lines in the hydrogen spectrum are the *Lyman series* (in the ultraviolet) and the *Paschen, Brackett, Pfund* and *Humphrey* series in the infrared.

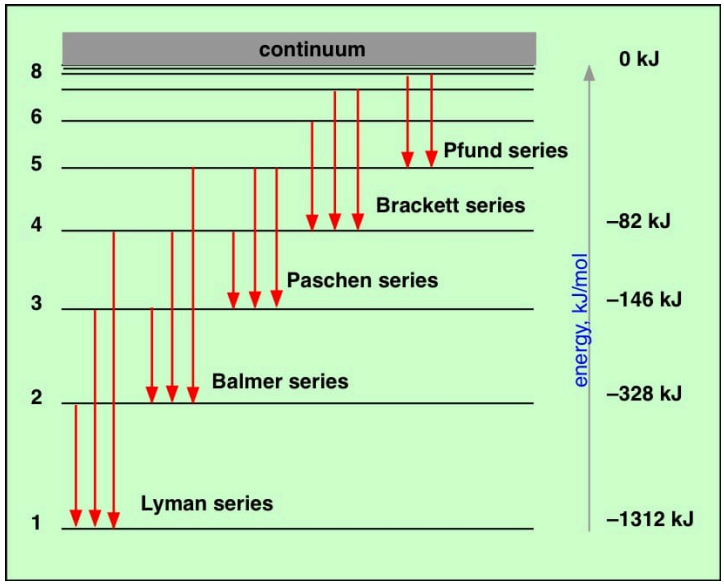
How the Bohr model explains the hydrogen line spectrum

Each spectral line represents an energy difference between two possible states of the atom. Each of these states corresponds to the electron in the hydrogen atom being in an "orbit" whose radius increases with the quantum number *n*. The lowest allowed value of *n* is 1; because the electron is as close to the nucleus as it can get, the energy of the system has its minimum (most negative) value. This is the "normal" (most stable) state of the hydrogen atom, and is called the *ground state*.



If a hydrogen atom absorbs radiation whose energy corresponds to the difference between that of n=1 and some higher value of n, the atom is said to be in an excited state. Excited states are unstable and quickly decay to the ground state, but not always in a single step. For example, if the electron is initially promoted to the n=3 state, it can decay either to the ground state or to the n=2 state, which then decays to n=1. Thus this single n=1→3 excitation can result in the three emission lines depicted in the diagram above, corresponding to n=3→1, n=3→2, and n=2→1.

If, instead, enough energy is supplied to the atom to completely remove the electron, we end up with a hydrogen ion and an electron. When these two particles recombine ($H^+ + e^- \rightarrow H$), the electron can initially find itself in a state corresponding to any value of n, leading to the emission of many lines.



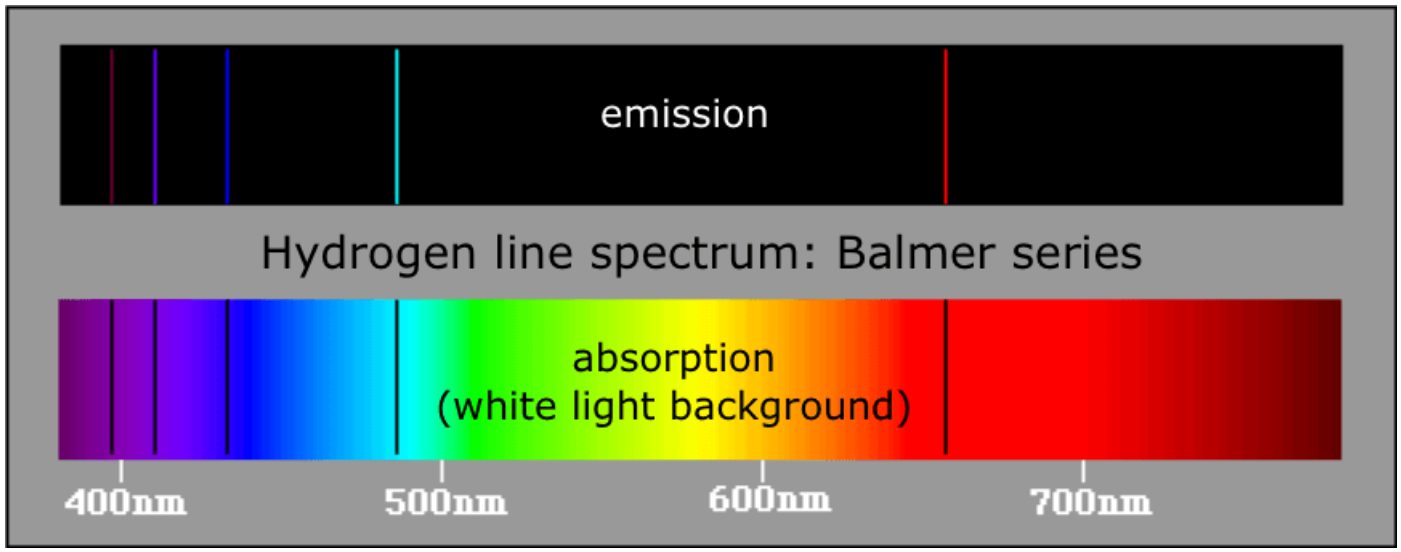
To provide a compact display, the vertical energy scale has been distorted and only the longest-wavelength transitions for each series are depicted (red arrows).

The lines of the hydrogen spectrum can be organized into different *series* according to the value of n at which the emission terminates (or at which absorption originates.) The first few series are named after their discoverers. The most well-known (and first-observed) of these is the Balmer series, which lies mostly in the visible region of the spectrum. The Lyman lines are in the ultraviolet, while the other series lie in the infrared. The lines in each series crowd together as they converge toward the *series limit* which corresponds to ionization of the atom and is observed as the beginning of the *continuum* emission. Note that the *ionization energy* of hydrogen (from its ground state) is 1312 kJ mol^{-1}. Although an infinite number of n-values are possible, the

number of observable lines is limited by our ability to resolve them as they converge into the continuum; this number is around a thousand.

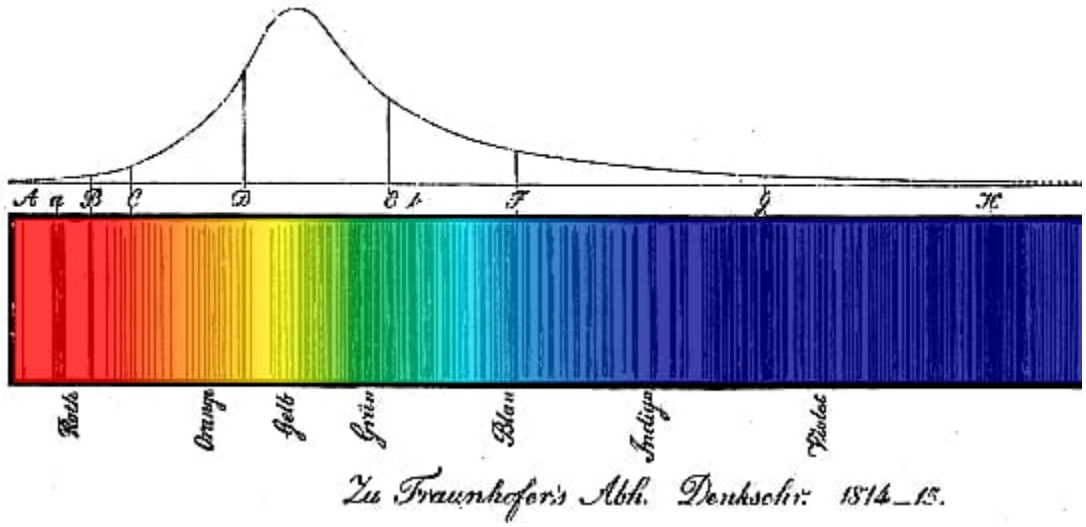
Emission and absorption spectra

The line *emission* spectra we have been discussing are produced when electrons which had previously been excited to values of n greater than 1 fall back to the n=1 ground state, either directly, or by way of intermediate-n states. But if light from a continuous source (a hot body such as a star) passes through an atmosphere of hydrogen (such as the star's outer atmosphere), those wavelengths that correspond to the allowed transitions are absorbed, and appear as dark lines superimposed on the continuous spectrum.



Emission and absorption spectra of the Balmer series of hydrogen

These dark absorption lines were first observed by William Wollaston in his study of the solar spectrum. In 1814, Joseph von Fraunhofner (1787-1826) re-discovered them and made accurate measurements of 814 lines, including the four most prominent of the Balmer lines.



This page titled 5.4: The Bohr Atom is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.5: The Quantum Atom

Learning Objectives

Make sure you thoroughly understand the following essential ideas

- State the fundamental distinction between Bohr's original model of the atom and the modern orbital model.
- Explain the role of the *uncertainty principle* in preventing the electron from falling into the nucleus.
- State the physical meaning of the *principal quantum number* of an electron orbital, and make a rough sketch of the shape of the probability-*vs*-distance curve for any value of *n*.
- Sketch out the shapes of an s, *p*, or a typical *d* orbital.
- Describe the significance of the *magnetic quantum number* as it applies to a *p* orbital.
- State the *Pauli exclusion principle*.

The picture of the atom that Niels Bohr developed in 1913 served as the starting point for modern atomic theory, but it was not long before Bohr himself recognized that the advances in quantum theory that occurred through the 1920's required an even more revolutionary change in the way we view the electron as it exists in the atom. This lesson will attempt to show you this view—or at least the portion of it that can be appreciated without the aid of more than a small amount of mathematics.

From Orbits to Orbitals

About ten years after Bohr had developed his theory, de Broglie showed that the electron should have wavelike properties of its own, thus making the analogy with the mechanical theory of standing waves somewhat less artificial. One serious difficulty with the Bohr model still remained, however: it was unable to explain the spectrum of any atom more complicated than hydrogen. A refinement suggested by Sommerfeld assumed that some of the orbits are elliptical instead of circular, and invoked a second quantum number, *l*, that indicated the degree of ellipticity. This concept proved useful, and it also began to offer some correlation with the placement of the elements in the periodic table.

By 1926, de Broglie's theory of the wave nature of the electron had been experimentally confirmed, and the stage was set for its extension to all matter. At about the same time, three apparently very different theories that attempted to treat matter in general terms were developed. These were Schrödinger's wave mechanics, Heisenberg's matrix mechanics, and a more abstract theory of P.A.M. Dirac. These eventually were seen to be mathematically equivalent, and all continue to be useful.

Of these alternative treatments, the one developed by Schrödinger is the most easily visualized. Schrödinger started with the simple requirement that the total energy of the electron is the sum of its kinetic and potential energies:

$$E = \underbrace{\frac{mv^2}{2}}_{\text{kinetic energy}} + \underbrace{\frac{-e^2}{r}}_{\text{potential energy}} \tag{5.5.1}$$

The second term represents the **potential energy** of an electron (whose charge is denoted by *e*) at a distance *r* from a proton (the nucleus of the hydrogen atom). In quantum mechanics it is generally easier to deal with equations that use momentum ($p = mv$) rather than velocity, so the next step is to make this substitution:

$$E = \frac{p^2}{2m} - \frac{e^2}{r} \tag{5.5.2}$$

This is still an entirely classical relation, as valid for the waves on a guitar string as for those of the electron in a hydrogen atom. The third step is the big one: in order to take into account the wavelike character of the hydrogen atom, a mathematical expression that describes the position and momentum of the electron at all points in space is applied to both sides of the equation. The function, denoted by Ψ, "modulates" the equation of motion of the electron so as to reflect the fact that the electron manifests itself with greater probability in some locations that at others. This yields the celebrated Schrödinger equation

$$\left(\frac{mv^2}{2} - \frac{e^2}{r}\right)\Psi = E\Psi \tag{5.5.3}$$

Physical significance of the wavefunction

How can such a simple-looking expression contain within it the quantum-mechanical description of an electron in an atom— and thus, by extension, of all matter? The catch, as you may well suspect, lies in discovering the correct form of Ψ , which is known as the *wave function*. As this names suggests, the value of Ψ is a function of location in space relative to that of the proton which is the source of the binding force acting on the electron. As in any system composed of standing waves, certain boundary conditions must be applied, and these are also contained in Ψ; the major ones are that the value of must approach zero as the distance from the nucleus approaches infinity, and that the function be continuous.

When the functional form of has been worked out, the Schrödinger equation is said to have been solved for a particular atomic system. The details of how this is done are beyond the scope of this course, but the consequences of doing so are extremely important to us. Once the form of is known, the allowed energies E of an atom can be predicted from the above equation. Soon after Schrödinger's proposal, his equation was solved for several atoms, and in each case the predicted energy levels agreed exactly with the observed spectra.

There is another very useful kind of information contained in Ψ. Recalling that its value depends on the location in space with respect to the nucleus of the atom, the square of this function Ψ^2, evaluated at any given point, represents the probability of finding the electron at that particular point. The significance of this cannot be overemphasized; although the electron remains a particle having a definite charge and mass, and the question of "where" it is located is no longer meaningful. Any single experimental observation will reveal a definite location for the electron, but this will in itself have little significance; only a large number of such observations (similar to a series of multiple exposures of a photographic film) will yield meaningful results which will show that the electron can "be" anywhere with at least some degree of probability. This does not mean that the electron is "moving around" to all of these places, but that (in accord with the uncertainty principle) the concept of location has limited meaning for a particle as small as the electron. If we count only those locations in space at which the probability of the electron manifesting itself exceeds some arbitrary value, we find that the Ψ function defines a definite three-dimensional region which we call an *orbital*.

Why doesn't the electron fall into the nucleus?

We can now return to the question which Bohr was unable to answer in 1912. Even the subsequent discovery of the wavelike nature of the electron and the analogy with standing waves in mechanical systems did not really answer the question; the electron is still a particle having a negative charge and is attracted to the nucleus.

The answer comes from the Heisenberg Uncertainty Principle, which says that a quantum particle such as the electron cannot simultaneously have sharply-defined values of location and of momentum (and thus kinetic energy). To understand the implications of this restriction, suppose that we place the electron in a small box. The walls of the box define the precision δx to which the location is known; the smaller the box, the more exactly will we know the location of the electron. But as the box gets smaller, the uncertainty in the electron's kinetic energy will increase. As a consequence of this uncertainty, the electron will at times possess so much kinetic energy (the "confinement energy") that it may be able to penetrate the wall and escape the confines of the box.

This process is known as *tunneling*; the *tunnel effect* is exploited in various kinds of semiconductor devices, and it is the mechanism whereby electrons jump between dissolved ions and the electrode in batteries and other electrochemical devices.

The region near the nucleus can be thought of as an extremely small funnel-shaped box, the walls of which correspond to the electrostatic attraction that must be overcome if an electron confined within this region is to escape. As an electron is drawn toward the nucleus by electrostatic attraction, the volume to which it is confined diminishes rapidly. Because its location is now more precisely known, its kinetic energy must become more uncertain; the electron's kinetic energy rises more rapidly than its potential energy falls, so that it gets ejected back into one of its allowed values of *n*.

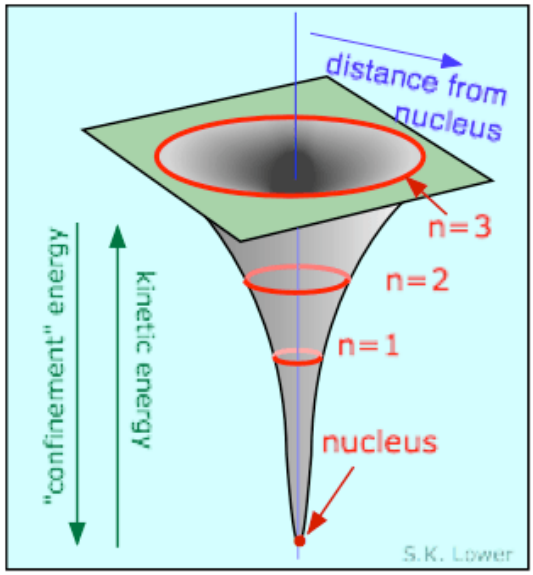


Figure 5.5.1: The electron well. The red circles show the average distance of the electron from the nucleus for the allowed quantum levels (standing wave patterns) of n=1 through n=3. As n decreases the potential energy of the system becomes more negative and the electron becomes more confined in space. According to the uncertainty principle, this increases the momentum of the electron, and hence its kinetic energy. The latter acts as a kind of "confinement energy" that restores the electron to one of the allowed levels.

We can also dispose of the question of why the orbiting electron does not radiate its kinetic energy away as it revolves around the nucleus. The Schrödinger equation completely discards any concept of a definite path or trajectory of a particle; what was formerly known as an "orbit" is now an "orbital", defined as the locations in space at which the probability of finding the electrons exceeds some arbitrary value. It should be noted that this wavelike character of the electron coexists with its possession of a momentum, and thus of an effective velocity, even though its motion does not imply the existence of a definite path or trajectory that we associate with a more massive particle.

Orbitals

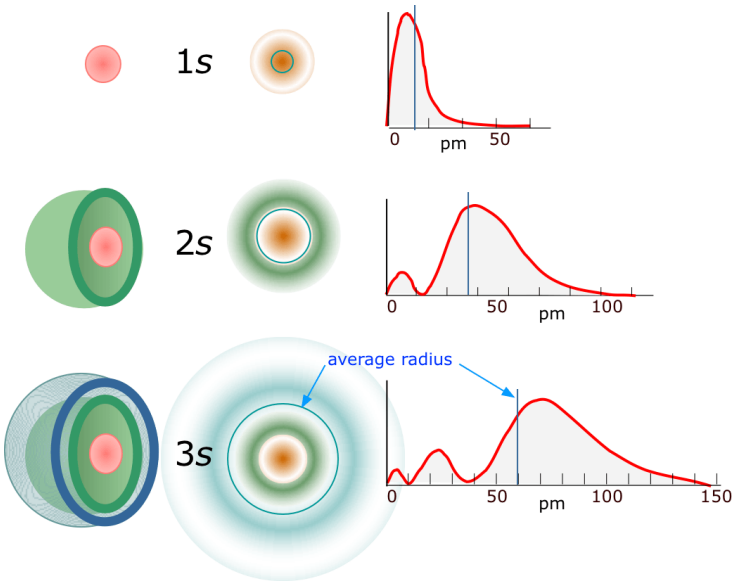
The modern view of atomic structure dismisses entirely the old but comfortable planetary view of electrons circling around the nucleus in fixed orbits. As so often happens in science, however, the old outmoded theory contains some elements of truth that are retained in the new theory. In particular, the old Bohr orbits still remain, albeit as spherical shells rather than as two-dimensional circles, but their physical significance is different: instead of defining the "paths" of the electrons, they merely indicate the locations in the space around the nucleus at which the probability of finding the electron has higher values. The electron retains its particle-like mass and momentum, but because the mass is so small, its wavelike properties dominate. The latter give rise to patterns of standing waves that define the possible states of the electron in the atom.

The quantum numbers

Modern quantum theory tells us that the various allowed states of existence of the electron in the hydrogen atom correspond to different standing wave patterns. In the preceding lesson we showed examples of standing waves that occur on a vibrating guitar string. The wave patterns of electrons in an atom are different in two important ways:

1. Instead of indicating displacement of a point on a vibrating string, the electron waves represent the *probability* that an electron will manifest itself (appear to be located) at any particular point in space. (Note carefully that this is *not* the same as saying that "the electron is smeared out in space"; at any given instant in time, it is either at a given point or it is not.)
2. The electron waves occupy all three dimensions of space, whereas guitar strings vibrate in only two dimensions.

Aside from this, the similarities are striking. Each wave pattern is identified by an integer number *n*, which in the case of the atom is known as the *principal quantum number*. The value of n tells how many peaks of amplitude (antinodes) exist in that particular standing wave pattern; the more peaks there are, the higher the energy of the state.



The three simplest orbitals of the hydrogen atom are depicted above in pseudo-3D, in cross-section, and as plots of probability (of finding the electron) as a function of distance from the nucleus. The average radius of the electron probability is shown by the blue circles or plots in the two columns on the right. These radii correspond exactly to those predicted by the Bohr model.

> The principal quantum number. The value of n tells how many peaks of amplitude (antinodes) exist in that particular standing wave pattern; the more peaks there are, the higher the energy of the state.

Physical Significance of *n*

The potential energy of the electron is given by the formula

$$E = \frac{-4\pi^2 e^4 m}{h^2 n^2} \tag{5.5.4}$$

with

- e is the charge of the electron, m is its mass,
- h is Planck's constant, and
- n is the principal quantum number.

The negative sign ensures that the potential energy is *always* negative. Notice that this energy in inversely proportional to the *square* of n, so that the energy rises toward zero as n becomes very large, but it can never exceed zero.

Equation 5.5.4 was actually part of Bohr's original theory and is still applicable to the hydrogen atom, although **not** to atoms with two or more electrons. In the Bohr model, each value of n corresponded to an orbit of a different radius. The larger the orbital radius, the higher the potential energy of the electron; the inverse square relationship between electrostatic potential energy and distance is reflected in the inverse square relation between the energy and n in the above formula. Although the concept of a definite trajectory or orbit of the electron is no longer tenable, the same orbital radii that relate to the different values of n in Bohr's theory now have a new significance: they give the average distance of the electron from the nucleus. The averaging process must encompass several probability peaks in the case of higher values of n. The spatial distribution of these probability maxima defines the particular orbital.

This physical interpretation of the principal quantum number as an index of the *average distance of the electron* from the nucleus turns out to be extremely useful from a chemical standpoint, because it relates directly to the tendency of an atom to lose or gain electrons in chemical reactions.

The Angular Momentum Quantum Number

The electron wave functions that are derived from Schrödinger's theory are characterized by several quantum numbers. The first one, *n*, describes the nodal behavior of the probability distribution of the electron, and correlates with its potential energy and average distance from the nucleus as we have just described.

The theory also predicts that orbitals having the same value of *n* can differ in shape and in their orientation in space. The quantum number *l*, known as the *angular momentum quantum number*, determines the shape of the orbital. (More precisely, *l* determines the number of angular nodes, that is, the number of regions of zero probability encountered in a 360° rotation around the center.)

When *l* = 0, the orbital is spherical in shape. If *l* = 1, the orbital is elongated into something resembling a figure-8 shape, and higher values of *l* correspond to still more complicated shapes— but note that the number of peaks in the radial probability distributions (below) decreases with increasing l. The possible values that *l* can take are strictly limited by the value of the principal quantum number; *l* can be no greater than *n* – 1. This means that for *n* = 1, *l* can only have the single value zero which corresponds to a spherical orbital. For historical reasons, the orbitals corresponding to different values of *l* are designated by letters, starting with *s* for *l* = 0, *p* for *l* = 1, *d* for *l* = 2, and *f* for *l* = 3.

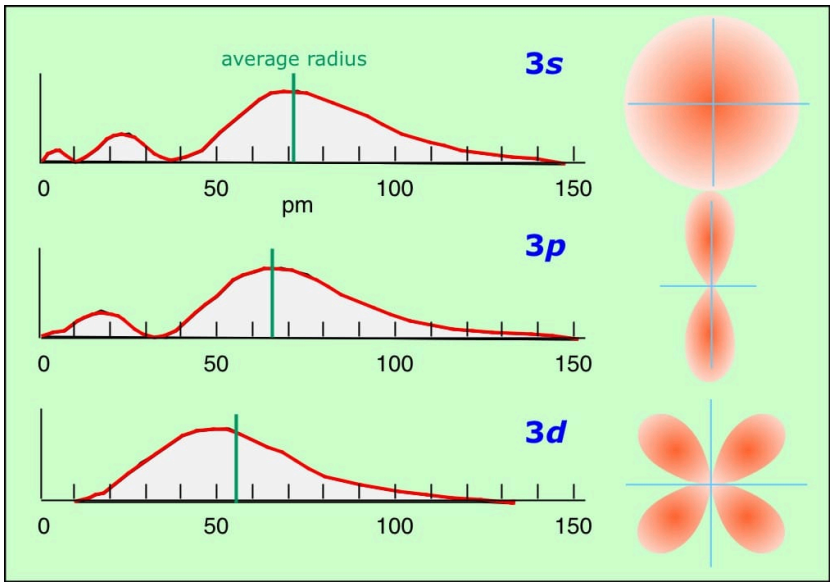


Figure 5.5.2: The shapes and radial distributions of the orbitals corresponding to the three allowed values of l for the n = 3 level of hydrogen are shown above. Notice that the average orbital radius r decreases somewhat at higher values of l.

The function relationship is given by

$$\bar{r} = (5.29\ pm)\frac{n^2}{Z}\left[\frac{3}{2} - \frac{l(l-1)}{2n^2}\right] \tag{5.5.5}$$

in which Z is the nuclear charge of the atom, which of course is unity for hydrogen.

The Magnetic Quantum Number

An s-orbital, corresponding to *l* = 0, is spherical in shape and therefore has no special directional properties. The probability cloud of a*p* orbital is aligned principally along an axis extending along any of the three directions of space. The additional quantum number *m* is required to specify the particular direction along which the orbital is aligned.

"Direction in space" has no meaning in the absence of a force field that serves to establish a reference direction. For an isolated atom there is no such external field, and for this reason there is no distinction between the orbitals having different values of m . If the atom is placed in an external magnetic or electrostatic field, a coordinate system is established, and the orbitals having different values of m will split into slightly different energy levels. This effect was first seen in the case of a magnetic field, and this is the origin of the term magnetic quantum number. In chemistry, however, electrostatic fields are much more important for defining directions at the atomic level because it is through such fields that nearby atoms in a molecule interact with each

other. The electrostatic field created when other atoms or ions come close to an atom can cause the energies of orbitals having different direction properties to split up into different energy levels; this is the origin of the colors seen in many inorganic salts of transition elements, such as the blue color of copper sulfate.

The quantum number m can assume $2l + 1$ values for each value of l, from $-l$ through 0 to $+l$. When $l = 0$ the only possible value of m will also be zero, and for the p orbital (l = 1), m can be –1, 0, and +1. Higher values of l introduce more complicated orbital shapes which give rise to more possible orientations in space, and thus to more values of m.

Electron Spin and the Exclusion Principle

Certain fundamental particles have associated with them a *magnetic moment* that can align itself in either of two directions with respect to an external magnetic field. The electron is one such particle, and the direction of its magnetic moment is called its *spin*.

A basic principle of modern physics states that for particles such as electrons that possess half-integral values of spin, no two of them can be in identical quantum states within the same system. The quantum state of a particle is defined by the values of its quantum numbers, so what this means is that **no two electrons in the same atom can have the same set of quantum numbers.** This is known as the Pauli exclusion principle, named after the German physicist Wolfgang Pauli (1900-1958, Nobel Prize 1945).

The Electrons is not Really Spinning

The mechanical analogy implied by the term spin is easy to visualize, but *should not be taken literally*. Physical rotation of an electron is meaningless. However, the coordinates of the electron's wave function can be rotated mathematically; when this is done, it is found that a rotation of 720° is required to restore the function to its initial value— rather weird, considering that a 360° rotation will leave any extended body unchanged! Electron spin is basically a relativistic effect in which the electron's momentum distorts local space and time. It has no classical counterpart and thus cannot be visualized other than through mathematics.

The exclusion principle was discovered empirically and was placed on a firm theoretical foundation by Pauli in 1925. A complete explanation requires some familiarity with quantum mechanics, so all we will say here is that if two electrons possess the same quantum numbers n, l, m and s (defined below), the wave function that describes the state of existence of the two electrons together becomes zero, which means that this is an "impossible" situation.

A given *orbital* is characterized by a fixed set of the quantum numbers n, l, and m. The electron spin itself constitutes a fourth quantum number s, which can take the two values +1 and –1. Thus **a given orbital can contain two electrons having opposite spins**, which "cancel out" to produce zero magnetic moment. Two such electrons in a single orbital are often referred to as an *electron pair*.

If it were not for the exclusion principle, the atoms of all elements would behave in the same way, and there would be no need for a science of Chemistry!

As we have seen, the lowest-energy standing wave pattern the electron can assume in an atom corresponds to n=1, which describes the state of the single electron in hydrogen, and of the two electrons in helium. Since the quantum numbers m and l are zero for n=1, the pair of electrons in the helium orbital have the values (n, l, m, s) = (1,0,0,+1) and (1,0,0,–1)— that is, they differ only in spin. These two sets of quantum numbers are the only ones that are possible for a n=1 orbital. The additional electrons in atoms beyond helium must go into higher-energy (n>1) orbitals. Electron wave patterns corresponding to these greater values of n are concentrated farther from the nucleus, with the result that these electrons are less tightly bound to the atom and are more accessible to interaction with the electrons of neighboring atoms, thus influencing their *chemical behavior*. If it were not for the Pauli principle, all the electrons of every element would be in the lowest-energy n=1 state, and the differences in the chemical behavior the different elements would be minimal. Chemistry would certainly be a simpler subject, but it would not be very interesting!

This page titled 5.5: The Quantum Atom is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.6: Atomic Electron Configurations

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

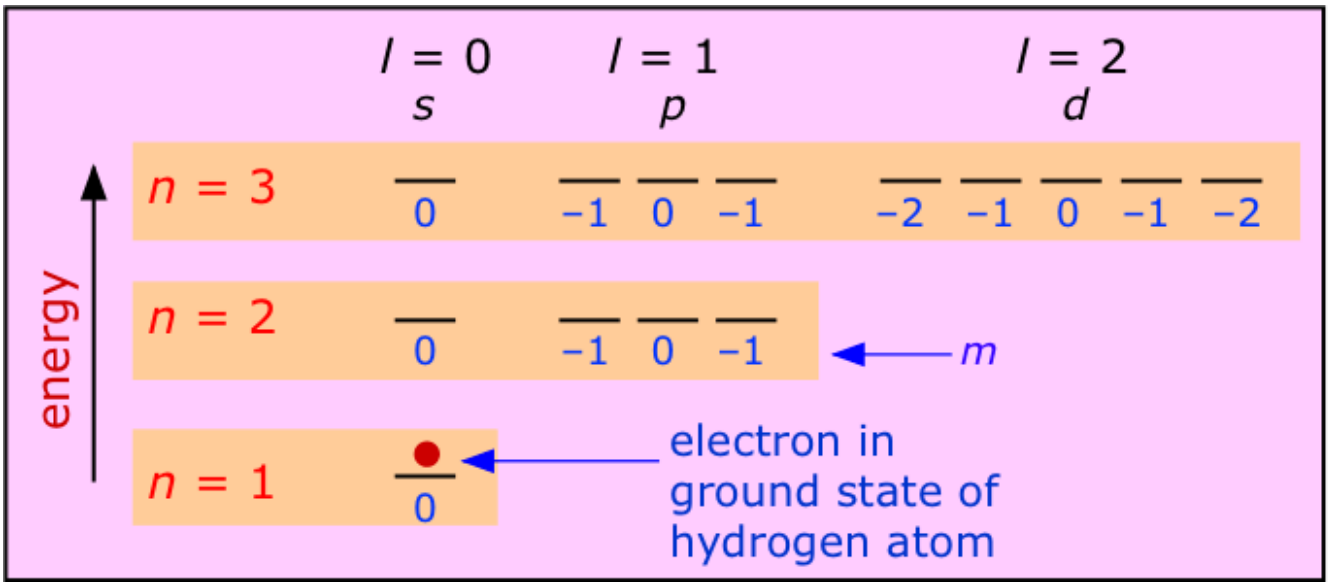
- State the principle feature of that distinguishes the energies of the excited states of a single-electron atom from atoms containing more than one electron.
- Explain why the first ionization energy of the helium atom is smaller than twice the first ionization of the hydrogen atom.
- Be able to write a plausible electron configuration for any atom having an atomic number less than 90.

In the previous section you learned that an electron standing-wave pattern characterized by the quantum numbers (*n*,*l*,*m*) is called an *orbital*. According to the *Pauli exclusion principle*, no two electrons in the same atom can have the same set of quantum numbers (*n*,*l*,*m*,*s*). This limits the number of electrons in a given orbital to two (s = ±1), and it requires that atom containing more then two electrons must place them in standing wave patterns corresponding to higher principal quantum numbers *n*, which means that these electrons will be farther from the nucleus and less tightly bound by it.

In this chapter, we will see how the Pauli restrictions on the allowable quantum numbers of electrons in an atom affect the electronic configuration of the different elements, and, by influencing their chemical behavior, governs the structure of the periodic table.

One-electron atoms

Let us begin with atoms that contain only a single electron. Hydrogen is of course the only *electrically neutral* species of this kind, but by removing electrons from heavier elements we can obtain one-electron ions such as He^{+} and Li^{2+}, etc. Each has a ground state configuration of $1s^1$, meaning that its single electron exhibits a standing wave pattern governed by the quantum numbers *n*=1, *m*=0 and *l*=0, with the spin quantum number s undefined because there is no other electron to compare it with. All have simple emission spectra whose major features were adequately explained by Bohr's model.

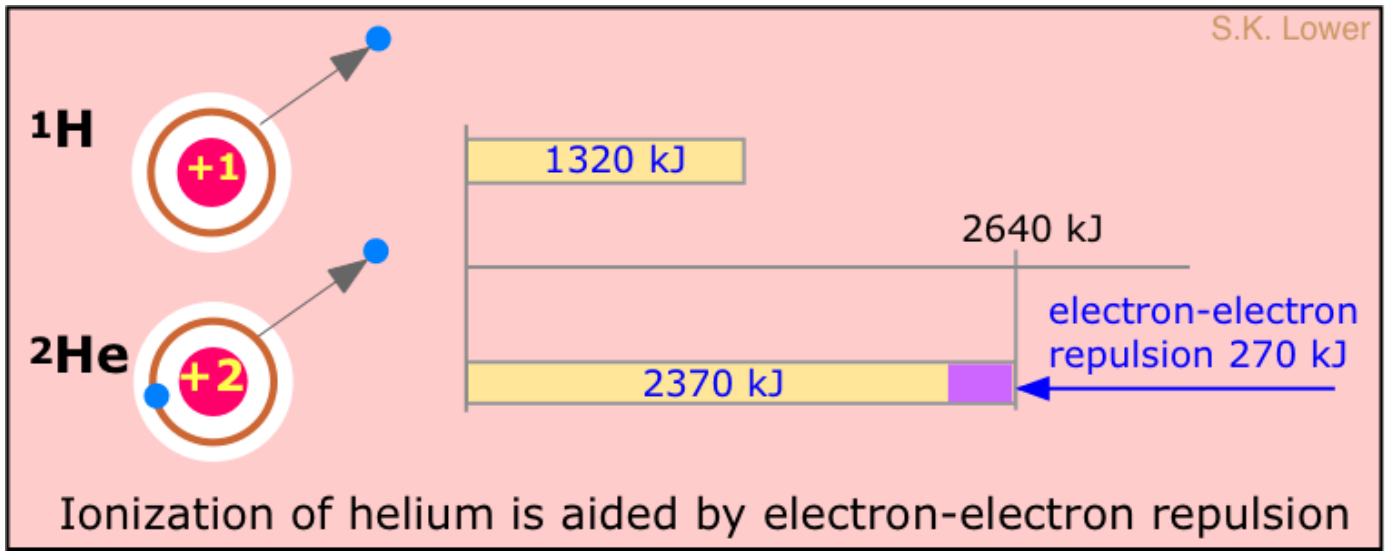


The most important feature of a single-electron atom is that the energy of the electron depends only on the principal quantum number *n*. As the above diagram shows, the quantum numbers *l* and *m* have no effect on the energy; we say that all orbitals having a given value of *n* are *degenerate*. Thus the emission spectrum produced by exciting the electron to the *n*=2 level consists of a single line, not four lines. The wavelength of this emission line for the atoms H, $He^{+}$and Li^{2+} will diminish with atomic number

because the greater nuclear charge will lower the energies of the various *n* levels. For the same reason, the energies required to remove an electron from these species increases rapidly as the nuclear charge increases, because the increasing attraction pulls the electron closer to the nucleus, thus producing an even greater attractive force.

Electron-Electron Repulsion

It takes 1312 kJ of energy to remove the electron from a mole of hydrogen atoms. What might we expect this value to be for helium? Helium contains two electrons, but its nucleus contains two protons; each electron "sees" both protons, so we might expect that the electrons of helium would be bound twice as strongly as the electron of hydrogen. The ionization energy of helium should therefore be twice 1312 kJ/mol, or 2612 kJ/mol. However, if one looks at the spectrum of helium, the continuum is seen to begin at a wavelength corresponding to an ionization energy of 2372 kJ/mol, or about 90% of the predicted value.



Why are the electrons in helium bound less tightly than the +2 nuclear charge would lead us to expect? The answer is that there is another effect to consider: the *repulsion* between the two electrons; the resulting electron-electron repulsion subtracts from the force holding the electron to the nucleus, reducing the local binding of each.

Electron-electron repulsion is a major factor in both the spectra and chemical behavior of the elements heavier than hydrogen. In particular, it acts to "break the degeneracy" (split the energies) of orbitals having the same value of *n* but different *l*.

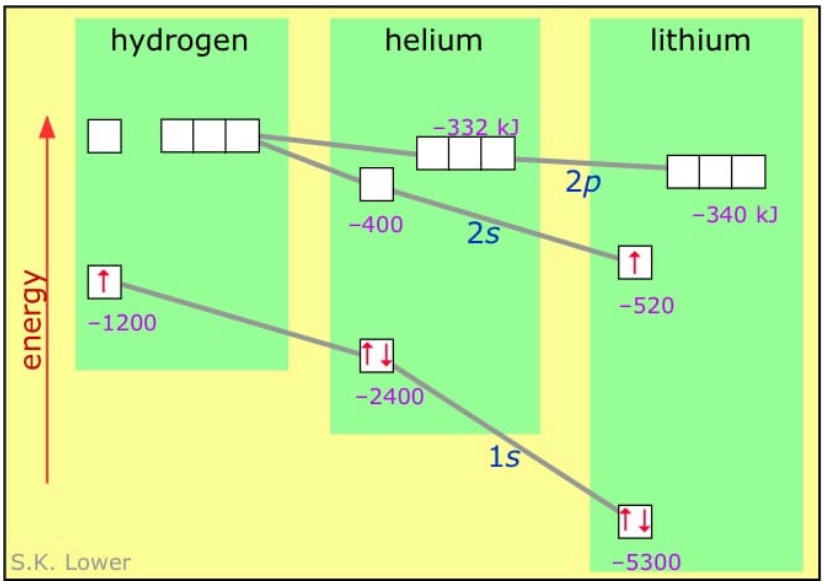
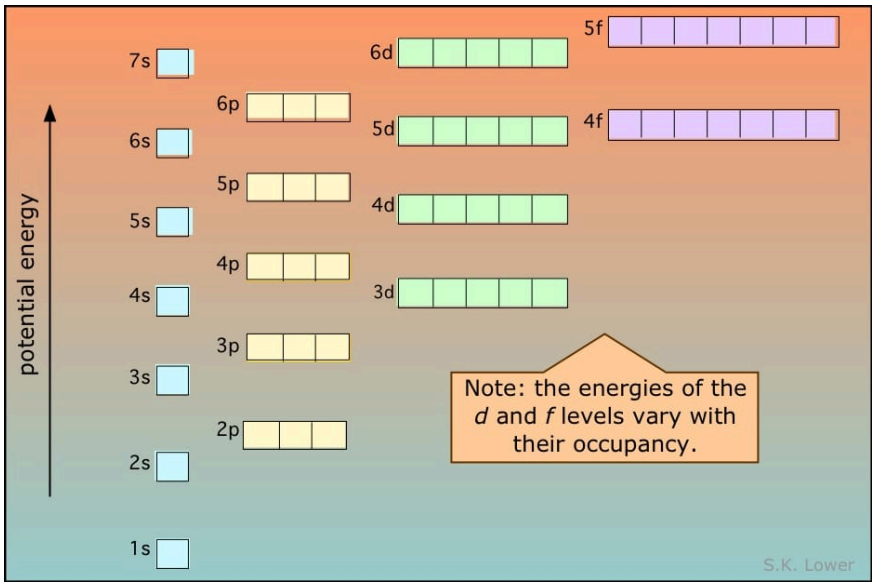


Figure 5.6.1: The numbers on this diagram show the energies of some orbitals in the atoms of H, He, and Li, relative to the 2s-2p orbitals of H. The energies of the 1s orbitals fall rapidly as Z increases. The fall in energies of the 2s orbitals is smaller owing to electron-electron repulsion. The 2p orbitals have a node at the nucleus and are less affected by Z but are more affected by electron repulsion, so they remain higher than the 2s orbitals.

The diagram below shows how the energies of the *s*- and *p*-orbitals of different principal quantum numbers get split as the result of electron-electron repulsion. Notice the contrast with the similar diagram for one-electron atoms near the top of this page. The fact that electrons preferentially fill the lowest-energy empty orbitals is the basis of the rules for determining the electron configuration of the elements and of the structure of the periodic table.



The *Aufbau* rules

The German word *Aufbau* means "building up", and this term has traditionally been used to describe the manner in which electrons are assigned to orbitals as we carry out the imaginary task of constructing the atoms of elements having successively larger atomic numbers. In doing so, we are effectively "building up" the periodic table of the elements, as we shall shortly see.

How to play the *Aufbau* game

- Electrons occupy the lowest-energy available orbitals; lower-energy orbitals are filled before the higher ones.

- No more than two electrons can occupy any orbital.
- For the lighter elements, electrons will fill orbitals of the same type only one electron at a time, so that their spins are all unpaired. They will begin to pair up only after all the orbitals are half-filled. This principle, which is a consequence of the electrostatic repulsion between electrons, is known as *Hund's rule*.
- For the first 18 elements, up to the point where the 3s and 3*p* levels are completely filled, this scheme is entirely straightforward and leads to electronic configurations which you are expected to be able to work out for each of these elements.

The preceding diagram illustrates the main idea here. Each orbital is represented as a little box that can hold up to two electrons having opposite spins, which we designated by upward- or downward-pointing arrows. Electrons fill the lowest-energy boxes first, so that additional electrons are forced into wave-patterns corresponding to higher (less negative) energies. Thus in the above diagram, the "third" electron of lithium goes into the higher-energy 2s orbital, giving this element an electron configuration which we write $1s^2\ 2s^1$.

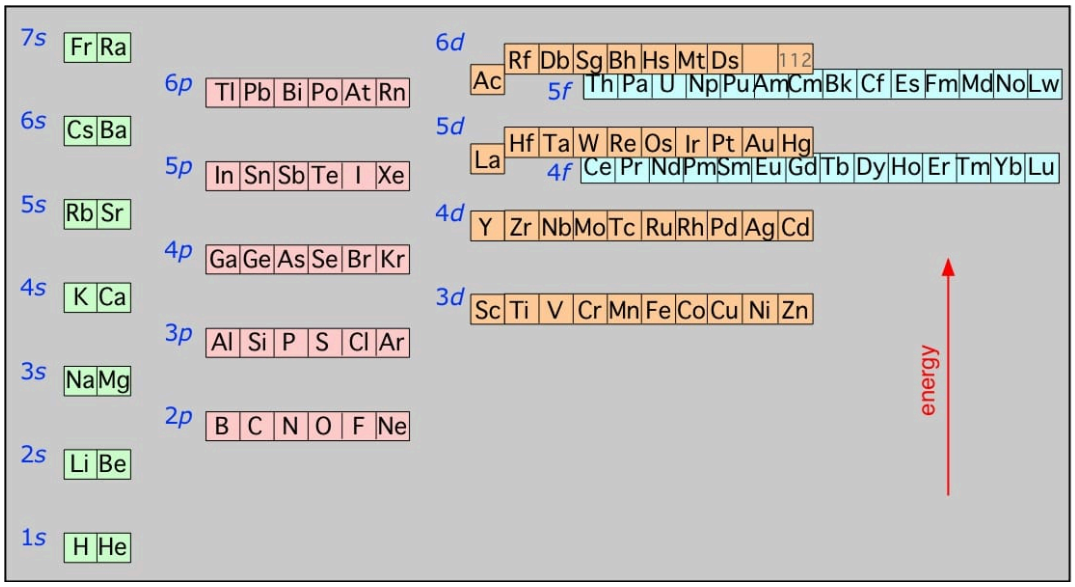
Example 5.6.1: Phosphorus

What is the electron configuration of the atom of phosphorus, atomic number 15?

The number of electrons filling the lowest-energy orbitals are:

1s: 2 electrons, 2s: 2 electrons; 2*p*: 6 electrons, 3s: 2 electrons. This adds up to 12 electrons. The remaining three electrons go into the 3*p* orbital, so the complete electron configuration of P is $1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^3$.

Energies of the highest occupied orbitals of the elements: This diagram illustrates the Aufbau rules as they are applied to all the elements. Note especially how the energies of the *nd* orbitals fall between the (*n*–1)*s* and (*n*–1)*p* orbitals so, for example the 3*d* orbitals begin to fill after the 4*s* orbital is filled, but before electrons populate the 4*p* orbitals. A similar relation exists with *d*- and *f*-orbitals.

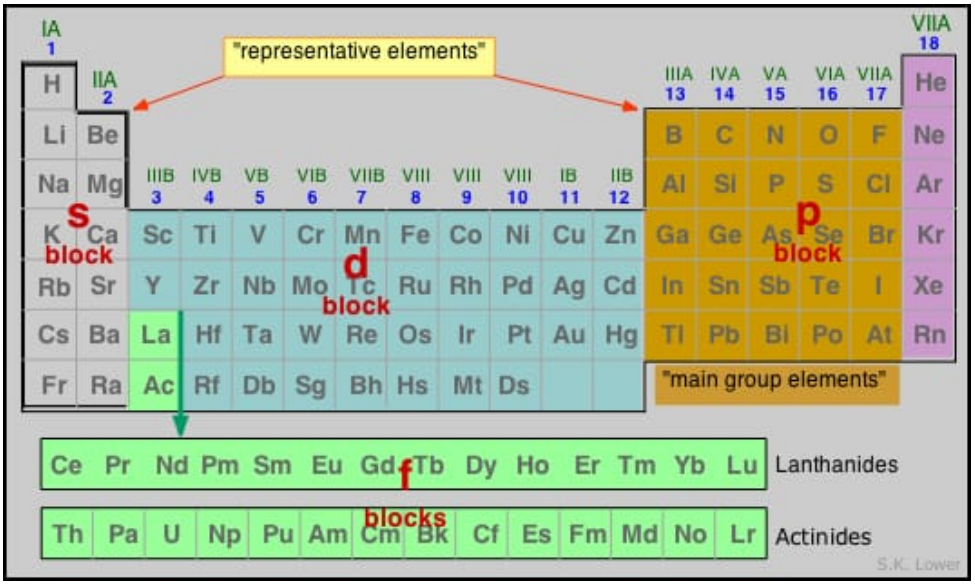


It is very important that you understand this diagram and how it follows from the Pauli exclusion principle: You should be able to reproduce it from memory up to the 6s level, because it forms the fundamental basis of the periodic table of the elements.

Bending the rules

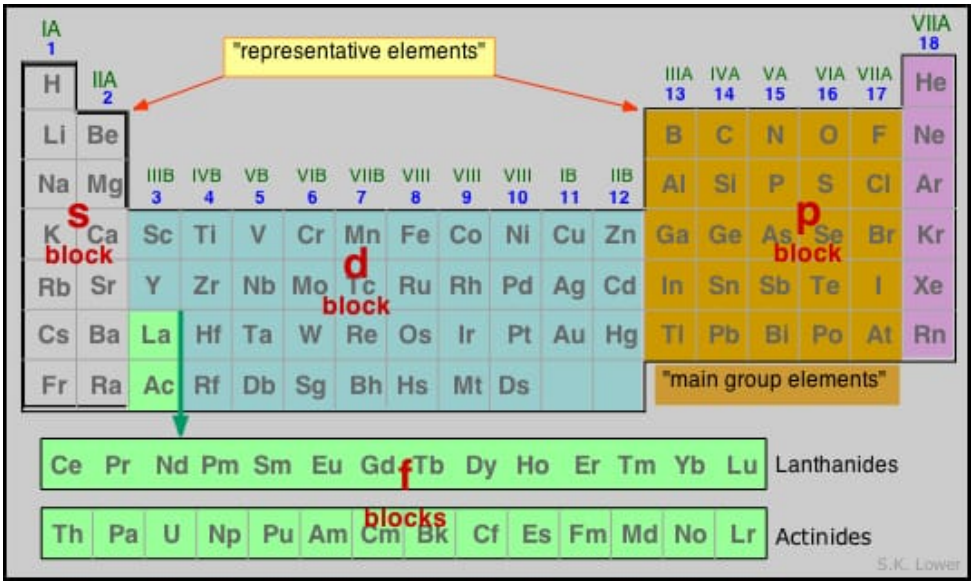
Inspection of a table of electron configurations of the elements reveals a few apparent non-uniformities in the filling of the orbitals, as is illustrated here for the elements of the so-called first transition series in which the 3*d* orbitals are being

populated. These anomalies are a consequence of the very small energy differences between some of the orbitals, and of the reduced electron-electron repulsion when electrons remain unpaired (Hund's rule), as is evident in chromium, which contains six unpaired electrons.



The other anomaly here is copper, which "should" have the outer-shell configuration $3d^9 4s^2$. The actual configuration of the Cu atom appears to be $3d^{10} 4s^1$. Although the 4*s* orbital is normally slightly below the 3*d* orbital energy, the two are so close that interactions between the two when one is empty and the other is not can lead to a reversal. Detailed calculations in which the shapes and densities of the charge distributions are considered predict that the relative energies of many orbitals can reverse in this way. It gets even worse when *f*-orbitals begin to fill!

Because these relative energies can vary even for the same atom in different chemical environments, most instructors will not expect you to memorize them.



This diagram shows how the atomic orbitals corresponding to different principal quantum numbers become interspersed with one another at higher values of *n*. The actual situation is more complicated than this; calculations show that the energies of *d* and *f* orbitals vary with the atomic number of the element.

The Periodic Table

The relative orbital energies illustrated above and the Pauli exclusion principle constitute the fundamental basis of the periodic table of the elements which was of course worked out empirically late in the 19th century, long before electrons had been heard of.

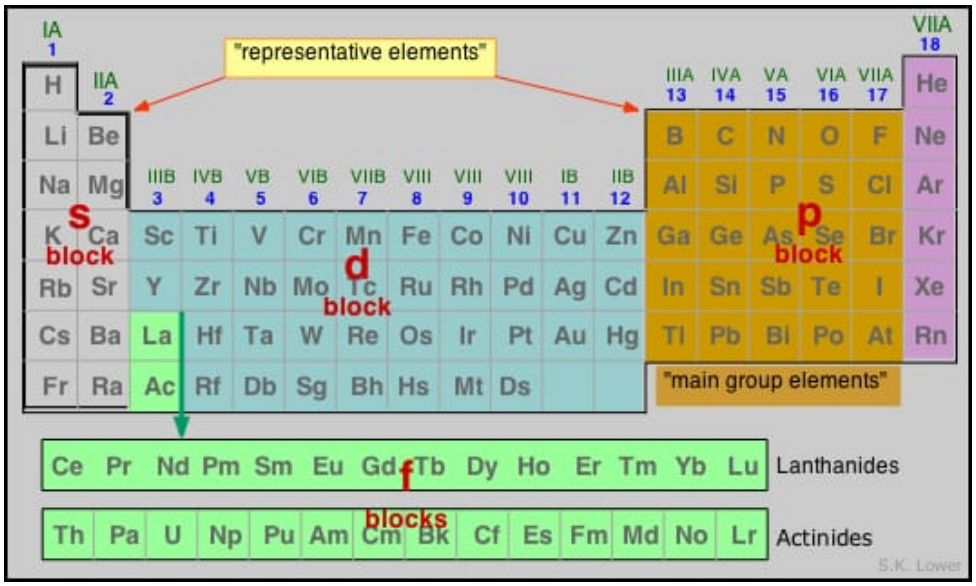


Figure 5.6.2: Structure of the periodic table

The periodic table of the elements is conventionally divided into sections called*blocks*, each of which designates the type of "sub-orbital" (*s, p, d, f*) which contains the highest-energy electrons in any particular element. Note especially that

- The non-metallic elements occur only in the *p*-block;
- The *d*-block elements contain the so-called transition elements;
- The *f*-block elements go in between Groups 3 and 4 of the *d*-block.

The above diagram illustrates the link between the electron configurations of the elements and the layout of the periodic table. Each row, also known as a *period*, commences with two s-block elements and continues through the *p* block. At the end of the rows corresponding to n>1 is an element having a p^6 configuration, a so-called *noble gas element*. At n values of 2 and 3, *d*- and *f*-block element sequences are added.

The table shown above is called the *long form* of the periodic table; for many purposes, we can use a "short form" table in which the *d*-block is shown below the *s*- and *p*- block "representative elements" and the *f*-block does not appear at all. Note that the "long form" would be even longer if the *f*-block elements were shown where they actually belong, between La-Hf and Ac-Db.

This page titled 5.6: Atomic Electron Configurations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.7: Periodic Properties of the Elements

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- You should be able to sketch out the general form of the periodic table and identify the various *blocks* and identify the *groups* corresponding to the *alkali metals*, the *transition elements*, the *halogens*, and the *noble gases*.
- For the first eighteen elements, you should be able to predict the formulas of typical binary compounds they can be expected to form with hydrogen and with oxygen.
- Comment on the concept of the "size" of an atom, and give examples of how radii are defined in at least two classes of of substances.
- Define *ionization energy* and *electron affinity*, and explain the periodic general trends.
- State the meaning and significance of *electronegativity*.

The periodic table in the form originally published by Dmitri Mendeleev in 1869 was an attempt to list the chemical elements in order of their atomic weights, while breaking the list into rows in such a way that elements having similar physical and chemical properies would be placed in each column. At that time, nothing was known about atoms; the development of the table was entirely empirical. Our goal in this lesson is to help you understand how the shape and organization of the modern periodic table are direct consequences of the atomic electronic structure of the elements.

Organization of the Periodic Table

To understand how the periodic table is organized, imagine that we write down a long horizontal list of the elements in order of their increasing atomic number. It would begin this way:

H He Li Be B C N O F Ne Na Mg Al Si P S Cl Ar K Ca...

Now if we look at the various physical and chemical properties of these elements, we would find that their values tend to increase or decrease with Z in a manner that reveals a repeating pattern— that is, a *periodicity*. For the elements listed above, these breaks can be indicated by the vertical bars shown here in color:

H **He** | Li Be B C N O F **Ne** | Na Mg Al Si P S Cl **Ar** | Ca ...

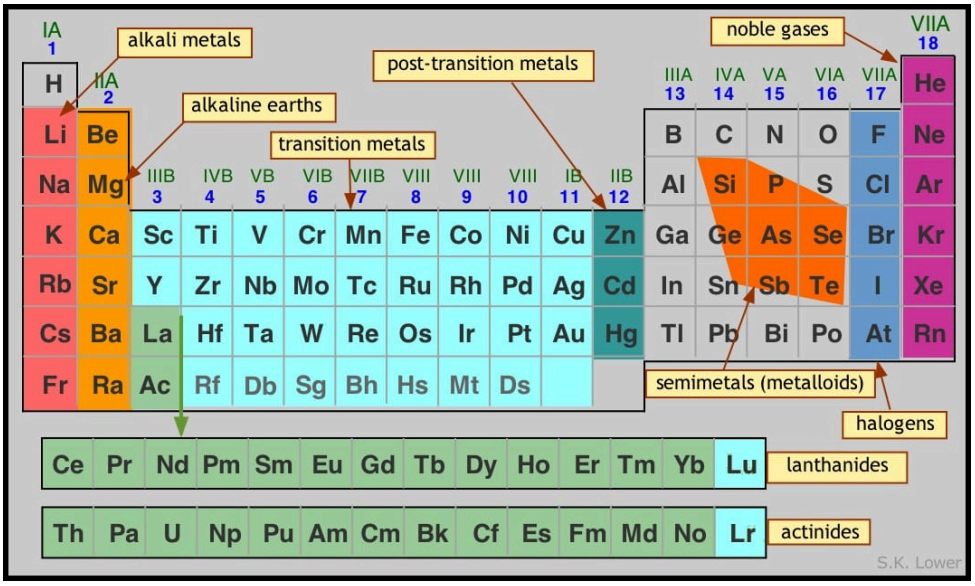
Periods: To construct the table, we place each sequence in a separate row, which we call a *period*. The rows are aligned in such a way that the elements in each vertical column possess certain similarities. Thus the first short-period elements H and He are chemically similar to the elements Li and Ne at the beginning and end of the second period. The first period is split in order to place H above Li and He above Ne.

The "block" nomenclature shown above refers to the sub-orbital type (quantum number l, or s-p-d-f classification) of the highest-energy orbitals that are occupied in a given element. For n=1 there is no p block, and the s block is split so that helium is placed in the same group as the other inert gases, which it resembles chemically. For the second period (n=2) there is a p block but no d block; in the usual "long form" of the periodic table it is customary to leave a gap between these two blocks in order to accommodate the d blocks that occur at n=3 and above. At n=6 we introduce an f block, but in order to hold the table to reasonable dimensions the f blocks are placed below the main body of the table.

Groups: Each column of the periodic table is known as a *group*. The elements belonging to a given group bear a strong similarity in their chemical behaviors.

In the past, two different systems of Roman numerals and letters were used to denote the various groups. North Americans added the letter B to denote the d-block groups and A for the others; this is the system shown in the table above. The the rest of the world used A for the d-block elements and B for the others. In 1985, a new international system was adopted in which the columns were simply labeled 1-18. Although this system has met sufficient resistance in North America to slow its incorporation into textbooks, it seems likely that the "one to eighteen" system will gradually take over.

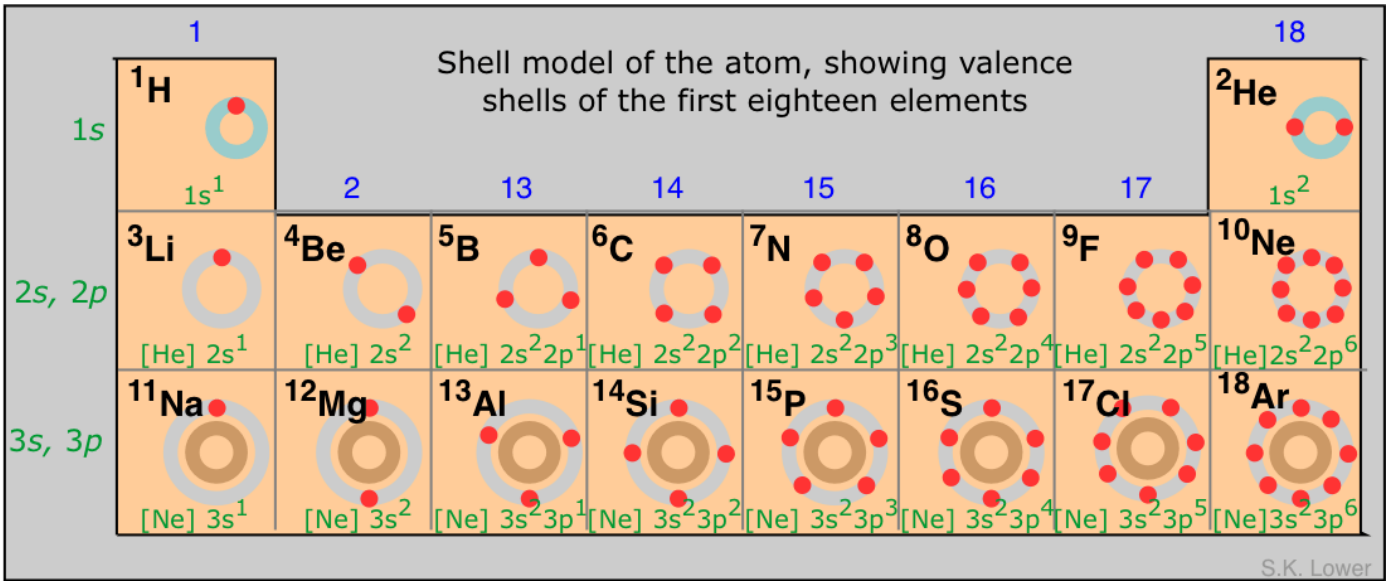
Families. Chemists have long found it convenient to refer to the elements of different groups, and in some cases of spans of groups by the names indicated in the table shown below. The two of these that are most important for you to know are the *noble gases* and the*transition metals*.



We begin with the image you saw in the preceding lesson, showing the long form of the table with the "block" structure emphasized. You will recall that the two f blocks are written at the bottom merely to keep the table from becoming inconveniently wide; these two blocks actually go in between La-Hf and Ac-Db, respectively, in the d block.

The shell model of the atom

The properties of an atom depend ultimately on the number of electrons in the various orbitals, and on the nuclear charge which determines the compactness of the orbitals. In order to relate the properties of the elements to their locations in the periodic table, it is often convenient to make use of a simplified view of the atom in which the nucleus is surrounded by one or more concentric spherical "shells", each of which consists of the highest-principal quantum number orbitals (always *s*- and *p*-orbitals) that contain at least one electron. The shell model (as with any scientific model) is less a description of the world than a simplified way of looking at it that helps us to understand and correlate diverse phenomena. The principal simplification here is that it deals only with the *main group elements* of the *s*- and *p*-blocks, omitting the *d*- and *f*-block elements whose properties tend to be less closely tied to their group numbers.



The electrons (denoted by the red dots) in the outer-most shell of an atom are the ones that interact most readily with other atoms, and thus play a major role in governing the chemistry of an element. Notice the use of noble-gas symbols to simplify the electron-configuration notation.

In particular, the number of outer-shell electrons (which is given by the rightmost digit in the *group number*) is a major determinant of an element's "combining power", or *valence*. The general trend is for an atom to gain or lose electrons, either directly (leading to formation of *ions*) or by sharing electrons with other atoms so as to achieve an outer-shell configuration of s^2p^6. This configuration, known as an *octet*, corresponds to that of one of the noble-gas elements of Group 18.

- the elements in Groups 1, 2 and 13 tend to give up their valence electrons to form positive ions such as Na^+, Mg^{2+} and Al^{3+}, as well as compounds NaH, MgH_2 and AlH_3. The outer-shell configurations of the metal atoms in these species correspond to that of neon.
- elements in Groups 15-17 tend to acquire electrons, forming ions such as P^{3-}, S^{2-} and Cl^- or compounds such as PH_3, H_2S and HCl. The outer-shell configurations of these elements correspond to that of argon.
- the Group 14 elements do not normally form ions at all, but share electrons with other elements in tetravalent compounds such as CH_4.

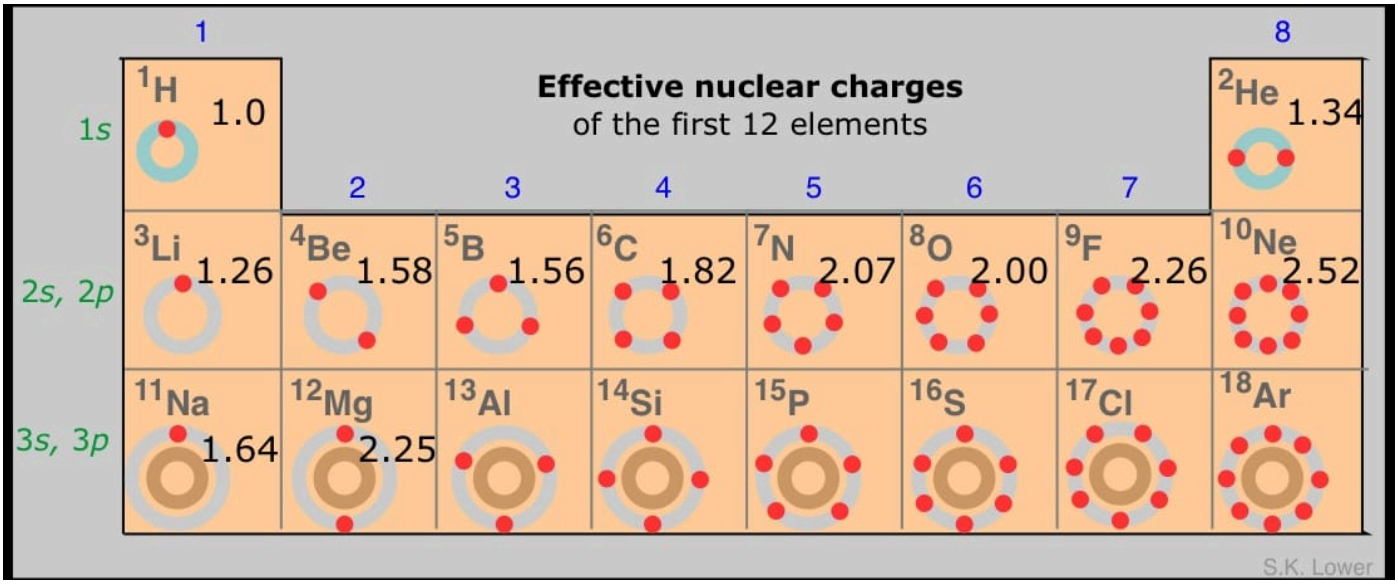
The above diagram shows the first three rows of what are known as the *representative elements*— that is, the s- and *p*-block elements only. As we move farther down (into the fourth row and below), the presence of *d*-electrons exerts a complicating influence which allows elements to exhibit multiple valances. This effect is especially noticeable in the transition-metal elements, and is the reason for not including the d-block with the representative elements at all.

Effective nuclear charge

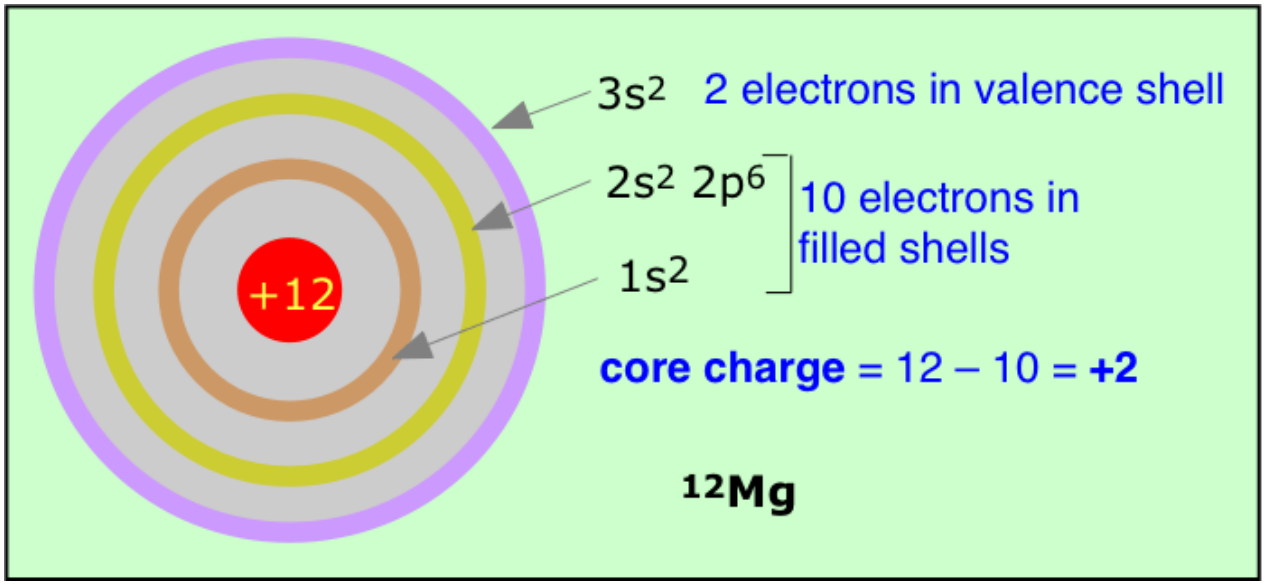
Those electrons in the outmost or valence shell are especially important because they are the ones that can engage in the sharing and exchange that is responsible for chemical reactions; how tightly they are bound to the atom determines much of the chemistry of the element. The degree of binding is the result of two opposing forces: the attraction between the electron and the nucleus, and the repulsions between the electron in question and all the other electrons in the atom. All that matters is the net force, the difference between the nuclear attraction and the totality of the electron-electron repulsions.

We can simplify the shell model even further by imagining that the valence shell electrons are the *only* electrons in the atom, and that the nuclear charge has whatever value would be required to bind these electrons as tightly as is observed experimentally. Because the number of electrons in this model is less than the atomic number *Z*, the required nuclear charge will also be smaller. and is known as the *effective nuclear charge*. Effective nuclear charge is essentially the positive charge that a valence electron "sees".

Part of the difference between Z and $Z_{effective}$ is due to other electrons in the valence shell, but this is usually only a minor contributor because these electrons tend to act as if they are spread out in a diffuse spherical shell of larger radius. The main actors here are the electrons in the much more compact inner shells which surround the nucleus and exert what is often called a shielding or "*screening*" effect on the valence electrons.

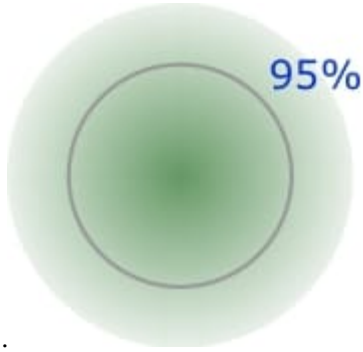


The formula for calculating effective nuclear charge is not very complicated, but we will skip a discussion of it here. An even simpler although rather crude procedure is to just subtract the number of inner-shell electrons from the nuclear charge; the result is a form of effective nuclear charge which is called the *core charge* of the atom.



Sizes of atoms and ions

The concept of "size" is somewhat ambiguous when applied to the scale of atoms and molecules. The reason for this is apparent when you recall that an atom has no definite boundary; there is a finite (but very small) probability of finding the electron of a hydrogen atom, for example, 1 cm, or even 1 km from the nucleus. It is not possible to specify a definite value for the radius of an isolated atom; the best we can do is to define a spherical shell within whose radius some arbitrary percentage of the electron density can be found



When an atom is combined with other atoms in a solid element or compound, an effective radius can be determined by observing the distances between adjacent rows of atoms in these solids. This is most commonly carried out by X-ray scattering experiments. Because of the different ways in which atoms can aggregate together, several different kinds of atomic radii can be defined.

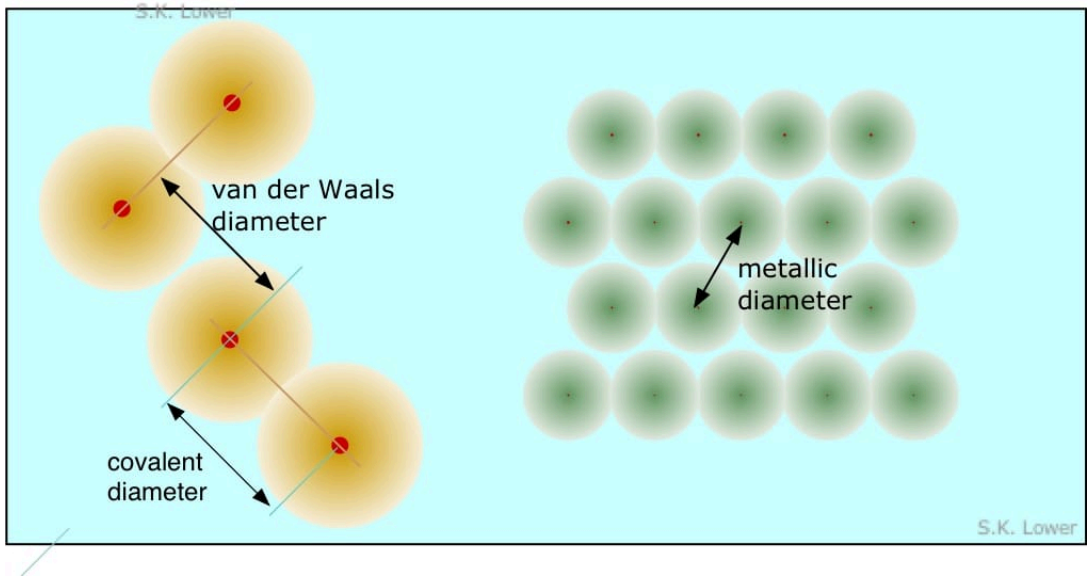
Distances on the atomic scale have traditionally been expressed in Ångstrom units (1Å = 10^{-8} cm = 10^{-10} m), but nowadays the picometer is preferred:

$$1\ \text{pm} = 10^{-12}\ \text{m} = 10^{-10}\ \text{cm} = 10^{-2}\ \text{Å},\ \text{or}\ 1\text{Å} = 100\ \text{pm}.$$

The radii of atoms and ions are typically in the range 70-400 pm.

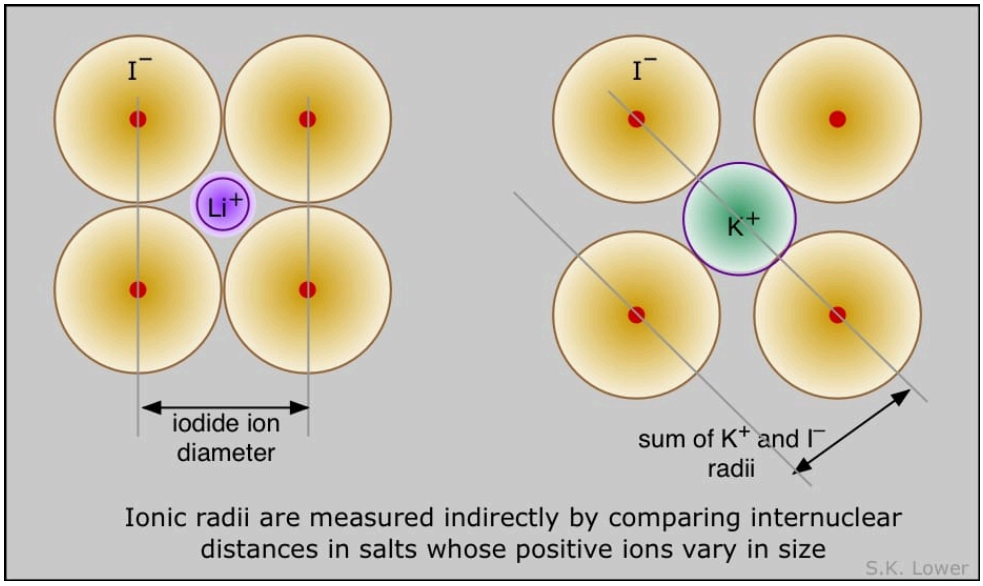
A rough idea of the size of a metallic atom can be obtained simply by measuring the density of a sample of the metal. This tells us the number of atoms per unit volume of the solid. The atoms are assumed to be spheres of radius *r* in contact with each other, each of which sits in a cubic box of edge length 2*r*. The volume of each box is just the total volume of the solid divided by the number of atoms in that mass of the solid; the atomic radius is the cube root of *r*.

Although the radius of an atom or ion cannot be measured directly, in most cases it can be inferred from measurements of the distance between adjacent nuclei in a crystalline solid. This is most commonly carried out by X-ray scattering experiments. Because such solids fall into several different classes, several kinds of atomic radius are defined. Many atoms have several different radii; for example, sodium forms a metallic solid and thus has a metallic radius, it forms a gaseous molecule Na_2 in the vapor phase (covalent radius), and of course it forms ionic solids such as NaCl.



molecules in crystalline iodine.

Ionic radius is the effective radius of ions in solids such as NaCl. It is easy enough to measure the distance between adjacent rows of Na^+ and Cl^- ions in such a crystal, but there is no unambiguous way to decide what portions of this distance are attributable to each ion. The best one can do is make estimates based on studies of several different ionic solids (LiI, KI, NaI, for example) that contain one ion in common. Many such estimates have been made, and they turn out to be remarkably consistent.

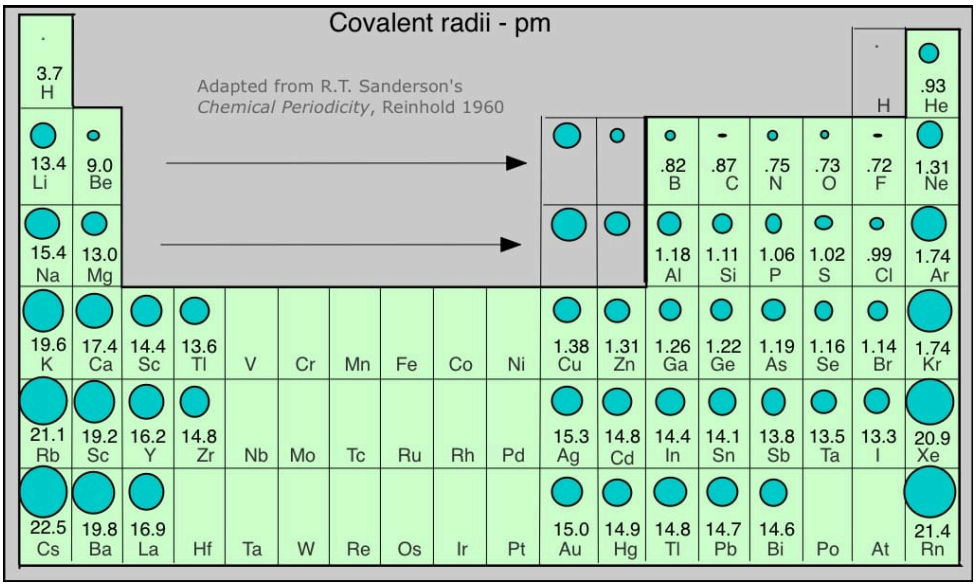


The lithium ion is sufficiently small that in LI, the iodide ions are in contact, so I-I distances are twice the ionic radius of I^-. This is not true for KI, but in this solid, adjacent potassium and iodide ions are in contact, allowing estimation of the K^+ ionic radius.

Many atoms have several different radii; for example, sodium forms a metallic solid and thus has a metallic radius, it forms a gaseous molecule Na_2 in the vapor phase (covalent radius), and of course it forms ionic solids as mentioned above.

Periodic trends in atomic size

We would expect the size of an atom to depend mainly on the principal quantum number of the highest occupied orbital; in other words, on the "number of occupied electron shells". Since each row in the periodic table corresponds to an increment in n, atomic radius increases as we move down a column. The other important factor is the nuclear charge; the higher the atomic number, the more strongly will the electrons be drawn toward the nucleus, and the smaller the atom. This effect is responsible for the contraction we observe as we move across the periodic table from left to right.



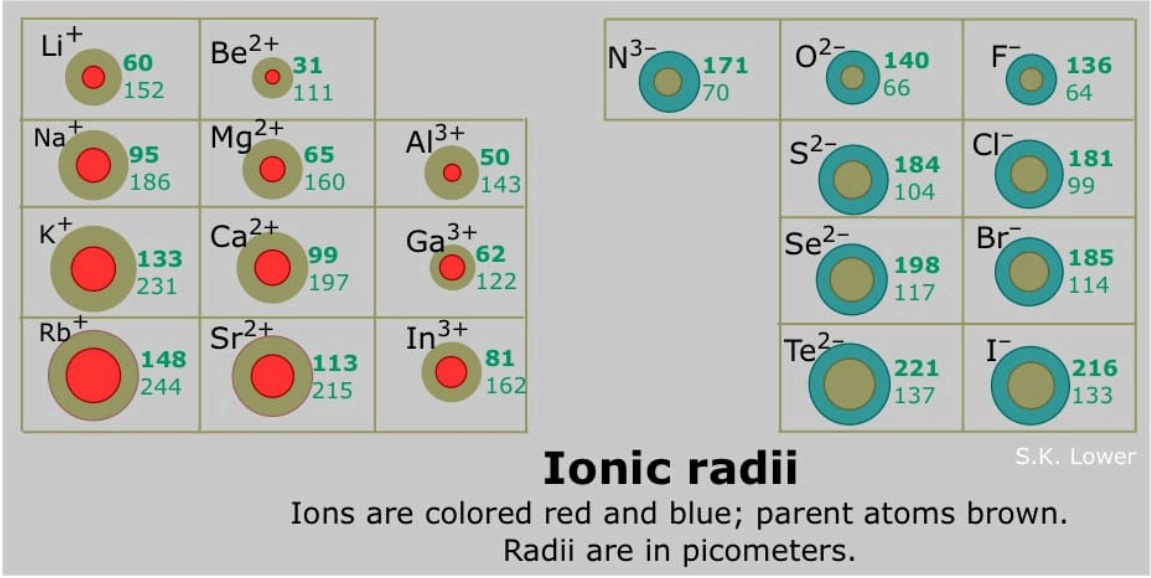
The figure shows a periodic table in which the sizes of the atoms are represented graphically. The apparent discontinuities in this diagram reflect the difficulty of comparing the radii of atoms of metallic and nonmetallic bonding types. Radii of the noble gas elements are estimates from those of nearby elements.

Ionic radii

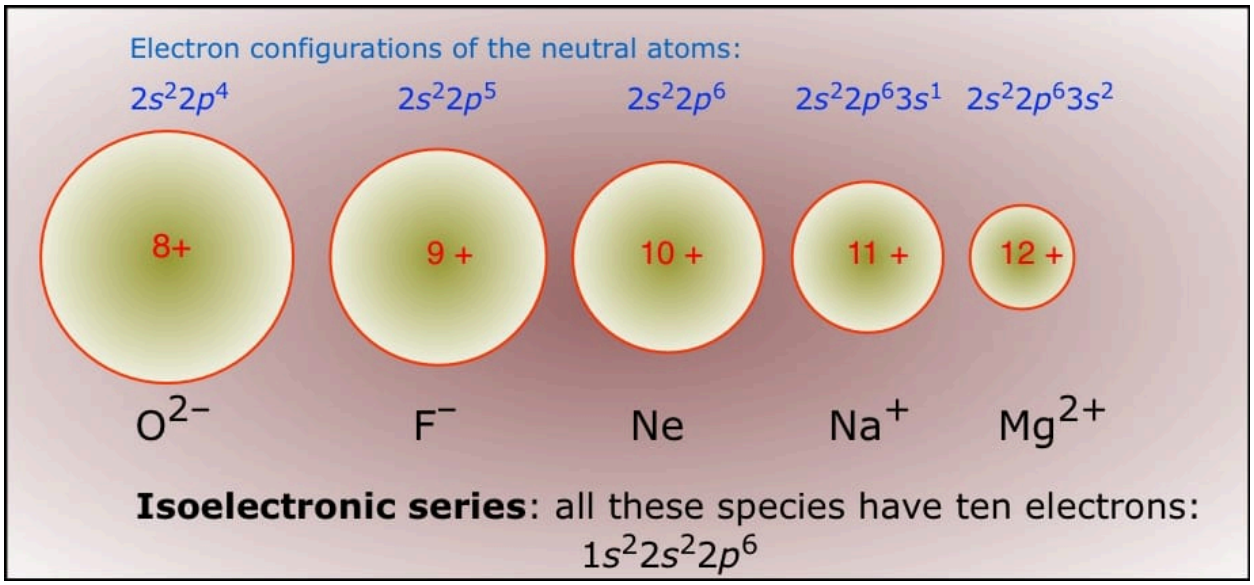
A positive ion is always smaller than the neutral atom, owing to the diminished electron-electron repulsion. If a second electron is lost, the ion gets even smaller; for example, the ionic radius of Fe^{2+} is 76 pm, while that of Fe^{3+} is 65 pm. If formation of the ion involves complete emptying of the outer shell, then the decrease in radius is especially great.

The hydrogen ion H^+ is in a class by itself; having no electron cloud at all, its radius is that of the bare proton, or about 0.1 pm— a contraction of 99.999%! Because the unit positive charge is concentrated into such a small volume of space, the *charge density* of the hydrogen ion is extremely high; it interacts very strongly with other matter, including water molecules, and in aqueous solution it exists only as the *hydronium ion* H_3O^+.

Negative ions are always larger than the parent ion; the addition of one or more electrons to an existing shell increases *electron-electron repulsion* which results in a general expansion of the atom.



An *isoelectronic series* is a sequence of species all having the same number of electrons (and thus the same amount of electron-electron repulsion) but differing in nuclear charge. Of course, only one member of such a sequence can be a neutral atom (neon in the series shown below.) The effect of increasing nuclear charge on the radius is clearly seen.



Periodic Trends in ion formation

Chemical reactions are based largely on the interactions between the most loosely bound electrons in atoms, so it is not surprising that the tendency of an atom to gain, lose or share electrons is one of its fundamental chemical properties.

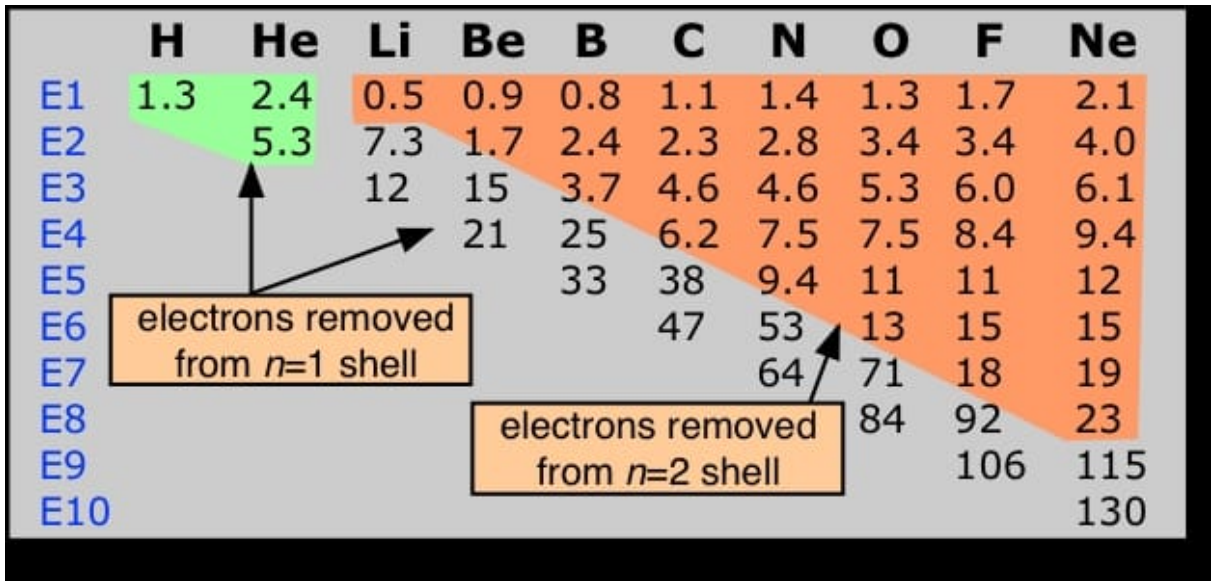
Ionization Energy

This term always refers to the formation of *positive* ions. In order to remove an electron from an atom, work must be done to overcome the electrostatic attraction between the electron and the nucleus; this work is called the *ionization energy* of the atom and corresponds to the exothermic process

$$M_{(g)} \rightarrow M^+_{(g)} + e^- \tag{5.7.1}$$

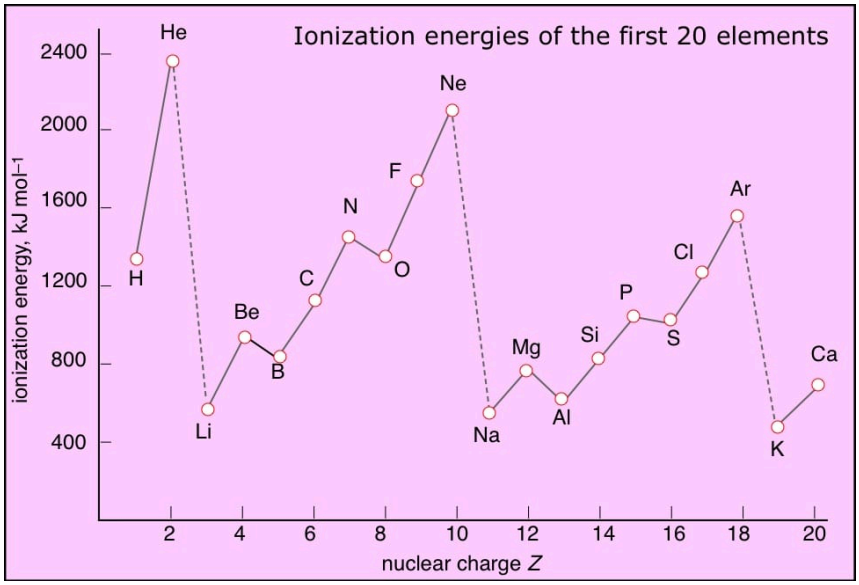
where $M_{(g)}$ stands for any isolated (gaseous) atom.

An atom has as many ionization energies as it has electrons. Electrons are always removed from the highest-energy occupied orbital. An examination of the successive ionization energies of the first ten elements (below) provides experimental confirmation that the binding of the two innermost electrons (1s orbital) is significantly different from that of the *n*=2 electrons.Successive ionization energies of an atom increase rapidly as reduced electron-electron repulsion causes the electron shells to contract, thus binding the electrons even more tightly to the nucleus.

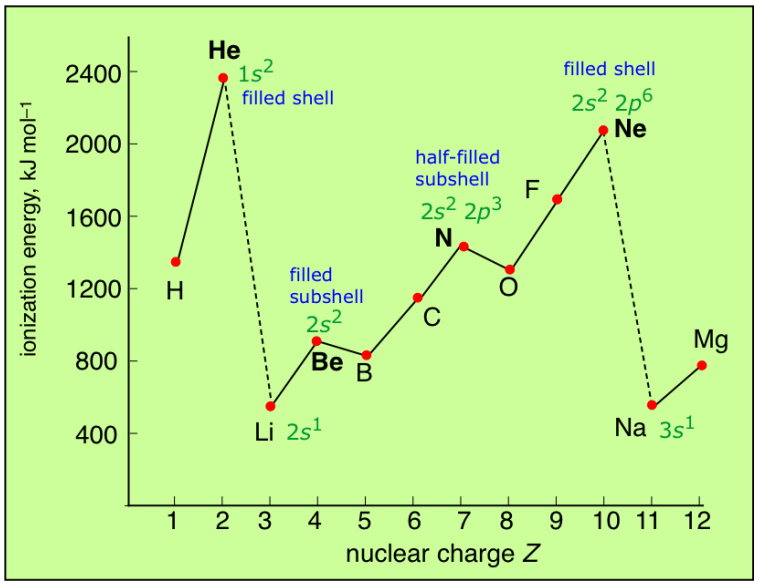


Successive ionizations of the first ten elements. Note the very large jumps in the energies required to remove electrons from the 1s orbitals of atoms of the second-row elements Li-Ne.

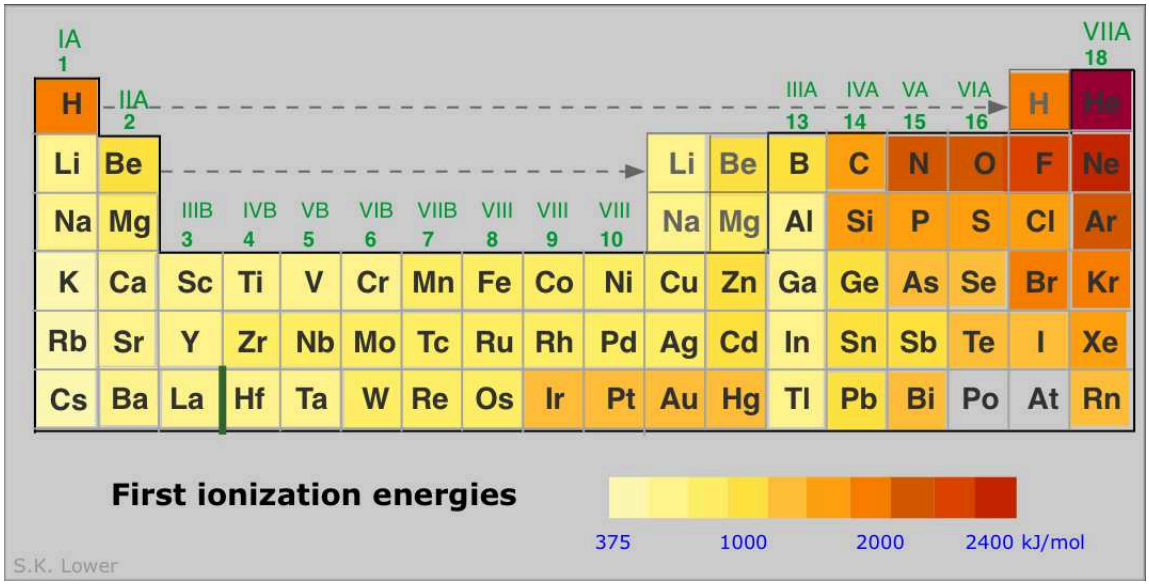
Ionization energies increase with the nuclear charge Z as we move across the periodic table. They decrease as we move down the table because in each period the electron is being removed from a shell one step farther from the nucleus than in the atom immediately above it. This results in the familiar zig-zag lines when the first ionization energies are plotted as a function of Z.



This more detailed plot of the ionization energies of the atoms of the first ten elements reveals some interesting irregularities that can be related to the slightly lower energies (greater stabilities) of electrons in half-filled (spin-unpaired) relative to completely-filled subshells.



Finally, a more comprehensive survey of the ionization energies of the main group elements is shown below.



Some points to note:

- The noble gases have the highest IE's of any element in the period. This has nothing to do with any mysterious "special stability" of the s^2p^6 electron configuration; it is simply a matter of the high nuclear charge acting on more contracted orbitals.
- IE's (as well as many other properties) tend not to vary greatly amongst the *d*-block elements. This reflects the fact that as the more-compact *d* orbitals are being filled, they exert a screening effect that partly offsets that increasing nuclear charge on the outermost *s* orbitals of higher principal quantum number.
- Each of the Group 13 elements has a lower first-IE than that of the element preceding it. The reversal of the IE trend in this group is often attributed to the more easy removal of the single outer-shell p electron compared to that of electrons contained in filled (and thus spin-paired) *s*- and *d*-orbitals in the preceding elements.

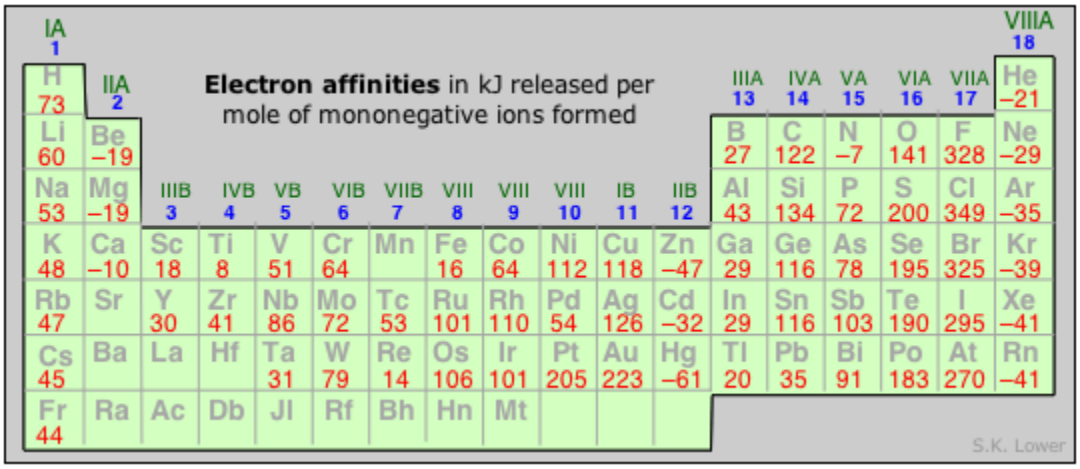
Electron affinity

Formation of a negative ion occurs when an electron from some external source enters the atom and become incorporated

into the lowest energy orbital that possesses a vacancy. Because the entering electron is attracted to the positive nucleus, the formation of negative ions is usually exothermic. The energy given off is the electron affinity of the atom. For some atoms, the electron affinity appears to be slightly negative, suggesting that electron-electron repulsion is the dominant factor in these instances.

In general, electron affinities tend to be much smaller than ionization energies, suggesting that they are controlled by opposing factors having similar magnitudes. These two factors are, as before, the nuclear charge and electron-electron repulsion. But the latter, only a minor actor in positive ion formation, is now much more significant. One reason for this is that the electrons contained in the inner shells of the atom exert a collective negative charge that partially cancels the charge of the nucleus, thus exerting a so-called shielding effect which diminishes the tendency for negative ions to form.

Because of these opposing effects, the periodic trends in electron affinities are not as clear as are those of ionization energies. This is particularly evident in the first few rows of the periodic table, in which small effects tend to be magnified anyway because an added electron produces a large percentage increase in the number of electrons in the atom.



In general, we can say that electron affinities become more exothermic as we move from left to right across a period (owing to increased nuclear charge and smaller atom size). There are some interesting irregularities, however:

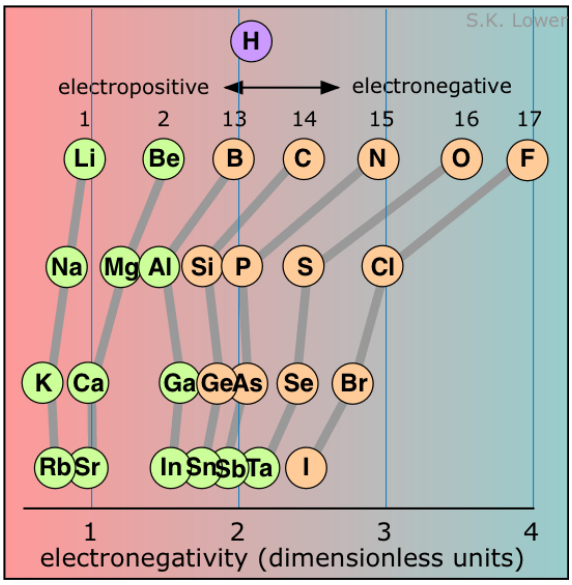
- In the Group 2 elements, the filled 2s orbital apparently shields the nucleus so effectively that the electron affinities are slightly endothermic.
- The Group 15 elements have rather low values, due possibly to the need to place the added electron in a half-filled *p* orbital; why the electron affinity of nitrogen should be endothermic is not clear. The vertical trend is for electron affinity to become less exothermic in successive periods owing to better shielding of the nucleus by more inner shells and the greater size of the atom, but here also there are some apparent anomalies.

Electronegativity

When two elements are joined in a chemical bond, the element that attracts the shared electrons more strongly is more *electronegative*. Elements with low electronegativities (the metallic elements) are said to be *electropositive*.

> Electronegativities are properties of atoms that are chemically bound to each other; there is no way of measuring the electronegativity of an isolated atom.

Moreover, the same atom can exhibit different electronegativities in different chemical environments, so the "electronegativity of an element" is only a general guide to its chemical behavior rather than an exact specification of its behavior in a particular compound. Nevertheless, electronegativity is eminently useful in summarizing the chemical behavior of an element. You will make considerable use of electronegativity when you study chemical bonding and the chemistry of the individual elements.



Because there is no single definition of electronegativity, any numerical scale for measuring it must of necessity be somewhat arbitrary. Most such scales are themselves based on atomic properties that are directly measurable and which relate in one way or the other to electron-attracting propensity. The most widely used of these scales was devised by Linus Pauling and is related to ionization energy and electron affinity. The Pauling scale runs from 0 to 4; the highest electron affinity, 4.0, is assigned to fluorine, while cesium has the lowest value of 0.7. Values less than about 2.2 are usually associated with electropositive, or metallic character. In the representation of the scale shown in figure, the elements are arranged in rows corresponding to their locations in the periodic table. The correlation is obvious; electronegativity is associated with the higher rows and the rightmost columns.

The location of **hydrogen** on this scale reflects some of the significant chemical properties of this element. Although it acts like a metallic element in many respects (forming a positive ion, for example), it can also form hydride-ion (H^-) solids with the more electropositive elements, and of course its ability to share electrons with carbon and other *p*-block elements gives rise to a very rich chemistry, including of course the millions of organic compounds.

This page titled 5.7: Periodic Properties of the Elements is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

5.8: Why Don't Electrons Fall into the Nucleus?

The picture of electrons "orbiting" the nucleus like planets around the sun remains an enduring one, not only in popular images of the atom but also in the minds of many of us who know better. The proposal, first made in 1913, that the centrifugal force of the revolving electron just exactly balances the attractive force of the nucleus (in analogy with the centrifugal force of the moon in its orbit exactly counteracting the pull of the Earth's gravity) is a nice picture, but is simply untenable.

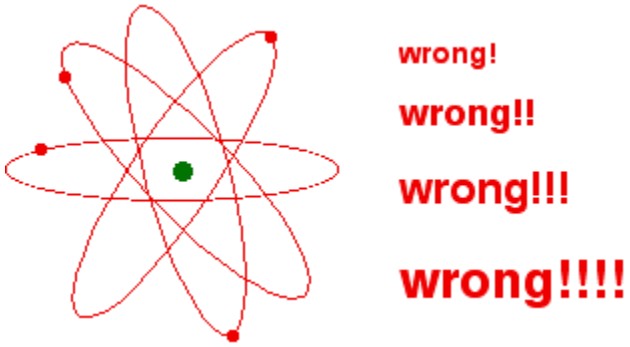


Figure 5.8.1: Most popular science pictures of the atom show electrons moving around a nucleus like planets around the sun. These pictures are quite simply wrong. They come from an old idea about the structure of the atom and have lasted, partly from habit, and partly because the modern view of the arrangement of electrons in the atom is too difficult to draw simple pictures of.

One origin for this hypothesis that suggests this perspective is plausible is the similarity of the gravity and Coulombic interactions. The expression for the force of gravity between two masses (Newton's Law of gravity) is

$$F_{gravity} \propto \frac{m_1 m_2}{r^2} \tag{5.8.1}$$

with

- m_1 and m_2 representing the **mass** of object 1 and 2, respectively and
- r representing the distance between the objects centers

The expression for the Coulomb force between two charged species is

$$F_{Coulomb} \propto \frac{q_1 q_2}{r^2} \tag{5.8.2}$$

with

- q_1 and q_2 representing the **charge** of object 1 and 2, respectively and
- r representing the distance between the objects centers

However, an electron, unlike a planet or a satellite, is electrically charged, and it has been known since the mid-19th century that an electric charge that undergoes acceleration (changes velocity and direction) will emit electromagnetic radiation, losing energy in the process. A revolving electron would transform the atom into a miniature radio station, the energy output of which would be at the cost of the potential energy of the electron; according to classical mechanics, the electron would simply spiral into the nucleus and the atom would collapse.

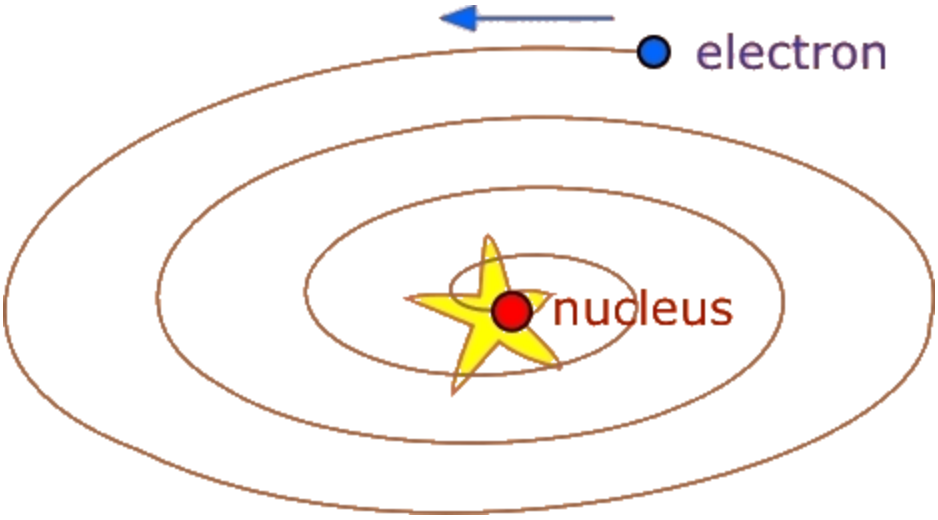
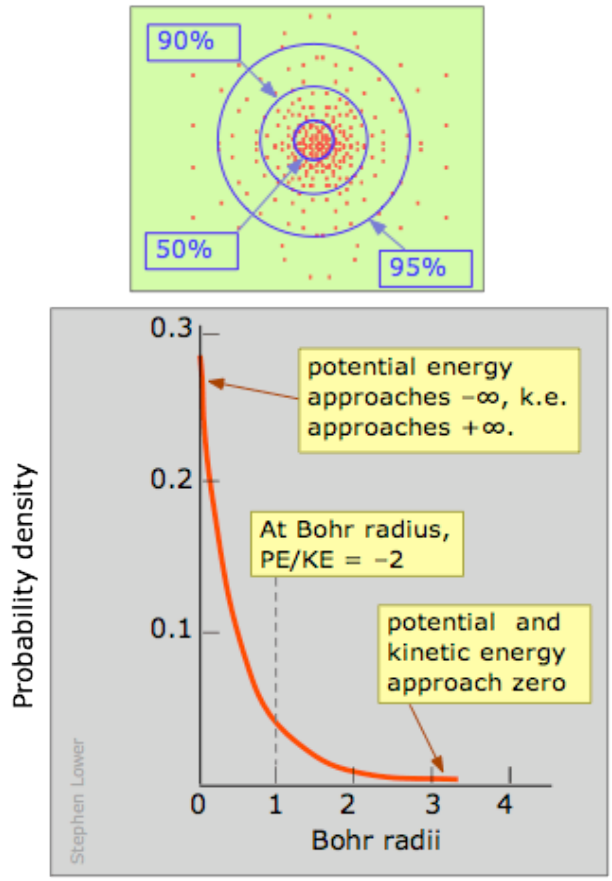


Figure 5.8.2: The classical death spiral of an electron around a nucleus.

Quantum theory to the Rescue!

By the 1920's, it became clear that a tiny object such as the electron cannot be treated as a classical particle having a definite position and velocity. The best we can do is specify the probability of its manifesting itself at any point in space. If you had a magic camera that could take a sequence of pictures of the electron in the 1s orbital of a hydrogen atom, and could combine the resulting dots in a single image, you would see something like this. Clearly, the electron is more likely to be found the closer we move toward the nucleus.



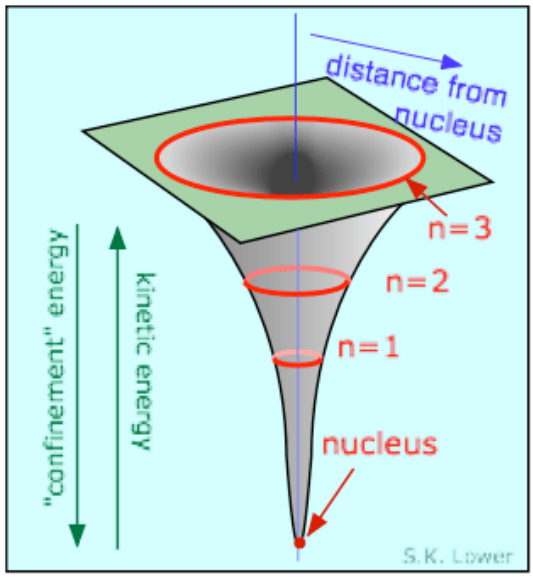
This is confirmed by this plot which shows the quantity of electron charge per unit volume of space at various distances from the

nucleus. This is known as a probability density plot. The per unit volume of space part is very important here; as we consider radii closer to the nucleus, these volumes become very small, so the number of electrons per unit volume increases very rapidly. In this view, it appears as if the electron does fall into the nucleus!

> According to classical mechanics, the electron would simply spiral into the nucleus and the atom would collapse. Quantum mechanics is a different story.

The Battle of the Infinities Saves the electron from its death spiral

As you know, the potential energy of an electron becomes more negative as it moves toward the attractive field of the nucleus; in fact, it approaches negative infinity. However, because the total energy remains constant (a hydrogen atom, sitting peacefully by itself, will neither lose nor acquire energy), the loss in potential energy is compensated for by an increase in the electron's kinetic energy (sometimes referred to in this context as "confinement" energy) which determines its momentum and its effective velocity.



So as the electron approaches the tiny volume of space occupied by the nucleus, its potential energy dives down toward minus-infinity, and its kinetic energy (momentum and velocity) shoots up toward positive-infinity. This "battle of the infinities" cannot be won by either side, so a compromise is reached in which theory tells us that the fall in potential energy is just twice the kinetic energy, and the electron dances at an average distance that corresponds to the Bohr radius.

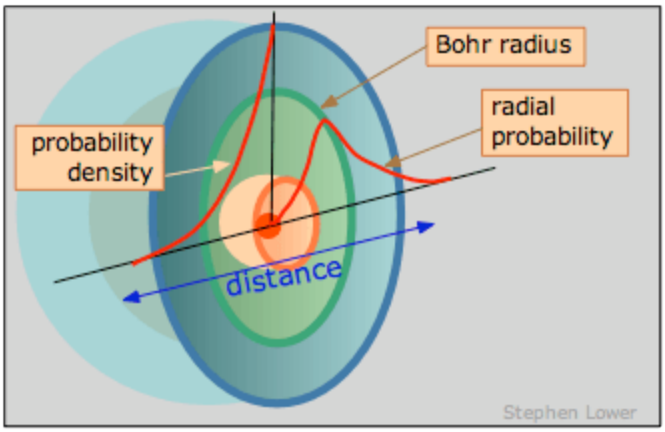
There is still one thing wrong with this picture; according to the Heisenberg uncertainty principle (a better term would be "indeterminacy"), a particle as tiny as the electron cannot be regarded as having either a definite location or momentum. The Heisenberg principle says that either the location or the momentum of a quantum particle such as the electron can be known as precisely as desired, but as one of these quantities is specified more precisely, the value of the other becomes increasingly indeterminate. It is important to understand that this is not simply a matter of observational difficulty, but rather a fundamental property of nature.

What this means is that within the tiny confines of the atom, the electron cannot really be regarded as a "particle" having a definite energy and location, so it is somewhat misleading to talk about the electron "falling into" the nucleus.

> Arthur Eddington, a famous physicist, once suggested, not entirely in jest, that a better description of the electron would be "wavicle"!

Probability Density vs. Radial probability

We can, however, talk about where the electron has the highest probability of manifesting itself— that is, where the maximum negative charge will be found.



This is just the curve labeled "probability density"; its steep climb as we approach the nucleus shows unambiguously that the electron is most likely to be found in the tiny volume element at the nucleus. But wait! Did we not just say that this does not happen? What we are forgetting here is that as we move out from the nucleus, the number of these small volume elements situated along any radius increases very rapidly with r, going up by a factor of $4\pi r^2$. So the probability of finding the electron somewhere on a given radius circle is found by multiplying the probability density by $4\pi r^2$. This yields the curve you have probably seen elsewhere, known as the radial probability, that is shown on the right side of the above diagram. The peak of the radial probability for principal quantum number $n = 1$ corresponds to the Bohr radius.

To sum up, the probability density and radial probability plots express two different things: the first shows the electron density at any single point in the atom, while the second, which is generally more useful to us, tells us the the relative electron density summed over all points on a circle of given radius.

References

- Why Doesn't the Electron Fall Into the Nucleus? Franklin Mason and Robert Richardson, J Chem. Ed. 1983 (40-42). See also the comment on this article by Werner Luck, J Chem Ed 1985 (914).
- For more detailed descriptions of these two kinds of plots, see this McMaster U. page by Richard Bader.
- The author is grateful to Robert Harrison of U. of Tennessee-Knoxville whose suggestions led to improving this article.

Contributors and Attributions

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

This page titled 5.8: Why Don't Electrons Fall into the Nucleus? is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

- Why do electrons not fall into the nucleus? by Stephen Lower is licensed CC BY 4.0. Original source: http://www.chem1.com/acad/webtext/virtualtextbook.html.

6: PROPERTIES OF GASES

The gaseous state of matter is the only one that is based on a simple model that can be developed from first principles. As such, it serves as the starting point for the study of the other states of matter.

6.1: Observable Properties of Gas

6.2: Ideal Gas Model - The Basic Gas Laws

6.3: Dalton's Law

6.4: Kinetic Molecular Theory (Overview)

6.5: More on Kinetic Molecular Theory

6.6: Real Gases and Critical Phenomena

This page titled 6: Properties of Gases is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Observable Properties of Gas

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

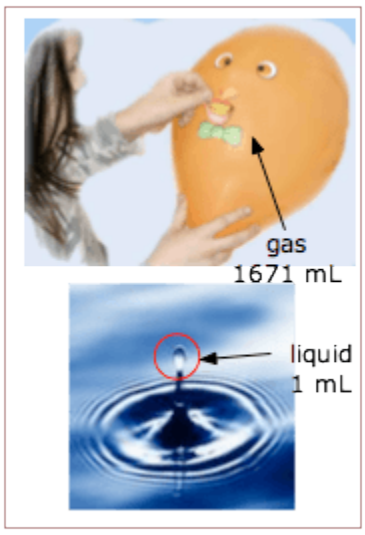
- State the ***three major properties of gases*** that distinguish them from condensed phases of matter.
- Define ***pressure***, and explain why a gas exerts pressure on the walls of a container.
- Explain the operation of a simple *barometer*, and why its invention revolutionized our understanding of gases.
- Explain why a barometer that uses water as the barometric fluid is usually less practical than one which employs mercury.
- How are the Celsius and Fahrenheit *temperature scale*s defined? How do the magnitudes of the "degree" on each scale related?
- Why must the temperature and pressure be specified when reporting the *volume* of a gas?

The invention of the sensitive *balance* in the early seventeenth century showed once and for all that gases have weight and are therefore matter. Guericke's invention of air pump (which led directly to his discovery of the *vacuum*) launched the “pneumatic era" of chemistry long before the existence of atoms and molecules had been accepted. Indeed, the behavior of gases was soon to prove an invaluable tool in the development of the atomic theory of matter.

Introduction

The study of gases allows us to understand the behavior of matter at its simplest: individual particles, acting independently, almost completely uncomplicated by interactions and interferences between each other. This knowledge of gases will serve as the pathway to our understanding of the far more complicated *condensed phases* (liquids and solids) in which the theory of gases will no longer give us correct answers, but it will still provide us with a useful *model* that will at least help us to rationalize the behavior of these more complicated states of matter.

First, we know that a gas has **no definite volume or shape**; a gas will fill whatever volume is available to it. Contrast this to the behavior of a liquid, which always has a distinct upper surface when its volume is less than that of the space it occupies. The other outstanding characteristic of gases is their **low *densities***, compared with those of liquids and solids. One mole of liquid water at 298 K and 1 atm pressure occupies a volume of 18.8 cm^3, whereas the same quantity of water vapor at the same temperature and pressure has a volume of 30200 cm^3, more than 1000 times greater.

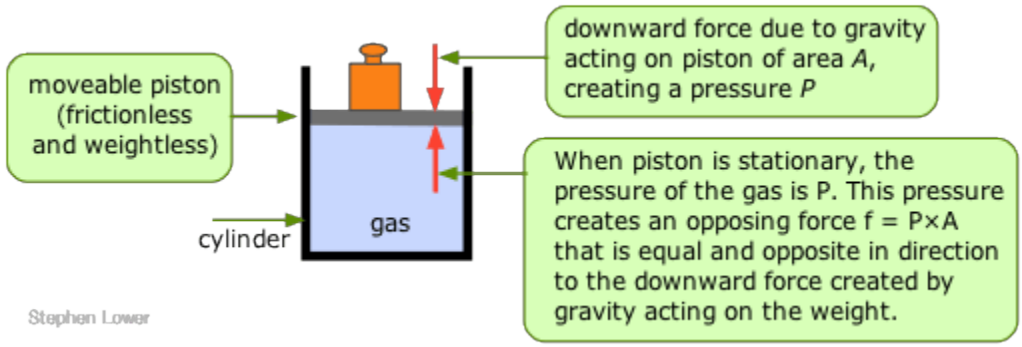


The most remarkable property of gases, however, is that to a very good approximation, **they all *behave the same way*** in response to changes in temperature and pressure, expanding or contracting by predictable amounts. This is very different from

the behavior of liquids or solids, in which the properties of each particular substance must be determined individually. We will see later that each of these three *macroscopic* characteristics of gases follows directly from the *microscopic* view— that is, from the *atomic* nature of matter.

The Pressure of a Gas

The molecules of a gas, being in continuous motion, frequently strike the inner walls of their container. As they do so, they immediately bounce off without loss of kinetic energy, but the reversal of direction (*acceleration*) imparts a *force* to the container walls. This force, divided by the total surface area on which it acts, is the *pressure* of the gas.



The pressure of a gas is observed by measuring the pressure that must be applied externally in order to keep the gas from expanding or contracting. To visualize this, imagine some gas trapped in a cylinder having one end enclosed by a freely moving piston. In order to keep the gas in the container, a certain amount of weight (more precisely, a *force, f*) must be placed on the piston so as to exactly balance the force exerted by the gas on the bottom of the piston, and tending to push it up. The pressure of the gas is simply the quotient *f/A*, where *A* is the cross-section area of the piston.

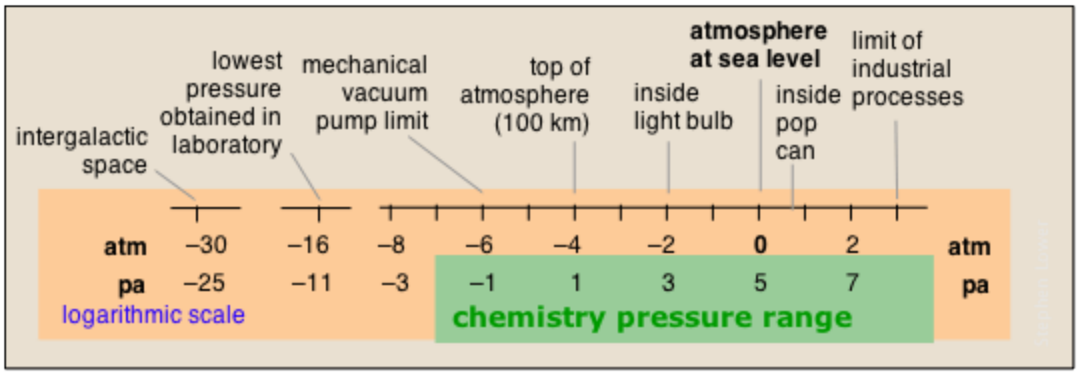
Pressure Units

The unit of pressure in the SI system is the *pascal* (Pa), defined as a force of one newton per square meter (1 Nm^{-2} = 1 kg m^{-1} s^{-2}.) At the Earth's surface, the force of gravity acting on a 1 kg mass is 9.81 N. Thus if the weight is 1 kg and the surface area of the piston is 1 M^2, the pressure of the gas would be 9.81 Pa. A 1-gram weight acting on a piston of 1 cm^2 cross-section would exert a pressure of 98.1 pA. (If you wonder why the pressure is higher in the second example, consider the number of cm^2 contained in 1 m^2.)

In chemistry, it is often common to express pressures in units of *atmospheres* or *torr*: 1 atm = 101325 Pa = 760 torr. The older unit *millimeter of mercury* (mm Hg) is almost the same as the torr and is defined as one mm of level difference in a mercury barometer at 0°C. In meteorology, the pressure unit most commonly used is the *bar*:

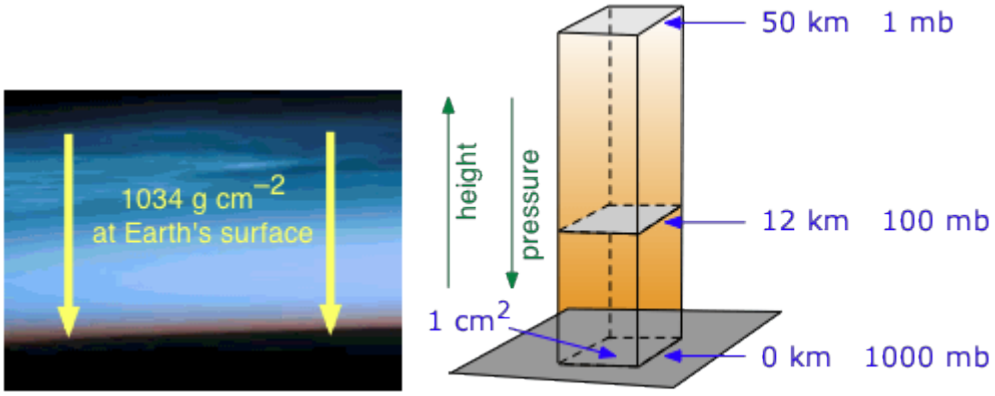
$$1 \text{ bar} = 10^5 \text{ N m}^{-2} = 750.06 \text{ torr} = 0.987 \text{ atm}.$$

The pressures of gases encountered in nature span an exceptionally wide range, only part of which is ordinarily encountered in chemistry. Note that in the chart below, the pressure scales are logarithmic; thus 0 on the atm scale means 10^0 = 1 atm.



Atmospheric Pressure

The column of air above us exerts a force on each 1-cm^2 of surface equivalent to a weight of about 1034 g. The higher into the air you go, the lower the mass of air above you, hence the lower the pressure (right).



This figure is obtained by solving Newton's law $\mathbf{F} = m\mathbf{a}$ for m, using the acceleration of gravity for $\mathbf{a}$:

$$m = \frac{101375\ kg\,m^{-1}\,s^{-2}}{9.8\,m\,s^{-2}} = 10340\,kg\,m^{-1} = 1034\ g\ cm^{-2} \tag{6.1.1}$$

Example 6.1.1

If several kilos of air are constantly pressing down on your body, why do you not feel it?

Solution

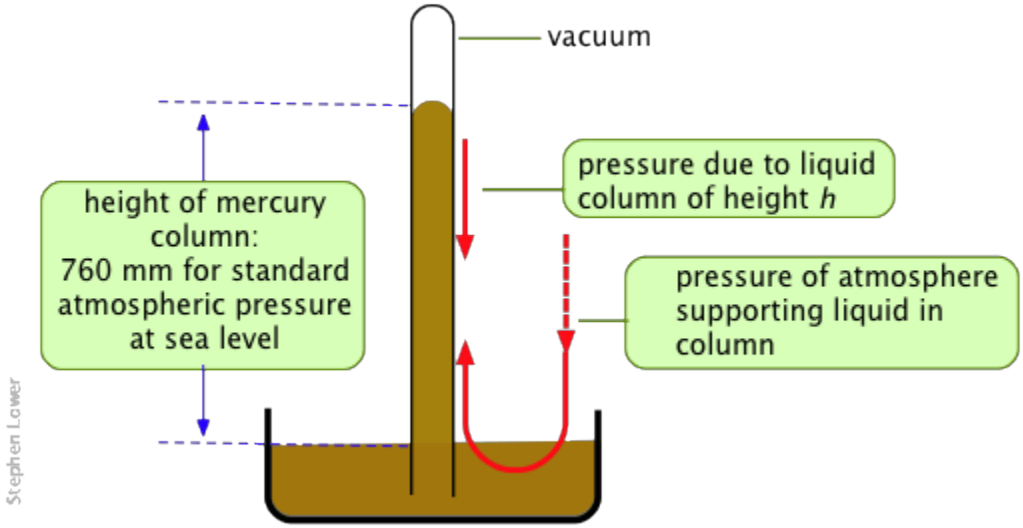
Because every other part of your body (including within your lungs and insides) also experiences the same pressure, so there is no net force (other than gravity) acting on you.

This was the crucial first step that led eventually to the concept of gases and their essential role in the early development of Chemistry. In the early 17th century the Italian Evangelista Torricelli invented a device — the *barometer* — to measure the pressure of the atmosphere. A few years later, the German scientist and some-time mayor of Magdeburg Otto von Guericke devised a method of pumping the air out of a container, thus creating which might be considered the opposite of air: the *vacuum*.

As with so many advances in science, idea of a vacuum — a region of nothingness — was not immediately accepted. Torricelli's invention overturned the then-common belief that air (and by extension, all gases) are weightless. The fact that we live at the bottom of a sea of air was most spectacularly demonstrated in 1654, when two teams of eight horses were unable to pull apart two 14-inch copper hemispheres (the "Magdeburg hemispheres") which had been joined together and then evacuated with Guericke's newly-invented vacuum pump.



The classical barometer, still used for the most accurate work, measures the height of a column of liquid that can be supported by the atmosphere. As indicated below, this pressure is exerted directly on the liquid in the reservoir, and is transmitted hydrostatically to the liquid in the column.



Metallic mercury, being a liquid of exceptionally high density and low vapor pressure, is the ideal barometric fluid. Its widespread use gave rise to the "millimeter of mercury" (now usually referred to as the "torr") as a measure of pressure.

Example 6.1.2: Water Barometer

How is the air pressure of 1034 g cm^{-3} related to the 760-mm height of the mercury column in the barometer? What if water were used in place of mercury?

Solution

The density of Hg is 13.6 g cm^{-3}, so in a column of 1-cm^2 cross-section, the height needed to counter the atmospheric pressure would be (1034 g × 1 cm^2) / (13.6 g cm^{-3}) = 76 cm.

The density of water is only 1/13.6 that of mercury, so standard atmospheric pressure would support a water column whose height is 13.6 x 76 cm = 1034 cm, or 10.3 m. You would have to read a water barometer from a fourth-story window!

Water barometers were once employed to measure the height of the ground and the heights of buildings before more modern methods were adopted.

A modification of the barometer, the *U-tube manometer*, provides a simple device for measuring the pressure of any gas in a container. The U-tube is partially filled with mercury, one end is connected to container, while the other end can either be opened to the atmosphere. The pressure inside the container is found from the difference in height between the mercury in the two sides of the U-tube. The illustration below shows how the two kinds of manometer work.

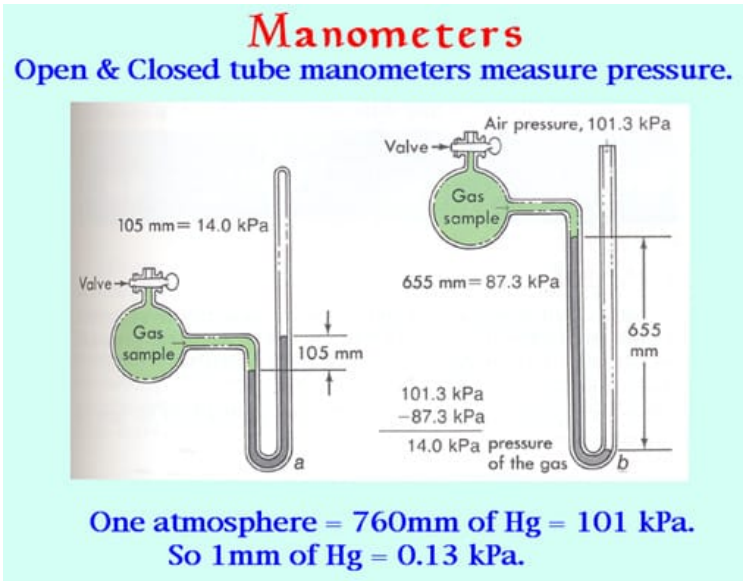


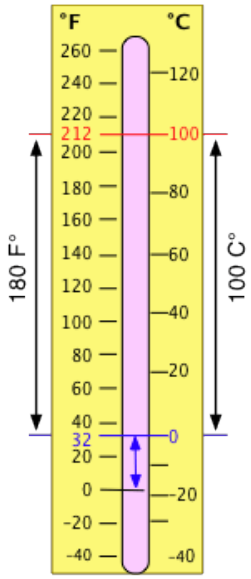
Figure 6.1.3: In the open-tube manometer on the right, the atmosphere pushes the mercury column down in the open side, and again the pressure in torr is found from the difference in the levels. When using this method, the atmospheric pressure must be determined by means of a barometer.

The manometers ordinarily seen in the laboratory come in two flavors: closed-tube and open-tube. In the closed-tube unit shown at the left, the longer limb of the J-tube is evacuated by filling it with mercury and then inverting it. If the sample container is also evacuated, the mercury level will be the same in both limbs. When gas is let into the container, its pressure pushes the mercury down on one side and up on the other; the difference in levels is the pressure in torr. For practical applications in engineering and industry, especially where higher pressures must be monitored, many types of mechanical and electrical *pressure gauges* are available.

The Temperature of a Gas

If two bodies are at different temperatures, heat will flow from the warmer to the cooler one until their temperatures are the same. This is the principle on which *thermometry* is based; the temperature of an object is measured indirectly by placing a calibrated device known as a *thermometer* in contact with it. When thermal equilibrium is obtained, the temperature of the thermometer is the same as the temperature of the object.

A *thermometer* makes use of some temperature-dependent quantity, such as the density of a liquid, to allow the temperature to be found indirectly through some easily measured quantity such as the length of a mercury column. The resulting scale of temperature is entirely arbitrary; it is defined by locating its zero point, and the size of the *degree* unit. At one point in the 18th century, 35 different temperature scales were in use! The Celsius temperature scale (formally called centigrade) locates the zero point at the freezing temperature of water; the Celsius degree is defined as 1/100 of the difference between the freezing and boiling temperatures of water at 1 atm pressure.



The older *Fahrenheit* scale placed the zero point at the coldest temperature it was possible to obtain at the time (by mixing salt and ice.) The 100° point was set with body temperature (later found to be 98.6°F.) On this scale, water freezes at 32°F and boils at 212°F. The Fahrenheit scale is a finer one than the Celsius scale; there are 180 Fahrenheit degrees in the same temperature interval that contains 100 Celsius degrees, so 1 F° = 5/9 C°. Since the zero points are also different by 32F°, conversion between temperatures expressed on the two scales requires the addition or subtraction of this offset, as well as multiplication by the ratio of the degree size.

You should be able to derive the formulas for this conversion.

Absolute Temperature

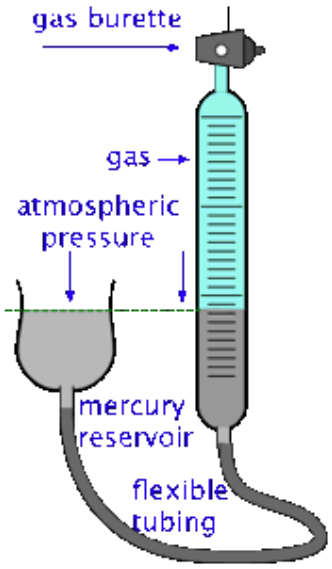
In 1787 the French mathematician and physicist Jacques Charles discovered that for each Celsius degree that the temperature of a gas is lowered, the volume of the gas will diminish by 1/273 of its volume at 0°C. The obvious implication of this is that if the temperature could be reduced to –273°C, the volume of the gas would contract to zero. Of course, all real gases condense to liquids before this happens, but at sufficiently low pressures their volumes are linear functions of the temperature (*Charles' Law*), and extrapolation of a plot of volume as a function of temperature predicts zero volume at -273°C. This temperature, known as *absolute zero*, corresponds to the total absence of thermal energy.

The temperature scale on which the zero point is –273.15°C was suggested by Lord Kelvin, and is usually known as the *Kelvin* scale. Since the size of the Kelvin and Celsius degrees are the same, conversion between the two scales is a simple matter of adding or subtracting 273.15; thus room temperature, 20°, is about 293 K.

Because the Kelvin scale is based on an absolute, rather than on an arbitrary zero of temperature, it plays a special significance in scientific calculations; most fundamental physical relations involving temperature are expressed mathematically in terms of absolute temperature. In engineering work, an absolute scale based on the Fahrenheit degree is sometimes used; this is known as the **Rankine scale**.

The volume occupied by a gas

The *volume* of a gas is simply the space in which the molecules of the gas are free to move. If we have a mixture of gases, such as air, the various gases will coexist within the same volume. In these respects, gases are very different from liquids and solids, the two condensed states of matter. The volume of a gas can be measured by trapping it above mercury in a calibrated tube known as a *gas burette*. The SI unit of volume is the cubic meter, but in chemistry we more commonly use the *liter* and the *milliliter* (mL). The *cubic centimeter* (cc) is also frequently used; it is very close to 1 milliliter (mL).



It's important to bear in mind, however, that the volume of a gas varies with both the temperature and the pressure, so reporting the volume alone is not very useful. A common practice is to measure the volume of the gas under the ambient temperature and atmospheric pressure, and then to correct the observed volume to what it would be at standard atmospheric pressure and some fixed temperature, usually 0° C or 25°C.

Contributors and Attributions

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

This page titled 6.1: Observable Properties of Gas is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

- Properties of Gas by Stephen Lower is licensed CC BY-NC-SA 4.0. Original source: http://www.chem1.com/acad/webtext/virtualtextbook.html.

6.2: Ideal Gas Model - The Basic Gas Laws

Learning Objectives

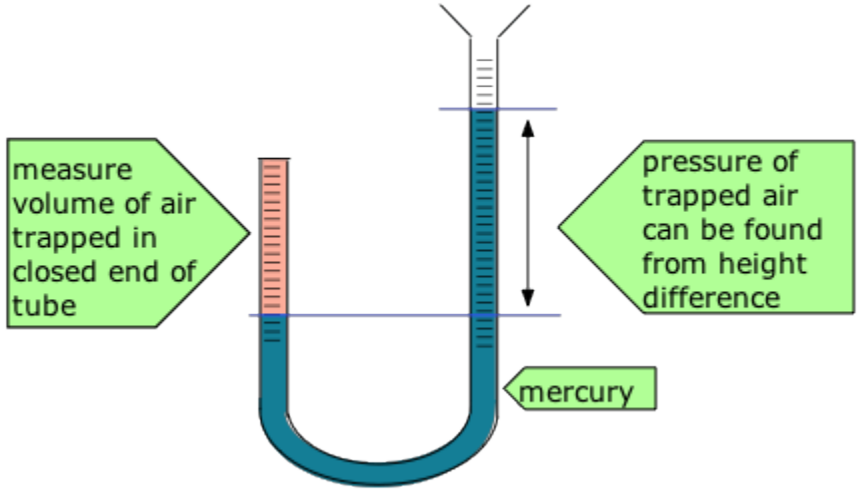
Make sure you thoroughly understand the following essential ideas which have been presented above, and be able to state them in your own words.

- **Boyle's Law** - The *PV* product for any gas at a fixed temperature has a constant value. Understand how this implies an *inverse* relationship between the pressure and the volume.
- ***Charles' Law*** - The volume of a gas confined by a fixed pressure varies *directly* with the absolute temperature. The same is true of the pressure of a gas confined to a fixed volume.
- ***Avogadro's Law*** - This is quite intuitive: the volume of a gas confined by a fixed pressure varies directly with the quantity of gas.
- The E.V.E.N. principle - this is just another way of expressing Avogadro's Law.
- Gay-Lussac's Law of Combining Volumes - you should be able to explain how this principle, that follows from the E.V.E.N. principle and the Law of Combining Weights,
- ***The ideal gas equation of state*** - this is one of the very few mathematical relations you *must* know. Not only does it define the properties of the hypothetical substance known as an *ideal gas*, but it's importance extends quite beyond the subject of gases.

The "pneumatic" era of chemistry began with the discovery of the *vacuum* around 1650 which clearly established that gases are a form of matter. The ease with which gases could be studied soon led to the discovery of numerous empirical (experimentally-discovered) laws that proved fundamental to the later development of chemistry and led indirectly to the atomic view of matter. These laws are so fundamental to all of natural science and engineering that everyone learning these subjects needs to be familiar with them.

Pressure-volume relations: Boyle's law

Robert Boyle (1627-91) showed that the volume of air trapped by a liquid in the closed short limb of a J-shaped tube decreased in exact proportion to the pressure produced by the liquid in the long part of the tube. The trapped air acted much like a spring, exerting a force opposing its compression. Boyle called this effect "*the spring of the air*", and published his results in a pamphlet of that title.



The difference between the heights of the two mercury columns gives the pressure (76 cm = 1 atm), and the volume of the air is calculated from the length of the air column and the tubing diameter. Some of Boyle's actual data are shown in Table 6.2.1.

Table 6.2.1: Volume vs. Pressure

volume	pressure	*P* × *V*
96.0	2.00	192
76.0	2.54	193
46.0	4.20	193
26.0	7.40	193

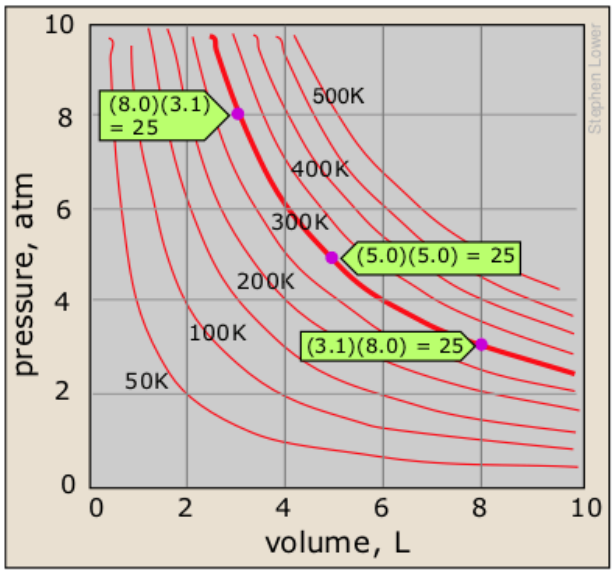
Boyle's law can be expressed as

$$PV = \text{constant}$$

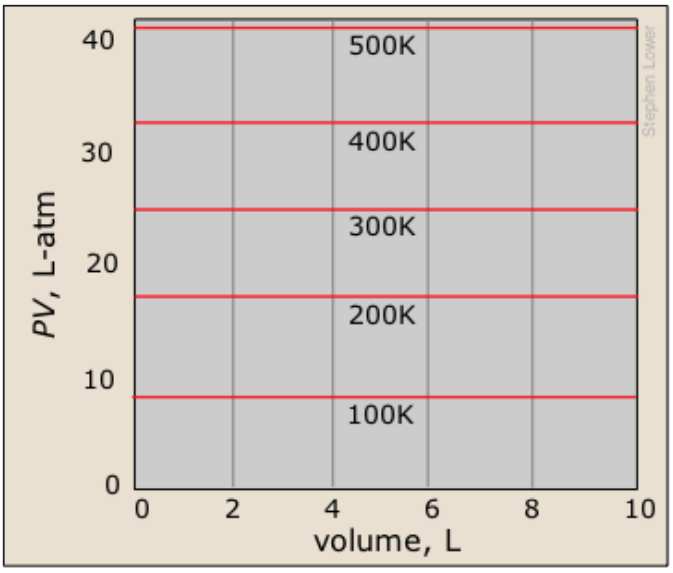
or, equivalently,

$$P_1V_1 = P_2V_2 \tag{6.2.1}$$

These relations hold true **only** if the number of molecules *n* and the temperature are constant. This is a relation of *inverse proportionality*; any change in the pressure is exactly compensated by an opposing change in the volume. As the pressure decreases toward zero, the volume will increase without limit. Conversely, as the pressure is increased, the volume decreases, but can never reach zero. There will be a separate *P*-*V* plot for each temperature; a single *P*-*V* plot is therefore called an *isotherm*.



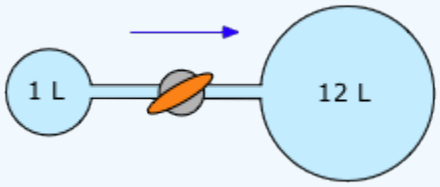
Shown here are some isotherms for one mole of an ideal gas at several different temperatures. Each plot has the shape of a hyperbola — the locus of all points having the property $x\,y = a$, where a is a constant. You will see later how the value of this constant (*PV*=25 for the 300K isotherm shown here) is determined. It is very important that you understand this kind of plot which governs any relationship of inverse proportionality. You should be able to sketch out such a plot when given the value of any one (*x*,*y*) pair.



A related type of plot with which you should be familiar shows the product *PV* as a function of the pressure. You should understand why this yields a straight line, and how this set of plots relates to the one immediately above.

Example 6.2.1

In an industrial process, a gas confined to a volume of 1 L at a pressure of 20 atm is allowed to flow into a 12-L container by opening the valve that connects the two containers. What will be the final pressure of the gas?



Solution

The final volume of the gas is (1 + 12)L = 13 L. The gas expands in inverse proportion two volumes

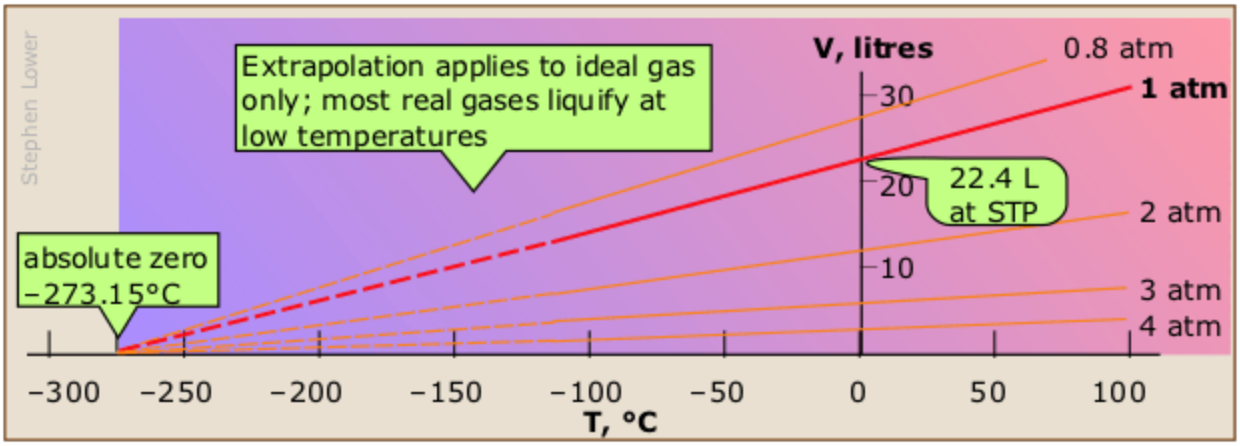
$$P_2 = (20\,atm)(1\,L \div 13\,L) = 1.5\,atm$$

Note that there is no need to make explicit use of any "formula" in problems of this kind!

How the temperature affects the volume: Charles' law

All matter expands when heated, but gases are special in that their degree of expansion is independent of their composition. The French scientists Jacques Charles (1746-1823) and Joseph Gay-Lussac (1778-1850) independently found that if the pressure is held constant, the volume of any gas changes by the same fractional amount (1/273 of its value) for each C° change in temperature.

The volume of a gas confined against a constant pressure is directly proportional to the absolute temperature. A graphical expression of the law of Charles and Gay-Lussac can be seen in these plots of the volume of one mole of an ideal gas as a function of its temperature at various constant pressures.



What do these plots show?

The straight-line plots show that the ratio V/T (and thus *dV*/*dT*) is a constant at any given pressure. Thus we can express the law algebraically as V/T = constant or $V_1/T_1 = V_2/T_2$

What is the significance of the extrapolation to zero volume?

If a gas contracts by 1/273 of its volume for each degree of cooling, it should contract to zero volume at a temperature of –273°C. This, of course, is the temperature of absolute zero, and this extrapolation of Charles' law is the first evidence of the special significance of this temperature.

Why do the plots for different pressures have different slopes?

The lower the pressure, the greater the volume (Boyle's law), so at low pressures the fraction (V/273) will have a larger value. You might say that the gas must "contract faster" to reach zero volume when its starting volume is larger.

Example 6.2.2

The air pressure in a car tire is 30 psi (pounds per square inch) at 10°C. What will be pressure be after driving has raised its temperature to 45°C ? (Assume that the volume remains unchanged.)

Solution

The gas expands in direct proportion to the ratio of the absolute temperatures:

$$P_2 = (30\,psi) \times (318\,K \div 283\,K) = 33.7\,psi$$

The relation between the temperature of a gas and its volume has long been known. In 1702, Guillaume Amontons (1163-1705), who is better known for his early studies of friction, devised a thermometer that related the temperature to the volume of a gas. Robert Boyle had observed this inverse relationship in 1662, but the lack of any uniform temperature scale at the time prevented them from establishing the relationship as we presently understand it. Jacques Charles discovered the law that is named for him in the 1780s, but did not publish his work. John Dalton published a form of the law in 1801, but the first thorough published presentation was made by Gay-Lussac in 1802, who acknowledged Charles' earlier studies.



The buoyancy that lifts a hot-air balloon into the sky depends on the difference between the density (mass ÷ volume) of the air entrapped within the balloon's envelope, compared to that of the air surrounding it. When a balloon on the ground is being prepared for flight, it is first partially inflated by an external fan, and possesses no buoyancy at all. Once the propane burners are started, this air begins to expand according to Charles' law. After the warmed air has completely inflated the balloon, further expansion simply forces excess air out of the balloon, leaving the weight of the diminished mass of air inside the envelope smaller than that of the greater mass of cooler air that the balloon displaces.

Jacques Charles collaborated with the Montgolfier brothers whose hot-air balloon made the world's first manned balloon flight in June, 1783. Ten days later, Charles himself co-piloted the first hydrogen-filled balloon. Gay-Lussac, who had a special interest in the composition of the atmosphere, also saw the potential of the hot-air balloon, and in 1804 he ascended to a then-record height of 6.4 km.

Volume and the Number of Molecules

Gay-Lussac's Law of Combining Volumes

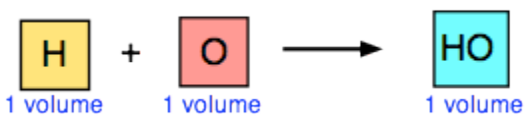
In the same 1808 article in which Gay-Lussac published his observations on the thermal expansion of gases, he pointed out that when two gases react, they do so in volume ratios that can always be expressed as small whole numbers. This came to be known as the *Law of combining volumes*. These "small whole numbers" are of course the same ones that describe the "combining weights" of elements to form simple compounds, as described in the lesson dealing with the simplest formulas. The Italian scientist Amedeo Avogadro (1776-1856) drew the crucial conclusion: these volume ratios must be related to the relative numbers of molecules that react, and thus the famous "E.V.E.N principle":

Equal **v**olumes of gases, measured at the same temperature and pressure, contain **e**qual **n**umbers of molecules

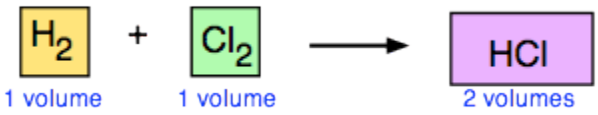
Avogadro's law thus predicts a *directly proportional* relation between the number of moles of a gas and its volume. This relationship, originally known as *Avogadro's Hypothesis*, was crucial in establishing the formulas of simple molecules at a time (around 1811) when the distinction between atoms and molecules was not clearly understood. In particular, the existence of diatomic molecules of elements such as H_2, O_2, and Cl_2 was not recognized until the results of combining-volume experiments such as those depicted below could be interpreted in terms of the E.V.E.N. principle.

How the E.V.E.N. principle led to the correct formula of water

Early chemists made the mistake of assuming that the formula of water is HO. This led them to miscalculate the molecular weight of oxygen as 8 (instead of 16). If this were true, the reaction H + O → HO would correspond to the following combining volumes results according to the E.V.E.N principle:



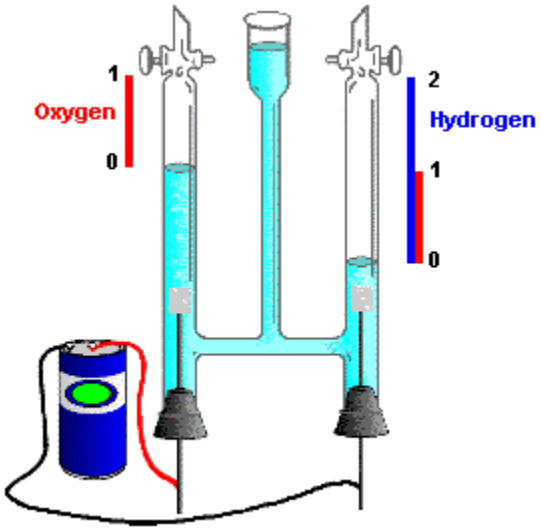
But a similar experiment on the formation of hydrogen chloride from hydrogen and chlorine yielded twice the volume of HCl that was predicted by the the assumed reaction H + Cl → HCl. This could be explained only if hydrogen and chlorine were diatomic molecules:



This made it necessary to re-visit the question of the formula of water. The experiment immediately confirmed that the correct formula of water is H_2O:



This conclusion was also seen to be consistent with the observation, made a few years earlier by the English chemists Nicholson and Carlisle that the reverse of the above reaction, brought about by the electrolytic decomposition of water, yields hydrogen and oxygen in a 2:1 volume ratio.



Putting it all together: The Ideal Gas Equation of State

If the variables *P, V, T* and *n* (the number of moles) have known values, then a gas is said to be in a definite *state*, meaning that all other physical properties of the gas are also defined. The relation between these *state variables* is known as an *equation of state*. By combining the expressions of Boyle's, Charles', and Avogadro's laws (you should be able to do this!) we can write the very important *ideal gas* equation of state

$$PV = NRT \tag{6.2.2}$$

where the proportionality constant *R* is known as the *gas constant*. This is one of the few equations you *must* commit to memory in this course; you should also know the common value and units of R.

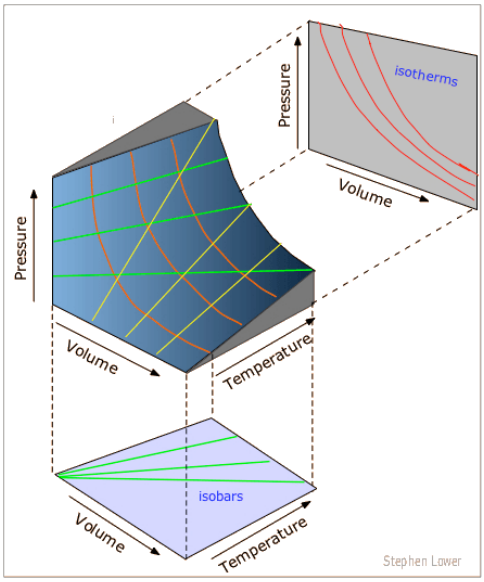
An *ideal gas* is defined as a hypothetical substance that obeys the ideal gas equation of state.

Take note of the word "hypothetical" here. No real gas (whose molecules occupy space and interact with each other) can behave in a truly ideal manner. But we will all *gases* behave more and more like an ideal gas as the pressure approaches zero. A pressure of only 1 atm is sufficiently close to zero to make this relation useful for most gases at this pressure.

Many textbooks show formulas, such as $P_1V_1 = P_2V_2$ for Boyle's law. *Don't bother memorizing them*; if you really understand the meanings of these laws as stated above, you can easily derive them on the rare occasions when they are needed. ***The ideal gas equation is the only one you need to know.***

PVT surface for an ideal gas

In order to depict the relations between the three variables *P, V* and *T* we need a three-dimensional graph.



Each point on the curved surface represents a possible combination of (P,V,T) for an arbitrary quantity of an ideal gas. The three sets of lines inscribed on the surface correspond to states in which one of these three variables is held constant. The red curved lines, being lines of constant temperature, or isotherms, are plots of Boyle's law. These isotherms are also seen projected onto the P-V plane at the top right. The yellow lines are isochors and represent changes of the pressure with temperature at constant volume. The green lines, known as isobars, and projected onto the V-T plane at the bottom, show how the volumes contract to zero as the absolute temperature approaches zero, in accordance with the law of Charles and Gay-Lussac.

Example 6.2.3

A biscuit made with baking powder has a volume of 20 mL, of which one-fourth consists of empty space created by gas bubbles produced when the baking powder decomposed to CO_2. What weight of $\mathrm{NaHCO_3}$ was present in the baking powder in the biscuit? Assume that the gas reached its final volume during the baking process when the temperature was 400°C.

(Baking powder consists of sodium bicarbonate mixed with some other solid that produces an acidic solution on addition of water, initiating the reaction

$$\mathrm{NaHCO_3\,(s) + H^+ \longrightarrow Na^+ + H_2O + CO_2}$$

Solution: Use the ideal gas equation to find the number of moles of CO_2 gas; this will be the same as the number of moles of $NaHCO_3$ (84 g mol^{-1}) consumed :

$$n = \frac{(1\mathrm{atm})(0.005\mathrm{L})}{\left(.082\mathrm{Latmmol^{-1}K^{-1}}\right)(673\mathrm{K})} = 9.1 \times 10^{-6}\,\mathrm{mol}$$

$$9.1\mathrm{E}-6\mathrm{mol} \times 84\mathrm{gmol^{-1}} = 0.0076\mathrm{g}$$

This page titled 6.2: Ideal Gas Model - The Basic Gas Laws is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

6.3: Dalton's Law

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented below.

- One mole of a gas occupies a volume of 22.4 L at *STP* (standard temperature and pressure, 273K, 1 atm = 101.3 kPa.)
- The above fact allows us to relate the measurable property of the **density of a gas** to its **molar mass**.
- The composition of a **mixture of gases** is commonly expressed in terms *mole fractions*; be sure you know how to calculate them.
- *Dalton*'s *Law of partial pressures* says that every gas in a mixture acts independently, so the total pressure a gas exerts against the walls of a container is just the sum of the partial pressures of the individual components.

Although all gases closely follow the ideal gas law *PV* = *nRT* under appropriate conditions, each gas is also a unique chemical substance consisting of molecular units that have definite masses. In this lesson we will see how these molecular masses affect the properties of gases that conform to the ideal gas law. Following this, we will look at gases that contain more than one kind of molecule— in other words, *mixtures* of gases. We begin with a review of molar volume and the E.V.E.N. principle, which is central to our understanding of gas mixtures.

The Molar Volume of a Gas

You will recall that the *molar mass* of a pure substance is the mass of 6.02 x 10^{23} (*Avogadro*'s *number*) of particles or molecular units of that substance. Molar masses are commonly expressed in units of grams per mole (g mol^{-1}) and are often referred to as *molecular weights*. As was explained in the preceding lesson, equal volumes of gases, measured at the same temperature and pressure, contain equal numbers of molecules (this is the "*EVEN*" *principle*, more formally known as *Avogadro's law*.) Standard temperature and pressure: 273K, 1 atm

The magnitude of this volume will of course depend on the temperature and pressure, so as a means of convenient comparison it is customary to define a set of conditions T = 273 K and P = 1 atm as *standard temperature and pressure*, usually denoted as *STP*. Substituting these values into the ideal gas equation of state and solving for *V* yields a volume of 22.414 liters for 1 mole.

Example 6.3.1

What would the volume of one mole of air be at 20°C on top of Mauna Kea, Hawa'ii (altitude 4.2 km) where the air pressure is approximately 60 kPa?



Scoria and cinder cones on Mauna Kea's summit in winter. (Public Domain; USGS)

Solution

Apply Boyle's and Charles' laws as successive correction factors to the standard sea-level pressure of 101.3 kPa:

$$V_2 = 22.4\ \text{L} \times \left(\frac{293\text{K}}{273\text{K}}\right) \times \left(\frac{101.3\ \text{kPa}}{60\ \text{kPa}}\right) = 41\ \text{L}$$

(V_M at STP; Charles' law; Boyle's law)

The *standard molar volume 22.4 L mol*$^{-1}$ is a value worth memorizing, but remember that it is valid only at STP. The molar volume at other temperatures and pressures can easily be found by simple proportion. The molar volume of a substance can tell us something about how much space each molecule occupies, as the following example shows.

Example 6.3.2

Estimate the average distance between the molecules in a gas at 1 atm pressure and 0°C.

Solution

Consider a 1-cm^3 volume of the gas, which will contain

$$\frac{6.02 \times 10^{23}\ mol^{-1}}{22,400\ cm^3\ mol^{-1}} = 2.69 \times 10^{19} cm^{-3}$$

The volume per molecule (not the same as the volume *of* a molecule, which for an ideal gas is zero!) is just the reciprocal of this, or $3.72 \times 10^{-20}\ cm^3$. Assume that the molecules are evenly distributed so that each occupies an imaginary box having this volume. The average distance between the centers of the molecules will be defined by the length of this box, which is the cube root of the volume per molecule:

$$(3.72 \times 10^{-20})^{1/3} = 3.38 \times 10^{-7}\, cm = 3.4\, nm$$

Under conditions at which the ideal gas model is applicable (that is, almost always unless you are a chemical engineer dealing with high pressures), "a molecule is a molecule", so the volume of Avogadro's number of molecules will be independent of the

composition of the gas. The reason, of course, is that the volume of the gas is mostly empty space; the volumes of the molecules themselves are negligible.

Molar Mass and Density of a Gas

The molecular weight (molar mass) of any gas is the mass, expressed in grams, of Avogadro's number of its molecules. This is true regardless of whether the gas is composed of one molecular species or is a mixture. For a mixture of gases, the molar mass will depend on the molar masses of its components, and on the fractional abundance of each kind of molecule in the mixture. The term "average molecular weight" is often used to describe the molar mass of a gas mixture.

The average molar mass ($\bar{m}$) of a mixture of gases is just the sum of the mole fractions of each gas, multiplied by the molar mass of that substance:

$$\bar{m} = \sum_i \chi_i m_i \tag{6.3.1}$$

Example 6.3.3

Find the average molar mass of dry air whose volume-composition is O_2 (21%), N_2 (78%) and Ar (1%).

Solution

The average molecular weight is the mole-fraction-weighted sum of the molecular weights of its components. The mole fractions, of course, are the same as the volume-fractions (E.V.E.N. principle.)

$$m = (0.21 \times 32) + (0.78 \times 28) + (0.01 \times 20) = 28$$

The molar volumes of all gases are the same when measured at the same temperature and pressure. However, the molar *masses* of different gases will vary. This means that different gases will have different *densities* (different masses per unit volume). If we know the molecular weight of a gas, we can calculate its density.

Example 6.3.4

Uranium hexafluoride UF_6 gas is used in the isotopic enrichment of natural uranium. Calculate its density at STP.

Solution

The molecular weight of UF_6 is 352.

$$\frac{352\ g\ mol^{-1}}{22.4\ L\ mol^{-1}} = 15.7\ g\ L^{-1}$$

Note: there is no need to look up a "formula" for this calculation; simply combine the molar mass and molar volume in such a way as to make the units come out correctly.

More importantly, if we can measure the density of an unknown gas, we have a convenient means of estimating its molecular weight. This is one of many important examples of how a *macroscopic* measurement (one made on bulk matter) can yield *microscopic* information (that is, about molecular-scale objects.)

Gas densities are now measured in industry by electro-mechanical devices such as vibrating reeds which can provide continuous, on-line records at specific locations, as within pipelines. Determination of the molecular weight of a gas from its density is known as the *Dumas method*, after the French chemist Jean Dumas (1800-1840) who developed it. One simply measures the weight of a known volume of gas and converts this volume to its STP equivalent, using Boyle's and Charles' laws. The weight of the gas divided by its STP volume yields the density of the gas, and the density multiplied by 22.4 mol^{-1} gives the molecular weight. Pay careful attention to the examples of gas density calculations shown here and in your textbook. You will be expected to carry out calculations of this kind, converting between molecular weight and gas density.

Example 6.3.5

Calculate the approximate molar mass of a gas whose measured density is 3.33 g/L at 30°C and 780 torr.

Solution

Find the volume that would be occupied by 1 L of the gas at STP; note that correcting to 273 K will reduce the volume, while correcting to 1 atm (760 torr) will increase it:

$$V = (1.00\text{L})\left(\frac{273}{303}\right)\left(\frac{780}{760}\right) = 0.924\text{L}$$

The number of moles of gas is

$$n = \frac{0.924\,L}{22.4\,L\,mol^{-1}} = 0.0412\,mol$$

The molecular weight is therefore

$$\frac{33\,g\,L^{-1}}{0.0412\,mol\,L^{-1}} = 80.7\,g\,mol^{-1}$$

Density of a Gas Mixture

Gas density measurements can be a useful means of estimating the composition of a mixture of two different gases; this is widely done in industrial chemistry operations in which the compositions of gas streams must be monitored continuously.

Example 6.3.6: Carbon Dioxide and Methane

Find the composition of a mixture of CO_2 (44 g/mol) and methane CH_4 (16 g/mol) that has a STP density of 1.214 g/L.

Solution

The density of a mixture of these two gases will be directly proportional to its composition, varying between that of pure methane and pure CO_2. We begin by finding these two densities:

For CO_2:

(44 g/mol) ÷ (22.4 L/mol) = 1.964 g/L

For CH_4:

(16 g/mol) ÷ (22.4 L/mol) = 0.714 g/L

If *x* is the mole fraction of CO_2 and (1–*x*) is the mole fraction of CH_4, we can write

1.964 x + 0.714 (1–x) = 1.214

(Does this make sense? Notice that if x = 0, the density would be that of pure CH_4, while if it were 1, it would be that of pure CO_2.)

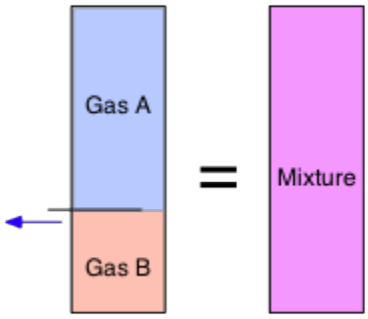
Expanding the above equation and solving for x yields the mole fractions of 0.40 for CO_2 and 0.60 for CH_4.

Expressing the Composition of a Gas Mixture

Because most of the volume occupied by a gas consists of empty space, there is nothing to prevent two or more kinds of gases from occupying the same volume. *Homogeneous mixtures* of this kind are generally known as *solutions*, but it is customary to refer to them simply as *gaseous mixtures*. We can specify the composition of gaseous mixtures in many different ways, but the most common ones are by *volumes* and by *mole fractions*.

Volume Fractions

From Avogadro's Law we know that "equal volumes contains equal numbers of molecules". This means that the volumes of gases, unlike those of solids and liquids, are additive. So if a partitioned container has two volumes of gas A in one section and one mole of gas B in the other (both at the same temperature and pressure), and we remove the partition, the volume remains unchanged.



We can specify the composition of this mixture by saying that the *volume fraction* of A is 1/3 and that of B is 2/3.

Volume fractions are often called partial volumes:

$$V_i = \frac{v_i}{\sum v_i} \tag{6.3.2}$$

Don't let this type of notation put you off! The summation sign Σ (Greek Sigma) simply means to add up the *v*'s (volumes) of every gas. Thus if Gas A is the "*i*-th" substance as in the expression immediately above, the summation runs from *i*=1 through *i*=2. Note that we can employ partial volumes to specify the composition of a mixture even if it had never actually been made by combining the pure gases.

When we say that air, for example, is 21 percent oxygen and 78 percent nitrogen by volume, this is the same as saying that these same percentages of the molecules in air consist of O_2 and N_2. Similarly, in 1.0 mole of air, there is 0.21 mol of O_2 and 0.78 mol of N_2 (the other 0.1 mole consists of various trace gases, but is mostly neon.) Note that you could never assume a similar equivalence with mixtures of liquids or solids, to which the E.V.E.N. principle does not apply.

Mole Fractions

These last two numbers (0.21 and 0.78) also express the mole fractions of oxygen and nitrogen in air. *Mole fraction* means exactly what it says: the fraction of the molecules that consist of a specific substance. This is expressed algebraically by

$$X_i = \frac{n_i}{\sum_i n_i} \tag{6.3.3}$$

so in the case of oxygen in the air, its mole fraction is

$$X_{O_2} = \frac{n_{O_2}}{n_{O_2} + n_{N_2} + n_{Ar}} = \frac{0.21}{1} = 0.21$$

Example 6.3.7

A mixture of O_2 and nitrous oxide, N_2O, is sometimes used as a mild anesthetic in dental surgery. A certain mixture of these gases has a density of 1.482 g L^{-1} at 25 and 0.980 atm. What was the mole-percent of N_2O in this mixture?

Solution

First, find the density the gas would have at STP:

$$\left(1.482 \mathrm{gL}^{-1}\right) \times \left(\frac{298}{273}\right)\left(\frac{1}{0.980}\right) = 1.65 \mathrm{gL}^{-1}$$

The molar mass of the mixture is (1.65 g L^{-1})(22.4 L mol^{-1}) = 37.0 g mol^{-1}. The molecular weights of O_2 and N_2 are 32 and 44, respectively. 37.0 is 5/12 of the difference between the molar masses of the two pure gases. Since the density of a gas mixture is directly proportional to its average molar mass, the mole fraction of the heavier gas in the mixture is also 5/12:

$$\frac{37-32}{44-32} = \frac{5}{12} = 0.42$$

Example 6.3.8

What is the mole fraction of carbon dioxide in a mixture consisting of equal masses of CO_2 (MW=44) and neon (MW=20.2)?

Solution

Assume any arbitrary mass, such as 100 g, find the equivalent numbers of moles of each gas, and then substitute into the definition of mole fraction:

- n_{CO2} = (100 g) ÷ (44 g mol^{-1}) = 2.3 mol
- n_{Ne}= (100 g) ÷ (20.2 g mol^{-1}) = 4.9 mol
- X_{Ne} = (2.3 mol) ÷ (2.3 mol + 4.9 mol) = 0.32

Dalton's Law of Partial Pressures

The ideal gas equation of state applies to mixtures just as to pure gases. It was in fact with a gas mixture, ordinary air, that Boyle, Gay-Lussac and Charles did their early experiments. The only new concept we need in order to deal with gas mixtures is the *partial pressure*, a concept invented by the famous English chemist John Dalton (1766-1844). Dalton reasoned that the low density and high compressibility of gases indicates that they consist mostly of empty space; from this it follows that when two or more different gases occupy the same volume, they behave entirely independently.

The contribution that each component of a gaseous mixture makes to the total pressure of the gas is known as the *partial pressure* of that gas. Dalton himself stated this law in the simple and vivid way shown at the left.

The usual way of stating *Dalton's Law of Partial Pressures* is

The total pressure of a gas is the sum of the partial pressures of its components

which is expressed algebraically as

$$P_{total} = P_1 + P_2 + P_3 ... = \sum_i P_i \tag{6.3.4}$$

or, equivalently

$$P_{total} = \frac{RT}{V}\sum_i n_i \tag{6.3.5}$$

There is also a similar relationship based on *volume fractions*, known as *Amagat's law of partial volumes*. It is exactly analogous to Dalton's law, in that it states that the total volume of a mixture is just the sum of the partial volumes of its components. But there are two important differences: Amagat's law holds only for ideal gases which must all be at the same temperature and pressure. Dalton's law has neither of these restrictions. Although Amagat's law seems intuitively obvious, it sometimes proves useful in chemical engineering applications. We will make no use of it in this course.

Example 6.3.9

Calculate the mass of each component present in a mixture of fluorine (MW and xenon (MW 131.3) contained in a 2.0-L flask. The partial pressure of Xe is 350 torr and the total pressure is 724 torr at 25°C.

Solution

From Dalton's law, the partial pressure of F_2 is (724 – 350) = 374 torr:

The mole fractions are

$$\chi_{Xe} = \frac{350}{724} = 0.48$$

and

$$\chi_{F_2} = \frac{374}{724} = 0.52$$

The total number of moles of gas is

$$n = \frac{PV}{RT} = \frac{(724/60)(2)}{(.082)(298)} = 0.078\text{mol}$$

The mass of Xe is

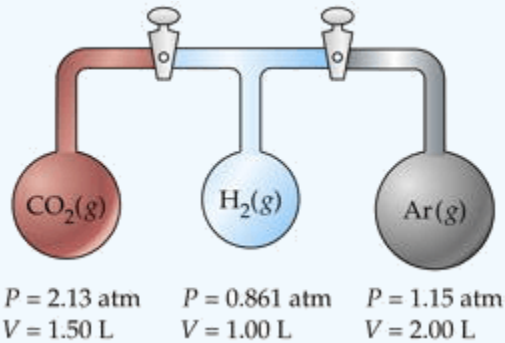
$$(131.3\,g\,mol^{-1}) \times (0.48 \times 0.078\,mol) = 4.9\,g$$

Example 6.3.10

Three flasks having different volumes and containing different gases at various pressures are connected by stopcocks as shown. When the stopcocks are opened,

a. What will be the pressure in the system?
b. Which gas will be most abundant in the mixture?

Assume that the temperature is uniform and that the volume of the connecting tubes is negligible.



Solution

The trick here is to note that the total number of moles n_T and the temperature remain unchanged, so we can make use of Boyle's law PV = constant. We will work out the details for CO_2 only, denoted by subscripts a.

For CO_2,

$$P_aV_a = (2.13\ atm)(1.50\ L) = 3.19\ L\text{-}atm.$$

Adding the PV products for each separate container, we obtain

$$\sum P_iV_i = 6.36L - atm = n_TRT.$$

We will call this sum P_1V_1. After the stopcocks have been opened and the gases mix, the new conditions are denoted by P_2V_2.

From Boyle's law,

= 6.36 L-atm. $V_2 = \Sigma V_i$ = 4.50 L.

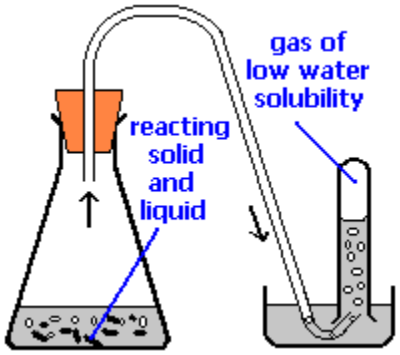
Solving for the final pressure P_2 we obtain

$$(6.36\ L\text{-}atm)/(4.50\ L) = \mathbf{1.41\ atm}.$$

For CO_2, this works out to (3.19/RT) / (6.36/RT) = 0.501. Because this exceeds 0.5, we know that this is the most abundant gas in the final mixture.

Application of Dalton's Law: Collecting Gases over Water

A common laboratory method of collecting the gaseous product of a chemical reaction is to conduct it into an inverted tube or bottle filled with water, the opening of which is immersed in a larger container of water. This arrangement is called a *pneumatic trough*, and was widely used in the early days of chemistry. As the gas enters the bottle it displaces the water and becomes trapped in the upper part.



The volume of the gas can be observed by means of a calibrated scale on the bottle, but what about its pressure? The total pressure confining the gas is just that of the atmosphere transmitting its force through the water. (An exact calculation would also have to take into account the height of the water column in the inverted tube.) But liquid water itself is always in equilibrium with its vapor, so the space in the top of the tube is a mixture of two gases: the gas being collected, and gaseous H_2O. The partial pressure of H_2O is known as the vapor pressure of water and it depends on the temperature. In order to determine the quantity of gas we have collected, we must use Dalton's Law to find the partial pressure of that gas.

Example 6.3.11

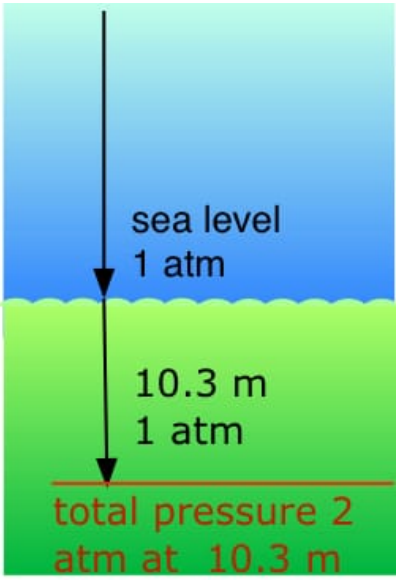
Oxygen gas was collected over water as shown above. The atmospheric pressure was 754 torr, the temperature was 22°C, and the volume of the gas was 155 mL. The vapor pressure of water at 22°C is 19.8 torr. Use this information to estimate the number of moles of O_2 produced.

Solution

From Dalton's law, $P_{O_2} = P_{total} – P_{H_2O} = 754 – 19.8 = 734 \ \ torr = 0.966 \ \ atm.$

$$n = \frac{PV}{RT} = \frac{0.966\text{atm} \times (0.155\text{L})}{\left(.082\text{Latmmol}^{-1}\text{K}^{-1}\right)(295\text{K})} = .00619\text{mol}$$

Our respiratory systems are designed to maintain the proper oxygen concentration in the blood when the partial pressure of O_2 is 0.21 atm, its normal sea-level value. Below the water surface, the pressure increases by 1 atm for each 10.3 m increase in depth; thus a scuba diver at 10.3 m experiences a total of 2 atm pressure pressing on the body. In order to prevent the lungs from collapsing, the air the diver breathes should also be at about the same pressure.



Figure

But at a total pressure of 2 atm, the partial pressure of O_2 in ordinary air would be 0.42 atm; at a depth of 100 ft (about 30 m), the O_2 pressure of .8 atm would be far too high for health. For this reason, the air mixture in the pressurized tanks that scuba divers wear must contain a smaller fraction of O_2. This can be achieved most simply by raising the nitrogen content, but high partial pressures of N_2 can also be dangerous, resulting in a condition known as nitrogen narcosis. The preferred diluting agent for sustained deep diving is helium, which has very little tendency to dissolve in the blood even at high pressures.

This page titled 6.3: Dalton's Law is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

6.4: Kinetic Molecular Theory (Overview)

Learning Objectives

Make sure you thoroughly understand the following essential ideas which are presented below. It is especially important that you know the principal assumptions of the kinetic-molecular theory. These can be divided into those that refer to the nature of the molecules themselves, and those that describe the nature of their motions:

- **The molecules** - Negligible volume, absence of inermolecular attractions (think of them as very hard, "non-sticky" objects.)
- **Their motions** - Completely random in direction, in straight lines only (this is a consequence of their lack of attractions), average velocities proportional to the absolute temperature..
- The idea that random motions of individual molecules can result in non-random (directed) movement of the gas as a whole is one of the most important concepts of chemistry, exemplified here as the principle of *diffusion*.
- In most courses you will be expected to know and be able to use (or misuse!) *Graham's law*.

The properties such as temperature, pressure, and volume, together with others dependent on them (density, thermal conductivity, etc.) are known as *macroscopic* properties of matter; these are properties that can be observed in bulk matter, without reference to its underlying structure or molecular nature. By the late 19th century the atomic theory of matter was sufficiently well accepted that scientists began to relate these macroscopic properties to the behavior of the individual molecules, which are described by the *microscopic* properties of matter. The outcome of this effort was the *kinetic molecular theory* of gases. This theory applies strictly only to a hypothetical substance known as an *ideal gas*; we will see, however, that under many conditions it describes the behavior of real gases at ordinary temperatures and pressures quite accurately, and serves as the starting point for dealing with more complicated states of matter.

The basic ideas of kinetic-molecular theory

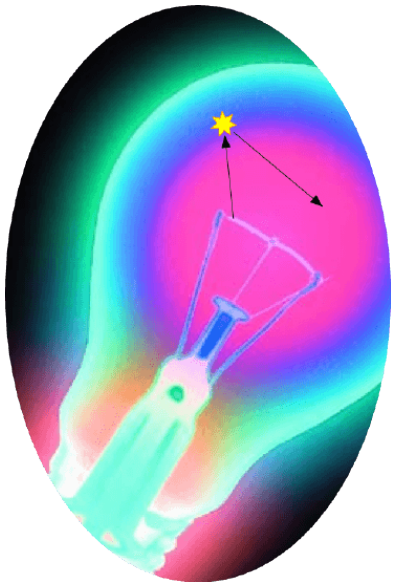
The "kinetic-molecular theory of gases" may sound rather imposing, but it is based on a series of easily-understood assumptions that, taken together, constitute a model that greatly simplifies our understanding of the gaseous state of matter. The five basic tenets of the kinetic-molecular theory are as follows:

1. A gas is composed of molecules that are separated by average distances that are much greater than the sizes of the molecules themselves. ***The volume occupied by the molecules of the gas is negligible compared to the volume of the gas itself.***
2. The molecules of an ideal gas exert **no attractive forces** on each other, or on the walls of the container.
3. The molecules are in **constant random motion**, and as material bodies, they obey Newton's laws of motion. This means that the molecules move in **straight lines** (see demo illustration at the left) until they collide with each other or with the walls of the container.
4. Collisions are perfectly *elastic*; when two molecules collide, they change their directions and kinetic energies, but the total **kinetic energy is conserved**. *Collisions are not "sticky"*.
5. **The average kinetic energy of the gas molecules is directly proportional to the absolute temperature**. Notice that the term "average" is very important here; the velocities and kinetic energies of individual molecules will span a wide range of values, and some will even have zero velocity at a given instant. This implies that all molecular motion would cease if the temperature were reduced to absolute zero.

According to this model, most of the volume occupied by a gas is *empty space*; this is the main feature that distinguishes gases from *condensed* states of matter (liquids and solids) in which neighboring molecules are constantly in contact. Gas molecules are in rapid and continuous motion; at ordinary temperatures and pressures their velocities are of the order of 0.1-1 km/sec and each molecule experiences approximately 10^{10} collisions with other molecules every second.

The Gas Laws explained

If gases do in fact consist of widely-separated particles, then the observable properties of gases must be explainable in terms of the simple mechanics that govern the motions of the individual molecules. The kinetic molecular theory makes it easy to see why a gas should exert a pressure on the walls of a container. Any surface in contact with the gas is constantly bombarded by the molecules.



At each collision, a molecule moving with momentum *mv* strikes the surface. Since the collisions are elastic, the molecule bounces back with the same velocity in the opposite direction. This change in velocity ΔV is equivalent to a*n acceleration a*; according to Newton's second law, a *force f = ma* is thus exerted on the surface of area *A* exerting a pressure *P = f/A*.

Kinetic Interpretation of Temperature

According to the kinetic molecular theory, the average kinetic energy of an ideal gas is directly proportional to the absolute temperature. Kinetic energy is the energy a body has by virtue of its motion:

$$K.E. = \frac{mv^2}{2} \tag{6.4.1}$$

As the temperature of a gas rises, the average velocity of the molecules will increase; a doubling of the temperature will increase this velocity by a factor of four. Collisions with the walls of the container will transfer more momentum, and thus more kinetic energy, to the walls. If the walls are cooler than the gas, they will get warmer, returning less kinetic energy to the gas, and causing it to cool until thermal equilibrium is reached. Because temperature depends on the *average* kinetic energy, the concept of temperature only applies to a statistically meaningful sample of molecules. We will have more to say about molecular velocities and kinetic energies farther on.

- **Kinetic explanation of Boyle's law**: Boyle's law is easily explained by the kinetic molecular theory. The pressure of a gas depends on the number of times per second that the molecules strike the surface of the container. If we compress the gas to a smaller volume, the same number of molecules are now acting against a smaller surface area, so the number striking per unit of area, and thus the pressure, is now greater.
- **Kinetic explanation of Charles' law:** Kinetic molecular theory states that an increase in temperature raises the average kinetic energy of the molecules. If the molecules are moving more rapidly but the pressure remains the same, then the molecules must stay farther apart, so that the increase in the rate at which molecules collide with the surface of the container is compensated for by a corresponding increase in the area of this surface as the gas expands.
- **Kinetic explanation of Avogadro's law**: If we increase the number of gas molecules in a closed container, more of them will collide with the walls per unit time. If the pressure is to remain constant, the volume must increase in proportion, so that the molecules strike the walls less frequently, and over a larger surface area.
- **Kinetic explanation of Dalton's law**: "Every gas is a vacuum to every other gas". This is the way Dalton stated what we

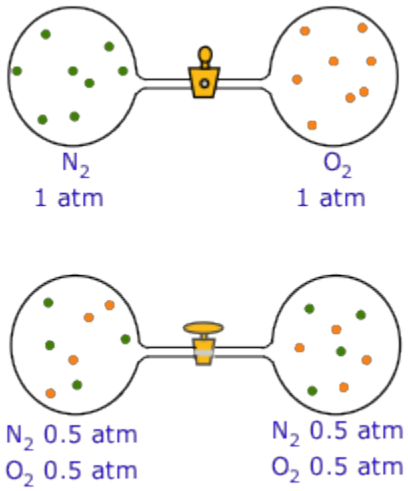
now know as his law of partial pressures. It simply means that each gas present in a mixture of gases acts independently of the others. This makes sense because of one of the fundamental tenets of KMT theory that gas molecules have negligible volumes. So Gas A in mixture of A and B acts as if Gas B were not there at all. Each contributes its own pressure to the total pressure within the container, in proportion to the fraction of the molecules it represents.

Some important practical applications of KMT

The molecules of a gas are in a state of perpetual motion in which the velocity (that is, the speed and direction) of each molecule is completely random and independent of that of the other molecules. This fundamental assumption of the kinetic-molecular model helps us understand a wide range of commonly-observed phenomena.

Diffusion: random motion with direction

Diffusion refers to the transport of matter through a *concentration gradient*; the rule is that substances move (or tend to move) from regions of higher concentration to those of lower concentration. The diffusion of tea out of a tea bag into water, or of perfume from a person, are common examples; we would not expect to see either process happening in reverse!



When the stopcock is opened, random motions cause each gas to diffuse into the other container. After diffusion is complete (bottom), individual molecules of both kinds continue to pass between the flasks in both directions.

It might at first seem strange that the *random* motions of molecules can lead to a completely predictable drift in their ultimate distribution. The key to this apparent paradox is the distinction between an *individual* and the *population*. Although we can say nothing about the fate of an individual molecule, the behavior of a large collection ("population") of molecules is subject to the laws of statistics. This is exactly analogous to the manner in which insurance actuarial tables can accurately predict the average longevity of people at a given age, but provide no information on the fate of any single person.

Effusion and Graham's law

If a tiny hole is made in the wall of a vessel containing a gas, then the rate at which gas molecules leak out of the container will be proportional to the number of molecules that collide with unit area of the wall per second, and thus with the **rms - average velocity** of the gas molecules. This process, when carried out under idealized conditions, is known as *effusion*.

Around 1830, the English chemist Thomas Graham (1805-1869) discovered that the relative rates at which two different gases, at the same temperature and pressure, will effuse through identical openings is inversely proportional to the square root of its molar mass.

$$v \propto \frac{1}{\sqrt{M}} \tag{6.4.2}$$

Graham's law, as this relation is known, is a simple consequence of the square-root relation between the velocity of a body and its kinetic energy.

$$HCl + NH_3 \rightarrow NH_4Cl_{(s)} \tag{6.4.3}$$

According to the kinetic molecular theory, the molecules of two gases at the same temperature will possess the same average kinetic energy. If $v_1$and v_2 are the average velocities of the two kinds of molecules, then at any given temperature $KE_1 = KE_2$ and

$$\frac{m_1 v_1^2}{2} = \frac{m_2 v_2^2}{2} \tag{6.4.4}$$

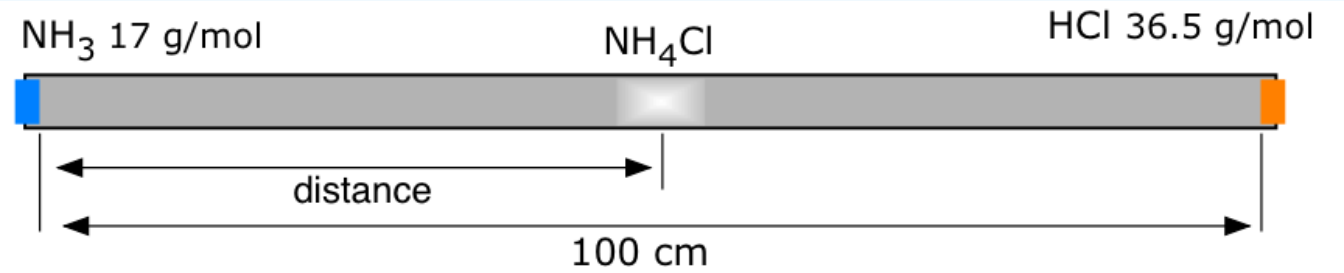
or, in terms of molar masses M,

$$\frac{v_1}{v_2} = \sqrt{\frac{M_2}{M_1}} \tag{6.4.5}$$

Thus the average velocity of the lighter molecules must be greater than those of the heavier molecules, and the ratio of these velocities will be given by the inverse ratio of square roots of the molecular weights. Although Graham's law applies exactly only when a gas diffuses into a vacuum, the law gives useful estimates of relative diffusion rates under more practical conditions, and it provides insight into a wide range of phenomena that depend on the relative average velocities of molecules of different masses.

Example 6.4.1

The glass tube shown above has cotton plugs inserted at either end. The plug on the left is moistened with a few drops of aqueous ammonia, from which NH_3 gas slowly escapes. The plug on the right is similarly moisted with a strong solution of hydrochloric acid, from which gaseous HCl escapes. The gases diffuse in opposite directions within the tube; at the point where they meet, they combine to form solid ammonium chloride, which appears first as a white fog and then begins to coat the inside of the tube.



The reaction is

$$NH_{3(g)} + HCl_{(g)} \rightarrow NH_4Cl_{(s)} \tag{6.4.6}$$

a. In what part of the tube (left, right, center) will the NH_4Cl first be observed?
b. If the distance between the two ends of the tube is 100 cm, how many cm from the left end of the tube will the NH_4Cl first form?

Solution

a) The lighter ammonia molecules will diffuse more rapidly, so the point where the two gases meet will be somewhere in the right half of the tube.

b) The ratio of the diffusion velocities of ammonia (v_1)and hydrogen chloride (v_2) can be estimated from Graham's law:

$$\frac{v_1}{v_2} = \sqrt{\frac{36.5}{17}} = 1.46 \tag{6.4.7}$$

We can therefore assign relative velocities of the two gases as $v_1 = 1.46$ and $v_2 = 1$. Clearly, the meeting point will be directly proportional to v_1. It will, in fact, be proportional to the ratio $v_1/(v_1+v_2)$*:

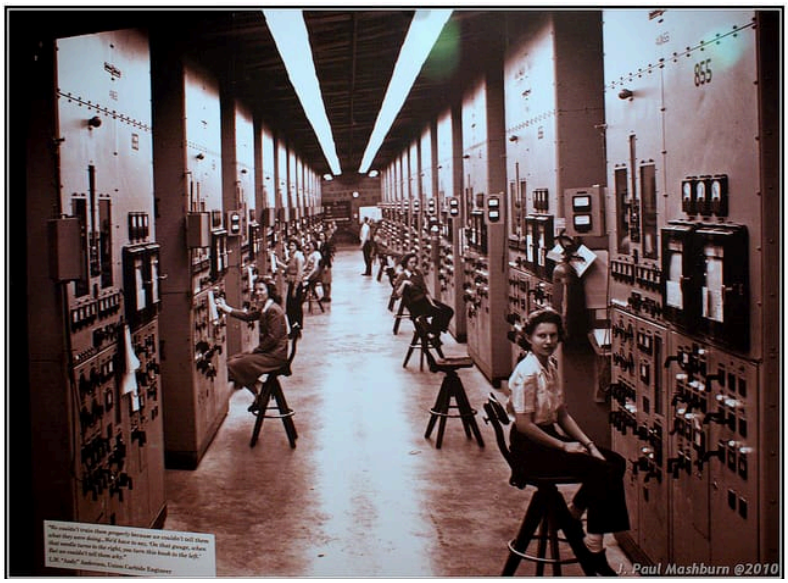
$$\frac{v_1}{v_1 + v_2} \times 100\ cm = \frac{1.46}{1.46 + 1.00} \times 100\,cm = 59\ cm \tag{6.4.8}$$

*In order to see how this ratio was deduced, consider what would happen in the three special cases in which v_1=0, v_2=0, and v_1=v_2, for which the distances (from the left end) would be 0, 50, and 100 cm, respectively. It should be clear that the simpler ratio v_1/v_2 would lead to absurd results.

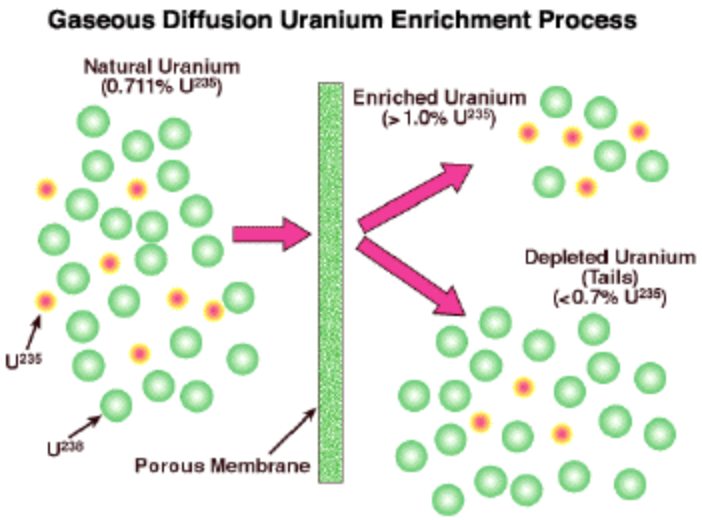
Note that the above calculation is only an estimate. Graham's law is strictly valid only under special conditions, the most important one being that no other gases are present. Contrary to what is written in some textbooks and is often taught, Graham's law does not accurately predict the relative rates of escape of the different components of a gaseous mixture into the outside air, nor does it give the rates at which two gases will diffuse through another gas such as air. See *Misuse of Graham*'s *Law*s by Stephen J. Hawkes, *J. Chem. Education* 1993 70(10) 836-837

Uranium enrichment

One application of this principle that was originally suggested by Graham himself but was not realized on a practical basis until a century later is the separation of isotopes. The most important example is the enrichment of uranium in the production of nuclear fission fuel.



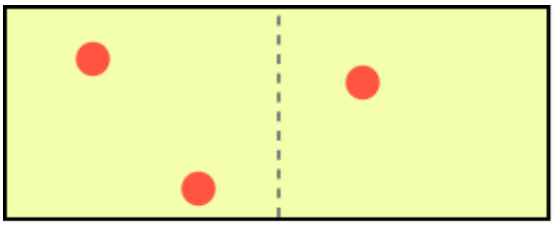
The K-25 Gaseous Diffusion Plant was one of the major sources of enriched uranium during World War II. It was completed in 1945 and employed 12,000 workers. Owing to the secrecy of the Manhattan Project, the women who operated the system were unaware of the purpose of the plant; they were trained to simply watch the gauges and turn the dials for what they were told was a "government project".

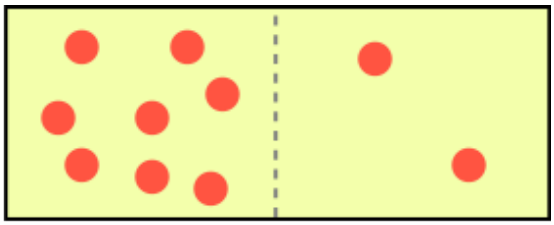


Uranium consists mostly of U^{238}, with only 0.7% of the fissionable isotope U^{235}. Uranium is of course a metal, but it reacts with fluorine to form a gaseous hexafluoride, UF_6. In the very successful *gaseous diffusion process* the UF_6 diffuses repeatedly through a porous wall. Each time, the lighter isotope passes through a bit more rapidly then the heavier one, yielding a mixture that is minutely richer in U^{235}. The process must be over a thousand times to achieve the desired degree of enrichment. The development of a large-scale diffusion plant was a key part of the U.S. development of the first atomic bomb in 1945. This process is now obsolete, having been replaced by other methods.

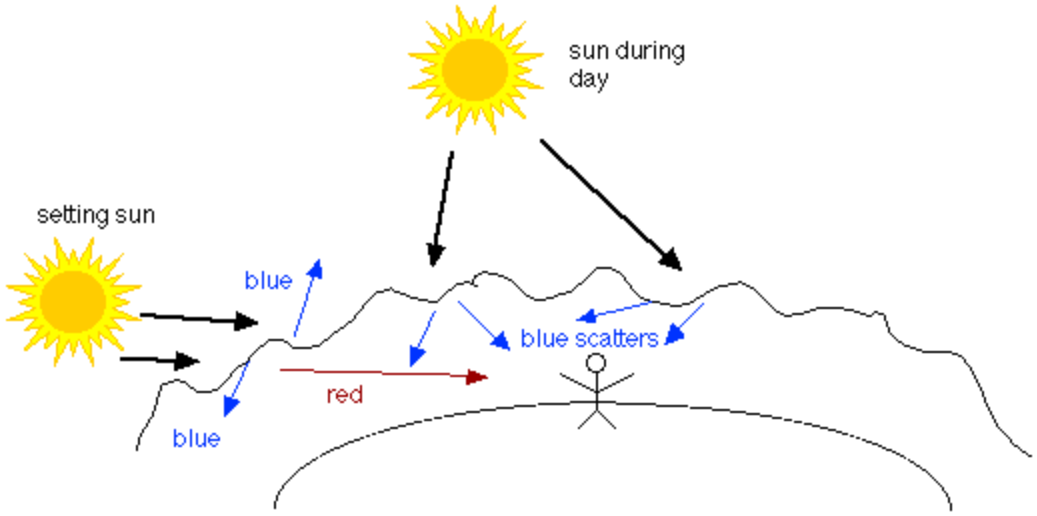
Density fluctuations: Why is the sky blue?

Diffusion ensures that molecules will quickly distribute themselves throughout the volume occupied by the gas in a thoroughly uniform manner. The chances are virtually zero that sufficiently more molecules might momentarily find themselves near one side of a container than the other to result in an observable temporary density or pressure difference. This is a result of simple statistics. But statistical predictions are only valid when the sample population is large.



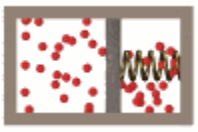


Consider what would happen if we consider extremely small volumes of space: cubes that are about 10^{-7} cm on each side, for example. Such a cell would contain only a few molecules, and at any one instant we would expect to find some containing more or less than others, although in time they would average out to the same value. The effect of this statistical behavior is to give rise to random fluctuations in the density of a gas over distances comparable to the dimensions of visible light waves. When light passes through a medium whose density is non-uniform, some of the light is *scattered*. The kind of scattering due to random density fluctuations is called *Rayleigh scattering*, and it has the property of affecting (scattering) shorter wavelengths more effectively than longer wavelengths. The clear sky appears blue in color because the blue (shorter wavelength) component of sunlight is scattered more. The longer wavelengths remain in the path of the sunlight, available to delight us at sunrise or sunset.



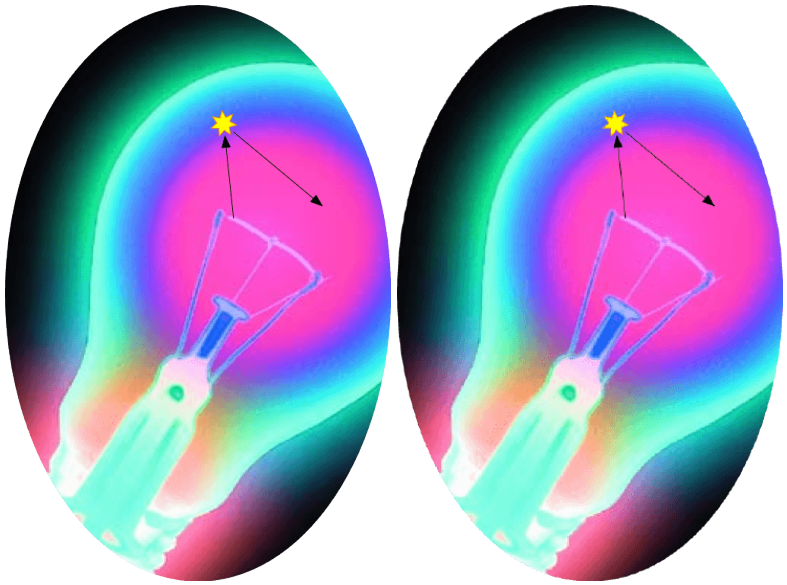
[source]

What we have been discussing is a form of what is known as *fluctuation phenomena*. As the animation shows, the random fluctuations in pressure of a gas on either side do not always completely cancel when the density of molecules (i.e., *pressures*) are quite small.



An interesting application involving several aspects of the kinetic molecular behavior of gases is the use of a gas, usually argon, to extend the lifetime of incandescent lamp bulbs. As a light bulb is used, tungsten atoms evaporate from the filament and condense on the cooler inner wall of the bulb, blackening it and reducing light output. As the filament gets thinner in certain spots, the increased electrical resistance results in a higher local power dissipation, more rapid evaporation, and eventually the filament breaks.

The pressure inside a lamp bulb must be sufficiently low for the *mean free path* of the gas molecules to be fairly long; otherwise heat would be conducted from the filament too rapidly, and the bulb would melt. (Thermal conduction depends on intermolecular collisions, and a longer mean free path means a lower collision frequency). A complete vacuum would minimize heat conduction, but this would result in such a long mean free path that the tungsten atoms would rapidly migrate to the walls, resulting in a very short filament life and extensive bulb blackening.



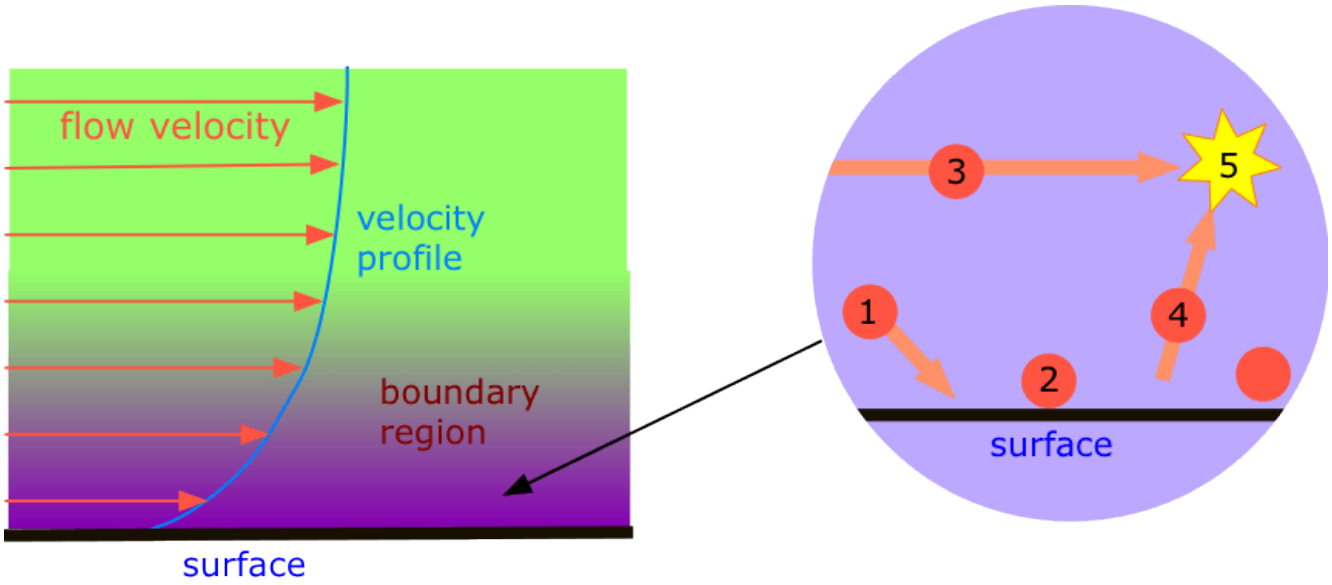
Around 1910, the General Electric Company hired Irving Langmuir as one of the first chemists to be employed as an industrial scientist in North America. Langmuir quickly saw that bulb blackening was a consequence of the long mean free path of vaporized tungsten atoms, and he showed that the addition of a small amount of argon will reduce the mean free path, increasing the probability that an outward-moving tungsten atom will collide with an argon atom. A certain proportion of these will eventually find their way back to the filament, partially reconstituting it.

Krypton would be a better choice of gas than argon, since its greater mass would be more effective in changing the direction of the rather heavy tungsten atom. Unfortunately, krypton, being a rarer gas, is around 50 times as expensive as argon, so it is used only in “premium” light bulbs. The more recently-developed *halogen-cycle lamp* is an interesting chemistry-based method of prolonging the life of a tungsten-filament lamp.

Viscosity of gases

Gases, like all fluids, exhibit a resistance to flow, a property known as *viscosity*. The basic cause of viscosity is the random nature of thermally-induced molecular motion. In order to force a fluid through a pipe or tube, an additional non-random translational motion must be superimposed on the thermal motion.

There is a slight problem, however. Molecules flowing near the center of the pipe collide mostly with molecules moving in the same direction at about the same velocity, but those that happen to find themselves near the wall will experience frequent collisions with the wall. Since the molecules in the wall of the pipe are not moving in the direction of the flow, they will tend to absorb more kinetic energy than they return, with the result that the gas molecules closest to the wall of the pipe lose some of their forward momentum. Their random thermal motion will eventually take them deeper into the stream, where they will collide with other flowing molecules and slow them down. This gives rise to a resistance to flow known as *viscosity*; this is the reason why long gas transmission pipelines need to have pumping stations every 100 km or so.

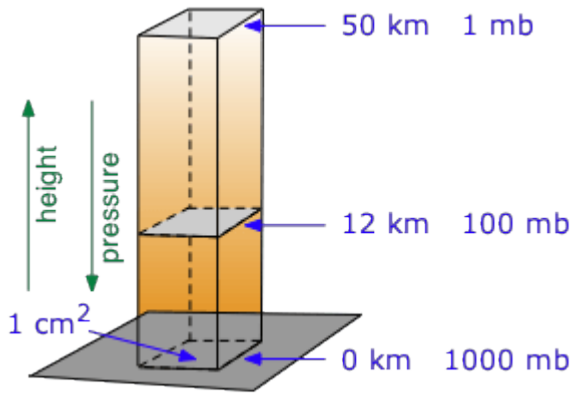


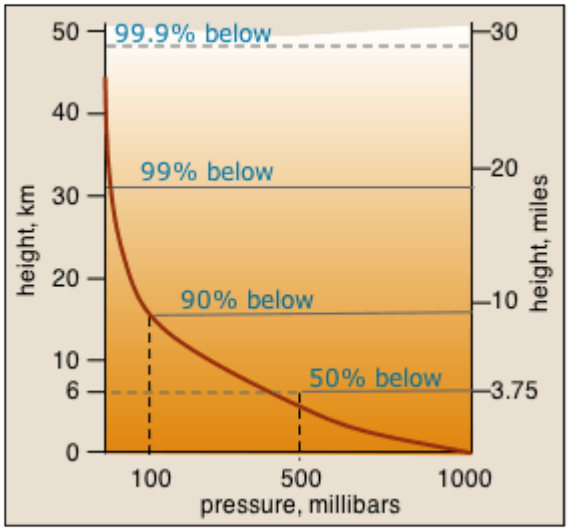
Origin of gas viscosity. This shows the boundary region where random movements of molecules in directions other than the one of the flow (1) move toward the confining surface and temporarily adsorb to it (2). After a short time, thermal energy causes the molecule to be released (4) with most of its velocity not in the flow direction. A rapidly-flowing molecule (**3**) collides with it (**5**) and loses some of its flow velocity.

As you know, liquids such as syrup or honey exhibit smaller viscosities at higher temperatures as the increased thermal energy reduces the influence of intermolecular attractions, thus allowing the molecules to slip around each other more easily. Gases, however, behave in just the opposite way; gas viscosity arises from collision-induced transfer of momentum from rapidly-moving molecules to slow ones that have been released from the boundary layer. The higher the temperature, the more rapidly the molecules move and collide with each other, so the higher the viscosity.

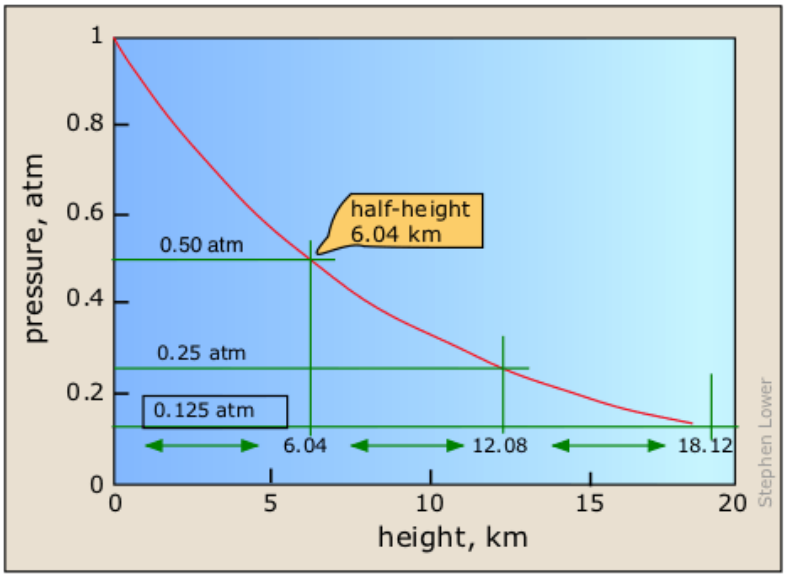
Distribution of gas molecules in a gravitational field

Everyone knows that the air pressure decreases with altitude. This effect is easily understood qualitatively through the kinetic molecular theory. Random thermal motion tends to move gas molecules in all directions equally. In the presence of a gravitational field, however, motions in a downward direction are slightly favored. This causes the concentration, and thus the pressure of a gas to be greater at lower elevations and to decrease without limit at higher elevations.





The pressure at any elevation in a vertical column of a fluid is due to the weight of the fluid above it. This causes the pressure to decrease exponentially with height.



This plot shows how the pressure of air at 25° C decreases with altitude. Note the constant increment of altitude (6.04 km, the "half height") required to reduce the pressure by half its value. This reflects the special property of an exponential function a= e^y, namely that the derivative da/dy is just e^y itself.

The exact functional relationship between pressure and altitude is known as the *barometric distribution law*. It is easily derived using first-year calculus. For air at 25°C the pressure P_h at any altitude is given by

$$P_h = P_o e^{-.011h} \tag{6.4.9}$$

in which P_o is the pressure at sea level.

This is a form of the very common *exponential decay law* which we will encounter in several different contexts in this course. An exponential decay (or growth) law describes any quantity whose rate of change is directly proportional to its current value, such as the amount of money in a compound-interest savings account or the density of a column of gas at any altitude. The most important feature of any quantity described by this law is that the fractional rate of change of the quantity in question (in this case, $\Delta P/P$ or in calculus, dP/P) is a constant. This means that the increase in altitude required to reduce the pressure by half is also a constant, about 6 km in the Earth's case.

Because heavier molecules will be more strongly affected by gravity, their concentrations will fall off more rapidly with elevation. For this reason the partial pressures of the various components of the atmosphere will tend to vary with altitude. The difference in pressure is also affected by the temperature; at higher temperatures there is more thermal motion, and hence a less rapid fall-off of pressure with altitude. Owing to atmospheric convection and turbulence, these effects are not observed in the lower part of the atmosphere, but in the uppermost parts of the atmosphere the heavier molecules do tend to drift downward.

The ionosphere and radio communication

At very low pressures, mean free paths are sufficiently great that collisions between molecules become rather infrequent. Under these conditions, highly reactive species such as ions, atoms, and molecular fragments that would ordinarily be destroyed on every collision, can persist for appreciable periods of time.

The most important example of this occurs at the top of the Earth's atmosphere, at an altitude of 200 km, at pressure of about 10^{-7} atm. Here the mean free path will be 10^7 times its value at 1 atm, or about 1 m. In this part of the atmosphere, known as the *thermosphere*, the chemistry is dominated by species such as O, O_2^+ and HO which are formed by the action of intense solar ultraviolet light on the normal atmospheric gases near the top of the stratosphere. The high concentrations of electrically charged species in these regions (sometimes also called the *ionosphere*) reflect radio waves and are responsible for around-the-world transmission of mid-frequency radio signals.

The ion density in the lower part of the ionosphere (about 80 km altitude) is so great that the radiation from broadcast-band radio stations is absorbed in this region before these waves can reach the reflective high-altitude layers. However, the pressure in this

region (known as the *D-layer*) is great enough that the ions recombine soon after local sunset, causing the D-layer to disappear and allowing the waves to reflect off of the upper (F-layer) part of the ionosphere. This is the reason that distant broadcast stations can only be heard at night.

This page titled 6.4: Kinetic Molecular Theory (Overview) is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

6.5: More on Kinetic Molecular Theory

Learning Objectives

Make sure you thoroughly understand the following essential ideas that are presented below. It is especially important that you know the precise meanings of all the italicized terms in the context of this topic.

- You should be able to sketch the general shape of the *Maxwell-Boltzmann plot* showing the distribution of molecular velocities. You should also show how these plots are affected by the temperature and by the molar mass.
- Although there is no need for you to be able to derive the *ideal gas equation of state*, you should understand that the equation $PV = nRT$ can be derived from the principles of kinetic molecular theory, as outlined above.
- Explain the concept of the *mean free path* of a gas molecule (but no need to reproduce the mathematics.)

In this section, we look in more detail at some aspects of the kinetic-molecular model and how it relates to our empirical knowledge of gases. For most students, this will be the first application of algebra to the development of a chemical model; this should be educational in itself, and may help bring that subject back to life for you! As before, your emphasis should on understanding these models and the ideas behind them, there is no need to memorize any of the formulas.

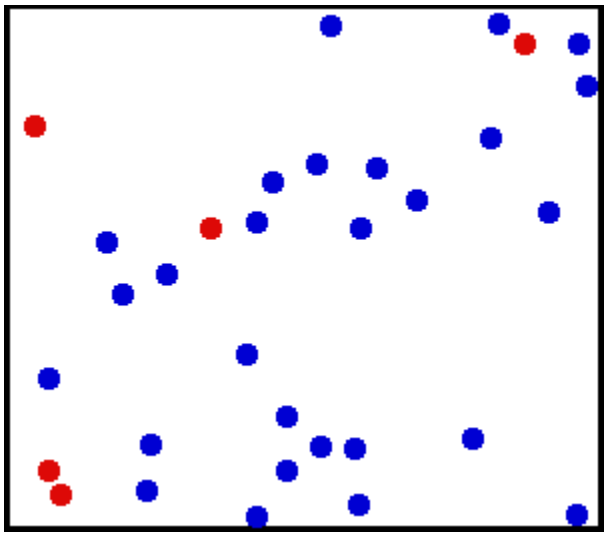


Image: Wikimedia Commons

The Velocities of Gas Molecules

At temperatures above absolute zero, all molecules are in motion. In the case of a gas, this motion consists of straight-line jumps whose lengths are quite great compared to the dimensions of the molecule. Although we can never predict the velocity of a particular individual molecule, the fact that we are usually dealing with a huge number of them allows us to know what fraction of the molecules have kinetic energies (and hence velocities) that lie within any given range.

The trajectory of an individual gas molecule consists of a series of straight-line paths interrupted by collisions. What happens when two molecules collide depends on their relative kinetic energies; in general, a faster or heavier molecule will impart some of its kinetic energy to a slower or lighter one. Two molecules having identical masses and moving in opposite directions at the same speed will momentarily remain motionless after their collision.

If we could measure the instantaneous velocities of all the molecules in a sample of a gas at some fixed temperature, we would obtain a wide range of values. A few would be zero, and a few would be very high velocities, but the majority would fall into a more or less well defined range. We might be tempted to define an average velocity for a collection of molecules, but here we would

need to be careful: molecules moving in opposite directions have velocities of opposite signs. Because the molecules are in a gas are in random thermal motion, there will be just about as many molecules moving in one direction as in the opposite direction, so the velocity vectors of opposite signs would all cancel and the average velocity would come out to zero. Since this answer is not very useful, we need to do our averaging in a slightly different way.

The proper treatment is to average the *squares* of the velocities, and then take the square root of this value. The resulting quantity is known as the *root mean square*, or *RMS* velocity

$$\nu_{rms} = \sqrt{\frac{\sum \nu^2}{n}} \tag{6.5.1}$$

which we will denote simply by $\bar{v}$. The formula relating the RMS velocity to the temperature and molar mass is surprisingly simple, considering the great complexity of the events it represents:

$$\bar{v} = v_{rms} = \sqrt{\frac{3RT}{m}} \tag{6.5.2}$$

in which m is the molar mass in kg mol^{-1}, and k = R÷6.02E23, the "gas constant per molecule", is known as the *Boltzmann constant*.

Example 6.5.1

What is the average velocity of a nitrogen molecules at 300 K?

Solution

The molar mass of N_2 is 28.01 g. Substituting in the above equation and expressing R in energy units, we obtain

$$v^2 = \frac{3 \times 8.31 \text{Jmol}^{-1}\text{K}^{-1} \times 300\text{K}}{28.01 \times 10^{-3}\text{kgmol}^{-1}} = 2.67 \times 10^5 \text{Jkg}^{-1}$$

Recalling the definition of the joule (1 J = 1 kg m^2 s^{-2}) and taking the square root,

$$\bar{v} = \sqrt{2.67 \times 10^5 \text{Jkg}^{-1} \times \frac{1\text{kgm}^2\text{s}^{-2}}{1\text{J}}} = 517\text{ms}^{-1}$$

or

$$517\text{ms}^{-1} \times \frac{1\text{km}}{10^3\text{m}} \times \frac{3600\text{s}}{1\text{h}} = 1860\text{kmh}^{-1}$$

Comment: this is fast! The velocity of a rifle bullet is typically 300-500 m s^{-1}; convert to common units to see the comparison for yourself.

A simpler formula for estimating average molecular velocities is

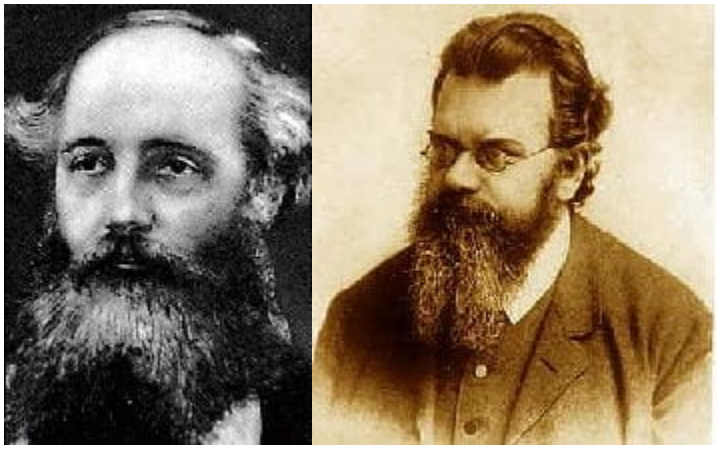
$$v = 157\sqrt{\frac{T}{m}} \tag{6.5.3}$$

in which v is in units of meters/sec, T is the absolute temperature and m the molar mass in grams.

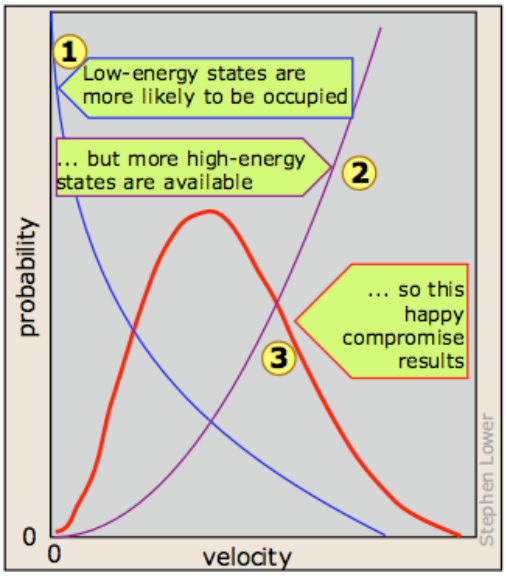
The Boltzmann Distribution

If we were to plot the number of molecules whose velocities fall within a series of narrow ranges, we would obtain a slightly asymmetric curve known as a *velocity distribution*. The peak of this curve would correspond to the *most probable* velocity. This velocity distribution curve is known as the *Maxwell-Boltzmann distribution*, but is frequently referred to only by Boltzmann's name. The **Maxwell-Boltzmann distribution law** was first worked out around 1850 by the great Scottish physicist, **James Clerk**

Maxwell (left, 1831-1879), who is better known for discovering the laws of electromagnetic radiation. Later, the Austrian physicist **Ludwig Boltzmann** (1844-1906) put the relation on a sounder theoretical basis and simplified the mathematics somewhat. Boltzmann pioneered the application of statistics to the physics and thermodynamics of matter, and was an ardent supporter of the atomic theory of matter at a time when it was still not accepted by many of his contemporaries.



The derivation of the Boltzmann curve is a bit too complicated to go into here, but its physical basis is easy to understand. Consider a large population of molecules having some fixed amount of kinetic energy. As long as the temperature remains constant, this total energy will remain unchanged, but it can be distributed among the molecules in many different ways, and this distribution will change continually as the molecules collide with each other and with the walls of the container.



It turns out, however, that kinetic energy is acquired and handed around only in discrete amounts which are known as *quanta*. Once the molecule has a given number of kinetic energy quanta, these can be apportioned amongst the three directions of motion in many different ways, each resulting in a distinct total *velocity state* for the molecule. The greater the number of quanta, (that is, the greater the total kinetic energy of the molecule) the greater the number of possible velocity states. If we assume that all velocity states are equally probable, then simple statistics predicts that higher velocities will be more favored simply because there are so many more of them (2).

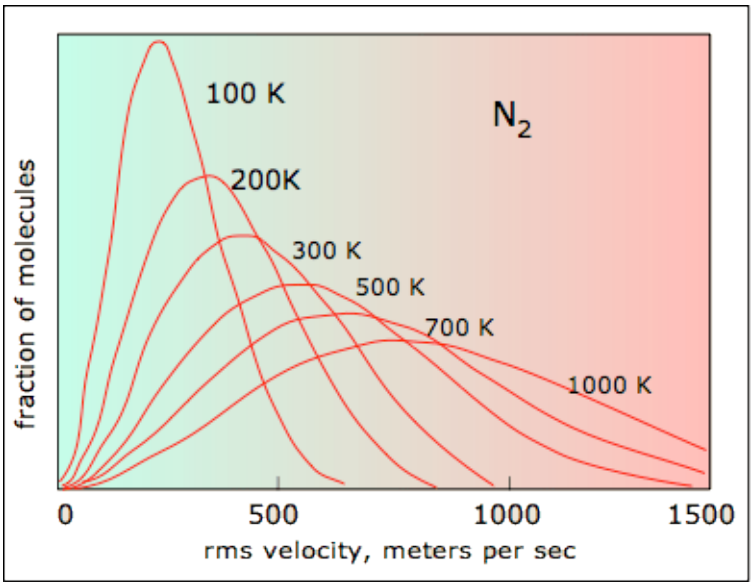
Although the number of possible higher-energy states is greater, the lower-energy states are more likely to be occupied (1). This is because only so much kinetic energy available to the gas as a whole; every molecule that acquires kinetic energy in a collision leaves behind another molecule having less. This tends to even out the kinetic energies in a collection of molecules, and ensures that there are always some molecules whose instantaneous velocity is near zero. The net effect of these two opposing tendencies,

one favoring high kinetic energies and the other favoring low ones, is the peaked curve (3) seen above. Notice that because of the asymmetry of this curve, the *mean* (rms average) velocity is not the same as the *most probable* velocity, which is defined by the peak of the curve.

At higher temperatures (or with lighter molecules) the latter constraint becomes less important, and the mean velocity increases. But with a wider velocity distribution, the number of molecules having any one velocity diminishes, so the curve tends to flatten out.

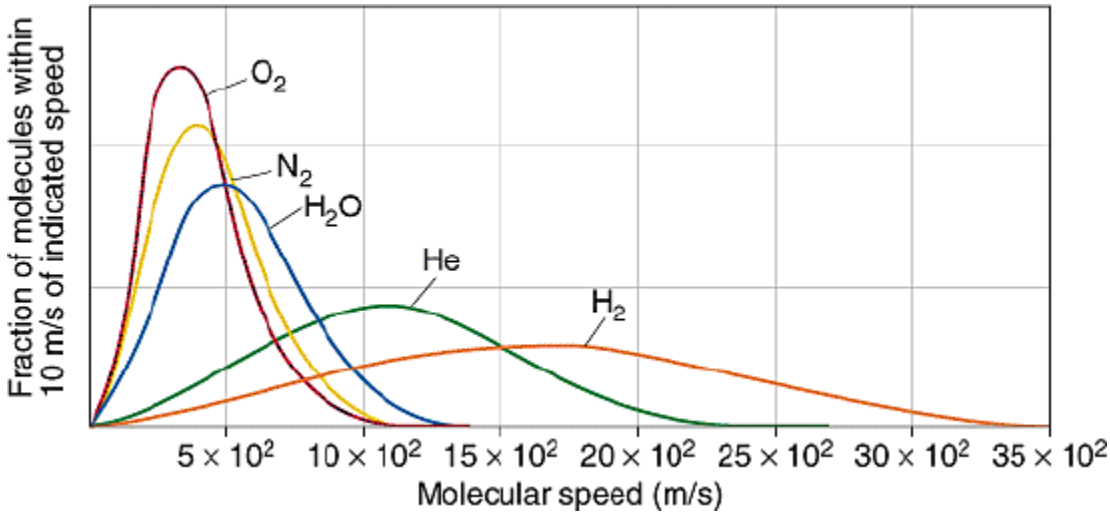
Velocity Distributions Depend on Temperature and Mass

Higher temperatures allow a larger fraction of molecules to acquire greater amounts of kinetic energy, causing the Boltzmann plots to spread out.



Notice how the left ends of the plots are anchored at zero velocity (there will always be a few molecules that happen to be at rest.) As a consequence, the curves flatten out as the higher temperatures make additional higher-velocity states of motion more accessible. The area under each plot is the same for a constant number of molecules.

All molecules have the same kinetic energy ($mv^2/2$) at the same temperature, so the fraction of molecules with higher velocities will increase as m, and thus the molecular weight, decreases.



The ability of a planet to retain an atmospheric gas depends on the average velocity (and thus on the temperature and mass) of the gas molecules and on the planet's mass, which determines its gravity and thus the escape velocity. In order to retain a gas for the age of the solar system, the average velocity of the gas molecules should not exceed about one-sixth of the escape velocity. The escape velocity from the Earth is 11.2 km/s, and 1/6 of this is about 2 km/s. Examination of the above plot reveals that hydrogen molecules can easily achieve this velocity, and this is the reason that hydrogen, the most abundant element in the universe, is almost absent from Earth's atmosphere.

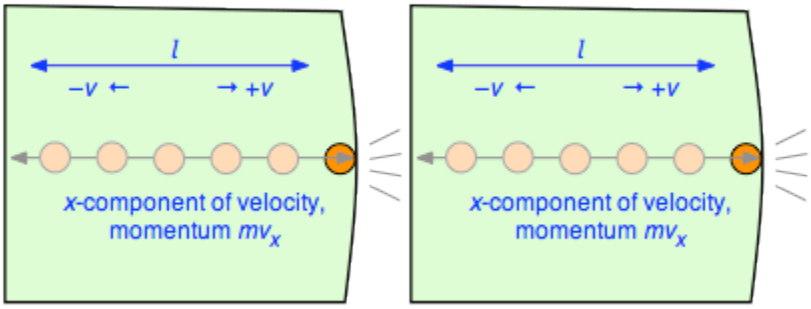
Although **hydrogen** is not a significant atmospheric component, water vapor is. A very small amount of this diffuses to the upper part of the atmosphere, where intense solar radiation breaks down the H_2O into H_2. Escape of this hydrogen from the upper atmosphere amounts to about 2.5×10^{10} g/year.

Derivation of the Ideal Gas Equation

The ideal gas equation of state came about by combining the empirically determined ("ABC") laws of Avogadro, Boyle, and Charles, but one of the triumphs of the kinetic molecular theory was the derivation of this equation from simple mechanics in the late nineteenth century. This is a beautiful example of how the principles of elementary mechanics can be applied to a simple model to develop a useful description of the behavior of macroscopic matter. We begin by recalling that the pressure of a gas arises from the force exerted when molecules collide with the walls of the container. This force can be found from Newton's law

$$f = ma = m\frac{dv}{dt} \tag{6.5.4}$$

in which v is the velocity component of the molecule in the direction perpendicular to the wall and m is its mass.



To evaluate the derivative in Equation 6.5.4, which is the velocity change per unit time, consider a single molecule of a gas contained in a cubic box of length \(l\). For simplicity, assume that the molecule is moving along the *x*-axis which is perpendicular to a pair of walls, so that it is continually bouncing back and forth between the same pair of walls. When the molecule of mass \(*m*\) strikes the wall at velocity \(+*v*\) (and thus with a momentum \(*mv*\)) it will rebound elastically and end up moving in the opposite direction with –*v*. The total change in velocity per collision is thus 2*v* and the change in momentum is $2mv$.

The Frequency of Collisions

After the collision the molecule must travel a distance l to the opposite wall, and then back across this same distance before colliding again with the wall in question. This determines the time between successive collisions with a given wall; the number of collisions per second will be $v/2l$. The *force* F exerted on the wall is the rate of change of the momentum, given by the product of the momentum change per collision and the collision frequency:

$$F = \frac{d(mv_x}{dt} = (2mv_x) \times \left(\frac{v_x}{2l}\right) = \frac{mv_x^2}{l} \tag{6.5.5}$$

Pressure is force per unit area, so the pressure P exerted by the molecule on the wall of cross-section l^2 becomes

$$P = \frac{mv^2}{l^3} = \frac{mv^2}{V} \tag{6.5.6}$$

in which V is the volume of the box.

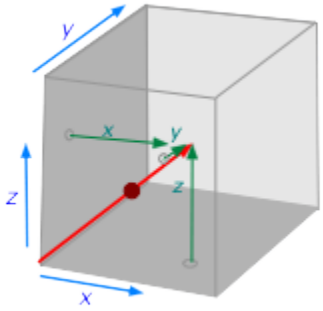
The pressure produced by *N* molecules

As noted near the beginning of this unit, any given molecule will make about the same number of moves in the positive and negative directions, so taking a simple average would yield zero. To avoid this embarrassment, we square the velocities before averaging them, and **then** take the square root of the average. This result is known as the *root mean square* (rms) velocity.

We have calculated the pressure due to a single molecule moving at a constant velocity in a direction perpendicular to a wall. If we now introduce more molecules, we must interpret v^2 as an average value which we will denote by $\bar{v^2}$. Also, since the molecules are moving randomly in all directions, only one-third of their total velocity will be directed along any one Cartesian axis, so the total pressure exerted by N molecules becomes

$$P = \frac{N}{3}\frac{m\bar{\nu}^2}{V} \tag{6.5.7}$$

The above statement that "one-third of the total velocity (of all the molecules together)..." does *not* mean that 1/3 of the molecules themselves are moving in each of these three directions; each individual particle is free to travel in any possible direction between collisions. However, any random trajectory can be regarded as composed of three *components* that correspond to these three axes.



The red arrow in the illustration depicts the path of a single molecule as it travels from the point of its last collision at the origin (lower left corner). The length of the arrow (which you may recognize as a vector) is proportional to its velocity. The three components of the molecule's velocity are indicated by the small green arrows. It should be clearly apparent the trajectory is mainly along the *x*,*z*axis. In the section that follows, Equation 6.5.8 contains another 1/3 factor that similarly divides the kinetic energy into components along the three axes. This makes sense because kinetic energy is partly determined by velocity.

The temperature of a gas is a measure of the average translational kinetic energy of its molecules, so we begin by calculating the latter. Recalling that $mv^2/2$ is the average translational kinetic energy ε, we can rewrite the Equation ??? as

$$PV = \frac{1}{3}Nm\bar{v^2} = \frac{2}{3}N\epsilon \tag{6.5.8}$$

The 2/3 factor in the proportionality reflects the fact that velocity components in each of the three directions contributes ½ *kT* to the kinetic energy of the particle. The average translational kinetic energy is directly proportional to temperature:

$$\epsilon = \frac{3}{2}kT \tag{6.5.9}$$

in which the proportionality constant *k* is known as the *Boltzmann constant*. Substituting this into Equation 6.5.8 yields

$$PV = \left(\frac{2}{3}N\right)\left(\frac{3}{2}kT\right) = NkT \tag{6.5.10}$$

Notice that Equation ??? looks very much like the ideal gas equation

$$PV = nRT$$

but is not quite the same, however; we have been using capital N to denote the number of *molecules*, whereas n stands for the

number of *moles*. And of course, the proportionality factor is not the gas constant R, but rather the Boltzmann constant, 1.381 × 10^{-23} J K^{-1}. If we multiply \(*k*\) by Avogadro's number (N_A

\[(1.381 \times 10^{–23}{\, J \,K^{–1}) (6.022 \times 10^{23}) = 8.314 \,J \,K^{–1}.\]

Hence, the Boltzmann constant k is just the gas constant per molecule. So for *n moles* of particles, the Equtation ??? turns into our old friend

$$PV = nRT \tag{6.5.11}$$

The ideal gas equation of state came about by combining the empirically determined laws of Boyle, Charles, and Avogadro, but one of the triumphs of the kinetic molecular theory was the derivation of this equation from simple mechanics in the late nineteenth century. This is a beautiful example of how the principles of elementary mechanics can be applied to a simple model to develop a useful description of the behavior of macroscopic matter, and it will be worth your effort to follow and understand the individual steps of the derivation. (But don't bother to memorize it!)

RT has the dimensions of energy

Since the product PV has the dimensions of energy, so does *RT*, and this quantity in fact represents the average translational kinetic energy per mole of molecular particles. The relationship between these two energy units can be obtained by recalling that 1 atm is $1.013 \times 10^5\ N\ m^{-2}$, so that

Extra close brace or missing open brace

The *gas constant* \(*R*\) is one of the most important fundamental constants relating to the macroscopic behavior of matter. It is commonly expressed in both pressure-volume and in energy units:

$$R = 0.082057 \text{ L atm mol}^{-1} \text{ K}^{-1} = 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$$

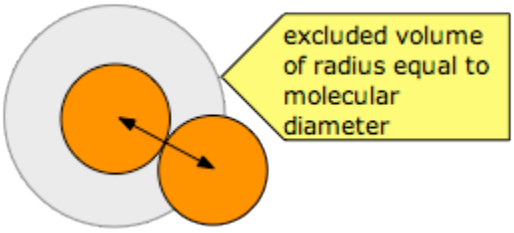
That is, *R* expresses the amount of energy per Kelvin degree. As noted above, the *Boltzmann constant k*, which appears in many expressions relating to the statistical treatment of molecules, is just

R ÷ 6.02E23 = 1.3807 × 10^{-23} J K^{-1},

the "gas constant per molecule "

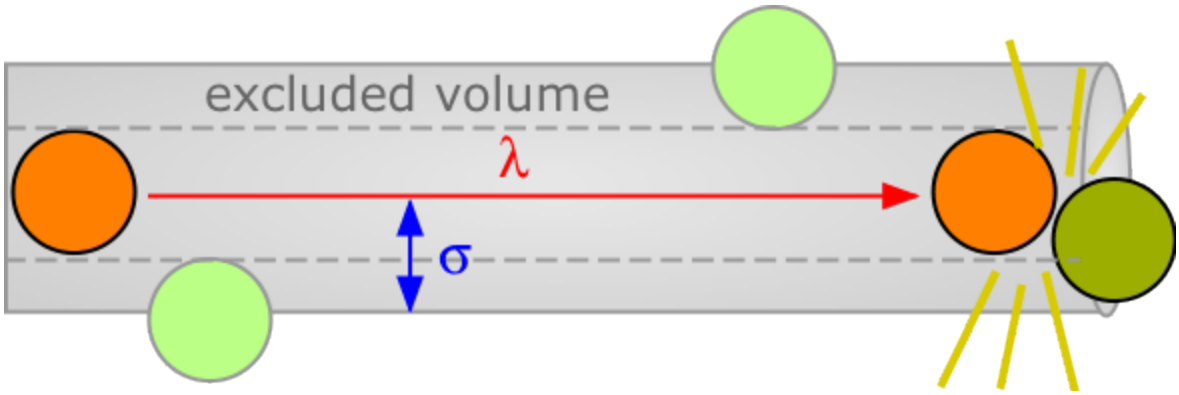
How Far does a Molecule travel between Collisions?

Molecular velocities tend to be very high by our everyday standards (typically around 500 meters per sec), but even in gases, they bump into each other so frequently that their paths are continually being deflected in a random manner, so that the net movement (diffusion) of a molecule from one location to another occurs rather slowly. How close can two molecules get?



How close can two molecules get? Each molecule is surrounded by an imaginary sphere (gray circle) whose radius σ is equal to the sum of the radii of the colliding molecules. This sphere defines the excluded volume, within which the center of another molecule cannot enter.

The average distance a molecule moves between such collisions is called the **mean free path** (λ), which depends on the number of molecules per unit volume and on their size. To avoid collision, a molecule of diameter σ must trace out a path corresponding to the axis of an imaginary cylinder whose cross-section is $\pi\sigma^2$. Eventually it will encounter another molecule (extreme right in the diagram below) that has intruded into this cylinder and defines the terminus of its free motion.



The volume of the cylinder is $\pi\sigma^2\lambda$. At each collision the molecule is diverted to a new path and traces out a new exclusion cylinder. After colliding with all *n* molecules in one cubic centimeter of the gas it will have traced out a total exclusion volume of $\pi\sigma^2\lambda$. Solving for λ and applying a correction factor $\sqrt{2}$ to take into account exchange of momentum between the colliding molecules (the detailed argument for this is too complicated to go into here), we obtain

$$\lambda = \frac{1}{\sqrt{2}\pi n\sigma^2} \tag{6.5.12}$$

Small molecules such as He, H_2 and CH_4 typically have diameters of around 30-50 pm. At STP the value of n, the number of molecules per cubic meter, is

$$\frac{6.022 \times 10^{23}\ mol^{-1}}{22.4 \times 10^{-3} m^3\ mol^{-1}} = 2.69 \times 10\ m^{-3} \tag{6.5.13}$$

Substitution into Equation 6.5.12 yields a value of around \(10^{-7}\; m (100\; nm)\) for the mean free path of most molecules under these conditions. Although this may seem like a very small distance, it typically amounts to 100 molecular diameters, and more importantly, about 30 times the average distance between molecules. This explains why so many gases conform very closely to the ideal gas law at ordinary temperatures and pressures.

On the other hand, at each collision the molecule can be expected to change direction. Because these changes are random, the net change in location a molecule experiences during a period of one second is typically rather small. Thus in spite of the high molecular velocities, the speed of molecular *diffusion* in a gas is usually quite small.

This page titled 6.5: More on Kinetic Molecular Theory is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

6.6: Real Gases and Critical Phenomena

Learning Objectives

Make sure you thoroughly understand the following essential concepts that have been presented above.

- Real gases are subject to the effects of molecular volume (intermolecular repulsive force) and intermolecular attractive forces.
- The behavior of a real gas approximates that of an ideal gas as the pressure approaches zero.
- The effects of non-ideal behavior are best seen when the *PV* product is plotted as a function of *P*. You should be able to identify the regions of such a plot in which attractive and repulsive forces dominate.
- Each real gas has its own unique equation of state. Various general equations of state have been devised in which adjustable constants are used to approximate the behavior of a particular gas.
- The most well-known equation of state is that of van der Waals. Although you need not memorize this equation, you should be able to explain the significance of its terms.

The "ideal gas laws" as we know them do a remarkably good job of describing the behavior of a huge number chemically diverse substances as they exist in the gaseous state under ordinary environmental conditions, roughly around 1 atm pressure and a temperature of 300 K. But when the temperature is reduced, or the pressure is raised, the relation PV = constant that defines the ideal gas begins to break down, and its properties become unpredictable; eventually the gas condenses into a liquid. Why is this important? It is of obvious interest to a chemical engineer who needs to predict the properties of the gases involved in a chemical reaction carried out at several hundred atmospheres pressure. This is especially so when we consider that some of the basic tenets of the ideal gas model have to be abandoned in order to explain such properties as

- the average distance between collisions (the molecules really *do* take up space!)
- at sufficiently high pressures and low temperatures, intermolecular attractions assume control and the gas condenses to a liquid;
- the viscosity of a gas flowing through a pipe (the molecules *do* get temporarily "stuck" on the pipe surface, and are therefore affected by intermolecular attractive forces.)

Even so, many of the common laws such as Boyle's and Charles' continue to describe these gases quite well even under conditions where these phenomena are evident. Under ordinary environmental conditions (moderate pressures and above 0°C), the isotherms of substances we normally think of as gases don't appear to differ very greatly from the hyperbolic form

$$\frac{PV}{RT} = \text{constant} \tag{6.6.1}$$

However, over a wider range of conditions, things begin to get more complicated. Thus isopentane (Figure $6.6.1$) behaves in a reasonably ideal manner above 210 K, but below this temperature the isotherms become somewhat distorted, and at 185 K and below they cease to be continuous, showing peculiar horizontal segments in which reducing the volume does not change the pressure.

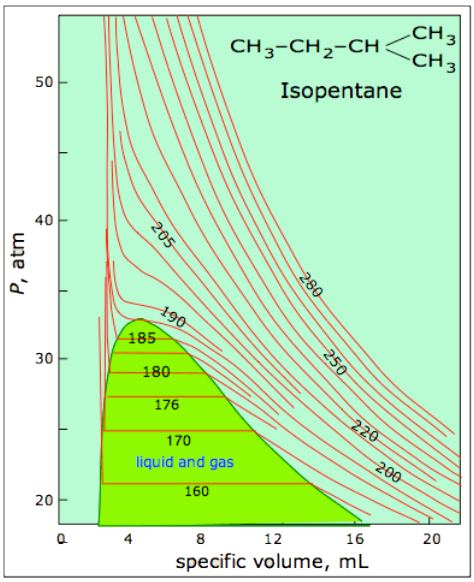


Figure 6.6.1: PV curves for the real gas Isopentane

Within this region, any attempt to compress the gas simply "squeezes" some of it into the liquid state whose greater density exactly compensates for the smaller volume, thus maintaining the pressure at a constant value. It turns out that real gases eventually begin to follow their own unique equations of state, and ultimately even cease to be gases. In this unit we will see why this occurs, what the consequences are, and how we might modify the ideal gas equation of state to extend its usefulness over a greater range of temperatures and pressures.

Effects of Intermolecular Forces

According to Boyle's law, the product *PV* is a constant at any given temperature, so a plot of *PV* as a function of the pressure of an ideal gas yields a horizontal straight line. This implies that any increase in the pressure of the gas is exactly counteracted by a decrease in the volume as the molecules are crowded closer together. But we know that the molecules themselves are finite objects having volumes of their own, and this must place a lower limit on the volume into which they can be squeezed. So we must reformulate the ideal gas equation of state as a relation that is true only in the limiting case of zero pressure:

$$\lim_{P \to 0} PV = nRT \tag{6.6.2}$$

So what happens when a real gas is subjected to a very high pressure? The outcome varies with both the molar mass of the gas and its temperature, but in general we can see the the effects of both repulsive and attractive intermolecular forces:

- **Repulsive forces**: As a gas is compressed, the individual molecules begin to get in each other's way, giving rise to a very strong repulsive force acts to oppose any further volume decrease. We would therefore expect the PV vs P line to curve upward at high pressures, and this is in fact what is observed for all gases at sufficiently high pressures.
- **Attractive forces**: At very close distances, all molecules repel each other as their electron clouds come into contact. At greater distances, however, brief statistical fluctuations in the distribution these electron clouds give rise to a universal attractive force between all molecules. The more electrons in the molecule (and thus the greater the molecular weight), the greater is this attractive force. As long as the energy of thermal motion dominates this attractive force, the substance remains in the gaseous state, but at sufficiently low temperatures the attractions dominate and the substance condenses to a liquid or solid.

The universal attractive force described above is known as the *dispersion*, or *London* force. There may also be additional (and usually stronger) attractive forces related to charge imbalance in the molecule or to hydrogen bonding. These various attractive forces are often referred to collectively as *van der Waals forces*. A plot of *PV/RT* as a function of pressure is a very sensitive indicator of deviations from ideal behavior, since such a plot is just a horizontal line for an ideal gas. Figures 6.6.2 and 6.6.3 demonstrate how these plots vary with the nature of the gas and with temperature, respectively.

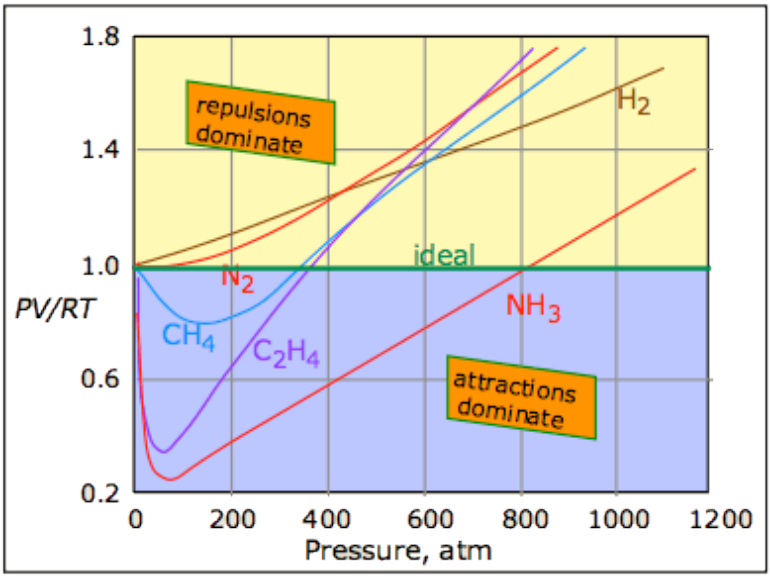


Figure 6.6.2

Intermolecular attractions, which generally increase with molecular weight, cause the *PV* product to decrease as higher pressures bring the molecules closer together and thus within the range of these attractive forces; the effect is to cause the volume to decrease more rapidly than it otherwise would. The repulsive forces always eventually win out. However, as the molecules begin to intrude on each others' territory, the stronger repulsive forces cause the curve to bend upward.

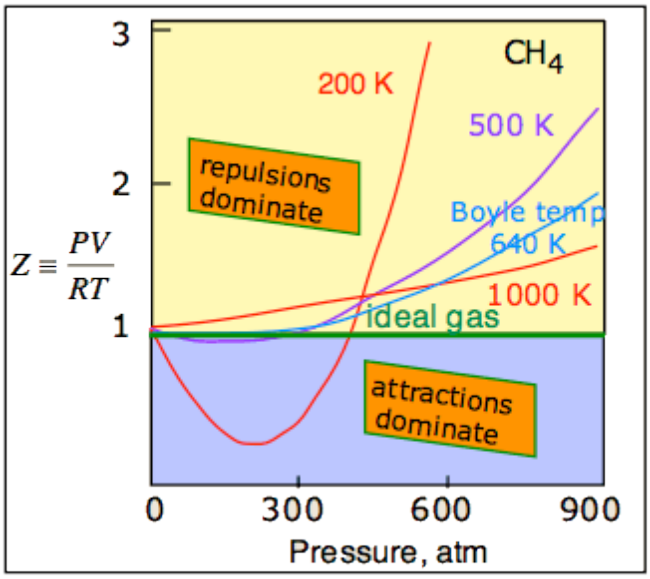


Figure 6.6.3: Temperature dependent compressibility vs. pressure curves for methane (CH_4).

The **temperature** makes a big difference! At higher temperatures, increased thermal motions overcome the effects of intermolecular attractions which normally dominate at lower pressures (Figure 6.6.3). So all gases behave more ideally at higher temperatures. For any gas, there is a special temperature (the **Boyle temperature**) at which attractive and repulsive forces exactly balance each other at zero pressure. As you can see in this plot for methane, some of this balance does remain as the pressure is increased.

The van der Waals Equation of State

How might we modify the ideal gas equation of state to take into account the effects of intermolecular interactions? The first and most well known answer to this question was offered by the Dutch scientist J.D. van der Waals (1837-1923) in 1873. The ideal gas model assumes that the gas molecules are merely points that occupy no volume; the "*V*" term in the equation is the volume of the

container and is independent of the nature of the gas.

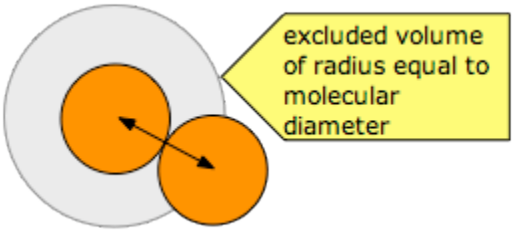


Figure 6.6.4

van der Waals recognized that the molecules themselves take up space that subtracts from the volume of the container (Figure 6.6.4), so that the "volume of the gas" *V* in the ideal gas equation should be replaced by the term $(V–b)$, in which b relates to the *excluded volume*, typically of the order of 20-100 cm^3 mol^{-1}. The excluded volume surrounding any molecule defines the closest possible approach of any two molecules during collision. Note that the excluded volume is greater then the volume of the molecule, its radius being half again as great as that of a spherical molecule.

The other effect that van der Waals needed to correct for are the intermolecular attractive forces. These are ignored in the ideal gas model, but in real gases they exert a small cohesive force between the molecules, thus helping to hold the gas together and reducing the pressure it exerts on the walls of the container.

Because this pressure depends on both the frequency *and* the intensity of collisions with the walls, the reduction in pressure is proportional to the *square* of the number of molecules per volume of space, and thus for a fixed number of molecules such as one mole, the reduction in pressure is inversely proportional to the square of the volume of the gas. The smaller the volume, the closer are the molecules and the greater will be the effect. The van der Walls equation replaces the P term in the ideal gas equation with $P + (a/V^2)$ in which the magnitude of the constant a increases with the strength of the intermolecular attractive forces.

The complete van der Waals equation of state can be written as

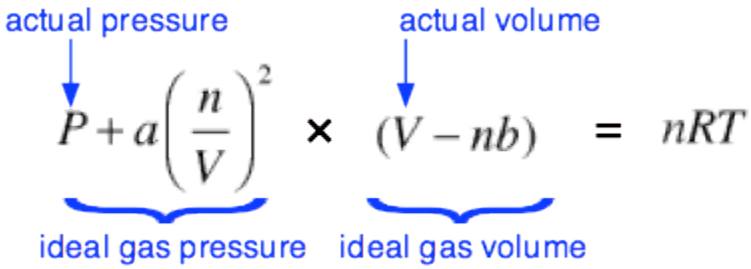


Figure 6.6.5: van der Waal's equation of state for a real gas

Although most students are not required to memorize this equation, you are expected to understand it and to explain the significance of the terms it contains. You should also understand that the van der Waals constants a and b must be determined empirically for every gas. This can be done by plotting the *P*-*V* behavior of the gas and adjusting the values of a and b until the van der Waals equation results in an identical plot. The constant a is related in a simple way to the molecular radius; thus the determination of a constitutes an indirect measurement of an important microscopic quantity.

Table 6.6.1: van der Waals constants for some gases

Substance	molar mass (g)	a (L^2-atm $mole^{-2}$)	b (L mol^{-1})
hydrogen H_2	2	0.244	0.0266
helium He	4	0.034	0.0237
methane CH_4	16	2.25	0.0428
water H_2O	18	5.46	0.0305
nitrogen N_2	28	1.39	0.0391

Substance	molar mass (g)	a (L^2-atm $mole^{-2}$)	b (L mol^{-1})
carbon dioxide CO_2	44	3.59	0.0427
carbon tetrachloride CCl_4	154	20.4	0.1383

The van der Waals equation is only one of many equations of state for real gases. More elaborate equations are required to describe the behavior of gases over wider pressure ranges. These generally take account of higher-order nonlinear attractive forces, and require the use of more empirical constants. Although we will make no use of them in this course, they are widely employed in chemical engineering work in which the behavior of gases at high pressures must be accurately predicted.

Condensation and the Critical Point

The most striking feature of real gases is that they cease to remain gases as the temperature is lowered and the pressure is increased. Figure 6.6.6 illustrates this behavior; as the volume is decreased, the lower-temperature isotherms suddenly change into straight lines. Under these conditions, the pressure remains constant as the volume is reduced. This can only mean that the gas is "disappearing" as we squeeze the system down to a smaller volume. In its place, we obtain a new state of matter, the liquid. In the green-shaded region, *two* phases, liquid, and gas, are simultaneously present. Finally, at very small volume all the gas has disappeared and only the liquid phase remains. At this point the isotherms bend strongly upward, reflecting our common experience that a liquid is practically incompressible.

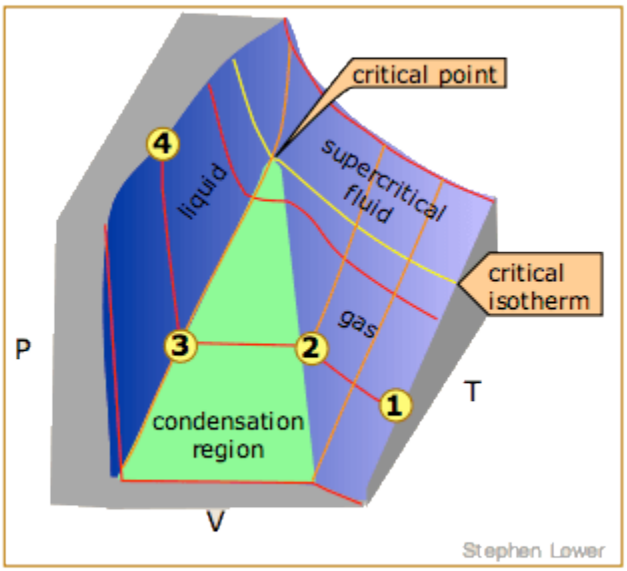


Figure 6.6.6

To better understand this plot, look at the isotherm labeled 1. As the gas is compressed from 1 to 2, the pressure rises in much the same way as Boyle's law predicts. Compression beyond 2, however, does not cause any rise in the pressure. What happens instead is that some of the gas condenses to a liquid. At 3, the substance is entirely in its liquid state. The very steep rise to 4 corresponds to our ordinary experience that liquids have very low compressibilities. The range of volumes possible for the liquid diminishes as the critical temperature is approached.

The Critical Point

Liquid and gas can coexist only within the regions indicated by the green-shaded area in the diagram above. As the temperature and pressure rise, this region becomes more narrow, finally reaching zero width at the *critical point*. The values of *P, T*, and *V* at this juncture are known as the *critical constants* P_c, T_c, and V_c. The isotherm that passes through the critical point is called the *critical isotherm*. Beyond this isotherm, the gas and liquids become indistinguishable; there is only a single fluid phase,

sometimes referred to as a *supercritical liquid* (Figure 6.6.7).

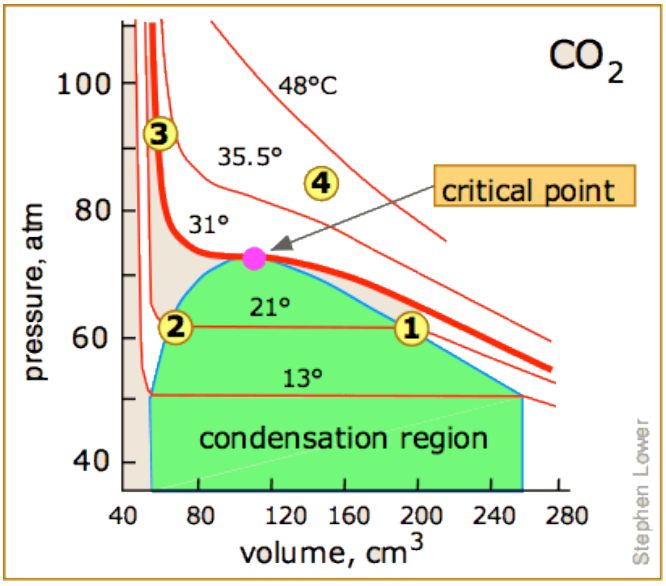


Figure 6.6.7

At temperatures below 31°C (the *critical temperature*), CO_2 acts somewhat like an ideal gas even at a rather high pressure (4). Below 31°, an attempt to compress the gas to a smaller volume eventually causes condensation to begin. Thus at 21°C, at a pressure of about 62 atm (1), the volume can be reduced from 200 cm^3 to about 55 cm^3 without any further rise in the pressure. Instead of the gas being compressed, it is replaced with the far more compact liquid as the gas is essentially being "squeezed" into its liquid phase. After all of the gas has disappeared (2), the pressure rises very rapidly because now all that remains is an almost incompressible liquid. Above this isotherm (3), CO_2 exists only as a *supercritical fluid*.

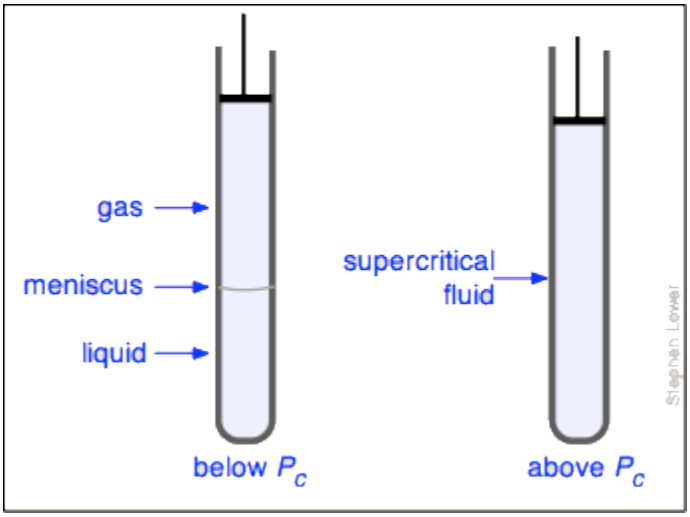


Figure 6.6.8

What happens if you have some liquid carbon dioxide in a transparent cylinder at just under its P_c of 62 atm, and you then compress it slightly? Nothing very dramatic until you notice that the meniscus has disappeared. By successively reducing and increasing the pressure, you can "turn the meniscus on and off".

One intriguing consequence of the very limited bounds of the liquid state is that you could start with a gas at large volume and low temperature, raise the temperature, reduce the volume, and then reduce the temperature so as to arrive at the liquid region at the lower left, without ever passing through the two-phase region, and thus *without undergoing condensation!*

Supercritical fluids

The *supercritical* state of matter, as the fluid above the critical point is often called, possesses the flow properties of a gas and the solvent properties of a liquid. The density of a supercritical fluid can be changed over a wide range by adjusting the pressure; this, in turn, changes its solubility, which can thus be optimized for a particular application. The picture at the right shows a commercial laboratory device used for carrying out chemical reactions under supercritical conditions.

Supercritical carbon dioxide is widely used to dissolve the caffeine out of coffee beans and as a dry-cleaning solvent. Supercritical water has recently attracted interest as a medium for chemically decomposing dangerous environmental pollutants such as PCBs. Supercritical fluids are being increasingly employed as as substitutes for organic solvents (so-called "green chemistry") in a range of industrial and laboratory processes. Applications that involve supercritical fluids include extractions, nano particle and nano structured film formation, supercritical drying, carbon capture and storage, as well as enhanced oil recovery studies.

This page titled 6.6: Real Gases and Critical Phenomena is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7: SOLIDS AND LIQUIDS

The pages present an overview of the condensed states of matter. Although there is more detail than can be found in standard textbooks, the level is still suitable for first-year college and advanced high school courses. These pages should also be helpful as review material for students in more advanced courses in chemistry, geology, and materials science.

7.1: Matter under the Microscope

7.2: Intermolecular Interactions

7.3: Hydrogen-Bonding and Water

7.4: Liquids and their Interfaces

7.5: Changes of State

7.6: Introduction to Crystals

7.7: Ionic and Ion-Derived Solids

7.8: Cubic Lattices and Close Packing

7.9: Polymers and Plastics

7.10: Colloids and their Uses

Index

This page titled 7: Solids and Liquids is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Matter under the Microscope

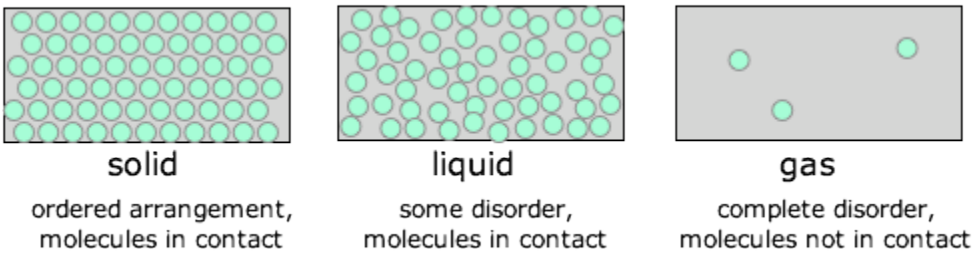
Learning Objectives

- State the major feature that characterizes a *condensed state* of matter.
- Describe some of the **major observable properties** that distinguish gases, liquid and solids, and state their relative magnitudes in these three states of matter.
- Describe the **dominant forces** and the resulting **physical properties** that distinguish ionic, covalent, metallic, and molecular solids.
- Explain the difference between *crystalline* and *amorphous* solids, and cite some examples of each.
- Name some of the basic **molecular units** from which solids of different type can be composed.
- What is meant by an "*extended*" or "infinite-molecule solid"?
- Describe some of the special properties of *graphite* and their structural basis.

Gases, liquids, and especially solids surround us and give form to our world. Chemistry at its most fundamental level is about atoms and the forces that act between them to form larger structural units. But the matter that we experience with our senses is far removed from this level. This unit will help you see how these *macroscopic* properties of matter depend on the *microscopic* particles of which it is composed.

Solids, Liquids and Gases

What distinguishes solids, liquids, and gases– the three major *states of matter*— from each other? Let us begin at the microscopic level, by reviewing what we know about gases, the simplest state in which matter can exist. At ordinary pressures, the molecules of a gas are so far apart that intermolecular forces have an insignificant effect on the random thermal motions of the individual particles. As the temperature decreases and the pressure increases, intermolecular attractions become more important, and there will be an increasing tendency for molecules to form temporary clusters. These are so short-lived, however, that even under extreme conditions, gases cannot be said to possess “structure” in the usual sense.



The contrast at the *microscopic* level between solids, liquids and gases is most clearly seen in the simplified schematic views above. The molecular units of crystalline solids tend to be highly ordered, with each unit occupying a fixed position with respect to the others. In liquids, the molecules are able to slip around each other, introducing an element of disorder and creating some void spaces that decrease the density. Gases present a picture of almost total disorder, with practically no restrictions on where any one molecule can be.

Having lived our lives in a world composed of solids, liquids, and gases, few of us ever have any difficulty deciding into which of these categories a given sample of matter falls. Our decision is most commonly based on purely visual cues:

- a gas has no definite boundaries other than those that might be imposed by the walls of a confining vessel.
- Liquids and solids possess a clearly delineated phase boundary that gives solids their definite shapes and whose light-reflecting properties enable us to distinguish one phase from another.
- Solids can have any conceivable shape, and their surfaces are usually too irregular to show *specular* (mirror-like) reflection of light. Liquids, on the other hand, are *mobile*; except when in the form of tiny droplets, liquids have no inherent shape of their own, but assume the shape of their container and show an approximately flat upper surface.

Our experience also tells us that these categories are quite distinct; a *phase*, which you will recall is a region of matter having uniform intensive properties, is either a gas, a liquid, or a solid. Thus the three states of matter are not simply three points on a continuum; when an ordinary solid melts, it usually does so at a definite temperature, without apparently passing through any states that are intermediate between a solid and a liquid.

Limiting Behavior

Although these common-sense perceptions are usually correct, they are not infallible, and in fact there are solids such as glasses and many plastics that do not have sharp melting points, but instead undergo a gradual transition from solid to liquid known as *softening*, and when subject to enough pressure, solids can exhibit something of the flow properties of liquids (glacial ice, for example).



Figure 7.1.1: Some solids can flow — slowly. Kluane Glacier in Canada's Yukon Territory. Typical glacial flow rates are 10-200 m/yr.

A more scientific approach would be to compare the macroscopic physical properties of the three states of matter, but even here we run into difficulty. It is true, for example, that the density of a gas is usually about a thousandth of that of the liquid or solid at the same temperature and pressure; thus one gram of water vapor at 100°C and 1 atm pressure occupies a volume of 1671 mL; when it condenses to liquid water at the same temperature, it occupies only 1.043 mL.

Table 7.1.1: Comparison of the molar volumes of neon in its three states.

Phase	Density
gas	22,400 cm^3/mol total volume (42 cm^3/mol excluded volume)
liquid	16.8 cm^3/mol
solid	13.9 cm^3/mol

Table 7.1.1 compares the molar volumes of **neon** in its three states. For the gaseous state, P = 1 atm and T = 0°C. The excluded volume is the volume actually taken up by the neon atoms according to the van der Waals equation of state model. It is this extreme contrast with the gaseous states that leads to the appellation "*condensed states of matter*" for liquids and solids. However, gases at very high pressures can have densities that exceed those of other solid and liquid substances, so density alone is not a sufficiently comprehensive criterion for distinguishing between the gaseous and condensed states of matter. Similarly, the density of a *solid* is usually greater than that of the corresponding liquid at the same temperature and pressure, but not always: you have certainly seen ice floating on water!

Example 7.1.1: Density of Xenon Gas

Compare the density of gaseous xenon (molar mass 131 g) at 100 atm and 0°C with that of a hydrocarbon liquid for which $\rho = 0.104\ g/mL$ at the same temperature.

Solution

For simplicity, we will assume that xenon approximates an ideal gas under these conditions, which it really does not.

The ideal molar volume at 0° C and 1 atm is 22.4 L; at 100 atm, this would be reduced to 0.22 L or 220 mL, giving a density

$$\rho = \frac{131\,g}{224\,mL} = 0.58\,g/mL.$$

In his autobiographical *Uncle Tungsten*, the physician/author Oliver Sacks describes his experience with xenon-filled balloons of "*astonishing density — as near to 'lead balloons" as could be* [imagined]. *If one twirled these xenon balloons in one's hand, then stopped, the heavy gas, by its own momentum, would continue rotating for a minute, almost as if it were a liquid.*"

Other physical properties, such as the compressibility, surface tension, and viscosity, are somewhat more useful for distinguishing between the different states of matter. Even these, however, provide no well-defined dividing lines between the various states. Rather than try to develop a strict scheme for classifying the three states of matter, it will be more useful to simply present a few generalizations.

Table 7.1.2: Relative magnitudes of some properties of the three states of matter

property	gas	liquid	solid
density	very small	large	large
thermal expansion coefficient	large (= R/P)	small	small
cohesiveness	nil	small	large
surface tension	nil	medium	very large
viscosity	small	medium	very large
kinetic energy per molecule	large	small	smaller
disorder	random	medium	small

Condensed States of Matter

Some of these deal with macroscopic properties (that is, properties such as the *density* that relate to *bulk matter*), and others with microscopic properties that refer to the individual molecular units. Even the most casual inspection of the above table shows that solids and liquids possess an important commonality that distinguishes them from gases: in solids and liquids, the molecules are in contact with their neighbors. As a consequence, these *condensed states of matter* generally possess much higher *densities* than gases.

In our study of gases, we showed that the macroscopic properties of a gas (the pressure, volume, and temperature) are related through an ***equation of state***, and that for the limiting case of an ideal gas, this equation of state can be derived from the relatively small set of assumptions of the kinetic molecular theory. To the extent that a volume of gas consists mostly of empty space, all gases have very similar properties. Equations of state work for gases because gases consist mostly of empty space, so intermolecular interactions can be largely neglected. In condensed matter, these interactions dominate, and they tend to be unique to each particular substance, so there is no such thing as a generally useful equation of state of liquids and solids.

Is there a somewhat more elaborate theory that can predict the behavior of the other two principal states of matter, liquids and solids? Very simply, the answer is "no"; despite much effort, no one has yet been able to derive a general equation of state for condensed states of matter. The best one can do is to construct models based on the imagined interplay of attractive and

repulsive forces, and then test these models by computer simulation. Nevertheless, the very factors that would seem to make an equation of state for liquids and solids impossibly complicated also give rise to new effects that are easily observed, and which ultimately define the distinguishing characteristics of the gaseous, liquid, and solid states of matter. In this unit, we will try to learn something about these distinctions, and how they are affected by the chemical constitution of a substance.

Liquids

Crystalline solids and gases stand at the two extremes of the spectrum of perfect order and complete chaos. Liquids display elements of both qualities, and both in limited and imperfect ways. Liquids and solids share most of the properties of having their molecular units in direct contact as discussed in the previous section on condensed states of matter. At they same time, liquids, like gases, are fluids, meaning that their molecular units can move more or less independently of each other. But whereas the *volume* of a gas depends entirely on the pressure (and thus generally on the volume within which it is confined), the volume of a liquid is largely independent of the pressure. Here we offer just enough to help you see how they relate to the other major states of matter.

Solids

Of the four ancient elements of "fire, air, earth and water", it is the many forms of solids ("earths") that we encounter in daily life and which give form, color and variety to our visual world. The solid state, being the form of any substance that prevails at lower temperatures, is one in which thermal motion plays an even smaller role than in liquids. The thermal kinetic energy that the individual molecular units do have at temperatures below their melting points allows them to oscillate around a fixed center whose location is determined by the balance between local forces of attraction and repulsion due to neighboring units, but only very rarely will a molecule jump out of the fixed space allotted to it in the lattice. Thus solids, unlike liquids, exhibit *long-range order*, *cohesiveness,* and *rigidity*, and possess definite *shapes*.

Classification of solids

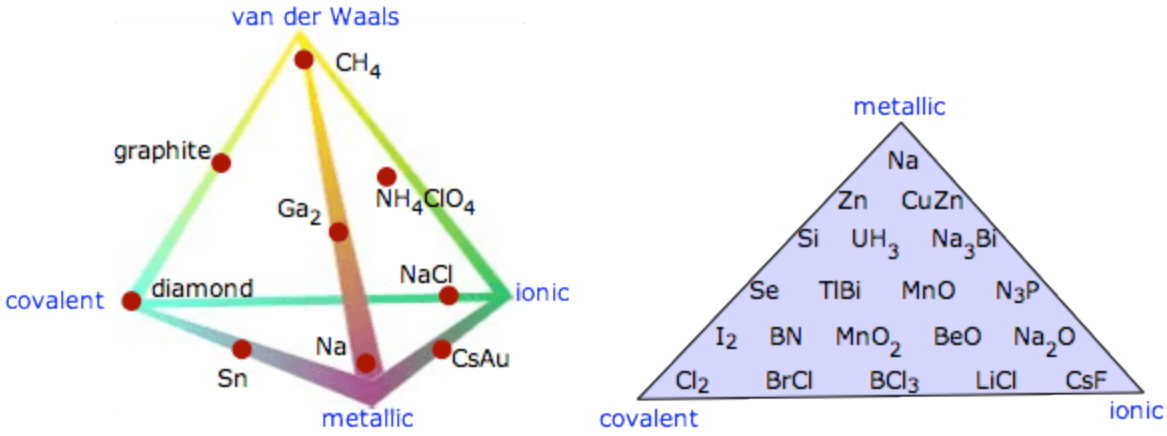
Most people who have lived in the world long enough to read this have already developed a rough way of categorizing solids on the basis of macroscopic properties they can easily observe; everyone knows that a piece of metal is fundamentally different from a rock or a chunk of wood. Unfortunately, nature's ingenuity is far too versatile to fit into any simple system of classifying solids, especially those composed of more than a single chemical substance.

Classification Scheme 1: According to bond type

The most commonly used classification is based on the kinds of forces that join the molecular units of a solid together. We can usually distinguish four major categories on the basis of properties such as general appearance, hardness, and melting point.

type of solid	molecular units	dominant forces	typical properties
ionic	ions	coulombic	high-melting, hard, brittle
covalent	atoms of electronegative elements	chemical bonds	non-melting (decompose), extremely hard
metallic	atoms of electropositive elements	mobile electrons	moderate-to-high melting, deformable, conductive, metallic luster
molecular	molecules	van der Waals	low-to-moderate mp, low hardness

It's important to understand that these four categories are in a sense idealizations that fail to reflect the diversity found in nature. The triangular diagram shown here illustrates this very nicely by displaying examples of binary compounds whose properties suggest that they fall somewhere other than at a vertex of the triangle.



The triangle shown above excludes what is probably the largest category: molecular solids that are bound by van der Waals forces. One way of including these is to expand the triangle to a tetrahedron (the so-called **Laing tetrahedron**). Although this illustrates the concept, it is visually awkward to include many examples of the intermediate cases.

Classification Scheme 2: By type of Molecular Unit

Solids, like the other states of matter, can be classified according to whether their fundamental molecular units are atoms, electrically-neutral molecules, or ions. But solids possess an additional property that gases and liquids do not: an enduring structural arrangement of their molecular units. Over-simplifying only a bit, we can draw up a rough classification of solids according to the following scheme:

structure	atoms	molecules
array of discrete units	noble gas solids, metals	molecular solids
array of linked units	metals and covalent solids	"extended molecule" compounds
disordered arrangement	alternative forms of some elements (e.g. S, Se)	polymers, glasses

Classification Scheme 3: Classification by Dominant Attractive Force

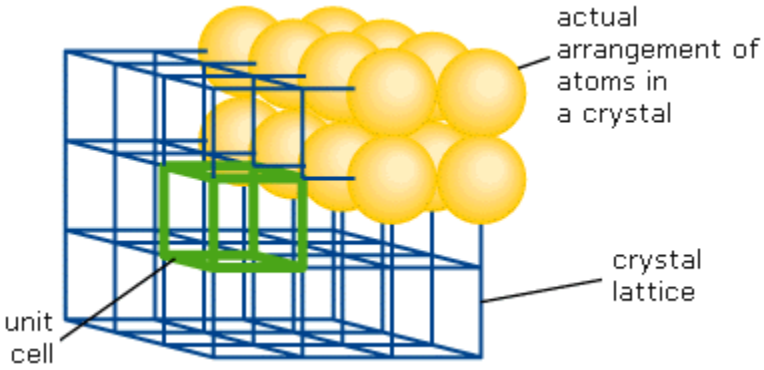
Notice how the boiling points in the following selected examples reflect the major type of attractive force that binds the molecular units together. Bear in mind, however, that more than one type of attractive force can be operative in many substances.

substance	bp °C	molecular units	dominant attractive force	separation distance (pm)	attraction energy (kJ/mol)
sodium fluoride	990	$Na^+ F^-$	coulombic	18.8	657
sodium hydroxide	318	$Na^+ OH^-$	ion-dipole	21.4	90.4
water	100	H_2O	dipole-dipole	23.7	20.2
neon	249	Ne	dispersion	33.0	0.26

Crystalline Solids

In a solid comprised of identical molecular units, the most favored (lowest potential energy) locations occur at regular intervals in space. If each of these locations is actually occupied, the solid is known as a *perfect* crystal. What really defines a crystalline solid is that its structure is composed of repeating *unit cells* each containing a small number of molecular units bearing a fixed geometric relation to one another. The resulting long-range order defines a three-dimensional geometric framework known as a

lattice.



Geometric theory shows that only fourteen different types of lattices are possible in three dimensions, and that just six different unit cell arrangements can generate these lattices. The regularity of the external faces of crystals, which in fact correspond to lattice planes, reflects the long-range order inherent in the underlying structure. Perfection is no more attainable in a crystal than in anything else; real crystals contain *defects* of various kinds, such as lattice positions that are either vacant or occupied by impurities, or by abrupt displacements or dislocations of the lattice structure. Most pure substances, including the metallic elements, form crystalline solids. But there are some important exceptions.

Metallic Solids

In metals the valence electrons are free to wander throughout the solid, instead of being localized on one atom and shared with a neighboring one. The valence electrons behave very much like a mobile fluid in which the fixed lattice of atoms is immersed. This provides the ultimate in electron sharing, and creates a very strong binding effect in solids composed of elements that have the requisite number of electrons in their valence shells. The characteristic physical properties of metals such as their ability to bend and deform without breaking, their high thermal and electrical conductivities and their metallic sheen are all due to the fluid-like behavior of the valence electrons.

Molecular solids

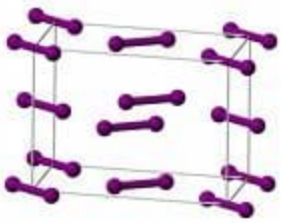
Recall that a "molecule" is defined as a discrete aggregate of atoms bound together sufficiently tightly (that is, by *directed covalent* forces) to allow it to retain its individuality when the substance is dissolved, melted, or vaporized.

The two words italicized in the preceding sentence are important; *covalent bonding* implies that the forces acting between atoms *within* the molecule are much stronger than those acting *between* molecules, and the *directional property* of covalent bonding confers on each molecule a distinctive shape which affects a number of its properties. Most compounds of carbon — and therefore, most chemical substances, fall into this category.

Many simpler compounds also form molecules; H_2O, NH_3, CO_2, and PCl_5 are familiar examples. Some of the elements, such as H_2, O_2, O_3, P_4 and S_8 also occur as discrete molecules. Liquids and solids that are composed of molecules are held together by *van der Waals forces*, and many of their properties reflect this weak binding. Thus molecular solids tend to be soft or deformable, have low melting points, and are often sufficiently volatile to evaporate (sublime) directly into the gas phase; the latter property often gives such solids a distinctive odor.

Iodine is a good example of a volatile molecular crystal. The solid (mp 114° C , bp 184°) consists of I_2 molecules bound together only by dispersion forces. If you have ever worked with solid iodine in the laboratory, you will probably recall the smell and sight of its purple vapor which is easily seen in a closed container.





Because dispersion forces and the other van der Waals forces increase with the number of atoms, larger molecules are generally less volatile, and have higher melting points, than do the smaller ones. Also, as one moves down a column in the periodic table, the outer electrons are more loosely bound to the nucleus, increasing the polarizibility of the atom and thus its susceptibility to van der Waals-type interactions. This effect is particularly apparent in the progression of the boiling points of the successively heavier noble gas elements.

Covalent Solids

These are a class of extended-lattice compounds (see Section 6 below) in which each atom is covalently bonded to its nearest neighbors. This means that the entire crystal is in effect one super-giant “molecule”. The extraordinarily strong binding forces that join all adjacent atoms account for the extreme hardness of such substances; these solids cannot be broken or abraded without cleaving a large number of covalent chemical bonds. Similarly, a covalent solid cannot “melt” in the usual sense, since the entire crystal is its own giant molecule. When heated to very high temperatures, these solids usually decompose into their elements.

Diamond

Diamond is the hardest material known, defining the upper end of the 1-10 scale known as **Mohs Hardness**. Diamond cannot be melted; above 1700°C it is converted to graphite the more stable form of carbon.

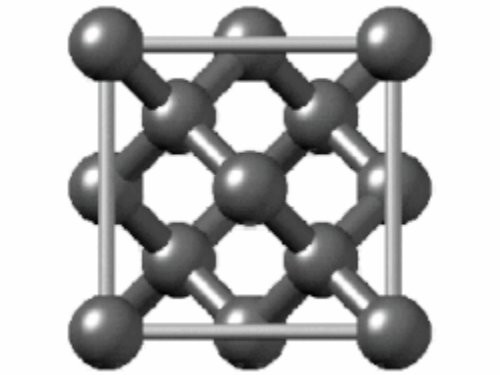


Figure Rotating model of the diamond cubic crystal structure. Image used wtih pmerission (CC0; MarinaVladivostok via Wikicommons)

The diamond unit cell is face-centered cubic and contains 8 carbon atoms. The four darkly shaded ones are contained within the cell and are completely bonded to other members of the cell. The other carbon atoms (6 in faces and 4 at corners) have some bonds that extend to atoms in other cells. (Two of the carbons nearest the viewer are shown as open circles in order to more clearly reveal the bonding arrangement.)

Other covalent solids

Boron nitride BN is similar to carbon in that it exists as a diamond-like cubic polymorph as well as in a hexagonal form analogous

to graphite. Cubic BN is the second hardest material after diamond, and finds use in industrial abrasives and cutting tools. Recent interest in BN has centered on its carbon-like ability to form nanotubes and related nanostructures.

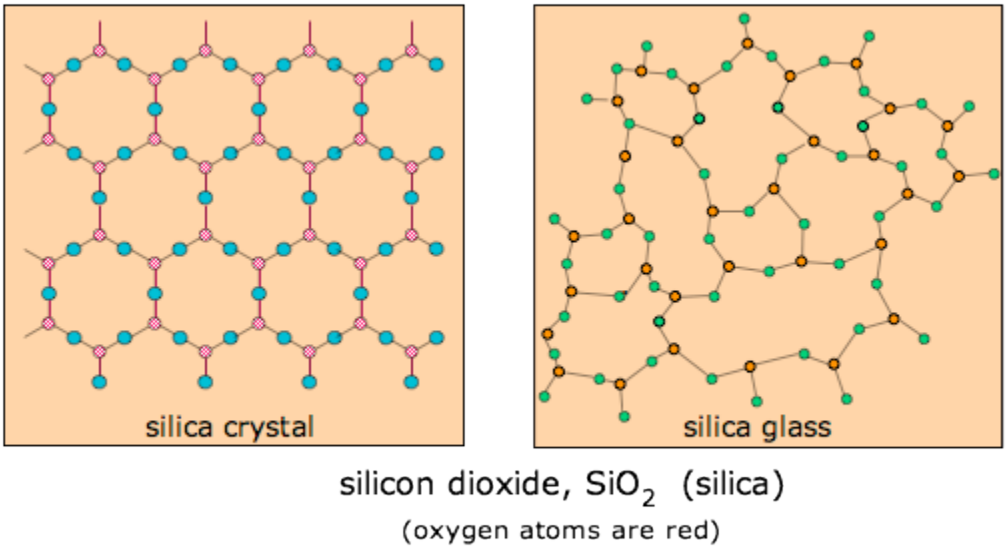
Silicon carbide SiC is also known as *carborundum*. Its structure is very much like that of diamond with every other carbon replaced by silicon. On heating at atmospheric pressure, it decomposes at 2700°C, but has never been observed to melt. Structurally, it is very complex; at least 70 crystalline forms have been identified. Its extreme hardness and ease of synthesis have led to a diversity of applications — in cutting tools and abrasives, high-temperature semiconductors, and other high-temperature applications, manufacture of specialty steels, jewelry, and many more. Silicon carbide is an extremely rare mineral on the earth, and comes mostly from meteorites which are believed to have their origins in carbonaceous stars. The first synthetic SiC was made accidently by E.G. Acheson in 1891 who immediately recognized its industrial prospects and founded the Carborundum Co.

Tungsten carbide WC is probably the most widely-encountered covalent solid owing to its use in "carbide" cutting tools and as the material used to make the rotating balls in ball-point pens. It's high-melting (2870°C) form has a structure similar to that of diamond and is only slightly less hard. In many of its applications it is embedded in a softer matrix of cobalt or coated with titanium compounds.

Amorphous Solids

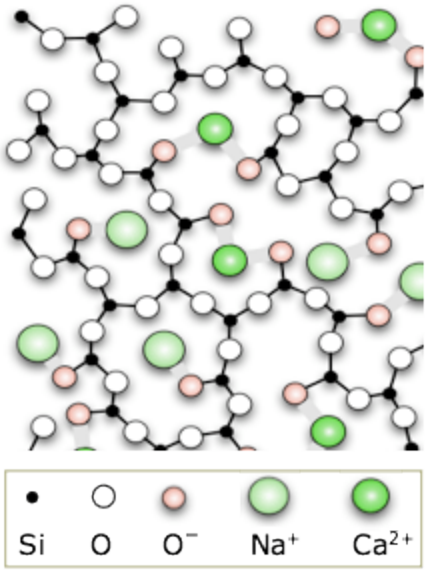
In some solids there is so little long-range order that the substance cannot be considered crystalline at all; such a solid is said to be *amorphous*. Amorphous solids possess short-range order but are devoid of any organized structure over longer distances; in this respect they resemble liquids. However, their rigidity and cohesiveness allow them to retain a definite shape, so for most practical purposes they can be considered to be solids.

Glasses refers generally to solids formed from their melts that do not return to their crystalline forms on cooling, but instead form hard, and often transparent amorphous solids. Although some organic substances such as sugar can form glasses ("rock candy"), the term more commonly describes inorganic compounds, especially those based on *silica*, SiO_2. Natural silica-based glasses, known as obsidian, are formed when certain volcanic magmas cool rapidly.



The difference between crystalline silica and silica glass is shown in these simplified two-dimensional projections. It is readily apparent that much of the ordered arrangement of crystalline silica is lost in the glassy form. But the glass retains enough Si–O bonds to form a hard, rigid material.

Ordinary glass is composed mostly of SiO_2, which usually exists in nature in a crystalline form known as *quartz*. If quartz (in the form of sand) is melted and allowed to cool, it becomes so viscous that the molecules are unable to move to the low potential energy positions they would occupy in the crystal lattice, so that the disorder present in the liquid gets “frozen into” the solid. In a sense, glass can be regarded as a supercooled liquid. Glasses are transparent because the distances over which disorder appears are small compared to the wavelength of visible light, so there is nothing to scatter the light and produce cloudiness.



Ordinary glass is made by melting silica sand to which has been added some calcium and sodium carbonates. These additives reduce the melting point and make it more difficult for the SiO2 molecules to arrange themselves into crystalline order as the mixture cools. Glass is believed to have first been made in the Middle East at least as early as 3000 BCE. Its workability and ease of coloring has made it one of mankind's most important and versatile materials.

Types of molecular units

Molecules

Molecules, not surprisingly, are the most common building blocks of pure substances. Most of the 15-million-plus chemical substances presently known exist as distinct molecules. Chemists commonly divide molecular compounds into "small" and "large-molecule" types, the latter usually falling into the class of *polymers* (see below.) The dividing line between the two categories is not very well defined, and tends to be based more on the properties of the substance and how it is isolated and purified.

Atoms

We usually think of atoms as the building blocks of molecules, so the only pure substances that consist of plain atoms are those of some of the elements — mostly the metallic elements, and also the noble-gas elements. The latter do form liquids and crystalline solids, but only at very low temperatures. Although the **metallic elements** form crystalline solids that are essentially atomic in nature, the special properties that give rise to their "metallic" nature puts them into a category of their own. Most of the non-metallic elements exist under ordinary conditions as small molecules such as O_2 or S_6, or as extended structures that can have a somewhat polymeric nature. Many of these elements can form more than one kind of structure, each one stable under different ranges of temperature and pressure. Multiple structures of the same element are known as *allotropes*, although the more general term *polymorph* is now preferred.

Ions

Ions, you will recall, are atoms or molecules that have one or more electrons missing (positive ions) or in excess (negative ions), and therefore possess an electric charge. A basic law of nature, the **electroneutrality principle**, states that bulk matter cannot acquire more than a trifling (and chemically insignificant) net electric charge. So one important thing to know about ions is that in ordinary matter, whether in the solid, liquid, or gaseous state, any positive ions must be accompanied by a compensating number of negative ions. Ionic substances such as sodium chloride form crystalline solids that can be regarded as made of ions. These solids tend to be quite hard and have high melting points, reflecting the strong forces between oppositely-charged ions. Solid metal oxides, such as CaO and MgO which are composed of doubly-charged ions don't melt at all, but simply dissociate into the elements at very high temperatures.

Polymers

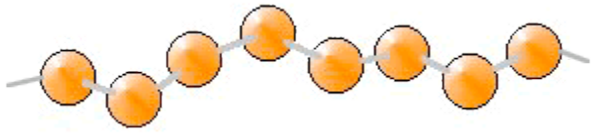
Plastics and natural materials such as rubber or cellulose are composed of very large molecules called *polymers*; many important biomolecules are also polymeric in nature. Owing to their great length, these molecules tend to become entangled in the liquid state, and are unable to separate to form a crystal lattice on cooling. In general, it is very difficult to get such substances to form anything other than amorphous solids.

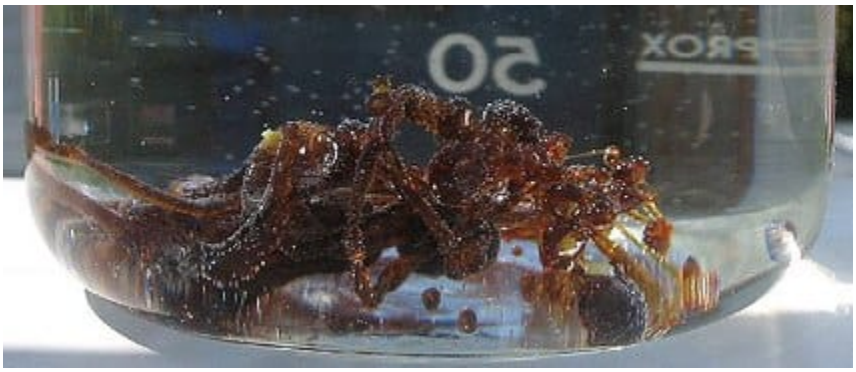
Extended solids

actually exist in their solid forms as linked assemblies of these basic units arranged in chains or layers that extend indefinitely in one, two, or three dimensions. Thus the very simple models of chemical bonding that apply to the isolated molecules in gaseous form must be modified to account for bonding in some of these solids. The terms "one-dimensional" and "two-dimensional", commonly employed in this context, should more accurately be prefixed by "*quasi*-"; after all, even a single atom occupies three-dimensional space!

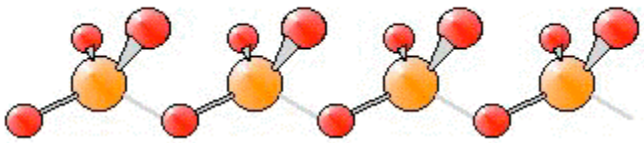
One-dimensional solids

Atoms of some elements such as sulfur and selenium can bond together in long chains of indefinite length, thus forming polymeric, amorphous solids. The most well known of these is the amorphous "plastic sulfur" formed when molten sulfur is cooled rapidly by pouring it into water.These are never the most common (or stable) forms of these elements, which prefer to form discrete molecules.



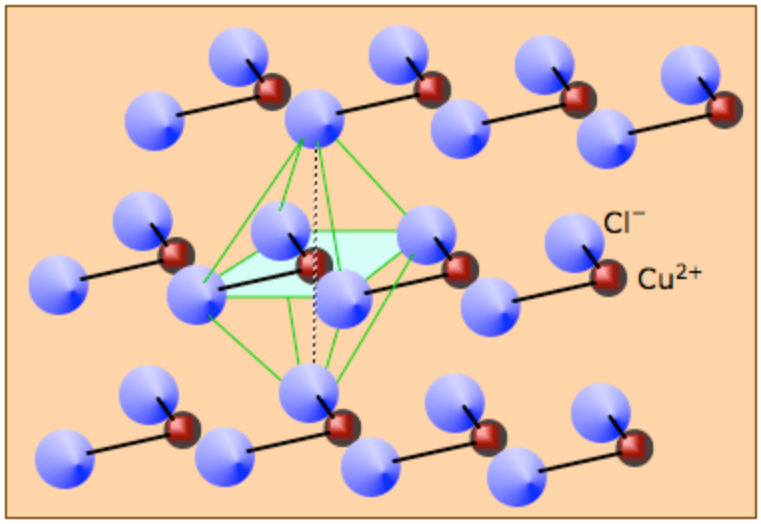


Rubber-like strands of plastic sulfur formed by pouring hot molten sulfur into cold water. After a few days, it will revert to ordinary crystalline sulfur. But small molecules can also form extended chains. Sulfur trioxide is a gas above room temperature, but when it freezes at 17°C the solid forms long chains in which each S atom is coordinated to four oxygen atoms.



Multi-dimensional solids

Many inorganic substances form crystalline solids which are built up from parallel chains in which the basic formula units are linked by weak bonds involving dipole-dipole and dipole-induced dipole interactions. Neighboring chains are bound mainly by dispersion forces.



Solid copper (II) chloride consists of multiple $CuCl_2$ units joined into long chains by covalent bonding. The chains are held together laterally by weaker van der Waals forces (mainly dispersion and ion-induced dipole). Each copper atom resides at the center of an imaginary octahedron whose vertices are defined by chlorine atoms in adjacent chains.

Layer or sheet-like structures

Solid cadmium chloride is a good example of a *layer* structure. The Cd and Cl atoms occupy separate layers; each of these layers extends out in a third dimension to form a *sheet*. The $CdCl_2$ crystal is built up from stacks of these layers held together by van der Waals forces.

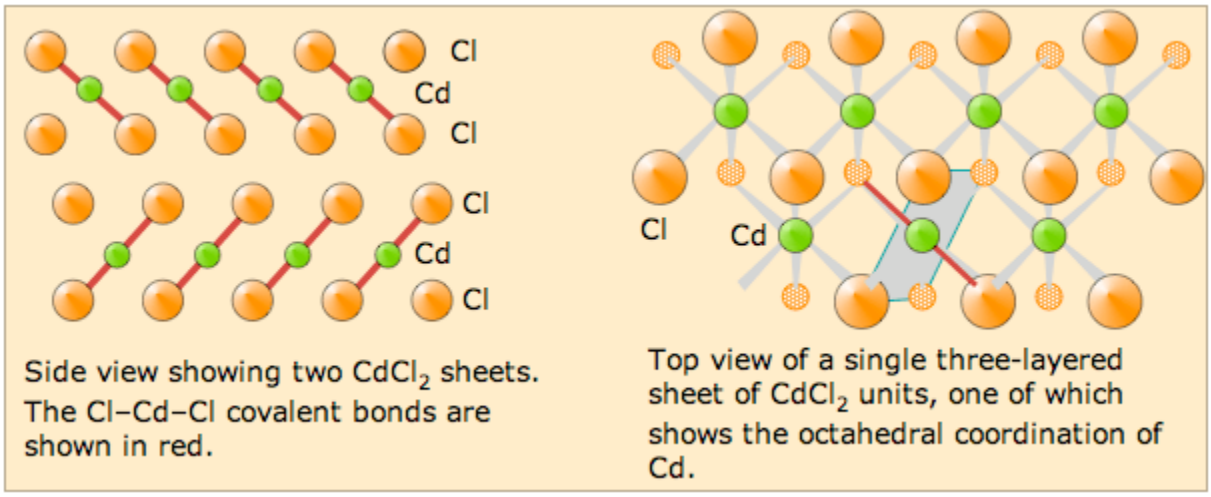
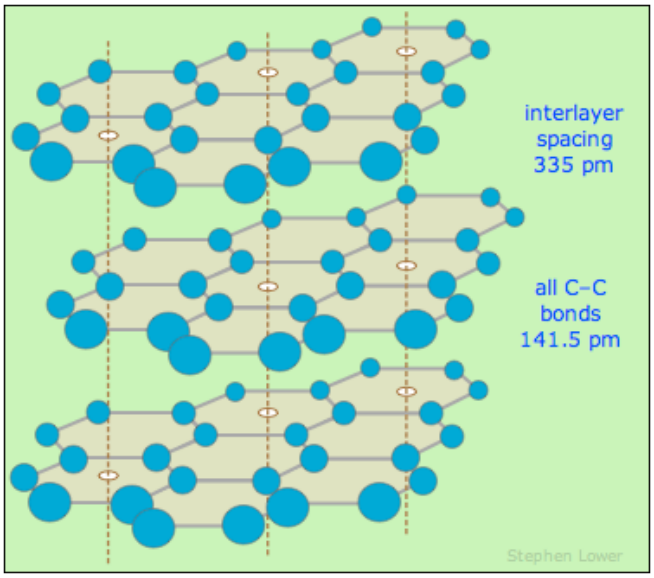


Figure: $CdCl_2$ lattice. Each Cd atom is covalently bonded to two chlorine atoms in adjacent layers; all other "bonds" are weak van der Waals attractions.

It's worth pointing out that although salts such as $CuCl_2$ and $CdCl_2$ are dissociated into ions when in aqueous solution, the solids themselves should not be regarded as "ionic solids". See also this section of the lesson on ionic solids.

Graphite

Graphite is a polymorph of carbon and its most stable form. It consists of sheets of fused benzene rings stacked in layers. The spacing between layers is sufficient to admit molecules of water vapor and other atmospheric gases which become absorbed in the interlamellar spaces and act as lubricants, allowing the layers to slip along each other. Thus graphite itself often has a flake-like character and is commonly used as a solid lubricant, although it loses this property in a vacuum.

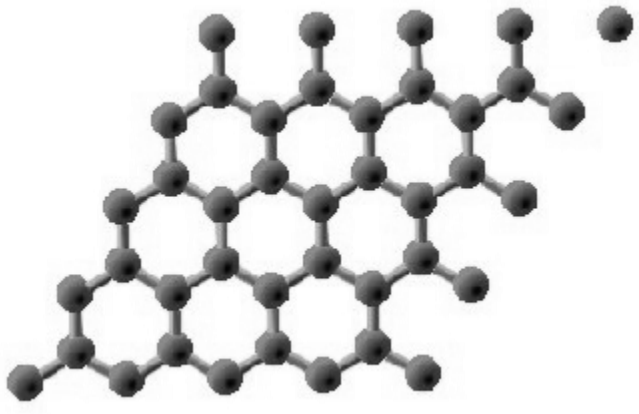


As would be expected from its anistropic structure, the electric and thermal conductivity of graphite are much greater in directions parallel to the layers than across the layers. The melting point of 4700-5000°C makes graphite useful as a high-temperature refractory material.

Graphite is the most common form of relatively pure carbon found in nature. Its name comes from the same root as the Greek word for "write" or "draw", reflecting its use as pencil "lead" since the 16th century. (The misnomer, which survives in common use, is due to its misidentification as an ore of the metallic element of the same name at a time long before modern chemistry had developed.)

Graphene

Graphene is a two-dimensional material consisting of a single layer of graphite — essentially "chicken wire made of carbon" that was discovered in 2004. Small fragments of graphene can be obtained by several methods; one is to attach a piece of Scotch Tape™ to a piece of graphite and then carefully pull it off (a process known as *exfoliation*.) Fragments of graphene are probably produced whenever one writes with a pencil.



Graphene has properties that are uniquely different from all other solids. It is the strongest known material, and it exhibits extremely high electrical conductivity due to its massless electrons which are apparently able to travel at relativistic velocities through the layer.

This page titled 7.1: Matter under the Microscope is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.2: Intermolecular Interactions

Learning Objectives

- Name five kinds of *molecular units* that condensed matter can be composed of.
- Sketch out a *potential energy curve*, showing clearly the equilibrium separation and potential energy minimum.
- State the difference between *bonded* and *non-bonded* attractions.
- Explain the meaning and significance of the *dipole moment* of a molecule.
- Define *induced dipole* and *polarizability*.
- State the six kinds of *intermolecular attractive forces* and their relative strengths.

Liquids and solids differ from gases in that they are held together by forces that act between the individual molecular units of which they are composed. In this lesson we will take a closer look at these forces so that you can more easily understand, and in many cases predict, the diverse physical properties of the many kinds of solids and liquids we encounter in the world.

The very existence of condensed states of matter suggests that there are attractive forces acting between the basic molecular units of solids and liquids. The term *molecular unit* refers to the smallest discrete structural unit that makes up the liquid or solid. In most of the over 15 million chemical substances that are presently known, these structural units are actual *molecules*—that is, aggregates of atoms that have their own distinguishing properties, formulas, and molecular weights. But the molecular units can also be individual *atoms*, *ions* and more extended units. As with most artificial classifications, these distinctions tend to break down in extreme cases: most artificial polymers ("plastics") are composed of molecules of various sizes and shapes, some metal alloys contain identifiable molecular units, and it is not too much of a stretch to regard a diamond or a crystal of NaCl as a "molecule" in itself.

Potential Energy Curves

On the atomic or molecular scale, all particles exert both attractive and repulsive forces on each other. If the attractive forces between two or more atoms are strong enough to make them into an enduring unit with its own observable properties, we call the result a "molecule" and refer to the force as a "chemical bond".

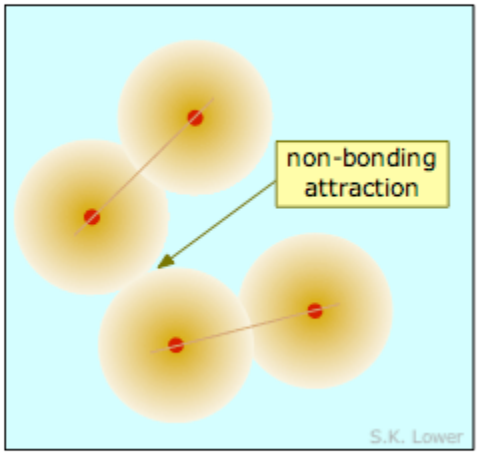


Figure 7.2.1: Bonding vs. non-bonding interactions

The two diatomic molecules depicted in Figure 7.2.1 have come into close contact with each other, but the attractive force that acts between them is not strong enough to bind them into a new molecular unit, so we call this force a *non-bonding attraction*. In the absence of these non-bonding attractions, all matter would exist in the gaseous state only; there would be no condensed phases.

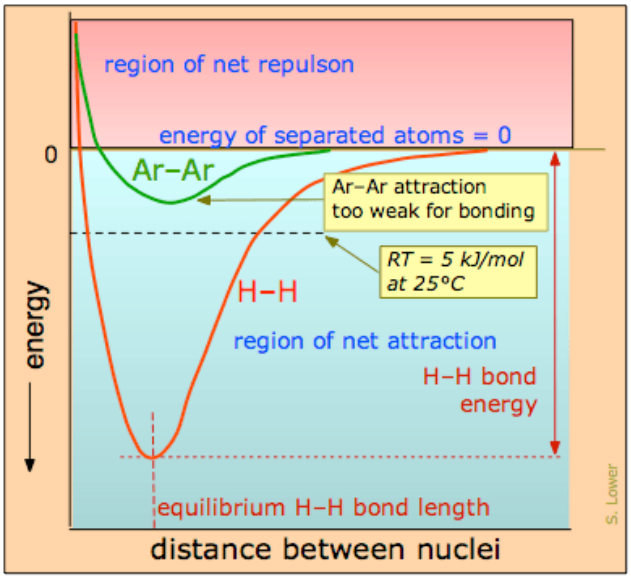


Figure 7.2.2: Potential Energy curve for bonding and non-bonding interactions

The distinction between bonding- and non-bonding attractions can be seen by comparing the potential energy plot for a pair of hydrogen atoms with that for two argon atoms (Figure 7.2.2). As two hydrogen atoms are brought together, the potential energy falls to a minimum and then rises rapidly as the two electron clouds begin to repel each other. The potential energy minimum defines the *energy* and the *average length* of the H–H bond — two of its unique measurable properties.

The potential energy of a pair of argon atoms also falls as they are brought together, but not enough to hold them together. (e.g., the laws of quantum mechanics do not allow this noble gas element to form stable Ar_2 molecules.) However, these non-bonding attractions enable argon to exist as a liquid and solid at low temperatures, but are unable to withstand disruptions caused by thermal energy at ordinary temperatures, so we commonly know argon as a gas.

From a classic pictures, at temperatures above absolute zero, all molecular-scale particles possess *thermal energy* that keeps them in constant motion (and from a quantum picture, motion does not stop even at absolute zero due to Heisenberg Uncertainly principle). The average thermal energy is given by the product of the gas constant *R* and the absolute temperature. At 25°C, this works out to

$$RT = (8.314\,J\,K^{-1}mol^{-1})(298\,K) = 2,480\,J\,mol^{-1} \approx 2.5\,kJ\,mol^{-1} \tag{7.2.1}$$

A functional chemical bond is much stronger than this (typically over 100 kJ/mol), so the effect of thermal motion is simply to cause the bond to vibrate; only at higher temperatures (where the value of *RT* is larger) will most bonds begin to break. Non-bonding attractive forces between pairs of atoms are generally too weak to sustain even a single vibration. In addition to unique distinguishing properties such as bond energy, bond length and stretching frequencies, covalent bonds usually have *directional properties* that depend on the orbital structures of the component atoms. The much-weaker non-bonding attractions possess none of these properties.

The shape of a potential energy curve (often approximated as a "Morse" curve) shows how repulsive and attractive forces affect the potential energy in opposite ways: repulsions always raise this energy, and attractions reduce it. The curve passes through a minimum when the attractive and repulsive forces are exactly in balance. As we stated above, all particles exert both kinds of forces on one another; these forces are all basically electrical in nature and they manifest themselves in various ways and with different strengths.

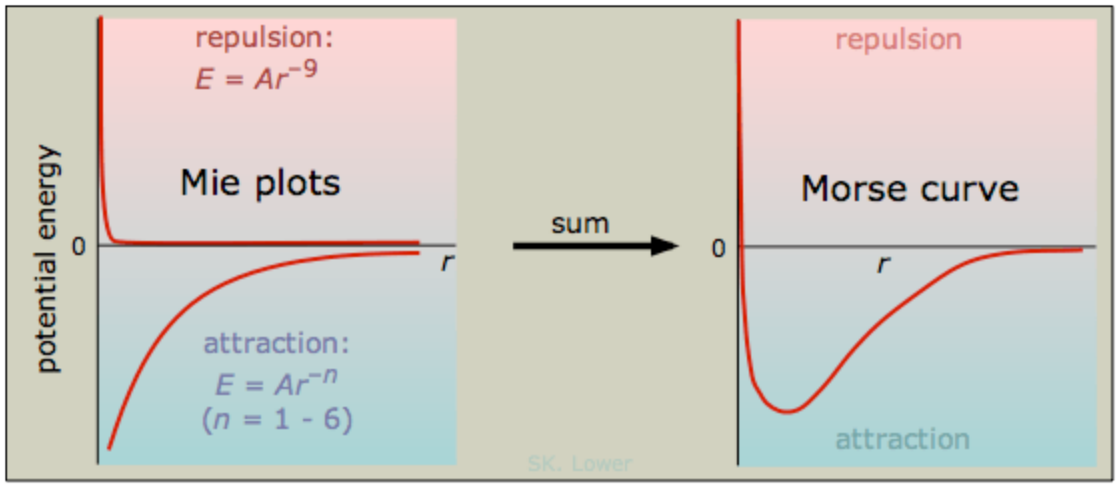


Figure 7.2.3: Intermolecular potential curves result from repulsion at short distances and attraction at longer distances

The distance corresponding to the minimum potential energy is known as the *equilibrium distance*. This is the average distance that will be maintained by the two particles if there are no other forces acting on them, such as might arise from the presence of other particles nearby. A general empirical expression for the interaction potential energy curve between two particles can be written as

$$E = Ar^{-n} + Br^{-m} \tag{7.2.2}$$

A and B are proportionality constants and n and m are integers. This expression is sometimes referred to as the *Mie equation*. The first term, A, corresponds to repulsion is always positive, and n must be larger than m, reflecting the fact that repulsion always dominates at small separations. The B coefficient is negative for attractive forces, but it will become positive for electrostatic repulsion between like charges. The larger the value of one of these exponents, the closer the particles must come before the force becomes significant. Table 7.2.1 lists the exponents for the types of interactions we will describe in this lesson.

Table 7.2.1: Classification of intermolecular forces

Species Involved	Type of Force	n	m
ions	Coulombic	-	1
ion - polar molecule	ion-dipole	-	2
two polar molecules	dipole-dipole	-	3
ion - nonpolar molecule	ion - induced dipole	-	4
polar and nonpolar molecule	dipole - induced dipole	-	6
nonpolar molecules	dispersion	-	6
repulsions	quantum	9	-
Note: the blue-shaded interactions are known collectively as van der Waals interactions			

The Universal Repulsive Force

The value of n for the repulsive force in Figure 7.2.3 is 9; this may be the highest inverse-power law to be found in nature. The magnitude of such a force is negligible until the particles are almost in direct contact, but once it kicks in, it becomes very strong; if you want to get a feel for it, try banging your head into a concrete wall. Because the repulsive force is what prevents two atoms from occupying the same space, this is just what you would expect. If the repulsive force did not always win out against all attractive forces, all matter would collapse into one huge glob! The universal repulsive force arises directly from two main aspects of quantum theory.

1. First, the Heisenberg uncertainty principle tells us that the electrons situated within the confines of an atom possess kinetic energy that would exert an outward pressure were it not for the compensating attractive force of the positively-charged

nucleus. But even the very slight decrease in volume that would result from squeezing the atom into a smaller space will raise this pressure so as to effectively resist this change in volume. This is the basic reason that condensed states of matter have extremely small compressibilities.

2. Working in concert with this is the Pauli exclusion principle each electron must have a different set of quantum numbers. So as two particles begin to intrude upon each other, the volume their electrons occupy gets divided up between each spin-pair, and the ones forced into higher quantum states would normally occupy even greater volumes. The effect is again to massively raise the potential energy as the particles begin to squeeze too close together.

In a wonderful article (*Science* 187 605-612 1975), the physicist Victor Weiskopf showed how these considerations, combined with a few fundamental constants, leads to realistic estimates of such things as the hardness and compressibility of solids, the heights of mountains, the lengths of ocean waves, and the sizes of stars.

Ion-Ion Interactions

Electrostatic attraction between electrically-charged particles is the strongest of all the intermolecular forces. These Coulombic forces (as they are often called) cause opposite charges to attract and like charges to repel.

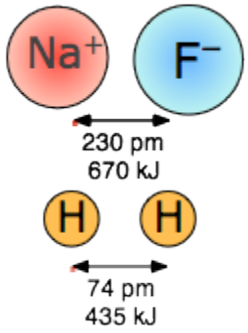


Figure 7.2.1: Bonding vs. non-bonding interactions

Coulombic forces are involved in all forms of chemical bonding; when they act between separate charged particles (ion-ion interactions) they are especially strong. Thus the energy required to pull a mole of Na^+ and Cl^- ions apart in the sodium chloride crystal is greater than that needed to break the covalent bond in H_2 (Figure 7.2.1). The effects of ion-ion attraction are seen most directly in solids such as NaCl which consist of oppositely-charged ions arranged in two inter-penetrating crystal lattices.

According to Coulomb's Law the force between two charged particles is given by

$$F = \frac{q_1 q_2}{4\pi\epsilon_0 r^2} \tag{7.2.3}$$

Instead of using SI units, chemists often prefer to express atomic-scale distances in picometers and charges as electron charge (±1, ±2, etc.) Using these units, the proportionality constant $1/4\pi\epsilon$ works out to $2.31 \times 10^{16}\ J\ pm$. The sign of F determines whether the force will be attractive (–) or repulsive (+); notice that the latter is the case whenever the two *q*'s have the same sign.

Equation 7.2.3 is an example of an *inverse square law*; the force falls off as the square of the distance. A similar law governs the manner in which the illumination falls off as you move away from a point light source; recall this the next time you walk away from a street light at night, and you will have some feeling for what an inverse square law means.

The stronger the attractive force acting between two particles, the greater the amount of work required to separate them. Work represents a flow of energy, so the foregoing statement is another way of saying that when two particles move in response to a force, their potential energy is lowered. This work, as you may recall if you have studied elementary mechanics, is found by integrating the negative force with respect to distance over the distance moved. Thus the energy that must be supplied in order to completely separate two oppositely-charged particles initially at a distance r_0 is given by

$$w = -\int_{r_o}^{\infty} \frac{q_1 q_2}{4\pi\epsilon_0 r} dr = -\frac{q_1 q_2}{4\pi\epsilon_0 r_o} \tag{7.2.4}$$

Example 7.2.1

When sodium chloride is melted, some of the ion pairs vaporize and form neutral NaCl molecules. How much energy would be released when one mole of Na^+ and Cl^- ions are brought together in this way?

Solution

The energy released will be the same as the work required to separate

$$E = \frac{(2.31 \times 10^{16} Jpm)(+1)(-1)}{276\ pm}$$
$$= -8.37 \times 10^{-19}\ J$$

The ion-ion interaction is the simplest of electrostatic interactions and other higher order interactions exists as discussed below.

Dipoles

According to Coulomb's law (Equation 7.2.2), the electrostatic force between an ion and an uncharged particle having Q = 0 should be zero. Bear in mind, however, that this formula assumes that the two particles are point charges having zero radii. A real particle such as an atom or a molecule occupies a certain volume of space. Even if the electric charges of the protons and electrons cancel out (as they will in any neutral atom or molecule), it is possible that the spatial distribution of the electron cloud representing the most loosely-bound [valence] electrons might be asymmetrical, giving rise to an *electric dipole moment*. There are two kinds of dipole moments:

- *Permanent* electric dipole moments can arise when bonding occurs between elements of differing electronegativities.
- *Induced* (temporary) dipole moments are created when an external electric field distorts the electron cloud of a neutral molecule.

An *electric dipole* refers to a separation of electric charge. An idealized electric dipole consists of two point charges of magnitude +*q* and –*q* separated by a distance *r*. Even though the overall system is electrically neutral, the charge separation gives rise to an electrostatic effect whose strength is expressed by the *electric dipole moment* given by

$$\mu = q \times r \tag{7.2.5}$$

Dipole moments possess both magnitude and direction, and are thus vectorial quantities; they are conventionally represented by arrows whose heads are at the negative end.

Permanent dipole moments

These are commonly referred to simply as "dipole moments". The most well-known molecule having a dipole moment is ordinary water. The charge imbalance arises because oxygen, with its nuclear charge of 8, pulls the electron cloud that comprises each O–H bond toward itself. These two "bond moments" add vectorially to produce the permanent dipole moment denoted by the red arrow. Note the use of the δ (Greek *delta*) symbol to denote the positive and negative ends of the dipoles.

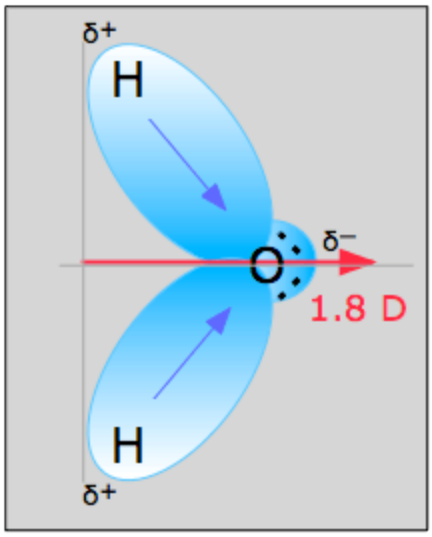


Figure 7.2.5: Dipole Moment in water

When an electric dipole is subjected to an external electric field, it will tend to orient itself so as to minimize the potential energy; that is, its negative end will tend to point toward the higher (more positive) electric potential. In liquids, thermal motions will act to disrupt this ordering, so the overall effect depends on the temperature. In condensed phases the local fields due to nearby ions or dipoles in a substance play an important role in determining the physical properties of the substance, and it is in this context that dipolar interactions are of interest to us here. We will discuss each kind of interaction in order of decreasing strength.

Induced dipoles

Even if a molecule is electrically neutral and possesses no permanent dipole moment, it can still be affected by an external electric field. Because all atoms and molecules are composed of charged particles (nuclei and electrons), the electric field of a nearby ion will cause the centers of positive and negative charges to shift in opposite directions. This effect, which is called *polarization*, results in the creation of a temporary, or induced dipole moment. The induced dipole then interacts with the species that produced it, resulting in a net attraction between the two particles.

The larger an atom or ion, the more loosely held are its outer electrons, and the more readily will the electron cloud by distorted by an external field. A quantity known as the *polarizability* expresses the magnitude of the temporary dipole that can be induced in it by a nearby charge.

Ion-Dipole interactions

A dipole that is close to a positive or negative ion will orient itself so that the end whose partial charge is opposite to the ion charge will point toward the ion. This kind of interaction is very important in aqueous solutions of ionic substances; H_2O is a highly polar molecule, so that in a solution of sodium chloride, for example, the Na^+ ions will be enveloped by a shell of water molecules with their oxygen-ends pointing toward these ions, while H_2O molecules surrounding the Cl^- ions will have their hydrogen ends directed inward. As a consequence of ion-dipole interactions, all ionic species in aqueous solution are hydrated; this is what is denoted by the suffix in formulas such as $K^+(aq)$, etc.

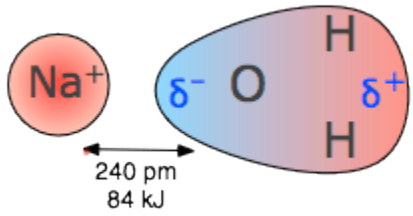


Figure 7.2.6: Ion-Dipole Interaction between a sodium ion and water

The strength of ion-dipole attraction depends on the magnitude of the dipole moment and on the charge density of the ion. This

latter quantity is just the charge of the ion divided by its volume. Owing to their smaller sizes, positive ions tend to have larger charge densities than negative ions, and they should be more strongly hydrated in aqueous solution. The hydrogen ion, being nothing more than a bare proton of extremely small volume, has the highest charge density of any ion; it is for this reason that it exists entirely in its hydrated form H_3O^+ in water.

Dipole-dipole interactions

As two dipoles approach each other, they will tend to orient themselves so that their oppositely-charged ends are adjacent. Two such arrangements are possible: the dipoles can be side by side but pointing in opposite directions, or they can be end to end. It can be shown that the end-to-end arrangement gives a lower potential energy.

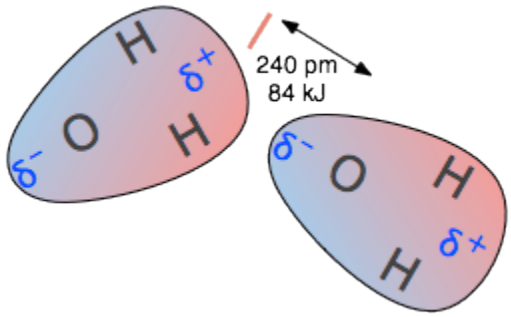


Figure 7.2.7: Dipole-Dipole Interaction between a water molecule and a water molecule

Dipole-dipole attraction is weaker than ion-dipole attraction, but it can still have significant effects if the dipole moments are large. The most important example of dipole-dipole attraction is hydrogen bonding.

Ion-induced dipole Interactions

The most significant induced dipole effects result from nearby ions, particularly cations (positive ions). Nearby ions can distort the electron clouds even in polar molecules, thus temporarily changing their dipole moments. The larger ions (especially negative ones such as SO_2^{2-} and ClO_4^{2-}) are highly polarizable, and the dipole moments induced in them by a cation can play a dominant role in compound formation.

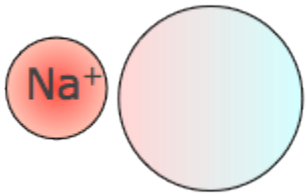


Figure 7.2.8: Ion-Dipole Interaction between a sodium ion and water

Dipole-induced dipole interactions

A permanent dipole can induce a temporary one in a species that is normally nonpolar, and thus produce a net attractive force between the two particles (Figure 7.2.9). This attraction is usually rather weak, but in a few cases it can lead to the formation of loosely-bound compounds. This effect explains the otherwise surprising observation that a wide variety of neutral molecules such as hydrocarbons, and even some of the noble gas elements, form stable hydrate compounds with water.

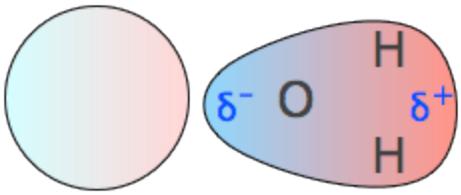


Figure 7.2.9: Dipole-induced dipole Interaction between a dipole and water

Dispersion (London) Forces

The fact that noble gas elements and completely non-polar molecules such as H_2 and N_2 can be condensed to liquids or solids tells us that there must be yet another source of attraction between particles that does not depend on the existence of permanent dipole moments in either particle (Figure 7.2.10). To understand the origin of this effect, it is necessary to realize that when we say a molecule is “nonpolar”, we really mean that the *time-averaged* dipole moment is zero. This is the same kind of averaging we do when we draw a picture of an orbital, which represents all the locations in space in which an electron can be found with a certain minimum probability. On a very short time scale, however, the electron must be increasingly localized; not even quantum mechanics allows it to be in more than one place at any given instant. As a consequence, there is no guarantee that the distribution of negative charge around the center of an atom will be perfectly symmetrical at every instant; every atom therefore has a weak, fluctuating dipole moment that is continually disappearing and reappearing in another direction.

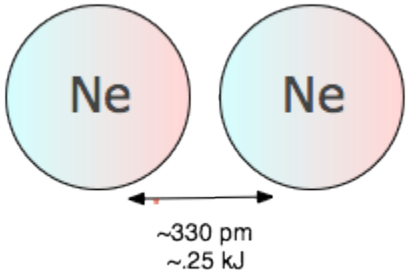


Figure 7.2.10: Dispersion Interaction between a dipole and water

Dispersion or London forces can be considered to be "spontaneous dipole - induced dipole" interactions.

Although these extremely short-lived fluctuations quickly average out to zero, they can still induce new dipoles in a neighboring atom or molecule, which helps sustain the original dipole and gives rise to a weak attractive force known as the *dispersion* or *London force*. Although dispersion forces are the weakest of all the intermolecular attractions, they are **universally present**. Their strength depends to a large measure on the number of electrons in a molecule. This can clearly be seen by looking at the noble gas elements in Table 7.2.2, whose ability to condense to liquids and freeze to solids is entirely dependent on dispersion forces.

Table 7.2.2: Intermolecular interactions in the noble gases

element	He	Ne	Ar	Kr	Xe
atomic number	2	10	18	36	54
boiling point, K	27	87	120	165	211
critical temperature, K	5	44	151	209.5	290
heat of vaporization, kJ mol^{-1}	0.08	1.76	6.51	59	12.6

It is important to note that dispersion forces are *additive*; if two elongated molecules find themselves side by side, dispersion force attractions will exist all along the regions where the two molecules are close. This can produce quite strong attractions between large polymeric molecules even in the absence of any stronger attractive forces.

"van der Waals" forces is a catch all Term

Although nonpolar molecules are by no means uncommon, many kinds of molecules possess permanent dipole moments, so

liquids and solids composed of these species will be held together by a combination of dipole-dipole, dipole-induced dipole, and dispersion forces. These weaker forces (that is, those other than Coulombic attractions) are known collectively as *van der Waals forces*. These include attraction and repulsions between atoms, molecules, and surfaces, as well as other intermolecular forces. The term includes:

- force between two permanent dipoles and higher order moments like quadrupole
- force between a permanent dipole and a corresponding induced dipole
- force between two instantaneously induced dipoles (dispersion forces)

Table 7.2.3 shows some estimates of the contributions of the various types of van der Waals forces that act between several different types of molecules. Note particularly how important dispersion forces are in all of these examples, and how this, in turn, depends on the polarizability.

Table 7.2.3: Intermolecular interactions in the noble gases

Dubstance	Boiling Point °C	Dipole Moment D	Polarizability	% Dipole-induced Dipole	% Dipole-Dipole	% Dispersion
Ar	–186	0	1.6	0	0	100
CO	–190	0.1	2.0	0	0	100
HCl	–84	1.0	2.6	4.2	14.4	81.4
HBr	–67	0.8	3.6	2.2	3.3	94.5
HI	–35	0.4	5.4	0.4	0.1	99.5
NH_3	–33	1.5	2.6	5.4	44.6	50.0
H_2O	100	1.8	1.5	4.0	77.0	19.0

This page titled 7.2: Intermolecular Interactions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

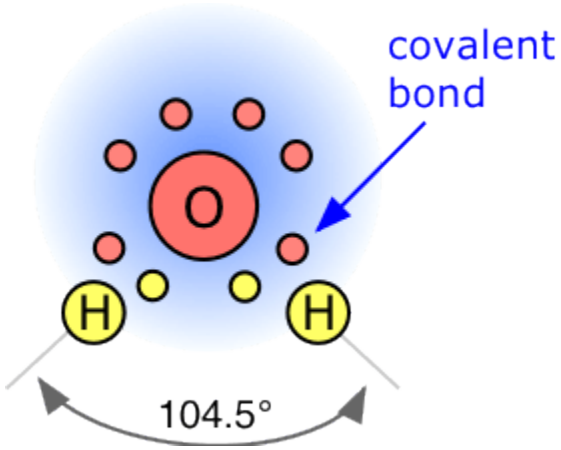
7.3: Hydrogen-Bonding and Water

Learning Objectives

- Identify three special properties of water that make it unusual for a molecule of its size, and explain how these result from hydrogen bonding.
- Explain what is meant by hydrogen bonding and the molecular structural features that bring it about.
- Describe the "structure", such as it is, of liquid water.
- Sketch out structural examples of hydrogen bonding in three small molecules other than H_2O.
- Describe the roles of hydrogen bonding in proteins and in DNA.

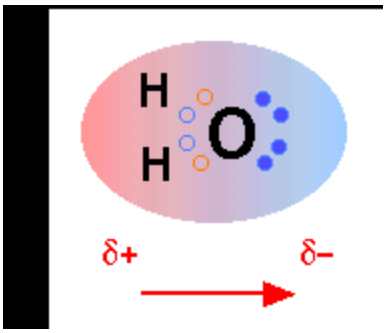
Most students of chemistry quickly learn to relate the structure of a molecule to its general properties. Thus we generally expect small molecules to form gases or liquids, and large ones to exist as solids under ordinary conditions. And then we come to H_2O, and are shocked to find that many of the predictions are way off, and that water (and by implication, life itself) should not even exist on our planet! In this section we will learn why this tiny combination of three nuclei and ten electrons possesses special properties that make it unique among the more than 15 million chemical species we presently know.

In water, each hydrogen nucleus is covalently bound to the central oxygen atom by a pair of electrons that are shared between them. In H_2O, only two of the six outer-shell electrons of oxygen are used for this purpose, leaving four electrons which are organized into two non-bonding pairs. The four electron pairs surrounding the oxygen tend to arrange themselves as far from each other as possible in order to minimize repulsions between these clouds of negative charge. This would ordinarily result in a tetrahedral geometry in which the angle between electron pairs (and therefore the H-O-H *bond angle*) is 109.5°. However, because the two non-bonding pairs remain closer to the oxygen atom, these exert a stronger repulsion against the two covalent bonding pairs, effectively pushing the two hydrogen atoms closer together. The result is a distorted tetrahedral arrangement in which the H—O—H angle is 104.5°.

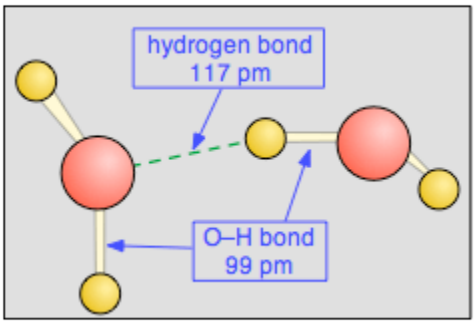


Water's large dipole moment leads to hydrogen bonding

The H_2O molecule is electrically neutral, but the positive and negative charges are not distributed uniformly. This is illustrated by the gradation in color in the schematic diagram here. The electronic (negative) charge is concentrated at the oxygen end of the molecule, owing partly to the nonbonding electrons (solid blue circles), and to oxygen's high nuclear charge which exerts stronger attractions on the electrons. This charge displacement constitutes an *electric dipole*, represented by the arrow at the bottom; you can think of this dipole as the electrical "image" of a water molecule.

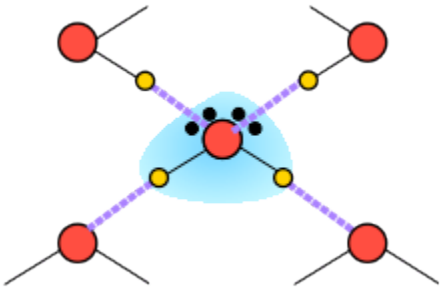


Opposite charges attract, so it is not surprising that the negative end of one water molecule will tend to orient itself so as to be close to the positive end of another molecule that happens to be nearby. The strength of this *dipole-dipole attraction* is less than that of a normal chemical bond, and so it is completely overwhelmed by ordinary thermal motions in the gas phase. However, when the H_2O molecules are crowded together in the liquid, these attractive forces exert a very noticeable effect, which we call (somewhat misleadingly) *hydrogen bonding*. And at temperatures low enough to turn off the disruptive effects of thermal motions, water freezes into ice in which the hydrogen bonds form a rigid and stable network.



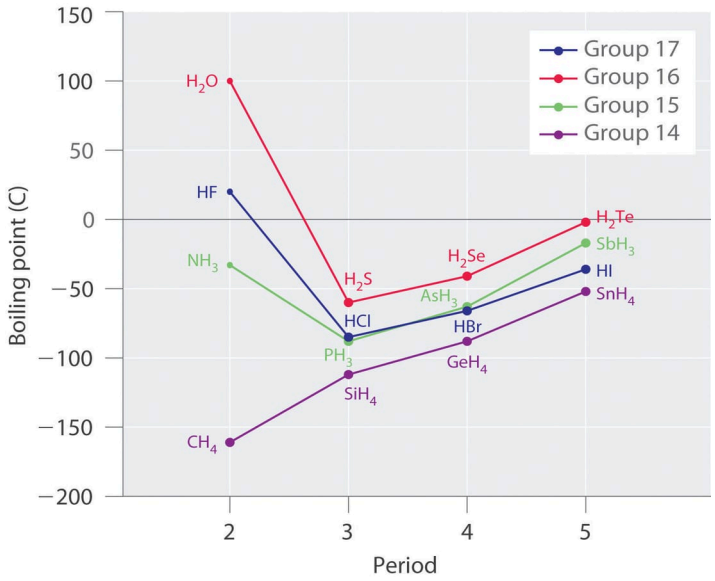
Notice that the hydrogen bond (shown by the dashed green line) is somewhat longer than the covalent O—H bond. It is also *much weaker*, about 23 kJ mol^{-1} compared to the O–H covalent bond strength of 492 kJ mol^{-1}.

Water has long been known to exhibit many physical properties that distinguish it from other small molecules of comparable mass. Although chemists refer to these as the "anomalous" properties of water, they are by no means mysterious; all are entirely predictable consequences of the way the size and nuclear charge of the oxygen atom conspire to distort the electronic charge clouds of the atoms of other elements when these are chemically bonded to the oxygen.



Boiling point

The most apparent peculiarity of water is its **very high boiling point** for such a light molecule. Liquid methane CH_4 (molecular weight 16) boils at –161°C. As you can see from this diagram, extrapolation of the boiling points of the various Group 16 hydrogen compounds to H_2O suggests that this substance should be a gas under normal conditions.

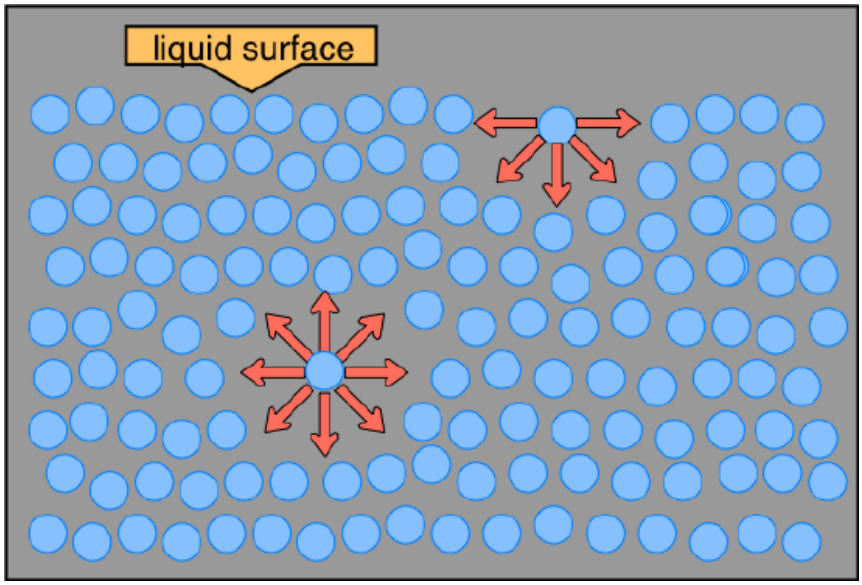


The Effects of Hydrogen Bonding on Boiling Points. These plots of the boiling points of the covalent hydrides of the elements of groups 14–17 show that the boiling points of the lightest members of each series for which hydrogen bonding is possible (HF, NH_3, and H_2O) are anomalously high for compounds with such low molecular masses.

Surface Tension

Compared to most other liquids, water also has a high **surface tension**. Have you ever watched an insect walk across the surface of a pond? The *water strider* takes advantage of the fact that the water surface acts like an elastic film that resists deformation when a small weight is placed on it. (If you are careful, you can also "float" a small paper clip or steel staple on the surface of water in a cup.) This is all due to the *surface tension* of the water. A molecule within the bulk of a liquid experiences attractions to neighboring molecules in all directions, but since these average out to zero, there is no net force on the molecule. For a molecule that finds itself *at* the surface, the situation is quite different; it experiences forces only sideways and downward, and this is what creates the stretched-membrane effect.





The distinction between molecules located at the surface and those deep inside is especially prominent in H_2O, owing to the strong hydrogen-bonding forces. The difference between the forces experienced by a molecule at the surface and one in the bulk liquid gives rise to the liquid's surface tension. This drawing highlights two H_2O molecules, one at the surface, and the other in the bulk of the liquid. The surface molecule is attracted to its neighbors below and to either side, but there are no attractions pointing in the 180° solid angle angle above the surface. As a consequence, a molecule at the surface will tend to be drawn into

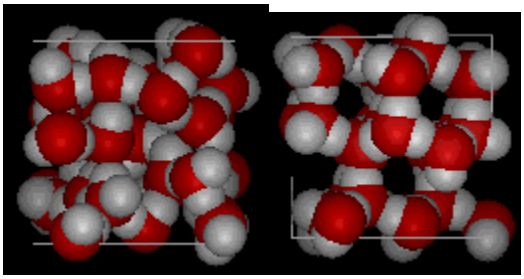
the bulk of the liquid. But since there must always be some surface, the overall effect is to minimize the surface area of a liquid.

The geometric shape that has the smallest ratio of surface area to volume is the *sphere*, so very small quantities of liquids tend to form spherical drops. As the drops get bigger, their weight deforms them into the typical tear shape.



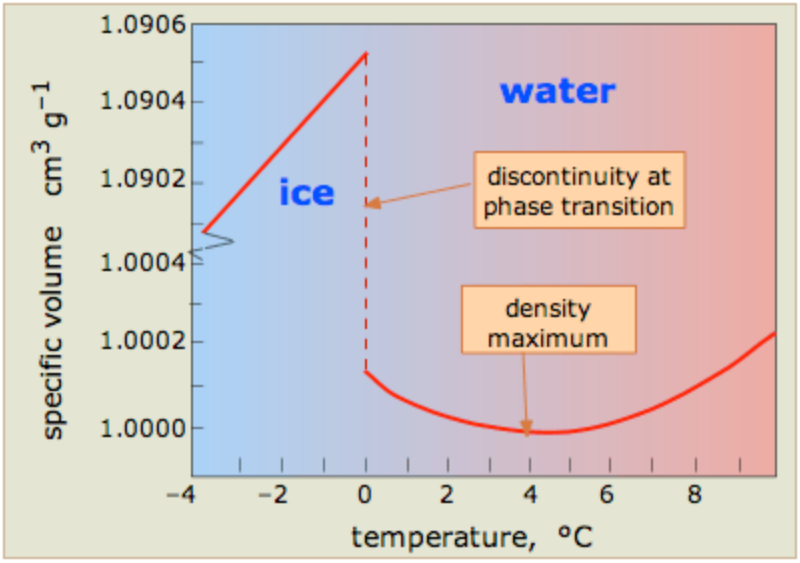
Ice floats on water

The most energetically favorable configuration of H_2O molecules is one in which each molecule is hydrogen-bonded to four neighboring molecules. Owing to the thermal motions described above, this ideal is never achieved in the liquid, but when water freezes to ice, the molecules settle into exactly this kind of an arrangement in the ice crystal. This arrangement requires that the molecules be somewhat farther apart then would otherwise be the case; as a consequence, ice, in which hydrogen bonding is at its maximum, has a more open structure, and thus a lower density than water.



Here are three-dimensional views of a typical local structure of water (left) and ice (right.) Notice the greater openness of the ice structure which is necessary to ensure the strongest degree of hydrogen bonding in a uniform, extended crystal lattice. The more crowded and jumbled arrangement in liquid water can be sustained only by the greater amount of thermal energy available above the freezing point.

When ice melts, the more vigorous thermal motion disrupts much of the hydrogen-bonded structure, allowing the molecules to pack more closely. Water is thus one of the very few substances whose solid form has a lower density than the liquid at the freezing point. Localized clusters of hydrogen bonds still remain, however; these are continually breaking and reforming as the thermal motions jiggle and shove the individual molecules. As the temperature of the water is raised above freezing, the extent and lifetimes of these clusters diminish, so the density of the water increases.



At higher temperatures, another effect, common to all substances, begins to dominate: as the temperature increases, so does the amplitude of thermal motions. This more vigorous jostling causes the average distance between the molecules to increase, reducing the density of the liquid; this is ordinary thermal expansion.

Because the two competing effects (hydrogen bonding at low temperatures and thermal expansion at higher temperatures) both lead to a decrease in density, it follows that there must be some temperature at which the density of water passes through a maximum. This temperature is 4° C; this is the temperature of the water you will find at the bottom of an ice-covered lake in which this most dense of all water has displaced the colder water and pushed it nearer to the surface.

Structure of Liquid Water

The nature of liquid water and how the H_2O molecules within it are organized and interact are questions that have attracted the interest of chemists for many years. There is probably no liquid that has received more intensive study, and there is now a huge literature on this subject. The following facts are well established:

- H_2O molecules attract each other through the special type of dipole-dipole interaction known as hydrogen bonding
- a hydrogen-bonded cluster in which four H_2Os are located at the corners of an imaginary tetrahedron is an especially favorable (low-potential energy) configuration, *but*...
- the molecules undergo rapid thermal motions on a time scale of picoseconds (10^{-12} second), so the lifetime of any specific clustered configuration will be fleetingly brief.

A variety of techniques including infrared absorption, neutron scattering, and nuclear magnetic resonance have been used to probe the microscopic structure of water. The information garnered from these experiments and from theoretical calculations has led to the development of around twenty "models" that attempt to explain the structure and behavior of water. More recently, computer simulations of various kinds have been employed to explore how well these models are able to predict the observed physical properties of water.

This work has led to a gradual refinement of our views about the structure of liquid water, but it has not produced any definitive answer. There are several reasons for this, but the principal one is that the very concept of "structure" (and of water "clusters") depends on both the time frame and volume under consideration. Thus, questions of the following kinds are still open:

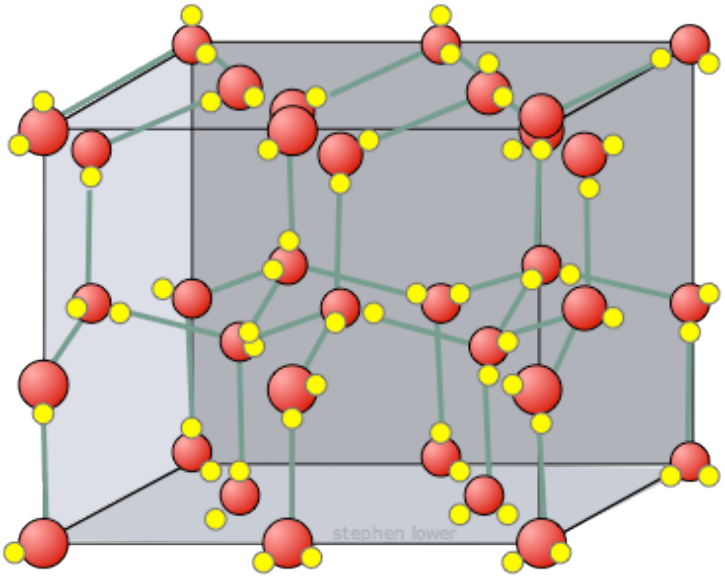
- How do you distinguish the members of a "cluster" from adjacent molecules that are not in that cluster?
- Since individual hydrogen bonds are continually breaking and re-forming on a picosecond time scale, do water clusters have any meaningful existence over longer periods of time? In other words, clusters are transient, whereas "structure" implies a molecular arrangement that is more enduring. Can we then legitimately use the term "clusters" in describing the structure of water?
- The possible locations of neighboring molecules around a given H_2O are limited by energetic and geometric considerations, thus giving rise to a certain amount of "structure" within any small volume element. It is not clear, however, to what extent

these structures interact as the size of the volume element is enlarged. And as mentioned above, to what extent are these structures maintained for periods longer than a few picoseconds?

In the 1950's it was assumed that liquid water consists of a mixture of hydrogen-bonded clusters $(H_2O)_n$ in which n can have a variety of values, but little evidence for the existence of such aggregates was ever found. The present view, supported by computer-modeling and spectroscopy, is that on a very short time scale, water is more like a "gel" consisting of a single, huge hydrogen-bonded cluster. On a 10^{-12}-10^{-9} sec time scale, rotations and other thermal motions cause individual hydrogen bonds to break and re-form in new configurations, inducing ever-changing local discontinuities whose extent and influence depends on the temperature and pressure.

Ice

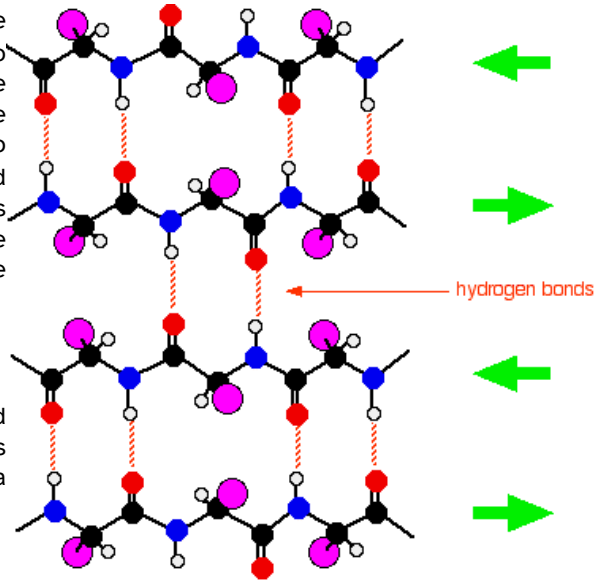
Ice, like all solids, has a well-defined structure; each water molecule is surrounded by four neighboring H_2Os. two of these are hydrogen-bonded to the oxygen atom on the central H_2O molecule, and each of the two hydrogen atoms is similarly bonded to another neighboring H_2O.



Ice forms crystals having a hexagonal lattice structure, which in their full development would tend to form hexagonal prisms very similar to those sometimes seen in quartz. This does occasionally happen, and anyone who has done much winter mountaineering has likely seen needle-shaped prisms of ice crystals floating in the air. Under most conditions, however, the snowflake crystals we see are flattened into the beautiful fractal-like hexagonal structures that are commonly observed.

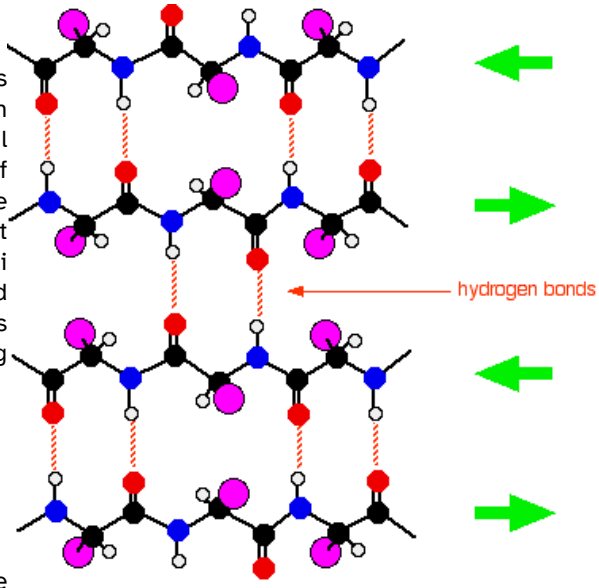
Snowflakes

The H_2O molecules that make up the top and bottom plane faces of the prism are packed very closely and linked (through hydrogen bonding) to the molecules inside. In contrast to this, the molecules that make up the sides of the prism, and especially those at the hexagonal corners, are much more exposed, so that atmospheric H_2O molecules that come into contact with most places on the crystal surface attach very loosely and migrate along it until they are able to form hydrogen-bonded attachments to these corners, thus becoming part of the solid and extending the structure along these six directions. This process perpetuates itself as the new extensions themselves acquire a hexagonal structure.



Why is ice slippery?

At temperatures as low as 200 K, the surface of ice is highly disordered and water-like. As the temperature approaches the freezing point, this region of disorder extends farther down from the surface and acts as a lubricant.



The illustration is taken from from an article in the April 7, 2008 issue of C&EN honoring the physical chemist Gabor Somorjai who pioneered modern methods of studying surfaces.

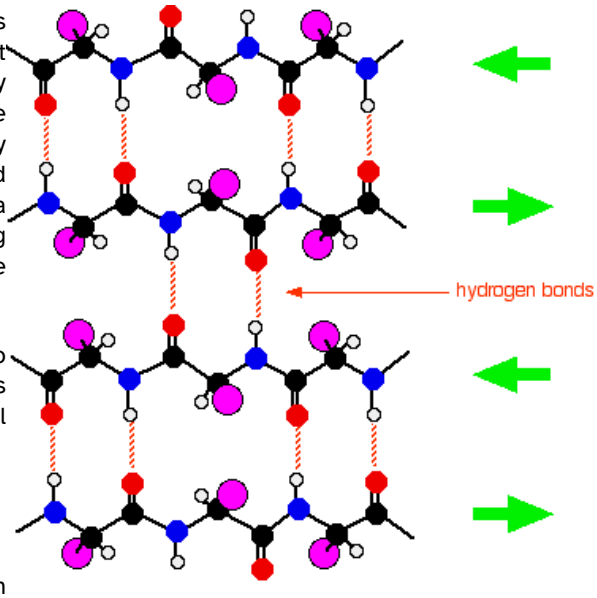
"Pure" water

To a chemist, the term "pure" has meaning only in the context of a particular application or process. The distilled or de-ionized water we use in the laboratory contains dissolved atmospheric gases and occasionally some silica, but their small amounts and relative inertness make these impurities insignificant for most purposes. When water of the highest obtainable purity is required for certain types of exacting measurements, it is commonly filtered, de-ionized, and triple-vacuum distilled. But even this "chemically pure" water is a mixture of isotopic species: there are two stable isotopes of both hydrogen (H^1 and H^2, the latter often denoted by D) and oxygen (O^{16} and O^{18}) which give rise to combinations such as H_2O^{18}, HDO^{16}, etc., all of which are readily identifiable in the infrared spectra of water vapor. And to top this off, the two hydrogen atoms in water contain protons whose magnetic moments can be parallel or antiparallel, giving rise to *ortho*- and *para*-water, respectively. The two forms are normally present in a *o/p* ratio of 3:1.

The amount of the **rare isotopes of oxygen and hydrogen in water** varies enough from place to place that it is now possible to determine the age and source of a particular water sample with some precision. These differences are reflected in the H and O isotopic profiles of organisms. Thus the isotopic analysis of human hair can be a useful tool for crime investigations and anthropology research.

More about hydrogen bonding

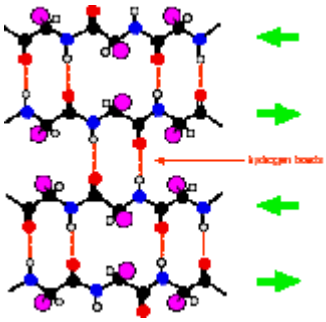
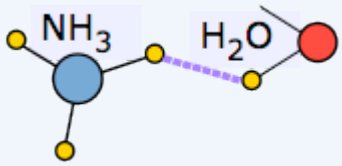
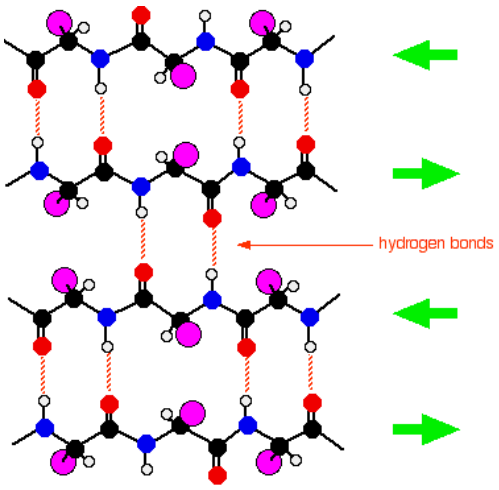
Hydrogen bonds form when the electron cloud of a hydrogen atom that is attached to one of the more electronegative atoms is distorted by that atom, leaving a partial positive charge on the hydrogen. Owing to the very small size of the hydrogen atom, the density of this partial charge is large enough to allow it to interact with the lone-pair electrons on a nearby electronegative atom. Although hydrogen bonding is commonly described as a form of dipole-dipole attraction, it is now clear that it involves a certain measure of electron-sharing (between the external non-bonding electrons and the hydrogen) as well, so these bonds possess some covalent character.

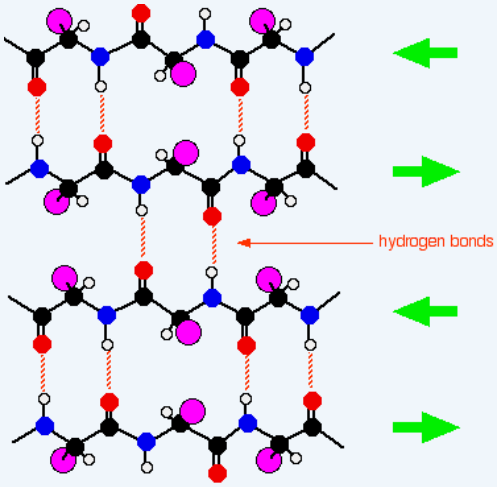
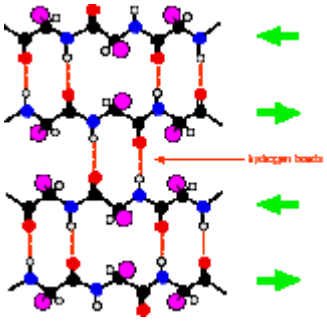
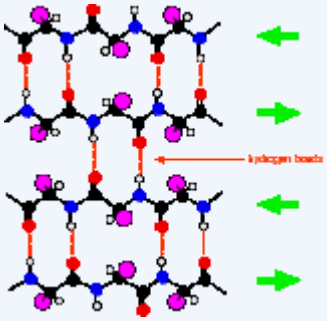
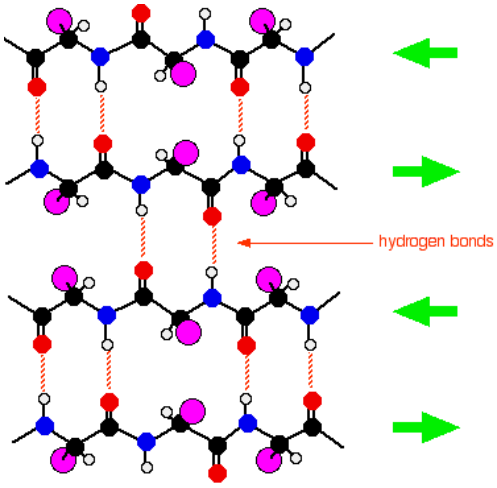
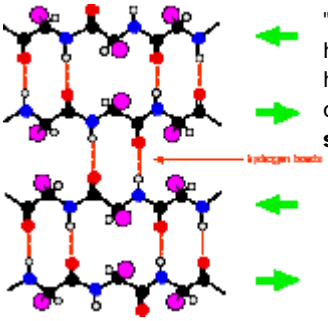


Hydrogen bonds are longer than ordinary covalent bonds, and they are also weaker. The experimental evidence for hydrogen bonding usually comes from X-ray diffraction studies on solids that reveal shorter-than-normal distances between hydrogen and other atoms.

Hydrogen bonding in small molecules

The following examples show something of the wide scope of hydrogen bonding in molecules.

	**Ammonia** (mp –78, bp –33°C) is hydrogen-bonded in the liquid and solid states.
	Hydrogen bonding is responsible for **ammonia**'s remarkably high solubility in water.
	Many **organic (carboxylic) acids** form hydrogen-bonded dimers in the solid state.

	Here the hydrogen bond acceptor is the π electron cloud of a benzene ring. This type of interaction is important in maintaining the shape of proteins.
	**Hydrogen fluoride** (mp –92, bp 33°C) is another common substance that is strongly hydrogen-bonded in its condensed phases.
	The **bifluoride ion** (for which no proper Lewis structure can be written) can be regarded as a complex ion held together by the strongest hydrogen bond known: about 155 kJ mol^{-1}.
	 "***As slow as molasses in the winter!***" Multiple hydroxyl groups provide lots of opportunities for hydrogen bonding and lead to the high viscosities of substances such as **glycerine** and **sugar syrups**.

Hydrogen bonding in biopolymers

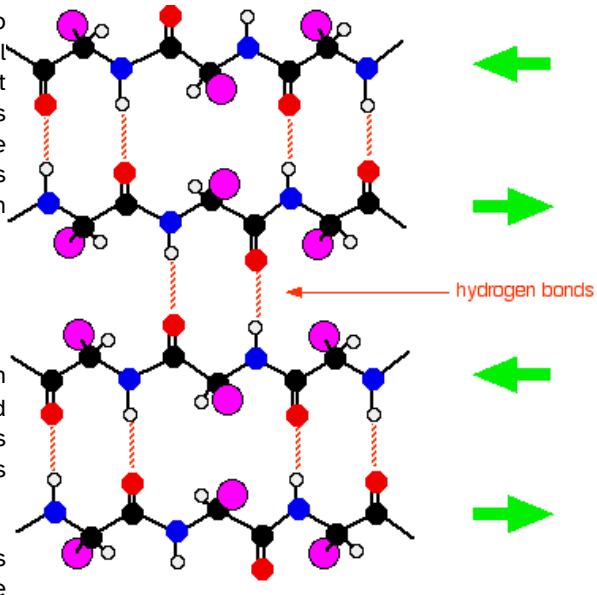
Hydrogen bonding plays an essential role in natural polymers of biological origin in two ways:

- Hydrogen bonding between adjacent polymer chains (intermolecular bonding);
- Hydrogen bonding between different parts of the same chain (intramolecular bonding;
- Hydrogen bonding of water molecules to –OH groups on the polymer chain ("bound water") that helps maintain the shape of the polymer.

The examples that follow are representative of several types of biopolymers.

Cellulose

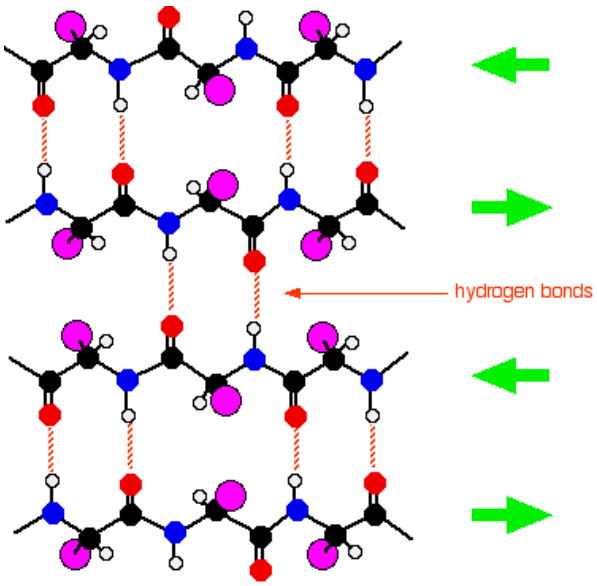
Cellulose is a linear polymer of glucose (see above), containing 300 to over 10,000 units, depending on the source. As the principal structural component of plants (along with lignin in trees), cellulose is the most abundant organic substance on the earth. The role of hydrogen bonding is to cross-link individual molecules to build up sheets as shown here. These sheets than stack up in a staggered array held together by van der Waals forces. Further hydrogen-bonding of adjacent stacks bundles them together into a stronger and more rigid structure.



Proteins

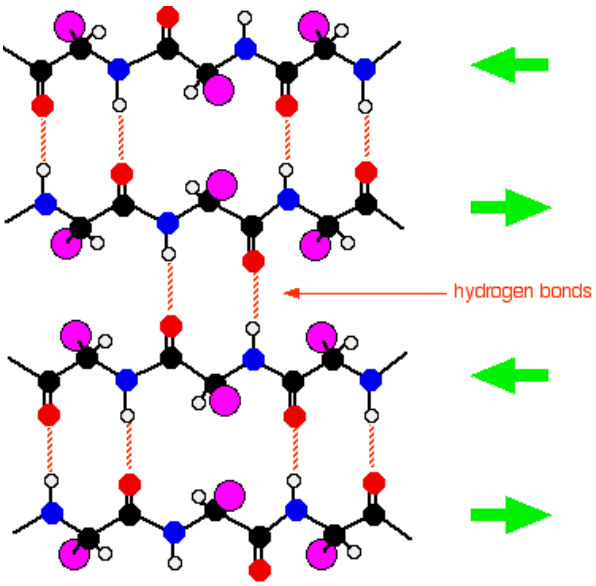
These polymers made from amino acids R—CH(NH_2)COOH depend on intramolecular hydrogen bonding to maintain their shape (secondary and tertiary structure) which is essential for their important function as biological catalysts (enzymes). Hydrogen-bonded water molecules embedded in the protein are also important for their structural integrity.

The principal hydrogen bonding in proteins is between the -N—H groups of the "amino" parts with the -C=O groups of the "acid" parts. These interactions give rise to the two major types of the secondary structure which refers to the arrangement of the amino acid polymer chain:



alpha-helix

[images]



beta-sheet

Although carbon is not usually considered particularly electronegative, C—H----X hydrogen bonds are also now known to be significant in proteins.

DNA (Deoxyribonucleic acid)

Who you are is totally dependent on hydrogen bonds! DNA, as you probably know, is the most famous of the biopolymers owing to its central role in defining the structure and function of all living organisms. Each strand of DNA is built from a sequence of four different *nucleotide monomers* consisting of a *deoxyribose sugar, phosphate group*s, and a *nitrogenous base* conventionally identified by the letters A,T, C and G. DNA itself consists of two of these *polynucleotide chains* that are coiled around a common axis in a configuration something like the protein alpha helix depicted above. The sugar-and-phosphate backbones are on the outside so that the nucleotide bases are on the inside and facing each other. The two strands are held together by hydrogen bonds that link a nitrogen atom of a nucleotide in one chain with a nitrogen or oxygen on the nucleotide that is across from it on the other chain.

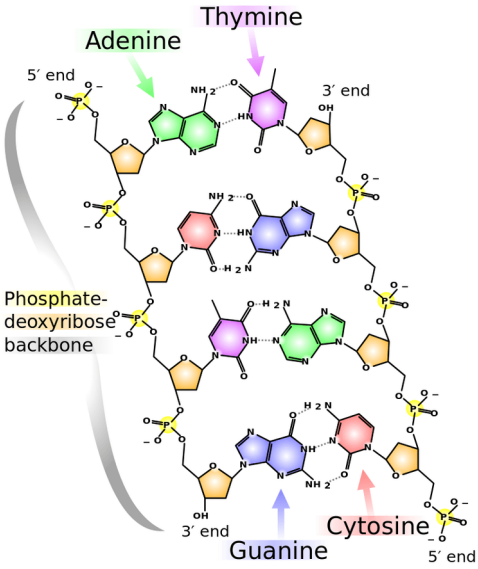


Figure: Chemical structure of DNA; hydrogen bonds shown as dotted lines. Image used with permisison from Wikipedia.

Efficient hydrogen bonding within this configuration can only occur between the pairs A-T and C-G, so these two complementary pairs constitute the "alphabet" that encodes the genetic information that gets transcribed whenever new protein molecules are built. Water molecules, hydrogen-bonded to the outer parts of the DNA helix, help stabilize it.

This page titled 7.3: Hydrogen-Bonding and Water is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.4: Liquids and their Interfaces

Learning Objectives

- Liquids are both *fluids* and *condensed phases*. Explain what this tells us about liquids, and what other states of matter fit into each of these categories.
- Define *viscosity*, and comment on the molecular properties that correlate with viscosity.
- Define *surface tension* and explain its cause.
- State the major factors that determine the extent to which a liquid will *wet* a solid surface.
- Explain what a *surfactant* is, and how it reduces the surface tension of water and aids in cleaning.
- Explain the origins of *capillary rise* and indicate the major factors that affect it.
- Describe the structure of a *soap bubble*, and comment on the role of the "soap" molecules in stabilizing it.
- Comment on the applicability of the term "structure" when describing a pure liquid phase.

The molecular units of a liquid, like those of solids, are in direct contact, but never for any length of time and in the same locations. Whereas the molecules or ions of a solid maintain the same average positions, those of liquids are continually jumping and sliding to new ones, giving liquids something of the mobility of gases. From the standpoint of chemistry, this represents the best of two worlds; rapid chemical change requires intimate contact between the agents undergoing reaction, but these agents, along with the reaction products, must be free to move away to allow new contacts and further reaction to take place. This is why so much of what we do with chemistry takes place in the liquid phase.

Liquids occupy a rather peculiar place in the trinity of solid, liquid and gas. A liquid is the preferred state of a substance at temperatures intermediate between the realms of the solid and the gas. Howeve, if one look at the melting and boiling points of a variety of substances Figure 7.4.1, you will notice that the temperature range within which many liquids can exist tends to be rather small. In this, and in a number of other ways, the liquid state appears to be somewhat tenuous and insecure, as if it had no clear right to exist at all, and only does so as an oversight of Nature. Certainly the liquid state is the most complicated of the three states of matter to analyze and to understand. But just as people whose personalities are more complicated and enigmatic are often the most interesting ones to know, it is these same features that make the liquid state of matter the most fascinating to study.

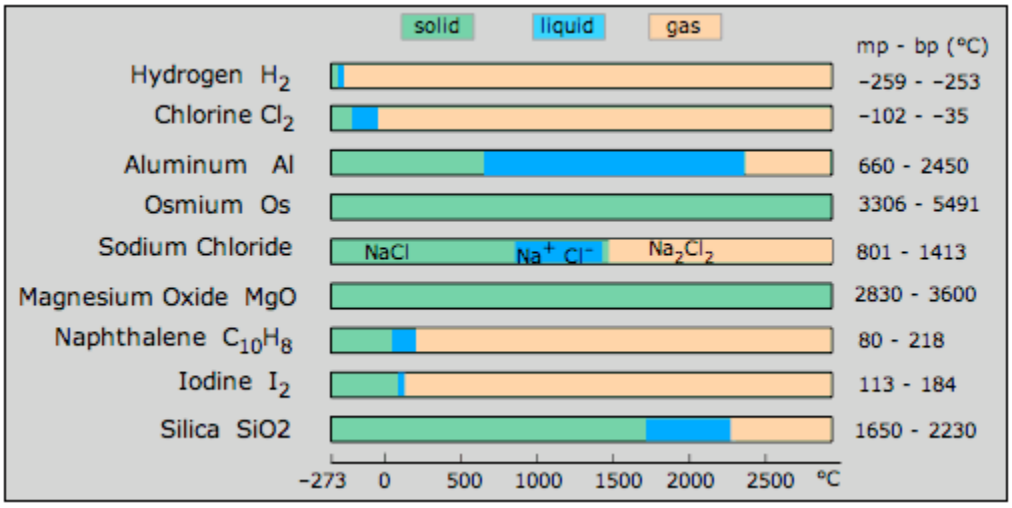


Figure 7.4.1: Phase ranges for materials under Atomispheric Pressure.

Anyone can usually tell if a substance is a liquid simply by looking at it. What special physical properties do liquids possess that make them so easy to recognize? One obvious property is their *mobility*, which refers to their ability to move around, to change their shape to conform to that of a container, to flow in response to a pressure gradient, and to be displaced by other objects. But these properties are shared by gases, the other member of the two *fluid states* of matter. The real giveaway is that a liquid occupies a fixed volume, with the consequence that a liquid possesses a definite *surface*. Gases, of course, do not; the volume and shape of a gas are simply those of the container in which it is confined. The higher density of a liquid also plays a role here;

it is only because of the large density difference between a liquid and the space above it that we can see the surface at all. (What we are really seeing are the effects of reflection and refraction that occur when light passes across the boundary between two phases differing in density, or more precisely, in their *refractive indexes*.)

Viscosity: Resistance to Flow

The term *viscosity* is a measure of resistance to flow. It can be measured by observing the time required for a given volume of liquid to flow through the narrow part of a *viscometer* tube. The viscosity of a substance is related to the strength of the forces acting between its molecular units. In the case of water, these forces are primarily due to hydrogen bonding. Liquids such as syrups and honey are much more viscous because the sugars they contain are studded with hydroxyl groups (–OH) which can form multiple hydrogen bonds with water and with each other, producing a sticky disordered network.

Table 7.4.1: Specific viscosity (i.e., relative to water) of some liquids at 20°C.

substance	viscosity
water H(OH)	1.00
diethyl ether $(CH_3\text{-}CH_2)_2O$	0.23
benzene C_6H_6	0.65
glycerin $C_3H_2(OH)_3$	280
mercury	1.5
motor oil, SAE30	200
honey	~10,000
molasses	~5000
pancake syrup	~3000

Even in the absence of hydrogen bonding, dispersion forces are universally present (as in mercury). Because these forces are additive, they can be very significant in long carbon-chain molecules such as those found in oils used in cooking and for lubrication. Most "straight-chain" molecules are really bent into complex shapes, and dispersion forces tend to preserve their spaghetti-like entanglements with their neighbors.

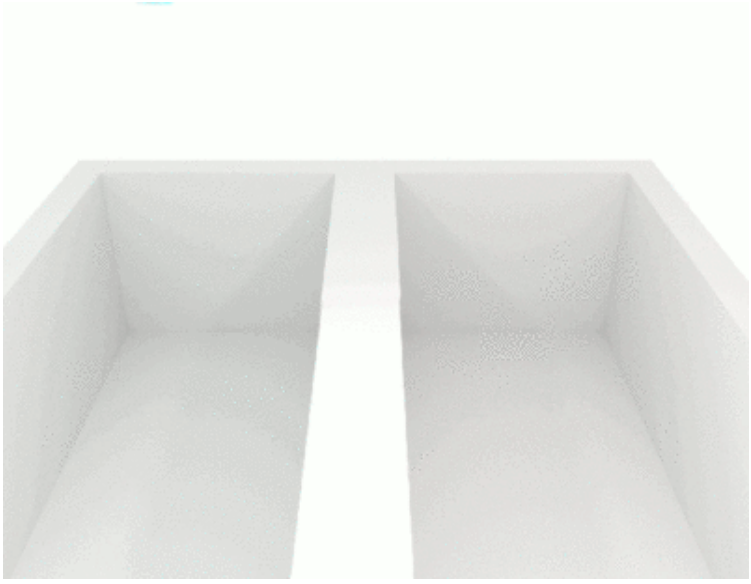


Figure 7.4.2: Viscosity demonstration. The fluid on the left has a lower viscosity than the fluid on the right. (CC SA-BY 4.0; Synapticrelay).

The temperature dependence of the viscosity of liquids is well known to anyone who has tried to pour cold syrup on a pancake. Because the forces that give rise to viscosity are weak, they are easily overcome by thermal motions, so it is no surprise that

viscosity decreases as the temperature rises.

Table 7.4.2: Viscosity of Water as a Function of Temperature

T/°C	0	10	20	40	60	80	100
viscosity/cP	1.8	1.3	1.0	0.65	0.47	0.36	0.28

Automotive lubricating oils can be too viscous at low temperatures (making it harder for your car to operate on a cold day), while losing so much viscosity at engine operating temperatures that their lubricating properties become impaired. These engine oils are sold in a wide range of viscosities; the higher-viscosity oils are used in warmer weather and the lower-viscosity oils in colder weather. The idea is to achieve a fairly constant viscosity that is ideal for the particular application. By blending in certain ingredients, lubricant manufacturers are able to formulate “multigrade” oils whose viscosities are less sensitive to temperatures, thus making a single product useful over a much wider temperature range.

The next time you pour a viscous liquid over a surface, notice how different parts of the liquid move at different rates and sometimes in different directions. To flow freely, the particles making up a fluid must be able to move independently. Intermolecular attractive forces work against this, making it difficult for one molecule to pull away from its neighbors and force its way in between new neighbors.

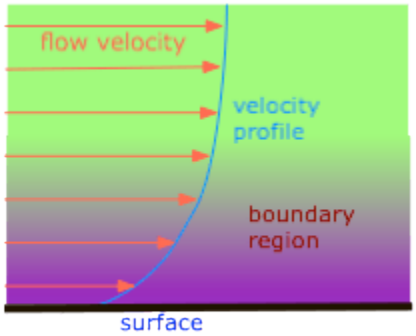


Figure 7.4.3: How viscosity Impedes Flow. via laminar flow.

The pressure drop that is observed when a liquid flows through a pipe is a direct consequence of viscosity. Those molecules that happen to find themselves near the inner walls of a tube tend to spend much of their time attached to the walls by intermolecular forces, and thus move forward very slowly. Movement of the next layer of molecules is impeded as they slip and slide over the slow-movers; this process continues across successive layers of molecules as we move toward the center of the tube, where the velocity is greatest. This effect is called *viscous drag*, and is directly responsible for the pressure drop that can be quite noticeable when you are taking a shower bath and someone else in the house suddenly turns on the water in the kitchen.

Liquids and gases are both *fluids* and exhibit resistance to flow through a confined space. However, it is interesting (and not often appreciated) that their viscosities have entirely different origins, and that they vary with temperature in opposite ways. Why should the viscosity of a gas *increase* with temperature?

Surface Tension

A molecule within the bulk of a liquid experiences attractions to neighboring molecules in all directions, but since these average out to zero, there is no net force on the molecule because it is, on the average, as energetically comfortable in one location within the liquid as in another. Liquids ordinarily do have surfaces, however, and a molecule that finds itself in such a location is attracted to its neighbors below and to either side, but there is no attraction operating in the 180° solid angle above the surface. As a consequence, a molecule at the surface will tend to be drawn into the bulk of the liquid. Conversely, *work* must be done in order to move a molecule within a liquid to its surface.

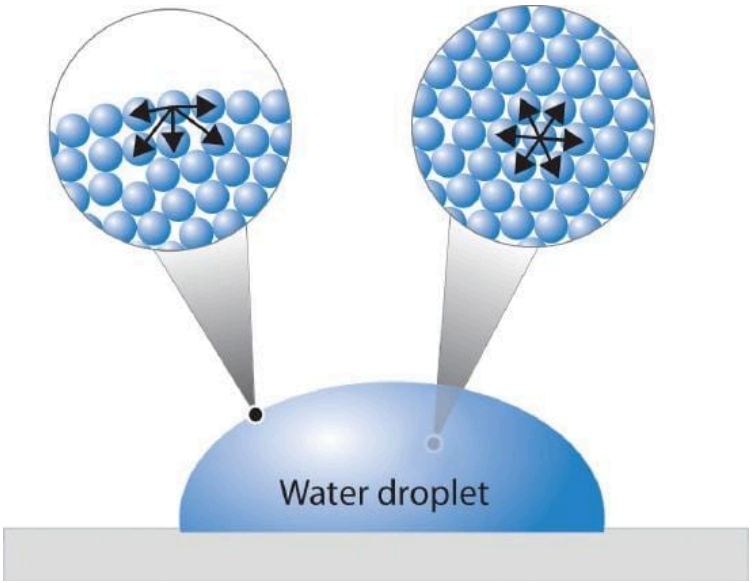


Figure 7.4.4: Molecules at the surface of water experience a net attraction to other molecules in the liquid, which holds the surface of the bulk sample together. In contrast, those in the interior experience uniform attractive forces..

Clearly there must always be some molecules at the surface, but the smaller the surface area, the lower the potential energy. Thus **intermolecular attractive forces act to minimize the surface area of a liquid**. The geometric shape that has the smallest ratio of surface area to volume is the *sphere*, so very small quantities of liquids tend to form spherical drops. As the drops get bigger, their weight deforms them into the typical tear shape.

> Think of a bubble as a hollow drop. Surface tension acts to minimize the surface, and thus the radius of the spherical shell of liquid, but this is opposed by the pressure of vapor trapped within the bubble.

The imbalance of forces near the upper surface of a liquid has the effect of an elastic film stretched across the surface. You have probably seen water striders and other insects take advantage of this when they walk across a pond. Similarly, you can carefully "float" a light object such as a steel paperclip on the surface of water in a cup.

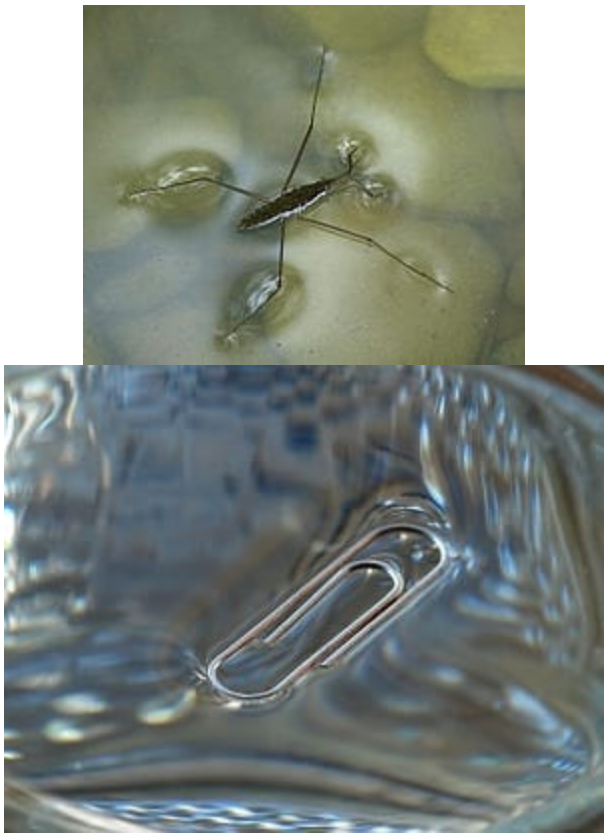


Figure 7.4.5: (left) Strider insect on a sruface of a pond. (right) A floating paperclip. Figures used with permission from Wikipedia.

Surface tension is defined as the amount of work that must be done in order to create unit area of surface. The SI units are J m^{-2} (or N m^{-1}), but values are more commonly expressed in mN m^{-1} or in cgs units of dyn cm^{-1} or erg cm^{-2}. Table 7.4.3 compares the surface tensions of several liquids at room temperature. Note especially that:

- hydrocarbons and non-polar liquids such as ether have rather low values
- one of the main functions of soaps and other surfactants is to reduce the surface tension of water
- mercury has the highest surface tension of any liquid at room temperature. It is so high that mercury does not flow in the ordinary way, but breaks into small droplets that roll independently.

Table 7.4.3*: Surface Tensions of select liquids*

substance	surface tension (dyne/cm)
water H(OH)	72.7
diethyl ether $(CH_3\text{-}CH_2)_2O$	17.0
benzene C_6H_6	40.0
glycerin $C_3H_2(OH)_3$	63
mercury (15°C)	487
n-octane	21.8
sodium chloride solution (6*M* in water)	82.5
sucrose solution (85% in water)	76.4
sodium oleate (soap) solution in water	25

Surface tension and viscosity are not directly related, as you can verify by noting the disparate values of these two quantities for mercury. Viscosity depends on intermolecular forces within the liquid, whereas surface tension arises from the difference in the magnitudes of these forces within the liquid and at the surface. Surface tension is also affected by the electrostatic charge of a body. This is most dramatically illustrated by the famous **"mercury beating heart"** demonstration..

Surface tension always decreases with temperature as thermal motions reduce the effect of intermolecular attractions (Table 7.4.4). This is one reason why washing with warm water is more effective; the lower surface tension allows water to more readily penetrate a fabric.

Table 7.4.4: Surface Tension of water as a Function of Temperature

°C	dynes/cm
0	75.9
20	72.7
50	67.9
100	58.9

Why do "tears" form inside a wine glass? You have undoubtedly noticed this; pour some wine into a glass, and after a few minutes, droplets of clear liquid can be seen forming on the inside walls of the glass about a centimeter above the level of the wine. This happens even when the wine and the glass are at room temperature, so it has nothing to do with condensation.



The explanation involves Raoult's law, hydrogen bonding, adsorption, and surface tension, so this phenomenon makes a good review of much you have learned about liquids and solutions. The tendency of a surface tension gradient to draw water into the region of higher surface tension is known as the Maringoni effect

First, remember that both water and alcohol are hydrogen-bonding liquids; as such, they are both strongly attracted to the oxygen atoms and -OH groups on the surface of the glass. This causes the liquid film to creep up the walls of the glass. Alcohol, the more volatile of the two liquids, vaporizes more readily, causing the upper (and thinnest) part of the liquid film to become enriched in water. Because of its stronger hydrogen bonding, water has a larger surface tension than alcohol, so as the alcohol evaporates, the surface tension of the upper part of the liquid film increases. This that part of the film draw up more liquid and assume a spherical shape which gets distorted by gravity into a "tear", which eventually grows so large that gravity wins out over adsorption, and the drop falls back into the liquid, soon to be replaced by another.

Interfacial effects in liquids

The surface tension discussed immediately above is an attribute of a liquid in contact with a gas (ordinarily the air or vapor) or a

vacuum. But if you think about it, the molecules in the part of a liquid that is in contact with any other phase (liquid or solid) will experience a different balance of forces than the molecules within the bulk of the liquid. Thus surface tension is a special case of the more general *interfacial tension* which is defined by the work associated with moving a molecule from within the bulk liquid to the interface with any other phase.

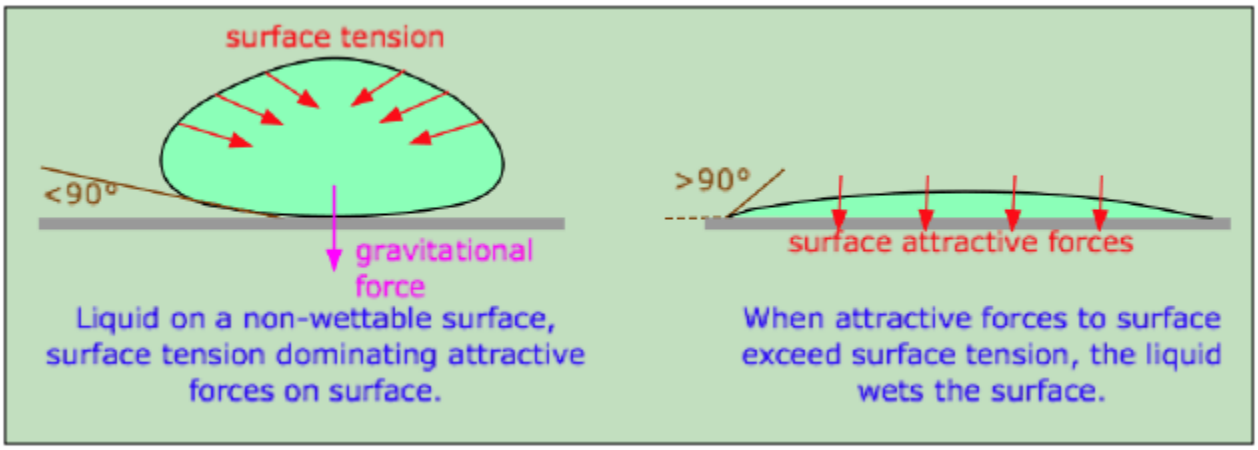
Wetting

Take a plastic mixing bowl from your kitchen, and splash some water around in it. You will probably observe that the water does not cover the inside surface uniformly, but remains dispersed into drops. The same effect is seen on a dirty windshield; running the wipers simply breaks hundreds of drops into thousands. By contrast, water poured over a clean glass surface will wet it, leaving a uniform film.



Figure 7.4.4: This photo of water drops on a leaf shows how the waxy coating naturally present on most leaves prevents its wetting. Images used with permission from Wikipedia.

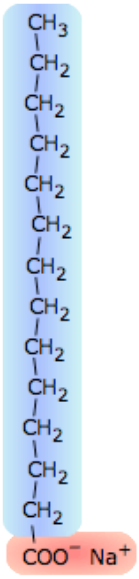
When a molecule of a liquid is in contact with another phase, its behavior depends on the relative attractive strengths of its neighbors on the two sides of the phase boundary. If the molecule is more strongly attracted to its own kind, then interfacial tension will act to minimize the area of contact by increasing the curvature of the surface. This is what happens at the interface between water and a hydrophobic surface such as a plastic mixing bowl or a windshield coated with oily material. A liquid will wet a surface if the angle at which it makes contact with the surface is less than 90°. The value of this contact angle can be predicted from the properties of the liquid and solid separately.

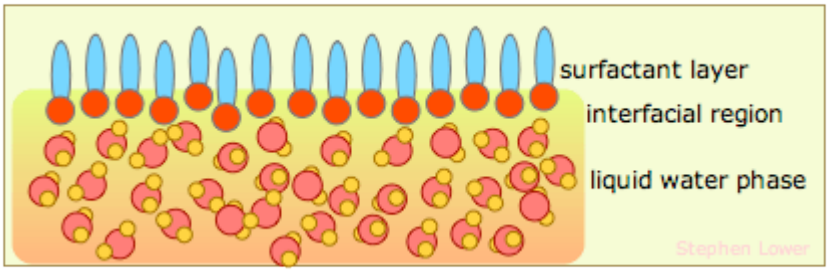


A clean glass surface, by contrast, has –OH groups sticking out of it which readily attach to water molecules through hydrogen bonding; the lowest potential energy now occurs when the contact area between the glass and water is maximized. This causes the water to spread out evenly over the surface, or to *wet* it.

Surfactants

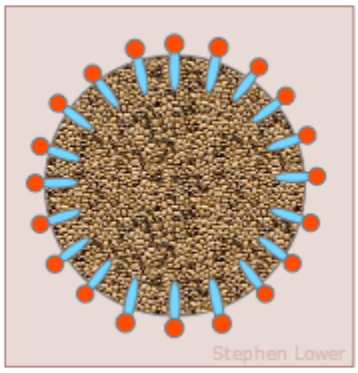
The surface tension of water can be reduced to about one-third of its normal value by adding some soap or synthetic detergent. These substances, known collectively as *surfactants*, are generally hydrocarbon molecules having an ionic group on one end. The ionic group, being highly polar, is strongly attracted to water molecules; we say it is *hydrophilic*. The hydrocarbon (*hydrophobic*) portion is just the opposite; inserting it into water would break up the local hydrogen-bonding forces and is therefore energetically unfavorable. What happens, then, is that the surfactant molecules migrate to the surface with their hydrophobic ends sticking out, effectively creating a new surface. Because hydrocarbons interact only through very weak dispersion forces, this new surface has a greatly reduced surface tension.





Washing

How do soaps and detergents help get things clean? There are two main mechanisms. First, by reducing water's surface tension, the water can more readily penetrate fabrics (see the illustration under "Water repellency" below.) Secondly, much of what we call "dirt" consists of non-water soluble oils and greasy materials which the hydrophobic ends of surfactant molecules can penetrate. When they do so in sufficient numbers and with their polar ends sticking out, the resulting aggregate can hydrogen-bond to water and becomes "solubilized".

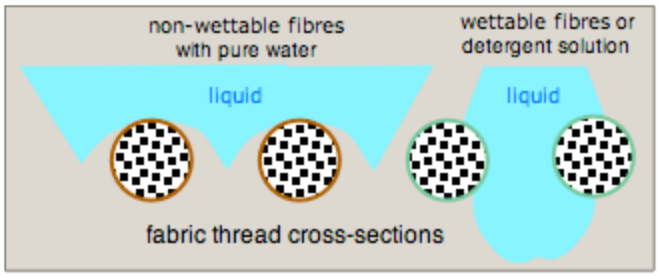


Washing is usually more effective in warm water; higher temperatures reduce the surface tension of the water and make it easier for the surfactant molecules to penetrate the material to be removed.

The answer is no, but claims that they can are widely circulated in promotions of dubious products such as "magnetic laundry disks" which are supposed to reduce the need for detergents.

Water repellency

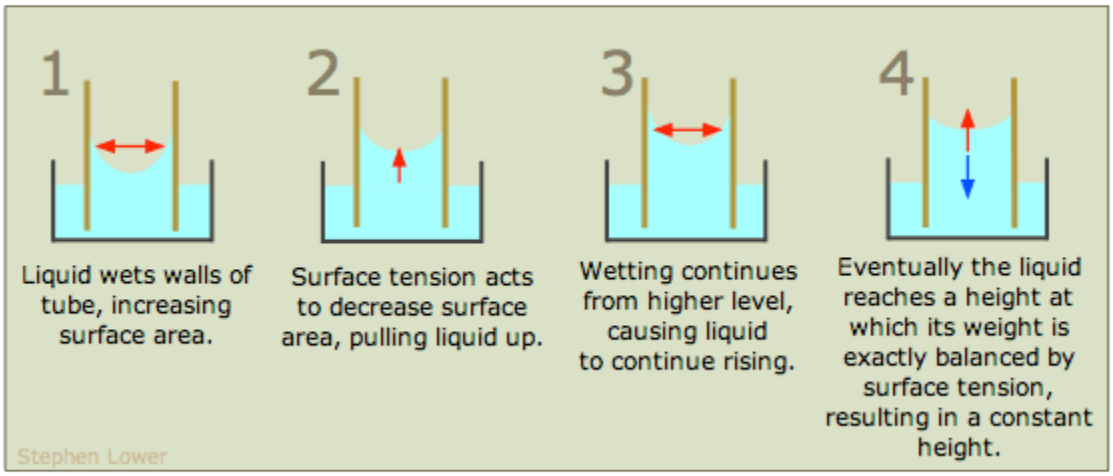
In Gore-Tex, one of the more successful water-proof fabrics, the fibers are made non-wettable by coating them with a Teflon-like fluoropolymer.



Water is quite strongly attracted to many natural fibers such as cotton and linen through hydrogen-bonding to their cellulosic hydroxyl groups. A droplet that falls on such a material will flatten out and be drawn through the fabric. One way to prevent this is to coat the fibers with a polymeric material that is not readily wetted. The water tends to curve away from the fibers so as to minimize the area of contact, so the droplets are supported on the gridwork of the fabric but tend not to fall through.

Capillary rise

If the walls of a narrow tube can be efficiently wetted by a liquid, then the the liquid will be drawn up into the tube by *capillary action*. This effect is only noticeable in narrow containers (such as burettes) and especially in small-diameter *capillary tubes*. The smaller the diameter of the tube, the higher will be the capillary rise. A clean glass surface is highly attractive to most molecules, so most liquids display a concave *meniscus* in a glass tube.



To help you understand capillary rise, the above diagram shows a glass tube of small cross-section inserted into an open container of water. The attraction of the water to the inner wall of the tube pulls the edges of the water up, creating a curved *meniscus* whose surface area is smaller than the cross-section area of the tube. The surface tension of the water acts against this enlargement of its surface by attempting to reduce the curvature, stretching the surface into a flatter shape by pulling the liquid farther up into the tube. This process continues until the weight of the liquid column becomes equal to the surface tension force, and the system reaches mechanical equilibrium.

Capillary rise results from a combination of two effects: the tendency of the liquid to wet (bind to) the surface of the tube (measured by the value of the contact angle), and the action of the liquid's surface tension to minimize its surface area.

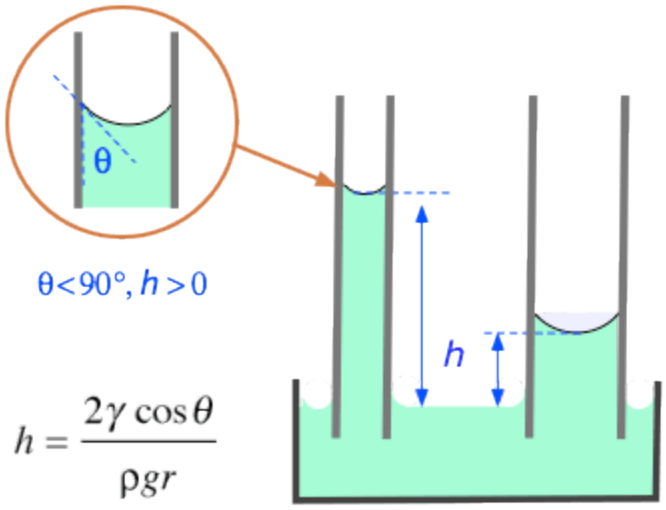
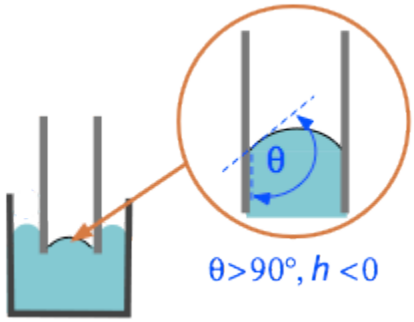


Figure 7.4.4:

In the formula shown at the left (which you need not memorize!)

h = elevation of the liquid (m)
γ = surface tension (N/m)
θ = contact angle (radians)
ρ = density of liquid (kg/m^3)
g = acceleration of gravity (m/s^{-2})
r = radius of tube (m)

The contact angle between water and ordinary soda-lime glass is essentially zero; since the cosine of 0 radians is unity, its capillary rise is especially noticable. In general, water can be drawn very effectively into narrow openings such as the channels between fibers in a fabric and into porous materials such as soils.



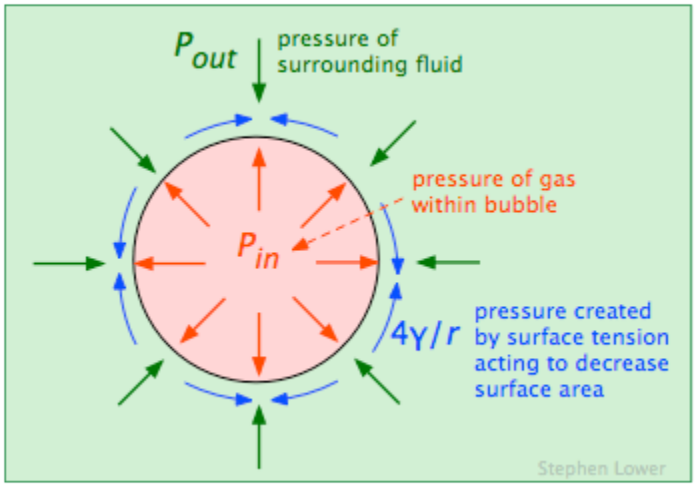
Note that if θ is greater than 90° (π/2 radians), the capillary "rise" will be negative — meaning that the molecules of the liquid are more strongly attracted to each other than to the surface. This is readily seen with mercury in a glass container, in which the meniscus is upwardly convex instead of concave.

Capillary rise is the principal mechanism by which water is able to reach the highest parts of trees. Water strongly bonds to the narrow (25 μM) cellulose channels in the xylem. (Osmotic pressure and "suction" produced by loss of water vapor through the leaves also contribute to this effect, and are the main drivers of water flow in smaller plants.)

Bubbles

Bubbles can be thought of as "negative drops" — spherical spaces within a liquid containing a gas, often just the vapor of the liquid. Bubbles within pure liquids such as water (which we see when water boils) are inherently unstable because the liquid's

surface tension causes them to collapse. But in the presence of a surfactant, bubbles can be stabilized and given an independent if evanescent existence.



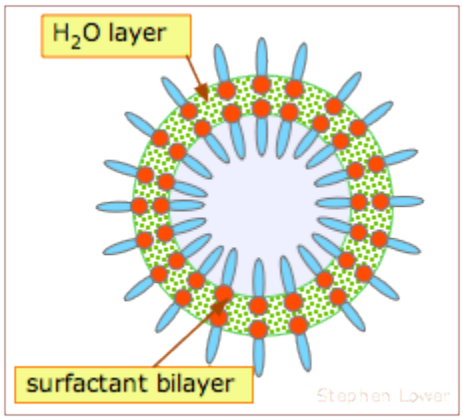
The pressure of the gas inside a bubble P_{in} must be sufficient to oppose the pressure outside of it (P_{out}, the atmospheric pressure plus the hydrostatic pressure of any other fluid in which the bubble is immersed. But the force caused by surface tension γ of the liquid boundary also tends to collapse the bubble, so P_{in} must be *greater* than P_{out} by the amount of this force, which is given by 4γ/*r*:

$$P_{in} = P_{out} + \frac{4\gamma}{r} \tag{7.4.1}$$

The most important feature of this relationship (known as *LaPlace's law*) is the that the pressure required to maintain the bubble is inversely proportional to its radius. This means that the *smallest* bubbles have the *greatest* internal gas pressures! This might seem counterintuitive, but if you are an experienced soap-bubble blower, or have blown up a rubber balloon (in which the elastic of the rubber has an effect similar to the surface tension in a liquid), you will have noticed that you need to puff harder to begin the expansion.

Soap bubbles

All of us at one time or another have enjoyed the fascination of creating soap bubbles and admiring their intense and varied colors as they drift around in the air, seemingly aloof from the constraints that govern the behavior of ordinary objects — but only for a while! Their life eventually comes to an abrupt end as they fall to the ground or pop in mid-flight.



The walls of these bubbles consist of a thin layer of water molecules sandwiched between two layers of surfactant molecules. Their spherical shape is of course the result of water's surface tension. Although the surfactant (soap) initially reduces the surface tension, expansion of the bubble spreads the water into a thinner layer and spreads the surfactant molecules over a wider area,

deceasing their concentration. This, in turn, allows the water molecules to interact more strongly, increasing its surface tension and stabilizing the bubble as it expands.

The bright colors we see in bubbles arises from interference between light waves that are reflected back from the inner and outer surfaces, indicating that the thickness of the water layer is comparable the range of visible light (around 400-600 nm).

Once the bubble is released, it can endure until it strikes a solid surface or collapses owing to loss of the water layer by evaporation. The latter process can be slowed by adding a bit of glycerine to the liquid. A variety of recipes and commercial "bubble-making solutions" are available; some of the latter employ special liquid polymers which slow evaporation and greatly extend the bubble lifetimes. Bubbles blown at very low temperatures can be frozen, but these eventually collapse as the gas diffuses out.

Bubbles, surface tension, and breathing

The sites of gas exchange with the blood in mammalian lungs are tiny sacs known as *alveoli*. In humans there are about 150 million of these, having a total surface area about the size of a tennis court. The inner surface of each alveolus is about 0.25 mm in diameter and is coated with a film of water, whose high surface tension not only resists inflation, but would ordinarily cause the thin-walled alveoli to collapse. In order to counteract this effect, special cells in the alveolar wall secrete a phospholipid *pulmonary surfactant* that reduces the surface tension of the water film to about 35% of its normal value. But there is another problem: the alveoli can be regarded physically as a huge collection of interconnected bubbles of varying sizes. As noted above, the surface tension of a surfactant-stabilized bubble increases with their size. So by making it easier for the smaller alveoli to expand while inhibiting the expansion of the larger ones, the surfactant helps to equalize the volume changes of all the alveoli as one inhales and exhales.

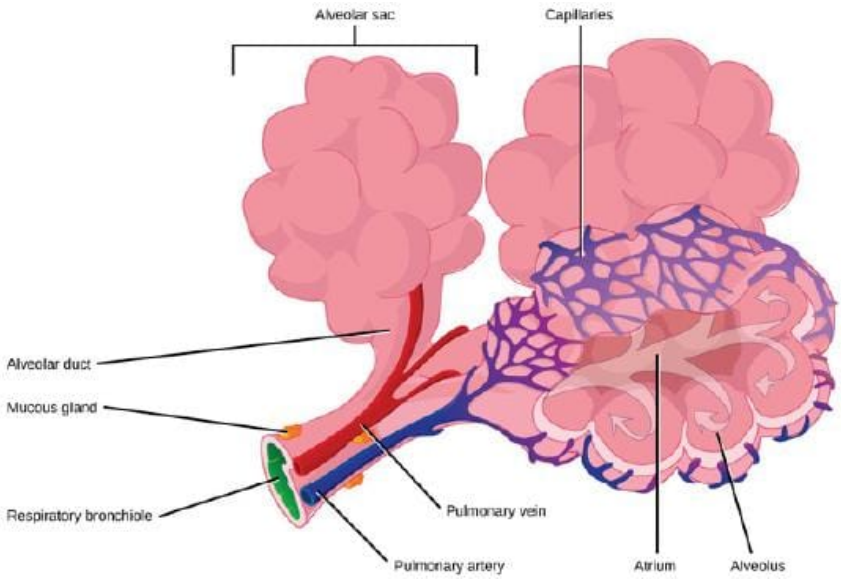
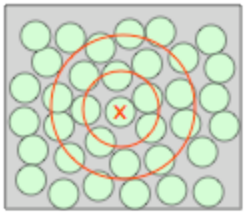


Figure 7.4.5:

Pulmonary surfactant is produced only in the later stages of fetal development, so premature infants often do not have enough and are subject to respiratory distress syndrome which can be fatal.

Structure of liquids

You can think of a simple liquid such as argon or methane as a collection of loosely-packed marbles that can assume various shapes.Although the overall arrangement of the individual molecular units is entirely random, there is a certain amount of short-range order: the presence of one molecule at a given spot means that the neighboring molecules must be at least as far away as the sum of the two radii, and this in turn affects the possible locations of more distant concentric shells of molecules.



An important consequence of the disordered arrangement of molecules in a liquid is the presence of void spaces. These, together with the increased kinetic energy of colliding molecules which helps push them apart, are responsible for the approximately 15-percent decrease in density that is observed when solids based on simple spherical molecules such as Ne and Hg melt into liquids. These void spaces are believed to be the key to the flow properties of liquids; the more “holes” there are in the liquid, the more easily the molecules can slip and slide over one another.

As the temperature rises, thermal motions of the molecules increase and the local structure begins to deteriorate, as shown in the plots below.

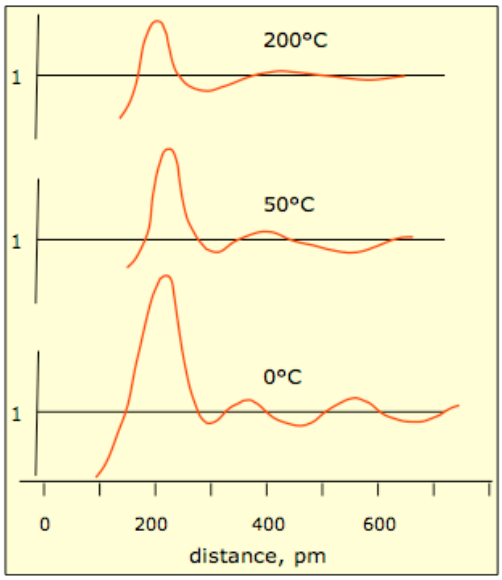
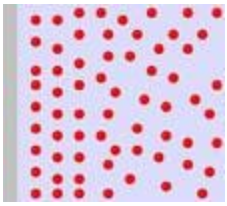


Figure 7.4.6: There is very little experimental information on the structure of liquids, other than the X-ray diffraction studies that yield plots such as this one for liquid mercury.

This plot shows the relative probability of finding a mercury atom at a given distance from another atom located at distance 0. You can see that as thermal motions increase, the probabilities even out at greater distances. It is very difficult to design experiments that yield the kind of information required to define the microscopic arrangement of molecules in the liquid state.

Many of our current ideas on the subject come from computer simulations based on hypothetical models. In a typical experiment, the paths of about 1000 molecules in a volume of space are calculated. The molecules are initially given random kinetic energies whose distribution is consistent with the Boltzmann distribution for a given temperature. The trajectories of all the molecules are followed as they change with time due to collisions and other interactions; these interactions must be calculated according to an assumed potential energy-vs.-distance function that is part of the particular model being investigated.

These computer experiments suggest that whatever structure simple liquids do possess is determined mainly by the repulsive forces between the molecules; the attractive forces act in a rather nondirectional, general way to hold the liquid together. It is also found that if spherical molecules are packed together as closely as geometry allows (in which each molecule would be in contact with twelve nearest neighbors), the collection will have a long-range order characteristic of a solid until the density is decreased by about ten percent, at which point the molecules can slide around and move past one another, thus preserving only short-range order. In recent years, experimental studies based on ultra-short laser flashes have revealed that local structures in liquids have extremely short lifetimes, of the order of picoseconds to nanoseconds.



It has long been suspected that the region of a liquid that bounds a solid surface is more ordered than within the bulk liquid. This has been confirmed for the case of water in contact with silicon, in which the liquid's layers form layers, similar to what is found in liquid crystals.

This page titled 7.4: Liquids and their Interfaces is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.5: Changes of State

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the terms in the context of this topic.

- Describe what is meant by the *escaping tendency* of molecules from a solid, liquid, or gas. What experimentally-observable quantity serves as its measure?
- The terms "*vapor pressure*" and "*pressure of the vapor* above a solid or liquid" are easily confused. Explain the difference between them, and state under what conditions they will have identical values.
- Define *relative humidity* and calculate its value, given the partial pressure of water vapor and a suitable vapor pressure table or plot for water.
- Explain the difference between evaporation and boiling, and why liquids may not begin to boil until the temperature exceeds the boiling point.
- Given a *phase diagram* of a pure substance, label all of the lines and the regions they enclose, identify the *normal melting and boiling points*, the *triple point* and the *critical point*, and state the physical significance of the latter two.
- Conversely, sketch out a properly-labeled phase diagram for a pure substance, given the parameters mentioned above, along with information about the relative densities of the solid and liquid phases.

A given substance will exist in the form of a solid, liquid, or gas, depending on the temperature and pressure. In this unit, we will learn what common factors govern the preferred state of matter under a particular set of conditions, and we will examine the way in which one phase gives way to another when these conditions change.

Phase Stability

Earlier in the morning, the droplets of water in Figure 7.5.1 were tiny crystals of ice, but even though the air temperature is still around 0°C and will remain so all day, the sun's warmth has rendered them into liquid form, bound up by surface tension into reflective spheres. By late afternoon, most of the drops will be gone, their H_2O molecules now dispersed as a tenuous atmospheric gas.



Figure 7.5.1: Dew on a spider's web in the morning. Credit (Luc Viatour / www.Lucnix.be).

Solid, liquid, and gas — these are the basic three states, or *phases*, in which the majority of small-molecule substances can exist. At most combinations of pressure and temperature, only one of these phases will be favored; this is the phase that is most

thermodynamically stable under these conditions. A proper explanation of why most substances have well-defined melting and boiling points needs to invoke some principles of thermodynamics and quantum mechanics. A full explanation of this would go beyond the scope of what most students who see this lesson are familiar with, but the following greatly over-simplified explanation should convince you that it is something more than black magic.

All atoms and molecules at temperatures above absolute zero possess thermal energy that keeps them in constant states of motion. A fundamental law of nature mandates that this energy tends to spread out and be shared as widely as possible. Within a single molecular unit, this spreading and sharing can occur by dispersing the energy into the many allowed states of motion (translation, vibration, rotation) of the molecules of the substance itself. There are a huge number of such states, and they are quantized, meaning that they all require different amounts of thermal energy to come into action. Temperature is a measure of the intensity of thermal energy, so the higher the temperature, the greater will be the number of states that can be active, and the more extensively will the energy be dispersed among these allowed states.

In solids, the molecular units are bound into fixed locations, so the kinds of motion (and thus the number of states) that can be thermally activated is relatively small. Because the molecules of solids possess the lowest potential energies, solids are the most stable states at low temperatures. At the other extreme are gas molecules which are not only free to vibrate and rotate, but are in constant translational motion. The corresponding number of quantum states is hugely greater for gases, providing a nearly-endless opportunity to spread energy. But this can only happen if the temperature is high enough to populate this new multitude of states. Once it does, the gaseous state wins out by a landslide.

Escaping Tendency and Vapor Pressure

Escaping tendency is more formally known as *free energy*. Bear in mind also that changes in state always involve changes in enthalpy and internal energy. In much the same way that tea spreads out from a tea bag into the larger space of the water in which it is immersed, molecules that are confined within a phase (liquid, solid, or gas) will tend to spread themselves (and the thermal energy they carry with them) as widely as possible. This fundamental law of nature is manifested in what we will call the *escaping tendency* of the molecules from the phase. The escaping tendency is a quantity of fundamental importance in understanding all chemical equilibria and transformations. We need not define the term in a formal way at this point. What is important for now is how we can observe and compare escaping tendencies.

Think first of a **gas**: what property of the gas constitutes the best measure of its tendency to escape from a container? It does not require much reflection to conclude that the greater the *pressure* of the gas, the more frequently will its molecules collide with the walls of the container and possibly find their way through an opening to the outside. **Thus the pressure confining a gas is a direct measure of the tendency of molecules to escape from a gaseous phase.**

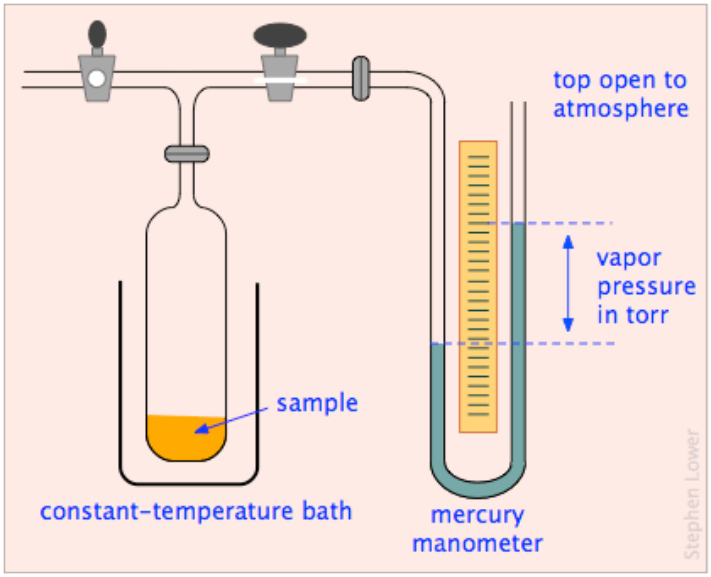


Figure 7.5.2: How to measure the Vapor Pressure of a sample.

What about liquids and solids? Although we think of the molecules of condensed phases as permanently confined within them, these molecules still possess some thermal energy, and there is always a chance that one that is near the surface will occasionally

fly loose and escape into the space outside the solid or liquid. We can observe the tendency of molecules to escape into the gas phase from a solid or liquid by placing the substance in a closed, evacuated container connected to a manometer for measuring gas pressure (Figure 7.5.2).

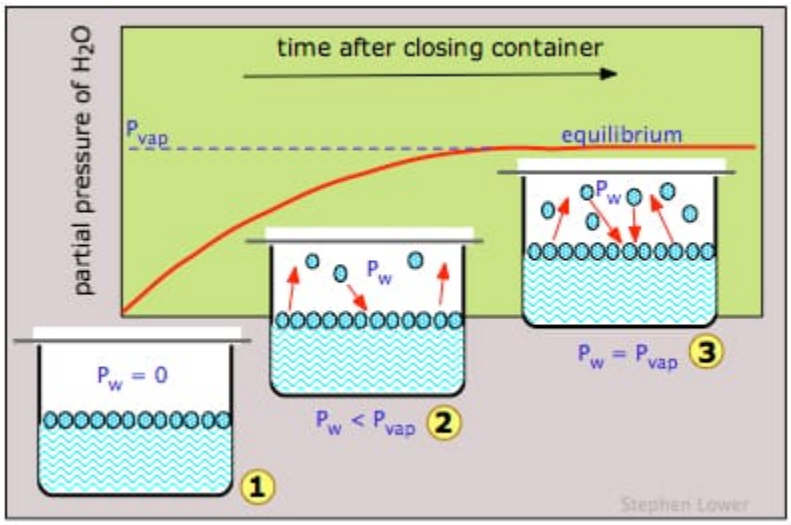


Figure 7.5.3:

If we do this for water (Figure 7.5.3), the partial pressure of water P_W in the vapor space will initially be zero (**1**). Gradually, P_W will rise as molecules escape from the substance and enter the vapor phase. But at the same time, some of the vapor molecules will "escape" back into the liquid phase (**2**). But because this latter process is less favorable (at the particular temperature represented here), P_W continues to rise. Eventually a balance is reached between the two processes (**3**), and P_W eventually stabilizes at a fixed value P_{vap} that depends on the substance and on the temperature and is known as the *equilibrium vapor pressure*, or simply as the "vapor pressure" of the liquid or solid. **The vapor pressure is a direct measure of the escaping tendency of molecules from a condensed state of matter.**

Note carefully that if the container is left open to the air, it is unlikely that many of the molecules in the vapor phase will return to the liquid phase. They will simply escape from the entire system and the partial pressure of water vapor P_W will never reach P_{vap}; the liquid will simply *evaporate* without any kind of equilibrium ever being achieved.

The escaping tendency of molecules from a phase always increases with the temperature; therefore the vapor pressure of a liquid or solid will be greater at higher temperatures. As Figure 7.5.4 shows, the variation of the vapor pressure with the temperature is not linear.

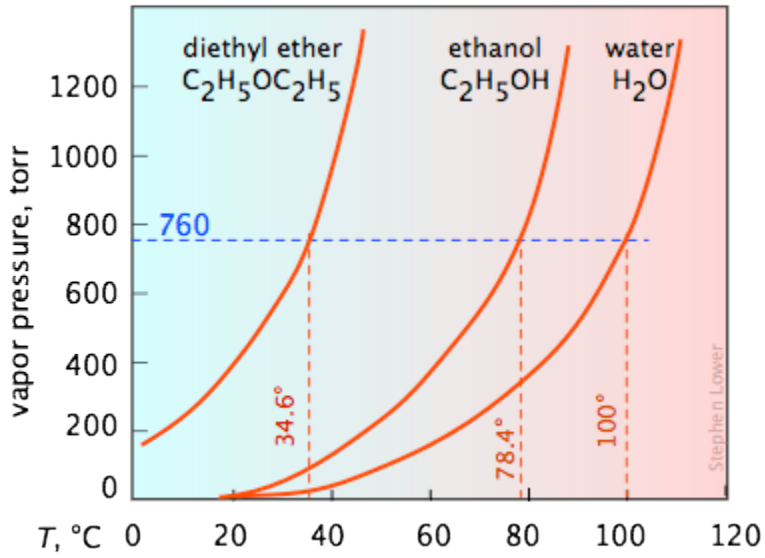


Figure 7.5.4:

It's important that you be able to interpret vapor pressure plots such as the three shown here. Take special note of how boiling points can be found from these plots. You will recall that the *normal boiling point* is the temperature at which the liquid is in equilibrium with its vapor at a partial pressure of 1 atm (760 torr). Thus the intercepts of each curve with the blue dashed 760-torr line indicate the normal boiling points of each liquid. Similarly, you can easily estimate the boiling points these liquids would have in Denver, Colorado where the atmospheric pressure is 630 torr by simply constructing a horizontal line corresponding to this pressure.

Definition: The Normal Boiling Point

The *normal boiling point* is the temperature at which the liquid is in equilibrium with its vapor at a partial pressure of 1 atm. This is when the vapor pressure is at atmospheric pressure.

Vapor pressure of water

The great importance of H_2O in our world merits a more detailed look at its vapor pressure properties. Because the vapor pressure of water varies greatly over the range of temperatures in which the liquid can exist (Figure 7.5.5).

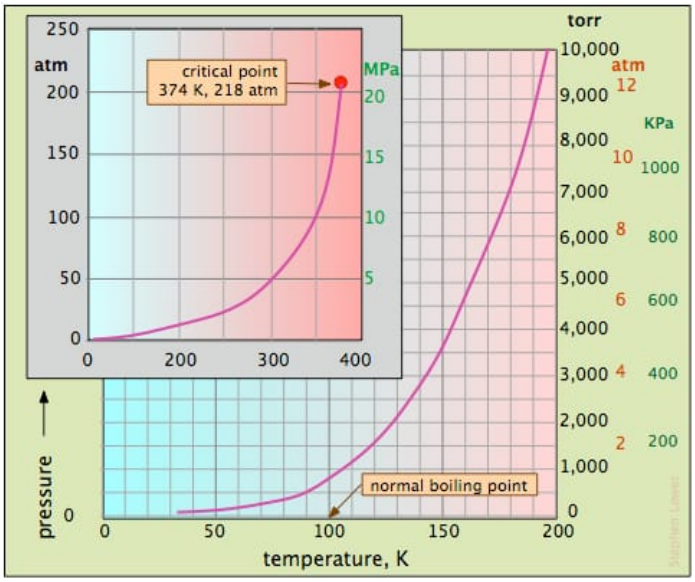


Figure 7.5.5:

The larger plot in Figure 7.5.5 covers the lowest temperatures, while the inset shows the complete range of pressure values. Note particularly that

- The *normal boiling point* is the temperature at which the vapor pressure is the same as that of the standard atmosphere, 760 torr.
- The boiling point at any other pressure can be found by dropping a vertical line from the curve to the temperature axis.
- As seen on the inset plot, the vapor pressure curve of water ends at the *critical point*.

Relative humidity

The vapor pressure of water at 22°C is about 20 torr, or around 0.026 atm (2.7 kPa). This is the partial pressure of H_2O that will be found in the vapor space within a closed container of water at this temperature; the air in this space is said to be *saturated* with water vapor. Humid air is sometimes described as "heavy", but this is misleading; the average molar mass of dry air is 29, but that of water is only 18, so humid air is actually less dense. The feeling of "heaviness" probably relates to the reduced ability of perspiration to evaporate in humid air. In ordinary air, the partial pressure of water vapor is normally less than its saturation or equilibrium value. The ratio of the partial pressure of H_2O in the air to its (equilibrium) vapor pressure at any given temperature is known as the ***relative humidity***. Water enters the atmosphere through evaporation from the ocean and other bodies of water, and from water-saturated soils. The resulting vapor tends to get dissipated and diluted by atmospheric circulation, so the relative

humidity rarely reaches 100 percent. When it does and the weather is warm, we are very uncomfortable because vaporization of water from the skin is inhibited; if the air is already saturated with water, then there is no place for our perspiration to go, other than to drip down our face.

Why the indoor air is so dry in the winter?

Because the vapor pressure increases with temperature, a parcel of air containing a fixed partial pressure of water vapor will have a larger relative humidity at low temperatures than at high temperatures. Thus when cold air enters a heated house, its water content remains unchanged but the relative humidity drops. In climates with cold winters, this promotes increased moisture loss from house plants and from mucous membranes, leading to wilting of the former and irritation of the latter.

Example 7.5.1

The vapor pressure of water is 3.9 torr at –2°C and 20 torr at 22°C. What will be the relative humidity inside a house maintained at 22°C when the outside air temperature is –2°C and the relative humidity is 70%?

Solution

At 70 percent relative humidity, the partial pressure of the –2° air is (0.7 × 3.9 torr) = 2.7 torr. When this air enters the house, its relative humidity will be (2.7 torr)/(20 torr) = 0.14 or **14%**.

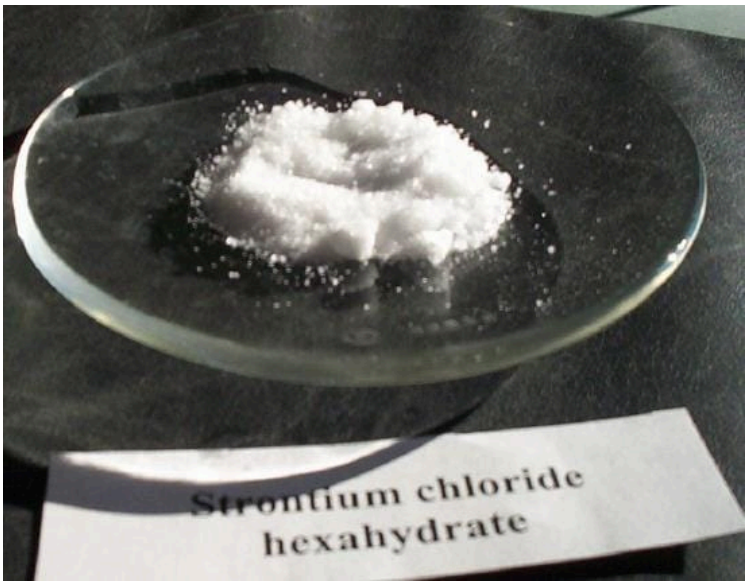
In the evening, especially on clear nights, solid objects (even spider webs!) lose heat to the sky more rapidly than does the air. It is often important to know what temperature such objects must drop to so that atmospheric moisture will condense out on them (Figure 7.5.1). The *dew point* is the temperature at which the relative humidity is 100 percent — that is, the temperature at which the vapor pressure of water becomes equal to its partial pressure at a given [higher] temperature and relative humidity. For water to condense directly out of the atmosphere as rain, the air must be **at or below** the dew point, but this is not of itself the only requirement for the formation of rain, as we will see shortly.

Vapor pressures of Solid Hydrates

Many solid salts incorporate water molecules into their crystal lattices; the resulting compounds are known as *hydrates*. These solid hydrates possess definite vapor pressures that correspond to an equilibrium between the hydrated and anhydrous compounds and water vapor. For example Strontium chloride hexahydrate:

$$\mathrm{SrCl_2 \cdot 6\,H_2O\,(s)} \rightarrow \mathrm{SrCl_2\,(s) + 6\,H_2O\,(g)} \tag{7.5.1}$$

The vapor pressure of this hydrate is 8.4 torr at 25°C. Only at this unique partial pressure of water vapor can the two solids coexist at 25°C. If the partial pressure of water in the air is greater than 8.4 torr, a sample of anhydrous $SrCl_2$ will absorb moisture from the air and change into the hydrate. In fact, when fully hydrates, water is responsible for 40% of the hydrate's mass.



Strontium chloride hexahydrate. from Wikipedia.

Example 7.5.2

What will be the relative humidity of air in an enclosed vessel containing solid $SrCl_2 \cdot 6H_2O$ at 25°C?

Solution

What fraction of the vapor pressure of water at this temperature (23.8 torr) is the vapor pressure of the hydrate (8.4 torr)? Expressed in percent, this is the relative humidity.

If the partial pressure of H_2O in the air is less than the vapor pressure of the hydrate, the latter will tend to lose moisture and revert back to its anhydrous form. This process is sometimes accompanied by a breakup of the crystal into a powdery form, an effect known as *efflorescence*.

Condensation and boiling: Nucleation

Evaporation and boiling of a liquid, and condensation of a gas (vapor) are such ordinary parts of our daily life that we hardly give them a thought. Every time we boil water to make a pot of tea and see the cloud of steam above the teapot, we are observing this most common of all phase changes. How can we understand these changes in terms of vapor pressure?

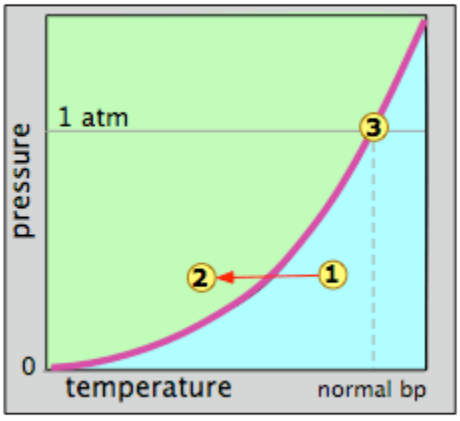


Figure 7.5.6: Condensation and boiling

Figure 7.5.6 plots the vapor-pressure as a function of temperature can represent water or any other liquid. When we say this is a

vapor-pressure plot, we mean that each point on the curve represents a combination of temperature and vapor pressure at which the liquid (green) and the vapor (blue) can coexist. Thus at the normal boiling point, defined as the temperature at which the vapor pressure is 1 atm, the state of the system corresponds to the point labeled **3**.

Suppose that we select an arbitrary point **1** at a temperature and pressure at which only the gaseous state is stable. We then decrease the temperature so as to move the state point toward point **2** in the liquid region. When the state point falls on the vapor pressure line, the two phases can coexist and we would expect some liquid to condense. Once the state point moves to the left of the vapor pressure line, the substance will be entirely in the liquid phase. This is supposedly what happens when "steam" (actually tiny water drops) forms above a pot of boiling water.

The reverse process should work the same way: starting with a temperature in the liquid region, nothing happens until we reach the vapor pressure line, at which point the liquid begins to change into vapor. At higher temperatures, only the vapor remains. This is the *theory*, but it is not complete. The *fact* is that a vapor will generally *not* condense to a liquid at the boiling point (also called the condensation point or dew point), and a liquid will generally *not* boil at its boiling point.

Bubbles and Drops

The reason for the discrepancy is that the vapor pressure, as we normally use the term and as it is depicted by the liquid-vapor line on the phase diagram, refers to the partial pressure of vapor in equilibrium with a liquid whose surface is reasonably flat, as it would be in a partially filled container. In a drop of liquid or in a bubble of vapor within a liquid, the surface of the liquid is not flat, but curved. For drops or bubbles that are of reasonable size, this does not make much difference, but these drops and bubbles must grow from smaller ones, and these from tinier ones still. Eventually, one gets down to the primordial drops and bubbles having only a few molecular dimensions, and it is here that we run into a problem: this is the problem of *nucleation* — the formation and growth of the first tiny drop (in the vapor) or of a bubble (in a liquid).

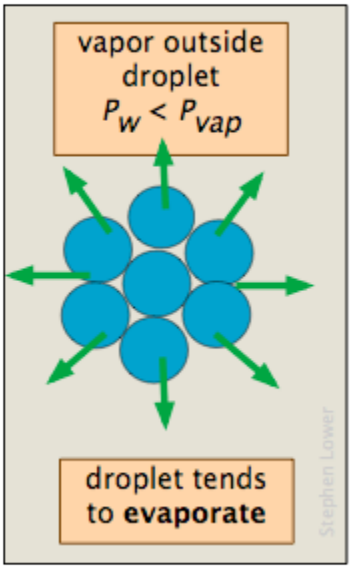


Figure $7.5.7a$

The vapor pressure of a liquid is determined by the attractive forces that act over a 180° solid angle at the surface of a liquid. In a very small drop, the liquid surface is curved in such a way that each molecule experiences fewer nearest-neighbor attractions than is the case for the bulk liquid. The outermost molecules of the liquid are bound to the droplet less tightly, and the drop has a larger vapor pressure than does the bulk liquid. If the vapor pressure of the drop is greater than the partial pressure of vapor in the gas phase, the drop will evaporate. Thus it is highly unlikely that a droplet will ever form within a vapor as it is cooled.

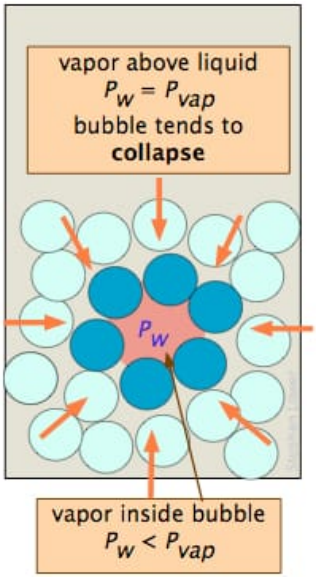


Figure 7.5.7*b*

A bubble, like a drop, must start small and grow larger, but there is a diffculty here that is similar to the one with bubbles. A bubble is a hole in a liquid; molecules at the liquid boundary are curved inward, so that they experience nearest-neighbor attractions over a solid angle greater than 180°. As a consequence, the vapor pressure of the liquid facing into a bubble P_w is always less than that of the bulk liquid P_w at the same temperature. When the bulk liquid is at its boiling point (that is, when its vapor pressure is 1 atm), the pressure of the vapor within the bubble will be *less* than 1 atm, so the bubble will tend to collapse. Also, since the bubble is formed within the liquid, the hydrostatic pressure of the overlaying liquid will add to this effect. For both of these reasons, a liquid will not boil until the temperature is raised slightly above the boiling point, a phenomenon known as ***superheating***. Once the boiling begins, it will continue to do so at the liquid's proper boiling point.

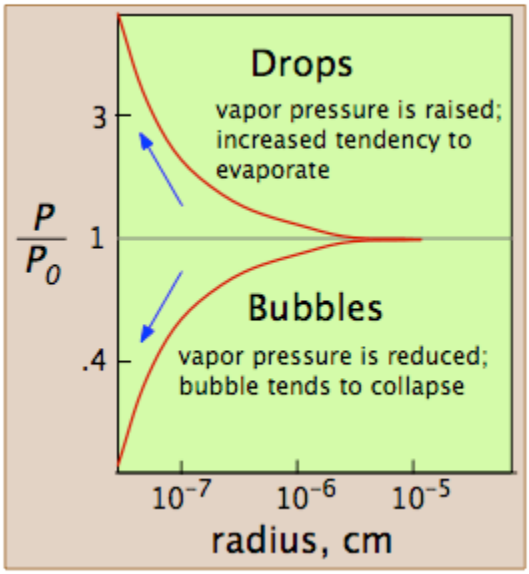


Figure 7.5.8:

These plots show how, in the case of water, the vapor pressure of a very small bubble or drop varies with its radius of curvature; the quantity being plotted is the ratio of the actual vapor pressure P to P_o, the vapor pressure of a flat surface.

Condensation of Liquids

If the tiniest of drops are destined to self-destruct, why do vapors ever condense (e.g., why does it rain)?

- If you cool a vapor in a container, condensation takes place not within the vapor itself, but on the inner surface of the container. What happens here is that intermolecular attractions between the solid surface will cause vapor molecules to adsorb to the surface and stabilize the incipient drop until it grows to a size at which it can be self-sustaining. This is the origin of the condensation on the outside of a cool drink, or of the dew that appears on the grass.
- In the case of the cloud of steam you see over the boiling water, the first few droplets form on tiny dust particles in the air — the ones you can see by scattered light when a sunbeam shines through a darkened room.

Clouds and precipitation In the region of the atmosphere where rain forms there are large numbers of solid particles, mostly of microscopic size. Some of these are particles of salt produced by evaporation of spray from the ocean surface. Many condensation nuclei are of biological origin; these include bacteria, spores, and particles of ammonium sulfate. There is volcanic and meteor dust, and of course there is dust and smoke due to the activities of humans. These particles tend to adsorb water vapor, and some may even dissolve to form a droplet of concentrated solution. In either case, the vapor pressure of the water is reduced below its equilibrium value, thus stabilizing the aggregate until it can grow to self-sustaining size and become fog, rain, or snow.

This, by the way, is why fog is an irritant to the nose and throat; each fog droplet carries within it a particle of dust or (in air polluted by the burning of sulfur-containing fossil fuels) a droplet of sulfuric acid, which it effciently deposits on your sensitive mucous membranes. If you own a car which is left outside on a foggy night, you may have noticed how dirty the windshield is in the morning.

Superheating and boiling of liquids

What is the difference between the evaporation and boiling of a liquid? When a liquid evaporates at a temperature below its boiling point, the molecules that enter the vapor phase do so directly from the surface. When a liquid boils, bubbles of vapor form in the interior of the liquid, and are propelled to the surface by their lower density (buoyancy). As they rise, the diminishing hydrostatic pressure causes the bubbles to expand, reducing their density (and increasing their buoyancy) even more.

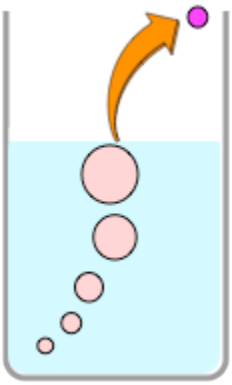


Figure 7.5.9:

But as we explained above, getting that first bubble to form and survive is often sufficiently difficult that liquids commonly superheat before they begin to boil. If you have had experience in an organic chemistry laboratory, you probably know this as “bumping”, and have been taught to take precautions against it. In large quantities, superheated liquids can be very dangerous, because the introduction of an impurity (such as release of an air bubble from the container surface) or even a mechanical disturbance can trigger nucleation and cause boiling to occur suddenly and almost explosively (Video 7.5.1).

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Video 7.5.1: Superheating water on a stovetop

Exploding water in the microwave

Many people have been seriously burned after attempting to boil water in a microwave oven, or after having added powdered material such as instant coffee to such water. When water is heated on a stove, the bottom of the container superheats only the thin layer of water immediately in contact with it, producing localized "microexplosions" that you can hear just before regular smooth boiling begins; these bubbles quickly disperse and serve as nucleation centers for regular boiling. In a microwave oven, however, the energy is absorbed by the water itself, so that the entire bulk of the water can become superheated. If this happens, the slightest disturbance can produce an explosive flash into vapor.

Sublimation

Some solids have such high vapor pressures that heating leads to a substantial amount of direct vaporization even before the melting point is reached. This is the case for solid iodine, for example. I_2 melts at 115°C and boils at 183°C, is easily sublimed at temperatures around 100°C. Even ice has a measurable vapor pressure near its freezing point, as evidenced by the tendency of snow to evaporate in cold dry weather. There are other solids whose vapor pressure overtakes that of the liquid before melting can occur. Such substances sublime without melting; a common example is solid carbon dioxide (“Dry Ice”) at 1 atm (see the CO_2phase diagram below).

Phase Diagrams

The temperatures and pressures at which a given phase of a substance is stable (that is, from which the molecules have the lowest escaping tendency) is an important property of any substance. Because both the temperature and pressure are factors, it is customary to plot the regions of stability of the various phases in *P* - *T* coordinates, as in this generic *phase diagram* for a hypothetical substance.

Because pressures and temperatures can vary over very wide ranges, it is common practice to draw phase diagrams with non-linear or distorted coordinates. This enables us to express a lot of information in a compact way and to visualize changes that could not be represented on a linearly-scaled plot. It is important that you be able to interpret a phase diagram, or alternatively, construct a rough one when given the appropriate data. Take special note of the following points:

1. The three colored regions on the diagram are the ranges of pressure and temperature at which the corresponding phase is the only stable one.
2. The three lines that bound these regions define all values of (*P,T*) at which *two* phases can coexist (i.e., be in equilibrium). Notice that one of these lines is the *vapor pressure curve* of the *liquid* as described above. The "*sublimation curve*" is just a vapor pressure curve of the *solid*. The slope of the line depends on the difference in density of the two phases.
3. In order to depict the important features of a phase diagram over the very wide range of pressures and temperatures they encompass, the axes are not usually drawn to scale, and are usually highly distorted. This is the reason that the "melting curve" looks like a straight line in most of these diagrams.
4. Where the three named curves intersect, all three phases can coexist. This condition can only occur at a unique value of (*P,T*) known as the *triple point*. Since all three phases are in equilibrium at the triple point, their vapor pressures will be identical at this temperature.
5. The line that separates the liquid and vapor regions ends at the *critical point*. At temperatures and pressures greater than the critical temperature and pressure, no separate liquid phase exists. We refer to this state simply as a *fluid*, although the term *supercritical liquid* is also commonly used.

The best way of making sure you understand a phase diagram is to imagine that you are starting at a certain temperature and pressure, and then change just one of these parameters, keeping the other constant. You will be traversing a horizontal or vertical path on the phase diagram, and there will be a change in state every time your path crosses a line. Of special importance is the horizontal path (shown by the blue line on the diagram above) corresponding to a pressure of 1 atmosphere; this line defines the *normal melting and boiling temperature*s of a substance.

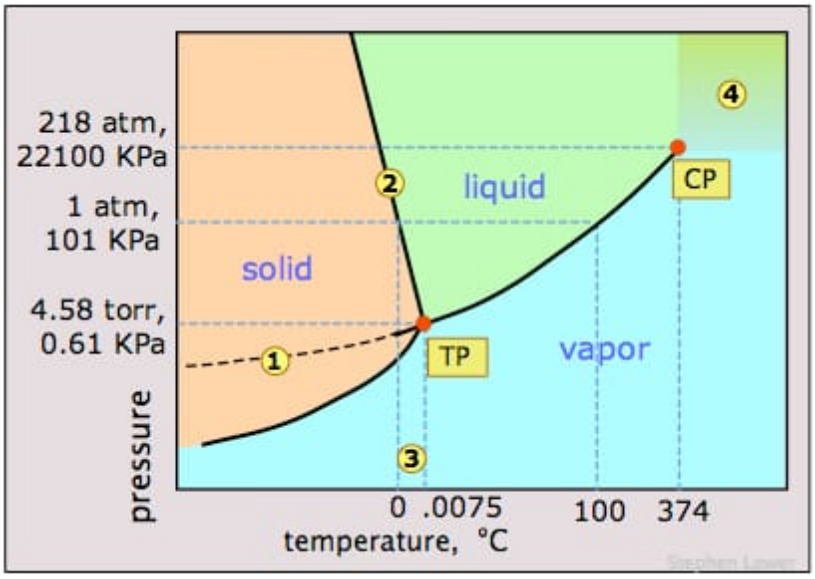


Figure 7.5.11: Phase diagram of H_2O

Notice the following features for the phase diagram of water (Figure 7.5.11):

- The slope of the line **2** separating the solid and liquid regions is negative; this reflects the unusual property that the density of the liquid is greater than that of the solid, and it means that the melting point of ice decreases as the pressure increases. Thus if ice at 0°C is subjected to a high pressure, it will find itself above its melting point and it will melt. (Contrary to what is sometimes said, however, this is not the reason that ice melts under the pressure of ice skates or skis, providing a lubricating film which makes these modes of transportation so enjoyable. The melting in these cases arises from frictional heating).
- The dashed line **1** is the extension of the liquid vapor pressure line below the freezing point. This represents the vapor pressure of *supercooled* water — a metastable state of water which can temporarily exist down to about –20°C. (If you live in a region subject to "freezing rain", you will have encountered supercooled water!)
- **3** The triple point (**TP**) of water is just 0.0075° above the freezing point; only at this temperature and pressure can all three phases of water coexist indefinitely.
- **4** Above the critical point (**CP**) temperature of 374°C, no separate liquid phase of water exists.

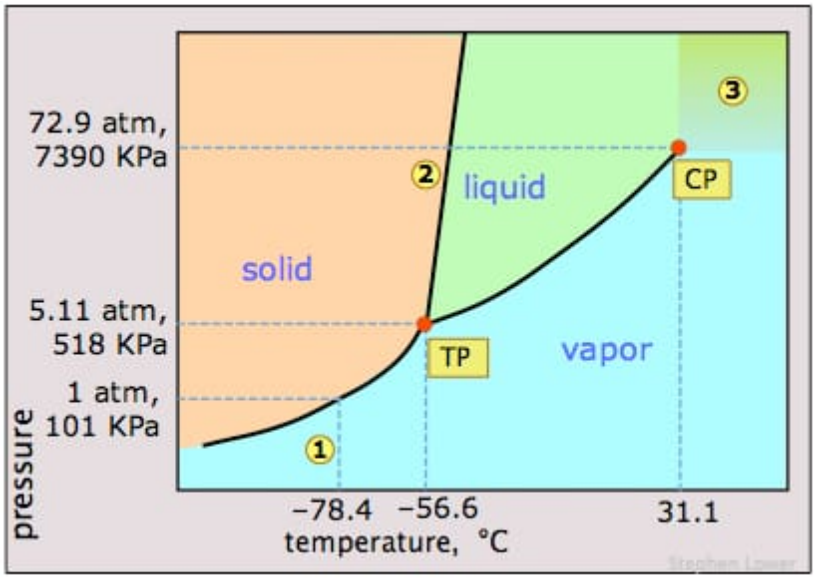


Figure 7.5.12: Phase diagram of CO_2

Dry ice, solid carbon dioxide, is widely used as a refrigerant and the phase diagram in Figure 7.5.12 shows why it is “dry”. The triple point pressure is at 5.11 atm, so below this pressure, liquid CO_2 cannot exist; the solid can only sublime directly to vapor. Gaseous carbon dioxide at a partial pressure of 1 atm is in equilibrium with the solid at 195K (–79 °C, **1**); this is the *normal sublimation*

temperature of carbon dioxide. The surface temperature of dry ice will be slightly less than this, since the partial pressure of CO_2 in contact with the solid will usually be less than 1 atm. Notice also that the critical temperature of CO_2 is only 31°C. This means that on a very warm day, the CO_2 in a fire extinguisher will be entirely vaporized; the vessel must therefore be strong enough to withstand a pressure of 73 atm.

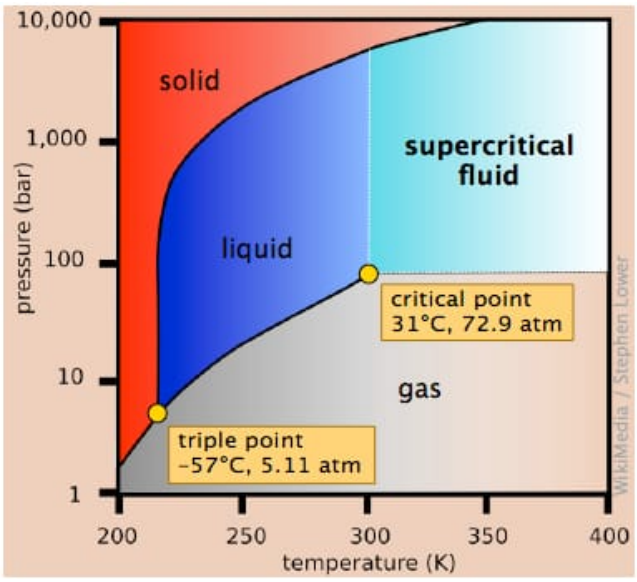


Figure 7.5.13: Phase diagram of CO_2 on greater temperature and pressure range

This view of the carbon dioxide phase diagram employs a logarithmic pressure scale and thus encompasses a much wider range of pressures, revealing the upper boundary of the fluid phase (liquid and supercritical). **Supercritical carbon dioxide** (CO_2 above its critical temperature) possesses the solvent properties of a liquid and the penetrating properties of a gas; one major use is to remove caffeine from coffee beans.

Elemental iodine, I_2, forms dark gray crystals that have an almost metallic appearance. It is often used in chemistry classes as an example of a solid that is easily sublimed; if you have seen such a demonstration or experimented with it in the lab, its phase diagram might be of interest.

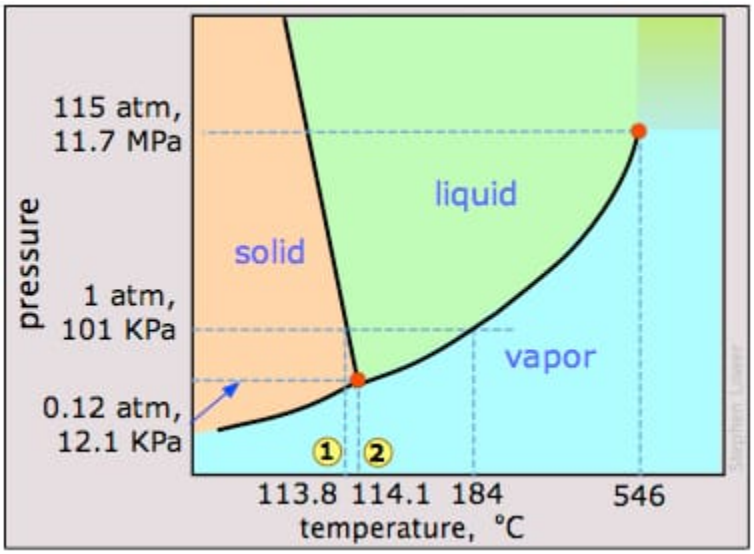


Figure 7.5.14: Phase diagram of I_2

The most notable feature of iodine's phase behavior is the small difference (less than a degree) between the temperatures of its triple point **1** and melting point **2**. Contrary to the impression many people have, there is nothing really special about iodine's tendency to sublime, which is shared by many molecular crystals including ice and naphthalene ("moth balls".) The vapor pressure

of iodine at room temperature is really quite small — only about 0.3 torr (40 Pa).The fact that solid iodine has a strong odor and is surrounded by a purple vapor in a closed container is mainly a consequence of its strong ability to absorb green light (this leaves blue and red which make purple) and the high sensitivity of our noses to its vapor.

Phase diagram of Sulfur

Sulfur exhibits a very complicated phase behavior that has puzzled chemists for over a century; what you see here is the greatly simplified phase diagram shown in most textbooks. The difficulty arises from the tendency of S_8 molecules to break up into chains (especially in the liquid above 159°C) or to rearrange into rings of various sizes (S_6 to S_{20}). Even the vapor can contain a mixture of species S_2 through S_{10}.

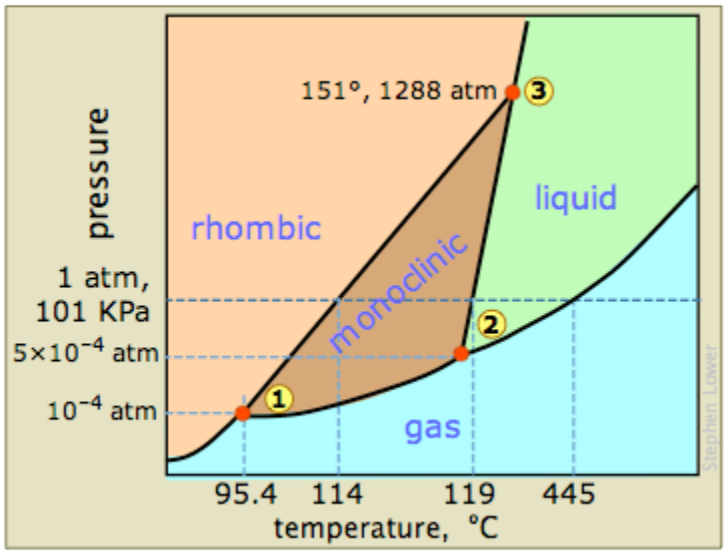


Figure 7.5.14: Phase diagram of Sulfur (S_8)

The phase diagram of sulfur contains a new feature: there are two solid phases, *rhombic* and *monoclinic*. The names refer to the crystal structures in which the S_8 molecules arrange themselves. This gives rise to **three triple points**, indicated by the numbers on the diagram.

Exercise 7.5.1

From the phase diagram in Figure 7.5.14 identify one example of three phases that *never* coexist in sulfur (S_8 (hint there are several correct answers)?

When rhombic sulfur (the stable low-temperature phase) is heated slowly, it changes to the monoclinic form at 114°C, which then melts at 119°. But if the monoclinic form is heated rapidly the molecules do not have time to rearrange themselves, so the rhombic arrangement persists as a metastable phase until it melts at 119-120°. Formation of more than one solid phase is not uncommon — in fact, if one explores into the very high pressures (see below), it seems to be the rule.

Extremes Pressure and Temperatures

We tend to think of the properties of substances as they exist under the conditions we encounter in everyday life, forgetting that most of the matter that makes up our world is situated inside the Earth, where pressures are orders of magnitude higher (Figure 7.5.15). Geochemists and planetary scientists need to know about the phase behavior of substances at high temperatures and pressures to develop useful models to test their theories about the structure and evolution of the Earth and of the solar system.

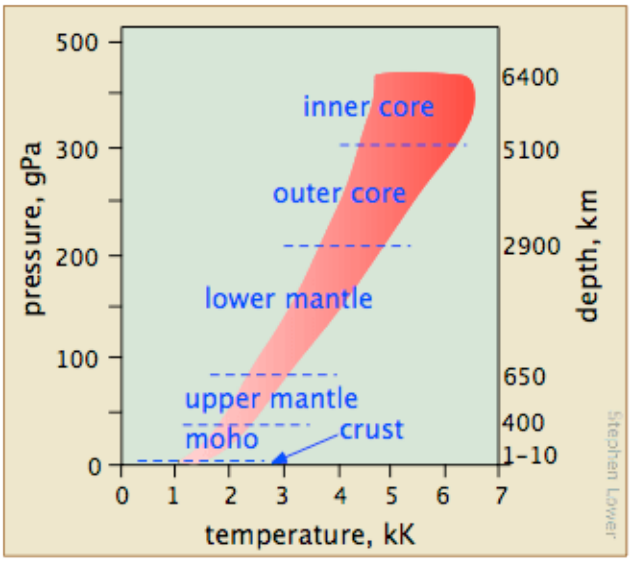


Figure 7.5.15: Pressure and temperatures in the Earth as a function of depth from the surface.

What ranges of temperatures and pressures are likely to be of interest — and more importantly, are experimentally accessible?

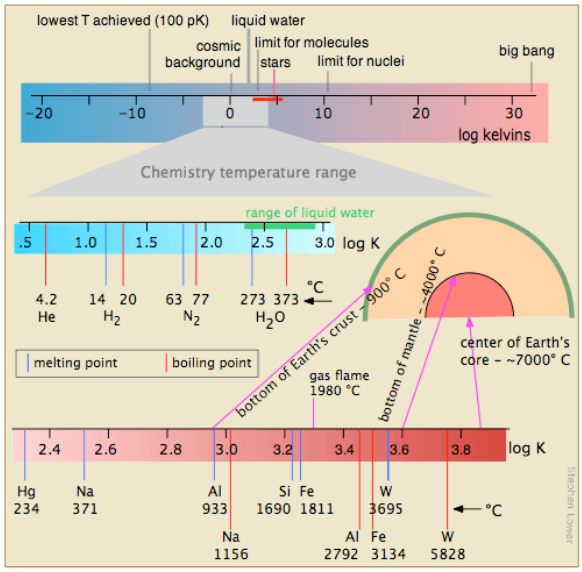


Figure 7.5.16:

Figure 7.5.16 shows several scales (all of which, please note, are logarithmic) that cover respectively the temperature range for the universe; the low temperatures of importance to chemistry (note the green line indicating the temperatures at which liquid water can exist); the higher temperatures, showing the melting and boiling points of several elements for reference. The highest temperatures that can be produced in the laboratory are achieved (but only for very short time intervals) by light pulses from laser or synchrotron radiation.

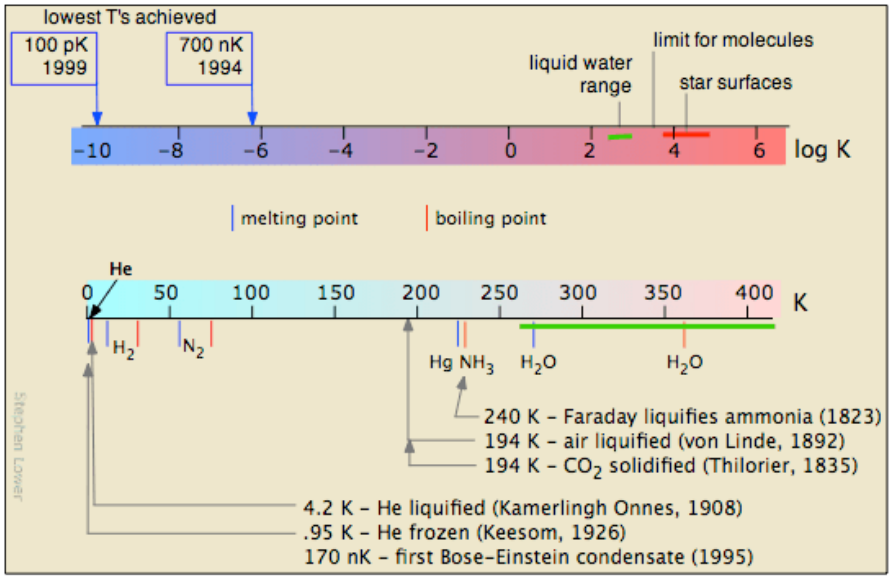


Figure 7.5.17:

The study of **low temperatures** is limited by the laws of physics that prohibit reaching absolute zero. But the fact that there is no limit to how close one can approach 0 K has encouraged a great deal of creative experimentation.

The study of matter at high pressures is not an easy task. The general techniques were pioneered between 1908-1960 by P.W. Bridgeman of Harvard University, whose work won him the 1946 Nobel Prize in physics. The more recent development of the diamond anvil cell has greatly extended the range of pressures attainable and the kinds of observations that can be made. Shock-wave techniques have made possible the production of short-lived pressures in the tPa range.

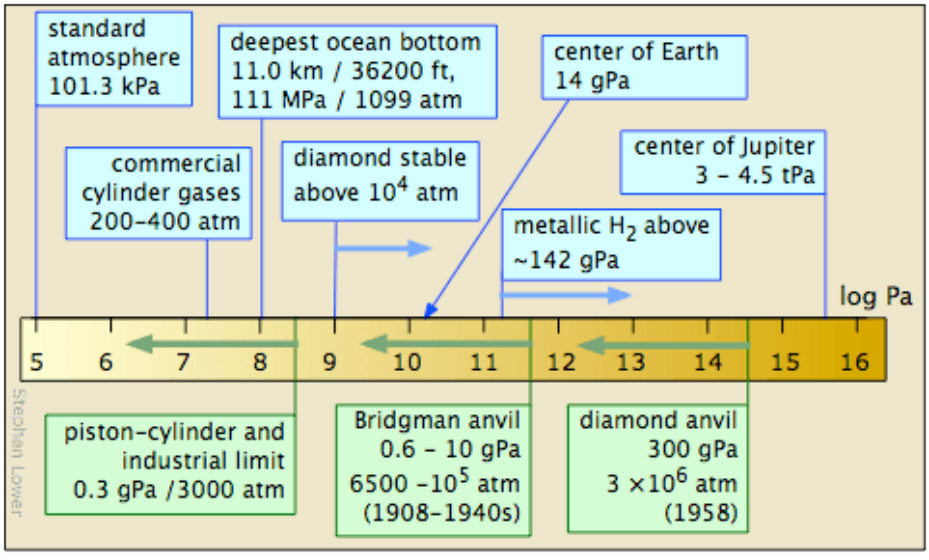


Figure 7.5.18:

High pressure laboratory studies have revealed that many molecular substances such as hydrogen and water change to solid phases having melting points well above room temperature at very high pressures; there is a solid form of ice that remains frozen even at 100°C. At still higher pressures, many of these substances become metals. It is believed that much of the inner portion of the largest planets consists of metallic hydrogen — and, in fact, that all substances can become metallic at sufficiently high pressures.

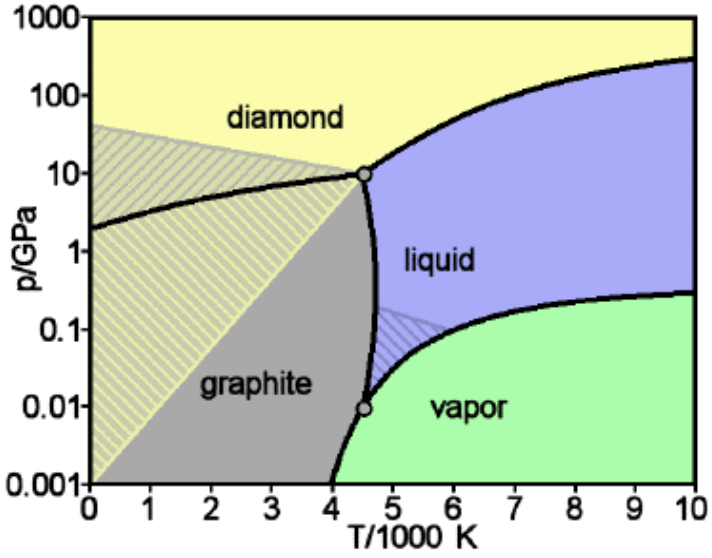


Figure 7.5.19: Phase diagram of carbon: diamond and graphite

Graphite is the stable form of solid carbon at low pressures; diamond is only stable above about 10^4 atm. But once it is in this form, the rate at which diamond converts back to graphite is immeasurably slow under ordinary environmental conditions; there is simply not enough thermal energy available to break all of those carbon-carbon bonds. So the diamonds we admire in jewelry and pay dearly for are said to be *metastable*. [image: WikiMedia]

Synthetic Diamonds

To transform graphite into diamond at a reasonable rate, a pressure of 200,000 atm and a temperature of about 4000 K would be required.Since no apparatus can survive these conditions, the process, known as is *high pressure high temperature synthesis* (HPHT) is carried out commercially at 70,000 atm and 2300 K in a solution of molten nickel, which also acts as a catalyst. Traces of Ni in the finished product serve to distinguish synthetic diamonds from natural ones. However, most synthetic diamonds are too small (only a few millimeters) and too flawed for gem quality, and are used mainly to fabricate industrial grinding and cutting tools.

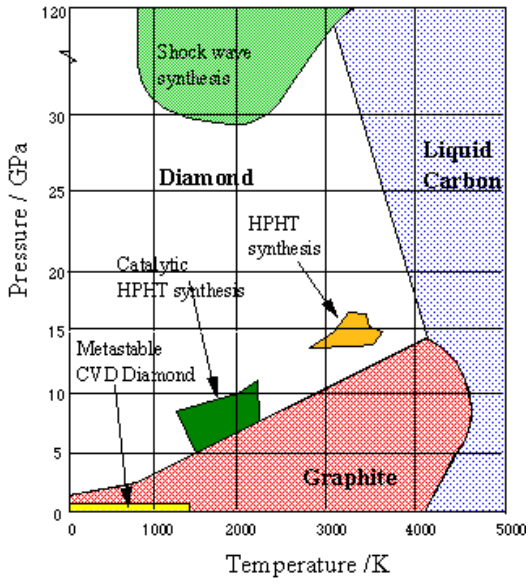


Figure 7.5.20: Phase diagram of carbon: diamond and graphite

More recently, thin diamond films have become important for engineering applications and semiconductor fabrication. These are most commonly made by condensation of gaseous carbon onto a suitable substrate (chemical vapor deposition, CVD). The conditions under which synthetic diamonds are made are depicted on the above phase diagram from Bristol University.

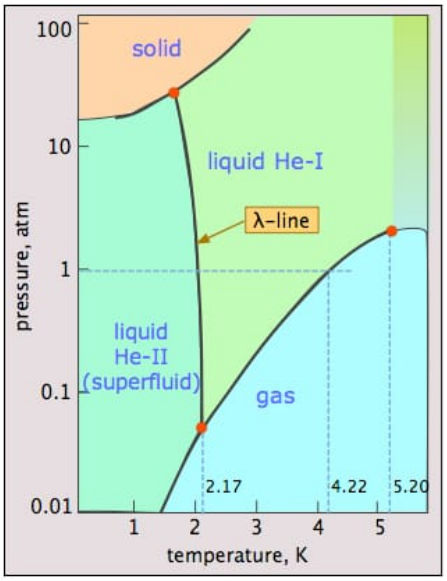
YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Video 7.5.2*: Chemist Roy Gat explains how he uses phase diagrams to synthesize synthetic diamonds at low pressure and temperatures.*

Helium Phase weirdness



Helium is unique in that quantum phenomena, which normally apply only to tiny objects such as atoms and electrons, extend to and dominate its macroscopic properties. A glance at the phase diagram of $^4\mathrm{He}$ reveals some of this quantum weirdness. The main points to notice are:

- Helium can be frozen only at high pressures;
- The solid and gas cannot coexist (be in equilibrium) under any conditions;
- There are *two liquid phases*, helium I (an ordinary liquid) and helium II (a highly unusual *ordered liquid*);
- The λ (*lambda*) line represents the (*P,T*) values at which the two phases can coexist;
- Helium-II behaves as a *superfluid* — essentially a *quantum liquid*.

Why cannot liquid helium freeze at low pressures? The low mass of the He atoms and their close confinement in the solid provides them with a very high *zero-point energy* (the Heisenberg uncertainty principle in action!) that allows them to vibrate with such amplitude that they overcome the dispersion forces that would otherwise hold the solid together, thus keeping the atoms too separated to form a solid. Only by applying a high pressure (25 atm) can this effect be overcome.

Helium-II, quantum liquids and superfluidity

We usually need quantum theory only to describe the properties of tiny objects such as electrons, but with liquid He-II, it extends

to the macroscopic scale of the bulk liquid. $^4\mathrm{He}$ atoms (99.99+ percent of natural helium) are *bosons*, which means that at low temperatures, they can all occupy the same quantum state (all other normal atoms are *fermions* and are subject to the Pauli exclusion principle).

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Video 7.5.3: Helium becomes superfluid and displays amazing properties.

All together, now

Objects that occupy the same quantum state all possess the same momentum. Thus when one atom moves, they all move together. In a sense, this means that the entire bulk of liquid He-II acts as a single entity. This property, known as *superfluidity*, gives rise to a number of remarkable effects, most notably:

- The liquid can flow through a narrow channel without friction, up to a critical velocity that depends on the ratio of flow rate to channel width;
- When placed in a container, they form a film that climbs up the walls and down the outside, seeking the same level outside the container as inside if this is possible;
- A small molecule dissolved in helium-II behaves the same as it would in a vacuum.

In liquid helium-II, only about 10% of the atoms are in such a state, but it is enough to give the liquid some of the weird properties of a quantum liquid.

He^3 exhibits similar properties, but owing to its low natural abundance, it was not extensively studied until the 1940s when large amounts became available as a byproduct of nuclear weapons manufacture. Finally, in 1996, its superfluidity was observed at a temperature of 2 nK. Although He^3 atoms are fermions, only those that pair up (and thus assume the properties of bosons) give rise to superfluidity.

Water at high pressures

Water, like most other substances, exhibits many solid forms at higher pressures (Figure 7.5.22). So far, fifteen distinct ice phases have been identified; these are designated by Roman numerals ice-I through ice-XV. Ice-I can exist in two modifications; the crystal lattice of ice-Ic is cubic, while that of ice-Ih is hexagonal. The latter corresponds to the ordinary ice that we all know. It's interesting to note that several high-pressure phases of ice can exist at temperatures in excess of 100°C.

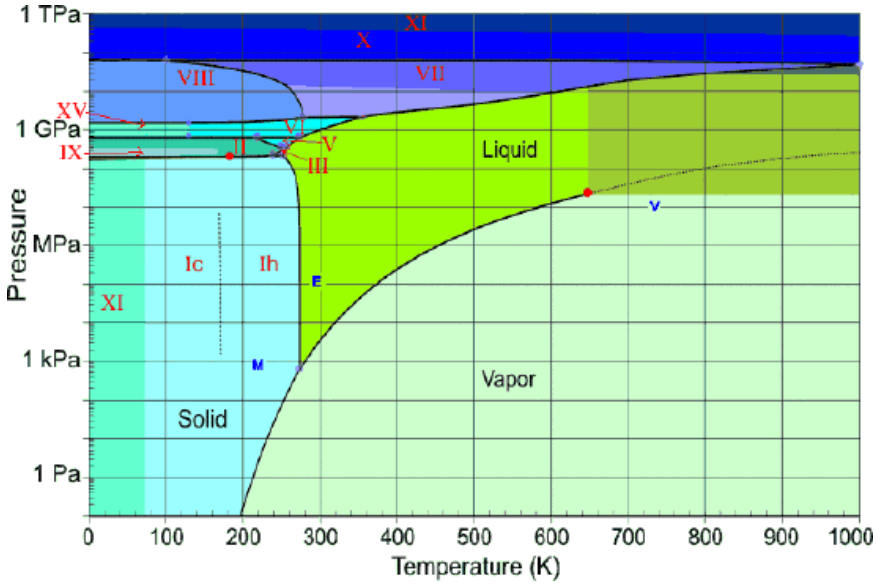


Figure 7.5.22: Higher pressure and temperature phase diagram of water. Image from Martin Chaplin's LSBU site.

This page titled 7.5: Changes of State is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.6: Introduction to Crystals

Learning Objectives

- Identify the three kinds of *rotational symmetry axes* of a cube.
- State what is meant by a crystal's *habit*, and identify some factors that might affect it.
- Explain why the *angles between adjacent faces* (of even a broken crystal) tend to have the same small set of values.
- What is a *unit cell*, and how does it relate to a crystal lattice?
- What are *Bravais lattices*, and why are they important?
- Find the *Miller index* of a line or plane in a unit cell, or sketch the line or plane having a given Miller index.

The delicately faceted surfaces of large crystals that occur in nature have always been a source of fascination and delight. In some ways they seem to represent a degree of perfection that is not apparent in other forms of matter. But in the realm of pure solid substances, crystals are the rule rather than the exception, although this may not be apparent unless they are observed under a hand-lens or a microscope. It is remarkable that the visual examination of crystals was able to establish a fairly mature science of crystallography (applied mainly to the study of minerals) by the end of the 19th Century, even before the atomic theory of matter had been universally accepted. Today this aspect of crystallography is of importance not only to chemists and physicists, but also to geologists, amateur minerologists and "rock-hounds" who maintain some of the best Web resources on crystals. In this lesson we will see how the external shape of a crystal can reveal much about the underlying arrangement of its constituent atoms, ions, or molecules.

The first thing we notice about a crystal is the presence of planes — called *faces* — which constitute the external boundaries of the solid. Of course, any solid, including non-crystalline glass, can be carved, molded or machined to display planar faces; examples of these can be found in any "dollar store" display of costume jewelry. What distinguishes and defines a true crystal is that these faces develop *spontaneously and naturally* as the solid forms from a melt or from solution. The multiple faces invariably display certain geometrical relationships to one another, resulting in a symmetry that attracts our attention and delights the eye.



Figure 7.6.1: Natural crystalline sulfur. (CC-BY-SA; http://images-of-elements.com/sulfur.php)

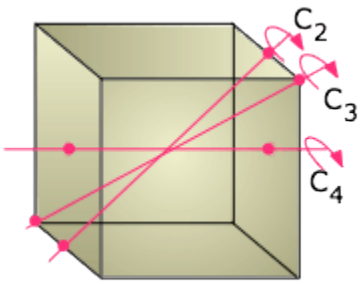
Symmetry elements

One of the most apparent elements of this geometrical regularity are the sets of parallel faces that many crystals display. Nowhere is this more apparent than in the cubes that develop when sodium chloride crystallizes from solution. We usually think of a cubic shape in terms of the equality of its edge lengths and the 90° angles between its sides, but there is a more fundamental way of

classifying shapes that chemists find very useful. This is to look at what *geometric transformations* (such as rotations around an axis) we can perform that leave the appearance unchanged.

Cubic symmetry

For example, you can rotate a cube 90° around an axis perpendicular to any pair of its six faces without making any apparent change to it. We say that the cube possesses three mutually perpendicular *four-fold rotational axes*, abbreviated C_4 axes. But if you think about it, a cube can also be rotated around an axis that extends between opposite corners; in this case, it takes three 120° rotations to go through a complete circle, so these axes (also four in number) are three-fold or C_3 axes. And finally, there are two-fold (C_2) axes that pass diagonally through the centers of the six pairs of opposite edges.



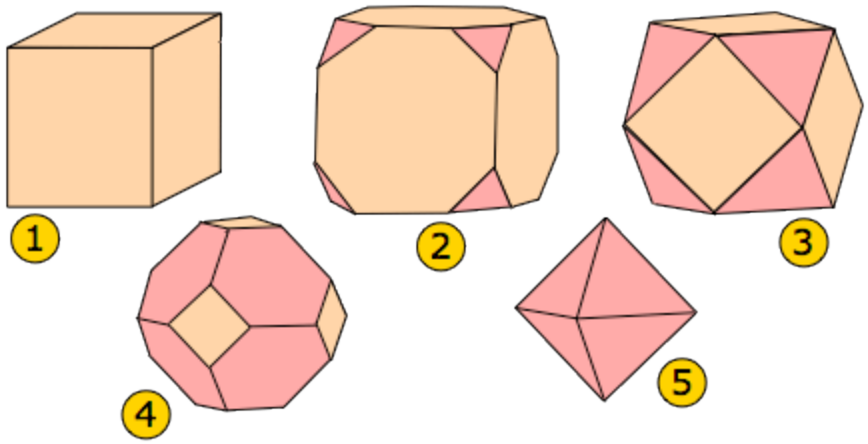
In addition, there are imaginary symmetry planes that mirror the portions of the cube that lie on either side of them. Three of these are parallel to the three major axes of the crystal, and an additional six pass diagonally through opposite edges.



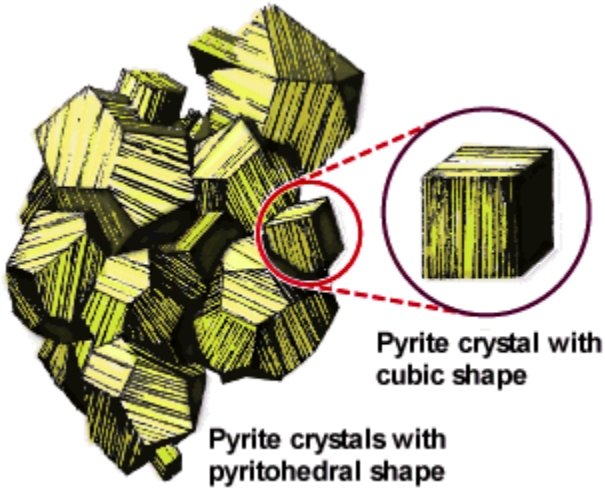
All told, there are 13 rotational axes and 9 mirror planes (only a few of which are shown above) that define cubic symmetry. Why is this important? Although anyone can recognize a cube when they see one, it turns out that many crystals, both natural and synthetic, are for one reason or another unable to develop all of their faces equally. Thus a crystal that forms on the bottom of a container will be unable to grow any faces that project downward for the simple reason that there is no supply of ions or molecules from that direction. The same effect occurs when a mineral crystal tries to grow in contact with other solids. Finally, the presence of certain impurities that selectively adsorb to one or more faces can block the addition of more material to them, thus either completely inhibiting their formation or forcing them to grow at slower rates. These alternative shapes that can develop from a single basic crystal type are known as *habits*.

Crystal habits

Sodium chloride grown from pure aqueous solution forms simple cubes, but the addition of various impurities can result in habits that can be regarded as cubes that have been truncated along planes normal to some of the symmetry axes. (The same effects can sometimes be seen as a crystal slowly dissolves and material is released more rapidly from some directions than others.)



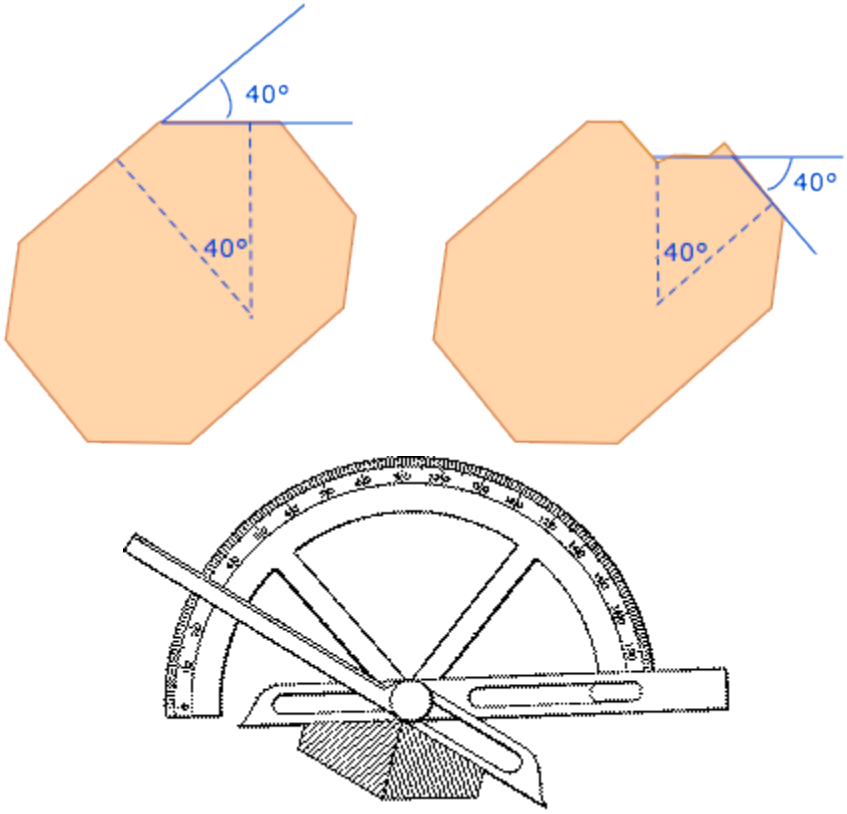
In this example, the perfect cube **1** develops triangular faces at the corners **2**. If these enlarge beyond their maximum size **3**, the triangular faces meet in a new set of edges that are hexagonal **4**. Eventually we are left with the eight faces of what is obviously a regular octahedron **5**. One might think that these five shapes bear no relationship to one another, but in fact they all possess the same set of set of symmetry elements as the simple cube and are thus various *habit*s of the same underlying cubic structure and belong to the *cubic crystal system* described further below.



This sketch of a natural formation of pyrite crystals (FeS_2) shows various habits of this cubic crystal growing together.

The Law of Constant Angles

Even though a given crystal may be distorted or broken, the angles between corresponding faces remain the same. Thus you can crush a crystal underfoot or break it up with a hammer, but you will always find that the fragments possess a limited set of interfacial angles.



This fundamental law, discovered by Nicholas Steno in 1669, was a major key development in crystallography. About 100 years later, a protractor-like device (the *contact goniometer*) was invented to enable more accurate measurements than the rather crude ones that had formerly been traced out on paper.

Cleavage Planes

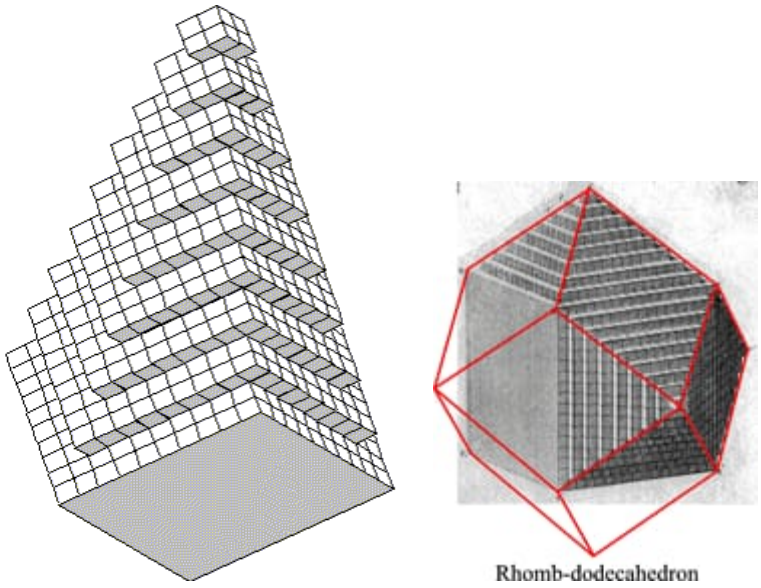
When a crystal is broken by applying a force in certain directions (as opposed to being pulverized by a hammer) it will often be seen to break cleanly into two pieces along what are known as cleavage planes. The new faces thus formed always correspond to the symmetry planes associated with a particular crystal type, and of course make constant angles with any other faces that may be present.



(left) Halite (natural NaCl) crystals easily cleave along planes parallel to the cubic faces. (right) This sample of gypsum (CaSO4) is easily cleaved with a razor blade.

Scientific crystallography began with an accident

Cleavage planes were first described in the late 17th century, but nothing much was thought about their significance until about a hundred years later when the Abbe Hauy accidently dropped a friend's sample of calcite and noticed how cleanly it broke. Further experimentation showed that other calcite crystals, even ones of different initial shapes (habits), displayed similar rhombohedral shapes upon cleavage, and that these in turn produced similar shapes when they were cleaved. This led Haüy to suggest that continued cleavages would ultimately lead to the smallest possible unit which would be the fundamental building block of the crystal. (Remember that the atomic theory of matter had not developed at this time.)



Haüy's elaborately drawn figures (published in 1784) showed how external faces of a crystal could be produced by stacking the units in various ways. For example, by omitting rows from a cubic stack of primal cubelets, one could arrive at the various stages between the cube and the octahedra for sodium chloride that we saw earlier on this page.

The modern interpretation of these observations replaces Haüy's primal shapes with atoms or molecules, or more generally with points in space that these define the possible locations of atoms or molecules. It is easy to see how plane faces can develop along some directions and not others if one assumes that the new faces must follow a linear sequence of points.

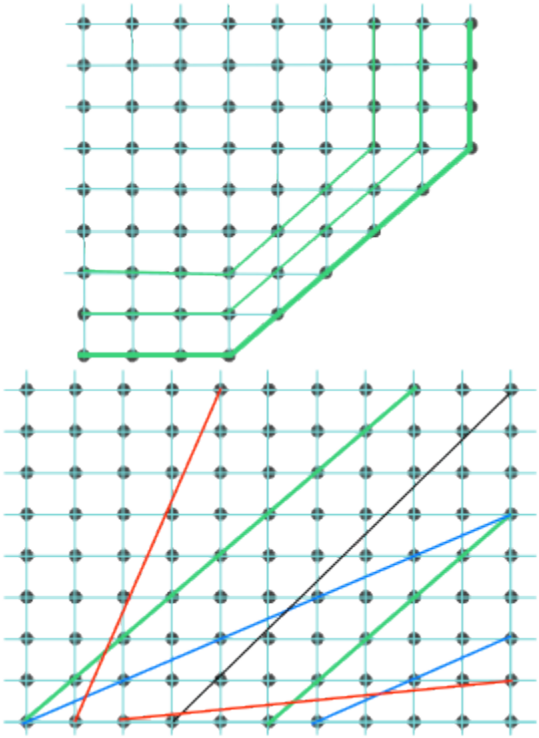
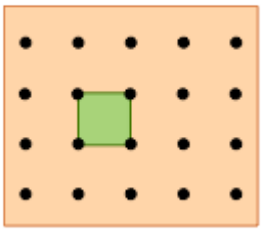
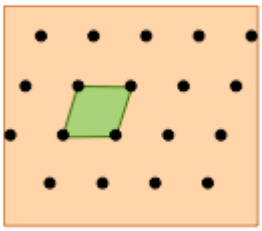
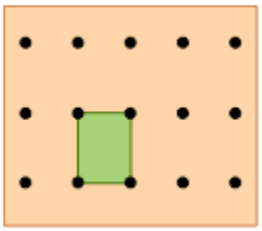
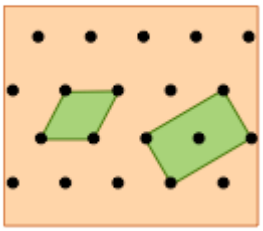
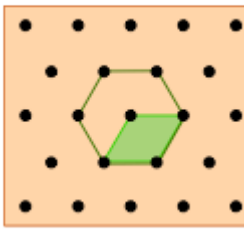
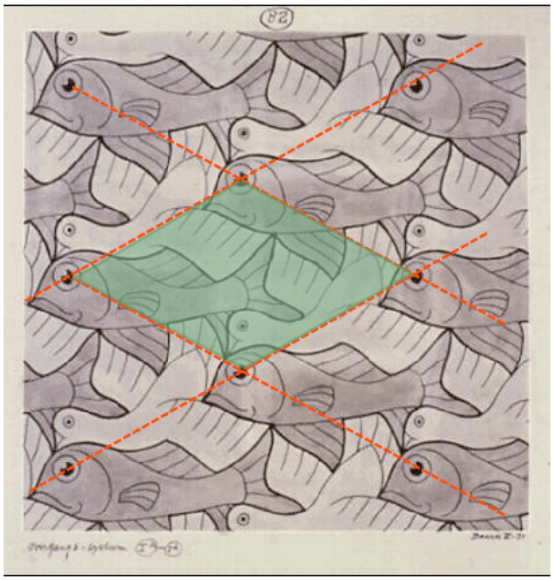


Figure 7.6.2: (left) Formation of plane faces exhibiting constant angles/ (right) Extended plane faces more readily form along the green and blue lines.



				
Square Lattice: $x = y$, 90° angles	**Parallelogram** lattice $x \neq y$, angles < 90°	**Rectangular** lattice 0°	**Rhombic** or **centered-rectangle** lattice: $x = y$, angles neither 60° or 90°;	**Hexagonal** lattice (but unit cell is a rhombus with $x = y$ and angles 60°)

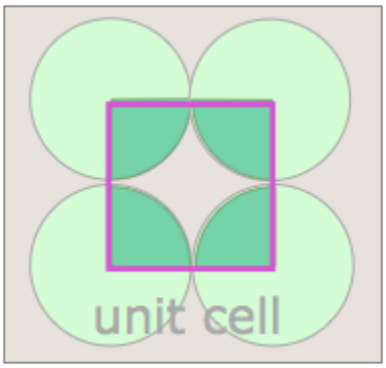
Although everyone has seen and admired the huge variety of patterns on printed fabrics or wallpapers, few are aware that these are all based on one of five types of two-dimensional "unit cells" that form the basis for these infinitely-extendable patterns. One of the most remarkable uses of this principle is in the work of the Dutch artist Maurits Escher (1888-1972).



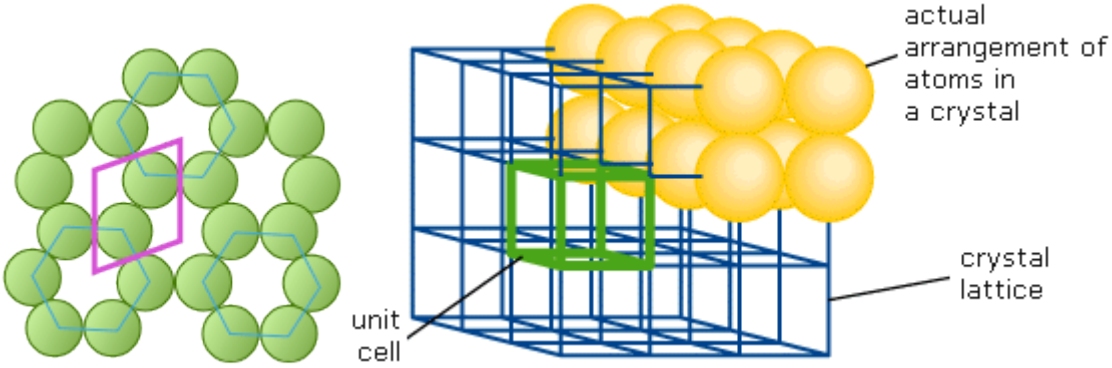
Shown below are two-dimensional views of the unit cells for two very common types of crystal lattices, one having cubic symmetry and the other being hexagonal. Although we could use a hexagon for the second of these lattices, the *rhombus* is preferred because it is simpler.



Notice that in both of these lattices, the corners of the unit cells are centered on a lattice point. This means that an atom or molecule located on this point in a real crystal lattice is shared with its neighboring cells. As is shown more clearly here for a two-dimensional square-packed lattice, a single unit cell can claim "ownership" of only one-quarter of each molecule, and thus "contains" 4 × ¼ = 1 molecule.



The unit cell of the graphite form of carbon is also a rhombus, in keeping with the hexagonal symmetry of this arrangement.Notice that to generate this structure from the unit cell, we need to shift the cell in both the *x*- and *y*- directions in order to leave empty spaces at the correct spots. We could alternatively use regular hexagons as the unit cells, but the *x*+*y* shifts would still be required, so the simpler rhombus is usually preferred.



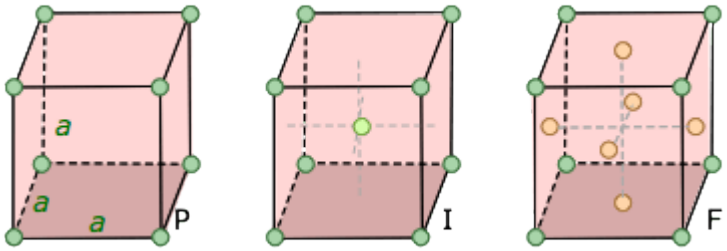
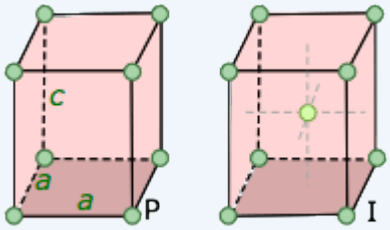
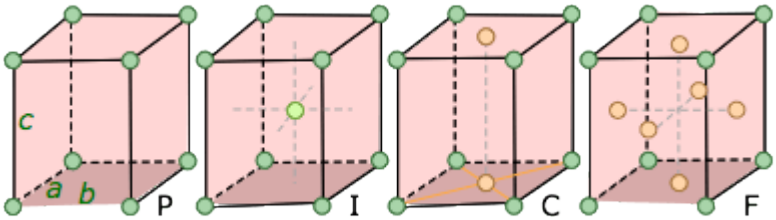
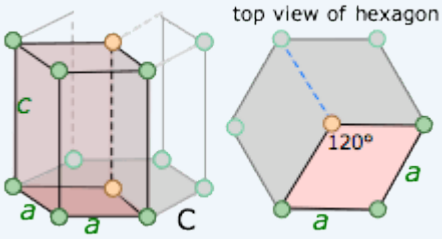
This image nicely illustrates the relations between the unit cell, the lattice structure, and the actual packing of atoms in a typical crystal.

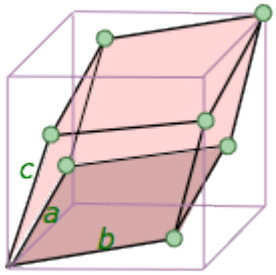
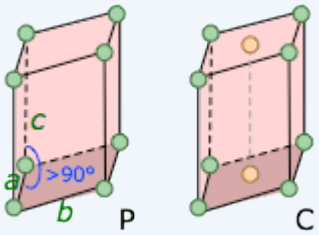
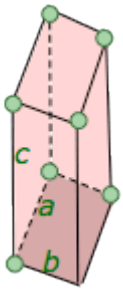
Crystal systems and Bravais lattices

We saw above that five basic cell shapes can reproduce any design motif in two dimensions. If we go to the three-dimensional world of crystals, there are just seven possible basic lattice types, known as *crystal systems*, that can produce an infinite lattice by successive translations in three-dimensional space so that each lattice point has an identical environment. Each system is defined by the relations between the axis lengths and angles of its unit cell. For example, if the three edge lengths are identical and all corner angles are 90°, a crystal belongs to the cubic system.

The simplest possible cube is defined by the eight lattice points at its corner, but variants are also possible in which additional

lattice points exist in the faces ("face-centered cubic") or in the center of the cube ("body-centered cubic"). If variants of this kind are taken into account, the total number of possible lattices is fourteen; these are known as the fourteen *Bravais lattices*.

Crystal System	Bravais lattices (P = primitive, I= body-centered, F = face-centered)
cubic $a = b = c$ $\alpha = \beta = \gamma = 90°$	 The F cell corresponds to *closest cubical packing*, a very common and important structure.
tetragonal $a = b \neq c$ $\alpha = \beta = \gamma = 90°$	 A cube that has been extended in one direction, creating a unique *c*-axis. An F cell would simply be a network of joined I cells.
orthorhombic $a \neq b \neq c$ $\alpha = \beta = \gamma = 90°$	 Three unequal axes at right angles. The "C" form has atoms in the two faces that cut the *c*-axis.
hexagonal $a = b \neq c$ $\alpha = \beta = 60°$, $\gamma = 120°$	 Just as in the 2-dimensional examples given above, the unit cell of the hexagonal lattice has a rhombic cross-section; the entire hexagonal unit is built from three of these rhombic prisms.

Crystal System	Bravais lattices (P = primitive, I= body-centered, F = face-centered)
trigonal(rhombohedral) $a = b = c$ $\alpha = \beta = \gamma \neq 90°$,	 Think of this as a cube that has been skewed or distorted to one side so that opposite faces remain parallel to each other. This can also be regarded as a special case of the hexagonal system, and is often classified as such by U.S. minerologists who recognize only six crystal systems. The rhombohedral form of the hexagonal system is difficult to visualize.
monoclinic $a \neq b \neq c$ $\alpha = \gamma = 90°$, $\beta > 90°$	 Two 90° angles, one > 90°, with all sides of different lengths. A *C* cell (also seen in the orthorhombic class) has additional points in the center of each end. Monoclinic I and F cells can be constructed from C cells.
triclinic $a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90°$	 This is the most generalized of the crystal systems, with all lengths and angles unequal, and no right angles.

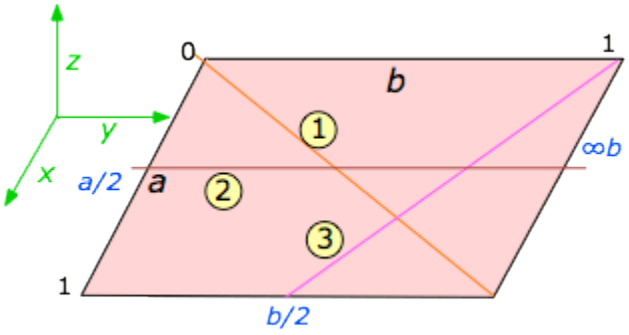
Notes on the above diagrams:

- The labels *a,b,c* along the unit cell axes represent the *dimensions* of the unit cell. Visual examination of a crystal does not allow us to determine their actual values, but merely to know whether any two (or all three) are the same.
- When $a = b$, both axes may be given "*a*" labels, since neither is unique.
- The angles α, β and γ are those between the *b*-*c*, *a*-*c*, and *a*-*b* axes, respectively. Similarly in the cube, all axes are *a* axes.

Lines and planes in unit cells: the Miller index

In any kind of repeating pattern, it is useful to have a convenient way of specifying the orientation of elements relative to the unit cell. This is done by assigning to each such element a set of integer numbers known as its *Miller index*.

Indexing lines in two-dimensions



To understand indexing, it will be easier to begin with a unit cell plane that we are viewing from above, along the [invisible] *z*-axis. The drawing shows such a plane with three lines crossing it at various slopes. The index of each line is found by first determining the points where it intersects the *x* and *y* axes as fractions of the unit cell parameters *a* and *b*. Thus in the above example:

- Line 1 starts at the origin and extends to the lower right-hand corner which corresponds to intersections of the *x* axis at one unit cell length of both *a* and *b* — that is, at 1×*a* and 1×*b*. We can abbreviate this set of intersections as [1,1].
- Line 2 intersects the *x* axis at one-half the unit cell distance *a*, or at *a*/2. This line is parallel to the *y* axis, so it never intersects it; this is mathematically the same as saying that it intersects it at infinity, or in terms of unit cell increments, at ∞×*b*. We can describe this line in terms of the unit cell intercepts by the pair of values [½,∞].
- Line 3 starts at the upper right corner of the cell which corresponds to the coordinates (1,1), but this is equivalent to the origin (0,0) of the neighboring unit cell on the right. So in terms of our coordinate system (which repeats for each unit cell), this line extends in the negative-*y* direction and intersects this axis at –*b*/2. The x-intercept is *a*. These intercepts correspond to [1,–2].

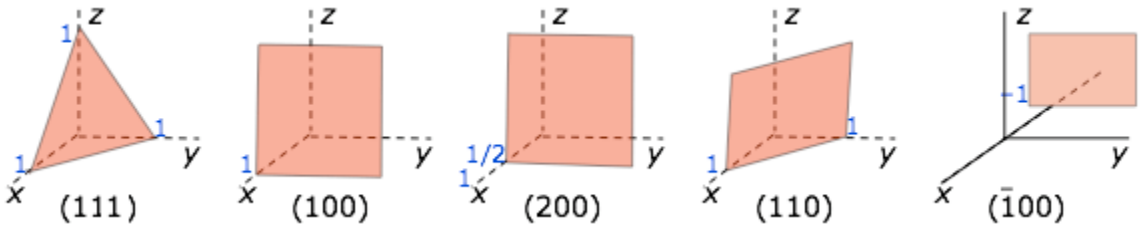
The *Miller indices* of the lines are given by the *reciprocals* of these values:

Line 1 [1,1] → **(11)**	Line 2 [½,∞] → **(20)**	Line 3 [1,–2] → **1,2)**

Miller indices are written in parentheses with no spaces between numbers. Negative values are indicated by an overbar as in 1.

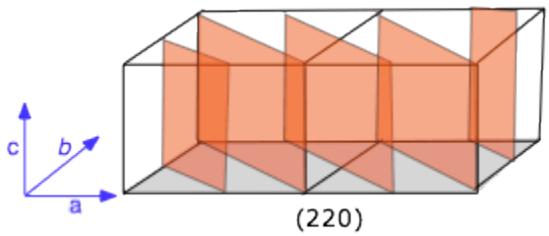
Indexing planes in three-dimensions

We proceed in exactly the same way, except that we now have 3-digit Miller indices corresponding to the axes *a, b* and *c*.



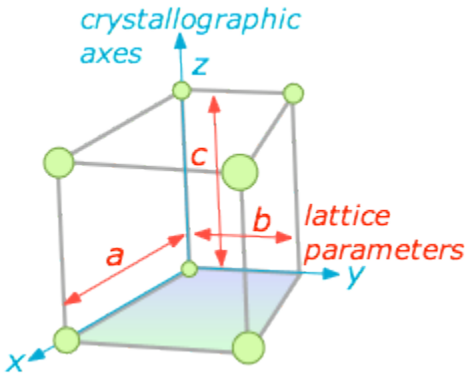
It is important to note that multiple parallel planes that repeat at the same interval have identical Miller indices. This simply

reflects the fact that we can repeat the coordinate axes at any regular interval.

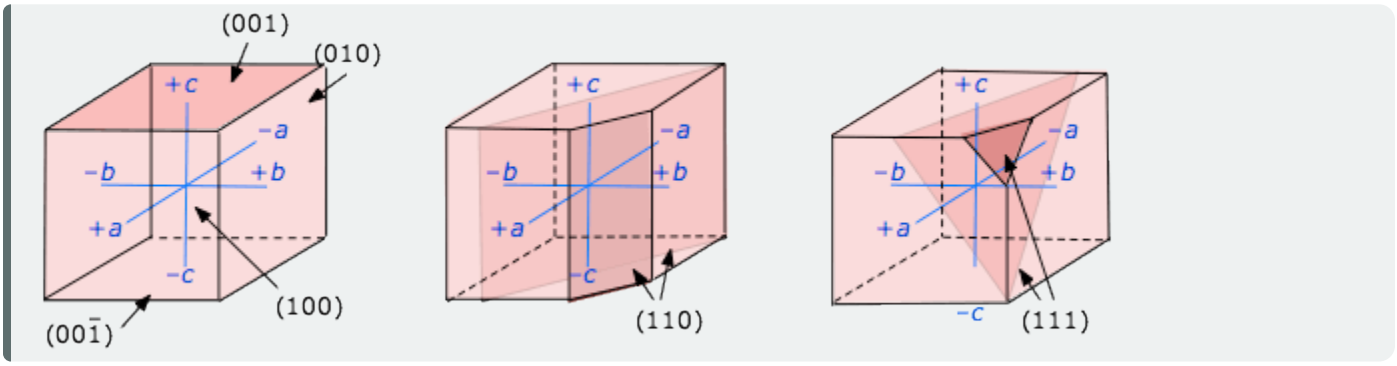


Identifying crystal faces

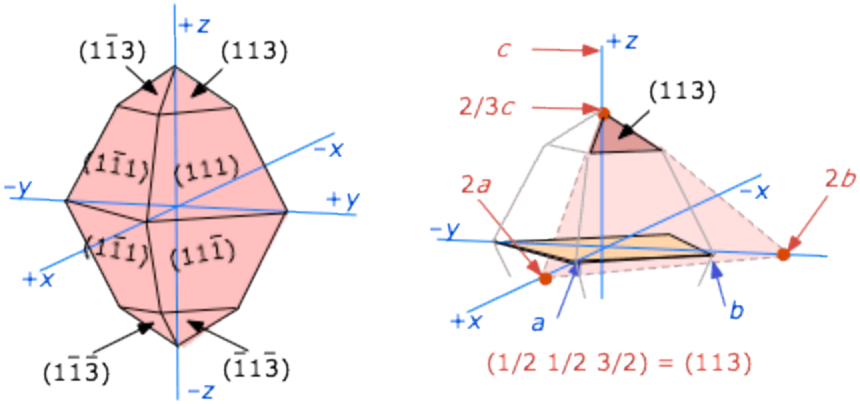
We mentioned previously that the plane faces of crystals are their most important visually-distinctive property, so it is important to have a convenient way of referring to any given face. First, we define a set of reference directions (*x,y,z*) which are known as the *crystallographic axes*. In most cases these axes correspond to directions that are fairly apparent on visual examination of one or more crystals of a given kind. They are parallel to actual or possible edges of the crystal, and they are not necessarily orthogonal. We now know, as Haüy first suggested, that these directions correspond to rows of lattice points in the underlying structure of the crystal.



We also define three *lattice parameters* (*a,b,c*) which mark out the boundaries of the unit cell along the crystallographic axes. The index of a particular face is determined by the fractional values of (a,b,c) at which the face intersects the axes (*x,y,z*). Study the examples shown below for three different habits of a cubic lattice.



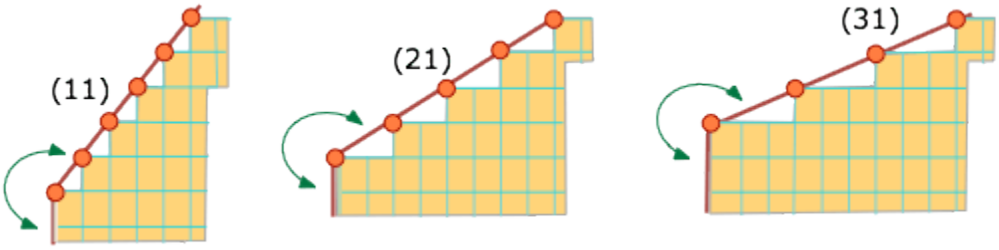
Below is a more complicated example of one particular habit of an orthorhombic crystal. The figure at the right shows how the (113) face is indexed.



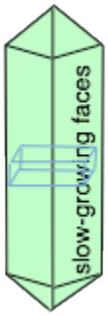
In this case, the plane at the top of the crystal is extended downward to the (*x*,*y*) plane. This extended plane cuts the (*x*,*y*,*z*) axes at (2*a*, 2*b*, 2/3*c*). The corresponding inverses would be (½,½,3/2). In order to make them into proper Miller indices (which should always be integers) we multiply everything by 2, yielding (113).

Why do Miller indices use mostly small numbers?

It is remarkable that the faces that bound real crystals generally have small Miller indices. The low values for the indices suggest that a given lattice plane has a high density of lattice points per unit area, a logical consequence of each molecule being surrounded and held by its closely-packed neighbors. In the 2-dimensional projection below, compare the facial lattice-point density in the (11) plane with that of the (31) plane.



Crystals with a single long unit-cell axis tend to form planes with the long axis normal to the plane, so that the major faces of the crystal are planes containing the short-axis translations. Similarly, crystals with a single, short unit-cell axis tend to be needles. The main faces, on the sides of the needles, contain the short lattice translation — a high density of lattice points. In general, it is found that crystals have linear dimensions that mirror the reciprocals of the lattice parameters.

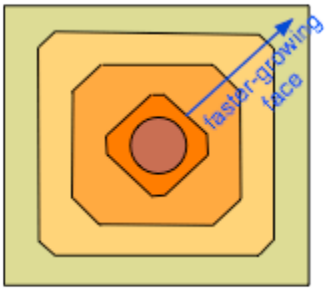


Factors affecting crystal growth habits

Faces having a lower density of lattice points (as in the (31) face shown above) can acquire new layers more rapidly, and thus grow more rapidly than faces having a high lattice-point density. The faces that can potentially develop in a crystal are determined entirely by the symmetry properties of the underlying lattice. But the faces that actually develop under specific conditions — and thus the overall *shape* of the crystal — is determined by the relative *rates* of growth of the various faces. The slower the growth

rate, the larger the face.

This relation can be understood by noting that faces that grow normal to shorter unit cell axes (as in the needle-shaped crystal shown above) present a larger density of lattice points to the surface (that is, more points per unit surface area.) This means that more time is required for diffusion of enough new particles to build out a new layer on such a surface.



An interesting experiment is to grind a large crystal of salt into a spherical shape and immerse it in a saturated solution of sodium chloride. At first, the most disturbed and exposed parts on the surface dissolve, revealing a large variety of underlying plane faces. As growth resumes, the smaller of these are rapidly replaced by larger faces. Eventually, the fast-growing faces eliminate themselves and the high-lattice point density faces that correspond to the sides of the cube win out.

In addition to these structural effects, the conditions under which a crystal is grown can affect its habit. Temperature, degree of supersaturation, nature of the solvent all have their effects, and these may affect the growth of different faces in different ways. The presence of impurities in the solution can radically alter the habit of a crystal, as seen in the following table for the growth of sodium chloride:

Table: Impurity Effect on the Growth of Sodium Chloride Crystals

no impurity	$Fe(CN)_6^{4-}$	formamide	Pb^{2+}, Cd^{2+}	polyvinyl alcohol
cubes	dendrites	large crystals	needles	

These effects presumably come about because these substances preferentially adsorb to certain faces, impeding their growth.

This page titled 7.6: Introduction to Crystals is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.7: Ionic and Ion-Derived Solids

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above.

- What is an ionic solid, what are its typical physical properties, and what kinds of elements does it contain?
- Define the lattice energy of an ionic solid in terms of the energetic properties of its component elements.
- Make a rough sketch that describes the structure of solid sodium chloride.
- Describe the role that the relative ionic radii play in contributing to the stability of an ionic solids.
- Give examples of some solids that can form when ionic solutions are evaporated, but which do not fall into the category of "ionic" solids.

In this section we deal mainly with a very small but important class of solids that are commonly regarded as composed of ions. We will see how the relative sizes of the ions determine the energetics of such compounds. And finally, we will point out that not all solids that are formally derived from ions can really be considered "ionic" at all.

Ionic Solids

The idealized ionic solid consists of two interpenetrating lattices of oppositely-charged point charges that are held in place by a balance of coulombic forces. But because real ions occupy space, no such "perfect" ionic solid exists in nature. Nevertheless, this model serves as a useful starting point for understanding the structure and properties of a small group of compounds between elements having large differences in electronegativity.

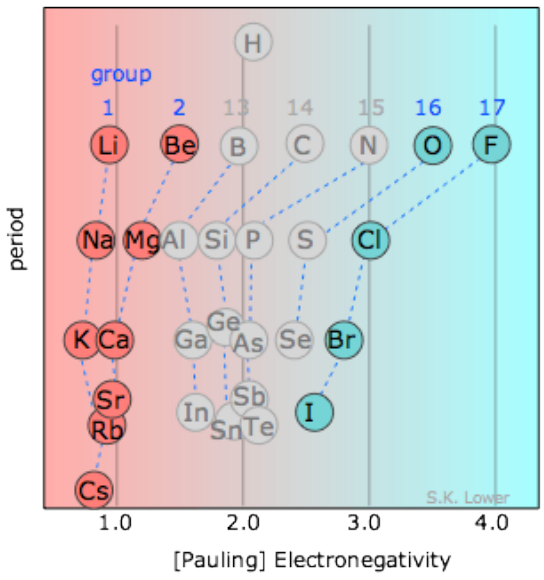


Figure 7.7.1

Chemists usually apply the term "ionic solid" to binary compounds of the metallic elements of Groups 1-2 with one of the halogen elements or oxygen. As can be seen from the diagram, the differences in electronegativity between the elements of Groups 1-2 and those of Group 17 (as well as oxygen in Group 16) are sufficiently great that the binding in these solids is usually dominated by Coulombic forces and the crystals can be regarded as built up by aggregation of oppositely-charged ions.

Sodium Chloride (rock-salt) Structure

The most well known ionic solid is sodium chloride, also known by its geological names as *rock-salt* or *halite*. We can look at this

compound in both structural and energetic terms.



Rock Salt

Structurally, each ion in sodium chloride is surrounded and held in tension by six neighboring ions of opposite charge. The resulting crystal lattice is of a type known as *simple cubic*, meaning that the lattice points are equally spaced in all three dimensions and all cell angles are 90°.

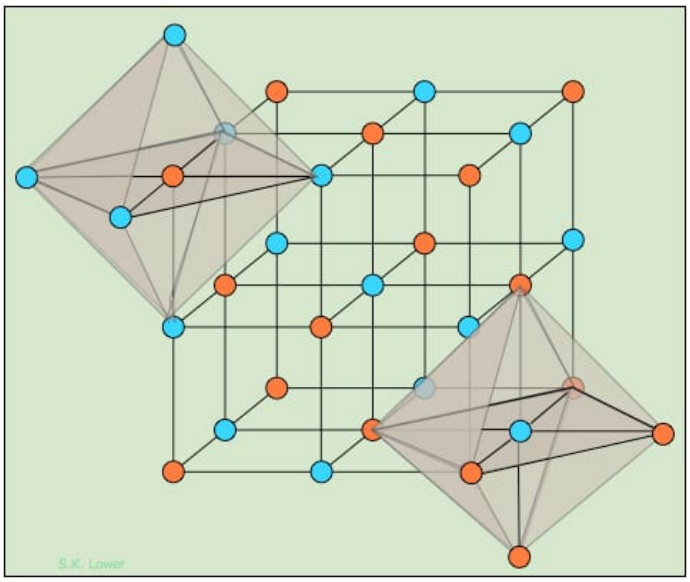


Figure 7.7.2: The differently-colored circles represent the Na^+ and Cl^- ions; because their locations are geometrically equivalent, it does not matter which color is assigned to which ion.

In Figure 7.7.2, we have drawn two imaginary octahedra centered on ions of different kinds and extending partially into regions outside of the diagram. (We could equally well have drawn them at any of the lattice points, but show only two in order to reduce clutter.) Our object in doing this is to show that each ion is surrounded by six other ions of opposite charge; this is known as *(6,6) coordination*. Another way of stating this is that each ion resides in an *octahedral hole* within the cubic lattice.

How can one sodium ion surrounded by six chloride ions (or *vice versa*) be consistent with the simplest formula NaCl? The answer is that each of those six chloride ions also sits at the center of its own octahedron defined by another six sodium ions. You might think that this corresponds to Na_6Cl_6, but note that the central sodium ion shown in the diagram can claim only a one-sixth share of each of its chloride ion neighbors, so the formula NaCl is not just the simplest formula, but correctly reflects the 1:1 stoichiometry of the compound. But of course, as in all ionic structures, there are no distinguishable "molecular" units that

correspond to the NaCl simplest formula. Bear in mind that large amount of empty space in diagrams depicting a crystal lattice structure can be misleading, and that the ions are really in direct contact with each other to the extent that this is geometrically possible.

Sodium Chloride Energetics

Sodium chloride, like virtually all salts, is a more energetically favored configuration of sodium and chlorine than are these elements themselves; in other words, the reaction

$$Na_{(s)} + ½Cl_{2(g)} \rightarrow NaCl_{(s)} \tag{7.7.1}$$

is accompanied by a release of energy in the form of heat. How much heat, and why? To help us understand, we can imagine the formation of one mole of sodium chloride from its elements proceeding in these hypothetical steps in which we show the energies explicitly:

Step 1: *Atomization* of sodium (breaking one mole of metallic sodium into isolated sodium atoms)

$$\mathrm{Na\,(s) + 108\,kJ \longrightarrow Na\,(g)} \tag{7.7.2}$$

Step 2: Same thing with chlorine. This requires more energy because it involves breaking a covalent bond.

$$\mathrm{½Cl_2\,(g) + 127\ \ kJ \longrightarrow Cl\,(g)} \tag{7.7.3}$$

Step 3: We strip an electron from one mole of sodium atoms (this costs a lot of energy!)

$$\mathrm{Na\,(g) + 496\ \ kJ \longrightarrow Na^+\,(g) + e^-} \tag{7.7.4}$$

Step 4: Feeding these electrons to the chlorine atoms gives most of this energy back.

$$\mathrm{Cl\,(g) + e^- \longrightarrow Cl^-\,(g) + 348\ \ kJ} \tag{7.7.5}$$

Step 5: Finally, we bring one mole of the ions together to make the crystal lattice — with a huge release of energy.

$$\mathrm{Na^+\,(g) + Cl^-\,(g) \longrightarrow NaCl\,(s) + 787\ \ kJ} \tag{7.7.6}$$

If we add all of these equations together, we get

$$\mathrm{Na\,(s) + \tfrac{1}{2}\,Cl_2\,(g) \longrightarrow NaCl\,(s) + 404\ kJ} \tag{7.7.7}$$

In other words, the formation of solid sodium chloride from its elements is highly *exothermic*. As this energy is released in the form of heat, it spreads out into the environment and will remain unavailable to push the reaction in reverse. We express this by saying that "sodium chloride is more *stable* than its elements".

Looking at the equations above, you can see that Equation 7.7.6 constitutes the big payoff in energy. The 787 kj/mol noted there is known as the NaCl *lattice energy*. Its large magnitude should be no surprise, given the strength of the coulombic force between ions of opposite charge.

It turns out that it is the lattice energy that renders the gift of stability to all ionic solids.Note that this lattice energy, while due principally to coulombic attraction between each ion and its eight nearest neighbors, is really the sum of *all* the interactions with the crystal. Lattice energies cannot be measured directly, but they can be estimated fairly well from the energies of the other processes described in the table immediately above.

How Geometry and Periodic Properties Interact

The most energetically stable arrangement of solids made up of identical molecular units (as in the noble gas elements and pure metals) are generally those in which there is a minimum of empty space; these are known as *close-packed* structures, and there are several kinds. In the case of ionic solids of even the simplest 1:1 stoichiometry, the positive and negative ions usually differ so much in size that packing is often much less efficient. This may cause the solid to assume lattice geometries that differ from the one illustrated above for sodium chloride.

By way of illustration, consider the structure of cesium chloride (the spelling *cæsium* is also used), CsCl. The radius of the Cs^+ ion is 168 pm compared to 98 pm for Na^+ and cannot possibly fit into the octahedral hole of a simple cubic lattice of chloride ions. The CsCl lattice therefore assumes a different arrangement.

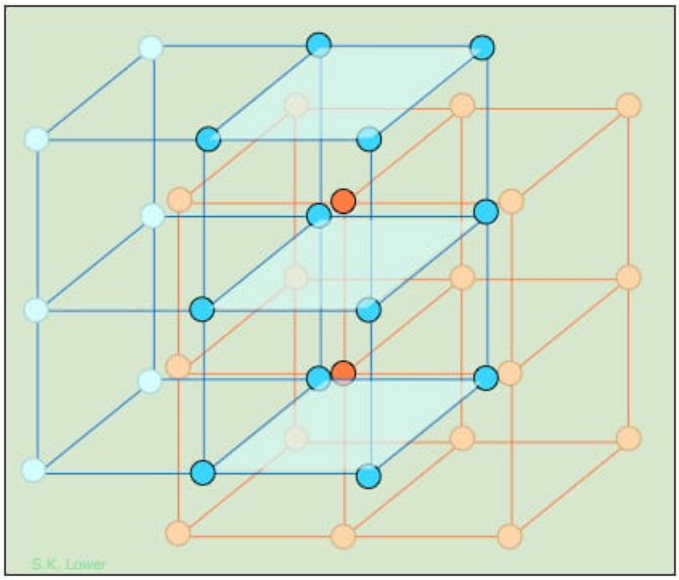


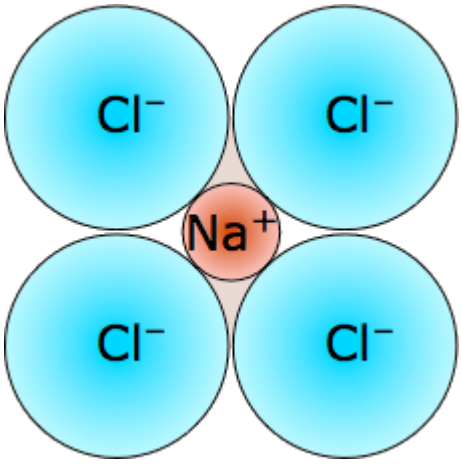
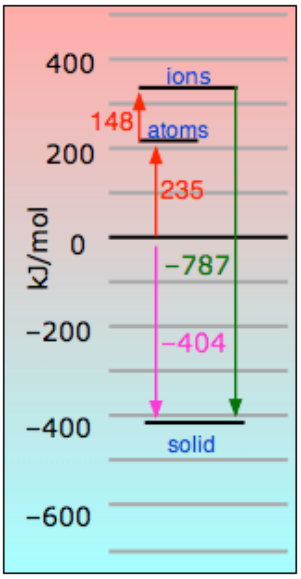
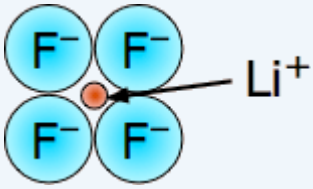
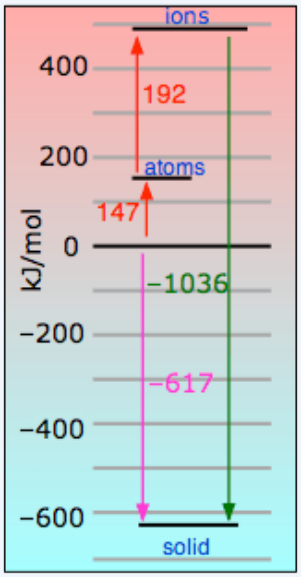
Figure 7.7.3: The CsCl structure, like that of NaCl, can be regarded as two interpenetrating lattices of the oppositely-charged ions. In CsCl, however, the metal ions, instead of being surrounded by six chloride ions, are shifted into the center of each cubic element of the Cl−-ion lattice.

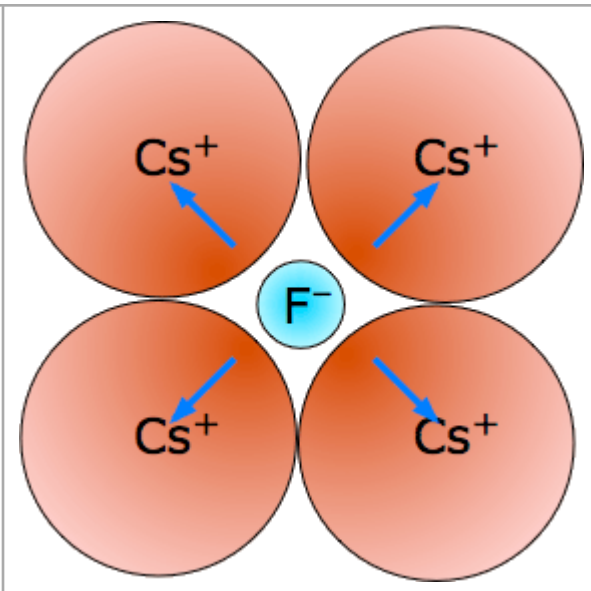
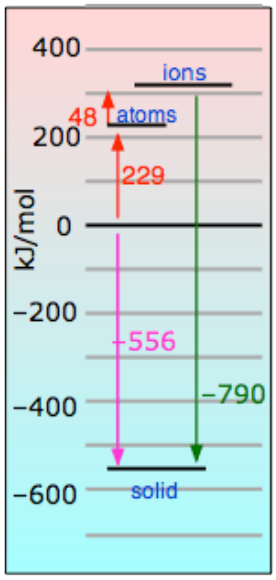
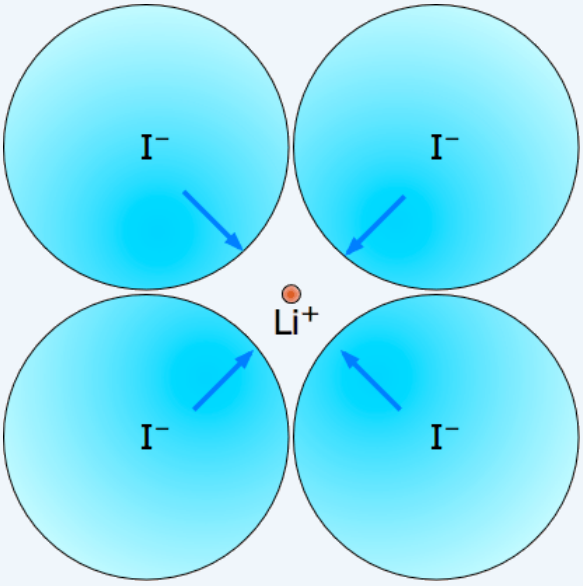
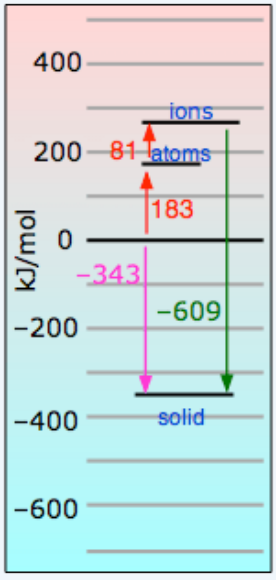
Figure 7.7.3 focuses on two of these cubic lattice elements whose tops and bottoms are shaded for clarity. It should be easy to see that each cesium ion now has eight nearest-neighbor chloride ions. Each chloride ion is also surrounded by eight cesium ions, so all the lattice points are still geometrically equivalent. We therefore describe this structure as having *(8,8) coordination*.

The two kinds of lattice arrangements exemplified by NaCl ("rock salt") and CsCl are found in a large number of other 1:1 ionic solids, and these names are used generically to describe the structures of these other compounds. There are of course many other fundamental lattice arrangements (not all of them cubic), but the two we have described here are sufficient to illustrate the point that the *radius ratio* (the ratio of the radii of the positive to the negative ion) plays an important role in the structures of simple ionic solids.

The Alkali Halides

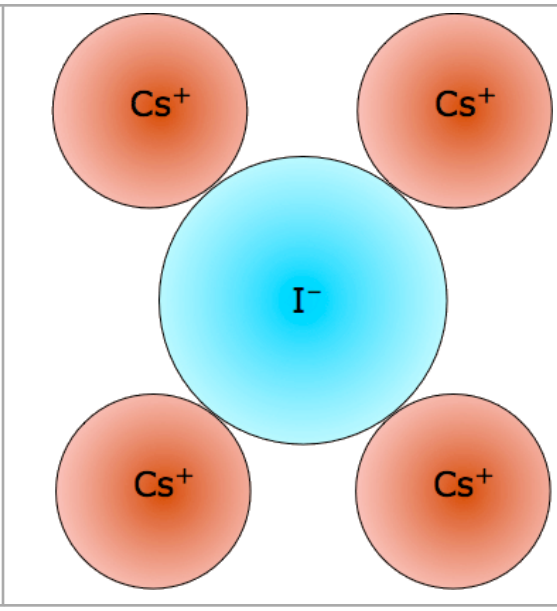
The interaction of the many atomic properties that influence ionic binding are nicely illustrated by looking at a series of alkali halides, especially those involving extreme differences in atomic radii. The latter are all drawn to the same scale. On the energetic plots at the right, the lattice energies are shown in green. We will start with the one you already know very well.

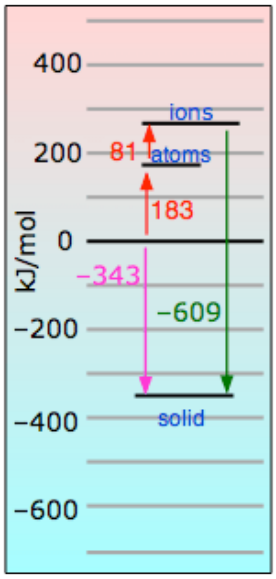
Sodium chloride - NaCl ("rock-salt") mp/bp 801/1413 °C; coordination (6,6)		
Lithium Fluoride - LiF - mp/bp 846/1676 °C, rock-salt lattice structure (6,6). Tiny-tiny makes strong-strong! This is the most "ionic" of the alkali halides, with the largest lattice energy and highest melting and boiling points. The small size of these ions (and consequent high charge densities) together with the large electronegativity difference between the two elements places a lot of electronic charge between the atoms. Even in this highly ionic solid, the electron that is "lost" by the lithium atom turns out to be closer to the Li nucleus than when it resides in the 2s shell of the neutral atom.		

Cesium Fluoride, CsF - mp/bp 703/ 1231 °C, (8,8) coordination. With five shells of electrons shielding its nucleus, the Cs^+ ion with its low charge density resembles a big puff-ball which can be distorted by the highly polarizing fluoride ion. The resulting ion-induced dipoles (blue arrows) account for much of the lattice energy here. The reverse of this would be a tiny metal ion trying to hold onto four relatively huge iodide ions like Lithium iodide.		
Lithium iodide, LiI - mp/bp 745/1410 °C. Negative ions can make even bigger puff-balls. The tiny lithium ion can't get very close to any of the iodides to generate a very strong coulombic binding, but does polarize them to create an ion-induced dipole component. It does not help that the negative ions are in contact with each other. The structural geometry is the same (6,6) coordination as NaCl.		

Cesium iodide, CsI - mp/bp 626/1280 °C.

Even with the (8,8) coordination afforded by the CsCl structure, this is a pretty sorry combination owing to the low charge densities. The weakness of coulombic- compared to van der Waals interactions makes this the least-"ionic" of all the alkali halide solids.

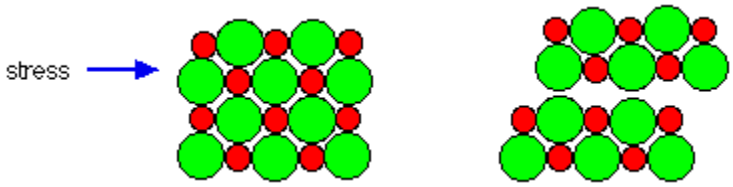




Conclusion: Many of the alkali halide solids are not all that "ionic" in the sense that coulombic forces are the predominant actors; in many, such as the CsI illustrated above, ion-induced dipole forces are more important.

Some Properties of Ionic Solids

As noted above, ionic solids are generally hard and brittle. Both of these properties reflect the strength of the coulombic force. *Hardness* measures resistance to *deformation*. Because the ions are tightly bound to their oppositely-charged neighbors and, a mechanical force exerted on one part of the solid is resisted by the electrostatic forces operating over an extended volume of the crystal.



But by applying sufficient force, one layer of ions can be made to slip over another; this is the origin of *brittleness*. This slippage quickly propagates along a plane of the crystal (more readily in some directions than in others), weakening their attraction and leading to physical *cleavage*. Because the "ions" in ionic solids lack *mobility*, the solids themselves are **electrical insulators**.

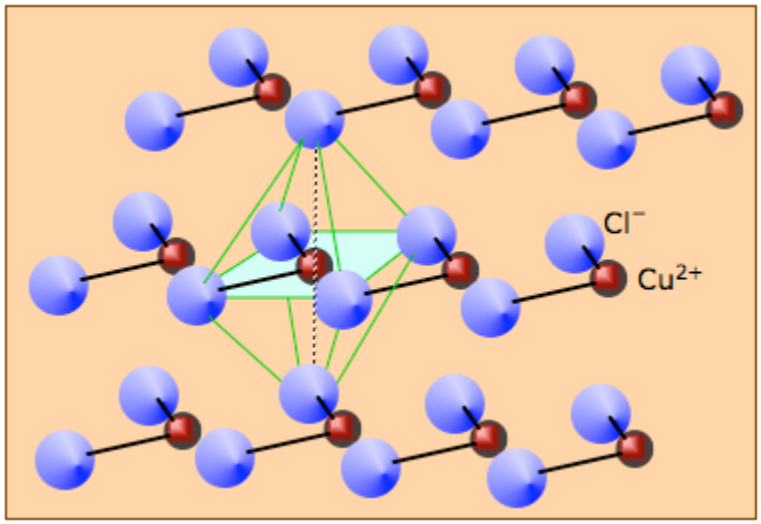
Not all ion-derived solids are "ionic".

Even within the alkali halides, the role of coulombic attraction diminishes as the ions become larger and more polarizable or differ greatly in radii. This is especially true of the anions, which tend to be larger and whose electron clouds are more easily distorted. In solids composed of polyatomic ions such as $(NH_4)_2SO_4$, $SrClO_4$, NH_4CO_3, ion-dipole and ion-induced dipole forces may actually be stronger than the coulombic force. Higher ionic charges help, especially if the ions are relatively small. This is especially evident in the extremely high melting points of the Group 2 and higher oxides:

MgO (magnesia)	CaO (lime)	SrO (strontia)	Al2O3 (alumina)	ZrO2 (zirconia)
2830 °C	2610 °C	2430 °C	2050 °C	2715 °C

These substances are known as *refractories*, meaning that they retain their essential properties at high temperatures. Magnesia, for example, is used to insulate electric heating elements and, in the form of fire bricks, to line high-temperature furnaces. No boiling points have been observed for these compounds; on further heating, they simply dissociate into their elements. Their

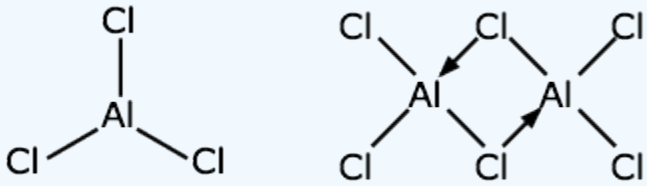
crystal structures can be very complex, and some (notably Al_2O_3) can have several solid forms. Even in the most highly ionic solids there is some electron sharing, so the idea of a “pure” ionic bond is an abstraction.



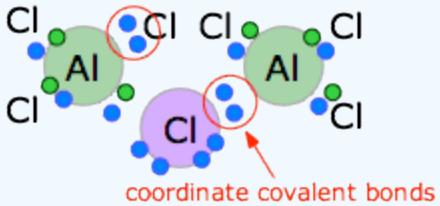
Many solids that are formally derived from ions cannot really be said to form "ionic" solids at all. For example, anhydrous copper(II) chloride consists of layers of copper atoms surrounded by four chlorine atoms in a square arrangement. Neighboring chains are offset so as to create an octahedral coordination of each copper atom. Similar structures are commonly encountered for other salts of transition metals. Similarly, most oxides and sulfides of metals beyond Group 2 tend to have structures dominated by other than ion-ion attractions.

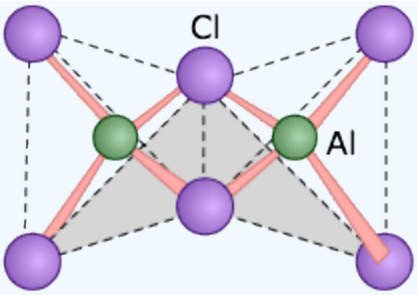
Aluminum Halides

The trihalides of aluminum offer another example of the dangers of assuming ionic character of solids that are formally derived from ions. Aqueous solutions of what we assume to be AlF_3, $AlCl_3$, $AlBr_3$, and AlI_3 all exhibit the normal properties ionic solutions (they are electrically conductive, for example), but the solids are quite different: the melting point of AlF_3 is 1290°C, suggesting that it is indeed ionic. But $AlCl_3$ melts at 192°C — hardly consistent with ionic bonding, and the other two halides are also rather low-melting. Structural studies show that when $AlCl_3$ vaporizes or dissolves in a non-polar solvent it forms a dimer Al_2Cl_6. The two other halides exist only as dimers in all states.



The structural formula of the Al_2Cl_6 molecule shows that the aluminum atoms are bonded to four chlorines, two of which are shared between the two metal atoms. The arrows represent coordinate covalent bonds in which the bonding electrons both come from the same atom (chlorine in this case.)





As shown at the right above, the aluminum atoms can be considered to be located at the centers of two tetrahedra that possess one edge in common.

This page titled 7.7: Ionic and Ion-Derived Solids is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.8: Cubic Lattices and Close Packing

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- The difference between square and hexagonal packing in two dimensions.
- The definition and significance of the unit cell.
- Sketch the three Bravais lattices of the cubic system, and calculate the number of atoms contained in each of these unit cells.
- Show how alternative ways of stacking three close-packed layers can lead to the hexagonal or cubic close packed structures.
- Explain the origin and significance of octahedral and tetrahedral holes in stacked close-packed layers, and show how they can arise.

Close-Packing of Identical Spheres

Crystals are of course three-dimensional objects, but we will begin by exploring the properties of arrays in two-dimensional space. This will make it easier to develop some of the basic ideas without the added complication of getting you to visualize in 3-D — something that often requires a bit of practice. Suppose you have a dozen or so marbles. How can you arrange them in a single compact layer on a table top? Obviously, they must be in contact with each other in order to minimize the area they cover. It turns out that there are two efficient ways of achieving this:



The essential difference here is that any marble within the interior of the square-packed array is in contact with four other marbles, while this number rises to six in the hexagonal-packed arrangement. It should also be apparent that the latter scheme covers a smaller area (contains less empty space) and is therefore a more efficient packing arrangement. If you are good at geometry, you can show that square packing covers 78 percent of the area, while hexagonal packing yields 91 percent coverage.

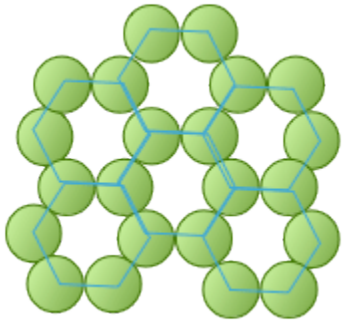
If we go from the world of marbles to that of atoms, which kind of packing would the atoms of a given element prefer?



If the atoms are identical and are bound together mainly by dispersion forces which are completely non-directional, they will favor a structure in which as many atoms can be in direct contact as possible. This will, of course, be the hexagonal arrangement.

Directed chemical bonds between atoms have a major effect on the packing. The version of hexagonal packing shown at the right

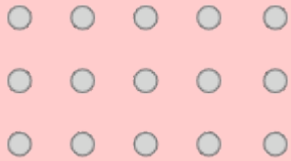
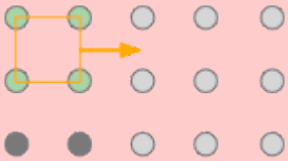
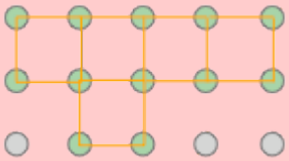
occurs in the form of carbon known as *graphite* which forms 2-dimensional sheets. Each carbon atom within a sheet is bonded to three other carbon atoms. The result is just the basic hexagonal structure with some atoms missing.



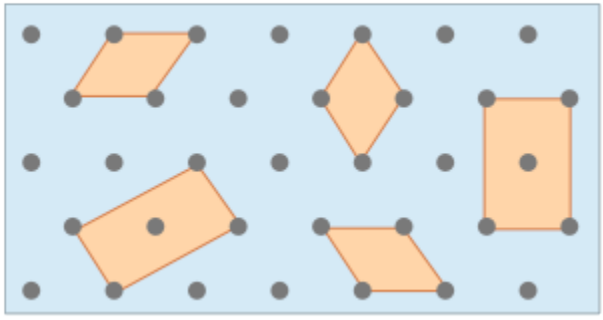
The coordination number of 3 reflects the sp^2-hybridization of carbon in graphite, resulting in plane-trigonal bonding and thus the sheet structure. Adjacent sheets are bound by weak dispersion forces, allowing the sheets to slip over one another and giving rise to the lubricating and flaking properties of graphite.

Lattices

The underlying order of a crystalline solid can be represented by an array of regularly spaced points that indicate the locations of the crystal's basic structural units. This array is called a crystal lattice. Crystal lattices can be thought of as being built up from repeating units containing just a few atoms. These repeating units act much as a rubber stamp: press it on the paper, move ("translate") it by an amount equal to the lattice spacing, and stamp the paper again.

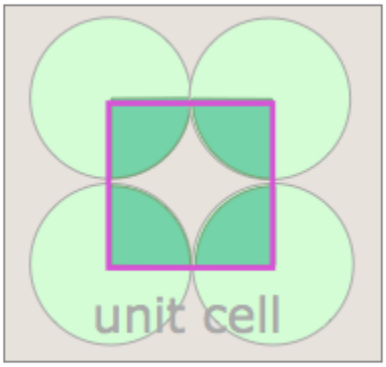
		
The gray circles represent a square array of lattice points.	The orange square is the simplest unit cell that can be used to define the 2-dimensional lattice.	Building out the lattice by moving ("translating") the unit cell in a series of steps,

Although real crystals do not actually grow in this manner, this process is conceptually important because it allows us to classify a lattice type in terms of the simple repeating unit that is used to "build" it. We call this shape the *unit cell*. Any number of primitive shapes can be used to define the unit cell of a given crystal lattice. The one that is actually used is largely a matter of convenience, and it may contain a lattice point in its center, as you see in two of the unit cells shown here. In general, the best unit cell is the simplest one that is capable of building out the lattice.

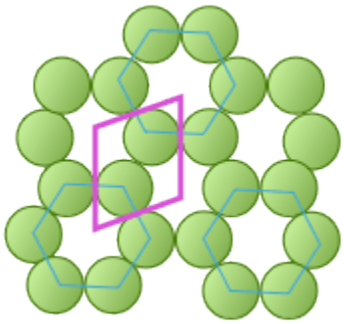




Shown above are unit cells for the close-packed square and hexagonal lattices we discussed near the start of this lesson. Although we could use a hexagon for the second of these lattices, the rhombus is preferred because it is simpler.



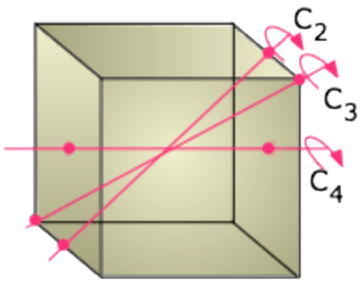
Notice that in both of these lattices, the corners of the unit cells are centered on a lattice point. This means that an atom or molecule located on this point in a real crystal lattice is shared with its neighboring cells. As is shown more clearly here for a two-dimensional square-packed lattice, a single unit cell can claim "ownership" of only one-quarter of each molecule, and thus "contains" 4 × ¼ = 1 molecule.



The unit cell of the graphite form of carbon is also a rhombus, in keeping with the hexagonal symmetry of this arrangement. Notice that to generate this structure from the unit cell, we need to shift the cell in both the *x*- and *y*- directions in order to leave empty spaces at the correct spots. We could alternatively use regular hexagons as the unit cells, but the *x*+*y* shifts would still be required, so the simpler rhombus is usually preferred. As you will see in the next section, the empty spaces within these unit cells play an important role when we move from two- to three-dimensional lattices.

Cubic crystals

In order to keep this lesson within reasonable bounds, we are limiting it mostly to crystals belonging to the so-called *cubic* system. In doing so, we can develop the major concepts that are useful for understanding more complicated structures (as if there are not enough complications in cubics alone!) But in addition, it happens that cubic crystals are very commonly encountered; most metallic elements have cubic structures, and so does ordinary salt, sodium chloride.

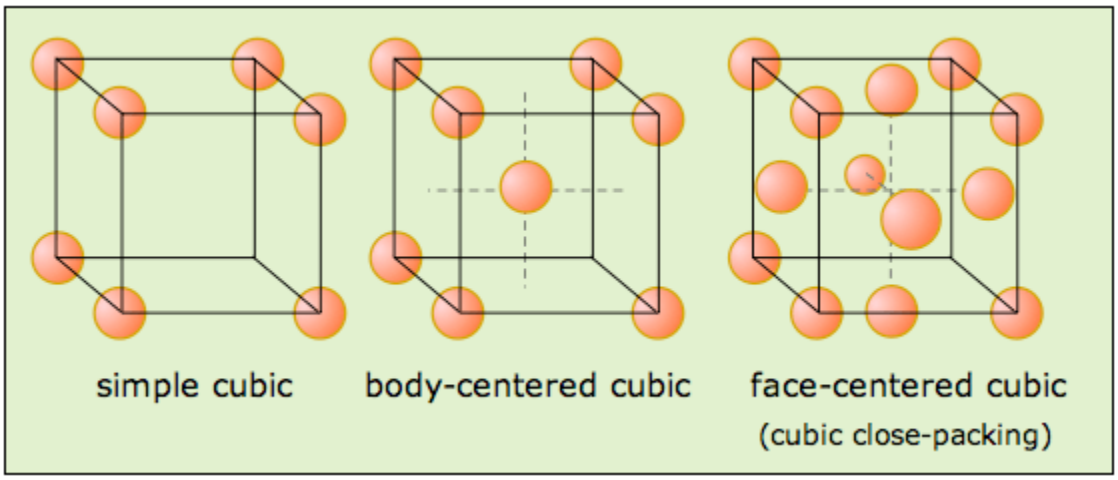


We usually think of a cubic shape in terms of the equality of its edge lengths and the 90° angles between its sides, but there is another way of classifying shapes that chemists find very useful. This is to look at what *geometric transformations* (such as rotations around an axis) we can perform that leave the appearance unchanged. For example, you can rotate a cube 90° around an axis perpendicular to any pair of its six faces without making any apparent change to it. We say that the cube possesses three mutually perpendicular *four-fold rotational axes*, abbreviated C_4 axes. But if you think about it, a cube can also be rotated around the axes that extend between opposite corners; in this case, it takes three 120° rotations to go through a complete circle, so these axes (also four in number) are three-fold or C_3 axes.

Cubic crystals belong to one of the seven crystal systems whose lattice points can be extended indefinitely to fill three-dimensional space and which can be constructed by successive translations (movements) of a primitive unit cell in three dimensions. As we will see below, the cubic system, as well as some of the others, can have variants in which additional lattice points can be placed at the center of the unit or at the center of each face.

The three types of cubic lattices

The three Bravais lattices which form the cubic crystal system are shown here.



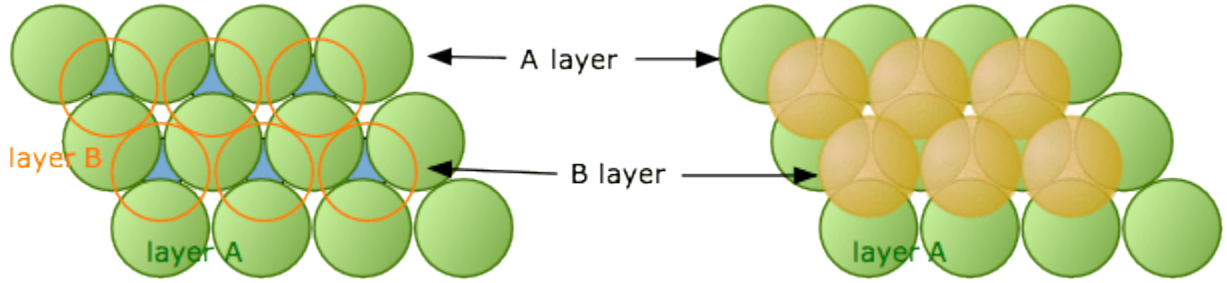
Structural examples of all three are known, with body- and face-centered (BCC and FCC) being much more common; most metallic elements crystallize in one of these latter forms. But although the simple cubic structure is uncommon by itself, it turns out that many BCC and FCC structures composed of ions can be regarded as interpenetrating combinations of two simple cubic lattices, one made up of positive ions and the other of negative ions. Notice that only the FCC structure, which we will describe below, is a close-packed lattice within the cubic system.

Close-packed lattices in three dimensions

Close-packed lattices allow the maximum amount of interaction between atoms. If these interactions are mainly attractive, then close-packing usually leads to more energetically stable structures. These lattice geometries are widely seen in metallic, atomic, and simple ionic crystals.

As we pointed out above, hexagonal packing of a single layer is more efficient than square-packing, so this is where we begin.

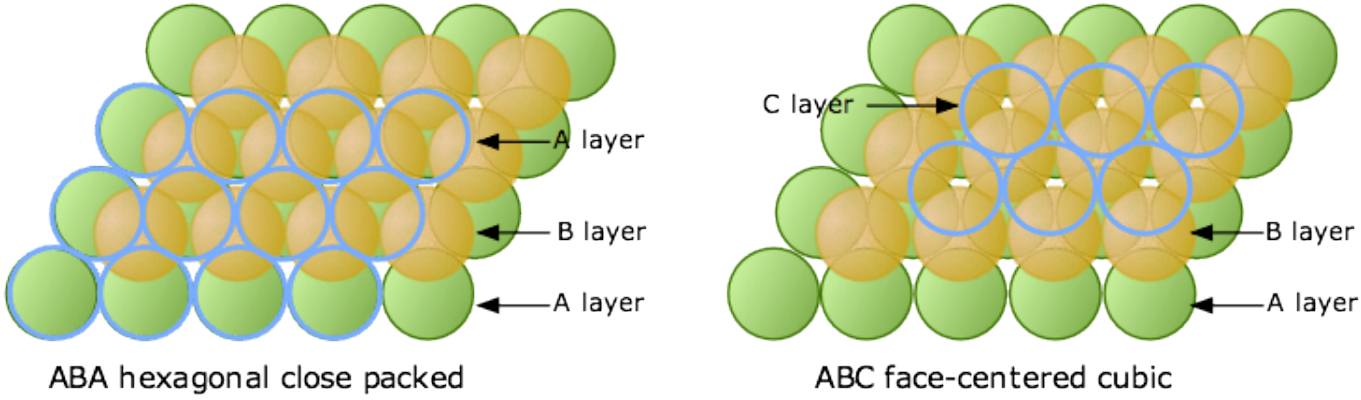
Imagine that we start with the single layer of green atoms shown below. We will call this the A layer. If we place a second layer of atoms (orange) on top of the A-layer, we would expect the atoms of the new layer to nestle in the hollows in the first layer. But if all the atoms are identical, only some of these void spaces will be accessible.



In the diagram on the left, notice that there are two classes of void spaces between the A atoms; one set (colored blue) has a vertex pointing up, while the other set (not colored) has down-pointing vertices. Each void space constitutes a depression in which atoms of a second layer (the B-layer) can nest. The two sets of void spaces are completely equivalent, but only one of these sets can be occupied by a second layer of atoms whose size is similar to those in the bottom layer. In the illustration on the right above we have arbitrarily placed the B-layer atoms in the blue voids, but could just as well have selected the white ones.

Two choices for the third layer lead to two different close-packed lattice types

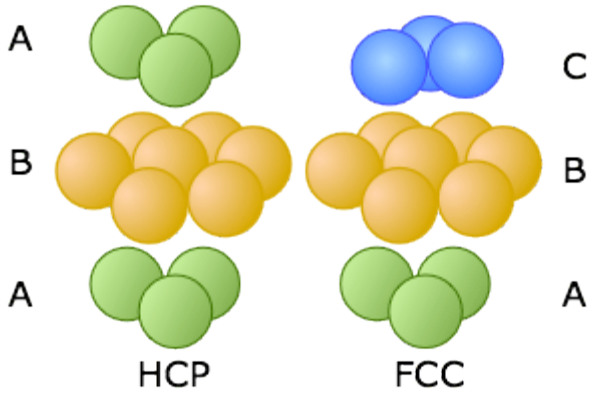
Now consider what happens when we lay down a third layer of atoms. These will fit into the void spaces within the B-layer. As before, there are two sets of these positions, but unlike the case described above, they are not equivalent.



The atoms in the third layer are represented by open blue circles in order to avoid obscuring the layers underneath. In the illustration on the left, this third layer is placed on the B-layer at locations that are directly above the atoms of the A-layer, so our third layer is just a another A layer. If we add still more layers, the vertical sequence A-B-A-B-A-B-A... repeats indefinitely.

In the diagram on the right above, the blue atoms have been placed above the white (unoccupied) void spaces in layer A. Because this third layer is displaced horizontally (in our view) from layer A, we will call it layer C. As we add more layers of atoms, the sequence of layers is A-B-C-A-B-C-A-B-C..., so we call it ABC packing.

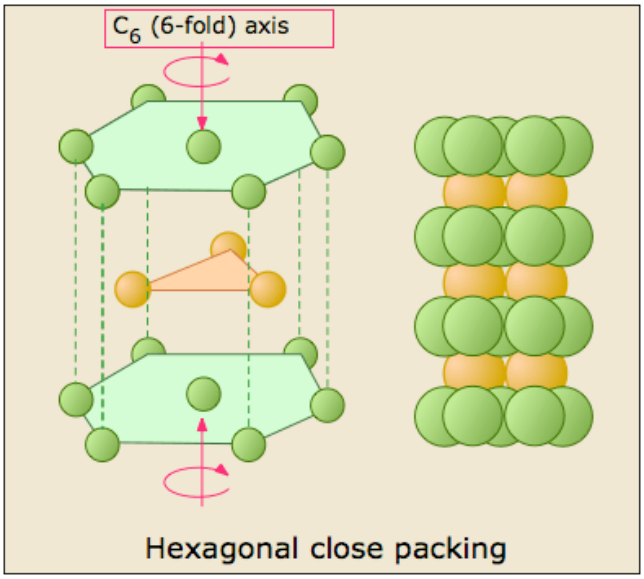
> For the purposes of clarity, only three atoms of the A and C layers are shown in these diagrams. But in reality, each layer consists of an extended hexagonal array; the two layers are simply displaced from one another.



These two diagrams that show exploded views of the vertical stacking further illustrate the rather small fundamental difference between these two arrangements— but, as you will see below, they have widely divergent structural consequences. Note the opposite orientations of the A and C layers

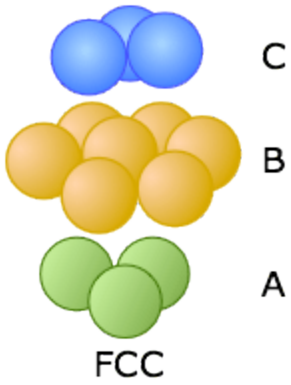
The Hexagonal closed-packed structure

The HCP stacking shown on the left just above takes us out of the cubic crystal system into the hexagonal system, so we will not say much more about it here except to point out each atom has 12 nearest neighbors: six in its own layer, and three in each layer above and below it.

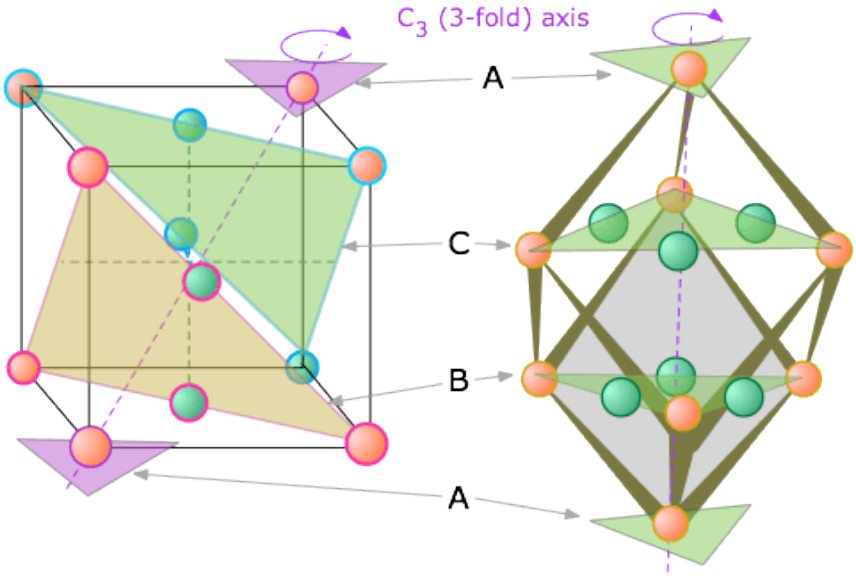


The cubic close-packed structure

Below we reproduce the FCC structure that was shown above.



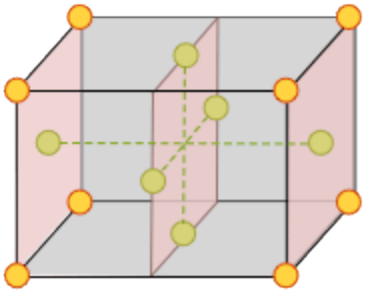
You will notice that the B-layer atoms form a hexagon, but this is a *cubic* structure. How can this be? The answer is that the FCC stack is inclined with respect to the faces of the cube, and is in fact coincident with one of the three-fold axes that passes through opposite corners. It requires a bit of study to see the relationship, and we have provided two views to help you. The one on the left shows the cube in the normal isometric projection; the one on the right looks down upon the top of the cube at a slightly inclined angle.



Both the CCP and HCP structures fill 74 percent of the available space when the atoms have the same size. You should see that the two shaded planes cutting along diagonals within the interior of the cube contain atoms of different colors, meaning that they belong to different layers of the CCP stack. Each plane contains three atoms from the B layer and three from the C layer, thus reducing the symmetry to C_3, which a cubic lattice must have.

The FCC unit cell

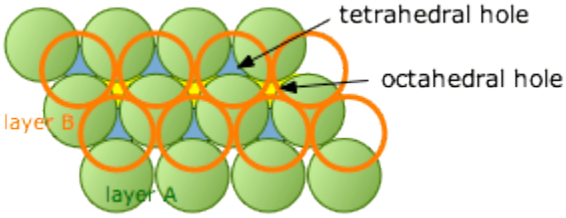
The figure below shows the the face-centered cubic unit cell of a cubic-close packed lattice.



How many atoms are contained in a unit cell? Each corner atom is shared with eight adjacent unit cells and so a single unit cell can claim only 1/8 of each of the eight corner atoms. Similarly, each of the six atoms centered on a face is only half-owned by the cell. The grand total is then (8 × 1/8) + (6 × ½) = 4 atoms per unit cell.

Interstitial Void Spaces

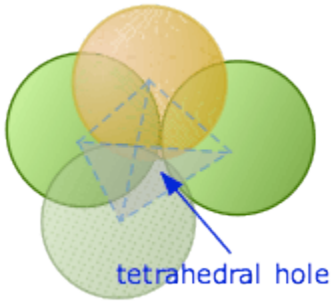
The atoms in each layer in these close-packing stacks sit in a depression in the layer below it. As we explained above, these void spaces are not completely filled. (It is geometrically impossible for more than two identical spheres to be in contact at a single point.) We will see later that these *interstitial void spaces* can sometimes accommodate additional (but generally smaller) atoms or ions.



If we look down on top of two layers of close-packed spheres, we can pick out two classes of void spaces which we call *tetrahedral* and *octahedral holes*.

Tetrahedral holes

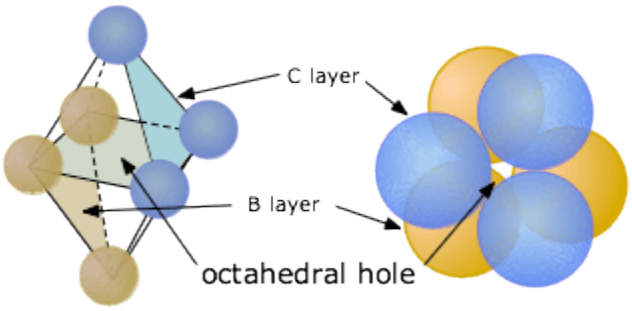
If we direct our attention to a region in the above diagram where a single atom is in contact with the three atoms in the layers directly below it, the void space is known as a *tetrahedral hole*. A similar space will be be found between this single atom and the three atoms (not shown) that would lie on top of it in an extended lattice. Any interstitial atom that might occupy this site will interact with the four atoms surrounding it, so this is also called a *four-coordinate interstitial space*.



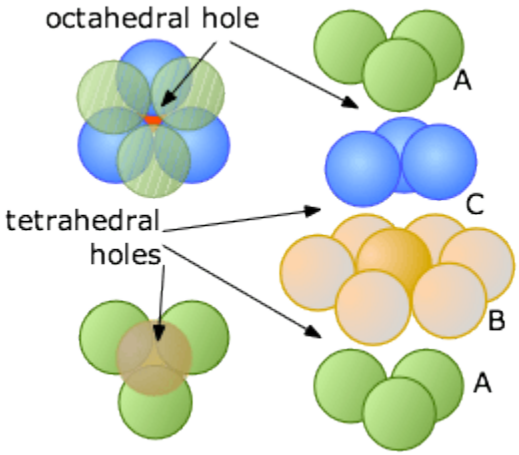
Don't be misled by this name; the boundaries of the void space are spherical sections, not tetrahedra. The tetrahedron is just an imaginary construction whose four corners point to the centers of the four atoms that are in contact.

Octahedral holes

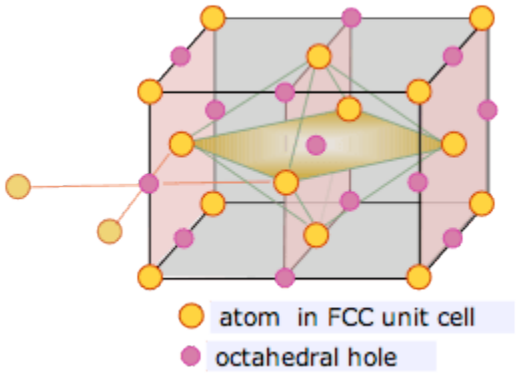
Similarly, when two sets of three trigonally-oriented spheres are in close-packed contact, they will be oriented 60° apart and the centers of the spheres will define the six corners of an imaginary octahedron centered in the void space between the two layers, so we call these *octahedral holes* or *six-coordinate interstitial sites*. Octahedral sites are larger than tetrahedral sites.



An octahedron has six corners and eight sides. We usually draw octahedra as a double square pyramid standing on one corner (left), but in order to visualize the octahedral shape in a close-packed lattice, it is better to think of the octahedron as lying on one of its faces (right).



Each sphere in a close-packed lattice is associated with one octahedral site, whereas there are only half as many tetrahedral sites. This can be seen in this diagram that shows the central atom in the B layer in alignment with the hollows in the C and A layers above and below.



The face-centered cubic unit cell contains a single octahedral hole within itself, but octahedral holes shared with adjacent cells exist at the centers of each edge. Each of these twelve edge-located sites is shared with four adjacent cells, and thus contributes (12 × ¼) = 3 atoms to the cell. Added to the single hole contained in the middle of the cell, this makes a total of 4 octahedral sites per unit cell. This is the same as the number we calculated above for the number of atoms in the cell.

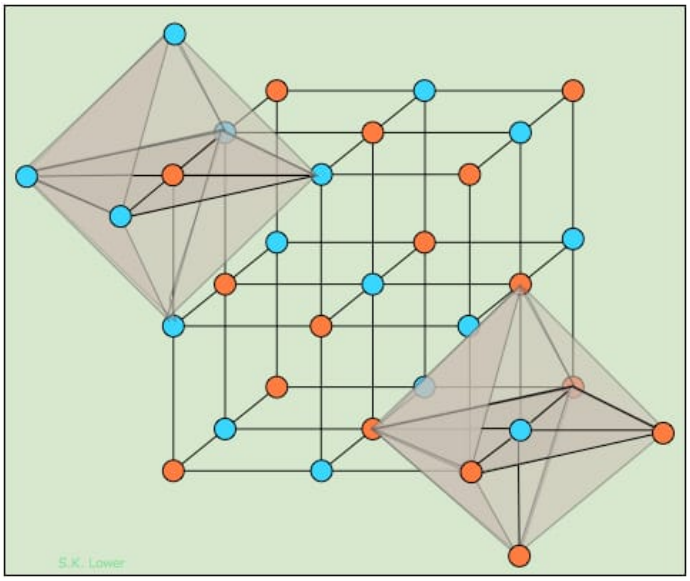
Common cubic close-packed structures

It can be shown from elementary trigonometry that an atom will fit exactly into an octahedral site if its radius is 0.414 as great as that of the host atoms. The corresponding figure for the smaller tetrahedral holes is 0.225.

Many pure metals and compounds form face-centered cubic (cubic close- packed) structures. The existence of tetrahedral and octahedral holes in these lattices presents an opportunity for "foreign" atoms to occupy some or all of these interstitial sites. In order to retain close-packing, the interstitial atoms must be small enough to fit into these holes without disrupting the host CCP lattice. When these atoms are too large, which is commonly the case in ionic compounds, the atoms in the interstitial sites will push the host atoms apart so that the face-centered cubic lattice is somewhat opened up and loses its close-packing character.

The rock-salt structure

Alkali halides that crystallize with the "rock-salt" structure exemplified by sodium chloride can be regarded either as a FCC structure of one kind of ion in which the octahedral holes are occupied by ions of opposite charge, or as two interpenetrating FCC lattices made up of the two kinds of ions. The two shaded octahedra illustrate the identical coordination of the two kinds of ions; each atom or ion of a given kind is surrounded by six of the opposite kind, resulting in a coordination expressed as (6:6).

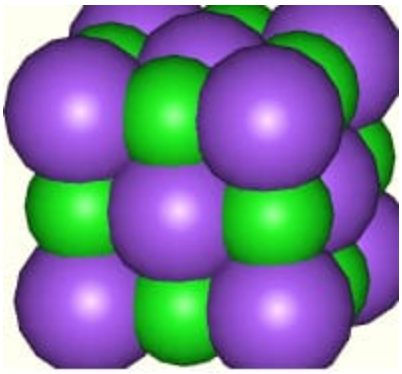


How many NaCl units are contained in the unit cell? If we ignore the atoms that were placed outside the cell in order to construct the octahedra, you should be able to count fourteen "orange" atoms and thirteen "blue" ones. But many of these are shared with adjacent unit cells.

An atom at the corner of the cube is shared by eight adjacent cubes, and thus makes a 1/8 contribution to any one cell. Similarly, the center of an edge is common to four other cells, and an atom centered in a face is shared with two cells. Taking all this into consideration, you should be able to confirm the following tally showing that there are four AB units in a unit cell of this kind.

Orange	Blue
8 at corners: 8 x 1/8 = 1	12 at edge centers: 12 x ¼ = 3
6 at face centers: 6 x ½ = 3	1 at body center = 1
total: 4	total: 4

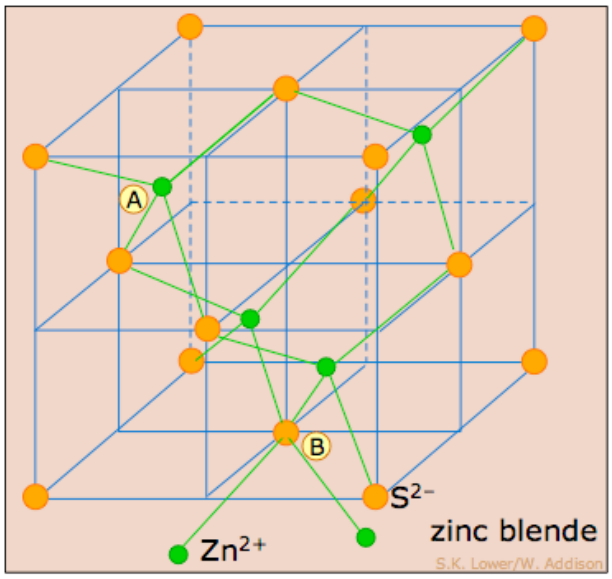
If we take into consideration the actual sizes of the ions (Na^+ = 116 pm, Cl^- = 167 pm), it is apparent that neither ion will fit into the octahedral holes with a CCP lattice composed of the other ion, so the actual structure of NaCl is somewhat expanded beyond the close-packed model.



The space-filling model on the right depicts a face-centered cubic unit cell of chloride ions (purple), with the sodium ions (green) occupying the octahedral sites.

The zinc-blende structure: using some tetrahedral holes

Since there are two tetrahedral sites for every atom in a close-packed lattice, we can have binary compounds of 1:1 or 1:2 stoichiometry depending on whether half or all of the tetrahedral holes are occupied. Zinc-blende is the mineralogical name for zinc sulfide, ZnS. An impure form known as *sphalerite* is the major ore from which zinc is obtained.

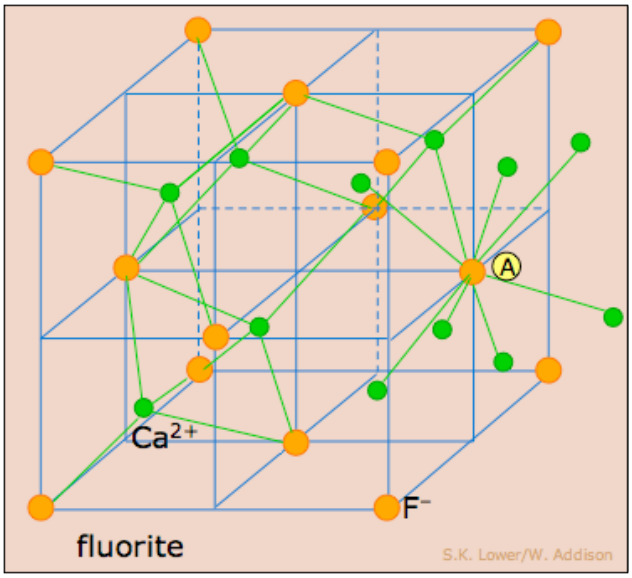


This structure consists essentially of a FCC (CCP) lattice of sulfur atoms (orange) (equivalent to the lattice of chloride ions in NaCl) in which zinc ions (green) occupy half of the tetrahedral sites. As with any FCC lattice, there are four atoms of sulfur per unit cell, and the the four zinc atoms are totally contained in the unit cell. Each atom in this structure has **four** nearest neighbors, and is thus tetrahedrally coordinated.

It is interesting to note that if all the atoms are replaced with carbon, this would correspond to the *diamond* structure.

The fluorite structure: all tetrahedral sites occupied

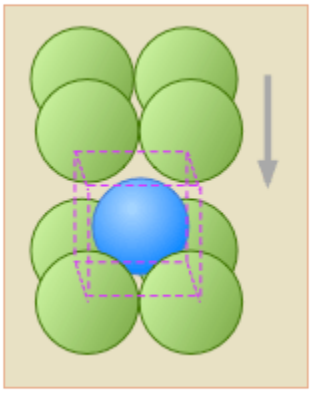
Fluorite, CaF_2, having twice as many ions of fluoride as of calcium, makes use of all eight tetrahedral holes in the CPP lattice of calcium ions (orange) depicted here. To help you understand this structure, we have shown some of the octahedral sites in the next cell on the right; you can see that the calcium ion at **A** is surrounded by eight fluoride ions, and this is of course the case for all of the calcium sites. Since each fluoride ion has four nearest-neighbor calcium ions, the coordination in this structure is described as (8:4).



Although the radii of the two ions (F^-= 117 pm, Ca^{2+} = 126 pm does not allow true close packing, they are similar enough that one could just as well describe the structure as a FCC lattice of fluoride ions with calcium ions in the octahedral holes.

Simple- and body-centered cubic structures

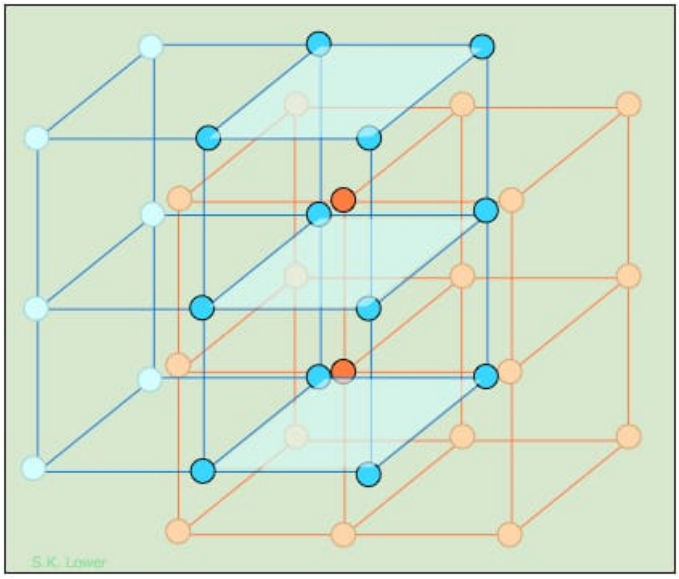
In Section 4 we saw that the only cubic lattice that can allow close packing is the face-centered cubic structure. The simplest of the three cubic lattice types, the *simple cubic lattice*, lacks the hexagonally-arranged layers that are required for close packing. But as shown in this exploded view, the void space between the two square-packed layers of this cell constitutes an octahedral hole that can accommodate another atom, yielding a packing arrangement that in favorable cases can approximate true close-packing. Each second-layer B atom (blue) resides within the unit cell defined the A layers above and below it.



The A and B atoms can be of the same kind or they can be different. If they are the same, we have a *body-centered cubic lattice*. If they are different, and especially if they are oppositely-charged ions (as in the CsCl structure), there are size restrictions: if the B atom is too large to fit into the interstitial space, or if it is so small that the A layers (which all carry the same electric charge) come into contact without sufficient A-B coulombic attractions, this structural arrangement may not be stable.

The cesium chloride structure

CsCl is the common model for the BCC structure. As with so many other structures involving two different atoms or ions, we can regard the same basic structure in different ways. Thus if we look beyond a single unit cell, we see that CsCl can be represented as two interpenetrating simple cubic lattices in which each atom occupies an octahedral hole within the cubes of the other lattice.



This page titled 7.8: Cubic Lattices and Close Packing is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.9: Polymers and Plastics

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the bold terms in the context of this topic.

- Aside from their high molar masses, how do synthetic polymers differ from ordinary molecular solids?
- Polymers can be classified according to their chemical composition, their physical properties, and their general application. For each of these three categories, name two examples that that might be considered when adapting a polymer to a particular end-use.
- and a ***thermoset***, and comment on the molecular basis for their different properties, including crystallinity.
- Describe the two general methods of polymer synthesis.
- Name two kinds each of commercially-important synthetic thermoplastics and thermosets, and specify some of their principal uses.
- Name two kinds of commercially-important natural polymers.
- Describe some of the concerns and sources of small-molecule release from polymers.
- What are some of the problems connected with recycling or re-use of polymeric materials?

Plastics and natural materials such as rubber or cellulose are composed of very large molecules called polymers. Polymers are constructed from relatively small molecular fragments known as monomers that are joined together. Wool, cotton, silk, wood and leather are examples of natural polymers that have been known and used since ancient times. This group includes biopolymers such as proteins and carbohydrates that are constituents of all living organisms.

Synthetic polymers, which includes the large group known as plastics, came into prominence in the early twentieth century. Chemists' ability to engineer them to yield a desired set of properties (strength, stiffness, density, heat resistance, electrical conductivity) has greatly expanded the many roles they play in the modern industrial economy. This Module deals mostly with synthetic polymers, but will include a synopsis of some of the more important natural polymers. It will close with a summary of some of the very significant environmental problems created by the wide use of plastics.

Polymers and "pure substances"

Let's begin by looking at an artificial polymer that is known to everyone in the form of flexible, transparent **plastic bags**: *polyethylene*. It is also the simplest polymer, consisting of random-length (but generally very long) chains made up of two-carbon units.

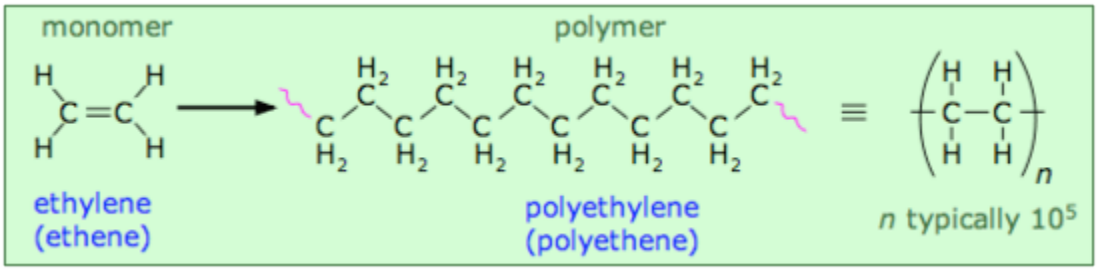


Figure 7.9.1: Polymers

You will notice some "fuzziness" in the way that the polyethylene structures are represented above. The squiggly lines at the ends of the long structure indicate that the same pattern extends indefinitely. The more compact notation on the right shows the minimal repeating unit enclosed in brackets overprinted with a dash; this means the same thing and is the preferred way of depicting polymer structures.

In most areas of chemistry, a "pure substance" has a definite structure, molar mass, and properties. It turns out, however, that few polymeric substances are uniform in this way. This is especially the case with synthetic polymers, whose molecular weights cover

a range of values, as may the sequence, orientation, and connectivity of the individual monomers. So most synthetic polymers are really *mixtures* rather than pure substances in the ordinary chemical sense of the term. Their molecular weights are typically distributed over a wide range.

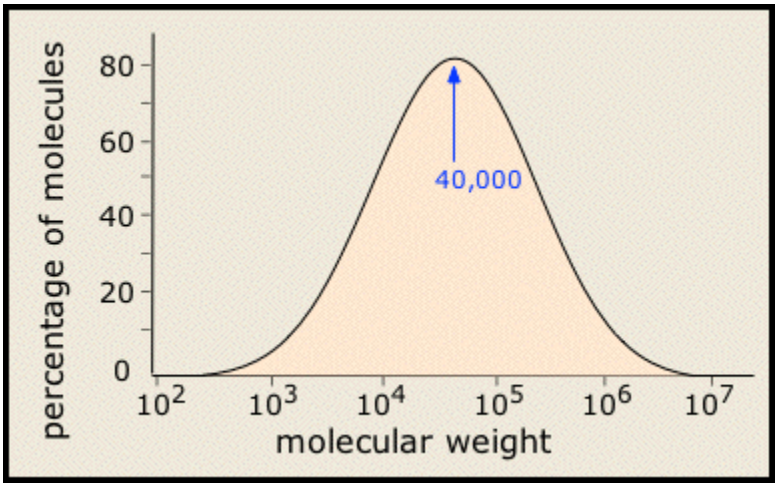


Figure 7.9.1: Polymers

Don't be misled by chemical formulas that depict polymers such as polyethylene as reasonably straight chains of substituted carbon atoms. Free rotation around C—C bonds allows long polymer molecules to curl up and and tangle very much like spaghetti (Figure 7.9.2). Thus polymers generally form *amorphous* solids. There are, however, ways in which certain polymers can be partially oriented.



Figure 7.9.2: The structure of polymers more closely resemble spaghetti than the linear structure in Figure 7.9.1. from Wikipedia

Classification of polymers

Polymers can be classified in ways that reflect their chemical makeup, or perhaps more importantly, their properties and applications. Many of these factors are strongly interdependent, and most are discussed in much more detail in subsequent sections of this page.

Chemistry

- Nature of the monomeric units
- Average chain length and molecular weight
- *Homopolymers* (one kind of monomeric unit) or *copolymers*;
- Chain topology: how the monomeric units are connected
- Presence or absence of cross-branching
- Method of polymerization

Properties

- Density
- Thermal properties — can they soften or melt when heated?
- Degree of crystallinity
- Physical properties such as hardness, strength, machineability.
- Solubility, permeability to gases

Applications

- molded and formed objects ("plastics")
- sheets and films
- elastomers (i.e., elastic polymers such as rubber)
- adhesives
- coatings, paints, inks
- fibers and yarns

Physical properties of polymers

The physical properties of a polymer such as its strength and flexibility depend on:

- *chain length* - in general, the longer the chains the stronger the polymer;
- *side groups* - polar side groups (including those that lead to hydrogen bonding) give stronger attraction between polymer chains, making the polymer stronger;
- *branching* - straight, unbranched chains can pack together more closely than highly branched chains, giving polymers that have higher density, are more crystalline and therefore stronger;
- *cross-linking* - if polymer chains are linked together extensively by covalent bonds, the polymer is harder and more difficult to melt.

Amorphous and crystalline polymers

The spaghetti-like entanglements of polymer molecules tend to produce amorphous solids, but it often happens that some parts can become sufficiently aligned to produce a region exhibiting crystal-like order, so it is not uncommon for some polymeric solids to consist of a random mixture of amorphous and crystalline regions. As might be expected, shorter and less-branched polymer chains can more easily organize themselves into ordered layers than have can long chains. Hydrogen-bonding between adjacent chains also helps, and is very important in fiber-forming polymers both synthetic (Nylon 6.6) and natural (cotton cellulose).

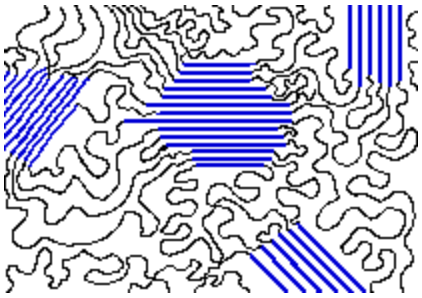


Figure 7.9.3: The crystalline parts of this polymer are shown in blue.

Pure crystalline solids have definite melting points, but polymers, if they melt at all, exhibit a more complex behavior. At low temperatures, the tangled polymer chains tend to behave as rigid glasses. For example, the natural polymer that we call *rubber* becomes hard and brittle when cooled to liquid nitrogen temperature. Many synthetic polymers remain in this state to well above room temperature.

The melting of a crystalline compound corresponds to a sudden loss of long-range order; this is the fundamental reason that such solids exhibit definite melting points, and it is why there is no intermediate form between the liquid and the solid states. In amorphous solids there is no long-range order, so there is no melting point in the usual sense. Such solids simply become less and less viscous as the temperature is raised.

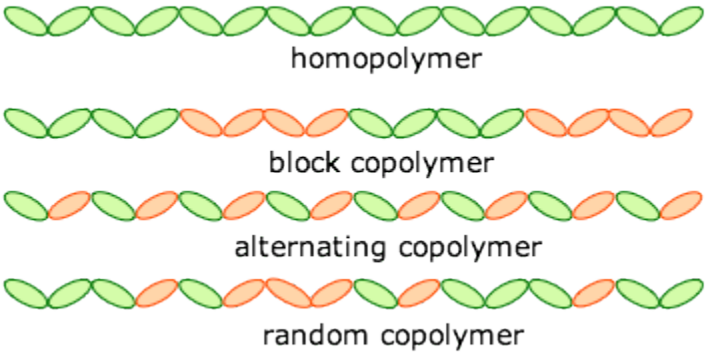
In some polymers (known as *thermoplastics*) there is a fairly definite softening point that is observed when the thermal kinetic energy becomes high enough to allow internal rotation to occur within the bonds and to allow the individual molecules to slide independently of their neighbors, thus rendering them more flexible and deformable. This defines the *glass transition temperature* t_g .

Depending on the degree of crystallinity, there will be a higher temperature, the *melting point* t_m , at which the crystalline regions come apart and the material becomes a viscous liquid. Such liquids can easily be injected into molds to manufacture objects of various shapes, or extruded into sheets or fibers. Other polymers (generally those that are highly cross-linked) do not melt at all; these are known as *thermosets*. If they are to be made into molded objects, the polymerization reaction must take place within the molds — a far more complicated process. About 20% of the commercially-produced polymers are thermosets; the remainder are thermoplastics.

2 Thermoplastic polymer structures

Homopolymers and heteropolymers

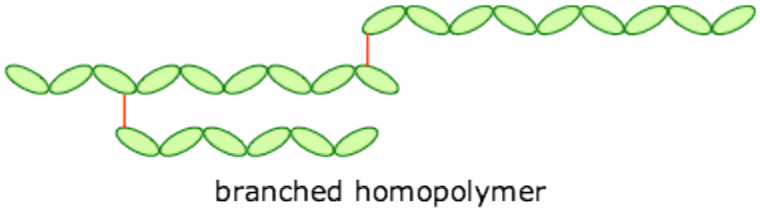
Copolymerization is an invaluable tool for "tuning" polymers so that they have the right combination of properties for an application. For example, homopolymeric polystyrene is a rigid and very brittle transparent thermoplastic with a glass transition temperature of 97°C. Copolymerizing it with acrylonitrile yields an alternating "SAN" copolymer in which t_g is raised to 107°, making it useable for transparent drink containers.



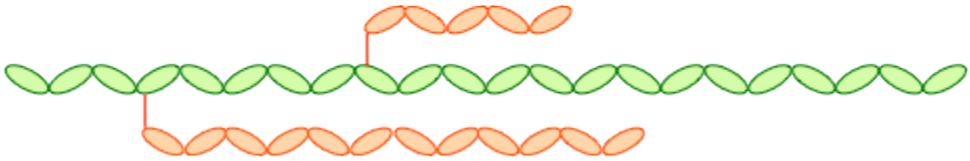
A polymer that is composed of identical monomeric units (as is polyethylene) is called a *homopolymer*. *Heteropolymers* are built up from more than one type of monomer. Artificial heteropolymers are more commonly known as copolymers.

Chain topology

Polymers may also be classified as straight-chained or branched, leading to forms such as these:

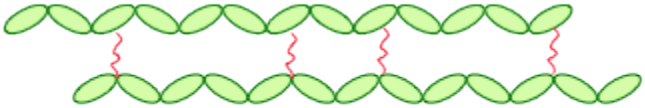


branched homopolymer



graft copolymer

The monomers can be joined end-to-end, and they can also be *cross-linked* to provide a harder material:



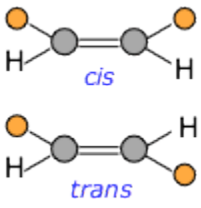
cross-linked homopolymer

If the cross-links are fairly long and flexible, adjacent chains can move with respect to each other, producing an *elastic* polymer or *elastomer*

Chain configuration and tacticity

In a linear polymer such as polyethylene, rotations around carbon-carbon single bonds can allow the chains to bend or curl up in various ways, resulting in the spaghetti-like mixture of these different *conformations* we alluded to above. But if one of the hydrogen atoms is replaced by some other entity such as a methyl group, the relative orientations of the individual monomer units that make up a linear section of any carbon chain becomes an important characteristic of the polymer.

***Cis-trans* isomerism** occurs because rotation around carbon-carbon double bonds is not possible — unlike the case for single bonds. Any pair of unlike substituents attached to the two carbons is permanently locked into being on the same side (*cis*) or opposite sides (*trans*) of the double bond.



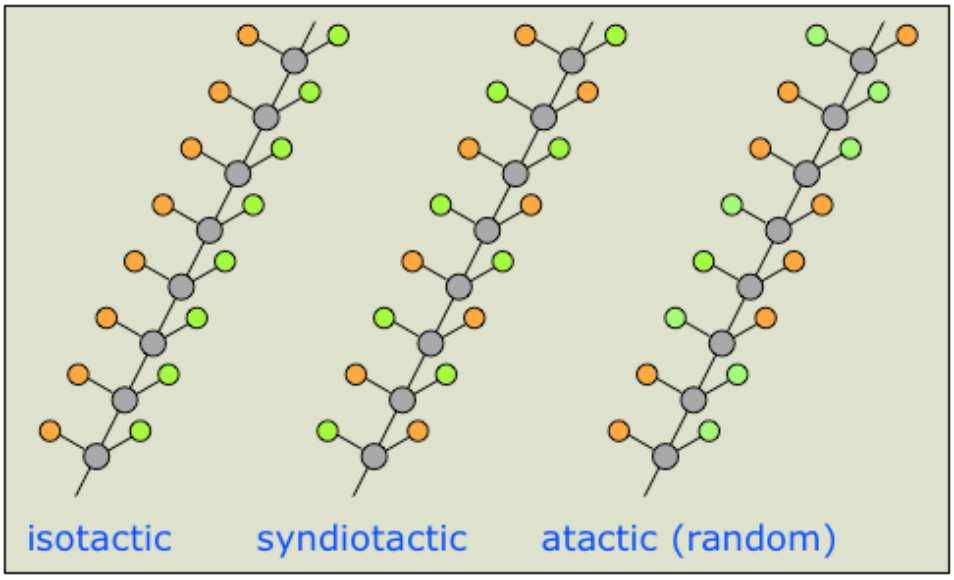
If the carbon chain contains double bonds, then *cis-trans isomerism* becomes possible, giving rise to two different possible configurations (known as diastereomers) at each unit of the chain. This seemingly small variable can profoundly affect the nature of the polymer. For example, the latex in natural rubber is made mostly of *cis*-polyisoprene, whereas the *trans* isomer (known as gutta percha latex) has very different (and generally inferior) properties.

Chirality

The tetrahedral nature of carbon bonding has an important consequence that is not revealed by simple two-dimensional structural formulas: atoms attached to the carbon can be on one side or on the other, and these will not be geometrically equivalent if all four of the groups attached to a single carbon atom are different. Such carbons (and the groups attached to them) are said to be *chiral,* and can exist in two different three-dimensional forms known as *enantiomers*.

For an individual carbon atom in a polymer chain, two of its attached groups will ordinarily be the chain segments on either side of the carbon. If the two remaining groups are different (say one hydrogen and the other methyl), then the above conditions are satisfied and this part of the chain can give rise to two enantiomeric forms.

A chain that can be represented as (in which the orange and green circles represent different groups) will have multiple chiral centers, giving rise to a huge number of possible enantiomers. In practice, it is usually sufficient to classify chiral polymers into the following three classes of *stereoregularity*, usually referred to as *tacticity*.



The tacticity of a polymer chain can have a major influence on its properties. Atactic polymers, for example, being more disordered, cannot crystallize.

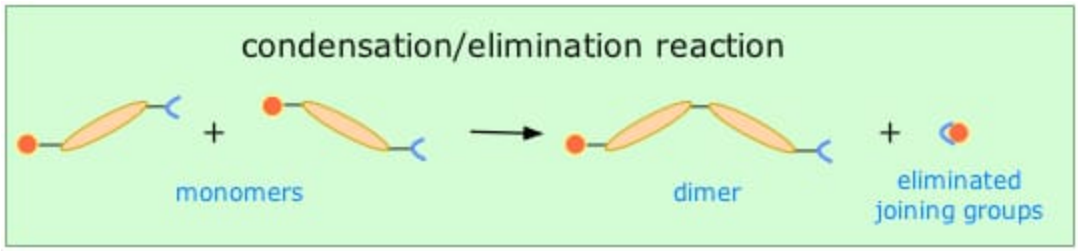
One of the major breakthroughs in polymer chemistry occurred in the early 1950s when the German chemist Karl Ziegler discovered a group of catalysts that could efficiently polymerize ethylene. At about the same time, Giulio Natta (Italian) made the first isotactic (and crystalline) polyethylene. The Ziegler-Natta catalysts revolutionized polymer chemistry by making it possible to control the stereoregularity of these giant molecules. The two shared the 1963 Nobel Prize in Chemistry

3 How polymers are made

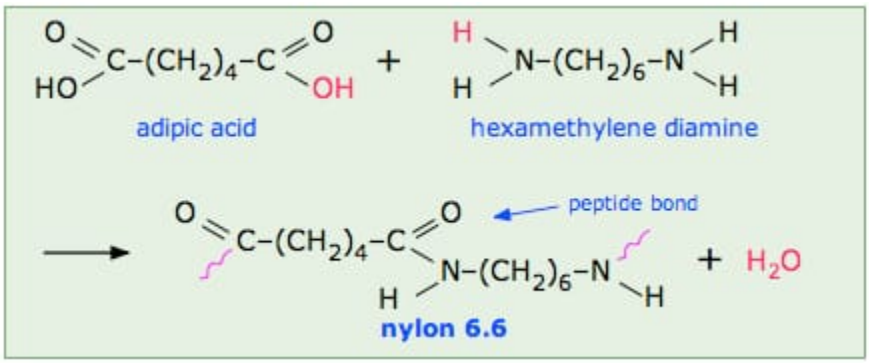
Polymers are made by joining small molecules into large ones. But most of these monomeric molecules are perfectly stable as they are, so chemists have devised two general methods to make them react with each other, building up the backbone chain as the reaction proceeds.

Condensation-elimination polymerization

This method (also known as *step-growth*) requires that the monomers possess two or more kinds of functional groups that are able to react with each other in such a way that parts of these groups combine to form a small molecule (often H_2O) which is eliminated from the two pieces. The now-empty bonding positions on the two monomers can then join together .



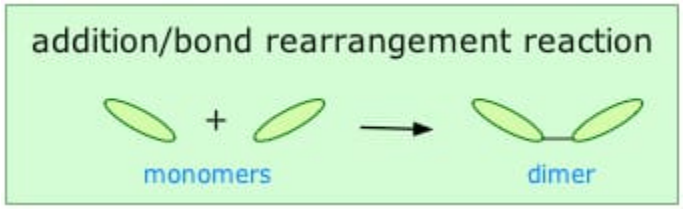
This occurs, for example, in the synthesis of the Nylon family of polymers in which the eliminated H_2O molecule comes from the hydroxyl group of the acid and one of the amino hydrogens:



Note that the monomeric units that make up the polymer are not identical with the starting components.

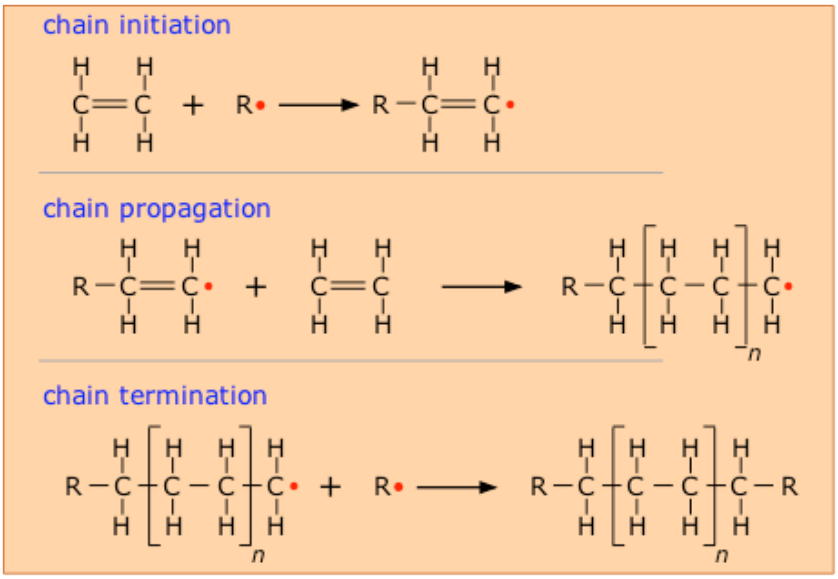
Addition polymerization

Addition or *chain-growth* polymerization involves the rearrangement of bonds within the monomer in such a way that the monomers link up directly with each other:



In order to make this happen, a chemically active molecule (called an *initiator*) is needed to start what is known as a *chain reaction*. The manufacture of polyethylene is a very common example of such a process. It employs a *free-radical* initiator that donates its unpaired electron to the monomer, making the latter highly reactive and able to form a bond with another monomer at this site.

In theory, only a single chain-initiation process needs to take place, and the chain-propagation step then repeats itself indefinitely, but in practice multiple initiation steps are required, and eventually two radicals react (*chain termination*) to bring the polymerization to a halt.



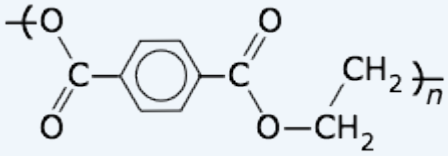
As with all polymerizations, chains having a range of molecular weights are produced, and this range can be altered by controlling the pressure and temperature of the process.

4 Gallery of common synthetic polymers

Thermoplastics

Note: the left panels below show the polymer name and synonyms, structural formula, glass transition temperature, melting point/ decomposition temperature, and (where applicable) the *resin identification symbol* used to facilitate recycling.

Polycarbonate (Lexan®)  T_g = 145°C, T_m = 225°C.	This polymer was discovered independently in Germany and the U.S. in 1953. Lexan is exceptionally hard and strong; we see it most commonly in the form of **compact disks**. It was once widely used in water bottles, but concerns about leaching of unreacted monomer (bisphenol-A) has largely suppressed this market.

Polyethylene terephthalate (PET, **Mylar**)  T_g = 76°C, T_m = 250°C. 	Thin and very strong films of this material are made by drawing out the molten polymer in both directions, thus orienting the molecules into a highly crystalline state that becomes "locked-in" on cooling. Its many applications include food packaging (in foil-laminated drink containers and microwaveable frozen-food containers), overhead-projector film, weather balloons, and as aluminum-coated reflective material in spacecraft and other applications.
Nylon (a *polyamide*)  T_g = 50°C, T_m = 255°C.	Nylon has a fascinating history, both scientific and cultural. It was invented by DuPont chemist Wallace Carothers (1896-1937). The common form **Nylon 6.6** has six carbon atoms in both parts of its chain; there are several other kinds. Notice that the two copolymer sub-units are held together by *peptide* bonds, the same kinds that join amino acids into proteins. Nylon 6.6 has good abrasion resistance and is self-lubricating, which makes it a good engineering material. It is also widely used as a fiber in carpeting, clothing, and tire cord. For an interesting account of the development of Nylon, see *Enough for One Liftetime: Wallace Carothers, Inventor of Nylon*by Ann Gaines (1971)

Polyacrylonitrile (Orlon, Acrilan, "acrylic" fiber)  T_g = 85°C, T_m = 318°C.	Used in the form of fibers in rugs, blankets, and clothing, especially cashmere-like sweaters. The fabric is very soft, but tends to "pill" — i.e., produce fuzz-like blobs. Owing to its low glass transition temperature, it requires careful treatment in cleaning and ironing.

Polyethylene



T_g = –78°C, T_m = 100°C.



LDPE

Control of polymerization by means of catalysts and additives has led to a large variety of materials based on polyethylene that exhibit differences in densities, degrees of chain branching and crystallinity, and cross-linking. Some major types are low-density (LDPE), linear low density (LLDPE), high-density (HDPE).

LDPE was the first commercial form (1933) and is used mostly for ordinary "plastic bags", but also for food containers and in six-pack soda can rings. Its low density is due to long-chain branching that inhibits close packing. LLDPE has less branching; its greater toughness allows its use in those annoyingly-thin plastic bags often found in food markets.

A "very low density" form (VLDPE) with extensive short-chain branching is now used for plastic stretch wrap (replacing the original component of Saran Wrap) and in flexible tubing.

HDPE has mostly straight chains and is therefore stronger. It is widely used in milk jugs and similar containers, garbage containers, and as an "engineering plastic" for machine parts.

 HDPE	
Polymethylmethacrylate (Plexiglass, Lucite, Perspex)  T_g = 114°C, T_m = 130-140°C.	This clear, colorless polymer is widely used in place of glass, where its greater impact resistance, lighter weight, and machineability are advantages. It is normally copolymerized with other substances to improve its properties. Aircraft windows, plastic signs, and lighting panels are very common applications. Its compatibility with human tissues has led to various medical applications, such as replacement lenses for cataract patients.

Polypropylene  T_g = –10°C, T_m = 173°C.  PP	Polypropylene is used alone or as a copolymer, usually with with ethylene. These polymers have an exceptionally wide range of uses — rope, binder covers, plastic bottles, staple yarns, non-woven fabrics, electric kettles. When uncolored, it is translucent but not transparent. Its resistance to fatigue makes it useful for food containers and their lids, and flip-top lids on bottled products such as ketchup.

polystyrene  T_g = 95°C, T_m = 240°C.  PS	Polystyrene is transparent but rather brittle, and yellows under uv light. Widely used for inexpensive packaging materials and "take-out trays", foam "packaging peanuts", CD cases, foam-walled drink cups, and other thin-walled and moldable parts.

polyvinyl acetate  T_g = 30°C	PVA is too soft and low-melting to be used by itself; it is commonly employed as a water-based emulsion in paints, wood glue and other adhesives.

polyvinyl chloride ("vinyl", "PVC")  T_g = 85°C, T_m = 240°C.  PVC	This is one of the world's most widely used polymers. By itself it is quite rigid and used in construction materials such as pipes, house siding, flooring. Addition of plasticizers make it soft and flexible for use in upholstery, electrical insulation, shower curtains and waterproof fabrics. There is some effort being made to phase out this polymer owing to environmental concerns (see below).

Synthetic rubbers

Neoprene (polychloroprene)
T_g = –70°C



Polybutadiene T_g < –90°C



Neoprene, invented in 1930, was the first mass-produced synthetic rubber. It is used for such things as roofing membranes and wet suits.

Polybutadiene substitutes a hydrogen for the chlorine; it is the major component (usually admixed with other rubbers) of tires. Synthetic rubbers played a crucial role in World War II

SBS (styrene-butadiene-styrene) rubber is a block copolymer whose special durability makes it valued for tire treads.

Polytetrafluroethylene (**Teflon**, PTFE)





Decomposes above 350°C.

This highly-crystalline fluorocarbon is exceptionally inert to chemicals and solvents. Water and oils do not wet it, which accounts for its use in cooking ware and other anti-stick applications, including personal care products.

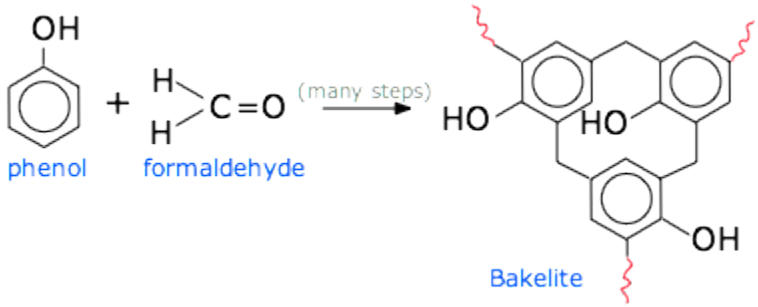
These properties — non-adhesion to other materials, non-wetability, and very low coefficient of friction ("slipperyness") — have their origin in the highly electronegative nature of fluorine whose atoms partly shield the carbon chain. Fluorine's outer electrons are so strongly attracted to its nucleus that they are less available to participate in London (dispersion force) interactions.

Polyaramid (Kevlar)  Sublimation temperature 450°C.	Kevlar is known for its ability to be spun into fibers that have five times the tensile strength of steel. It was first used in the 1970s to replace steel tire cords. Bullet-proof vests are one of it more colorful uses, but other applications include boat hulls, drum heads, sports equipment, and as a replacement for asbestos in brake pads. It is often combined with carbon or glass fibers in composite materials. The high tensile strength is due in part to the extensive hydrogen bonding between adjacent chains. Kevlar also has the distinction of having been invented by a woman chemist, Stephanie Kwolek.

Thermosets

The thermoplastic materials described above are chains based on relatively simple monomeric units having varying degrees of polymerization, branching, bending, cross-linking and crystallinity, but with each molecular chain being a discrete unit. In thermosets, the concept of an individual molecular unit is largely lost; the material becomes more like a gigantic extended molecule of its own — hence the lack of anything like a glass transition temperature or a melting point.

These properties have their origins in the nature of the monomers used to produce them. The most important feature is the presence of multiple reactive sites that are able to form what amount to cross-links at every center. The phenolic resins, typified by the reaction of phenol with formaldehyde, illustrate the multiplicity of linkages that can be built.



Phenolic resins

These are made by condensing one or more types of phenols (hydroxy-substituted benzene rings) with formaldehyde, as illustrated above. This was the first commercialized synthetic molding plastic. It was developed in 1907-1909 by the Belgian chemist Leo Baekeland, hence the common name bakelite. The brown material (usually bulked up with wood powder) was valued for its electrical insulating properties (light fixtures, outlets and other wiring devices) as well as for consumer items prior to the mid-century. Since that time, more recently developed polymers have largely displaced these uses. Phenolics are still extensively used as adhesives in plywood manufacture, and for making paints and varnishes.



Urea resins



Condensation of formaldehyde with urea yields lighter-colored and less expensive materials than phenolics. The major use if urea-formaldehyde resins is in bonding wood particles into particle board. Other uses are as baked-on enamel coatings for kitchen appliances and to coat cotton and rayon fibers to impart wrinkle- water-, and stain-resistance to the finished fabrics.

Melamine resins

Melamine, with even more amino ($-NH_2$) groups than urea, reacts with formaldehyde to form colorless solids that are harder then urea resins. The are most widely encountered in dinner-ware (plastic plates, cups and serving bowls) and in plastic laminates such as Formica.



Alkyd-polyester resins



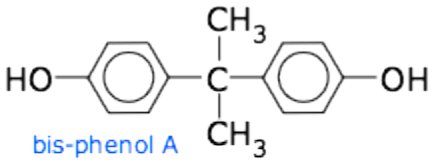
An ester is the product of the reaction of an organic acid with an alcohol, so polyesters result when multifunctional acids such as phthalic acid react with polyhydric alcohols such as glycerol. The term *alkyd* derives from the two words alcohol and acid.

Alkyd resins were first made by Berzelius in 1847, and they were first commercialized as ***Glyptal*** (glycerine + phthalic acid) varnishes for the paint industry in 1902.

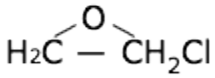
The later development of other polyesters greatly expanded their uses into a wide variety of fibers and molded products, ranging from clothing fabrics and pillow fillings to glass-reinforced plastics.

Epoxy resins

This large and industrially-important group of resins typically starts by condensing bisphenol-A with epichlorohydrin in the presence of a catalyst. (The -*epi*prefix refers to the *epoxide* group in which an oxygen atom that bridges two carbons.) These resins are usually combined with others to produce the desired properties. Epoxies are especially valued as glues and adhesives, as their setting does not depend on evaporation and the setting time can be varied over a wide range. In the *two-part resins* commonly sold for home use, the unpolymerized mixture and the hardener catalyst are packaged separately for mixing just prior to use. In some formulations the polymerization is initiated by heat ("heat curing"). Epoxy *dental fillings* are cured by irradiation with uv light.



bis-phenol A



epichlorohydrin

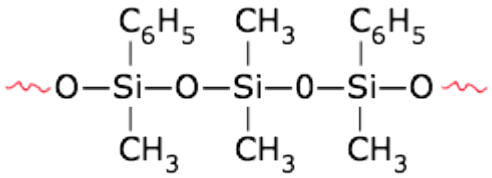
Polyurethanes

Organic isocyanates R–NCO react with multifunctional alcohols to form polymeric *carbamates*, commonly referred to as *polyurethanes*. Their major use is in plastic foams for thermal insulation and upholstery, but a very large number of other applications, including paints and varnishes and plastic wheels used in fork-lift trucks, shopping carts and skateboards.



Silicones

Polysiloxanes (–Si–O–Si-) are the most important of the small class *inorganic polymers*. The commercial silicone polymers usually contained attached organic side groups that aid to cross-linking. Silicones can be made in a wide variety of forms; those having lower molecular weights are liquids, while the more highly polymerized materials are rubbery solids. These polymers have a similarly wide variety of applications: lubricants, caulking materials and sealants, medical implants, non-stick cookware coatings, hair-conditioners and other personal-care products.



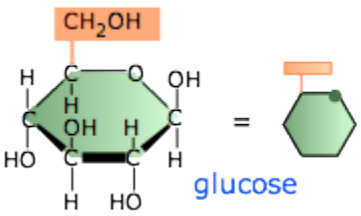
a typical silicone polymer

Natural Polymers

Polymers derived from plants have been essential components of human existence for thousands of years. In this survey we will look at only those that have major industrial uses, so we will not be discussing the very important biopolymers *proteins* and *nucleic acids*.

Polysaccharides

Polysaccharides are polymers of *sugars*; they play essential roles in energy storage, signaling, and as structural components in all living organisms. The only ones we will be concerned with here are those composed of *glucose*, the most important of the six-carbon *hexoses*. Glucose serves as the primary fuel of most organisms.

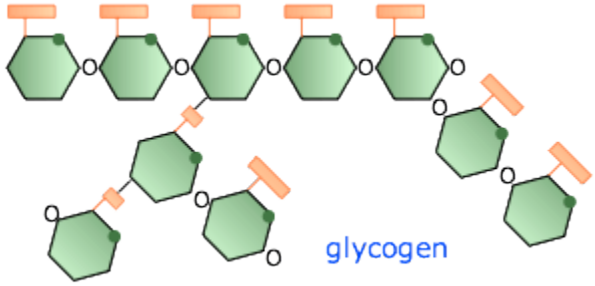


glucose

Glucose, however, is highly soluble and cannot be easily stored, so organisms make polymeric forms of glucose to set aside as *reserve storage*, from which glucose molecules can be withdrawn as needed.

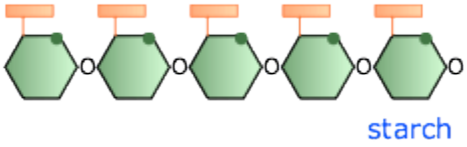
Glycogen

In humans and higher animals, the reserve storage polymer is *glycogen*. It consists of roughly 60,000 glucose units in a highly branched configuration. Glycogen is made mostly in the liver under the influence of the hormone *insulin* which triggers a process in which digested glucose is polymerized and stored mostly in that organ. A few hours after a meal, the glucose content of the blood begins to fall, and glycogen begins to be broken down in order to maintain the body's required glucose level.



Starch

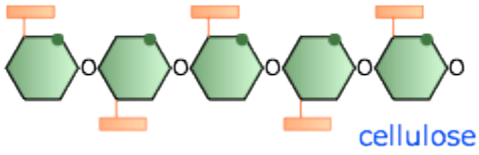
In plants, these glucose-polymer reserves are known as *starch*. Starch granules are stored in seeds or tubers to provide glucose for the energy needs of newly-germinated plants, and in the twigs of deciduous plants to tide them over during the winter when *photosynthesis* (the process in which glucose is synthesized from CO_2 and H_2O) does not take place. The starches in food grains such as rice and wheat, and in tubers such as potatoes, are a major nutritional source for humans.



Plant starches are mixtures of two principal forms, *amylose* and *amylopectin*. Amylose is a largely-unbranched polymer of 500 to 20,000 glucose molecules that curls up into a helical form that is stabilized by internal hydrogen bonding. Amylopectin is a much larger polymer having up to two million glucose residues arranged into branches of 20 to 30 units.

Cellulose and its derivatives

Cellulose is the most abundant organic compound on the earth. Extensive hydrogen bonding between the chains causes native cellulose to be about 70% crystalline. It also raises the melting point (>280°C) to above its combustion temperature. The structures of starch and cellulose appear to be very similar; in the latter, every other glucose molecule is "upside-down". But the consequences of this are far-reaching; starch can dissolve in water and can be digested by higher animals including humans, whereas cellulose is insoluble and undigestible. Cellulose serves as the principal structural component of green plants and (along with lignin) in wood.



Cotton is one of the purest forms of cellulose and has been cultivated since ancient times. Its ability to absorb water (which increases its strength) makes cotton fabrics especially useful for clothing in very warm climates.



Cotton also serves (along with treated wood pulp) as the source the industrial production of cellulose-derived materials which were the first "plastic" materials of commercial importance.

- Nitrocellulose was developed in the latter part of the 19th Century. It is prepared by treating cotton with nitric acid, which reacts with the hydroxyl groups in the cellulose chain. It was first used to make molded objects the first material used for a photographic film base by Eastman Kodak. Its extreme flammability posed considerable danger in movie theaters, and its spontaneous slow decomposition over time had seriously degraded many early films before they were transferred to more stable media. Nitrocellulose was also used as an explosive and propellant, for which applications it is known as *guncotton*.
- **Cellulose acetate** was developed in the early 1900s and became the first artificial fiber that was woven into fabrics that became prized for their lustrous appearance and wearing comfort. Kodak developed it as a "safety film" base in the 1930's to replace nitrocellulose, but it did not come into wide use for this purpose until 1948. A few years later, is became the base material for magnetic recording tape.
- **Viscose** is the general term for "regenerated" forms of cellulose made from solutions of the polymer in certain strong solvents. When extruded into a thin film it becomes cellophane which has been used as a food wrapping since 1912 and is the base for transparent adhesive tapes such as Scotch Tape. Viscose solutions extruded through a spinneret produce fibers known as rayon. Rayon (right) was the first "artificial silk" and has been used for tire cord, apparel, and carpets. It was popular for womens' stockings before Nylon became available for this purpose.



Rubber

A variety of plants produce a sap consisting of a colloidal dispersion of *cis*-polyisoprene. This milky fluid is especially abundant in the rubber tree (*Hevea*), from which it drips when the bark is wounded. After collection, the latex is coagulated to obtain the solid rubber. Natural rubber is thermoplastic, with a glass transition temperature of –70°C.





cis-polyisoprene

Raw natural rubber tends to be sticky when warm and brittle when cold, so it was little more than a novelty material when first introduced to Europe around 1770. It did not become generally useful until the mid-nineteenth century when Charles Goodyear found that heating it with sulfur — a process he called *vulcanization* — could greatly improve its properties.



Why does a rubber band heat up when it is stretched, and why does it spontaneously snap back? It all has to do with *entropy*.

Vulcanization creates disulfide cross-links that prevent the polyisoprene chains from sliding over each other. The degree of cross-linking can be controlled to produce a rubber having the desired elasticity and hardness. More recently, other kinds of chemical treatment (such as epoxidation) have been developed to produce rubbers for special purposes.

> *Better things for better living... through chemistry*" is a famous commercial slogan that captured the attitude of the public around 1940 when synthetic polymers were beginning to make a major impact in people's lives. What was not realized at the time, however, were some of the problems these materials would create as their uses multiplied and the world became more wary of "chemicals". (DuPont dropped the "through chemistry" part in 1982.)

Small-molecule release

Many kinds of polymers contain small molecules — either unreacted monomers, or substances specifically added (plasticizers, uv absorbers, flame retardants, etc.) to modify their properties. Many of these smaller molecules are able to diffuse through the material and be released into any liquid or air in contact with the plastic — and eventually into the aquatic environment. Those that are used for building materials (in mobile homes, for example) can build up in closed environments and contribute to indoor air pollution.

Residual monomer

Formation of long polymer chains is a complicated and somewhat random process that is never perfectly stoichiometric. It is therefore not uncommon for some unreacted monomer to remain in the finished product. Some of these monomers, such as formaldehyde, styrene (from polystyrene, including polystyrene foam food take-out containers), vinyl chloride, and bisphenol-A (from polycarbonates) are known carcinogens. Although there is little evidence that the small quantities that diffuse into the air or leach out into fluids pose a quantifiable health risk, people are understandably reluctant to tolerate these exposures, and public policy is gradually beginning to regulate them.

Perfluorooctanoic acid (PFOA), the monomer from which Teflon is made, has been the subject of a 2004 lawsuit against a DuPont factory that contaminated groundwater. Small amounts of PFOA have been detected in gaseous emissions from hot fluorocarbon products.

Plasticizers

These substances are compounded into certain types of plastics to render them more flexible by lowering the glass transition temperature. They accomplish this by taking up space between the polymer chains and acting as lubricants to enable the chains to more readily slip over each other. Many (but not all) are small enough to be diffusible and a potential source of health problems.



Polyvinyl chloride polymers are one of the most widely-plasticized types, and the odors often associated with flexible vinyl materials such as garden hoses, waterbeds, cheap shower curtains, raincoats and upholstery are testament to their ability to migrate into the environment.

The well-known "new car smell" is largely due to plasticizer release from upholstery and internal trim.

There is now an active movement to develop non-diffusible and "green" plasticizers that do not present these dangers.

Endocrine disrupters

Plastics-related compounds are not the only kind of endocrine disrupters found in the environment. Others include pesticide and fungicide residues, and industrial chemicals such as polychlorinated biphenols (PCBs).

To complicate matters even further, many of these small molecules have been found to be physiologically active owing to their ability to mimic the action of hormones or other signaling molecules, probably by fitting into and binding with the specialized receptor sites present in many tissues. The evidence that many of these chemicals are able to act in this way at the cellular level is fairly clear, but there is still some dispute whether many of these pose actual health risks to adult humans at the relatively low concentrations in which they commonly occur in the environment.

There is, however, some concern about the effects of these substances on non-adults and especially on fetuses, given that endocrines are intimately connected with sexual differentiation and neurological development which continues up through the late teens.

Decomposition products

Most commonly-used polymers are not readily biodegradable, particularly under the anaerobic conditions of most landfills. And what decomposition does occur will combine with rainwater to form *leachate*s that can contaminate nearby streams and groundwater supplies. Partial photodecomposition, initiated by exposure to sunlight, is a more likely long-term fate for exposed plastics, resulting in tiny broken-up fragments. Many of these materials are less dense than seawater, and once they enter the oceans through coastal sewage outfalls or from marine vessel wastes, they tend to remain there indefinitely.

Open burning of polymeric materials containing chlorine (polyvinyl chloride, for example) is known to release compounds such as dioxins that persist in the environment. Incineration under the right conditions can effectively eliminate this hazard.

Disposed products containing **fluorocarbons** (Teflon-coated ware, some personal-care, waterproofing and anti-stick materials) break down into perfluorooctane sulfonate which has been shown to damage aquatic animals.

Hazards to animals

There are two general types of hazards that polymers can introduce into the aquatic environment. One of these relates to the release of small molecules that act as hormone disrupters as described above. It is well established that small aquatic animals such as fish are being seriously affected by such substances in many rivers and estuarine systems, but details of the sources and identities of these molecules have not been identified. One confounding factor is the release of sewage water containing human birth-control drugs (which have a feminizing effect on sexual development) into many waterways.

The other hazard relates to pieces of plastic waste that aquatic animals mistake for food or become entangled in.

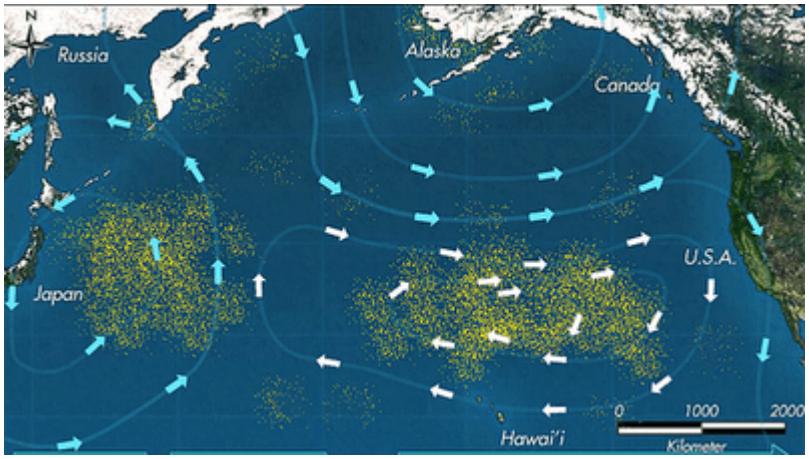


This plastic bag (probably mistaken for a jellyfish, the sea turtle's only food) cannot be regurgitated and leads to intestinal blockage and a slow death.



Remains of an albatross that mistook bits of plastic junk for food

These dangers occur throughout the ocean, but are greatly accentuated in regions known as gyres. These are regions of the ocean in which a combination of ocean currents drives permanent vortices that tend to collect and concentrate floating materials. The most notorious of these are the Great Pacific Gyres that have accumulated astounding quantities of plastic waste.





Recycling



The huge quantity (one estimate is 10^8 metric tons per year) of plastic materials produced for consumer and industrial use has created a gigantic problem of what to do with plastic waste which is difficult to incinerate safely and which, being largely non-biodegradable, threatens to overwhelm the capacity of landfills. An additional consideration is that *de novo* production most of the major polymers consumes non-renewable hydrocarbon resources.

Plastic water bottles (left) present a special recycling problem because of their widespread use in away-from-home locations.

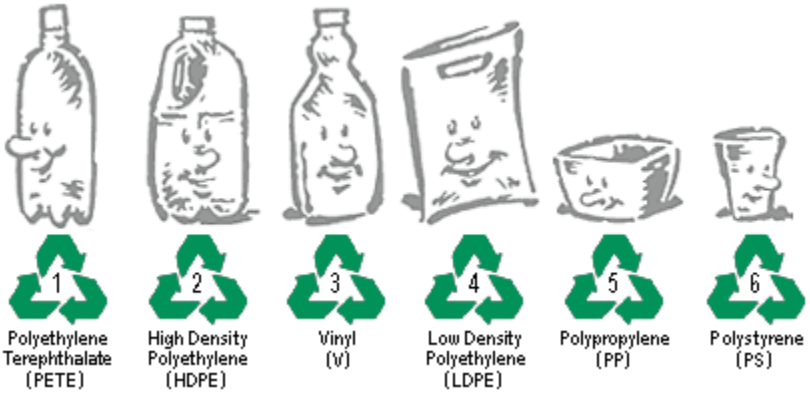
Plastics recycling has become a major industry, greatly aided by enlightened trash management policies in the major developed nations. However, it is plagued with some special problems of its own:

- Recycling is only profitable when there is a market for the regenerated material. Such markets vary with the economic cycle (they practically disappeared during the recession that commenced in 2008.)
- The energy-related costs of collecting and transporting plastic waste, and especially of processing it for re-use, are frequently the deciding factor in assessing the practicability of recycling.
- Collection of plastic wastes from diverse sources and locations and their transport to processing centers consumes energy and presents numerous operational problems.
- Most recycling processes are optimized for particular classes of polymers. The diversity of plastic types necessitates their separation into different waste streams — usually requiring manual (i.e., low-cost) labor. This in turn encourages shipment of these wastes to low-wage countries, thus reducing the availability of recycled materials in the countries in which the plastics originated.

Some of the major recycling processes include

- Thermal decomposition processes that can accommodate mixed kinds of plastics and render them into fuel oil, but the large inputs of energy they require have been a problem.
- A very small number of condensation polymers can be depolymerized so that the monomers can be recovered and re-used.
- Thermopolymers can be melted and pelletized, but those of widely differing types must be treated separately to avoid incompatability problems.
- Thermosets are usually shredded and used as filler material in recycled thermopolymers.

In order to facilitate efficient recycling, a set of seven *resin idenfication codes* has been established (the seventh, not shown below, is "other").



These codes are stamped on the bottoms of many containers of widely-distributed products. Not all categories are accepted by all local recycling authorities, so residents need to be informed about which kinds should be placed in recycling containers and which should be combined with ordinary trash.

Tire recycling

The large number of rubber tires that are disposed of, together with the increasing reluctance of landfills to accept them, has stimulated considerable innovation in the re-use of this material, especially in the construction industry.

This page titled 7.9: Polymers and Plastics is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

7.10: Colloids and their Uses

Learning Objectives

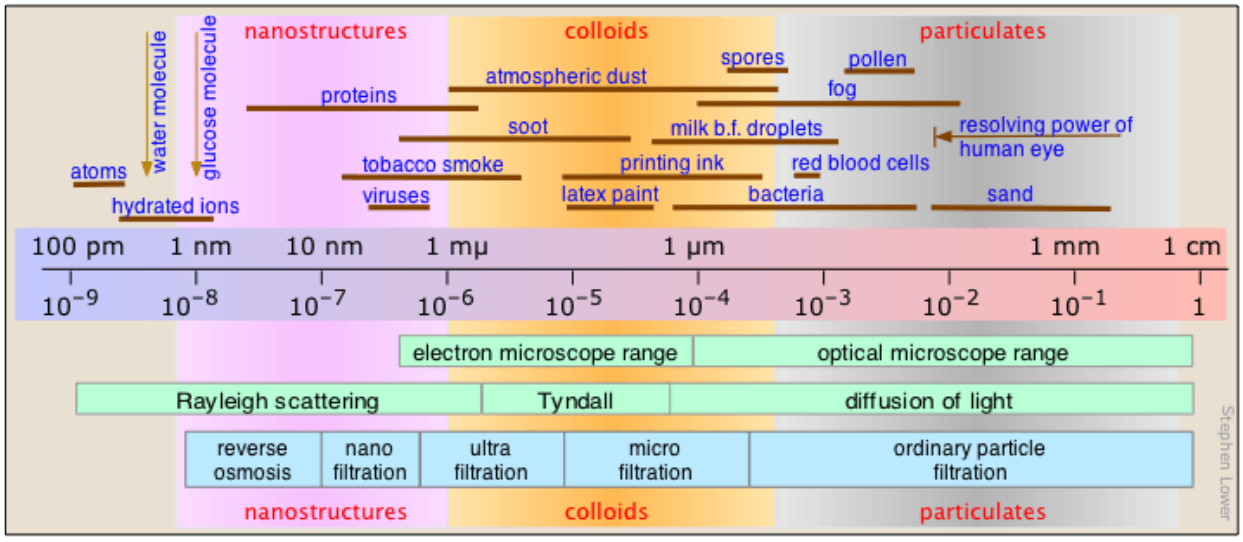
- Summarize the principal distinguishing properties of solutions, colloidal dispersions, and suspensions.
- For the various dispersion types (emulsion, gel, sol, foam, etc.), name the type (gas, liquid, or solid) of both the dispersed phase and the dispersions phase.
- Describe the origins of Brownian motion and how it can observed.
- Describe the electric double layer that surrounds many colloidal particles.
- Explain the mechanisms responsible for the stability of lyophilic and lyophobic colloidal dispersions.
- Define: surfactant, detergent, emulsifier, micelle.
- Give some examples of how colloidal dispersions can be made.
- Explain why freezing or addition of an electrolyte can result in the coagulation of an emulsion.
- Describe some of the colloid-related principles involved in food chemistry, such as the stabilization of milk and mayonaisse, the preparation of butter, and the various ways of cooking eggs.
- Describe the role of colloids in wastewater treatment.

Sand, salt, and chalk dust are made up of chunks of solid particles, each containing huge numbers of molecules. You can usually see the individual particles directly, although the smallest ones might require some magnification. At the opposite end of the size scale, we have individual molecules which dissolve in liquids to form homogeneous solutions. There is, however, a vast but largely hidden world in between: particles so tiny that they cannot be resolved by an optical microscope, or molecules so large that they begin to constitute a phase of their own when they are suspended in a liquid. This is the world of **colloids** which we will survey in this lesson. As you will see, we encounter colloids in the food we eat, the consumer products we buy... and we ourselves are built largely of colloidal materials.

Introducing Colloids

Colloids occupy an intermediate place between [particulate] suspensions and solutions, both in terms of their observable properties and particle size. In a sense, they bridge the microscopic and the macroscopic. As such, they possess some of the properties of both, which makes colloidal matter highly adaptable to specific uses and functions. Colloid science is central to biology, food science and numerous consumer products.

- Solutions are homogeneous mixtures whose component particles are individual molecules whose smallest dimension is generally less than 1 nm. Within this size range, thermal motions maintain homogeneity by overcoming the effects of gravitational attraction.
- Colloidal dispersions appear to be homogeneous, and the colloidal particles they contain are small enough (generally between 1-1000 nm) to exhibit Brownian motion, cannot be separated by filtration, and do not readily settle out. But these dispersions are inherently unstable and under certain circumstances, most colloidal dispersions can be "broken" and will "flocculate" or settle out.
- Suspensions are heterogeneous mixtures in which the suspended particles are sufficiently large (> 1000 nm in their smallest dimension) to settle out under the influence of gravity or centrifugal force. The particles that form suspensions are sometimes classified into various size ranges.



Colloidal particles need not fall within the indicated size range in all three dimensions; thus fibrous colloids such as many biopolymers may be very extended sizes along one direction.

The nature of colloidal particles

To begin, you need to recall two important definitions:

- a *phase* is defined as a region of matter in which the composition and physical properties are uniform. Thus ice and liquid water, although two forms of the single substance H_2O, constitute two separate phases within a *heterogeneous mixture*.
- A *solution* is a *homogeneous mixture* of two or more substances consisting of a single phase. (Think of sugar dissolved in water).

But imagine that you are able to shrink your view of a solution of sugar in water down to the sub-microscopic level at which individual molecules can be resolved: you would see some regions of space occupied by H_2O molecules, others by sugar molecules, and likely still others in which sugar and H_2O molecules are momentarily linked together by hydrogen bonding— not to mention the void spaces that are continually appearing and disappearing between molecules as they are jumbled about by thermal motions. As with so many simple definitions, the concept of **homogeneity** (and thus of a *solution*) breaks down as we move from the macro-scale into the molecular scale. And it is the region in between these two extremes that constitutes the realm of the colloid.

Smaller is bigger

What makes colloidal particles so special is not so much their sizes as it is the manner in which their surface areas increase as their sizes decrease. If we take a sample of matter and cut it up into smaller and smaller chunks, the total surface area will increase very rapidly. Although mass is conserved, surface area is not; as a solid is sliced up into smaller bits, more surfaces are created. These new surfaces are smaller, but there are many more of them; the ratio of surface area to mass can become extremely large.

Example 7.10.1

a. Consider a cube of material having a length of exactly 1 cm. What will be the surface area of this cube?
b. Now let us cut this cube into smaller cubes by making 10 slices in each direction. How many smaller cubes wil this make, and what will be the total surface area?

Solution

a. A cube possesses six square surfaces, so the total surface area is 6 × (1 cm^2) = 6 cm^2.
b. Each new cube has a face length of 0.10 cm, and thus a surface area of 6 × (0.1 cm^2) = 0.6 cm^2. But there are 10^3 of these smaller cubes, so the total surface area is now 60 cm^2 -- quite a bit larger than it was originally!

The total surface area increases as the inverse cube of the the face length, so as we make our slices still smaller, the total surface area grows rapidly. In practical situations with real colloids, surface areas can reach hectares (or acres) per mole!

number of slices per cube face	length of each face (cm)	surface area per face	number of cubes	total surface area
0	1	1 cm^2	1	6 cm^2
10	0.1	0.01 cm^2	1000	60 cm^2
100	0.01	10^{-4} cm^2	10^6	600 cm^2
1000	10^{-3}	10^{-6} cm^2	10^9	0.6 m^2
n	1/*n*	n^{-2} cm^2	n^3	6*n* cm^2

Why do we focus so much attention on surface area? The general answer is that surfaces (or more generally, *interfaces* between phases) possess physical and chemical properties of their own. In particular,

- Surfaces can exert *van der Waals attractive forces* on other molecules near them, and thus loosely bind other particles by *adsorption*
- Interfaces between different phases usually give rise to imbalances in electrical charge which can cause them to interact with nearby ions.
- The surfaces of many solids present "broken bonds" which are chemically active.

In normal "bulk" matter, these properties are mostly hidden from us owing to the small amount of surface area in relation to the quantity of matter. But as the particle size diminishes, surface phenomena begin to dominate their properties. The small sizes of colloidal solids allows the properties of their surfaces to dominate their behavior.

Colloidal Dispersions

Colloidal matter commonly exists in the form of colloidal-sized phases of solids, liquids, or gases that are uniformly dispersed in a separate *medium*(sometimes called the *dispersions phase*) which may itself be a solid, liquid, or gas. Colloids are often classified and given special names according to the particular kinds of phases involved.

dispersed phase	medium	dispersion type	examples
gas	liquid	*foam*	whipped cream
gas	solid	solid foam	pumice[1], aerogels[2]
liquid	gas	liquid *aerosol*	fog, clouds
liquid	liquid	*emulsion*	milk[3], mayonaisse, salad dressing
liquid	solid	*gel*	Jell-O, lubricating greases, opal[4]
solid	gas	solid aerosol	smoke
solid	liquid	*sol*	paints, some inks, blood
solid	solid	solid sol	bone, colored glass, many alloys

Notes on this table:

1. *Pumice* is a volcanic rock formed by the rapid depressurization and cooling of molten lava. The sudden release of pressure as the lava is ejected from the volcano allows dissolved gases to expand, producing tiny bubbles that get frozen into the matrix. Pumice is distinguished from other rocks by its very low density.
2. *Aerogels* are manufactured rigid solids made by removing the liquid from gels, leaving a solid, porous matrix that can have remarkable and useful physical properties. Aerogels based on silica, carbon, alumina and other substances are available.
3. Milk is basically an emulsion of butterfat droplets dispersed in an aqueous solution of carbohydrates.
4. Opal consists of droplets of liquid water dispersed in a silica (SiO_2) matrix.

Large molecules can behave as colloids

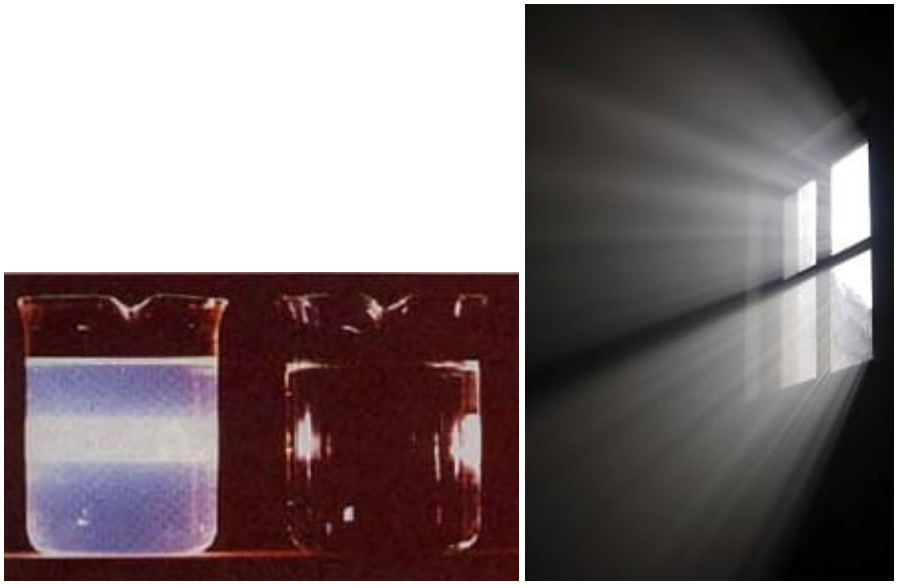
Very large polymeric molecules such as proteins, starches and other biological polymers, as well as many natural polymers, exhibit colloidal behavior. There is no clear point at which a molecule becomes sufficiently large to behave as a colloidal particle.

Colloidal dispersions behave very much like solutions in that they appear to be homogeneous on a macroscopic scale. They are often said to be *microheterogeneous*. The most important feature that distinguishes them from other particulate matter is that:

Colloids dispersed in liquids or gases are sufficiently small that they do not settle out under the influence of gravity. This, together with the their small sizes which allows them to pass through most filters, makes it difficult to separate colloidal matter from the phase in which it is dispersed.

Optical properties of colloidal dispersions

Colloidal dispersions are distinguished from true solutions by their light-scattering properties. The nature of this scattering depends on the ratio of the particle size in the medium to the wavelength of the light. A collimated beam of light passing through a solution composed of ordinary molecules (*r*) tends retain its shape. When such a beam is directed through a colloidal dispersion, it spreads out (*left container*).→



John Tyndall discovered this effect in 1869.*Tyndall scattering* (as it is commonly known) scatters all wavelengths equally. This is in contrast to *Rayleigh scattering*, which scatters shorter wavelengths more, bringing us blue skies and red sunsets. Tyndall scattering can be seen even in dispersions that are transparent. As the density of particles (or the particle size) increases, the light scattering may become great enough to produce a "cloudy" effect, as in this image of a smoke-filled room. This is the reason that milk, fog, and clouds themselves appear to be white. The individual water droplets in clouds (or the butterfat droplets in milk) are actually transparent, but the intense light scattering disperses the light in all directions, preventing us from seeing through them.

The Ultramicroscope

Colloidal particles are, like molecules, too small to be visible though an ordinary optical microscope. However, if one looks in a direction perpendicular to the light beam, a colloidal particle will "appear" over a dark background as a tiny speck due to the Tyndall scattering. A microscope specially designed for this application is known as an *ultramicroscope*. Bear in mind that the ultramicroscope (invented in Austria in 1902) does not really allow us to "see" the particle; the scattered light merely indicates where it is at any given instant.

Brownian motion

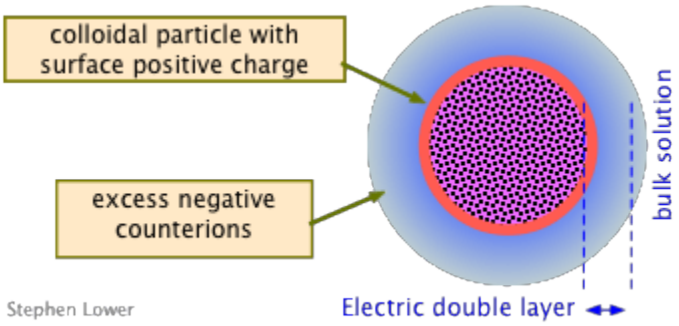
If you observe a single colloidal particle through the ultramicroscope, you will notice that it is continually jumping around in an irregular manner. These movements are known as Brownian motion. Scottish botanist Robert Brown discovered this effect in 1827 when observing pollen particles floating in water through a microscope. (Pollen particles are larger than colloids, but they are still small enough to exhibit some Brownian motion.)

It is worth noting that Albert Einstein's analysis of Brownian motion in 1901 constituted the first proof of the molecular theory of matter. Brownian motion arises from collisions of the liquid molecules with the solid particle. For large particles, the millions of collisions from different directions cancel out, so they remain stationary. The smaller the particle, the smaller the number of surrounding molecules able to collide with it, and more likely that random fluctuations will occur in the number of collisions from different sides. Simple statistics predicts that every once in a while, the imbalance in collisions from different directions will become great enough to give the particle a real kick!

Electrical Properties of Colloids

In general, differences in electric potential exist between all phase boundaries. If you have studied electrochemistry, you will know that two dissimilar metals in contact exhibit a "contact potential", and that similar potential differences exist between a metal and a solution in which it is immersed. But this principle extends well beyond ordinary electrochemistry; there are small potential differences even at the water-glass interface in a drinking glass, and the water-air interface above it.

Colloids are no exception to this rule; there is always a difference in electric potential between the colloid "phase" and that of the surrounding liquid. Even if the liquid consists of pure water, the polar H_2O molecules at the colloid's surface are likely to be predominantly oriented with either their oxygen (negative) or hydrogen (positive) ends facing the interface, depending on the electrical properties of the colloid particle itself.



Interfacial electrical potential differences can have a variety of origins:

- Particles composed of ionic or ionizable substances usually have surface charges due to adsorption of an ion (usually an anion) from the solution, or to selective loss of one kinds of ion from the crystal surface. For example, Ag+ ions on the surface of a silver iodide crystal go into solution more readily than the Br- ions, leaving a negatively-charged surface.
- The charges of amphiprotic groups such as those on the surfaces of metal oxides and hydroxides will vary with the **pH** of the aqueous medium. Thus a particle of a metal oxide M–O will become positive in acidic solution due to formation of $M–OH^+$, while that of a sparingly soluble hydroxide M–OH will become negative at low pH as it changes to $M–O^-$. Colloidal-sized protein molecule can behave in a similar manner owing to the behavior of amphiprotic carboxylate-, amino- and sulfhydryl

groups.
- Non-ionic particles or droplets such as oils or latex will tend to selectively adsorb positive or negative ions present in solution, thus "coating themselves" with electrical charge.
- In clays and other complex structures, isomorphous replacement of one ion by another having a different charge will leave a net electric charge on the particle. Thus particles of kaolinite clay become negatively charged due to replacement of some of the Si^{4+} ions by Al^{3+}.

Charged colloidal particles will attract an excess of oppositely-charged *counter-ions* to their vicinity from the bulk solution, forming a localized "cloud" of compensating charge around each particle. The entire assembly is called an *electric double layer.* Electric double layers of one kind or another exist at all phase boundaries, but those associated with colloids are especially important.

Stability of colloidal dispersions

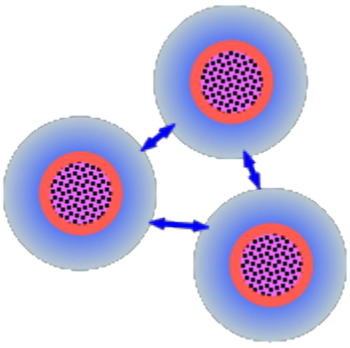
What keeps the colloidal particles suspended in the dispersion medium? How can we force the particles to settle out? These are very important practical matters:

- Colloidal products such as paints and many foods (e.g., milk) must remain in dispersed form if they are to be useful;
- Other dispersions, often those formed as by-products of operations such as mining, water treatment, paper-manufacture, or combustion are environmental nuisances. The only practical way of disposing them is to separate the colloidal material from the much greater volume of the dispersion medium (most commonly water). Simple evaporation of the water is usually not a practical option; it is generally too slow, or too expensive if forced by heating.

You will recall that weak attractive forces act between matter of all kinds. These are known generally as *van der Waals* and *dispersion* forces, and they only "take hold" at very close distances. Countering these is the *universal repulsive force* that acts at even shorter distances, but is far stronger; it is the basic reason why two atoms cannot occupy the same space. For very small atomic and molecular sized particles, another thing that keeps them apart is thermal motion. Thus when two molecules in a gas collide, they do so with more than enough kinetic energy to overcome the weak attractive forces between them. As the temperature of the gas is reduced, so is the collisional energy; below its boiling point, the attractive forces dominate and the gas condenses into a liquid.

Electrical forces help keep colloids dispersed

When particles of colloidal dimension suspended in a liquid collide with each other, they do so with much smaller kinetic energies than is the case for gases, so in the absence of any compensating repulsion forces, we might expect van der Waals or dispersion attractions to win out. This would quickly result in the growth of aggregates sufficiently large to exceed colloidal size and to fall to the bottom of the container. This process is called *coagulation*.



So how do stable dispersions such as sols manage to survive? In the preceding section, we saw that each particle with its double layer is more or less electrically neutral. However, when two particles approach each other, each one "sees" mainly the outer part [shown here in blue] of the double layer of the other. These will always have the same charge sign (which depends on the type of colloid and the nature of the medium), so there will be an electrostatic repulsive force that opposes the dispersion force attractions.

Electrostatic (coulombic) forces have a strong advantage in this respect because they act over much greater distances do van

der Waals forces. But as we will see further on, electrostatic repulsion can lose its effectiveness if the ionic concentration of the medium is too great, or if the medium freezes. Under these conditions, there are other mechanisms that can stabilize colloidal dispersions.

Interactions with the solvent

Colloids can be divided into two general classes according to how the particles interact with the dispersions medium (often referred to as the "solvent").

Lyophilic colloids

In one class of colloids, called *lyophilic* ("solvent loving") colloids, the particles contain chemical groups that interact strongly with the solvent, creating a sheath of solvent molecules that physically prevent the particles from coming together. Ordinary *gelatine* is a common example of a lyophilic colloid. It is in fact *hydrophilic*, since it forms strong hydrogen bonds with water. When you mix Jell-O or tapioca powder to make a gelatine dessert, the material takes up water and forms a stable colloidal gel. Lyophilic (hydrophilic) colloids are very common in biological systems and in foods.

Lyophobic colloids

Most of the colloids in manufactured products exhibit very little attraction to water: think of oil emulsions or glacially-produced rock dust in river water. These colloids are said to be *lyophobic*. Lyophobic colloids are all inherently unstable; they will eventually *coagulate*. However, "eventually" can be a very long time (the settling time for some clay colloids in the ocean is 200-600 years!).

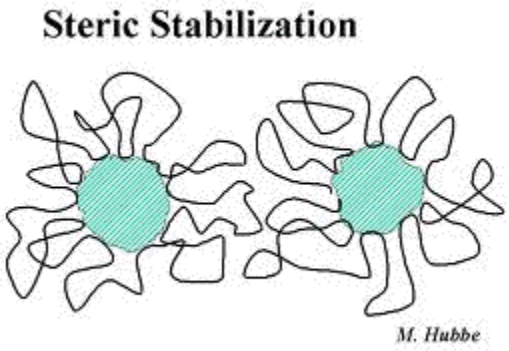
For systems in which coagulation proceeds too rapidly, the process can be slowed down by adding a stabilizer. Stabilizers can act by coating the particles with a protective layer such as a polymer as described immediately below, or by providing an ion that is selectively adsorbed by the particle, thereby surrounding it with a charged sheath that will repel similar particles it collides with. Dispersions of these colloids are stabilized by electrostatic repulsion between the electric double layers surrounding the particles which we discussed in the preceding section.

Stabilization by cloaking

"Stabilization by stealth" has unwittingly been employed since ancient times through the use of natural gums to stabilize pigment particles in inks, paints, and pottery glazes. These gums are also widely used to stabilize foods and personal care products. A lyophobic colloid can be made to masquerade as lyophilic by coating it with something that itself possesses suitable lyophilic properties.

Steric stabilization

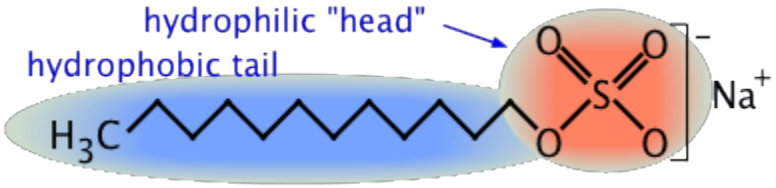
Alternatively, attaching a lyophobic material to a colloid of any type can surround the particles with a protective shield that physically prevents the particles from approaching close enough to join together. This method usually employs synthetic polymers and is often referred to as *steric stabilization*.



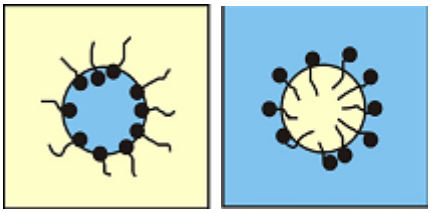
Synthetic polymers, which can be tailor-made for specific applications, are now widely employed for both purposes. The polymer can be attached to the central particle either by simple adsorption or by chemical bond formation.

Surfactants and micelle formation

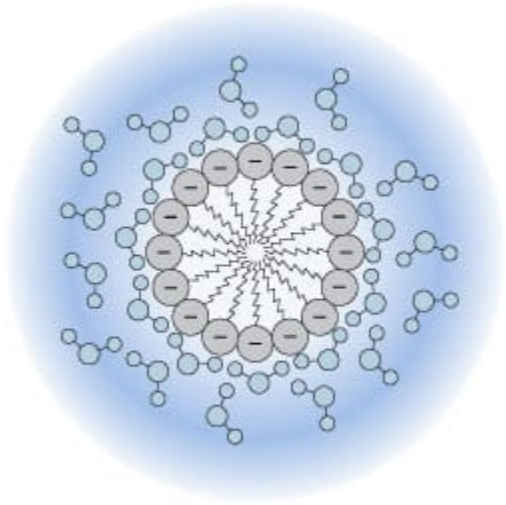
Surfactants and *detergents* are basically the same thing. Surfactants that serve as cleaning agents are commonly called *detergents* (from L. *detergere* "to wipe away, cleanse"). Surfactants are molecules consisting of a hydrophylic "head" connected to a hydrophobic chain. Because such molecules can interact with both "oil" and water phases, they are often said to be *amphiphilic*. Typical of these is the well known cleaning detergent *sodium dodecyl sulfonate* ("sodium laurel sulfate") $CH_3(CH_2)_{11}OSO_3^-\ Na^+$.



Amphiphiles possess the very important property of being able to span an oil-water interface. By doing so, they can stabilize emulsions of both the water-in-oil and oil-in-water types. Such molecules are essential components of the *lipid bilayers* that surround the cells and cellular organelles of living organisms.



Emulsions are inherently unstable; left alone, they tend to separate into "oil" and "water" phases. Think of a simple salad dressing made by shaking vegetable oil and vinegar. When a detergent-like molecule is employed to stabilize an emulsion, it is often referred to as an *emulsifier*. The resulting structure (left) is known as a *micelle*.

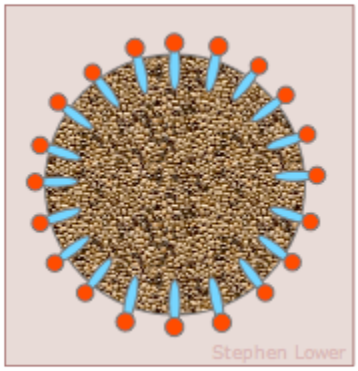


Emulsifiers are essential components of many foods. They are widely employed in pharmaceuticals, consumer goods such as lotions and other personal care products, paints and printing inks, and numerous industrial processes.

How detergents remove "dirt"

The "dirt" we are trying to remove consists of oily or greasy materials whose hydrophobic nature makes them resistant to the

action of pure water. If the water contains amphiphilic molecules such as soaps or cleaning detergents that can embed their hydrophobic ends in the particles, the latter will present a hydrophilic interface to the water and will thus become "solubilized".



Soaps and detergents can also disrupt the cell membranes of many types of bacteria, for which they serve as *disinfectants*. However, they are generally ineffective against viruses, which do not possess cell membranes.

Bile: your body's own detergent

Oils and fats are important components of our diets, but being insoluble in water, they are unable to mix intimately with the aqueous fluid in the digestive tract in which the digestive enzymes are dissolved. In order to enable the lipase enzymes (produced by the pancreas) to break down these lipids into their component fatty acids, our livers produce a mixture of surfactants known as bile. The great surface area of the micelles in the resulting emulsion enables efficient contact between the lipase enzymes and the lipid materials.

The liver of the average adult produces about 500 mL of bile per day. Most of this is stored in the gall bladder, where it is concentrated five-fold by removal of water. As partially-digested material exits the stomach, the gall bladder squeezes bile into the top of the small intestine (the *duodenum*).

In addition to its action as a detergent (which also aids in the destruction of bacteria that may have survived the high acidity of the gastric fluid), the alkaline nature of the bile salts neutralizes the acidity of the stomach exudate. The bile itself consists of of salts of a variety of bile acids, all of which are derived from cholesterol. The cholesterol-like part of the structure is hydrophobic, while the charged end of the salt is hydrophilic.

Microemulsions

Ordinary emulsions are inherently unstable; they do not form spontaneously, and once formed, the drop sizes are sufficiently large to scatter light, producing a milky appearance. As time passes, the average drop size tends to increase, eventually resulting in gravitational separation of the phases.



Microemulsions, in contrast, are thermodynamically stable and can form spontaneously. The drop radii are at the very low end of the colloidal scale, often 100 nm or smaller. This is too small to appreciably scatter visible light, so microemulsions appear visually to be homogenous systems.

Microemulsions require the presence of one or more surfactants which increase the flexibility and stability of the boundary regions. This allows them to vary form smaller micelles than surface tension forces would ordinarily allow; in some cases they can form sponge-like bicontinuous mixtures in which "oil" and "water" phases extend throughout the mixture, affording more contact area between the phases.

The uses of microemulsions are quite wide-ranging, with drug delivery, polymer synthesis, enzyme-assisted synthesis, coatings, and enhanced oil recovery being especially prominent.

4 Making and breaking colloidal dispersions

Particles of colloidal size can be made in two general ways:

- Start with larger particles and break them down into smaller ones (***Dispersion***).
- Build up molecular-sized particles (atoms, ions, or small molecules) into aggregates within the colloidal size range. (***Condensation***)

Dispersion processes all require an input of energy as new surfaces are created. For **solid particles**, this is usually accomplished by some kind of grinding process such as in a ball- or roller-mill. Solids and liquids can also be broken into colloidal dimensions by injecting them into the narrow space between a rapidly revolving shaft and its enclosure, thus subjecting them to a strong *shearing force* that tends to pull the two sides of a particle in opposite directions.

The application of ultrasound (at about 20 kHz) to a mixture of two immiscible **liquids** can create liquid-in-liquid dispersions; the process is comparable to what we do when we shake a vinegar-and-oil salad dressing in order to create a more uniform distribution of the two liquids.

Condensation

Numerous methods exist for building colloidal particles from sub-colloidal entities.

$$S_2O_3^{2-} + H_2O \rightarrow S + SO_4^{2-} + 2H^+ + 2e^- \tag{7.10.1}$$

$$2Fe^{3+} + 3H_2O \rightarrow Fe_2O_3 + 6H^+ \tag{7.10.2}$$

$$Fe^{3+} + 2H_2O \rightarrow FeO(OH) + 3H^+ \tag{7.10.3}$$

Dissolution followed by precipitation

This method is useful for dispersing hydrophobic organic substances in water. For example, a sample of paraffin wax is dissolved in ethanol, and the resulting solution is carefully added to a container of boiling water.

Formation of precipitates under controlled conditions

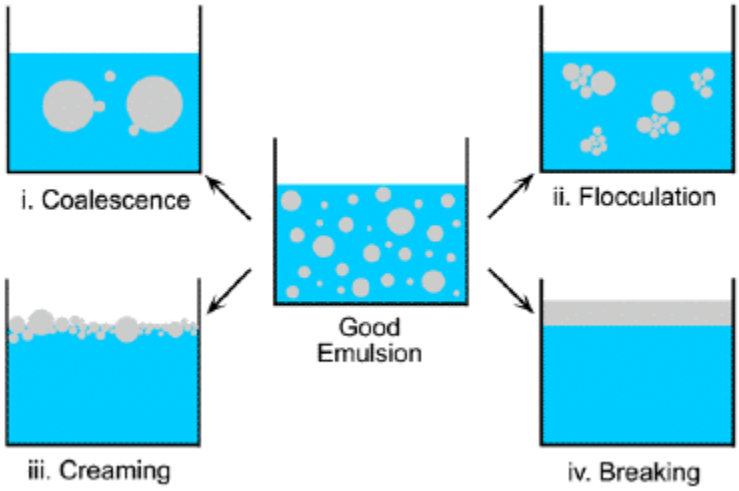
The trick here is to prevent the initial colloidal particles of the newly-formed compound from coalescing into an ordinary precipitate, as will ordinarily occur when solutions of two dissolved salts are combined directly. An alternative that is sometimes useful is to form the sol by a chemical process that procedes more slowly than direct precipitation:

- Sulfur sols are readily formed by oxidation of thiosulfate ions in acidic solution:
- Sols of oxides or hydrous oxides of transition metals can often be formed by boiling a soluble salt in water under slightly acidic conditions to prevent formation of insoluble hydroxides:
- Addition of a dispersant (usually a surfactant) can sometimes prevent colloidal particles from precipitating. Thus barium sulfate sols can be prepared from barium thiocyanate and $(NH_4)_2SO_4$ in the presence of potassium citrate.
- Ionic solids can often selectively adsorb cations or anions from solutions containing the same kinds of ions that are present in the crystal lattice, thus coating the particles with protective electric charges. This is probably what happens in the example of Fe^{3+}-ion hydrolysis given above.
- Similarly, if a solution of $AgNO_3$ is added to a dilute solution of potassium iodide, the AgI will form as a negatively-charged sol $(AgI)\cdot I^-$. But if the AgI is precipitated by adding KI to a solution of $AgNO_3$, the excess Ag^+ will adsorb to the new particles, giving a positively-charged sol of $(AgI)\cdot Ag^+$.

How Dispersions are Broken

That oil-in-vinegar salad dressing you served at dinner the other day has now mostly separated into two layers, with unsightly globs of one phase floating in the other. This is surface chemistry in action! Emulsions are fundamentally unstable because molecules near surfaces (i.e., interfaces between phases) are no longer surrounded by their own kind on all sides. The resulting repulsions between like and unlike exact an energetic cost that must eventually be repaid through processes that reduce the interfacial area.

The consequent breakup of the emulsion can proceed through various stages:



- *Coalescence* - smaller drops join together to form larger ones;
- *Flocculation* - the small drops stick together without fully coalescing;
- *Creamin*g - Most oils have lower densities than water, so the drops float to the surface, but may not completely coalesce;
- *Breaking* - the ultimate thermodynamic fate and end result of the preceding steps.

The time required for these processes to take place is highly variable, and can be extended by the presence of *stabilizer* substances. Thus *milk*, an emulsion of butterfat in water, is stabilized by some of its natural components.

Coagulation and flocculation

The processes described above that allow colloids to remain suspended sometimes fail when conditions change, or equally troublesome, they work entirely too well and make it impossible to separate the colloidal particles from the medium; this is an especially serious problem in wastewater settling basins associated with sewage treatment and operations such as mining and the pulp-and-paper industries.

Coagulation is the general term that refers to the "breaking" of dispersions so that the colloidal particles can be collected, usually by settling out. The term *Flocculation* is often used as a synonym for coagulation, but it is more properly reserved for a special method of effecting coagulation which is described further on. Most coagulation processes act by disrupting the outer (diffuse) part of the electric double layer that gives rise to the electrostatic repulsion between them.

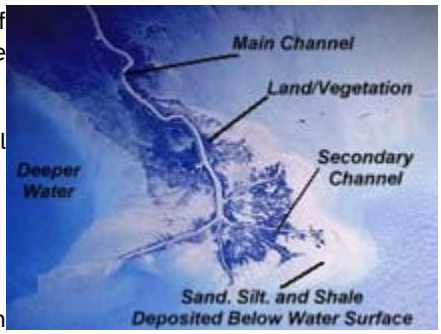
"Do not freeze"

Have you ever encountered milk that had previously been frozen? Not likely something you would want to drink! You will see "Do not freeze" labels on many foodstuffs and on colloidal consumer products such as latex house paint. Freezing disrupts the double layer by causing the ions within it to combine into neutral species so that the particles can now approach closely enough for attractive forces to take over, and once they do so, they never let go: coagulation is definitely an irreversible process!

Addition of an electrolyte

Coagulation of water-suspended dispersions can be brought about by raising the ionic concentration of the medium. The added ions will migrate to the oppositely-charged regions of the double layer, thus neutralizing its charges; this effectively reduces the thickness of the double layer, eventually allowing the attractive forces to prevail.

Rivers carry millions of tons of colloidal clay into the oceans. If you fly over the mouth of a river such as the Mississippi (shown here in a satellite image), you can sometimes see the difference in color as the clay colloids coagulate due to the action of the salt water.



The coagulated clay accumulates as sediments which eventually form a geographical feature called a river delta.

5 Gels

A liquid phase dispersed in a solid medium is known as a *gel*, but this formal definition does not always convey the full sense of the nature of the "solid". The latter may start out as a powdery or granulated material such as natural gelatin or a hydrophilic polymer, but once the gel has formed, the "solid" part is less a "phase" than a cross-linked network that extends throughout the volume of the liquid, whose quantity largely defines the volume of the entire gel.

Hydrogels can contain up to 90% water by weight

Most of the gels we commonly encounter have water as the liquid phase, and thus are called *hydrogels*; ordinary gelatin deserts are well known examples.

The "solid" components of hydrogels are usually polymeric materials that have an abundance of hydrophilic groups such as hydroxyl (–OH) that readily hydrogen-bond to water and also to each other, creating an irregular, flexible, and greatly-extendable network. These polymers are sometimes synthesized for this purpose, but are more commonly formed by processing natural materials, including natural polymers such as cellulose.

- Gelatine is a protein-like material made by breaking down the connective tissues of animal skins, organs, and bones. The many polar groups on the resulting protein fragments bind them together, along with water molecules, to form a gel.
- A number of so-called super-absorbant polymers derived from cellulose, polyvinyl alcohol and other materials can absorb huge quantities of water, and have found uses for products such as disposable diapers, environmental spill control, water retention media for plants, surgical pads and wound dressings, and protective inner coatings and water-blockers in fiber optics and electrical cables.

Gels are essential components of a huge variety of consumer products ranging from thickening agents in foods and personal care products to cushioning agents in running shoes.

Gels can be fragile!

You may have noticed that a newly-opened container of yogurt or sour cream appears to be smooth and firm, but once some of the material has been spooned out, little puddles of liquid appear in the hollowed-out depressions.

As the spoon is plunged into the material, it pulls the nearby layers of the gel along with it, creating a shearing action that breaks it apart, releasing the liquid. Anyone who has attacked an egg yolk with a cook's whisk, written with a ball-point pen, or spread latex paint on a wall has made use of this phenomenon which is known as *shear thinning*.

Our bodies are mostly gels



The interior (the cytoplasm) of each cell in the soft tissues of our bodies consists of a variety of inclusions (organelles) suspended in a gel-like liquid phase called the cytosol. Dissolved in the cytosol are a variety of ions and molecules varying from the small to the large; among the latter, proteins and carbohydrates make up the "solid" portion of the gel structure.

Embedded within the cytosol is the filament-like *cytoskeleton* which controls the overall shape of the cell and holds the organelles in place.

(In free-living cells such as the amoeba, changes in the cytoskeleton enable the organism to alter its shape and move around to engulf food particles.)

Be thankful for the gels in your body; without them, you would be little more than a bag of gunge-filled liquid, likely to end up as a puddle on the floor!

The individual cells are bound into tissues by the extracellular matrix (ECM) which — on a much larger scale, holds us together and confers an overall structure and shape to the body. The ECM is made of a variety of structural fibers (collagens, elastins) embedded in a gel-like matrix.

6 Applications of colloids

Thickening agents

The usefulness of many industrial and consumer products is strongly dependent on their viscosity and flow properties. Toothpastes, lotions, lubricants, coatings are common examples. Most of the additives that confer desirable flow properties on these products are colloidal in nature; in many cases, they also provide stabilization and prevent phase separation. Since ancient times, various natural gums have been employed for such purposes, and many remain in use today.

More recently, manufactured materials whose properties can be tailored for specific applications have become widely available. Examples are colloidal microcrystalline cellulose, carboxymethyl cellulose, and fumed silica.

Fumed silica is a fine (5-50 nm), powdery form of SiO_2 of exceptionally low bulk density (as little as 0.002 g cm^{-3}); the total surface area of one Kg can be as great as 60 hectares (148 acres). It is made by spraying $SiCl_4$(a liquid) into a flame. It is used as a filler, for viscosity and flow control, a gelling agent, and as an additive for strenghthening concrete.

Food colloids

Most of the foods we eat are largely colloidal in nature. The function of food colloids generally has less to do with nutritional value than appearance, texture, and "mouth feel". The latter two terms relate to the flow properties of the material, such as spreadability and the ability to "melt" (transform from gel to liquid emulsion) on contact with the warmth of the mouth.

Dairy products

Milk is basically an emulsion of lipid oils ("butterfat") dispersed in water and stabilized by phospholipids and proteins. Most of the protein content of milk consists of a group known as *caseins* which aggregate into a complex micellar structure which is bound together by calcium phosphate units.



Homogenizer



The stabilizers present in fresh milk will maintain its uniformity for 12-24 hours, but after this time the butterfat globules begin to coalesce and float to the top ("creaming"). In order to retard this process, most milk sold after the early 1940's undergoes *homogenization* in which the oil particles are forced through a narrow space under high pressure. This breaks up the oil droplets into much smaller ones which remain suspended for the useful shelf life of the milk.

Before homogenization become common, milk bottles commonly had enlarged tops ↑ to make it easier to skim off the cream that would separate out.

The structures of cream, yogurt and ice cream are dominated by the casein aggregates mentioned above.

Ice cream is a complex mixture of several colloid types:

- an *emulsion* (of butterfat globules in a highly viscous aquatic phase);
- a semisolid *foam* consisting of small (100 μ) air bubbles which are beat into the mixture as it is frozen. Without these bubbles, the frozen mixture would be too hard to conveniently eat;
- a *gel* in which a network of tiny (50 μ) ice crystals are dispersed in a semi-glassy aqueous phase containing sugars and dissolved macromolecules.



Whereas milk is an oil (butterfat)-in-water dispersion, **butter** and **margarine** have a "reversed" (water-in-oil) arrangement. This transformation is accomplished by subjecting the butterfat droplets in cream to violent agitation (*churning*) which forces the droplets to coalesce into a semisolid mass within which remnants of the water phase are embedded. The greater part of this phase ends up as the by-product *buttermilk*.

Eggs: colloids for breakfast, lunch, and dessert



A detailed study of eggs and their many roles in cooking can amount to a mini-course in colloid chemistry in itself. There is something almost magical in the way that the clear, viscous "white" of the egg can be transformed into a white, opaque semi-

solid by brief heating, or rendered into more intricate forms by poaching, frying, scrambling, or baking into custards, soufflés, and meringues, not to mention tasty omelettes, quiches, and more exotic delights such as the *eggah* (Arabic) and *kuku* (Persian) dishes of the Middle-East.

The raw eggwhite is basically a colloidal sol of long-chain protein molecules, all curled up into compact folded forms due to hydrogen bonding between different parts of the same molecule. Upon heating, these bonds are broken, allowing the proteins to unfold. The denuded chains can now tangle and bind to each other, transforming the sol into a cross-linked hydrogel, now so dense that scattered light changes its appearance to opaque white.

What happens next depends very much on the skill of the cook. The idea is to drive out enough of the water entrapped within the gel network to achieve the desired density while retaining enough gel structure to prevent it from forming a rubbery mass, as usually happens with hard-boiled eggs. This is especially important when the egg structure is to be incorporated into other food components as in baked dishes.

The key to all this is temperature control; the eggwhite proteins begin to coagulate at 65°C and if yolk proteins are present, the mixture is nicely set at about 73°; by 80° the principal (albumin) protein has set, and at much above this the gel network will collapse into an overcooked mass. The temperature limit required to avoid this disaster can be raised by adding milk or sugar; the water part of the milk dilutes the proteins, while sugar molecules hydrogen-bond to them, forming a protective shield that keeps the proton strand separated. This is essential when baking custards, but incorporating a bit of cream into scrambled eggs can similarly help them retain their softness.

Whipped cream and meringues



The other colloidal personalities eggs can display are liquid and solid foams. Instead of applying heat to unfold the proteins, we "beat" them; the shearing force of a whisk or egg-beater helps pull them apart, and the air bubbles that get entrapped in the mixture attract the hydrophobic parts of the unfolded proteins and help hold them in place. Sugar will stabilize the foam by raising its viscosity, but will interfere with protein folding if added before the foam is fully formed. Sugar also binds the residual water during cooking, retarding its evaporation until after the proteins not broken up by beating can be thermally coagulated.

Paints and inks

Paints have been used since ancient times for both protective and decorative purposes. They consist basically of pigment particles dispersed in vehicle — a liquid capable for forming a stable solid film as the paint "dries".

The earliest protective coatings were made by dissolving plant-derived natural polymers (resins) in an oil such as that of linseed. The double-bonds in these oils tends to oxidize when exposed to air, causing it to polymerize into an impervious film. The colloidal pigments were stabilized with naturally-occurring surfactants such as polysaccharide gums.

Present-day paints are highly-engineered products specialized for particular industrial or architectural coatings and for marine or domestic use. For environmental reasons, water-based ("latex") vehicles are now preferred.

Inks

The most critical properties of inks relate to their drying and surface properties; they must be able to flow properly and attach to the surface without penetrating it — the latter is especially critical when printing on a porous material such as paper.

Many inks consist of organic dyes dissolved in a water-based solvent, and are not colloidal at all. The ink used in printing newspapers employs colloidal carbon black dispersed in an oil vehicle. The pressure applied by the printing press forces the vehicle into the pores of the paper, leaving most of the pigment particles on the surface.



The inks employed in **ball-point pens** are gels, formulated in such a way that the ink will only flow over the ball and onto the paper when the shearing action of the ball (which rotates as it moves across the paper) "breaks" the gel into a liquid; the resulting liquid coats the ball and is transferred to the paper. As in conventional printing, the pigment particles remain on the paper surface, while the liquid is pressed into the pores and gradually evaporates.

Water and wastewater treatment



Turbidities of 5, 50, and 500 units. [WikiMedia]

Water, whether intended specifically for drinking, or wastewaters such as sewage or from industrial operations such as from pulp-and-paper manufacture (most of which are likely to end up being re-used elsewhere) usually contains colloidal matter that cannot be removed by ordinary sand filters, as evidenced by its turbidity. Even "pristine" surface waters often contain suspended soil sediments that can harbor infectious organisms and may provide them with partial protection from standard disinfection treatments.

The sulfates of aluminum (alum) and of iron(III) have long been widely employed for this purpose. Synthetic polymers tailored specifically for these applications have more recently come into use.

The usual method of removing turbidity is to add a flocculating agent (flocculant). These are most often metallic salts that can form gel-like hydroxide precipitates, often with the aid of added calcium hydroxide (quicklime) if pH of the water must be raised.



The flocculant salts neutralize the surface charges of the colloids, thus enabling them to coagulate; these are engulfed and trapped by fragments of gelatinous precipitate, which are drawn together into larger aggregates by gentle agitation until they become sufficiently large to form flocs which can be separated by settling or filtration.

Soil colloids

The four major components of soils are mineral sediments, organic matter, water, and air. The water is primarily adsorbed to the mineral and organic materials, but may also share pore spaces with air; pore spaces constitute about half the bulk volume of typical solid.

The principal colloidal components of soils are mineral sediments in the form of clays, and the humic materials in the organic matter. In addition to influencing the consistency of soil by binding water molecules, soil colloids play an essential role in storing and exchanging the mineral ions required by plants.

Most soil colloids are negatively charged, and therefor attract cations such as Ca^{2+}, Mg^{2+}, and K^{+} into the outer parts of their double layers. Because these ions are loosely bound, they constitute a source from which plant roots can draw these essential nutrients. Conversely, they can serve as a sink for these same ions when they are released after the plant dies.

Clays

These are layered structures based on alumino-silicates or hydrous oxides, mostly of iron or aluminum. Each layer is built of two or three sheets of extended silica or alumina structures linked together by shared oxygen atoms. These layers generally have an overall negative charge owing to the occasional replacement of a Si^{4+} ion by one of Al^{3+}.

Adjacent layers are separated by a region of adsorbed cations (to neutralize the negative charges) and water molecules, and thus are held together relatively loosely. It is these interlayer regions that enable clays to work their magic by exchanging ions with both the soil water and the roots of plants.

Humic substances

The principal organic components of soil are complex substances of indeterminate structure that present –OH and –COOH groups

which become increasingly dissociated as the pH increases. This allows them to bind and exchange cations in much the same way as described above.

This page titled 7.10: Colloids and their Uses is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8: SOLUTIONS

Solutions are *homogeneous* (single-phase) mixtures of two or more *components*. For convenience, we often refer to the majority component as the *solvent*; minority components are *solutes*; there is really no fundamental distinction between them. Solutions play a very important role in Chemistry because they allow intimate and varied encounters between molecules of different kinds, a condition that is essential for rapid chemical reactions to occur.

8.1: Solutions and their Concentrations

8.2: Thermodynamics of Solutions

8.2.2A: Solutions of Gaseous Solutes in Gaseous Solvents

8.2.2B: Solutions of Gaseous Solutes in Liquid Solvents

8.2.2C: Solutions of Liquid Solutes in Liquid Solvents

8.2.2D: Solutions of Solid Solutes in Liquid Solvents

8.3: Colligative Properties- Raoult's Law

8.4: Colligative Properties- Boiling Point Elevation and Freezing Point Depression

8.5: Colligative Properties - Osmotic Pressure

8.6: Reverse Osmosis

8.7: Colligative Properties and Entropy

8.8: Ideal vs. Real Solutions

8.9: Distillation

8.10: Ions and Electrolytes

8.10.9A: Electrolytes and Electrolytic Solutions

8.10.9B: The nature of ions in aqueous solution

8.10.9C: Weak and Strong Electrolytes

8.10.9D: Ionic migration

8.10.9E: Some applications of electrolytic conduction

This page titled 8: Solutions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Solutions and their Concentrations

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Describe the major reasons that solutions are so important in the *practical* aspects of chemistry.
- Explain why expressing a concentration as "*x*-percent" can be ambiguous.
- Explain why the molarity of a solution will vary with its temperature, whereas molality and mole fraction do not.
- Given the necessary data, convert (in either direction) between any two concentration units, e.g. molarity - mole fraction.
- Show how one can prepare a given volume of a solution of a certain molarity, molality, or percent concentration from a solution that is more concentrated (expressed in the same units.)
- Calculate the concentration of a solution prepared by mixing given volumes to two solutions whose concentrations are expressed in the same units.

Solutions are *homogeneous* (single-phase) mixtures of two or more *components*. For convenience, we often refer to the majority component as the *solvent*; minority components are *solutes*; there is really no fundamental distinction between them. Solutions play a very important role in Chemistry because they allow intimate and varied encounters between molecules of different kinds, a condition that is essential for rapid chemical reactions to occur. Several more explicit reasons can be cited for devoting a significant amount of effort to the subject of solutions:

- For the reason stated above, most chemical reactions that are carried out in the laboratory and in industry, and that occur in living organisms, take place in solution.
- Solutions are so common; very few pure substances are found in nature.
- Solutions provide a convenient and accurate means of introducing known small amounts of a substance to a reaction system. Advantage is taken of this in the process of titration, for example.
- The physical properties of solutions are sensitively influenced by the balance between the intermolecular forces of like and unlike (solvent and solute) molecules. The physical properties of solutions thus serve as useful experimental probes of these intermolecular forces.

We usually think of a solution as a liquid made by adding a gas, a solid or another liquid *solute* in a liquid *solvent*. Actually, solutions can exist as gases and solids as well.

Solid solutions are very common; most natural minerals and many metallic alloys are solid solutions.

Still, it is liquid solutions that we most frequently encounter and must deal with. Experience has taught us that sugar and salt dissolve readily in water, but that “oil and water don’t mix”. Actually, this is not strictly correct, since all substances have at least a slight tendency to dissolve in each other. This raises two important and related questions: why do solutions tend to form in the first place, and what factors limit their mutual solubilities?

Understanding Concentrations

Concentration is a general term that expresses the quantity of solute contained in a given amount of solution. Various ways of expressing concentration are in use; the choice is usually a matter of convenience in a particular application. You should become familiar with all of them.

Parts-per concentration

In the consumer and industrial world, the most common method of expressing the concentration is based on the quantity of solute in a fixed quantity of solution. The “quantities” referred to here can be expressed in weight, in volume, or both (i.e., the *weight* of solute in a given *volume* of solution.) In order to distinguish among these possibilities, the abbreviations (w/w), (v/v) and (w/v) are used.

In most applied fields of Chemistry, (w/w) measure is often used, and is commonly expressed as weight-percent concentration, or simply "percent concentration". For example, a solution made by dissolving 10 g of salt with 200 g of water contains "1 part of salt per 20 g of water".

"*Cent*" is the Latin-derived prefix relating to the number 100 (L. *centum*), as in *century* or *centennial*. It also denotes 1/100th (from L. *centesimus*) as in *centimeter* and the monetary unit *cent*. It is usually more convenient to express such concentrations as "parts per 100", which we all know as "percent". So the solution described above is a "5% (w/w) solution" of NaCl in water. In clinical chemistry, (w/v) is commonly used, with weight expressed in grams and volume in mL (Example 8.1.1).

Example 8.1.1

The normal saline solution used in medicine for nasal irrigation, wound cleaning and intravenous drips is a 0.91% (w/v) solution of sodium chloride in water. How would you prepare 1.5 L of this solution?

Solution

The solution will contain 0.91 g of NaCl in 100 mL of water, or 9.1 g in 1 L. Thus you will add (1.5 × 9.1g) = 13.6 g of NaCl to 1.5 L of water.

Percent means parts per 100; we can also use parts per thousand (ppt) for expressing concentrations in grams of solute per kilogram of solution. For more dilute solutions, parts per million (ppm) and parts per billion (10^9; ppb) are used. These terms are widely employed to express the amounts of trace pollutants in the environment.

Example 8.1.2

Describe how you would prepare 30 g of a 20 percent (w/w) solution of KCl in water.

Solution

The weight of potassium chloride required is 20% of the total weight of the solution, or 0.2 × (30 g) = 6.0 g of KCl. The remainder of the solution (30 – 6 = 24) g consists of water. Thus you would dissolve 6.0 g of KCl in 24 g of water.

Weight/volume and volume/volume basis

It is sometimes convenient to base concentration on a fixed volume, either of the solution itself, or of the solvent alone. In most instances, a 5% by volume solution of a solid will mean 5 g of the solute dissolved in 100 ml of the solvent.

Example 8.1.3

Fish, like all animals, need a supply of oxygen, which they obtain from oxygen dissolved in the water. The minimum oxygen concentration needed to support most fish is around 5 ppm (w/v). How many moles of O_2 per liter of water does this correspond to?

Solution

5 ppm (w/v) means 5 grams of oxygen in one million mL (1000 L) of water, or 5 mg per liter. This is equivalent to (0.005 g) / (32.0 g mol^{-1}) = 1.6 × 10^{-4} mol.

If the solute is itself a liquid, volume/volume measure usually refers to the volume of solute contained in a fixed volume of *solution* (not *solvent*). The latter distinction is important because volumes of mixed substances are not strictly additive. These kinds of concentration measure are mostly used in commercial and industrial applications. The "proof" of an alcoholic beverage is the (v/v)-percent, multiplied by two; thus a 100-proof vodka has the same alcohol concentration as a solution made by adding sufficient water to 50 ml of alcohol to give 100 ml of solution.

Molarity: mole/volume basis

This is the method most used by chemists to express concentration, and it is the one most important for you to master. Molar concentration (molarity) is the number of moles of solute per liter of solution.

The important point to remember is that the volume of the *solution* is different from the volume of the *solvent*; the latter quantity can be found from the molarity only if the densities of both the solution and of the pure solvent are known. Similarly, calculation of the weight-percentage concentration from the molarity requires density information; you are expected to be able to carry out these kinds of calculations, which are covered in most texts.

Example 8.1.4

How would you make 120 mL of a 0.10 M solution of potassium hydroxide in water?

Solution

The amount of KOH required is

$$(0.120\ \text{L}) \times (0.10\ \text{mol L}^{-1}) = 0.012\ \text{mol}.$$

The molar mass of KOH is 56.1 g, so the weight of KOH required is

$$(0.012\ mol) \times (56.1\ g\ mol^{-1}) = 0.67\ g \tag{8.1.1}$$

We would dissolve this weight of KOH in a volume of water that is less than 120 mL, and then add sufficient water to bring the volume of the solution up to 120 mL.

Note: if we had simply added the KOH to 120 mL of water, the molarity of the resulting solution would not be the same. This is because volumes of different substances are not strictly additive when they are mixed. Without actually measuring the volume of the resulting solution, its molarity would not be known.

Example 8.1.5

Calculate the molarity of a 60-% (w/w) solution of ethanol (C_2H_5OH) in water whose density is 0.8937 g mL^{-1}.

Solution

One liter of this solution has a mass of 893.7 g, of which

$$0.60 \times (893.7\ g) = 536.2\ g \tag{8.1.2}$$

consists of ethanol. The molecular weight of C_2H_5OH is 46.0, so the number of moles of ethanol present in one liter (that is, the molarity) will be

$$\frac{\frac{536.2\ g}{46.0\ g\ mol^{-1}}}{1L} = 11.6\ mol\ L^{-1} \tag{8.1.3}$$

Normality and Equivalents

Normality is a now-obsolete concentration measure based on the number of *equivalents* per liter of solution. Although the latter term is now also officially obsolete, it still finds some use in clinical- and environmental chemistry and in electrochemistry. Both terms are widely encountered in pre-1970 textbooks and articles.

The *equivalent weight* of an *acid* is its molecular weight divided by the number of titratable hydrogens it carries. Thus for sulfuric acid H_2SO_4, one mole has a mass of 98 g, but because both hydrogens can be neutralized by strong base, its equivalent weight

is 98/2 = 49 g. A solution of 49 g of H_2SO_4per liter of water is 0.5 molar, but also "1 normal" (1*N* = 1 eq/L). Such a solution is "equivalent" to a 1*M* solution of HCl in the sense that each can be neutralized by 1 mol of strong base.

solution of $FeCl_3$ is said to be "3 normal" (3 *N*) because it dissociates into three moles/L of chloride ions.

Although molar concentration is widely employed, it suffers from one serious defect: since volumes are temperature-dependent (substances expand on heating), so are molarities; a 0.100 M solution at 0° C will have a smaller concentration at 50° C. For this reason, molarity is not the preferred concentration measure in applications where physical properties of solutions and the effect of temperature on these properties is of importance.

Mole fraction: mole/mole basis

This is the most fundamental of all methods of concentration measure, since it makes no assumptions at all about volumes. The mole fraction of substance *i* in a mixture is defined as

$$X_i = \frac{n_i}{\sum_j n_j} \tag{8.1.4}$$

in which n_j is the number of moles of substance *j*, and the summation is over all substances in the solution. Mole fractions run from zero (substance not present) to unity (the pure substance). The sum of all mole fractions in a solution is, by definition, unity:

$$\sum_i X_i = 1 \tag{8.1.5}$$

Example 8.1.6

What fraction of the molecules in a 60-% (w/w) solution of ethanol in water consist of H_2O?

Solution

From the previous problem, we know that one liter of this solution contains 536.2 g (11.6 mol) of C_2H_5OH. The number of moles of H_2O is

$$(\ (893.7 - 536.2)\ g) / (18.0\ g\ mol^{-1}) = 19.9\ mol.$$

The mole fraction of water is thus

$$\frac{19.9}{19.9 + 11.6} = 0.63 \tag{8.1.6}$$

Thus 63% of the molecules in this solution consist of water, and 37% are ethanol.

In the case of ionic solutions, each kind of ion acts as a separate component.

Example 8.1.7

Find the mole fraction of water in a solution prepared by dissolving 4.5 g of $CaBr_2$ in 84.0 mL of water.

Solution

The molar mass of $CaBr_2$ is 200 g, and 84.0 mL of H_2O has a mass of very close to 84.0 g at its assumed density of 1.00 g mL^{-1}. Thus the number of moles of $CaBr_2$ in the solution is

$$\frac{4.50\ g}{200\ g/mol} = 0.0225\ mol \tag{8.1.7}$$

Because this salt is completely dissociated in solution, the solution will contain 0.268 mol of Ca^{2+} and (2 × .268) = 0.536 of Br^-. The number of moles of water is

$$(84\ g) / (18\ g\ mol^{-1}) = 4.67\ mol.$$

The mole fraction of water is then

$$\frac{0.467\ \cancel{mol}}{0.268 + 0.536 + 4.67\ \cancel{mol}} = \frac{0.467}{5.47} = 0.854 \tag{8.1.8}$$

Thus H_2O constitutes 85 out of every 100 molecules in the solution.

Molality: mole/weight basis

A 1-molal solution contains one mole of solute per 1 kg of solvent. Molality is a hybrid concentration unit, retaining the convenience of mole measure for the solute, but expressing it in relation to a temperature-independent mass rather than a volume. Molality, like mole fraction, is used in applications dealing with certain physical properties of solutions; we will see some of these in the next lesson.

Example 8.1.8

Calculate the molality of a 60-% (w/w) solution of ethanol in water.

Solution

From the above problems, we know that one liter of this solution contains 11.6 mol of ethanol in

$$(893.7 - 536.2) = 357.5\ g$$

of water. The molarity of ethanol in the solution is therefore

$$(11.6\ mol) / (0.3575\ kg) = 32.4\ mol\ kg^{-1}.$$

Conversion between Concentration Measures

Anyone doing practical chemistry must be able to convert one kind of concentration measure into another. The important point to remember is that any conversion involving molarity requires a knowledge of the *density* of the solution.

Example 8.1.9

A solution prepared by dissolving 66.0 g of urea $(NH_2)_2CO$ in 950 g of water had a density of 1.018 g mL^{-1}. Express the concentration of urea in

a. weight-percent
b. mole fraction
c. molarity
d. molality

Solution

a) The weight-percent of solute is (100%) $^{-1}$ (66.0 g) / (950 g) = 6.9%

The molar mass of urea is 60, so the number of moles is

$$(66\ g)\ /(60\ g\ mol^{-1}) = 1.1\ mol.$$

The number of moles of H_2O is

$$(950\ g)\ /\ (18\ g\ mol^{-1}) = 52.8\ mol.$$

b) Mole fraction of urea:

$$(1.1\ mol)\ /\ (1.1 + 52.8\ mol) = 0.020$$

c) molarity of urea: the volume of 1 L of solution is

$$(66 + 950)g\ /\ (1018\ g\ L^{-1}) = 998\ mL.$$

The number of moles of urea (from a) is 1.1 mol.
Its molarity is then

$$(1.1\ mol)\ /\ (0.998\ L) = 1.1\ mol\ L^{-1}.$$

d) The molality of urea is (1.1 mol) / (.066 + .950) kg = 1.08 mol kg^{-1}.

Example 8.1.10

Ordinary dry air contains 21% (v/v) oxygen. About many moles of O_2 can be inhaled into the lungs of a typical adult woman with a lung capacity of 4.0 L?

Solution

The number of molecules (and thus the number of moles) in a gas is directly proportional to its volume (Avogadro's law), so the mole fraction of O_2 is 0.21. The molar volume of a gas at 25° C is

$$(298/271) \times 22.4\ L\ mol^{-1} = 24.4\ L\ mol^{-1}$$

so the moles of O_2 in 4 L of air will be

$$(4\ /\ 24.4) \times (0.21\ mol) \times (24.4\ L\ mol^{-1}) = 0.84\ mol\ O_2.$$

Dilution calculations

These kinds of calculations arise frequently in both laboratory and practical applications. If you have a thorough understanding of concentration definitions, they are easily tackled. The most important things to bear in mind are

- Concentration is inversely proportional to volume;
- Molarity is expressed in mol L^{-1}, so it is usually more convenient to express volumes in liters rather than in mL;
- Use the principles of unit cancelations to determine what to divide by what.

Example 8.1.11

Commercial hydrochloric acid is available as a 10.17 molar solution. How would you use this to prepare 500 mL of a 4.00 molar solution?

Solution

The desired solution requires (0.50 L) × (4.00 M $L^{-1)}$ = 2.0 mol of HCl. This quantity of HCl is contained in (2.0 mol) / (10.17 M L^{-1}) = 0.197 L of the concentrated acid. So one would measure out 197 mL of the concentrated acid, and then add water to make the total volume of 500 mL.

Example 8.1.12

Calculate the molarity of the solution produced by adding 120 mL of 6.0 M HCl to 150 mL of 0.15 M HCl. What important assumption must be made here?

Solution

The assumption, of course, is that the density of HCl within this concentration range is constant, meaning that their volumes will be additive.

Moles of HCl in first solution:

$$(0.120\ \text{L}) \times (6.0\ \text{mol}\ L^{-1}) = 0.72\ \text{mol HCl}$$

Moles of HCl in second solution:

$$(0.150\ \text{L}) \times (0.15\ \text{mol}\ L^{-1}) = 0.02\ \text{mol HCl}$$

Molarity of mixture:

$$(0.72 + 0.02)\ \text{mol} / (.120 + .150)\ \text{L} = 4.3\ \text{mol}\ L^{-1}.$$

This page titled 8.1: Solutions and their Concentrations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

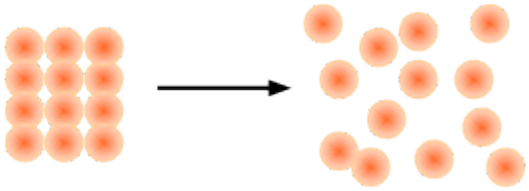
8.2: Thermodynamics of Solutions

Learning Objectives

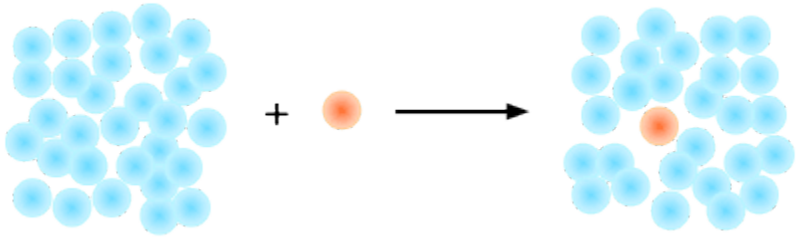
Make sure you thoroughly understand the following essential ideas:

- Describe the two fundamental processes that must occur whenever a solute dissolves in a solvent, and discuss the effects of the absorption or release of energy on the extent of these processes.
- Another factor entering into the process of solution formation is the increase (or occasionally, the decrease) in the entropy — that is, the degree to which thermal energy is dispersed or "diluted". Explain this in your own terms.
- Explain how the adage "like dissolves like" reflects the effects mentioned above. What is the principal physical property of a molecule that defines this "likeness"?
- What do we mean when we describe a liquid such as water as "associated"? Explain how this relates to the the solubility of solutes in such liquids.

You may recall that in the earlier unit on phase equilibria, we pointed out that aggregations of molecules that are more disordered tend to be the ones that are favored at higher temperature, whereas those that possess the lowest potential energy are favored at lower temperatures. This is a general principle that applies throughout the world of matter; the stable form at any given temperature will always be that which leads to the best balance between low potential energy and high molecular disorder. To see how these considerations are applied to solutions, think about the individual steps that must be carried out when a solute is dissolved in a solvent:



1. If the solute is a solid or liquid, it must first be *dispersed* — that is, its molecular units must be pulled apart. This requires energy, and so this step always works against solution formation.



2. The solute must then be introduced into the solvent. Whether this is energetically favorable or unfavorable depends on the nature of the solute and solvent. If the solute is A and the solvent is B, then what is important is the strength of the attractive forces between A-A and B-B molecules, compared to those between A-B pairs; if the latter are greater, then the potential energy will be lower when the substances are mixed and solution formation will be favored.

If step **2** releases more energy than is consumed in step **1**, this will favor solution formation, and we can generally expect the solute to be soluble in the solvent. Even if the dissolution process is slightly endothermic, there is a third important factor, the *entropy* increase, that will very often favor the dissolved state.

Entropy of Solution

As anyone who has shuffled a deck of cards knows, disordered arrangements of objects are statistically more favored simply because there are more ways in which they can be realized. And as the number of objects increases, the more does statistics

govern their most likely arrangements. The numbers of objects (molecules) we deal with in Chemistry is so huge that their tendency to become as spread out as possible becomes overwhelming. However, in doing so, the thermal energy they carry with them is also spread and dispersed, so the availability of this energy, as measured by the temperature, is also of importance. Chemists use the term "entropy" to denote this aspect of molecular randomness.

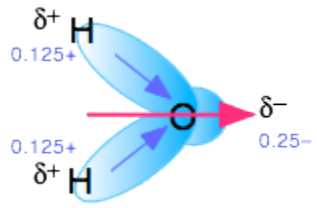
Readers of this section who have had some exposure to thermodynamics will know that solubility, like all equilibria, is governed by the *Gibbs free energy change* for the process, which incorporates the entropy change at a fundamental level. A proper understanding of these considerations requires some familiarity with thermodynamics, which most students do not encounter until well into their second semester of Chemistry. If you are not there yet, do not despair; you are hereby granted temporary permission to think of molecular "disorder" and entropy simply in terms of "spread-outedness".

Thus in the very common case in which a small quantity of solid or liquid dissolves in a much larger volume of solvent, the solute becomes more spread out in space, and the number of equivalent ways in which the solute can be distributed within this volume is greatly increased. This is the same as saying that the entropy of the solute increases.

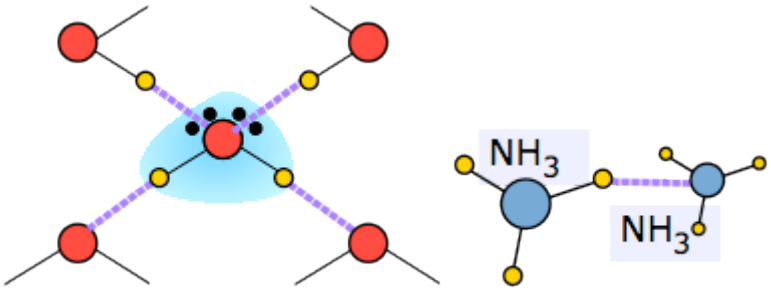
If the energetics of dissolution are favorable, this increase in entropy means that the conditions for solubility will always be met. Even if the energetics are slightly endothermic, the entropy effect can still allow the solution to form, although perhaps limiting the maximum concentration that can be achieved. In such a case, we may describe the solute as being *slightly soluble* in a certain solvent. What this means is that a greater volume of solvent will be required to completely dissolve a given mass of solute.

Enthalpy of Solution

Polar molecules are those in which electric charge is distributed asymmetrically. The most familiar example is ordinary water, in which the highly electronegative oxygen atom pulls part of the electric charge cloud associated with each O–H bond closer to itself. Although the H_2O molecule is electrically neutral overall, this charge imbalance gives rise to a permanent *electric dipole moment*.

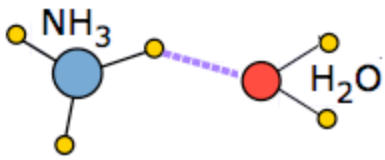


Chemists use the term "Associated" liquids to refer to liquids in which the effects of hydrogen bonding dominate the local structure. Water is the most important of these, but ammonia NH_3 and hydrogen cyanide HCN are other common examples.



Thus liquid water consists of an extended network of H_2O molecules linked together by dipole-dipole attractions that we call *hydrogen bonds*. Because these are much weaker than ordinary chemical bonds, they are continually being disrupted by thermal forces. As a result, the extended structure is highly disordered (in contrast to that of solid ice) and continually changing.

When a solute molecule is introduced into an associated liquid, a certain amount of energy must be expended in order to break the local hydrogen-bond structure and make space for the new molecule. If the solute is itself an ion or a polar molecule, new ion-dipole or dipole-dipole attractions come into play. In favorable cases these may release sufficient potential energy to largely compensate for the energy required to incorporate the solute into the structure.

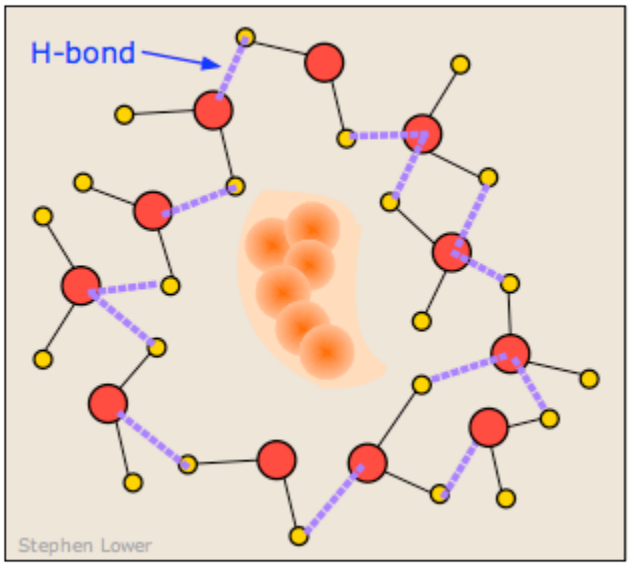


An extreme example of this occurs when ammonia dissolves in water. Each NH_3 molecule can form three hydrogen bonds, so the resulting solution is even more hydrogen-bonded than is pure water — accounting for the considerable amount of heat released in the process and the extraordinarily large solubility of ammonia in water.

Nonpolar solutes are Sparingly Soluble in Water: The Hydrophobic effect

When a nonpolar solute such as oxygen or hexane is introduced into an associated liquid, we might expect that the energy required to break the hydrogen bonds to make space for the new molecule is not compensated by the formation of new attractive interactions, suggesting that the process will be energetically unfavorable. We can therefore predict that solutes of these kinds will be only sparingly soluble in water, and this is indeed the case.

It turns out, however, that this is not an entirely correct explanation for the small solubility of non polar solutes in water. It is now known that the H_2O molecules that surround a non-polar intruder and find themselves unable to form energy-lowering polar or hydrogen-bonded interactions with it will rearrange themselves into a configuration that maximizes the hydrogen bonding between the water molecules themselves. In doing so, this creates a cage-like shell around the solute molecule. In terms of the energetics of the process, these new H_2O-H_2O interactions largely compensate for the lack of solute-H_2O interactions.



However, this shell of highly organized water molecules exacts its own toll on the solubility by reducing the entropy of the system. Dissolution of a solute normally increases the entropy by spreading the solute molecules (and the thermal energy they contain) through the larger volume of the solvent. But in this case, the H_2O molecules within the highly structured shell surrounding the solute molecule are themselves constrained to this location, and their number is sufficiently great to reduce the entropy by far more than the dissolved solute increases it.

The implications of the hydrophobic effect extend far beyond the topic of solubility. It governs the way that proteins fold, the formation of soap bubbles, and the formation of cell membranes. The small solubility of a non polar solute in an associated liquid such as water results more from the negative entropy change rather than from energetic considerations. This phenomenon is known as the *hydrophobic effect*. In the next section, we will explore the ways in which these energy-and-entropy considerations come together in various kinds of solutions.

This page titled 8.2: Thermodynamics of Solutions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

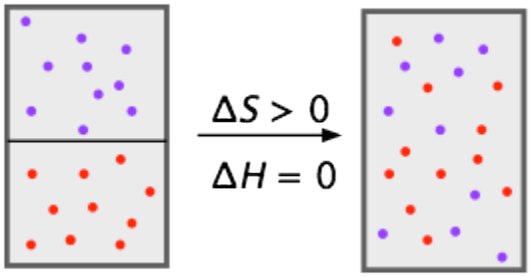
8.2.2A: Solutions of Gaseous Solutes in Gaseous Solvents

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

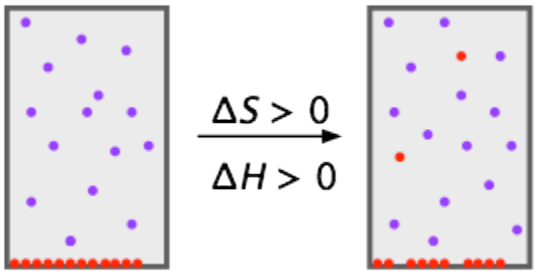
- Why are gases the only state of matter that never fail to form solutions?

Mixtures of gases are really solutions, but we tend not to think of them this way because they mix together freely and with no limits to their compositions; we say that gases are *miscible in all proportions*.



To the extent that gases behave ideally (because they consist mostly of empty space), their mixing does not involve energy changes at all; the mixing of gases is driven entirely by the increase in entropy (S) as each kind of molecule occupies and shares the space and kinetic energy of the other. Your nose can be a remarkably sensitive instrument for detecting components of gaseous solutions, even at the parts-per-million level. The olfactory experiences resulting from cooking cabbage, eating asparagus, and bodily emanations that are not mentionable in polite society are well known.

Can solids or liquids "dissolve" in a gaseous solvent? In a very narrow sense they can, but only to a very small extent. Dissolution of a condensed phase of matter into a gas is formally equivalent to evaporation (of a liquid) or sublimation (of a solid), so the process really amounts to the mixing of gases.



The energy required to remove molecules from their neighbors in a liquid or solid and into the gaseous phase is generally too great to be compensated by the greater entropy they enjoy in the larger volume of the mixture, so solids tend to have relatively low vapor pressures. The same is true of liquids at temperatures well below their boiling points. These two cases of gaseous solutions can be summarized as follows:

gaseous solvent, solute →	gas
energy to disperse solute	nil
energy to introduce into gas	nil
increase in entropy	large
miscibility	complete

This page titled 8.2.2A: Solutions of Gaseous Solutes in Gaseous Solvents is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.2B: Solutions of Gaseous Solutes in Liquid Solvents

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Explain why, in contrast, gases tend to be only slightly soluble in liquids and solids. And why are some combinations, such as the dissolution of ammonia or hydrogen chloride in water, significant exceptions to this rule?
- State Henry's law and explain
- Why do fish in rivers and streams sometimes become asphyxiated (oxygen-starved) in hot weather?

Solutions of Gases in Liquids

Gases dissolve in liquids, but usually only to a small extent. When a gas dissolves in a liquid, the ability of the gas molecules to move freely throughout the volume of the solvent is greatly restricted. If this latter volume is small, as is often the case, the gas is effectively being compressed. Both of these effects amount to a decrease in the entropy of the gas that is not usually compensated by the entropy increase due to mixing of the two kinds of molecules. Such processes greatly restrict the solubility of gases in liquids.

liquid solvent, solute →	gas
energy to disperse solute	nil
energy to introduce into solvent	medium to large
increase in entropy	negative
miscibility	**usually very limited**

Solubility of gases in water

One important consequence of the entropy decrease when a gas dissolves in a liquid is that the solubility of a gas decreases at higher temperatures; this is in contrast to most other situations, where a rise in temperature usually leads to increased solubility. Bringing a liquid to its boiling point will completely remove a gaseous solute. Some typical gas solubilities, expressed in the number of moles of gas at 1 atm pressure that will dissolve in a liter of water at 25° C, are given below:

solute	formula	solubility, mol L^{-1} atm^{-1}
ammonia	NH_3	57
carbon dioxide	CO_2	0.0308
methane	CH_4	0.00129
nitrogen	N_2	0.000661
oxygen	O_2	0.00126
sulfur dioxide	SO_2	1.25

As we indicated above, the only gases that are readily soluble in water are those whose polar character allows them to interact strongly with it.

Ammonia is remarkably soluble in water

Inspection of the above table reveals that ammonia is a champion in this regard. At 0° C, one liter of water will dissolve about 90 g (5.3 mol) of ammonia. The reaction of ammonia with water according to

$$NH_3 + H_2O \longrightarrow NH_4^+ + OH^- \tag{8.2.2B.1}$$

makes no significant contribution to its solubility; the equilibrium lies heavily on the left side (as evidenced by the strong odor of ammonia solutions). Only about four out of every 1000 NH_3molecules are in the form of ammonium ions at equilibrium. This is truly impressive when one calculates that this quantity of NH_3 would occupy (5.3 mol) × (22.4 L mol^{-1}) = 119 L at STP. Thus one volume of water will dissolve over 100 volumes of this gas. It is even more impressive when you realize that in order to compress 119 L of an ideal gas into a volume of 1 L, a pressure of 119 atm would need to be applied! This, together with the observation that dissolution of ammonia is accompanied by the liberation of a considerable amount of heat, tells us that the high solubility of ammonia is due to the formation of more hydrogen bonds (to H_2O) than are broken within the water structure in order to accommodate the NH_3 molecule.

If we actually compress 90 g of pure NH_3 gas to 1 L, it will liquefy, and the vapor pressure of the liquid would be about 9 atm. In other words, the escaping tendency of NH_3 molecules from H_2O is only about 1/9th of what it is from liquid NH_3. One way of interpreting this is that the strong intermolecular (dipole-dipole) attractions between NH_3 and the solvent H_2O give rise to a force that has the effect of a negative pressure of 9 atm.

This classic experiment nicely illustrates the high solubility of gaseous ammonia in water. A flask fitted with a tube as shown is filled with ammonia gas and inverted so that the open end of tube is submerged in a container of water. A small amount of water is pushed up into the flask to get the process started. As the gas dissolves in the water, its pressure is reduced, creating a partial vacuum that draws additional water into the flask. The effect can be made more dramatic by adding an indicator dye such as phenolphthalein to the water, which turns pink as the water emerges from the "fountain" and becomes alkaline.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

In old textbooks, ammonia's extraordinarily high solublility in water was incorrectly attributed to the formation of the non-existent compound "ammonium hydroxide" NH_4OH. Although this formula is still occasionally seen, the name ammonium hydroxide is now used as a synonym for "aqueous ammonia" whose formula is simply NH_3*(aq)*.

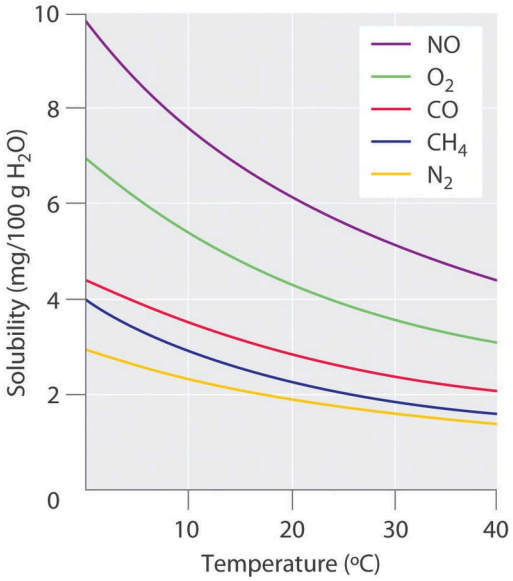
As can also be seen in the above table, the gases CO_2 and SO_2 also exhibit higher solubilities in water. The main product in each case is a loosely-bound *hydrate* of the gas, denoted by CO_2*(aq)* or SO_2*(aq)*. A very small fraction of the hydrate $CO_2 \cdot H_2O$ then combines to form *carbonic acid* H_2CO_3.

Solubility of gases decreases with Temperature

Recall that entropy is a measure of the ability of thermal energy to spread and be shared and exchanged by molecules in the system. Higher temperature exerts a kind of multiplying effect on a positive entropy change by increasing the amount of thermal energy available for sharing. Have you ever noticed the tiny bubbles that form near the bottom of a container of water when it is placed on a hot stove? These bubbles contain air that was previously dissolved in the water, but reaches its solubility limit as the water is warmed. You can completely rid a liquid of any dissolved gases (including unwanted ones such as Cl_2 or H_2S) by boiling it in an open container.

This is quite different from the behavior of most (but not all) solutions of solid or liquid solutes in liquid solvents. The reason for this behavior is the very large entropy increase that gases undergo when they are released from the confines of a condensed phase .

Fresh water at sea level dissolves 14.6 mg of oxygen per liter at 0°C and 8.2 mg/L at 25°C. These saturation levels ensure that fish and other gilled aquatic animals are able to extract sufficient oxygen to meet their respiratory needs. But in actual aquatic environments, the presence of decaying organic matter or nitrogenous runoff can reduce these levels far below saturation. The health and survival of these organisms is severely curtailed when oxygen concentrations fall to around 5 mg/L.



The temperature dependence of the solubility of oxygen in water is an important consideration for the well-being of aquatic life; thermal pollution of natural waters (due to the influx of cooling water from power plants) has been known to reduce the dissolved oxygen concentration to levels low enough to kill fish. The advent of summer temperatures in a river can have the same effect if the oxygen concentration has already been partially depleted by reaction with organic pollutants.

Solubility of gases increases with pressure: Henry's Law

The pressure of a gas is a measure of its "escaping tendency" from a phase. So it stands to reason that raising the pressure of a gas in contact with a solvent will cause a larger fraction of it to "escape" into the solvent phase. The direct-proportionality of gas solubility to pressure was discovered by William Henry (1775-1836) and is known as *Henry's Law*. It is usually written as

$$P = k_H C \tag{8.2.2B.2}$$

with

- P is the partial pressure of the gas above the liquid,
- C is the concentration of gas dissolved in the liquid, and
- k_H is the *Henry's law constant*, which can be expressed in various units, and in some instances is defined in different ways, so be very careful to note these units when using published values.

For Table 7b.2.X, k_H is given as

$$k_H = \frac{\text{partial pressure of gas in atm}}{\text{concentration in liquid } mol\ L^{-1}} \tag{8.2.2B.3}$$

Table 7b.2.X: *Henry's law constants in water at 25° C, L atm mol^{-1}*

gas	He	N_2	O_2	CO_2	CH_4	NH_3
K_H	2703	1639	769	29.4	0.00129	57

Example $8.2.2B.1$

Some vendors of bottled waters sell pressurized "oxygenated water" that is (falsely) purported to enhance health and athletic performance by supplying more oxygen to the body.

a. How many moles of O_2 will be in equilibrium with one liter of water at 25° C when the partial pressure of O_2 above the water is 2.0 atm?
b. How many mL of air (21% O_2 v/v) must you inhale in order to introduce an equivalent quantity of O_2 into the lungs (where it might actually do some good?)

Solution:

a. Solving Henry's law for the concentration, we get

$$C = \frac{P}{k_H} = \frac{2.0\ atm}{769\ L\ atm\ mol^{-1}} = 0.0026\ mol\ L^{-1} \tag{8.2.2B.4}$$

a. At 25° C, 0.0026 mol of O_2 occupies (22.4 L) × (.0026 mol) × (298/273) = 0.063 L. The equivalent volume of air would be (0.063 L) (.21) = 0.303 L. Given that the average tidal volume of the human lung is around 400 mL, this means that taking one extra breath would take in more O_2 than is present in 1 L of "oxygenated water".

Artificially carbonated water was first prepared by Joseph Priestley (who later discovered oxygen) in 1767 and was commercialized in 1783 by Joseph Schweppe, a Swiss-German jeweler. Naturally-carbonated spring waters have long been reputed to have curative values, and these became popular tourist destinations in the 19th century. The term "seltzer water" derives from one such spring in Niederselters, Germany. Of course, carbonation produced by fermentation has been known since ancient times. The tingling sensation that carbonated beverages produce in the mouth comes from the carbonic acid produced when bubbles of carbon dioxide come into contact with the mucous membranes of the mouth and tongue:

$$CO_2 + H_2O \rightarrow H_2CO_3 \tag{8.2.2B.5}$$

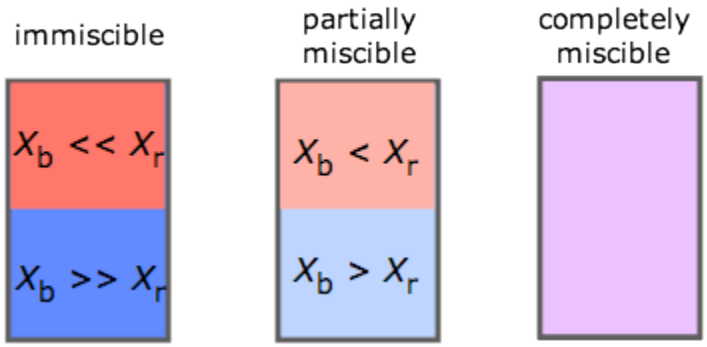
This page titled 8.2.2B: Solutions of Gaseous Solutes in Liquid Solvents is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.2C: Solutions of Liquid Solutes in Liquid Solvents

Whereas all gases will mix to form solutions regardless of the proportions, liquids are much more fussy. Some liquids, such as ethyl alcohol and water, are miscible in all proportions. Others, like the proverbial oil and water, are not; each liquid has only a limited solubility in the other, and once either of these limits is exceeded, the mixture separates into two phases.

solute →	
energy to disperse solute	
energy to introduce into solvent	
increase in entropy	
miscibility	

The reason for this variability is apparent from the table. Mixing of two liquids can be exothermic, endothermic, or without thermal effect, depending on the particular substances. Whatever the case, the energy factors are not usually very large, but neither is the increase in randomness; the two factors are frequently sufficiently balanced to produce limited miscibility.



The range of possibilities is shown here in terms of the mole fractions *X* of two liquids A and B. If A and B are only slightly miscible, they separate into two layers according to their relative densities. Note that when one takes into account trace levels, no two liquids are totally immiscible.

A useful general rule is that liquids are completely miscible when their intermolecular forces are very similar in nature; "like dissolves like". Thus water is miscible with other liquids that can engage in hydrogen bonding, whereas a hydrocarbon liquid in which London or dispersion forces are the only significant intermolecular effect will only be completely miscible with similar kinds of liquids.

Substances such as the alcohols, $CH_3(CH_2)_nOH$, which are hydrogen-bonding (and thus hydrophilic) at one end and hydrophobic at the other, tend to be at least partially miscible with both kinds of solvents. If *n* is large, the hydrocarbon properties dominate and the alcohol has only a limited solubility in water. Very small values of *n* allow the –OH group to dominate, so miscibility in water increases and becomes unlimited in ethanol ($n = 1$) and methanol ($n = 0$), but miscibility with hydrocarbons decreases owing to the energy required to break alcohol-alcohol hydrogen bonds when the non polar liquid is added.

These considerations have become quite important in the development of alternative automotive fuels based on mixing these alcohols with gasoline. At ordinary temperatures the increased entropy of the mixture is great enough that the unfavorable energy factor is entirely overcome, and the mixture is completely miscible. At low temperatures, the entropy factor becomes less predominant, and the fuel mixture may separate into two phases, presenting severe problems to the fuel filter and carburetor.

This page titled 8.2.2C: Solutions of Liquid Solutes in Liquid Solvents is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.2.2D: Solutions of Solid Solutes in Liquid Solvents

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- A key to understanding the solubility of ionic solids in water are the concepts of lattice energy and hydration energy. Explain the meaning of these terms, and sketch out a diagram that shows how these are related to the "heat of solution".

Molecular Solids in Liquid Solvents

The stronger intermolecular forces in solids require more input of energy to disperse the molecular units into a liquid solution, but there is also a considerable increase in entropy that can more than compensate if the intermolecular forces are not too strong, and if the solvent has no strong hydrogen bonds that must be broken in order to introduce the solute into the liquid.

solvent →	non polar liquid	polar liquid
energy to disperse solute	moderate	moderate
energy to introduce into solvent	small	moderate
increase in entropy	moderate	moderate
miscibility	**moderate**	**small**

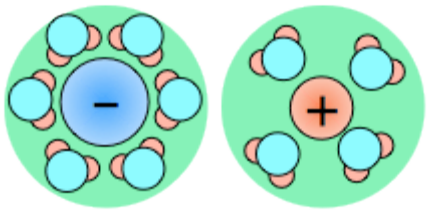
For example, at 25° C and 1 atm pressure, 20 g of iodine crystals will dissolve in 100 ml of ethyl alcohol, but the same quantity of water will dissolve only 0.30 g of iodine. As the molecular weight of the solid increases, the intermolecular forces holding the solid together also increase, and solubilities tend to fall off; thus the solid linear hydrocarbons $CH_3(CH_2)_nCH_3$ ($n > 20$) show diminishing solubilities in hydrocarbon liquids.

Ionic Solids in Liquid Solvents

Since the Coulombic forces that bind ions and highly polar molecules into solids are quite strong, we might expect these solids to be insoluble in just about any solvent. Ionic solids are insoluble in most non-aqueous solvents, but the high solubility of some (including NaCl) in water suggests the need for some further explanation.

solvent →	non polar	polar (water)
energy to disperse solute	large	large (endothermic)
energy to introduce into liquid	small	highly negative (exothermic)
increase in entropy	moderate	moderate to slightly negative
miscibility	very small	small to large

The key factor here turns out to be the interaction of the ions with the solvent. The electrically-charged ions exert a strong coulombic attraction on the end of the water molecule that has the opposite partial charge.



As a consequence, ions in solution are always *hydrated*; that is, they are quite tightly bound to water molecules through ion-dipole interaction. The number of water molecules contained in the *primary hydration shell* varies with the radius and charge of the ion.

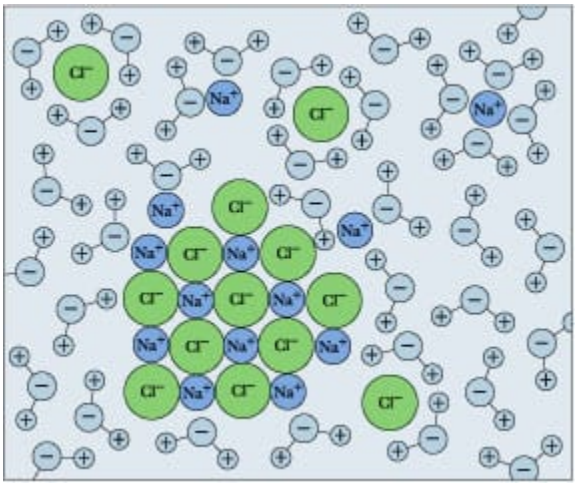
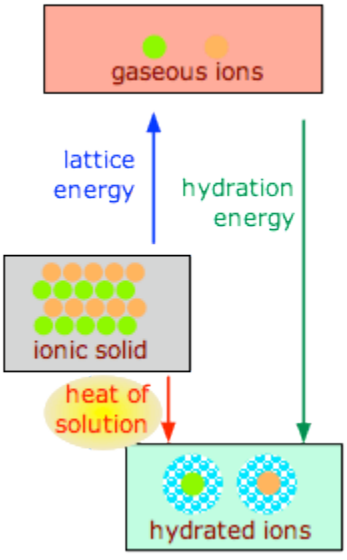


Figure $8.2.2D.1$: Hydration shells around some ions in a sodium chloride solution. The average time an ion spends in a shell is about 2-4 nanoseconds. But this is about two orders of magnitude longer than the lifetime of an individual H_2O–H_2O hydrogen bond.

Lattice and Hydration Energies



The dissolution of an ionic solid MX in water can be thought of as a sequence of two (hypothetical) steps:

$$MX(s) \rightarrow M^+(g) + X^-(g) \tag{8.2.2D.1}$$

$$M^+(g) + X^-(g) + H_2O(l) \rightarrow M^+(aq) + X{-}(aq) \tag{8.2.2D.2}$$

The enthalpy difference of the first step is the lattice energy and is always positive; the enthalpy difference of the second step is the hydration energy and is always negative)

- The first reaction is always endothermic; it takes a lot of work to break up an ionic crystal lattice (Table $8.2.2D.1$).
- The hydration step is always exothermic as H_2O molecules are attracted into the electrostatic field of the ion (Table $8.2.2D.2$).
- The heat (enthalpy) of solution is the sum of the lattice and hydration energies, and can have either sign.

Table $8.2.2D.1$***: Hydration Energies (kJ mol^{-1})***

$H^+(g)$	–1075	**$F^-(g)$**	–503
$Li^+(g)$	–515	**$Cl^-(g)$**	–369
$Na^+(g)$	–405	**$Br^-(g)$**	–336
$K^+(g)$	–321	**$I^-(g)$**	–398
$Mg^{2+}(g)$	–1922	**$OH^-(g)$**	–460
$Ca^{2+}(g)$	–1592	**NO_3^-**	–328
$Sr^{2+}(g)$	–1445	**SO_4^{2-}**	–1145

Single-ion hydration energies (Table $8.2.2D.1$) cannot be observed directly, but are obtained from the differences in hydration energies of salts having the given ion in common. When you encounter tables such as the above in which numeric values are related to different elements, you should always stop and see if you can make sense of any obvious trends. In this case, the things to look for are the size and charge of the ions as they would affect the electrostatic interaction between two ions or between an ion and a [polar] water molecule.

Table $8.2.2D.2$***: Lattice Energies (kJ/mol^{-1})***

F^-	Cl^-	Br^-	I^-	
Li^+	+1031	+848	+803	+759
Na^+	+918	+780	+742	+705
K^+	+817	+711	+679	+651
Mg^{2+}	+2957	+2526	+2440	+2327
Ca^{2+}	+2630	+2258	+2176	+2074
Sr^{2+}	+2492	+2156	+2075	+1963

Lattice energies are not measured directly, but are estimates based on electrostatic calculations which are reliable only for simple salts. Enthalpies of solution are observable either directly or (for sparingly soluble salts,) indirectly. Hydration energies are not measurable; they are estimated as the sum the other two quantities. It follows that any uncertainty in the lattice energies is reflected in those of the hydration energies. For this reason, tabulated values of the latter will vary depending on the source.

Example $8.2.2D.1$: Calcium Chloride

When calcium chloride, $CaCl_2$, is dissolved in water, will the temperature immediately after mixing rise or fall?

Solution:

Estimate the heat of solution of $CaCl_2$.

- lattice energy of solid $CaCl_2$: +2258 kJ mol^{-1}
- hydration energy of the three gaseous ions: (–1562 –381 – 381) = –2324 kJ mol^{-1}
- heat of solution:

$$(2258 - 2324)\ \text{kJ mol}^{-1} = \mathbf{-66\ kJ\ mol^{-1}}$$

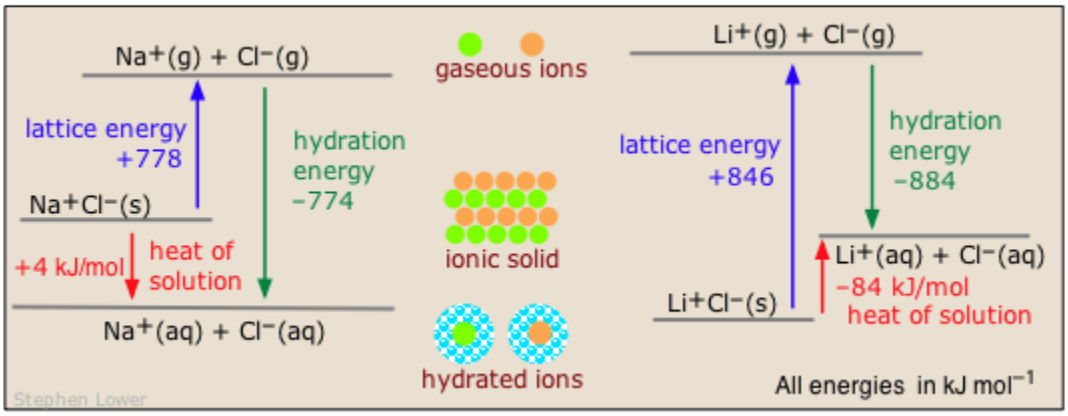
Since the process is exothermic, this heat will be released to warm the solution.

As often happens for a quantity that is the sum of two large terms having opposite signs, the overall dissolution process can come out as either endothermic or exothermic, and examples of both kinds are common.

Table 8.2.2D.3: Energy terms associated with the dissolution of some salts (kJ mol^{-1})

substance →	LiF	NaI	KBr	CsI	LiCl	NaCl	KCl	AgCl
lattice energy	1021	682	669	586	846	778	707	910
hydration energy	1017	686	649	552	884	774	690	844
enthalpy of solution	+3	–4	+20	+34	–38	+4	+17	+66

Two common examples illustrate the contrast between exothermic and endothermic heats of solution of ionic solids:



Hydration Entropy can make a Difference!

Hydration shells around some ions in a sodium chloride solution. The average time an ion spends in a shell is about 2-4 nanoseconds. But this is about two orders of magnitude longer than the lifetime of an individual H_2O–H_2O hydrogen bond. The balance between the lattice energy and hydration energy is a major factor in determining the solubility of an ionic crystal in water, but there is another factor to consider as well. We generally assume that there is a rather large increase in the entropy when a solid is dispersed into the liquid phase. However, in the case of ionic solids, each ion ends up surrounded by a shell of oriented water molecules. These water molecules, being constrained within the hydration shell, are unable to participate in the spreading of thermal energy throughout the solution, and reduce the entropy. In some cases this effect predominates so that dissolution of the salt leads to a net decrease in entropy. Recall that any process in which the the entropy diminishes becomes less probable as the temperature increases; this explains why the solubilities of some salts decrease with temperature.

This page titled 8.2.2D: Solutions of Solid Solutes in Liquid Solvents is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

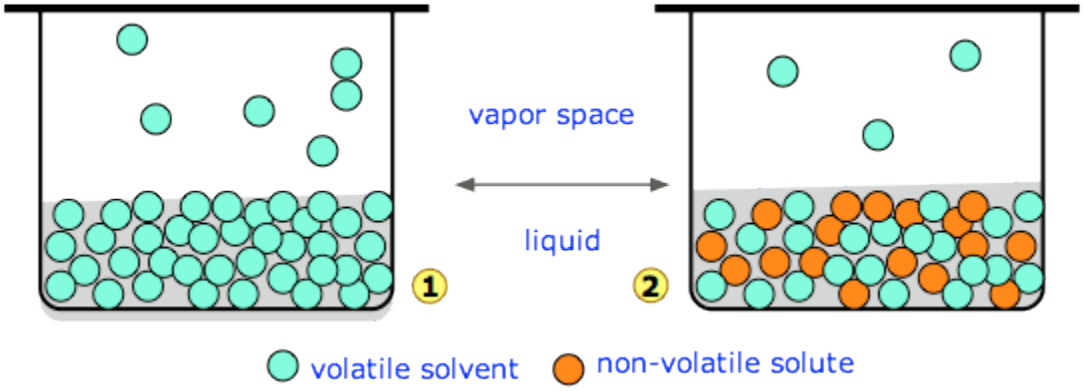
8.3: Colligative Properties- Raoult's Law

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- State *Raoult's law* in your own words, and explain why it makes sense.
- What do we mean by the *escaping tendency* of a molecule from a phase? How might we be able to observe or measure it?
- Explain why *boiling point elevation* follows naturally from Raoult's law.
- Explaining *freezing point depression* is admittedly a bit more difficult, but you should nevertheless be able to explain how the application of salt on an ice-covered road can cause the ice to melt.

The tendency of molecules to escape from a liquid phase into the gas phase depends in part on how much of an increase in entropy can be achieved in doing so. Evaporation of solvent molecules from the liquid always leads to a large increase in entropy because of the greater volume occupied by the molecules in the gaseous state. But if the liquid solvent is initially "diluted" with solute, its entropy is already larger to start with, so the amount by which it can increase on entering the gas phase will be less. There will accordingly be less tendency for the solvent molecules to enter the gas phase, and so the vapor pressure of the solution diminishes as the concentration of solute increases and that of solvent decreases.



The number 55.5 mol L^{-1} (= 1000 g L^{-1} ÷ 18 g mol^{-1}) is a useful one to remember if you are dealing a lot with aqueous solutions; this represents the concentration of water in pure water. (Strictly speaking, this is the *molal* concentration of H_2O; it is only the molar concentration at temperatures around 4° C, where the density of water is closest to 1.000 g cm^{-1}.)

Diagram 1 (above left) represents pure water whose concentration in the liquid is 55.5 M. A tiny fraction of the H_2O molecules will escape into the vapor space, and if the top of the container is closed, the pressure of water vapor builds up until equilibrium is achieved. Once this happens, water molecules continue to pass between the liquid and vapor in both directions, but at equal rates, so the partial pressure of H_2O in the vapor remains constant at a value known as the *vapor pressure* of water at the particular temperature.

In Figure 8.3.1, we have replaced a fraction of the water molecules with a substance that has zero or negligible vapor pressure — a *nonvolatile solute* such as salt or sugar. This has the effect of diluting the water, reducing its escaping tendency and thus its vapor pressure.

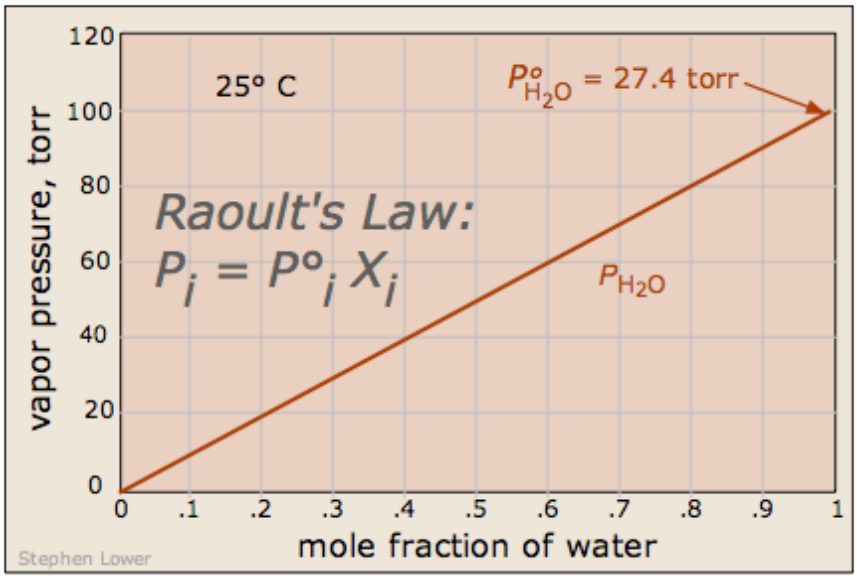
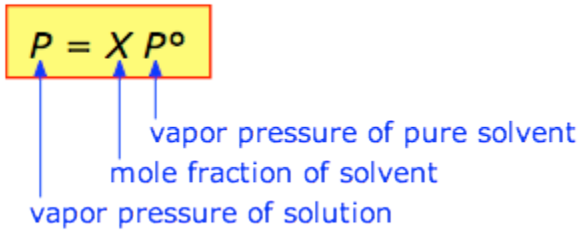


Figure 8.3.1:

What's important to remember is that the reduction in the vapor pressure of a solution of this kind is directly proportional to the fraction of the [volatile] solute molecules in the liquid — that is, to the mole fraction of the solvent. The reduced vapor pressure is given by *Raoult's law*(1886):



From the definition of mole fraction, you should understand that in a two-component solution (i.e., a solvent and a single solute),

$$\chi_{solvent} = 1 - \chi_{solute}. \tag{8.3.1}$$

Example 8.3.1

Estimate the vapor pressure of a 40 % (W/W) solution of ordinary cane sugar ($C_{22}O_{11}H_{22}$, 342 g mol^{-1}) in water. The vapor pressure of pure water at this particular temperature is 26.0 torr.

Solution

100 g of solution contains (40 g) ÷ (342 g mol^{-1}) = 0.12 mol of sugar and (60 g) ÷ (18 g mol^{-1}) = 3.3 mol of water. The mole fraction of water in the solution is

$$\frac{3.3}{3.3 + 12} = 0.96 \tag{8.3.2}$$

and its vapor pressure will be 0.96 × 26.0 torr = 25.1 torr.

Example 8.3.2

The vapor pressure of water at 10° C is 9.2 torr. Estimate the vapor pressure at this temperature of a solution prepared by dissolving 1 mole of $CaCl_2$ in 1 L of water.

Solution

Each mole of $CaCl_2$ dissociates into one mole of Ca^{2+} and two moles of Cl^{1-}, giving a total of three moles of solute particles. The mole fraction of water in the solution will be

$$\frac{55.5}{3+55.5} = 0.95 \tag{8.3.3}$$

The vapor pressure will be 0.95 × 9.2 torr = 8.7 torr.

Since the sum of all mole fractions in a mixture must be unity, it follows that the more moles of solute, the smaller will be the mole fraction of the solvent. Also, if the solute is a salt that dissociates into ions, then the proportion of solvent molecules will be even smaller.

This page titled 8.3: Colligative Properties- Raoult's Law is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.4: Colligative Properties- Boiling Point Elevation and Freezing Point Depression

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Explain why boiling point elevation follows naturally from Raoult's law.
- Explaining freezing point depression is admittedly a bit more difficult, but you should nevertheless be able to explain how the application of salt on an ice-covered road can cause the ice to melt.

The colligative properties really depend on the escaping tendency of solvent molecules from the liquid phase. You will recall that the vapor pressure is a direct measure of escaping tendency, so we can use these terms more or less interchangeably.

Boiling Point Elevation

If addition of a nonvolatile solute lowers the vapor pressure of the solution via Raoult's law , then it follows that the temperature must be raised to restore the vapor pressure to the value corresponding to the pure solvent. In particular, the temperature at which the vapor pressure is 1 atm will be higher than the normal boiling point by an amount known as the boiling point elevation . The exact relation between the boiling point of the solution and the mole fraction of the solvent is rather complicated, but for dilute solutions the elevation of the boiling point is directly proportional to the *molal* concentration of the solute:

$$\Delta_B = K_B \frac{\text{moles of solute}}{\text{kg of solution}}$$

Δ_B: boiling point elevation (K)

K_B: molal boiling point elevation constant (K mol^{-1} kg)

Bear in mind that the proportionality constant K_B is a property of the ***solvent*** because this is the only component that contributes to the vapor pressure in the model we are considering in this section.

Table 8.4.1*: Boiling point elevation constants*

solvent	normal bp, °C	K_b , K mol^{-1} kg
water	100	0.514
ethanol	79	1.19
acetic acid	118	2.93
carbon tetrachloride	76.5	5.03

Example 8.4.1

Sucrose ($C_{22}O_{11}H_{22}$, 342 g mol^{-1}), like many sugars, is highly soluble in water; almost 2000 g will dissolve in 1 L of water, giving rise to what amounts to pancake syrup. Estimate the boiling point of such a sugar solution.

Solution

moles of sucrose:

$$\frac{2000\,g}{342\,g\,mol^{-1}} = 5.8\ mol \tag{8.4.1}$$

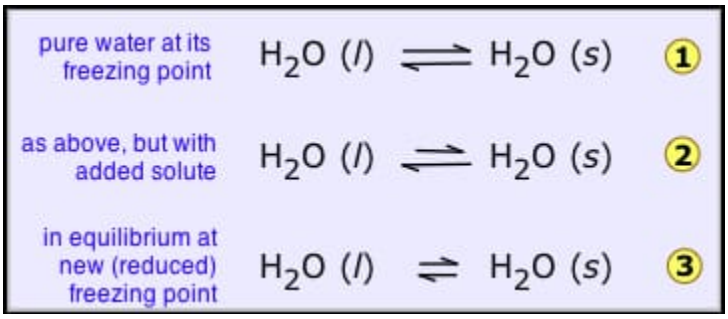
mass of water: assume 1000 g (we must know the density of the solution to find its exact value)

The molality of the solution is (5.8 mol) ÷ (1.0 kg) = 5.8 m.

Using the value of K_b from the table, the boiling point will be raised by (0.514 K mol^{-1} kg) × (5.8 mol kg^{-1}) = 3.0 K, so the boiling point will be 103° C.

Freezing Point Depression

The freezing point of a substance is the temperature at which the solid and liquid forms can coexist indefinitely — that is, they are in equilibrium. Under these conditions molecules pass between the two phases at equal rates because their escaping tendencies from the two phases are identical. Suppose that a liquid solvent and its solid (water and ice, for example) are in equilibrium (1 below), and we add a non-volatile solute (such as salt, sugar, or automotive antifreeze liquid) to the water. This will have the effect of reducing the mole fraction of H_2O molecules in the liquid phase, and thus reduce the tendency of these molecules to escape from it, not only into the vapor phase (as we saw above), but also into the solid (ice) phase. This will have no effect on the rate at which H_2O molecules escape from the ice into the water phase, so the system will no longer be in equilibrium and the ice will begin to melt 2.



If we wish to keep the solid from melting, the escaping tendency of molecules from the solid must be reduced. This can be accomplished by reducing the *temperature*; this lowers the escaping tendency of molecules from both phases, but it affects those in the solid more than those in the liquid, so we eventually reach the new, lower freezing point where the two quantities are again in exact balance and both phases can coexist 3.

If you prefer to think in terms of vapor pressures, you can use the same argument if you bear in mind that the vapor pressures of the solid and liquid must be the same at the freezing point. Dilution of the liquid (the solvent) by the nonvolatile solute reduces the vapor pressure of the solvent according to Raoult's law, thus reducing the temperature at which the vapor pressures of the liquid and frozen forms of the solution will be equal. As with boiling point elevation, in dilute solutions there is a simple linear relation between the freezing point depression and the molality of the solute:

$$\Delta T_f = K_f \frac{\text{moles of solute}}{\text{kg of solvent}} \tag{8.4.2}$$

Note that K_f values are all *negative*!

Table 8.4.2: Freezing point depression constants

Solvent	Normal Freezing Point (°C)	K_f (K mol^{-1} kg)
water	0.0	–1.86
acetic acid	16.7	–3.90

Solvent	Normal Freezing Point (°C)	K_f (K mol^{-1} kg)
benzene	5.5	–5.10
camphor	180	–40.0
cyclohexane	6.5	–20.2
phenol	40	–7.3

Salting Roads

The use of salt to de-ice roads is a common application of this principle. The solution formed when some of the salt dissolves in the moist ice reduces the freezing point of the ice. If the freezing point falls below the ambient temperature, the ice melts. In very cold weather, the ambient temperature may be below that of the salt solution, and the salt will have no effect. The effectiveness of a de-icing salt depends on the number of particles it releases on dissociation and on its solubility in water:

name	Formula	lowest practical T, °C
ammonium sulfate	$(NH_4)_2SO_4$	–7
calcium chloride	$CaCl_2$	–29
potassium chloride	KCl	–15
sodium chloride	NaCl	–9
urea	$(NH_2)_2CO$	–7

Automotive radiator antifreezes are mostly based on ethylene glycol, $(CH_2OH)_2$. Owing to the strong hydrogen-bonding properties of this double alcohol, this substance is miscible with water in all proportions, and contributes only a very small vapor pressure of its own. Besides lowering the freezing point, antifreeze also raises the boiling point, increasing the operating range of the cooling system. The pure glycol freezes at –12.9°C and boils at 197°C, allowing water-glycol mixtures to be tailored to a wide range of conditions.

Example 8.4.2

Estimate the freezing point of an antifreeze mixture is made up by combining one volume of ethylene glycol (MW = 62, density 1.11 g cm^{-3}) with two volumes of water.

Solution

Assume that we use 1 L of glycol and 2 L of water (the actual volumes do not matter as long as their ratios are as given.) The mass of the glycol will be 1.10 kg and that of the water will be 2.0 kg, so the total mass of the solution is 3.11 kg. We then have:

- *number of moles of glycol:* (1110 g) ÷ (62 g mol^{-1}) = 17.9 mol
- *molality of glycol:* (17.9 mol) ÷ (2.00 kg) = 8.95 mol kg^{-1}
- *freezing point depression:* ΔT_F = (–1.86 K kg^{-1} mol) × (8.95 mol kg^{-1}) = –16.6 K so the solution will freeze at about –17°C.

Any ionic species formed by dissociation will also contribute to the freezing point depression. This can serve as a useful means of determining the fraction of a solute that is dissociated.

Example 8.4.3

An aqueous solution of nitrous acid (HNO_2, MW = 47) freezes at –0.198 .C. If the solution was prepared by adding 0.100 mole of the acid to 1000 g of water, what percentage of the HNO_2 is dissociated in the solution?

Solution

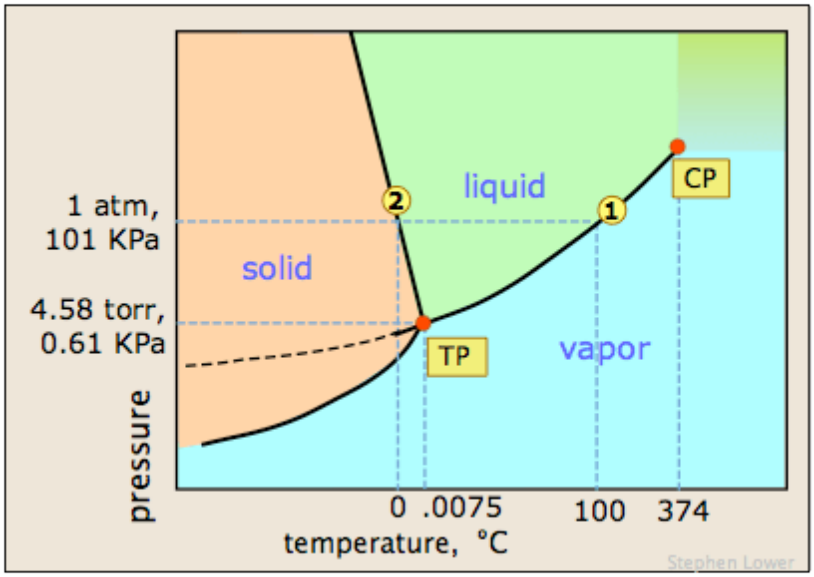
The nominal molality of the solution is (.001 mol) ÷ (1.00 kg) = 0.001 mol kg^{-1}.

But the *effective* molality according to the observed ΔT_F value is given by

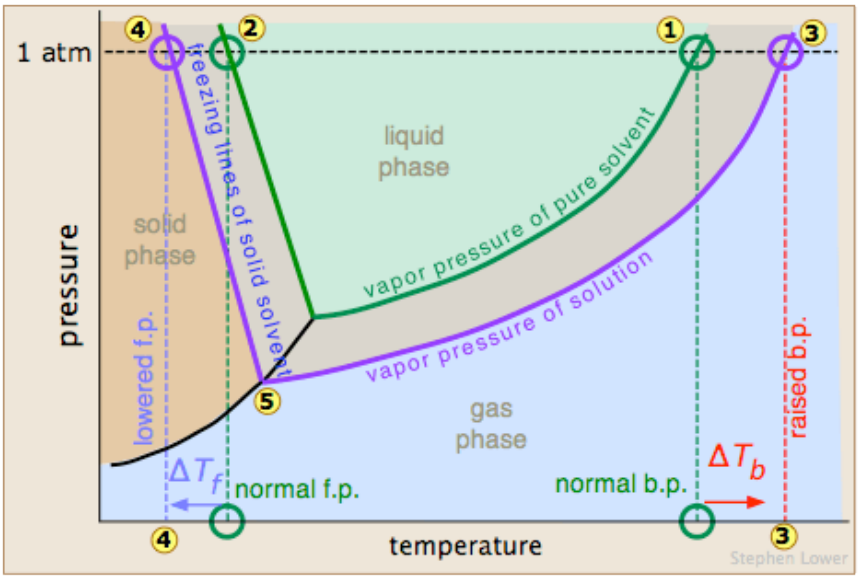
$\Delta T_F \div K_F$ = (–.198 K) ÷(–1.86 K kg mol^{-1}) = 0.106 mol kg^{-1}; this is the total number of moles of species present after the dissociation reaction $HNO_2 \rightarrow H^+ + NO^-$ has occurred. If we let $x = [H^+] = [NO_2^-]$, then by stoichiometry, $[HNO_2]$ = 0.100 - x and .106 - x = 2x and x = .0355. The fraction of HNO_2 that is dissociated is .0355 ÷ 0.100 = .355, corresponding to 35.5% dissociation of the acid.

Another Perspective of Freezing Point Depression and Boiling Point Elevation

A simple *phase diagram* can provide more insight into these phenomena. You may already be familiar with the phase map for water below.



The one shown below expands on this by plotting lines for both pure water and for its "diluted" state produced by the introduction of a non-volatile solute.



The normal boiling point of the pure solvent is indicated by point 1 where the vapor pressure curve intersects the 1-atm line — that is, where the escaping tendency of solvent molecules from the liquid is equivalent to 1 atmosphere pressure. Addition of a non-volatile solute reduces the vapor pressures to the values given by the blue line. This shifts the boiling point to the right 3, corresponding to the increase in temperature ΔT_b required to raise the escaping tendency of the H_2O molecules back up to 1 atm.

To understand freezing point depression, notice that the vapor pressure line intersects the curved black vapor pressure line of the solid (ice), which corresponds to a new triple point 5 at which all three phases (ice, water vapor, and liquid water) are in equilibrium and thus exhibit equal escaping tendencies. This point is by definition the origin of the freezing (solid-liquid) line, which intersects the 1-atm line at a reduced freezing point ΔT_f, indicated by 4.

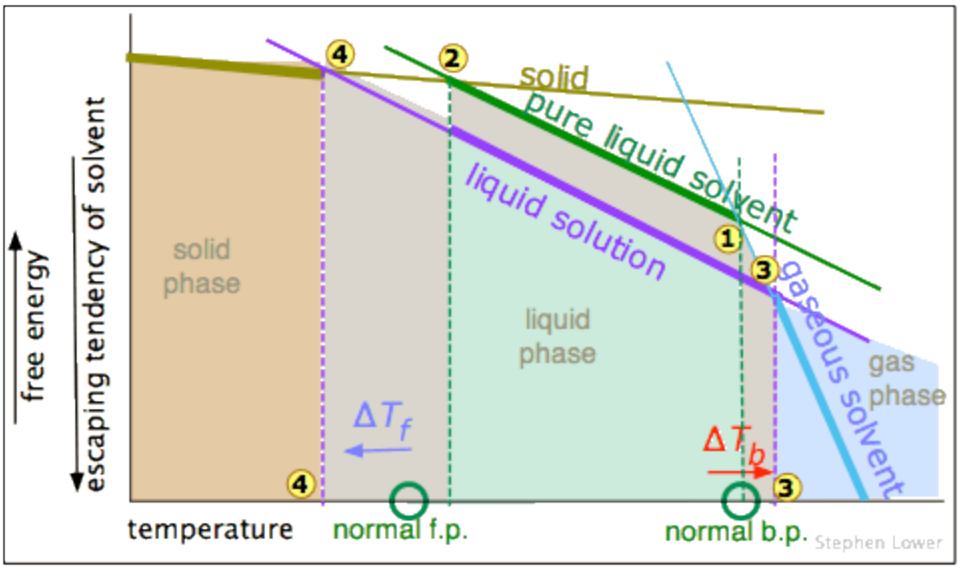
Note that the above analysis assumes that the solute is soluble only in the liquid solvent, but not in its solid form. This is generally more or less true. For example, when arctic ice forms from seawater, the salts get mostly "squeezed" out. This has the interesting effect of making the water that remains more saline, and hence more dense, causing it to sink to the bottom part of the ocean where it gets taken up by the south-flowing deep current.

A Thermodynamics Perspective on Freezing and Boiling

Those readers who have some knowledge of thermodynamics will recognize that what we have been referring to as "escaping" tendency is really a manifestation of the *Gibbs Energy*. This schematic plot shows how the *G*'s for the solid, liquid, and gas phases of a typical substance vary with the temperature.

The rule is that the phase with the most negative free energy rules.

The phase that is most stable (and which therefore is the only one that exists) is always the one having the most negative free energy (indicated here by the thicker portions of the plotted lines.) The melting and boiling points correspond to the respective temperatures where the solid and liquid 2 and liquid and vapor 1 have identical free energies.



As we saw above, adding a solute to the liquid dilutes it, making its free energy more negative, with the result that the freezing and boiling points are shifted to the left and right, respectively.

The relationships shown in these plots depend on the differing slopes of the lines representing the free energies of the phases as the temperature changes. These slopes are proportional to the entropy of each phase. Because gases have the highest entropies, the slope of the "gaseous solvent" line is much greater than that of the others. Note that this plot is not to scale.

This page titled 8.4: Colligative Properties- Boiling Point Elevation and Freezing Point Depression is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.5: Colligative Properties - Osmotic Pressure

Learning Objectives

- Define a semipermeable membrane in the context of osmotic flow.
- Explain, in simple terms, what fundamental process "drives" osmotic flow.
- What is osmotic pressure, and how is it measured?
- Osmotic pressure can be a useful means of estimating the molecular weight of a substance, particularly if its molecular weight is quite large. Explain in your own words how this works.
- What is reverse osmosis, and what is its principal application?
- Explain the role of osmotic pressure in food preservation, and give an example.
- Describe the role osmosis plays in the rise of water in plants (where is the semipermeable membrane?), and why it cannot be the only cause in very tall trees.

Osmosis is the process in which a liquid passes through a membrane whose pores permit the passage of solvent molecules but are too small for the larger solute molecules to pass through.

Semipermeable Membranes and Osmotic flow

Figure $8.5.1$ shows a simple osmotic cell. Both compartments contain water, but the one on the right also contains a solute whose molecules (represented by green circles) are too large to pass through the membrane. Many artificial and natural substances are capable of acting as semi-permeable membranes. The walls of most plant and animal cells fall into this category.

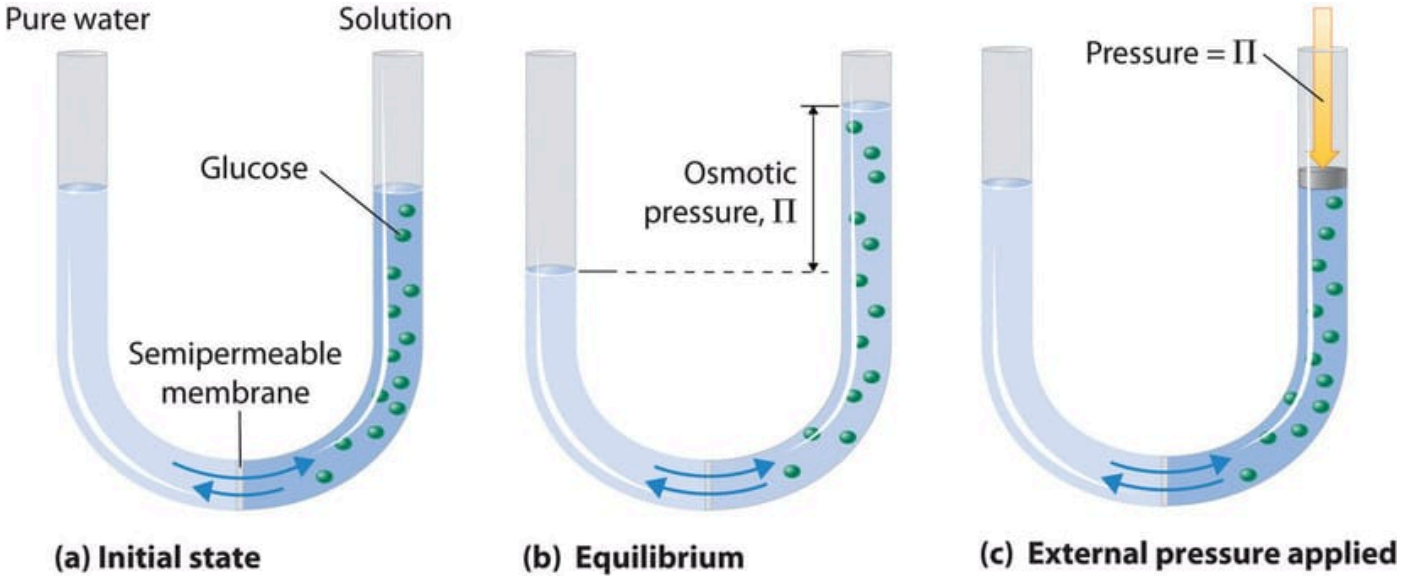


Figure $8.5.1$: Osmotic Pressure. (a) A dilute solution of glucose in water is placed in the right arm of a U-tube, and the left arm is filled to the same height with pure water; a semipermeable membrane separates the two arms. Because the flow of pure solvent through the membrane from left to right (from pure water to the solution) is greater than the flow of solvent in the reverse direction, the level of liquid in the right tube rises. (b) At equilibrium, the pressure differential, equal to the osmotic pressure of the solution (Π_{soln}), equalizes the flow rate of solvent in both directions. (c) Applying an external pressure equal to the osmotic pressure of the original glucose solution to the liquid in the right arm reverses the flow of solvent and restores the original situation.

If the cell is set up so that the liquid level is initially the same in both compartments, you will soon notice that the liquid rises in the left compartment and falls in the right side, indicating that water molecules from the right compartment are migrating through the semipermeable membrane and into the left compartment. This migration of the solvent is known as osmotic flow, or simply osmosis.

The escaping tendency of a substance from a phase increases with its concentration in the phase. What is the force that drives

the molecules through the membrane? This is a misleading question, because there is no real “force” in the physical sense other than the thermal energies all molecules possess. Osmosis is a consequence of simple statistics: the randomly directed motions of a collection of molecules will cause more to leave a region of high concentration than return to it; the escaping tendency of a substance from a phase increases with its concentration in the phase.

Diffusion and Osmotic Flow

Suppose you drop a lump of sugar into a cup of tea, without stirring. Initially there will be a very high concentration of dissolved sugar at the bottom of the cup, and a very low concentration near the top. Since the molecules are in random motion, there will be more sugar molecules moving from the high concentration region to the low concentration region than in the opposite direction. The motion of a substance from a region of high concentration to one of low concentration is known as *diffusion*. Diffusion is a consequence of a *concentration gradient* (which is a measure of the difference in escaping tendency of the substance in different regions of the solution).

There is really no special force on the individual molecules; diffusion is purely a consequence of statistics. Osmotic flow is simply diffusion of a solvent through a membrane impermeable to solute molecules. Now take two solutions of differing solvent concentration, and separate them by a semipermeable membrane (Figure 8.5.2). Being semipermeable, the membrane is essentially invisible to the solvent molecules, so they diffuse from the high concentration region to the low concentration region just as before. This flow of solvent constitutes *osmotic flow*, or *osmosis*.

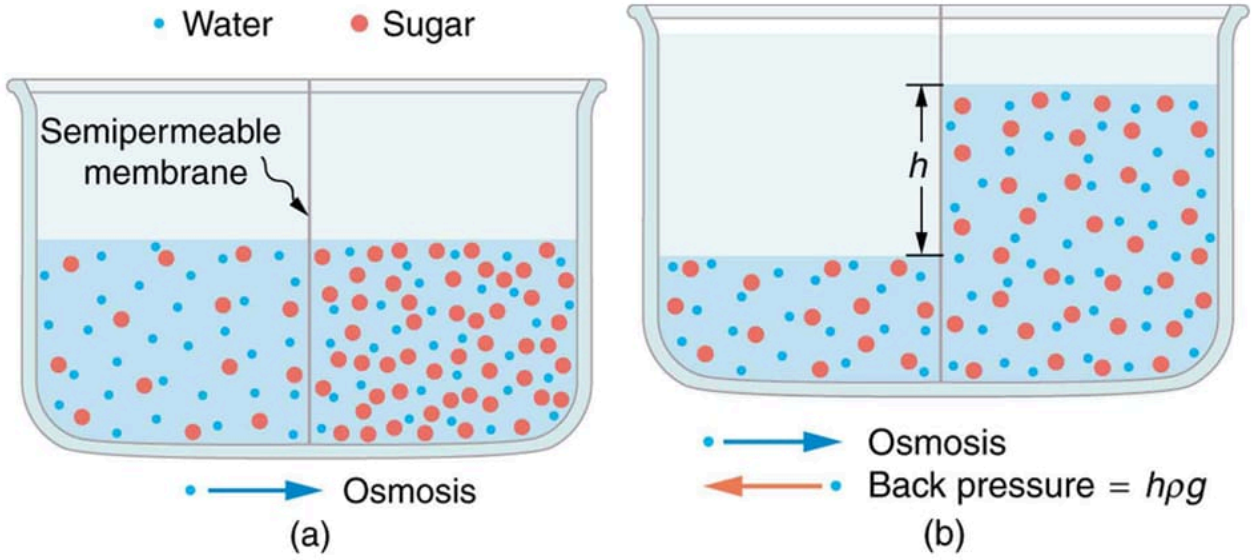


Figure 8.5.2***:*** *Osmosis osmotic flow(a) Two sugar-water solutions of different concentrations, separated by a semipermeable membrane that passes water but not sugar. Osmosis will be to the right, since water is less concentrated there. (b) The fluid level rises until the back pressure ρgh equals the relative osmotic pressure; then, the net transfer of water is zero. (CC-BY; OpenStax).*

Figure 8.5.2 shows water molecules (blue) passing freely in both directions through the semipermeable membrane, while the larger solute molecules remain trapped in the left compartment, diluting the water and reducing its escaping tendency from this cell, compared to the water in the right side. This results in a net osmotic flow of water from the right side which continues until the increased hydrostatic pressure on the left side raises the escaping tendency of the diluted water to that of the pure water at 1 atm, at which point osmotic equilibrium is achieved.

> Osmotic flow is simply diffusion of a solvent through a membrane impermeable to solute molecules.

In the absence of the semipermeable membrane, diffusion would continue until the concentrations of all substances are uniform throughout the liquid phase. With the semipermeable membrane in place, and if one compartment contains the pure solvent, this can never happen; no matter how much liquid flows through the membrane, the solvent in the right side will always be more concentrated than that in the left side. Osmosis will continue indefinitely until we run out of solvent, or something else stops it.

Osmotic equilibrium and osmotic pressure

One way to stop osmosis is to raise the hydrostatic pressure on the solution side of the membrane. This pressure squeezes the solvent molecules closer together, raising their escaping tendency from the phase. If we apply enough pressure (or let the pressure build up by osmotic flow of liquid into an enclosed region), the escaping tendency of solvent molecules from the solution will eventually rise to that of the molecules in the pure solvent, and osmotic flow will case. The pressure required to achieve *osmotic equilibrium* is known as the *osmotic pressure*. Note that the osmotic pressure is the pressure required to *stop* osmosis, not to sustain it.

Osmotic pressure is the pressure required to *stop* osmotic flow It is common usage to say that a solution “has” an osmotic pressure of "*x* atmospheres". It is important to understand that this means nothing more than that a pressure of this value must be applied to the solution to **prevent** flow of pure solvent into this solution through a semipermeable membrane separating the two liquids.

Osmotic Pressure and Solute Concentration

The Dutch scientist Jacobus Van't Hoff (1852-1911) was one of the giants of physical chemistry. He discovered this equation after a chance encounter with a botanist friend during a walk in a park in Amsterdam; the botanist had learned that the osmotic pressure increases by about 1/273 for each degree of temperature increase. van’t Hoff immediately grasped the analogy to the ideal gas law. The osmotic pressure Π of a solution containing n moles of solute particles in a solution of volume V is given by the *van 't Hoff equation*:

$$\Pi = \frac{nRT}{V} \tag{8.5.1}$$

in which

- R is the gas constant (0.0821 L atm mol^{-1} K^{-1}) and
- T is the absolute temperature.

In contrast to the need to employ solute molality to calculate the effects of a non-volatile solute on changes in the freezing and boiling points of a solution, we can use solute molarity to calculate osmotic pressures.

Note that the fraction n/V corresponds to the molarity (M) of a solution of a non-dissociating solute, or to twice the molarity of a totally-dissociated solute such as $NaCl$. In this context, molarity refers to the summed total of the concentrations of all solute species. Hence, Equation 8.5.1 can be expressed as

$$\Pi = MRT \tag{8.5.2}$$

Recalling that Π is the Greek equivalent of *P*, the re-arranged form $\Pi V = nRT$ of the above equation should look familiar. Much effort was expended around the end of the 19th century to explain the similarity between this relation and the *ideal gas law*, but in fact, the Van’t Hoff equation turns out to be only a very rough approximation of the real osmotic pressure law, which is considerably more complicated and was derived after van 't Hoff's formulation. As such, this equation gives valid results only for extremely dilute ("ideal") solutions.

According to the Van't Hoff equation, an ideal solution containing 1 mole of dissolved particles per liter of solvent at 0° C will have an osmotic pressure of 22.4 atm.

Example 8.5.1

Sea water contains dissolved salts at a total ionic concentration of about 1.13 mol L^{-1}. What pressure must be applied to prevent osmotic flow of pure water into sea water through a membrane permeable only to water molecules?

Solution

This is a simple application of Equation 8.5.2.

$$\begin{aligned}\Pi &= MRT \\ &= (1.13\ mol/L)(0.0821\ L\,atm\,mol^{-1}\ K^{-1})(298\ K) \\ &= 27.6\ atm\end{aligned}$$

Molecular Weight Determination by Osmotic Pressure

Since all of the colligative properties of solutions depend on the concentration of the solvent, their measurement can serve as a convenient experimental tool for determining the concentration, and thus the molecular weight, of a solute. Osmotic pressure is especially useful in this regard, because a small amount of solute will produce a much larger change in this quantity than in the boiling point, freezing point, or vapor pressure. even a 10^{-6} molar solution would have a measurable osmotic pressure. Molecular weight determinations are very frequently made on proteins or other high molecular weight polymers. These substances, owing to their large molecular size, tend to be only sparingly soluble in most solvents, so measurement of osmotic pressure is often the only practical way of determining their molecular weights.

Example 8.5.2: average molecular weight

The osmotic pressure of a benzene solution containing 5.0 g of polystyrene per liter was found to be 7.6 torr at 25°C. Estimate the average molecular weight of the polystyrene in this sample.

Solution:

osmotic pressure:

$$\begin{aligned}\Pi &= \frac{7.6\,torr}{760\,torr\,atm^{-1}} \\ &= 0.0100\,atm\end{aligned}$$

Using the form of the van 't Hoff equation (Equation 8.5.1), *PV* = *nRT*, the number of moles of polystyrene is

$$n = (0.0100\ \text{atm})(1\ \text{L}) \div (0.0821\ \text{L atm mol}^{-1}\ \text{K}^{-1})(298\ \text{K}) = 4.09 \times 10^{-4}\ \text{mol}$$

Molar mass of the polystyrene:

$$(5.0\ \text{g}) \div (4.09 \times 10^{-4}\ \text{mol}) = \mathbf{12200\ g\ mol^{-1}}.$$

The experiment to demonstrate this is quite simple: pure solvent is introduced into one side of a cell that is separated into two parts by a semipermeable membrane. The polymer solution is placed in the other side, which is enclosed and connected to a manometer or some other kind of pressure gauge. As solvent molecules diffuse into the solution cell the pressure builds up; eventually this pressure matches the osmotic pressure of the solution and the system is in osmotic equilibrium. The osmotic pressure is read from the measuring device and substituted into the van't Hoff equation to find the number of moles of solute.

This page titled 8.5: Colligative Properties - Osmotic Pressure is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

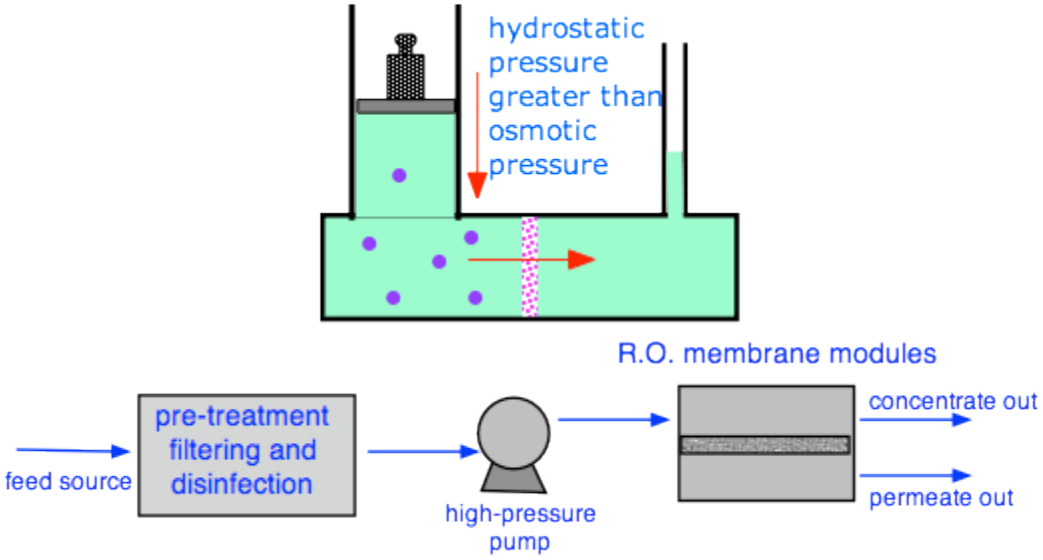
8.6: Reverse Osmosis

Learning Objectives

- What is reverse osmosis, and what is its principal application?
- Explain the role of osmotic pressure in food preservation, and give an example.
- Describe the role osmosis plays in the rise of water in plants (where is the semipermeable membrane?), and why it cannot be the only cause in very tall trees.

$$E = mc^2 \tag{8.6.1}$$

If it takes a pressure of Π atm to bring about osmotic equilibrium, then it follows that applying a hydrostatic pressure greater than this to the high-solute side of an osmotic cell will force water to flow back into the fresh-water side. This process, known as *reverse osmosis*, is now the major technology employed to desalinate ocean water and to reclaim "used" water from power plants, runoff, and even from sewage. It is also widely used to deionize ordinary water and to purify it for for industrial uses (especially beverage and food manufacture) and drinking purposes.



Pre-treatment commonly employs activated-carbon filtration to remove organics and chlorine (which tends to damage RO membranes). Although bacteria are unable to pass through semipermeable membranes, the latter can develop pinhole leaks, so some form of disinfection is often advised. The efficiency and cost or RO is critically dependent on the properties of the semipermeable membrane.

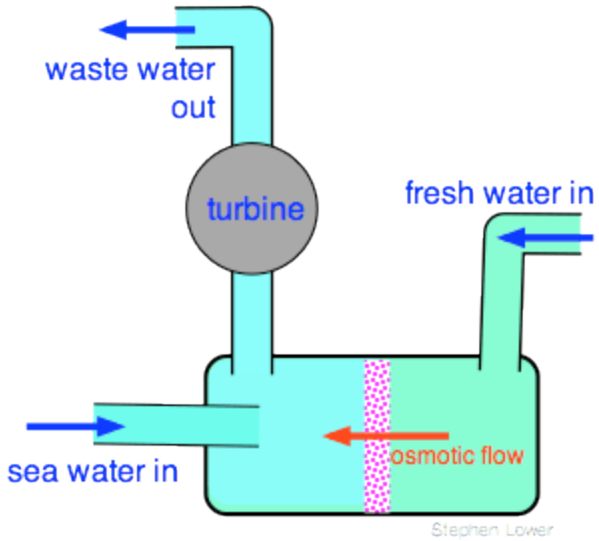


(left) Large-scale RO plants use multiple membrane cartridges. This one, in Tampa Bay FL, supplies desalinated drinking water to 2.4 million residents. (right) Many smaller RO units, suitable for home use, can fit under a kitchen sink.

Osmotic Generation of Electric Power

The osmotic pressure of seawater is almost 26 atm. Since a pressure of 1 atm will support a column of water 10.6 m high, this means that osmotic flow of fresh water through a semipermeable membrane into seawater could in principle support a column of the latter by 26 x 10.3 = 276 m (904 ft)!

So imagine an osmotic cell in which one side is supplied with fresh water from a river, and the other side with seawater. Osmotic flow of fresh water into the seawater side forces the latter up through a riser containing a turbine connected to a generator, thus providing a constant and fuel-less source of electricity. The key component of such a scheme, first proposed by an Israeli scientist in 1973 and known as *pressure-retarded osmosis* (PRO) is of course a semipermeable membrane capable of passing water at a sufficiently high rate.



The world's first experimental PRO plant was opened in 2009 in Norway. Its capacity is only 4 kW, but it serves as proof-in-principle of a scheme that is estimated capable of supplying up to 2000 terawatt-hours of energy worldwide. The semipermeable membrane operates at a pressure of about 10 atm and passes 10 L of water per second, generating about 1 watt per m^2 of membrane. PRO is but one form of *salinity gradient power* that depends on the difference between the salt concentrations in different bodies of water.

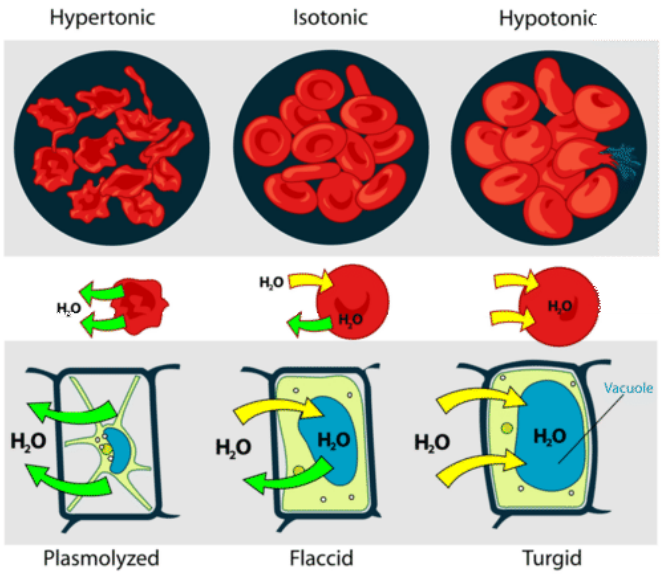
> 1 atm is equivalent to 1034 g cm^{-2}, so from the density of water we get (1034 g cm^{-2}) ÷ (1 g cm^{-3}) = 1034 cm = 10.3 m.

Osmosis in Biology and Physiology

Because many plant and animal cell membranes and tissues tend to be permeable to water and other small molecules, osmotic flow plays an essential role in many physiological processes.

Normal saline solution

The interiors of cells contain salts and other solutes that dilute the intracellular water. If the cell membrane is permeable to water, placing the cell in contact with pure water will draw water into the cell, tending to rupture it. This is easily and dramatically seen if red blood cells are placed in a drop of water and observed through a microscope as they burst. This is the reason that "normal saline solution", rather than pure water, is administered in order to maintain blood volume or to infuse therapeutic agents during medical procedures.



In order to prevent irritation of sensitive membranes, one should always add some salt to water used to irrigate the eyes, nose, throat or bowel. Normal saline contains 0.91% w/v of sodium chloride, corresponding to 0.154 M, making its osmotic pressure close to that of blood.

Food preservation

The drying of fruit, the use of sugar to preserve jams and jellies, and the use of salt to preserve certain meats, are age-old methods of preserving food. The idea is to reduce the water concentration to a level below that in living organisms. Any bacterial cell that wanders into such a medium will have water osmotically drawn out of it, and will die of dehydration. A similar effect is noticed by anyone who holds a hard sugar candy against the inner wall of the mouth for an extended time; the affected surface becomes dehydrated and noticeably rough when touched by the tongue.

In the food industry, what is known as *water activity* is measured on a scale of 0 to 1, where 0 indicates no water and 1 indicates all water. Food spoilage micro-organisms, in general, are inhibited in food where the water activity is below 0.6. However, if the pH of the food is less than 4.6, micro-organisms are inhibited (but not immediately killed] when the water activity is below 0.85.

Diarrhea

The presence of excessive solutes in the bowel draws water from the intestinal walls, giving rise to diarrhea. This can occur when a food is eaten that cannot be properly digested (as, for example, milk in lactose-intolerant people). The undigested material contributes to the solute concentration, raising its osmotic pressure. The situation is made even worse if the material undergoes bacterial fermentation which results in the formation of methane and carbon dioxide, producing a frothy discharge.

Water Transport in Plants

Osmotic flow plays an important role in the transport of water from its source in the soil to its release by transpiration from the leaves, it is helped along by hydrogen-bonding forces between the water molecules. Capillary rise is not believed to be a significant factor.

Water enters the roots via osmosis, driven by the low water concentration inside the roots that is maintained by both the active [non-osmotic] transport of ionic nutrients from the soil and by the supply of sugars that are photosynthesized in the leaves. This generates a certain amount of *root pressure* which sends the water molecules on their way up through the vascular channels of the stem or trunk. But the maximum root pressures that have been measured can push water up only about 20 meters, whereas the tallest trees exceed 100 meters. Root pressure can be the sole driver of water transport in short plants, or even in tall ones such as trees that are not in leaf. Anyone who has seen apparently tender and fragile plants pushing their way up through asphalt pavement cannot help but be impressed!

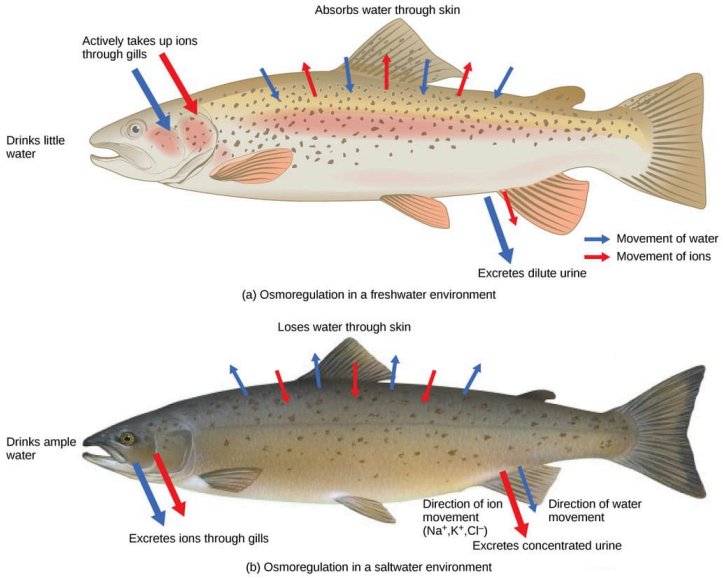
But when taller plants are actively transpiring (losing water to the atmosphere], osmosis gets a boost from what plant physiologists call *cohesion tension* or *transpirational pull*. As each H_2O molecule emerges from the opening in the leaf it pulls along the chain of molecules beneath it. So hydrogen-bonding is no less important than osmosis in the overall water transport process. If the soil becomes dry or saline, the osmotic pressure outside the root becomes greater than that inside the plant, and the plant suffers from “water tension”, i.e., wilting.

The following section is a bit long, but for those who are interested in biology it offers a beautiful example of how the constraints imposed by osmosis have guided the evolution of ocean-living creatures into fresh-water species . It concerns ammonia NH_3, a product of protein metabolism that is generated within all animals, but is highly toxic and must be eliminated.

Marine invertebrates (those that live in seawater) are covered in membranes that are fairly permeable to water and to small molecules such as ammonia. So water can diffuse in either direction as required, and ammonia can diffuse out as quickly as it forms. Nothing special here.

Invertebrates that live in fresh water do have problem: the salt concentrations within their bodies are around 1%, much greater than in fresh water. For this reason they have evolved surrounding membranes that are largely impermeable to salts (to prevent their diffusion out of the body) and to water (to prevent osmotic flow in.) But these organisms must also be able to exchange oxygen and carbon dioxide with their environment. The special respiratory organs (gills) that mediate this process, as a consequence of being permeable to these two gases, will also allow water molecules (whose sizes are comparable to those of the respiratory gases) to pass through. In order to protect fresh-water invertebrates from the disastrous effects of unlimited water inflow through the gill membranes, these animals possess special excretory organs that expel excess water back into the environment. Thus in such animals, there is a constant flow of water passing through the body. Ammonia and other substances that need to be excreted are taken up by this stream which constitutes a continual flow of dilute urine.

Fishes fall into two general classes: most fish have bony skeletons and are known as teleosts. Sharks and rays have cartilage instead of bones, and are called elasmobranchs. For the teleosts that live in fresh water, the situation is very much the same as with fresh-water invertebrates; they take in and excrete water continuously. The fact that an animal lives in the water does not mean that it enjoys an unlimited supply of water. Marine teleosts have a more difficult problem. Their gills are permeable to water, as are those of marine invertebrates. But the salt content of seawater (about 3%), being higher than the about 1% in the fish’s blood, would draw water out of the fish. Thus these animals are constantly losing water, and would be liable to desiccation if water could freely pass out of their gills. Some does, of course, and with it goes most of its nitrogen in the form of NH_3.



Salmon physiology responds to freshwater and seawater to maintain osmotic balance. Fish are osmoregulators, but must use different mechanisms to survive in (a) freshwater or (b) saltwater environments. Image used withpermission (CC BY 3.0; OpenStax).

Thus most of the waste nitrogen exits not through the usual excretory organs as with most vertebrates, but through the gills. But in order to prevent excessive loss of water, the gills have reduced permeability to this water, and with it, to comparably-sized

NH_3. So in order to prevent ammonia toxicity, the remainder of it is converted to a non-toxic substance (trimethylamine oxide $(CH_3)_3NO$) which is excreted via the kidneys.

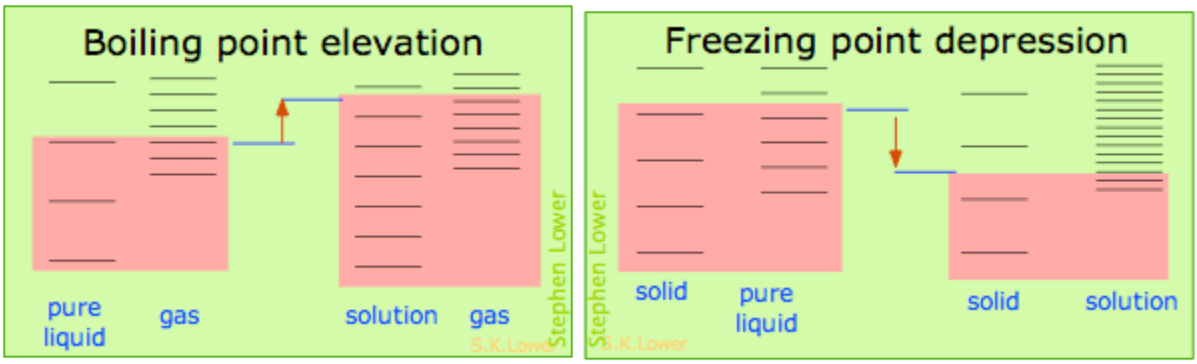
The marine elasmobranchs solve the loss-of-water problem in another way: they convert waste ammonia to urea $(NH_3)_2CO$ which is highly soluble and non-toxic. Their kidneys are able to control the quantity of urea excreted so that their blood retains about 2-2.5 percent of this substance. Combined with the 1 percent of salts and other substances in their blood, this raises the osmotic pressure within the animal to slightly above that of seawater, Thus the same mechanism that protects them from ammonia poisoning also ensures them an adequate water supply.

This page titled 8.6: Reverse Osmosis is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.7: Colligative Properties and Entropy

All four solution effects (Reduced Vapor Pressure, Freezing Point depression, Boiling Point Elevation, Osmotic pressure) result from “dilution” of the solvent by the added solute. Because of this commonality they are referred to as colligative properties (Lat. *co ligare*, connected to.) The key role of the solvent concentration is obscured by the greatly-simplified expressions used to calculate the magnitude of these effects, in which only the solute concentration appears. The details of how to carry out these calculations and the many important applications of colligative properties are covered elsewhere. Our purpose here is to offer a more complete explanation of *why* these phenomena occur.

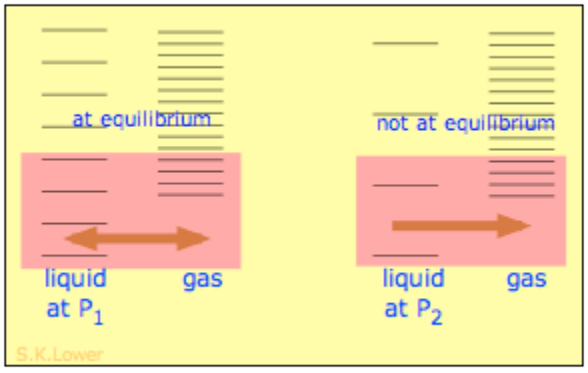
Basically, these all result from the effect of **dilution of the solvent** on its entropy, and thus in the increase in the density of energy states of the system in the solution compared to that in the pure liquid. Equilibrium between two phases (liquid-gas for boiling and solid-liquid for freezing) occurs when the energy states in each phase can be populated at equal densities. The temperatures at which this occurs are depicted by the shading.



- Dilution of the solvent adds new energy states to the liquid, but does not affect the *vapor phase*. This raises the temperature required to make equal numbers of microstates accessible in the two phases.
- Dilution of the solvent adds new energy states to the liquid, but does not affect the *solid phase*. This reduces the temperature required to make equal numbers of states accessible in the two phases.

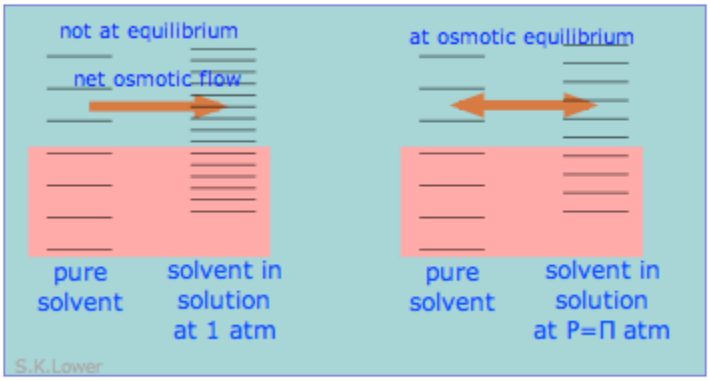
Effects of pressure on the entropy: Osmotic Pressure

When a liquid is subjected to hydrostatic pressure— for example, by an inert, non-dissolving gas that occupies the vapor space above the surface, the vapor pressure of the liquid is raised. The pressure acts to compress the liquid very slightly, effectively narrowing the potential energy well in which the individual molecules reside and thus increasing their tendency to escape from the liquid phase. (Because liquids are not very compressible, the effect is quite small; a 100-atm applied pressure will raise the vapor pressure of water at 25°C by only about 2 torr.) In terms of the entropy, we can say that the applied pressure reduces the dimensions of the "box" within which the principal translational motions of the molecules are confined within the liquid, thus reducing the density of energy states in the liquid phase.



Applying hydrostatic pressure to a liquid increases the spacing of its microstates, so that the number of energetically accessible states in the gas, although unchanged, is relatively greater— thus increasing the tendency of molecules to escape into the vapor phase. In terms of free energy, the higher pressure raises the free energy of the liquid, but does not affect that of the gas phase.

This phenomenon can explain **osmotic pressure**. Osmotic pressure, students must be reminded, is not what drives osmosis, but is rather the hydrostatic pressure that must be applied to the more concentrated solution (more dilute solvent) in order to *stop* osmotic flow of solvent into the solution. The effect of this pressure Π is to slightly increase the spacing of solvent energy states on the high-pressure (dilute-solvent) side of the membrane to match that of the pure solvent, restoring osmotic equilibrium.



Osmotic pressure does not drive osmosis, but is rather the hydrostatic pressure that must be applied to the more concentrated solution (more dilute solvent) in order to **stop** osmotic flow of solvent into the solution.

This page titled 8.7: Colligative Properties and Entropy is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.8: Ideal vs. Real Solutions

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Describe the physical reasons that a binary liquid solution might exhibit non-ideal behavior

The popular liquor vodka consists mainly of ethanol (ethyl alcohol) and water in roughly equal portions. Ethanol and water both have substantial vapor pressures, so both components contribute to the total pressure of the gas phase above the liquid in a closed container of the two liquids. One might expect the vapor pressure of a solution of ethanol and water to be directly proportional to the sums of the values predicted by Raoult's law for the two liquids individually, but in general, this does not happen. The reason for this can be understood if you recall that Raoult's law reflects a single effect: the smaller proportion of vaporizable molecules (and thus their reduced escaping tendency) when the liquid is diluted by otherwise "inert" (non-volatile) substance.

Ideal Solutions

There are some solutions whose components follow Raoult's law quite closely. An example of such a solution is one composed of hexane C_6H_{14} and heptane C_7H_{16}. The total vapor pressure of this solution varies in a straight-line manner with the mole fraction composition of the mixture.

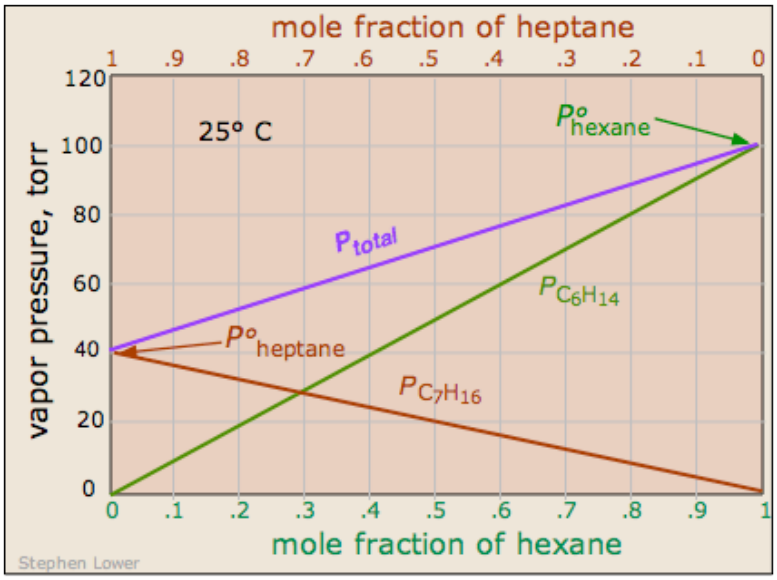


Figure 8.8.1: Raoult's law plot for a mixture of hexane and heptane.

Note that the mole fraction scales at the top and bottom run in opposite directions, since by definition,

$$\chi_{hexane} = 1 - \chi_{heptane} \tag{8.8.1}$$

If this solution behaves ideally, then is the sum of the Raoult's law plots for the two pure compounds:

$$P_{total} = P_{heptane} + P_{hexane} \tag{8.8.2}$$

An *ideal solution* is one whose vapor pressure follows Raoult's law throughout its range of compositions. Experience has shown solutions that approximate ideal behavior are composed of molecules having very similar structures. Thus hexane and heptane are both linear hydrocarbons that differ only by a single $-CH_2$ group. This provides a direct clue to the underlying cause of non-

ideal behavior in solutions of volatile liquids. In an ideal solution, the interactions are there, but they are all energetically identical. Thus in an ideal solution of molecules A and B, A—A and B—B attractions are the same as A—B attractions. This is the case only when the two components are chemically and structurally very similar.

The ideal solution differs in a fundamental way from the definition of an ideal gas, defined as a hypothetical substance that follows the ideal gas law. The kinetic molecular theory that explains ideal gas behavior assumes that the molecules occupy no space and that intermolecular attractions are totally absent.

The definition of an ideal gas is clearly inapplicable to liquids, whose volumes directly reflect the volumes of their component molecules. And of course, the very ability of the molecules to form a condensed phase is due to the attractive forces between the molecules. So the most we can say about an ideal solution is that the attractions between its all of its molecules are identical — that is, A-type molecules are as strongly attracted to other A molecules as to B-type molecules. Ideal solutions are perfectly democratic: there are no favorites.

Real Solutions

Real solutions are more like real societies, in which some members are "more equal than others." Suppose, for example, that unlike molecules are more strongly attracted to each other than are like molecules. This will cause A–B pairs that find themselves adjacent to each other to be energetically more stable than A–A and B–B pairs. At compositions in which significant numbers of both kind of molecules are present, their tendencies to escape the solution — and thus the vapor pressure of the solution, will fall below what it would be if the interactions between all the molecules were identical. This gives rise to a *negative deviation* from Raoult's law. The chloroform-acetone system, illustrated above, is a good example.

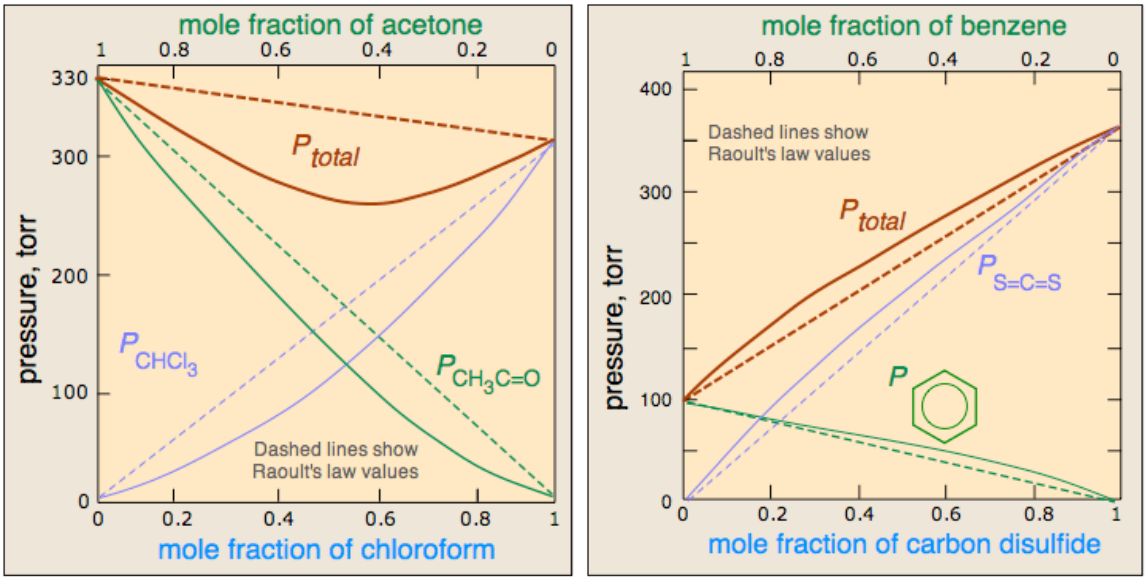


Figure 8.8.2: Systems with negative deviation (left) and positive deviation (right) of Raoult's law. (CC BY 4.0; Stephen Lower via Chem1)

Conversely, if like molecules of each kind are more attracted to each other than to unlike ones, then the molecules that happen to be close to their own kind will be stabilized. At compositions approaching 50 mole-percent, A and B molecules near each other will more readily escape the solution, which will therefore exhibit a higher vapor pressure than would otherwise be the case. It should not be surprising molecules as different as benzene and CS_2 should interact more strongly with their own kind, hence the positive deviation illustrated here.

You will recall that all gases approach ideal behavior as their pressures approach zero. In the same way, as the mole fraction of either component approaches unity, the behavior of the solution approaches ideality. This is a simple consequence of the fact that at these limits, each molecule is surrounded mainly by its own kind, and the few A-B interactions will have little effect. Raoult's law is therefore a limiting law:

$$P_i = \lim_{x_i \to 0} P^o \chi_i \tag{8.8.3}$$

it gives the partial pressure of a substance in equilibrium with the solution more and more closely as the mole fraction of that substance approaches unity.

This page titled 8.8: Ideal vs. Real Solutions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.9: Distillation

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Sketch out a typical *boiling point diagram* for a binary liquid solution, and use this to show how a simple one-stage *distillation* works.
- Explain the role of the *lever rule* in fractional distillation
- Describe the purpose and function of a *fractionating column*
- Sketch out boiling point diagrams for high- and low-boiling *azeotropes*
- Describe the role of distillation in *crude oil refining*, and explain, in a very general way, how further processing is used to increase the yield of gasoline motor fuel.

Distillation is a process whereby a mixture of liquids having different vapor pressures is separated into its components. At first one might think that this would be quite simple: if you have a solution consisting of liquid A that boils at 50°C and liquid B with a boiling point of 90°C, all that would be necessary would be to heat the mixture to some temperature between these two values; this would boil off all the A (whose vapor could then be condensed back into pure liquid A), leaving pure liquid B in the pot. But that overlooks that fact that these liquids will have substantial vapor pressures at *all* temperatures, not only at their boiling points.

Vapor Pressure vs. Composition Phase Diagrams

To fully understand distillation, we will consider an ideal binary liquid mixture of A and B. If the mole fraction of A in the mixture is χ_A, then by the definition of mole fraction, that of B is

$$\chi_B = 1 - \chi_A \tag{8.9.1}$$

Since distillation depends on the different vapor pressures of the components to be separated, let's first consider the vapor pressure vs. composition plots (Figure 8.9.1) for a hypothetical mixture at some arbitrary temperature at which both liquid and gas phases can exist, depending on the total pressure.

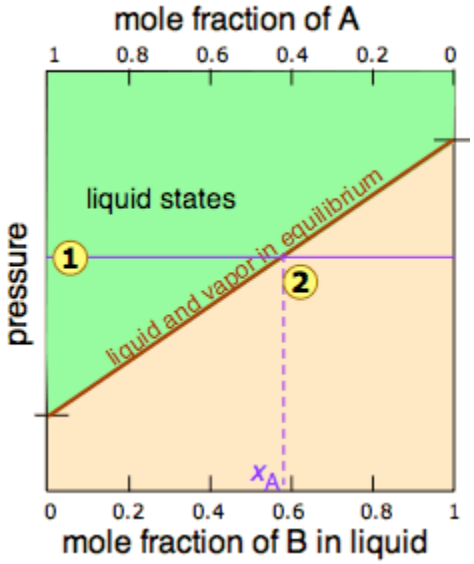


Figure 8.9.1: Liquid-Vapor Equilibrium curve.

In Figure 8.9.2, all states of the system (i.e., combinations of pressure and composition) in which the solution exists solely as a **liquid** are shaded in green. Since liquids are more stable at higher pressures, these states occupy the upper part of the diagram.

At any given total vapor pressure such as at 1, the composition of the vapor in equilibrium with the liquid (designated by x_A) corresponds to the intercept with the diagonal equilibrium line at 2. The diagonal line is just an expression of the linearity between vapor pressure and composition according to Raoult's law.

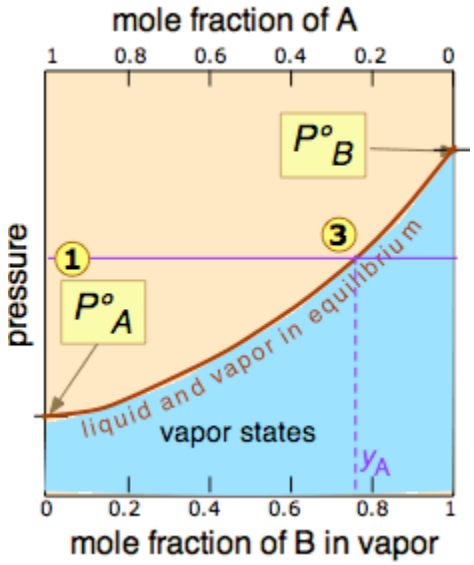


Figure 8.9.2: Liquid-Vapor Equilibrium curve.

3

of the pressure line

1

with the equilibrium curve defines the mole fractions of

and

present in the vapor. (Note that mole fractions of gases, which we are dealing with here, are conventionally represented by

y

, hence

y_A

and

y_B

). The curvature of the equilibrium line arises from the need to combine

Raoult's law

with

Dalton's law of partial pressures

which applies to gaseous mixtures.

Figure 8.9.3.

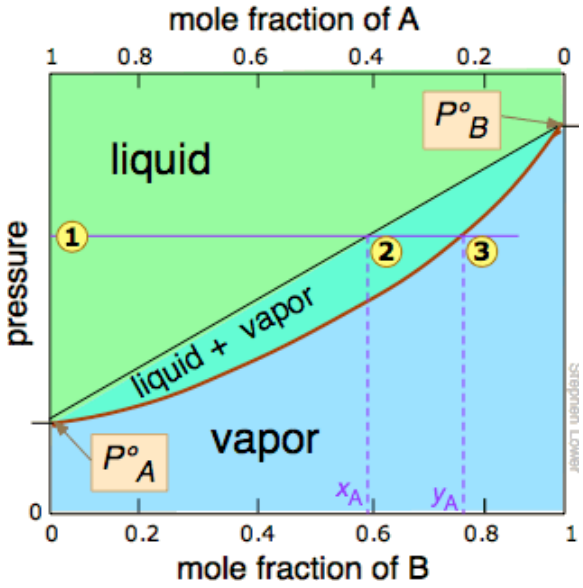


Figure 8.9.3

The two liquid-vapor equilibrium lines (one curved, the other straight) now enclose an area in which liquid and vapor can **coexist**; outside of this region, the mixture will consist entirely of liquid or of vapor. At this particular pressure (1), the intercept (2) with the upper boundary of the two-phase region gives the mole fractions of A and B in the liquid phase, while the intercept (3) with the lower boundary gives the mole fractions of the two components in the vapor.

Exercise 8.9.1

Take a moment to study Figure 8.9.5 and to confirm that

- because both intercepts occur on equilibrium lines, they describe the *compositions* of the liquid and vapor that can simultaneously exist;
- the compositions of the vapor and liquid are not the same;
- in the vapor, the mole fraction of B (the more volatile component of the solution) is greater than that in the liquid;
- in the liquid, the mole fraction of A (the less volatile component) is smaller than that of the vapor.

The vapor in equilibrium with a solution of two or more liquids is **always** richer in the more volatile component.

Temperatures vs. Composition Phase Diagrams (Boiling Point Diagrams)

The rule shown above suggests that if we heat a mixture sufficiently to bring its total vapor pressure into the two-phase region, we will have a means of separating the mixture into two portions which will be enriched in the more volatile and less volatile components respectively. This is the principle on which distillation is based. But what temperature is required to achieve this? Again, we will spare you the mathematical details, but it is possible to construct a plot similar to the Figure 8.9.4 except that the vertical axis represents *temperature* rather than pressure. This kind of plot is called a *boiling point diagram*.

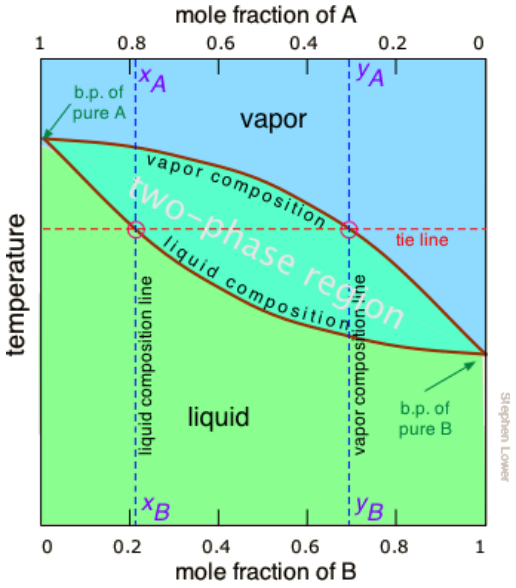


Figure 8.9.4: Boiling point diagram of a volatile two-component system.

Some important things to understand about Figure 8.9.6:

- The shape of the two-phase region is bi-convex, as opposed to the half-convex shape of the pressure-composition plot.
- The slope of the two-phase region is opposite to what we saw in the previous plot, and the areas corresponding to the single-phase regions are reversed. This simply reflects the fact that liquids having a higher vapor pressure boil at lower temperatures, and *vice versa*.
- The horizontal line that defines the temperature is called the *tie line*. Its intercepts with the two equilibrium curves specify the composition of the liquid and vapor in equilibrium with the mixture at the given temperature.
- The vapor composition line is also known as the *dew point line* — the temperature at which condensation begins on cooling.
- The liquid composition line is also called the *bubble point line* — the temperature at which boiling begins on heating.

The tie line shown in Figure 8.9.6 is for one particular temperature. But when we heat a liquid to its boiling point, the composition will change as the more volatile component (B in these examples) is selectively removed as vapor. The remaining liquid will be enriched in the less volatile component, and its boiling point will consequently rise. To understand this process more thoroughly, let us consider the situation at several points during the distillation of an equimolar solution of A and B.

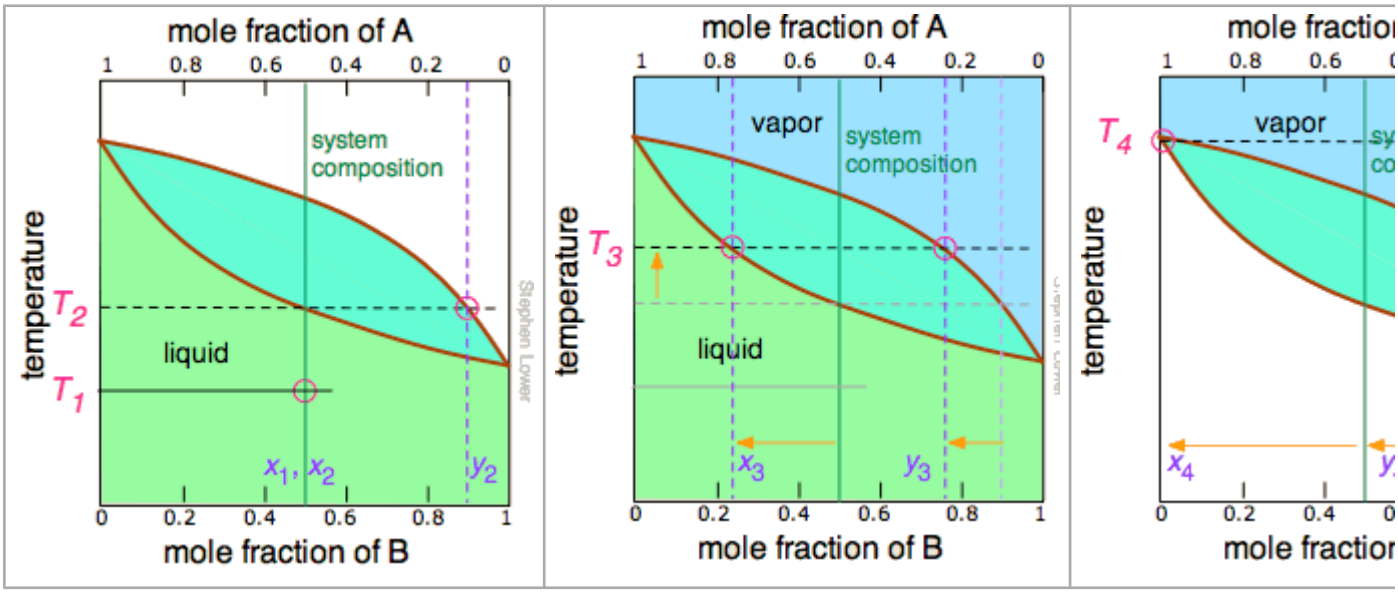


Figure 8.9.5A: We begin with the liquid at T_1, below its boiling point. When the temperature rises to T_2, boiling begins and the first vapor (and thus the first drop of condensate) will have the composition y_2.	***Figure*** 8.9.5B**:** As the more volatile component B is boiled off, the liquid and vapor/condensate compositions shift to the left (orange arrows).	***Figure*** 8.9.5C**:** At T_4, the las disappears. The system is nc of composition

Notice that the vertical green *system composition* line remains in the same location in the three plots because the "system" is defined as consisting of both the liquid in the "pot" and that in the receiving container which was condensed from the vapor. The principal ideas you should take away from this are that

- distillation can never completely separate two volatile liquids;
- the composition of the vapor and thus of the condensed distillate changes continually as each drop forms, starting at y_2 and ending at y_4 in this example;
- if the liquid is completely boiled away, the composition of the distillate will be the same as that of the original solution.

Laboratory Distillation Setup

The apparatus used for a simple laboratory batch distillation is shown here. The purpose of the thermometer is to follow the progress of the distillation; as a rough rule of thumb, the distillation should be stopped when the temperature rises to about half-way between the boiling points of the two pure liquids, which should be at least 20-30 C° apart (if they are closer, then *fractional distillation*, described below, becomes necessary).

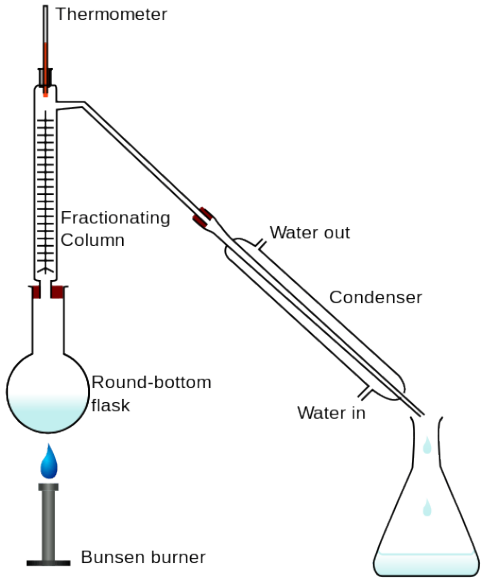


Figure 8.9.6: Fractional distillation setup. An Erlenmeyer flask is used as a receiving flask. Here the distillation head and fractionating column are combined in one piece. from Wikipedia

Condensers are available in a number of types. The simple *Liebig condenser* shown above is the cheapest and therefore most commonly used in student laboratories. Several other classic designs increase the surface area separating the vapor/distillate and cooling water, leading to greater heat exchange efficiency and allowing higher throughput.

Fractional Distillation

Although distillation can never achieve complete separation of volatile liquids, it can in principal be carried out in such a way that any desired degree of separation can be achieved if the solution behaves ideally and one is willing to go to the trouble. The general procedure is to distill only a fraction of the liquid, the smaller the better. The condensate, now enriched in the more volatile component, is then collected and re-distilled (again, only a small fraction), thus obtaining a condensate even-more-enriched in the more volatile component. If we repeat this sequence many times, we can eventually obtain almost-pure, if minute,

samples of the two components.

But since this would hardly be practical, there is a better way. In order to understand it, you need to know about the *lever rule*, which provides a simple way of determining the relative *quantities* (not just the compositions) of two phases in equilibrium. The lever rule is easily derived from **Raoult's** and **Dalton's laws**, but we will simply illustrate it graphically (Figure 8.9.7). The plot shows the boiling point diagram of a simple binary mixture of composition 1. At the temperature corresponding to the tie line, the composition of the liquid corresponds to 2 and that of the vapor to 3.

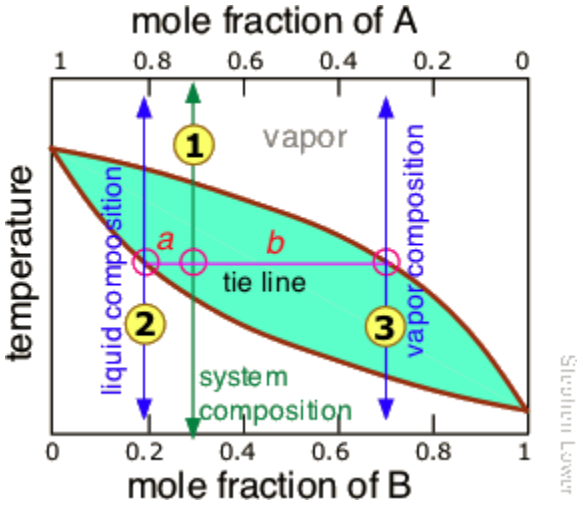


Figure 8.9.7:

So now for the lever rule: the relative quantities of the liquid and the vapor we identified above are given by the lengths of the tie-line segments labeled *a* and *b*. Thus in this particular example, in which *b* is about four times longer than *a*, we can say that the mole ratio of vapor (of composition 3) to liquid (composition 2) is 4.

Steps in Fractional Distillation

It is not practical to carry out an almost-infinite number of distillation steps to obtain nearly-infinitesimal quantities of the two pure liquids we wish to separate. So instead of collecting each drop of condensate and re-distilling it, we will distil half of the mixture in each step. Suppose you want to separate a liquid mixture composed of 20 mole-% B and 80 mole-% A, with A being the more volatile.

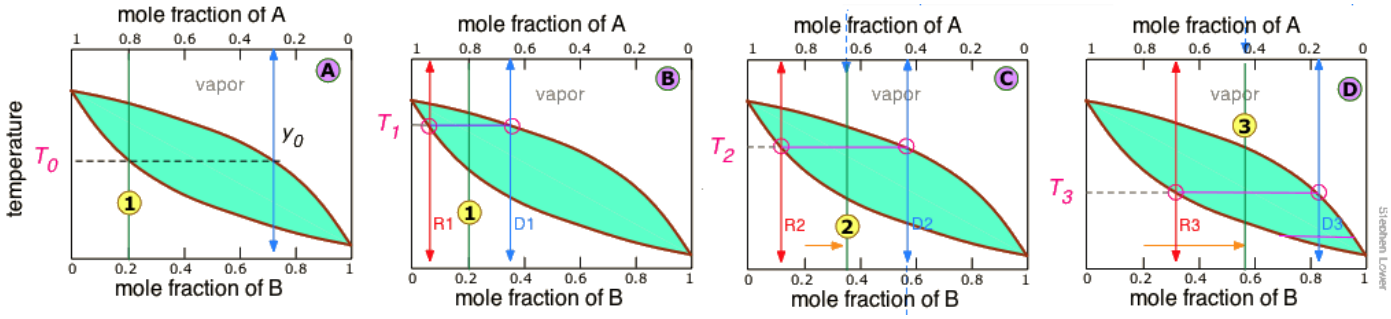


Figure 8.9.8: Steps in Fractional Distillation

A As we heat the mixture whose overall composition is indicated by 1, the first vapor is formed at $T_0$and has the composition y_0, found by extending the horizontal dashed line until it meets the vapor curve. This vapor is clearly enriched in B; if it is condensed, the resulting liquid will have a mole fraction x_B approaching that of A in the original liquid. But this is only the first drop, we don't want to stop there!

B As the liquid continues to boil, the boiling temperature rises. When it reaches T_1, we will have boiled away half of the liquid. At this point, the "system" composition (liquid plus vapor) is still the same (1), but is now equally divided between the liquid, which we call "residue" R_1, and the condensed vapor, the *distillate* D_1.

> *How do we know it is equally divided? We have picked T_1 so that the tie line is centered on the system concentration, so by the lever rule, R_1 and D_1 contain equal numbers of moles.*

C We now take the condensed liquid D_1 having the composition 2 , and distill half of it, obtaining distillate of composition D_2.

D .. and then carry out yet another distillation, this time using D_3 as our feedstock.

E Our four-stage fractionation has enriched the more volatile solute from 20 to slightly over 80 mole-percent in D_4. The less volatile component A is most concentrated in R_1. R_2 through R_4 are thrown away (but not down the sink, please!)

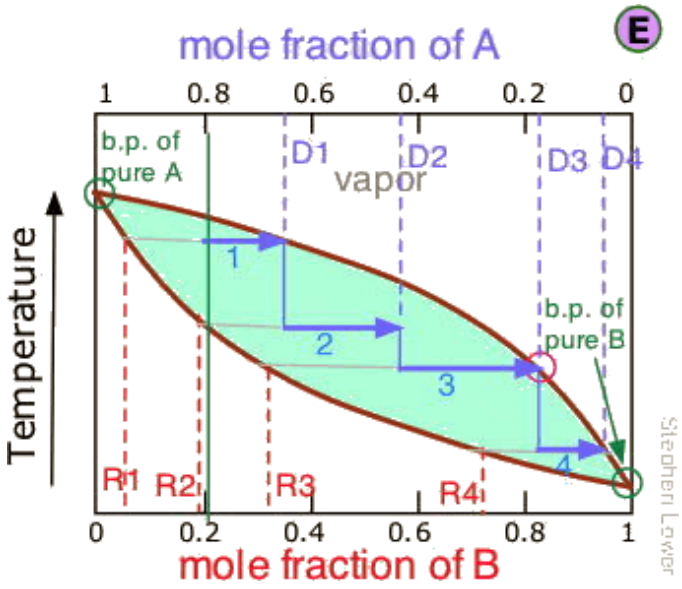


Figure 8.9.9: Boiling point diagrams illustrating fractional distillation"

This may be sufficient for some purposes, but we might wish to do much better, using perhaps 1000 stages instead of just 4. What could be more tedious?

Fractionation with reflux

Not to worry! The multiple successive distillations can be carried out "virtually" by inserting a *fractionating column* between the boiling flask and the condenser.

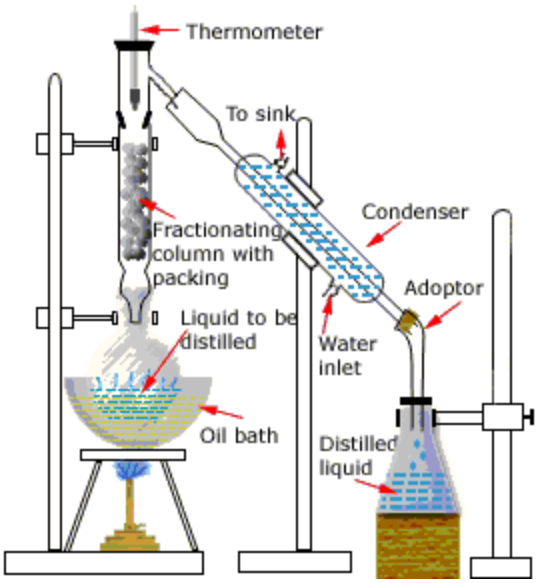
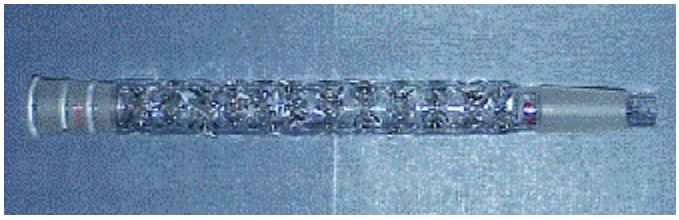


Figure 8.9.10:

These columns are made with indentations or are filled with materials that provide a large surface area extending through the vertical temperature gradient (higher temperature near the bottom, lower temperature at the top.) The idea is that hot vapors condense at various levels in the column and the resulting liquid drips down (*refluxes*) to a lower level where it is vaporized, which corresponds roughly to a re-distillation.



Vigreux columns having multiple indentations are widely used (above right). Simple columns can be made by filling a glass tube with beads, short glass tubes, or even stainless steel kitchen-type scouring pads. More elaborate ones have spinning steel ribbons.

Separation efficiency: theoretical plates

The operation of fractionating columns can best be understood by reference to a bubble-cap column. The one shown here consists of four sections, or "plates" through which hot vapors rise and bubble up through pools of condensate that collect on each plate. The intimate contact between vapor and liquid promotes equilibration and re-distillation at successively higher temperatures at each higher plate in the column. Unlike the case of the step-wise fractional distillation we discussed above, none of the intermediate residues is thrown away; they simply drip down back into the pot where their fractionation journey begins again, always leading to a further concentration of the less-volatile component in the remaining liquid. At the same time, the vapor emerging from the top plate (5) provides a continuing flow of volatile-enriched condensate, although in diminishing quantities as it is depleted in the boiling pot.

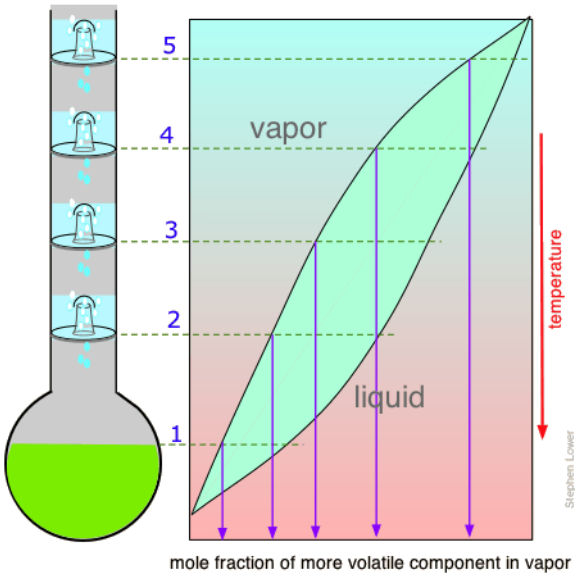


Figure 8.9.11:

If complete equilibrium is attained between the liquid and vapor at each stage, then we can describe the system illustrated above as providing "five theoretical plates" of separation (remember that the pot represents the first theoretical plate.) Equilibrium at each stage requires a steady-state condition in which the quantity of vapor moving upward at each stage is equal to the quantity of liquid draining downward — in other words, the column should be operating in total reflux, with no net removal of distillate. So any real distillation process will be operated at a reflux ratio that provides optimum separation in a reasonable period of time.

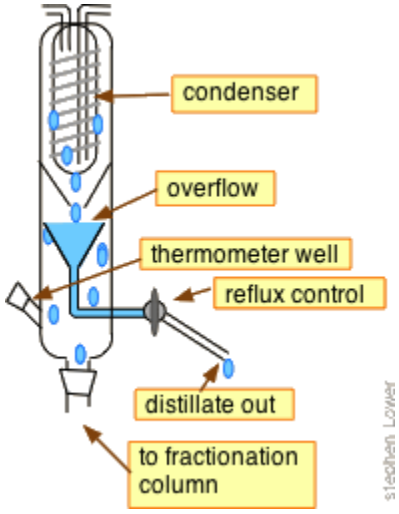


Figure 8.9.11: Reflux ratio can be adjusted by placing a special control head on top of the fractionation column. The laboratory-type device shown here illustrates the general concept. Reflux control is mainly of importance in large-scale industrial operations.

Some of the more advanced laboratory-type devices (such as some spinning-steel band columns) are said to offer up to around 200 theoretical plates of separating power.

Azeotropes: the Limits of Distillation

The boiling point diagrams presented in the foregoing section apply to solutions that behave in a reasonably ideal manner — that is, to solutions that do not deviate too far from Raoult's law. As we explained above, mixtures of liquids whose intermolecular interactions are widely different do not behave ideally, and may be impossible to separate by ordinary distillation. The reason for this is that under certain conditions, the compositions of the liquid and of the vapor in equilibrium with it become identical, precluding any further separation. These cross-over points appear as "kinks" in the boiling point diagrams.

High- and low-boiling azeotropes

Thus in this boiling point diagram for a mixture exhibiting a positive deviation from Raoult's law, successive fractionations of mixtures correspond to either (1) or (2) bring the distillation closer to the azeotropic composition indicated by the dashed vertical line. Once this point is reached, further distillation simply yields more of the same "high-boiling" azeotrope.

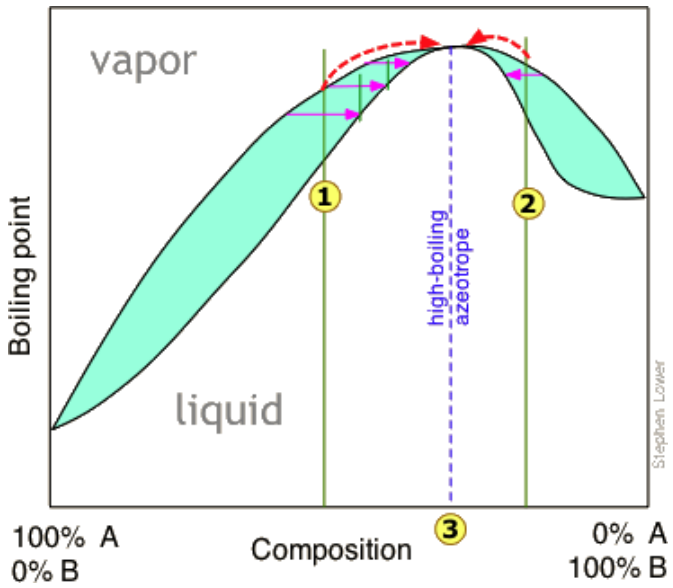


Figure 8.9.12: Maximum Boiling Azeotrope

Distillation of a mixture having a negative deviation from Raoult's law leads to a similar stalemate, in this case yielding a "low-boiling" azeotrope. High- and low-boiling azeotropes are commonly referred to as *constant-boiling mixtures*, and they are more common than most people think.

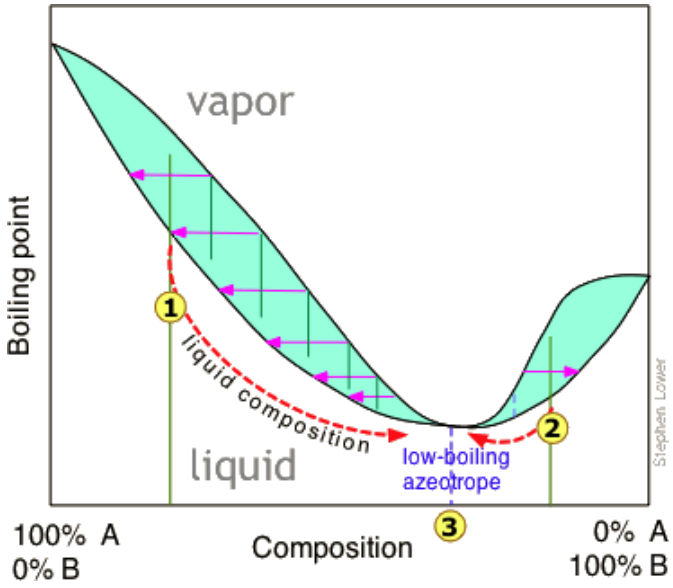


Figure 8.9.13: Maximum Boiling Azeotrope

"Breaking" an azeotrope

There are four general ways of dealing with azeotropes. The first two of these are known collectively as *azeotropic distillation*.

- **Addition of a third substance** that alters the intermolecular attractions is the most common trick. The drawback is that another procedure is usually needed to remove this other substance.
- **Pressure-swing distillation** takes advantage of the fact that boiling point (*T,X*) diagrams are two-dimensional slices of a (*T,X,P*) diagram in which the pressure is the third variable. This means that the azeotropic composition depends on the pressure, so distillation at some pressure other than 1 atm may allow one to "jump" the azeotrope.
- **Use of a molecular sieve** — a porous material that selectively absorbs one of the liquids, most commonly water when the latter is present at a low concentration.
- **Give up.** It often happens that the azeotropic composition is sufficiently useful that it's not ordinarily worth the trouble of obtaining a more pure product. This accounts for the concentrations of many commercial chemicals such as mineral acids.

Some constant-boiling mixtures with water

mixture	azeotrope
Ethanol	98%, high, 78.1°C
Hydrochloric acid	20.2% high, 108.6°C
Hydrofluoric acid	35.6%, 111.3°C
Nitric acid	68%, 120.5°C
Sulfuric acid	98.3%, 338°C

Ethanol is one of the major industrial chemicals, and is of course the essential component of beverages that have been a part of civilization throughout recorded history. Most ethanol is produced by fermentation of the starch present in food grains, or of sugars formed by the enzymatic degradation of cellulose. Because ethanol is toxic to the organisms whose enzymes mediate the fermentation process, the ethanol concentration in the fermented mixture is usually limited to about 15%. The liquid phase of the mixture is then separated and distilled.

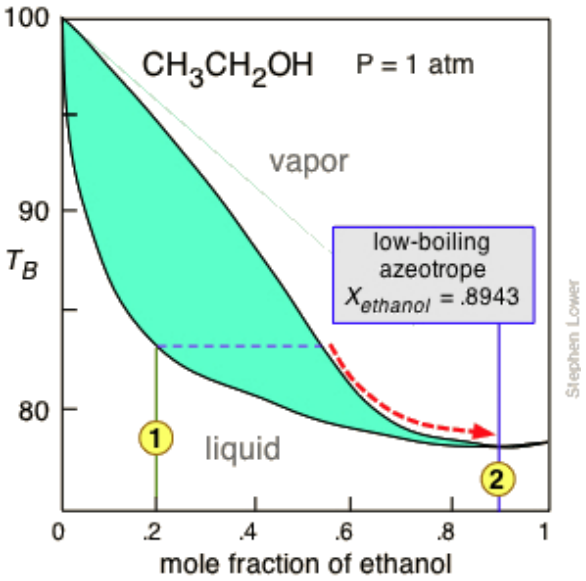


Figure 8.9.8**:** *The normal boiling point of ethanol is 78.4°C, but its formation of a low-boiling* azeotrope *with water limits the maximum concentration that can be obtained by ordinary distillation to 95.6 volume-% (89.5 mole-%). This concentration is adequate for many applications, for which the product is commonly sold as "95% ethanol".*

For applications requiring anhydrous ethanol ("absolute ethanol"), the most common method is the use of zeolite-based molecular sieves to absorb the remaining water. Addition of benzene can break the azeotrope, and this was the most common production method in earlier years. For certain critical uses where the purest ethanol is required, it is synthesized directly from ethylene.

Here be briefly discuss two distillation methods that students are likely to encounter in more advanced organic lab courses.

Vacuum distillation: Many organic substances become unstable at high temperatures, tending to decompose, polymerize or react with other substances at temperatures around 200° C or higher. A liquid will boil when its vapor pressure becomes equal to the pressure of the gas above it, which is ordinarily that of the atmosphere. If this pressure is reduced, boiling can take place at a lower temperature. (Even pure water will boil at room temperature under a partial vacuum.) "Vacuum distillation" is of course a misnomer; a more accurate term would be "reduced-pressure distillation". Vacuum distillation is very commonly carried out in the laboratory and will be familiar to students who take more advanced organic lab courses. It is also sometimes employed on a large industrial scale.

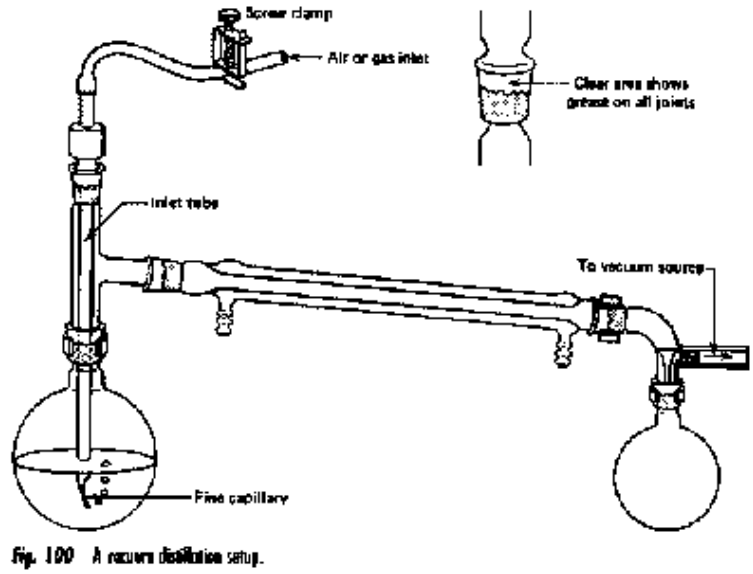


Fig. 100 A vacuum distillation setup.



Figure 8.9.9: vacuum connection at lower right

The vacuum distillation setup is similar that employed in ordinary distillation, with a few additions:

- The vacuum line is connected to the bent adaptor above the receiving flask.
- In order to avoid uneven boiling and superheating ("bumping"), the boiling flask is usually provided with a fine capillary ("ebulliator") through which an air leak produces bubbles that nucleate the boiling liquid.
- The vacuum is usually supplied by a mechanical pump, or less commonly by a water aspirator or a "house vacuum" line.
- The boiling flask is preferably heated by a water- or steam bath, which provides more efficient heat transfer to the flask and avoids localized overheating. Prior to about 1960, open flames were commonly used in student laboratories, resulting in occasional fires that enlivened the afternoon, but detracted from the student's lab marks.
- A Claisen-type distillation head (below) provides a convenient means of accessing the boiling flask for inserting an air leak capillary or introducing additional liquid through a separatory funnel. This Claisen-Vigreux head includes a fractionation column.



Figure 8.9.10: A Claisen-type distillation head

Steam Distillation: Strictly speaking, this topic does not belong in this unit, since steam distillation is used to separate *immiscible* liquids rather than solutions. But because immiscible liquid mixtures are not treated in elementary courses, we present a brief description of steam distillation here for the benefit of students who may encounter it in an organic lab course. A mixture of immiscible liquids will boil when their *combined* vapor pressures reach atmospheric pressure. This combined vapor pressure is just the sum of the vapor pressures of each liquid individually, and is independent of the quantities of each phase present.

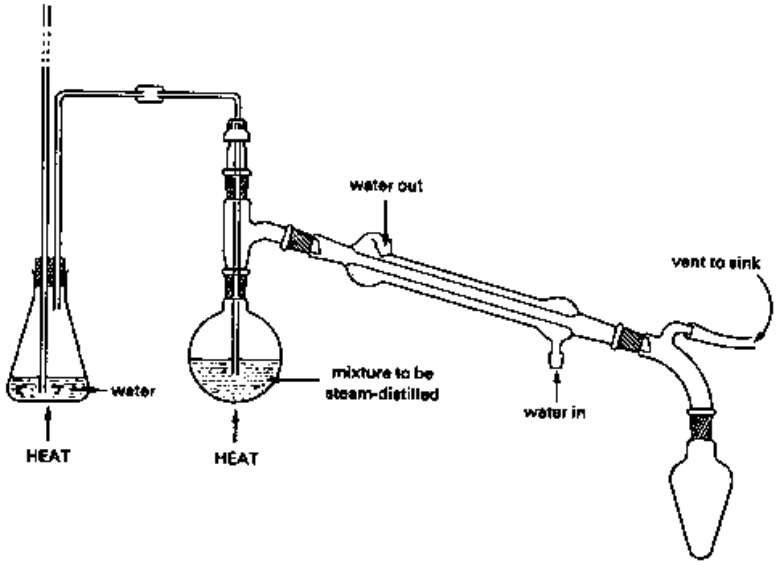


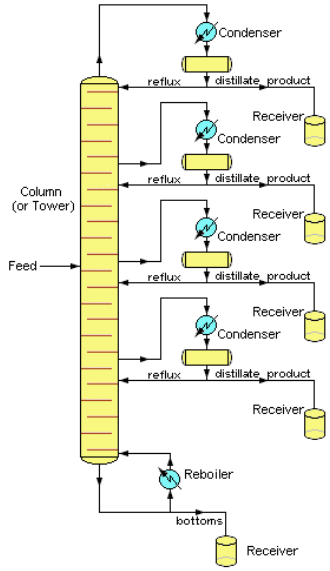
Figure 8.9.11

Because water boils at 100° C, a mixture of water and an immiscible liquid (an "oil"), even one that has a high boiling point, is guaranteed to boil *below* 100°, so this method is especially valuable for separating high boiling liquids from mixtures containing non-volatile impurities. Of course the water-oil mixture in the receiving flask must itself be separated, but this is usually easily accomplished by means of a separatory funnel since their densities are ordinarily different.

There is a catch, however: the lower the vapor pressure of the oil, the greater is the quantity of water that co-distills with it. This is the reason for using steam: it provides a source of water able to continually restore that which is lost from the boiling flask. Steam distillation from a water-oil mixture without the introduction of additional steam will also work, and is actually used for some special purposes, but the yield of product will be very limited. Steam distillation is widely used in industries such as petroleum refining (where it is often called "steam stripping") and in the flavors-and-perfumes industry for the isolation of essential oils

The term *essential oil* refers to the aromas ("essences") of these [mostly simple] organic liquids which occur naturally in plants, from which they are isolated by steam distillation or solvent extraction. Steam distillation was invented in the 13th Century by Ibn al-Baiter, one of the greatest of the scientists and physicians of the Islamic Golden Age in Andalusia.

Distillation is one of the major "unit operations" of the chemical process industries, especially those connected with petroleum and biofuel refining, liquid air separation, and brewing. Laboratory distillations are typically batch operations and employ relatively simple fractionating columns to obtain a pure product. In contrast, industrial distillations are most often designed to produce mixtures having a desired boiling range rather than pure products.



Industrial operations commonly employ bubble-cap fractionating columns (seldom seen in laboratories), although packed columns are sometimes used. Perhaps the most distinctive feature of large scale industrial distillations is that they usually operate on a continuous basis in which the preheated crude mixture is preheated in a furnace and fed into the fractionating column at some intermediate point. A reboiler unit maintains the bottom temperature at a constant value. The higher-boiling components then move down to a level at which they vaporize, while the lighter (lower-boiling) material moves upward to condense at an appropriate point.

Petroleum is a complex mixture of many types of organic molecules, mostly hydrocarbons, that were formed by the effects of heat and pressure on plant materials (mostly algae) that grew in regions that the earth's tectonic movements buried over periods of millions of years. This mixture of liquid and gases migrates up through porous rock until it s trapped by an impermeable layer of sedimentary rock. The molecular composition of *crude oil* (the liquid fraction of petroleum) is highly variable, although its overall elemental makeup generally reflects that of typical plants.

element	carbon	hydrogen	nitrogen	oxygen	sulfur	metals
amount	83-87%	10-14%	0.1-2%	0.1-1.5%	0.5-6%	

The principal molecular constituents of crude oil are

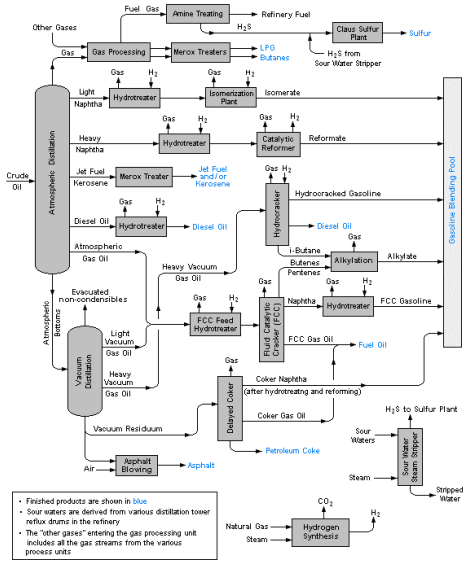
- alkanes: Also known as *paraffins*, these are saturated linear- or branched-chain molecules having the general formula C_nH_{2n+2} in which n is mostly between 5 and 40.
- unsaturated aliphatic: Linear- or branched chain molecules containing one or more double or triple bonds (alkenes or alkynes).
- Cycloalkanes:Also known as naphthenes these are saturated hydrocarbons C_nH_{2n} containing one or more ring structures.
- Aromatic hydrocarbons:These contain one or more fused benzene rings C_nH_n, often with hydrocarbon side-chains.

The word *gasoline* predates its use as a motor fuel; it was first used as a topical medicine to rid people of head lice, and to remove grease spots and stains from clothing. The first major step of refining is to fractionate the crude oil into various boiling ranges.

boiling range	fraction name	further processing
	butane and propane	gas processing
30 - 210°	straight-run gasoline	blending into motor gasoline
100 - 200°	naphtha	reforming into gasoline components
150 - 250°	kerosene	jet fuel blending
160 -400°	light gas oil	distillate fuel blending into diesel or fuel oil
315 - 540°	heavy gas oil	catalytic cracking: large molecules are broken up into smaller ones and recycled
>450°	asphalts, bottoms	may be vacuum-distilled into more fractions

Further processing and blending

About 16% of crude oil is diverted to the petrochemical industry where it is used to make ethylene and other feedstocks for plastics and similar products. Because the fraction of straight-run gasoline is inadequate to meet demand, some of the lighter fractions undergo reforming and the heavier ones cracking and are recycled into the gasoline stream. These processes necessitate a great amount of recycling and blending, into which must be built a considerable amount of flexibility in order to meet seasonal needs (more volatile gasolines and heating fuel oil in winter, more total gasoline volumes in the summer.)



This page titled 8.9: Distillation is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.10: IONS AND ELECTROLYTES

Electrolytic solutions are those that are capable of conducting an electric current. A substance that, when added to water, renders it conductive, is known as an *electrolyte*. A common example of an electrolyte is ordinary salt, sodium chloride. Solid NaCl and pure water are both non-conductive, but a solution of salt in water is readily conductive. A solution of sugar in water, by contrast, is incapable of conducting a current; sugar is therefore a *non-electrolyte*.



-  Topic hierarchy

8.10.9A: Electrolytes and Electrolytic Solutions

8.10.9B: The nature of ions in aqueous solution

8.10.9C: Weak and Strong Electrolytes

8.10.9D: Ionic migration

8.10.9E: Some applications of electrolytic conduction

This page titled 8.10: Ions and Electrolytes is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.10.9A: Electrolytes and Electrolytic Solutions

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Describe the properties of water that make it an ideal electrolytic solvent.
- Describe the general structure of ionic hydration shells.
- Explain why all cations act as acids in water.
- Describe some of the major ways in which the conduction of electricity through a solution differs from metallic conduction.
- Define resistance, resistivity, conductance, and conductivity.
- Define molar conductivity and explain its significance.
- Explain the major factors that cause molar conductivity to diminish as electrolyte concentrations increase.
- Describe the contrasting behavior of strong, intermediate, and weak electrolytes.
- Explain the distinction between ionic diffusion and ionic migration.
- Define the limiting ionic conductivity, and comment on some of its uses.
- Explain why hydrogen- and hydroxide ions exhibit exceptionally large ionic mobilities.

Electrolytic solutions are those that are capable of conducting an electric current. A substance that, when added to water, renders it conductive, is known as an *electrolyte*. A common example of an electrolyte is ordinary salt, sodium chloride. Solid NaCl and pure water are both non-conductive, but a solution of salt in water is readily conductive. A solution of sugar in water, by contrast, is incapable of conducting a current; sugar is therefore a *non-electrolyte*.

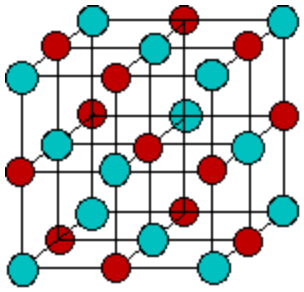
These facts have been known since 1800 when it was discovered that an electric current can decompose the water in an electrolytic solution into its elements (a process known as *electrolysis*). By mid-century, Michael Faraday had made the first systematic study of electrolytic solutions. Faraday recognized that for a sample of matter to conduct electricity, two requirements must be met:

1. The matter must be composed of, or contain, electrically charged particles.
2. These particles must be *mobile*; that is, they must be free to move under the influence of an external applied electric field.

In metallic solids, the charge carriers are electrons rather than ions; their mobility is a consequence of the quantum-mechanical uncertainty principle which promotes the escape of the electrons from the confines of their local atomic environment. In the case of electrolytic solutions, Faraday called the charge carrier *ions* (after the Greek word for "wanderer"). His most important finding was that each kind of ion (which he regarded as an electrically-charged atom) carries a definite amount of charge, most commonly in the range of ±1-3 units.

The fact that the smallest charges observed had magnitudes of ±1 unit suggested an "atomic" nature for electricity itself, and led in 1891 to the concept of the "electron" as the unit of electric charge — although the identification of this unit charge with the particle we now know as the electron was not made until 1897.

An ionic solid such as NaCl is composed of charged particles, but these are held so tightly in the crystal lattice that they are unable to move about, so the second requirement mentioned above is not met and solid salt is not a conductor. If the salt is melted or dissolved in water, the ions can move freely and the molten liquid or the solution becomes a conductor.

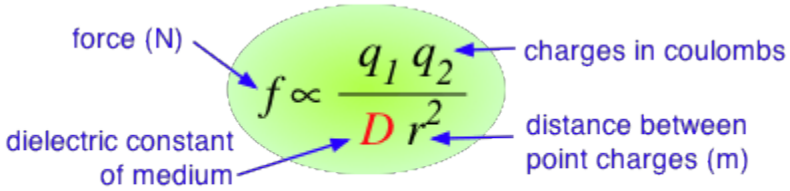


Since positively-charged ions are attracted to a negative electrode that is traditionally known as the *cathode*, these are often referred to as *cations*. Similarly, negatively-charged ions, being attracted to the positive electrode, or *anode*, are called *anions*. (These terms were all coined by Faraday.)

The role of the solvent: what's special about water

Although we tend to think of the solvent (usually water) as a purely passive medium within which ions drift around, it is important to understand that electrolytic solutions would not exist without the **active involvement of the solvent** in reducing the strong attractive forces that hold solid salts and molecules such as HCl together. Once the ions are released, they are stabilized by interactions with the solvent molecules. Water is not the only liquid capable of forming electrolytic solutions, but it is by far the most important. It is therefore essential to understand those properties of water that influence the stability of ions in aqueous solution.

According to Coulomb's law, the force between two charged particles is directly proportional to the product of the two charges, and inversely proportional to the square of the distance between them:



Note: a positive value of f indicates repulsion

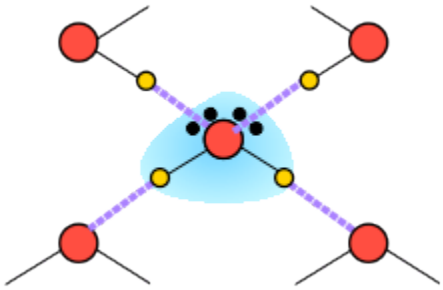
The proportionality constant *D* is the dimensionless *dielectric constant*. Its value in empty space is unity, but in other media it will be larger. Since *D* appears in the denominator, this means that the force between two charged particles within a gas or liquid will be less than if the particles were in a vacuum. Water has one of the highest dielectric constants of any known liquid; the exact value varies with the temperature, but 80 is a good round number to remember. When two oppositely-charged ions are immersed in water, the force acting between them is only 1/80 as great as it would be between the two gaseous ions at the same distance. It can be shown that in order to separate one mole of Na^+ and Cl^- ions at their normal distance of 23.6 pm in solid sodium chloride, the work required will be 586 J in a vacuum, but only 7.3 J in water

The dielectric constant is a *bulk property* of matter, rather than being a property of the molecule itself, as is the dipole moment. It is a cooperative effect of all the molecules in the liquid, and is a measure of the extent to which an applied electric field will cause the molecules to line up with the negative ends of their dipoles pointing toward the positive direction of the electric field. The high dielectric constant of water is a consequence of the small size of the H_2O molecule in relation to its large dipole moment.

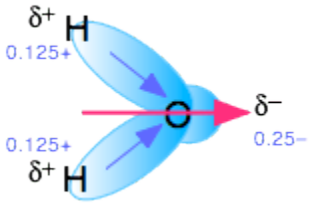
When one molecule is reoriented by the action of an external electric field, local hydrogen bonding tends to pull neighboring molecules into the same alignment, thus producing an additive effect that would be absent if the molecules were all acting independently.

Water's dipole moment stabilizes dissolved ions through hydrogen bonding

When an ion is introduced into a solvent, the attractive interactions between the solvent molecules must be disrupted to create space for the ion. This costs energy and would by itself tend to inhibit dissolution. However, if the solvent has a high permanent dipole moment, the energy cost is more than recouped by the ion-dipole attractions between the ion and the surrounding solvent molecules.



Water, as you know, has a sizeable dipole moment that is the source of the hydrogen bonding that holds the liquid together. The greater strength of ion-dipole attractions compared to hydrogen-bonding (dipole-dipole) attractions stabilizes the dissolved ion.



Water is not the only electrolytic solvent, but it is by far the best. For some purposes, chemists occasionally need to employ non-aqueous solvents when studying electrolytes. Here are a few examples:

solvent	Melting Point (°C)	Boiling Point °C	D	dipole moment	specific conductivity, S/cm
water	0	100	80.1	1.87	5.5×10^{-8}
methanol	-98	64.7	32.7	1.7	1.5×10^{-7}
ethanol	-114	78.3	24.5	1.66	1.35×10^{-9}
acetonitrile	-43.8	81.6	37.5	3.45	7×10^{-6}
dimethyl sulfoxide	18.5	189	46.7	3.96	3×10^{-8}
ethylene carbonate	36.4	238	89.6	4.87	$< 1 \times 10^{-9}$

This page titled 8.10.9A: Electrolytes and Electrolytic Solutions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.10.9B: The nature of ions in aqueous solution

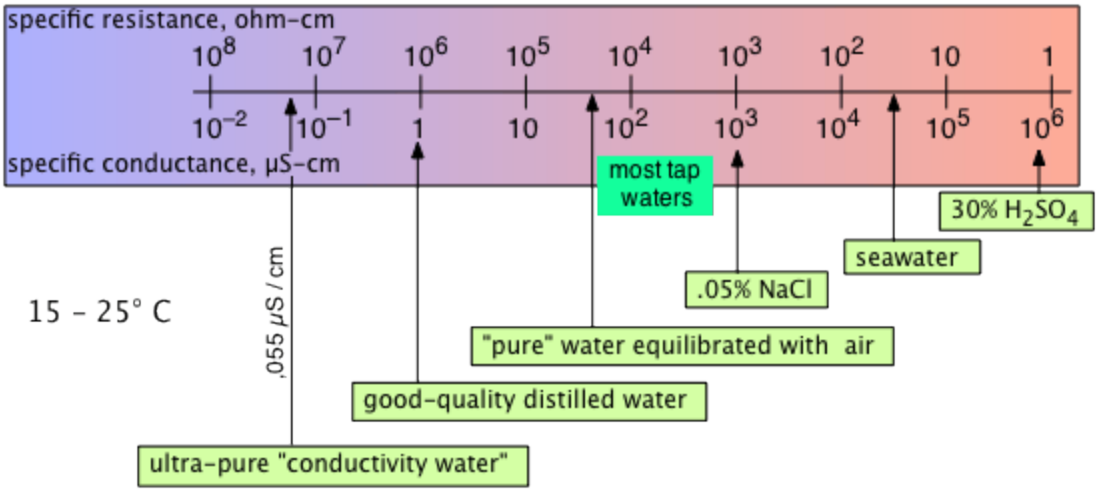
The kinds of ions we will consider in this lesson are mostly those found in solutions of common acids or salts. As is evident from the image below, most of the metallic elements form monatomic cations, but the number of monatomic anions is much smaller. This reflects the fact that many single-atom anions such as hydride H^-, oxide O^{2-}, sulfide S^{2-} and those in Groups 15 and 16, are unstable in (*i.e.*, react with) water, and their major forms are those in which they are combined with other elements, particularly oxygen. Some of the more familiar *oxyanions* are hydroxide OH^-, carbonate CO_3^{2-}, nitrate NO_3^-, sulfate SO_4^{2-}, chlorate ClO_4^{2-}, and arsenate AsO_4^{2-}.

Conductivity water" with $\kappa = 0.043 \times 10^{-6}$ S cm^{-1} at 18°C. Ordinary distilled water in equilibrium with atmospheric CO_2 has a conductivity that is 16 times greater.



It is now known that ordinary distillation cannot entirely remove all impurities from water. Ionic impurities get entrained into the fog created by breaking bubbles and are carried over into the distillate by capillary flow along the walls of the apparatus. Organic materials tend to be steam-volatile ("steam-distilled").

The best current practice is to employ a special still made of fused silica in which the water is volatilized from its surface without boiling. Complete removal of organic materials is accomplished by passing the water vapor through a column packed with platinum gauze heated to around 800°C through which pure oxygen gas is passed to ensure complete oxidation of carbon compounds.



Conductance measurements are widely used to gauge water quality, especially in industrial settings in which concentrations of dissolved solids must be monitored in order to schedule maintenance of boilers and cooling towers.

The conductance of a solution depends on **1)** the concentration of the ions it contains, **2)** on the number of charges carried by each ion, and **3)** on the *mobilities* of these ions. The latter term refers to the ability of the ion to make its way through the solution, either by ordinary thermal diffusion or in response to an electric potential gradient.

The first step in comparing the conductances of different solutes is to reduce them to a common concentration. For this, we define the *conductance per unit concentration* which is known as the *molar conductivity*, denoted by the upper-case Greek *lambda*:

$$\Lambda = \frac{\kappa}{c} \tag{8.10.9B.1}$$

When κ is expressed in S cm^{-1}, *C* should be in mol cm^{-3}, so Λ will have the units S cm^2. This is best visualized as the conductance of a cell having 1-cm^2 electrodes spaced 1 cm apart — that is, of a 1 cm cube of solution. But because chemists generally prefer to express concentrations in mol L^{-1} or mol dm^{-3} (mol/1000 cm^3) , it is common to write the expression for molar conductivity as

$$\Lambda = \frac{1000\kappa}{c} \tag{8.10.9B.2}$$

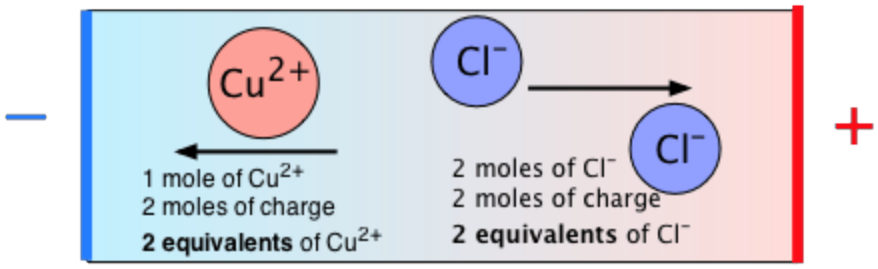
electrodes, separated again by 1 cm.

But if *c* is the concentration in moles per liter, this will still not fairly compare two salts having different stoichiometries, such as $AgNO_3$ and $FeCl_3$, for example. If we assume that both salts dissociate completely in solution, each mole of $AgNO_3$ yields two moles of charges, while $FeCl_3$ releases six (i.e., one Fe^{3+} ion, and three Cl^- ions.) So if one neglects the [rather small] differences in the ionic mobilities, the molar conductivity of $FeCl_3$ would be three times that of $AgNO_3$.

The most obvious way of getting around this is to note that one mole of a 1:1 salt such as $AgNO_3$ is "equivalent" (in this sense) to 1/3 of a mole of $FeCl_3$, and of ½ a mole of $MgBr_2$. To find the number of equivalents that correspond a given quantity of a salt, just divide the number of moles by the total number of positive charges in the formula unit. (If you like, you can divide by the number of negative charges instead; because these substances are electrically neutral, the numbers will be identical.)

We can refer to equivalent concentrations of individual ions as well as of neutral salts. Also, since acids can be regarded as salts of H^+, we can apply the concept to them; thus a 1M L^{-1} solution of sulfuric acid H_2SO_4 has a concentration of 2 eq L^{-1}.

The following diagram summarizes the relation between moles and equivalents for $CuCl_2$:



Example 8.10.9B.1

What is the concentration, in equivalents per liter, of a solution made by dissolving 4.2 g of chromium(III) sulfate pentahydrate $Cr_2(SO_4)_3 \cdot 5H_2O$ in sufficient water to give a total volume of 500 mL? (The molar mass of the hydrate is 482 g)

Solution

Assume that the salt dissociates into 6 positive charges and 6 negative charges.

- Number of moles of the salt: (4.2 g) / (482 g mol^{-1}) = 0.00871 mol
- Number of equivalents: (.00871 mol) / (6 eq mol^{-1}) = 0.00146 eq
- Equivalent concentration: (0.00146 eq) / (0.5 L) = 0.00290 eq L^{-1}

The concept of equivalent concentration allows us to compare the conductances of different salts in a meaningful way. *Equivalent conductivity* is defined similarly to molar conductivity

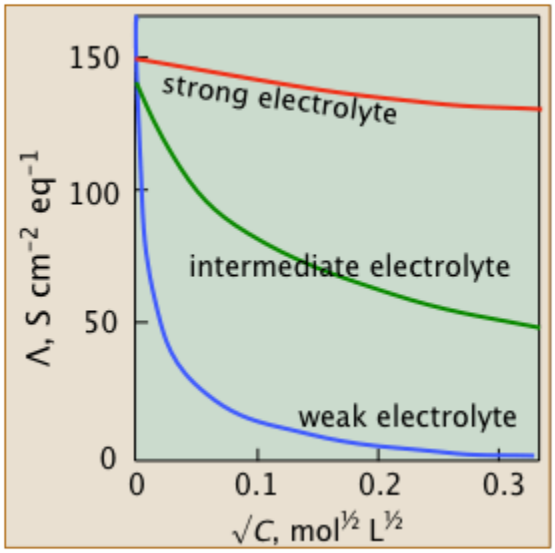
$$\Lambda = \frac{\kappa}{c} \tag{8.10.9B.3}$$

except that the concentration term is now expressed in equivalents per liter instead of moles per liter. (In other words, the equivalent conductivity of an electrolyte is the conductance per equivalent per liter.)

This page titled 8.10.9B: The nature of ions in aqueous solution is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.10.9C: Weak and Strong Electrolytes

The serious study of electrolytic solutions began in the latter part of the 19th century, mostly in Germany — and before the details of dissociation and ionization were well understood. These studies revealed that the equivalent conductivities of electrolytes all diminish with concentration (or more accurately, with the square root of the concentration), but they do so in several distinct ways that are distinguished by their behaviors at very small concentrations. This led to the classification of electrolytes as weak, intermediate, and strong.



You will notice that plots of conductivities vs. $\sqrt{c}$ start at c=0. It is of course impossible to measure the conductance of an electrolyte at vanishingly small concentrations (not to mention zero!), but for strong and intermediate electrolytes, one can extrapolate a series of observations to zero. The resulting values are known as *limiting equivalent conductances* or sometimes as "equivalent conductances at infinite dilution", designated by Λ°.

Strong electrolytes

These well-behaved systems include many simple salts such as NaCl, as well as all strong acids.
The Λ vs. $\sqrt{c}$ plots closely follow the linear relation

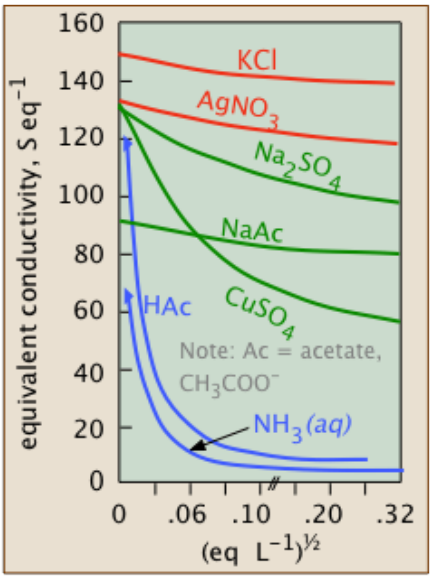
$$\Lambda = \Lambda^\circ - b\sqrt{c}$$

Intermediate electrolytes

These "not-so-strong" salts can't quite conform to the linear equation above, but their conductivities can be extrapolated to infinite dilution.

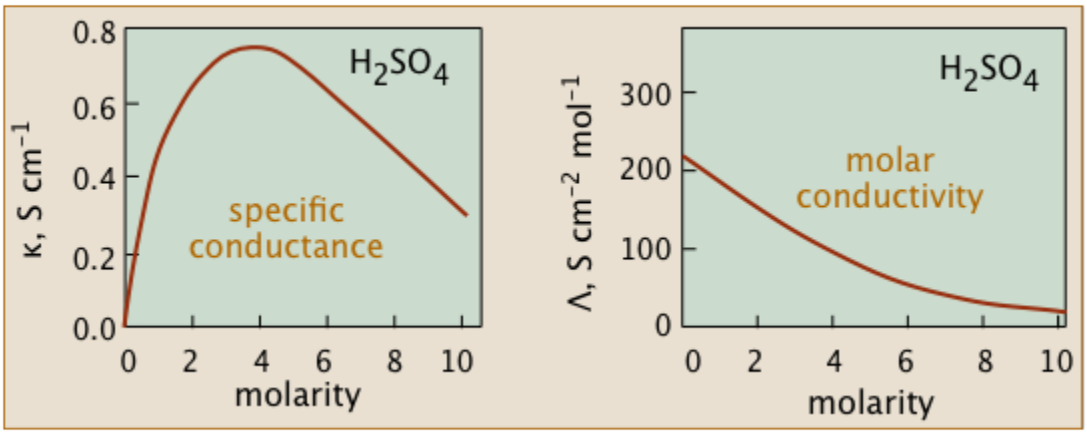
Weak electrolytes

"Less is more" for these oddities which possess the remarkable ability to exhibit infinite equivalent conductivity at infinite dilution. Although Λ° cannot be estimated by extrapolation, there is a clever work-around.



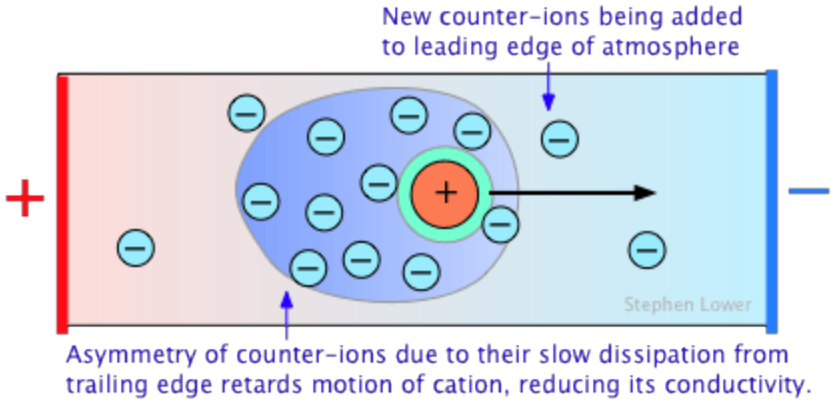
Conductivity diminishes as concentrations increase

Since ions are the charge carriers, we might expect the conductivity of a solution to be directly proportional to their concentrations in the solution. So if the electrolyte is totally dissociated, the conductivity should be directly proportional to the electrolyte concentration. But this *ideal behavior* is never observed; instead, the conductivity of electrolytes of all kinds diminishes as the concentration rises.



The non-ideality of electrolytic solutions is also reflected in their colligative properties, especially freezing-point depression and osmotic pressure. The primary cause of this is the presence of the ionic atmosphere that was introduced above. To the extent that ions having opposite charge signs are more likely to be closer together, we would expect their charges to partially cancel, reducing their tendency to migrate in response to an applied potential gradient.

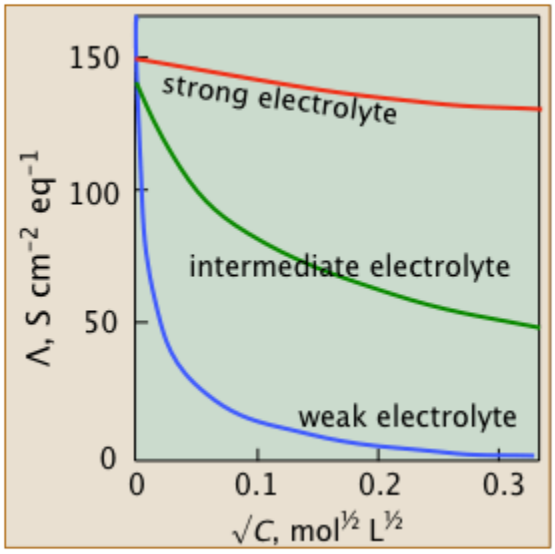
A secondary effect arises from the fact that as an ion migrates through the solution, its counter-ion cloud does not keep up with it. Instead, new counter-ions are continually acquired on the leading edge of the motion, while existing ones are left behind on the opposite side. It takes some time for the lost counter-ions to dissipate, so there are always more counter-ions on the trailing edge. The resulting asymmetry of the counter-ion field exerts a retarding effect on the central ion, reducing its rate of migration, and thus its contribution to the conductivity of the solution.



The quantitative treatment of these effects was first worked out by P. Debye and W. Huckel in the early 1920's, and was improved upon by Ostwald a few years later. This work represented one of the major advances in physical chemistry in the first half of the 20th Century, and put the behavior of electrolytic solutions on a sound theoretical basis. Even so, the Debye-Huckel theory breaks down for concentrations in excess of about 10^{-3} M L^{-1} for most ions.

Not all Electrolytes Totally Dissociate in Solution

plots for *strong* electrolytes is largely explained by the effects discussed immediately above. The existence of *intermediate* electrolytes served as the first indication that many salts are not completely ionized in water; this was soon confirmed by measurements of their colligative properties.



The curvature of the plots for intermediate electrolytes is a simple consequence of the Le Chatelier effect, which predicts that the equilibrium

$$MX_{(aq)} = M^{+}_{(aq)} + X^{-}_{(aq)} \tag{8.10.9C.1}$$

will shift to the left as the concentration of the "free" ions increases. In more dilute solutions, the actual *concentrations* of these ions is smaller, but their fractional abundance in relation to the undissociated form is greater. As the solution approaches zero concentration, virtually all of the \(MX_{(aq)}\) becomes dissociated, and the conductivity reaches its limiting value.

Weak electrolytes are dissociated only at extremely high dilution

hydrofluoric acid	HF	$K_a = 10^{-3.2}$

acetic acid	CH_3COOH	$K_a = 10^{-6.3}$
bicarbonate ion	HCO_3^-	$K_a = 10^{-10.3}$
ammonia	NH_3	$K_b = 10^{-4.7}$

Dissociation, of course, is a matter of degree. The equilibrium constants for the dissociation of an intermediate electrolyte salt MX are typically in the range of 1-200. This stands in contrast to the large number of weak acids (as well as weak bases) whose dissociation constants typically range from 10^{-3} to smaller than 10^{-10}.

These weak electrolytes, like the intermediate ones, will be totally dissociated at the limit of zero concentration; if the scale of the weak-electrolyte plot (blue) shown above were magnified by many orders of magnitude, the curve would resemble that for the intermediate electrolyte above it, and a value for $\Lambda°$ could be found by extrapolation. But at such a high dilution, the conductivity would be so minute that it would be masked by that of water itself (that is, by the H^+ and OH^- ions in equilibrium with the massive 55.6 M L^{-1} concentration of water) — making values of Λ in this region virtually unmeasurable.

This page titled 8.10.9C: Weak and Strong Electrolytes is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.10.9D: Ionic migration

The motion of ions in solution is mainly random

The conductance of an electrolytic solution results from the movement of the ions it contains as they migrate toward the appropriate electrodes. But the picture we tend to have in our minds of these ions moving in a orderly, direct march toward an electrode is wildly mistaken. The thermally-induced random motions of molecules is known as *diffusion*. The term *migration* refers specifically to the movement of ions due to an externally-applied electrostatic field.

The average thermal energy at temperatures within water's liquid range (given by RT) is sufficiently large to dominate the movement of ions even in the presence of an applied electric field. This means that the ions, together with the water molecules surrounding them, are engaged in a wild dance as they are buffeted about by thermal motions (which include Brownian motion).

If we now apply an external electric field to the solution, the chaotic motion of each ion is supplemented by an occasional jump in the direction dictated by the interaction between the ionic charge and the field. But this is really a surprisingly tiny effect:

It can be shown that in a typical electric field of 1 volt/cm, a given ion will experience only about one field-directed (non-random) jump for every 10^5 random jumps. This translates into an average migration velocity of roughly 10^{-7} m sec^{-1} (10^{-4} mm sec^{-1}). Given that the radius of the H_2O molecule is close to 10^{-10} m, it follows that about 1000 such jumps are required to advance beyond a single solvent molecule!

The ions migrate Independently

All ionic solutions contain at least two kinds of ions (a cation and an anion), but may contain others as well. In the late 1870's, the physicist Friedrich Kohlrausch noticed that the limiting equivalent conductivities of salts that share a common ion exhibit constant differences.

electrolyte	Λ_0(25°C)	difference	electrolyte	Λ_0 (25°C)	difference
KCl LiCl	149.9 115.0	34.9	HCl HNO_3	426.2 421.1	4.9
KNO_3 $LiNO_3$	145.0 140.1	34.9	LiCl $LiNO_3$	115.0 110.1	4.9

These differences represent the differences in the conductivities of the ions that are *not* shared between the two salts. The fact that these differences are identical for two pairs of salts such as KCl/LiCl and KNO_3 /$LiNO_3$ tells us that the mobilities of the non-common ions K^+ and LI^+ are not affected by the accompanying anions.

Kohlrausch's law greatly simplifies estimates of Λ_0

This principle is known as *Kohlrausch's law of independent migration*, which states that *in the limit of infinite dilution*,

Each ionic species makes a contribution to the conductivity of the solution that depends only on the nature of that particular ion, and is independent of the other ions present.

Kohlrausch's law can be expressed as

$$\Lambda_0 = \Sigma\, \lambda^0_+ + \Sigma\, \lambda^0_-$$

This means that we can assign a *limiting equivalent conductivity* λ_0 to each kind of ion:

Limiting ionic equivalent conductivities at 25°C, S cm^{-1} eq^{-1}

cation	H_3O^+	NH_4^+	K^+	Ba^{2+}	Ag^+	Ca^{2+}	Sr^{2+}	Mg^{2+}	Na^+	Li^+
λ_0	349.98	73.57	73.49	63.61	61.87	59.47	59.43	53.93	50.89	38.66
anion	OH^-	SO_4^{2-}	Br^-	I^-	Cl^-	NO_3^-	ClO_3^-	CH_3COO^-	$C_2H_5COO^-$	$C_3H_7COO^-$
λ_0	197.60	80.71	78.41	76.86	76.30	71.80	67.29	40.83	35.79	32.57

Just as a compact table of thermodynamic data enables us to predict the chemical properties of a very large number of compounds, this compilation of equivalent conductivities of twenty different species yields reliable estimates of the of Λ_0 values for five times that number of salts.

We can now estimate weak electrolyte limiting conductivities

One useful application of Kohlrausch's law is to estimate the limiting equivalent conductivities of weak electrolytes which, as we observed above, cannot be found by extrapolation. Thus for acetic acid CH_3COOH ("HAc"), we combine the λ_0 values for H_3O^+ and CH_3COO^- given in the above table:

$$\Lambda^0_{HAc} = \lambda^0_{H+} + \lambda^0_{Ac-}$$

How fast do ions migrate in solution?

Movement of a migrating ion through the solution is brought about by a force exerted by the applied electric field. This force is proportional to the field strength and to the ionic charge. Calculations of the frictional drag are based on the premise that the ions are spherical (not always true) and the medium is continuous (never true) as opposed to being composed of discrete molecules. Nevertheless, the results generally seem to be realistic enough to be useful.

According to Newton's law, a constant force exerted on a particle will *accelerate* it, causing it to move faster and faster unless it is restrained by an opposing force. In the case of electrolytic conductance, the opposing force is frictional drag as the ion makes its way through the medium. The magnitude of this force depends on the radius of the ion and its primary hydration shell, and on the viscosity of the solution.

Eventually these two forces come into balance and the ion assumes a constant average velocity which is reflected in the values of λ_0 tabulated in the table above.

The relation between λ_0 and the velocity (known as the *ionic mobility* μ_0) is easily derived, but we will skip the details here, and simply present the results:

Anions are conventionally assigned negative μ_0 values because they move in opposite directions to the cations; the values shown here are absolute values $|\mu_0|$. Note also that the units are cm/sec per volt/cm, hence the cm^2 term.

Absolute limiting mobilities of ions at 25°C, (cm^2 $volt^{-1}$ sec^{-1}) × 100

cation	H_3O^+	NH_4^+	K^+	Ba^{2+}	Ag^+	Ca^{2+}	Sr^{2+}	Mg^{2+}	Na^+	Li^+
μ_0	0.362	0.0762	0.0762	0.0659	0.0642	0.0616	0.0616	0.0550	0.0520	0.0388
anion	OH^-	SO_4^{2-}	Br^-	I^-	Cl^-	NO_3^-	ClO_3^-	CH_3COO^-	$C_2H_5COO^-$	$C_3H_7COO^-$
μ_0	.2050	0.0827	0.0812	0.0796	0.0791	0.0740	0.0705	0.0461	0.0424	0.0411

As with the limiting conductivities, the trends in the mobilities can be roughly correlated with the charge and size of the ion. (Recall that negative ions tend to be larger than positive ions.)

Cations and anions carry different fractions of the current

In electrolytic conduction, ions having different charge signs move in opposite directions. Conductivity measurements give only the sum of the positive and negative ionic conductivities according to Kohlrausch's law, but they do not reveal how much of the charge is carried by each kind of ion. Unless their mobilities are the same, cations and anions do not contribute equally to the total electric current flowing through the cell.

Recall that an electric current is defined as a flow of electric charges; the current in amperes is the number of coulombs of charge moving through the cell per second. Because ionic solutions contain equal quantities of positive and negative charges, it follows that the current passing through the cell consists of positive charges moving toward the cathode, and negative charges moving toward the anode. But owing to mobility differences, cations and ions do not usually carry identical fractions of the charge.

Transference numbers are often referred to as *transport numbers*; either term is acceptable in the context of electrochemistry. The fraction of charge carried by a given kind of ion is known as the *transference number* $t_{\pm}$. For a solution of a simple binary salt,

Extra close brace or missing open brace

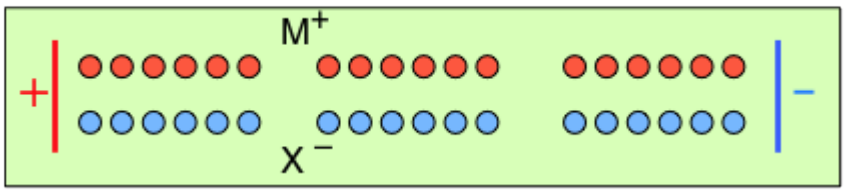
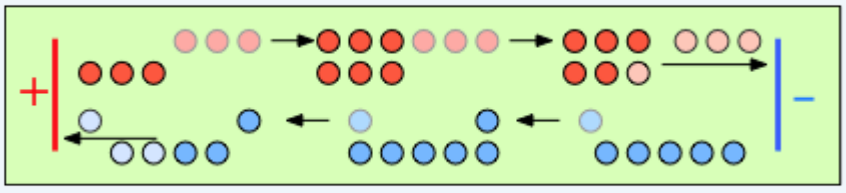
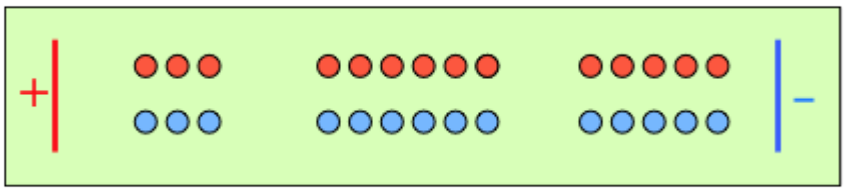
and

Extra close brace or missing open brace

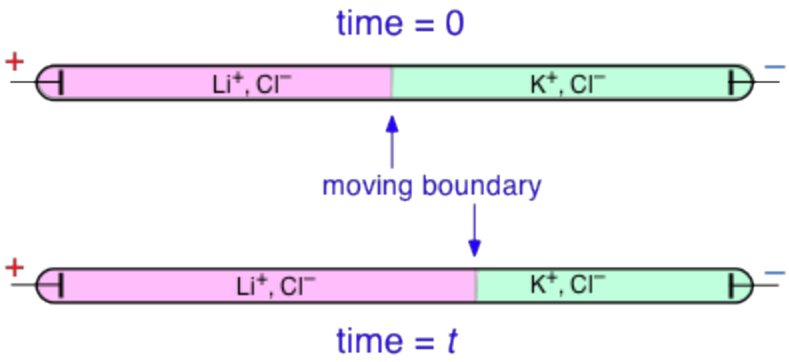
By definition,

$$t_+ + t_- = 1. \tag{8.10.9D.1}$$

To help you visualize the effects of non-identical transference numbers, consider a solution of M^+ X^- in which t_+ = 0.75 and t_- = 0.25. Let the cell be divided into three [imaginary] sections as we examine the distribution of cations and anions at three different stages of current flow.

	Initially, the concentrations of M^+ and X^- are the same in all parts of the cell.
	After 4 faradays of charge have passed through the cell, 3 eq of cations and 1 eq of anions have crossed any given plane parallel to the electrodes. Note that 3 anions are discharged at the anode, exactly balancing the number of cations discharged at the cathode.
	In the absence of diffusion, the ratio of the ionic concentrations near the electrodes equals the ratio of their transport numbers.

Transference numbers can be determined experimentally by observing the movement of the boundary between electrolyte solutions having an ion in common, such as LiCl and KCl:



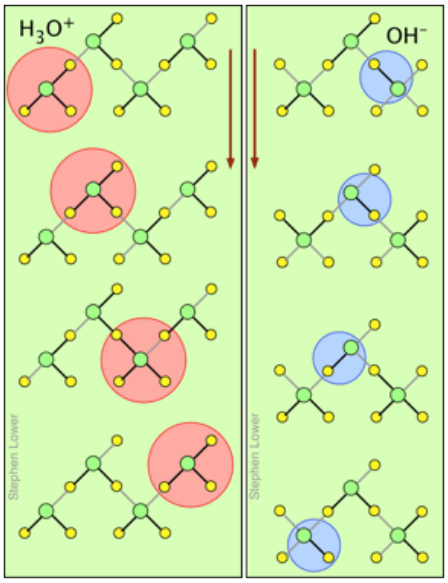
In this example, K^+ has a higher transference number than Li^+, but don't try to understand why the KCl boundary move to the left; the details of how this works are rather complicated and not important for the purposes of this this course.

H^+ and OH^- ions "migrate" without moving, and rapidly!

You may have noticed from the tables above that the hydrogen- and hydroxide ions have extraordinarily high equivalent conductivities and mobilities. This is a consequence of the fact that unlike other ions which need to bump and nudge their way through the network of hydrogen-bonded water molecules, these ions are *participants* in this network. By simply changing the H_2O partners they hydrogen-bond with, they can migrate "virtually". In effect, what migrates is the hydrogen-bonds, rather than the physical masses of the ions themselves.

This process is known as the *Grothuss Mechanism*. The shifting of the hydrogen bonds occurs when the rapid thermal motions of adjacent molecules brings a particular pair into a more favorable configuration for hydrogen bonding within the local molecular network. Bear in mind that what we refer to as "hydrogen ions" $H^+(aq)$ are really *hydronium ions* H_3O^+. It has been proposed that the larger aggregates $H_5O_2^+$and $H_9O_4^+$ are important intermediates in this process.

It is remarkable that this virtual migration process was proposed by Theodor Grotthuss in 1805 — just five years after the discovery of electrolysis, and he didn't even know the correct formula for water; he thought its structure was H–O–O–H.



These two diagrams will help you visualize the process. The successive downward rows show the first few "hops" made by the virtual H^+ and OH^-ions as they move in opposite directions toward the appropriate electrodes. (Of course, the same mechanism is operative in the absence of an external electric field, in which case *all* of the hops will be in random directions.)

Covalent bonds are represented by black lines, and hydrogen bonds by gray lines.

This page titled 8.10.9D: Ionic migration is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

8.10.9E: Some applications of electrolytic conduction

From the chemist's standpoint, the most important examples of conduction are in connection with electrochemical cells, electrolysis and batteries.

Determination of equilibrium constants

Owing to their high sensitivity, conductivity measurements are well adapted to the measurement of equilibrium constants for processes that involve very small ion concentrations. These include

- K_s values for sparingly soluble solids
- Autoprotolysis constants for solvents (such as K_w)
- Acid dissociation constants for weak acids

As long as the ion concentrations are so low, their values can be taken as activities, and limiting equivalent conductivities Λ_0 can be used directly.

Ion product of water

The very small conductivity of pure water makes it rather difficult to obtain a precise value for K_w; better values are obtained by measuring the potential of an appropriate galvanic cell. But the principle illustrated here might be applicable to other autoprotolytic solvents such as H_2SO_4.

Example 8.10.9E.1

Use the appropriate limiting molar ionic conductivities to estimate the autoprotolysis constant K_w of water at 25° C. Use the reaction equation

$$2\ H_2O \rightarrow H_3O^+ + OH^-.$$

Solution: The data we need are $\lambda_{H+} = 349.6$ and $\lambda_{OH-} = 199.1$ S cm^2 mol^{-1}.

The conductivity of water is $\kappa = [H^+]\ \lambda_{H+} + [OH^-]\ \lambda_{OH-}$ whose units work out to (mol cm^{-3}) (S cm^2 mol^{-1}). In order to express the ionic concentrations in molarities, we multiply the cm^{-3} term by (1 L / 1000 cm^{-3}), yielding

$$1000\ \kappa = [H^+]\ \lambda_{H+} + [OH^-]\ \lambda_{OH-} \text{ with units S cm}^{-1}\text{ L}^{-1}.$$

Recalling that in pure water, $[H^+] = [OH^-] = K_w^{½}$, we obtain

$$1000\ \kappa = (K_w^{½})(\lambda_{H+} + \lambda_{OH-}\ ) = (K_w^{½})(548.7\text{ S cm}^2\text{ mol}^{-1}).$$

Solving for K_w:

$$K_w = (1000\ \kappa / 548.7\text{ S cm}^2\text{ mol}^{-1})^2$$

Substituting Kohlrausch's water conductivity value of 0.043×10^{-6} S cm^{-1}) for κ gives

$$K_w = (1000 \times 0.043 \times 10^{-6}\text{ S cm}^{-1} / 548.7\text{ S cm}^2\text{ mol}^{-1})^2 = 0.614 \times 10^{-14}\text{ mol}^2\text{ cm}^{-6}\text{ (i.e., mol}^2\text{ L}^{-2}).$$

The accepted value for the autoprotolysis constant of water at 25° C is $K_w = 1.008 \times 10^{-14}$ mol^2 L^{-2}.

Example $8.10.9E.2$: Solubility products

A saturated solution of silver chloride AgCl has a conductance 2.28 x 10^{-6} S cm^{-1} at 25°C. The conductance of the water used to make up this solution is 0.116 x 10^{-6} S cm^{-1}. The limiting ionic conductivities of the two ions are λ_{Ag+}= 61.9 and λ_{Cl-} = 76.3 S cm^2 mol^{-1}. Use this information to estimate the molar solubility of AgCl.

The limiting molar conductivity of the solution is

$$\Lambda_o = \lambda_{Ag+} + \lambda_{Cl-} = 138.2 \text{ S cm}^{-1}.$$

The conductance due to the dissociated salt alone is the difference

$$(2.785 - 0.043) = 2.16 \text{ x } 10^{-6} \text{ S cm}^{-1}.$$

Substituting into the expression $\Lambda = 1000\kappa/c$ yields

= 1.56 x 10^{-5} mol/1000 cm^3or 1.56 x 10^{-5} mol L^{-1}.

Conductometric titration

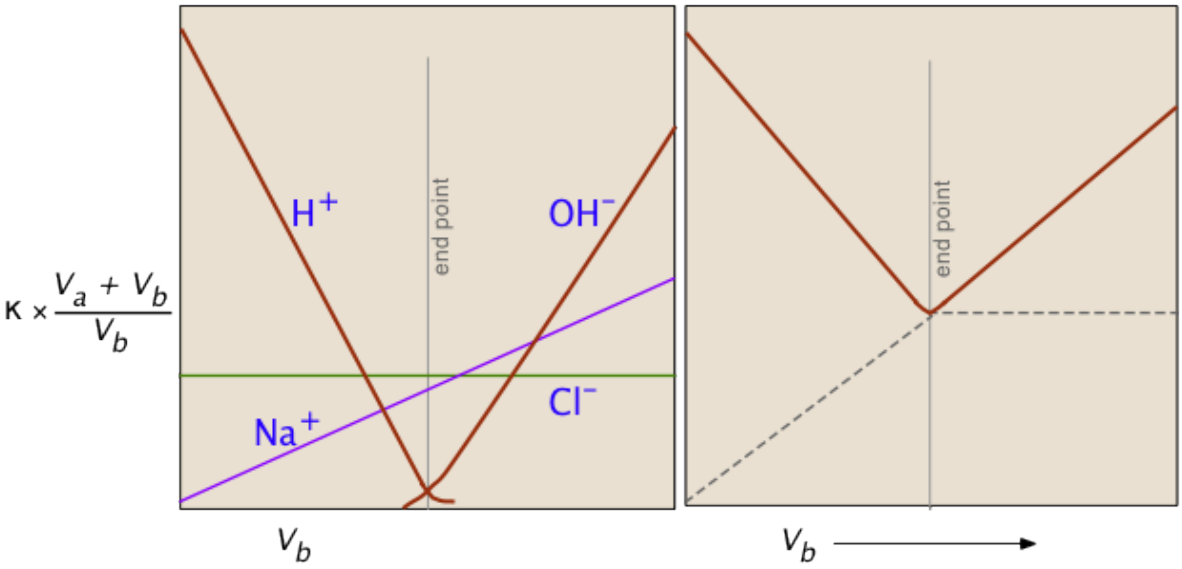
A chemical reaction in which there is a significant change in the number or mobilities of ionic species can be followed by monitoring the change in conductance. Many acid-base reactions fall into this category. In *conductometric titration*, conductometry is employed to detect the end-point of a titration.



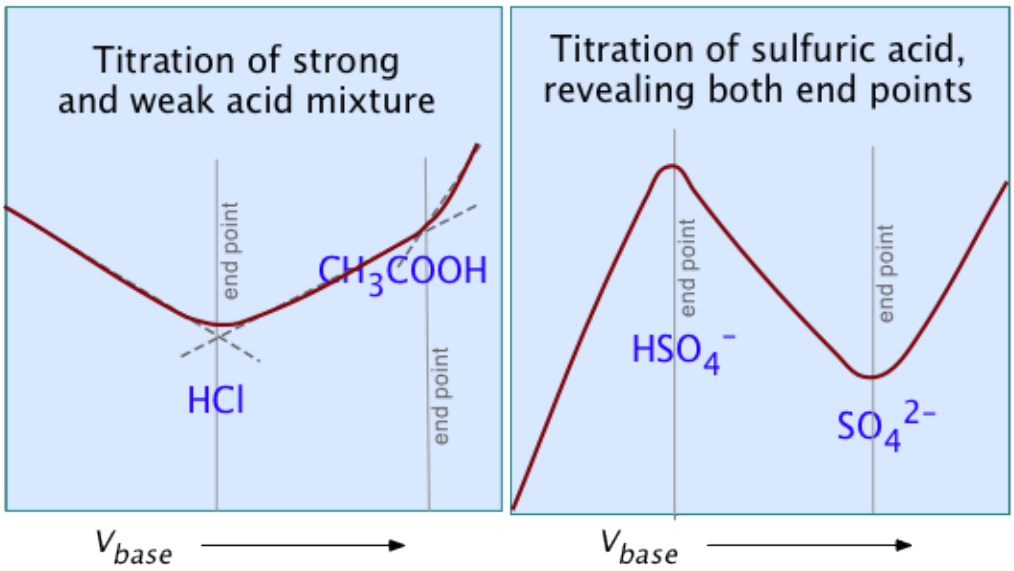
Consider, for example, the titration of the strong acid HCl by the strong base NaOH. In ionic terms, the process can be represented as

$$\mathbf{H^+} + Cl^- + Na^+ + \mathbf{OH^-} \rightarrow \mathbf{H_2O} + Na^+ + Cl^-$$

At the end point, only two ionic species remain, compared to the four during the initial stages of the titration, so the conductivity will be at a minimum. Beyond the end point, continued addition of base causes the conductivity to rise again. The very large mobilities of the $H^+$and OH^- ions cause the conductivity to rise very sharply on either side of the end point, making it quite easy to locate.



The plot on the left depicts the conductivities due to the individual ionic species. But of course the conductivity we measure is just the sum of all of these (Kohlrausch's law of independent migration), so the plot on the right corresponds to what you actually see when the titration is carried out. The factor $(V_a + V_b)/V_b$ compensates for the dilution of the solution as more base is added.



For most ordinary acid-base titrations, conductometry rarely offers any special advantage over regular volumetric analysis or potentiometry. But in some special cases such as those illustrated here, it is the only method capable of yielding useful results.

Ground- and soil conduction

Most people think of electrolytic conduction as something that takes place mainly in batteries, electrochemical plants and in laboratories, but by far the biggest and most important electrolytic system is the earth itself, or at least the thin veneer of soil sediments that coat much of its surface.

Soils are composed of sand and humic solids within which are embedded pore spaces containing gases (air and microbial respiration products) and water. This water, both that which fills the pores as well as that which is adsorbed to soil solids, contains numerous dissolved ions on which soil conductivity mainly depends. These ions include exchangeable cations present on the surfaces of the clay components of soils. Although these electrolyte channels are small and tortuously connected, they are present in huge numbers and offer an abundance of parallel conduction paths, so the ground itself acts as a surprisingly efficient

conductor.



There is no better illustration of this than the use of a ground path to transmit a current of up to 1800 amperes along the 1360-km path of the Pacific DC Intertie that extends from the Celilo Converter Station (view below) in northern Oregon to Los Angeles. This system normally employs a two-conductor power line that operates at ±500,000 volts DC, but when one of these conductors is out of service, the ground path acts as a substitute conductor. This alternative path is said to have a lower resistance than the normal metallic conductor!

From an electrochemical standpoint, the most interesting aspect of this system is the manner in which the current flows into or out of the ground. The grounding system at Celilo is composed of over 1000 cast-iron electrodes buried in a circular 3.3-km trench of petroleum coke which acts as the working electrode. At the Los Angeles end, grounding is achieved by means of 24 silicon-iron alloy electrodes submerged in the nearby Pacific Ocean.

On a much smaller scale, single-wire earth return systems are often employed to supply regular single-phase ac power to remote regions, or as the return path for high-voltage DC submarine cables. For direct current submarine systems, a copper cable laid on the bottom is suitable for the cathode. The anodes are normally graphite rods surrounded by coke.

You may have noticed that the pole-mounted step-down transformers used to distribute single-phase ac power in residential neighborhoods are connected to the high-voltage (10 Kv or so) primary line by only a single wire. The return circuit back to the local substation is completed by a ground connection which runs down the pole to a buried electrode.



The Celilo Converter Station is located at The Dalles, Oregon

Other applications of ground conductivity

Ground-wave radio propagation

During the daytime, radio transmission at frequencies below about 5 Mhz (such as in the old standard AM broadcast band) depends entirely on so-called ground waves that follow the curvature of the earth. This occurs because that portion of the vertically-polarized wavefronts in contact with the earth induce an electrolytic current in the ground, causing their lower portions to travel more slowly, bending their pathways in toward the earth.

Agricultural and environmental soils assessment

Conductivity has long been used as a tool to assess the salinity of agricultural soils — a serious problem in irrigated regions, where evaporation of irrigation water over the years can raise salinity to levels that can reduce crop yields. Because other soil characteristics (moisture content, density, and mineralogy, fertilization) also play important roles, some care is required to correctly interpret measurements. Measuring devices that can be drawn behind a tractor and are equipped with GPS receivers allow the production of conductivity maps for entire fields.

Archaeological exploration

Buried artifacts such as stone walls and foundations and large metallic objects can be located by a series of conductivity measurements at archaeological sites.

This page titled 8.10.9E: Some applications of electrolytic conduction is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9: CHEMICAL BONDING AND MOLECULAR STRUCTURE

Why do some atoms join together to form molecules, but others do not? Why is the CO_2 molecule linear whereas H_2O is bent? How can we tell? How does hemoglobin carry oxygen through our bloodstream? There is no topic more fundamental to Chemistry than the nature of the chemical bond, and the introduction you find here will provide you with an overview of the fundamentals and a basis for further study.

9.1: Three Views of Chemical Bonding

9.2: Molecules - Properties of Bonded Atoms

9.3: Models of Chemical Bonding

9.4: Polar Covalence

9.5: Molecular Geometry

9.6: The Hybrid Orbital Model

9.7: The Hybrid Orbital Model II

9.8: Molecular Orbital Theory

9.9: Bonding in Coordination Complexes

9.10: Bonding in Metals

9.11: Bonding in Semiconductors

9.12: The Shared-Electron Covalent Bond

This page titled 9: Chemical Bonding and Molecular Structure is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Three Views of Chemical Bonding

Chemical bonds form when electrons can be simultaneously close to two or more nuclei, but beyond this, there is no simple, easily understood theory that would not only explain why atoms bind together to form molecules, but would also predict the three-dimensional structures of the resulting compounds as well as the energies and other properties of the bonds themselves. Unfortunately, no one theory exists that accomplishes these goals in a satisfactory way for all of the many categories of compounds that are known. Moreover, it seems likely that if such a theory does ever come into being, it will be far from simple.

When we are faced with a scientific problem of this complexity, experience has shown that it is often more useful to concentrate instead on developing **models**. A scientific model is something like a theory in that it should be able to explain observed phenomena and to make useful predictions. But whereas a theory can be discredited by a single contradictory case, a model can be useful even if it does not encompass all instances of the phenomena it attempts to explain. We do not even require that a model be a credible representation of reality; all we ask is that be able to explain the behavior of those cases to which it is applicable in terms that are consistent with the model itself. An example of a model that you may already know about is the kinetic molecular theory of gases. Despite its name, this is really a model (at least at the level that beginning students use it) because it does not even try to explain the observed behavior of real gases. Nevertheless, it serves as a tool for developing our understanding of gases, and as a starting point for more elaborate treatments.

Given the extraordinary variety of ways in which atoms combine into aggregates, it should come as no surprise that a number of useful bonding models have been developed. Most of them apply only to certain classes of compounds, or attempt to explain only a restricted range of phenomena. In this section we will provide brief descriptions of some of the bonding models; the more important of these will be treated in much more detail in later parts of this chapter.

Classical models

By classical, we mean models that do not take into account the quantum behavior of small particles, notably the electron. These models generally assume that electrons and ions behave as point charges which attract and repel according to the laws of electrostatics. Although this completely ignores what has been learned about the nature of the electron since the development of quantum theory in the 1920's, these classical models have not only proven extremely useful, but the major ones also serve as the basis for the chemist's general classification of compounds into "covalent" and "ionic" categories.

Electrostatic (Ionic Bonding)

Ever since the discovery early in the 19th century that solutions of salts and other electrolytes conduct electric current, there has been general agreement that the forces that hold atoms together must be electrical in nature. Electrolytic solutions contain ions having opposite electrical charges, opposite charges attract, so perhaps the substances from which these ions come consist of positive and negatively charged atoms held together by electrostatic attraction.

It turns out that this is not true generally, but a model built on this assumption does a fairly good job of explaining a rather small but important class of compounds that are called ionic solids. The most well known example of such a compound is sodium chloride, which consists of two interpenetrating lattices of Na^+ and Cl^- ions arranged in such as way that every ion of one type is surrounded (in three dimensional space) by six ions of opposite charge.

The main limitation of this model is that it applies really well only to the small class of solids composed of Group 1 and 2 elements with highly electronegative elements such as the halogens. Although compounds such as $CuCl_2$ dissociate into ions when they dissolve in water, the fundamental units making up the solid are more like polymeric chains of covalently-bound $CuCl_2$ molecules that have little ionic character.

Shared-Electrons (Covalent Bonding)

This model originated with the theory developed by G.N. Lewis in 1916, and it remains the most widely-used model of chemical bonding. The essential element s of this model can best be understood by examining the simplest possible molecule. This is the **hydrogen molecule ion** H_2^+, which consists of two nuclei and one electron. First, however, think what would happen if we tried

to make the even simpler molecule H_2^{2+}. Since this would consist only of two protons whose electrostatic charges would repel each other at all distances, it is clear that such a molecule cannot exist; something more than two nuclei are required for bonding to occur.

In the hydrogen molecule ion H_2^+ we have a third particle, an electron. The effect of this electron will depend on its location with respect to the two nuclei. If the electron is in the space between the two nuclei, it will attract both protons toward itself, and thus toward each other. If the total attraction energy exceeds the internuclear repulsion, there will be a net bonding effect and the molecule will be stable. If, on the other hand, the electron is off to one side, it will attract both nuclei, but it will attract the closer one much more strongly, owing to the inverse-square nature of Coulomb's law. As a consequence, the electron will now help the electrostatic repulsion to push the two nuclei apart.

We see, then, that the electron is an essential component of a chemical bond, but that it must be in the right place: between the two nuclei. Coulomb's law can be used to calculate the forces experienced by the two nuclei for various positions of the electron. This allows us to define two regions of space about the nuclei, as shown in the figure. One region, the binding region, depicts locations at which the electron exerts a net binding effect on the new nuclei. Outside of this, in the antibinding region, the electron will actually work against binding.

This simple picture illustrates the number one rule of chemical bonding: chemical bonds form when electrons can be simultaneously close to two or more nuclei. It should be pointed out that this principle applies also to the ionic model; as will be explained later in this chapter, the electron that is "lost" by a positive ion ends up being closer to more nuclei (including the one from whose electron cloud it came) in the compound.

- **The polar covalent model**: A purely covalent bond can only be guaranteed when the electronegativities (electron-attracting powers) of the two atoms are identical. When atoms having different electronegativities are joined, the electrons shared between them will be displaced toward the more electronegative atom, conferring a polarity on the bond which can be described in terms of percent ionic character. The polar covalent model is thus an generalization of covalent bonding to include a very wide range of behavior.
- **The Coulombic model**: This is an extension of the ionic model to compounds that are ordinarily considered to be non-ionic. Combined hydrogen is always considered to exist as the **hydride ion** H^-, so that methane can be treated as if it were $C^{4+}H^{-}_4$. This is not as bizarre as it might seem at first if you recall that the proton has almost no significant size, so that it is essentially embedded in an electron pair when it is joined to another atom in a covalent bond. This model, which is not as well known as it deserves to be, has considerable predictive power, both as to bond energies and structures.
- **The VSEPR model**: The "valence shell electron repulsion" model is not so much a model of chemical bonding as a scheme for explaining the shapes of molecules. It is based on the quantum mechanical view that bonds represent electron clouds-physical regions of negative electric charge that repel each other and thus try to stay as far apart as possible.

Quantum Models

Quantum models of bonding take into account the fact that a particle as light a the electron cannot really be said to be in any single location. The best we can do is define a region of space in which the probability of finding the electron has some arbitrary value which will always be less than unity. The shape of this volume of space is called an *orbital* and is defined by a mathematical function that relates the probability to the (x,y,z) coordinates of the molecule. Like other models of bonding, the quantum models attempt to show how more electrons can be simultaneously close to more nuclei. Instead of doing so through purely geometrical arguments, they attempt this by predicting the nature of the orbitals which the valence electrons occupy in joined atoms.

- **The hybrid orbital model**: This was developed by Linus Pauling in 1931 and was the first quantum-based model of bonding. It is based on the premise that if the atomic s, p, and d orbitals occupied by the valence electrons of adjacent atoms are combined in a suitable way, the *hybrid orbitals* that result will have the character and directional properties that are consistent with the bonding pattern in the molecule. The rules for bringing about these combinations turn out to be remarkably simple, so once they were worked out it became possible to use this model to predict the bonding behavior in a wide variety of molecules. The hybrid orbital model is most usefully applied to the p-block elements the first two rows of the periodic table, and is especially important in organic chemistry.
- **The molecular orbital model**: This model takes a more fundamental approach by regarding a molecule as a collection of valence electrons and positive cores. Just as the nature of atomic orbitals derives from the spherical symmetry of the atom, so will the properties of these new *molecular orbitals* be controlled by the interaction of the valence electrons with the multiple positive centers of these atomic cores. These new orbitals, unlike those of the hybrid model, are *delocalized*; that is, they do not "belong" to any one atom but extend over the entire region of space that encompasses the bonded atoms. The

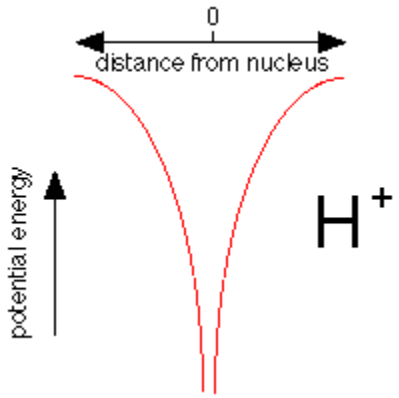
available (valence) electrons then fill these orbitals from the lowest to the highest, very much as in the *Aufbau principle* that you learned for working out atomic electron configurations. For small molecules (which are the only ones we will consider here), there are simple rules that govern the way that atomic orbitals transform themselves into molecular orbitals as the separate atoms are brought together. The real power of molecular orbital theory, however, comes from its mathematical formation which lends itself to detailed predictions of bond energies and other properties.

- **The electron-tunneling model**: A common theme uniting all of the models we have discussed is that bonding depends on the fall in potential energy that occurs when opposite charges are brought together. In the case of covalent bonds, the shared electron pair acts as a kind of "electron glue" between the joined nuclei. In 1962, however, it was shown that this assumption is not strictly correct, and that instead of being concentrated in the space between the nuclei, the electron orbitals become even more concentrated around the bonded nuclei. At the same time however, they are free to "move" between the two nuclei by a process known as *tunneling*. This refers to a well-known quantum mechanical effect that allows electrons (or other particles small enough to exhibit wavelike properties) to pass ("tunnel") through a barrier separating two closely adjacent regions of low potential energy. One result of this is that the effective volume of space available to the electron is increased, and according to the *uncertainty principle* this will reduce the kinetic energy of the electron.

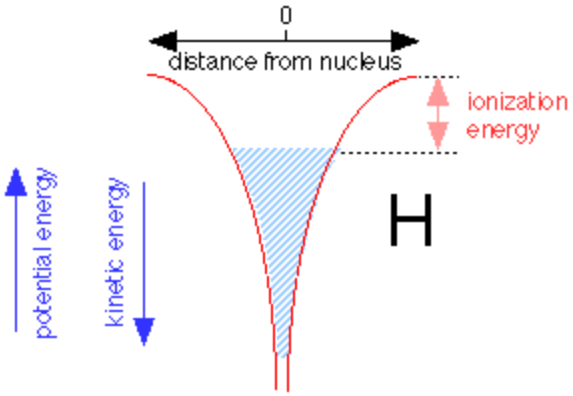
The electron-tunneling model

According to this model, the bonding electrons act as a kind of fluid that concentrates in the region of each nucleus (lowering the potential energy) and at the same time is able to freely flow between them (reducing the kinetic energy). A summary of the concept, showing its application to a simple molecule, is shown on the next page. Despite its conceptual simplicity and full acknowledgment of the laws of quantum mechanics, this model is not widely known and is rarely taught.

Chemical bonding occurs when one or more electrons can be simultaneously close to two nuclei. But how can this be arranged? The conventional picture of the shared electron bond places the bonding electrons in the region **between** the two nuclei. This makes a nice picture, but it is not consistent with the principle that opposite charges attract. This would imply that the electrons would be "happiest" (at the lowest potential energy) when they are very close to a nucleus, not half a bond-length away from two of them!



This plot shows how the potential energy of an electron in the hydrogen atom varies with its distance from the nucleus. Notice how the energy falls without limit as the electron approaches the nucleus, represented here as a proton, H^+. If potential energy were the only consideration, the electron would fall right into the nucleus where its potential energy would be minus infinity.

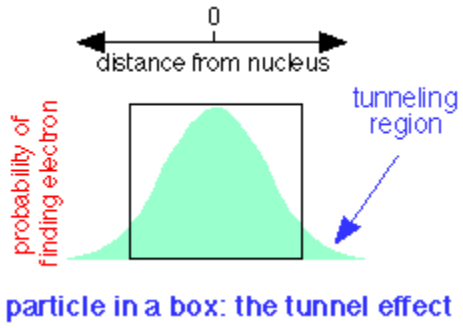


When an electron is added to the proton to make a neutral hydrogen atom, it tries to get as close to the nucleus as possible. The Heisenberg uncertainty principle requires the total energy of the electron energy to increase as the volume of space it occupies diminishes. As the electron gets closer to the nucleus, the nuclear charge confines the electron to such a tiny volume of space that its energy rises, allowing it to "float" slightly away from the nucleus without ever falling into it.

The shaded region above shows the range of energies and distances from the nucleus the electron is able to assume within the 1s orbital. The electron can thus be regarded as a fluid that occupies a vessel whose walls conform to the red potential energy curves shown above. Note that as the potential energy falls, the kinetic energy increases, but only half as fast (this is called the *virial theorem*). Thus close to the nucleus, the kinetic energy is large and so is the electron's effective velocity. The top of the shaded area defines the work required to raise its potential energy to zero, thus removing it from the atom; this corresponds, of course, to the ionization energy.

The Tunneling Effect

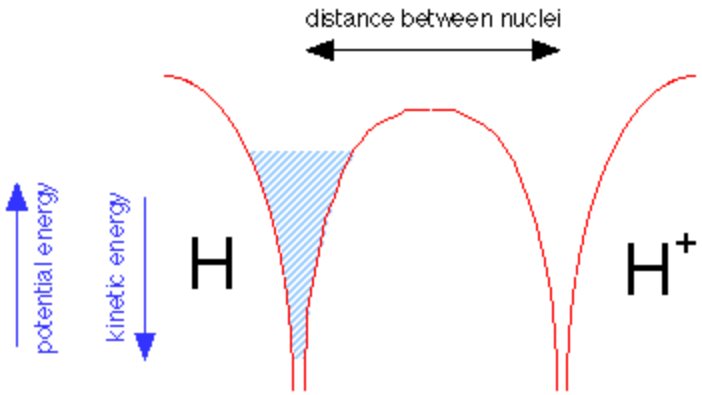
A quantum particle can be described by a waveform which is the plot of a mathematical function related to the probability of finding the particle at a given location at any time. If the particle is confined to a box, it turns out that the wave does not fall to zero at the walls of the box, but has a finite probability of being found outside it. This means that a quantum particle is able to penetrate, or "tunnel through" its confining boundaries. This remarkable property is called the *tunnel effect*.



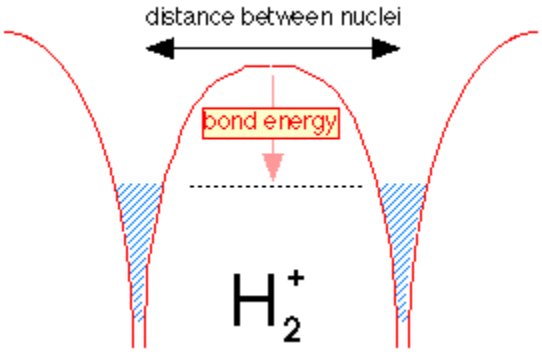
In terms of the electron fluid model introduced above, the fluid is able to "leak out" of the atom if another low-energy location can be found nearby.

Electron tunneling in the simplest molecule

Suppose we now bring a bare proton up close to a hydrogen atom. Each nucleus has its own potential well, but only that of the hydrogen atom is filled, as indicated by the shading in the leftmost potential well.



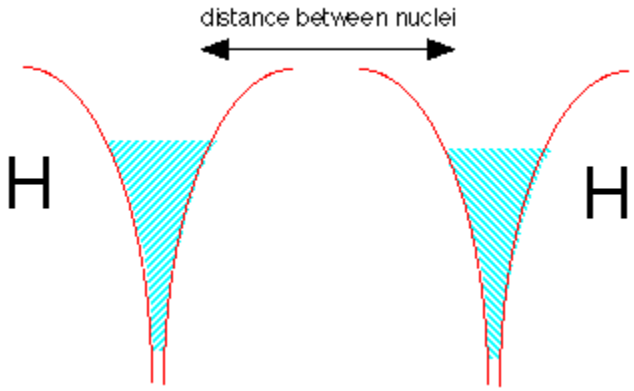
But the electron fluid is able to tunnel through the potential energy barrier separating the two wells; like any liquid, it will seek a common level in the two sides of the container as shown below. The electron is now "simultaneously close to two nuclei" while never being in between them. Bear in mind that this would be physically impossible for a real liquid composed of real molecules; this is purely a quantum effect that is restricted to a low-mass particle such as the electron.



Because the same amount of electron fluid is now shared between the two wells, its level in both is lower. The difference between what it is now and what is was before corresponds to the **bond energy** of the hydrogen molecule ion.

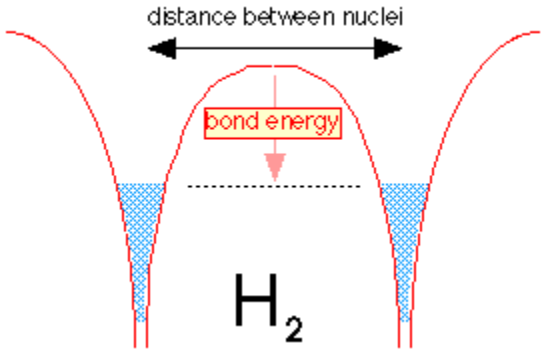
The dihydrogen molecule

Now let's make a molecule of **dihydrogen**. We start with two hydrogen atoms, each with one electron. But there is a problem here: both potential energy wells are already filled with electron fluid; there is no room for any more without pushing the energy way up.



But quantum theory again comes to the rescue! If the two electrons have opposite spins, the two fluids are able to interpenetrate each other, very much as two gases are able to occupy the same container. This is depicted by the double

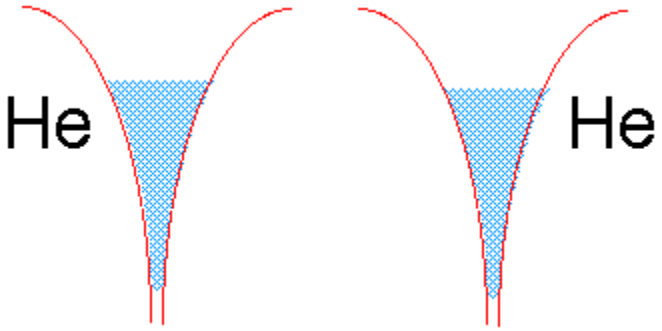
shading in the diagram below.



When the two hydrogen atoms are within tunneling distance, half of the electron fluid (really the probability of finding the electron) from each well flows into the other well. Because the two fluids are able to interpenetrate, the level is not much different from what it was in the H_2^+ ion, but the greater *density* of the electron-fluid between the two nuclei makes H_2 a strongly bound molecule.

So why does dihelium not exist?

If we tried to join two helium atoms in this way, we would be in trouble. The electron well of He already contains two electrons of opposite spin. There is no room for more electron fluid (without raising the energy), and thus no way the electrons in either He atom can be simultaneously close to two nuclei.



Ionic Bonding

Even before G.N.Lewis developed his theory of the shared electron pair bond, it was believed that bonding in many solid salts could be satisfactorily explained on the basis of simple electrostatic forces between the positive and negative ions which are assumed to be the basic molecular units of these compounds. Lewis himself continued to recognize this distinction, which has continued to be a part of the tradition of chemistry; the shared electron pair bond is known as the covalent bond, while the other type is the ionic or electrovalent bond.

The covalent bond is formed when two atoms are able to share electrons:



whereas the ionic bond is formed when the "sharing" is so unequal that an electron from atom A is completely lost to atom B, resulting in a pair of ions:

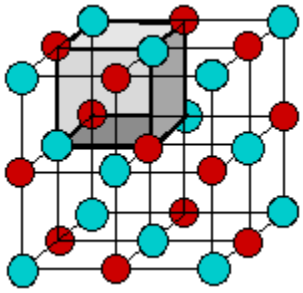


The two extremes of electron sharing represented by the covalent and ionic models appear to be generally consistent with the

observed properties of molecular and ionic solids and liquids. But does this mean that there are really two kinds of chemical bonds, ionic and covalent?

Bonding in ionic solids

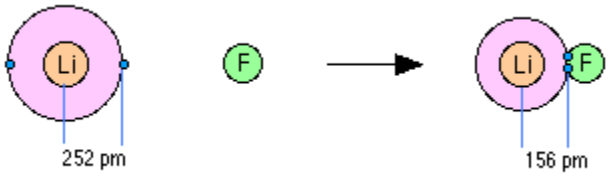
According to the ionic electrostatic model, solids such as NaCl consist of positive and negative ions arranged in a crystal lattice. Each ion is attracted to neighboring ions of opposite charge, and is repelled by ions of like charge; this combination of attractions and repulsions, acting in all directions, causes the ion to be tightly fixed in its own location in the crystal lattice.



Since electrostatic forces are nondirectional, the structure of an ionic solid is determined purely by geometry: two kinds of ions, each with its own radius, will fall into whatever repeating pattern will achieve the lowest possible potential energy. Surprisingly, there are only a small number of possible structures.

Is there such as thing as a "purely" ionic bond?

When two elements form an ionic compound, is an electron really lost by one atom and transferred to the other one? In order to deal with this question, consider the data on the ionic solid LiF. The average radius of the neutral Li atom is about 2.52Å.



Now if this Li atom reacts with an atom of F to form LiF, what is the average distance between the Li nucleus and the electron it has "lost" to the fluorine atom? The answer is 1.56Å; *the electron is now closer to the lithium nucleus than it was in neutral lithium!* So the answer to the above question is both yes and no: yes, the electron that was now in the 2*s* orbital of Li is now within the grasp of a fluorine 2*p* orbital, but no, the electron is now even **closer** to the Li nucleus than before, so how can it be "lost"? The one thing that is inarguably true about LiF is that there are more electrons closer to positive nuclei than there are in the separated Li and F atoms. But this is just the condition that gives rise to all forms of chemical bonding:

Chemical bonds form when electrons can be simultaneously near two or more nuclei

It is obvious that the electron-pair bond brings about this situation, and this is the reason for the stability of the covalent bond. What is not so obvious (until you look at the numbers such as were quoted for LiF above) is that the "ionic" bond results in the same condition; even in the most highly ionic compounds, both electrons are close to both nuclei, and the resulting mutual attractions bind the nuclei together. This being the case, is there really any fundamental difference between the ionic and covalent bond?

The answer, according to modern chemical thinking is probably "no"; in fact, there is some question as to whether it is realistic to consider that these solids consist of "ions" in the usual sense. The preferred picture that seems to be emerging is one in which the electron orbitals of adjacent atom pairs are simply skewed so as to place more electron density around the "negative" element

than around the "positive" one.

This being said, it must be reiterated that the ionic model of bonding is a useful one for many purposes, and there is nothing wrong with using the term "ionic bond" to describe the interactions between the atoms in "ionic solids" such as LiF and NaCl.

Polar covalence

If there is no such thing as a "completely ionic" bond, can we have one that is completely covalent? The answer is yes, if the two nuclei have equal electron attracting powers. This situation is guaranteed to be the case with homonuclear diatomic molecules—molecules consisting of two identical atoms. Thus in Cl_2, O_2, and H_2, electron sharing between the two identical atoms must be exactly even; in such molecules, the center of positive charge corresponds exactly to the center of negative charge: halfway between the two nuclei.

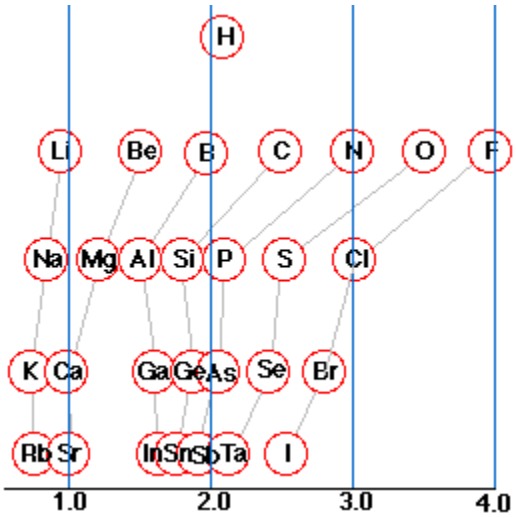
Electronegativity

This term was introduced earlier in the course to denote the relative electron-attracting power of an atom. The electronegativity is not the same as the electron affinity; the latter measures the amount of energy released when an electron from an external source "falls into" a vacancy within the outermost orbital of the atom to yield an isolated negative ion.

The products of bond formation, in contrast, are not ions and they are not isolated; the two nuclei are now drawn closely together by attraction to the region of high electron density between them. Any shift of electron density toward one atom takes place at the energetic expense of stealing it from the other atom.

It is important to understand that electronegativity is not a measurable property of an atom in the sense that ionization energies and electron affinity are; electronegativity is a property that an atom displays when it is bonded to another. Any measurement one does make must necessarily depend on the properties of both of the atoms.

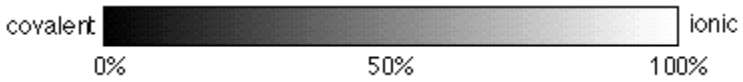
By convention, electronegativities are measured on a scale on which the highest value, 4.0, is arbitrarily assigned to fluorine. A number of electronegativity scales have been proposed, each based on slightly different criteria. The most well known of these is due to Linus Pauling, and is based on a study of bond energies in a variety of compounds.



The periodic trends in electronegativity are about what one would expect; the higher the nuclear charge and the smaller the atom, the more strongly attractive will it be to an outer-shell electron of an atom within binding distance. The division between the metallic and nonmetallic elements is largely that between those that have Pauling electronegativies greater than about 1.7, and those that have smaller electronegativities.

The greater the electronegativity difference between two elements **A** and **B**, the more **polar** will be their molecule **AB**. It is important to point out, however, that the pairs having the greatest electronegativity differences, the alkali halides, are solids order ordinary conditions and exist as molecules only in the rarefied conditions of the gas phase. Even these **ionic solids** possess a certain amount of covalent character, so, as discussed above, there is no such thing as a "purely ionic" bond. It has become

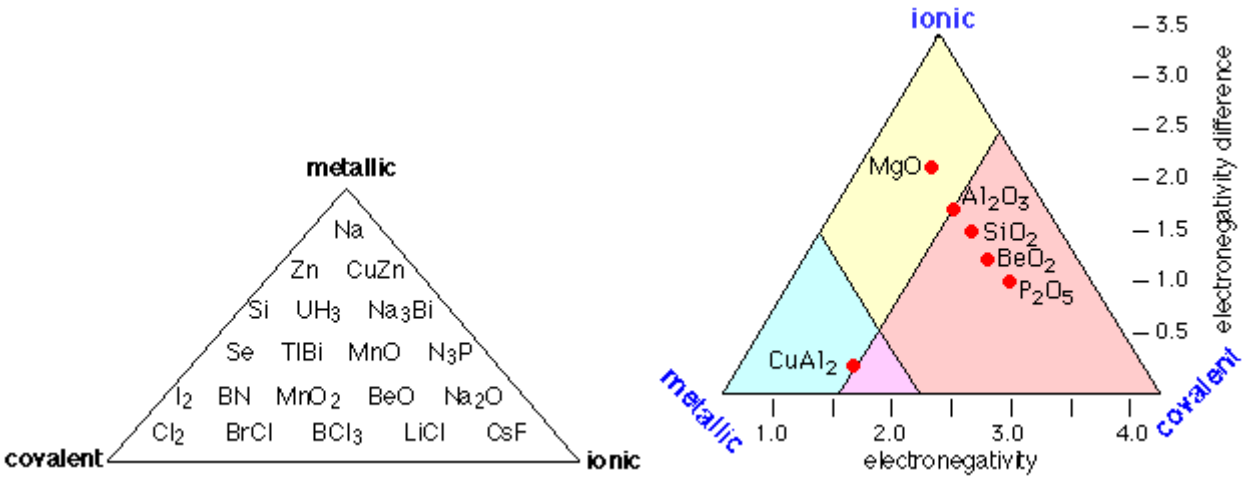
more common to place binary compounds on a scale something like that shown here, in which the degree of shading is a rough indication of the number of compounds at any point on the covalent-ionic scale.



Covalent or ionic: a false dichotomy

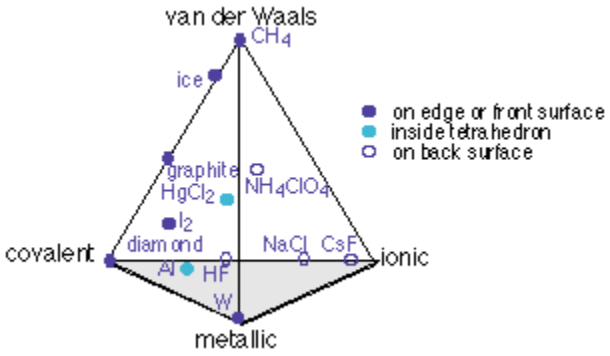
The covalent-ionic continuum described above is certainly an improvement over the old covalent *-versus* - ionic dichotomy that existed only in the textbook and classroom, but it is still only a one-dimensional view of a multidimensional world, and thus a view that hides more than it reveals. The main thing missing is any allowance for the type of bonding that occurs between more pairs of elements than any other: metallic bonding. Intermetallic compounds are rarely even mentioned in introductory courses, but since most of the elements are metals, there are a lot of them, and many play an important role in metallurgy. In metallic bonding, the valence electrons lose their association with individual atoms; they form what amounts to a mobile "electron fluid" that fills the space between the crystal lattice positions occupied by the atoms, (now essentially positive ions.) The more readily this electron delocalization occurs, the more "metallic" the element.

Thus instead of the one-dimension chart shown above, we can construct a triangular diagram whose corners represent the three extremes of "pure" covalent, ionic, and metallic bonding.



We can take this a step farther by taking into account collection of weaker binding effects known generally as van der Waals forces. Contrary to what is often implied in introductory textbooks, these are the major binding forces in most of the common salts that are not alkali halides; these include NaOH, $CaCl_2$, $MgSO_4$. They are also significant in solids such as $CuCl_2$ and solid SO_3 in which infinite covalently-bound chains are held together by ion-induced dipole and similar forces.

The only way to represent this four-dimensional bonding-type space in two dimensions is to draw a projection of a tetrahedron, each of its four corners representing the "pure" case of one type of bonding.



Note that some of the entries on this diagram (ice, CH_4, and the two parts of NH_4ClO_4) are covalently bound units, and their placement refers to the binding between these units. Thus the H_2O molecules in ice are held together mainly by hydrogen bonding, which is a van der Waals force, with only a small covalent contribution.

Note: the triangular and tetrahedral diagrams above were adapted from those in the excellent article by William B. Jensen, "Logic, history and the chemistry textbook", Part II, J. Chemical Education 1998: 817-828.

This page titled 9.1: Three Views of Chemical Bonding is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.2: Molecules - Properties of Bonded Atoms

Learning Objectives

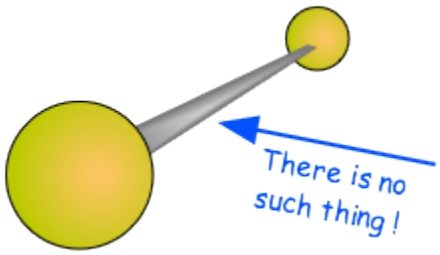
Make sure you thoroughly understand the following essential ideas which are presented below.

- How would *you* define a *chemical bond*?
- What is meant by the *connectivity* of a molecule? What additional information might be needed in order to specify its structure?
- Explain the difference between *stability* and *reactivity*, and how these factors might prevent a given structure from existing long enough to qualify as a *molecule*.
- Sketch out a potential energy curve for a typical diatomic molecule, and show how it illustrates the *bond energy* and *bond length*.
- Explain how the heat released or absorbed in a chemical reaction can be related to the bond energies of the reactants and products.
- State the major factors that determine the distance between two bonded atoms.
- Describe, in a general way, how the *infrared spectrum* of a substance can reveal details about its molecular structure.

The concept of chemical bonding lies at the very core of Chemistry; it is what enables about one hundred elements to form the more than fifty million known chemical substances that make up our physical world. Before we get into the theory of chemical bonding, we need to define what we are talking about: Exactly what is a chemical bond? And what observable properties can we use to distinguish one kind of bond from another? This is the first of ten lessons that will help familiarize you with the fundamental concepts of this very broad subject.

What is a chemical bond?

You probably learned some time ago that chemical bonds are what hold atoms together to form the more complicated aggregates that we know as molecules and extended solids. Chemists talk about bonds all the time, and draw pictures of them as lines joining atom symbols. Teachers often identify them as the little sticks that connect the spheres that represent atoms in a plastic molecular model. So it's not surprising that we sometimes tend to think of chemical bonds as “things”. But no one has ever *seen* a chemical bond, and there is no reason to believe that they really even exist as physical objects.



"SOMETIMES IT SEEMS to me that a bond between two atoms has become so real, so tangible, so friendly, that I can almost see it. Then I awake with a little shock, for a chemical bond is not a real thing. It does not exist. No one has ever seen one. No one ever can. It is a figment of our own imagination." **C.A. Coulson** (1910-1974) was an English theoretical chemist who played a central role in the development of quantum theories of chemical bonding.

It is more useful to regard a chemical bond as an *effect* that causes certain atoms to join together to form enduring structures that have unique physical and chemical properties.

So although the "chemical bond" is no more than a convenient fiction, *chemical bonding*, which leads to the near-infinity of substances (31 million in mid-2007), lies at the very core of chemistry. The forces that hold bonded atoms together are basically just the same kinds of electrostatic attractions that bind the electrons of an atom to its positively-charged nucleus;

Chemical bonding occurs when one or more electrons are simultaneously attracted to *two* (or more) nuclei.

This is the most important fact about chemical bonding that you should know, but it is not of itself a workable *theory* of bonding because it does not describe the conditions under which bonding occurs, nor does it make useful predictions about the properties of the bonded atoms.

What is a molecule?

Even at the end of the 19th century, when compounds and their formulas had long been in use, some prominent chemists doubted that molecules (or atoms) were any more than a convenient model. Molecules suddenly became real in 1905, when Albert Einstein showed that Brownian motion, the irregular microscopic movements of tiny pollen grains floating in water, could be directly attributed to collisions with molecule-sized particles.

Most people think of molecules as the particles that result when atoms become joined together in some way. This conveys the general picture, but a somewhat better definition that we will use in these lessons is:

A molecule is an aggregate of atoms that possesses distinctive observable properties

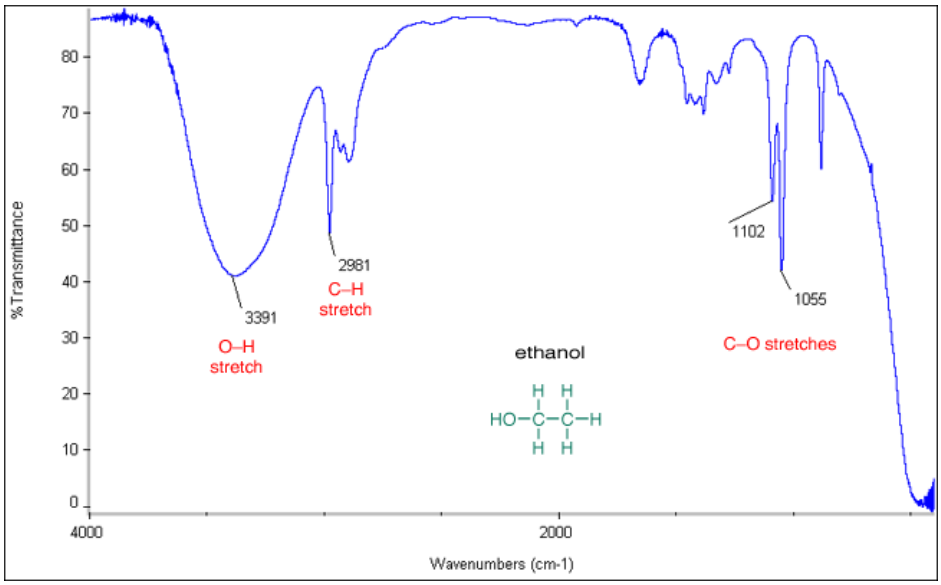
A more restrictive definition distinguishes between a "true" molecule that exists as an independent particle, and an *extended solid* that can only be represented by its simplest formula. Methane, CH_4, is an example of the former, while sodium chloride, which does not contain any discrete NaCl units, is the most widely-known extended solid. But because we want to look at chemical bonding in the most general way, we will avoid making this distinction here except in a few special cases. In order to emphasize this "aggregate of atoms" definition, we will often use terms such as "chemical species" and "structures" in place of "molecules" in this lesson.

The definition written above is an *operational* one; that is, it depends on our ability to observe and measure the molecule's properties. Clearly, this means that the molecule must retain its identity for a period of time long enough to carry out the measurements. For most of the molecules that chemistry deals with, this presents no difficulty. But it does happen that some structures that we can write formulas for, such as He_2, have such brief lives that no significant properties have been observed. So to some extent, what we consider to be a molecule depends on the technology we use to observe them, and this will necessarily change with time.

Structure, structure, structure!

And what are those properties that characterize a particular kind of molecule and distinguish it from others? Just as real estate is judged by "location, location, location", the identity of a chemical species is defined by its *structure*. In its most fundamental sense, the structure of a molecule is specified by the identity of its constituent atoms and the sequence in which they are joined together, that is, by the *bonding connectivity*. This, in turn, defines the *bonding geometry*— the spatial relationship between the bonded atoms.

The importance of bonding connectivity is nicely illustrated by the structures of the two compounds *ethanol* and *dimethyl ether*, both of which have the *simplest formula* C_2H_6O. The *structural formulas* reveal the very different connectivities of these two molecules whose physical and chemistry properties are quite different:



Structures without molecules: stability and reactivity

The precise definition of bonding energy is described in another lesson and is not important here. For the moment you only need to know that in any stable structure, the potential energy of its atoms is lower than that of the individual isolated atoms. Thus the formation of methane from its gaseous atoms (a reaction that cannot be observed under ordinary conditions but for which the energetics are known from indirect evidence)

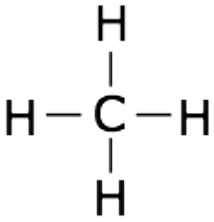
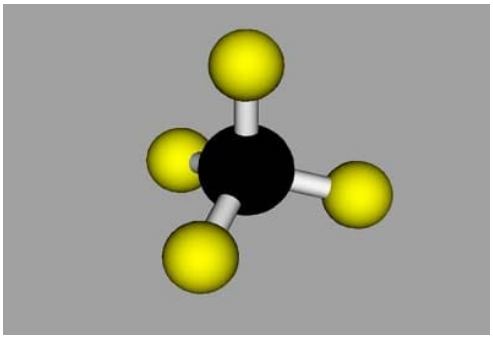
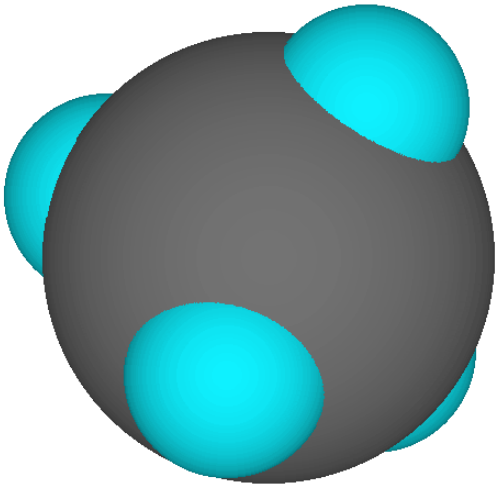
$$4\,\mathrm{H}\,(\mathrm{g}) + \mathrm{C}\,(\mathrm{g}) \longrightarrow \mathrm{CH_4} \tag{9.2.1}$$

is accompanied by the release of heat, and is thus an *exothermic* process. The quantity of heat released is related to the stability of the molecule. The smaller the amount of energy released, the more easily can the molecule absorb thermal energy from the environment, driving the above reaction in reverse and leading to the molecule's decomposition. A highly stable molecule such as methane must be subjected to temperatures of more than 1000°C for significant decomposition to occur. But the noble-gas molecule KrF_2 is so weakly bound that it decomposes even at 0°C, and the structure He_2 has never been observed. If a particular arrangement of atoms is too unstable to reveal its properties at any achievable temperature, then it does not qualify to be called a molecule.

There are many molecules that are energetically stable enough to meet the above criterion, but are so ***reactive*** that their lifetimes are too brief to make their observation possible. The molecule CH_3, *methyl*, is a good example: it can be formed by electrical discharge in gaseous CH_4, but it is so reactive that it combines with almost any molecule it strikes (even another CH_3)within a few collisions. It was not until the development of spectroscopic methods (in which a molecule is characterized by the wavelengths of light that it absorbs or emits) that methyl was recognized as a stable albeit shamelessly promiscuous molecule that is an important intermediate in many chemical processes ranging from flames to atmospheric chemistry.

How we Depict Chemical Structures

Chemical species are traditionally represented by *structural formulas* such as the one for ascorbic acid (vitamin C) which we show here. The lines, of course, represent the "chemical bonds" of the molecule. More importantly, the structural formula of a molecule defines its *connectivity*, as was illustrated in the comparison of ethanol and dimethyl ether shown above.

 **methane**		 Note how this shows CH_4 to be roughly spherical.
Ordinary structural formula, showing connectivity only.	Ball-and-stick model, showing the "chemical bonds" and bonding geometry, but with the individual atoms unrealistically separated.	Space-filling model, showing relative sizes of the atoms and general shape of the molecule, but not all atoms visible. No obvious "chemical bonds" here!

One limitation of such formulas is that they are drawn on a two-dimensional paper or screen, whereas most molecules have a three-dimensional shape. The wedge-shaped lines in the structural formula are one way of indicating which bonds extend above or below the viewing plane. You will probably be spared having to learn this convention until you get into second-year Chemistry. Three-dimensional models (either real plastic ones or images that incorporate perspective and shading) reveal much more about a molecule's structure. The ball-and-stick and space-filling renditions are widely employed, but each has its limitations, as seen in the following examples:

But what would a molecule "really" look like if you could view it through a magical microscope of some kind? A possible answer would be this computer-generated view of nicotine. At first you might think it looks more like a piece of abstract sculpture than a molecule, but it does reveal the shape of the negative charge-cloud that envelops the collection of atom cores and nuclei hidden within. This can be very important for understanding how the molecule interacts with the similar charge-clouds that clothe solvent and bioreceptor molecules.

Finally, we get to see one! In 2009, IBM scientists in Switzerland succeeded in imaging a real molecule, using a technique known as atomic force microscopy in which an atoms-thin metallic probe is drawn ever-so-slightly above the surface of an immobilized pentacene molecule cooled to nearly absolute zero. In order to improve the image quality, a molecule of carbon monoxide was placed on the end of the probe. The image produced by the AFM probe is shown at the very bottom. What is actually being imaged is the surface of the electron clouds of the molecule, which consists of five fused hexagonal rings of carbon atoms with hydrogens on its periphery. The tiny bumps that correspond to these hydrogen atom attest to the remarkable resolution of this experiment.

Visualization of molecular structures

The purpose of rendering a molecular structure in a particular way is not to achieve "realism" (whatever that might be), but rather to convey useful information of some kind. Modern computer rendering software takes its basic data from various kinds of standard structural databases which are compiled either from experimental X-ray scattering data, or are calculated from theory.

As was mentioned above, it is often desirable to show the "molecular surface"— the veil of negative charge that originates in the valence electrons of the atoms but which tends to be spread over the entire molecule to a distance that can significantly affect *van der Waals interactions* with neighboring molecules. It is often helpful to superimpose images representing the atoms within the molecule, scaled to their average covalent radii, and to draw the "bonding lines" expressing their connectivity.

Knowing the properties of molecular surfaces is vitally important to understanding any process that depends on one molecule

remaining in physical contact with another. Catalysis is one example, but one of the main interests at the present time is *biological signaling*, in which a relatively small molecule binds to or "docks" with a receptor site on a much larger one, often a protein. Sophisticated molecular modeling software such as was used to produce these images is now a major tool in many areas of research biology.

Visualizing **very large molecules** such as carbohydrates and proteins that may contain tens of thousands of atoms presents obvious problems. The usual technique is to simplify the major parts of the molecule, representing major kinds of extended structural units by shapes such as ribbons or tubes which are twisted or bent to approximate their conformations. These are then gathered to reveal the geometrical relations of the various units within the overall structure. Individual atoms, if shown at all, are restricted to those of special interest.

Study of the surface properties of large molecules is crucial for understanding how proteins, carbohydrates, and DNA interact with smaller molecules, especially those involved in transport of ions and small molecule across cell membranes, immune-system behavior, and signal transduction processes such as the "turning on" of genes.

Observable Properties of Bonded Atom Pairs

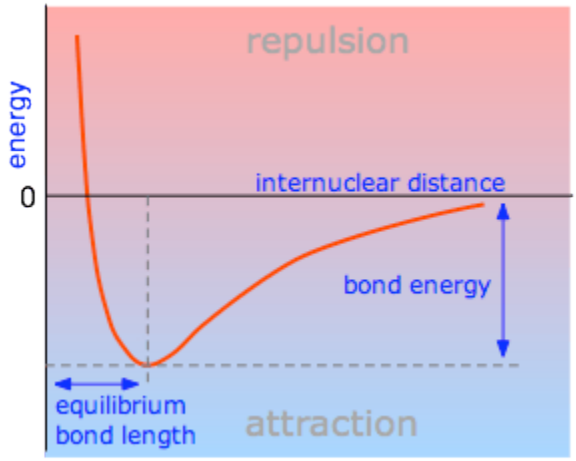
When we talk about the properties of a particular chemical bond, we are really discussing the relationship between two adjacent atoms that are part of the molecule.*Diatomic* molecules are of course the easiest to study, and the information we derive from them helps us interpret various kinds of experiments we carry out on more complicated molecules.

It is important to bear in mind that the exact properties of a specific kind of bond will be determined in part by the nature of the other bonds in the molecule; thus the energy and length of the C–H bond will be somewhat dependent on what other atoms are connected to the carbon atom. Similarly, the C-H bond length can vary by as much a 4 percent between different molecules. For this reason, the values listed in tables of bond energy and bond length are usually averages taken over a variety of compounds that contain a specific atom pair..

In some cases, such as C—O and C—C, the variations can be much greater, approaching 20 percent. In these cases, the values fall into groups which we interpret as representative of single- and *multiple* bonds: double, and triple.

Potential energy curves

The energy of a system of two atoms depends on the distance between them. At large distances the energy is zero, meaning “no interaction”. At distances of several atomic diameters attractive forces dominate, whereas at very close approaches the force is repulsive, causing the energy to rise. The attractive and repulsive effects are balanced at the minimum point in the curve. Plots that illustrate this relationship are known as *Morse curves*, and they are quite useful in defining certain properties of a chemical bond.



The internuclear distance at which the potential energy minimum occurs defines the **bond length**. This is more correctly known as the *equilibrium* bond length, because thermal motion causes the two atoms to vibrate about this distance. In general, the stronger the bond, the smaller will be the bond length.

Attractive forces operate between all atoms, but unless the potential energy minimum is at least of the order of RT, the two atoms will not be able to withstand the disruptive influence of thermal energy long enough to result in an identifiable molecule. Thus we can say that a chemical bond exists between the two atoms in H_2. The weak attraction between argon atoms does not allow Ar_2 to exist as a molecule, but it does give rise to the van Der Waals force that holds argon atoms together in its liquid and solid forms.

Potential energy and kinetic energy Quantum theory tells us that an electron in an atom possesses kinetic energy K as well as potential energy P, so the total energy E is always the sum of the two: E = P + K. The relation between them is surprisingly simple: K = −0.5 P. This means that when a chemical bond forms (an exothermic process with ΔE < 0), the decrease in potential energy is accompanied by an increase in the kinetic energy (embodied in the momentum of the bonding electrons), but the magnitude of the latter change is only half as much, so the change in potential energy always dominates. The bond energy −ΔE has half the magnitude of the fall in potential energy.

Bond energies

The *bond energy* is the amount of work that must be done to pull two atoms completely apart; in other words, it is the same as the depth of the "well" in the potential energy curve shown above. This is almost, but not quite the same as the *bond dissociation energy* actually required to break the chemical bond; the difference is the very small *zero-point energy*. This relationship will be clarified below in the section on bond vibrational frequencies. Bond energies are usually determined indirectly from thermodynamic data, but there are two main experimental ways of measuring them directly:

1. The **direct thermochemical method** involves separating the two atoms by an electrical discharge or some other means, and then measuring the heat given off when they recombine. Thus the energy of the C—C single bond can be estimated from the heat of the recombination reaction between methyl radicals, yielding ethane:

$$CH_3 + CH_3 \rightarrow H_3C\text{–}CH_3 \tag{9.2.2}$$

Although this method is simple in principle, it is not easy to carry out experimentally. The highly reactive components must be prepared in high purity and in a stream of moving gas.

2. The **spectroscopic method** is based on the principle that absorption of light whose wavelength corresponds to the bond energy will often lead to the breaking of the bond and dissociation of the molecule. For some bonds, this light falls into the green and blue regions of the spectrum, but for most bonds ultraviolet light is required. The experiment is carried out by observing the absorption of light by the substance being studied as the wavelength is decreased. When the wavelength is sufficiently small to break the bond, a characteristic change in the absorption pattern is observed.

Spectroscopy is quite easily carried out and can yield highly precise results, but this method is only applicable to a relatively small number of simple molecules. The major problem is that the light must first be *absorbed* by the molecule, and relatively few molecules happen to absorb light of a wavelength that corresponds to a bond energy.

Experiments carried out on diatomic molecules such as O_2 and CS yield unambiguous values of bond energy, but for more complex molecules there are complications. For example, the heat given off in the CH_3 combination reaction written above will also include a small component that represents the differences in the energies of the C-H bonds in methyl and in ethane. These can be corrected for by experimental data on reactions such as

$$CH_3 + H \rightarrow CH_4 \tag{9.2.3}$$

$$CH_2 + H \rightarrow CH_3 \tag{9.2.4}$$

By assembling a large amount of experimental information of this kind, a consistent set of average bond energies can be obtained (see table below.) The energies of double bonds are greater than those of single bonds, and those of triple bonds are higher still.

Use of bond energies in estimating heats of reaction

One can often get a very good idea of how much heat will be absorbed or given off in a reaction by simply finding the difference in the total bond energies contained in the reactants and products. The strength of an individual bond such as O–H depends to some extent on its environment in a molecule (that is, in this example, on what other atom is connected to the oxygen atom), but tables of "average" energies of the various common bond types are widely available and can provide useful estimates of the quantity of heat absorbed or released in many chemical reaction.

Table 1: Average Bond Energies (kJ/mol)

Single Bonds						Multiple Bonds	
H—H	432	N—H	391	I—I	149	C = C	614
H—F	565	N—N	160	I—Cl	208	C ≡ C	839
H—Cl	427	N—F	272	I—Br	175	O = O	495
H—Br	363	N—Cl	200		C = O*	745	
H—I	295	N—Br	243	S—H	347	C ≡ O	1072
	N—O	201	S—F	327	N = O	607	
C—H	413	O—H	467	S—Cl	253	N = N	418
C—C	347	O—O	146	S—Br	218	N ≡ N	941
C—N	305	O—F	190	S—S	266	C ≡ N	891
C—O	358	O—Cl	203		C = N	615	
C—F	485	O—I	234	Si—Si	340		
C—Cl	339		Si—H	393			
C—Br	276	F—F	154	Si—C	360		
C—I	240	F—Cl	253	Si—O	452		
C—S	259	F—Br	237				
	Cl—Cl	239					
	Cl—Br	218					
	Br—Br	193					

*C == O(CO_2) = 799

Average bond energies are the averages of bond dissociation energies (see Table T3 for more complete list). For example the average bond energy of O-H in H_2O is 464 kJ/mol. This is due to the fact that the H-OH bond requires 498.7 kJ/mol to dissociate, while the O-H bond needs 428 kJ/mol.

$$\frac{498.7\ kJ/mol + 428\ kJ/mol}{2} = 464\ kJ/mol \tag{9.2.5}$$

Example 1: Dichloromethane

Consider the reaction of chlorine with methane to produce dichloromethane and hydrogen chloride:

$$\mathrm{CH_4\,(g) + 2\,Cl_2\,(g) \longrightarrow CH_2Cl_2\,(g) + 2\,HCl\,(g)} \tag{9.2.6}$$

In this reaction, two C–Cl bonds and two H–Cl bonds are broken, and two new C–Cl and H–Cl bonds are formed. The net change associated with the reaction is

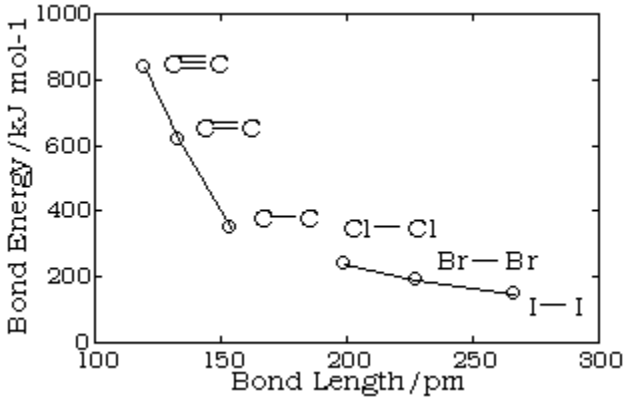
2(C–H) + 2(Cl–Cl) – 2(C–Cl) – 2(H–Cl) = (830 + 486 –660 – 864) kJ

which comes to –208 kJ per mole of methane; this agrees quite well with the observed heat of reaction, which is –202 kJ/mol.

Bond lengths and angles

bond	Ave. Length	Ave. Energy/kJ mol^{-1}
H—H	74 pm	432
H—C	109 pm	415
H—N	101	390
H—O	96	460
H—Cl	127	428
H—Br	141	362
C—C	154	345
C=C	133	615
C≡C	120	835
N≡N	110	942
Cl—Cl	199	240
Br—Br	228	190
I—I	267	149

The length of a chemical bond the distance between the centers of the two bonded atoms (the *internuclear distance*.) Bond lengths have traditionally been expressed in Ångstrom units, but picometers are now preferred (1Å = 10^{-8} cm = 100 pm.) Bond lengths are typically in the range 1-2 Å or 100-200 pm. Even though the bond is vibrating, equilibrium bond lengths can be determined experimentally to within ±1 pm.



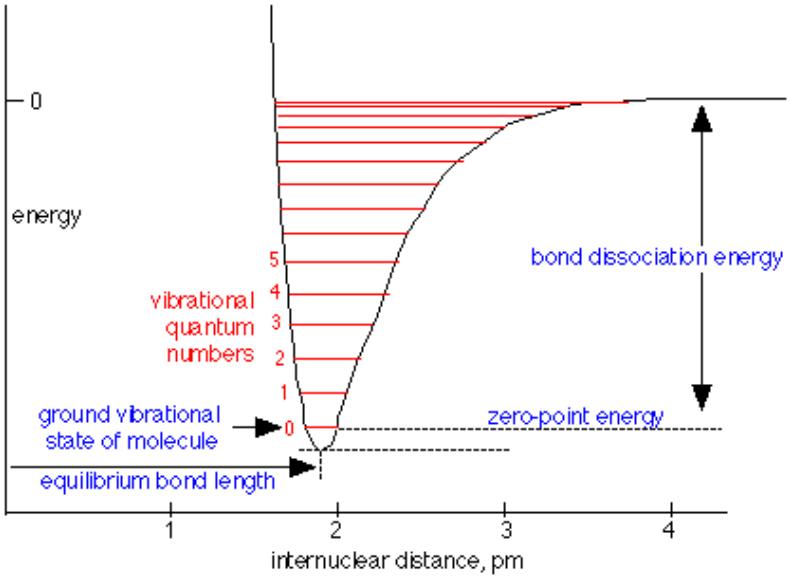
Bond lengths depend mainly on the sizes of the atoms, and secondarily on the bond strengths, the stronger bonds tending to be shorter. Bonds involving hydrogen can be quite short; The shortest bond of all, H–H, is only 74 pm. Multiply-bonded atoms are closer together than singly-bonded ones; this is a major criterion for experimentally determining the *multiplicity* of a bond. This trend is clearly evident in the above plot which depicts the sequence of carbon-carbon single, double, and triple bonds.

The most common method of measuring bond lengths in solids is by analysis of the diffraction or scattering of X-rays when they pass through the regularly-spaced atoms in the crystal. For gaseous molecules, neutron- or electron-diffraction can also be used.

The complete structure of a molecule requires a specification of the coordinates of each of its atoms in three-dimensional space. This data can then be used by computer programs to construct *visualizations* of the molecule as discussed above. One such visualization of the water molecule, with bond distances and the HOH bond angle superimposed on a space-filling model, is shown here. (It is taken from an excellent reference source on water). The colors show the results of calculations that depict the way in which electron charge is distributed around the three nuclei.

Bond stretching and infrared absorption

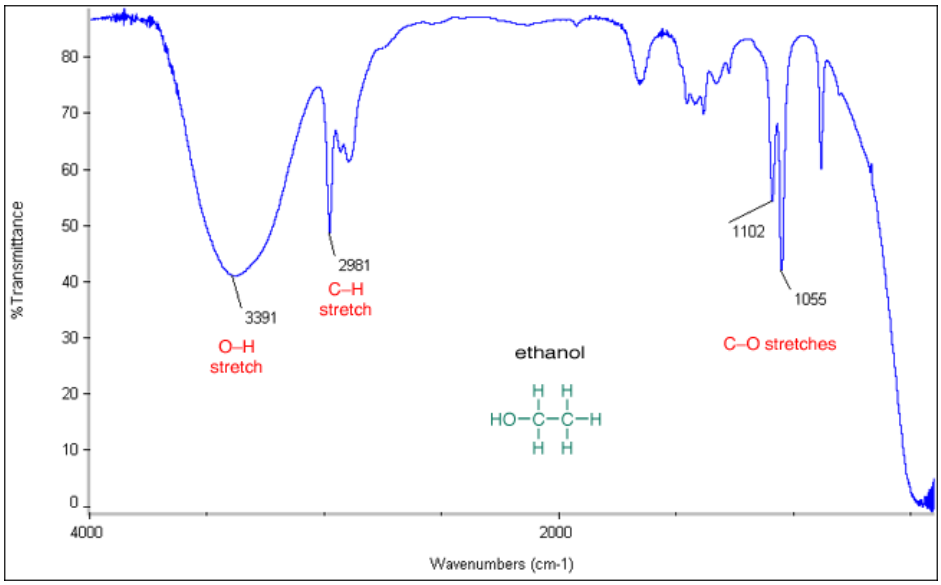
When an atom is displaced from its equilibrium position in a molecule, it is subject to a restoring force which increases with the displacement. A spring follows the same law (Hooke's law); a chemical bond is therefore formally similar to a spring that has weights (atoms) attached to its two ends. A mechanical system of this kind possesses a natural vibrational frequency which depends on the masses of the weights and the stiffness of the spring. These vibrations are initiated by the thermal energy of the surroundings; chemically-bonded atoms are never at rest at temperatures above absolute zero.



On the atomic scale in which all motions are *quantized*, a vibrating system can possess a series of vibrational frequencies, or *states*. These are depicted by the horizontal lines in the potential energy curve shown here. Notice that the very bottom of the curve does not correspond to an allowed state because at this point the positions of the atoms are precisely specified, which would violate the uncertainty principle. The lowest-allowed, or *ground* vibrational state is the one denoted by 0, and it is normally the only state that is significantly populated in most molecules at room temperature. In order to jump to a higher state, the molecule must absorb a photon whose energy is equal to the distance between the two states.

For ordinary chemical bonds, the energy differences between these natural vibrational frequencies correspond to those of *infrared light*. Each wavelength of infrared light that excites the vibrational motion of a particular bond will be absorbed by the molecule. In general, the stronger the bond and the lighter the atoms it connects, the higher will be its natural stretching frequency and the shorter the wavelength of light absorbed by it. Studies on a wide variety of molecules have made it possible to determine the wavelengths absorbed by each kind of bond. By plotting the degree of absorption as a function of wavelength, one obtains the *infrared spectrum* of the molecule which allows one to "see" what kinds of bonds are present.

Infrared spectrum of alcohol The low points in the plot below indicate the frequencies of infrared light that are absorbed by ethanol (ethyl alcohol), CH_3CH_2OH. Notice how stretching frequencies involving hydrogen are higher, reflecting the smaller mass of that atom. Only the most prominent absorption bands are noted here.

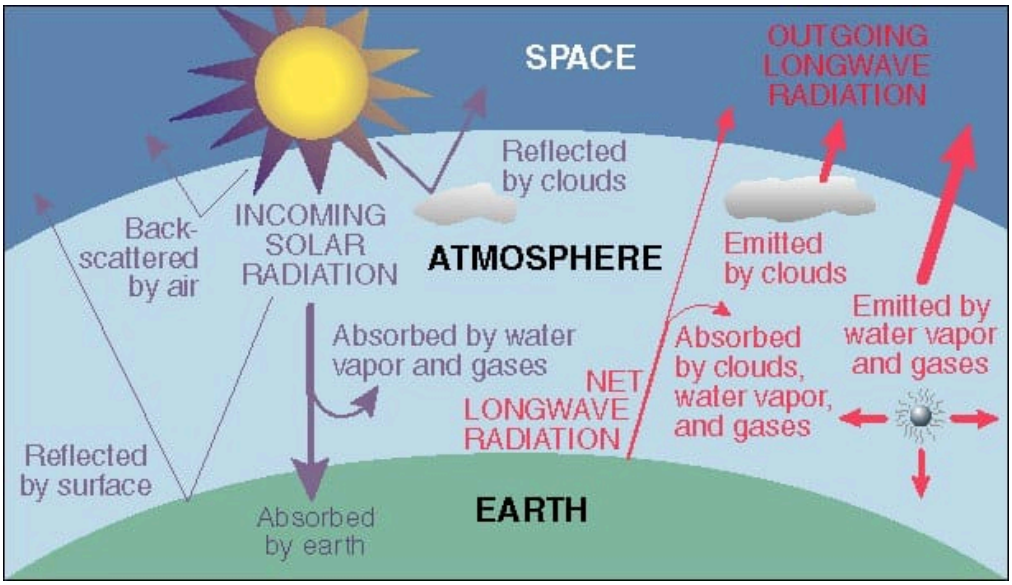


Actual infrared spectra are complicated by the presence of more complex motions (stretches involving more than two atoms, wagging, etc.), and absorption to higher quantum states (overtones), so infrared spectra can become quite complex. This is not necessarily a disadvantage, however, because such spectra can serve as a "fingerprint" that is unique to a particular molecule and can be helpful in identifying it. Largely for this reason, infrared spectrometers are standard equipment in most chemistry laboratories. Now that you know something about bond stretching vibrations, you can impress your friends by telling them **why water is blue!**

Infrared Absorption and Global Warming

The aspect of bond stretching and bending frequencies that impacts our lives most directly is the way that some of the gases of the atmosphere absorb infrared light and thus affect the heat balance of the Earth. Owing to their symmetrical shapes, the principal atmospheric components N_2 and O_2 do not absorb infrared light, but the minor components water vapor and carbon dioxide are strong absorbers, especially in the long-wavelength region of the infrared. Absorption of infrared light by a gas causes its temperature to rise, so any source of infrared light will tend to warm the atmosphere; this phenomenon is known as the *greenhouse effect*.

The incoming radiation from the Sun (which contains relatively little long-wave infrared light) passes freely through the atmosphere and is absorbed by the Earth's surface, warming it up and causing it to re-emit some of this energy as long-wavelength infrared. Most of the latter is absorbed by the H_2O and CO_2 , the major greenhouse gas is in the unpolluted atmosphere, effectively trapping the radiation as heat. Thus the atmosphere is heated by the Earth, rather than by direct sunlight. Without the "*greenhouse gases*" in the atmosphere, the Earth's heat would be radiated away into space, and our planet would be too cold for life.



Radiation balance of the Earth In order to maintain a constant average temperature, the quantity of radiation (sunlight) absorbed by the surface must be exactly balanced by the quantity of long-wavelength infrared emitted by the surface and atmosphere and radiated back into space. Atmospheric gases that absorb this infrared light (depicted in red on the right part of this diagram) partially block this emission and become warmer, raising the Earth's temperature. This diagram is from the U. of Oregon Web page referenced below.

Since the beginning of the Industrial Revolution in the 19th century, huge quantities of additional greenhouse gases have been accumulating in the atmosphere. Carbon dioxide from fossil fuel combustion has been the principal source, but intensive agriculture also contributes significant quantities of methane (CH_4) and nitrous oxide (N_2O) which are also efficient far-infrared absorbers. The measurable increase in these gases is believed by many to be responsible for the increase in the average temperature of the Earth that has been noted over the past 50 years— a trend that could initiate widespread flooding and other disasters if it continues.

This page titled 9.2: Molecules - Properties of Bonded Atoms is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.3: Models of Chemical Bonding

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented.

- Comment on the distinction between a *theory* and a *model* in the context of chemical bonding.
- What is meant by a *classical* model of chemical bonding?

Why do atoms bind together— sometimes? The answer to this question would ideally be a simple, easily understood theory that would not only explain why atoms bind together to form molecules, but would also predict the three-dimensional structures of the resulting compounds as well as the energies and other properties of the bonds themselves. Unfortunately, no one theory exists that accomplishes these goals in a satisfactory way for all of the many categories of compounds that are known. Moreover, it seems likely that if such a theory does ever come into being, it will be far from simple.

About Models in Science

When we are faced the need to find a scientific explanation for a complex phenomenon such as bonding, experience has shown that it is often best to begin by developing a *model*. A *scientific model* is something like a theory in that it should be able to explain observations and to make useful predictions. But whereas a theory can be discredited by a single contradictory case, a model can be useful even if it does not encompass all instances of the effects it attempts to explain. We do not even require that a model be a credible representation of reality; all we ask is that it be able to explain the behavior of those cases to which it is applicable in terms that are consistent with the model itself.

An example of a model that you may already know about is the kinetic molecular theory of gases. Despite its name, this is really a model (at least at the level that beginning students use it) because it does not even try to explain the observed behavior of real gases. Nevertheless, it serves as a tool for developing our understanding of gases, and as an essential starting point for more elaborate treatments.

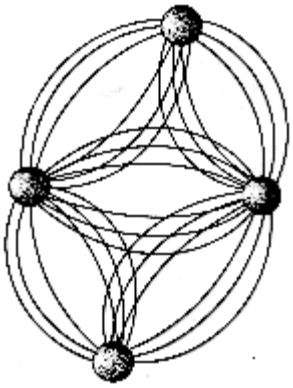
One thing is clear: chemical bonding is basically *electrical* in nature, the result of attraction between bodies of opposite charge; bonding occurs when outer-shell electrons are simultaneously attracted to the positively-charged nuclei of two or more nearby atoms. The need for models arises when we try to understand why

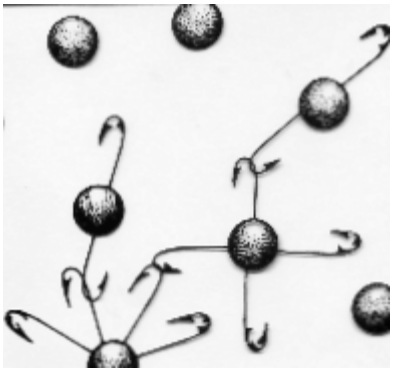
- Not all pairs of atoms can form stable bonds
- Different elements can form different numbers of bonds (this is expressed as "combining power" or "valence".)
- The geometric arrangement of the bonds ("bonding geometry") around a given kind of atom is a property of the element.

Given the extraordinary variety of ways in which atoms combine into aggregates, it should come as no surprise that a number of useful bonding models have been developed. Most of them apply only to certain classes of compounds or attempt to explain only a restricted range of phenomena. In this section we will provide brief descriptions of some of the bonding models; the more important of these will be treated in much more detail in later lessons in this unit.

Some early views of chemical bonding

Intense speculation about “chemical affinity” began in the 18th century. Some likened the tendency of one atom to “close” with another as an expression of a human-like kind of affection. Others attributed bonding to magnetic-like forces (left) or to varying numbers of “hooks” on different kinds of atoms (right). The latter constituted a primitive (and extremely limited) way of explaining the different *combining powers* (*valances*) of the different elements.





"There are no such things..."

Napoleon's definition of history as a set of lies agreed on by historians seems to have a parallel with chemical bonding and chemists. At least in Chemistry, we can call the various explanations "models" and get away with it even if they are demonstrably wrong, as long as we find them useful. In a provocative article (*J Chem Educ* 1990 67(4) 280-298), J. F. Ogilvie tells us that there are no such things as orbitals, or, for that matter, non-bonding electrons, bonds, or even uniquely identifiable atoms within molecules. This idea disturbed a lot of people (teachers and textbook authors preferred to ignore it) and prompted a spirited rejoinder (*J Chem Ed* 1992 69(6) 519-521) from Linus Pauling, father of the modern quantum-mechanical view of the chemical bond.

But the idea has never quite gone away. Richard Bader of McMaster University has developed a quantitative "atoms in molecules" model that depicts molecules as a collection of point-like nuclei embedded in a diffuse cloud of electrons. There are no "bonds" in this model, but only "bond paths" that correspond to higher values of electron density along certain directions that are governed by the manner in which the positive nuclei generate localized distortions of the electron cloud.

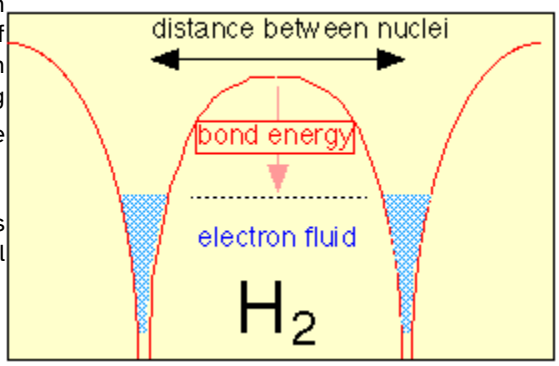
Classical models of the chemical bond

By *classical*, we mean models that do not take into account the quantum behavior of small particles, notably the electron. These models generally assume that electrons and ions behave as point charges which attract and repel according to the laws of electrostatics. Although this completely ignores what has been learned about the nature of the electron since the development of quantum theory in the 1920's, these classical models have not only proven extremely useful, but the major ones also serve as the basis for the chemist's general classification of compounds into "covalent" and "ionic" categories.

The Ionic Model

Ever since the discovery early in the 19th century that solutions of salts and other electrolytes conduct electric current, there has been general agreement that the forces that hold atoms together must be electrical in nature. Electrolytic solutions contain *ions* having opposite electrical charges; opposite charges attract, so perhaps the substances from which these ions come consist of positive and negatively charged atoms held together by electrostatic attraction.

It turns out that this is not true generally, but a model built on this assumption does a fairly good job of explaining a rather small but important class of compounds that are called *ionic solids*. The most well known example of such a compound is sodium chloride, which consists of two interpenetrating lattices of Na^+ and Cl^- ions arranged in such as way that every ion of one type is surrounded (in three dimensional space) by six ions of opposite charge.



One can envision the formation of a solid NaCl unit by a sequence of events in which one mole of gaseous Na atoms lose electrons to one mole of Cl atoms, followed by condensation of the resulting ions into a crystal lattice:

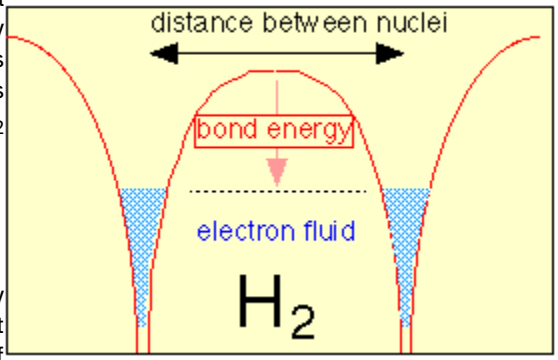
$Na(g) \rightarrow Na^+(g) + e^-$	+494 kJ	ionization energy
$Cl(g) + e^- \rightarrow Cl^-(g)$	−368 kJ	electron affinity
$Na^+(g) + Cl^-(g)$	−498 kJ	lattice energy
$Na(g) + Cl(g) \rightarrow NaCl(s)$	−372 kJ	Sum: Na-Cl bond energy

Note: positive energy values denote endothermic processes, while negative ones are exothermic.

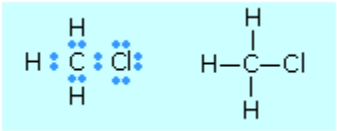
Since the first two energies are known experimentally, as is the energy of the sum of the three processes, the lattice energy can be found by difference. It can also be calculated by averaging the electrostatic forces exerted on each ion over the various directions in the solid, and this calculation is generally in good agreement with observation, thus lending credence to the model. The sum of the three energy terms is clearly negative, and corresponds to the liberation of heat in the net reaction (bottom row of the table), which defines the Na–Cl "bond" energy.

The ionic solid is more stable than the equivalent number of gaseous atoms simply because the three-dimensional NaCl structure allows more electrons to be closer to more nuclei. This is the criterion for the stability of any kind of molecule; all that is special about the "ionic" bond is that we can employ a conceptually simple electrostatic model to predict the bond strength.

The main limitation of this model is that it applies really well only to the small class of solids composed of Group 1 and 2 elements with highly electronegative elements such as the halogens. Although compounds such as $CuCl_2$ dissociate into ions when they dissolve in water, the fundamental units making up the solid are more like polymeric chains of covalently-bound $CuCl_2$ molecules that have little ionic character.



Shared-electron (covalent) model



This model originated with the theory developed by G.N. Lewis in 1916, and it remains the most widely-used model of chemical bonding. It is founded on the idea that a pair of electrons shared between two atoms can create a mutual attraction, and thus a chemical bond.

Usually each atom contributes one electron (one of its *valence electrons*) to the pair, but in some cases both electrons come from one of the atoms. For example, the bond between hydrogen and chlorine in the hydrogen chloride molecule is made up of the single 1s electron of hydrogen paired up with one of chlorine's seven valence (3*p*) electrons. The stability afforded by this sharing is thought to derive from the noble gas configurations (helium for hydrogen, argon for chlorine) that surround the bound atoms.

The origin of the electrostatic binding forces in this model can best be understood by examining the simplest possible molecule. This is the **hydrogen molecule ion** H_2^+, which consists of two nuclei and one electron.

First, however, think what would happen if we tried to make the even simpler molecule H_2^{2+}. Since this would consist only of two protons whose electrostatic charges would repel each other at all distances, it is clear that such a molecule cannot exist; something more than two nuclei are required for bonding to occur.

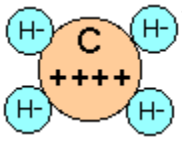
H_2^+ we have a third particle, the electron. The effect of this electron will depend on its location with respect to the two nuclei. If the electron is in the space between the two nuclei (the *binding region* in the diagram), it will attract both protons toward itself, and thus toward each other. If the total attraction energy exceeds the internuclear repulsion, there will be a net bonding effect and the molecule will be stable. If, on the other hand, the electron is off to one side (in an *antibinding region*), it will attract both nuclei, but it will attract the closer one much more strongly, owing to the inverse-square nature of Coulomb's law. As a consequence, the electron will now actively work ag against bonding by helping to push the two nuclei apart.

Polar covalent model

A purely covalent bond can only be guaranteed when the *electronegativities* (electron-attracting powers) of the two atoms are

identical. When atoms having different electronegativities are joined, the electrons shared between them will be displaced toward the more electronegative atom, conferring a polarity on the bond which can be described in terms of percent ionic character.

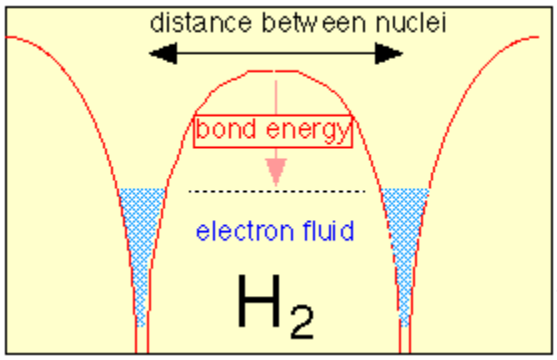
Coulombic model



This is an extension of the ionic model to compounds that are ordinarily considered to be non-ionic. Combined hydrogen is always considered to exist as the hydride ion H^-, so that methane can be treated as if it were C^{4+} H^-_4.

This is not as bizarre as it might seem at first if you recall that the proton has almost no significant size, so that it is essentially embedded in an electron pair when it is joined to another atom in a covalent bond. This model, which is not as well known as it deserves to be, has surprisingly good predictive power, both as to bond energies and structures.

VSEPR model



The "valence shell electron repulsion" model is not so much a model of chemical bonding as a scheme for explaining the shapes of molecules. It is based on the quantum mechanical view that bonds represent electron clouds— physical regions of negative electric charge that repel each other and thus try to stay as far apart as possible. We will explore this concept in much greater detail in a later unit.

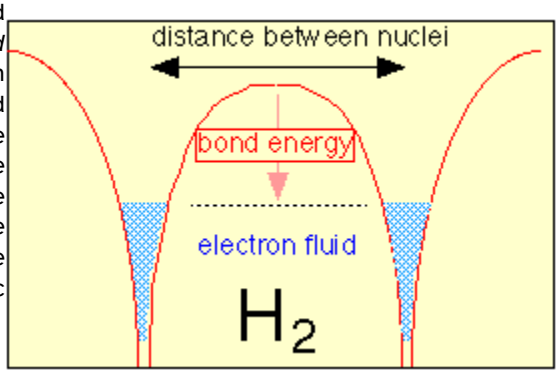
3 Quantum-mechanical models

These models of bonding take into account the fact that a particle as light as the electron cannot really be said to be in any single location. The best we can do is define a region of space in which the probability of finding the electron has some arbitrary value which will always be less than unity. The shape of this volume of space is called an *orbital* and is defined by a mathematical function that relates the probability to the *(x,y,z)* coordinates of the molecule.

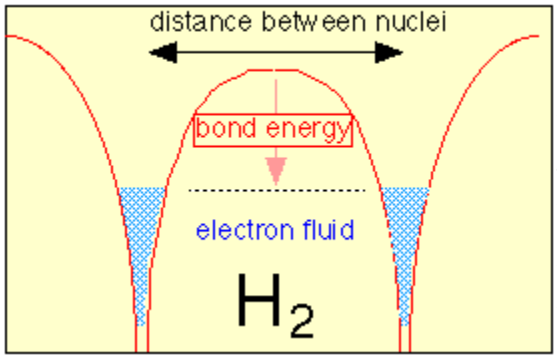
Like other models of bonding, the quantum models attempt to show how more electrons can be simultaneously close to more nuclei. Instead of doing so through purely geometrical arguments, they attempt this by predicting the nature of the orbitals which the valence electrons occupy in joined atoms.

The hybrid orbital model

This was developed by Linus Pauling in 1931 and was the first quantum-based model of bonding. It is based on the premise that if the atomic *s*,*p*, and *d* orbitals occupied by the valence electrons of adjacent atoms are combined in a suitable way, the hybrid orbitals that result will have the character and directional properties that are consistent with the bonding pattern in the molecule. The rules for bringing about these combinations turn out to be remarkably simple, so once they were worked out it became possible to use this model to predict the bonding behavior in a wide variety of molecules. The hybrid orbital model is most usefully applied to the *p*-block elements in the first few rows of the periodic table, and is especially important in organic chemistry.



The molecular orbital model

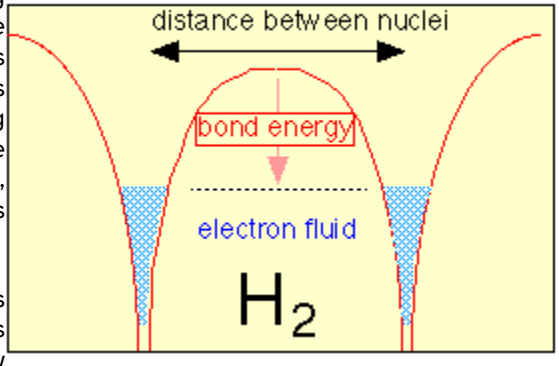


This model takes a more fundamental approach by regarding a molecule as a collection of valence electrons and positive cores. Just as the nature of atomic orbitals derives from the spherical symmetry of the atom, so will the properties of these new molecular orbitals be controlled by the interaction of the valence electrons with the multiple positive centers of these atomic cores.

These new orbitals, unlike those of the hybrid model, are delocalized; that is, they do not "belong" to any one atom but extend over the entire region of space that encompasses the bonded atoms. The available (valence) electrons then fill these orbitals from the lowest to the highest, very much as in the Aufbau principle that you learned for working out atomic electron configurations. For small molecules (which are the only ones we will consider here), there are simple rules that govern the way that atomic orbitals transform themselves into molecular orbitals as the separate atoms are brought together. The real power of molecular orbital theory, however, comes from its mathematical formulation which lends itself to detailed predictions of bond energies and other properties.

The electron-tunneling model

A common theme uniting all of the models we have discussed is that bonding depends on the fall in potential energy that occurs when opposite charges are brought together. In the case of covalent bonds, the shared electron pair acts as a kind of "electron glue" between the joined nuclei. In 1962, however, it was shown that this assumption is not strictly correct, and that instead of being concentrated in the space between the nuclei, the electron orbitals become even more concentrated around the bonded nuclei. At the same time however, they are free to "move" between the two nuclei by a process known as *tunneling*.



This refers to a well-known quantum mechanical effect that allows electrons (or other particles small enough to exhibit wavelike properties) to pass ("tunnel") through a barrier separating two closely adjacent regions of low potential energy. One result of this is that the effective volume of space available to the electron is increased, and according to the *uncertainty principle* this will reduce the kinetic energy of the electron.

According to this model, the bonding electrons act as a kind of fluid that concentrates in the region of each nucleus (lowering the potential energy) and at the same time is able to freely flow between them (reducing the kinetic energy). Despite its conceptual simplicity and full acknowledgment of the laws of quantum mechanics, this model is less known than it deserves to be and is unfortunately absent from most textbooks.

This page titled 9.3: Models of Chemical Bonding is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.4: Polar Covalence

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented below.

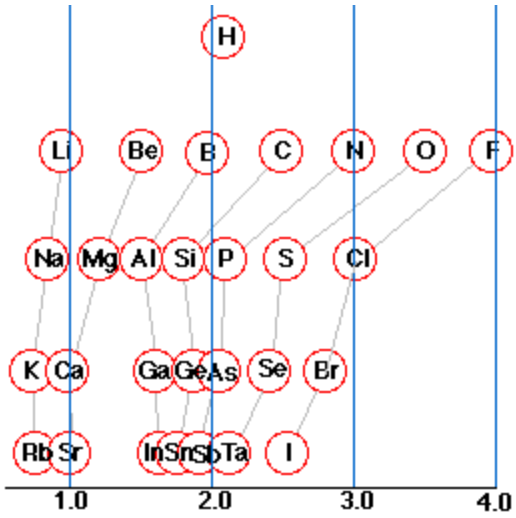
- Define electronegativity, and describe the general way in which electronegativities depend on the location of an element in the periodic table.
- Explain how electronegativity values relate to the polar nature of a chemical bond.
- What two factors determine the magnitude of an *electric dipole moment* ?
- Sketch structural diagrams that illustrate how the presence of polar bonds in a molecule can either increase or diminish the magnitude of the molecular dipole moment.
- Calculate the formal charge of each atom in a structure, and comment on its significance in relation to the polarity of that structure.
- Select the more likely of two or more electron-dot structures for a given species.
- Explain the difference between formal charge and oxidation number.
- Describe the limiting cases of covalent and ionic bonds. Explain what both categories of bonds have in common.

The electrons constituting a chemical bond are simultaneously attracted by the electrostatic fields of the nuclei of the two bonded atoms. In a homonuclear molecule such as O_2 the bonding electrons will be shared equally by the two atoms. In general, however, differences in the sizes and nuclear charges of the atoms will cause one of them to exert a greater attraction on the bonding pair, causing the electron cloud to be displaced toward the more strongly-attracting atom.

Electronegativity

The electronegativity of an atom denotes its relative electron-attracting power in a chemical bond. It is important to understand that electronegativity is not a measurable property of an atom in the sense that ionization energies and electron affinities are, although it can be correlated with both of these properties. The actual electron-attracting power of an atom depends in part on its chemical environment (that is, on what other atoms are bonded to it), so tabulated electronegativities should be regarded as no more than predictors of the behavior of electrons, especially in more complicated molecules. There are several ways of computing electronegativities, which are expressed on an arbitrary scale. The concept of electronegativity was introduced by Linus Pauling and his 0-4 scale continues to be the one most widely used.

The 0-4 electronegativity scale of Pauling is the best known of several arbitrary scales of this kind. Electronegativity values are not directly observable, but are derived from measurable atomic properties properties such as ionization energy and electron affinity. The place of any atom on this scale provides a good indication of its ability to compete with another atom in attracting a shared electron pair to it, but the presence of bonds to other atoms, and of multiple- or nonbonding electron pairs may make predictions about the nature a given bond less reliable.



An atom that has a small electronegativity is said to be *electropositive*. As the diagram shows, the metallic elements are generally electropositive. The position of hydrogen in this regard is worth noting; although physically a nonmetal, much of its chemistry is metal-like.

Dipole moments

When non-identical atoms are joined in a covalent bond, the electron pair will be attracted more strongly to the atom that has the higher electronegativity. As a consequence, the electrons will not be shared equally; the center of the negative charges in the molecule will be displaced from the center of positive charge. Such bonds are said to be *polar* and to possess *partial ionic character*, and they may confer a polar nature on the molecule as a whole.

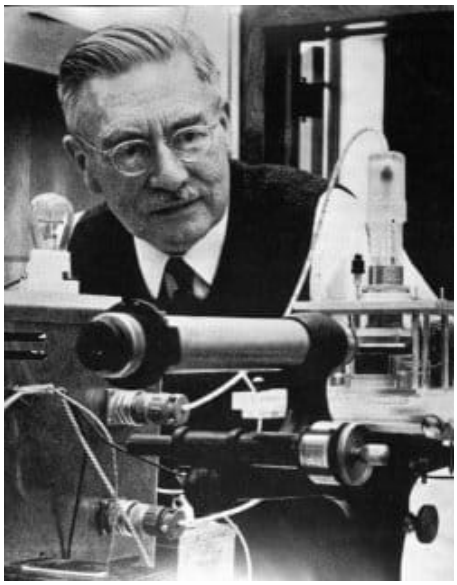
A polar molecule acts as an *electric dipole* which can interact with electric fields that are created artificially or that arise from nearby ions or polar molecules. Dipoles are conventionally represented as arrows pointing in the direction of the negative end. The magnitude of interaction with the electric field is given by the *permanent electric dipole moment* of the molecule. The dipole moment corresponding to an individual bond (or to a diatomic molecule) is given by the product of the quantity of charge displaced *q* and the bond length *r*:

$$\mu = q \times r \tag{9.4.1}$$

In SI units, *q* is expressed in coulombs and *r* in meters, so μ has the dimensions of $C \cdot m$. If one entire electron charge is displaced by 100 pm (a typical bond length), then

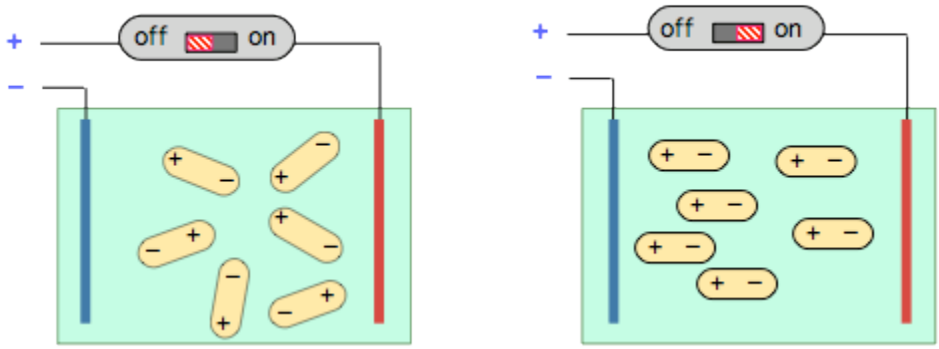
$$\mu = (1.6022 \times 10^{-19}\ C) \times (10^{-10}\ m) = 1.6 \times 10^{-29}\ C \cdot m = 4.8\ D \tag{9.4.2}$$

The quantity denoted by D, the *Debye* unit, is still commonly used to express dipole moments. It was named after *Peter Debye* (1884-1966), the Dutch-American physicist who pioneered the study of dipole moments and of electrical interactions between particles; he won the Nobel Prize for Chemistry in 1936.



How dipole moments are measured

When a solution of polar molecules is placed between two oppositely-charged plates, they will tend to align themselves along the direction of the field. This process consumes energy which is returned to the electrical circuit when the field is switched off, an effect known as electrical *capacitance*.



Measurement of the capacitance of a gas or solution is easy to carry out and serves as a means of determining the magnitude of the dipole moment of a substance.

Example 9.4.1

Estimate the percent ionic character of the bond in the hydrogen fluoride molecule from the experimental data shown at the right.

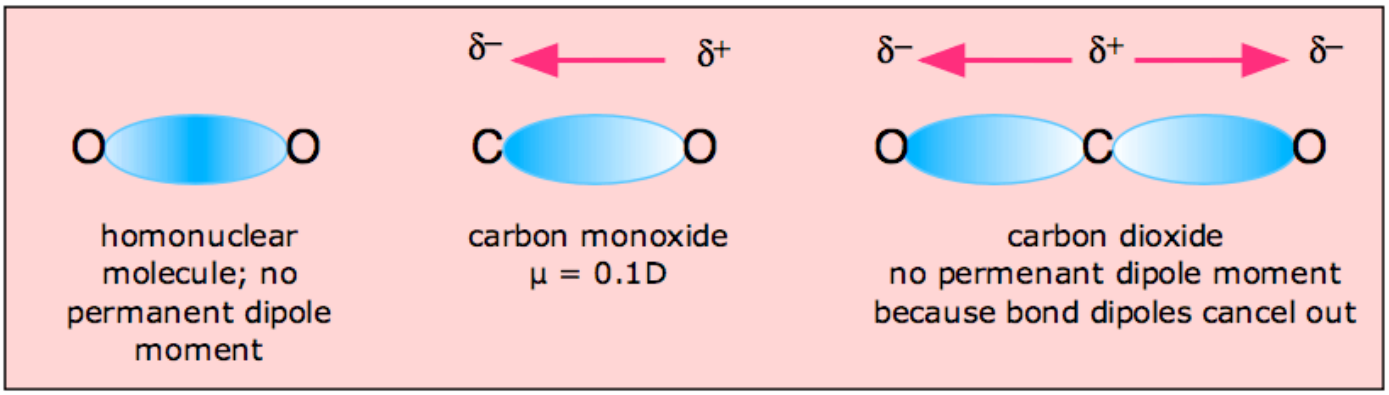
Solution

$$\frac{6.37\text{E-}30\ \text{C-m}}{92\text{E-}12\ \text{m}} = 6.92\ \text{E-}20\ \text{C}\quad \text{(charge displaced)}$$

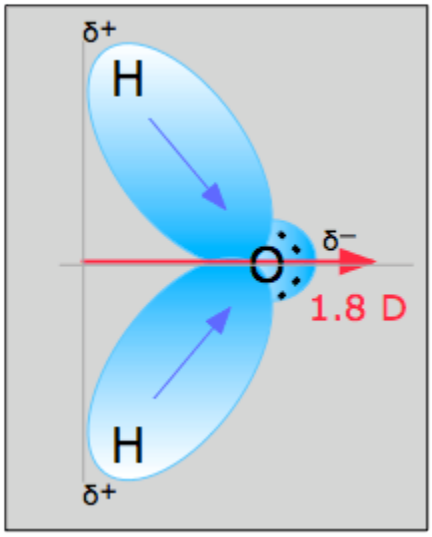
$$\frac{6.92\ \text{E-}20\ \text{C}}{1.60\text{E-}19\ \text{C/electron}} = 0.43\ \text{(ionic character of bond)}$$

Dipole moments as vector sums

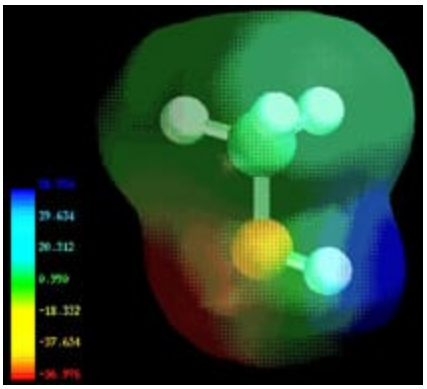
In molecules containing more than one polar bond, the molecular dipole moment is just the vector combination of what can be regarded as individual "bond dipole moments". Being vectors, these can reinforce or cancel each other, depending on the geometry of the molecule; it is therefore not uncommon for molecules containing polar bonds to be nonpolar overall, as in the example of carbon dioxide:



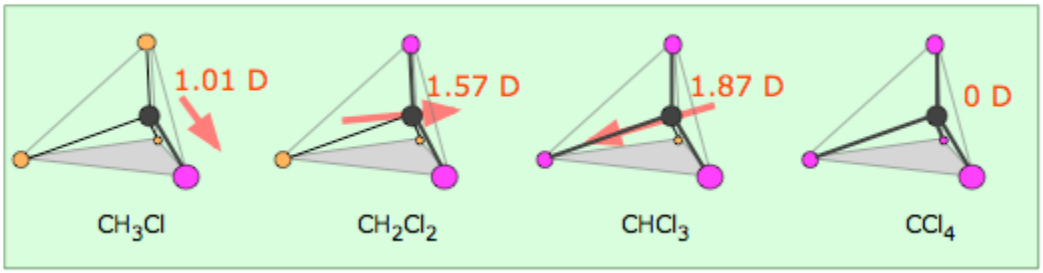
The zero dipole moment of CO_2 is one of the simplest experimental methods of demonstrating the linear shape of this molecule.



H_2O, by contrast, has a very large dipole moment which results from the two polar H–O components oriented at an angle of 104.5°. The nonbonding pairs on oxygen are a contributing factor to the high polarity of the water molecule. In molecules containing nonbonding electrons or multiple bonds, the elecronegativity difference may not correctly predict the bond polarity. A good example of this is carbon monoxide, in which the partial negative charge resides on the carbon, as predicted by its negative formal charge (below.)



Electron densities in a molecule (and the dipole moments that unbalanced electron distributions can produce) are now easily calculated by molecular modeling programs. In this example, for methanol CH_3OH, the blue area centered on hydrogen represents a positive charge, the red area centered where we expect the lone pairs to be located represents a negative charge, while the light green around methyl is approximately neutral. The manner in which the individual bonds contribute to the dipole moment of the molecule is nicely illustrated by the series of chloromethanes:



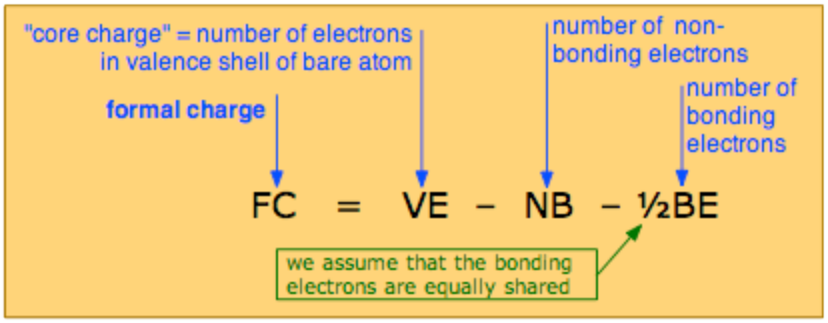
(Bear in mind that all four positions around the carbon atom are equivalent in this tetrahedral molecule, so there are only four chloromethanes.)

Formal charge and oxidation number

Although the total number of valence electrons in a molecule is easily calculated, there is not always a simple and unambiguous way of determining how many reside in a particular bond or as non-bonding pairs on a particular atom. For example, one can write valid Lewis octet structures for carbon monoxide showing either a double or triple bond between the two atoms, depending on how many nonbonding pairs are placed on each: **C::O:::** and **:C:::O:** (see Problem Example 3 below). The choice between structures such as these is usually easy to make on the principle that the more electronegative atom tends to surround itself with the greater number of electrons. In cases where the distinction between competing structures is not all that clear, an arbitrarily-calculated quantity known as the *formal charge* can often serve as a guide.

> The formal charge on an atom is the electric charge it would have if all bonding electrons were shared equally with its bonded neighbors.

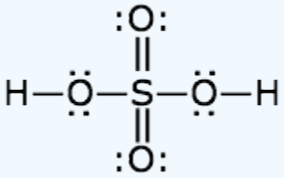
The formal charge on an atom is calculated by the following formula:



in which the core charge is the electric charge the atom would have if all its valence electrons were removed. In simple cases, the formal charge can be worked out visually directly from the Lewis structure, as is illustrated farther on.

Example 9.4.2: Formal Charge

Find the formal charges of all the atoms in the sulfuric acid structure shown here.



Solution

The atoms here are hydrogen, sulfur, and double- and single-bonded oxygens. Remember that a double bond is made up of two electron-pairs.

- hydrogen: FC = 1 – 0 – 1 = 0
- sulfur: FC = 6 – 0 – 6 = 0
- hydroxyl oxygen: FC = 6 – 4 – 2 = 0
- double-bonded oxygen: FC = 6 – 4 – 2 = 0

Using formal charge to select the best Lewis structure

The general rule for choosing between alternative structures is that the one involving the *smallest* formal charges is most favored, although the following example shows that this is not always the case.

Example 9.4.1: Carbon Monoxide

Write out some structures for carbon monoxide CO, both those that do and do not obey the octet rule, and select the "best" on the basis of the formal charges.

Solution

Structure that obeys the octet rule:

a) For **:C:::O:** Carbon: 4 – 2 – 3 = **–1**; Oxygen: 6 – 2 – 3 = **+1**

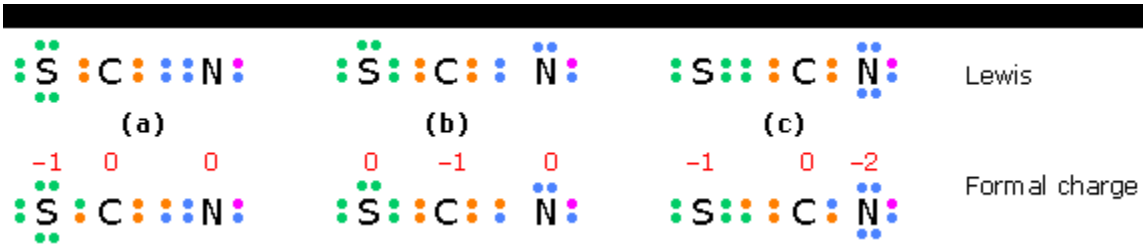
Structures that do not obey the octet rule (for carbon):

b) For :C:O::: Carbon: 4 – 2 – 1 = **+1**; Oxygen: 6 – 6 – 1 = **–1**

c) For :C::O:: Carbon: 4 – 2 –2 = **0**; Oxygen: 6 – 4 – 2 = **0**

Comment: All three structures are acceptable (because the formal charges add up to zero for this neutral molecule) and contribute to the overall structure of carbon monoxide, although not equally. Both experiment and more advanced models show that the triple-bonded form *(a)* predominates. Formal charge, which is no more than a bookkeeping scheme for electrons, is by itself unable to predict this fact.

In a species such as the thiocyanate ion SCN^- in which two structures having the same minimal formal charges can be written, we would expect the one in which the negative charge is on the more electronegative atom to predominate.



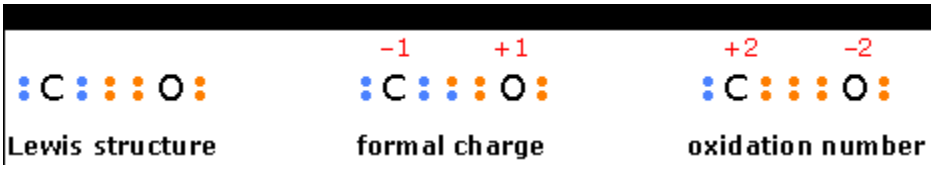
The electrons in the structures of the top row are the valence electrons for each atom; an additional electron (purple) completes the nitrogen octet in this negative ion. The electrons in the bottom row are divided equally between the bonded atoms; the difference between these numbers and those above gives the formal charges.

Formal charge can also help answer the question “where is the charge located?” that is frequently asked about polyatomic ions. Thus by writing out the Lewis structure for the ammonium ion NH_4^+, you should be able to convince yourself that the nitrogen atom has a formal charge of +1 and each of the hydrogens has **0**, so we can say that the positive charge is localized on the central atom.

Oxidation number

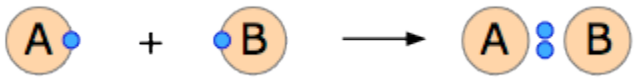
This is another arbitrary way of characterizing atoms in molecules. In contrast to formal charge, in which the electrons in a bond are assumed to be shared equally, *oxidation number* is the electric charge an atom would have if the bonding electrons were assigned exclusively to the more electronegative atom. Oxidation number serves mainly as a tool for keeping track of electrons in reactions in which they are exchanged between reactants, and for characterizing the “combining power” of an atom in a molecule or ion.

The following diagram compares the way electrons are assigned to atoms in calculating formal charge and oxidation number in carbon monoxide.

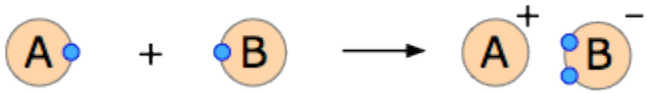


Ionic compounds

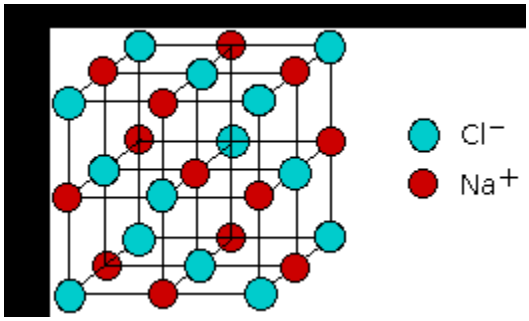
The *shared-electron pair model* introduced by G.N. Lewis showed how chemical bonds could form in the absence of electrostatic attraction between oppositely-charged ions. As such, it has become the most popular and generally useful model of bonding in all substances other than metals. A chemical bond occurs when electrons are simultaneously attracted to two nuclei, thus acting to bind them together in an energetically-stable arrangement. The *covalent bond* is formed when two atoms are able to share a pair of electrons:



In general, however, different kinds of atoms exert different degrees of attraction on their electrons, so in most cases the sharing will not be equal. One can even imagine an extreme case in which the sharing is so unequal that the resulting "molecule" is simply a pair of ions:



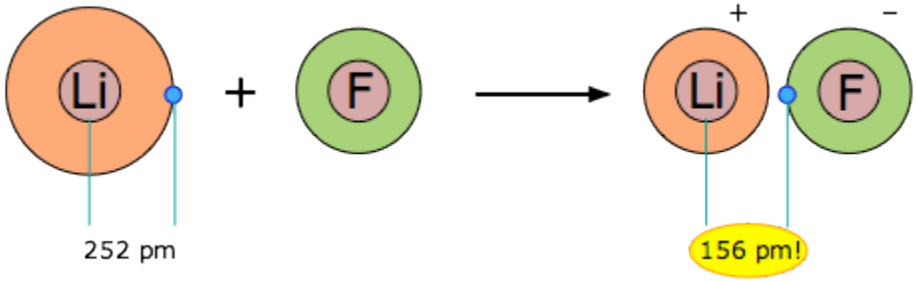
The resulting substance is sometimes said to contain an *ionic bond*. Indeed, the properties of a number of compounds can be adequately explained using the ionic model. But does this mean that there are really two kinds of chemical bonds, ionic and covalent? According to the ionic electrostatic model, solids such as NaCl consist of positive and negative ions arranged in a crystal lattice. Each ion is attracted to neighboring ions of opposite charge, and is repelled by ions of like charge; this combination of attractions and repulsions, acting in all directions, causes the ion to be tightly fixed in its own location in the crystal lattice.



Since electrostatic forces are nondirectional, the structure of an ionic solid is determined purely by geometry: two kinds of ions, each with its own radius, will fall into whatever repeating pattern will achieve the lowest possible potential energy. Surprisingly, there are only a small number of possible structures; one of the most common of these, the *simple cubic* lattice of NaCl, is shown here.

Is there such as thing as an ionic bond?

When two elements form an ionic compound, is an electron really lost by one atom and transferred to the other one? In order to deal with this question, consider the data on the ionic solid LiF. The average radius of the neutral Li atom is about 2.52Å. Now if this Li atom reacts with an atom of F to form LiF, what is the average distance between the Li nucleus and the electron it has "lost" to the fluorine atom? The answer is 1.56Å; the electron is now closer to the lithium nucleus than it was in neutral lithium!



So the answer to the above question is both yes and no: yes, the electron that was now in the 2*s* orbital of Li is now within the grasp of a fluorine 2*p* orbital, but no, the electron is now even closer to the Li nucleus than before, so how can it be "lost"? The one thing that is inarguably true about LiF is that there are more electrons closer to positive nuclei than there are in the separated Li and F atoms. But this is just the rule we stated at the beginning of this unit: **chemical bonds form when electrons can be simultaneously near two or more nuclei.**

It is obvious that the electron-pair bond brings about this situation, and this is the reason for the stability of the covalent bond. What is not so obvious (until you look at the numbers such as are quoted for LiF above) is that the "ionic" bond results in the same condition; even in the most highly ionic compounds, both electrons are close to both nuclei, and the resulting mutual attractions

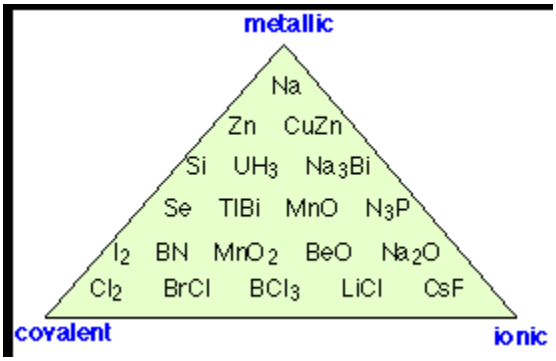
bind the nuclei together. This being the case, is there really any fundamental difference between the ionic and covalent bond?

The answer, according to modern chemical thinking is probably “no”; in fact, there is some question as to whether it is realistic to consider that these solids consist of “ions” in the usual sense. The preferred picture that seems to be emerging is one in which the electron orbitals of adjacent atom pairs are simply skewed so as to place more electron density around the “negative” element than around the “positive” one.

This being said, it must be reiterated that the ionic model of bonding is a useful one for many purposes, and there is nothing wrong with using the term “ionic bond” to describe the interactions between the atoms in the very small class of “ionic solids” such as LiF and NaCl.

"Covalent, ionic or metallic" is an oversimplification!

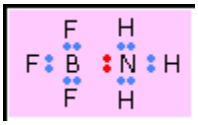
If there is no such thing as a “completely ionic” bond, can we have one that is completely covalent? The answer is yes, if the two nuclei have **equal electron attracting powers**. This situation is guaranteed to be the case with homonuclear diatomic molecules-- molecules consisting of two identical atoms. Thus in Cl_2, O_2, and H_2, electron sharing between the two identical atoms must be exactly even; in such molecules, the center of positive charge corresponds exactly to the center of negative charge: halfway between the two nuclei.



Categorizing all chemical bonds as either ionic, covalent, or metallic is a gross oversimplification; as this diagram shows, there are examples of substances that exhibit varying degrees of all three bonding characteristics.

Dative (Coordinate) covalent Bonds

In most covalent bonds, we think of the electron pair as having a dual parentage, one electron being contributed by each atom. There are, however, many cases in which both electrons come from only one atom. This can happen if the donor atom has a non-bonding pair of electrons and the acceptor atom has a completely empty orbital that can accommodate them. These are called dative or coordinate covalent bonds.



This is the case, for example, with boron trifluoride and ammonia. In BF_3, one the 2p orbitals is unoccupied and can accommodate the lone pair on the nitrogen atom of ammonia. The electron acceptor, BF_3, acts as a *Lewis acid* here, and NH_3 is the *Lewis base*. Bonds of this type (sometimes known as *coordinate covalent* or *dative* bonds) tend to be rather weak (usually 50-200kJ/mol); in many cases the two joined units retain sufficient individuality to justify writing the formula as a molecular *complex* or *adduct*.

This page titled 9.4: Polar Covalence is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.5: Molecular Geometry

Learning Objectives

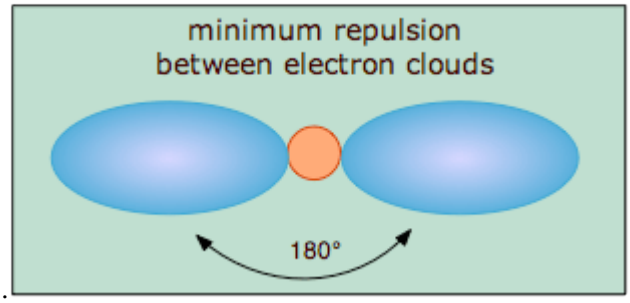
Make sure you thoroughly understand the following essential ideas:

- Describe the manner in which repulsion between electron-pairs affects the orientation of the regions that contain them.
- Define coordination geometry, and describe the particular geometry associated with electron-pair repulsion between two, three, four, five, or six identical bonding regions.
- Explain the distinction between coordination geometry and molecular geometry, and provide an illustration based on the structure of water or ammonia.
- Draw a diagram of a tetrahedral or octahedral molecule.

The Lewis electron-dot structures you have learned to draw have no geometrical significance other than depicting the order in which the various atoms are connected to one another. Nevertheless, a slight extension of the simple shared-electron pair concept is capable of rationalizing and predicting the geometry of the bonds around a given atom in a wide variety of situations.

Electron-pair repulsion

The **valence shell electron pair repulsion (*VSEPR*) model** that we describe here focuses on the bonding and nonbonding electron pairs present in the outermost ("valence") shell of an atom that connects with two or more other atoms. Like all electrons, these occupy regions of space which we can visualize as *electron clouds*— regions of negative electric charge, also known as *orbitals*— whose precise character can be left to more detailed theories



The covalent model of chemical bonding assumes that the electron pairs responsible for bonding are concentrated into the region of apace between the bonded atoms. The fundamental idea of VSEPR thoery is that these regions of negative electric charge will repel each other, causing them (and thus the chemical bonds that they form) to stay as far apart as possible. Thus the two electron clouds contained in a simple triatomic molecule AX_2 will extend out in opposite directions; an angular separation of 180° places the two bonding orbitals as far away from each other they can get. We therefore expect the two chemical bonds to extend in opposite directions, producing a linear molecule.

If the central atom also contains one or more pairs of **nonbonding** electrons, these additional regions of negative charge will behave very much like those associated with the bonded atoms. The orbitals containing the various bonding and nonbonding pairs in the valence shell will extend out from the central atom in directions that minimize their mutual repulsions. If the central atom possesses partially occupied *d*-orbitals, it may be able to accommodate five or six electron pairs, forming what is sometimes called an "**expanded octet**".

Digonal and trigonal coordination

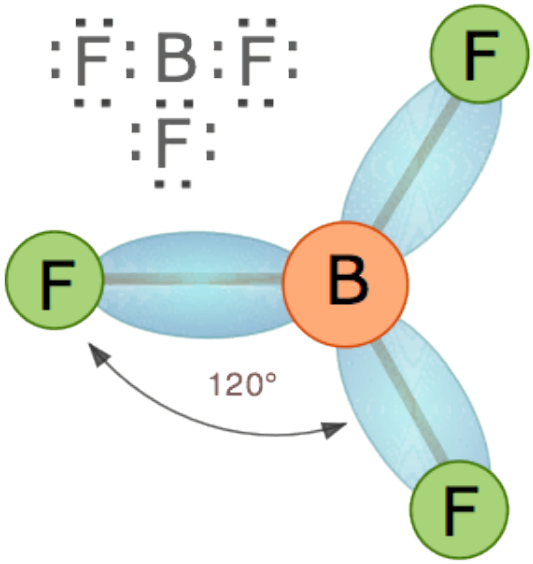
Linear molecules

As we stated above, a simple triatomic molecule of the type AX_2 has its two bonding orbitals 180° apart, producing a molecule that we describe as having *linear* geometry. Examples of triatomic molecules for which VSEPR theory predicts a linear shape are $BeCl_2$ (which, you will notice, doesn't possess enough electrons to conform to the octet rule) and CO_2. If you write out the electron dot formula for carbon dioxide, you will see that the C-O bonds are double bonds. This makes no difference to VSEPR theory; the central carbon atom is still joined to two other atoms, and the electron clouds that connect the two oxygen atoms are 180° apart.



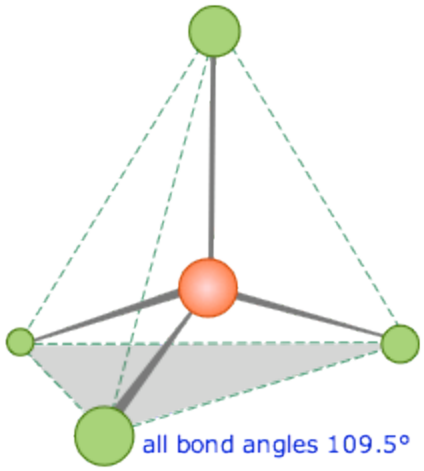
Trigonal molecules

In an AX_3 molecule such as BF_3, there are three regions of electron density extending out from the central atom. The repulsion between these will be at a minimum when the angle between any two is (360° ÷ 3) = 120°. This requires that all four atoms be in the same plane; the resulting shape is called *trigonal planar*, or simply *trigonal*.



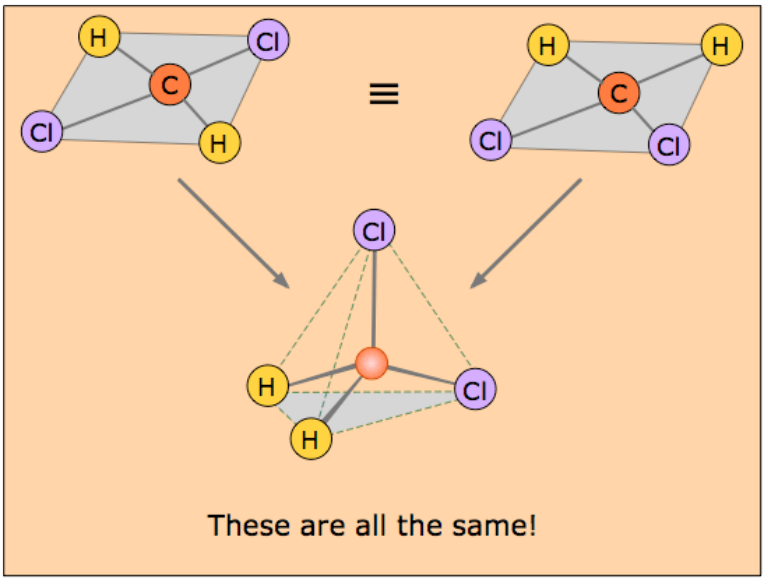
Tetrahedral coordination

Methane, CH_4, contains a carbon atom bonded to four hydrogens. What bond angle would lead to the greatest possible separation between the electron clouds associated with these bonds? In analogy with the preceding two cases, where the bond angles were 360°/2=180° and 360°/3=120°, you might guess 360°/4=90°; if so, you would be wrong. The latter calculation would be correct if all the atoms were constrained to be in the same plane (we will see cases where this happens later), but here there is no such restriction. Consequently, the four equivalent bonds will point in four **geometrically equivalent directions** in three dimensions corresponding to the four corners of a *tetrahedron* centered on the carbon atom. The angle between any two bonds will be 109.5°. This is called *tetrahedral coordination*.



This is the most important coordination geometry in Chemistry: it is imperative that you be able to sketch at least a crude perspective view of a tetrahedral molecule.

It is interesting to note that the tetrahedral coordination of carbon in most of its organic compounds was worked out in the nineteenth century on purely geometrical grounds and chemical evidence, long before direct methods of determining molecular shapes were developed. For example, it was noted that there is only one dichloromethane, CH_2Cl_2.

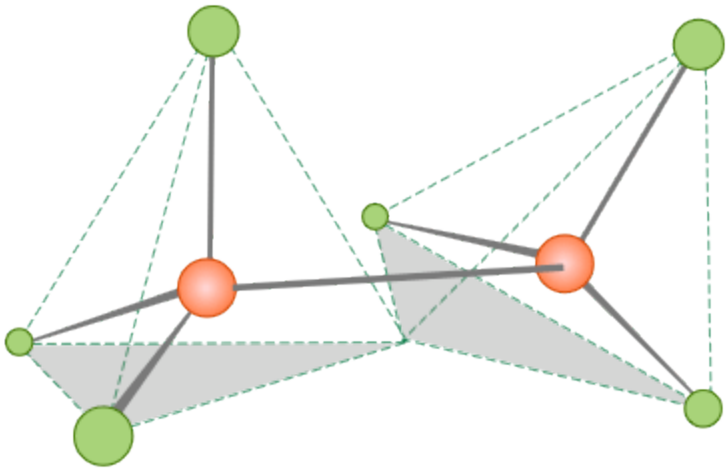


If the coordination around the carbon were square, then there would have to be two isomers of CH_2Cl_2, as shown in the pair of structures here. The distances between the two chlorine atoms would be different, giving rise to differences in physical properties would allow the two isomers to be distinguished and separated.

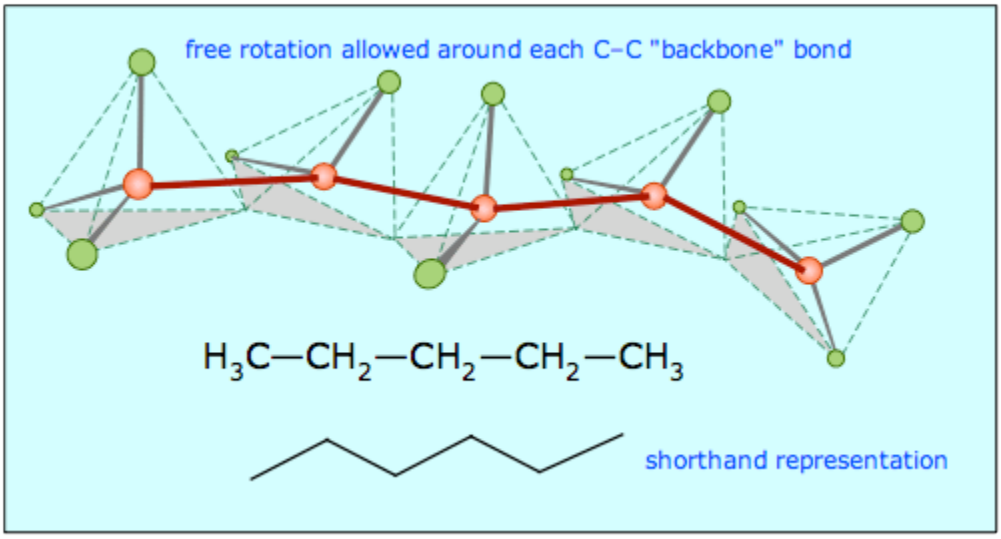
The existence of only one kind of CH_2Cl_2 molecule means that all four positions surrounding the carbon atom are geometrically equivalent, which requires a tetrahedral coordination geometry. If you study the tetrahedral figure closely, you may be able to convince yourself that it represents the connectivity shown on both of the "square" structures at the top. A three-dimensional ball-and-stick mechanical model would illustrate this very clearly.

Tetrahedrally-coordinated carbon chains

tetrahedra joined end-to-end.



Similar *alkane* chains having the general formula $H_3C–(CH_2)_n–CH_3$ (or C_nH_{2n+2}) can be built up; a view of *pentane*, C_5H_{12}, is shown below.



Notice that these "straight chain hydrocarbons" (as they are often known) have a carbon "backbone" structure that is not really straight, as is illustrated by the zig-zag figure that is frequently used to denote hydrocarbon structures.

Coordination geometry and molecular geometry

Coordination number refers to the number of electron pairs that surround a given atom; we often refer to this atom as the *central atom* even if this atom is not really located at the geometrical center of the molecule. If all of the electron pairs surrounding the central atom are shared with neighboring atoms, then the *coordination geometry* is the same as the *molecular geometry*. The application of VSEPR theory then reduces to the simple problem of naming (and visualizing) the geometric shapes associated with various numbers of points surrounding a central point (the central atom) at the greatest possible angles. Both classes of geometry are named after the shapes of the imaginary geometric figures (mostly regular solid polygons) that would be centered on the central atom and would have an electron pair at each vertex.

If one or more of the electron pairs surrounding the central atom is not shared with a neighboring atom (that is, if it is a lone pair), then the molecular geometry is simpler than the coordination geometry, and it can be worked out by inspecting a sketch of the coordination geometry figure.

Tetrahedral coordination with lone pairs

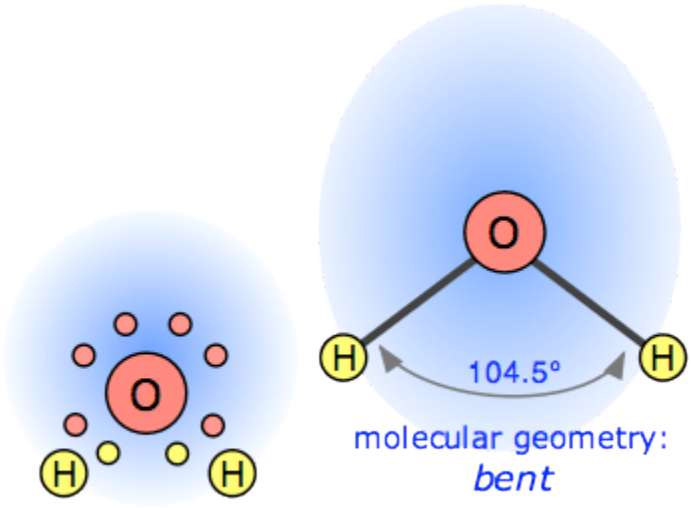
In the examples we have discussed so far, the shape of the molecule is defined by the coordination geometry; thus the carbon in methane is tetrahedrally coordinated, and there is a hydrogen at each corner of the tetrahedron, so the molecular shape is also tetrahedral.

It is common practice to represent bonding patterns by "generic" formulas such as AX_4, AX_2E_2, etc., in which "X" stands for bonding pairs and "E" denotes lone pairs. This convention is known as the "AXE Method."

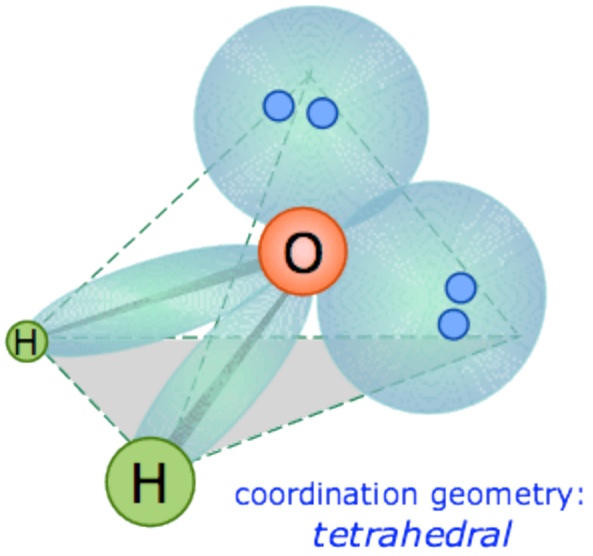
The bonding geometry will not be tetrahedral when the valence shell of the central atom contains nonbonding electrons, however. The reason is that the *nonbonding electrons* are also in orbitals that occupy space and repel the other orbitals. This means that in figuring the coordination number around the central atom, we must count both the bonded atoms and the nonbonding pairs.

The water molecule: AX_2E_2

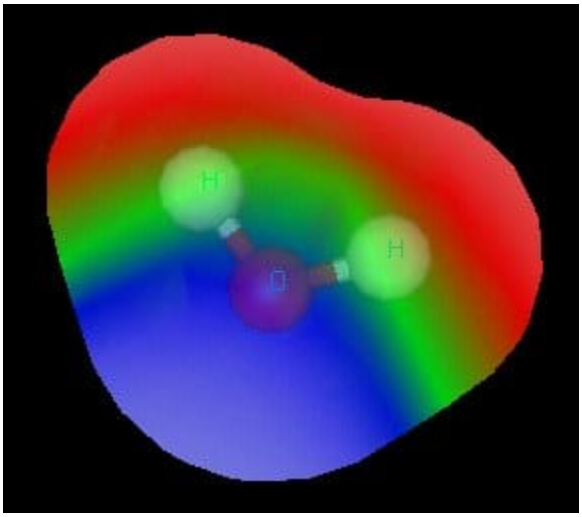
In the water molecule, the central atom is O, and the Lewis electron dot formula predicts that there will be two pairs of nonbonding electrons. The oxygen atom will therefore be tetrahedrally coordinated, meaning that it sits at the center of the tetrahedron as shown below.



Two of the coordination positions are occupied by the shared electron-pairs that constitute the O–H bonds, and the other two by the non-bonding pairs. Thus although the oxygen atom is tetrahedrally coordinated, the bonding geometry (shape) of the H_2O molecule is described as*bent*.



There is an important difference between bonding and non-bonding electron orbitals. Because a nonbonding orbital has no atomic nucleus at its far end to draw the electron cloud toward it, the charge in such an orbital will be concentrated closer to the central atom. As a consequence, nonbonding orbitals exert more repulsion on other orbitals than do bonding orbitals. Thus in H_2O, the two nonbonding orbitals push the bonding orbitals closer together, making the H–O–H angle 104.5° instead of the tetrahedral angle of 109.5°.



This image was produced by a computer simulation based on the more complete molecular orbital model that we describe in the next lesson. Although the water molecule is electrically neutral, it is not electrically uniform; the non-bonding electrons create a higher concentration of negative charge (blue color) at the oxygen end, making the hydrogen side relatively positive (red). This charge unbalance gives rise to many of the so-called anomalous properties of water.

Ammonia: AX_3E

The electron-dot structure of NH_3 places one pair of nonbonding electrons in the valence shell of the nitrogen atom. This means that there are three bonded atoms and one lone pair, for a coordination number of four around the nitrogen, the same as occurs in H_2O. We can therefore predict that the three hydrogen atom will lie at the corners of a tetrahedron centered on the nitrogen atom. The lone pair orbital will point toward the fourth corner of the tetrahedron, but since that position will be vacant, the NH_3molecule itself cannot be tetrahedral. Instead, it assumes a *pyramidal* shape. More precisely, the shape is that of a *trigonal pyramid* (i.e., a pyramid having a triangular base). The hydrogen atoms are all in the same plane, with the nitrogen above (or below, or to the side; molecules of course don't know anything about "above" or "below"!) The fatter orbital containing the non-bonding electrons pushes the bonding orbitals together slightly, making the H–N–H bond angles about 107°.

Computer-generated image of NH_3 molecule showing electrostatic potential (red=+, blue=–.)

Central atoms with five bonds

Compounds of the type AX_5 are formed by some of the elements in Group 15 of the periodic table; PCl_5 and AsF_5 are examples.

In what directions can five electron pairs arrange themselves in space so as to minimize their mutual repulsions? In the cases of coordination numbers 2, 3, 4, and 6, we could imagine that the electron pairs distributed themselves as far apart as possible on the surface of a sphere; for the two higher numbers, the resulting shapes correspond to the regular polyhedron having the same number of sides. The problem with coordination number 5 is that there is no such thing as a regular polyhedron with five vertices.

In 1758, the great mathematician Euler proved that there are only five regular convex polyhedra, known as the platonic solids: tetrahedron (4 triangular faces), octahedron (6 triangular faces), icosahedron (20 triangular faces), cube (6 square faces), and dodecahedron (12 pentagonal faces). Chemical examples of all are known; the first icosahedral molecule, LaC_{60} (in which the La atom has 20 nearest C neighbors) was prepared in 1986.

Besides the five regular solids, there can be 15 semi-regular isogonal solids in which the faces have different shapes, but the vertex angles are all the same. These geometrical principles are quite important in modern structural chemistry.

The shape of PCl_5 and similar molecules is a *trigonal bipyramid*. This consists simply of two triangular-base pyramids joined base-to-base. Three of the chlorine atoms are in the plane of the central phosphorus atom (*equatorial* positions), while the other two atoms are above and below this plane (*axial* positions). Equatorial and axial atoms have different geometrical relationships to their neighbors, and thus differ slightly in their chemical behavior.

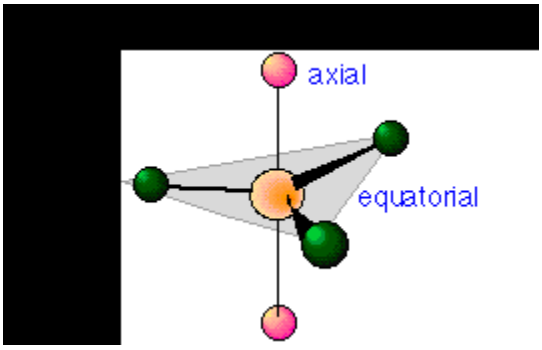
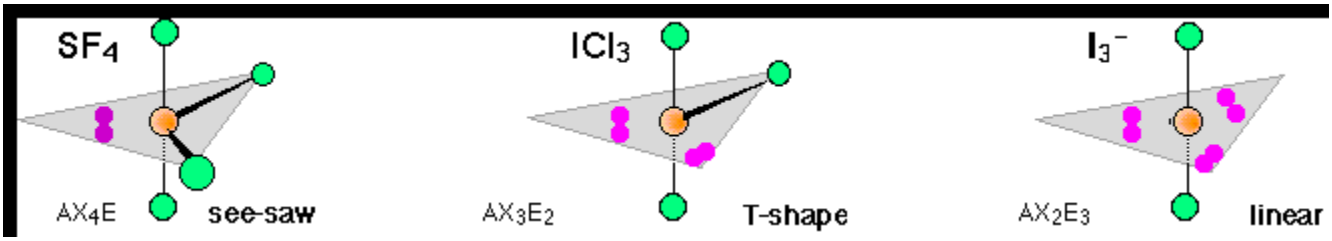


Figure: trigonal bipyramid structure

In 5-coordinated molecules containing lone pairs, these non-bonding orbitals (which you will recall are closer to the central atom and thus more likely to be repelled by other orbitals) will preferentially reside in the equatorial plane. This will place them at 90° angles with respect to no more than two axially-oriented bonding orbitals.

Using this reasoning, we can predict that an AX_4E molecule (that is, a molecule in which the central atom A is coordinated to four other atoms "X" and to one nonbonding electron pair) such as SF_4 will have a "see-saw" shape; substitution of more nonbonding pairs for bonded atoms reduces the triangular bipyramid coordination to even simpler molecular shapes, as shown below.



Octahedral coordination

Just as four electron pairs experience the minimum repulsion when they are directed toward the corners of a tetrahedron, six electron pairs will try to point toward the corners of an *octahedron*. An octahedron is not as complex a shape as its name might imply; it is simply two square-based pyramids joined base to base. You should be able to sketch this shape as well as that of the tetrahedron.

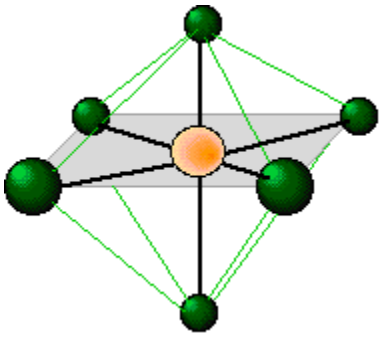


Figure: octahedrally-coordinated molecule

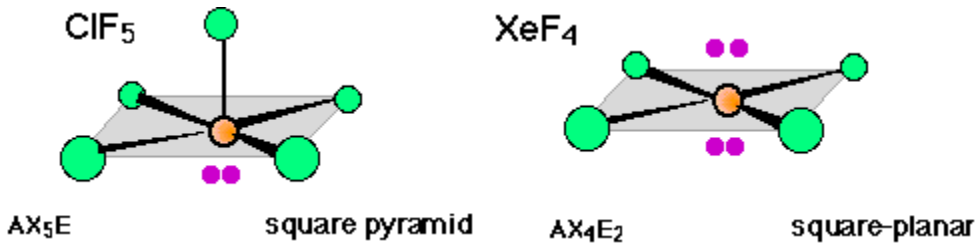
The shaded plane shown in this octahedrally-coordinated molecule is only one of three equivalent planes defined by a four-fold symmetry axis. All the ligands are geometrically equivalent; there are no separate axial and equatorial positions in an AX_6 molecule.

At first, you might think that a coordination number of six is highly unusual; it certainly violates the octet rule, and there are only a few molecules (SF_6 is one) where the central atom is hexavalent. It turns out, however, that this is one of the most commonly encountered coordination numbers in inorganic chemistry. There are two main reasons for this:

- Many transition metal ions form coordinate covalent bonds with lone-pair electron donor atoms such as N (in NH_3) and O (in H_2O). Since transition elements can have an outer configuration of $d^{10}s^2$, up to six electron pairs can be accommodated around the central atom. A coordination number of 6 is therefore quite common in transition metal hydrates, such as $Fe(H_2O)_6^{3+}$.
- Although the central atom of most molecules is bonded to fewer than six other atoms, there is often a sufficient number of lone pair electrons to bring the total number of electron pairs to six.

Octahedral coordination with lone pairs

There are well known examples of 6-coordinate central atoms with 1, 2, and 3 lone pairs. Thus all three of the molecules whose shapes are depicted below possess octahedral coordination around the central atom. Note also that the orientation of the shaded planes shown in the two rightmost images are arbitrary; since all six vertices of an octahedron are identical, the planes could just as well be drawn in any of the three possible vertical orientations.



Summary of VSEPR theory

The VSEPR model is an extraordinarily powerful one, considering its great simplicity. Its application to predicting molecular structures can be summarized as follows:

1. 1. Electron pairs surrounding a central atom repel each other; this repulsion will be minimized if the orbitals containing these electron pairs point as far away from each other as possible.
2. 2. The coordination geometry around the central atom corresponds to the polyhedron whose number of vertices is equal to the number of surrounding electron pairs (coordination number). Except for the special case of 5, and the trivial cases of 2 and 3, the shape will be one of the regular polyhedra.
3. 3. If some of the electron pairs are nonbonding, the shape of the molecule will be simpler than that of the coordination polyhedron.
4. 4. Orbitals that contain nonbonding electrons are more concentrated near the central atom, and therefore offer more repulsion than bonding pairs to other orbitals.

While VSEPR theory is quite good at predicting the general shapes of most molecules, it cannot yield exact details. For example, it does not explain why the bond angle in H_2O is 104.5°, but that in H_2S is about 90°. This is not surprising, considering that the emphasis is on electronic repulsions, without regard to the detailed nature of the orbitals containing the electrons, and thus of the bonds themselves.

The Valence Shell Electron Repulsion theory was developed in the 1960s by Ronald Gillespie of McMaster University (Hamilton, Ontario, Canada) and Ronald Nyholm (University College, London). It is remarkable that what seems to be a logical extension of the 1916 Lewis shared-electron pair model of bonding took so long to be formulated; it was first presented in the authors' classic article *Inorganic Stereochemistry* published in the 1957 *Chemical Society of London Quarterly Reviews* (Vol 11, pg 339). Although it post-dates the more complete quantum mechanical models, it is easy to grasp and within a decade had become a staple of every first-year college chemistry course.

This page titled 9.5: Molecular Geometry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.6: The Hybrid Orbital Model

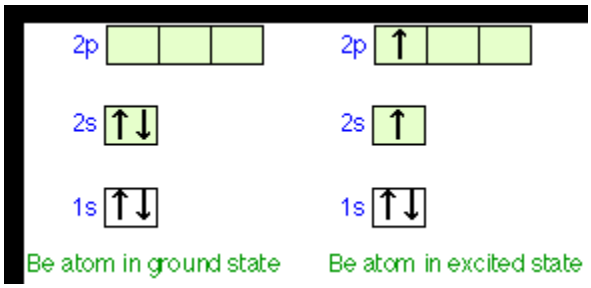
Learning Objectives

- Explain why the sharing of atomic orbitals (as implied in the Lewis model) cannot adequately account for the observed bonding patterns in simple molecules.
- Sketch out a diagram illustrating how the plots of atomic s- and *p*- orbital wave functions give rise to a pair of hybrid orbitals.
- Draw "orbital box" diagrams showing how combinations of an atomic *s* orbital and various numbers of *p* orbitals create *sp*, sp^2, and sp^3 hybrid orbitals.
- Show how hybrid orbitals are involved in the molecules methane, water, and ammonia.

As useful and appealing as the concept of the *shared-electron pair bond* is, it raises a somewhat troubling question that we must sooner or later face: what is the nature of the **orbitals** in which the shared electrons are contained? Up until now, we have been tacitly assuming that each valence electron occupies the same kind of atomic orbital as it did in the isolated atom. As we shall see below, his assumption very quickly leads us into difficulties.

Atomic orbitals alone do not work for Molecules

Consider how we might explain the bonding in a compound of divalent beryllium, such as beryllium hydride, BeH_2. The beryllium atom, with only four electrons, has a configuration of $1s^2 2s^2$. Note that the two electrons in the 2*s* orbital have opposite spins and constitute a stable pair that has no tendency to interact with unpaired electrons on other atoms.



The only way that we can obtain two unpaired electrons for bonding in beryllium is to promote one of the 2*s* electrons to the 2*p* level. However, the energy required to produce this excited-state atom would be sufficiently great to discourage bond formation. It is observed that Be does form reasonably stable bonds with other atoms. Moreover, the two bonds in BeH_2 and similar molecules are completely equivalent; this would not be the case if the electrons in the two bonds shared Be orbitals of different types, as in the "excited state" diagram above.

These facts suggest that it is incorrect to assume that the distribution of valence electrons that are shared with other atoms can be described by atomic-type *s*, *p*, and *d* orbitals at all.

Remember that these different orbitals arise in the first place from the interaction of the electron with the *single central* electrostatic force field associated with the positive nucleus. An outer-shell electron in a bonded atom will be under the influence of a force field emanating from *two* positive nuclei, so we would expect the orbitals in the bonded atoms to have a somewhat different character from those in free atoms. In fact, as far as valence electrons are concerned, we can throw out the concept of atomic orbital altogether and reassign the electrons to a new set of *molecular orbitals* that are characteristic of each molecular configuration. This approach is indeed valid, but we will defer a discussion of it until a later unit.

For now, we will look at a less-radical model that starts out with the familiar valence-shell atomic orbitals, and allows them to combine to form *hybrid orbitals* whose shapes conform quite well to the bonding geometry that we observe in a wide variety of molecules.

What are hybrid orbitals?

First, recall that the electron, being a quantum particle, cannot have a distinct location; the most we can do is define the region of space around the nucleus in which the probability of finding the electron exceeds some arbitrary value, such as 90% or 99%. This region of space is the **orbital**. Because of the wavelike character of matter, the orbital corresponds to a standing wave pattern in 3-dimensional space which we can often represent more clearly in 2-dimensional cross section. The quantity that is varying ("waving") is a **number** denoted by ψ (*psi*) whose value varies from point to point according to the wave function for that particular orbital.

Orbitals of all types are simply mathematical functions that describe particular standing-wave patterns that can be plotted on a graph but have no physical reality of their own. Because of their wavelike nature, two or more orbitals (i.e., two or more functions ψ) can be combined both in-phase and out-of-phase to yield a pair of resultant orbitals which, to be useful, must have squares that describe actual electron distributions in the atom or molecule.

The *s,p,d* and *f* orbitals that you are familiar with are the most convenient ones for describing the electron distribution in isolated atoms because assignment of electrons to them according to the usual rules always yields an overall function Ψ^2 that predicts a spherically symmetric electron distribution, consistent with all physical evidence that atoms are in fact spherical. For atoms having more than one electron, however, the *s,p,d*, *f* basis set is only one of many possible ways of arriving at the same observed electron distribution. We use it not because it is unique, but because it is the simplest.

In the case of a molecule such as BeH_2, we know from experimental evidence that the molecule is linear and therefore the electron density surrounding the central atom is no longer spherical, but must be concentrated along two directions 180° apart, and we need to construct a function Ψ^2 having these geometrical properties. There are any number of ways of doing this, but it is convenient is to use a particular set of functions ψ (which we call *hybrid orbitals*) that are constructed by combining the atomic *s,p,d,*and *f* functions that are already familiar to us.

You should understand that *hybridization is not a physical phenomenon*; it is merely a *mathematical operation* that combines the atomic orbitals we are familiar with in such a way that the new (hybrid) orbitals possess the geometric and other properties that are reasonably consistent with what we observe in a wide range (but certainly not in all) molecules. In other words, hybrid orbitals are abstractions that describe reality fairly well in certain classes of molecules (and fortunately, in much of the very large class of organic substances) and are therefore a useful means of organizing a large body of chemical knowledge... but they are far from infallible.

> *Hybridization is not a physical phenomenon*; it is merely a *mathematical operation* that combines the atomic orbitals we are familiar with in such a way that the new (hybrid) orbitals possess the geometric and other properties that are reasonably consistent with what we observe in a wide range (but certainly not in all) molecules.

This approach, which assumes that the orbitals remain more or less localized on one central atom, is the basis of the theory which was developed in the early 1930s, mainly by *Linus Pauling*.

Linus Pauling

Linus Pauling (1901-1994) was the most famous American chemist of the 20th century and the author of the classic book *The Nature of the Chemical Bond*. His early work pioneered the application of X-ray diffraction to determine the structure of complex molecules; he then went on to apply quantum theory to explain these observations and predict the bonding patterns and energies of new molecules. Pauling, who spent most of his career at Cal Tech, won the Nobel Prize for Chemistry in 1954 and the Peace Prize in 1962.

"In December 1930 Pauling had his famous 'breakthrough' where, in a rush of inspiration, he '*stayed up all night, making, writing out, solving the equations, which were so simple that I could solve them in a few minutes*'. This flurry of calculations would eventually become the first of Pauling's germinal series of papers on the nature of the chemical bond. '*I just kept getting more and more euphorious as time went by*', Pauling would recall. "

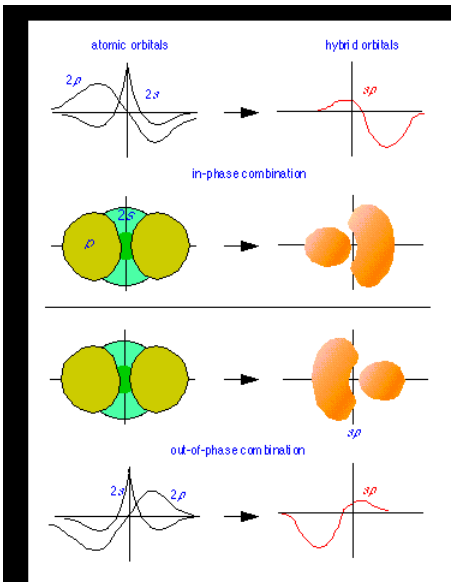
Although the hybrid orbital approach has proven very powerful (especially in organic chemistry), it does have its limitations. For example, it predicts that both H_2O and H_2S will be tetrahedrally coordinated bent molecules with bond angles slightly smaller than the tetrahedral angle of 109.5° owing to greater repulsion by the nonbonding pair. This description fits water (104.5°) quite

well, but the bond angle in hydrogen sulfide is only 92°, suggesting that atomic *p* orbitals (which are 90° apart) provide a better description of the electron distribution about the sulfur atom than do sp^3 hybrid orbitals.

The hybrid orbital model is fairly simple to apply and understand, but it is best regarded as one special way of looking at a molecule that can often be misleading. Another viewpoint, called the *molecular orbital theory*, offers us a complementary perspective that it is important to have if we wish to develop a really thorough understanding of chemical bonding in a wider range of molecules.

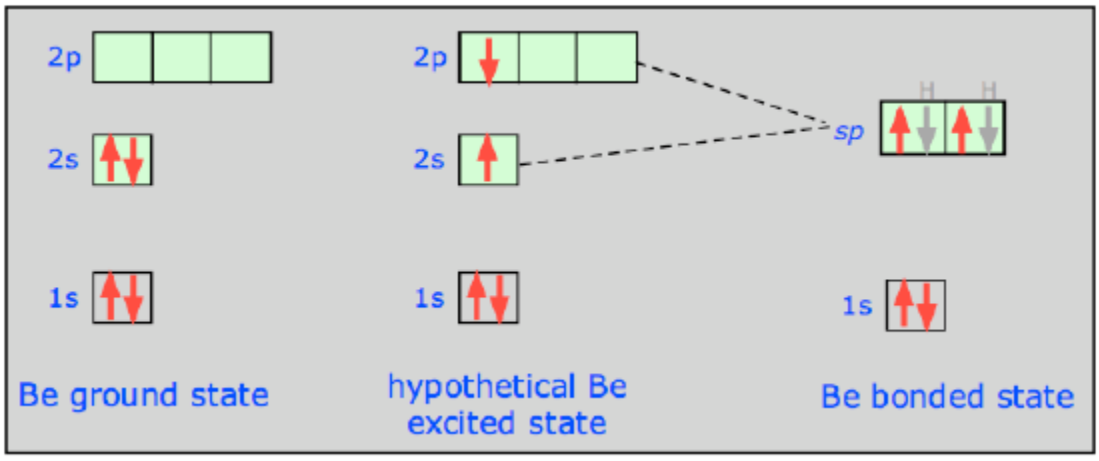
Constructing hybrid orbitals

Below: "Constructive" and "destructive" combinations of 2*p* and 2*s* wave functions (line plots) give rise to the sp hybrid function shown at the right. The solid figures depict the corresponding probability functions ψ^2.

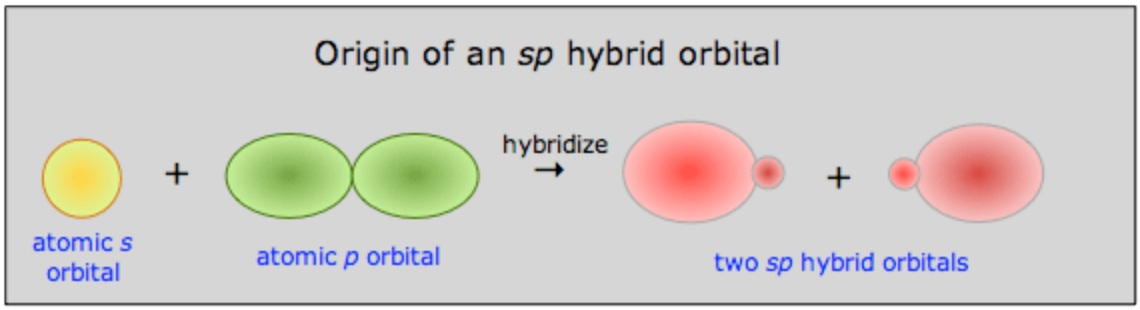


Hybrid orbitals are constructed by combining the ψ functions for atomic orbitals. Because wave patterns can combine both constructively and destructively, a pair of atomic wave functions such as the s- and *p*- orbitals shown at the left can combine in two ways, yielding the *sp hybrids* shown.

From an energy standpoint, we can represent the transition from atomic s- and *p*-orbitals to an *sp* hybrid orbital in this way:



Notice here that 1) the total number of occupied orbitals is conserved, and 2) the two *sp* hybrid orbitals are intermediate in energy between their parent atomic orbitals. In terms of plots of the actual orbital functions ψ we can represent the process as follows:

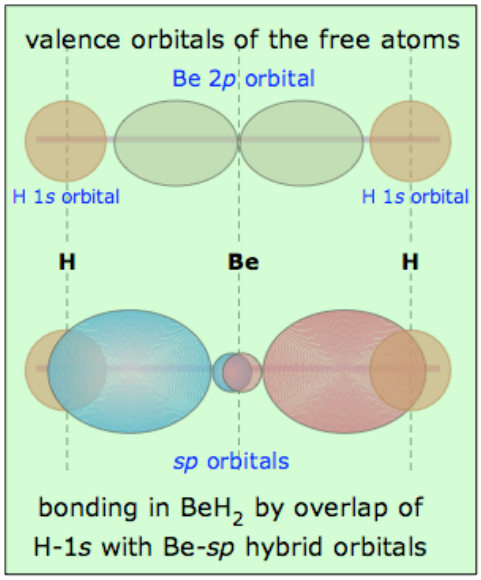


The probability of finding the electron at any location is given not by ψ, but by ψ^2, whose form is roughly conveyed by the solid figures in this illustration.

Hybrids derived from atomic s- and *p* orbitals

Digonal bonding: sp-hybrid orbitals

Returning to the example of BeH_2, we can compare the valence orbitals in the free atoms with those in the beryllium hydride molecule as shown here. It is, of course, the *overlap* between the hydrogen-1s orbitals and the two lobes of the beryllium *sp*-hybrid orbitals that constitutes the two Be—H "bonds" in this molecule. Notice that whereas a single *p*-orbital has lobes on both sides of the atom, a single *sp*-hybrid has most of its electron density on one side, with a minor and more spherical lobe on the other side. This minor lobe is centered on the central atom (some textbook illustrations don't get this right.)

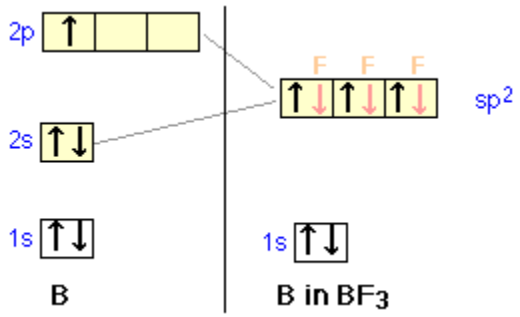


As far as the *shape* of the molecule is concerned, the result is exactly the same as predicted by the VSEPR model (although hybrid orbital theory predicts the same result in a more fundamental way.) We can expect any central atom that uses *sp*-hybridization in bonding to exhibit linear geometry when incorporated into a molecule.

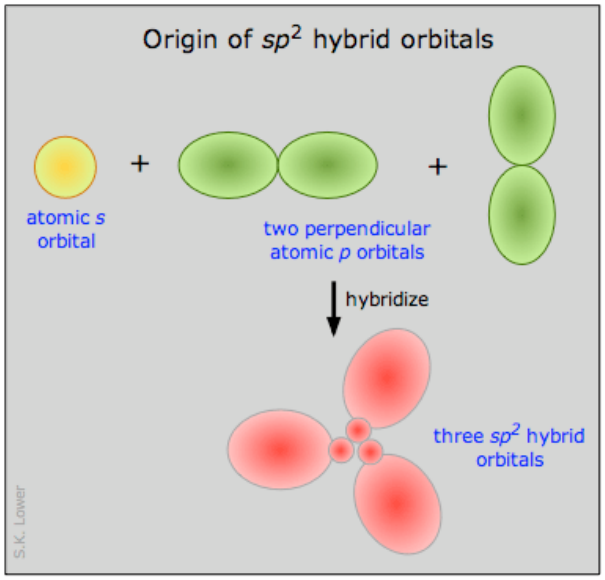
Trigonal (sp^2) hybridization

We can now go on to apply the same ideas to some other simple molecules. In **boron trifluoride**, for example, we start with the boron atom, which has three outer-shell electrons in its normal or ground state, and three fluorine atoms, each with seven outer electrons. As is shown in this configuration diagram, one of the three boron electrons is unpaired in the ground state. In order to explain the trivalent bonding of boron, we postulate that the atomic s- and *p*- orbitals in the outer shell of boron mix to form three equivalent hybrid orbitals. These particular orbitals are called sp^2 *hybrids*, meaning that this set of orbitals is derived from one s-

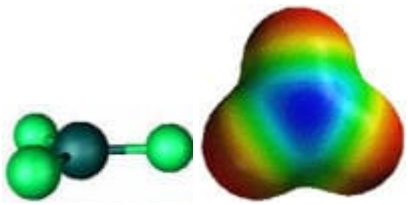
orbital and two p-orbitals of the free atom.



This illustration shows how an s-orbital mixes with two *p* orbitals to form a set of three sp^2 hybrid orbitals. Notice again how the three atomic orbitals yield the same number of hybrid orbitals.

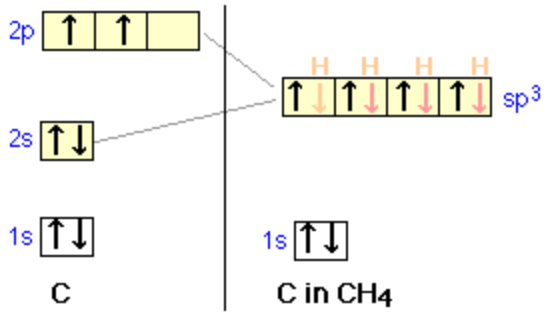


Boron Trifluoride BF_3 is a common example of sp^2 hybridization. The molecule has plane trigonal geometry.

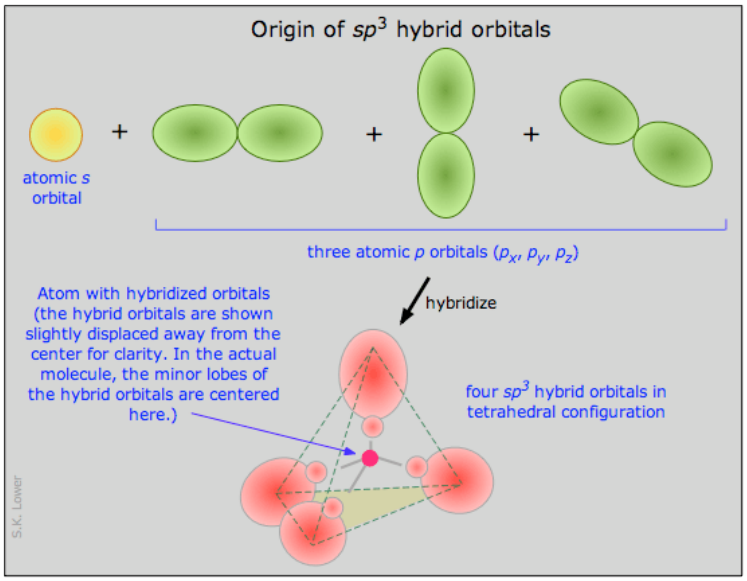


Tetrahedral (sp^3) hybridization

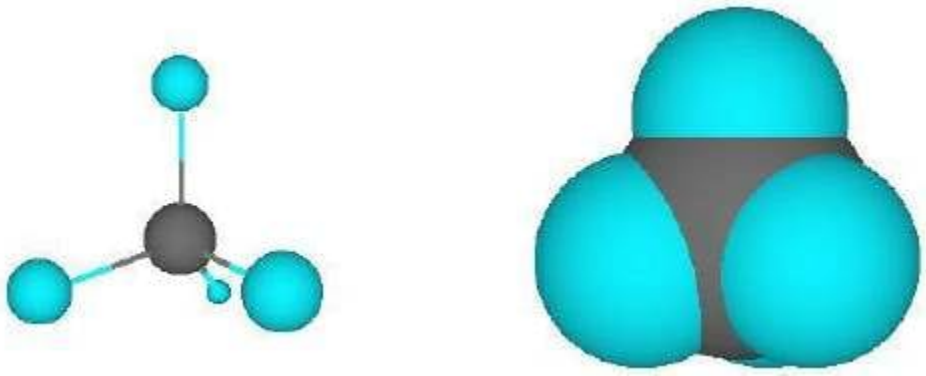
Let us now look at several tetravalent molecules, and see what kind of hybridization might be involved when four outer atoms are bonded to a central atom. Perhaps the commonest and most important example of this bond type is methane, CH_4.



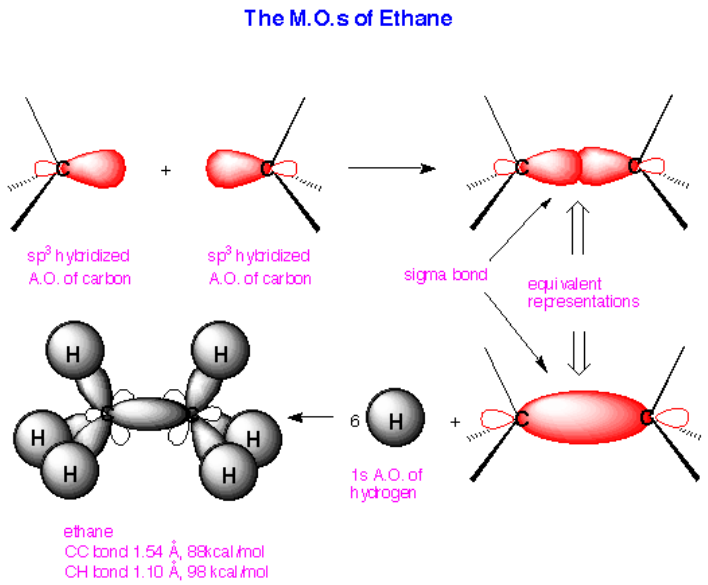
orbitals of carbon mix into four sp^3 hybrid orbitals which are chemically and geometrically identical; the latter condition implies that the four hybrid orbitals extend toward the corners of a tetrahedron centered on the carbon atom.



Methane is the simplest hydrocarbon; the molecule is approximately spherical, as is shown in the space-filling model:



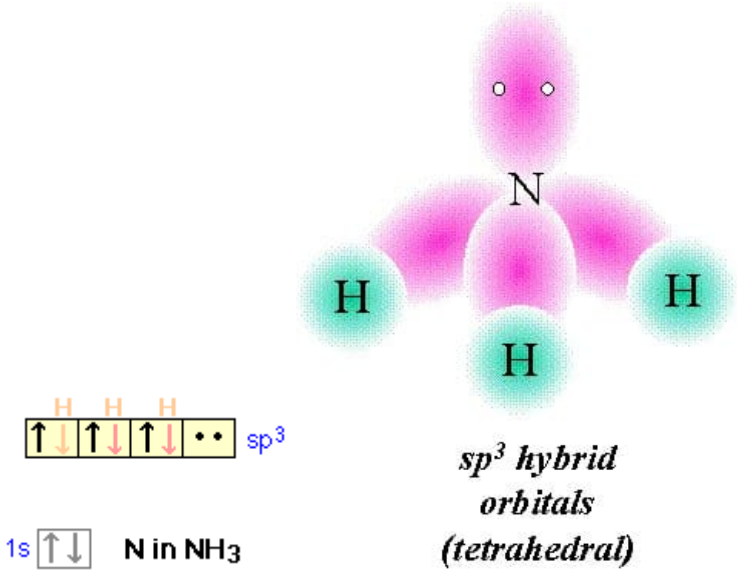
By replacing one or more of the hydrogen atoms in CH_4 with another sp^3 hybridized carbon fragments, hydrocarbon chains of any degree of complexity can be built up. The simplest of these is ethane:



This shows how an sp^3 orbital on each of two two carbon atoms join (overlap) to form a carbon-carbon bond, and then the remaining carbon sp^3 orbital overlaps with six hydrogen 1s orbitals to form the ethane molecule.

Lone pair electrons in hybrid orbitals

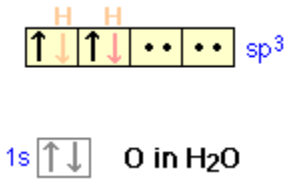
If lone pair electrons are present on the central atom, these can occupy one or more of the sp^3 orbitals. This causes the molecular geometry to be different from the coordination geometry, which remains tetrahedral. In the ammonia molecule, for example, the nitrogen atom normally has three unpaired p electrons, but by mixing the 2*s* and 2*p* orbitals, we can create four sp^3-hybrid orbitals just as in carbon. Three of these can form shared-electron bonds with hydrogen, resulting in ammonia, NH_3. The fourth of the sp^3 hybrid orbitals contains the two remaining outer-shell electrons of nitrogen which form a non-bonding lone pair. In acidic solutions these can coordinate with a hydrogen ion, forming the ammonium ion NH_4^+.



Although no bonds are formed by the lone pair in NH_3, these electrons do give rise to a charge cloud that takes up space just like any other orbital.

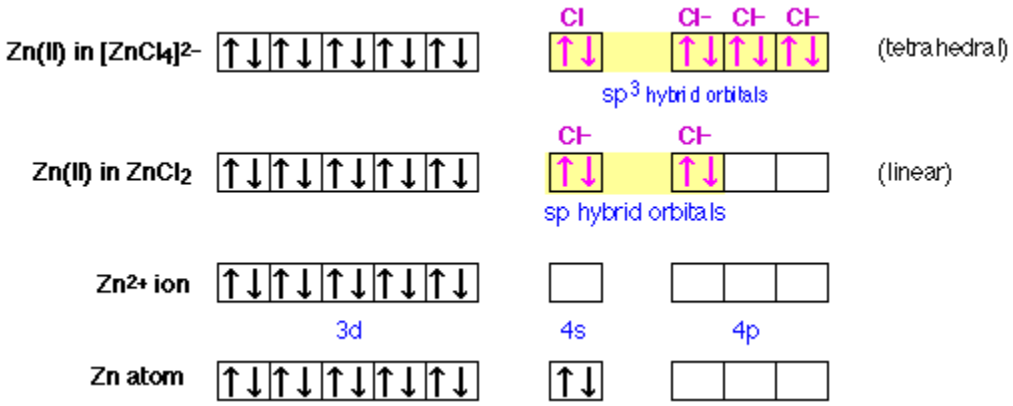
In the **water** molecule, the oxygen atom can form four sp^3 orbitals. Two of these are occupied by the two lone pairs on the oxygen

atom, while the other two are used for bonding. The observed H-O-H bond angle in water (104.5°) is less than the tetrahedral angle (109.5°); one explanation for this is that the non-bonding electrons tend to remain closer to the central atom and thus exert greater repulsion on the other orbitals, thus pushing the two bonding orbitals closer together.



Molecular ions

Hybridization can also help explain the existence and structure of many **inorganic molecular ions**. Consider, for example, electron configurations of **zinc** in the compounds in the illustrations below. The **tetrachlorozinc ion** (top row) is another structure derived from zinc and chlorine. As we might expect, this ion is tetrahedral; there are four chloride ions surrounding the central zinc ion. The zinc ion has a charge of +2, and each chloride ion is –1, so the net charge of the complex ion is –2.



At the bottom is shown the electron configuration of atomic zinc, and just above it, of the divalent zinc ion. Notice that this ion has no electrons at all in its 4-shell. In zinc chloride, shown in the next row up, there are two equivalent chlorine atoms bonded to the zinc. The bonding orbitals are of sp character; that is, they are hybrids of the 4s and one 4p orbital of the zinc atom. Since these orbitals are empty in the isolated zinc ion, the bonding electrons themselves are all contributed by the chlorine atoms, or rather, the chloride ions, for it is these that are the bonded species here. Each chloride ion possesses a complete octet of electrons, and two of these electrons occupy each sp bond orbital in the zinc chloride complex ion. This is an example of a *coordinate covalent bond*, in which the bonded atom contributes both of the electrons that make up the shared pair.

This page titled 9.6: The Hybrid Orbital Model is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.7: The Hybrid Orbital Model II

Learning Objectives

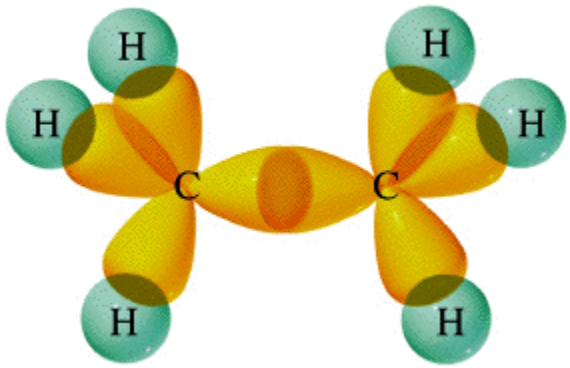
Make sure you thoroughly understand the following essential ideas:

- Sketch out diagrams showing the hybridization and bonding in compounds containing single, double, and triple carbon-carbon bonds.
- Define *sigma* and *pi* bonds.
- Describe the hybridization and bonding in the *benzene* molecule.

This is a continuation of the previous page which introduced the *hybrid orbital* model and illustrated its use in explaining how valence electrons from atomic orbitals of s and p types can combine into equivalent shared-electron pairs known as *sp*, sp^2, and sp^3 hybrid orbitals. In this lesson, we extend this idea to compounds containing double and triple bonds, and to those in which atomic d electrons are involved (and which do not follow the octet rule.)

Hybrid types and Multiple bonds

We have already seen how *sp* hybridization in carbon leads to its combining power of four in the methane molecule. Two such tetrahedrally coordinated carbons can link up together to form the molecule **ethane C_2H_6**. In this molecule, each carbon is bonded in the same way as the other; each is linked to four other atoms, three hydrogens and one carbon. The ability of carbon-to-carbon linkages to extend themselves indefinitely and through all coordination positions accounts for the millions of organic molecules that are known.

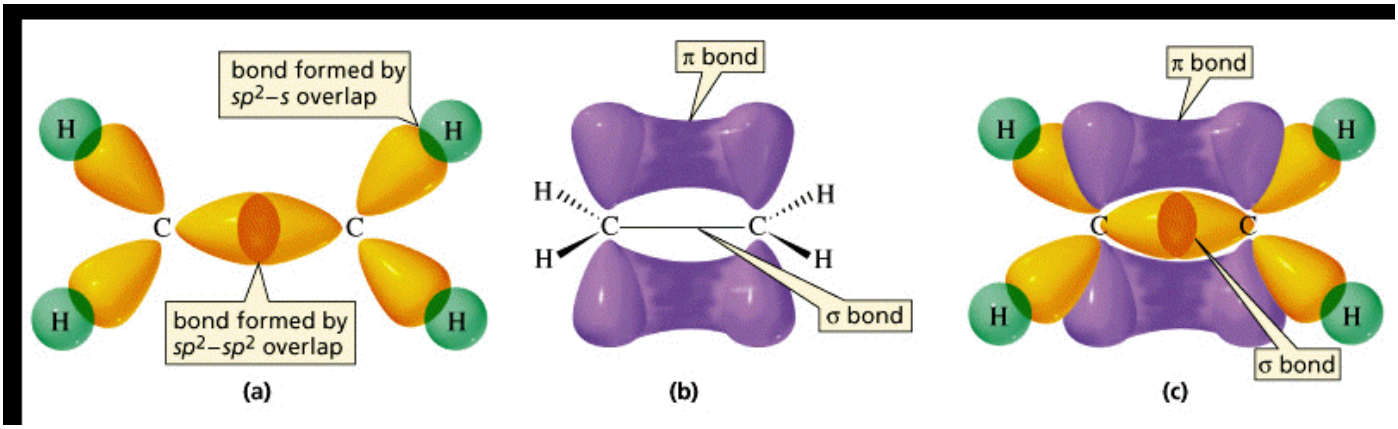


Trigonal hybridization in carbon: the double bond

Carbon and hydrogen can also form a compound **ethylene (ethene)** in which each carbon atom is linked to only three other atoms. Here, we can regard carbon as being trivalent. We can explain this trivalence by supposing that the orbital hybridization in carbon is in this case not sp^3, but is sp^2 instead; in other words, only two of the three *p*orbitals of carbon mix with the 2*s* orbital to form hybrids; the remaining p-orbital, which we will call the i orbital, remains unhybridized. Each carbon is bonded to three other atoms in the same kind of plane trigonal configuration that we saw in the case of boron trifluoride, where the same kind of hybridization occurs. Notice that the bond angles around each carbon are all 120°.

This alternative hybridization scheme explains how carbon can combine with four atoms in some of its compounds and with three other atoms in other compounds. You may be aware of the conventional way of depicting carbon as being tetravalent in all its compounds; it is often stated that carbon always forms four bonds, but that sometimes, as in the case of ethylene, one of these may be a double bond. This concept of the multiple bond preserves the idea of tetravalent carbon while admitting the existence

of molecules in which carbon is clearly combined with fewer than four other atoms.



These three views of the ethylene molecule emphasize different aspects of the disposition of shared electron pairs in the various bonding orbitals of ethene (ethylene). **(a)** The "backbone" structure consisting of σ (*sigma*) bonds formed from the three sp^2-hybridized orbitals on each carbon. **(b)** The π (*pi*) bonding system formed by overlap of the unhybridized p_z orbital on each carbon. The π orbital has two regions of electron density extending above and below the plane of the molecule. **(c)** A cutaway view of the combined σ and π system.

orbital that is perpendicular to the molecular plane. These two parallel p_z orbitals will interact with each other; the two orbitals merge, forming a sausage-like charge cloud (the π bond) that extends both above and below the plane of the molecule. It is the pair of electrons that occupy this new extended orbital that constitutes the “fourth” bond to each carbon, and thus the “other half” of the double bond in the molecule.

More about *sigma* and *pi* bonds

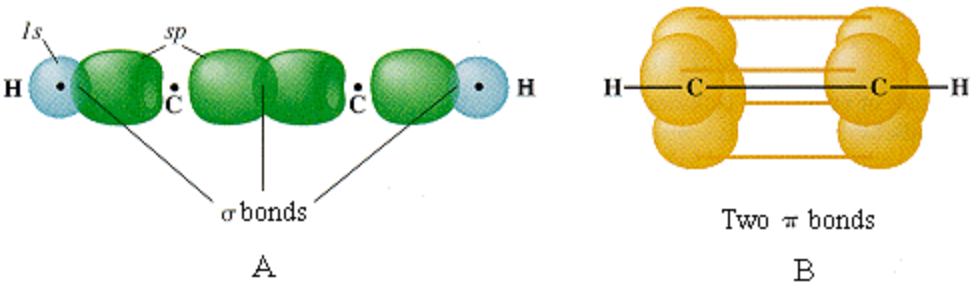
The σ (*sigma*) bond has its maximum electron density along the line-of-centers joining the two atoms (below left). Viewed end-on, the σ bond is cylindrically symmetrical about the line-of-centers. It is this symmetry, rather than its parentage, that defines the σ bond, which can be formed from the overlap of two s-orbitals, from two *p*-orbitals arranged end-to-end, or from an s- and a *p*-orbital. They can also form when *sp* hybrid orbitals on two atoms overlap end-to-end.

Pi orbitals, on the other hand, require the presence of two atomic *p* orbitals on adjacent atoms. Most important, the charge density in the π orbital is concentrated above and below the molecular plane; it is almost zero along the line-of-centers between the two atoms. It is this perpendicular orientation with respect to the molecular plane (and the consequent lack of cylindrical symmetry) that defines the π orbital. The combination of a σ bond and a π bond extending between the same pair of atoms constitutes the double bond in molecules such as ethylene.

Carbon-carbon triple bonds: *sp* hybridization in acetylene

We have not yet completed our overview of multiple bonding, however. Carbon and hydrogen can form yet another compound, acetylene (ethyne), in which each carbon is connected to only two other atoms: a carbon and a hydrogen. This can be regarded as an example of divalent carbon, but is usually rationalized by writing a triple bond between the two carbon atoms.

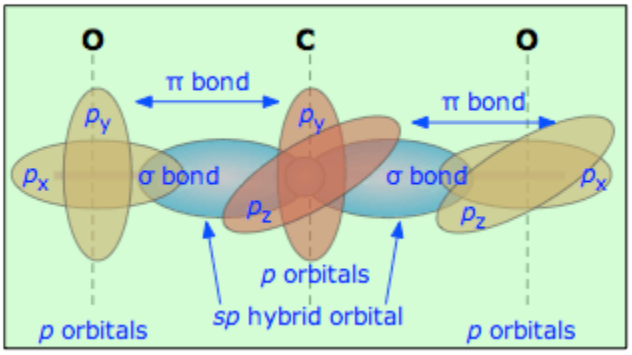
We assume here that since two geometrically equivalent bonds are formed by each carbon, this atom must be *sp*-hybridized in acetylene. On each carbon, one *sp* hybrid bonds to a hydrogen and the other bonds to the other carbon atom, forming the σ bond skeleton of the molecule. In addition to the *sp* hybrids, each carbon atom has two half-occupied *p* orbitals oriented at right angles to each other and to the interatomic axis. These two sets of parallel and adjacent *p* orbitals can thus merge into two sets of π orbitals.



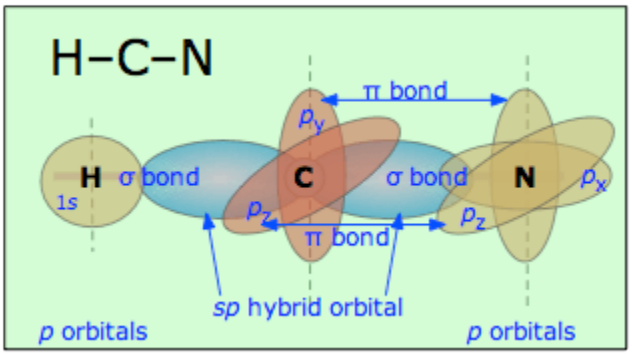
The triple bond in acetylene is seen to consist of one σ bond joining the line-of-centers between the two carbon atoms, and two π bonds whose lobes of electron density are in mutually-perpendicular planes. The acetylene molecule is of course linear, since the angle between the two *sp* hybrid orbitals that produce the *s* skeleton of the molecule is 180°.

Multiple bonds between unlike atoms

Multiple bonds can also occur between dissimilar atoms. For example, in **carbon dioxide** each carbon atom has two unhybridized atomic *p* orbitals, and each oxygen atom still has one *p* orbital available. When the two O-atoms are brought up to opposite sides of the carbon atom, one of the *p* orbitals on each oxygen forms a π bond with one of the carbon *p*-orbitals. In this case, *sp*-hybridization is seen to lead to two double bonds. Notice that the two C–O π bonds are mutually perpendicular.



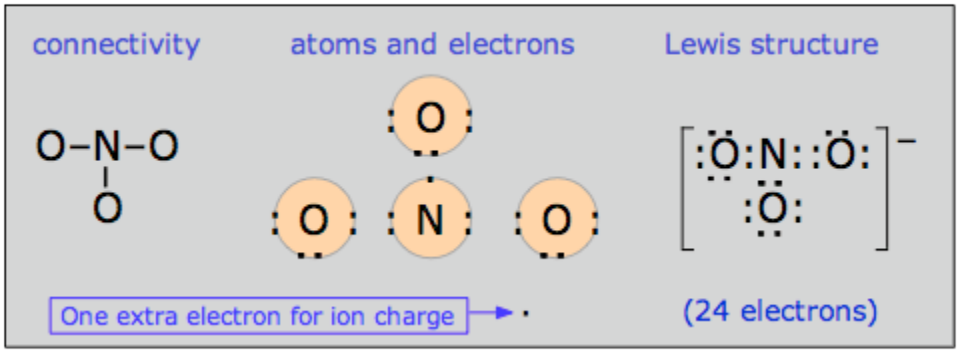
Similarly, in **hydrogen cyanide**, HCN, we assume that the carbon is *sp*-hybridized, since it is joined to only two other atoms, and is hence in a divalent state. One of the *sp*-hybrid orbitals overlaps with the hydrogen 1s orbital, while the other overlaps end-to-end with one of the three unhybridized *p* orbitals of the nitrogen atom. This leaves us with two nitrogen *p*-orbitals which form two mutually perpendicular π bonds to the two atomic *p* orbitals on the carbon. Hydrogen cyanide thus contains one single and one triple bond. The latter consists of a σ bond from the overlap of a carbon *sp* hybrid orbital with a nitrogen *p* orbital, plus two mutually perpendicular π bonds deriving from parallel atomic *p* orbitals on the carbon and nitrogen atoms.



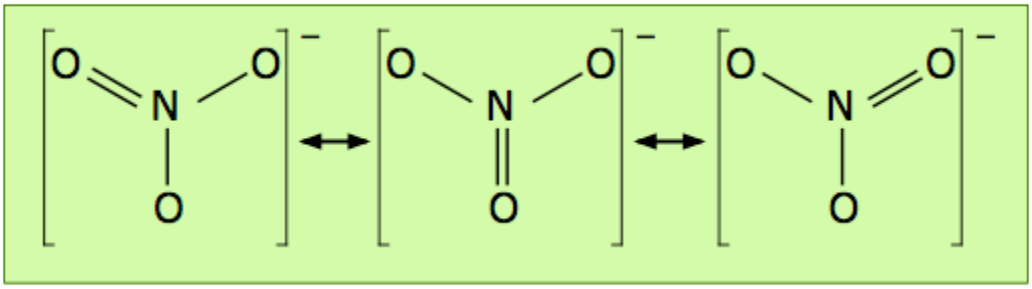
The nitrate ion

Pi bond delocalization furnishes a means of expressing the structures of other molecules that require more than one electron-dot

or structural formula for their accurate representation. A good example is the nitrate ion, which contains 24 electrons:

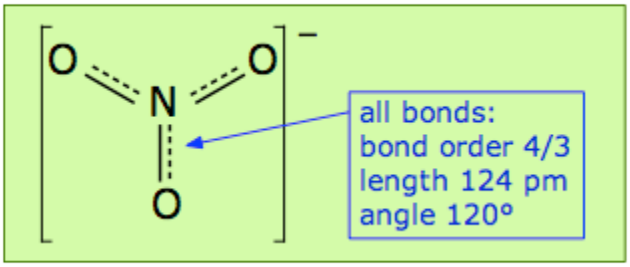


The electron-dot formula shown above is only one of three equivalent resonance structures that are needed to describe trigonal symmetry of this ion.

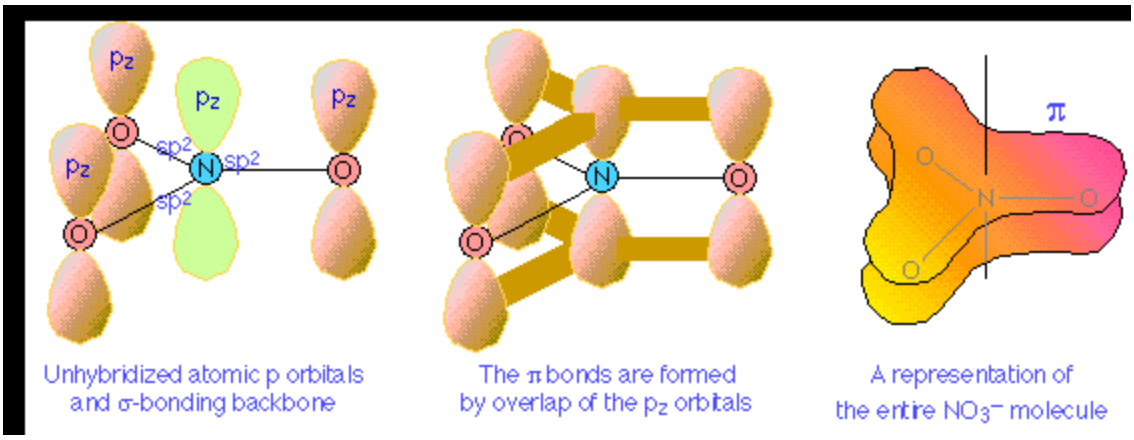


Nitrogen has three half-occupied *p* orbitals available for bonding, all perpendicular to one another. Since the nitrate ion is known to be planar, we are forced to assume that the nitrogen outer electrons are sp^2-hybridized. The addition of an extra electron fills all three hybrid orbitals completely. Each of these filled sp^2 orbitals forms a σ bond by overlap with an empty oxygen $2p_z$ orbital; this, you will recall, is an example of*coordinate covalent bonding*, in which one of the atoms contributes both of the bonding electrons. The empty oxygen 2*p* orbital is made available when the oxygen electrons themselves become *sp* hybridized; we get three filled *sp* hybrid orbitals, and an empty 2*p* atomic orbital, just as in the case of nitrogen.

The π bonding system arises from the interaction of one of the occupied oxygen *sp*orbitals with the unoccupied $2p_x$ orbital of the nitrogen. Notice that this, again, is a coordinate covalent sharing, except that in this instance it is the oxygen atom that donates both electrons.

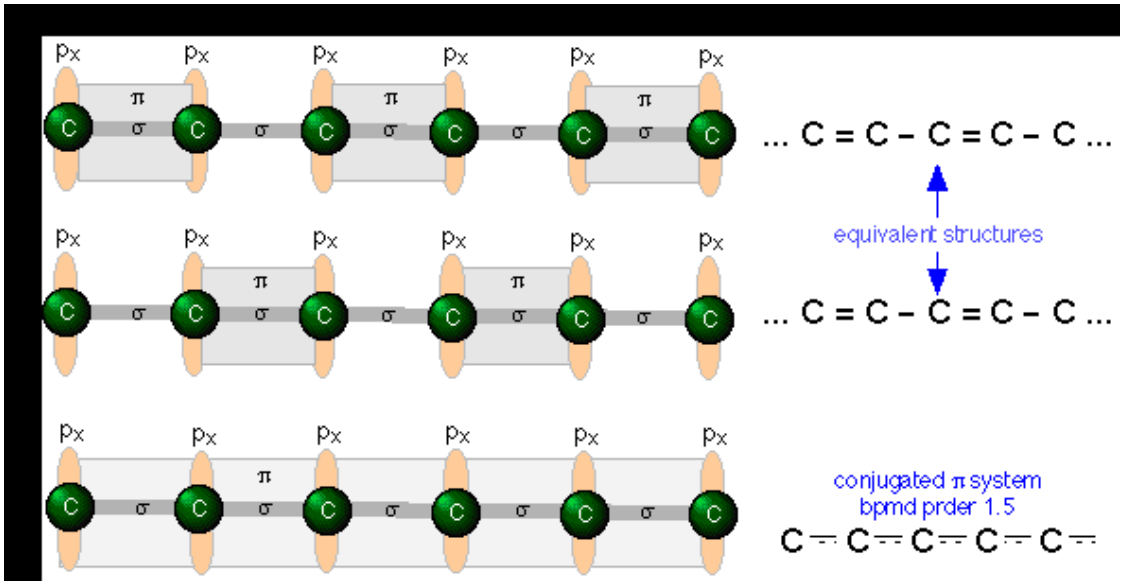


Pi bonds can form in this way between the nitrogen atom and any of the three oxygens; there are thus three equivalent π bonds possible, but since nitrogen can only form one complete π bond at a time, the π bonding is divided up three ways, so that each N–O bond has a bond order of 4/3.



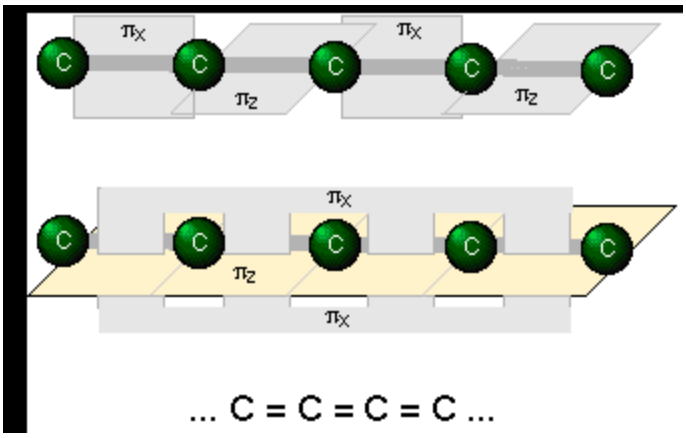
Conjugated Double Bonds

We have seen that the π bonding orbital is distinctly different in shape and symmetry from the σ bond. There is another important feature of the π bond that is of far-reaching consequence, particularly in organic and coordination chemistry. Consider, for example, an extended hydrocarbon molecule in which alternate pairs of carbon atoms are connected by double and single bonds. Each non-terminal carbon atom forms two σ bonds to two other carbons and to a hydrogen (not shown.) This molecule can be viewed as a series of ethylene units joined together end-to-end. Each carbon, being *sp* hybridized, still has a half-filled atomic *p* orbital. Since these *p* orbitals on adjacent carbons are all parallel, we can expect them to interact with each other to form π bonds between alternate pairs of carbon atoms as shown below.



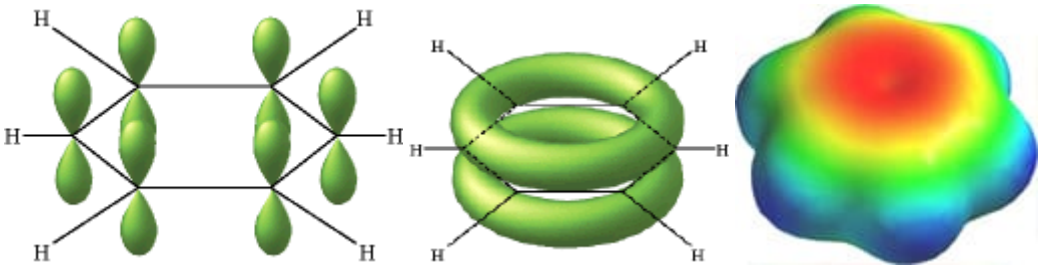
But since each carbon atom possesses a half-filled *p* orbital, there is nothing unique about the π bond arrangement; an equally likely arrangement might be one in which the π bonding orbitals are shifted to neighboring pairs of carbons (middle illustration above). You will recall that when there are two equivalent choices for the arrangements single and double bonds in a molecule, we generally consider the structure to be *aresonance hybrid*. In keeping with this idea, we would expect the electron density in a π system of this kind to be extended or shared out evenly along the entire molecular framework, as shown in the bottom figure.

A system of alternating single and double bonds, as we have here, is called a *conjugated system*. Chemists say that the π bonds in a conjugated system are *delocalized*; they are, in effect, “smeared out” over the entire length of the conjugated part of the molecule. Each pair of adjacent carbon atoms is joined by a σ bond and "half" of a π bond, resulting in an a C-C *bond order* of 1.5. An even higher degree of conjugation exists in compounds containing extended $(C{=}C)_n$ chains. These compounds, known as **cumulenes**, exhibit interesting electrical properties, and whose derivatives can act as "organic wires".



Benzene

The classic example of π bond delocalization is found in the cyclic molecule *benzene* (C_6H_6) which consists of six carbon atoms bound together in a hexagonal ring. Each carbon has a single hydrogen atom attached to it. The lines in this figure represent the σ bonds in benzene. The basic ring structure is composed of σ bonds formed from overlap of sp^2 hybrid orbitals on adjacent carbon atoms. The unhybridized carbon p_z orbitals project above and below the plane of the ring. They are shown here as they might appear if they did not interact with one another.



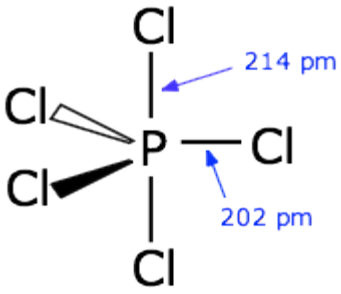
But what happens, of course, is that the lobes of these atomic orbitals meld together to form circular rings of electron density above and below the plane of the molecule. The two of these together constitute the "second half" of the carbon-carbon double bonds in benzene. This computer-generated plot of electron density in the benzene molecule is derived from a more rigorous theory that does not involve hybrid orbitals; the highest electron density (blue) appears around the periphery of the ring, while the lowest (red) is in the "doughnut hole" in the center.

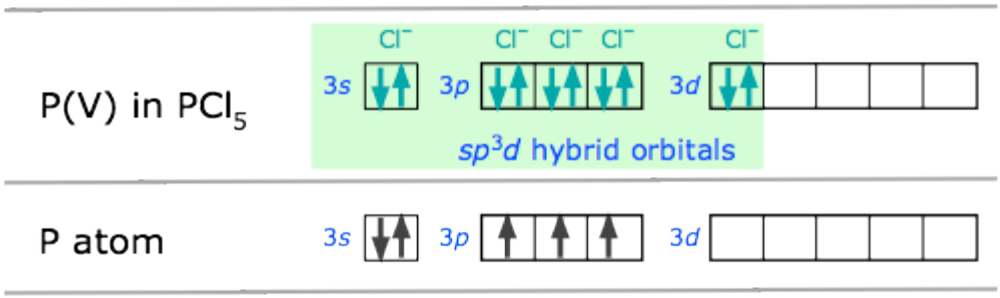
Hybrids involving *d* orbitals

In atoms that are below those in the first complete row of the periodic table, the simple octet rule begins to break down. For example, we have seen that PCl_3 does conform to the octet rule but PCl_5 does not. We can describe the bonding in PCl_3 very much as we do NH_3: four sp^3-hybridized orbitals, three of which are shared with electrons from other atoms and the fourth containing a nonbonding pair.

Pentagonal bipyramid molecules: sp^3d hybridization

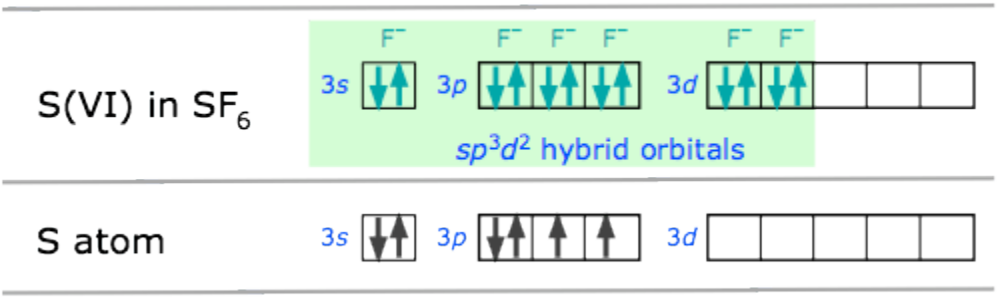
hybrid orbitals directed toward the corners of a trigonal bipyramid, as is predicted by VSEPR theory.





Octahedral coordination: sp^3d^2 hybridization

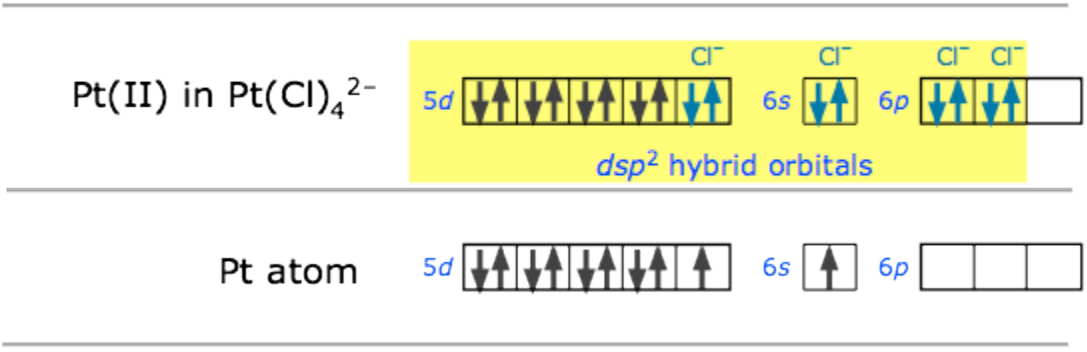
The molecule **sulfur hexafluoride SF_6** exemplifies one of the most common types of *d*-orbital hybridization. The six bonds in this octahedrally-coordinated molecule are derived from mixing six atomic orbitals into a hybrid set. The easiest way to understand how these come about is to imagine that the molecule is made by combining an imaginary S^{6+} ion (which we refer to as the *S(VI) valence state*) with six F^- ions to form the neutral molecule. These now-empty 3*s* and 3*p* orbitals then mix with two 3*d* orbitals to form the sp^3d^2 hybrids.



Some of the most important and commonly encountered compounds which involve the *d*orbitals in bonding are the *transition metal complexes*. The term “complex” in this context means that the molecule is composed of two or more kinds of species, each of which can have an independent existence.

Square-planar molecules: dsp^2 hybridization

For example, the ions Pt^{2+} and Cl^- can form the ion $[PtCl_4]^{2-}$. To understand the hybridization scheme, it helps to start with the neutral Pt atom, then imagine it losing two electrons to become an ion, followed by grouping of the two unpaired 5*d* electrons into a single *d* orbital, leaving one vacant.This vacant orbital, along with the 6s and two of the 6*p* orbitals, can then accept an electron pair from four chlorines.

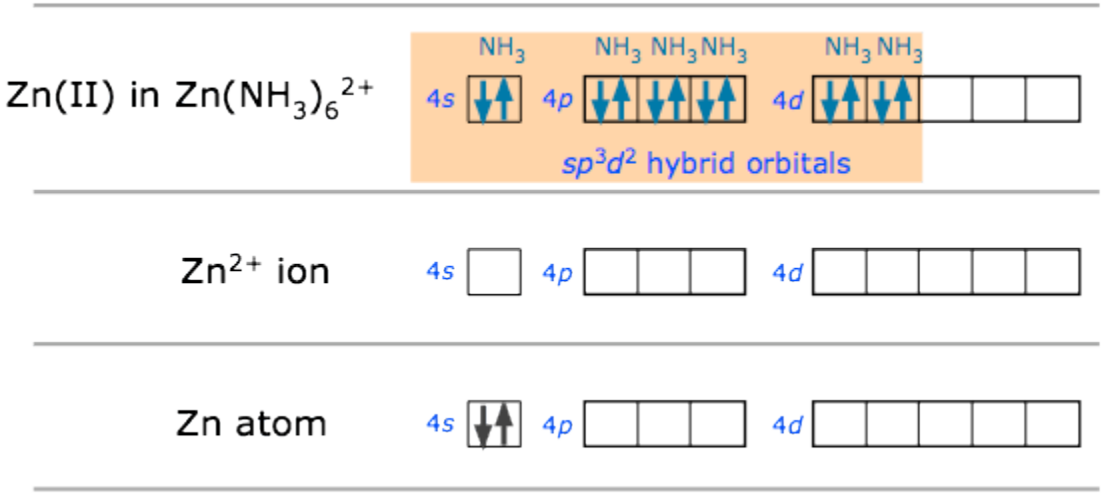


All of the four-coordinated molecules we have discussed so far have tetrahedral geometry around the central atom. Methane, CH_4,

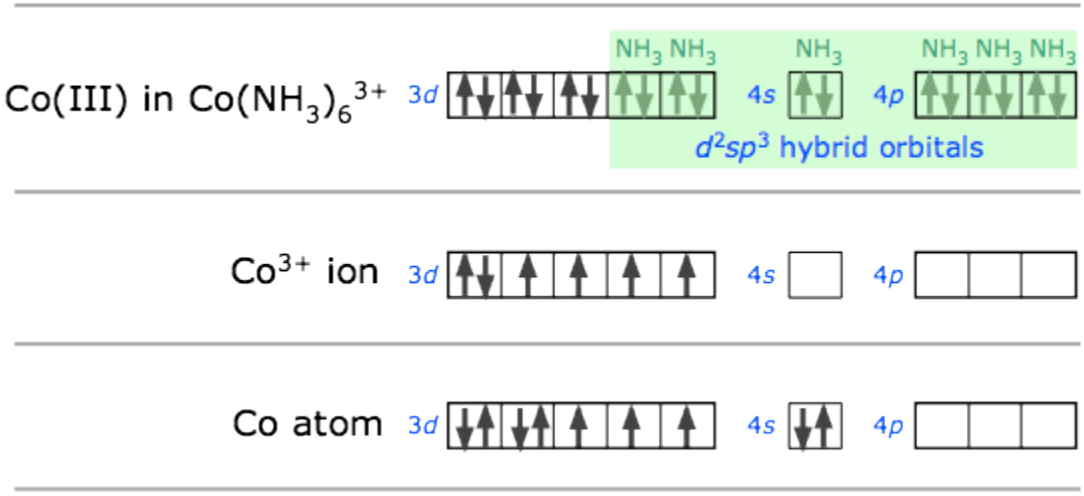
is the most well known example. It may come as something as a surprise, then, to discover that the **tetrachlorplatinum (II) ion** $\mathbf{[PtCl_4]^{2-}}$ has an essentially two-dimensional *square-planar* configuration. This type of bonding pattern is quite common when the parent central ion (Pt^{2+} in this case) contains only eight electrons in its outmost *d*-subshell.

Octahedral coordination: sp^3d^2 and d^2sp^3

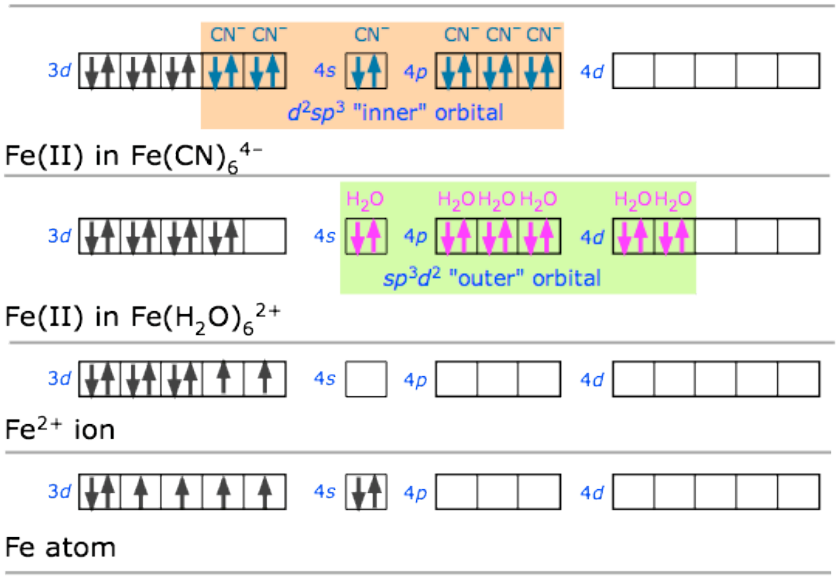
Many of the most commonly encountered transition metal ions accept electron pairs from donors such as CN^- and NH_3 (or lacking these, even from H_2O) to form octahedral coordination complexes. The **hexaminezinc(II)** cation depicted below is typical.



In sp^3d^2 hybridization the bonding orbitals are derived by mixing atomic orbitals having the same principal quantum number (n = 4 in the preceding example). A slightly different arrangement, known as d^2sp^3 hybridization, involves *d* orbitals of lower principal quantum number. This is possible because of the rather small energy differences between the *d* orbitals in one "shell" with the s and *p* orbitals of the next higher one — hence the term "inner orbital" complex which is sometimes used to describe ions such as **hexaminecobalt(III)**, shown below.. Both arrangements produce octahedral coordination geometries.



In some cases, the same central atom can form either inner or outer complexes depending on the particular ligand and the manner in which its electrostatic field affects the relative energies of the different orbitals.Thus the **hexacyanoiron(II)** ion utilizes the iron 3*d* orbitals, whereas **hexaaquoiron(II)** achieves a lower energy by accepting two H_2O molecules in its 4*d* orbitals.



Final remarks about hybrid orbitals

As is the case with any scientific model, the hybridization model of bonding is useful only to the degree to which it can predict phenomena that are actually observed. Most models contain weaknesses that place limits on their general applicability. The need for caution in accepting this particular model is made more apparent when we examine the shapes of the molecules below the first full row of the periodic table. For example, we would expect the bonding in hydrogen sulfide to be similar to that in water, with tetrahedral geometry around the sulfur atom. Experiments, however, reveal that the H–S–H bond angle is only 92°. Hydrogen sulfide thus deviates much more from tetrahedral geometry than does water, and there is no apparent and clear reason why it should. It is certainly difficult to argue that electron-repulsion between the two nonbonding orbitals is pushing the H–S bonds closer together (as is supposed to happen to the H–O bonds in water); many would argue that this repulsion would be less in hydrogen sulfide than in water, since sulfur is a larger atom and is hence less electronegative.

orbitals does not apply to H_2S. It looks like the “simple” explanation that bonding occurs through two half occupied atomic *p* orbitals 90° apart comes closer to the mark. Perhaps hybridization is not an all-or-nothing phenomenon; perhaps the two 3*p* orbitals are substantially intact in hydrogen sulfide, or are hybridized only slightly. In general, the hybridization model does not work very well with nonmetallic elements farther down in the periodic table, and there is as yet no clear explanation why. We must simply admit that we have reached one of the many points in chemistry where our theory is not sufficiently developed to give a clear and unequivocal answer. This does not detract, however, from the wide usefulness of the hybridization model in elucidating the bond character and bond shapes in the millions of molecules based on first-row elements, particularly of carbon.

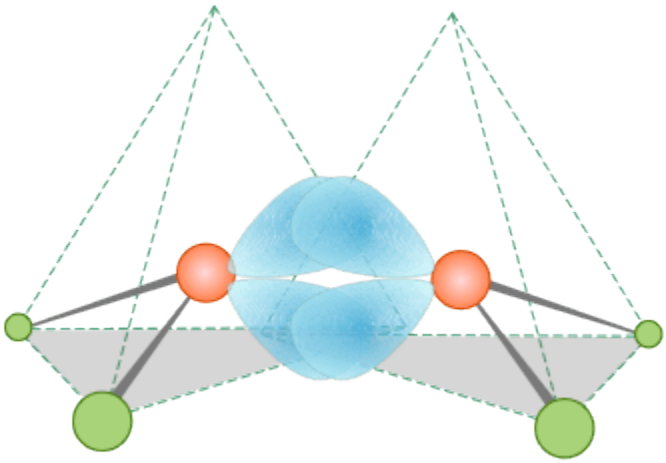
The justification we gave for invoking hybridization in molecules such as BeH_2, BF_3 and CH_4 was that the bonds in each are geometrically and chemically equivalent, whereas the atomic s- and *p*-orbitals on the central atoms are not. By combining these into new orbitals of *sp*, sp^2 and sp^3 types we obtain the required number of completely equivalent orbitals. This seemed easy enough to do on paper; we just drew little boxes and wrote “sp^2” or whatever below them. But what is really going on here?

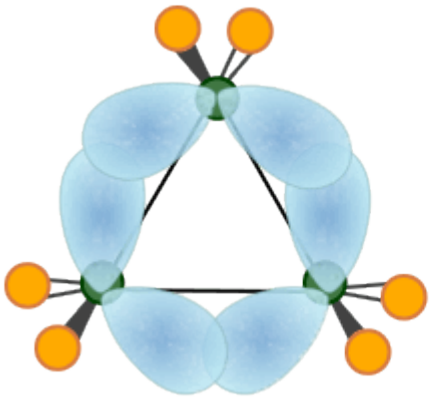
The full answer is beyond the scope of this course, so we can only offer the following very general explanation. First, recall what we mean by “orbital”: a mathematical function ψ having the character of a standing wave whose square ψ^2 is proportional to the probability of finding the electron at any particular location in space. The latter, the *electron density distribution*, can be observed (by X-ray scattering, for example), and in this sense is the only thing that is “real”.

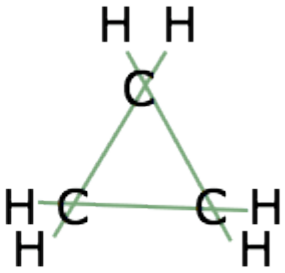
A given standing wave (ψ-function) can be synthesized by combining all kinds of fundamental wave patterns (that is, atomic orbitals) in much the same way that a color we observe can be reproduced by combining different sets of primary colors in various proportions. In neither case does it follow that these original orbitals (or colors) are actually present in the final product. So one could well argue that hybrid orbitals are not “real”; they simply turn out to be convenient for understanding the bonding of simple molecules at the elementary level, and this is why we use them.

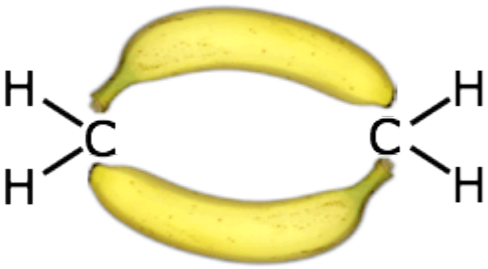
An alternative to hybrids: the Bent-Bond model

It turns out, in fact, that the electron distribution and bonding in ethylene can be equally well described by assuming no hybridization at all. The "bent bond" model requires only that the directions of some of the atomic-*p* orbitals be distorted sufficiently to provide the overlap needed for bonding; these are sometimes referred to as "*banana bonds*".









The smallest of the closed-ring hydrocarbons is cyclopropane, a planar molecule in which the C–C bond angles are 120°— quite a departure from the tetrahedral angle of 109.5° associated with sp^3 hybridization! Theoretical studies suggest that the bent-bond model does quite well in predicting its properties.

This page titled 9.7: The Hybrid Orbital Model II is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.8: Molecular Orbital Theory

Learning Objectives

Make sure you thoroughly understand the following essential ideas

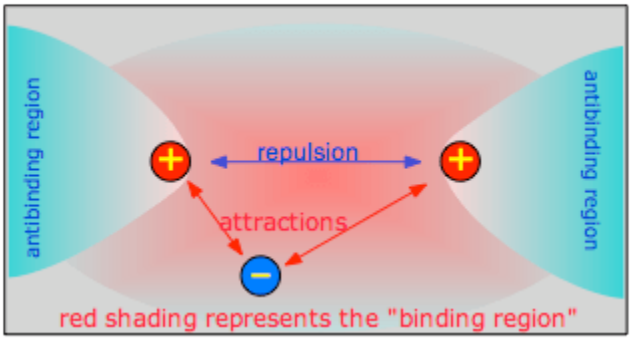
- In what fundamental way does the molecular orbital model differ from the other models of chemical bonding that have been described in these lessons?
- Explain how bonding and antibonding orbitals arise from atomic orbitals, and how they differ physically.
- Describe the essential difference between a *sigma* and a *pi* molecular orbital.
- Define *bond order*, and state its significance.
- Construct a "molecular orbital diagram" of the kind shown in this lesson for a simple diatomic molecule, and indicate whether the molecule or its positive and negative ions should be stable.

The molecular orbital model is by far the most productive of the various models of chemical bonding, and serves as the basis for most quantiative calculations, including those that lead to many of the computer-generated images that you have seen elsewhere in these units. In its full development, molecular orbital theory involves a lot of complicated mathematics, but the fundamental ideas behind it are quite easily understood, and this is all we will try to accomplish in this lesson.

This is a big departure from the simple *Lewis* and *VSEPR* models that were based on the one-center orbitals of individual atoms. The more sophisticated *hybridization* model recognized that these orbitals will be modified by their interaction with other atoms. But all of these *valence-bond* models, as they are generally called, are very limited in their applicability and predictive power, because they fail to recognize that distribution of the pooled valence electrons is governed by the totality of positive centers.

Molecular Orbitals

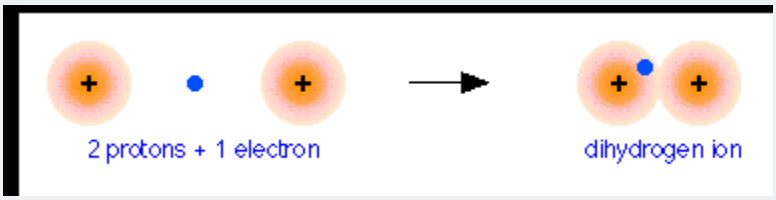
Chemical bonding occurs when the net attractive forces between an electron and two nuclei exceeds the electrostatic repulsion between the two nuclei. For this to happen, the electron must be in a region of space which we call the *binding region*. Conversely, if the electron is off to one side, in an *anti-binding region*, it actually adds to the repulsion between the two nuclei and helps push them away.



The easiest way of visualizing a molecular orbital is to start by picturing two isolated atoms and the electron orbitals that each would have separately. These are just the orbitals of the separate atoms, by themselves, which we already understand. We will then try to predict the manner in which these atomic orbitals interact as we gradually move the two atoms closer together. Finally, we will reach some point where the internuclear distance corresponds to that of the molecule we are studying. The corresponding orbitals will then be the *molecular orbitals* of our new molecule.

The hydrogen molecule ion: the simplest molecule

To see how this works, we will consider the simplest possible molecule, H_2^+. This is the hydrogen molecule ion, which consists of two nuclei of charge +1, and a single electron shared between them.

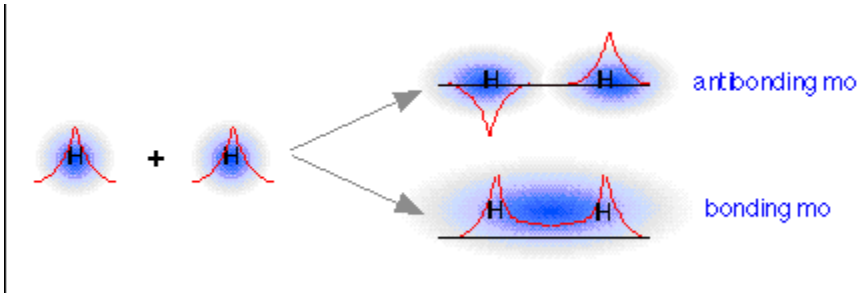


As two H nuclei move toward each other, the 1s atomic orbitals of the isolated atoms gradually merge into a new molecular orbital in which the greatest electron density falls between the two nuclei. Since this is just the location in which electrons can exert the most attractive force on the two nuclei simultaneously, this arrangement constitutes a *bonding molecular orbital*. Regarding it as a three- dimensional region of space, we see that it is symmetrical about the line of centers between the nuclei; in accord with our usual nomenclature, we refer to this as a *σ (sigma) orbital*.

Bonding and Antibonding Molecular Orbitals

There is one minor difficulty: we started with two orbitals (the 1s atomic orbitals), and ended up with only one orbital. Now according to the rules of quantum mechanics, orbitals cannot simply appear and disappear at our convenience. For one thing, this would raise the question of at just what internuclear distance do we suddenly change from having two orbitals, to having only one? It turns out that when orbitals interact, they are free to change their forms, but there must always be the same number. This is just another way of saying that there must always be the same number of possible allowed sets of electron quantum numbers.

How can we find the missing orbital? To answer this question, we must go back to the wave-like character of orbitals that we developed in our earlier treatment of the hydrogen atom. You are probably aware that wave phenomena such as sound waves, light waves, or even ocean waves can combine or interact with one another in two ways: they can either reinforce each other, resulting in a stronger wave, or they can interfere with and partially destroy each other. A roughly similar thing occurs when the "matter waves" corresponding to the two separate hydrogen 1s orbitals interact; both in-phase and out-of-phase combinations are possible, and both occur. The in-phase, reinforcing interaction yields the *bonding orbital* that we just considered. The other, corresponding to out-of-phase combination of the two orbitals, gives rise to a molecular orbital that has its greatest electron probability in what is clearly the antibonding region of space. This second orbital is therefore called an *antibonding* orbital.



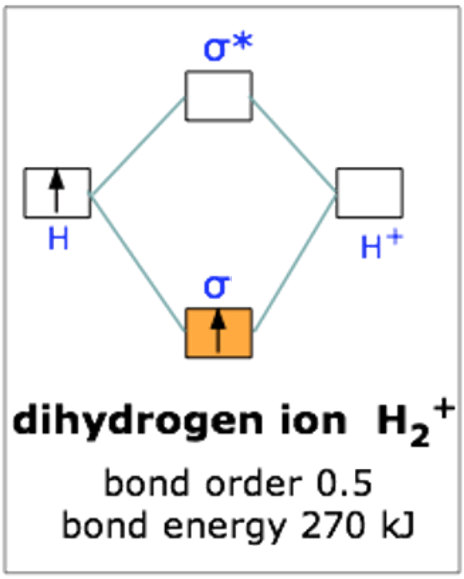
When the two 1s wave functions combine out-of-phase, the regions of high electron probability do not merge. In fact, the orbitals act as if they actually repel each other. Notice particularly that there is a region of space exactly equidistant between the nuclei at which the probability of finding the electron is zero. This region is called a *nodal surface*, and is characteristic of antibonding orbitals. It should be clear that any electrons that find themselves in an antibonding orbital cannot possibly contribute to bond formation; in fact, they will actively oppose it.

We see, then, that whenever two orbitals, originally on separate atoms, begin to interact as we push the two nuclei toward each other, these two atomic orbitals will gradually merge into a pair of molecular orbitals, one of which will have bonding character, while the other will be antibonding. In a more advanced treatment, it would be fairly easy to show that this result follows quite naturally from the wave-like nature of the combining orbitals.

What is the difference between these two kinds of orbitals, as far as their potential energies are concerned? More precisely, which kind of orbital would enable an electron to be at a lower potential energy? Clearly, the potential energy decreases as the electron moves into a region that enables it to “see” the maximum amount of positive charge. In a simple diatomic molecule, this will be in the internuclear region— where the electron can be simultaneously close to two nuclei. The bonding orbital will therefore have the lower potential energy.

Molecular Orbital Diagrams

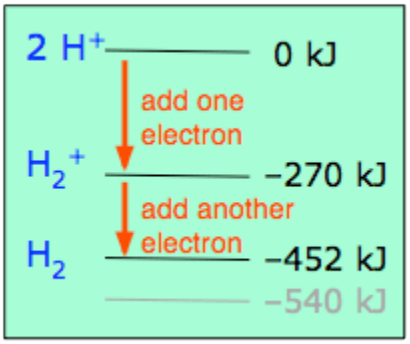
This scheme of bonding and antibonding orbitals is usually depicted by a molecular orbital diagram such as the one shown here for the **dihydrogen ion H_2^+**. Atomic valence electrons (shown in boxes on the left and right) fill the lower-energy molecular orbitals before the higher ones, just as is the case for atomic orbitals. Thus, the single electron in this simplest of all molecules goes into the bonding orbital, leaving the antibonding orbital empty.



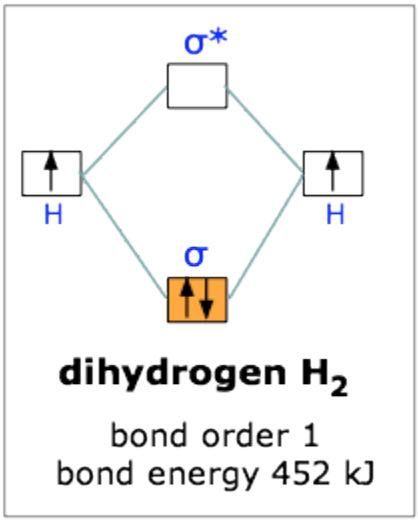
Since any orbital can hold a maximum of two electrons, the bonding orbital in $H_2^+$is only half-full. This single electron is nevertheless enough to lower the potential energy of one mole of hydrogen nuclei pairs by 270 kJ— quite enough to make them stick together and behave like a distinct molecular species. Although H_2^+ is stable in this energetic sense, it happens to be an extremely reactive molecule— so much so that it even reacts with itself, so these ions are not commonly encountered in everyday chemistry.

Dihydrogen

If one electron in the bonding orbital is conducive to bond formation, might two electrons be even better? We can arrange this by combining two hydrogen atoms-- two nuclei, and two electrons. Both electrons will enter the bonding orbital, as depicted in the Figure.



We recall that one electron lowered the potential energy of the two nuclei by 270 kJ/mole, so we might expect two electrons to produce twice this much stabilization, or 540 kJ/mole.

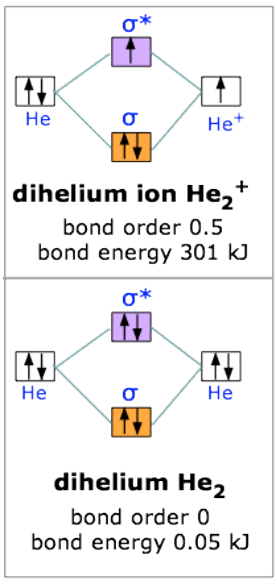


Bond order is defined as the difference between the number of electron pairs occupying bonding and nonbonding orbitals in the molecule. A bond order of unity corresponds to a conventional "single bond".

Experimentally, one finds that it takes only 452 kJ to break apart a mole of hydrogen molecules. The reason the potential energy was not lowered by the full amount is that the presence of two electrons in the same orbital gives rise to a repulsion that acts against the stabilization. This is exactly the same effect we saw in comparing the ionization energies of the hydrogen and helium atoms.

Dihelium

With two electrons we are still ahead, so let's try for three. The dihelium positive ion is a three-electron molecule. We can think of it as containing two helium nuclei and three electrons. This molecule is stable, but not as stable as dihydrogen; the energy required to break He_2+ is 301 kJ/mole. The reason for this should be obvious; two electrons were accommodated in the bonding orbital, but the third electron must go into the next higher slot— which turns out to be the sigma antibonding orbital. The presence of an electron in this orbital, as we have seen, gives rise to a repulsive component which acts against, and partially cancels out, the attractive effect of the filled bonding orbital.



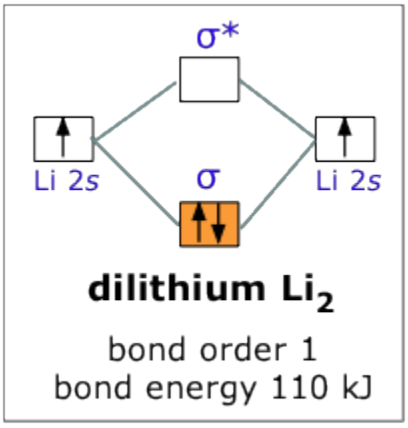
Taking our building-up process one step further, we can look at the possibilities of combining to helium atoms to form dihelium. You should now be able to predict that He_2 cannot be a stable molecule; the reason, of course, is that we now have four electrons—two in the bonding orbital, and two in the antibonding orbital. The one orbital almost exactly cancels out the effect of the other. Experimentally, the bond energy of dihelium is only .084 kJ/mol; this is not enough to hold the two atoms together in the presence of random thermal motion at ordinary temperatures, so dihelium dissociates as quickly as it is formed, and is therefore not a distinct chemical species.

Diatomic molecules containing second-row atoms

The four simplest molecules we have examined so far involve molecular orbitals that derived from two 1s atomic orbitals. If we wish to extend our model to larger atoms, we will have to contend with higher atomic orbitals as well. One greatly simplifying principle here is that only the *valence-shell* orbitals need to be considered. Inner atomic orbitals such as 1s are deep within the atom and well-shielded from the electric field of a neighboring nucleus, so that these orbitals largely retain their atomic character when bonds are formed.

Dilithium

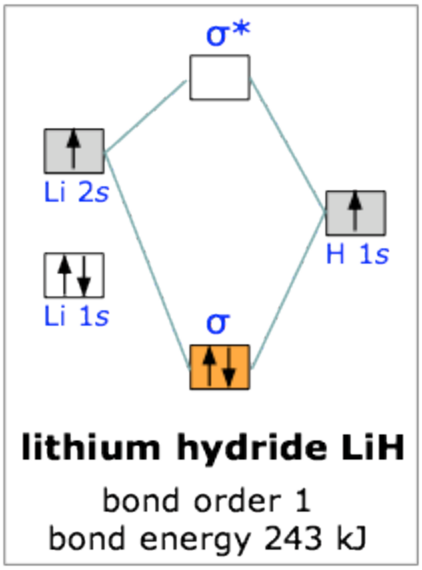
For example, when lithium, whose configuration is $1s^2 2s^1$, bonds with itself to form Li_2, we can forget about the 1s atomic orbitals and consider only the σ bonding and antibonding orbitals. Since there are not enough electrons to populate the antibonding orbital, the attractive forces win out and we have a stable molecule.



The bond energy of dilithium is 110 kJ/mole; notice that this value is less than half of the 270 kJ bond energy in dihydrogen, which also has two electrons in a bonding orbital. The reason, of course, is that the 2*s* orbital of Li is much farther from its nucleus than is the 1s orbital of H, and this is equally true for the corresponding molecular orbitals. It is a general rule, then, that the larger the parent atom, the less stable will be the corresponding diatomic molecule.

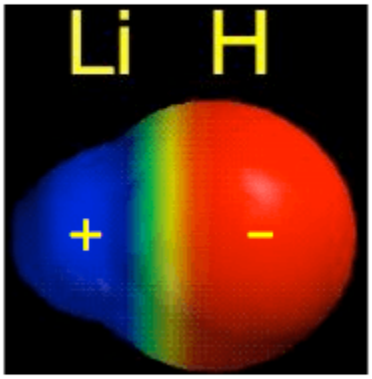
Lithium hydride

All the molecules we have considered thus far are *homonuclear*; they are made up of one kind of atom. As an example of a *heteronuclear* molecule, let's take a look at a very simple example— lithium hydride. Lithium hydride is a stable, though highly reactive molecule. The diagram shows how the molecular orbitals in lithium hydride can be related to the atomic orbitals of the parent atoms. One thing that makes this diagram look different from the ones we have seen previously is that the parent atomic orbitals have widely differing energies; the greater nuclear charge of lithium reduces the energy of its 1s orbital to a value well below that of the 1s hydrogen orbital.



There are two occupied atomic orbitals on the lithium atom, and only one on the hydrogen. With which of the lithium orbitals does the hydrogen 1s orbital interact? The lithium 1s orbital is the lowest-energy orbital on the diagram. Because this orbital is so small and retains its electrons so tightly, it does not contribute to bonding; we need consider only the 2s orbital of lithium which combines with the 1s orbital of hydrogen to form the usual pair of sigma bonding and antibonding orbitals. Of the four electrons in lithium and hydrogen, two are retained in the lithium 1s orbital, and the two remaining ones reside in the σ orbital that constitutes the Li–H covalent bond.

The resulting molecule is 243 kJ/mole more stable than the parent atoms. As we might expect, the bond energy of the heteronuclear molecule is very close to the average of the energies of the corresponding homonuclear molecules. Actually, it turns out that the correct way to make this comparison is to take the geometric mean, rather than the arithmetic mean, of the two bond energies. The geometric mean is simply the square root of the product of the two energies.



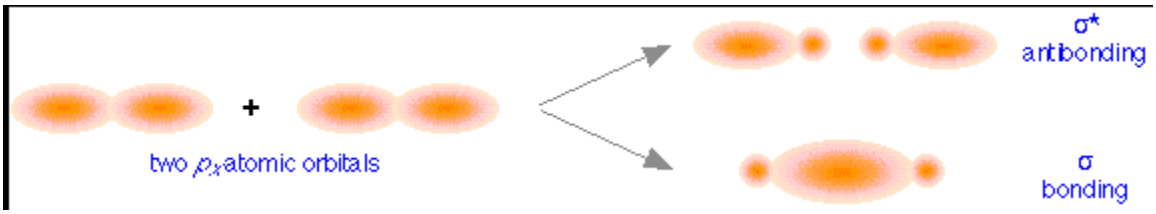
The geometric mean of the H_2 and Li_2 bond energies is 213 kJ/mole, so it appears that the lithium hydride molecule is 30 kJ/mole more stable than it "is supposed" to be. This is attributed to the fact that the electrons in the 2σ bonding orbital are not equally shared between the two nuclei; the orbital is skewed slightly so that the electrons are attracted somewhat more to the hydrogen atom. This *bond polarity*, which we considered in some detail near the beginning of our study of covalent bonding, arises from the greater electron-attracting power of hydrogen— a consequence of the very small size of this atom. The electrons can be at a lower potential energy if they are slightly closer to the hydrogen end of the lithium hydride molecule. It is worth pointing out, however, that the electrons are, on the average, also closer to the lithium nucleus, compared to where they would be in the 2s orbital of the isolated lithium atom. So it appears that everyone gains and no one loses here!

Σ and π orbitals

The molecules we have considered thus far are composed of atoms that have no more than four electrons each; our molecular

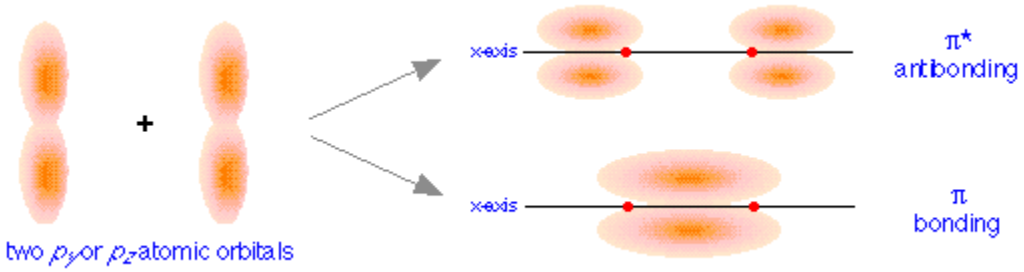
orbitals have therefore been derived from s-type atomic orbitals only. If we wish to apply our model to molecules involving larger atoms, we must take a close look at the way in which *p*-type orbitals interact as well. Although two atomic *p* orbitals will be expected to split into bonding and antibonding orbitals just as before, it turns out that the extent of this splitting, and thus the relative energies of the resulting molecular orbitals, depend very much on the nature of the particular *p* orbital that is involved.

You will recall that there are three possible *p* orbitals for any value of the principal quantum number. You should also recall that *p* orbitals are not spherical like s orbitals, but are elongated, and thus possess definite directional properties. The three *p* orbitals correspond to the three directions of Cartesian space, and are frequently designated p_x, p_y, and p_z, to indicate the axis along which the orbital is aligned. Of course, in the free atom, where no coordinate system is defined, all directions are equivalent, and so are the *p* orbitals. But when the atom is near another atom, the electric field due to that other atom acts as a point of reference that defines a set of directions. The line of centers between the two nuclei is conventionally taken as the *x* axis. If this direction is represented horizontally on a sheet of paper, then the *y* axis is in the vertical direction and the *z* axis would be normal to the page.



These directional differences lead to the formation of two different classes of molecular orbitals. The above figure shows how two p_x atomic orbitals interact. In many ways the resulting molecular orbitals are similar to what we got when s atomic orbitals combined; the bonding orbital has a large electron density in the region between the two nuclei, and thus corresponds to the lower potential energy. In the out-of-phase combination, most of the electron density is away from the internuclear region, and as before, there is a surface exactly halfway between the nuclei that corresponds to zero electron density. This is clearly an *antibonding orbital*— again, in general shape, very much like the kind we saw in hydrogen and similar molecules. Like the ones derived from s-atomic orbitals, these molecular orbitals are σ (*sigma*) orbitals.

Sigma orbitals are cylindrically symmetric with respect to the line of centers of the nuclei; this means that if you could look down this line of centers, the electron density would be the same in all directions.



orbitals, we get the bonding and antibonding pairs that we would expect, but the resulting molecular orbitals have a different symmetry: rather than being rotationally symmetric about the line of centers, these orbitals extend in both perpendicular directions from this line of centers. Orbitals having this more complicated symmetry are called π (*pi*) orbitals. There are two of them, π_y and π_z differing only in orientation, but otherwise completely equivalent.

The different geometric properties of the π and σ orbitals causes the latter orbitals to split more than the π orbitals, so that the σ* antibonding orbital always has the highest energy. The σ bonding orbital can be either higher or lower than the π bonding orbitals, depending on the particular atom.

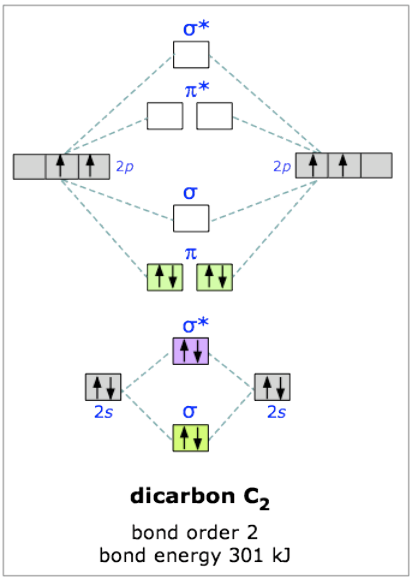
Second-Row Diatomics

If we combine the splitting schemes for the 2*s* and 2*p* orbitals, we can predict bond order in all of the diatomic molecules and ions composed of elements in the first complete row of the periodic table. Remember that only the *valence orbitals* of the atoms

need be considered; as we saw in the cases of lithium hydride and dilithium, the inner orbitals remain tightly bound and retain their localized atomic character.

Dicarbon

Carbon has four outer-shell electrons, two 2*s* and two 2*p*. For two carbon atoms, we therefore have a total of eight electrons, which can be accommodated in the first four molecular orbitals. The lowest two are the 2*s*-derived bonding and antibonding pair, so the "first" four electrons make no net contribution to bonding. The other four electrons go into the pair of *pi*bonding orbitals, and there are no more electrons for the antibonding orbitals— so we would expect the dicarbon molecule to be stable, and it is. (But being extremely reactive, it is known only in the gas phase.)

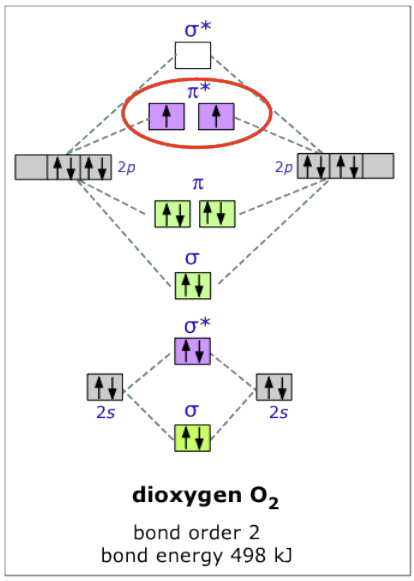


You will recall that one pair of electrons shared between two atoms constitutes a "single" chemical bond; this is Lewis' original definition of the covalent bond. In C_2 there are two paris of electrons in the π bonding orbitals, so we have what amounts to a double bond here; in other words, the bond order in dicarbon is two.

Dioxygen

The electron configuration of oxygen is $1s^2 2s^2 2p^4$. In O_2, therefore, we need to accommodate twelve valence electrons (six from each oxygen atom) in molecular orbitals. As you can see from the diagram, this places two electrons in antibonding orbitals. Each of these electrons occupies a separate π* orbital because this leads to less electron-electron repulsion (Hund's Rule).

The bond energy of molecular oxygen is 498 kJ/mole. This is smaller than the 945 kJ bond energy of N_2— not surprising, considering that oxygen has two electrons in an antibonding orbital, compared to nitrogen's one.



The two unpaired electrons of the dioxygen molecule give this substance an unusual and distinctive property: O_2 is *paramagnetic*. The paramagnetism of oxygen can readily be demonstrated by pouring liquid O_2 between the poles of a strong permanent magnet; the liquid stream is trapped by the field and fills up the space between the poles.

Since molecular oxygen contains two electrons in an antibonding orbital, it might be possible to make the molecule more stable by removing one of these electrons, thus increasing the ratio of bonding to antibonding electrons in the molecule. Just as we would expect, and in accord with our model, O_2^+ has a bond energy higher than that of neutral dioxygen; removing the one electron actually gives us a more stable molecule. This constitutes a very good test of our model of bonding and antibonding orbitals. In the same way, adding an electron to O_2 results in a weakening of the bond, as evidenced by the lower bond energy of O_2^-. The bond energy in this ion is not known, but the length of the bond is greater, and this is indicative of a lower bond energy. These two dioxygen ions, by the way, are highly reactive and can be observed only in the gas phase.

This page titled 9.8: Molecular Orbital Theory is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.9: Bonding in Coordination Complexes

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Define the terms *coordination complex, ligand, polydentate*, and *chelate*.
- Explain the origins of *d*-orbital splitting; that is, why the energies of certain atomic-*d* orbitals are more affected by some ligands than others in an octahedral complex.
- Why are many coordination complexes highly *colored*?
- Explain the meaning of high-spin and low-spin complexes, and illustrate in a general way how a particular set of ligands can change one kind into another. Also, describe how these differences are observed experimentally.
- Describe the role of iron in heme and the general structural components of hemoglobin.

Complexes such as $Cu(NH_3)_6^{2+}$ have been known and studied since the mid-nineteenth century. and their structures had been mostly worked out by 1900. Although the hybrid orbital model was able to explain how neutral molecules such as water or ammonia could bond to a transition metal ion, it failed to explain many of the special properties of these complexes. Finally, in 1940-60, a model known as *ligand field theory* was developed that is able to organize and explain most of the observed properties of these compounds. Since that time, coordination complexes have played major roles in cellular biochemistry and inorganic catalysis.

What is a Complex?

If you have taken a lab course in chemistry, you have very likely admired the deep blue color of **copper sulfate** crystals, $CuSO_4{\cdot}5H_2O$. The proper name of this substance is copper(II) sulfate pentahydrate, and it is typical of many salts that incorporate *waters of hydration* into their crystal structures. It is also a *complex*, a term used by chemists to describe a substance composed of two other substances (in this case, $CuSO_4$ and H_2O) each of which is capable of an independent existence.



Figure 9.9.1: (left) Crystals of $\mathrm{CuSO_4 \cdot 5\,H_2O}$. from Stephanb (via Wikipedia). (right) Anhydrous $\mathrm{CuSO_4}$ after the water was removed.

The binding between the components of a complex is usually weaker than a regular chemical bond; thus most solid hydrates can be decomposed by heating, driving off the water and yielding the *anhydrous* salt:

$$\underbrace{\mathrm{CuSO_4 \cdot 5\,H_2O}}_{\text{blue}} \rightarrow \underbrace{\blacksquare}_{\text{white}} + 5H_2O \tag{9.9.1}$$

Driving off the water in this way also destroys the color, turning it from a beautiful deep blue to a nondescript white. If the anhydrous salt is now dissolved in water, the blue color now pervades the entire solution. It is apparent that the presence of water is somehow necessary for the copper(II) ion to take on a blue color, but why should this be?



Figure 9.9.2: Difference between anhydrous and hydrated copper sulfate after a few drops of water was added to anhydrous $CuSO_4$.

A very common lab experiment that most students carry out is to add some dilute ammonia to a copper sulfate solution. At first, the solution turns milky as the alkaline ammonia causes the precipitation of copper hydroxide:

$$\mathrm{Cu^{2+} + 2\,OH^-} \rightarrow \mathrm{Cu(OH)_2\,(s)} \tag{9.9.2}$$

However, if more ammonia is added, the cloudiness disappears and the solution assumes an intense deep blue color that makes the original solution seem pale by comparison. The equation for this reaction is usually given as

$$\mathrm{Cu^{2+} + 6\,NH_3} \rightarrow \mathrm{Cu(NH_3)_6{}^{2+}} \tag{9.9.3}$$

The new product is commonly known as the *copper-ammonia complex ion*, or more officially, hexamminecopper(II) complex ion.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Equation 9.9.3 is somewhat misleading, however, in that it implies the formation of a new complex where none existed before. In fact, since about 1895, it has been known that the ions of most transition metals dissolve in water to form complexes with water itself, so a better representation of the reaction of dissolved copper with ammonia would be

$$\mathrm{Cu(H_2O)_6{}^{2+} + 6\,NH_3} \rightarrow \mathrm{Cu(NH_3)_6{}^{2+} + 6\,H_2O} \tag{9.9.4}$$

In effect, the ammonia binds more tightly to the copper ion than does water, and it thus displaces the latter when it comes into contact with the hexaaquocopper(II) ion, as the dissolved form of Cu^{2+} is properly known.

Most transition metals dissolve in water to form complexes with water itself.

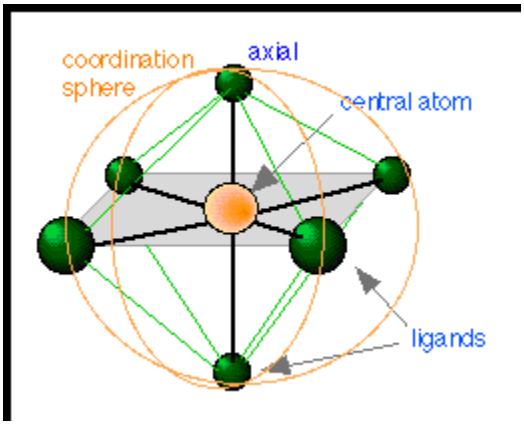
The basics of Coordination Complexes

Although our primary focus in this unit is on bonding, the topic of coordination complexes is so important in chemistry and biochemistry that some of their basic features are worth knowing about, even if their detailed chemistry is beyond the scope

of this course. These complexes play an especially crucial role in physiology and biochemistry. Thus heme, the oxygen-carrying component of red blood cells (and the source of the red color) is basically a complex of iron, and the part of chlorophyll that converts sunlight into chemical energy within green plants is a magnesium complex.

Some Definitions

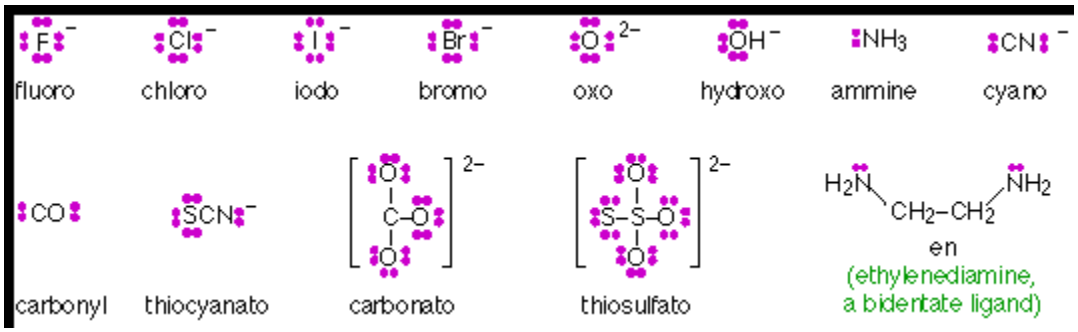
We have already defined a *complex* as a substance composed of two or more components capable of an independent existence. A *coordination complex* is one in which a *central atom* or ion is joined to one or more *ligands* (Latin *ligare*, to tie) through what is called a *coordinate covalent bond* in which both of the bonding electrons are supplied by the ligand. In such a complex the central atom acts as an electron-pair acceptor (*Lewis acid* — think of H^+ which has no electrons at all, but can accept a pair from something like Cl^-) and the ligand as an electron-pair donor (*Lewis base*). The central atom and the ligands coordinated to it constitute the *coordination sphere*. Thus the salt $[Co(NH_3)_5Cl]Cl_2$ is composed of the complex ion $[Co(NH_3)_5Cl]^{2+}$ and two Cl^- ions; components within the square brackets are inside the coordination sphere, whereas the two chloride ions are situated outside the coordination sphere. These latter two ions could be replaced by other ions such as NO_3^- without otherwise materially changing the nature of the salt.



The central atoms of coordination complexes are most often *cations* (positive ions), but may in some cases be neutral atoms, as in nickel carbonyl $Ni(CO)_4$.

Ligands composed of ions such as F^- or small molecules such as H_2O or CN^- possess more than one set of lone pair electrons, but only one of these pairs can coordinate with a central ion. Such ligands are said to be *monodentate* ("one tooth".) Larger ligands may contain more than one atom capable of coordinating with a single central ion, and are described as *polydentate*. Thus ethylenediamine (shown below) is a *bidentate* ligand. Polydentate ligands whose geometry enables them to occupy more than one coordinating position of a central ion act as *chelating agents* (Greek χελος, *chelos*, claw) and tend to form extremely stable complexes known as *chelates*.

Chelation is widely employed in medicine, water-treatment, analytical chemistry and industry for binding and removing metal ions of particular kinds. Some of the more common ligands (chelating agents) are shown here:

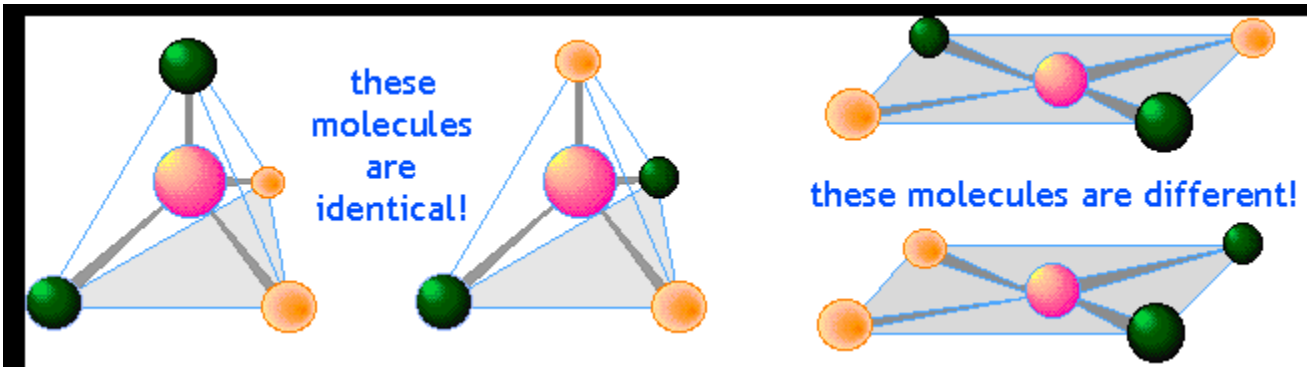


Structure and bonding in transition metal complexes

Complexes such as $Cu(NH_3)_6^{2+}$ have been known and studied since the mid-nineteenth century. Why they should form, or what their structures might be, were complete mysteries. At that time all inorganic compounds were thought to be held together by ionic charges, but ligands such as water or ammonia are of course electrically neutral. A variety of theories such as the existence of "secondary valences" were concocted, and various chain-like structures such as $CuNH_3\text{-}NH_3\text{-}NH_3\text{-}NH_3\text{-}NH_3\text{-}NH_3$ were proposed. Finally, in the mid-1890s, after a series of painstaking experiments, the chemist Alfred Werner (Swiss, 1866-1919) presented the first workable theory of complex ion structures.

Werner claimed that his theory first came to him in a flash after a night of fitful sleep; by the end of the next day he had written his landmark paper that eventually won him the 1913 Nobel Prize in Chemistry.

Werner was able to show, in spite of considerable opposition, that transition metal complexes consist of a central ion surrounded by ligands in a square-planar, tetrahedral, or octahedral arrangement. This was an especially impressive accomplishment at a time long before X-ray diffraction and other methods had become available to observe structures directly. His basic method was to make inferences of the structures from a careful examination of the chemistry of these complexes and particularly the existence of structural isomers. For example, the existence of two different compounds AX_4 having the same composition shows that its structure must be square-planar rather than tetrahedral.



What holds them together?

An understanding of the nature of the bond between the central ion and its ligands would have to await the development of Lewis' *shared-electron pair theory* and Pauling's *valence-bond* picture. We have already shown how *hybridization* of the *d* orbitals of the central ion creates vacancies able to accommodate one or more pairs of unshared electrons on the ligands. Although these models correctly predict the structures of many transition metal complexes, they are by themselves unable to account for several of their special properties:

- The metal-to-ligand bonds are generally much weaker than ordinary covalent bonds;
- Some complexes utilize "inner" *d* orbitals of the central ion, while others are "outer-orbital" complexes;
- Transition metal ions tend to be intensely colored.

Paramagnetism of coordination complexes

Unpaired electrons act as tiny magnets; if a substance that contains unpaired electrons is placed near an external magnet, it will undergo an attraction that tends to draw it into the field. Such substances are said to be *paramagnetic*, and the degree of paramagnetism is directly proportional to the number of unpaired electrons in the molecule. Magnetic studies have played an especially prominent role in determining how electrons are distributed among the various orbitals in transition metal complexes.

Studies of this kind are carried out by placing a sample consisting of a solution of the complex between the poles of an electromagnet. The sample is suspended from the arm of a sensitive balance, and the change in apparent weight is measured with the magnet turned on and off. An increase in the weight when the magnet is turned on indicates that the sample is attracted to the magnet (*paramagnetism*) and must therefore possess one or more unpaired electrons. The precise number can be determined by calibrating the system with a substance whose electron configuration is known.

Crystal field theory

The current model of bonding in coordination complexes developed gradually between 1930-1950. In its initial stages, the model was a purely electrostatic one known as *crystal field theory* which treats the ligand ions as simple point charges that interact with the five atomic *d* orbitals of the central ion. It is this theory which we describe below.

It is remarkable that this rather primitive model, quite innocent of quantum mechanics, has worked so well. However, an improved and more complete model that incorporates molecular orbital theory is known as *ligand field theory*. In an isolated transition metal atom the five outermost *d* orbitals all have the same energy which depends solely on the spherically symmetric electric field due to the nuclear charge and the other electrons of the atom. Suppose now that this atom is made into a *cation* and is placed in solution, where it forms a hydrated species in which six H_2O molecules are coordinated to the central ion in an *octahedral*arrangement. An example of such an ion might be **hexaaquotitanium(III), $Ti(H_2O)_6^{3+}$**.

The ligands (H_2O in this example) are bound to the central ion by electron pairs contributed by each ligand. Because the six ligands are located at the corners of an octahedron centered around the metal ion, these electron pairs are equivalent to clouds of negative charge that are directed from near the central ion out toward the corners of the octahedron. We will call this an octahedral electric field, or the *ligand field*.

d-orbital splitting

The differing shapes of the five kinds of d orbitals cause them to interact differently with the electric fields created by the coordinated ligands. This diagram (from a Purdue U. chemistry site) shows outlines of five kinds of d orbitals.

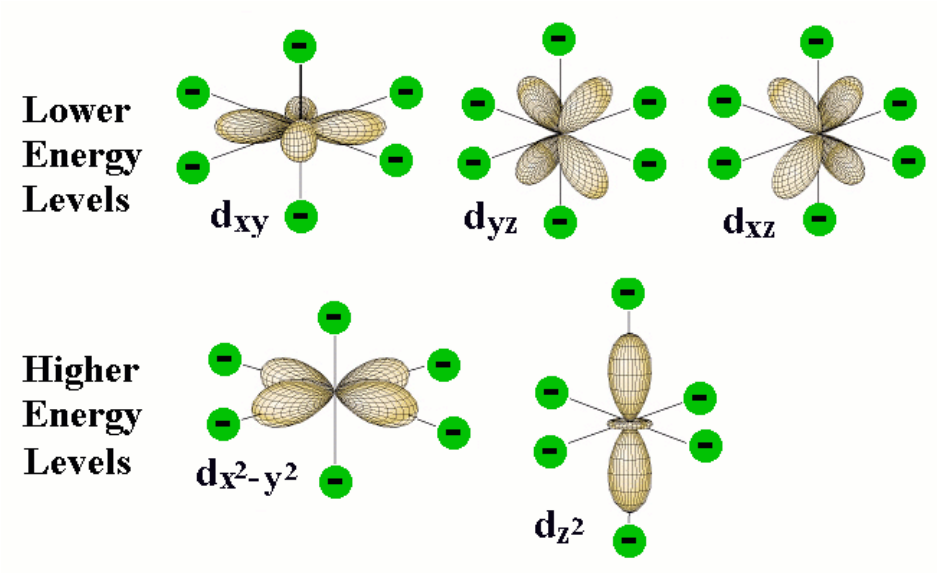


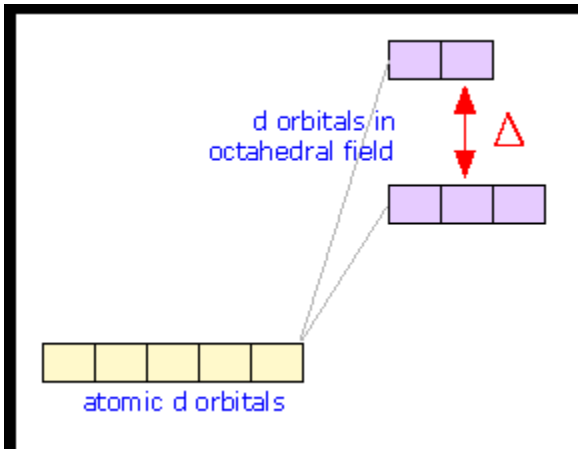
Figure 9.9.3:

The green circles represent the coordinating electron-pairs of the ligands located at the six corners of the octahedron around the central atom. The two d orbitals at the bottom have regions of high electron density pointing directly toward the ligand orbitals; the resulting electron-electron repulsion raises the energy of these *d* orbitals.

Although the five *d* orbitals of the central atom all have the same energy in a spherically symmetric field, their energies will not all be the same in the octahedral field imposed by the presence of the ligands. The reason for this is apparent when we consider the different geometrical properties of the five d orbitals. Two of the *d* orbitals, designated d_{x2} and d_{x2-y2}, have their electron clouds pointing directly toward ligand atoms. We would expect that any electrons that occupy these orbitals would be subject to repulsion by the electron pairs that bind the ligands that are situated at corresponding corners of the octahedron. As a consequence, the energies of these two *d* orbitals will be raised in relation to the three other *d* orbitals whose lobes are not directed toward the octahedral positions.

d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd
La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg
3	4	5	6	7	8	9	10	11	12

The number of electrons in the *d* orbital of the central atom is easily determined from the location of the element in the periodic table, taking in account, of course, of the number of electrons removed in order to form the positive ion.



The effect of the octahedral ligand field due to the ligand electron pairs is to split the *d* orbitals into two sets whose energies differ by a quantity denoted by Δ ("delta") which is known as the *d orbital splitting energy*. Note that both sets of central-ion *d* orbitals are repelled by the ligands and are both raised in energy; the upper set is simply raised by a greater amount. Both the total energy shift and Δ are strongly dependent on the particular ligands.

Why are transition metal complexes often highly colored?

Returning to our example of $Ti(H_2O)_6^{3+}$, we note that Ti has an outer configuration of $4s^2 3d^2$, so that Ti^{3+} will be a d^1 ion. This means that in its ground state, one electron will occupy the lower group of *d* orbitals, and the upper group will be empty. The *d*-orbital splitting in this case is 240 kJ per mole which corresponds to light of blue-green color; absorption of this light promotes the electron to the upper set of *d* orbitals, which represents the *exited state* of the complex. If we illuminate a solution of $Ti(H_2O)_6^{3+}$ with white light, the blue-green light is absorbed and the solution appears violet in color.

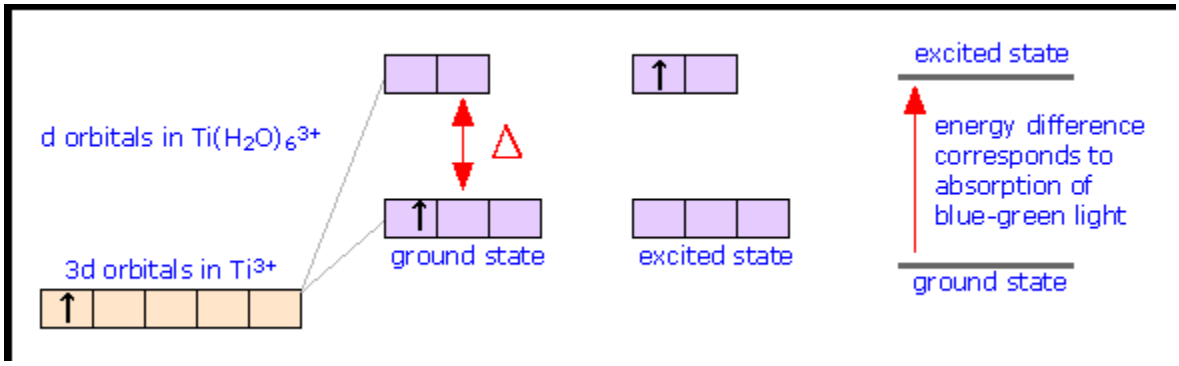


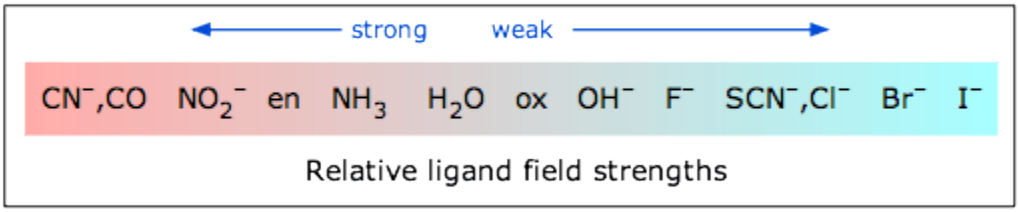
Figure 9.9.4 shows how the colors of hexaamminenickel II complexes are affected by different ligands.



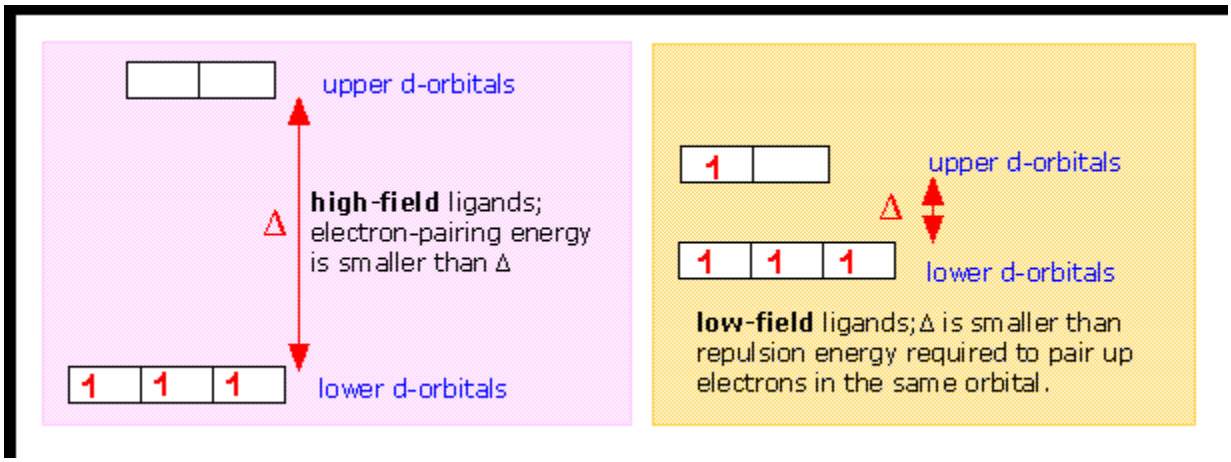
Figure 9.9.4: Color of various Ni(II) complexes in aqueous solution. From left to right, hexaamminenickel(II), tris(ethylenediamine)nickel(II), tetrachloronickelate(II) and hexaaquanickel(II). from LHcheM (via Wikipedia)

High- and low spin complexes

The magnitude of the *d* orbital splitting depends strongly on the nature of the ligand and in particular on how strong an electrostatic field is produced by its electron pair bond to the central ion.



If Δ is not too large then the electrons that occupy the *d* orbitals do so with their spins unpaired until a d^5 configuration is reached, just as occurs in the normal *Aufbau* sequence for atomic electron configurations. Thus a weak-field ligand such as H_2O leads to a "high spin" complex with Fe(II).



In contrast to this, the cyanide ion acts as a strong-field ligand; the *d* orbital splitting is so great that it is energetically more favorable for the electrons to pair up in the lower group of *d* orbitals rather than to enter the upper group with unpaired spins. Thus hexacyanoiron(II) is a "low spin" complex— actually zero spin, in this particular case.

Different *d* orbital splitting patterns occur in square planar and tetrahedral coordination geometries, so a very large number of arrangements are possible. In most complexes the value of Δ corresponds to the absorption of visible light, accounting for the colored nature of many such compounds in solution and in solids such as $\mathrm{CuSO_4 \cdot 5\,H_2O}$ ()Figure 9.9.1.

Approximately one-third of the chemical elements are present in living organisms. Many of these are metallic ions whose function within the cell depends on the formation of *d*-orbital coordination complexes with small molecules such as *porphyrins* (see below). These complexes are themselves bound within proteins (*metalloproteins*) which provide a local environment that is essential for

their function, which is either to transport or store diatomic molecule (oxygen or nitric oxide), to transfer electrons in oxidation-reduction processes, or to catalyze a chemical reaction. The most common of these utilize complexes of Fe and Mg, but other micronutrient metals including Cu, Mn, Mo, Ni, Se, and Zn are also important.

Hemoglobin

Hemoglobin is one of a group of heme proteins that includes myoglobin, cytochrome-c, and catalase. Hemoglobin performs the essential task of transporting dioxygen molecules from the lungs to the tissues in which it is used to oxidize glucose, this oxidation serving as the source of energy required for cellular metabolic processes.

Hemoglobin consists of four *globin* protein subunits (depicted by different colors in this diagram) joined together by weak intermolecular forces. Each of these subunits contains, buried within it, a molecule of *heme*, which serves as the active site of oxygen transport. The presence of hemoglobin increases the oxygen carrying capacity of 1 liter of blood from 5 to 250 ml. Hemoglobin is also involved in blood pH regulation and CO_2 transport.

Heme itself consists of an iron atom coordinated to a tetradentate *porphyrin*. When in the ferrous (Fe^{2+} state) the iron binds to oxygen and is converted into Fe^{3+}. Because a bare heme molecule would become oxidized by the oxygen without binding to it, the adduct must be stabilized by the surrounding globin protein. In this environment, the iron becomes octahedrally-coordinated through binding to a component of the protein in a fifth position, and in the sixth position either by an oxygen molecule or by a water molecule, depending on whether the hemoglobin is in its oxygenated state (in arteries) or deoxygenated state (in veins).

The heme molecule (purple) is enfolded within the polypeptide chain as shown here. The complete hemoglobin molecule contains four of these subunits, and all four must be present for it to function. The binding of O_2 to heme in hemoglobin is not a simple chemical equilibrium; the binding efficiency is regulated by the concentrations of H^+, CO_2, and organic phosphates. It is remarkable that the binding sites for these substances are on the outer parts of the globin units, far removed from the heme. The mechanism of this exquisite molecular-remote-control arises from the fact that the Fe^{2+}ion is too large to fit inside the porphyrin, so it sits slightly out of the porphyrin plane. This Fe radius diminishes when it is oxygenated, allowing it to move into the plane. In doing so, it pulls the protein component to which it is bound with it, triggering a sequence of structural changes that extend throughout the protein.

Myoglobin is another important heme protein that is found in muscles. Unlike hemoglobin, which consists of four protein subunits, myoglobin is made up of only one unit. Its principal function is to act as an oxygen storage reservoir, enabling vigorous muscle activity at a rate that could not be sustained by delivery of oxygen through the bloodstream. Myoglobin is responsible for the red color of meat. Cooking of meat releases the O_2 and oxidizes the iron to the +3 state, changing the color to brown.

Carbon monoxide poisoning

Other ligands, notably cyanide ion and **carbon monoxide**, are able to bind to hemoglobin much more strongly than does iron, thereby displacing it and rendering hemoglobin unable to transport oxygen. Air containing as little as 1 percent CO will convert hemoglobin to carboxyhemoglobin in a few hours, leading to loss of consciousness and death. Even small amounts of carbon monoxide can lead to substantial reductions in the availability of oxygen. The 400-ppm concentration of CO in cigarette smoke will tie up about 6% of the hemoglobin in heavy smokers; the increased stress this places on the heart as it works harder to compensate for the oxygen deficit is believed to be one reason why smokers are at higher risk for heart attacks. CO binds to hemoglobin 200 times more tightly than does O_2.

Chlorophyll

Chlorophyll is the light-harvesting pigment present in green plants. Its name comes from the Greek word χλορος (*chloros*), meaning “green”- the same root from which chlorine gets its name. Chlorophyll consists of a ring-shaped tetradentate ligand known as a *porphin* coordinated to a central magnesium ion. A histidine residue from one of several types of associated proteins forms a fifth coordinate bond to the Mg atom.

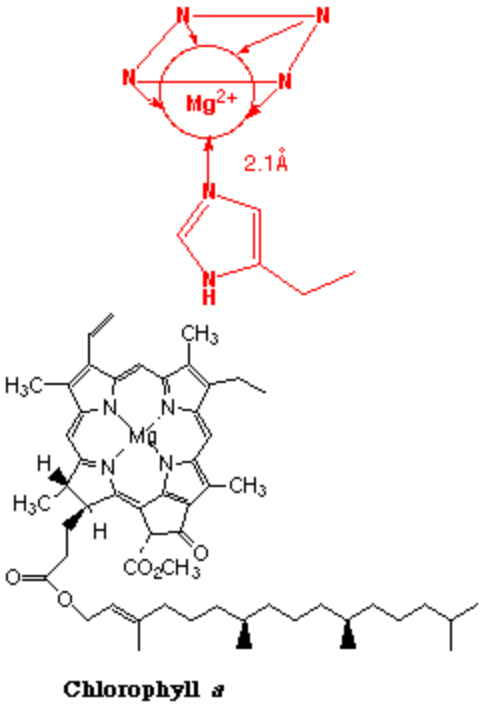


Figure 9.9.5: Left: detail of Mg coordination; notice that the metal is slightly out of the plane of the porphin ring. Right: plan view of Mg within the porphin ligand. A histidine residue from an associated protein forms the fifth coordination point to the Mg atom.

The light energy trapped by chlorophyll is utilized to drive a sequence of reactions whose net effect is to bring about the reduction of CO_2 to glucose ($C_6H_{12}O_6$) in a process known as photosynthesis which serves as the fuel for all life processes in both plants and animals.

This page titled 9.9: Bonding in Coordination Complexes is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.10: Bonding in Metals

Learning Objectives

- Explain the fundamental difference between the bonding in metallic solids compared to that in other types of solids and within molecules. Name some physical properties of metals that reflect this difference.
- Sketch out a diagram illustrating how a simple molecular-orbital approach to bonding in metals of Groups 1 and 2 always leaves some upper MO's empty.
- Describe, at the simplest level, the origin of electron "bands" in metals.
- Describe how the electrical and thermal conductivity of metals can be explained according to band theory.
- Explain why the electrical conductivity of a metal decreases with temperature, whereas that of a semiconductor increases.

Most of the known chemical elements are metals, and many of these combine with each other to form a large number of *intermetallic compounds*. The special properties of metals— their bright, lustrous appearance, their high electrical and thermal conductivities, and their malleability— suggest that these substances are bound together in a very special way.

Properties of metals

The fact that the metallic elements are found on the left side of the periodic table offers an important clue to the nature of how they bond together to form solids.

- These elements all possess *low electronegativities* and readily form positive ions M^{n+}. Because they show no tendency to form negative ions, the kind of bonding present in ionic solids can immediately be ruled out.
- The metallic elements have empty or nearly-empty outer *p*-orbitals, so there are never enough outer-shell electrons to place an octet around an atom.

These points lead us to the simplest picture of metals, which regards them as a lattice of positive ions immersed in a “sea of electrons” which can freely migrate throughout the solid. In effect the electropositive nature of the metallic atoms allows their valence electrons to exist as a mobile fluid which can be displaced by an applied electric field, hence giving rise to their high *electrical conductivities*. Because each ion is surrounded by the electron fluid in all directions, the bonding has no directional properties; this accounts for the high *malleability* and *ductility* of metals.

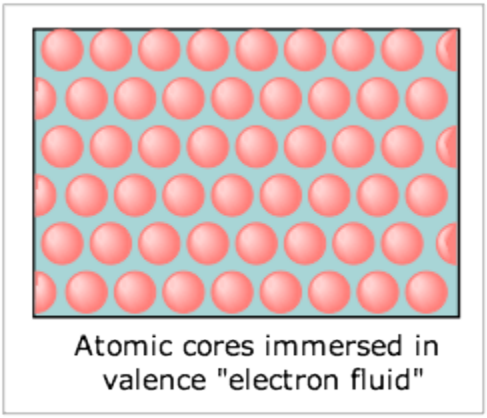


Figure 9.10.1: Atomic Cores Immersed in a Valence "Electron Fluid"

This view is an oversimplification that fails to explain metals in a quantitative way, nor can it account for the differences in the properties of individual metals. A more detailed treatment, known as the bond theory of metals, applies the idea of resonance hybrids to metallic lattices. In the case of an alkali metal, for example, this would involve a large number of hybrid structures in which a given Na atom shares its electron with its various neighbors.

Molecular orbitals in metals

will look. These are all constructed by combining the individual atomicMO's that are so closely spaced in energy that they form what is known as a *band* of allowed energies. In metallic lithium only the lower half of this band is occupied.

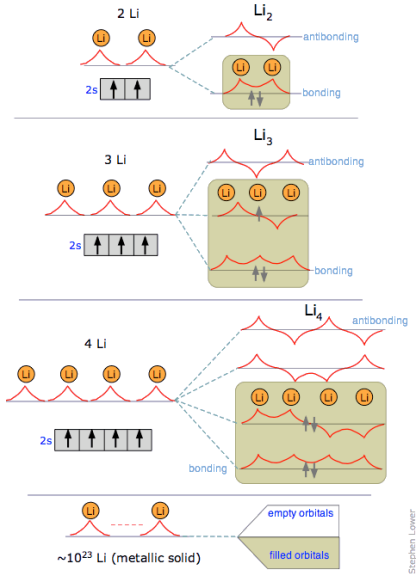


Figure 9.10.2: molecular orbitals in lithium molecules

Origin of Metallic Properties

Metallic solids possess special properties that set them apart from other classes of solids and make them easy to identify and familiar to everyone. All of these properties derive from the liberation of the valence electrons from the control of individual atoms, allowing them to behave as a highly mobile fluid that fills the entire crystal lattice. What were previously valence-shell orbitals of individual atoms become split into huge numbers of closely-spaced levels known as bands that extend throughout the crystal.

Melting Point and Strength

The strength of a metal derives from the electrostatic attraction between the lattice of positive ions and the fluid of valence electrons in which they are immersed. The larger the nuclear charge (atomic number) of the atomic kernel and the smaller its size, the greater this attraction. As with many other periodic properties, these work in opposite ways, as is seen by comparing the melting points of some of the Group 1-3 metals (right). Other factors, particularly the lattice geometry are also important, so exceptions such as is seen in Mg are not surprising.

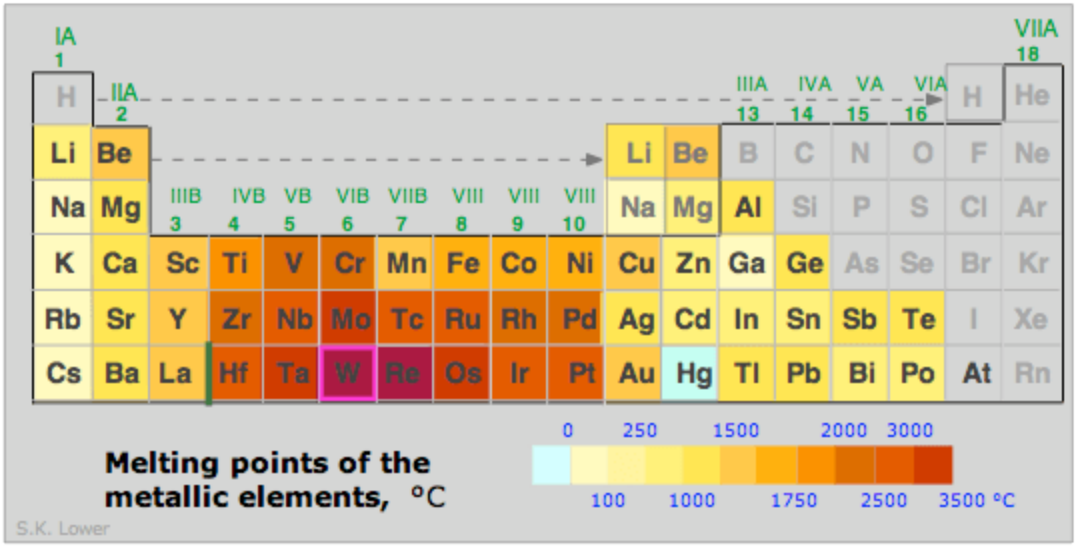


Figure 9.10.3: melting points of metals

In general, the transition metals with their valence-level *d* electrons are stronger and have higher melting points: Fe, 1539°C; Re 3180, Os 2727; W 3380°C.

Malleability and ductility

These terms refer respectively to how readily a solid can be shaped by pressure (forging, hammering, rolling into a sheet) and by being drawn out into a wire. Metallic solids are known and valued for these qualities, which derive from the non-directional nature of the attractions between the kernel atoms and the electron fluid. The bonding within ionic or covalent solids may be stronger, but it is also directional, making these solids subject to fracture (brittle) when struck with a hammer, for example. A metal, by contrast, is more likely to be simply deformed or dented.

Electrical conductivity: why are metals good conductors?

In order for a substance to conduct electricity, it must contain charged particles (*charge carriers*) that are sufficiently mobile to move in response to an applied electric field. In the case of ionic solutions and melts, the ions themselves serve this function. (Ionic solids contain the same charge carriers, but because they are fixed in place, these solids are insulators.) In metals the charge carriers are the electrons, and because they move freely through the lattice, metals are highly conductive. The very low mass and inertia of the electrons allows them to conduct high-frequency alternating currents, something that electrolytic solutions are incapable of. In terms of the band structure, application of an external field simply raises some of the electrons to previously unoccupied levels which possess greater momentum.

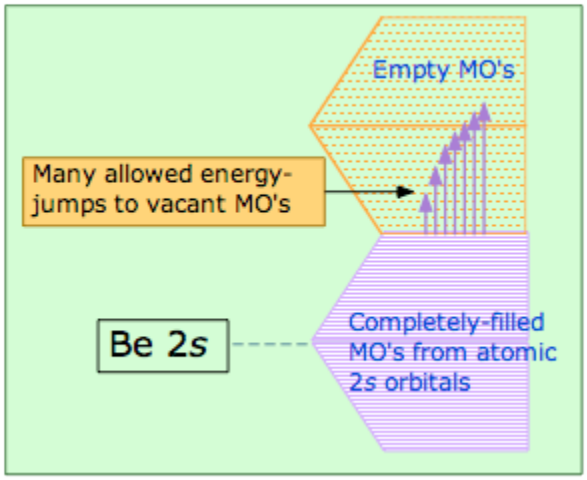


Figure 9.10.4: origin of metal conductivity

The conductivity of an electrolytic solution **decreases** as the temperature **falls** due to the decrease in "viscosity" which inhibits ionic mobility. The mobility of the electron fluid in metals is practically unaffected by temperature, but metals do suffer a slight conductivity decrease (opposite to ionic solutions) as the temperature rises; this happens because the more vigorous thermal motions of the kernel ions disrupts the uniform lattice structure that is required for free motion of the electrons within the crystal. Silver is the most conductive metal, followed by copper, gold, and aluminum.

Metals conduct electricity readily because of the essentially infinite supply of higher-energy empty MOs that electrons can populate as they acquire higher kinetic energies. This diagram illustrates the overlapping band structure (explained farther on) in beryllium. The MO levels are so closely spaced that even thermal energies can provide excitation and cause heat to rapidly spread through the solid.

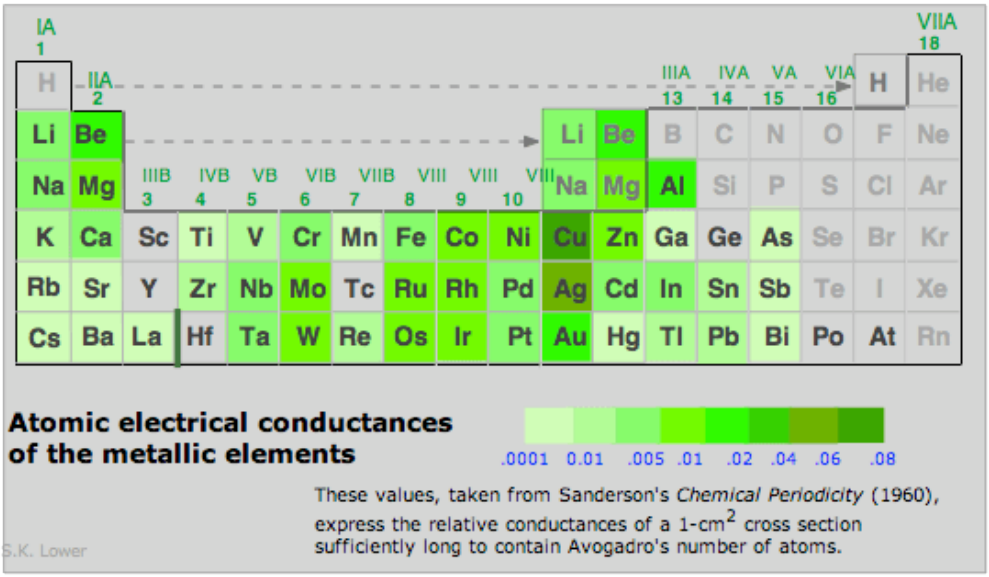


Figure 9.10.5: Electrical Conductances of the metals

Electrical conductivities of the metallic elements vary over a wide range. Notice that those of silver and copper (the highest of any metal) are in classes by themselves. Gold and aluminum follow close behind.

Thermal conductivity: why do metals conduct heat?

Everyone knows that touching a metallic surface at room temperature produces a colder sensation than touching a piece of wood or plastic at the same temperature. The very high thermal conductivity of metals allows them to draw heat out of our bodies very efficiently if they are below body temperature. In the same way, a metallic surface that is above body temperature will feel much warmer than one made of some other material. The high thermal conductivity of metals is attributed to vibrational excitations of the fluid-like electrons; this excitation spreads through the crystal far more rapidly than it does in non-metallic solids which depend on vibrational motions of atoms which are much heavier and possess greater inertia.

Appearance: Why are metals shiny?

We usually recognize a metal by its “metallic luster”, which refers to its ability of reflect light. When light falls on a metal, its rapidly changing electromagnetic field induces similar motions in the more loosely-bound electrons near the surface (this could not happen if the electrons were confined to the atomic valence shells.) A vibrating charge is itself an emitter of electromagnetic radiation, so the effect is to cause the metal to re-emit, or *reflect*, the incident light, producing the shiny appearance. What color is a metal? With the two exceptions of copper and gold, the closely-spaced levels in the bands allow metals to absorb all wavelengths equally well, so most metals are basically black, but this is ordinarily evident only when the metallic particles are so small that the band structure is not established.

The distinctive color of gold is a consequence of Einstein's theory of special relativity acting on the extremely high momentum of the inner-shell electrons, increasing their mass and causing the orbitals to contract. The outer (5d) electrons are less affected, and this gives rise to increased blue-light absorption, resulting in enhanced reflection of yellow and red light.

Thermionic Effect

The electrons within the electron fluid have a distribution of velocities very much like that of molecules in a gas. When a metal is heated sufficiently, a fraction of these electrons will acquire sufficient kinetic energy to escape the metal altogether; some of the electrons are essentially “boiled out” of the metal. This *thermionic effect*, which was first observed by Thomas Edison, was utilized in vacuum tubes which served as the basis of electronics from its beginning around 1910 until semiconductors became dominant in the 1960’s.

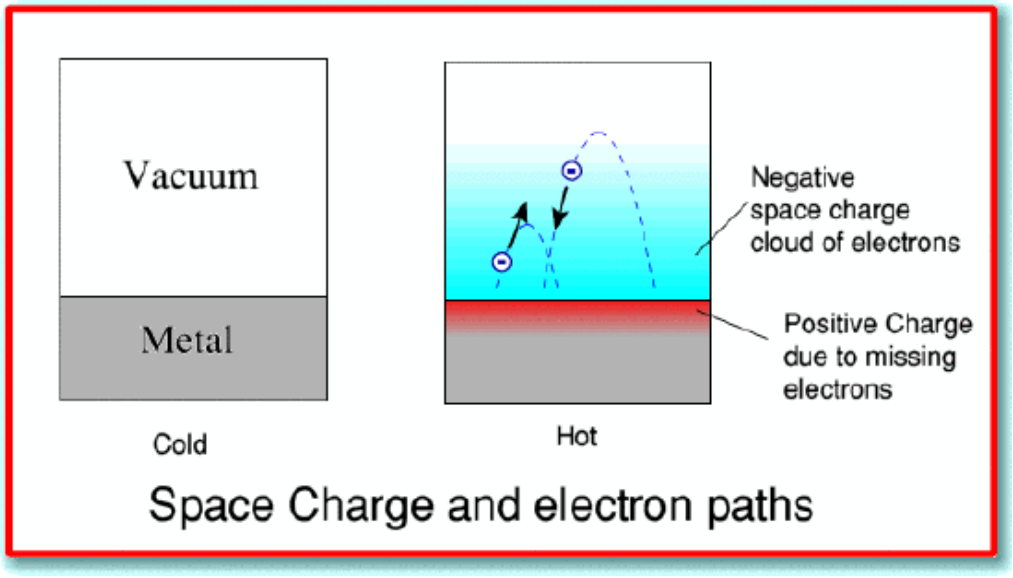


Figure 9.10.6: The Thermionic Effect

Band Structure of Metals

Most metals are made of atoms that have an outer configuration of s^2, which we would expect to completely fill the band of MO’s we have described. With the band completely filled and no empty levels above, we would not expect elements such as beryllium to be metallic. What happens is that the empty *p* orbitals also split into a band. Although the energy of the 2*p* orbital of an isolated Be atom is about 160 kJ greater than that of the 2*s* orbital, the bottom part of the 2*p* band overlaps the upper part of the 2*s* band, yielding a continuous *conduction band* that has plenty of unoccupied orbitals. It is only when these bands become filled with 2*p* electrons that the elements lose their metallic character.

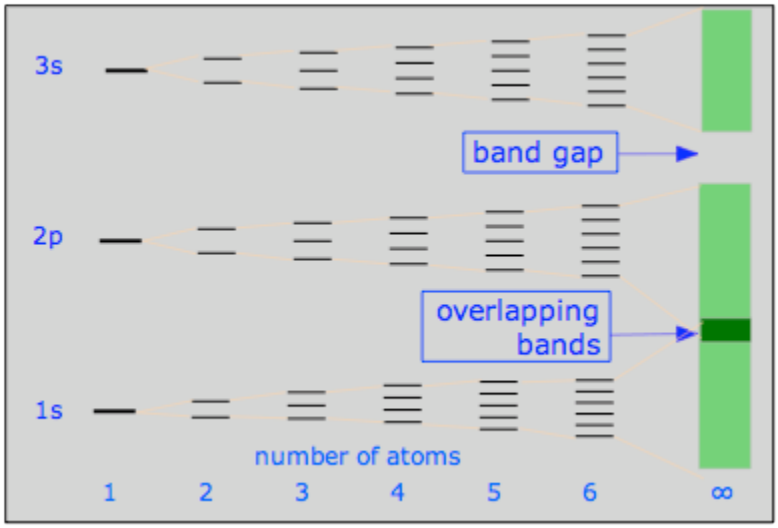


Figure 9.10.7: Origin of metal band structures

This diagram illustrates the band structure in a 3rd-row metal such as Na or Mg, and how it arises from MO splitting in very small units M_2 - M_6. The conduction bands for the "infinite" molecule M_N are shaded.

In most metals there will be bands derived from the outermost s-, *p*-, and *d* atomic levels, leading to a system of bands, some of which will overlap as described above. Where overlap does not occur, the almost continuous energy levels of the bands are separated by a forbidden zone, or *band gap*. Only the outermost atomic orbitals form bands; the inner orbitals remain localized on the individual atoms and are not involved in bonding.

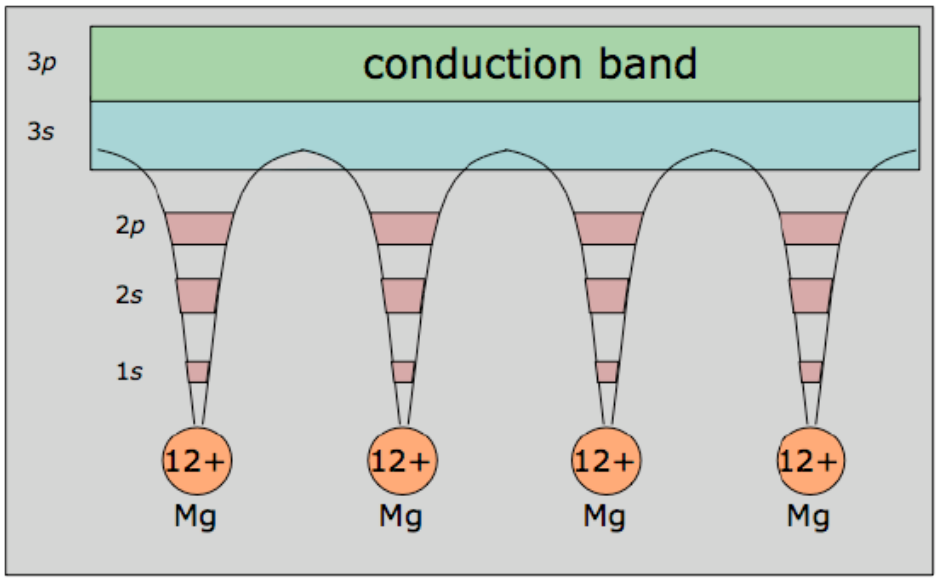


Figure 9.10.8: Conduction Bands in Magnesium solids

In its mathematical development, the band model relies strongly on the way that the free electrons within the metal interact with the ordered regularity of the crystal lattice. The alternative view shown here emphasizes this aspect by showing the inner orbitals as localized to the atomic cores, while the valence electrons are delocalized and belong to the metal as a whole, which in a sense constitutes a huge molecule in its own right.

This page titled 9.10: Bonding in Metals is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.11: Bonding in Semiconductors

Learning Objectives

- With the aid of simple diagrams, show how different band energy ranges in solids can produce conductors, insulators, and semiconductors.
- Describe the nature and behavior of a simple PN junction.

The band theory of solids provides a clear set of criteria for distinguishing between conductors (metals), insulators and semiconductors. As we have seen, a conductor must posses an upper range of allowed levels that are only partially filled with valence electrons. These levels can be within a single band, or they can be the combination of two overlapping bands. A band structure of this type is known as a *conduction band*.

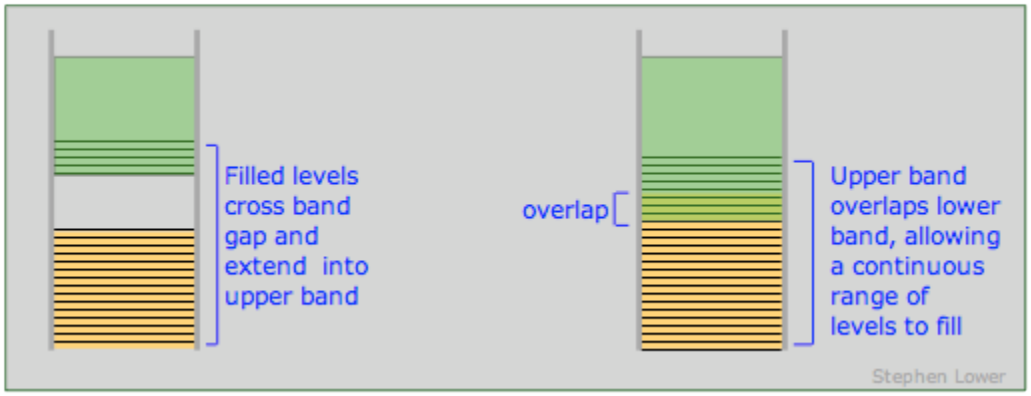


Figure 9.11.1: Conduction band structures in metals

Band arrangements in conductors. Metallic conduction requires the presence of empty levels into which electrons can move as they acquire momentum. This can be achieved when a band is only partially occupied or overlaps an empty band (right), or when the gap between a filled band and an upper empty one is sufficiently small (left) to allow ordinary thermal energy to supply the promotion energy.

Insulators and semiconductors

An *insulator* is characterized by a large band gap between the highest filled band and an even higher empty band. The band gap is sufficiently great to prevent any significant population of the upper band by thermal excitation of electrons from the lower one. The presence of a very intense electric field may be able to supply the required energy, in which case the insulator undergoes *dielectric breakdown*. Most molecular crystals are insulators, as are covalent crystals such as diamond.

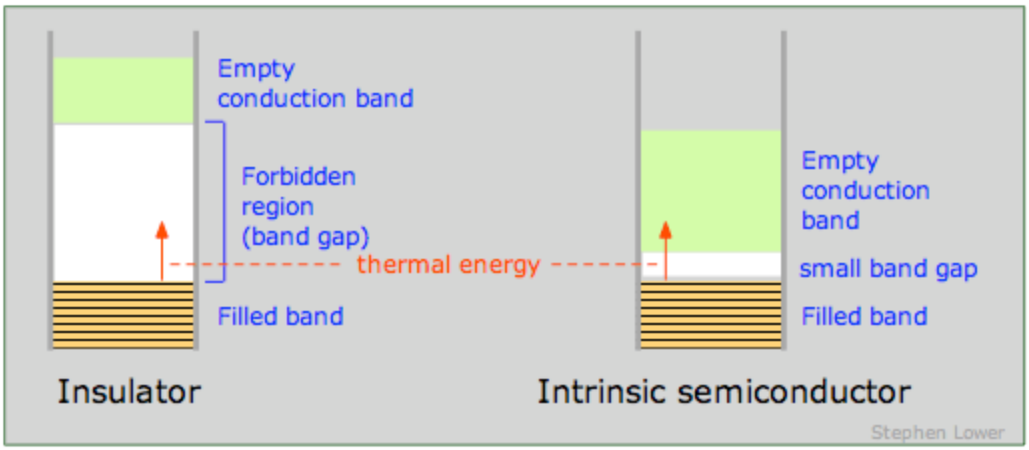


Figure 9.11.2: bands in insulators and semiconductors

If the band gap is sufficiently small to allow electrons in the filled band below it to jump into the upper empty band by thermal excitation, the solid is known as a *semiconductor*. In contrast to metals, whose electrical conductivity decreases with temperature (the more intense lattice vibrations interfere with the transfer of momentum by the electron fluid), the conductivity of semiconductors increases with temperature. In many cases the excitation energy can be provided by absorption of light, so most semiconductors are also *photoconductors*. Examples of semiconducting elements are Se, Te, Bi, Ge, Si, and graphite.

impurity band in semiconductors

The presence of an impurity in a semiconductor can introduce a new band into the system. If this new band is situated within the forbidden region, it creates a new and smaller band gap that will increase the conductivity. The huge semiconductor industry is based on the ability to tailor the band gap to fit the desired application by introducing an appropriate impurity atom (*dopant*) into the semiconductor lattice. The dopant elements are normally atoms whose valance shells contain one electron more or less than the atoms of the host crystal.

Semiconductor materials have traditionally been totally inorganic, composed mostly of the lighter P-block elements. More recently, organic semiconductors have become an important field of study and development.

Thermal properties of Semiconductors

At absolute zero, all of the charge carriers reside in lower of the bands below the small band gap in a semiconductor (that is, in the valence band of the illustration on the left above, or in the impurity band of the one on the right.) At higher temperatures, thermal excitation of the electrons allows an increasing fraction jump across this band gap and populate either the empty impurity band or the conduction band as shown at the right. The effect is the same in either case; the semiconductor becomes more conductive as the temperature is raised. Note that this is just the opposite to the way temperature affects the conductivity of metals.

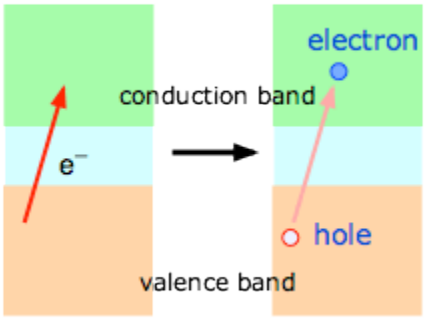


Figure 9.11.4: Thermal excitation into semiconductor conduction band

N- and P-type materials

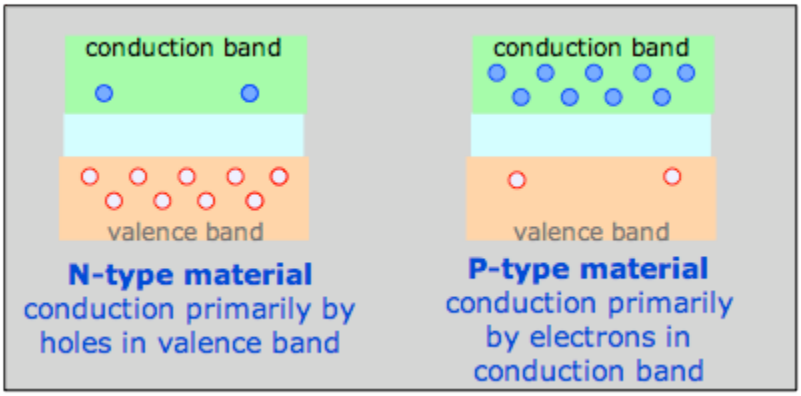


Figure 9.11.5: N and P type materials

For example, a phosphorus atom introduced as an impurity into a silicon lattice possesses one more valence electron than Si. This electron is delocalized within the impurity band and serves as the charge carrier in what is known as an *N-type semiconductor*. In a semiconductor of the P-type, the dopant might be arsenic, which has only three valence electrons. This creates what amounts to an electron deficiency or *hole* in the electron fabric of the crystal, although the solid remains electrically neutral overall. As this vacancy is filled by the electrons from silicon atoms the vacancy hops to another location, so the charge carrier is effectively a positively charged hole, hence the *P-type* designation.

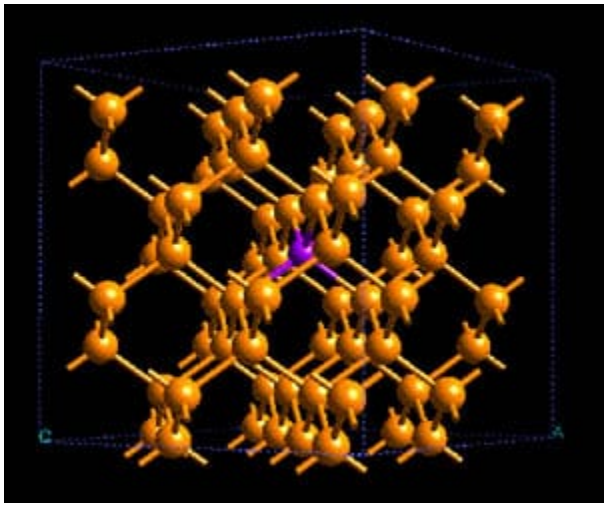


Figure 9.11.6: Structure of As (magenta) doped into crystalline Si (yellow.)

Substitution of just one dopant atom into 10^7 atoms of Si can increase the conductivity by a factor of 100,000.

The PN junction

When P- and N-type materials are brought into contact, creating a *PN junction*. Holes in the P material and electrons in the N material drift toward and neutralize each other, creating a depletion region that is devoid of charge carriers. But the destruction of these carriers leaves immobile positive ions in the N material and negative ions in the P material, giving rise to an interfacial potential difference ("*space charge*") as depicted here.

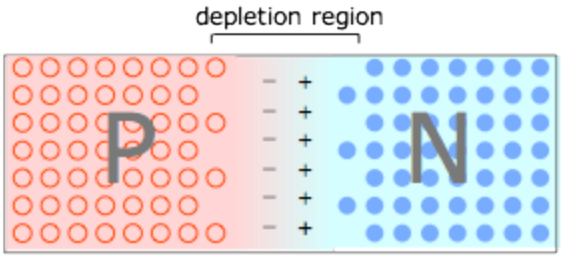


Figure 9.11.7: PN junction and depletion region

As this charge builds up, it acts to resist the further diffusion of electrons and holes, leaving a carrier-free *depletion region*, which acts as a barrier at the junction interface.

This page titled 9.11: Bonding in Semiconductors is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

9.12: The Shared-Electron Covalent Bond

More than one non-equivalent structure

It sometimes happens that the octet rule can be satisfied by arranging the electrons in different ways. For example, there are three different ways of writing valid electron dot structures for the thiocyanate ion SCN^-. Some of these structures are more realistic than others; to decide among them, you need to know something about the concepts of *formal charge* and *electronegativity*. These topics are discussed in the lesson that follows this one, where examples of working out such structures are given.

This page titled 9.12: The Shared-Electron Covalent Bond is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10: FUNDAMENTALS OF ACIDS AND BASES

Acids and bases touch upon virtually all areas of chemistry, biochemistry, and physiology. This set of lessons will get you started by presenting the underlying concepts in a systematic way. Aside from the section on pH which presumes an elementary knowledge of logarithms. The subject of acid-base equilibrium *calculations* is not covered in this lesson.

10.1: Introduction to Acids and Bases

10.2: Aqueous Solutions- pH and Titrations

10.3: Acid-base reactions à la Brønsted

10.4: Acid-Base Reactions

10.5: Lewis Acids and Bases

10.6: Types of Acids and Bases

10.7: Acid-Base Gallery

This page titled 10: Fundamentals of Acids and Bases is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Introduction to Acids and Bases

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above.

- Suggest simple tests you could carry out to determine if an unknown substance is an acid or a base.
- State the chemical definitions of an acid and a base in terms of their behavior in water.
- Write the formula of the salt formed when a given acid and base are combined.

The concepts of an acid, a base, and a salt are very old ones that have undergone several major refinements as chemical science has evolved. Our treatment of the subject at this stage will be mainly conceptual and qualitative, emphasizing the definitions and fundamental ideas associated with acids and bases. We will not cover calculations involving acid-base equilibria in these lessons.

1 Acids

The term *acid* was first used in the seventeenth century; it comes from the Latin root *ac-*, meaning “sharp”, as in *acetum*, vinegar. Some early writers suggested that acidic molecules might have sharp corners or spine-like projections that irritate the tongue or skin. Acids have long been recognized as a distinctive class of compounds whose aqueous solutions exhibit the following properties:

- A characteristic sour taste (think of lemon juice!);
- ability to change the color of litmus from blue to red;
- react with certain *metals* to produce gaseous H_2;
- react with *bases* to form a salt and water.

Litmus is a natural dye found in certain lichens. The name is of Scandinavian origin, e.g. *lit* (color) + *mosi* (moss) in Icelandic. "Litmus test" has acquired a meaning that transcends both Chemistry and science to denote any kind of test giving a yes/no answer.

How oxygen got mis-named

The first chemistry-based definition of an acid turned out to be wrong: in 1787, Antoine Lavoisier, as part of his masterful classification of substances, identified the known acids as a separate group of the “complex substances” (compounds). Their special nature, he postulated, derived from the presence of some common element that embodies the “acidity” principle, which he named *oxygen*, derived from the Greek for “acid former”.

Lavoisier had recently assigned this name to the new gaseous element that Joseph Priestly had discovered a few years earlier as the essential substance that supports combustion. Many combustion products (oxides) do give acidic solutions, and oxygen is in fact present in most acids, so Lavoisier’s mistake is understandable. In 1811 Humphrey Davy showed that muriatic (hydrochloric) acid (which Lavoisier had regarded as an element) does not contain oxygen, but this merely convinced some that chlorine was not an element but an oxygen-containing compound. Although a dozen oxygen-free acids had been discovered by 1830, it was not until about 1840 that the hydrogen theory of acids became generally accepted. By this time, the misnomer *oxygen* was too well established a name to be changed. The root *oxy* comes from the Greek word οξνς, which means "sour".

2 Acids and the hydrogen ion

The key to understanding acids (as well as bases and salts) had to await Michael Faraday's mid-nineteenth century discovery that solutions of salts (known as electrolytes) conduct electricity. This implies the existence of charged particles that can migrate under the influence of an electric field. Faraday named these particles *ions* (“wanderers”). Later studies on electrolytic solutions suggested that the properties we associate with acids are due to the presence of an excess of *hydrogen ions* in the solution. By 1890 the Swedish chemist Svante Arrhenius (1859-1927) was able to formulate the first useful theory of acids:

"an acidic substance is one whose molecular unit contains at least one hydrogen atom that can dissociate, or *ionize*, when dissolved in water, producing a hydrated hydrogen ion and an anion."

hydrochloric acid	$HCl \rightarrow H^+(aq) + Cl^-(aq)$
sulfuric acid	$H_2SO_4 \rightarrow H^+(aq) + HSO_4^-(aq)$
hydrogen sulfite ion	$HSO_3^-(aq) \rightarrow H^+(aq) + SO_3^{2-}(aq)$
acetic acid	$H_3CCOOH \rightarrow H^+(aq) + H_3CCOO^-(aq)$

Strictly speaking, an “Arrhenius acid” must contain hydrogen. However, there are substances that do not themselves contain hydrogen, but still yield hydrogen ions when dissolved in water; the hydrogen ions come from the water itself, by reaction with the substance.

An acid is a substance that yields an excess of hydrogen ions when dissolved in water.

There are three important points to understand about hydrogen in acids:

- Although all Arrhenius acids contain hydrogen, not all hydrogen atoms in a substance are capable of dissociating; thus the $-CH_3$hydrogens of acetic acid are “non-acidic”. An important part of knowing chemistry is being able to predict which hydrogen atoms in a substance will be able to dissociate into hydrogen ions.
- Those hydrogens that do dissociate can do so to different degrees. The *strong acids* such as HCl and HNO_3 are effectively 100% dissociated in solution. Most organic acids, such as acetic acid, are *weak*; only a small fraction of the acid is dissociated in most solutions. HF and HCN are examples of weak inorganic acids.
- Acids that possess more than one dissociable hydrogen atom are known as *polyprotic* acids; H_2SO_4 and H_3PO_4 are well-known examples. Intermediate forms such as HPO_4^{2-}, being capable of both accepting and losing protons, are called *ampholytes*.

H_2SO_4 sulfuric acid	→	**HSO_4^-** hydrogen sulfate ("bisulfate") ion	→	SO_4^{2-} sulfate ion			
H_2S hydrosulfuric acid	→	**HS^-** hydrosulfide ion	→	S^{2-} sulfide ion			
H_3PO_4 phosphoric acid	→	**$H_2PO_4^-$** dihydrogen phosphate ion	→	**HPO_4^{2-}** hydrogen phosphate ion	→	PO_4^{3-} phosphate ion	
HOOC-COOH oxalic acid	→	**$HOOC\text{-}COO^-$** hydrogen oxalate ion	→	$^-OOC\text{-}COO^-$ oxalate ion			

You will find out in a later section of this lesson that hydrogen ions cannot exist as such in water, but don't panic! It turns out that

chemists still find it convenient to pretend as if they are present, and to write reactions that include them.

3 Bases

The name *base* has long been associated with a class of compounds whose aqueous solutions are characterized by:

- a bitter taste;
- a “soapy” feeling when applied to the skin
- ability to restore the original blue color of litmus that has been turned red by acids
- ability to react with acids to form salts.
- react with certain *metals* to produce gaseous H_2

The word “alkali” is often applied to strong inorganic bases. It is of Arabic origin, from *al-kali* ("the ashes") which refers to the calcined wood ashes that were boiled with water to obtain *potash* which contains the strong base KOH, used in soap making. The element name *potassium* and its symbol K (from the Latin *kalium)* derive from these sources.

Just as an acid is a substance that liberates hydrogen ions into solution, a *base*yields *hydroxide ions* when dissolved in water:

$NaOH_{(s)} \rightarrow Na^{+}_{(aq)} + OH^{-(aq)}$

Sodium hydroxide is an *Arrhenius base* because it contains hydroxide ions. However, other substances which do not contain hydroxide ions can nevertheless produce them by reaction with water, and are therefore also classified as bases. Two classes of such substances are the metal oxides and the hydrogen compounds of certain nonmetals:

$Na_2O_{(s)} + H_2O \rightarrow [2\ NaOH] \rightarrow 2\ Na^{+}(aq) + 2\ OH^{-}(aq)$

$NH_3 + H_2O \rightarrow NH_4^{+}(aq) + OH^{-}(aq)$

A base is a substance that yields an excess of hydroxide ions when dissolved in water.

4 Neutralization

Acids and bases react with one another to yield two products: water, and an ionic compound known as a *salt*. This kind of reaction is called a *neutralization reaction*.

$$\underset{\text{acid}}{HCl} + \underset{\text{base}}{NaOH} \rightarrow \underset{\text{salt}}{NaCl} + \underset{\text{water}}{H_2O}$$

This "molecular" equation is convenient to write, but we need to re-cast it as a net ionic equation to reveal what is really going on here when the reaction takes place in water, as is almost always the case.

$$H^{+} + Cl^{-} + Na^{+} + OH^{-} \rightarrow Na^{+} + Cl^{-} + H_2O$$

If we cancel out the ions that appear on both sides (and therefore don't really participate in the reaction), we are left with the net equation

$H^{+}_{(aq)} + OH^{-}_{(aq)} \rightarrow H_2O$ (**1**)

which is the fundamental process that occurs in all neutralization reactions.

Confirmation that this equation describes all neutralization reactions that take place in water is provided by experiments indicating that no matter what acid and base are combined, all liberate the same amount of heat (57.7 kJ) per mole of H^+neutralized.

In the case of a weak acid, or a base that is not very soluble in water, more than one step might be required. For example, a similar reaction can occur between acetic acid and calcium hydroxide to produce *calcium acetate*:

$$2\ CH_3COOH + Ca(OH)_2 \rightarrow CH_3COOCa + 2\ H_2O$$

If this takes place in aqueous solution, the reaction is really between the very small quantities of H^+ and OH^- resulting from the dissociation of the acid and the dissolution of the base, so the reaction is identical with Equation 1:

$$H^+_{(aq)} + OH^-_{(aq)} \rightarrow H_2O$$

If, on the other hand, we add *solid* calcium hydroxide to pure liquid acetic acid, the net reaction would include both reactants in their "molecular" forms:

$$2\ CH_3COOH_{(l)} + Ca(OH)_{2\ (s)} \rightarrow 2\ CH_3COO^- + Ca^{2+} + 2\ H_2O$$

The “salt” that is produced in a neutralization reaction consists simply of the anion and cation that were already present. The salt can be recovered as a solid by evaporating the water.

This page titled 10.1: Introduction to Acids and Bases is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10.2: Aqueous Solutions- pH and Titrations

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Define the *ion product of water*, and know its room-temperature value.
- State the criteria, in terms of H^+ and OH^- concentrations, of an *acidic*, *alkaline*, and a *neutral* solution.
- Given the effective hydrogen ion concentration in a solution, calculate the *pH*. Conversely, find the hydrogen- or hydroxide ion concentration in a solution having a given pH.
- Find the *pH* or *pOH* of a solution when either one is known.
- Describe the process of *acid-base titration*, and explain the significance of the equivalence point.
- Sketch a typical *titration curve* for a *monoprotic* or *polyprotic acid*.

As you will see in the lesson that follows this one, *water* plays an essential role in acid-base chemistry as we ordinarily know it. To even those who know very little about chemistry, the term *pH* is recognized as a measure of "acidity", so the major portion of this unit is devoted to the definition of pH and of the pH scale. But since these topics are intimately dependant on the properties of water and its ability do dissociate into hydrogen and hydroxyl ions, we begin our discussion with this topic. We end this lesson with a brief discussion of acid-base titration— probably the most frequently carried-out chemistry laboratory operation in the world.

Dissociation of water

The ability of acids to react with bases depends on the tendency of hydrogen ions to combine with hydroxide ions to form water:

$$H^+_{(aq)} + OH^-_{(aq)} \rightarrow H_2O \ (\mathbf{1})$$

This tendency happens to be very great, so the reaction is practically complete— but not "completely" complete; a few stray H^+ and OH^- ions will always be present. What's more, this is true even if you start with the purest water attainable. This means that in pure water, the *reverse* reaction, the "dissociation" of water

$$H_2O \rightarrow H^+_{(aq)} + OH^-_{(aq)} \ \mathbf{(2)}$$

will proceed to a very slight extent. Both reactions take place simultaneously, but **(1)** is so much faster than **(2)** that only a minute fraction of H_2O molecules are dissociated.

Liquids that contain ions are able to conduct an electric current. Pure water is practically an insulator, but careful experiments show that even the most highly purified water exhibits a very slight conductivity that corresponds to a concentration of both the H^+ ion and OH^- ions of almost exactly 1.00×10^{-7} mol L^{-1} at 25°C.

> All chemical reactions that take place in a single phase (such as in a solution) are theoretically "incomplete" and are said to be reversible.

Example

What *fraction* of water molecules in a liter of water are dissociated

Solution: 1 L of water has a mass of 1000 g. The number of moles in 1000 g of H_2O is (1000 g)/(18 g mol^{-1}) = 55.5 mol. This corresponds to (55.5 mol)(6.02E23 mol-1) = 3.34E25 H_2O molecules.

An average of 10^{-7} mole, or $(10^{-7})(6.02E23) = 6.0E16$ H_2O molecules will be dissociated at any time. The fraction of dissociated water molecules is therefore (6.0E16)/(3.3E25) = 1.8E–9.

Thus we can say that only about two out of every billion (10^9) water molecules will be dissociated.

Ion Product of water

The degree of dissociation of water is so small that you might wonder why it is even mentioned here. The reason stems from an important relationship that governs the concentrations of H^+ and OH^- ions in aqueous solutions:

$[H^+][OH^-] = 1.00 \times 10^{-14}$ **(3)**

must know this!

in which the square brackets [ ] refer to the concentrations (in moles per litre) of the substances they enclose.

The quantity $1.00 \times 10^{-14}$is commonly denoted by K_w. Its value varies slightly with temperature, pressure, and the presence of other ions in the solution.

This expression is known as the *ion product of water*, and it applies to *all* aqueous solutions, not just to pure water. The consequences of this are far-reaching, because it implies that if the concentration of H^+ is large, that of OH^- will be small, and vice versa. This means that H^+ ions are present in *all* aqueous solutions, not just acidic ones.

This leads to the following **important definitions**, which you must know:

acidic solution	$[H^+] > [OH^-]$
alkaline ("basic") solution	$[H^+] < [OH^-]$
neutral solution	$[H^+] = [OH^-] = 1.00 \times 10^{-7}$ mol L^{-1}

Take special note of the following definition:

A *neutral solution* is one in which the concentrations of H^+ and OH^- ions are identical.

The values of these concentrations are constrained by Eq. **3**. Thus, in a *neutral solution*, both the hydrogen- and hydroxide ion concentrations are 1.00×10^{-7} mol L^{-1}:

$[H^+][OH^-] = [1.00 \times 10^{-7}][1.00 \times 10^{-7}] = 1.00 \times 10^{-14}$

Hydrochloric acid is a typical strong acid that is totally dissociated in solution:

$HCl \rightarrow H^+(aq) + Cl^-(aq)$

A 1.0*M* solution of HCl in water therefore does not really contain any significant concentration of HCl molecules at all; it is a solution in of H^+ and Cl^- in which the concentrations of both ions are 1.0 mol L^{-1}. The concentration of hydroxide ion in such a solution, according to Eq **2**, is

$[OH^-] = (K_w)/[H^+] = (1.00 \times 10^{-14}) / (1 \text{ mol } L^{-1}) = 1.00 \times 10^{-14}$ mol L^{-1}.

Similarly, the concentration of hydrogen ion in a solution made by dissolving 1.0 mol of sodium hydroxide in water will be 1.00 x

10^{-14} mol L^{-1}.

2 pH

When dealing with a range of values (such as the variety of hydrogen ion concentrations encountered in chemistry) that spans many powers of ten, it is convenient to represent them on a more compressed logarithmic scale. By convention, we use the *pH scale* to denote hydrogen ion concentrations:

$$\mathbf{pH = -\log_{10}[H^+]}\ \mathbf{(4)}$$ must know this!

or conversely, $\mathbf{[H^+] = 10^{-pH}}$.

This notation was devised by the Danish chemist Soren Sorenson (1868-1939) in 1909. There are several accounts of why he chose "pH"; a likely one is that the letters stand for the French term *pouvoir hydrogène*, meaning "power of hydrogen"— "power" in the sense of an exponent. It has since become common to represent other small quantities in "p-notation". Two that you need to know in this course are the following:

$pOH = -\log_{10}[OH^-]$

$pK_w = -\log K_w$ (= 14 when $K_w = 1.00 \times 10^{-14}$)

Note that pH and pOH are expressed as numbers without any units, since logarithms must be dimensionless.

Recall from Eq **3** that $[H^+][OH^-] = 1.00 \times 10^{-14}$; if we write this in "p-notation" it becomes

$$\mathbf{pH + pOH = 14}$$

(5) must know this!

In a neutral solution at 25°C, pH = pOH = 7.0. As pH increases, pOH diminishes; a higher pH corresponds to an alkaline solution, a lower pH to an acidic solution. In a solution with $[H^+] = 1$ M , the pH would be 0; in a 0.00010 *M* solution of H^+, it would be 4.0. Similarly, a 0.00010 M solution of NaOH would have a pOH of 4.0, and thus a pH of 10.0. It is very important that you thoroughly understand the pH scale, and be able to convert between $[H^+]$ or $[OH^-]$ and pH in both directions.

Example

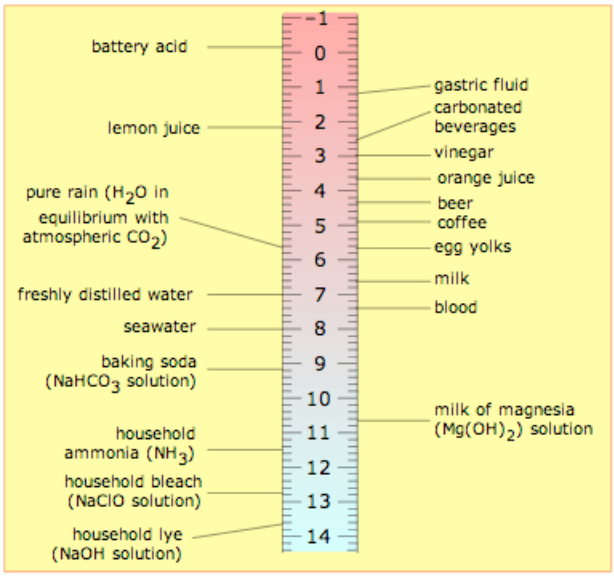
The pH of blood must be held very close to 7.40. Find the hydroxide ion concentration that corresponds to this pH.

Solution: The pOH will be (14.0 – 7.40) = 6.60.
$[OH^-] = 10^{-pOH} = 10^{-6.6} = 2.51 \times 10^{-7}$ *M*

The pH scale

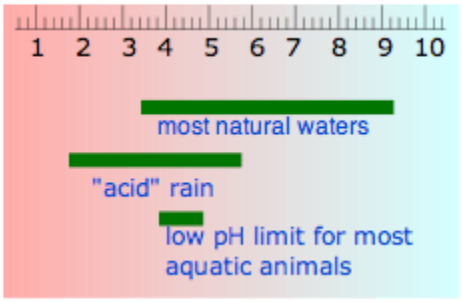
The range of possible pH values runs from about 0 to 14.

The word "about" in the above statement reflects the fact that at very high concentrations (10 M hydrochloric acid or sodium hydroxide, for example,) a significant fraction of the ions will be associated into neutral pairs such as $H^+ \cdot Cl^-$, thus reducing the concentration of “available” ions to a smaller value which we will call the *effective concentration*. It is the *effective concentration* of H^+ and OH^- that determines the pH and pOH. For solutions in which ion concentrations don't exceed 0.1 M, the formulas pH = –log $[H^+]$ and pOH = –log$[OH^-]$ are generally reliable, but don't expect a 10.0 M solution of a strong acid to have a pH of exactly –1.00!



The table shown here will help give you a general feeling for where common substances fall on the pH scale. Notice especially that

- most foods are slightly acidic;
- the principal "bodily fluids" are slightly alkaline, as is seawater— not surprising, since early animal life began in the oceans.
- the pH of freshly-distilled water will drift downward as it takes up carbon dioxide from the air; CO_2 reacts with water to produce carbonic acid, H_2CO_3.
- the pH of water that occurs in nature varies over a wide range. Groundwaters often pick up additional CO_2 respired by organisms in the soil, but can also become alkaline if they are in contact with carbonate-containing sediments. "Acid" rain is by definition more acidic than pure water in equilibrium with atmospheric CO_2, owing mainly to sulfuric and nitric acids that originate from fossil-fuel emissions of nitrogen oxides and SO_2.



pH indicators

The colors of many dye-like compounds depend on the pH, and can serve as useful *indicators* to determine whether the pH of a solution is above or below a certain value.

Natural indicator dyes



The best known of these is of course litmus, which has served as a means of distinguishing between acidic and alkaline substances since the early 18th century.



Many *flower pigments* are also dependent on the pH. You may have noticed that the flowers of some hydrangea shrub species are blue when grown in acidic soils, and white or pink in alkaline soils.

Red cabbage is a popular make-it-yourself indicator.

Universal indicators

Most indicator dyes show only one color change, and thus are only able to determine whether the pH of a solution is greater or less than the value that is characteristic of a particular indicator. By combining a variety of dyes whose color changes occur at different pHs, a "universal" indicator can be made. Commercially-prepared pH test papers of this kind are available for both wide and narrow pH ranges.



Titration

Since acids and bases readily react with each other, it is experimentally quite easy to find the amount of acid in a solution by determining how many moles of base are required to neutralize it. This operation is called *titration*, and you should already be familiar with it from your work in the Laboratory.

We can titrate an acid with a base, or a base with an acid. The substance whose concentration we are determining (the *analyte*) is the substance being titrated; the substance we are adding in measured amounts is the *titrant*. The idea is to add titrant until the titrant has reacted with all of the analyte; at this point, the number of moles of titrant added tells us the concentration of base (or acid) in the solution being titrated.

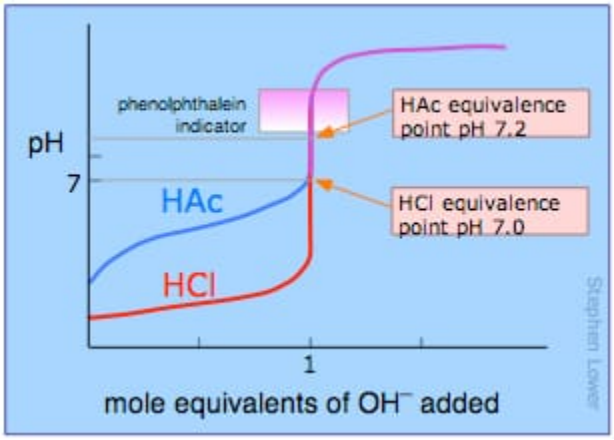
Example 10.2.1:

36.00 ml of a solution of HCl was titrated with 0.44 M KOH. The volume of KOH solution required to neutralize the acid solution was 27.00 ml. What was the concentration of the HCl?

Solution: The number of moles of titrant added was
(.027 L)(.44 mol L^{-1}) = .0119 mol. Because one mole of KOH reacts with one mole of HCl, this is also the number of moles of HCl; its concentration is therefore
(.0119 mol) ÷ (.036 L) = 0.33 *M* .

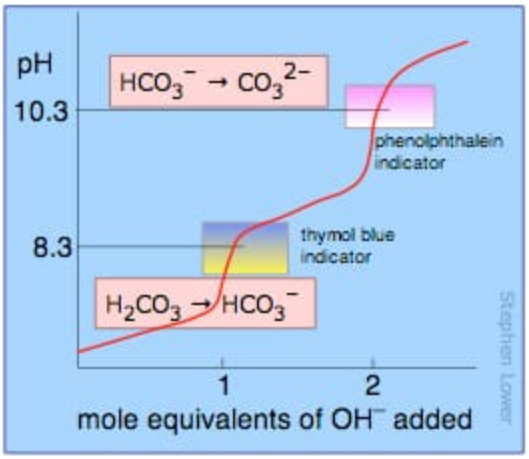
Titration curves

The course of a titration can be followed by plotting the pH of the solution as a function of the quantity of titrant added. The figure shows two such curves, one for a strong acid (HCl) and the other for a weak acid, acetic acid, denoted by HAc. Looking first at the HCl curve, notice how the pH changes very slightly until the acid is almost neutralized. At that point, which corresponds to the vertical part of the plot, just one additional drop of NaOH solution will cause the pH to jump to a very high value— almost as high as that of the pure NaOH solution.



Compare the curve for HCl with that of HAc. For a weak acid, the pH jump near the neutralization point is less steep. Notice also that the pH of the solution at the neutralization point is greater than 7. These two characteristics of the titration curve for a weak acid are very important for you to know.

If the acid or base is polyprotic, there will be a jump in pH for each proton that is titrated. In the example shown here, a solution of carbonic acid H_2CO_3 is titrated with sodium hydroxide. The first equivalence point (at which the H_2CO_3 has been converted entirely into bicarbonate ion HCO_3^-) occurs at pH 8.3. The solution is now identical to one prepared by dissolving an identical amount of sodium bicarbonate in water.



Addition of another mole equivalent of hydroxide ion converts the bicarbonate into carbonate ion and is complete at pH 10.3; an identical solution could be prepared by dissolving the appropriate amount of sodium carbonate in water.

Finding the equivalence point: indicators

When enough base has been added to react completely with the hydrogens of a monoprotic acid, the equivalence point has been reached. If a strong acid and strong base are titrated, the pH of the solution will be 7.0 at the equivalence point. However, if the acid is a weak one, the pH will be greater than 7; the “neutralized” solution will not be “neutral” in terms of pH. For a polyprotic acid, there will be an equivalence point for each titratable hydrogen in the acid. These typically occur at pH values that are 4-5 units apart, but they are occasionally closer, in which case they may not be readily apparent in the titration curve.

The key to a successful titration is knowing when the equivalence point has been reached. The easiest way of finding the equivalence point is to use an *indicator* dye; this is a substance whose color is sensitive to the pH. One such indicator that is commonly encountered in the laboratory is *phenolphthalein*; it is colorless in acidic solution, but turns intensely red when the solution becomes alkaline. If an acid is to be titrated, you add a few drops of phenolphthalein to the solution before beginning the titration. As the titrant is added, a local red color appears, but quickly dissipates as the solution is shaken or stirred. Gradually, as the equivalence point is approached, the color dissipates more slowly; the trick is to stop the addition of base after a single drop results in a permanently pink solution.

Different indicators change color at different pH values. Since the pH of the equivalence point varies with the strength of the acid being titrated, one tries to fit the indicator to the particular acid. One can titrate polyprotic acids by using a suitable combination of several indicators, as is illustrated above for carbonic acid.

This page titled 10.2: Aqueous Solutions- pH and Titrations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10.3: Acid-base reactions à la Brønsted

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Explain the difference between the *Arrhenius* and *Bronsted-Lowry concepts* of acids and bases, and give examples of an acid and a base in which the Arrhenius concept is inapplicable.
- Explain why a *hydrogen ion* cannot exist in water.
- Given the formula of an acid or base, write the formula of its *conjugate*.
- State the fundamental difference between a *strong acid* and a *weak acid*.
- Describe the *leveling effect*, and explain its origin.
- State the factors that determine whether a solution of a *salt* will be acidic or alkaline.
- Write the equation for the *autoprotolysis* of a given *ampholyte*.

In this lesson we develop this concept and illustrate its applications to "strong" and "weak" acids and bases, emphasizing the common theme that acid-base chemistry is always a competition between two bases for the proton. In the final section, we show how the concept of "proton energy" can help us understand and predict the direction and extent of common types of acid-base reactions without the need for calculations.

Proton donors and acceptors

The older Arrhenius theory of acids and bases viewed them as substances which produce hydrogen ions or hydroxide ions on dissociation. As useful a concept as this has been, it was unable to explain why NH_3, which contains no OH^- ions, is a base and not an acid, why a solution of $FeCl_3$ is acidic, or why a solution of Na_2S is alkaline. A more general theory of acids and bases was developed by Franklin in 1905, who suggested that the *solvent* plays a central role. According to this view, an acid is a solute that gives rise to a *cation* (positive ion) characteristic of the solvent, and a base is a solute that yields a *anion* (negative ion) which is also characteristic of the solvent. The most important of these solvents is of course H_2O, but Franklin's insight extended the realm of acid-base chemistry into non-aqueous systems as we shall see in a later lesson.

Brønsted acids and bases

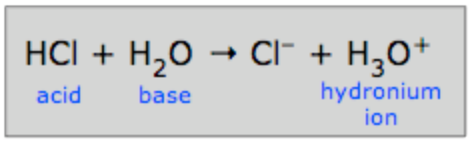
In 1923, the Danish chemist J.N. Brønsted, building on Franklin's theory, proposed that an acid is a proton donor; a base is a proton acceptor. In the same year the English chemist T.M. Lowry published a paper setting forth some similar ideas without producing a definition; in a later paper Lowry himself points out that Brønsted deserves the major credit, but the concept is still widely known as the Brønsted-Lowry theory.

An acid is a proton donor and a base is a proton acceptor.

These definitions carry a very important implication: a substance cannot act as an acid without the presence of a base to accept the proton, and *vice versa*. As a very simple example, consider the equation that Arrhenius wrote to describe the behavior of hydrochloric acid:

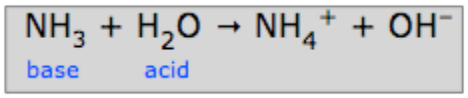
$$HCl \rightarrow H^+ + A^- \tag{10.3.1}$$

This is fine as far as it goes, and chemists still write such an equation as a shortcut. But in order to represent this more realistically as a proton donor-acceptor reaction, we now depict the behavior of HCl in water by



in which the acid HCl donates its proton to the acceptor (base) H_2O.

"Nothing new here", you might say, noting that we are simply replacing a shorter equation by a longer one. But consider how we might explain the alkaline solution that is created when ammonia gas NH_3 dissolves in water. An alkaline solution contains an excess of hydroxide ions, so ammonia is clearly a base, but because there are no OH^- ions in NH_3, it is clearly not an Arrhenius base. It is, however, a Brønsted base:



In this case, the water molecule acts as the acid, donating a proton to the base NH_3 to create the *ammonium ion* NH_4^+.

The foregoing examples illustrate several important aspects of the Brønsted-Lowry concept of acids and bases:

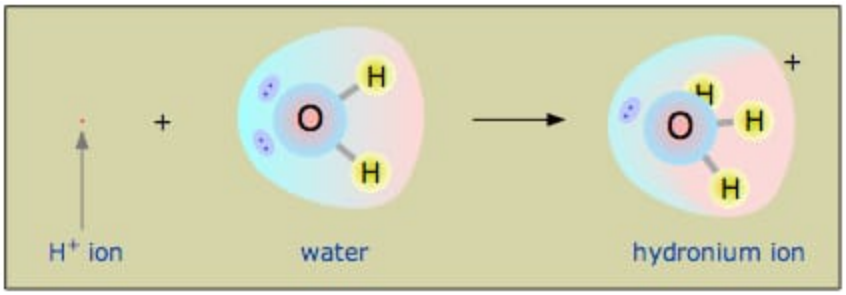
- A substance cannot act as an acid unless a proton acceptor (base) is present to receive the proton;
- A substance cannot act as a base unless a proton donor (acid) is present to supply the proton;
- Water plays a dual role in many acid-base reactions; H_2O can act as a proton acceptor (base) for an acid, or it can serve as a proton donor (acid) for a base (as we saw for ammonia.
- The *hydronium ion* H_3O^+ plays a central role in the acid-base chemistry of aqueous solutions.

Brønsted (1879-1947) was a Danish physical chemist. Although he is now known mainly for his proton donor-acceptor theory of acids and bases (see his original article), he published numerous earlier papers on chemical affinity, and later on the catalytic effects of acids and bases on chemical reactions. In World War II he opposed the Nazis, and this led to his election to the Danish parliament in 1947, but he was unable to take his seat because of illness and he died later in that year.

Lowry (1874-1936) was the first holder of the chair in physical chemistry at Cambridge University. His extensive studies of the effects of acids and bases on the optical behavior of camphor derivatives (specifically, how they rotate the plane of polarized light) led him to formulate a theory of acids and bases similar to and simultaneously with that of Brønsted.

The Hydronium ion

There is another serious problem with the Arrhenius view of an acid as a substance that dissociates in water to produce a hydrogen ion. The hydrogen ion is no more than a proton, a bare nucleus. Although it carries only a single unit of positive charge, this charge is concentrated into a volume of space that is only about a hundred-millionth as large as the volume occupied by the smallest atom. (Think of a pebble sitting in the middle of a sports stadium!) The resulting extraordinarily high *charge density* of the proton strongly attracts it to any part of a nearby atom or molecule in which there is an excess of negative charge. In the case of water, this will be the lone pair (unshared) electrons of the oxygen atom; the tiny proton will be buried within the lone pair and will form a shared-electron (coordinate) bond with it, creating a *hydronium ion*, H_3O^+. In a sense, H_2O is acting as a base here, and the product H_3O^+ is the conjugate acid of water:



Owing to the overwhelming excess of H_2O molecules in aqueous solutions, a bare hydrogen ion has no chance of surviving in water.

Although other kinds of dissolved ions have water molecules bound to them more or less tightly, the interaction between H^+ and H_2O is so strong that writing "$H^+(aq)$" hardly does it justice, although it is formally correct. The formula H_3O^+ more adequately conveys the sense that it is both a molecule in its own right, and is also the conjugate acid of water.

The equation "HA → H^+ + A^-" is so much easier to write that chemists still use it to represent acid-base reactions in contexts in which the proton donor-acceptor mechanism does not need to be emphasized. Thus it is permissible to talk about "hydrogen ions" and use the formula H^+ in writing chemical equations as long as you remember that they are not to be taken literally in the context of aqueous solutions.

Interestingly, experiments indicate that the proton does not stick to a single H_2O molecule, but changes partners many times per second. This molecular promiscuity, a consequence of the uniquely small size and mass the proton, allows it to move through the solution by rapidly hopping from one H_2O molecule to the next, creating a new H_3O^+ ion as it goes. The overall effect is the same as if the H_3O^+ ion itself were moving. Similarly, a hydroxide ion, which can be considered to be a "proton hole" in the water, serves as a landing point for a proton from another H_2O molecule, so that the OH^- ion hops about in the same way.

Because hydronium- and hydroxide ions can "move without actually moving" and thus without having to plow their way through the solution by shoving aside water molecules as do other ions, solutions which are acidic or alkaline have extraordinarily high *electrical conductivities*.

Acid-base reactions *à la* Brønsted

According to the Brønsted concept, the process that was previously written as a simple *dissociation* of a generic acid HA ("HA → H^+ + A^-)" is now an *acid-base reaction* in its own right:

$$HA + H_2O \rightarrow A^- + H_3O^+ \tag{10.3.2}$$

The idea, again, is that the proton, once it leaves the acid, must end up somewhere; it cannot simply float around as a free hydrogen ion.

Conjugate pairs

A reaction of an acid with a base is thus a *proton exchange reaction*; if the acid is denoted by AH and the base by B, then we can write a generalized acid-base reaction as

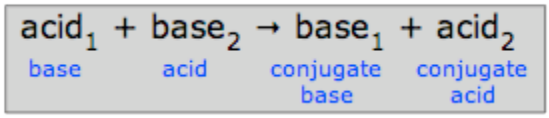
$$AH + B \rightarrow A^- + BH^+ \tag{10.3.3}$$

Notice that the *reverse* of this reaction,

$$BH^+ + A^- \rightarrow B + AH^+ \tag{10.3.4}$$

is also an acid-base reaction. Because *all* simple reactions can take place in both directions to some extent, it follows that transfer of a proton from an acid to a base must necessarily create a new pair of species that can, at least in principle, constitute an

acid-base pair of their own.



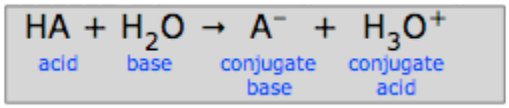
In this schematic reaction, $base_1$ is said to be *conjugate* to $acid_1$, and $acid_2$ is conjugate to $base_2$. The term *conjugate* means "connected with", the implication being that any species and its conjugate species are related by the gain or loss of one proton. The table below shows the conjugate pairs of a number of typical acid-base systems.

Some common conjugate acid-base pairs

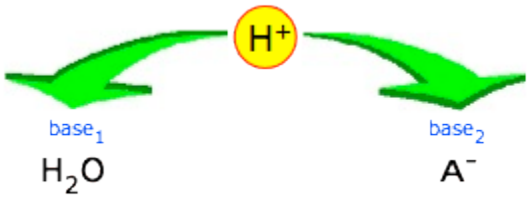
acid		base	
hydrochloric acid	HCl	Cl^-	chloride ion
acetic acid	CH_3CH_2COOH	$CH_3CH_2COO^-$	acetate ion
nitric acid	HNO_3	NO_3^-	nitrate ion
dihydrogen phosphate ion	$H_2PO_4^-$	HPO_4^-	monohydrogen phosphate ion
hydrogen sulfate ion	HSO_4^-	SO_4^{2-}	sulfate ion
hydrogen carbonate ("bicarbonate") ion	HCO_3^-	CO_3^{2-}	carbonate ion
ammonium ion	NH_4^+	NH_3	ammonia
iron(III) ("ferric") ion	$Fe(H_2O)_6^{3+}$	$Fe(H_2O)_5OH^{2+}$	
water	H_2O	OH^-	hydroxide ion
hydronium ion	H_3O+	H_2O	water

Strong acids and weak acids

We can look upon the generalized acid-base reaction



as a competition of two bases for a proton:



If the base H_2O overwhelmingly wins this tug-of-war, then the acid HA is said to be a *strong acid*. This is what happens with hydrochloric acid and the other common strong "mineral acids" H_2SO_4, HNO_3, and $HClO_4$:

hydrochloric acid	$HCl + H_2O \rightarrow \mathbf{Cl^- + H_3O^+}$

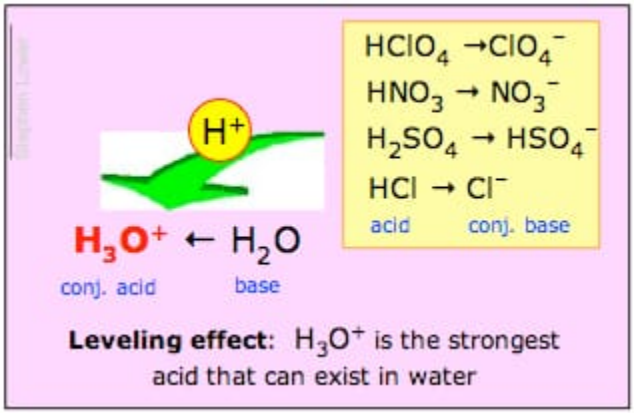
sulfuric acid	$H_2SO_4 + H_2O \rightarrow \mathbf{HSO_4^- + H_3O^+}$
nitric acid	$HNO_3 + H_2O \rightarrow \mathbf{NO_3^- + H_3O^+}$
perchloric acid	$HClO_4 + H_2O \rightarrow \mathbf{ClO_4^- + H_3O^+}$

Solutions of these acids in water are really solutions of the ionic species shown in heavy type on the right. This being the case, it follows that what we call a 1 M solution of "hydrochloric acid" in water, for example, does not really contain a significant concentration of HCl at all; the only real a acid present in such a solution is H_3O^+!

1. H_3O^+ is the **strongest** acid that can exist in water;
2. All strong acids appear to be equally strong in water.

The Leveling Effect

The second of these statements is called the *leveling effect*. It means that although the inherent proton-donor strengths of the strong acids differ, they are all completely dissociated in water. Chemists say that their strengths are "leveled" by the solvent water.



A comparable effect would be seen if one attempted to judge the strengths of several adults by conducting a series of tug-of-war contests with a young child. One would expect the adults to win overwhelmingly on each trial; their strengths would have been "leveled" by that of the child.

Weak acids

Most acids, however, are able to hold on to their protons more tightly, so only a small fraction of the acid is dissociated. Thus hydrocyanic acid, HCN, is a weak acid in water because the proton is able to share the lone pair electrons of the cyanide ion CN^- more effectively than it can with those of H_2O, so the reaction

$$HCN + H_2O \rightarrow H_3O^+ + CN^- \tag{10.3.5}$$

proceeds to only a very small extent. Since a strong acid binds its proton only weakly, while a weak acid binds it tightly, we can say that

> Strong acids are "weak" and weak acids are "strong." If you are able to explain this apparent paradox, you understand one of the most important ideas in acid-base chemistry!

Examples of proton donor-acceptor reactions

reaction	acid	base	conjugate acid	conjugate base
autoionization of water H_2O	H_2O	H_2O	H_3O^+	OH^-
ionization of hydrocyanic acid HCN	HCN	H_2O	H_3O^+	CN^-
ionization of ammonia NH^3 in water	NH_3	H_2O	NH_4^+	OH^-
hydrolysis of ammonium chloride NH^4Cl	NH_4^+	H_2O	H_3O^+	NH_3
hydrolysis of sodium acetate $CH3COO^- Na^+$	H_2O	CH_3COO^-	CH_3COOH	OH^-
neutralization of HCl by NaOH	HCl	OH^-	H_2O	Cl^-
neutralization of NH^3 by acetic acid	CH_3COOH	NH_3	NH_4^+	CH_3COO^-
dissolution of BiOCl (bismuth oxychloride) by HCl	$2\ H_3O^+$	BiOCl	$Bi(H_2O)^{3+}$	H_2O, Cl^-
decomposition of $Ag(NH3)^{2+}$ by HNO^3	$2\ H_3O^+$	$Ag(NH_3)_2^+$	NH_4^+	H_2O
displacement of HCN by CH^3COOH	CH_3COOH	CN^-	HCN	CH_3COO^-

Strong acids have weak conjugate bases

This is just a re-statement of what is implicit in what has been said above about the distinction between strong acids and weak acids. The fact that HCl is a strong acid implies that its conjugate base Cl^- is too weak a base to hold onto the proton in competition with either H_2O or H_3O^+. Similarly, the CN^- ion binds strongly to a proton, making HCN a weak acid.

Salts of weak acids give alkaline solutions

The fact that HCN is a weak acid implies that the cyanide ion CN^- reacts readily with protons, and is thus is a relatively good base. As evidence of this, a salt such as KCN, when dissolved in water, yields a slightly alkaline solution:

$$CN^- + H_2O \rightarrow HCN + OH^-$$

This reaction is still sometimes referred to by its old name *hydrolysis* ("water splitting"), which is literally correct but tends to obscure its identity as just another acid-base reaction. Reactions of this type take place only to a small extent; a 0.1M solution of KCN is still, for all practical purposes, 0.1M in cyanide ion.

In general, the weaker the acid, the more alkaline will be a solution of its salt. However, it would be going to far to say that "ordinary weak acids have strong conjugate bases." The only really strong base is hydroxide ion, OH^-, so the above statement would be true only for the *very* weak acid H_2O.

Strong bases and weak bases

The only really strong bases you are likely to encounter in day-to-day chemistry are alkali-metal hydroxides such as NaOH and KOH, which are essentially solutions of the hydroxide ion. Most other compounds containing hydroxide ions such as $Fe(OH)_3$ and $Ca(OH)_2$ are not sufficiently soluble in water to give highly alkaline solutions, so they are not usually thought of as strong bases.

There are actually a number of bases that are stronger than the hydroxide ion — best known are the *oxide ion* O^{2-} and the *amide ion* NH_2^-, but these are so strong that they can rob water of a proton:

$$O^{2-} + H_2O \rightarrow 2\ OH^-$$

$NH_2^- + H_2O \rightarrow NH_3 + OH^-$

This gives rise to the same kind of leveling effect we described for acids, with hydroxide ion as the strongest base in water.

> Hydroxide ion is the strongest base that can exist in aqueous solution.

The most common example of this is ammonium chloride, NH_4Cl, whose aqueous solutions are distinctly acidic:

$NH_4^+ + H_2O \rightarrow NH_3 + H_3O^+$

Because this (and similar) reactions take place only to a small extent, a solution of ammonium chloride will only be slightly acidic.

Autoprotolysis

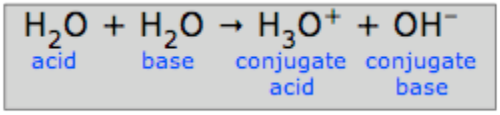
From some of the examples given above, we see that water can act as an acid

$CN^- + H_2O \rightarrow HCN + OH^-$

and as a base

$NH_4^+ + H_2O \rightarrow NH_3 + H_3O^+$

If this is so, then there is no reason why "water-the-acid" cannot donate a proton to "water-the-base":



This reaction is known as the *autoprotolysis of water*.

Chemists still often refer to this reaction as the "dissociation" of water and use the Arrhenius-style equation $H_2O \rightarrow H^+ + OH^-$as a kind of shorthand. As discussed in the previous lesson, this process occurs to only a tiny extent. It does mean, however, that hydronium and hydroxide ions are present in any aqueous solution.

Other liquids also exhibit autoprotolysis with the most well-known example is liquid ammonia:

$2\ NH_3 \rightarrow NH_4^+ + NH_2^-$

Even pure liquid sulfuric acid can play the game:

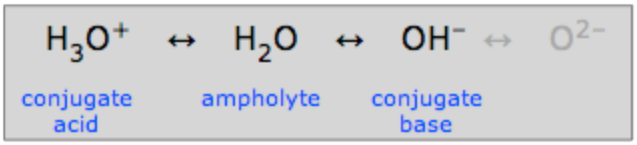
$2\ H_2SO_4 \rightarrow H_3SO_4^+ + HSO_4^-$

Each of these solvents can be the basis of its own acid-base "system", parallel to the familiar "water system".

Ampholytes

Water, which can act as either an acid or a base, is said to be *amphiprotic*: it can "swing both ways". A substance such as water that is amphiprotic is called an *ampholyte*.

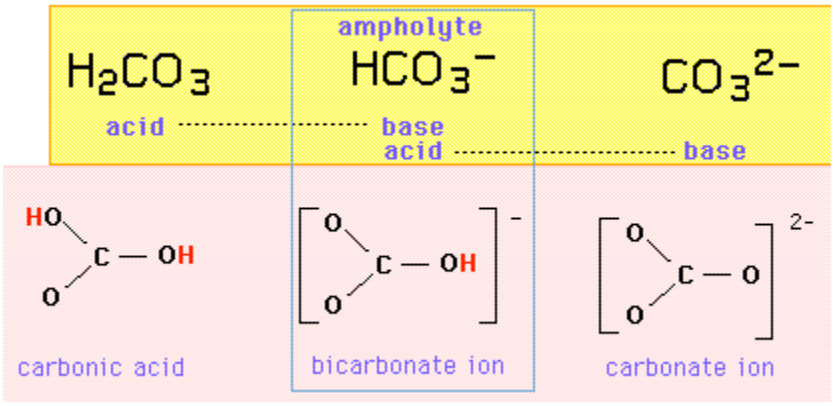
As indicated here, the hydroxide ion can also be an ampholyte, but not in aqueous solution in which the oxide ion cannot exist.



It is of course the amphiprotic nature of water that allows it to play its special role in ordinary aquatic acid-base chemistry. But many other amphiprotic substances can also exist in aqueous solutions. Any such substance will always have a conjugate acid and a conjugate base, so if you can recognize these two conjugates of a substance, you will know it is amphiprotic.

The carbonate system

For example, the triplet set {carbonic acid, bicarbonate ion, carbonate ion} constitutes an amphiprotric series in which the bicarbonate ion is the ampholyte, differing from either of its neighbors by the addition or removal of one proton:



If the bicarbonate ion is both an acid and a base, it should be able to exchange a proton with itself in an autoprotolysis reaction:

$$HCO_3^- + HCO_3^- \rightarrow H_2CO_3 + CO_3^{2-} \tag{10.3.6}$$

Your very life depends on the above reaction! CO_2, a metabolic by-product of every cell in your body, reacts with water to form carbonic acid H_2CO_3 which, if it were allowed to accumulate, would make your blood fatally acidic. However, the blood also contains carbonate ion, which reacts according to the reverse of the above equation to produce bicarbonate which can be safely carried by the blood to the lungs. At this location the autoprotolysis reaction runs in the forward direction, producing H_2CO_3 which loses water to form CO_2 which gets expelled in the breath. The carbonate ion is recycled back into the blood to eventually pick up another CO_2 molecule.

If you can write an autoprotolysis reaction for a substance, then that substance is amphiprotic.

This page titled 10.3: Acid-base reactions à la Brønsted is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10.4: Acid-Base Reactions

Learning Objectives

It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Sketch out a *proton-energy diagram* for a *strong acid*, a *weak acid or base*, and for a *strong base*.
- Describe how the *pH* affects the relative concentrations of a *conjugate acid-base pair*.

Will this acid react with that base? And if so, to what extent? These questions can be answered quantitatively by carrying out the detailed equilibrium calculations you will learn about in another lesson. However, modern acid-base chemistry offers a few simple principles that can enable you to make a qualitative decision at a glance. More importantly, the ideas which we develop in this section are guaranteed to give you a far better conceptual understanding of proton-based acid-base reactions in general.

Which base gets the proton?

Will acid HA react with base B? We stated above that the outcome of any acid-base reaction depends on how well two different bases can compete in the tug-of-war for the proton

$$A^- \leftarrow H^+ \rightarrow B^- \tag{10.4.1}$$

The proton will always go to the stronger base. Some insight into this can be had by thinking of the proton as having different potential energies when it is bound to different acceptors. We can draw a useful analogy with the electrons in an atom, which, you will recall, will always fall into the lowest-potential energy orbitals available, filling them from the bottom up. In a similar way, protons will "fall" into the lowest-energy empty spots (bases) they can find.

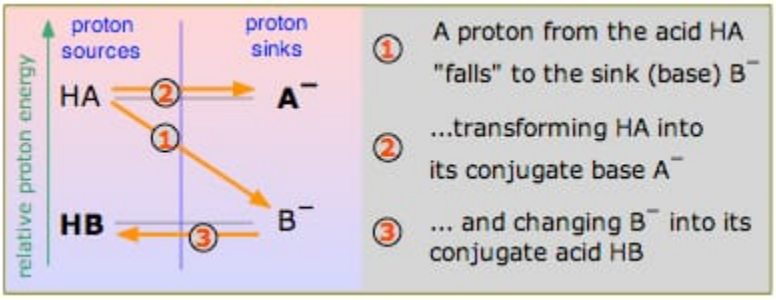


Figure 10.4.1: A generalized proton soure-sink acid-base reaction.

Consider the scheme shown here, which depicts two hypothetical acid-base conjugate pairs. Take careful note of the labeling of this diagram: the acids HA and HB are *proton sources* and the conjugate bases A^- and B^- are *proton sinks*. This "source-sink" terminology is synonymous with the "donor-acceptor" language that Brønsted taught us, but it also carries an implication about the relative energies of the proton as it exists in the two molecules HA and HB. If, as is indicated here, the proton is at a higher "potential energy" when it is in the form of HA than in HB, the reaction HA + B^- → HB + A^- will be favored compared to the reverse process HB + A^- → HA + B^-, which would require elevating the proton up to the A^- level. In this example, HA is the stronger acid because its proton can fall to a lower potential energy when it joins with B^- to form HB.

We will refer to diagrams such as the one in Figure 10.4.1 as "proton-energy diagrams", which is not quite correct, but we do not want to get into thermodynamics at this point. (If you already know something about chemical thermodynamics, we are really referring to *Gibbs energy*.)

It follows, then, that if we can arrange all the common acid-base conjugate pairs on this kind of a scale, we can predict the direction of any simple acid-base reaction without resorting to numbers. This will be illustrated further on, but in order to keep things simple, let's look at a few proton-energy diagrams that illustrate some of the acid-base chemistry that we discussed in the

preceding section.

Strong acids and weak acids

The hydronium ion is the dividing line; a strong acid, you will recall, is one whose conjugate base A^- loses out to the "stronger" base H_2O in the competition for the proton:

$$A^- \leftarrow H^+ \rightarrow H_2O \tag{10.4.2}$$

An acid that is a stronger proton donor than hydronium ion is a "strong" acid; if it is a weaker proton donor than H_3O^+, it is by definition "weak".This is seen most clearly in the diagram here, which contrasts the strong acid HA with the weak acid HB. HB "dissociates" to only a tiny extent because it is energetically unfavorable to promote its proton up to the H_2O-H_3O^+ level (process **3** in the diagram).

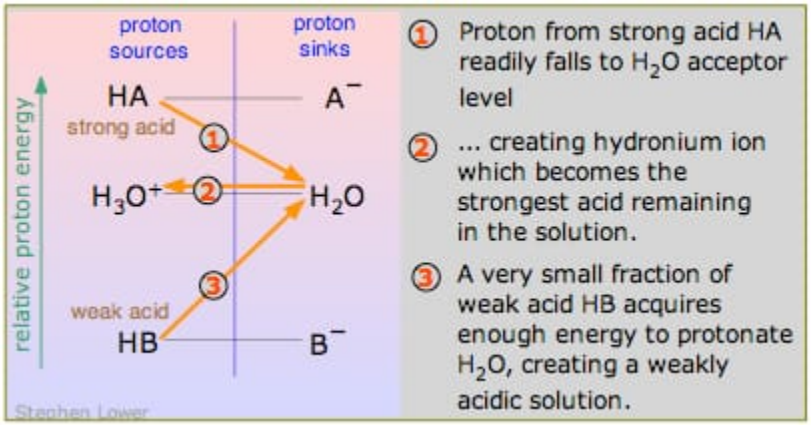


Figure 10.4.2: proton-energy diagram for strong and weak acids

Strong acids and leveling

A strong acid, you will recall, is one whose conjugate base A^- loses out to the "stronger" base H_2O in the competition for the proton:

$$A^- \leftarrow H^+ \rightarrow H_2O \tag{10.4.3}$$

Because the reaction

$$HA + H_2O \rightarrow A^- + H_3O^+ \tag{10.4.4}$$

for *any* strong acid HA is virtually complete, all strong acids appear to be equally strong in water (the leveling effect.)

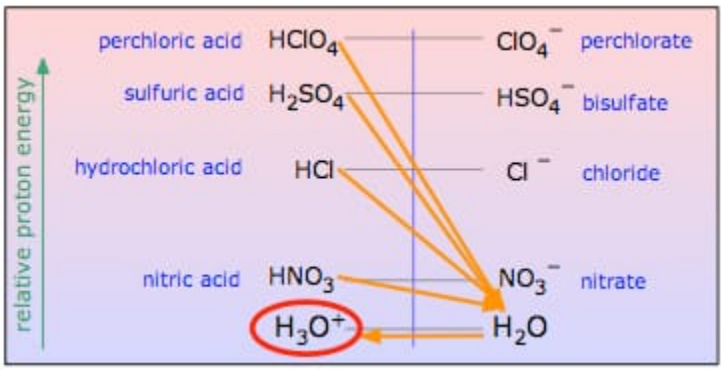


Figure 10.4.3: proton-energy diagram of leveling effect

From the proton-energy standpoint, a strong acid is one in which the energy of the proton is substantially greater when attached to the anion A^-than when it is attached to H_2O. Adding a strong acid HA to water will put it in contact with a huge proton sink that drains off the protons from any such acid, leaving the conjugate base A^- along with hydronium ion, the strongest acid that can exist in water.

Weak bases

Conjugate bases of weak acids tend to accept protons from water, leaving a small excess of OH^- ions and thus an alkaline solution. As you can see in the diagram, the weak base ammonia accepts a proton from water:

$$NH_3 + H_2O \rightarrow NH_4^+ + OH^- \tag{10.4.5}$$

The "weakness" of such a base is a consequence of the energetically unfavorable process (**1**) in which a proton must be raised up from the low-lying H_2O-OH^- level. From the standpoint of the "proton sources" column on the left, you can think of this as similar to the situation for weak acids that we discussed above; it can be considered a special case in which the weak acid is H_2O.

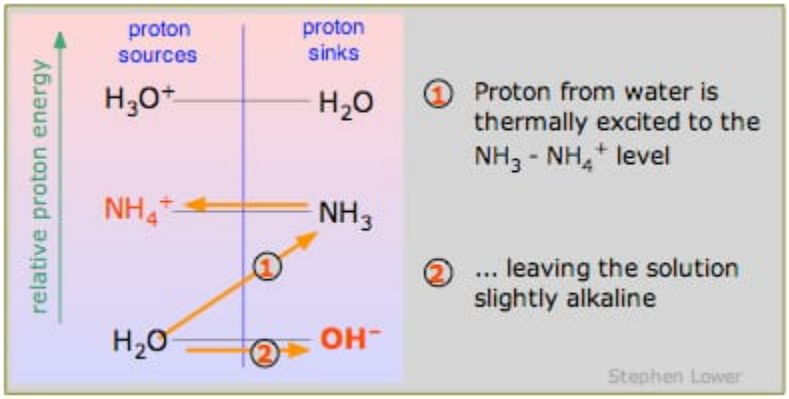


Figure 10.4.5: proton-energy diagram for weak base

For a very long time, chemists had regarded methane, CH_4, as the weakest acid, making the *methide ion* CH_3^- (which is also the simplest *carbanion*) the strongest base. Methane still holds its position as the weakest acid, but in 2008, the ion $LiO^-$was found to be an even stronger base than CH_4^-. Because both of these bases are observable only in the gas phase, these facts have little obvious import on aqueous-solution chemistry.

Autoprotolysis

Because water is amphiprotic, one H_2O molecule can donate a proton to another, as explained above. In this case the proton has to acquire considerable energy to make the jump (**1**) from the H_2O-OH^- level to the H_3O^+-H_2O level, so the reaction

$$2H_2O \rightarrow H_3O^+ + OH^- \tag{10.4.6}$$

occurs only to a minute extent. Think of this as the special case of the "weakest" acid H_2O reacting with the "weakest" base H_2O.

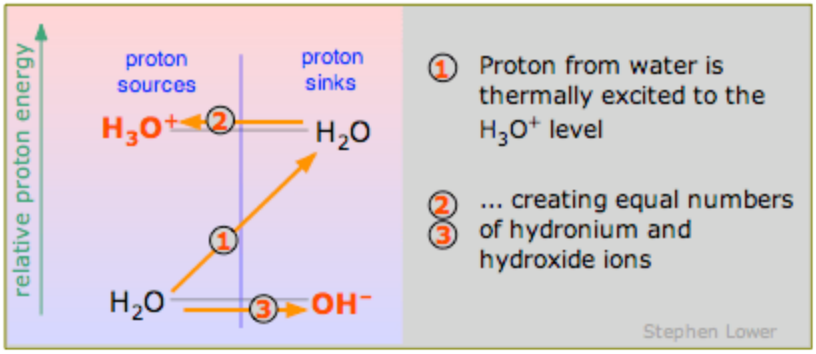


Figure 10.4.6: proton-energy diagram of autoprotolysis

Strong bases

Finally, what is a strong base? Just as a strong acid lies above the H_3O^+-H_2O level, so does a strong base lie below the H_2O-OH^- level. And for the same reason that H_3O^+ is the strongest *acid* that can exist in water, OH^- is the strongest *base* that can exist in water. The example of the oxide ion O^{2-} is shown here. Sodium oxide Na_2O is a white powder that dissolves in water to give oxide ions which immediately decompose into hydroxide ions

$$O^{2-} + H_2O \rightarrow 2OH^- \tag{10.4.7}$$

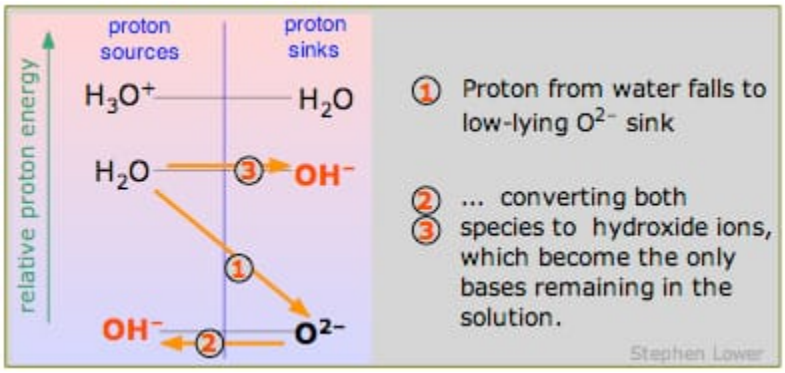


Figure 10.4.7: proton-energy diagram of strong base

Putting it all together, and the meaning of pH

This table combines common examples covering the entire range of acid-base strengths, from the strong to the very weak. The energy scale at the left gives you some idea of the relative proton-energy levels for each conjugate pair; notice that the zero is arbitrarily set to that of the H_3O^+-H_2O pair.

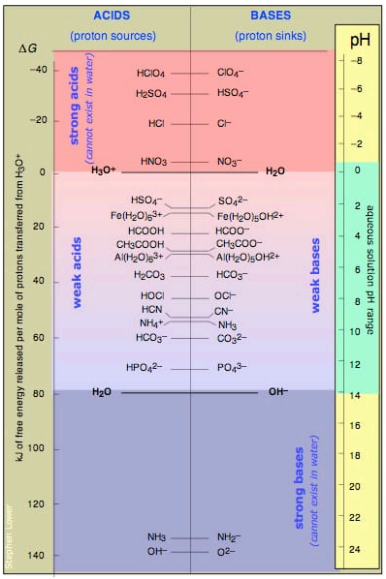


Figure 10.4.8: proton-free energy diagram for common acids and bases and pH

Of more importance is the pH scale on the right. The pH that corresponds to any conjugate pair is the pH at which equal concentrations of that pair are in their acid and base forms. For example, acetic acid CH_3COOH is "half ionized" at a pH of 4.7. If another strong acid such as HCl is added so as to reduce the pH, the proportion of acetate ion decreases, while if sodium hydroxide is added to force the pH higher, a larger fraction of the acetic acid will be "dissociated".

This illustrates another aspect of pH: at its most fundamental level, pH is an inverse measure of the "proton intensity" in the solution. The lower the pH, the higher the proton intensity, and the greater will be the fraction of higher-energy proton levels populated— which translates to higher acid-to-conjugate base concentration ratios. It is easy to see why acids such as H_2SO_4 and bases such as the amide ion NH_2^- cannot exist in aqueous solution; the pH would have to be at the impossible level of –6 for the former and +23 for the latter!

Why acids are titrated with hydroxide ion

When you titrate an acid with a base, you want virtually every molecule of the acid to react with the base. In the case of a weak acid such as hypochlorous, the reaction would be

$$HOCl + OH^- \rightarrow OCl^- + H_2O \tag{10.4.8}$$

Because the proton level in HOCl is considerably above that in H_2O, titration with NaOH solution will ensure that every last proton is eaten up by the hydroxide ion. If, instead, you used ammonia NH_3 as a titrant, the closeness of the two proton levels would cause the reaction to be incomplete, yielding a less distinct equivalence point. And, of course, titration with a base that is weaker then hypochlorite ion (such as sodium bicarbonate) would be hopeless.

As a practical matter, you can usually estimate that when the pH differs by more than about two units from the value that corresponds to the conjugate-pair for a monoprotic acid, the concentration of the non-favored species will be down by a factor of around 1000.

Contributor

- Stephen Lower, Professor Emeritus (Simon Fraser U.) Chem1 Virtual Textbook

This page titled 10.4: Acid-Base Reactions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10.5: Lewis Acids and Bases

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Write the equation for the proton transfer reaction involving a Brønsted-Lowry acid or base, and show how it can be interpreted as an electron-pair transfer reaction, clearly identifying the donor and acceptor.
- Give an example of a *Lewis acid-base reaction* that does not involve protons.
- Write equations illustrating the behavior of a given non-aqueous acid-base system.

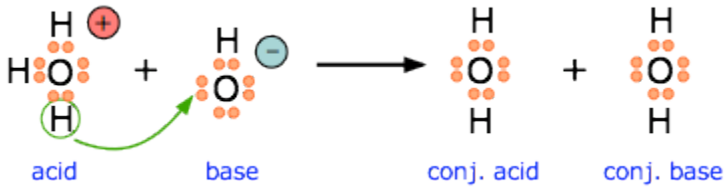
The Brønsted-Lowry proton donor-acceptor concept has been one of the most successful theories of Chemistry. But as with any such theory, it is fair to ask if this is not just a special case of a more general theory that could encompass an even broader range of chemical science. In 1916, G.N. Lewis of the University of California proposed that the *electron pair* is the dominant actor in acid-base chemistry. The Lewis theory did not become very well known until about 1923 (the same year that Brønsted and Lowry published their work), but since then it has been recognized as a very powerful tool for describing chemical reactions of widely different kinds and is widely used in organic and inorganic chemistry. According to Lewis,

- An *acid* is a substance that **accepts** a pair of electrons, and in doing so, forms a covalent bond with the entity that supplies the electrons.
- A *base* is a substance that **donates** an unshared pair of electrons to a recipient species with which the electrons can be shared.

In modern chemistry, electron donors are often referred to as nucleophiles, while acceptors are electrophiles.

Proton-Transfer Reactions Involve Electron-Pair Transfer

Just as any Arrhenius acid is also a Brønsted acid, any Brønsted acid is also a Lewis acid, so the various acid-base concepts are all "upward compatible". Although we do not really need to think about electron-pair transfers when we deal with ordinary aqueous-solution acid-base reactions, it is important to understand that it is the opportunity for electron-pair sharing that enables proton transfer to take place.

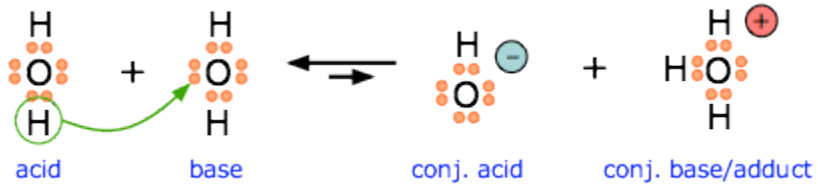


This equation for a simple acid-base neutralization shows how the Brønsted and Lewis definitions are really just different views of the same process. Take special note of the following points:

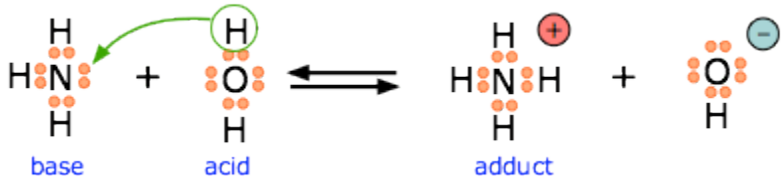
- The arrow shows the movement of a proton from the hydronium ion to the hydroxide ion.
- Note carefully that the electron-pairs themselves do not move; they remain attached to their central atoms. The electron pair on the base is "donated" to the acceptor (the proton) only in the sense that it ends up being *shared* with the acceptor, rather than being the exclusive property of the oxygen atom in the hydroxide ion.
- Although the hydronium ion is the nominal Lewis acid here, it does not itself accept an electron pair, but acts merely as the source of the proton that coordinates with the Lewis base.

The point about the electron-pair remaining on the donor species is especially important to bear in mind. For one thing, it distinguishes a *Lewis acid-base reaction* from an *oxidation-reduction reaction*, in which a physical transfer of one or more electrons from donor to acceptor does occur. The product of a Lewis acid-base reaction is known formally as an "adduct" or

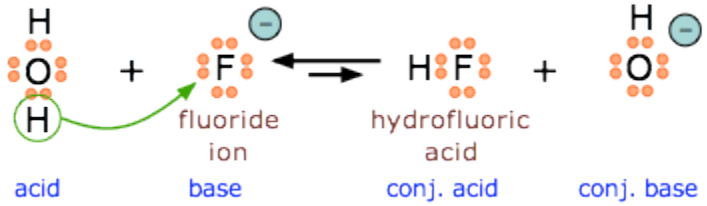
"complex", although we do not ordinarily use these terms for simple proton-transfer reactions such as the one in the above example. Here, the proton combines with the hydroxide ion to form the "adduct" H_2O. The following examples illustrate these points for some other proton-transfer reactions that you should already be familiar with.



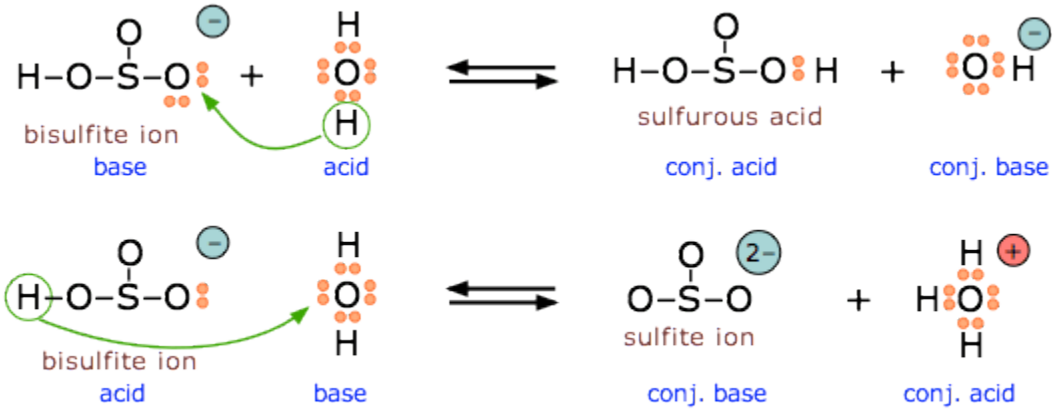
Another example, showing the autoprotolysis of water. Note that the conjugate base is also the adduct.



Ammonia is both a Brønsted and a Lewis base, owing to the unshared electron pair on the nitrogen. The reverse of this reaction represents the *hydrolysis* of the ammonium ion.



Because HF is a weak acid, fluoride salts behave as bases in aqueous solution. As a Lewis base, F^- accepts a proton from water, which is transformed into a hydroxide ion.



The bisulfite ion is *amphiprotic* and can act as an electron donor or acceptor.

Acid-base Reactions without Transferring Protons

The major utility of the Lewis definition is that it extends the concept of acids and bases beyond the realm of proton transfer reactions. The classic example is the reaction of boron trifluoride with ammonia to form an *adduct*:

$$BF_3 + NH_3 \rightarrow F_3B{-}NH_3 \tag{10.5.1}$$

One of the most commonly-encountered kinds of Lewis acid-base reactions occurs when electron-donating ligands form coordination complexes with transition-metal ions.

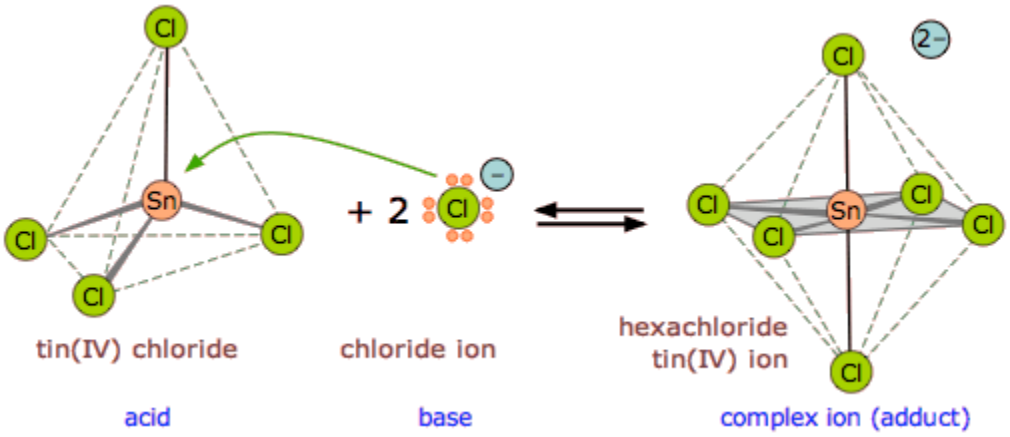


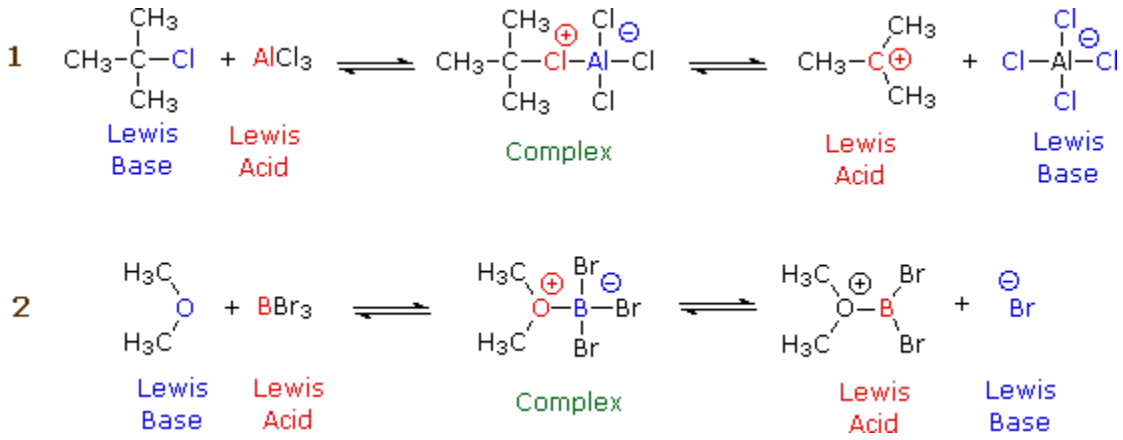
Figure 10.5.1: The tin atom in $SnCl_4$ can expand its valence shell by utilizing a pair of d-orbitals, changing its hybridization from sp^3 to sp^3d^2.

Exercise 10.5.1

Here are several more examples of Lewis acid-base reactions that **cannot** be accommodated within the Brønsted or Arrhenius models. Identify the Lewis acid and Lewis base in each reaction.

a. $Al(OH)_3 + OH^- \rightarrow Al(OH)_4{}^-$
b. $SnS_2 + S^{2-} \rightarrow SnS_3{}^{2-}$
c. $Cd(CN)_2 + 2\,CN^- \rightarrow Cd(CN)_4{}^{2+}$
d. $AgCl + 2\,NH_3 \rightarrow Ag(NH_3)_2{}^+ + Cl^-$
e. $Fe^{2+} + NO \rightarrow Fe(NO)^{2+}$
f. $Ni^{2+} + 6\,NH_3 \rightarrow Ni(NH_3)_5{}^{2+}$

Although organic chemistry is beyond the scope of these lessons, it is instructive to see how electron donors and acceptors play a role in chemical reactions. The following two diagrams show the mechanisms of two common types of reactions initiated by simple inorganic Lewis acids:



In each case, the species labeled "Complex" is an intermediate that decomposes into the products, which are conjugates of the original acid and base pairs. The electric charges indicated in the complexes are formal charges, but those in the products are

"real".

In reaction 1, the incomplete octet of the aluminum atom in $AlCl_3$ serves as a better electron acceptor to the chlorine atom than does the isobutyl part of the base. In reaction 2, the pair of non-bonding electrons on the dimethyl ether coordinates with the electron-deficient boron atom, leading to a complex that breaks down by releasing a bromide ion.

Non-aqueous Protonic Acid-Base Systems

We ordinarily think of Brønsted-Lowry acid-base reactions as taking place in aqueous solutions, but this need not always be the case. A more general view encompasses a variety of acid-base *solvent systems*, of which the *water system* is only one (Table 10.5.1). Each of these has as its basis an amphiprotic solvent (one capable of undergoing autoprotolysis), in parallel with the familiar case of water.

The ammonia system is one of the most common non-aqueous system in Chemistry. Liquid ammonia boils at –33° C, and can conveniently be maintained as a liquid by cooling with dry ice (–77° C). It is a good solvent for substances that also dissolve in water, such as ionic salts and organic compounds since it is capable of forming hydrogen bonds. However, many other familiar substances can also serve as the basis of protonic solvent systems as Table 10.5.1 indicates:

Table 10.5.1: Popular Solvent systems

solvent	autoprotolysis reaction	pK_{ap}
water	$2\ H_2O \rightarrow H_3O^+ + OH^-$	14
ammonia	$2\ NH_3 \rightarrow NH_4^+ + NH_2^-$	33
acetic acid	$2\ CH_3COOH \rightarrow CH_3COOH_2^+ + CH_3COO^-$	13
ethanol	$2\ C_2H_5OH \rightarrow C_2H_5OH_2^+ + C_2H_5O^-$	19
hydrogen peroxide	$2\ HO\text{-}OH \rightarrow HO\text{-}OH_2^+ + HO\text{-}O^-$	13
hydrofluoric acid	$2\ HF \rightarrow H_2F^+ + F^-$	10
sulfuric acid	$2\ H_2SO_4 \rightarrow H_3SO_4^+ + HSO_4^-$	3.5

One use of nonaqueous acid-base systems is to examine the relative strengths of the strong acids and bases, whose strengths are "leveled" by the fact that they are all totally converted into H_3O^+ or OH^- ions in water. By studying them in appropriate non-aqueous solvents which are poorer acceptors or donors of protons, their relative strengths can be determined.

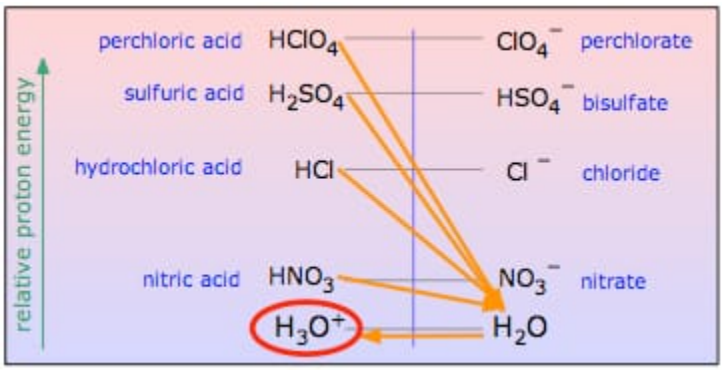


Figure 10.5.2: Use of non-aqueous solvents allows the study of strong acids that are hindered by the "leveling" of the solvent.

This page titled 10.5: Lewis Acids and Bases is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10.6: Types of Acids and Bases

Learning Objectives

- Given the formula of a **binary compound of hydrogen** with an element of Z<18, predict whether its aqueous solution be acidic or basic, and write an appropriate equation.
- Do the same for an **oxygen- or hydroxy compound** of a similar element.
- Define an **oxyacid**, and explain why many of these are very **strong acids**.
- Write an equation describing the **amphoteric** nature of zinc or aluminum hydroxide.
- Define an **acid anhydride**, and write an equation describing its behavior.
- Explain how **metal cations** can give acidic solutions.
- Write equations showing why aqueous solutions of some **salts** are acidic, while others are alkaline.
- Write the formulas of an **organic acid** and an **organic base**, and write an equation showing why the latter gives an alkaline solution in water.

You will already have noticed that not every compound that contains hydrogen atoms is acidic; ammonia NH_3, for example, gives an alkaline aqueous solution. Similarly, some compounds containing the group -OH are basic, but others are acidic. An important part of understanding chemistry is being able to recognize what substances will exhibit acidic and basic properties in aqueous solution. Fortunately, most of the common acids and bases fall into a small number of fairly well-defined groups, so this is not particularly difficult.

Binary Hydrides

Strictly speaking, the term *hydride* refers to ionic compounds of hydrogen with the electropositive metals of Groups 1-2; these contain the *hydride ion*, H^-, and are often referred to as "true" hydrides. However, the term is often used in its more general sense to refer to any binary compound MH_n in which M stands for any element. The hydride ion is such a strong base that it cannot exist in water, so salts such as sodium hydride react with water to yield hydrogen gas and an alkaline solution:

$$NaH + H_2O \rightarrow Na^+ + OH^- + H_2 \tag{10.6.1}$$

The more electronegative elements form covalent hydrides which generally react as acids, a well-known example being hydrogen chloride, a gas which dissolves readily in water to give the solution we know as hydrochloric acid

$$HCl_{(g)} + H_2O_{(l)} \rightarrow H_3O^+ + Cl^- \tag{10.6.2}$$

Most of the covalent hydrogen compounds are weak acids— in some cases, such as methane, CH_4, so weak that their acidic properties are rarely evident. Many, such as H_2O and NH_3, are amphiprotic. The latter compound, ammonia, is a weaker acid than H_2O, so it exhibits basic properties in water

$$NH_3 + H_2O \rightarrow NH_4^+ + OH^- \tag{10.6.3}$$

but behaves as an acid in non-aqueous solvents such as liquid ammonia itself:

$$NH_3 + NH_3 \rightarrow NH_4^+ + NH_2^- \tag{10.6.4}$$

In general, the acidity of the non-metallic hydrides increases with the atomic number of the element to which it is connected. Thus as the element M moves from left to right across the periodic table or down within a group, the acids MH become stronger, as indicated by the acid dissociation constants shown at the right. Note that

- The formulas shown in red represent "strong" acids (that is, acids stronger than H_3O^+.)
- Hydrofluoric acid is the only weak member of the hydrohalogen acids.
- Acids weaker than water do not behave as acids in aqueous solution. Thus for most practical purposes, methane and ammonia are not commonly regarded as acids. H_2O itself is treated as an acid only in the narrow context of aqueous solution chemistry.

CH_4 10^{-46}	NH_3 10^{-35}	H_2O 10^{-16}	HF 10^{-3}
	PH_3 10^{-27}	H_2S 10^{-7}	HCl 10^{7}
		H_2Te 10^{-3}	HBr 10^{9}
			HI 10^{10}

Attempts to explain these trends in terms of a single parameter such as the electronegativity of M tend not to be very useful. The reason is that acid strengths depend on a number of factors such as the strength of the M-H bond and the energy released when the resultant ions become hydrated in solution. This last factor plays a major role in making HF something of an anomaly amongst the strong acids of Group 17.



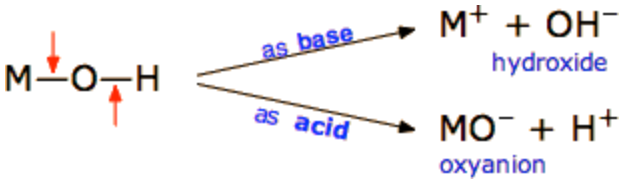
Ammonia is such a weak acid that its conjugate base, amide ion NH_2^-, cannot exist in water. In aqueous solution, NH_3 acts as a weak base, accepting a proton from water and leaving a OH^- ion. An aqueous solution of NH_3 is sometimes called “ammonium hydroxide”. This misnomer reflects the pre-Brønsted view that all bases contain –OH units that yield hydroxide ions on dissociation according to the Arrhenius scheme

$$NH_4OH \rightleftharpoons NH_4^+ + OH^- \tag{10.6.5}$$

A solution of ammonia in water is more correctly referred to as "aqueous ammonia" and represented by the formula $NH_3(aq)$. There is no physical evidence for the existence of NH_4OH, but the name seems to remain forever etched on reagent bottles in chemical laboratories and in the vocabularies of chemists.

Hydroxy Compounds

Compounds containing the *hydroxyl group* –OH constitute the largest category of acids, especially if the organic acids (discussed separately farther on) are included. M–OH compounds also include many of the most common bases.

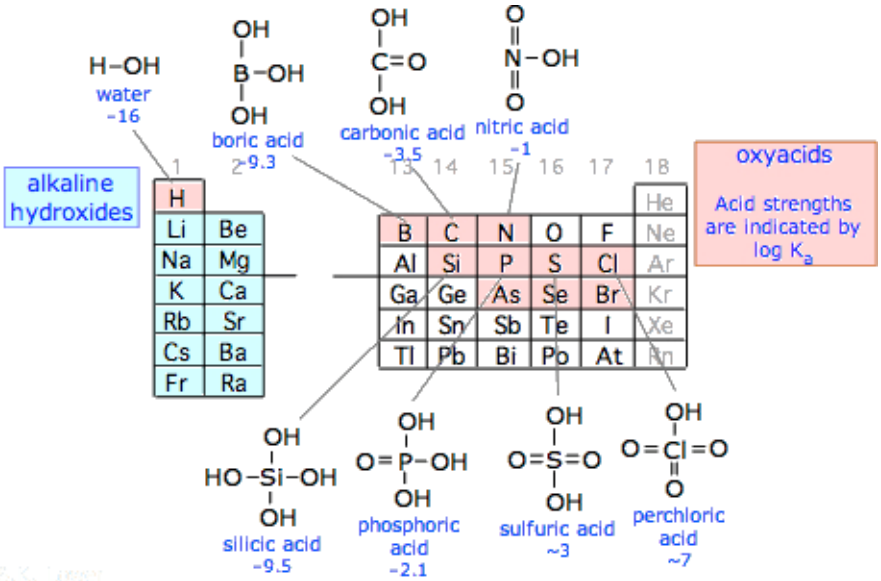


Whether a compound of the general type M–O–H will act as an acid or a base depends is influenced by the relative tendencies of the M–O and the O–H bonds to break apart in water. If the M–O bond cleaves more readily, then the –OH part will tend to retain its individuality and with its negative charge will become a hydroxide ion. If the O–H bond breaks, the MO-part of the molecule will remain intact as an *oxyanion* MO^- and release of the proton will cause the MOH compound to act as an acid.

This is not solely a matter of the relative strengths of the two bonds; the energy change that occurs when the resulting ions interact with water molecules is also an important factor.

In general, if M is a metallic element, the metal hydroxide compound MOH will be basic. The case of the highly electropositive elements of Groups 1 and 2 is somewhat special in that their solid MOH compounds exist as interpenetrating lattices of metal cations and OH^- ions, so those that can dissolve readily in water form strongly alkaline solutions; KOH and NaOH are well known examples of strong bases. From the Brønsted standpoint, these different "bases" are really just different sources for the single strong base OH^-. As one moves into Group 2 of the periodic table the M-OH compounds become less soluble; thus a saturated solution of $Ca(OH)_2$ (commonly known as *limewater*) is only weakly alkaline. Hydroxides of the metallic elements of the p-block and of the transition metals are so insoluble that their solutions are not alkaline at all. Nevertheless these solids dissolve readily in acidic solutions to yield a salt plus water, so they are formally bases.

The acidic character of hydroxy compounds of the nonmetals, known collectively as *oxyacids*, is attributed to the displacement of negative charge from the hydroxylic oxygen atom by the electronegative central atom. The net effect is to make the oxygen slightly more positive, thus easing the departure of the hydrogen as H^+ . The presence of other electron-attracting groups on the central atom has a marked effect on the strength of an oxyacid. Of special importance is the doubly-bonded oxygen atom. With the exception of the halogen halides, all of the common strong acids contain one or more of these oxygens, as in sulfuric acid $SO_2(OH)$, nitric acid $NO_2(OH)$ and phosphoric acid $PO(OH)_3$. In general the strengths of these acids depends more on the number of oxygens than on any other factor, so periodic trends are not so important.



Most of the halogen elements form more than one oxyacid. Fluorine is an exception; being more electronegative than oxygen, no oxyacids of this element are known. Chlorine is the only halogen for which all four oxyacids are known, and the K_a values for this series show how powerfully the Cl–O oxygen atoms affect the acid strength.

Oxygen compounds

Binary oxides that contain no hydrogen atoms can exhibit acid-base behavior when they react with water. The division between acidic and basic oxygen oxides largely parallels that between the hydroxy compounds. The oxygen compounds of the highly electropositive metals of Groups 1-2 actually contain the *oxide ion* O^-. This ion is another case of a proton acceptor that is stronger than OH^-, and thus cannot exist in aqueous solution. Ionic oxides therefore tend to give strongly alkaline solutions:

$$O^- + H_2O \longrightarrow 2\,OH^-\,(aq) \tag{10.6.6}$$

In some cases, such as that of MgO, the solid is so insoluble that little change in pH is noticed when it is placed in water. CaO, however, which is known as *quicklime*, is sufficiently soluble to form a strongly alkaline solution with the evolution of considerable heat; the result is the slightly-soluble *slaked lime*, $Ca(OH)_2$. Oxygen compounds of the transition metals are generally insoluble solids having rather complex extended structures. Although some will dissolve in acids, they display no acidic properties in water.

Amphoteric oxides and hydroxides

The oxides and hydroxides of the metals of Group 3 and higher tend to be only weakly basic, and most display an amphoteric nature. Most of these compounds are so slightly soluble in water that their acidic or basic character is only obvious in their reactions with strong acids or bases. In general, these compounds tend to be more basic than acidic; thus the oxides and hydroxides of aluminum, iron, and zinc all dissolve in mildly acidic solutions, whereas they require treatment with concentrated hydroxide ion solutions to react as acids.

$Al(OH)_3 + 3\,H^+ \rightarrow Al^{3+}(aq) + 3H_2O$	$Al(OH)_3(s) + OH^- \rightarrow Al(OH)_3^-\,3\,(aq)$
$Zn(OH)_2 + 3\,H^+ \rightarrow Zn^{2+}(aq) + 2H_2O$	$Zn(OH)_2(s) + 2\,OH^- \rightarrow Zn(OH)_3^{4-}(aq)$
$FeO(OH) + 3\,H^+ \rightarrow Fe^{3+}(aq) + 3H_2O$	$Fe_2O_3(s) + 3\,OH^- \rightarrow 2\,FeO_2{+}(aq) + 3\,H_2O$

The product ions in the second column are known as aluminate, zincate, and ferrate. Other products, in which only some of the –OH groups of the parent hydroxides are deprotonated, are also formed, so there are actually whole series of these oxyanions for most metals.

An *amphoteric* substance is one that can act as either an acid or a base. An *amphiprotic* substance can act as either a proton donor or a proton acceptor. So all amphiprotic compounds are also amphoteric. An example of an amphoteric compound that is not amphiprotic is ZnO, which can act as an acid even though it has no protons to donate:

$$ZnO\,(s) + 4\,OH^-\,(aq) \longrightarrow Zn(OH)_4^{2-}\,(aq) \tag{10.6.7}$$

As a base, it "accepts" protons but does not retain them:

$$ZnO\,(s) + 2\,H^+ \rightleftharpoons Zn^{2+} + H_2O \tag{10.6.8}$$

The same remarks can be made about the other compounds shown in the table above. For most practical purposes, the distinction between amphiprotic and amphoteric is not worth worrying about.

Acid anhydrides

The binary oxygen compounds of the non-metallic elements tend to produce acidic solutions when they are added to water. Such compounds are sometimes referred to as *acid anhydrides* (“acids without water”.)

$$CO_2 + H_2O \rightarrow H_2CO_3 \tag{10.6.9}$$

$$SO_2 + H_2O \rightarrow [H_2SO_3] \tag{10.6.10}$$

$$SO_3 + H_2O \rightarrow H_2SO_4 \tag{10.6.11}$$

$$P_4O_{10} + 6H_2O \rightarrow 4H_3PO_4 \tag{10.6.12}$$

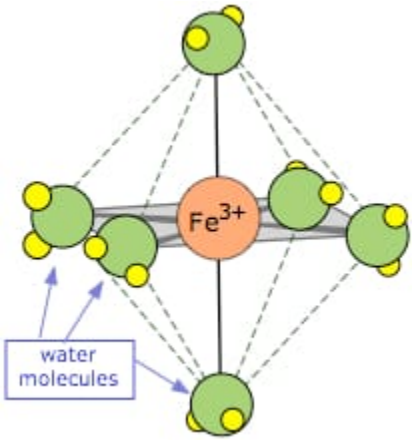
In some cases, the reaction involves more than simply incorporating the elements of water. Thus nitrogen dioxide, used in the commercial preparation of nitric acid, is not an anhydride in the strict sense:

$$3NO_2 + H_2O \rightarrow 2HNO + 3 + NO \tag{10.6.13}$$

Metal cations as acids

When sodium chloride is dissolved in pure water, the pH remains unchanged because neither ion reacts with water. However, a solution of magnesium chloride will be faintly acidic, and a solution of iron(III) chloride $FeCl_3$ will be distinctly so. How can this be? Since none of these cations contains hydrogen, we can only conclude that the protons come from the water.

The water molecules in question are those that find themselves close to any cation in aqueous solution; the positive field of the metal ion interacts with the polar H_2O molecule through ion-dipole attraction, and at the same time increases the acidity of these loosely-bound waters by making facilitating the departure H^+ ion. In general, the smaller and more highly charged the cation, the more acidic will it be; the acidity of the alkali metals and of ions like $Ag^+(aq)$ is negligible, but for more highly-charged ions such as Mg^{2+}, Pb^{2+} and Al^{3+}, the effect is quite noticeable.



Most of the transition-metal cations form organized coordination complexes in which four or six H_2O molecules are chemically bound to the metal ion where they are well within the influence of the coulombic field of the cation, and thus subject to losing a proton. Thus an aqueous solution of "Fe^{3+}" is really a solution of the ion *hexaaquo iron III*, whose first stage of "dissociation" can be represented as

$$Fe(H_2O)_6^{3+} + H_2O \rightarrow Fe(H_2O)_5(OH)^{2+} + H_3O^+ \tag{10.6.14}$$

As a consequence of this reaction, a solution of $FeCl_3$ turns out to be a stronger acid than an equimolar solution of acetic acid. A solution of $FeCl_2$, however, will be a much weaker acid; the +2 charge is considerably less effective in easing the loss of the proton.

It should be possible for a hydrated cation to lose more than one proton. For example, an $Al(H_2O)_6^{3+}$ ion should form, successively, the following species:

$AlOH(H_2O)_5^{2+} \rightarrow Al(OH)_2(H_2O)_4^+ \rightarrow Al(OH)_3(H_2O)_3^0 \rightarrow Al(OH)_4(H_2O)_2^- \rightarrow Al(OH)_5(H_2O)^{2-} \rightarrow Al(OH)_6^{3-}$

However, removal of protons becomes progressively more difficult as the charge decreases from a high positive value to a negative one; the last three species have not been detected in solution. In dilute solutions of aluminum chloride the principal species are actually $Al(H_2O)_6^{3+}$ (commonly represented simply as Al^{3+}) and $AlOH(H_2O)_5^{2+}$ ("$AlOH^{2+}$").

Salts

When salts dissolve in water, they yield solutions of anions and cations, so their effects on the pH of the solution will depend on the properties of the particular pair of ions. For a salt such as sodium chloride, the solution wil remain neutral because sodium ions have no acidic properties and chloride ions, being conjugate to the strong acid HCl have negligible proton-accepting tendencies. Ions of this kind are often referred to as "strong" ions (that is, derived from a strong acid and a strong base— HCl and NaOH in the case of NaCl.) The possible outcomes for the other three possibilities are shown below.

salt derived from	example	pH	reaction
weak acid + strong base	NaF	>7	$F^- + H_2O \rightarrow HF + OH^-$
strong acid + weak base	NH_4Cl	<7	$NH^{4+} + H_2O \rightarrow NH_3 + H_3O^+$
weak acid + weak base	NH_4F	?	depends on competition between above two reactions; need to do calculation

The reactions that cause salt solutions to have non-neutral pH values are sometimes still referred to by the older term *hydrolysis* ("water splitting")— a reminder of times before the concept of proton transfer acid-base reactions had developed.

Organic Acids and Bases

The *carboxyl group* –CO(OH) is the characteristic functional group of the organic acids. The acidity of the carboxylic hydrogen atom is due almost entirely to electron-withdrawal by the non-hydroxylic oxygen atom; if it were not present, we would have an alcohol –COH whose acidity is smaller even than that of H_2O. This partial electron withdrawal from one atom can affect not only a neighboring atom, but that atom's neighbor as well. Thus the strength of a carboxylic acid will be affected by the bonding environment of the carbon atom to which it is connected. This propagation of partial electron withdrawal through several adjacent atoms is known as the *inductive effect* and is extremely important in organic chemistry.

A very good example of the inductive effect produced by chlorine (another highly electronegative atom) is seen by comparing the strengths of acetic acid and of the successively more highly substituted chloroacetic acids:

$CH_3–COOH$ acetic acid 1.8×10^{-5}	$ClCH_2–COOH$ monochloroacetic acid 0.0014	$Cl_2CH–COOH$ dichloroacetic acid 0.055	$Cl_3C–COOH$ trichloroacetic acid 0.63

Phenols

The acidic character of the carboxyl group is really a consequence of the enhanced acidity of the –OH group as influenced by the second oxygen atom that makes up the –COOH group. The benzene ring has a similar although weaker electron-withdrawing effect, so hydroxyl groups that are attached to benzene rings also act as acids. The most well known example of such an acid is phenol, C_6H_5OH, also known as *carbolic acid*. Compared to carboxylic acids, phenolic acids are quite very weak, as indicated by the acid dissociation constants listed below:

$CH_3–COOH$ acetic acid 1.8×10^{-5}	$C_6H_5–OH$ phenol 1.1×10^{-10}	$C_6H_5–COOH$ benzoic acid 6.3×10^{-5}

Amines and organic bases

We have already discussed organic acids, so perhaps a word about organic bases would be in order. The –OH group, when bonded to carbon, is acidic rather than basic, so alcohols are not the analogs of the inorganic hydroxy compounds. The amines, consisting of the $–NH_2$ group bonded to a carbon atom, are the most common class of organic bases. Amines give weakly alkaline solutions

in water:

$$CH_3NH_2 + H_2O \rightarrow CH_3NH_3^+ + OH^- \tag{10.6.15}$$

Amines are end products of the bacterial degradation of nitrogenous organic substances such as proteins. They tend to have rather unpleasant “rotten fish” odors. This is no coincidence, since seafood contains especially large amounts of nitrogen-containing compounds which begin to break down very quickly. *Methylamine* CH_3NH_2, being a gas at room temperature, is especially apt to make itself known to us. Addition of lemon juice or some other acidic substance to fish will convert the methylamine to the methylaminium ion $CH_3NH_3^+$. Because ions are not volatile they have no odor.

This page titled 10.6: Types of Acids and Bases is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

10.7: Acid-Base Gallery

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Describe some of the special properties of *sulfuric acid* that make it especially important both in the laboratory and in industry.
- Name the major acids and bases that are important to the **fertilizer** industry
- Name the natural sources of any three of the major **organic** acids.
- What are *fatty acids*? In what major way do the physical properties of saturated and unsaturated fatty acids differ?
- Describe the general structure of an *amino acid*, and state why they are important.

Acids and bases are of interest not only to the chemically inclined; they play a major role in our modern industrial society — so anyone who participates in it, or who is interested in its history and development, needs to know something about them. Five of the major acids and bases fall into the "Top 20" industrial chemicals manufactured in the world. The following table shows year-2000 figures for the U.S:

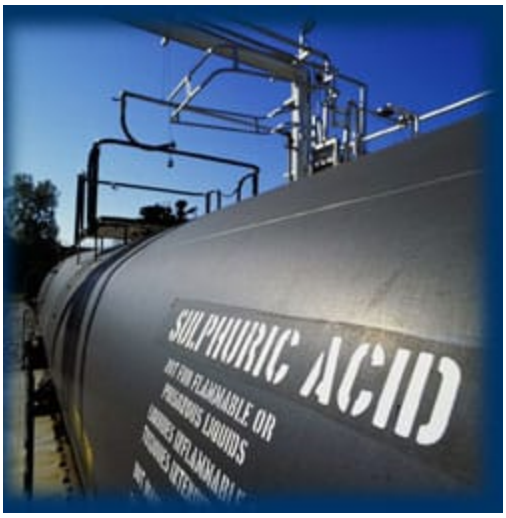
cal and rank	Sulfuric acid - 1	Lime (CaO) 3	Phosphoric acid - 4	Ammonia 5	Sodium hydroxide - 9	Nitric acid - 11
production in 10^9kg	40	20	16	15	11	8
major use	chemicals	cement	fertilizers	fertilizers	chemicals	chemicals

The mineral acids

This term refers to any inorganic acid, but its common use is usually limited to the major strong acids plus phosphoric acid. The major mineral acids— sulfuric, nitric, and hydrochloric— have been known since medieval times. Their discovery is usually credited to the Persian alchemist Abu Musa Jabir ibn Hayyan, known in the West by his Latinized name Geber. Jabir also invented *aqua regia*, the mixture of nitric and hydrochloric acids that has the unique ability to dissolve gold.

Sulfuric acid

More sulfuric acid is manufactured than any other industrial chemical, and it is the cheapest industrial acid worldwide. It has been continuously manufactured in the U.S. since 1793 and in Europe for much longer.



What you should know about it

- Pure anhydrous H_2SO_4 is a dense, viscous liquid which melts at 10.4°C and boils at about 300°C, decomposing back into its constituents, H_2O and SO_3.
- The acid undergoes **autoprotolysis**

$2\ H_2SO_4 \rightarrow H_3SO_4^+ + HSO_4^-$

$H_2SO_4 + 2\ NaCl \rightarrow 2\ Na^+ + SO_4^{2-} + HCl(g)$

$C_{12}H_{22}O_{11}(s) \rightarrow 12\ C(s) + 11\ H_2O$

$H_2SO_4 + H_2SO_4 \rightarrow HS_2O_7^- + H_3O^+$

- Its **high boiling point** makes the acid ideal for making other acids, such as nitric and hydrochloric, which are more volatile; removal of the gaseous product drives the reaction to the right as predicted by the *Le Chatelier Principle*:
- Sulfuric acid has a voracious appetite for water, and thus is an excellent **dehydrating agent**. This is seen most spectacularly if some concentrated acid is poured onto a small pile of table sugar; after a short time, a vigorous reaction ensues resulting in a pile of porous, steaming carbon
- Sulfuric acid can even dehydrate itself!
, some of which have been detected on the surface of Jupiter's moon Europa.
- Owing to the autoprotolysis and self-dehydration reactions described above, "pure" sulfuric acid contains at least six minority species in addition to H_2SO_4.



How it is made

Sulfur trioxide, the anhydride of sulfuric acid, is the immediate precursor. Gaseous SO_3 reacts vigorously with water, liberating much heat in the process:

$$SO_{3(g)} + H_2O_{(l)} \rightarrow H_2SO_{4(l)} \tag{10.7.1}$$

Industrial manufacture of the acid starts with sulfur dioxide, prepared from burning elemental sulfur or obtained as a byproduct from roasting sulfide ores. The oxidation of SO_2 to SO_3 looks simple

$$SO_{2(g)} + ½O_{2(g)} \rightarrow SO_{3(g)} \tag{10.7.2}$$

but there are several complications:

- All chemical reactions take place more rapidly at higher temperatures, but because this reaction is highly exothermic, raising the temperature decreases the yield.
- Because of the decrease in volume (1.5 moles of gases to 1 mole), raising the pressure will increase the yield, so the reaction is carried out at a temperature below 600°C but at very high pressure.
- Specialized catalysts are used to speed up the reaction at these lower temperatures.
- Dissolving the SO_3 directly in water would release large amounts of heat, creating a mist of fine acid droplets that would escape into the atmosphere. The SO_3 is instead dissolved in sulfuric acid to form *pyrosulfuric acid* or *oleum*, sometimes known as *fuming sufuric acid*:

$$H_2SO_{4(l)} + SO_{3(g)} \rightarrow H_2S_2O_{7(l)} \tag{10.7.3}$$

$$H_2S_2O_{7(l)} + H_2O_{(l)} \rightarrow 2H_2SO_{4(l)} \tag{10.7.4}$$

- The oleum is then treated with water to form industrial grade (96-98%) sulfuric acid:

What it is used for

Sulfuric acid has a broad spectrum of industrial uses, and the annual tonnage follows the economic cycle quite closely.

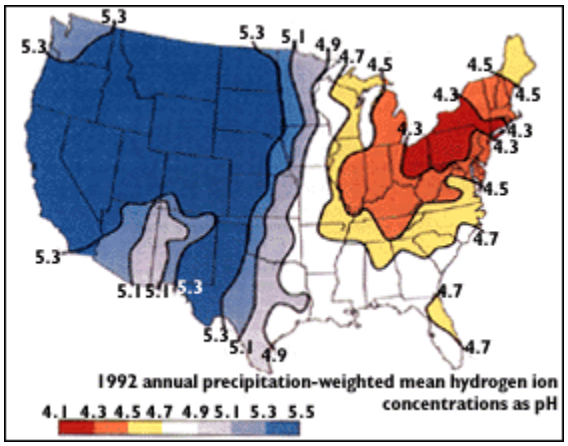
- Sixty percent of worldwide product goes into the manufacture of *phosphoric acid* H_3PO_4 which is used to make phosphate fertilizers and phosphate-based household detergents.
- An important nitrogen fertilizer, ammonium sulfate $(NH_4)_2SO_4$, is made by reacting sulfuric acid with ammonia, the latter often obtained from the thermal decomposition of coal.
- Sulfuric acid is the major component of pickling acid that is used to remove surface oxide scale from steel before it is

fabricated into steel product for the automobile and other industries.

- Aluminum sulfate (made from bauxite Al_2O_3 with H_2SO_4), is widely used in the papermaking industry to coagulate the cellulose fibers, producing a smooth, hard paper surface. Another major use is to make aluminum hydroxide which is used to filter out particulate matter in water treatment facilities.
- The familiar lead-acid storage battery employs sulfuric acid as an electrolyte. As the battery discharges, the concentraton of sulfuric acid in the electrolyte decreases as sulfate ions are taken up as $PbSO_4$. Owing to the high density of the acid, the state of charge of the battery can be measured by means of a hydrometer.

Sulfuric acid in the environment

Acid Rain - Combustion of fossil fuels which contain organic sulfur compounds releases SO_2 into the atmosphere. Photochemical oxidation of this compound to SO_3, which rapidly takes up moisture, leads to the formation of H_2SO_4, a major component of acid rain.



Acid mine drainage results when sediments of the very common *iron pyrite*, FeS2, are exposed to air and are oxidized:

$$FeS_2(s) + 7/2\ O_2 + H_2O \rightarrow Fe_2 + 2\ SO_4^{2-} + 2\ H^+$$

further oxidation of the iron to Fe^{3+} results in additional reactions. The resulting drainage liquid is often orange-brown in color and can a have a pH of below zero.

Nitric acid

Anhydrous HNO_3 is a colorless liquid boiling at 82.6°C, but "pure" HNO_3 only exists as the solid which melts at –41.6°C. In its liquid and gaseous states, the acid is always partially decomposed into nitrogen dioxide:

$$2\ HNO_3 \rightarrow 2\ NO_2 + ½\ O_2 + H_2O$$

This reaction, which is catalyzed by light, accounts for the brownish color of HNO_3 solutions.

What you should know about it

- HNO_2 undergoes autoprotolysis to a greater extent than any other liquid. Further reactions of the conjugate acid $H_2NO_3^+$ with HNO_3 lead to a complicated mixture of species in the liquid.
- Dilute nitric acid can be concentrated by distillation up to a maximum of 68%, at which point it forms a constant-boiling (azeotropic) mixture with water. Higher concentrations require dehydration with sulfuric acid; the result is fuming nitric acid.
- "Concentrated nitric acid" is sold as a 70% solution in water, corresponding to a concentration of about 16M.
- Nitric acid is a very strong **oxidizing agent**, which adds to its corrosive behavior with organic materials including, of course, skin, which it turns yellow owing to a reaction with the protein *keratin*. Reactions with many organic compounds are highly exothermic and often violent. The well-known reaction of nitric acid with metallic copper produces copious amounts of brown

nitrogen dioxide gas.



http://www.angelo.edu/ [image link]

How it is made

The simplest method, which was used industrially before 1900, was by treatment of sodium nitrate ("Chile saltpeter", $NaNO_3$) with sulfuric acid. The direct synthesis of the acid from atmospheric nitrogen and oxygen is thermodynamically favorable

$$½\ N_2 + 5/4\ O_2 + ½\ H_2O \rightarrow HNO_3$$

but is kinetically hindered by an extremely high activation energy, a fact for which we can be most thankful (see sidebar.) The first industrial nitrogen fixation process, developed in 1903, used this reaction to produce nitric acid, but it required the use of an electric arc to supply the activation energy and was therefore too energy-intensive to be economical.

If it were not for the high activation energy required to sustain this reaction, all of the oxygen in the atmosphere would be consumed and the oceans would be a dilute solution of nitric acid.

The modern Ostwald process involves the catalytic oxidation of ammonia to nitric oxide NO, which is oxidized in a further step to NO_2; reaction of the latter with water yields HNO_3. This route, first developed in 1901, did not become practical until the large-scale production of ammonia by the Haber-Bosch process in 1910.

What it is used for

The major industrial uses of nitric acid are for the production of ammonium nitrate **fertilizer**, and in the manufacture of **explosives**. On a much small scale, the acid is used in metal pickling, etching semiconductors and electronic circuit boards, and in the manufacture of acrylic fibers.

In the laboratory, the acid finds use in a wide variety of roles.

In the environment

High-temperature combustion processes (in internal combustion engines, power plants, and incinerators) can oxidize atmospheric nitrogen to nitric oxide (NO) and other oxides ("NO_x"); the NO is then photooxidized to NO_2, which reacts with water to form HNO_3 which is a major component of acid rain. $NO_2$is the major precursor of photochemical smog.

Hydrochloric acid

Unlike the other major acids, there is no such substance as "pure" hydrochloric acid; what we call "hydrochloric acid" is just an aqueous solution of hydrogen chloride gas (bp –84°C). But in a sense it is more "pure" than the acids discussed above, since there is no autoprotolysis; hydronium and chloride ions are the only significant species in the solution. Hydrochloric acid is usually sold as a 32-38% (12M) solution of HCl in water; concentrations greater than this are known as *fuming hydrochloric acid*.

Hydrochloric acid is still sometimes sold under its older name muriatic acid for cleaning bricks and other household purposes. The name comes from the same root as marine, reflecting its preparation from salt.

The acid has been known to chemists (and alchemists), and used for industrial purposes since the middle ages. Its composition HCl was demonstrated by Humphrey Davy in 1816.

What you should know about it

- Hydrochloric acid is the least hazardous of the strong mineral acids to work with because unlike the other ones, it is not an oxidizing agent. It is usually the acid of choice for titrations and other operations in which the main requirement is simply a strong source of hydronium ions.
- The concentrated acid boils at 48°C. As boiling continues, it loses HCl and the boiling point rises to 109°C, at which point a constant-boiling (azeotropic) solution remains, consisting of 20.2% HCl.



What it is used for

The uses of hydrochloric acid are far too many to enumerate individually, but the following stand out:

- A major industrial use is to **remove surface scale** from iron or steel ("pickling") before it is processed into sheets or other forms, or galvanized or coated.
- Production of **chlorinated organic chemicals**, particularly vinyl chloride, polyurethanes, and other construction polymers, consumes huge amounts of HCl.
- The acid is widely used for pH control of water, including neutralization of wastwater streams, and for regenerating ion-exchange water softeners.

How it is made

The ancient method of treating salt with sulfuric acid to release HCl has long since been supplanted by more efficient processes, including direct synthesis by "burning" hydrogen gas in chlorine:

$$H_{2(g)} + Cl_{2(g)} \rightarrow 2HCl_{(g)} \tag{10.7.5}$$

Most hydrochloric acid production now comes from reclaiming byproduct hydrogen chloride gas from other processes, especially those associated with the production of industrial organic compounds.

2 The Alkali Metals

The term *alkali* usually means a basic salt of a Group 1 or 2 ("alkali" or "alkaline earth") metal. All alkalies are of course bases, but the latter term is much more general, whether defined according to the Arrhenius, Brønsted-Lowry, or Lewis concepts. The word *alkali* comes from the Arabic *al-qali*, which refers to the ashes from which sodium and potassium hydroxides (*potash*, "ashes remaining in the pot", and the origin of the element name *potassium*) were extracted as a step in the making of soap.

Sodium hydroxide

Pure sodium hydroxide is a white solid consisting of Na^+ and OH^- ions in a crystal lattice. Although it is widely thought of as an ionic solid, *van der Waals forces* make a substantial contribution to its stability.

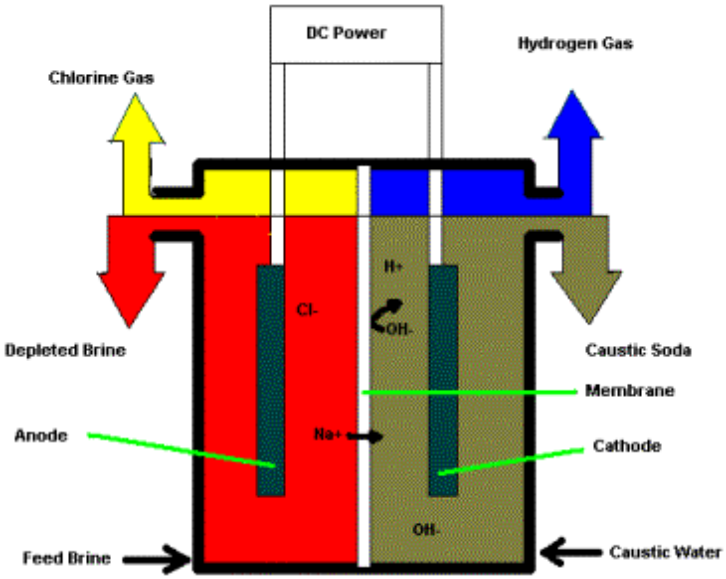


What you should know about it

- Owing to its *deliquescence* (ability to take up moisture) and its tendency to react with carbon dioxide, the solid must be stored in a closed container.
- In industry, sodium hydroxide is commonly known as *caustic soda* or simply as ***caustic***; NaOH sold for household purposes is usually known as *lye*.
- Sodium hydroxide slowly attacks **glass** to form sodium silicate. Glass vessels used to store concentrated solutions gradually develop a cloudy coating on the inside.
- Some **metals**, notably aluminum, zinc, and titanium, react with strongly alkaline solutions, but iron and copper are immune to this kind of attack.
- Highly alkaline solutions also soften and dissolve **skin**, accounting for the slippery feeling associated with strong bases. Sodium hydroxide was once used to dispose of animal carcasses, digesting them into an easily disposable liquid form.

How it is made

Sodium hydroxide is now manufactured by the electrolysis of brine solutions, and along with chlorine, is one of the two major products of the *chloralkali industry*.



Electrolysis of aqueous NaCl produces Cl_2 at the anode, but because H_2O can be reduced more readily than Na^+, the water is decomposed to H_2 and OH^- at the cathode, leaving a solution of NaOH. An older *mercury cell process* reduces the Na^+ to Na within a mercury amalgam (alloy), and the metallic sodium is then combined with water to produce NaOH and hydrogen. The net reaction for the reduction step is the same for both methods:

$$\[2 Na^+ + 2 H_2O + 2e^- \rightarrow H_{2(g)] + 2 NaOH\]$$

The resulting solution is usually evaporated to such a high concentration that it solidifies at ordinary temperatures. It is commonly shipped in rail cars or barges which can be heated with steam to liquefy the mixture for removal. (It is obviously uneconomical to ship large quantities of water across the country!)

What it is used for

- Sodium hydroxide is one of the most diverse industrial chemicals in terms of its applications. Most householders know it as the active ingredient of **drain cleaning agents**.
- Huge quantities are consumed by the **pulp and paper** industry, which is probably its single largest specific industrial application. It is used to remove the lignin component of wood pulp from the cellulose so that the latter can be processed into paper.
- About half of the NaOH output goes into the production of a wide variety of other **industrial chemicals**, and in degreasing steel drums and other industrial surfaces.
- "Lye" plays a role in the processing of many types of **foods**, including chocolate, olives, pretzels, and the "hominy" and "grits" corn products used in the Southern U.S. It also acts as a chemical peeling agent for fruits and vegetables.



In contrast to the extremely diverse applications of sodium hydroxide which makes the demand for this commodity relatively immune to the ups and downs of the economic cycle, the consumption of chlorine is directly dependent on the economy as reflected in the demand for polyvinyl chloride products that are now widely used in the the construction and home furnishings industries. Because chlorine, being a gas, is expensive to store, the output of the chloralkali industry is governed largely by the demand for this commodity. When times are good this presents no problem; caustic is then largely a by-product and can easily be stockpiled if supply exceeds demand. But during an economic downturn, the demand for chlorine declines, limiting its production along with that of caustic. But because the demand for caustic tends to decline much less, it becomes scarce and its price rises, thus tending to drive the industrial economy into even deeper trouble.

Sodium carbonate

This compound is known industrially as ***soda ash***, and domestically as *washing soda*. The common form is the *heptahydrate*, $Na_2CO_3 \cdot 7\ H_2O$. The white crystals of this substance spontaneously lose water (*effloresce*) when exposed to the air, forming the monohydrate.

What it is used for

- Although carbonates are much weaker bases than hydroxides, a solution of sodium carbonate can still have a pH of 11 or so, sufficiently high to allow it to substitute for sodium hydroxide in many applications— especially when the price of caustic is high.
- The single most important use of soda ash is in the manufacture of glass, where it serves to lower the melting point of the principal component, SiO_2.
- Another emerging major use is to neutralize the SO_2 emissions of fossil fuel-burning power plants.
- The older use of sodium carbonate as a **cleaning agent** (hence the name *washing soda*) was based partly on the ability of its alkaline solutions to emulsify grease, but mainly as a means of precipitating the insoluble carbonates of calcium and magnesium before these ions (commonly present in hard water) could form undesirable preciptates with soaps. The use of modern detergents has largely eliminated this once important market.



How it is made

Most of the world's sodium carbonate is made by the ammonia-soda "Solvay" process developed in 1861 by the Belgian chemist Ernest Solvay (1838-1922) whose patents made him into a major industrialist and a rich philanthropist. This process involves a set of simple reactions that essentially converts limestone ($CaCO_3$), ammonia NH_3 and brine (NaCl) into sodium bicarbonate $NaHCO_3$ and eventually Na_2CO_3, recycling several of the intermediate products in an ingenious way.

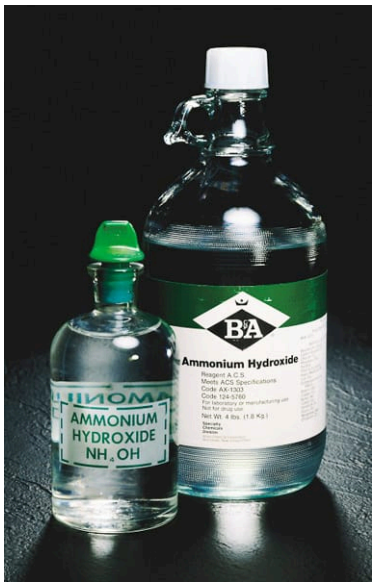


A minor source of soda ash (but quite significant in some countries, such as the U.S.) is the mining of natural *evaporites* (the

remains of ancient lakes), such as the *trona* found in Southern California.

Ammonia

Ammonia NH_3 is of course not a true alkali, but it is conveniently included in this section for discussion purposes. Most people are familiar with the pungent odor of this gas, which can be detected at concentrations as low as 20-50 ppm.



Tradition dies slowly: a non-existent chemical available in bottles!

What you should know about it

- More moles of ammonia are manufactured than of any other industrial chemical.
- Ammonia is extremely soluble in water. An aqueous solution of ammonia is still sometimes referred to in commerce as "ammonium hydroxide", but this term is no longer favored by chemists because no such compound as NH_4OH has ever been shown to exist. At neutral pH, about 99% of the ammonia in water exists as NH_4^+ ions.
- Ammonia is an end product of nitrogen metabolism in most organisms. One source that may be familiar to parents of infants is the bacterial decomposition of the contents of diapers.
- Liquid ammonia (bp –33°C) is often used as an ionizing laboratory solvent.

What it is used for

- The major use of ammonia (about 80%) is as a fertilizer, most commonly as anhydrous ammonia (the gas is injected directly into the soil) or after conversion to $(NH_4)_2SO_4$, NH_4NO_3, or urea $O–C(NH_3)_2$.
- Ammonia is used in the production of numerous polymers, including nylons and polyurethanes.
- Explosives manufacture accounts for about 5% of ammonia production.
- Beyond these, there are hundreds of minor uses, including as a household cleaning agent (aqua ammonia). a refrigerant, and as a laboratory reagent.

How it is made

Ammonia is made by direct synthesis from the elements:

$$N_{2(g)} + 3H_{2(g)} \rightarrow 2NH_{3(g)} \tag{10.7.6}$$

... a simple-looking reaction, but one that required some very creative work to implement; the *Haber-Bosch process* is considered to be the most important chemical synthesis of the 20th Century.

3 Some important organic acids

Most acids are organic— there are millions of them. The acidic function is usually a hydroxyl group connected to a carbon that is bonded to an electron-withdrawing oxygen atom; the combination is the well-known *carboxyl group*, –COOH. Here are a few that are part of everyone's life.

Acetic acid

This is next to formic acid in being the simplest of the organic acids, and in the form of vinegar (a 5-8% solution in water) its characteristic odor is known to everyone. The pure acid is a colorless liquid above 16.7°C; below this temperature it forms a crystalline solid, hence the term "glacial acetic acid" that is commonly applied to the pure substance. The name of the acid comes from *acetum*, the Latin word for vinegar.





What you should know about it

- The pure acid, although quite weak in the proton-donor sense, is quite corrosive and its vapors are very irritating.
- A 1.0M solution of the acid has a pH of about 2.4, corresponding to only four out of every thousand CH_3COOH molecules being dissociated.

What it is used for

Slightly less than half of the world production of acetic acid goes into the production of **polymers**. The end product visible to most people would be the flexible plastic bottles in which drinking water is sold. Other uses are related mostly to the production of other chemicals, mainly *acetic anhydride*, but also including *aspirin*.



How it is made

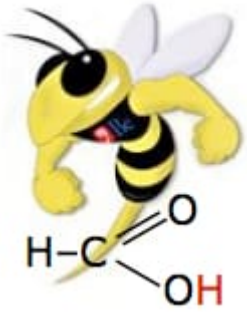
Bacterial fermentation of sugars has been the source of vinegar since ancient times, and is still accounts for most food-grade acetic acid and vinegar, but it now amounts to only about 10% of total acetic acid production:

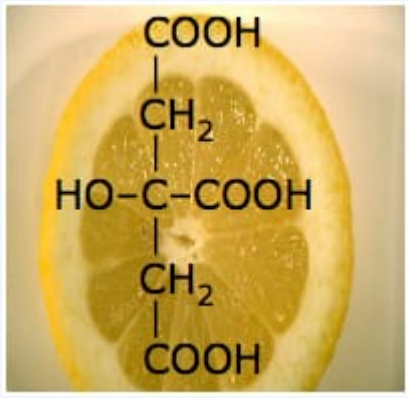
$C_6H_{12}O_6 \rightarrow 3\ CH_3COOH$

There are several important synthetic routes to acetic acid production, but the major one is by treating methanol with carbon monoxide:

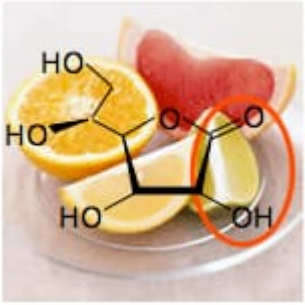
$CH_3OH + CO \rightarrow CH_3COOH$

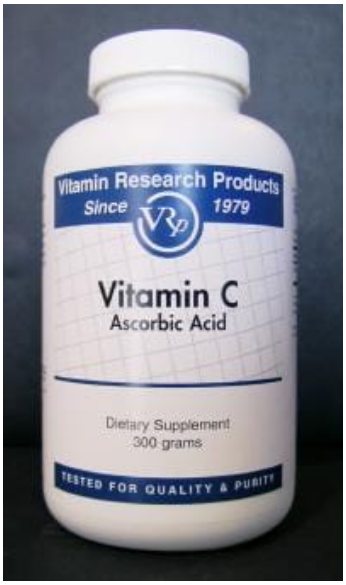
A brief look at some other interesting organic acids

 **Formic acid** (Methanoic acid) HCOOH mp 8.4°C.	This, the simplest of the carboxylic acids, is the chemical weapon that Nature has given ants and bees (the Latin word for ant is *formicus*.) Known since the 15th Century, it was first distilled from ants, but is now made synthetically. Its main uses are as a preservative and antibacterial agent in livestock feed. Chemists find it a useful source of carbon monoxide (just add sulfuric acid.)
 **Oxalic acid** HOOC–COOH	Just two carboxyl groups joined together, this acid has a special ability to grab up divalent metal ions and bind them into five-membered MO_2C_2rings. Because many such metal ions (Ca^{2+}, for example) are essential nutrients, this can be dangerous to your health. The acid occurs in many plants, notably rhubarb (it is what makes the leaves poisonous), parsley and spinach. If you have ever noticed a funny feeling in your mouth after drinking milk with a rhubarb desert, it is due to precipitation of calcium oxalate. This same solid is often a major component of kidney stones, and it contributes to the miseries of gout.

Lactic acid 	Structurally, lactic acid is both an alcohol and a carboxylic acid— not all that uncommon. We know it as the acid found in sour milk, yogurt, and in tired muscles. When the blood cannot deliver enough oxygen to oxidize glucose all the way to CO_2 and water, your muscles go into a much-less-energy-efficient "anaerobic" mode that generates lactate. In milk products, lactic acid is produced in about the same way by *acidophilus* bacteria that eat the *lactose* ("milk sugar") and who don't know how to utilize oxygen.
Citric acid  Known by Arabic alchemists in the 8th Century; first isolated by Scheele in 1784	Although it is one part alcohol and three parts acid, citric acid is quite weak, but nevertheless strong enough to make it unpleasant to suck on a lemon, of which it can comprise as much as 8% of the dry weight of this fruit. Biochemists know it as part of the *citric acid cycle,* a sequence of reactions involved in extracting energy from the oxidative metabolism of foods. Its major industrial use is as a food flavoring and preservative agent; large quantities are used to make soft drink beverages. The acid also finds use in cleaning agents.

Ascorbic acid

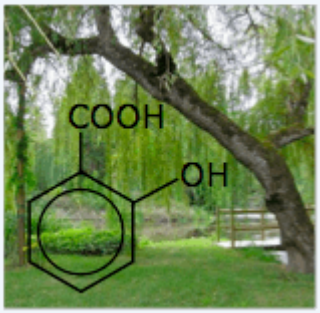




No carboxyl groups here, but the –OH group one carbon away from the double bond is still fairly acidic. One of its geometric isomers, L-ascorbic acid, is more widely known as **Vitamin C**; discovery of the essential role of this substance in preventing the disease *scurvy* (from which its name is derived) yielded two Nobel prizes in 1937. Most animals are capable of synthesizing their own Vitamin C, but primates (along with guinea pigs) seem to have lost the required gene somewhere along the way, so we must depend on fruits and veggies for our supply. (Because ascorbic acid is soluble in water, it tends to be leached out of vegetables when they are boiled, so it is much healthier to steam them.)

The major industrial use of ascorbic acid is as an *antioxidant*, so it is often added to foods and other materials as a preservative.

Salicylic acid



2-hydroxybenzoic acid is found in willow trees

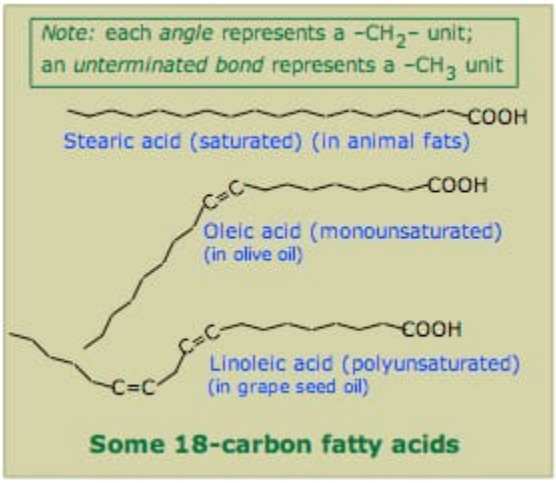
Does chemistry give you a headache? If so, pull a small shoot off of a willow tree and chew on it for a while. This has been a folk-remedy for pains and fever since ancient times. Willow bark (*salix* in Latin) contains, and lent its name to the active principle, salicylic acid.



The acid itself turned out to be a bit too much of an irritant to the stomach's lining, so the German firm Bayer began marketing a tamed version, *acetylsalicylic acid* (ASA) under the name Aspirin in 1899, and it has been going strong ever since. Interestingly, the detailed chemistry of its pain-and-fever reliving action was not discovered until 1970.

Fatty acids

This generic term refers to carboxylic acids built from a chain of 4 to 22 carbon atoms. Fatty acids, as the name implies, are derived from fats, in which they are bound to glycerol in the form of triglycerides. Fats, which occur in all animal tissues, are the most efficient means of storing metabolic energy. Vegetable oils are another source.



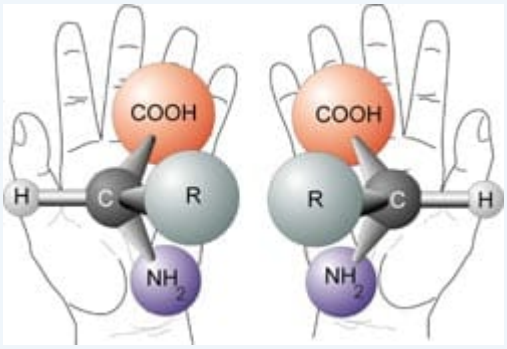
There are two general categories of fatty acids:

- **Saturated fatty acids** consist of straight chains of carbon atoms with a methyl group on one end and a carboxyl (acidic) group on the other, with *methylene groups* -CH_2- in between. Thus stearic (octadecanoic) acid is $CH_3(CH_2)_{16}COOH$. The linear shape of these molecules allows them to pack together very efficiently, contributing to their ability to store energy in a small space; the higher saturated acids tend to be waxy solids.
- **Unsaturated fatty acids** contain one or more carbon-carbon double bonds, which introduces a complication: because free rotation around the axis of a double bond is not possible, the neighboring hydrogens can be either on the same side of the bond (*cis*) or on opposite sites (*trans*)— thus the *cis*- and *trans*-fatty acids. These double bonds introduce kinks into the carbon chain, especially in the case of the *cis*-acids. The bent and curved carbon chains that result cannot pack together compactly enough to interact very strongly with each other, so unsaturated fatty acids tend to be liquids.

The human body is able to synthesize most of the fatty acids it needs, but two classes of unsaturated acids, known as the **essential fatty acids**, can be obtained only from foods. The compounds are known as ω-3 and ω-6 fatty acids; the ω (omega)-*n* notation means that the double bond is located *n* carbons away from the terminal CH_3group of the molecule.

Most of the unsaturated fatty acids found in nature have cis configurations around their double bonds. The trans fatty acids in our foods that we hear much about are made from vegetable oils that have had some of their double bonds changed to single bonds by a chemical process called hydrogenation. The object of this is to make the original liquid fatty acids into solid forms that are more suitable for use in foods (as in margarine) and particularly as shortenings in baking. Unfortunately, the hydrogenation process also changes the remaining cis double bonds into their trans forms which are believed to be implicated in cardiac disease.

Amino acids

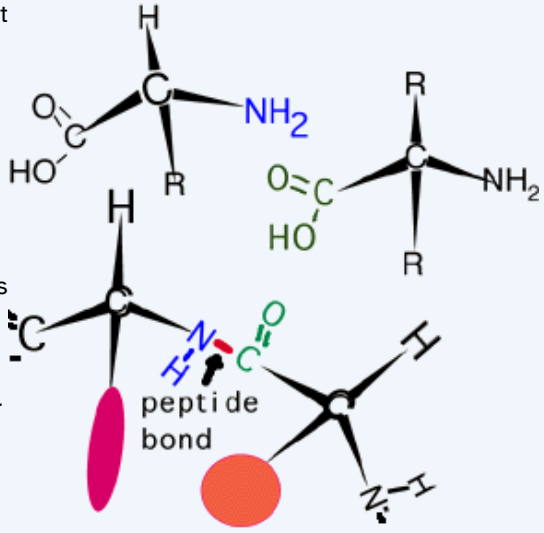


Amino acids are the building blocks of proteins, which is what you are largely made of. There are twenty of them, of which about half can be manufactured by the body; we must depend on our food to supply the remainder, which are known as the *essential amino acids*. Each one comes in two mirror-image forms; only the "left" forms are found in most proteins.

All amino acids have the basic structure shown above; they differ in the nature of the group of atoms designated by "R" in the diagram.

Although we call them "acids", the amino acids are really chemical hermaphrodites; you will recall that amines are weak bases. The balance between their acidic and basic properties can be shifted simply by changing the pH.

The carboxylic acid part of one amino acid can react with the amine part of another to form a peptide bond (an *amide* linkage) shown here.



The polymeric chains that result are known as peptides if they are fairly short (2 to about 20-50 amino acid units); longer chains, or aggregate made up of multiple peptide units, are proteins. Owing to their very large size (500 amino acid residues is quite common, but some have as many as 1500), proteins are able to fold in various ways, so the amino acid sequence alone is not sufficient to determine their properties.

This page titled 10.7: Acid-Base Gallery is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

11: CHEMICAL EQUILIBRIUM

The laws of chemical equilibrium define the direction in which a chemical reaction will proceed, as well as the quantities of reactants and products that will remain after the reaction comes to an end. An understanding of chemical equilibrium and how it can be manipulated is essential for anyone involved in Chemistry and its applications.

11.1: Introduction to Chemical Equilibrium

11.2: Le Chatelier's Principle

11.3: Reaction Quotient

11.4: Equilibrium Expressions

11.5: Equilibrium Calculations

11.6: Phase Distribution Equilibria

This page titled 11: Chemical Equilibrium is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Introduction to Chemical Equilibrium

Learning Objectives

- Define "*the equilibrium state of a chemical reaction system*". What is its practical significance?
- State the meaning and significance of the following terms:
  - *reversible reaction*
  - *quantitative reaction*
  - *kinetically inhibited reaction*
- Explain the meaning of the statement "*equilibrium is macroscopically static, but microscopically dynamic*". Very important!
- Explain how the relatve magnitudes of the forward and reverse *reaction rate constants* in the Mass Action expression affect the equilibrium composition of a reaction system.
- Describe several things you might look for during an experiment that would help determine if a reaction system is in its equilibrium state.

Chemical change is one of the two central concepts of chemical science, the other being s*tructure*. The very origins of Chemistry itself are rooted in the observations of transformations such as the combustion of wood, the freezing of water, and the winning of metals from their ores that have always been a part of human experience. It was, after all, the quest for some kind of constancy underlying change that led the Greek thinkers of around 200 BCE to the idea of elements and later to that of the atom. It would take almost 2000 years for the scientific study of matter to pick up these concepts and incorporate them into what would emerge, in the latter part of the 19th century, as a modern view of chemical change.

Chemical change: how far, how fast?

Chemical change occurs when the atoms that make up one or more substances rearrange themselves in such a way that new substances are formed. These substances are the components of the chemical reaction system; those components which decrease in quantity are called reactants, while those that increase are products. A given chemical reaction system is defined by a *balanced net chemical equation* which is conventionally written as

$$\text{reactants} \rightarrow \text{products} \tag{11.1.1}$$

The first thing we need to know about a chemical reaction represented by a balanced equation is whether it can actually take place. If the reactants and products are all substances capable of an independent existence, then in principle, the answer is always "yes". This answer must be qualified, however, by the following considerations:

How complete is the reaction?
: That is, what fraction of the reactants are converted into products? Some reactions convert essentially 100% of reactants to products, while for others the quantity of products may be undetectable. Many are somewhere in between, meaning that significant quantities of all components remain at the end. Later on, in another part of the course, you will learn that the tendency of a reaction to occur can be predicted entirely from the properties of the reactants and products through the laws of thermodynamics.

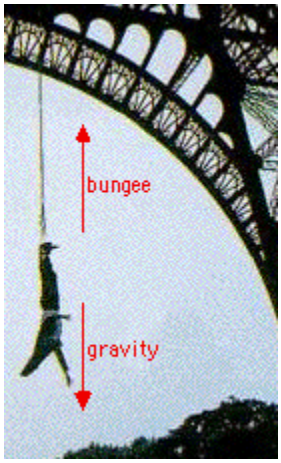
How fast does the reaction occur?
: Some reactions are over in microseconds; others take years. The speed of any one reaction can vary over a huge range depending on the temperature, the state of matter (gas, liquid, solid) and the presence of a catalyst. Unlike the question of completeness, there is no simple way of predicting reaction speed.

What is the mechanism of the reaction?
: What happens, at the atomic or molecular level, when reactants are transformed into products? What intermediate species (those that are produced but later consumed so that they do not appear in the net reaction equation) are involved? This is the *microscopic*, or kinetic view of chemical change, and cannot be predicted by theory as it is presently developed and must be inferred from the results of experiments.

A reaction that is thermodynamically possible but for which no reasonably rapid mechanism is available is said to be *kinetically limited*. Conversely, one that occurs rapidly but only to a small extent is *thermodynamically limited*. As you will see later, there are often ways of getting around both kinds of limitations, and their discovery and practical applications constitute an important area of industrial chemistry.

What is equilibrium?

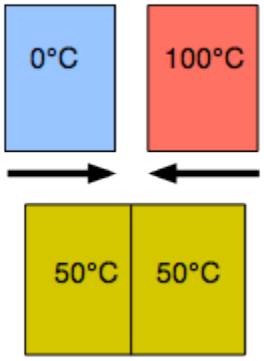
Basically, the term refers to what we might call a "balance of forces". In the case of mechanical equilibrium, this is its literal definition. A book sitting on a table top remains at rest because the downward force exerted by the earth's gravity acting on the book's mass (this is what is meant by the "weight" of the book) is exactly balanced by the repulsive force between atoms that prevents two objects from simultaneously occupying the same space, acting in this case between the table surface and the book. If you pick up the book and raise it above the table top, the additional upward force exerted by your arm destroys the state of equilibrium as the book moves upward. If you wish to hold the book at rest above the table, you adjust the upward force to exactly balance the weight of the book, thus restoring equilibrium.



An object is in a state of **mechanical equilibrium** when it is either static (motionless) or in a state of unchanging motion. From the relation $f = ma$, it is apparent that if the net force on the object is zero, its acceleration must also be zero, so if we can see that an object is not undergoing a change in its motion, we know that it is in mechanical equilibrium.

Thermal equilibrium

Another kind of equilibrium we all experience is **thermal equilibrium**. When two objects are brought into contact, heat will flow from the warmer object to the cooler one until their temperatures become identical. Thermal equilibrium arises from the tendency of thermal energy to become as dispersed or "diluted" as possible.



A metallic object at room temperature will feel cool to your hand when you first pick it up because the thermal sensors in your skin detect a flow of heat from your hand into the metal, but as the metal approaches the temperature of your hand, this sensation diminishes. The time it takes to achieve thermal equilibrium depends on how readily heat is conducted within and between the

objects; thus a wooden object will feel warmer than a metallic object even if both are at room temperature because wood is a relatively poor thermal conductor and will therefore remove heat from your hand more slowly.

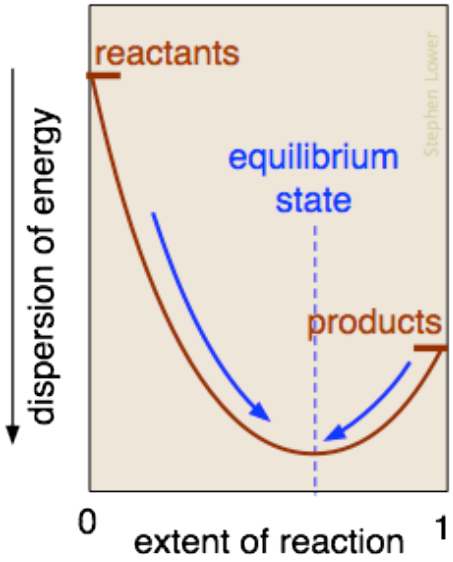
Thermal equilibrium is something we often want to avoid, or at least postpone; this is why we insulate buildings, perspire in the summer and wear heavier clothing in the winter.

Chemical equilibrium

When a chemical reaction takes place in a container which prevents the entry or escape of any of the substances involved in the reaction, the quantities of these components change as some are consumed and others are formed. Eventually this change will come to an end, after which the composition will remain unchanged as long as the system remains undisturbed. The system is then said to be in its *equilibrium state*, or more simply, "at equilibrium".

Why reactions go toward equilibrium

What is the nature of the "balance of forces" that drives a reaction toward chemical equilibrium? It is essentially the balance struck between the tendency of energy to reside within the chemical bonds of stable molecules, and its tendency to become dispersed and diluted. Exothermic reactions are particularly effective in this, because the heat released gets dispersed in the infinitely wider world of the surroundings.



In the reaction represented here, this balance point occurs when about 60% of the reactants have been converted to products. Once this equilibrium state has been reached, no further net change will occur. The only spontaneous changes that are allowed follow the arrows pointing toward maximum dispersal of energy.

Chemical equilibrium is something you definitely want to avoid for yourself as long as possible. The myriad chemical reactions in living organisms are constantly moving *toward* equilibrium, but are prevented from getting there by input of reactants and removal of products. So rather than being in equilibrium, we try to maintain a "steady-state" condition which physiologists call *homeostasis* — maintenance of a constant internal environment. **Equilibrium is death!**

For the time being, it's very important that you know this definition:

A chemical reaction is in equilibrium when there is no tendency for the quantities of reactants and products to change.

The **direction** in which a chemical reaction is written (and thus which components are considered reactants and which are

products) is arbitrary. Consider the following two reactions:

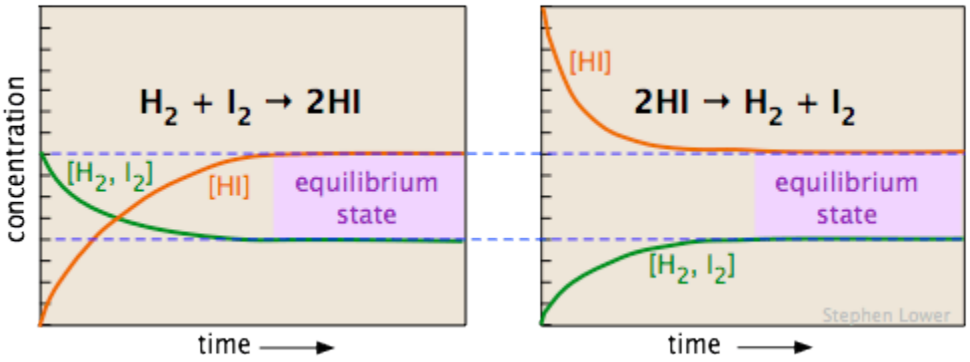
$$\underset{\text{synthesis of hydrogen iodide}}{H_2 + I_2 \rightarrow 2HI} \tag{11.1.2}$$

$$\underset{\text{dissociation of hydrogen iodide}}{2HI \rightarrow H_2 + I_2} \tag{11.1.3}$$

Equations 11.1.2 and 11.1.3 represent the same chemical reaction system in which the roles of the components are reversed, *and both yield the same mixture of components when the change is completed.* This is central to the concept of chemical equilibrium. It makes no difference whether we start with two moles of HI or one mole each of H_2 and I_2; once the reaction has run to completion, the quantities of these two components will be the same. In general, then, we can say that the composition of a chemical reaction system will tend to change in a direction that brings it closer to its equilibrium composition**.** Once this equilibrium composition has been attained, no further change in the quantities of the components will occur as long as the system remains undisturbed.

> The composition of a chemical reaction system will tend to change in a direction that brings it closer to its equilibrium composition.

The two diagrams below show how the concentrations of the three components of this chemical reaction change with time. Examine the two sets of plots carefully, noting which substances have zero initial concentrations, and are thus "products" of the reaction equations shown. Satisfy yourself that these two sets represent the same *chemical reaction system*, but with the reactions occurring in opposite directions. Most importantly, note how the final (equilibrium) concentrations of the components are the same in the two cases.



Whether we start with an equimolar mixture of H_2 and I_2 (left) or a pure sample of hydrogen iodide (shown on the right, using twice the initial concentration of HI to keep the number of atoms the same), the composition after equilibrium is attained (shaded regions on the right) will be the same.

> The **equilibrium** composition is independent of the direction from which it is approached (i.e., the initial conditions).

A chemical equation of the form A → B represents the transformation of A into B, but it does not imply that *all* of the reactants will be converted into products, or that the *reverse* reaction B → A cannot also occur. In general, *both* processes (forward and reverse) can be expected to occur, resulting in an *equilibrium mixture* containing finite amounts of *all* of the components of the reaction system. (We use the word *components* when we do not wish to distinguish between reactants and products.)

If the equilibrium state is one in which significant quantities of both reactants and products are present (as in the hydrogen iodide example given above), then the reaction is said to *incomplete* or *reversible*. The latter term is preferable because it avoids confusion with "complete" in its other sense of being completed or finished, implying that the reaction has run its course and is now at equilibrium.

Note that there is no fundamental difference between the meanings of A → B and A $\rightleftharpoons$ B. Some older textbooks just use A = B.

- If it is desired to emphasize the reversibility of a reaction, the single arrow in the equation is replaced with a pair of hooked lines pointing in opposite directions, as in A $\rightleftharpoons$ B.
- A reaction is said to be *complete* or *quantitative* when the equilibrium composition contains no significant amount of the

reactants. However, a reaction that is complete when written in one direction is said "not to occur" when written in the reverse direction.

In principle, all chemical reactions are reversible, but this reversibility may not be observable if the fraction of products in the equilibrium mixture is very small, or if the reverse reaction is very slow (the chemist's term is "*kinetically inhibited*")

How did Napoleon Bonaparte help discover reversible reactions?

We can thank Napoleon for bringing the concept of reaction reversibility to Chemistry. Napoleon recruited the eminent French chemist Claude Louis Berthollet (1748-1822) to accompany him as scientific advisor on the most far-flung of his campaigns, the expedition in Egypt in 1798. Once in Egypt, Berthollet noticed deposits of sodium carbonate around the edges of some the salt lakes found there. He was already familiar with the reaction

$$Na_2CO_3 + CaCl_2 \rightarrow CaCO_3 + 2NaCl \tag{11.1.4}$$

which was known to proceed to completion in the laboratory. He immediately realized that the Na_2CO_3 must have been formed by the reverse of this process brought about by the very high concentration of salt in the slowly-evaporating waters. This led Berthollet to question the belief of the time that a reaction could only proceed in a single direction. His famous textbook *Essai de statique chimique*(1803) presented his speculations on chemical affinity and his discovery that an excess of the product of a reaction could drive it in the reverse direction.

Unfortunately, Berthollet got a bit carried away by the idea that a reaction could be influenced by the amounts of substances present, and maintained that the same should be true for the compositions of individual compounds. This brought him into conflict with the recently accepted Law of Definite Proportions (that a compound is made up of fixed numbers of its constituent atoms), so his ideas (the good along with the bad) were promptly discredited and remained largely forgotten for 50 years. (Ironically, it is now known that certain classes of compounds do in fact exhibit variable composition of the kind that Berthollet envisioned.)

What is the law of mass action?

Berthollet's ideas about reversible reactions were finally vindicated by experiments carried out by others, most notably the Norwegian chemists (and brothers-in-law) Cato Guldberg and Peter Waage. During the period 1864-1879 they showed that an equilibrium can be approached from either direction (see the hydrogen iodide illustration above), implying that any reaction

$$aA + bB \rightarrow cC + dD \tag{11.1.5}$$

is really a competition between a "forward" and a "reverse" reaction. When a reaction is at equilibrium, the rates of these two reactions are identical, so no *net* (macroscopic) change is observed, although individual components are actively being transformed at the microscopic level.

Chemical Equilibrium is *dynamic*!

> The "active masses" are essentially the *concentrations* of the reactants and products that combine directly in the manner represented by the reaction equation. The meaning of the terms on the right sides of the equations is simply that the rate is proportional to the concentrations of the components.

Guldberg and Waage showed that for a reaction

$$aA + bB \rightarrow cC + dD \tag{11.1.6}$$

the rate (speed) of the reaction in either direction is proportional to what they called the "active masses" of the various components:

- rate of forward reaction = k_f [A]^a [B]^b
- rate of reverse reaction = k_r [C]^c [D]^d

in which the proportionality constants k are called *rate constants* and the quantities in square brackets represent concentrations.

If we combine the two reactants A and B, the forward reaction starts immediately; then, as the products C and D begin to build up, the reverse process gets underway. As the reaction proceeds, the rate of the forward reaction diminishes while that of the reverse reaction increases. Eventually the two processes are proceeding at the same rate, and the reaction is at equilibrium:

rate of forward reaction = rate of reverse reaction

$$k_f\,[A]^a\,[B]^b = k_r\,[C]^c\,[D]^d \tag{11.1.7}$$

It is very important that you understand the significance of this relation. The equilibrium state is one in which there is no **net** change in the quantities of reactants and products. But do not confuse this with a state of "no change"; at equilibrium, the forward and reverse reactions continue, but *at identical rates*, essentially cancelling each other out.

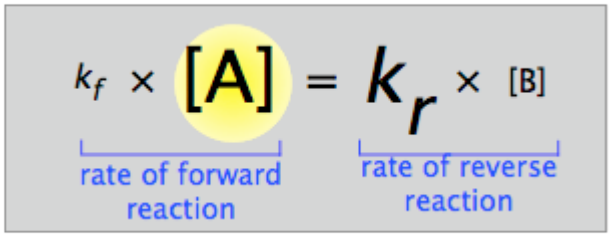
Equilibrium is macroscopically static, but is microscopically dynamic! To further illustrate the dynamic character of chemical equilibrium, suppose that we now change the composition of the system previously at equilibrium by adding some C or withdrawing some A (thus changing their "active masses"). The reverse rate will temporarily exceed the forward rate and a change in composition ("a shift in the equilibrium") will occur until a new equilibrium composition is achieved.

Composition of the Equilibrium State Depends on the ratio of the forward- and reverse rate constants.

Be sure you understand the difference between the *rate* of a reaction and a *rate constant*. The latter, usually designated by *k*, relates the reaction rate to the concentration of one or more of the reaction components — for example, *rate* = *k* [A]. At equilibrium the rates of the forward and reverse processes are identical, but the rate *constants* are generally different. To see how this works, consider the simplified reaction A → B in the following three scenarios.

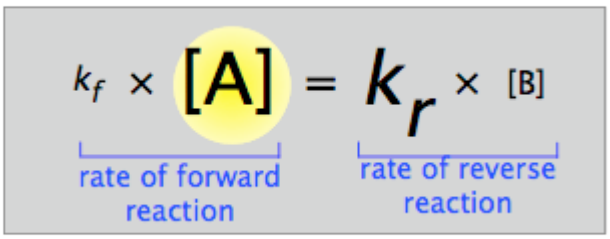
k_f >> k_r

If the rate constants are greatly different (by many orders of magnitude), then this requires that the equilibrium concentrations of products exceed those of the reactants by the same ratio. Thus the equilibrium composition will lie strongly on the "right"; the reaction can be said to be "complete" or "quantitative".



k_f << k_r

The rates can only be identical (equilibrium achieved) if the concentrations of the products are very small. We describe the resulting equilibrium as strongly favoring the left; very little product is formed. In the most extreme cases, we might even say that "the reaction does not take place".

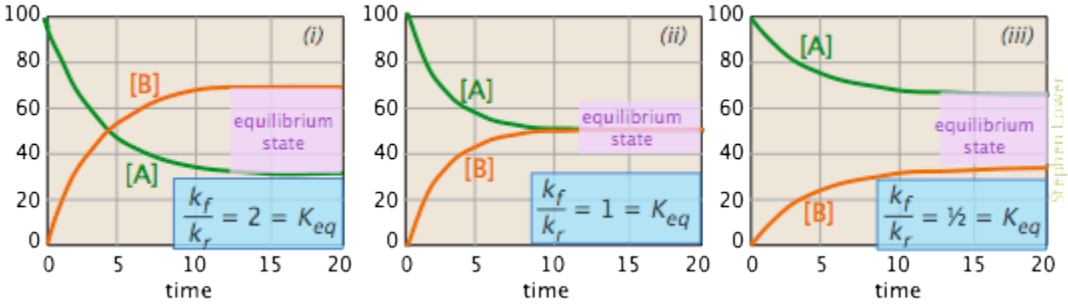


k_f ≈ k_r

If k_f and k_r have comparable values (within, say, several orders of magnitude), then significant concentrations of products and reactants are present at equilibrium; we say the the reaction is "incomplete" and "reversible".

The images shown below offer yet another way of looking at these three cases. The plots show how the relative concentrations of the reactant and product change during the course of the reaction. The plots differ in the assumptions we make about the ratio of

k_f to k_r. The equilibrium composition of the system is illustrated by the proportions of A and B in the horizontal parts of each plot where the composition remains unchanged. In each case, the two rate constants are sufficiently close in magnitude that each reaction can be considered "incomplete".



- In plot *(i)* the forward rate constant is twice as large as the reverse rate constant, so product (B) is favored, but there is sufficient reverse reaction to maintain a significant quantity of A.
- In *(ii)*, the forward and reverse rate constants have identical magnitudes. Not surprisingly, so are the equilibrium values of [A] and [B].
- In *(iii)*, the reverse rate constant exceeds the forward rate constant, so the equilibrium composition is definitely "on the left".

The Law of Mass Action is thus essentially the statement that the equilibrium composition of a reaction mixture can vary according to the quantities of components that are present. This of course is just what Berthollet observed in his Egyptian salt ponds, but we now understand it to be a consequence of the dynamic nature of chemical equilibrium.

When is a Reaction at Equilibrium?

Clearly, if we observe some change taking place— a change in color, the release of gas bubbles, the appearance of a precipitate, or the release of heat, we know the reaction is not yet at equilibrium. However, **the absence of any apparent change does not by itself establish that the reaction is at equilibrium**. The equilibrium state is one in which not only no change in composition take place, but also one in which no energetic tendency for further change is present. Unfortunately, "tendency" is not a property that is directly observable! Consider, for example, the reaction representing the synthesis of water from its elements:

$$2H_{2(g)} + O_{2(g)} \rightarrow 2H_2O_{(g)} \tag{11.1.8}$$

You can store the two gaseous reactants in the same container indefinitely without any observable change occurring. But if you create an electrical spark in the container or introduce a flame, *bang*! After you pick yourself up off the floor and remove the shrapnel from what's left of your body, you will know very well that the system was not initially at equilibrium! It happens that this particular reaction has a tremendous tendency to take place, but for reasons that we will discuss in a later chapter, nothing can happen until we "set it off" in some way— in this case by exposing the mixture to a flame or spark, or (in a more gentle way) by introducing a platinum wire, which acts as a *catalyst*.

A reaction of this kind is said to be highly favored thermodynamically, but inhibited kinetically. The similar reaction of hydrogen and iodine

$$H_{2(g)} + I_{2(g)} \rightarrow 2HI_{(g)} \tag{11.1.9}$$

by contrast is only moderately favored thermodynamically (and is thus incomplete), but its kinetics are both unspectacular and reasonably facile.

- As we explained above in the context of the law of mass action, addition or removal of one component of the reaction will affect the amounts of all the others. For example, if we add more of a reactant, we would expect to see the concentration of a product change. If this does not happen, then it is likely that the reaction is kinetically inhibited and that the system is unable to attain equilibrium.
- It is almost always the case, however, that once a reaction actually starts, it will continue on its own until it reaches

equilibrium, so if we can observe the change as it occurs and see it slow down and stop, we can be reasonably certain that the system is in equilibrium. This is by far the chemist's most common criterion.
- There is one other experimental test for equilibrium in a chemical reaction, although it is really only applicable to the kind of reactions we described above as being reversible. As we shall see later, the equilibrium state of a system is always sensitive to the temperature, and often to the pressure, so any changes in these variables, however, small, will temporarily disrupt the equilibrium, resulting in an observable change in the composition of the system as it moves toward its new equilibrium state.

Summary

Make sure you thoroughly understand the following essential ideas which have been presented above.

- Any reaction that can be represented by a balanced chemical equation can take place, at least in principle. However, there are two important qualifications:
  - The tendency for the change to occur may be so small that the quantity of products formed may be very low, and perhaps negligible.
    A reaction of this kind is said to be ***thermodynamically inhibited***. The tendency for chemical change is governed solely by the properties of the reactants and products, and can be predicted by applying the laws of thermodynamics.
  - The rate at which the reaction proceeds may be very small, or even zero, in which case we say the reaction is *kinetically inhibited*. Reaction rates depend on the *mechanism* of the reaction— that is, on what actually happens to the atoms as reactants are transformed into products. Reaction mechanisms cannot generally be predicted, and must be worked out experimentally. Also, the same reaction may have different mechansims under different conditions.
- As a chemical change proceeds, the quantities of the components on one side of the reaction equation will decrease, and those on the other side will increase. Eventually the reaction slows down and the composition of the system stops changing. At this point the reaction is in its *equilibrium state*, and no further change in composition will occur as long as the system is left undisturbed.
- For many reactions, the equilibrium state is one in which components on both sides of the equation (that is, both reactants and products) are present in significant amounts. Such a reaction is said to be *incomplete* or *reversible*.
- **The equilibrium composition is independent of the direction from which it is approached**; the labeling of substances as "reactants" or "products" is entirely a matter of convenience. (See the hydrogen iodide reaction plots above.)
- The *law of mass action* states that any chemical change is a competition between a *forward* reaction (left-to-right in the chemical equation) and a *reverse* reaction. The rate of each of these processes is governed by the concentrations of the substances reacting; as the reaction proceeds, these rates approach each other and at equilibrium they become identical.
- From the above, it follows that **equilibrium is a** *dynamic process* in which *microscopic change* (the forward and reverse reactions) continues to occur, but *macroscopic change* (changes in the quantities of substances) is absent.
- When a chemical reaction is at equilibrium, any disturbance of the system, such as a change in temperature, or addition or removal of one of the reaction components, will "shift" the composition to a new equilibrium state. This is the only unambiguous way of verifying that a reaction is at equilibrium. The fact that the composition remains static does not in itself prove that a reaction is at equilibrium, because the change may be kinetically inhibited.

This page titled 11.1: Introduction to Chemical Equilibrium is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Le Chatelier's Principle

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- A system in its *equilibrium state* will remain in that state indefinitely as long as it is undisturbed. If the equilibrium is destroyed by subjecting the system to a change of pressure, temperature, or the number of moles of a substance, then a net reaction will tend to take place that moves the system to a new equilibrium state. *Le Chatelier's principle* says that this net reaction will occur in a direction that partially offsets the change.
- The Le Chatelier Principle has practical effect only for reactions in which signficant quantities of both reactants and products are present at equilibrium— that is, for reactions that are *thermodynamically reversible*.
- Addition of more product substances to an equilibrium mixture will shift the equilibrium to the left; addition of more reactant substances will shift it to the right. These effects are easily explained in terms of competing forward- and reverse reactions— that is, by the *law of mass action*.
- If a reaction is *exothermic* (releases heat), an increase in the **temperature** will force the equilibrium to the left, causing the system to absorb heat and thus partially ofsetting the rise in temperature. The opposite effect occurs for *endothermic* reactions, which are shifted to the right by rising temperature.
- The effect of **pressure** on an equilibrium is significant only for reactions which involve different numbers of moles of gases on the two sides of the equation. If the number of moles of gases increases, than an increase in the total pressure will tend to initiate a reverse reaction that consumes some the products, partially reducing the effect of the pressure increase.
- The classic example of the practical use of the Le Chatelier principle is the *Haber-Bosch process* for the synthesis of ammonia, in which a balance between low temperature and high pressure must be found.

The previous Module emphasized the dynamic character of equilibrium as expressed by the Law of Mass Action. This law serves as a model explaining how the composition of the equilibrium state is affected by the "active masses" (concentrations) of reactants and products. In this lesson, we develop the consequences of this law to answer the very practical question of how an existing equilibrium composition is affected by the addition or withdrawal of one of the components.

Le Chatelier's Principle

If a reaction is at equilibrium and we alter the conditions so as to create a new *equilibrium state*, then the composition of the system will tend to change until that new equilibrium state is attained. (We say "tend to change" because if the reaction is *kinetically inhibited*, the change may be too slow to observe or it may never take place.) In 1884, the French chemical engineer and teacher **Henri Le Chatelier** (1850-1936) showed that in every such case, the new equilibrium state is one that partially reduces the effect of the change that brought it about.

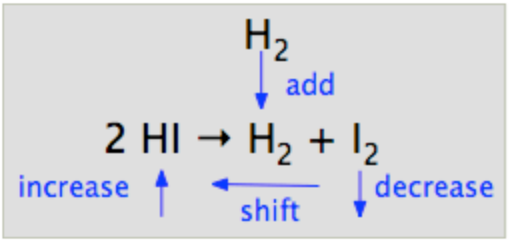
This law is known to every Chemistry student as the *Le Chatelier principle*. His original formulation was somewhat complicated, but a reasonably useful paraphrase of it reads as follows:

> **Le Chatelier principle**: If a system at equilibrium is subjected to a change of pressure, temperature, or the number of moles of a component, there will be a tendency for a net reaction in the direction that reduces the effect of this change.

To see how this works (and you *must* do so, as this is of such fundamental importance that you simply cannot do any meaningful chemistry without a thorough working understanding of this principle), look again at the hydrogen iodide dissociation reaction

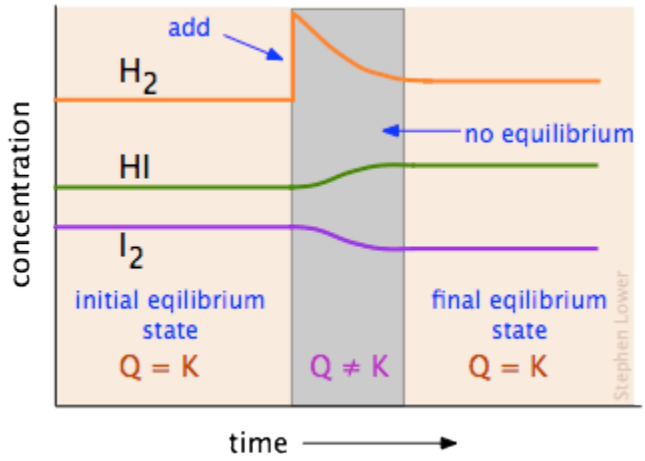
$$2HI \rightarrow H_2 + I_2 \tag{11.2.1}$$

Consider an arbitrary mixture of these three components at equilibrium, and assume that we inject more hydrogen gas into the container. Because the H_2 concentration now exceeds its new equilibrium value, the system is no longer in its equilibrium state, so a net reaction now ensues as the system moves to the new state.



The Le Chatelier principle states that the net reaction will be in a direction that tends to reduce the effect of the added H_2. This can occur if some of the H_2 is consumed by reacting with I_2 to form more HI; in other words, a net reaction occurs in the reverse direction. Chemists usually simply say that "the equilibrium shifts to the left".

To get a better idea of how this works, carefully examine the diagram below which follows the concentrations of the three components of this reaction as they might change in time (the time scale here will typically be about an hour):



Disruption and restoration of equilibrium. At the left, the concentrations of the three components do not change with time because the system is at equilibrium. We then add more hydrogen to the system, disrupting the equilibrium. A net reaction then ensues that moves the system to a new equilibrium state (right) in which the quantity of hydrogen iodide has increased; in the process, some of the I_2 and H_2 are consumed. Notice that the new equilibrium state contains more hydrogen than did the initial state, but not as much as was added; as the Le Chatelier principle predicts, the change we made (addition of H_2) has been partially counteracted by the "shift to the right". Table 11.2.1 Contains several examples showing how changing the quantity of a reaction component can shift an established equilibrium.

Table 11.2.1: Example of Le Chatelier's principle

system	change	result
$CO_2 + H_2 \rightarrow H_2O(g) + CO$	a drying agent is added to absorb H_2O	**Shift to the right**. Continuous removal of a product will force any reaction to the right
$H_2(g) + I_2(g) \rightarrow 2HI(g)$	Some nitrogen gas is added	**No change**; N_2 is not a component of this reaction system.
$NaCl(s) + H_2SO_4(l) \rightarrow Na_2SO_4(s) + HCl(g)$	reaction is carried out in an open container	Because HCl is a gas that can escape from the system, the reaction is forced to the **right**. This is the basis for the commercial production of hydrochloric acid.
$H_2O(l) \rightarrow H_2O(g)$	water evaporates from an open container	Continuous removal of water vapor forces the reaction to the **right**, so equilibrium is never achieved.

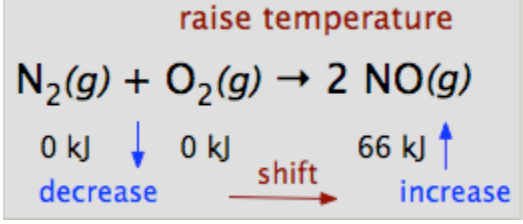
system	change	result
$HCN(aq) \rightarrow H^+(aq) + CN^-(aq)$	the solution is diluted	Shift to **right**; the product $[H^+][CN^-]$ diminishes more rapidly than does [HCN].
$AgCl(s) \rightarrow Ag^+(aq) + Cl^-(aq)$	some NaCl is added to the solution	Shift to **left** due to increase in Cl^- concentration. This is known as the common ion effect on solubility.
$N_2 + 3\ H_2 \rightarrow 2\ NH_3$	a catalyst is added to speed up this reaction	**No change**. Catalysts affect only the rate of a reaction; the have no effect at all on the composition of the equilibrium state.

How do changes in temperature affect Equilibria?

Virtually all chemical reactions are accompanied by the liberation or uptake of heat. If we regard heat as a "reactant" or "product" in an endothermic or exothermic reaction respectively, we can use the Le Chatelier principle to predict the direction in which an increase or decrease in temperature will shift the equilibrium state. Thus for the oxidation of nitrogen, an endothermic process, we can write

$$[\text{heat}] + N_2 + O_2 \rightleftharpoons 2NO \tag{11.2.2}$$

Suppose this reaction is at equilibrium at some temperature T_1 and we raise the temperature to T_2. The Le Chatelier principle tells us that a net reaction will occur in the direction that will partially counteract this change. Since the reaction is endothermic, a shift of the equilibrium to the right will take place.



Nitric oxide, the product of this reaction, is a major air pollutant which initiates a sequence of steps leading to the formation of atmospheric smog. Its formation is an unwanted side reaction which occurs when the air (which is introduced into the combustion chamber of an engine to supply oxygen) gets heated to a high temperature. Designers of internal combustion engines now try, by various means, to limit the temperature in the combustion region, or to restrict its highest-temperature part to a small volume within the combustion chamber.

How do changes in pressure affect equilibria?

You will recall that if the pressure of a gas is reduced, its volume will increase; pressure and volume are inversely proportional. With this in mind, suppose that the reaction

$$2NO_{2(g)} \rightleftharpoons N_2O_{4(g)} \tag{11.2.3}$$

is in equilibrium at some arbitrary temperature and pressure, and that we double the pressure, perhaps by compressing the mixture to a smaller volume. From the Le Chatelier principle we know that the equilibrium state will change to one that tends to counteract the increase in pressure. This can occur if some of the NO_2 reacts to form more of the dinitrogen tetroxide, since two moles of gas are being removed from the system for every mole of N_2O_4 formed, thereby decreasing the total volume of the system. Thus increasing the pressure will shift this equilibrium to the right.

It is important to understand that changing the pressure will have a significant effect only on reactions in which there is a change in the number of moles of gas. For the above reaction, this change

$$\Delta n_g = (n_{products} - n_{reactants}) = 1 - 2 = -1. \tag{11.2.4}$$

In the case of the nitrogen oxidation reaction

$$N_2 + O_2 \rightleftharpoons 2NO \tag{11.2.5}$$

Δn_g = 0 and changing pressure will have no effect on the equilibrium.

The volumes of solids and liquids are hardly affected by the pressure at all, so for reactions that do not involve gaseous substances, the effects of pressure changes are ordinarily negligible. Exceptions arise under conditions of very high pressure such as exist in the interior of the Earth or near the bottom of the ocean. A good example is the dissolution of calcium carbonate

$$CaCO_{3(s)} \rightleftharpoons Ca^{2+} + CO_3^{2-} \tag{11.2.6}$$

There is a slight decrease in the volume when this reaction takes place, so an increase in the pressure will shift the equilibrium to the right, with the results that calcium carbonate becomes more soluble at higher pressures.

The skeletons of several varieties of microscopic organisms that inhabit the top of the ocean are made of $CaCO_3$, so there is a continual rain of this substance toward the bottom of the ocean as these organisms die. As a consequence, the floor of the Atlantic ocean is covered with a blanket of calcium carbonate. This is not true for the Pacific ocean, which is deeper; once the skeletons fall below a certain depth, the higher pressure causes them to dissolve. Some of the seamounts (undersea mountains) in the Pacific extend above the solubility boundary so that their upper parts are covered with $CaCO_3$ sediments.

The effect of pressure on a reaction involving substances whose boiling points fall within the range of commonly encountered temperature will be sensitive to the states of these substances at the temperature of interest. For reactions involving gases, **only** changes in the partial pressures of those gases directly involved in the reaction are important; the presence of other gases has no effect.

Example 11.2.1: steam reforming of methane

The commercial production of hydrogen is carried out by treating natural gas with steam at high temperatures and in the presence of a catalyst ("steam reforming of methane"):

$$CH_4 + H_2O \rightleftharpoons CH_3OH + H_2$$

Given the following boiling points: CH_4 (methane) = –161°C, H_2O = 100°C, CH_3OH = 65°, H_2 = –253°C, predict the effects of an increase in the total pressure on this equilibrium at 50°, 75° and 120°C.

Solution

Calculate the change in the moles of gas for each process:

temp	equation	Δn_g	shift
50°	$CH_4(g) + H_2O(l) \rightarrow CH_3OH(l) + H_2(g)$	0	none
75°	$CH_4(g) + H_2O(l) \rightarrow CH_3OH(g) + H_2(g)$	+1	to left
120°	$CH_4(g) + H_2O(g) \rightarrow CH_3OH(g) + H_2(g)$	0	none

The Haber Process and why is it important

The Haber process for the synthesis of ammonia is based on the exothermic reaction

$N_2(g) + 3\ H_2(g) \rightarrow 2\ NH_3(g)\ \Delta H = -92$ kJ/mol

The Le Chatelier principle tells us that in order to maximize the amount of product in the reaction mixture, it should be carried out at high pressure and low temperature. However, the lower the temperature, the slower the reaction (this is true of virtually all chemical reactions.) As long as the choice had to be made between a low yield of ammonia quickly or a high yield over a long period of time, this reaction was infeasible economically.

Nitrogen is available for free, being the major component of air, but the strong triple bond in N_2 makes it extremely difficult to incorporate this element into species such as NO_3^- and NH_4^+ which serve as the starting points for the wide variety of nitrogen-containing compounds that are essential for modern industry. This conversion is known as *nitrogen fixation*, and because nitrogen is an essential plant nutrient, modern intensive agriculture is utterly dependent on huge amounts of fixed nitrogen in the form of fertilizer. Until around 1900, the major source of fixed nitrogen was the $NaNO_3$ found in extensive deposits in South America. Several chemical processes for obtaining nitrogen compounds were developed in the early 1900's, but they proved too inefficient to meet the increasing demand.

Although the direct synthesis of ammonia from its elements had been known for some time, the yield of product was found to be negligible. In 1905, Fritz Haber (1868-1934) began to study this reaction, employing the thinking initiated by Le Chatelier and others, and the newly-developing field of thermodynamics that served as the basis of these principles. From the Le Chatelier law alone, it is apparent that this exothermic reaction is favored by low temperature and high pressure. However, it was not as simple as that: the rate of any reaction increases with the temperature, so working with temperature alone, one has the choice between a high product yield achieved only very slowly, or a very low yield quickly. Further, the equipment and the high-strength alloy steels need to build it did not exist at the time. Haber solved the first problem by developing a *catalyst* that would greatly speed up the reaction at lower temperatures.

The second problem, and the development of an efficient way of producing hydrogen, would delay the practical implementation of the process until 1913, when the first plant based on the Haber-Bosch process (as it is more properly known, Carl Bosch being the person who solved the major engineering problems) came into operation. The timing could not have been better for Germany, since this country was about to enter the First World War, and the Allies had established a naval blockade of South America, cutting off the supply of nitrate for the the German munitions industry.

Bosch's plant operated the ammonia reactor at 200 atm and 550°C. Later, when stronger alloy steels had been developed, pressures of 800-1000 atm became common. The source of hydrogen in modern plants is usually natural gas, which is mostly methane:

$CH_4 + H_2O \rightarrow CO + 3\ H_2$	formation of synthesis gas from methane
$CO + H_2O \rightarrow CO_2 + H_2$	*shift reaction* carried out in reformer

The Haber-Bosch process is considered the most important chemical synthesis developed in the 20th century. Besides its scientific importance as the first large-scale application of the laws of chemical equilibrium, it has had tremendous economic and social impact; without an inexpensive source of fixed nitrogen, the intensive crop production required to feed the world's growing population would have been impossible. Haber was awarded the 1918 Nobel Prize in Chemistry in recognition of his work. Carl Bosch, who improved the process, won the Nobel Prize in 1931.

Many of the chemical reactions that occur in living organisms are regulated through the Le Chatelier principle.

Oxygen transport by the blood

Few of these are more important to warm-blooded organisms than those that relate to aerobic respiration, in which oxygen is transported to the cells where it is combined with glucose and metabolized to carbon dioxide, which then moves back to the lungs from which it is expelled.

hemoglobin + O_2 $\rightleftharpoons$ oxyhemoglobin

The partial pressure of O_2 in the air is 0.2 atm, sufficient to allow these molecules to be taken up by hemoglobin (the red pigment

of blood) in which it becomes loosely bound in a complex known as oxyhemoglobin. At the ends of the capillaries which deliver the blood to the tissues, the O_2 concentration is reduced by about 50% owing to its consumption by the cells. This shifts the equilibrium to the left, releasing the oxygen so it can diffuse into the cells.

Maintence of blood pH

Carbon dioxide reacts with water to form a weak acid H_2CO_3 which would cause the blood pH to fall to dangerous levels if it were not promptly removed as it is excreted by the cells. This is accomplished by combining it with carbonate ion through the reaction

$$H_2CO_3 + CO_3^{2-} \rightleftharpoons 2HCO_3^-$$

which is forced to the right by the high local CO_2 concentration within the tissues. Once the hydrogen carbonate (bicarbonate) ions reach the lung tissues where the CO_2 partial pressure is much smaller, the reaction reverses and the CO_2 is expelled.

Carbon monoxide poisoning

Carbon monoxide, a product of incomplete combustion that is present in automotive exhaust and cigarette smoke, binds to hemoglobin 200 times more tightly than does O_2. This blocks the uptake and transport of oxygen by setting up a competing equilibrium

O_2-hemoglobin $\rightleftharpoons$ hemoglobin $\rightleftharpoons$ CO-hemoglobin

Air that contains as little as 0.1 percent carbon monoxide can tie up about half of the hemoglobin binding sites, reducing the amount of O_2 reaching the tissues to fatal levels. Carbon monoxide poisoning is treated by administration of pure O_2 which promotes the shift of the above equilibrium to the left. This can be made even more effective by placing the victim in a hyperbaric chamber in which the pressure of O_2 can be made greater than 1 atm.

This page titled 11.2: Le Chatelier's Principle is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

11.3: Reaction Quotient

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- When arbitrary quantities of the different *components* of a *chemical reaction system* are combined, the overall system composition will not likely correspond to the equilibrium composition. As a result, a net change in composition ("a shift to the right or left") will tend to take place until the *equilibrium state* is attained.
- The status of the reaction system in regard to its equilibrium state is characterized by the value of the *equilibrium expression* whose formulation is defined by the coefficients in the balanced reaction equation; it may be expressed in terms of concentrations, or in the case of gaseous components, as partial pressures.
- The various terms in the equilibrium expression can have any arbitrary value (including zero); the value of the equilibrium expression itself is called the *reaction quotient* **Q.**
- If the concentration or pressure terms in the equilibrium expression correspond to the equilibrium state of the system, then **Q** has the special value ***K***, which we call the *equilibrium constant*.
- The ratio of Q/K (whether it is 1, >1 or <1) thus serves as an index of how far the system is from its equilibrium composition, and its value indicates the direction in which the net reaction must proceed in order to reach its equilibrium state.
- When *Q* = *K*, then the equilibrium state has been reached, and no further net change in composition will take place as long as the system remains undisturbed.

Consider a simple reaction such as the gas-phase synthesis of hydrogen iodide from its elements:

$$H_2 + I_2 \rightarrow 2HI \tag{11.3.1}$$

Suppose you combine arbitrary quantities of H_2, I_2 and HI. Will the reaction create more HI, or will some of the HI be consumed as the system moves toward its equilibrium state? The concept of the *reaction quotient*, which is the focus of this short lesson, makes it easy to predict what will happen.

What is the Equilibrium Quotient?

In the previous section we defined the *equilibrium expression* for the reaction

$$aA + bB \rightleftharpoons cC + dD \tag{11.3.2}$$

as

$$K = \underbrace{\frac{a_C^c a_D^d}{a_A^a a_b^b}}_{\substack{\text{in terms} \\ \text{of activities}}} \approx \underbrace{\frac{[C]^c[D]^d}{[A]^a[B]^b}}_{\substack{\text{in terms} \\ \text{of concetrations}}} \tag{11.3.3}$$

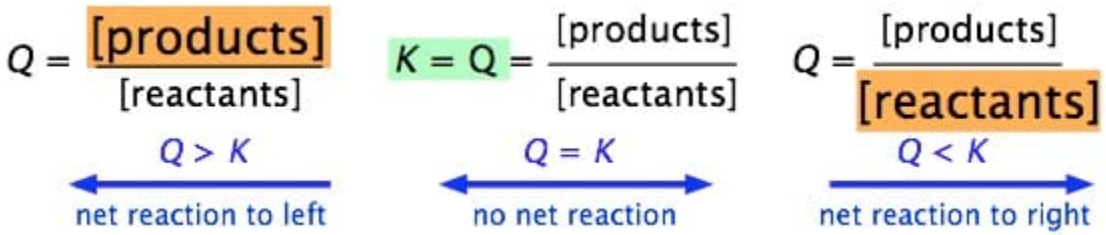
In the general case in which the concentrations can have any arbitrary values (including zero), this expression is called the *reaction quotient* (the term *equilibrium quotient* is also commonly used.) and its value is denoted by Q (or Q_c or Q_p if we wish to emphasize that the terms represent molar concentrations or partial pressures.) If the terms correspond to *equilibrium* concentrations, then the above expression is called the *equilibrium constant* and its value is denoted by K (or K_c or K_p).

> K is thus the special value that Q has when the reaction is at equilibrium

The value of Q in relation to *K* serves as an index how the composition of the reaction system compares to that of the equilibrium state, and thus it indicates the direction in which any net reaction must proceed. For example, if we combine the two reactants A and B at concentrations of 1 mol L^{-1} each, the value of Q will be 0÷1=0. The only possible change is the conversion of some of these reactants into products. If instead our mixture consists only of the two products C and D, Q will be indeterminately large (1÷0) and the only possible change will be in the reverse direction.

For example, if we combine the two reactants A and B at concentrations of 1 mol L^{-1} each, the value of Q will be 0÷1=0. The only possible change is the conversion of some of these reactants into products. If instead our mixture consists only of the two products C and D, Q will be indeterminately large (1÷0) and the only possible change will be in the reverse direction.

It is easy to see (by simple application of the Le Chatelier principle) that the ratio of *Q*/*K* immediately tells us whether, and in which direction, a net reaction will occur as the system moves toward its equilibrium state. A schematic view of this relationship is shown below:



It is very important that you be able to work out these relations for yourself, not by memorizing them, but from the definitions of Q and K.

Table 11.3.1*: Comparison of Q and K*

Condition	Status of System
Q > K	Product concentration too high for equilibrium; net reaction proceeds to *left*.
Q = K	System is at equilibrium; no net change will occur.
Q < K	Product concentration too low for equilibrium; net reaction proceeds to *right*.

It is very important that you be able to work out these relations for yourself, not by memorizing them, but from the definitions of Q and K.

Example 11.3.1

The equilibrium constant for the oxidation of sulfur dioxide is K_p = 0.14 at 900 K.

$$2\,SO_2\,(g) + O_2\,(g) \rightleftharpoons 2\,SO_3\,(g)$$

If a reaction vessel is filled with SO_3 at a partial pressure of 0.10 atm and with O_2 and SO_2 each at a partial pressure of 0.20 atm, what can you conclude about whether, and in which direction, any net change in composition will take place?

Solution:

The value of the equilibrium quotient Q for the initial conditions is

$$Q = \frac{p^2_{SO_3}}{p_{O_2}p^2_{SO_2}} = \frac{(0.10\ atm)^2}{(0.20\ atm)(0.20\ atm)^2} = 1.25\ atm^{-1}$$

Since *Q* > *K*, the reaction is not at equilibrium, so a net change will occur in a direction that decreases Q. This can only occur if some of the SO_3 is converted back into products. In other words, the reaction will "shift to the left".

The formal definitions of Q and *K* are quite simple, but they are of limited usefulness unless you are able to relate them to real chemical situations. The following diagrams illustrate the relation between Q and *K* from various standpoints. Take some time to study each one carefully, making sure that you are able to relate the description to the illustration.

Example 11.3.2: Dissociation of dinitrogen tetroxide

For the reaction

$$N_2O_{4(g)} \rightleftharpoons 2NO_{2(g)}$$

K_c = 0.0059 at 298 K.

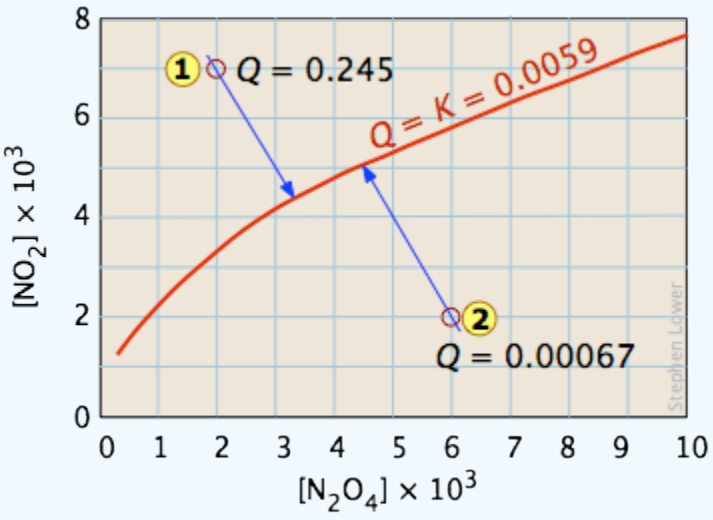
This equilibrium condition is represented by the red curve that passes through all points on the graph that satisfy the requirement that

$$Q = \frac{[NO_2]^2}{[N_2O_4]} = 0.0059$$

There are of course an infinite number of possible *Q*'s of this system within the concentration boundaries shown on the plot. Only those points that fall on the red line correspond to *equilibrium states* of this system (those for which $Q = K_c$). The line itself is a plot of [NO_2] that we obtain by rearranging the equilibrium expression

$$[NO_2] = \sqrt{[N_2O_4]K_c}$$

If the system is initially in a non-equilibrium state, its composition will tend to change in a direction that moves it to one that is on the line. Two such non-equilibrium states are shown. The state indicated by (1) has $Q > K$, so we would expect a net reaction that reduces Q by converting some of the NO_2 into N_2O_4; in other words, the equilibrium "shifts to the left". Similarly, in state (2), *Q* < *K*, indicating that the forward reaction will occur.



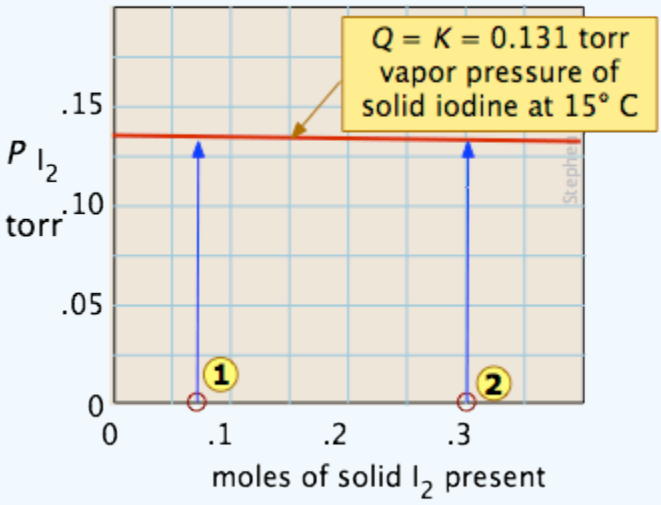
The blue arrows in the above diagram indicate the successive values that Q assumes as the reaction moves closer to equilibrium. The slope of the line reflects the stoichiometry of the equation. In this case, one mole of reactant yields two moles of products, so the slopes have an absolute value of 2:1.

Example 11.3.3: Phase-change equilibrium

One of the simplest equilibria we can write is that between a solid and its vapor. In this case, the equilibrium constant is just the vapor pressure of the solid. Thus for the process

$$I_{2(s)} \rightleftharpoons I_{2(g)}$$

all possible equilibrium states of the system lie on the horizontal red line and is independent of the quantity of solid present (as long as there is at least enough to supply the relative tiny quantity of vapor.)

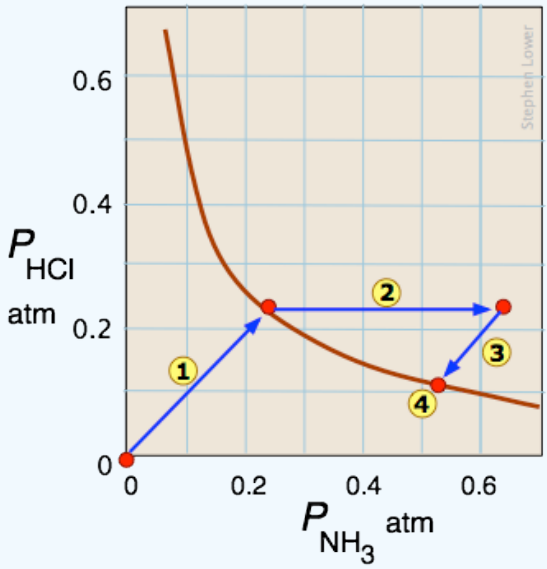


So adding various amounts of the solid to an empty closed vessel (states 1 and 2) causes a gradual buildup of iodine vapor. Because the equilibrium pressure of the vapor is so small, the amount of solid consumed in the process is negligible, so the arrows go straight up and all lead to the same equilibrium vapor pressure.

Example 11.3.4: Heterogeneous chemical reaction

The decomposition of ammonium chloride is a common example of a heterogeneous (two-phase) equilibrium. Solid ammonium chloride has a substantial vapor pressure even at room temperature:

$$NH_4Cl_{(s)} \rightleftharpoons NH_{3(g)} + HCl_{(g)} \tag{11.3.4}$$



Arrow 1 traces the states the system passes through when solid NH_4Cl is placed in a closed container. Arrow 2 represents the addition of ammonia to the equilibrium mixture; the system responds by following the path 3 back to a new equilibrium state 4 which, as the Le Chatelier principle predicts, contains a smaller quantity of ammonia than was added. The unit slopes of the paths 1 and 3 reflect the 1:1 stoichiometry of the gaseous products of the reaction.

This page titled 11.3: Reaction Quotient is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

11.4: Equilibrium Expressions

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- The *equilibrium quotient* Q is the value of the *equilibrium expression* of a reaction for any arbitrary set of concentrations or partial pressures of the reaction components.
- The *equilibrium constant K* is the value of Q when the reaction is at equilibrium. K has a unique value for a given reaction at a fixed temperature and pressure.
- Q and *K* can be expressed in terms of concentrations, partial pressures, or, when appropriate, in some combination of these.
- For a reaction in which all the components are gases, Q_c and K_c will have different values except in the special case in which the total number of moles of gas does not change.
- Concentration terms for substances whose concentrations do not change in the reaction do not appear in equilibrium expressions. The most common examples are [H_2O] when the reaction takes place in aqueous solution (so that [H_2O] is effectively constant at 55.6 *M*), and in heterogeneous reactions involving solids, in which the concentration of the solid is determined by the density of the solid itself.
- A reaction whose equilibrium constant is in the range of about 0.01 to 100 is said to be incomplete or [thermodynamically] *reversible*.
- Q and *K* are conventionally treated as dimensionless quantities, and need not ordinarily have *units* associated with them.
- *Heterogeneous* reactions are those in which two or more phases are involved; *homogeneous* reactions take place in a single phase. A common type of heterogeneous reaction is the loss of water of crystallization by a solid hydrate such as $CuSO_4 \cdot 5H_2O$.
- The equilibrium expression can be manipulated and combined in the following ways:
  - If the reaction is written in reverse, Q becomes Q^{-1};
  - If the coefficients of an equation are multiplied by *n*, Q becomes Q^n;
  - Q for the sum of two reactions (that is, for two reactions that take place in sequence) is the *product* $(Q_1)(Q_2)$.

You know that an equilibrium constant expression looks something like K = [products] / [reactants]. But how do you translate this into a format that relates to the actual chemical system you are interested in? This lesson will show you how to write the equilibrium constant expressions that you will need to use when dealing with the equilibrium calculation problems in the chapter that follows this one.

Pressures can Express Concentrations

Although we commonly write equilibrium quotients and equilibrium constants in terms of molar concentrations, any concentration-like term can be used, including mole fraction and molality. Sometimes the symbols K_c, K_x, and K_m are used to denote these forms of the equilibrium constant. Bear in mind that the *numerical values* of *K*'s and Q's expressed in these different ways will not generally be the same.

Most of the equilibria we deal with in this course occur in liquid solutions and gaseous mixtures. We can express K_c values in terms of moles per liter for both, but when dealing with gases it is often more convenient to use partial pressures. These two measures of concentration are of course directly proportional:

$$c = \frac{n}{V} = \frac{\frac{PV}{RT}}{V} = \frac{P}{RT} \tag{11.4.1}$$

so for a reaction $A_{(g)} \rightarrow B_{(g)}$ we can write the equilibrium constant as

$$K_p = \frac{P_B}{P_A} \tag{11.4.2}$$

Owing to interactions between molecules, especially when ions are involved, all of these forms of the equilibrium constant are only approximately correct, working best at low concentrations or pressures. The only equilibrium constant that is truly "constant" (except that it still varies with the temperature!) is expressed in terms of *activities*, which you can think of as "effective concentrations" that allow for interactions between molecules. In practice, this distinction only becomes important for equilibria involving gases at very high pressures (such as are often encountered in chemical engineering) and in ionic solutions more concentrated than about 0.001 *M*. We will not deal much with activities in this course.

For a reaction such as

$$CO_{2\,(g)} + OH^-_{(aq)} \rightleftharpoons HCO^-_{3\,(aq)} \tag{11.4.3}$$

that involves both gaseous and dissolved components, a "hybrid" equilibrium constant is commonly used:

$$K = \frac{[HCO_3^-]}{P_{CO_2}[OH^-]} \tag{11.4.4}$$

Clearly, it is essential to be sure of the units when you see an equilibrium constant represented simply by "K".

A note about pressure and concentration units

In this lesson (and in most of the others in this set,) we express concentrations in mol L^{-1} and pressures in atmospheres. Although this reflects common usage among chemists (older ones, especially!), these units are not part of the SI system which has been the international standard since the latter part of the 20th Century.

Molar concentrations are now more properly expressed in mol dm^{-3} and the "standard **atmosphere**" corresponds to a pressure of 101.325 kPa. Until 1990, 1 atm was the "standard pressure" employed in calculations involving the gas laws, and also in thermodynamics. Since that date, "standard pressure" has been 100.000 kPa, also expressed as 1 bar.

For most practical purposes, the differences between these values are so small that they can be neglected.

Converting between K_p and K_c

It is sometimes necessary to convert between equilibrium constants expressed in different units. The most common case involves pressure- and concentration equilibrium constants. Note that when *V* is expressed in liters and *P* in atmospheres, *R* must have the value 0.08206 L-atm/mol K.). The ideal gas law relates the partial pressure of a gas to the number of moles and its volume:

$$PV = nRT \tag{11.4.5}$$

Concentrations are expressed in moles/unit volume *n/V*, so by rearranging the above equation we obtain the explicit relation of pressure to concentration:

$$P = \left(\frac{n}{V}\right) RT \tag{11.4.6}$$

Conversely,

$$c = \frac{n}{V} = \frac{P}{RT} \tag{11.4.7}$$

so a concentration of a gas [A] can be expressed as $\frac{P_A}{RT}$.

For a reaction of the form $A + 3B \rightleftharpoons 2C$, we can write

$$K_p = \frac{(P_C)^2}{(P_A)\,(P_B)^3} = \frac{([C]RT)^2}{([A]RT)\,([B]RT)^3} = \underbrace{\frac{[C]^2}{[A]\,[B]^3}}_{K_c}\,(RT)^{\overbrace{-2}^{\Delta n_g}}$$

Assuming that all of the components are gases, the difference

$$(\text{moles of gas in products})–(\text{moles of gas in reactants}) = \Delta n_g$$

is given by

$$K_p = K_c(RT)^{\Delta n_g} \tag{11.4.8}$$

Do not show unchanging concentrations!

Substances whose concentrations undergo no significant change in a chemical reaction do not appear in equilibrium constant expressions. How can the concentration of a reactant or product *not* change when a reaction involving that substance takes place? There are two general cases to consider.

The substance is also the Solvent

This happens all the time in acid-base chemistry. Thus for the hydrolysis of the cyanide ion

$$\mathrm{CN^- + H_2O \rightleftharpoons HCN + OH^-}$$

we write

$$K_c = \frac{[\mathrm{HCN}][\mathrm{OH^-}]}{[\mathrm{CN^-}]}$$

in which no $[H_2O]$ term appears. The justification for this omission is that water is both the solvent and reactant, but only the tiny portion that acts as a reactant would ordinarily go in the equilibrium expression. The amount of water consumed in the reaction is so minute (because K is very small) that any change in the concentration of H_2O from that of pure water (55.6 mol L^{-1}) will be negligible.

Similarly, for the "autodissociation" of water

$$H_2O = H^+ + OH^-$$

the equilibrium constant is expressed as the "ion product"

$$K_w = [H^+][OH^-]$$

Be careful about throwing away H_2O whenever you see it. In the esterification reaction

$$\mathrm{CH_3COOH + C_2H_5OH \rightleftharpoons CH_3COOC_2H_5 + H_2O}$$

that we discussed in a previous section, a [H_2O] term must be present in the equilibrium expression if the reaction is assumed to be between the two liquids acetic acid and ethanol. If, on the other hand, the reaction takes place between a dilute aqueous solution of the acid and the alcohol, then the [H_2O] term would not be included.

The substance is a solid or a pure liquid phase

This is most frequently seen in solubility equilibria, but there are many other reactions in which solids are directly involved:

$$CaF_{2\,(s)} \rightarrow Ca^{2+}_{(aq)} + 2F^{-}_{(aq)} \tag{11.4.9}$$

$$Fe_3O_4(s) + 4H_{2\,(g)} \rightarrow 4H_2O_{(g)} + 3Fe_{(s)} \tag{11.4.10}$$

These are *heterogeneous reactions* (meaning reactions in which some components are in different phases), and the argument here is that concentration is only meaningful when applied to a substance within a single phase.

Thus the term $[CaF_2]$ would refer to the "concentration of calcium fluoride within the solid CaF_2", which is a constant depending on the molar mass of CaF_2 and the density of that solid. The concentrations of the two ions will be independent of the quantity of solid CaF_2 in contact with the water; in other words, the system can be in equilibrium as long as any CaF_2 at all is present. Throwing out the constant-concentration terms can lead to some rather sparse-looking equilibrium expressions. For example, the equilibrium expression for each of the processes shown in the following table consists solely of a single term involving the partial pressure of a gas:

Table 11.4.1*:*

$CaCO_{3(s)} \rightleftharpoons CaO_{(s)} + CO_{2(g)}$	$K_p = P_{CO_2}$	Thermal decomposition of limestone, a first step in the manufacture of cement.
$Na_2SO_4 \cdot 10H_2O_{(s)} Na_2SO_{4(s)} + 10H_2O_{(g)}$	$K_p = P^{10}_{H_2O}$	Sodium sulfate decahydrate is a solid in which H2O molecules ("waters of hydration") are incorporated into the crystal structure.)
$I_{2(s)} \rightleftharpoons I_{2(g)}$	$K_p = P_{I_2}$	sublimation of solid iodine; this is the source of the purple vapor you can see above solid iodine in a closed container.
$H_2O_{(l)} \rightleftharpoons H_2O_{(g)}$	$K_p = P_{H_2O}$	Vaporization of water. When the partial pressure of water vapor in the air is equal to K, the relative humidity is 100%.

The last two processes represent changes of state (*phase changes*) which can be treated exactly the same as chemical reactions. In each of the heterogeneous processes shown in Table 11.4.1, the reactants and products can be in equilibrium (that is, permanently coexist) only when the partial pressure of the gaseous product has the value consistent with the indicated K_p. Bear in mind also that these K_p's all increase with the temperature.

Example 11.4.1

What are the values of K_p for the equilibrium between liquid water and its vapor at 25°C, 100°C, and 120°C? The vapor pressure of water at these three temperatures is 23.8 torr, 760 torr (1 atm), and 1489 torr, respectively.

Comment: These vapor pressures are the partial pressures of water vapor in equilibrium with the liquid, so they are identical with the K_p's when expressed in units of atmospheres.

Solution

25°C

100°C

120°C

$K_p = \frac{23.8\ torr}{760\ torr/atm} = 0.031\ \ atm$ $K_p = 1.00\ \ atm$ $K_p = \frac{1489\ torr}{760\ torr/atm} = 1.96\ \ atm$ The partial pressure of H2O above the surface of liquid water in a closed container at 25°C will build up to this value. If the cover is removed so that this pressure cannot be maintained, the system will cease to be at equilibrium and the water will evaporate.

This temperature corresponds, of course, to the boiling point of water. The normal boiling point of a liquid is the temperature at

which the partial pressure of its vapor is 1 atm.

The only way to heat water above its normal boiling point is to do so in a closed container that can withstand the increased vapor pressure. Thus a pressure cooker that operates at 120°C must be designed to withstand an internal pressure of at least 2 atm.

Values of Equilibrium Constants

or "***reversible***".

As an equilibrium constant approaches the limits of zero or infinity, the reaction can be increasingly characterized as a one-way process; we say it is "***complete***" or "***irreversible***". The latter term must of course not be taken literally; the Le Chatelier principle still applies (especially insofar as temperature is concerned), but addition or removal of reactants or products will have less effect.

Although it is by no means a general rule, it frequently happens that reactions having very large equilibrium constants are ***kinetically hindered***, often to the extent that the reaction essentially does not take place.

The examples in the following table are intended to show that numbers (values of *K*), no matter how dull they may look, do have practical consequences!

Table 11.4.2*: Examples of Reversible Reactions*

Reaction	K	remarks
$N_{2(g)} + O_{2(g)} \rightleftharpoons 2NO_{(g)}$	5×10^{-31} at 25°C, 0.0013 at 2100°C	These two very different values of *K* illustrate very nicely why reducing combustion-chamber temperatures in automobile engines is environmentally friendly.
$3H_{2(g)} + N_{2(g)} \rightleftharpoons 2NH_{3(g)}$	7×10^{5} at 25°C, 56 at 1300°C	See the discussion of this reaction in the section on the Haber process.
$H_{2(g)} \rightleftharpoons 2H_{(g)}$	10^{-36} at 25°C, 6×10^{-5} at 5000°	Dissociation of any stable molecule into its atoms is endothermic. This means that all molecules will decompose at sufficiently high temperatures.
$H_2O_{(g)} \rightleftharpoons H_{2(g)} + ½O_{2(g)}$	8×10^{-41} at 25°C	You won't find water a very good source of oxygen gas at ordinary temperatures!
$CH_3COOH_{(l)} \rightleftharpoons 2H_2O_{(l)} + 2C_{(s)}$	$K_c = 10^{13}$ at 25°C	This tells us that acetic acid has a great tendency to decompose to carbon, but nobody has ever found graphite (or diamonds!) forming in a bottle of vinegar. A good example of a super kinetically-hindered reaction!

Do Equilibrium Constants have Units?

The equilibrium expression for the synthesis of ammonia

$$3H_{2(g)} + N_{2(g)} \rightarrow 2NH_{3(g)} \tag{11.4.11}$$

can be expressed as

$$K_p = \frac{P_{NH_3}^2}{P_{N_2} P_{H_2}^3} \tag{11.4.12}$$

or

$$K_c = \frac{[NH_3]^2}{[N_2][H_2]^3} \tag{11.4.13}$$

so K_p for this process would appear to have units of atm^{-1}, and K_c would be expressed in mol^{-2} L^2. And yet these quantities are often represented as being dimensionless. Which is correct? The answer is that both forms are acceptable. There are some situations (which you will encounter later) in which *K*'s must be considered dimensionless, but in simply quoting the value of an equilibrium constant it is permissible to include the units, and this may even be useful in order to remove any doubt about the units of the individual terms in equilibrium expressions containing both pressure and concentration terms. In carrying out your own calculations, however, there is rarely any real need to show the units.

Strictly speaking, equilibrium expressions do not have units because the concentration or pressure terms that go into them are really *ratios* having the forms (*n* mol L^{-1})/(1 mol L^{-1}) or (*n* atm)/(1 atm) in which the unit quantity in the denominator refers to the *standard state* of the substance; thus the units always cancel out. (But first-year students are not expected to know this!)

For substances that are liquids or solids, the standard state is just the concentration of the substance within the liquid or solid, so for something like CaF*(s)*, the term going into the equilibrium expression is [CaF_2]/[CaF_2] which cancels to unity; this is the reason we don't need to include terms for solid or liquid phases in equilibrium expressions. The subject of standard states would take us beyond where we need to be at this point in the course, so we will simply say that the concept is made necessary by the fact that energy, which ultimately governs chemical change, is always relative to some arbitrarily defined zero value which, for chemical substances, is the standard state.

How the Reaction Equation affects K

It is important to remember that an equilibrium quotient or constant is always tied to a specific chemical equation, and if we write the equation in reverse or multiply its coefficients by a common factor, the value of Q or K will change. The rules are very simple:

- Writing the equation in **reverse** will invert the equilibrium expression;
- **Multiplying** the coefficients by a common factor will raise *Q* or *K* to the corresponding power.

Here are some of the possibilities for the reaction involving the equilibrium between gaseous water and its elements:

Example 1: $2\,H_2 + O_2 \rightleftharpoons 2\,H_2O$ with equilibrium expression

$$K_p = \frac{P_{H_2O}^2}{P_{H_2}^2 P_{O_2}}$$

Example 2: $10\,H_2 + 5\,O_2 \rightleftharpoons 10\,H_2O$ with equilibrium expression

$$K_p = \frac{P_{H_2O}^{10}}{P_{H_2}^{10} P_{O_2}^5} = \left(\frac{P_{H_2O}^2}{P_{H_2}^2 P_{O_2}} \right)^5$$

Example 3: $H_2 + \frac{1}{2}\,O_2 \rightleftharpoons H_2O$ with equilibrium expression

$$K_p = \frac{P_{H_2O}}{P_{H_2} P_{O_2}^{1/2}}$$

$$= \left(\frac{P_{H_2O}^2}{P_{H_2}^2 P_{O_2}} \right)^{1/2}$$

Example 4: $\mathrm{H_2O} \rightleftharpoons \mathrm{H_2} + \frac{1}{2}\,\mathrm{O_2}$ with equilibrium expression

$$K_p = \frac{P_{H_2} P_{O_2}^{1/2}}{P_{H_2O}}$$

$$= \left(\frac{P_{H_2O}^2}{P_{H_2}^2 P_{O_2}} \right)^{-1/2}$$

Many chemical changes can be regarded as the sum or difference of two or more other reactions. If we know the equilibrium constants of the individual processes, we can easily calculate that for the overall reaction according to the following rule.

Multi-step Equilibria Rule

The equilibrium constant for the **sum** of two or more reactions is the **product** of the equilibrium constants for each of the steps.

Example 11.4.2

Calculate the value of K for the reaction

$$\mathrm{CaCO_3\,(s) + H^+\,(aq) \rightleftharpoons Ca^{2+}\,(aq) + HCO_3^-\,(aq)}$$

given the following equilibrium constants:

$CaCO_{3(s)} \rightleftharpoons Ca^{2+}_{(aq)} + CO^{2-}_{3(aq)}$	$K_1 = 10^{-6.3}$
$HCO^-_{3(aq)} \rightleftharpoons H^+_{(aq)} + CO^{2-}_{3(aq)}$	$K_2 = 10^{-10.3}$

Solution

The net reaction is the sum of reaction 1 and the reverse of reaction 2:

$CaCO_{3(s)} \rightleftharpoons Ca^{2+}_{(aq)} + CO^{2-}_{3(aq)}$	$K_1 = 10^{-6.3}$
$H^+_{(aq)} + CO^{2-}_{3(aq)} \rightleftharpoons HCO^-_{3(aq)}$	$K_{-2} = 10^{-(-10.3)}$
$CaCO_{3(s)} + H^+_{(aq)} \rightarrow Ca^{2+}_{(aq)} + HCO^-_{3(aq)}$	$K = \frac{K_1}{K_2} = 10^{(-6.4+10.3)} = 10^{+3.9}$

Comment:

This net reaction describes the dissolution of limestone by acid; it is responsible for the eroding effect of acid rain on buildings and statues. This an example of a reaction that has practically no tendency to take place by itself (small K_1) being "driven" by a second reaction having a large equilibrium constant (K_{-2}). From the standpoint of the Le Chatelier principle, the first reaction is "pulled to the right" by the removal of carbonate by hydrogen ion. *Coupled* reactions of this type are widely encountered in all areas of chemistry, and especially in biochemistry, in which a dozen or so reactions may be linked.

Example 11.4.3

The synthesis of HBr from hydrogen and liquid bromine has an equilibrium constant $K_p = 4.5 \times 10^{15}$ at 25°C. Given that the vapor pressure of liquid bromine is 0.28 atm, find K_p for the homogeneous gas-phase reaction at the same temperature.

Solution

The net reaction we seek is the sum of the heterogeneous synthesis of HBr and the *reverse* of the vaporization of liquid bromine:

$H_{2(g)} + Br_{2(l)} \rightleftharpoons 2HBr_{(g)}$	$K_p = 4.5 \times 10^{15}$
$Br_{2(g)} \rightleftharpoons Br_{2(l)}$	$K_p = (0.28)^{-1}$
$H_{2(g)} + Br_{2(g)} \rightleftharpoons 2HBr_{(g)}$	$K_p = 1.6 \times 10^{16}$

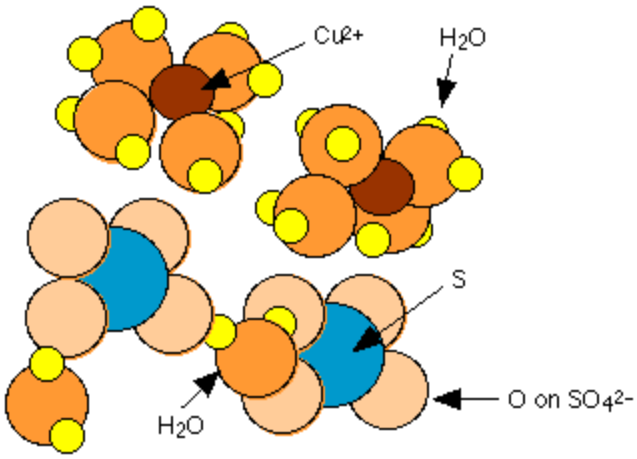
More on heterogeneous reactions

Heterogeneous reactions are those involving more than one phase. Some examples:

$Fe(s) + O_2(g) \rightleftharpoons FeO_2(s)$	**air-oxidation of metallic iron (formation of rust)**
\(CaF_2(s) \rightleftharpoons Ca(aq) + F^+(aq)_	dissolution of calcium fluoride in water
$H_2O(s) \rightleftharpoons H_2O(g)$	sublimation of ice (a *phase change*)
$NaHCO_3(s) + H^+(aq) \rightleftharpoons CO_2(g) + Na^+(aq) + H_2O(g)$	formation of carbon dioxide gas from sodium bicarbonate when water is added to baking powder (the hydrogen ions come from tartaric acid, the other component of baking powder.)

The vapor pressure of solid hydrates

A particularly interesting type of heterogeneous reaction is one in which a solid is in equilibrium with a gas. The sublimation of ice illustrated in the above table is a very common example. The equilibrium constant for this process is simply the partial pressure of water vapor in equilibrium with the solid— the *vapor pressure* of the ice.



Many common inorganic salts form solids which incorporate water molecules into their crystal structures. These water molecules are usually held rather loosely and can escape as water vapor. Copper(II) sulfate, for example forms a pentahydrate in which four of the water molecules are coordinated to the Cu^{2+} ion while the fifth is hydrogen-bonded to SO_4^{2-}. This latter water is more tightly bound, so that the pentahydrate loses water in two stages on heating:

$$\mathrm{CuSO_4 \cdot 5\,H_2O} \xrightarrow{140^{\circ}\mathrm{C}} \mathrm{CuSO_4 \cdot 5\,H_2O} \xrightarrow{400^{\circ}\mathrm{C}} \mathrm{CuSO_4}$$

These dehydration steps are carried out at the temperatures indicated above, but at any temperature, some moisture can escape from a hydrate. For the complete dehydration of the pentahydrate we can define an equilibrium constant:

$$\mathrm{CuSO_4 \cdot 5\,H_2O\,(s)} \rightleftharpoons \mathrm{CuSO_4\,(s) + 5\,H_2O\,(g)} \quad K_p = 1.14 \times 10^{10}$$

The vapor pressure of the hydrate (for this reaction) is the partial pressure of water vapor at which the two solids can coexist indefinitely; its value is $K_p^{1/5}$ atm. If a hydrate is exposed to air in which the partial pressure of water vapor is less than its vapor pressure, the reaction will proceed to the right and the hydrate will lose moisture. Vapor pressures always increase with temperature, so any of these compounds can be dehydrated by heating.

Loss of water usually causes a breakdown in the structure of the crystal; this is commonly seen with sodium sulfate, whose vapor pressure is sufficiently large that it can exceed the partial pressure of water vapor in the air when the relative humidity is low. What one sees is that the well-formed crystals of the decahydrate undergo deterioration into a powdery form, a phenomenon known as *efflorescence*.

Table 11.4.3*:*

name	formula	vapor pressure, torr	
		25°C	30°C
sodium sulfate decahydrate	$Na_2SO_4{\cdot}10H_2O$	19.2	25.3
copper(II) sulfate pentahydrate	$CuSO_4{\cdot}5H_2O$	7.8	12.5
calcium chloride monohydrate	$CaCl_2{\cdot}H_2O$	3.1	5.1
(water)	H_2O	23.5	31.6

Example 11.4.4

At what relative humidity will copper sulfate pentahydrate lose its waters of hydration when the air temperature is 30°C? What is K_p for this process at this temperature?

Solution

From the Table $11.4.3$, we see that the vapor pressure of the hydrate is 12.5 torr, which corresponds to a *relative humidity* of 12.5/31.6 = 0.40 or 40%. This is the humidity that will be maintained if the hydrate is placed in a closed container of dry air

For this hydrate, $K_p = \sqrt{p_{H_2O)}}$, so the partial pressure of water vapor that will be in equilibrium with the hydrate and the dehydrated solid (remember that both solids must be present to have equilibrium!), expressed in atmospheres, will be

$$\left(\frac{12.5}{760}\right)^5 = 1.20 \times 10^{-9}.$$

One of the first hydrates investigated in detail was calcium sulfate hemihydrate ($CaSO_4 \cdot ½\ H_2O$) which Le Chatelier (he of the "principle") showed to be the hardened form of $CaSO_4$ known as plaster of Paris. Anhydrous $CaSO_4$ forms compact, powdery crystals, whereas the elongated crystals of the hemihydrate bind themselves into a cement-like mass that makes this material useful for making art objects, casts for immobilizing damaged limbs, and as a construction material (fireproofing, drywall).

This page titled 11.4: Equilibrium Expressions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

11.5: Equilibrium Calculations

Learning Objectives

The five Examples presented above were carefully selected to span the range of problem types that students enrolled in first-year college chemistry courses are expected to be able to deal with. If you are able to reproduce these solutions on your own, you should be well prepared on this topic.

- The first step in the solution of all but the simplest equilibrium problems is to sketch out a table showing for each component the *initial* concentration or pressure, the *change* in this quantity (for example, +2*x*), and the *equilibrium* values (for example, .0036 + 2*x*). In doing so, the sequence of calculations required to get to the answer usually becomes apparent.
- Equilibrium calculations often involve quadratic- or higher-order equations. Because concentrations, pressures, and equilibrium constants are seldom known to a precision of more than a few significant figures, there is no need to seek exact solutions. Iterative approximations (as in Example $11.5.3$) or use of a graphical calculator (Example $11.5.4$) are adequate and convenient.
- Phase distribution equilibria play an important role in chemical separation processes on both laboratory and industrial scales. They are also involved in the movement of chemicals between different parts of the environment, and in the bioconcentration of pollutants in the food chain.

This page presents examples that cover most of the kinds of equilibrium problems you are likely to encounter in a first-year university course. Reading this page will ***not*** teach you how to work equilibrium problems! The only one who can teach you how to interpret, understand, and solve problems is **yourself**. So don't just "read" this and think you are finished. You need to find and solve similar problems on your own. Look over the problems in your homework assignment or at the end of the appropriate chapter in a textbook, and see how they fit into the general types described below. When you can solve them without looking at the examples below, you will be well on your way!

Calculating Equilibrium Constants

Clearly, if the concentrations or pressures of all the components of a reaction are known, then the value of K can be found by simple substitution. Observing individual concentrations or partial pressures directly may be not always be practical, however. If one of the components is colored, the extent to which it absorbs light of an appropriate wavelength may serve as an index of its concentration. Pressure measurements are ordinarily able to measure only the total pressure of a gaseous mixture, so if two or more gaseous products are present in the equilibrium mixture, the partial pressure of one may need to be inferred from that of the other, taking into account the stoichiometry of the reaction.

Example $11.5.1$

In an experiment carried out by Taylor and Krist (*J. Am. Chem. Soc.* 1941: 1377), hydrogen iodide was found to be 22.3% dissociated at 730.8 K. Calculate K_c for

$$2\,\mathrm{HI\,(g)} \rightleftharpoons \mathrm{H_2\,(g)} + \mathrm{I_2}$$

Solution

No explicit molar concentrations are given, but we do know that for every n moles of HI, $0.223n$ moles of each product is formed and $(1–0.223)n = 0.777n$ moles of HI remains. For simplicity, we assume that $n = 1$ and that the reaction is carried out in a 1.00-L vessel, so that we can substitute the required concentration terms directly into the equilibrium expression for K_c.

$$K_c = \frac{[\mathrm{H_2}][\mathrm{I_2}]}{[\mathrm{HI}]^2}$$
$$= \frac{(0.223)(0.223)}{(0.777)^2}$$
$$= 0.082$$

Example 11.5.2: Evaluating the Equilibrium Constant

Ordinary white phosphorus, $\mathrm{P_4}$, forms a vapor which dissociates into diatomic molecules at high temperatures:

$$\mathrm{P_4\,(g)} \rightleftharpoons 2\,\mathrm{P_2\,(g)}$$

A sample of white phosphorus, when heated to 1000°C, formed a vapor having a total pressure of 0.200 atm and a density of 0.152 g L^{-1}. Use this information to evaluate the equilibrium constant K_p for this reaction.

Solution

Before worrying about what the density of the gas mixture has to do with K_p, start out in the usual way by laying out the information required to express K_p in terms of an unknown x

ICE Table	$P_{4(g)}$	$P_{2(g)}$	comment
Initial (moles)	1	0	Since K is independent of the number of moles, assume the simplest initial case.
Change (moles)	$-x$	$2x$	*x* is the fraction of $\mathrm{P_4}$ that dissociates.
Equilibrium (moles)	$1-x$	$2x$	The denominator is the total number of moles:
Mole Fraction	$\chi_{P_4} = \frac{1-x}{1+x}$	$\chi_{P_2} = \frac{2x}{1+x}$	The denominator is the total number of moles:
Equilibrium (pressures)	$p_{P_4} = \chi_{P_4} 0.2 = \left(\frac{1-x}{1+x}\right) 0.2$	$p_{P_2} = \chi_{P_2} 0.2 = \left(\frac{2x}{1+x}\right) 0.2$	Partial pressure is the mole fraction times the total pressure.

The partial pressures in the bottom row were found by multiplying the mole fraction of each gas by the total pressure:

$$P_i = \chi_i P_{tot}$$

with the term in the denominator of each mole fraction is the total number of moles of gas present at equilibrium:

$$P_{tot} = 1 - x + 2x = 1 + x$$

Expressing the equilibrium constant in terms of x gives

$$K_p = \frac{p_{P_2}^2}{p_{P_4}}$$
$$= \frac{\left(\frac{2x}{1+x}\right)^2 0.2^2}{\left(\frac{1-x}{1+x}\right) 0.2}$$
$$= \left(\frac{4x^2}{(1-x)(1+x)}\right) 0.2$$
$$= \left(\frac{4x^2}{1+x^2}\right) 0.2$$

Now we need to find the dissociation fraction x of P_4, and at this point we hope you remember those gas laws that you were told you would be needing later in the course! The density of a gas is directly proportional to its molecular weight, so you need to calculate the densities of pure P_4 and pure P_2 vapors under the conditions of the experiment. One of these densities will be greater than 0.152 gL^{-1} and the other will be smaller; all you need to do is to find where the measured density falls in between the two limits, and you will have the dissociation fraction.

The molecular weight of phosphorus is 31.97, giving a molar mass of 127.9 g for $\mathrm{P_4}$. This mass must be divided by the volume to find the density; assuming ideal gas behavior, the volume of 127.9 g (1 mole) of $\mathrm{P_4}$ is given by *RT/P*, which works out to 522 L (remember to use the absolute temperature here.) The density of pure $\mathrm{P_4}$ vapor under the conditions of the experiment is then

$$\rho = \frac{m}{V} = (128\ g\ mol^{-1}) \times x = (522\ L mol^{-1}) = 0.245\ g\ L^{-1}$$

The density of pure P_2 would be half this, or 0.122 g L^{-1}. The difference between these two limiting densities is 0.123 g L^{-1}, and the difference between the density of pure $\mathrm{P_4}$ and that of the equilibrium mixture is (0.245 – 0.152) g L^{-1} or 0.093 g L^{-1}. The ratio 0.093/0.123 = 0.76 is therefore the fraction of $\mathrm{P_4}$ that remains and its fractional dissociation is (1 – 0.76) = 0.24. Substituting into the equilibrium expression above gives $K_p = 1.2$.

Exercise 11.5.2

Solve Example 11.5.2 by using a different set of initial conditions to demonstrated that the initial conditions indeed have no effect on determining the Equilibrium state and K_p.

Calculating Equilibrium Concentrations

This is by far the most common kind of equilibrium problem you will encounter: starting with an arbitrary number of moles of each component, how many moles of each will be present when the system comes to equilibrium? The principal source of confusion and error for beginners relates to the need to determine the values of several unknowns (a concentration or pressure for each component) from a single equation, the equilibrium expression. The key to this is to make use of the stoichiometric relationships between the various components, which usually allow us to express the equilibrium composition in terms of a single variable. The easiest and most error-free way of doing this is adopt a systematic approach in which you create and fill in a small table as shown in the following problem example. You then substitute the equilibrium values into the equilibrium constant expression, and solve it for the unknown.

This very often involves solving a quadratic or higher-order equation. Quadratics can of course be solved by using the familiar quadratic formula, but it is often easier to use an algebraic or graphical approximation, and for higher-order equations this is the only practical approach. There is almost never any need to get an exact answer, since the equilibrium constants you start with are rarely known all that precisely anyway.

Example 11.5.3

Phosgene ($\mathrm{COCl_2}$) is a poisonous gas that dissociates at high temperature into two other poisonous gases, carbon monoxide and chlorine. The equilibrium constant K_p = 0.0041 at 600°K. Find the equilibrium composition of the system after 0.124 atm of $\mathrm{COCl_2}$ is allowed to reach equilibrium at this temperature.

Solution

First we need a balanced chemical reaction

$$\mathrm{COCl_2} \rightleftharpoons \mathrm{CO\,(g) + Cl_2\,(g)}$$

Start by drawing up a table showing the relationships between the components:

ICE Table	$COCl_2$	$CO_{(g)}$	$Cl_{2(g)}$
Initial (pressure)	0.124 atm	0	0
Change (pressure)	-x	+x	+x
Equilibrium (pressure)	0.124 – *x*	+x	+x

Substitution of the equilibrium pressures into the equilibrium expression gives

$$\frac{x^2}{0.124 - x} = 0.0041$$

This expression can be rearranged into standard polynomial form

$$x^2 + 0.0041x–0.00054 = 0$$

and solved by the quadratic formula, but we will simply obtain an approximate solution by iteration. Because the equilibrium constant is small, we know that *x* will be rather small compared to 0.124, so the above relation can be approximated by

$$\frac{x^2}{0.124 - x} \approx \frac{x^2}{0.124} = 0.0041$$

which gives *x* = 0.0225. To see how good this is, substitute this value of x into the denominator of the original equation and solve again:

$$\frac{x^2}{0.124 - 0.0225} = \frac{x^2}{0.102} = 0.0041$$

This time, solving for x gives 0.0204. Iterating once more, we get

$$\frac{x^2}{0.124 - 0.0204} = \frac{x^2}{0.104} = 0.0041$$

and *x* = 0.0206 which is sufficiently close to the previous to be considered the final result. The final partial pressures are then 0.104 atm for $COCl_2$, and 0.0206 atm each for CO and Cl_2.

Comment: Using the quadratic formula to find the exact solution yields the two roots –0.0247 (which we ignore) and 0.0206, which show that our approximation is quite good.

Example 11.5.4

The gas-phase dissociation of phosphorus pentachloride to the trichloride has $K_p = 3.60$ at 540°C:

$$\mathrm{PCl_5\,(g) \rightleftharpoons PCl_3\,(g) + Cl_2\,(g)}$$

What will be the partial pressures of all three components if 0.200 mole of $\mathrm{PCl_5}$ and 3.00 moles of $\mathrm{PCl_3}$ are combined and brought to equilibrium at this temperature and at a total pressure of 1.00 atm?

Solution

As always, set up a table showing what you know (first two rows) and then expressing the equilibrium quantities:

ICE Table	$\mathrm{PC_5\,(g)}$	$\mathrm{PCl_3\,(g)}$	$\mathrm{Cl_2\,(g)}$
Initial (moles)	0.200	3.00	0
Change (moles)	–x	+x	+x
Equilibrium (moles)	0.200 – x	3.00 + x	x

ICE Table	$PC_5\,(g)$	$PCl_3\,(g)$	$Cl_2\,(g)$
Equilibrium (partial pressures)	$\frac{2.00-x}{3.20+x}$	$\frac{3.00+x}{3.20+x}$	$\frac{x}{3.20+x}$

The partial pressures in the bottom row were found by multiplying the mole fraction of each gas by the total pressure:

$$P_i = \chi_i P_{tot}$$

with the term in the denominator of each mole fraction is the total number of moles of gas present at equilibrium:

$$P_{tot} = (0.200–x) + (3.00 + x) + x = 3.20 + x$$

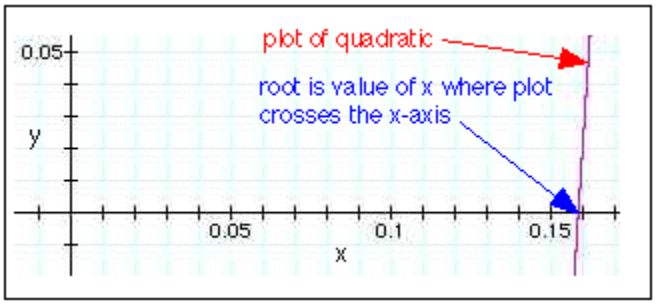
Substituting the equilibrium partial pressures into the equilibrium expression, we have

$$\frac{(3.00+x)(x)}{(0.200-x)(3.20+x)} = 3.60$$

whose polynomial form is

$$4.60x^2 + 13.80x–2.304 = 0.$$

You can use the quadratic question to solve this or you can do it graphically (more useful for higher order equations). Plotting this on a graphical calculator yields $x = 0.159$ as the positive root:



Substitution of this root into the expressions for the equilibrium partial pressures in the table yields the following values:

- P_{PCl_5} = 0.012 atm,
- P_{PCl_3} = 0.94 atm,
- P_{Cl_2} = 0.047 atm.

Effects of Dilution on Equilibrium

In the section that introduced the Le Chatelier principle, it was mentioned that diluting a weak acid such as acetic acid CH_3COOH ("HAc") will shift the dissociation equilibrium to the right:

$$HAc + H_2O \rightleftharpoons H_3O^+ + Ac^-$$

Thus a $0.10\ M$ solution of acetic acid is 1.3% ionized, while in a 0.01 M solution, 4.3% of the HAc molecules will be dissociated. This is because as the solution becomes more dilute, the product [H_3O^+][Ac^-] decreases more rapidly than does the $[HAc]$ term. At the same time the concentration of H_2O becomes greater, but because it is so large to start with (about 55.5 M), any effect this might have is negligible, which is why no $[H_2O]$ term appears in the equilibrium expression.

For a reaction such as

$$CH_3COOH\,(l) + C_2H_5OH\,(l) \rightleftharpoons CH_3COOC_2H_5\,(l) + H_2O\,(l) \tag{11.5.1}$$

(in which the water concentration does change), dilution will have no effect on the equilibrium; the situation is analogous to the

way the pressure dependence of a gas-phase reaction depends on the number of moles of gaseous components on either side of the equation (i.e., Δn_g).

Example 11.5.5

The biochemical formation of a disaccharide (double) sugar from two monosaccharides is exemplified by the reaction

$$\text{fructose}(aq) + \text{glucose-6-phosphate}(aq) \rightarrow \text{sucrose-6-phosphate}(aq) + \text{H}_2\text{O}(l) \tag{11.5.2}$$

(Sucrose is ordinary table sugar.) To what volume should a solution containing 0.050 mol of each monosaccharide be diluted in order to bring 5% conversion to sucrose phosphate? The equilibrium constant for this reaction is $K_c = 7.1 \times 10^{-6}$ at room temperature.

Solution

The initial and final numbers of moles in this equation are as follows:

ICE Table	fructose (fruc)	glucose-6-P (gluc6P)	sucrose-6-P (suc6P)	water H_2O)
Initial (moles)	0.05	0.05	0	-
Change (moles)	-x	-x	+x	-
Equilibrium (moles)	0.05-x	0.05-x	+x	-

What is the value of x? That is when 5% of the reaction has proceeded or when 5% of the fructose (or glucose-6-P) is consumed:

$$\frac{x}{0.05} = 0.05$$

so $x = 0.0025$. The equilibrated concentrations are then

- $[\text{suc6P}]_{equil} = \dfrac{0.0025}{V}$
- $[\text{fruc}]_{equil} = \dfrac{0.0475}{V}$
- $[\text{glu6P}]_{equil} = \dfrac{0.0475}{V}$

Substituting into the values in for the expression of K_c (in which the solution volume is the unknown), we have

$$K_c = \frac{[\text{suc6P}]_{equil}}{[\text{fruc}]_{equil}[\text{gluc6P}]_{equil}}$$

$$= \frac{\left(\dfrac{0.0025}{V}\right)}{\left(\dfrac{0.0475}{V}\right)^2} = 7.1 \times 10^{-6}$$

$$V = (7.1 \times 10^{-6})\frac{(0.0475)^2}{0.0025}$$

Solving for V gives a final solution volume of 6.4×10^{-4} L or $640 \mu L$. Why so small? The reaction is not favored and to push it forward, large concentrations of reactants are needed (Le Chatelier principle in action).

This page titled 11.5: Equilibrium Calculations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

11.6: Phase Distribution Equilibria

Learning Objectives

- Phase distribution equilibria play an important role in chemical separation processes on both laboratory and industrial scales. They are also involved in the movement of chemicals between different parts of the environment, and in the bioconcentration of pollutants in the food chain.

It often happens that two immiscible liquid phases are in contact, one of which contains a solute. How will the solute tend to distribute itself between the two phases? One's first thought might be that some of the solute will migrate from one phase into the other until it is distributed equally between the two phases, since this would correspond to the maximum dispersion (randomness) of the solute. This, however, does not take into the account the differing solubilities the solute might have in the two liquids; if such a difference does exist, the solute will preferentially migrate into the phase in which it is more soluble.

For a solute S distributed between two phases a and b the process $S_a = S_b$ is defined by the distribution law

$$K_{a,b} = \frac{[S]_a}{[S]_b} \tag{11.6.1}$$

in which

- $K_{a,b}$ is the *distribution ratio* (also called the distribution coefficient) and
- $[S]_i$ is the solubility of the solute in the phase.

The transport of substances between different phases is of immense importance in such diverse fields as pharmacology and environmental science. For example, if a drug is to pass from the aqueous phase with the stomach into the bloodstream, it must pass through the lipid (oil-like) phase of the epithelial cells that line the digestive tract. Similarly, a pollutant such as a pesticide residue that is more soluble in oil than in water will be preferentially taken up and retained by marine organism, especially fish, whose bodies contain more oil-like substances; this is basically the mechanism whereby such residues as DDT can undergo *biomagnification* as they become more concentrated at higher levels within the food chain. For this reason, environmental regulations now require that oil-water distribution ratios be established for any new chemical likely to find its way into natural waters. The standard "oil" phase that is almost universally used is octanol, $C_8H_{17}OH$.

In preparative chemistry it is frequently necessary to recover a desired product present in a reaction mixture by extracting it into another liquid in which it is more soluble than the unwanted substances. On the laboratory scale this operation is carried out in a separatory funnel as shown below. The two immiscible liquids are poured into the funnel through the opening at the top. The funnel is then shaken to bring the two phases into intimate contact, and then set aside to allow the two liquids to separate into layers, which are then separated by allowing the more dense liquid to exit through the stopcock at the bottom.



If the distribution ratio is too low to achieve efficient separation in a single step, it can be repeated; there are automated devices that can carry out hundreds of successive extractions, each yielding a product of higher purity. In these applications our goal is to exploit the Le Chatelier principle by repeatedly upsetting the phase distribution equilibrium that would result if two phases were to remain in permanent contact.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Video 11.6.1*: How to perform a liquid-liquid extraction using a separating funnel.*

Example 11.6.1

The distribution ratio for iodine between water and carbon disulfide is 650. Calculate the concentration of I_2 remaining in the aqueous phase after 50.0 mL of 0.10M I_2 in water is shaken with 10.0 mL of CS_2.

Solution

The equilibrium constant is

$$K_d = \frac{C_{CS_2}}{C_{H_2O}} = 650$$

mmol, so m_2 = (5.00 – m_1) mmol and we now have only the single unknown m_1. The equilibrium constant then becomes

$$((5.00–m_1)mmol/10mL) \div (m_1 mmol/50mL) = 650$$

Simplifying and solving for m_1 yields

$$\frac{(0.50–0.1)m_1}{(0.02m_1} = 650$$

with m_1 = 0.0382 mmol.

The concentration of solute in the water layer is (0.0382 mmol) / (50 mL) = **0.000763 M**, showing that almost all of the iodine has moved into the CS_2 layer.

This page titled 11.6: Phase Distribution Equilibria is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by LibreTexts.

12: Solubility Equilibria

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Discuss the roles of *lattice- and hydration energy* in determining the solubility of a salt in water.
- Explain what a *qualitative analysis separation scheme* is, and how it works.
- Write the *solubility product* expression for a salt, given its formula.
- Explain the distinction between an *ion product* and a solubility product.
- Given the formula of a salt and its K_s value, calculate the *molar solubility*.
- Explain the Le Chatelier principle leads to the *common ion effect*.
- Explain why a strong acid such as HCl will dissolve a sparingly soluble salt of a weak acid, but not a salt of a strong acid.
- Describe what happens (and why) when aqueous ammonia is slowly added to a solution of silver nitrate

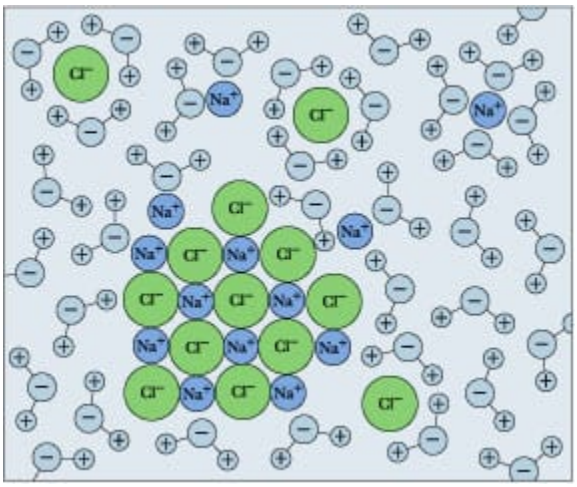
Dissolution of a salt in water is a chemical process that is governed by the same laws of chemical equilibrium that apply to any other reaction. There are, however, a number of special aspects of of these equilibria that set them somewhat apart from the more general ones that are covered in the lesson set devoted specifically to chemical equilibrium. These include such topics as the common ion effect, the influence of pH on solubility, supersaturation, and some special characteristics of particularly important solubility systems.

Solubility: the dissolution of salts in water

Drop some ordinary table salt into a glass of water, and watch it "disappear". We refer to this as *dissolution*, and we explain it as a process in which the sodium and chlorine units break away from the crystal surface, get surrounded by H_2O molecules, and become *hydrated ions*.

$$NaCl_{(s)} \rightarrow Na^+_{(aq)} + Cl^-_{(aq)} \tag{12.1}$$

The designation *(aq)* means "aqueous" and comes from *aqua*, the Latin word for water. It is used whenever we want to emphasize that the ions are hydrated — that H_2O molecules are attached to them.



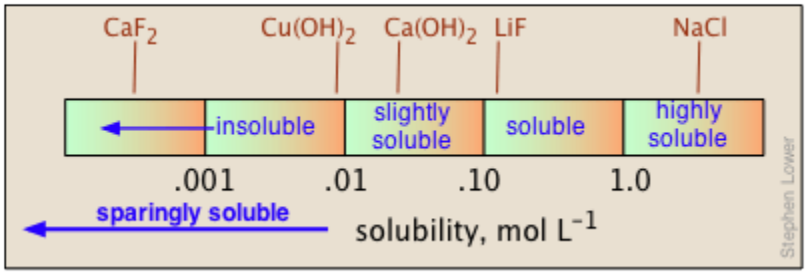
Remember that solubility equilibrium and the calculations that relate to it are only meaningful when *both* sides (solids and dissolved ions) are simultaneously present. But if you keep adding salt, there will come a point at which it no longer seems to dissolve. If this condition persists, we say that the salt has reached its *solubility limit*, and the solution is *saturated* in NaCl. The situation is now described by

$$NaCl_{(s)} \rightleftharpoons Na^{+}_{(aq)} + Cl^{-}_{(aq)} \tag{12.2}$$

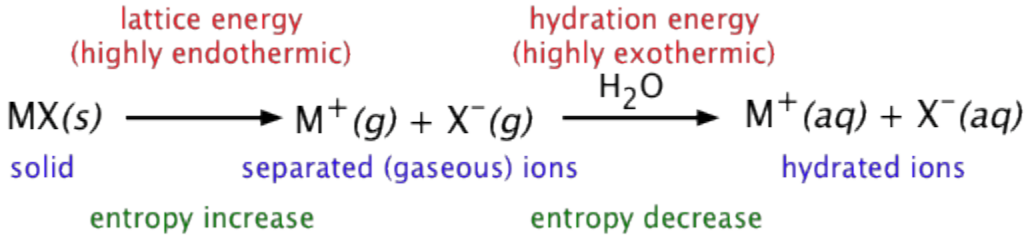
in which the solid and its ions are *in equilibrium*.

Salt solutions that have reached or exceeded their solubility limits (usually 36-39 g per 100 mL of water) are responsible for prominent features of the earth's geochemistry. They typically form when NaCl leaches from soils into waters that flow into salt lakes in arid regions that have no natural outlets; subsequent evaporation of these brines force the above equilibrium to the left, forming natural salt deposits. These are often admixed with other salts, but in some cases are almost pure NaCl. Many parts of the world contain buried deposits of NaCl (known as halite) that formed from the evaporation of ancient seas, and which are now mined.

Solubilities are most fundamentally expressed in molar (mol L^{-1} of solution) or molal (mol kg^{-1} of water) units. But for practical use in preparing stock solutions, chemistry handbooks usually express solubilities in terms of grams-per-100 ml of water at a given temperature, frequently noting the latter in a superscript. Thus 6.9^{20} means 6.9 g of solute will dissolve in 100 mL of water at 20° C. When quantitative data are lacking, the designations "soluble", "insoluble", "slightly soluble", and "highly soluble" are used. There is no agreed-on standard for these classifications, but a useful guideline might be that shown below.



The solubilities of salts in water span a remarkably large range of values, from almost completely insoluble to highly soluble. Moreover, there is no simple way of predicting these values, or even of explaining the trends that are observed for the solubilities of different anions within a given group of the periodic table.



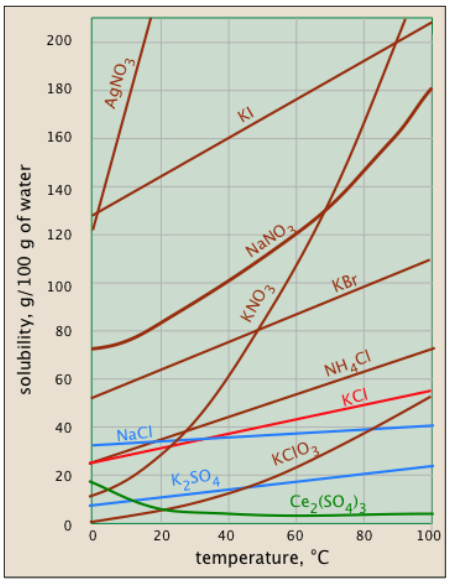
Ultimately, the driving force for dissolution (and for *all* chemical processes) is determined by the Gibbs free energy change. But because many courses cover solubility before introducing free energy, we will not pursue this here. Dissolution of a salt is conceptually understood as a sequence of the two processes depicted above:

1. breakup of the ionic lattice of the solid,
2. followed by attachment of water molecules to the released ions.

The first step consumes a large quantity of energy, something that by itself would strongly discourage solubility. But the second step *releases* a large amount of energy and thus has the opposite effect. Thus the net energy change depends on the sum of two large energy terms (often approaching 1000 kJ/mol) having opposite signs. Each of these terms will to some extent be influenced by the size, charge, and polarizability of the particular ions involved, and on the lattice structure of the solid. This large number of variables makes it impossible to predict the solubility of a given salt. Nevertheless, there are some clear trends for how the solubilities of a series of salts of a given anion (such as hydroxides, sulfates, etc.) change with a periodic table group. And of course, there are a number of general *solubility rules* — for example, that all nitrates are soluble, while most sulfides are insoluble.

Solubility and temperature

Solubility usually increases with temperature - but not always. This is very apparent from the solubility-vs.-temperature plots shown here. (Some of the plots are colored differently in order to make it easier to distinguish them where they crowd together.) The temperature dependence of any process depends on its entropy change — that is, on the degree to which thermal kinetic energy can spread throughout the system. When a solid dissolves, its component molecules or ions diffuse into the much greater volume of the solution, carrying their thermal energy along with them. So we would normally expect the entropy to increase — something that makes any process take place to a greater extent at a higher temperature.



So why does the solubility of cerium sulfate (green plot) diminish with temperature? Dispersal of the Ce^{3+} and SO_4^{2-} ions themselves is still associated with an entropy increase, but in this case the entropy of the *water* decreases even more owing to the ordering of the H_2O molecules that attach to the Ce^{3+} ions as they become hydrated. It's difficult to predict these effects, or explain why they occur in individual cases — but they do happen.

The Importance of Sparingly Soluble Solids

All solids that dissociate into ions exhibit some limit to their solubilities, but those whose saturated solutions exceed about 0.01 mol L^{-1} cannot be treated by simple equilibrium constants owing to ion-pair formation that greatly complicates their behavior. For this reason, most of what follows in this lesson is limited to salts that fall into the "sparingly soluble" category. The importance of sparingly soluble solids arises from the fact that formation of such a product can effectively remove the corresponding ions from the solution, thus driving the reaction to the right. Consider, for example, what happens when we mix solutions of strontium nitrate and potassium chloride in a 1:2 mole ratio. Although we might represent this process by

$$Sr(NO_3)_{2(aq)} + 2KCl_{(aq)} \rightarrow SrCl_{(aq)} + 2KNO_{3(aq)} \tag{12.3}$$

the net ionic equation

$$Sr^{2+} + 2NO_3^- + 2K^+ + 2Cl^- \rightarrow Sr^{2+} + 2NO_3^- + 2K^+ + 2Cl^- \tag{12.4}$$

indicates that no net change at all has taken place! Of course if the solution were than evaporated to dryness, we would end up with a mixture of the four salts shown in Equation 12.3, so in this case we might say that the reaction is half-complete. Contrast this with what happens if we combine equimolar solutions of barium chloride and sodium sulfate:

$$BaCl_{2(aq)} + Na_2SO_{4(aq)} \rightarrow 2NaCl_{(aq)} + BaSO_{4(s)} \tag{12.5}$$

whose net ionic equation is

$$Ba^{2+} + \cancel{2Cl^-} + \cancel{2Na^+} + SO_4^{2-} \rightarrow \cancel{2Na^+} + \cancel{2Cl^-} + BaSO_{4(s)} \tag{12.6}$$

which after canceling out like terms on both sides, becomes simply

$$Ba^{2+} + SO_4^{2-} \rightarrow BaSO_{4(s)} \tag{12.7}$$

Because the formation of sparingly soluble solids is "complete" (that is, equilibria such as the one shown above for barium sulfate lie so far to the right), virtually all of one or both of the contributing ions are essentially removed from the solution. Such reactions are said to be *quantitative*, and they are especially important in analytical chemistry:

- **Qualitative analysis:** This most commonly refers to a procedural scheme, widely encountered in first-year laboratory courses, in which a mixture of cations (usually in the form of their dissolved nitrate salts) is systematically separated and identified on the basis of the solubilities of their various anion salts such as chlorides, carbonates, sulfates, and sulfides. Although this form of qualitative analysis is no longer employed by modern-day chemists (instrumental techniques such as atomic absorption spectroscopy are much faster and comprehensive), it is still valued as an educational tool for familiarizing students with some of the major classes of inorganic salts, and for developing basic skills relating to observing, organizing, and interpreting results in the laboratory.
- **Quantitative gravimetric analysis:** In this classical form of chemical analysis, an insoluble salt of a cation is prepared by precipitating it by addition of a suitable anion. The precipitate is then collected, dried, and weighed ("gravimetry") in order to determine the concentration of the cation in the sample. For example, a gravimetric procedure for determining the quantity of barium in a sample might involve precipitating the metal as the sulfate according to Equation 12.7 above, using an excess of sulfate ion to ensure complete removal of the barium. This method of quantitative analysis became extremely important in the latter half of the nineteenth century, by which time reasonably accurate atomic weights had become available, and sensitive analytical balances had been developed. It was not until the 1960's that it became largely supplanted by instrumental techniques which were much quicker and accurate. Gravimetric analysis is still usually included as a part of more advanced laboratory instruction, largely as a means of developing careful laboratory technique.

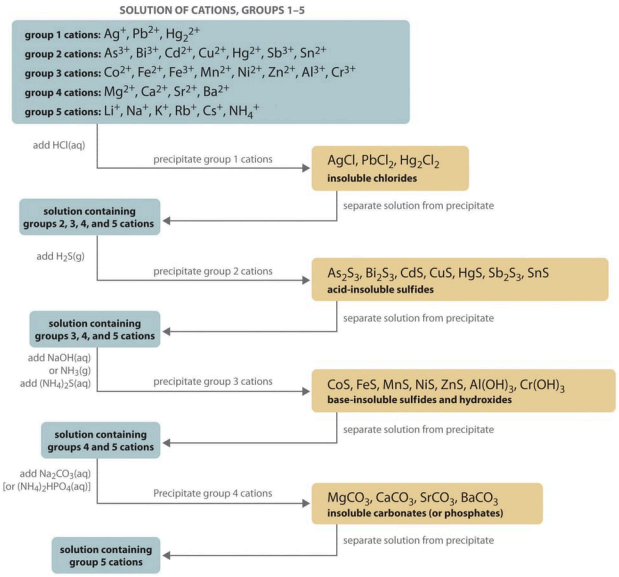


***Figure** 12.1**:** The standard qualitative analysis separation scheme. Steps in a Typical Qualitative Analysis Scheme for a Solution That Contains Several Metal Ions*

Solubility Products and Equilibria

Some salts and similar compounds (such as some metal hydroxides) dissociate completely when they dissolve, but the extent to which they dissolve is so limited that the resulting solutions exhibit only very weak conductivities. In these salts, which otherwise act as strong electrolytes, we can treat the dissolution-dissociation process as a true equilibrium. Although this seems almost trivial now, this discovery, made in 1900 by Walther Nernst who applied the Law of Mass Action to the dissociation scheme of Arrhenius, is considered one of the major steps in the development of our understanding of ionic solutions.

Using silver chromate as an example, we express its dissolution in water as

$$Ag_2CrO_{4(s)} \rightarrow 2Ag^+_{(aq)} + CrO^{2-}_{4(aq)} \tag{12.8}$$

When this process reaches equilibrium (*which requires that some solid be present*), we can write (leaving out the "*(aq)*s" for simplicity)

$$Ag_2CrO_{4(s)} \rightleftharpoons 2Ag^+ + CrO^{2-}_4 \tag{12.9}$$

The equilibrium constant is formally

$$K = \frac{[Ag^+]^2[CrO^{2-}_4]}{[Ag_2CrO_{4(s)}]} = [Ag^+]^2[CrO^{2-}_4] \tag{12.10}$$

But because solid substances do not normally appear in equilibrium expressions, the equilibrium constant for this process is

$$[Ag^+]^2[CrO^{2-}_4] = K_s = 2.76 \times 10^{-12} \tag{12.11}$$

Because equilibrium constants of this kind are written as products, the resulting *K*'s are commonly known as *solubility product*s, denoted by K_s or K_{sp}.

> Strictly speaking, concentration units do not appear in equilibrium constant expressions. However, many instructors prefer that students show them anyway, especially when using solubility products to calculate concentrations. If this is done, K_s in Equation 12.11 would have units of $mol^3\ L^{-3}$.

Equilibrium and non-equilibrium in solubility systems

An expression such as $[Ag^+]^2\ [CrO_4^{2-}]$ in known generally as an *ion product* — this one being the ion product for silver chromate. An ion product can in principle have *any* positive value, depending on the concentrations of the ions involved. Only in the special case when its value is identical with K_S does it become the solubility product. A solution in which this is the case is said to be *saturated*. Thus when

$$[Ag^+]^2\ [CrO^{2-}_4] = 2.76 \times 10^{-12} \tag{12.12}$$

at the temperature and pressure at which this value K_s of applies, we say that the "solution is saturated in silver chromate".

A solution must be saturated to be in equilibrium with the solid. This is a *necessary* condition for solubility equilibrium, but it is not by itself *sufficient*. True chemical equilibrium can only occur when all components are simultaneously present. A solubility system can be in equilibrium only when some of the solid is in contact with a saturated solution of its ions. Failure to appreciate this is a very common cause of errors in solving solubility problems.

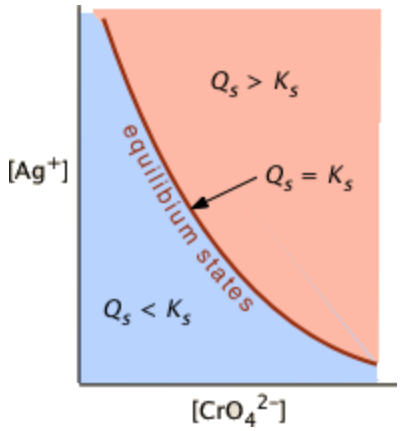
If the ion product is smaller than the solubility product, the system is not in equilibrium and no solid can be present. Such a solution is said to be *undersaturated*. A *supersaturated* solution is one in which the ion product exceeds the solubility product. A supersaturated solution is not at equilibrium, and no solid can ordinarily be present in such a solution. If some of the solid is added, the excess ions precipitate out and until solubility equilibrium is achieved.

How to know the saturation status of a solution

This is just a simple matter of comparing the ion product Q_s with the solubility product K_s. So for the system

$$Ag_2CrO_{4(s)} \rightleftharpoons 2Ag^+ + CrO^{2-}_4 \tag{12.13}$$

a solution in which $Q_s < K_s$ (i.e., $K_s\ /Q_s > 1$) is undersaturated (blue shading) and the no solid will be present. The combinations of $[Ag^+]$ and $[CrO_4^{2-}]$ that correspond to a saturated solution (and thus to equilibrium) are limited to those described by the curved line. The pink area to the right of this curve represents a supersaturated solution.



Example 12.1

A sample of groundwater that has percolated through a layer of gypsum ($CaSO_4$, K_s = 4.9E–5 = $10^{-4.3}$) is found to have be 8.4E–5 *M* in Ca^{2+} and 7.2E–5 *M* in SO_4^{2-}. What is the equilibrium state of this solution with respect to gypsum?

Solution

The ion product

$$Q_s = (8.4 \times 10^{-5})(7.2 \times 10^{-5}) = 6.0 \times 10^{-4} \tag{12.14}$$

exceeds K_s, so the ratio $K_s / Q_s > 1$ and the solution is *supersaturated* in $CaSO_4$.

There are two principal methods, neither of which is all that reliable for sparingly soluble salts:

- Evaporate a saturated solution of the solid to dryness, and weigh what's left.
- Measure the electrical conductivity of the saturated solution, which will be proportional to the concentrations of the ions.

How Solubilities relate to solubility products

The *solubility* (by which we usually mean the *molar solubility*) of a solid is expressed as the concentration of the "dissolved solid" in a *saturated solution*. In the case of a simple 1:1 solid such as AgCl, this would just be the concentration of Ag^+ or Cl^- in the saturated solution. However, for a more complicated stoichiometry such as as silver chromate, the solubility would be only one-half of the Ag^+ concentration.

For example, let us denote the solubility of Ag_2CrO_4 as S mol L^{-1}. Then for a saturated solution, we have

- $[Ag^+] = 2S$
- $[CrO_4^{2-}] = S$

Substituting this into Equation 12.11 above,

$$(2S)^2(S) = 4S^3 = 2.76 \times 10^{-12} \tag{12.15}$$

$$S = \left(\frac{K_s}{4}\right)^{1/3} = (6.9 \times 10^{-13})^{1/3} = 0.88 \times 10^{-4} \tag{12.16}$$

thus the solubility is $8.8 \times 10^{-5}\ M$.

Note that the relation between the solubility and the solubility product constant depends on the stoichiometry of the dissolution reaction. For this reason it is meaningless to compare the solubilities of two salts having the formulas A_2B and AB_2, on the basis of their K_s values.

It is meaningless to compare the solubilities of two salts having different formulas on the basis of their K_s values.

Example 12.2

under these conditions.

Solution

moles of solute in 100 mL; S = 0.0016 g / 78.1 g/mol = 2.05E-5 mol

$$S = 2.05E{-}5 \text{ mol}/0.100 \text{ L} = 2.05E{-}4 \text{ } \underline{M}$$

$$K_S = [Ca^{2+}][F^-]^2 = (S)(2S)^2 = 4 \times (2.05E{-}4)^3 = 3.44E{-}11$$

Example 12.3

Estimate the solubility of $La(IO_3)_3$ and calculate the concentration of iodate in equilibrium with solid lanthanum iodate, for which $K_S = 6.2 \times 10^{-12}$.

Solution

The equation for the dissolution is

$$La(IO_3)_3 \rightleftharpoons La^{3+} + 3IO_3^- \tag{12.17}$$

If the solubility is S, then the equilibrium concentrations of the ions will be

$$[La^{3+}] = S \text{ and } [IO_3^-] = 3S. \text{ Then } K_S = [La^{3+}][IO_3^-]^3 = S(3S)^3 = 27S^4$$

$$27S^4 = 6.2 \times 10^{-12}, \; S = ((6.2 \div 27) \times 10^{-12})^{¼} = 6.92 \times 10^{-4} \text{ } \underline{M}$$

$$[IO_3^-] = 3S = 2.08 \times 10^{-5} \text{ } (\underline{M})$$

Example 12.4 : Cadmium

Cadmium is a highly toxic environmental pollutant that enters wastewaters associated with zinc smelting (Cd and Zn commonly occur together in ZnS ores) and in some electroplating processes. One way of controlling cadmium in effluent streams is to add sodium hydroxide, which precipitates insoluble $Cd(OH)_2$ (K_S = 2.5E-14). If 1000 L of a certain wastewater contains Cd^{2+} at a concentration of 1.6E-5 $\underline{M}$, what concentration of Cd^{2+} would remain after addition of 10 L of 4 $\underline{M}$ NaOH solution?

Solution

As with most real-world problems, this is best approached as a series of smaller problems, making simplifying approximations as appropriate.

Volume of treated water: 1000 L + 10 L = 1010 L

Concentration of OH^- on addition to 1000 L of pure water:

(4 *M*) × (10 L)/(1010 L) = .040 *M*

Initial concentration of Cd^{2+} in 1010 L of water:

(1.6E–5 *M*) x (100/101) ≈ 1.6E–5 *M*

The easiest way to tackle this is to start by assuming that a stoichiometric quantity of $Cd(OH)_2$ is formed — that is, *all* of the Cd^{2+} gets precipitated.

Concentrations	$[Cd^{2+}]$, M	$[OH^-]$, M
initial	1.6E–5	0.04
change	–1.6E–5	–3.2E–5
final:	0	0.04 – 3.2E–5 ≈ .04

Now "turn on the equilibrium" — find the concentration of Cd^{2+} that can exist in a 0.04*M* OH^- solution:

Concentrations	$[Cd^{2+}]$, M	$[OH^-]$, M
initial	o	0.04
change	+*x*	+2*x*
at equilibrium	*x*	0.04 + 2*x* ≈ .04

Substitute these values into the solubility product expression:

$$Cd(OH)_{2(s)} = [Cd^{2+}][OH^-]^2 = 2.5 \times 10^{-14} \tag{12.18}$$

$$[Cd^{2+}] = \frac{2.5 \times 10^{-14}}{16 \times 10^{-4}} = 1.6 \times 10^{-13}\ M \tag{12.19}$$

Note that the effluent will now be very alkaline:

$$pH = 14 + \log 0.04 = 12.6 \tag{12.20}$$

so in order to meet environmental standards an equivalent quantity of strong acid must be added to neutralize the water before it is released.

The simple relations between K_S and molar solubility outlined above, and the calculation examples given here, cannot be relied upon to give correct answers. Some of the reasons for this are explained in Part 2 of this lesson, and have mainly to do with incomplete dissociation of many salts and with complex formation in the presence of anions such as Cl^- and OH^-. The situation is nicely described in the article *What Should We Teach Beginners about Solubility and Solubility Products?* by Stephen Hawkes (*J Chem Educ.*1998 75(9) 1179-81). See also the earlier article by Meites, Pode and Thomas *Are Solubilities and Solubility Products Related?*(*J Chem Educ.* 1966 43(12) 667-72).

It turns out that solubility equilibria more often than not involve many competing processes and their rigorous treatment can be quite complicated. Nevertheless, it is important that students master these over-simplified examples. However, it is also important that they are not taken too seriously!

The Common Ion Effect

It has long been known that the solubility of a sparingly soluble ionic substance is markedly decreased in a solution of another ionic compound when the two substances have an ion in common. This is just what would be expected on the basis of the Le

Chatelier Principle; whenever the process

$$CaF_{2(s)} \rightleftharpoons Ca^{2+} + 2F^{-} \tag{12.21}$$

is in equilibrium, addition of more fluoride ion (in the form of highly soluble NaF) will shift the composition to the left, reducing the concentration of Ca^{2+}, and thus effectively reducing the solubility of the solid. We can express this quantitatively by noting that the *solubility product expression*

$$[Ca^{2+}][F^{-}]^2 = 1.7 \times 10^{-10} \tag{12.22}$$

must always hold, even if some of the ionic species involved come from sources other than CaF_2*(s)*. For example, if some quantity *x* of fluoride ion is added to a solution initially in equilibrium with solid CaF_2, we have

- $[Ca^{2+}] = S$
- $[F^{-}] = 2S + x$

so that

$$K_s = [Ca^{2+}][F^{-}]^2 = S(2S + x)^2. \tag{12.23}$$

University-level students should be able to derive these relations for ion-derived solids of any stoichiometry. In most practical cases *x* will be large compared to S so that the 2S term can be dropped and the relation becomes

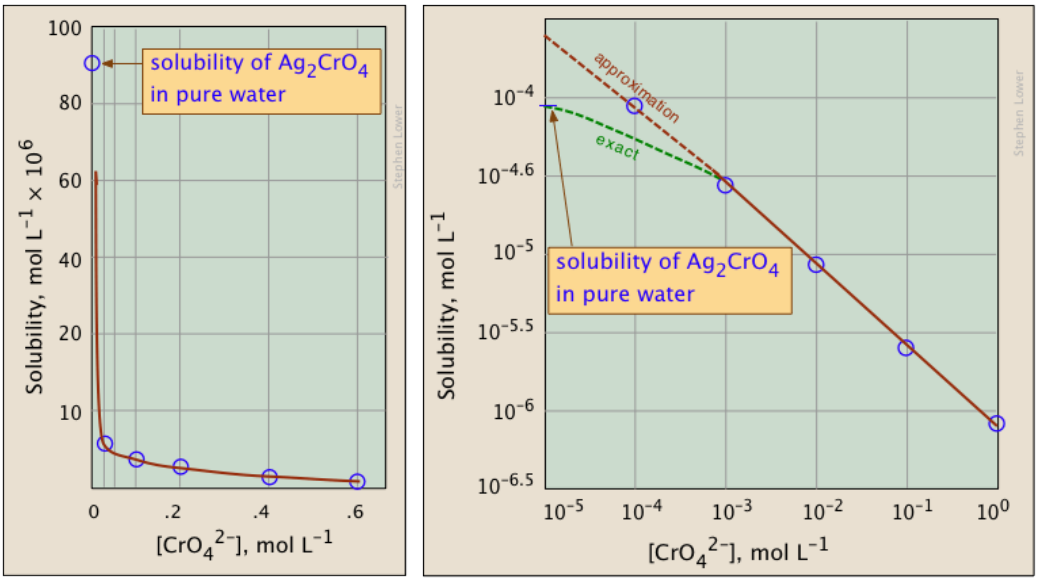
$$K_s \approx S\,x^2 \tag{12.24}$$

or

$$S \approx \frac{K_s}{x^2} \tag{12.25}$$

University-level students should be able to derive these relations for ion-derived solids of any stoichiometry.

The plots shown below illustrate the common ion effect for silver chromate as the chromate ion concentration is increased by addition of a soluble chromate such as Na_2CrO_4.



What's different about the plot on the right? If you look carefully at the scales, you will see that this one is plotted logarithmically (that is, in powers of 10.) Notice how a much wider a range of values can display on a logarithmic plot. The point of showing this pair of plots is to illustrate the great utility of log-concentration plots in equilibrium calculations in which simple approximations (such as that made in Equation 12.25 can yield straight-lines within the range of values for which the approximation is valid.

Example 12.5

Calculate the solubility of strontium sulfate (K_s = 2.8 × 10^{-7}) in

a. pure water and
b. in a 0.10 mol L^{-1} solution of Na_2SO_4.

= [Sr^{2+}][SO$_4^{2-}$] = S^2

$$S = \sqrt{K_s} = \sqrt{2.8 \times 10^{-7}} = 5.3 \times 10^{-4} \tag{12.26}$$

(b) In 0.10 mol L^{-1} Na_2SO_4, we have

= [Sr^{2+}][SO$_4^{2-}$] = S × (0.10 + S) = 2.8 × 10^{-7}

Because S is negligible compared to 0.10 M, we make the approximation

= [Sr^{2+}][SO$_4^{2-}$] ≈ S × (0.10 M) = 2.8 × 10^{-7}

so

This is roughly 100 times smaller than the result from *(a)*.

Selective Precipitation and Separations

Differences in solubility are widely used to selectively remove one species from a solution containing several kinds of ions.

Example 12.6

The solubility products of AgCl and Ag_2CrO_4 are 1.8E–10 and 2.0E–12, respectively. Suppose that a dilute solution of $AgNO_3$ is added dropwise to a solution containing 0.001*M* Cl$^-$ and 0.01*M* CrO$_4^{2-}$.

a. *Which solid,* AgCl *or* Ag_2CrO_4, will precipitate first?
b. What fraction of the first anion will have been removed when the second just begins to precipitate? Neglect any volume changes.

Solution

The silver ion concentrations required to precipitate the two salts are found by substituting into the appropriate solubility product expressions:

- *to precipitate AgCl:* [Ag$^+$] = 1.8E-10 / .001 = 1.8E-7 *M*
- *to precipitate Ag_2CrO_4:* [Ag$^+$] = (2.0E-12 / .01)^{½} = 1.4E–5 *M*

The first solid to form as the concentration of Ag$^+$ increases will be AgCl. Eventually the Ag$^+$ concentration reaches 1.4E-5 *M* and Ag_2CrO_4 begins to precipitate. At this point the concentration of chloride ion in the solution will be 1.3E-5 *M* which is about 13% of the amount originally present.

The preceding example is the basis of the Mohr titration of chloride by Ag$^+$, commonly done to determine the salinity of water samples. The equivalence point of this precipitation titration occurs when no more AgCl is formed, but there is no way of observing this directly in the presence of the white AgCl which is suspended in the container. Before the titration is begun, a small amount of K_2CrO_4 is added to the solution. Ag_2CrO_4 is red-orange in color, so its formation, which signals the approximate

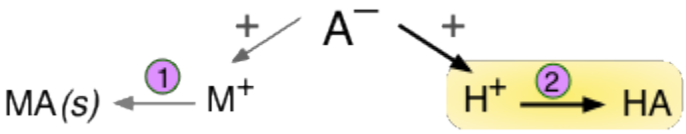
end of AgCl precipitation, can be detected visually.

Competing Equilibria involving solids

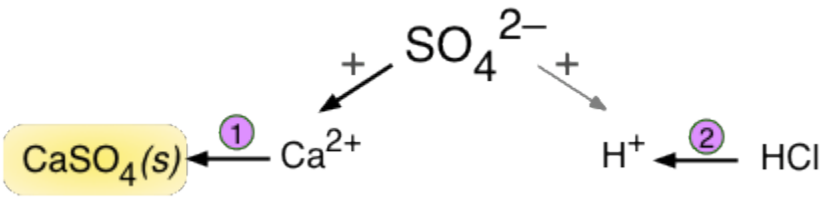
Solubility expression are probably the exception rather than the rule. Such equilibria are often in competition with other reactions with such species as H^+or OH^-, complexing agents, oxidation-reduction, formation of other sparingly soluble species or, in the case of carbonates and sulfites, of gaseous products. The exact treatments of these systems can be extremely complicated, involving the solution of large sets of simultaneous equations. For most practical purposes it is sufficient to recognize the general trends, and to carry out approximate calculations.

Salts of weak acids are soluble in strong acids, but strong acids will not dissolve salts of strong acids

The solubility of a sparingly soluble salt of a weak acid or base will depend on the pH of the solution. To understand the reason for this, consider a hypothetical salt MA which dissolves to form a cation M^+ and an anion A^-which is also the conjugate base of a weak acid HA. The fact that the acid is weak means that hydrogen ions (always present in aqueous solutions) and M^+ cations will both be competing for the A^-:



The weaker the acid HA, the more readily will reaction (2) take place, thus gobbling up A^- ions. If an excess of H^+ is made available by addition of a strong acid, even more A^- ions will be consumed, eventually reversing reaction (1), causing the solid to dissolve.



In (1), for example, sulfate ions react with calcium ions to form insoluble $CaSO_4$. Addition of a strong acid such as HCl (which is totally dissociated (2)) has no effect because $CaCl_2$ is soluble. Although H^+ can protonate some SO_4^{2-} ions to form hydrogen sulfate ("bisulfate") HSO_4^-, this ampholyte acid is too weak to reverse (1) by drawing a significant fraction of sulfate ions out of $CaSO_4(s)$.

Example 12.7: Aluminum Ion

Calculate the concentration of aluminum ion in a solution that is in equilibrium with aluminum hydroxide when the pH is held at 6.0.

The equilibria are

$$Al(OH)_3 \rightleftharpoons Al^{3+} + 3OH^- \tag{12.27}$$

with

$$K_s = 1.4 \times 10^{-34} \tag{12.28}$$

and

$$H_2O \rightleftharpoons H^+ + OH^- \tag{12.29}$$

with

$$K_w = 1 \times 10^{-14} \tag{12.30}$$

Substituting the equilibrium expression for the second of these into that for the first, we obtain

$$[OH^-]^3 = \left(\frac{K_w}{[H^+]}\right)^3 = \frac{K_s}{[Al^{3+}]} \tag{12.31}$$

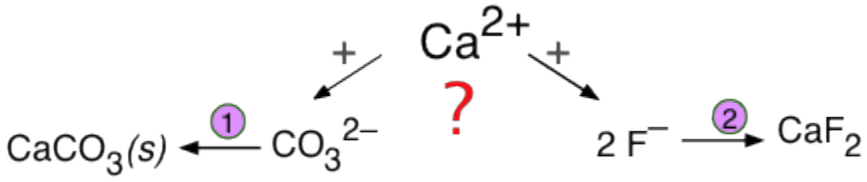
$$(1.0 \times 10^{-14}) / (1.0 \times 10^{-6})^3 = (1.4 \times 10^{-24}) / [Al^{3+}]$$

from which we find

$$[Al^{3+}] = 1.4 \times 10^{-10}\ M \tag{12.32}$$

Formation of a Competing Precipitate

If two different anions compete with a single cation to form two possible precipitates, the outcome depends not only on the solubilities of the two solids, but also on the concentrations of the relevant ions.



These kinds of competitions are especially important in groundwaters, which acquire solutes from various sources as they pass through sediment layers having different compositions. As the following example shows, competing equilibria of these kinds are very important for understanding geochemical processes involving the formation and transformation of mineral deposits.

Example 12.8

Suppose that groundwater containing 0.001*M* F^- and 0.0018*M* CO_3^{2-} percolates through a sediment containing calcite, $CaCO_3$. Will the calcite be replaced by fluorite, CaF_2?

The two solubility equilibria are

\[\ce{CaCO3 <=> Ca^{2+} + CO3^{2-} \quad K_s = 10^{-8.1}\]

\[\ce{CaF2 <=> Ca^{2+} + 2 F^{-} \quad K_s = 10^{-10.4}\]

Solution:

The equilibrium between the two solids and the two anions is

$$CaCO_3 + 2F^- \rightleftharpoons CaF_2 + CO_3^{2-} \tag{12.33}$$

This is just the sum of the dissolution reaction for $CaCO_3$ and the reverse of that for CaF_2, so the equilibrium constant is

$$K = \frac{[CO_3^{2-}]}{[F^-]^2} = \frac{10^{-8.1}}{10^{-10.4}} = 200 \tag{12.34}$$

That is, the two solids can coexist only if the reaction quotient Q ≤ 200. Substituting the given ion concentrations we find that

$$Q = \frac{0.0018}{0.0012} = 1800 \tag{12.35}$$

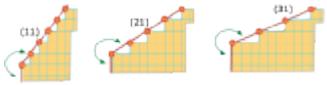
Since Q>*K*, we can conclude that the calcite will not change into fluorite.

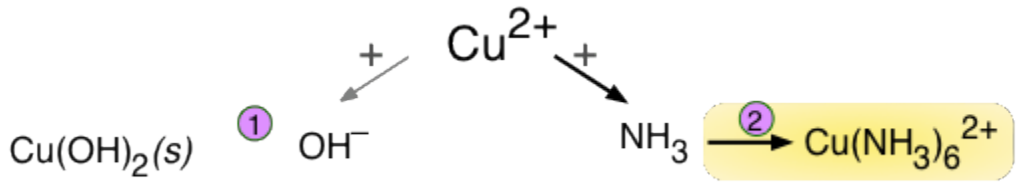
Complex Ion Formation

Most transition metal ions possess empty *d* orbitals that are sufficiently low in energy to be able to accept electron pairs from electron donors from cations, resulting in the formation of a covalently-bound complex ion. Even neutral species that have a nonbonding electron pair can bind to ions in this way. Water is an active electron donor of this kind, so aqueous solutions of ions such as $Fe^{3+}(aq)$ and $Cu^{2+}(aq)$ exist as the octahedral complexes $Fe(H_2O)_6^{3+}$ and $Cu(H_6O)_6^{2+}$, respectively.

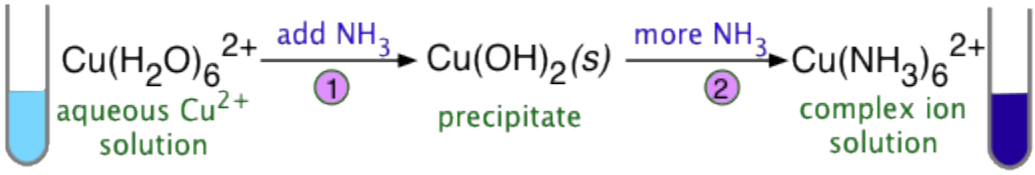
Many of the remarks made above about the relation between K_s and solubility also apply to calculations involving complex formation. See Stephen Hawkes' article *Complexation Calculations are Worse Than Useless* ("... to the point of absurdity...and should not be taught" in introductory courses.) (*J Chem Educ.* 1999 76(8) 1099-1100). However, it is very important that you understand the principles outlined in this section.

H_2O is only one possible electron donor; NH_3, CN^- and many other species (known collectively as *ligands*) possess lone pairs that can occupy vacant*d* orbitals on a metallic ion. Many of these bind much more tightly to the metal than does H_2O, which will undergo displacement and substitution by one or more of these ligands if they are present in sufficiently high concentration.

If a sparingly soluble solid is placed in contact with a solution containing a ligand that can bind to the metal ion much more strongly than H_2O, then formation of a complex ion will be favored and the solubility of the solid will be greater. Perhaps the most commonly seen example of this occurs when ammonia is added  to a solution of copper(II) nitrate, in which the $Cu^{2+}(aq)$ ion is itself the complex hexaaquo complex ion shown at the left:



Because ammonia is a weak base, the first thing we observe is formation of a cloudy precipitate of $Cu(OH)_2$ in the blue solution. As more ammonia is added (2), this precipitate dissolves, and the solution turns an intense deep blue, which is the color of hexamminecopper(II) and the various other related species such as $Cu(H_2O)_5(NH_3)^{2+}$, $Cu(H_2O)_4(NH_3)_2^{2+}$, etc.



In many cases, the complexing agent and the anion of the sparingly soluble salt are identical. This is particularly apt to happen with insoluble chlorides, and it means that addition of chloride to precipitate a metallic ion such as Ag^+ will produce a precipitate at first, but after excess Cl^- has been added the precipitate will redissolve as complex ions are formed.

Some important solubility systems

In this section, we discuss solubility equilibria that relate to some very commonly-encountered anions of metallic salts. These are especially pertinent to the kinds of separations that most college students are required to carry out (and understand!) in their first-year laboratory courses.

Solubility of oxides and hydroxides

Metallic oxides and hydroxides both form solutions containing OH^- ions. For example, the solubilities of the [sparingly soluble] oxide and hydroxide of magnesium are represented by

$$Mg(OH)_{2(s)} \rightarrow Mg^{2+} + 2OH^- \tag{12.36}$$

$$MgO_{(S)} + H_2O \rightarrow Mg^{2+} + 2OH^- \tag{12.37}$$

If you write out the solubility product expressions for these two reactions, you will see that they are identical in form and value.

Recall that pH = $-\log_{10}[H^+]$, so that $[H^+] = 10^{-pH}$.

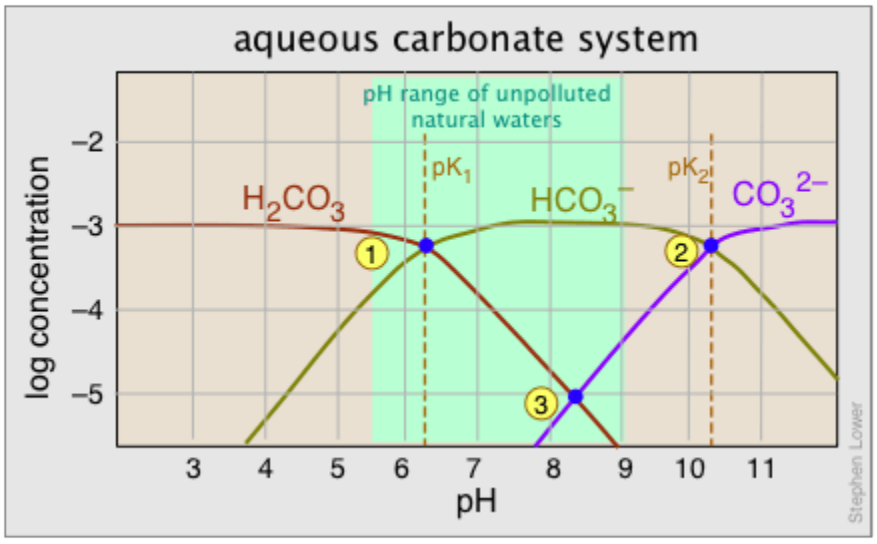
One might naïvely expect that the dissolution of an oxide such as MgO would yield as one of its products the oxide ion O^{2+}. But the oxide ion is such a strong base that it grabs a proton from water, forming two hydroxide ions instead:

$$O^{2+} + H_2O \rightarrow 2OH^- \tag{12.38}$$

This is an example of the rule that the hydroxide ion is the strongest base that can exist in aqueous solution.2" is an equilibrium mixture of hydrated CO_2molecules and carbonic acid. To keep things as simple as possible, we will not distinguish between them in what follows, and just use the formula H_2CO_3 to represent the two species collectively.

The other Group 2 metals, especially Mg, along with iron and several other transition elements are also found in carbonate sediments. When rain falls through the air, it absorbs atmospheric carbon dioxide, a small portion of which reacts with the water to form carbonic acid. Thus all pure water in contact with the air becomes acidic, eventually reaching a pH of 5.6.

As noted above, the equilibrium between bicarbonate and carbonate ions depends on the pH. Since the pH scale is logarithmic, it makes sense (and greatly simplifies the construction of the plot) to employ a log scale for the concentrations. The plot shown below corresponds to a total carbonate-system concentration of 10^{-3} *M*, which is representative of many ground waters. For river and lake waters, 10^{-5} *M* would be more typical; this would simply shift the curves downward without affecting their shapes.



Points 1 and 2 where adjacent curves overlap correspond to the two pK's. Recall that when the pH is the same as the pK, the

concentrations of the two conjugate species are identical and half of the total system concentration. This places the crossover points at log 0.5 = –0.3 below the system concentration level.

A 10^{-3} *M* solution of sodium bicarbonate would have a pH denoted by point 3, with [H_2CO_3] and [CO_3^{2-}] constituting only 1% (10^{-5} *M*) of the system. This corresponds to the equilibrium

$$2HCO_3^- \rightleftharpoons H_2CO_3 + CO_3^{2-} \tag{12.39}$$

Carbonates act as bases and, as such, react with acids. Thus, the portion of the global water cycle that transports carbon from the air into natural waters constitutes a gigantic acid-base reaction that yields hydrogen carbonate ions, commonly referred to as bicarbonate. The natural waters that result have pH values between 6 and 10 and are essentially solutions of bicarbonates.

When rainwater permeates into the soil, it can become even more acidic owing to the additional CO_2 produced by soil organisms. Also, the deeper the water penetrates, the greater its hydrostatic pressure and the more CO_2 it can hold, further increasing its acidity. If this water then works its way down through the fissures and cracks within a limestone layer, it will dissolve some of limestone, leaving void spaces which may eventually grow into limestone caves or form sinkholes that can swallow up cars or houses.

A well-known feature of limestone caves is the precipitated carbonate formations that decorate the ceilings and floors. These are known as *stalactite*s and *stalagmite*s, respectively. When water emerges from the ceiling of a cave that is open to the atmosphere, some of the excess CO_2 it contains is released as it equilibrates with the air. This raises its pH and thus reduces the solubility of of the carbonates, which precipitate as stalactites. Some of the water remains supersaturated and does not precipitate until it drips to the cave floor, where it builds up the stalagmite formations.

Hard Water

This term refers to waters that, through contact with rocks and sediments in lakes, streams, and especially in soils (groundwaters), have acquired metallic cations such as Ca^{2+}, Mg^{2+}, Fe^{2+}, Fe^{3+}, Zn^{2+} Mn^{2+}, etc. Owing to the ubiquity of carbonate sediments, the compensating negative charge is frequently supplied by the bicarbonate ion HCO_3^-, but other anions such as SO_4^{2-}, F^-, Cl^-, PO_4^{3-} and SiO_4^{2-} may also be significant.

Solid bicarbonates are formed only by Group 1 cations and all are readily soluble in water. But because HCO_3^- is amphiprotic, it can react with itself to yield carbonate:

Extra close brace or missing open brace

If bicarbonate-containing water is boiled, the CO_2 is driven off, and the equilibrium shifts to the right, causing any Ca^{2+} or similar ions to form a cloudy precipitate. If this succeeds in removing the "hardness cations", the water has been "softened". Such water is said to possess *carbonate hardness*, sometimes known as "*temporary hardness*". Waters in which anions other than HCO_3^- predominate cannot be softened by boiling, and thus possess *non-carbonate hardness* or "*permanent hardness*".

Hard waters present several kinds of problems, both in domestic and industrial settings:

- Waters containing dissolved salts leave solid deposits when they evaporate. Residents of areas having hard water (about 85 percent of the U.S.) notice evaporative deposits on shower walls, in teakettles, and on newly-washed windows, glassware, and vehicles.
- Much more seriously from an economic standpoint, evaporation of water in boilers used for the production of industrial steam leaves coatings on the heat exchanger surfaces that impede the transfer of heat from the combustion chamber, reducing the thermal transfer efficiency. The resultant overheating of these surfaces can lead to their rupture, and in the case of high-pressure boilers, to disastrous explosions. In the case of calcium and magnesium carbonates, the process is exacerbated by the reduced solubility of these salts at high temperatures. Removal of boiler scales is difficult and expensive.
- Municipal water supplies in hard-water areas tend to be supersaturated in hardness ions. As this water flows through distribution pipes and the plumbing of buildings, these ions often tend to precipitate out on their interior surfaces. Eventually,

this scale layer can become thick enough to restrict or even block the flow of water through the pipes. When scale deposits within appliances such as dishwashers and washing machines, it can severely degrade their performance.
- Cations of Group 2 and above react with soaps, which are sodium salts of fatty acids such as stearic acid, $C_{17}H_{35}COOH$. The sodium salts of such acids are soluble in water, which allows them to dissociate and act as surfactants:

$$C_{17}H_{35}COONa \rightarrow C_{17}H_{35}COO^{-}Na^{+} \tag{12.40}$$

but the presence of polyvalent ions causes them to form precipitates

$$2C_{17}H_{35}COO^{-} + Ca^{2+} \rightarrow (C_{17}H_{35}COO^{-})_2Ca_{(s)} \tag{12.41}$$

Calcium stearate is less dense than water, so it forms a scum that floats on top of the water surface; anyone who lives in a hard-water area is likely familiar with the unsightly "bathtub rings" it leaves around the high-water mark or the shower-wall stains.

Solubility Complications

All heterogeneous equilibria, on close examination, are beset with complications. But solubility equilibria are somewhat special in that there are more of them. Back in the days when the principal reason for teaching about solubility equilibria was to prepare chemists to separate ions in quantitative analysis procedures, these problems could be mostly ignored. But now that the chemistry of the environment has grown in importance — especially that relating to the ocean and natural waters — there is more reason for chemical scientists to at least know about the limitations of simple solubility products. This section will offer a quick survey of the most important of these complications, while leaving their detailed treatment to more advanced courses.

Many of the K_s values found in tables were determined prior to 1940 (some go back to the 1880s!) at a time before highly accurate methods became available. Especially suspect are many of those for highly insoluble salts which are more difficult to measure. A table showing the variations in K_{sp} values for the same salts among ten textbooks was published by Clark and Bonikamp in J Chem Educ. 1998 75(9) 1183-85.A good An example that used a variety of modern techniques to measure the solubility of silver chromate was published by A.L. Jones et al in the Australian J. of Chemistry, 1971 24 2005-12.

Generations of chemistry students have amused themselves by comparing the disparate K_S values to be found in various textbooks and table. In some cases, they differ by orders of magnitude. There are several reasons for this in addition to the ones described in detail further on.

- The most direct methods of measuring solubilities tend to not be very accurate for sparingly soluble salts. Two-significant figure precision is about the best one can hope in a single measurement.
- Many insoluble salts can exist in more than one crystalline form (*polymorphs*), and in some cases also as amorphous solids. Precipitation under different conditions (in the presence of different ions, at different temperatures, etc.) can yield different or mixed polymorphs.
- Other ions present in the solution can often get incorporated into the crystalline solid, usually replacing an ion of similar size (*substitutional solid solutions*). When this happens, it is no longer valid to write the equilibrium condition as a simple "product". This is very common in mineral deposits, and an important consideration in geochemistry,

Most salts are not Completely Dissociated in Water

The dissolution of cadmium iodide is water is commonly represented as

$$CdI_{2(s)} \rightarrow Cd^{2+} + 2I^{-} \tag{12.42}$$

What you were likely taught about the dissociation of salts in water is wrong! To most students (and to most of their teachers!), this would imply that a 0.1*M* solution of this salt would contain 0.1*M* of $Cd^{2+}(aq)$ — and this would be seriously in error because it fails to take into account that the two ions react with each other.

Firstly, they combine to form neutral, largely-covalent molecular species:

$$Cd^{2+}_{(aq)} + 2I^{-}_{(aq)} \rightarrow CdI_{2(aq)} \tag{12.43}$$

This non-ionic form accounts for 78% of the Cd present in the solution! In addition, they form a molecular ion $CdI^-_{(aq)}$ according to the following scheme:

Table 12.2: CdI2 species in solution

$CdI_{2(s)} \rightleftharpoons Cd^{2+} + 2I^-$	$K_1 = 10^{-3.9}$
$Cd^{2+} + I^- \rightleftharpoons CdI^+$	$K_2 = 10^{+2.3}$
$CdI2_{(s)} \rightleftharpoons CdI^+ + I^-$	$K = 10^{-1.6} = 0.023$

The data shown Tables 12.1 and 12.2 are taken from the article *Salts are Mostly NOT Ionized* by Stephen Hawkes: 1996 *J Chem Educ.* 73(5) 421-423. This fact was stated by Arrhenius in 1887, but has been largely ignored and is rarely mentioned in standard textbooks.

As a consequence, the concentration of "free" Cd^{2+}(aq) in an aqueous cadmium iodide solution is only about 2% of the value you would calculate by taking K_1 as the solubility product. The principal component of such as solution is actually [covalently-bound] CdI_2(aq). It turns out that many salts, especially those of metals beyond Group 2, are similarly only partially ionized in aqueous solution:

Table 12.2: Incomplete Dissociation of Select Salts

salt	molarity	% cation	other species
KCl	0.52	95	KCl(aq) 5%
$MgSO_4$	0.04	58	$MgSO_4$(aq) 42%
$CaCl_2$	0.44	70	$CaCl^+$(aq) 30%
$CuSO_4$	0.045	56	$CuSO_4$(aq) 44%
CdI_2	0.50	2	CdI_2(aq) 76%, CdI^-(aq) 22%
$FeCl_3$	0.1	10	$FeCl^{2+}$(aq) 42%, $FeCl_2$(aq) 40%, $FeOH^{2+}$(aq) 6%, $Fe(OH)_2^+$(aq) 2%

If you are enrolled in an introductory course and do not plan on taking more advanced courses in chemistry or biochemistry, you can probably be safe in ignoring this, since your instructor and textbook likely do so. However, if you expect to do more advanced work or teach, you really should take note of these points, since few textbooks mention them.

Formation of Hydrous Complexes

Transition metal ions form a large variety of complexes with H_2O and OH^-, both of which have electron-pairs available to coordinate with the central ion. This gives rise to a large variety of soluble species that are in competition with an insoluble solid. Because of this, a single equilibrium constant (solubility product) cannot describe the behavior of a solid such as $Fe(OH)_3$, which we summarize here as an example.

Aquo complexes: The electrostatic field of the positively-charged metal ion enhances the acidic nature of these H_2O molecules, encouraging them to shed a proton and leaving OH^- groups in their place.

$$Fe(H_2O)_6^{3+} \rightarrow Fe(H_2O)_5(OH)^{2+} + H^+ \tag{12.44}$$

This is just the first of a series of similar reactions, each one having a successively smaller equilibrium constant:

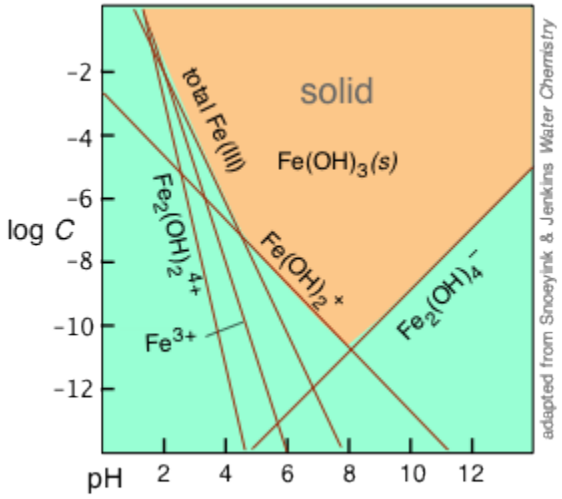
$$Fe(H_2O)_5(OH)^{2+} \rightarrow Fe(H_2O)_4(OH)_2^+ \rightarrow Fe(H_2O)_3(OH)_3 \rightarrow Fe(H_2O)_2(OH)_4^- \tag{12.45}$$

Hydroxo complexes: But there's more: when the hydroxide ion acts as a ligand, it gives rise to a series of hydroxo complexes, of which the insoluble $Fe(OH)_3$ can be considered a member:

Fe^{3+} + 3 OH^- → $Fe(OH)_3$(s)	$1/K_s = 10^{38}$

$Fe^{3+} + H_2O \rightarrow Fe(OH)_2^+ + H^+$	$K = 10^{-2.2}$
$Fe^{3+} + 2H_2O \rightarrow FeOH^+ + 2H^+$	$K = 10^{-6.7}$
$Fe^{3+} + 4H_2O \rightarrow Fe(OH)_4^- + 4H^+$	$K = 10^{-23}$
$2Fe^{3+} + 2H_2O \rightarrow Fe_2(OH)_2^{4+} + 2H^+$	$K = 10^{-2.8}$

Because five equilibria are involved in the hydroxy-complex behavior alone, a complete treatment of this system for Fe^{3+} requires the solution of five simultaneous equations, which is not a lot of fun. Fortunately, it is possible to make some approximations which greatly simplify the construction of a log-concentration vs. pH plot as shown below.



The equilibria listed above all involve H^+ and OH^- ions, and are therefore pH dependent, as illustrated by the straight lines in the plot, whose slopes reflect the pH dependence of the corresponding ionic species. At any given pH, the equilibrium with solid $Fe(OH)_3$ is controlled by the ionic species having the highest concentration at any given pH. The corresponding lines in the plot therefore delineate the region (indicated by the orange shading) at which the solid can exist.

Ionic interactions: The "non-common ion effect"

A sparingly-soluble salt will be more soluble in a solution that contains non-participating ions. This is just the opposite of the common ion effect, and it might at first seem rather counter-intuitive: why would adding more ions of any kind make a salt more likely to dissolve?

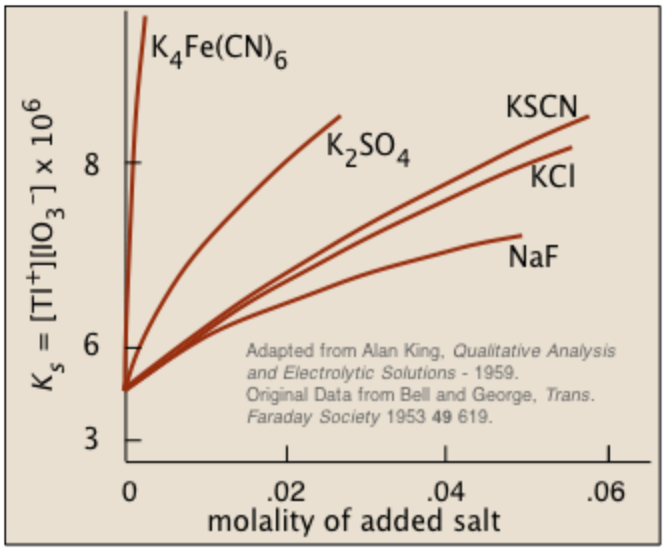
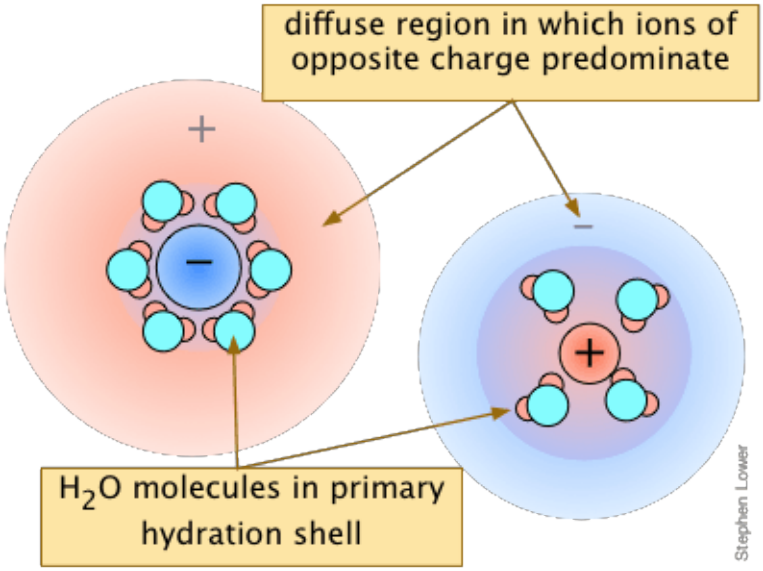


Figure 12.2***:*** *Solubility of thallium iodate in solutions containing dissolved salts*

A clue to the answer can be found in another fact: the higher the charge of the foreign ion, the more pronounced is the effect. This tells us that inter-ionic (and thus electrostatic) interactions must play a role. The details are rather complicated, but the general idea is that all ions in solution, besides possessing tightly-held waters of hydration, tend to attract oppositely-charged ions ("*counter-ions*") around them. This "atmosphere" of counterions is always rather diffuse, but much less so (and more tightly bound) when one or both kinds of ions have greater charges. From a distance, these ion-counterion bodies appear to be almost electrically neutral, which keeps them from interacting with each other (as to form a precipitate).



The overall effect is to reduce the concentrations of the less-shielded ions that are available to combine to form a precipitate. We say that the *thermodynamically-effective concentrations* of these ions are less than their "analytical" concentrations. Chemists refer to these effective concentrations as ionic *activities*, and they denote them by curly brackets $\{Ag^+\}$ as opposed to square brackets $[Ag^+]$ which refer to the nominal or analytical concentrations.

Although the concentrations of ions in equilibrium with a sparingly soluble solid are so low that they are essentially the same as the activities, the presence of other ions at concentrations of about 0.001M or greater can materially reduce the activities of the dissolution products, permitting the solubilities to be greater than what simple equilibrium calculations would predict.

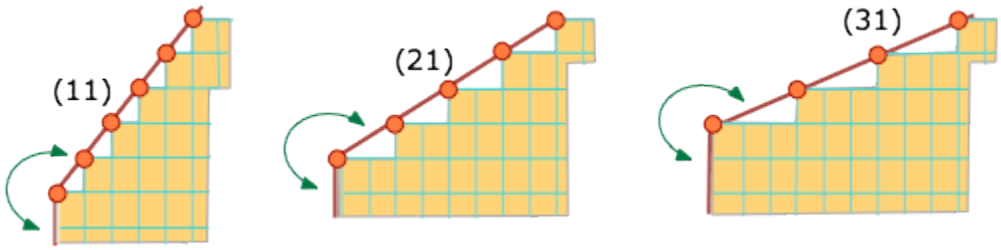
Measured solubilities are averages and depend on size

The heterogeneous nature of dissolution reactions leads to a number of peculiar effects relating to the nature of equilibria involving surfaces. These arise from the fact that the tendency of a crystalline solid to dissolve will depend on the particular face or location from which dissolution occurs. Since all crystals present a variety of faces to the solution, a measured K_S is really an average of values for these various faces.

And because many salts can exhibit different external shapes depending on the conditions under which they are formed, solubility products are similarly dependent on these conditions.

Very small crystals are more soluble than big ones

Molecules or ions that are situated on edges or corners are less strongly bound to the remainder of the solid than those on plane surfaces, and will consequently tend to dissolve more readily. Thus the leftmost face in the schematic lattice below will have more edge-bound molecular units than the other two, and this face (11) will be more soluble.



This means, among other things, that smaller crystals, in which the ratio of edges and corners is greater, will tend to have greater K_S values than larger ones. As a consequence, smaller crystals will tend to disappear in favor of larger ones. Practical use is sometimes made of this when the precipitate initially formed in a chemical analysis or separation is too fine to be removed by filtration. The suspension is held at a high temperature for several hours, during with time the crystallites grow in size. This procedure is sometimes referred to as digestion.

Formation of supersaturated solutions

Contrary to what you may have been taught, precipitates do not form when the ion concentration product reaches the solubility product of a salt in a solution that is pure and initially unsaturated; to form a precipitate from a homogeneous solution, a certain degree of supersaturation is required. The extent of supersaturation required to initiate precipitation can be surprisingly great. Thus formation of barium sulfate $BaSO_4$ by combining the two kinds of ions does not occur until Q_S exceeds K_S by a factor of 160 or more. In part, this reflects the fact that precipitation proceeds by a series of reactions beginning with formation of an ion-pair which eventually becomes an ion cluster:

$$Ba^{2+} + SO_4^{2-} \rightarrow (BaSO_4)^0 \rightarrow (BaSO_4)_2^0 \rightarrow (BaSO_4)_3^0 \rightarrow \text{etc.}$$

Owing to their overall neutrality, these aggregates are not stabilized by hydration, so they are more likely to break up than not. But a few may eventually survive until they are large enough (but still submicroscopic in size) to serve as precipitation nuclei.

Many substances other than salts form supersaturated solutions, and some salts form them more readily than others. Supersaturated solutions are easily made by dissolving the solid to near its solubility limit in a warmed solvent, and then letting it cool.

K_S) and are inherently unstable; dropping a "seed" crystal of the solid into such a solution will usually initiate rapid precipitation. But as is explained below, even a tiny dust particle may be enough. An old chemist's trick is to use the tip of a glass stirring rod to scrape the inner surface of a container holding a supersaturated solution; the minute particles of glass that are released presumably serve as precipitation nuclei.

The nucleation problem: precipitation is [theoretically] impossible!

Any process in which a new phase forms within an existing homogeneous phase is beset by the **nucleation problem**: the smallest of these new phases — raindrops forming in air, tiny bubbles forming in a liquid at its boiling point — are inherently less stable than larger ones, and therefore tend to disappear. The same is true of precipitate formation: if smaller crystals are more soluble, then how can the tiniest, first crystal, form at all?

In any ionic solution, small clumps of oppositely-charged ions are continually forming by ordinary collisional processes. The smallest of these aggregates possess a higher free energy than the isolated solvated ions, and they rapidly dissociate. Occasionally, however, one of these proto-crystallites reaches a critical size whose stability allows it to remain intact long enough to serve as a surface (a "nucleus") onto which the deposition of additional ions can lead to still greater stability. At this point, the process passes from the nucleation to the growth stage.

Theoretical calculations predict that nucleation from a perfectly homogeneous solution is a rather unlikely process; tenfold supersaturation should produce only one nucleus per cm^3 per year. Most nucleation is therefore believed to occur heterogeneously on the surface of some other particle, possibly a dust particle. The efficiency of this process is critically dependent on the nature and condition of the surface that gives rise to the nucleus.

This page titled 12: Solubility Equilibria is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13: ACID-BASE EQUILIBRIA

Acid-base chemistry can be extremely confusing, particularly when dealing with weak acids and bases. This set of lessons presents an updated view of the Brønsted-Lowry theory that makes it easy to understand answers to common questions: What's the fundamental difference between a strong acid and a weak acid? Can acid A neutralize base B? Why are some salts acidic and others alkaline? How do buffers work? What governs the shapes of titration curves?

13.1: Introduction to Acid/Base Equilibria

13.2: Strong Monoprotic Acids and Bases

13.3: Finding the pH of weak Acids, Bases, and Salts

13.4: Conjugate Pairs and Buffers

13.5: Acid/Base Titration

13.6: Applications of Acid-Base Equilibria

13.7: Exact Calculations and Approximations

This page titled 13: Acid-Base Equilibria is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13.1: Introduction to Acid/Base Equilibria

Learning Objectives

Make sure you thoroughly understand the following essential concepts:

- Write a chemical equation for the *autoprotolysis* of a species such as HCO_3^-; write an equation for the equilibrium constant for this reaction.
- Write equations that express the chemical reactions and equilibrium constant expressions that define the *strength of an acid or base* in terms of proton transfer processes (Brønsted-Lowry concept).
- Write expressions defining the *equilibrium constants* K_a *and* K_b.
- Given the value of any two of K_a, K_b, and K_w, evaluate the third one. (This requires that you can show that $K_a \times K_b = K_w$.)
- Explain the *significance of the value of* pK_a for a given monoprotic acid-base system — that is, what it tells us about that particular acid and its conjugate base.
- Sketch out a simple proton free energy (PFE) diagrams illustrating the distinction between monoprotic *strong and weak acids*.
- State the meaning of the *leveling effect*, and construct a simple PFE diagram that illustrates its action.
- Describe, in your own words, the *meaning of PFE*, how it relates to pH, and in its significance in terms of the relative concentrations of a conjugate acid-base pair in a solution.

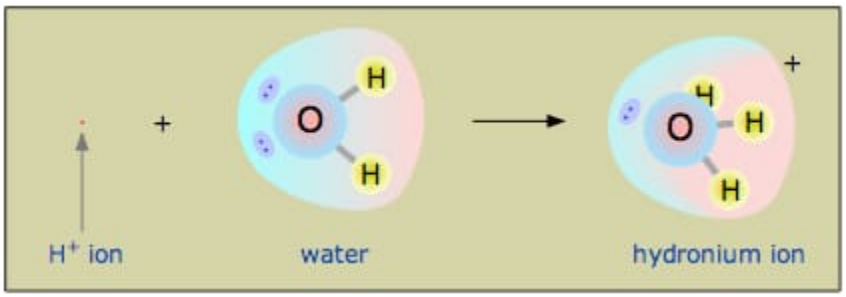
Acid-base reactions, in which protons are exchanged between donor molecules (acids) and acceptors (bases), form the basis of the most common kinds of equilibrium problems which you will encounter in almost any application of chemistry. In order to thoroughly understand the material in this unit, you are expected to be familiar with the following topics which were covered in the separate unit Introduction to Acid-Base Chemistry:

- The Arrhenius concept of acids and bases
- The Brønsted-Lowry concept, conjugate acids and bases
- Definition of pH and the pH scale
- Strong vs. weak acids and bases

You should also have some familiarity with the principles of chemical equilibrium, and know how to write and use equilibrium constants.

Proton transfer reactions in acid-base equilibria

As you should recall from your earlier introduction to acids and bases, the +1 electric charge of the tiny proton (a bare hydrogen nucleus) is contained in such a miniscule volume of space that the resulting *charge density* is far too large to enable its independent existence in solution; it will always attach to, and essentially bury itself in, the non-bonding orbitals of a solvent. Thus in aqueous solution, what we commonly represent as the "hydrogen ion" H^+ is more accurately described as the *hydronium ion* H_3O^+.



Acid-base chemistry is a *transactional process* in which protons are exchanged between two chemical species. A molecule or ion

that loses or "donates" a proton is acting as an *acid*; a species that receives or "accepts" that proton plays the role of a *base*.

Defining an acid as a substance that simply "dissociates" into its component parts

$$HA \rightarrow H^+ + AB^- \tag{13.1.1}$$

fails to capture the concept outlined in the above box. It is therefore important that you become comfortable with the Brønsted-Lowry model of acid-base behavior that represents the generalized acid-base reaction as a *proton transfer process*:

$$\underbrace{HA}_{\text{acid}} + \underbrace{B}_{base} \rightarrow \underbrace{A^-}_{\text{conjugate base}} + \underbrace{HB^+}_{\text{conjugate acid}} \tag{13.1.2}$$

The actual electric charges of the reaction products will of course depend on the particular nature of the species A, but the base will always have one more negative charge than the acid HA. Although anyone familiar with chemicals tends to regard certain substances as "acids" and others as "bases", it is important to bear in mind that these labels have meaning only in terms of the particular process being considered. For example, in the absence of water, ammonia, which we usually think of as a base, can lose a proton (and thus act as an acid) to a "stronger" base B to form the amide ion:

$$NH_3 + B \rightarrow NH_2^- + HB^+ \tag{13.1.3}$$

Similarly, anhydrous sulfuric acid can accept a proton from a stronger acid AH:

$$H_2SO_4 + AH \rightarrow H_3SO_4^+ + A^- \tag{13.1.4}$$

However, this is rather exotic stuff that is beyond the scope of this introductory lesson. The acid-base processes we describe in this set of lessons take place in aqueous solution, so we will always assume that H_2O, which can act either as a proton donor or proton acceptor, plays an active role:

- A substance that acts as an acid in aqueous solution donates a proton to H_2O, yielding the hydronium ion H_3O^+.

$$HA + H_2O \rightarrow H_3O^+ + A^- \tag{13.1.5}$$

- Similarly, a base will remove a proton from H_2O, yielding the hydroxide ion OH^-.

$$B + H_2O \rightarrow BH^+ + OH^- \tag{13.1.6}$$

Autoprotolysis

A small number of acids are themselves amphiprotic, meaning that one molecule can transfer a proton to another molecule of the same kind. The best-known example of this process, known as *autoprotolysis*, is of course that of water:

$$2H_2O_{(l)} \rightleftharpoons OH^- + H_3O^+ \tag{13.1.7}$$

Pure liquid ammonia and sulfuric acid are other well-known examples:

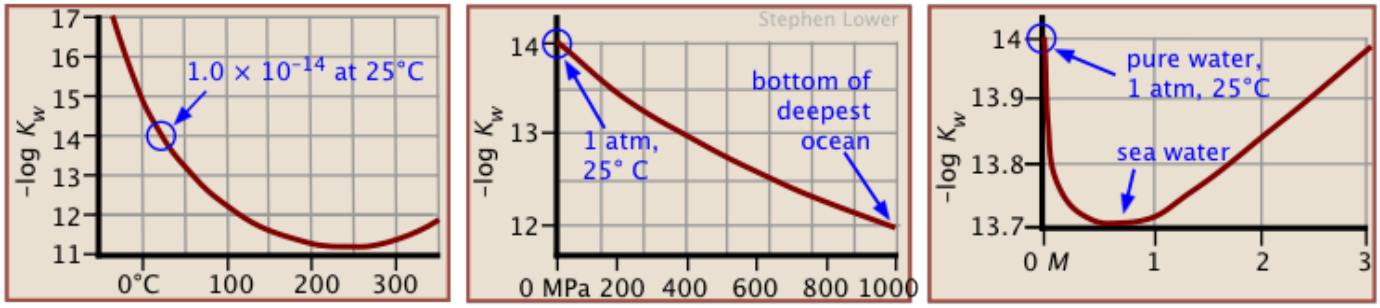
$$2NH_{3(l)} \rightleftharpoons NH_2^- + NH_4^+ \tag{13.1.8}$$

$$2H_2SO_{4(l)} \rightleftharpoons HSO_4^- + H_3SO_4^+ \tag{13.1.9}$$

Autoprotolysis reactions such as these typically proceed to only a very small extent. The equilibrium constant for Equation 13.1.7 is commonly written as

$$K_w = [H^+][OH^-] \tag{13.1.10}$$

and is usually known as the *dissociation constant* or *ion product* of water. For most purposes in elementary courses, we can use the value 1.0×10^{-14} for K_w. But you should understand that this value is correct only for pure water at 25° C and 1 atm pressure. K_w varies considerably with temperature, pressure, and the ionic content of water:



It should be obvious that anyone working in the fields of physiological or oceanographic chemistry **cannot** routinely use 1.0×10^{-14} for K_w!

The equilibrium constants K_a and K_b

Acids and bases vary greatly in their "strength" — that is, their tendencies to donate or accept protons. As with any chemical reaction, we can define an equilibrium constant K_c whose numerical value reflects the extent of the reaction — that is, the relative concentrations of the products and reactants when the reaction reaches equilibrium. The strength of an acid HA in water can be defined by the equilibrium

$$HA + H_2O \rightleftharpoons H_3O^+ + A^- \tag{13.1.11}$$

$$K_a = \frac{[H_3O^+][A^-]}{[HA]} \tag{13.1.12}$$

Similarly, the strength of the base A^- in water is defined by

$$A^- + H_2O \rightleftharpoons HA + OH^- \tag{13.1.13}$$

$$K_b = \frac{'[HA][OH^-]}{[A^-]} \tag{13.1.14}$$

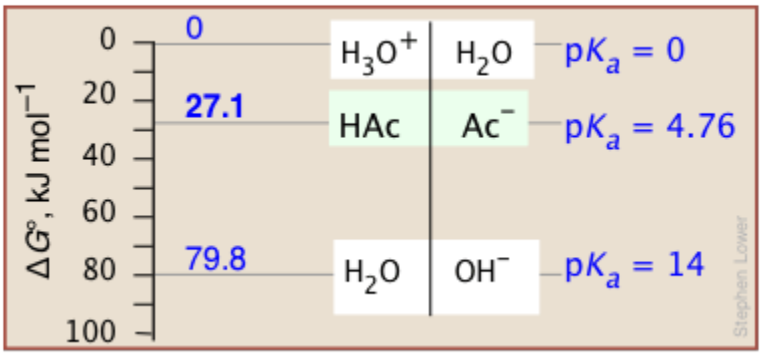
Note carefully that reaction 13.1.12 is *not* the reverse of 13.1.14.

K_a and K_b are inversely proportional

How are K_a and K_b related? The answer can found by multiplying the above two expressions for K_a and K_b:

$$K_a\,K_b = \frac{[H_3O^+]\,[A^-]}{[HA]}\,\frac{[HA]\,[OH^-]}{[A^-]} = K_w \tag{1-3}$$

Alternatively, we can simply add the two corresponding reaction equations:



and — as you should recall from your earlier study of equilibrium, the equilibrium constant of the sum of two reactions is the product of the two K's. So either way you figure it, the K_a for an acid and K_b for its conjugate base are related in a very simple way:

$$K_a K_b = K_w \tag{13.1.15}$$

Clearly, as the strength of a series of acids increases, the strengths of their conjugate bases will decrease, this inverse relation between K_a and K_b is implicit in Equation 13.1.15. You will recall that the pH scale serves as a convenient means of compressing a wide range of $[H^+]$-values into a small range of numbers. Just as we defined the pH as the negative logarithm of the hydrogen ion concentration, we can define

$$pK = -\log K \tag{13.1.16}$$

for any equilibrium constant. Acid and base strengths are very frequently expressed in terms of pK_a and pK_b. From Equation 13.1.15, it should be apparent that

$$pK_a + pK_b = pK_w \ (14.0 \ at \ 25^oC) \tag{13.1.17}$$

pK values of some common acids and bases

In order to convert *pK_a* into *K_a*, simply take the negative antilog of *pK_a*. Thus, for acetic acid with

$$pK_a = 4.7 \tag{13.1.18}$$

$$K_a = 10^{-4.7} = 2.0 \times 10^{-5} \tag{13.1.19}$$

The following table gives the p*Ka* values for a number of commonly-encountered acid-base systems which are listed in order of decreasing acid strength. It's worth taking some time to study this table in some detail.

lease take special note of the following points:

- Acids **(1-4)** are the major strong acids. All dissociate in water to produce the hydronium ion **(5)**, *the strongest acid that can exist in water*.
- **(6)** H_2SO_3 has never been detected; "sulfurous acid" and its formula refer to a solution of SO_2 in water.
- The names *bisulfate* **(3, 7)**, *bisulfite* **(6, 16)**, and *bicarbonate* **(13)** are still commonly used by chemists, but the preferred names are now *hydrogen sulfate*, *hydrogen sulfite*, and *hydrogen carbonate*, respectively. Similarly, *bisulfide* **(14, 26)** is also known as *hydrosulfide*.
- The ion **(9)** as it exists in aqueous solution is often represented as Fe^{3+} and is known more commonly as the ferrous or iron(III) ion. The aqueous aluminum ion Al^{3+} **(12)** as well as many transition metal ions have similarly complex structures and give acidic solutions in water because the high positive charge increases the acidity of the attached H_2O units. For more on these aquo complexes, see this brief discussion, or this much more detailed one by Jim Cook of the UK.
- The molecule H_2CO_3 **(13)** exists as a minority species in an aqueous solution of CO_2 and has never been isolated. In most contexts, the formula H_2CO_3, the name *carbonic acid* and the *pK_a* 6.3 refer to "total dissolved CO_2" in the water. "True" H_2CO_3 is about a thousand times stronger than the *pK_a* of its equilibrium mixture with CO_2(aq).
- The sulfide, amide and oxide ions **(26-28)**, being stronger bases than OH^-, cannot exist in water.

Common acids and bases

ref	acid	*pK_a*	conjugate base	*pK_b*	
1	$HClO_4$	perchloric acid	~ –7	ClO_4^- chlorate	~ 21
2	HCl	hydrochloric acid	~ –3	Cl^- chloride	~ 17
3	H_2SO_4	sulfuric acid	~ –3	HSO_4^- bisulfate	~ 17
4	HNO_3	nitric acid	~ –1	NO_3^- nitrate	15

ref	acid	pK_a	conjugate base	pK_b	
5	**H_3O^+**	**hydronium ion**	**0**	**H_2O water**	**14.00**
6	H_2SO_3	sulfurous acid	1.8	HSO_3^- bisulfite	12.2
7	HSO_4^-	bisulfate ion	1.9	SO_4^{2-} sulfate	12.1
8	H_3PO_4	phosphoric acid	2.12	$H_2PO_4^-$ dihydrogen phosphate	11.88
9	$[Fe(H_2O)_6]^{3+}$	hexaaquoiron(III)	2.10	$[Fe(H_2O)_5OH]^{2+}$	11.90
10	HF	hydrofluoric acid	3.2	F^- fluoride	10.8
11	CH_3COOH	acetic acid	4.76	CH_3COO^- acetate	9.3
12	$[Al(H_2O)_6]^{3+}$	hexaaquoaluminum(III)	4.9	$[Al(H_2O)_5OH]^{2+}$	9.1
13	H_2CO_3	carbonic acid	6.3	HCO_3^- bicarbonate	7.7
14	H_2S	hydrogen sulfide	7.04	HS^- bisulfide	6.96
15	$H_2PO_4^-$	dihydrogen phosphate	7.2	HPO_4^- monohydrogen phosphate	6.8
16	HSO_3^-	bisulfite ion	7.21	SO_3^{2-} sulfite	6.79
17	HOCl	hypochlorous acid	8.0	OCl^- hypochlorite	6.0
18	HCN	hydrocyanic acid	9.2	CN^- cyanide	4.8
19	H_3BO_4	boric acid	9.30	$B(OH)_4^-$ borate	4.79
20	NH_4^+	ammonium ion	9.25	NH_3 ammonia	4.75
21	$Si(OH)_4$	*ortho*-silicic acid	9.50	$SiO(OH)_3^-$ silicate	4.50
22	HCO_3^-	bicarbonate ion	10.33	CO_3^{2-} carbonate	3.67
23	HPO_4^{2-}	monohydrogen phosphate ion	12.32	PO_4^{3-} phosphate	1.67
24	$SiO(OH)_3^-$	silicate ion	12.6	$SiO_2(OH)_2^{2-}$	1.4
25	H_2O	water	14.00	OH^- hydroxide	0
26	HS^-	bisulfide ion	~ 19	S^{2-} sulfide	~ –5
27	NH_3	ammonia	~ 23	NH_2^- amide	~ -9
28	OH^-	hydroxide ion	~ 24	O^{2-} oxide	~ –10

The Fall of the proton

The concept of the "*potential energy of the proton*" (more properly, *proton free energy*), is not nearly as intimidating as it might sound; it is easy to grasp, requires no mathematics, and will enable you to achieve a far better understanding of acid-base chemistry.

Most courses (and their textbooks) do a good job of explaining the concept of proton transfer, but few of them provide you with a clear basis for understanding and making useful predictions about the *direction* and *extent* this transfer will take. For example, when the gas HCl is added to water, you probably know that the reaction proceeds strongly to the right:

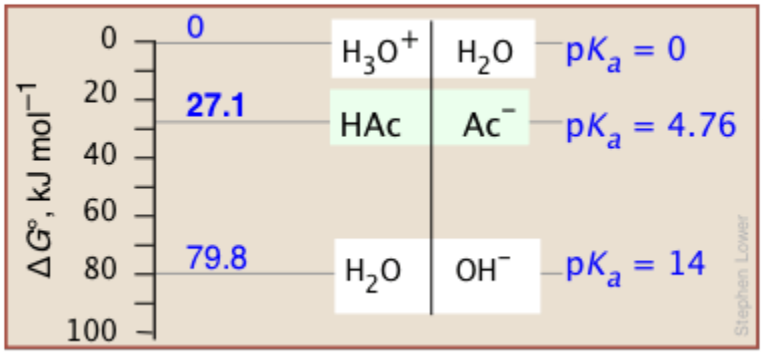
$$HCl + H_2O \rightleftharpoons H_3O^+ + Cl^- \tag{13.1.20}$$

...but what principle determines this, and what is to prevent the hydronium ion (also an acid) from pushing one of its own protons back onto a chloride ion, and thus reversing the reaction?

The concept of the *proton free energy level* was introduced in R.W. Gurney's classic *Ionic Processes in Solution* (1953). The best detailed exposition can be found in the first (1970) edition of *Aquatic Chemistry* by W. Stumm and J.L. Morgan.

The answer is that a proton is never really "pushed"; it will spontaneously seek out the lowest potential energy state (that is, the strongest base) it can find. Think how a macroscopic object such as a book will tend to fall to a location where its [gravitational] potential energy is as low as possible. Of course, the potential energy of a proton has nothing to do with gravity, but rather with how tightly it can bind to a base. This kind of potential energy is known as *proton free energy* (PFE) — but in the context of first-year chemistry, we can also just call it "proton energy".

You are not sure what free energy is? Don't worry about it for the time being; just think of it as you would any other form of potential energy: something that falls when chemical reactions take place. See here for the lesson on free energy.



For the present, all you need to know is that a proton will tend to "fall" to the lowest potential energy state it can find. In keeping with this idea, you can think of an acid as a *proton source*, and a base as a *proton sink*. Using this terminology, we can depict the generalized process HA + B → AB^- + HB^+ as in this plot, which depicts the proton on HA "falling" to the lower-PFE "sink" provided by the base B, leaving AB^- and HB^+ as products.

We will be making a lot of use of schematics like this farther on, so take a moment to familiarize yourself with information it depicts.

This "source/sink" nomenclature recalls the tendency of water to flow down from a high elevation to a lower one; this tendency (which is related to the amount of energy that can be extracted in the form of electrical work if the water flows through a power station at the bottom of the dam) will be directly proportional to the difference in elevation (difference in *potential energy*) between the source (top of the dam) and the sink (bottom of the dam).

Thus you can regard proton free energy as entirely analogous to the gravitational potential energy of the water contained in a dam; in this case, we are dealing with *chemical* potential energy, more commonly known as *free energy*. Take particular note of the following:

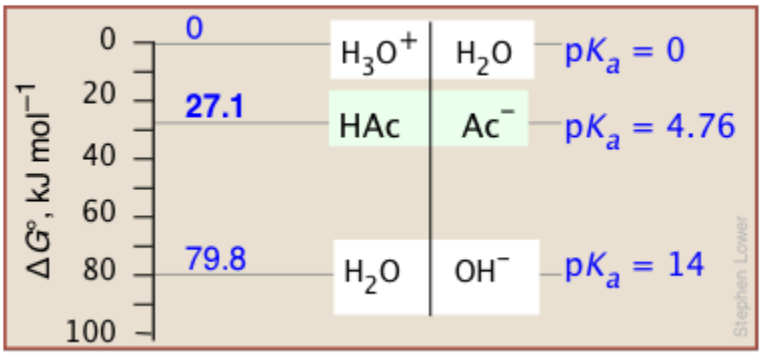
- The ordinate represents the relative free energy of a mole of protons when attached to a given base; we label it simply "proton energy" for brevity.
- The gray vertical line at the center of the diagram separates the acidic form of each conjugate pair from its base form; the terms "proton sources" and "proton sinks" are there to remind us of the role these forms play in acid-base chemistry.
- The red diagonal arrow represents the actual process in which a mole of protons "falls" from the acid HA to the base B. The fact that the arrow slants downward indicates that this process is spontaneous, meaning that the reaction proceeds to completion. In other words, the equilibrium constant for this reaction is very large.
- The green horizontal arrows depict the changes (that is, the formation of the two new conjugate species) that result from the proton transfer as the protons "fall" (in free energy) from the acid HA to the base B.

Strong acids in water

When you were first introduced to the Brønsted-Lowry model, you learned that the proper way to represent the "dissociation" of a strong acid like hydrochloric is

$$HCl + H_2O \rightarrow H_3O^+ + Cl^- \tag{13.1.21}$$

Thus water is revealed as an active participant in the behavior of an acid, rather than as an inert spectator or solvent. Although analogous acid-base reactions can take place in other solvents such as liquid ammonia, virtually all of the acid-base chemistry we encounter in daily life is based on water. Since water is acting as the proton acceptor here, it must be fulfilling the role of the base B in the schematic given above. We can therefore construct a similar plot specifically for HCl (or, in fact, for any strong acid).



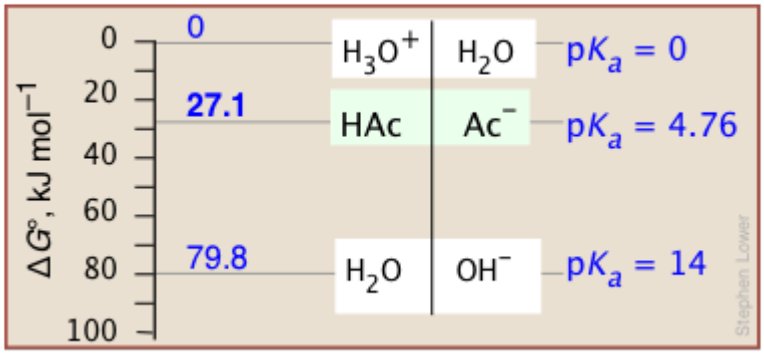
Recall that, for strong acids, this reaction is essentially complete. This means that a "solution of hydrochloric acid" is in reality a solution of hydronium and chloride ions; except in solutions much more concentrated than 1 M, the species HCl is only present in trace amounts.

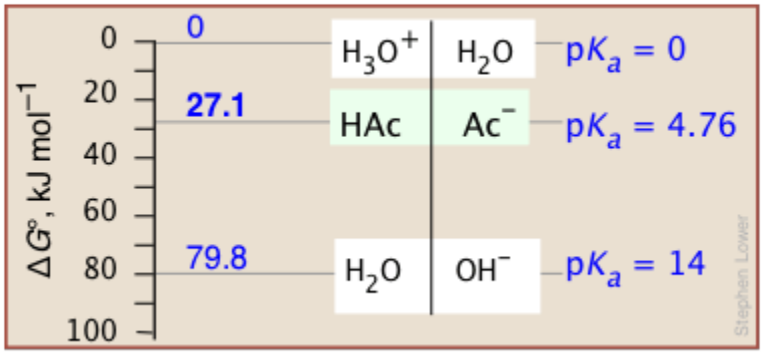
Weak acids in water

When acetic acid CH_3COOH is dissolved in water, the resulting solution displays only very mild acidic properties. For simplicity we will represent acetic acid and the acetate ion CH_3COO^- by HAc and Ac^-, respectively:

$$HAc + H_2O \rightleftharpoons H_3O^+ + Ac^- \tag{13.1.22}$$

But although this equation has the same form as (2-1) above, it hides an important qualitative difference: HAc is a *weak* acid, meaning that the proton transfer takes place to only a tiny extent; the equilibrium strongly favors the left side:

HAc + H_2O  H_3O^+ + Ac^-



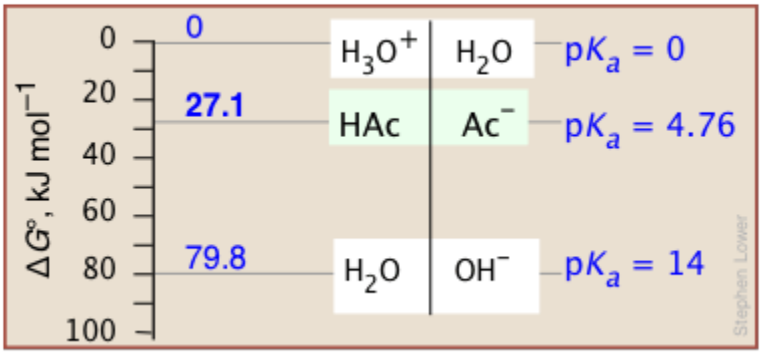
The corresponding proton free energy diagram now shows the H_3O^+-H_2O system at the top, with the proton-transfer arrow pointing upward, indicating that the protons get kicked up to a *higher* free-energy level in this process.

Where does this additional energy come from? Simply from the random thermal energy in the solution. As you might expect, this is not a thermodynamically favorable process, so it only happens to a very small extent. In contrast to the HCl example we described previously in which 99.99+ percent of the HCl molecules end up as hydronium ions, almost all of the acetic acid molecules remain unchanged, and only tiny amounts of H_3O^+ and Ac^- are produced.

Strong and weak acids both donate protons to water, yielding hydrogen ions and thus rendering the solution acidic.

What's different is that virtually *all* of the protons of the strong acid "fall" to the H_2O level, whereas only a tiny fraction of the weak acid molecules acquire enough thermal energy to promote their protons up to the H_2O level.

The fundamental difference between strong acids and weak acids is clearly evident when we combine the preceding two proton-free energy diagrams into one. Strong acids (strong "conjugate-pairs") are always above the PFE of the H_3O^+-H_2O system, while the weak conjugate pairs are below it.



The other big difference with acetic acid is that water is acting as a *base* here, rather than as an *acid* as it did in the HCl example. So it's not surprising that water, the basis of almost all acid-base chemistry, should appear in the middle of these diagrams.

Water can swing both ways

Substances that can both donate and accept protons are said to be *amphiprotic*. Many oxides and hydroxides behave in this way, and since H_2O falls into both of these classes, it is not surprising that water is amphiprotic:

- *Water ("the-acid")*:

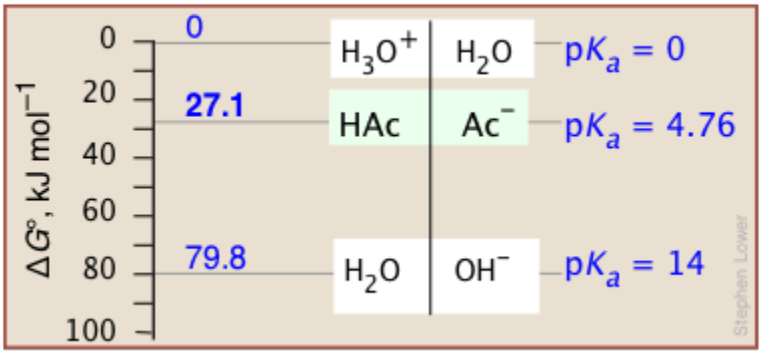
$$H_2O + B \rightleftharpoons BH^+ + OH^- \tag{13.1.23}$$

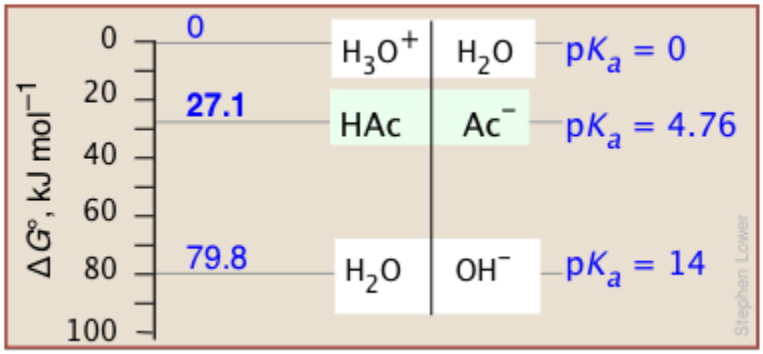
- *Water ("the base")*:

$$H_2O + AH \rightleftharpoons H_3O^+ + A^- \tag{13.1.24}$$

("AH" above represents any acid; B stands for any base)

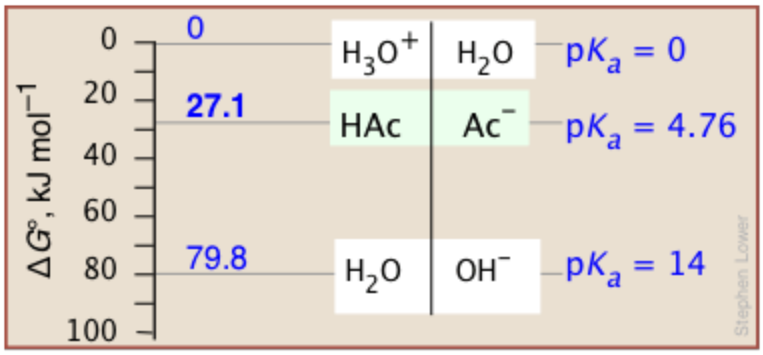
These two opposing tendencies can be represented by





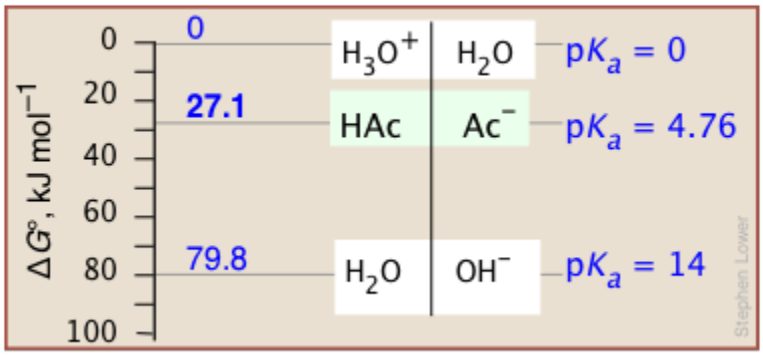
In the case of water, the contest is a draw; not only does neither side win, but the fraction of H_2O molecules that break up in this way is miniscule: In pure water, only about one H_2O molecule out of 10^9 is "dissociated" at any instant:

H_2O



$H^+ + OH^-$ (2-4)

The actual reaction, of course, is the proton transfer



(2-5)

for which the equilibrium constant

$K_w = [H_3O^+][OH^-]$ (2-6)

is known as the *ion product of water*. The value of K_w at room temperature is 1.008×10^{-14}. We can also express this in logarithmic terms: $pK_w = 14.0$.

As with any equilibrium constant, the value of K_w is affected by the

- *temperature*: K_w undergoes a 10-fold increase between 0°C and 60°C
- *pressure*: K_w is about doubled at 1000 atm
- *presence of ionic species in the solution*.

Because most practical calculations involving K_w refer to ionic solutions rather than to pure water, the common practice of using 10^{-14} as if it were a universal constant is unwise; under the conditions commonly encountered in the laboratory, pK_w can vary from about 11 to almost 15. In seawater, K_w is 6.3×10^{-12}. (See Stephen J. Hawkes: "*pK_w is almost never 14.0*", *J. Chem. Education* 1995: 72(9) 799-802)

Notice that under conditions when K_w differs significantly from 1.0×10^{-14}, the pH of a neutral solution will not be 7.0.

For example, at a pressure of 93 kbar and 527°C, $K_w = 10^{-3.05}$, the pH of pure water would be 1.5. Similarly unusual conditions apply to deposits of water in geological formations and in undersea vents.

Example

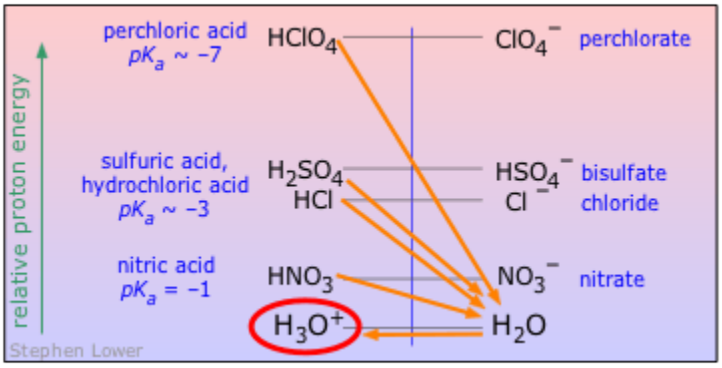
At 60° C, the ion product of water is 9.6×10^{-14}. What is the pH of a neutral solution at this temperature?

Solution:

Under these conditions, $[H^+][OH^-]$ = 9.6E–14. If the solution is neutral, $[H^+] = [OH^-] = (9.6E–14)^{½}$ = 3.1E–7, corresponding to a pH of log 3.1E-7 = **6.5**.

The leveling effect

The strengths of strong acids and bases are "leveled" in water. This means that **all strong acids and strong bases are equally strong in water**. To see this more clearly, notice that the free energies of protons in all of these acids (even nitric) are sufficiently above the H_3O^+/H_2O level that their "dissociation" is essentially complete. Thus



The strongest acid that can exist in water is H_3O^+ ;

The strongest base that can exist in water is OH^-

Similarly, the hydroxide ion is the strongest base that can exist in water.

So what happens if you add a soluble oxide such as Na_2O to water? Since O_2^- is a proton sink to H_2O, it will react with the solvent, leaving OH^- as the strongest base present: $Na_2O + H_2O \rightarrow 2\ OH^- + Na^+$.

Thus all bases stronger than OH^- appear equally strong in water — simply because they are all converted to OH^- .

3 Why acids react with bases, but not always completely

PFE's are the key to understanding acid-base reactions

In the preceding section, we have employed simple PFE (proton free energy) diagrams to help you better understand the concepts of strong and weak acids, water as an acid or a base, and the leveling effect. But we can gain a more generalized and comprehensive view of acid-base chemistry with the aid of a PFE diagram that shows a whole series of acid-base systems, arranged in order of decreasing acid strengths.

How acid strengths relate to proton free energies

As we explained near the beginning of this lesson, we ordinarily express the strength of an acid by quoting its "dissociation constant" K_a, which is the equilibrium constant for transfer of a proton to the base H_2O.

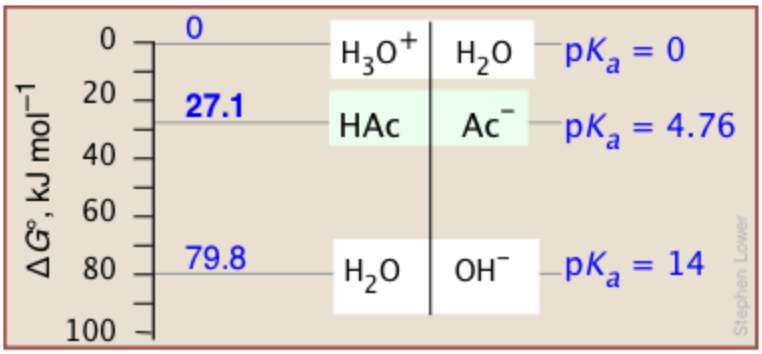
$HA + H_2O \rightarrow H_3O^+ + A^-$

We also introduced the concept of PFE as an alternative way of expressing the strength of an acid. If both PFE and K_a express the same thing, you may wonder how these two quantities are related.

If you *have* had an introduction to thermodynamics and specifically to free energy *G*, you might wish to look at the Appendix section of this lesson to review the relationship between the fall in free energy and the value of pK_a.

A thorough answer to this question requires a bit of thermodynamics, which you may not have yet studied, but you don't really need any of the math here. It should suffice to show the results in the form of a plot, which yields a simple straight-line relationship when plotted on a semi-logarithmic scale.

Note that



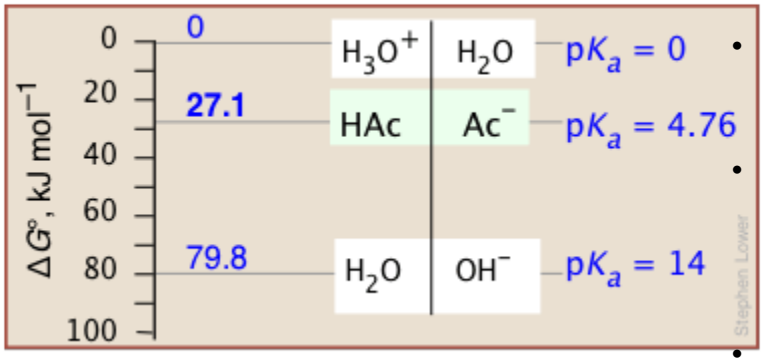
- The extent of *any* chemical reaction is proportional to the fall in free energy when it takes place. So negative ΔG values correspond to larger equilibrium constants.
- For acid-base reactions specifically, K_a is inversely proportional to ΔG, and thus to the fall in PFE.
- This means that the proton on a strong acid ($K_a > 1$) has a large negative PFE; it has a "long way to fall" to the H_2O level.
- Conversely, a proton on a weak acid ($K_a < 1$) can only move up to the H_2O level when supplied with thermal energy from the surroundings.
- A K value of $10^0 = 1$ corresponds to a zero-free energy change, meaning that the reaction is at equilibrum when [HA] = [A^-].
- Proton-free energy changes exceeding –20 kJ mol^{-1} result in K_a values so large that the proton transfer can be considered "complete" for most practical purposes.

Understanding the PFE chart

On the PFE diagram shown below, each acid-base system is represented by a horizontal line connecting an acidic species with its conjugate base. Each line is placed at a height on the diagram that is proportional to its pK_a, as measured on the logarithmic scale (labeled "pH") at the right. What we have, then, is a list of acid-base systems arranged in order of decreasing acid strength.

Think of this chart as an alternative view of the table of acid strengths displayed near the top of this page, in which some of the information that is "hidden" in the table is now revealed in an easily-grasped way.

Take particular note of the following points:



- The ΔG scale on the left measures the fall in potential energy when one mole of protons is transferred from a given acid to H_2O. Negative values correspond to stronger acids.
- Any acid shown on the left side of the vertical line running down the center of the diagram can donate protons to any base (on the right side of the line) that appears *below* it. The value of the equilibrium constant for such a reaction will exceed unity; for ΔG separations of more than about 10 kJ/mol, the reaction can be considered to be complete.
- The zero free energy assigned to the hydronium ion corresponds to the standard free energy of formation $\Delta G_f°$ of this ion.

Notice how the two conjugate pairs H_3O^+ / H_2O and H_2O/OH^- bracket the 80-kJ/mol region that comprises the weak acids. The weakness of these acids reflects the energetic cost of moving a proton "up-hill" to the H_2O level.

The strong acids at the top of the table can all be regarded as totally "dissociated" in water. This means that these acids cannot exist in water, as they are totally transformed into H_3O^+. Of course, they will also react with any bases below the H_2O level that may be available. (See "leveling effect" below.)

Similarly, the two strong bases NH_2^- and O_2^- shown at the bottom of the table cannot exist in water. Thus if solid sodium oxide Na_2O is added to water, the oxide ions are such a low-energy sink that will totally fill up with protons from the H_2O acid level which defines the top of the strong-base range.

Notice the pH scale along the right side, and note the range of pH values possible in water. We will say more about the significance of pH in these diagrams farther on.

The PFE chart helps you visualize (and understand!) acid-base chemistry

The virtue of a PFE chart is that you can see, at a glance, the relations between conjugate acid-base pairs that enables you

to make quick, qualitative determinations of what is going on even in complex mixtures of acids and bases. The speed and convenience this affords enables you to build a much deeper conceptual understanding of acid-base chemistry.

In order to do the same thing by consulting a simple list of acid K_a's or pK_a's, you also need to be familiar with the principles of chemical equilibrium and thermodynamics — and then take the time to work through the arithmetic.

In the preceding sections we have shown how the PFE-chart approach can clarify the distinction between strong and weak acids. We will now go on to describe a variety of other principles that are easily grasped by consulting a PFE chart.

Significance of the pH scale

You will have noticed the scale labeled "pH" at the left side of the PFE plot.

First, please understand what this scale does NOT do; it does not tell you what the pH of the solution will be when you add an acid to pure water. This pH will depend on both the strength (K_a or pK_a) of the acid, and on its concentration in the resulting solution; you will learn how to carry out pH calculations in later lessons.

All but the very weakest acids can drive the pH down to near the bottom of the scale if their solutions are sufficiently concentrated. The only thing you can be sure of is that the pH of a solution of a solution of an acid in pure water will never be greater than 7.

pH sets the average proton free energy

When pH was first introduced in 1909, it was defined in terms of the "hydrogen ion concentration" [H^+]. This has since been amended to the hydrogen ion *activity* {H^+}. The latter term refers to the "effective" concentration of these ions — that is to the "availability" of protons (regardless of whether they physically exist as H_3O^+ units or in other forms) to react with bases. But from the standpoint of thermodynamics, "availability" is just a loose term for "free energy". So we can also say that

pH is a measure of the average PFE in an aqueous solution

The more negative the pH, the higher the proton free energy in the solution. (The inverse relation is a consequence of the negative-logarithm definition of pH.)

The pH has the dual advantages of being both dimensionless and readily observable. More importantly, we can easily alter the pH of a solution by adding some strong acid or base. And when we do this, we are in effect using the pH as a tool for controlling the average PFE in the solution.

More on this in a moment; first, let us introduce an important relation that you may have seen in an earlier chemistry course, and which we will discuss more in a later lesson in this group.

pH can control the relative abundance of conjugate acid-base pairs

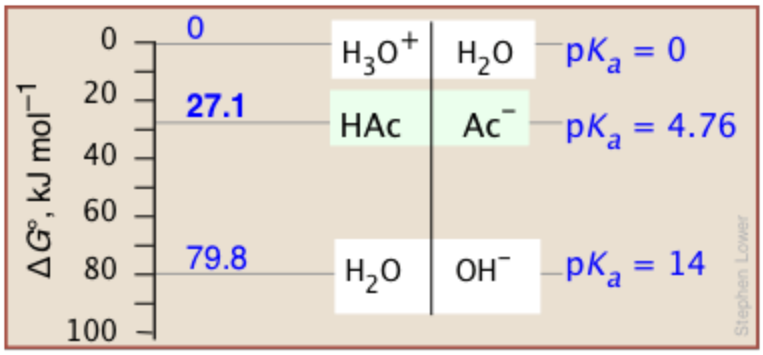
Recalling the equilibrium expression for a weak acid

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{13.1.25}$$

We can solve this for $[H^+]$:

$$H^+] = K_a \frac{[HA]}{[A^-]} \tag{13.1.26}$$

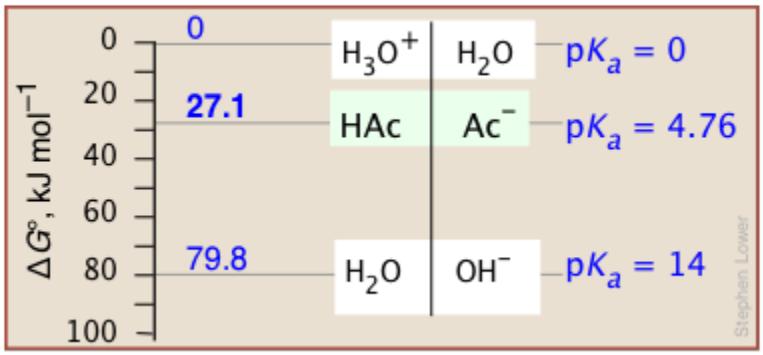
Re-writing this in terms of negative logarithms, this becomes



(3-3)

or, since pK_a = – log K_a, we invert the ratio to preserve the positive sign:

We will make extensive use of this equation in a later lesson on acid-base buffering and titration.



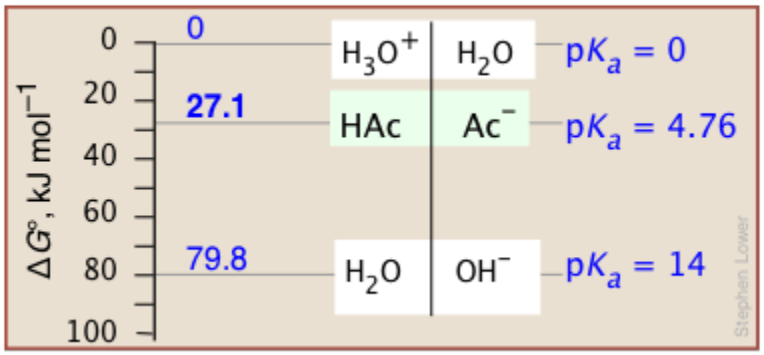
(3-4)

This extremely important relation tells us that the pH of a solution containing a weak acid-base system controls the relative concentrations of the acid and base forms of that system. Of special interest is the case in which the pH of a solution of an acid-base system is set to the value of its pK_a. According to the above equation, when pH = pK_a, the log term becomes zero, so that the ratio

$[AB^-] / [HA] = 10^0 = 1$, meaning that $[HA] = [AB^-]$. In other words,

This pH adjustment can be easily made by adding appropriate amounts of a strong acid or strong base to the solution.

When the pH of a solution is set to the value of the pK_a of an acid-base pair, the concentrations of the acid- and base forms will be identical. This condition can be represented schematically:



This is the basis for the arrangement of the various acid-base systems depicted in the PFE diagram; each conjugate pair is placed at a location on the vertical pH scale that corresponds to the acid's pK_a. It is instructive to consider how sensitively the concentration ratio $[A^-] / [HA]$ depends on the pH.

Example 13.1.2

Hypochlorous acid, HOCl, has a pK_a of 8.0.

a. Estimate the ratio R = [OCl^-] / [HOCl] in a solution of sodium hypochlorite whose pH has been adjusted to 8.2.
b. If the initial concentration of the hypochlorite solution was 0.10*M*, what will be the concentrations of the acid and base forms at pH 8.2?

Solution

a) Substituting in Eq 3-4, the log term becomes (8.2 – 8.0) = 0.2,
so $R = 10^{0.2} = 1.6$.

Note that it does not matter whether the initial solution consists of HOCl or NaOCl or some mixture thereof; adjusting the pH *forces* the system composition to the ratio calculated above.)

b) In the resulting solution, the base form OCl^- is the predominant species, the mole ratio R being 1.6/1. The mole fraction of OCl^- will therefore be 1.6/2.6 = 0.62, making that of HOCl (1.0 – .62) = 0.38. The actual concentrations will be 0.1 times these values, reflecting the 0.10*M* concentration of the solution.

Example 13.1.3

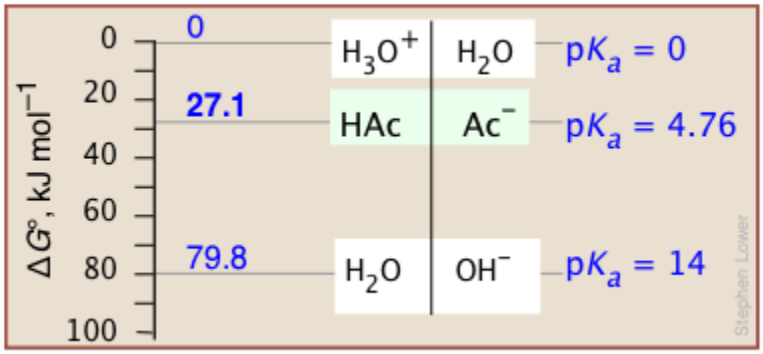
To what pH must we adjust a solution of acetic acid (pK_a 4.7) in order to convert 20 percent of the acid into its base form?

Solution:

In the final solution, the mole fractions will be [AB^-] = .20, [HA] = 0.80; the ratio [AB^-]/[HA] = 0.20/0.80 = 0.25. Substituting into Eq 2-9, we have

$$pH = 4.7 + \log \backslash '0.25 = 4.7–0.6 = 4.1 \tag{13.1.27}$$

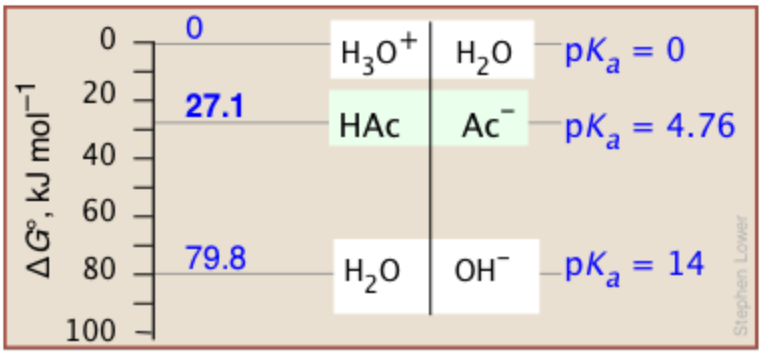
We can explore the relation between the pH and the distribution of conjugate species between the acid and base forms by carrying out calculations similar to those in Problem Example 2 for an acid-base system at different pH values.

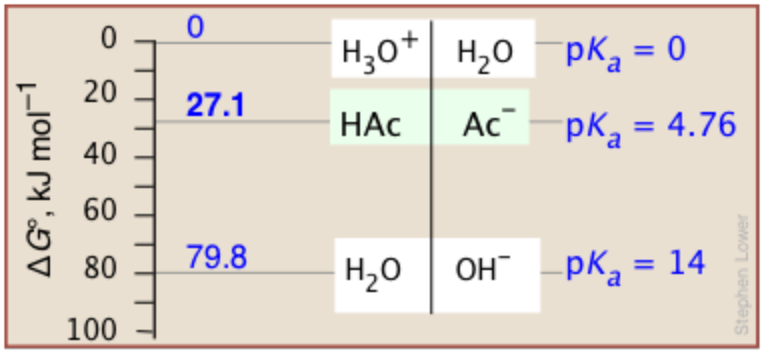


The orange bars and blue numbers show the mole fractions of the acid and base forms at pH values of the pK_a and the pK_a ± 1. The green numbers give the percent dissociation of the acid at each pH.

The apparent symmetry between the two extreme pH values suggests a simple relation between the log of the ratio

[AB^-] / [HA] and the pH. This is seen very clearly when this ratio is plotted as a function of pH.





Solutions containing many different weak acid-base systems are very commonly encountered in nature, especially in biological fluids or natural waters, including the ocean. When the pH of such solutions is altered by the addition of strong acid or base (as can occur, for example, when wastes from industrial processes or mine drainage, or acidic rain, enter a lake or stream); the distribution of all conjugate acid-base systems will change.

It is important to note that the [AB^-] / [HA] ratios (calculated here from Eq 3-4) yield only a qualitative picture and do not take into account the concentrations of the different acid-base systems or proton transfers between them. The latter processes would cause the lowest proton-vacant levels (that is, the strongest bases) to be completely filled from the bottom up, depending on the number of protons available. An exact calculation would require solving a set of simultaneous equations and is beyond the scope of this lesson.

A. What do we mean by the "strength" of an acid ?

The Brønsted-Lowry model defines the *strength* of an acid HA as its tendency to donate protons to a base B. But bases vary in their abilities to *accept* protons, so the tendency for the complete transaction

$HA + B \rightarrow BH^+ + A^-$(A-1)

to occur depends on the *two* processes

$HA \rightarrow H^+ + A^-$ K_1 (A-2)

$B + H^+ \rightarrow BH^+$ K_2(A-3)

K_a's are always expressed on a scale relative to the K_b of the solvent

Because it is impossible to study either of these steps independently, what we might call the "absolute" strength of an acid (Eq A-2) cannot be measured. For this reason, we define a "standard" base, usually the solvent, and most commonly, water. Thus we replace Eq A-3 with

$H_2O + H^+ \rightarrow H_3O^+ K_3$(A-4)

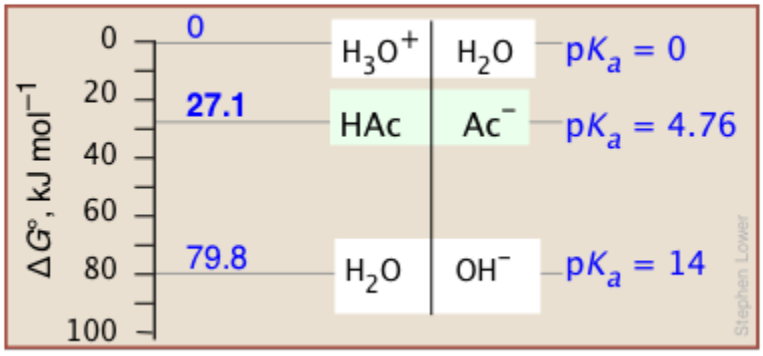
so that Eq A$^-$1 now expresses the strength of HA *relative to the base* H_2O:

$HA + H_2O \rightarrow H_3O^+ + A^-$ (A-5)

which, for convenience, we write in its abbreviated form

$HA \rightarrow H^+ + A^-$ (A-5)

whose equilibrium constant is



(A-6)

This value of K_a is, of course, numerically identical to that for Eq A-5.

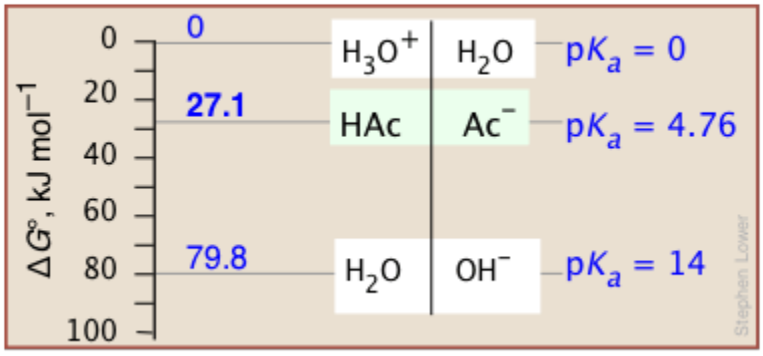
B. pH and proton free energy

This section may be of interest to readers who have studied elementary thermodynamics, and have some familiarity with Gibbs free energy. it is not required for understanding and using the concept of PFE.

From elementary thermodynamics, the driving force of a chemical reaction is given by the standard free energy change:

$\Delta G° = -RT \ln K$ (B-1)

Solving this equation for K yields

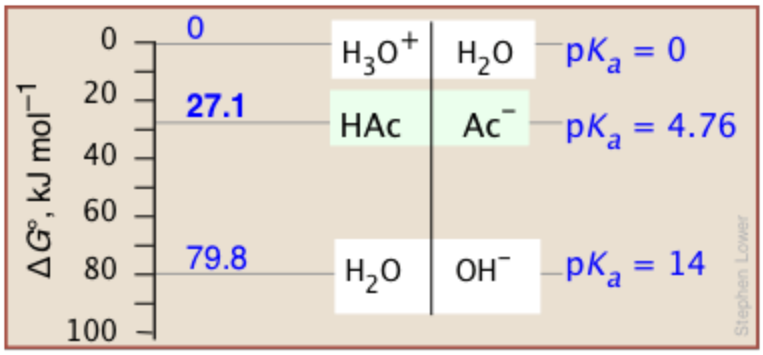


(B-2)

For a weak acid we know that

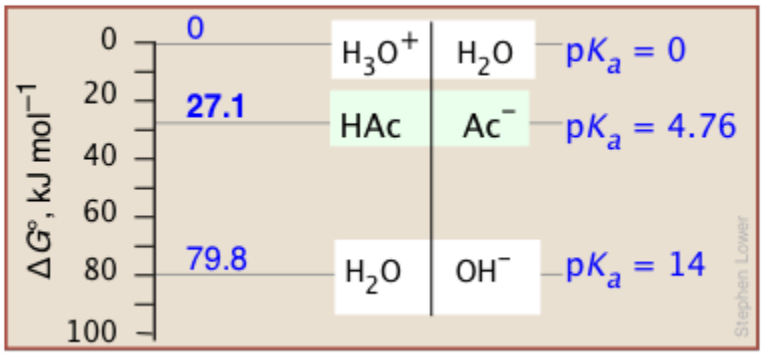
$$[H^+] = K_a \frac{[HA]}{[A^-]} \quad \text{(B-3)}$$

Substituting Eq B-2 into this expression, we get



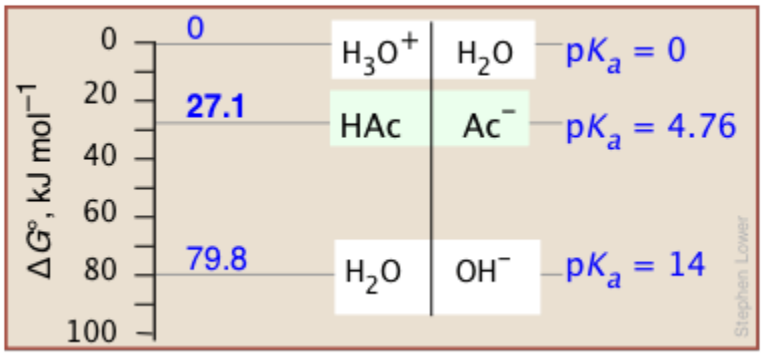
(B-4)

Next, we take the negative log of each term, recalling that ln x = 2.3 log x:



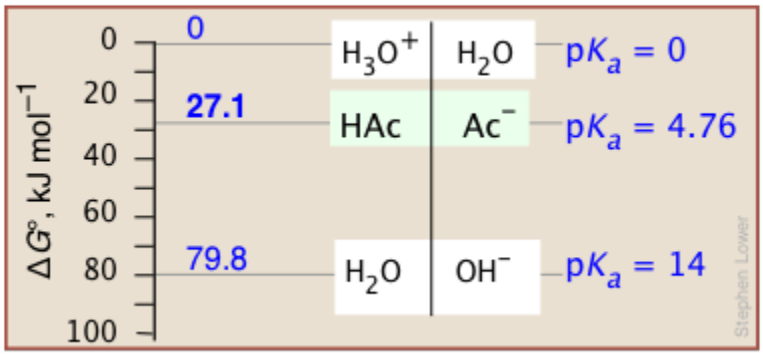
(B-5)

or



(B-6)

For the special case in which the concentrations of the conjugate pairs [H^+] and [HA] are equal, pH = pK_a, and the rightmost term above disappears, leaving



(B-7)

Thus the pH is a direct measure of the average proton free energy in a solution.

Example 13.1.1: **PFE level of acetic acid**

Construct a PFE diagram showing *a)* H_3O^+, *b)* acetic acid HAc (pK_a = 4.76), and *c)* H_2O.

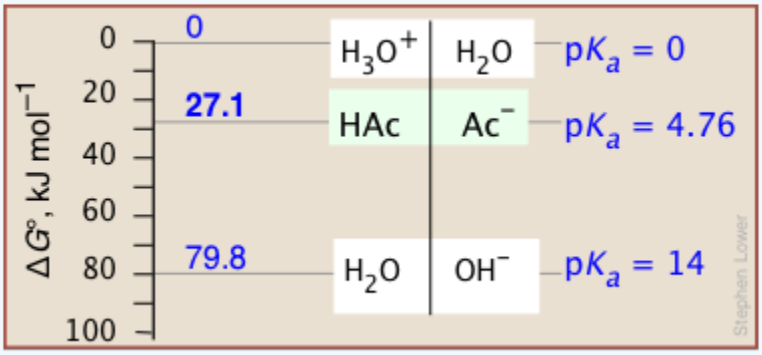
Solution

Solving Eq A⁻13 for ΔG° yields ΔG° = (2.3RT) × pK_a, or, at 25°C, ΔG° = (2.3 × 8.314 J K^{-1} mol^{-1}) × pK_a = 5698 J K^{-1} mol^{-1} × pK_a

a) The proton in H_3O^+ is already at the level of that in the standard base H_2O (Eq A⁻5), so for this transfer is [by definition] ΔG° = 5698 J K^{-1} mol^{-1} × 0 = **0**

b) The increase in the free energy required to transfer a proton from its level in HAc to that in H_2O is ΔG° = 5698 J K^{-1} mol^{-1} × 4.76 = **27.1 kJ mol^{-1}**

c) To transfer a proton from H_2O to another H_2O (autoprotolysis), its free energy must increase by 5698 J K^{-1} mol^{-1} × 14.0 = **79.8 kJ mol^{-1}**



This page titled 13.1: Introduction to Acid/Base Equilibria is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13.2: Strong Monoprotic Acids and Bases

Learning Objectives

Make sure you thoroughly understand the following essential concepts that have been presented above.

- *Estimate the pH* of a solution of a strong acid or base, given its concentration.
- Explain the distinction between the concentration of a substance and its *activity* in a solution.
- Describe the origin and effects of *ion-pairing* in concentrated solutions of strong acids.
- Name and write formulas for the *four major strong acids*.

To a good approximation, strong acids, in the forms we encounter in the laboratory and in much of the industrial world, have no real existence; they are all really solutions of $\mathrm{H_3O^+}$. So if you think about it, the labels on those reagent bottles you see in the lab are not strictly true! However, if the strong acid is highly diluted, the amount of $\mathrm{H_3O^+}$ it contributes to the solution becomes comparable to that which derives from the autoprotolysis of water. Under these conditions, we need to develop a more systematic way of working out equilibrium concentrations.

At Moderate Concentrations, Forget About Equilibria

A *strong acid*, you will recall, is one that is stronger than the hydronium ion $\mathrm{H_3O^+}$. This means that in the presence of water, the proton on a strong acid such as HCl will "fall" into the "sink" provided by H_2O, converting the latter into its conjugate acid $\mathrm{H_3O^+}$. In other words, $\mathrm{H_3O^+}$ is the strongest acid that can exist in aqueous solution.

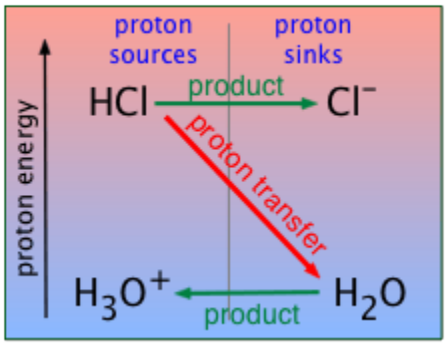


Figure 13.2.1

As we explained in the preceding lesson, all strong acids appear to be *equally* strong in aqueous solution because there are always plenty of H_2O molecules to accept their protons. This is called the "leveling effect".

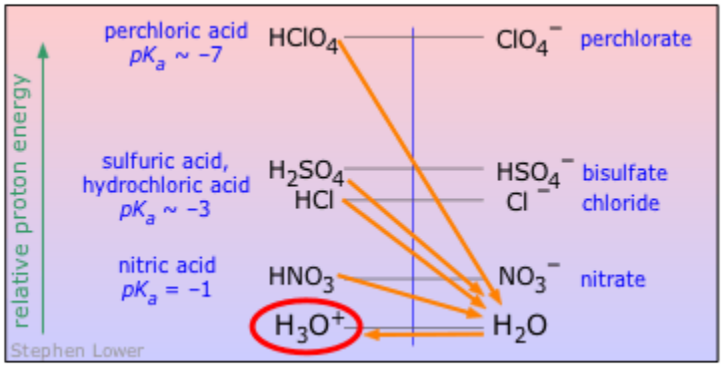


Figure 13.2.2

This greatly simplifies our treatment of strong acids because there is no need to deal with equilibria such as for hydrochloric acid

$$\mathrm{HCl} + \mathrm{H_2O} \longrightarrow \mathrm{H_3O^+} + \mathrm{Cl^-} \tag{13.2.1}$$

The equilibrium constants for such reactions are so overwhelmingly large that we can usually consider the concentrations of acid species such as "HCl" to be indistinguishable from zero. As we will see further on, this is not strictly true for highly-concentrated solutions of strong acids (Figure 13.2.3).

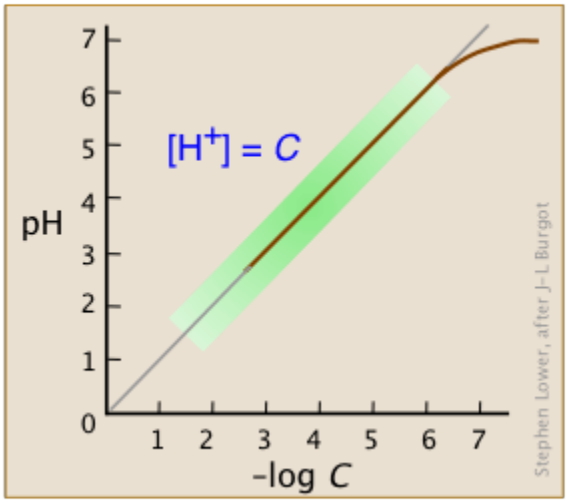


Figure 13.2.3

Over the normal range of concentrations we commonly work with (indicated by the green shading on this plot), the pH of a strong acid solution is given by the negative logarithm of its concentration in mol L^{-1}. Note that in very dilute solutions, the plot levels off, showing that this simple relation breaks down; **there is no way you can make the solution alkaline by diluting an acid!**

Example 13.2.1

What will be the pH of a 0.025 mol/l solution of hydrochloric acid?

Solution

If we assume all the hydronium concentration originates from the added acid. So

$$\mathrm{H_3O^+} \tag{13.2.2}$$

and we just find the negative logarithm of the concentration

$$pH = -\log_{10}[\mathrm{H_3o^+}] = -\log_{10} 0.025 = 1.6 \tag{13.2.3}$$

Since this pH is so far away from 7, our assumption is reasonable.

Major Strong Acids

Mineral acids are those that are totally inorganic. Not all mineral acids are strong; boric and carbonic acids are common examples of very weak ones. However, in ordinary usage, the term often implies one of those described below. With the exception of perchloric acid, which requires special handling, these are all widely used in industry and are almost always found in chemistry laboratories. Most have been known since ancient times

Table 13.2.1*: Mineral Acid properties*

acid	name	pK_a	base	pK_b
$HClO_4$	perchloric acid	~ –7	ClO_4^-	~ 21
HCl	hydrochloric acid	~ –7	Cl^-	~ 17
H_2SO_4	sulfuric acid	~ –7	HSO_4^-	~ 17
HNO_3	nitric acid	–1	NO_3^-	15
H_3O^+	**hydronium ion**	**H_2O**	14	

You should know the names and formulas of all four of these widely-encountered strong mineral acids.

- **Hydrochloric acid** HCl**:** In contrast to the other strong mineral acids, no pure compound "hydrochloric acid" exists. One of the most commonly used acids, hydrochloric acid is usually sold as a constant-boiling 37% solution of hydrogen chloride gas in water, making its concentration about 10 *M*. Although higher concentrations are possible, the high partial pressures of HCl in equilibrium with the solution makes them difficult to store, ship and work with. Commercial-grade hydrochloric acid is often called *muriatic acid*.
- **Nitric acid** HNO_3**:** Concentrated nitric acid is a 68% constant-boiling solution of HNO_3 in water, corresponding to about 11M. Although the acid itself is colorless, its slow decomposition into NO_2 (especially in the presence of light) often results in a yellow or orange color. More concentrated solutions are sold as "fuming nitric acid"; one form, available as "white fuming nitric acid" or "anhydrous nitricacid" contains 97.5% HNO_3 and only 2% water; the remainder consists of dissolved NO_2. Pure HNO_3 forms crystals that melt at –42°C. In addition to being a strong acid, nitric acid at high concentrations acts as a powerful oxidizing agent, a property that accounts for its use as a rocket fuel.
- **Perchloric acid** $HClO_4$: This is the strongest of the mineral acids, thanks to the electron-withdrawing action of the oxygen atoms which make it energetically easier for $HClO_4$ to lose its proton. Its usual commercial form is a constant-boiling 72.5% solution in water. Concentrated perchloric acid is a powerful oxidizing agent that can react explosively with organic materials and certain metals. Its use in the laboratory requires special handling. A 1947 explosion of a 1000-L vat of $HClO_4$ in a Los Angeles electroplating plant killed 17 people and damaged over 250 homes and other buildings.
- **Sulfuric acid** H_2SO_4**:** Sulfuric acid is by far the most industrially important acid, and is also the most concentrated one available. "Concentrated" sulfuric acid contains 98% by weight of H_2SO_4; its density of 1.83 kg/L and oil-like viscosity reflect this high concentration. "100% H_2SO_4" (which can be prepared However, is not stable) actually contains a variety of other species, all in equilibrium with H_2SO_4. These include $H_2S_2O_7$, $H_2S_4O_{13}$, and H_2SO_4's autoprotolysis products $H_3SO_4^-$ and HSO_4^+.

There is a class of super acids that are stronger than some of the common mineral acids. According to the classical definition, a superacid is an acid with an acidity greater than that of 100% pure sulfuric acid. Some, like fluorosulfuric acid FSO_3H are commercially available. Strong superacids are prepared by the combination of a strong Lewis acid and a strong Brønsted acid. The strongest known super acid is fluoroantimonic acid (H_2FSbF_6). This acid is so corrosive that the fumes alone will dissolve entire fume hoods, glass and plastic beakers, human skin, bone and most synthetic compounds.

Strong bases

The only strong bases that are commonly encountered are solutions of Group 1 hydroxides, mainly NaOH and KOH. Unlike most metal hydroxides, these solids are highly soluble in water, and can thus yield concentrated solutions of hydroxide ion, the strongest base that can exist in water — the ultimate aquatic proton sink. Sodium hydroxide is by far the most important of these strong bases; its common names "lye", "caustic soda" (or, in industry, often just "caustic"), reflect the diverse uses of NaOH.

Solid NaOH is usually sold in the form of pellets. When exposed to air, they become wet (*deliquescence*) and absorb CO_2, becoming contaminated with sodium carbonate. NaOH is the most soluble of the Group 1 hydroxides, dissolving in less than its own weight of water (111 g / 100 ml) to form a 2.8 M/L solution at 20°C. However, as with strong acids, the pH of such a solution cannot be reliably calculated from such a high concentration.

Acids at High Concentrations

At higher concentrations, intermolecular interactions and ion-pairing can cause the effective concentration (known as the *activity*) of $\mathrm{H_3O^+}$ to deviate from the value corresponding to the nominal or "analytical" concentration of the acid. Activities are important because only these work properly in equilibrium calculations. Also, pH is defined as the negative logarithm of the hydrogen ion activity, not its concentration. The relation between the concentration of a species and its activity is expressed by the *activity coefficient* γ:

$$a = \gamma C \tag{13.2.4}$$

As a solution becomes more dilute, γ approaches unity. At ionic concentrations not exceeding about 2 *M*, concentrations of typical strong acids can generally be used in place of activities without serious error. Note that activities of single ions other than $\mathrm{H_3O^+}$ cannot be determined, so activity coefficients in ionic solutions are always the average, or *mean*, of those for the ionic species present. This quantity is denoted as $\gamma_{\pm}$.

Table 13.2.2: Mean ionic activity coefficient of HCl at 25° C

molarity	$\gamma_{\pm}$
0.0005	0.975
0.01	0.904
0.10	0.796
1	0.809
2	1.01
5	2.38
10	10.44
12	17.25

Because activities of single ions cannot be measured, these mean values are the closest we can get to $\{H^+\}$ in solutions of strong acids.

Activity is a practical consideration when dealing with strong mineral acids which are available at concentrations of 10 M or greater.

In a 12 M solution of hydrochloric acid, for example, the mean ionic activity coefficient is 17.25. This means that under these conditions with [H_3O^+] = 12 M, the *activity* {H_3O^+} = 12 × 17.25 = 207, corresponding to a pH of about –2.3, instead of –1.1 as might be predicted if concentrations were being used.

These very high activity coefficients also explain another phenomenon: why you can detect the odor of HCl(g) over a concentrated hydrochloric acid solution even though this acid is supposedly 100% dissociated. It turns out that the activity {HCl} (which represents the "escaping tendency" of HCl(g) from the solution) is almost 49,000 for a 10 M solution! The source of this gas is best described as the result of ion pairing.

Similarly, in a solution prepared by adding 0.5 mole of the very strong acid $HClO_4$ to sufficient water to make the volume 1 liter, freezing-point depression measurements indicate that the concentrations of hydronium and perchlorate ions are only about 0.4 M. This does not mean that the acid is only 80% dissociated; there is no evidence of $HClO_4$ molecules in the solution. What has happened is that about 20% of the $\mathrm{H_3O^+}$ and ClO_4^- ions have formed ion-pair complexes in which the oppositely-charged species are loosely bound by electrostatic forces (Figure 13.2.4).

Ion-pairing reduces effective dissociation at high concentrations. If you have worked with concentrated hydrochloric acid in the lab, you will have noticed that the choking odor of hydrogen chloride gas is very apparent. How can this happen if this strong acid is really "100 percent dissociated" as all strong acids are said to be?

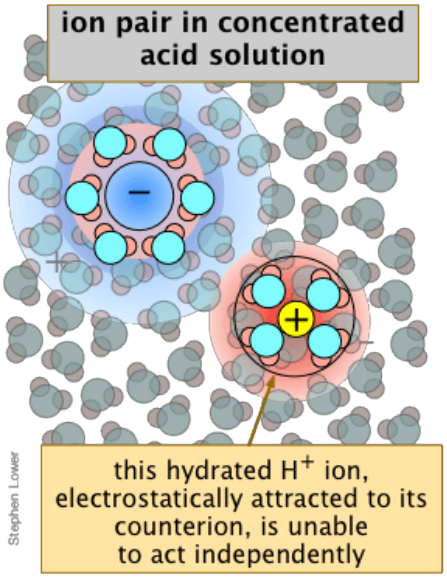


Figure 13.2.4:

At very high concentrations, too few H_2O molecules are available to completely fill the extended hydration shells that normally help keep the ions apart, reducing the fraction of "free" H_3O^+ ions capable of acting independently. Under these conditions, the term "dissociation" begins to lose its meaning. Although the concentration of HCl(aq) is never very high, its own activity coefficient can be as great as 2000 (Table 13.2.2), which means that its escaping tendency from the solution is extremely high, so that the presence of even a tiny amount is very noticeable.

Example 13.2.2

Estimate the pH of a 10.0 M solution of hydrochloric acid in which the mean ionic activity coefficient - is 10.4.

Solution

$$pH = - \log \{H^+\} \approx - \log (10.4 \times 10.0) = - \log 104 = \mathbf{- 2.0}$$

Compare this result with what you would get by using – log[H^+]

Systematic treatment of strong acids

In this section, we will derive expressions that relate the pH of solution of a strong acid to its concentration in a solution of pure water. We will use hydrochloric acid as an example. When HCl gas is dissolved in water, the resulting solution contains the ions H_3O^+, OH^-, and Cl–, However, except in very concentrated solutions, the concentration of HCl is negligible; for all practical purposes, molecules of “hydrochloric acid”, HCl, do not exist in dilute aqueous solutions.

To specify the concentrations of the three species present in an aqueous solution of HCl, we need three independent relations between them. These relations are obtained by observing that certain conditions must always be true in any solution of HCl. These are:

1. The *autoprotolysis* equilibrium of water must always be satisfied:

$$[H_3O^+][OH^-] = K_w \tag{13.2.5}$$

2. For any acid-base system, one can write a *mass balance* equation that relates the concentrations of the various dissociation products of the substance to its “nominal concentration”, which we designate here as C_a. For a solution of HCl, this equation would be

$$[HCl] + [Cl^-] = C_a \tag{13.2.6}$$

However, since HCl is a strong acid and therefore no "HCl" exists in the solution, we can neglect the first term, so the mass balance equation becomes simply

$$[Cl^-] = C_a \tag{13.2.7}$$

3. In any ionic solution, the sum of the positive and negative electric charges must be zero; in other words, all solutions are electrically neutral. This is known as the *electroneutrality principle*.

$$[H_3O^+] = [OH^-] + [Cl^-] \tag{13.2.8}$$

The next step is to combine these three equations into a single expression that relates the hydronium ion concentration to C_a. This is best done by starting with an equation that relates several quantities, such as Equation 13.2.8, and substituting the terms that we want to eliminate. Thus we can get rid of the [Cl–] term by substituting Equation 13.2.7 into Equation 13.2.8 :

$$[H_3O^+] = [OH^-] + C_a \tag{13.2.9}$$

The [OH^-] term can be eliminated by use of Equation 13.2.7:

$$[H_3O^+] = C_a + \frac{K_w}{[H_3O^+]} \tag{13.2.10}$$

This equation tells us that the hydronium ion concentration will be the same as the nominal concentration of a strong acid as long as the solution is not very dilute. Notice that Equation 13.2.10 is a quadratic equation. Recalling that $K_w = 10^{-14}$, it is apparent that the final term of the above equation will ordinarily be very small in comparison to the other terms, so it can be ordinarily be dropped, yielding the simple relation

$$[H_3O^+] \approx C_a \tag{13.2.11}$$

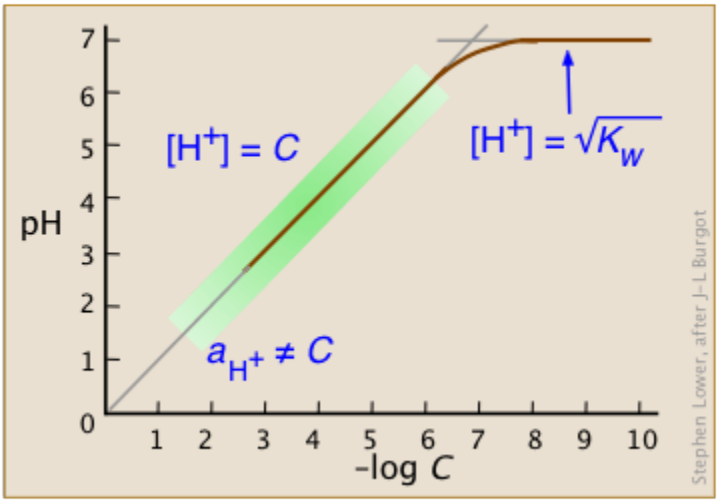


Figure 13.2.5

Only in extremely dilute solutions, around 10^{-6} *M* or below (where the plot curves), does this approximation become untenable. However, even then, the effect is tiny. After all, the hydronium ion concentration in a solution of a strong acid can never fall below 10^{-7} *M*; no amount of dilution can make the solution alkaline! So for almost all practical purposes, Equaton 13.2.11 is all you will ever need for a solution of a strong acid.

This page titled 13.2: Strong Monoprotic Acids and Bases is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13.3: Finding the pH of weak Acids, Bases, and Salts

The numerical examples and the images in this section have some problems (source text with comments).

Learning Objectives

Make sure you thoroughly understand the following essential concepts that have been presented above.

- Explain the difference between the terms C_a *and [HA]* as they relate to an aqueous solution of the acid HA.
- Define *degree of dissociation* and sketch a plot showing how the values of $alpha$ for a conjugate pair HA and A^- relate to each other and to the pK_a.
- Derive *the relationship between* acid concentration and dissociation constant to the hydrogen ion concentration in a solution of a weak acid.
- Re-write the above relation in *polynomial form*.
- *Derive the approximation* that is often used to estimate the pH of a solution of a weak acid in water.
- *Calculate the pH* of a solution of a weak monoprotic weak acid or base, employing the "*five-percent rule*" to determine if the approximation 2-4 is justified.
- Predict whether an aqueous *solution of a salt* will be acidic or alkaline, and explain why by writing an appropriate equation.

Most acids are weak; there are hundreds of thousands of them, whereas there are fewer than a dozen strong acids. We can treat weak acid solutions in much the same general way as we did for strong acids. The only difference is that we must now take into account the incomplete "dissociation"of the acid. We will start with the simple case of the pure acid in water, and then go from there to the more general one in which salts of the acid are present. The latter mixtures are known as buffer solutions and are extremely important in chemistry, physiology, industry and in the environment. To keep our notation as simple as possible, we will refer to “hydrogen ions” and [H^+] for brevity, and, wherever it is practical to do so, will assume that the acid HA "ionizes" or "dissociates" into H^+ and its conjugate base A–.

1 Aqueous solutions of weak acids or bases

A weak acid (represented here as HA) is one in which the reaction

$$HA \rightleftharpoons A^- + H^+ \tag{13.3.1}$$

is incomplete. This means that if we add 1 mole of the pure acid HA to water and make the total volume 1 L, the equilibrium concentration of the conjugate base A^- will be smaller (often *much* smaller) than 1 *M*/L, while that of undissociated HA will be only slightly less than 1 *M*/L. Equation 13.3.1 tells us that dissociation of a weak acid HA in pure water yields identical concentrations of its conjugate species. Let us represent these concentrations by *x*. Then, in our "1 *M* " solution, the concentration of each species is as shown here:

$$\underset{1-x}{HA} \rightarrow \underset{x}{A^-} + \underset{x}{H^+} \quad \leftarrow \text{moles per liter at equilibrium} \tag{1-2}$$

When dealing with problems involving acids or bases, bear in mind that when we speak of "the concentration", we usually mean the *nominal* or *analytical* concentration which is commonly denoted by C_a. For example, for a solution made by combining 0.10 mol of pure formic acid HCOOH with sufficient water to make up a volume of 1.0 L, C_a = 0.10 *M*. However, we know that the concentration of the actual species [HCOOH] will be smaller the 0.10 M because some it ends up as the formate ion $HCOO^-$. It will, of course, always be the case that the sum

$$[HCOOH] + [HCOO^-] = C_a$$

For the general case of an acid HA, we can write a *mass balance equation*

$$C_a = [HA] + [A^-] \tag{13.3.2}$$

which reminds us the "A" part of the acid must always be somewhere!

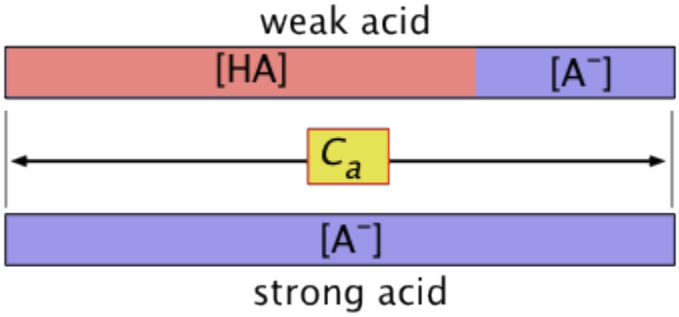


Figure 13.3.1: For a *strong acid* such as hydrochloric, its total dissociation means that [HCl] = 0, so the mass balance relationship in Equation 13.3.2 reduces to the trivial expression C_a = [Cl-]. Any acid for which [HA] > 0 is by definition a weak acid.

Similarly, for a base B we can write

$$C_b = [B] + [HB^+] \tag{13.3.3}$$

Equilibrium concentrations of the acid and its conjugate base

According to the above equations, the equilibrium concentrations of A^- and H^+ will be identical (as long as the acid is not so weak or dilute that we can neglect the small quantity of H^+ contributed by the autoprotolysis of H_2O). Let us represent these concentrations by *x*. Then, in a solution containing 1 *M*/L of a weak acid, the concentration of each species is as shown here:

$$\underset{1-x}{HA} \rightarrow \underset{x}{A^-} + \underset{x}{H^+} \quad \leftarrow \text{moles per liter at equilibrium} \tag{1-5}$$

Substituting these values into the equilibrium expression for this reaction, we obtain

$$\frac{[A^-][H^+]}{[HA]} = \frac{x^2}{1-x} \tag{13.3.4}$$

In order to predict the pH of this solution, we must solve for *x*. The presence of terms in both *x* and x^2 here tells us that this is a quadratic equation. In most practical cases in which K_a is 10^{-4} or smaller, we can assume that *x* is much smaller than 1 *M*, allowing us to make the simplifying approximation

$$(1–x) \approx 1 \tag{13.3.5}$$

so that

$$x^2 \approx K_a$$

and thus

$$x = \sqrt{K_a} \tag{13.3.6}$$

This approximation will not generally be valid when the acid is **very weak** or **very dilute.**

Solutions of arbitrary concentration

The above development was for a solution made by taking 1 mole of acid and adding sufficient water to make its volume 1.0 L. In

such a solution, the nominal concentration of the acid, denoted by C_a, is 1 *M*.

We can easily generalize this to solutions in which C_a has any value:

$$\underset{C_a - x}{HA} \rightarrow \underset{x}{A^-} + \underset{x}{H^+} \quad \leftarrow \text{moles per liter at equilibrium} \tag{1-9}$$

The above relation is known as a "mass balance on A". It expresses the simple fact that the "A" part of the acid must always be *somewhere* — either attached to the hydrogen, or in the form of the hydrated anion A^-.

The corresponding equilibrium expression is

$$\frac{[H^+]\,[A^-]}{[HA]} = \frac{x^2}{(C_a - x)} = K_a \tag{13.3.7}$$

and the approximations (when justified) Equation ??? and ??? become

$$(C_a – x) \approx C_a \tag{13.3.8}$$

$$x \approx \sqrt{C_a K_a} \tag{13.3.9}$$

Example 13.3.1: Aproximate pH of an Acetic Acid Solution

Estimate the pH of a 0.20 M solution of acetic acid, $K_a = 1.8 \times 10^{-5}$.

Solution

For brevity, we will represent acetic acid CH_3COOH as HAc, and the acetate ion by Ac^-. As before, we set x = $[H^+]$ = $[Ac^-]$, neglecting the tiny quantity of H^+ that comes from the dissociation of water.

$$\frac{[H^+]\,[Ac^-]}{[HAc]} = \frac{x^2}{(1 - x)}$$

Substitution into the equilibrium expression yields

$$\frac{x^2}{(0.20 - x)} = 1.8 \times 10^{-5}$$

The rather small value of K_a suggests that we can drop the x term in the denominator, so that

$$\frac{x^2}{0.20} \approx 1.8 \times 10^{-5}$$

or

$$x \approx \sqrt{(0.20)(1.8 \times 10^{-5})} = 1.9 \times 10^{-3}\ M$$

The pH of the solution is

$$-\log(1.9 \times 10^{-3}) = 2.7$$

Degree of dissociation

Even though we know that the process HA → H^+ + A^- does not correctly describe the transfer of a proton to H_2O, chemists still find it convenient to use the term "ionization" or "dissociation". The "degree of dissociation" (denoted by α of a weak acid is just

the fraction

$$\alpha = \frac{[\mathrm{A}^-]}{C_a} \tag{13.3.10}$$

which is often expressed as a per cent ($\alpha \times 100$).

Degree of dissociation depends on the concentration

It's important to understand that whereas K_a for a given acid is essentially a constant, α *will depend on the concentration* of the acid. Note that these equations are also valid for weak bases if K_b and C_b are used in place of K_a and C_a.

This can be shown by substituting Eq 5 into the expression for K_a:

$$K_a = \frac{C_a \alpha^2}{1 - \alpha} \quad \text{(1-14)}$$

Solving this for α results in a quadratic equation, but if the acid is sufficiently weak that we can say (1 – α) ≈ 1, the above relation becomes

$$\alpha \approx \sqrt{K_a / C_a} \quad \text{(1-15)}$$

the amount of HA that dissociates varies inversely with the square root of the concentration; as C_a approaches zero, α approaches unity and [HA] approaches C_a.

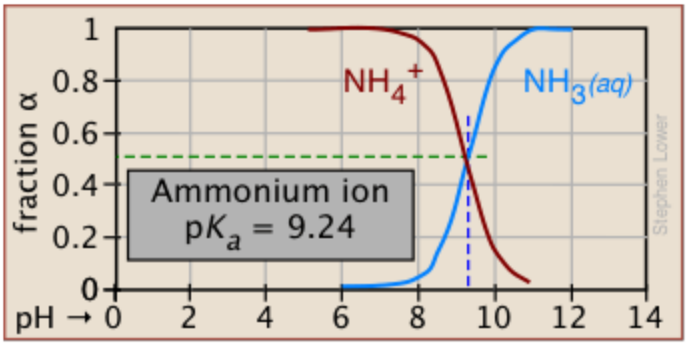
This principle is an instance of the *Ostwald dilution law* which relates the dissociation constant of a weak electrolyte (in this case, a weak acid), its degree of dissociation, and the concentration.

A common, but incorrect explanation of this law in terms of the Le Chatelier principle states that dilution increases the concentration of water in the equation HA + H_2O → H_3O^+ + A^-, thereby causing this equilibrium to shift to the right. The error here is that [H_2O] in most aqueous solutions is so large (55.5 M) that it can be considered constant; this is the reason the [H_2O] term does not appear in the expression for K_a. Another common explanation is that dilution reduces [H_3O^+] and [A^-], thus shifting the dissociation process to the right. However, dilution similarly reduces [HA], which would shift the process to the left. In fact, these two processes compete, but the former has greater effect because two species are involved.

It is probably more satisfactory to avoid Le Chatelier-type arguments altogether, and regard the dilution law as an entropy effect, a consequence of the greater dispersal of thermal energy throughout the system. This energy is carried by the molecular units within the solution; dissociation of each HA unit produces two new particles which then go their own ways, thus spreading (or "diluting") the thermal energy more extensively and massively increasing the number of energetically-equivalent microscopic states, of which entropy is a measure. (More on this here)

Degree of dissociation depends on the pH

Plots of this kind are discussed in more detail in the next lesson in this set under the heading ionization fractions.



The Le Chatelier principle predicts that the extent of the reaction

$$\mathrm{HA} \longrightarrow \mathrm{H^+ + A^-}$$

will be affected by the hydrogen ion concentration, and thus by the pH. This is illustrated here for the ammonium ion. Notice that when the pH is the same as the pK_a, the concentrations of the acid- and base forms of the conjugate pair are identical.

Example 13.3.1: percent dissociation

A 0.75 M solution of an acid HA has a pH of 1.6. What is its percent dissociation?

Solution

The dissociation stoichiometry HA → H^+ + AB^- tells us the concentrations [H^+] and [A^-] will be identical. Thus [H^+] = $10^{-1.6}$ = 0.025 M = [A^-]. The dissociation fraction

$$\alpha = \frac{[\mathrm{A^-}]}{[\mathrm{HA}]} = \frac{0.025}{0.75} = 0.033$$

and thus the acid is **3.3%** dissociated at 0.75 M concentration.

Sometimes the percent dissociation is given, and K_a must be evaluated.

Example 13.3.3: Ka from degree of dissociation

A weak acid HA is 2 percent dissociated in a 1.00 M solution. Find the value of K_a.

Solution

The equilibrium concentration of HA will be 2% smaller than its nominal concentration, so [HA] = 0.98 M, [A^-] = [H^+] = 0.02 M. Substituting these values into the equilibrium expression gives

$$\frac{[\mathrm{H^+}]\,[\mathrm{A^-}]}{[\mathrm{HA}]} = \frac{.0004}{.98} = 4.01 \times 10^{-4}$$

"Concentration of the acid" and [HA] are not the same

When dealing with problems involving acids or bases, bear in mind that when we speak of "the concentration", we usually mean the *nominal* or *analytical* concentration which is commonly denoted by C_a.

So for a solution made by combining 0.10 mol of pure formic acid HCOOH with sufficient water to make up a volume of 1.0 L, C_a

= 0.10 *M*. However, we know that the concentration of the actual species [HCOOH] will be smaller the 0.10 M because some it ends up as the formate ion $HCOO^-$. However, it will always be the case that the sum

$[HCOOH] + [HCOO^-] = C_a$.

For the general case of an acid HA, we can write a *mass balance equation*

$C_a = [HA] + [A^-]$(1-16)

which reminds us the "A" part of the acid must always be somewhere!

Similarly, for a base B we can write

$C_b = [B] + [HB^+]$(1-17)

Degree of dissociation varies inversely with the concentration

If we represent the dissociation of a C_a *M* solution of a weak acid by

$$\underset{C_a - x}{HA} \rightarrow \underset{x}{A^-} + \underset{x}{H^+} \quad \leftarrow \text{moles per liter at equilibrium} \qquad (1\text{-}18)$$

then its dissociation constant is given by

$$K_a = \frac{x^2}{C_a - x} \qquad (1\text{-}19)$$

Because the C_a term is in the denominator here, we see that the amount of HA that dissociates varies inversely with the concentration; as C_a approaches zero, [HA] approaches C_a.

If we represent the fraction of the acid that is dissociated as

$$\alpha = \frac{[A^-]}{C_a} \qquad (1\text{-}20)$$

then Eq 8 becomes

$$K_a = \frac{C_a \alpha^2}{1 - \alpha} \qquad (1\text{-}21)$$

If the acid is sufficiently weak that x does not exceed 5% of C_a,

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

the -term in the denominator can be dropped, yielding

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

$K_a \approx C_a$ ^{2} (1-22)

Note that the above equations are also valid for weak bases if Kb and C_b are used in place of K_a and C_a.

Example 13.3.1: effects of dilution

Compare the percent dissociation of 0.10 M and .0010 M solutions of boric acid ($K_a = 3.8 \times 10^{-10}$).

Solution

Boric acid is sufficiently weak that we can use the approximation of Eq 1-22 to calculate a:

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

= (5.8E–10 / .1)^{½} = 7.5E-5; multiply by 100 to get .0075 % diss. For the more dilute acid, a similar calculation yields 7.6E–4, or 0.76%.

2 Carrying out acid-base calculations

In Problem Example 1, we calculated the pH of a monoprotic acid solution, making use of an approximation in order to avoid the need to solve a quadratic equation. This raises the question: how "exact" must calculations of pH be? It turns out that the relation between pH and the nominal concentration of an acid, base, or salt (and especially arbitrary mixtures of these) can become quite complicated, requiring the solution of sets of simultaneous equations.

However, for almost all practical applications, one can make some approximations that simplify the math without detracting significantly from the accuracy of the results.

Equilibrium constants are rarely exactly known

As we pointed out in the preceding lesson, the "effective" value of an equilibrium constant (the *activity*) will generally be different from the value given in tables in all but the most dilute ionic solutions. Even if the acid or base itself is dilute, the presence of other "spectator" ions such as Na^+ at concentrations much in excess of 0.001 M can introduce error.

The usual advice is to consider K_a values to be accurate to ±5 percent at best, and even more uncertain when total ionic concentrations exceed 0.1 M. As a consequence of this uncertainty, there is generally little practical reason to express the results of a pH calculation to more than two significant digits.

Finding the pH of a solution of a weak monoprotic acid

This is by far the most common type of problem you will encounter in a first-year Chemistry class. You are given the concentration of the acid, expressed as C_a moles/L, and are asked to find the pH of the solution. The very important first step is to make sure you understand the problem by writing down the equation expressing the concentrations of each species in terms of a single unknown, which we represent here by *x*:

$$\underset{C_a - x}{HA} \rightarrow \underset{x}{A^-} + \underset{x}{H^+} \quad \leftarrow \text{moles per liter at equilibrium} \qquad (2\text{-}1)$$

Substituting these values into the expression for K_a, we obtain

$$\frac{[H^+]\,[A^-]}{[HA]} = \frac{x^2}{(C_a - x)} = K_a \qquad (2\text{-}2)$$

Don't bother to memorize these equations! If you understand the concept of mass balance on "A" expressed in (2-1), and can

write the expression for K_a, you just substitute the x's into the latter, and you're off! If you feel the need to memorize stuff you don't need, it is likely that you don't really understand the material — and that should be a *real* worry!

In order to predict the pH of this solution, we must first find $[H^+]$, that is, x. The presence of terms in both x and x^2 here tells us that this is a quadratic equation. This can be rearranged into $x^2 = K_a\ (1 - x)$ which, when written in standard polynomial form, becomes the quadratic

$$[\mathrm{H}^+]^2 - C_a[H^+] - K_w = 0 \tag{13.3.11}$$

However, don't panic! As we will explain farther on, in most practical cases we can make some simplifying approximations which eliminate the need to solve a quadratic. And when, as occasionally happens, a quadratic is unavoidable, we will show you some relatively painless ways of dealing with it.

How to deal with Quadratic Equations

What you do will depend on what tools you have available. If you are only armed with a simple calculator, then there is always the venerable quadratic formula that you may have learned about in high school, but if at all possible, you should avoid it: its direct use in the present context is somewhat laborious and susceptible to error.

Use of the standard quadratic formula on a computer or programmable calculator can lead to weird results!

The reason for this is that if $b^2 >> |4ac|$, one of the roots will require the subtraction of two terms whose values are very close; this can lead to considerable error when carried out by software that has finite precision.

One can get around this by computing the quantity

$$Q = -\frac{b + \pm(b)\sqrt{b^2 - 4ac}}{2}$$

from which the roots are $x_1 = Q/a$ and $x_2 = c/Q$. (See any textbook on numerical computing for more on this and other metnods.)

However, who want's to bother with this stuff in order to solve typical chemistry problems? Better to avoid quadratics altogether if at all possible!

Remember: there are always *two* values of x (two *roots*) that satisfy a quadratic equation. For all acid-base equilibrium calculations that are properly set up, these roots will be *real*, and only one will be positive; this is the one you take as the answer.

Approximations, judiciously applied, simplify the math

We have already encountered two of these approximations in the examples of the preceding section:

1. In all but the most dilute solutions, we can assume that the dissociation of the acid HA is the sole source of H^+, with the contribution due to water autoprotolysis being negligible.
2. We were able to simplify the equilibrium expressions by assuming that the x-term, representing the quantity of acid dissociated, is so small compared to the nominal concentration of the acid C_a that it can be neglected. Thus in Problem Example 1, the term in the denominator that has the form $(0.1 - x)$, representing the equilibrium concentration of the undissociated acid, is replaced by 0.1.

Most people working in the field of practical chemistry will never encounter situations in which the first of these approximations is invalid. This is not the case, however, for the second one.

Should I drop the *x*, or forge ahead with the quadratic form?

If the acid is fairly concentrated (usually with $C_a > 10^{-3}$ M) and sufficiently weak that most of it remains in its protonated form HA, then the concentration of H^+ it produces may be sufficiently small that the expression for K_a reduces to

$K_a \approx [H^+]^2 / C_a$

so that

$[H^+] \approx (K_a C_a)^{½}$(2-4)

This can be a great convenience because it avoids the need to solve a quadratic equation. However, it also exposes you to the danger that this approximation may not be justified.

The usual advice is that if this first approximation of *x* exceeds 5 percent of the value it is being subtracted from (0.10 in the present case), then the approximation is *not* justified. We will call this the "five percent rule".

Example 13.3.5:

Problem Example 5 - pH and degree of dissociation

a. Estimate the pH of a 0.20 M solution of acetic acid, $K_a = 1.8 \times 10^{-5}$.
b. What percentage of the acid is dissociated?

Solution

For brevity, we will represent acetic acid CH_3COOH as HAc, and the acetate ion by Ac^-. As before, we set x = $[H^+]$ = $[Ac^-]$, neglecting the tiny quantity of H^+ that comes from the dissociation of water.

Substitution into the equilibrium expression yields

\[K_a = x^2/(0.20 - x)

Can we simplify this by applying the approximation 0.20 – x ≈ 0.20 ? Looking at the number on the right side of this equation, we note that it is quite small. This means the left side must be equally small, which requires that the denominator be fairly large, so we can probably get away with dropping x. Doing so yields

$(x^2 / 0.20) = 1.8E\text{-}5$ or $x = (0.20 \times 1.8E{-}5)^{½} = 1.9E\text{-}3$ M

The "5 per cent rule" requires that the above result be no greater than 5% of 0.20, or 0.010. Because 0.0019 meets this condition, we can set
$x = [H^+] \approx 1.9 \times 10^{-3}$ M, and the pH will be $-\log (1.9 \times 10^{-3})$ = **2.7**

b) Percent dissociation: 100% × $(1.9 \times 10^{-3}$ M) / (0.20 M) = **0.95%**

This plot shows the combinations of K_a and C_a that generally yield satisfactory results with the approximation of Eq 4.

Weak bases are treated in an exactly analogous way:

Example 13.3.1: Weak base

Methylamine CH_3NH_2 is a gas whose odor is noticed around decaying fish. A solution of CH_3NH_2 in water acts as a weak base. A 0.10 M solution of this amine in water is found to be 6.4% ionized. Use this information to find \Kb and pK_b for methylamine.

Solution: When methylamine "ionizes", it takes up a proton from water, forming the methylaminium ion:

$CH_3NH_2 + H_2O \rightarrow CH_3NH_3^+ + OH^-$

Let x = $[CH_3NH_3^+]$ = $[OH^-]$ = .064 × 0.10 = 0.0064

$[CH_3NH_2]$ = (0.10 – .064) = 0.094

Substitute these values into equilibrium expression for \Kb:

$$K_b = \frac{0.0064^2}{0.094}$$

To make sure we can stop here, we note that (3.6E4 / .01) = .036; this is smaller than .05, so we pass the 5% rule and can use the approximation and drop the
x-term in the denominator.

pK_b = – log \Kb = – log (4.4×10^{-10}) = **3.36**

However, one does not always get off so easily!

Example 13.3.6: pH of a chloric acid solution

With a K_a of 0.010, $HClO_2$ is one of the "stronger" weak acids, thanks to the two oxygen atoms whose electronegativity withdraws some negative charge from the chlorine atom, making it easier for the hydrogen to depart as a proton.

Find the pH of a 0.10 M solution of chloric acid in pure water.

$$0.010 = \frac{x^2}{0.10 - x}$$

(i)

The approximation 0.10 – x ≈ 0.10 gives us

$x \approx (K_a\ C_a)^{½}$ = $(0.010 \times 0.10)^{½}$ = $(.001)^{½}$ = .032 (ii)

The difficulty, in this case, arises from the numerical value of K_a differing from the nominal concentration 0.10 M by only a factor of 10. As a result,
x / C_a = .032 / 0.10 = 0.32 which clearly exceeds the 5% limit; we have no choice but to face the full monte of the quadratic solution. (see Problem Example 8 below).

Successive approximations will get you there with minimal math

In the *method of successive approximations*, you start with the value of $[H^+]$ (that is, x) you calculated according to (2-4), which becomes the first approximation. You then substitute this into (2-2), which you solve to get a second approximation. This cycle is repeated until differences between successive answers become small enough to ignore.

Example 13.3.1: Method of successive approximations

Estimate the pH of a 0.10 M aqueous solution of $HClO_2$, K_a = 0.010, using the method of successive approximations.

Solution: The equilibrium condition is

$$K_a = \frac{x \cdot x}{c_a - x}$$

We solve this for x (neglecting the x in the denominator), resulting in the first approximation x_1, and then successively plug each result into the previous equation, yielding approximations x_2 and x_3:

$$x_1 = \sqrt{K_a \cdot c_a} = 0.032$$

$$x_2 = \sqrt{K_a \cdot (c_a - x_1)} = 0.026$$

$$x_3 = \sqrt{K_a \cdot (c_a - x_2)} = 0.027$$

The last two approximations x_2 and x_3 are within 5% of each other.

Note that if we had used x_1 as the answer, the error would have been 18%.

(An exact numeric solution yields the roots 0.027016 and –0.037016)

Use a graphic calculator or computer to find the positive root

The real roots of a polynomial equation can be found simply by plotting its value as a function of one of the variables it contains.

In this example, the pH of a 10^{-6} M solution of hypochlorous acid (HOCl, K_a = 2.9E–8) was found by plotting the value of $y = ax^2 + bx + c$, whose roots are the two values of x that correspond to y = 0.

This method generally requires a bit of informed trial-and-error to make the locations of the roots visible within the scale of the axes.

Be lazy, and use an on-line quadratic equation solver

If you google "quadratic equation solver", you will find numerous on-line sites that offer quick-and-easy "fill-in-the-blanks" solutions.

Unfortunately, few of these will be useful for acid-base problems involving numbers that must be expressed in "E-notation" (*e.g.*, 2.7E-11.) Of those that do, the one at the MathIsFun site is highly recommended; others can be found here and at the Quad2Deg site.

If you can access a quad equation solver on your personal electronic device or through the Internet, this is quick and painless. All you need to do is write the equation in polynomial form $ax^2 + bx + c = 0$, insert values for *a, b*, and *c*, and away you go!

This is so easy, that many people prefer to avoid the "5% test" altogether, and go straight to an exact solution. *However, make sure can do the 5%-thing for exams where Internet-accessible devices are not permitted!*

Example 13.3.1: chloric acid, again

Estimate the pH of a 0.10 M aqueous solution of $HClO_2$, K_a = 0.010.

Solution

The reaction equation $HClO_2 \rightarrow H^+ + ClO_2^-$ defines the equilibrium expression

$$0.010 = \frac{x \cdot x}{0.10 - x}$$

Multiplying by the denominator yields

$x^2 = 0.010 \times (0.10 - x) = .0010 - .01\,x$ which we arrange into standard polynomial form:

$x^2 + 0.01\,x - 0.0010 = 0$

Entering the coefficients {1 .01 –.001} into an online quad solver yields the roots
.027 and –.037. Taking the positive one, we have $[H^+]$ = .027 M;
the solution pH is – log .027 = **1.6**.

Note: a common error is to forget to enter the minus sign for the last term; try doing this and watch the program blow up!

Avoid math altogether and make a log-C vs pH plot

This is not only simple to do (all you need is a scrap of paper and a straightedge), but it will give you far more insight into what's going on, especially in polyprotic systems. All explained in Section 3 of the next lesson.

Solutions of salts

Most salts do *not* form pH-neutral solutions

Salts such as sodium chloride that can be made by combining a strong acid (HCl) with a strong base (NaOH, KOH) have a neutral pH, but these are exceptions to the general rule that solutions of most salts are mildly acidic or alkaline.

Salts of a strong base and a weak acid yield alkaline solutions.

"Hydro-lysis" literally means "water splitting", as exemplified by the reaction $A^- + H_2O \rightarrow HA + OH^-$. The term describes what was believed to happen prior to the development of the Brønsted-Lowry proton transfer model.

This important property has historically been known as *hydrolysis* — a term still used by chemists.

Some examples:

Potassium Cyanide

KCN can be thought of the salt made by combining the strong base KOH with the weak acid HCN: $K^+ + OH^- \rightarrow$ KOH. When solid KCN dissolves in water, this process is reversed, yielding a solution of the two ions. K^+, being a "strong ion", does not react with water. However, the "weak ion" CN^-, being the conjugate base of a weak acid, will have a tendency to abstract protons from water: $CN^- + H_2O \rightarrow HCN + OH^-$, causing the solution to remain slightly alkaline.

Sodium bicarbonate

$NaHCO_3$ (more properly known as sodium hydrogen carbonate) dissolves in water to yield a solution of the hydrogen carbonate ion HCO_3^-. This ion is an *ampholyte* — that is, it is both the conjugate base of the weak carbonic acid H_2CO_3, as well as the conjugate acid of the carbonate ion CO_3^{2-}:

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

The HCO_3^- ion is therefore *amphiprotic*: it can both accept and donate protons, so both processes take place:

$$\begin{array}{lll} \text{as base} & HCO_3^- + H_2O \rightarrow H_2CO_3 + OH^- & K_b = 10^{-3.7} \\ \text{as acid} & HCO_3^- + H_2O \rightarrow CO_3^{2-} + H_3O^+ & K_a = 10^{-10.3} \\ \hline & 2\,HCO_3^- + 2\,H_2O \rightarrow H_2CO_3 + CO_3^{2-} + \underbrace{OH^- + H_3O^+}_{2\,H_2O} & \\ \text{net reaction} & \mathbf{2\,HCO_3^- \rightarrow H_2CO_3 + CO_3^{2-}} & \end{array}$$

However, if we compare the K_a and K_b of HCO_3^-, it is apparent that its basic nature wins out, so a solution of $NaHCO_3$ will be slightly alkaline. (The value of pK_b is found by recalling that $K_a + K_b = 14$.)

Sodium acetate

A solution of CH_3COONa serves as a model for the strong-ion salt of any organic acid, since all such acids are weak: $CH_3COO^- + H_2O \rightarrow CH_3COOH + OH^-$

Salts of most cations (positive ions) give acidic solutions

The protons can either come from the cation itself (as with the ammonium ion NH_4^+), or from waters of hydration that are attached to a metallic ion.

This latter effect happens with virtually all salts of metals beyond Group I; it is especially noticeable for salts of transition ions such as hexaaquoiron(III) ("ferric ion"):

$Fe(H_2O)_6^{3+} \rightarrow Fe(H_2O)_5OH^{2+} + H^+$ (2-5)

This comes about because the positive field of the metal enhances the ability of H_2O molecules in the hydration shell to lose protons. In the case of the hexahyrated ion shown above, a whose succession of similar steps can occur, but for most practical purposes only the first step is significant.

Example 13.3.10: Aluminum chloride solution

Find the pH of a 0.15 M solution of aluminum chloride.

Solution

The aluminum ion exists in water as hexaaquoaluminum $Al(H_2O)_6^{3+}$, whose $pK_a = 4.9$, $K_a = 10^{-4.9} = 1.3E{-}5$. Setting $x = [H^+] = [Al(H_2O)_5OH^{2+}]$, the equilibrium expression is

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

Using the above approximation, we get
$x \approx (1.96E{-}6)^{½} = 1.4E{-}3$, corresponding to **pH = 2.8**.

Finally, we compute $x/C_a = 1.4E{-}3 \div 0.15 = .012$ confirming that we are within the "5% rule".

The only commonly-encountered salts in which the proton is donated by the cation itself are those of the *ammonium ion*:

\[\ce{NH_4^{+} → NH)3(aq) + H^{+}\lable{2-6} \nonumber \]

Example 13.3.11: Ammonium chloride solution

Calculate the pH of a 0.15 M solution of NH4Cl. The ammonium ion K_a is 5.5E−10.

Solution

According to Eq 6 above, we can set [NH3] = [H^+] = *x*, obtaining the same equilibrium expression as in the preceding problem.

Because K_a is quite small, we can safely use the approximation 0.15 - 1 ≈ .015, which yields pH = −log 0.90E−5 = **5.0**

Most salts of weak acids form alkaline solutions

As indicated in the example, such equilibria strongly favor the left side; the stronger the acid HA, the less alkaline the salt solution will be.

Because an ion derived from a weak acid such as HF is the conjugate base of that acid, it should not surprise you that a salt such as NaF forms an alkaline solution, even if the equilibrium greatly favors the left side::

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

F^- + H_2O HF + OH^-(2-7)

Example 13.3.1:

Solution of sodium fluoride

Find the pH of a 0.15 M solution of NaF. (HF K_a = 6.7E−4)

Solution: The reaction is F- + H_2O = HF + OH^-; because HF is a weak acid, the equilibrium strongly favors the right side. The usual approximation yields

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

However, on calculating x/C_a = .01 ÷ 0.15 = .07, we find that this does not meet the "5% rule" for the validity of the approximation. We therefore expand the equilibrium expression

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

into standard polynomial form x^2 + 6.7E−4 *x* − 1.0E−4 = 0 and enter the coefficients {1 6.7E−4 −.0001} into a quadratic solver. This yields the positive root *x* = 0.0099 which turns out to be sufficiently close to the approximation that we could have retained it after all.. perhaps 5% is a bit too restrictive for 2-significant digit calculations!

What about a salt of a weak acid *and* a weak base?

A salt of a weak acid gives an alkaline solution, while that of a weak base yields an acidic solution. What happens if we dissolve a salt of a weak acid and a weak base in water? Ah, this can get a bit tricky! Nevertheless, this situation arises very frequently in applications as diverse as physiological chemistry and geochemistry.

As an example of how one might approach such a problem, consider a solution of ammonium formate, which contains the ions NH_4^+ and HCOO-. Formic acid, the simplest organic acid, has a pK_a of 3.7; for NH_4^+, pK_a = 9.3.

Three equilibria involving these ions are possible here; in addition to the reactions of the ammonium and formate ions with water, we must also take into account their tendency to react with each other to form the parent neutral species:

$NH_4^+ \rightarrow NH_3 + H^+$	$K_1 = 10^{-9.3}$
$HCOO^- + H_2O \rightarrow HCOOH + OH^-$	$K_2 = (10^{-14}/10^{-3.7}) = 10^{-10.3}$
$NH_4^+ + HCOO^- \rightarrow NH_3 + HCOOH$	K_3

Inspection reveals that the last equation above is the sum of the first two,plus the reverse of the dissociation of water

$H^+ + OH^- \rightarrow H_2O$	$K_4 = 1/K_w$

The value of K_3 is therefore

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7} \quad (2\text{-}8)$$

A rigorous treatment of this system would require that we solve these equations simultaneously with the charge balance and the two mass balance equations. However, because K_3 is several orders of magnitude greater than K_1 or K_2, we can greatly simplify things by neglecting the other equilibria and considering only the reaction between the ammonium and formate ions.

Notice that the products of this reaction will tend to suppress the extent of the first two equilibria, reducing their importance even more than the relative values of the equilibrium constants would indicate.

Example 13.3.1:

ammonium formate

Estimate the pH of a 0.0100 M solution of ammonium formate in water.

Solution: From the stoichiometry of $HCOONH_4$,

$[NH_4^+] = [HCOO^-]$ and $[NH_3] = [HCOOH]$ (i)

then, from Eq 8 above,

$$K_3 = \frac{[NH_3]\,[HCOOH]}{[NH_4^+][HCOO^-]} = \frac{[HCOOH]^2}{[HCOO^-]^2} = \frac{K_w}{K_a\,K_b} \quad (ii)$$

in which K_b is the base constant of ammonia, $K_w/10^{-9.3}$.

From the formic acid dissociation equilibrium we have

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7} \quad (iii)$$

We now rewrite the expression for K_3

$$K_a = \frac{[HCOOH]^2}{[HCOO^-]^2} = \frac{[H^+]^2}{K_a{}^2} = \frac{K_w}{K_a K_b} \quad \text{(iv)}$$

which yields

$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7} \quad \text{(v)}$$

and thus the pH is **6.5**

What is interesting about this last example is that the pH of the solution is apparently independent of the concentration of the salt. If $K_a = K_b$, then this is always true and the solution will be neutral (neglecting activity effects in solutions of high ionic strength). Otherwise, it is only an approximation that remains valid as long as the salt concentration is substantially larger than the magnitude of either equilibrium constant. Clearly, the pH of any solution must approach that of pure water as the solution becomes more dilute.

Salts of analyte ions

Polyprotic acids form multiple anions; those that can themselves donate protons, and are thus amphiprotic, are called *analytes*. The most widely known of these is the *bicarbonate* (hydrogen carbonate) ion, HCO_3^-, which we commonly know in the form of its sodium salt $NaHCO_3$ as *baking soda*.

The other analyte series that is widely encountered, especially in biochemistry, is those derived from phosphoric acid:

$$\mathrm{H_3PO_4} \xrightarrow[10^{-2.12}]{K_1} \mathrm{H_2PO_4^-} \xrightarrow{10^{-7.2}} \mathrm{KPO_4}^{2-} \frac{K_3}{10^{-12.3}} \mathrm{PO_4^{3-}}$$

The solutions of analyte ions we most often need to deal with are the of "strong ions", usually Na^+, but sometimes those of Group 2 cations such as Ca^{2+}.

The exact treatment of these systems is generally rather complicated, but for the special cases in which the successive K_a's of the parent acid are separated by several orders of magnitude (as in the two systems illustrated above), a series of approximations reduces the problem to the simple expression

$$[\mathrm{H^+}] \approx \sqrt{K_1 K_2} \tag{13.3.12}$$

which, you will notice, as with the salt of a weak acid and a weak base discussed in the preceding subsection predicts that the pH is independent of the salt's concentration. This, of course, is a sure indication that this treatment is incomplete. Fortunately, however, it works reasonably well for most practical purposes, which commonly involve buffer solutions.

Exact treatment of solutions of weak acids and bases

When dealing with acid-base systems having very small K_a's, and/or solutions that are extremely dilute, it may be necessary to consider *all* the species present in the solution, including minor ones such as OH^-.

This is almost never required in first-year courses. However, for students in more advanced courses, this "comprehensive approach" (as it is often called) illustrates the important general methodology of dealing with more complex equilibrium problems. It also shows explicitly how making various approximations gradually simplifies the treatment of more complex systems.

Solutions of polyprotic acids

A diprotic acid H_2A can donate its protons in two steps:

$$H_2A \longrightarrow HA^- \longrightarrow HA^-$$

and similarly, for a tripotic acid H3A:

$$H_3A \longrightarrow H_2A^- \longrightarrow HA^{2-} \longrightarrow A^{3-}$$

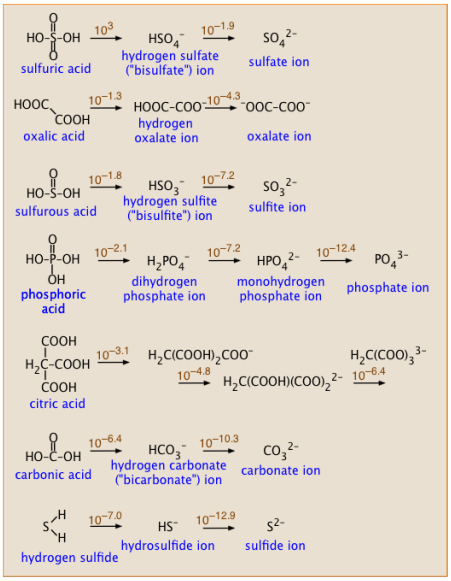
In general, we can expect K_a2 for the "second ionization" to be smaller than K_a1 for the first step because it is more difficult to remove a proton from a negatively charged species. The magnitude of this difference depends very much on whether the two removable protons are linked to the same atom, or to separate atoms spaced farther apart.

Some polyprotic acids you should know

These acids are listed in the order of decreasing K_{a1}. The numbers above the arrows show the successive K_a's of each acid.

- Notice how successive K_a's of each acid become smaller, and how their ratios relate to the structures of each acid.
- **Sulfuric acid** is the only *strong* polyprotic acid you are likely to encounter.
- **Sulfurous acid** has never been detected; what we refer to as H_2SO_3 is more properly represented as a hydrated form of sulfur dioxide: which dissociates into the hydrosulfite ion:

  $SO_2{\cdot}H_2O \rightarrow HSO_3^- + H^+$. The ion HSO_3^- exists only in solution; solid bisulfite salts are not known.
- Similarly, pure **carbonic acid** has never been isolated, but it does exist as a minority species in an aqueous solution of CO_2: $[CO_2(aq)] = 650[H_2CO_3]$. The formula H_2CO_3 ordinarily represents the combination $CO_2(aq)$ and "true" H_2CO_3. The latter is actually about a thousand times stronger than is indicated by the pK_a of 6.3, which is the weighted average of the equilibrium mixture.



$$[Gly^-] = \frac{10^{-9.8} \times 0.10}{[H^+]} = 10^{-4.7}$$

pH of a polyprotic acid (LindaHanson, 17 min)

Example 13.3.1: Comparison of two diprotic acids

Compare the successive pK_a values of sulfuric and oxalic acids (see their structures in the box, above right), and explain why they should be so different.

Solution

The two pK_a values of sulfuric acid differ by 3.0 – (–1.9) = 4.9, whereas for oxalic acid the difference is 1.3 – (–4.3) = 3.0. That's a difference of almost 100 between the two K_a's. Removal of a second proton from a molecule that already carries some negative charge is always expected to be less favorable energetically. In sulfuric acid, the two protons come from –OH groups connected to the same sulfur atom, so the negative charge that impedes loss of the second proton is more localized near the site of its removal. In oxalic acid, the two protons are removed from –OH groups attached to separate carbon atoms, so the negative charge of the mono-negative ions will exert less restraint on loss of the second proton.

Solutions of polyprotic acids in pure water

With the exception of sulfuric acid (and some other seldom-encountered strong diprotic acids), most polyprotic acids have sufficiently small K_{a1} values that their aqueous solutions contain significant concentrations of the free acid as well as of the various dissociated anions.

Thus for a typical diprotic acid H_2A, we must consider the three coupled equilibria

$$\mathrm{H_2A} \longrightarrow \mathrm{H^+ + HA^-} \tag{K1}$$

$$\mathrm{HA^-} \longrightarrow \mathrm{H^+ + HA^{2-}} \tag{K2}$$

$$\mathrm{H_2O} \longrightarrow \mathrm{H^+ + OH^-} \tag{Kw}$$

An exact treatment of such a system of four unknowns $[\mathrm{H_2A}]$, $[\mathrm{HA^-}]$, $[\mathrm{A^{2-}}]$ and $[\mathrm{H^+}]$ requires the solution of a quartic equation. If we include $[\mathrm{OH^-}]$, it's even worse!

In most practical cases, we can make some simplifying approximations:

- Unless the solution is extremely dilute or K_1 (and all the subsequent *K*'s) are extremely small, we can forget that any hydroxide ions are present at all.
- If K_1 is quite small, and the ratios of succeeding *K*'s are reasonably large, we may be able, without introducing too much error, to neglect the other *K*'s and treat the acid as monoprotic.

In addition to the three equilibria listed above, a solution of a polyprotic acid in pure water is subject to the following two conditions:

Material balance: although the distribution of species between the acid form $\mathrm{H_2A}$ and its base forms $\mathrm{HAB^-}$ and $\mathrm{A^{2-}}$ may vary, their sum (defined as the "total acid concentration" C_a is a constant for a particular solution:

$$C_a = [H_2A] + [HA^-] + [A^{2-}]$$

Charge balance: The solution may not possess a net electrical charge:

$$[\mathrm{H_3O^+}] = [\mathrm{OH^-}] + [\mathrm{HA^-}] + 2[\blacksquare]$$

Why do we multiply $[\mathrm{A^{2-}}]$ by 2? Two moles of $\mathrm{H_3O^+}$ are needed in order to balance out the charge of 1 mole of $\mathrm{A^{2-}}$.

Simplified treatment of polyprotic acid solutions

The calculations shown in this section are all you need for the polyprotic acid problems encountered in most first-year college chemistry courses. More advanced courses may require the more exact methods in Lesson 7.

Because the successive equilibrium constants for most of the weak polyprotic acids we ordinarily deal with diminish by several orders of magnitude, we can usually get away with considering only the first ionization step. To the extent that this is true, there is nothing really new to learn here. However, without getting into a lot of complicated arithmetic, we can often go farther and estimate the additional quantity of H^+ produced by the second ionization step.

Example 13.3.14: Solution of CO_2 in water

a) Calculate the pH of a 0.050 *M* solution of CO_2 in water.

For H_2CO_3, $K_1 = 10^{-6.4} = 4.5E{-}7$, $K_2 = 10^{-10.3} = 1.0E{-}14$.

b) Estimate the concentration of carbonate ion CO_3^{2-} in the solution.

Solution

a) Because K_1 and K_2 differ by almost four orders of magnitude, we will initially neglect the second dissociation step. Because this latter step produces only a tiny additional concentration of H^+, we can assume that $[H^+] = [HCO_3^-] = x$:

$$K_1 = \frac{[H^+]\,[HCO_3^-]}{[H_2CO_3]} = \frac{(x)^2}{.050 - x} = 4.5 \times 10^{-7} \quad \text{(i)}$$

Can we further simplify this expression by dropping the x in the denominator? Let's try:

$x = [0.05 \times (4.5E{-}7)]^{½} = 1.5E{-}4$.

Applying the "5-percent test", the quotient x/C_a must not exceed 0.05. Working this out yields (1.5E–4)/(.05) = .003, so we can avoid a quadratic.

$x = [H^+] \approx (K_aC_a)^{½} = [(4.5E{-}7) \times .01]^{½} = (.001)^{½} = 0.032$ M,

and the

pH = – log .032 = **1.5**.

b) We now wish to estimate $[CO_3^{2-}] \equiv x$.

Examining the second dissociation step, it is evident that this will consume x mol/L of HCO_3^-, and produce an equivalent amount of H^+ which adds to the quantity we calculated in *(a)*.

$$\underset{-x}{HCO_3^-} + H_2O \rightarrow \underset{x}{CO_3^{2-}} + \underset{+x}{H_3O^+} \qquad K_a = 4.7E{-}11 \quad \text{(ii)}$$

Owing to the very small value of K_2 compared to K_1, we can assume that the concentrations of HCO_3^- and H^+ produced in the first dissociation step will not be significantly altered in this step. This allows us to simplify the equilibrium constant expression and solve directly for $[CO_3^{2-}]$:

$$K_2 = \frac{[H^+]\,[CO_3^{2-}]}{[HCO_3^-]} = \frac{(x + 1.5E\text{-}4)\,x}{(1.5E\text{-}4) - x} \approx \frac{(1.5E\text{-}4)\,x}{(1.5E\text{-}4)} = x = 4.7E\text{-}11 \quad \text{(iii)}$$

It is of course no coincidence that this estimate of $[CO_3^{2-}]$ yields a value identical with K_2; this is entirely a consequence of the simplifying assumptions we have made. Nevertheless, as long as $K_2 << K_1$ and the solution is not highly dilute, the result will be sufficiently accurate for most purposes.

Sulfuric acid: A Special Case

Although this is a strong acid, it is also diprotic, and in its second dissociation step it acts as a weak acid. Because sulfuric acid is so widely employed both in the laboratory and industry, it is useful to examine the result of taking its second dissociation into consideration.

Example 13.3.1: Solution of H_2SO_4

Estimate the pH of a 0.010 M solution of H_2SO_4 with $K_1 = 10^3$ and $K_2 = 0.012$

Solution:

Because $K_1 > 1$, we can assume that a solution of this acid will be completely dissociated into H_3O^+ and bisulfite ions HSO_4^-. Thus the only equilibrium we need to consider is the dissociation of a 0.010 M solution of bisulfite ions.

$$HSO_4^- + H_2O \longrightarrow SO_4^{2-} + H_3O^+$$

$$K_2 = \frac{[H^+]\,[SO_4^{2-}]}{[HSO_4^-]} = \frac{(x)^2}{.010 - x} = 0.012$$

Setting $[H^+] = [SO_4^{2-}] = x$, and dropping x from the denominator, yields

$$x \approx \sqrt{(0.010)(0.012)} = \sqrt{1.2 \times 10^{-4}} = 0.0011$$

Applying the "five percent rule", we find that x / C_a = .0011/.01 = .11 which is far over the allowable error, so we must proceed with the quadratic form. Rewriting the equilibrium expression in polynomial form gives

$$x^2 + 0.022x - 1.2 \times 10^{-4} = 0$$

Inserting the coefficients {1 .022 .000012} into a quad-solver utility yields the roots 4.5×10^{-3} and -0.0027. Taking the positive root, we obtain

$$pH = -\log(0.0045) = 2.3$$

Thus the second "ionization" of H_2SO_4 has only reduced the pH of the solution by 0.1 unit from the value (2.0) it would theoretically have if only the first step were considered.

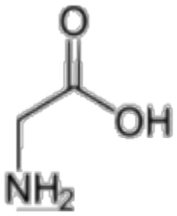
Amino acids and Zwitterions

Zwitterions: molecular hermaphrodites

Amino acids, the building blocks of proteins, contain amino groups $-NH_2$ that can accept protons, and carboxyl groups –COOH that can lose protons. Under certain conditions, these events can occur simultaneously, so that the resulting molecule becomes a “double ion” which goes by its German name *Zwitterion*.

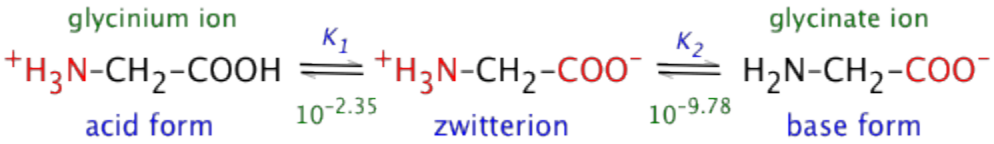
Glycine: the simplest amino acid

The simplest of the twenty natural amino acids that occur in proteins is *glycine* $H_2N–CH_2–COOH$, which we use as an example here.



glycine

Solutions of glycine are distributed between the acidic-, zwitterion-, and basic species:



(3-1)

Figure: Although the zwitterionic species is amphiprotic, it differs from a typical ampholyte such as HCO_3^- in that it is electrically neutral owing to the cancellation of the opposite electrical charges on the amino and carboxyl groups.

In the following development, we use the abbreviations H_2Gly^+ (glycinium), HGly (zwitterion), and Gly^- (glycinate) to denote the dissolved forms.

The two acidity constants are

$$K_1 = \frac{[\mathrm{H^+}]\,[\mathrm{HGly}\,]}{[\mathrm{H_2Gly^+}]} = 10^{-2.25} \tag{13.3.13}$$

and

$$K_2 = \frac{[\mathrm{H^+}]\,[\mathrm{Gly^-}]}{[\mathrm{HGly}]} = 10^{-9.78} \tag{13.3.14}$$

If glycine is dissolved in water, charge balance requires that

$$H_2Gly^+ + [H^+] \rightleftharpoons [Gly^-] + [OH^-] \tag{13.3.15}$$

Substituting the equilibrium constant expressions (including that for the autoprotolysis of water) into the above relation yields

$$\frac{[\mathrm{H^+}]\,[\mathrm{HGly}]}{K_1} + [\mathrm{H^+}] = \frac{K_2[\mathrm{HGly}]}{[\mathrm{H^+}]} + \frac{K_w}{[\mathrm{H^+}]} \tag{13.3.16}$$

Example 13.3.3: Glycine solution speciation

Calculate the pH and the concentrations of the various species in a 0.100 M solution of glycine.

Solution

Substitution into Eq 4 above yields

$$[\mathrm{H^+}] = \sqrt{\frac{10^{-9.8} \times 0.10 + 10^{-14}}{0.1/\left(10^{-2.3} + 1\right)}} = 10^{-6.1} \tag{i}$$

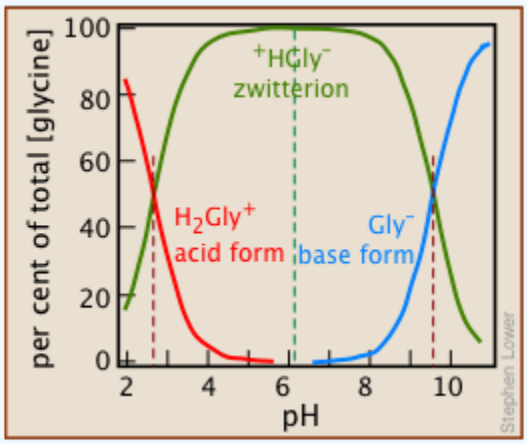
The concentrations of the acid and base forms are found from their respective equilibrium constant expressions (Eqs 2):

$$[\mathrm{H_2Gly^+}] = \frac{[\mathrm{H^+}] \times 0.10}{10^{-2.3}} = 10^{-4.7} \tag{ii}$$

$$[\mathrm{Gly^-}] = \frac{10^{-9.8} \times 0.10}{[\mathrm{H^+}]} = 10^{-4.7} \tag{iii}$$

The small concentrations of these singly-charged species in relation to $C_a = 0.10$ shows that the zwitterion is the only significant glycine species in the solution.

If Equations ii and iii in this Example are recalculated for a range of pH values, one can plot the concentrations of each species against pH for 0.10 M glycine in water:



This distribution diagram shows that the zwitterion is the predominant species between pH values corresponding to the two pKas given in Eq 3-1.

This page titled 13.3: Finding the pH of weak Acids, Bases, and Salts is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13.4: Conjugate Pairs and Buffers

Learning Objectives

Make sure you thoroughly understand the following essential concepts:

- State the values of C_a *and* C_b after a weak monoprotic acid or base has been partially neutralized with strong base or acid.
- Define the *ionization fraction*, and calculate the percent ionization, of a C_a *M* solution of a weak monoprotic acid.
- Sketch out a rough plot showing the *distribution of conjugate species* in a solution of a weak acid having a given value of K_a.
- Define a *buffer solution*, and explain how it works.
- Calculate the approximate pH of a solution of a weak monoprotic acid having a given pK_a that has been *partially neutralized* by addition of a strong base.
- Specify how to make a given volume of a buffer solution having a certain pH, starting with a weak acid and sodium hydroxide.
- Define the *buffer index*, and explain its significance and the conditions under which its value is maximized.
- Sketch out a *–log C vs pH (Sillén) plot* for a weak monoprotic acid/base system having a given nominal concentration and pK_a, and use it to estimate the pH of the solution.

We often tend to regard the pH as a quantity that is dependent on other variables such as the concentration and strength of an acid, base or salt. But in much of chemistry (and especially in biochemistry), we find it more useful to treat pH as the "master" variable that controls the relative concentrations of the acid- and base-forms of one or more sets of conjugate acid-base systems. In this lesson, we will explore this approach in some detail, showing its application to the very practical topics of buffer solutions, as well as the use of a simple graphical approach that will enable you to estimate the pH of a weak monoprotic or polyprotic acid or base without doing any arithmetic at all!

When the pH takes control

If we add 0.2 mol of sodium hydroxide to a solution containing 1.0 mol of a weak acid HA, then an equivalent number of moles of that acid will be converted into its base form A^-. The resulting solution will contain 0.2 mol of A^- and 0.8 mol of HA. Note that because we are discussing stoichiometry here, we are interested in quantities (moles) of reactants, not concentrations of reactants.

The important point to understand here is that we will end up with a “partly neutralized” solution in which both the acid and its conjugate base are present in significant amounts. Solutions of this kind are far more common than those of a pure acid or a pure base, and it is very important that you have a thorough understanding of them.

Example 13.4.1: Partly neutralized acid

To a solution containing 0.010 mole of acetic acid (HAc), we add 0.002 mole of sodium hydroxide. If the volume of the final solution is 100 ml, find the values of C_a, C_b, and the total system concentration C_t.

Solution

The added hydroxide ion, being a strong base, reacts completely with the acetic acid, leaving 0.010 – 0.002 = 0.008 mole of HAc and 0.002 mole of acetate ion Ac^-. The final concentrations are

$$C_a = \frac{0.008\ mol}{0.10\ L} = 0.08\ M$$

$$C_b = \frac{0.002\ mol}{0.10\ L} = 0.02M$$

$$C_t = \frac{0.010\ mol}{0.10\ L} = 0.10\ M.$$

Note that this solution would be indistinguishable from one prepared by combining C_a = 0.080 mole of acetic acid with C_b = 0.020 mole of sodium acetate and adjusting the volume to 100 ml.

Thus starting with a solution of a pure weak acid or weak base in water, we can add sufficient strong base or strong acid, respectively, to adjust the ratio of the conjugate species — that is, the ratios [HA]/[A^-] in the case of an acid, or [B]/[BH^+] for a base, to any value we want.

Ionization Fractions

To express the relative concentrations of the protonated and deprotonated forms of an acid-base system present in a solution, we could use the simple ratio [HA]/[A] (or its inverse), but this suffers from the drawback of yielding an indeterminate result when the concentration in the denominator is zero. For many purposes it is more convenient to use the *ionization fractions*

$$\alpha_0 = \frac{[HA]}{[HA] + A^-]} = \frac{[HA]}{C_a} \tag{13.4.1}$$

$$\alpha_1 = \frac{[A^-]}{[HA] + A^-]} = \frac{[A^-]}{C_a} \tag{13.4.2}$$

The fraction α_1 is also known as the *degree of dissociation* of the acid. By making appropriate substitutions using the relation

$$[H^+] = K_a \frac{[HA]}{[A^-]} \tag{13.4.3}$$

we can express the ionization fractions as functions of the pH:

$$\alpha_0 = \frac{[H^+]}{K_a + [H^+]} \tag{13.4.4}$$

$$\alpha_1 = \frac{K_a}{K_a + [H^+]} \tag{13.4.5}$$

Notice that the the values for both of these functions are close to zero or unity except within the pH range pK_a ± 1 (Figure 13.4.1).

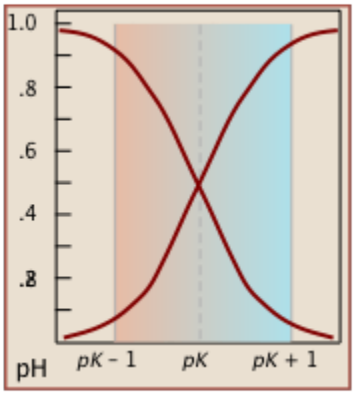
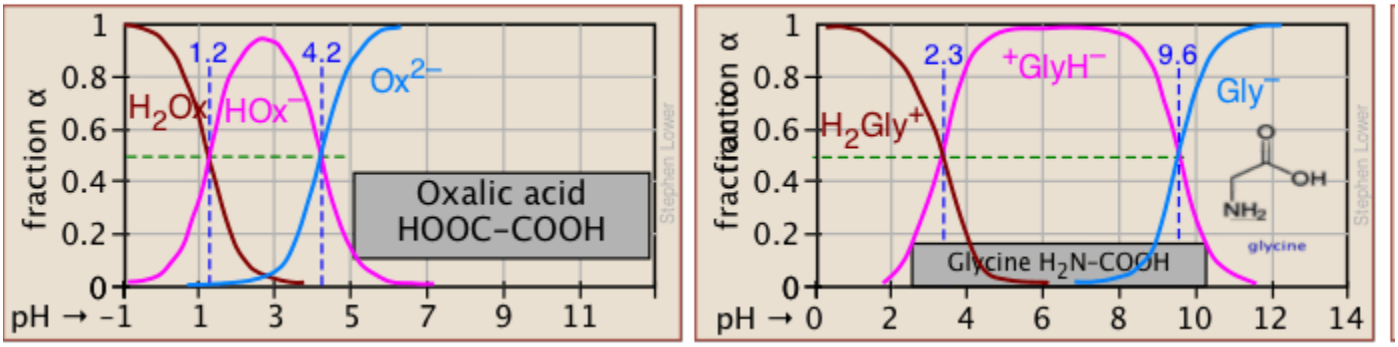


Figure 13.4.1: Partial Ionization of a weak acid

Plots of the α functions vs. pH for several systems are shown below. Notice the crossing points where [HA] = [A] when [H^+] = K_a that corresponds to unit value of the quotient in the Henderson-Hasselbalch approximation.



α functions vs. pH for several systems

The ionization fractions of a series of acids over a wide pH range can conveniently be summarized as shown below.

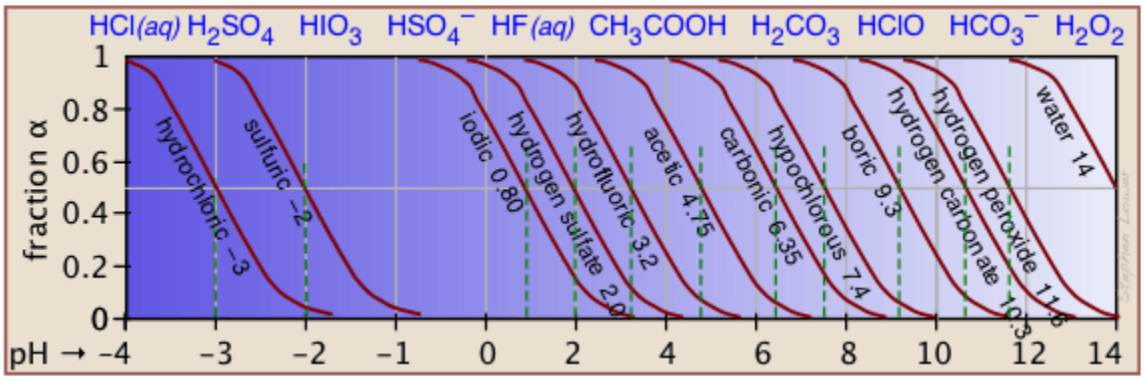
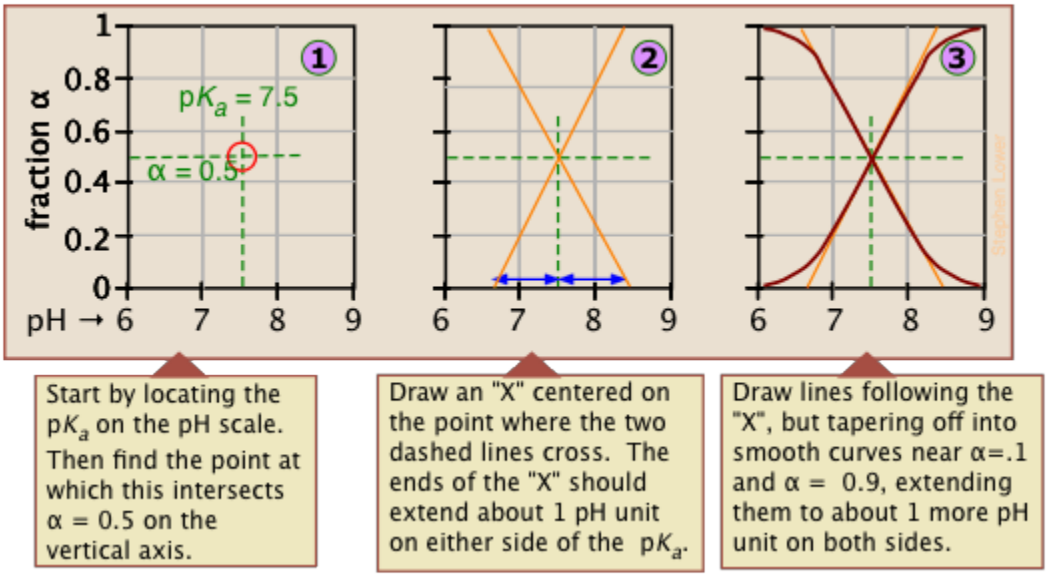


Figure 13.4.2: Ionization fractions of a series of acids

If you know the pK_a of an acid, you can easily sketch out its ionization fraction plot. The extreme tops and bottoms can only be estimated, but the rest of the plots are essentially straight lines.



As valuable as these plots are for showing how the distribution of conjugate species varies with the pH, they suffer from two drawbacks:

- Plots ofα cover only a single order of magnitude, but the actual concentrations, which we are often more interested in, vary over a far greater range from pH 0 to 14.

- They are of little help if one wishes to estimate the pH of a solution of the acid or base, or of one of its salts.

Both of these limitations are readily overcome by the use of easily-constructed logarithmic plots which we describe in the following section.

Buffer Solutions

The word *buffer* has several meanings in English, most of them referring (in its verb form) to cushion, shield, protect, or counteract an adverse effect. In chemistry, it refers specifically to a solution that resists a change in pH when acid or base is added. A *buffer* (or *buffered*) solution is one that resists a change in its pH when H^+ or OH^- ions are added or removed owing to some other reaction taking place in the same solution. Buffer solutions are essential components of all living organisms.

- Our blood is buffered to maintain a pH of 7.4 that must remain unchanged as metabolically-generated CO_2 (carbonic acid) is added and then removed by our lungs.
- Buffers in the oceans, in natural waters such as lakes and streams, and within soils help maintain their environmental stability against acid rain and increases in atmospheric CO_2.
- Many industrial processes, such as brewing, require buffer control, as do research studies in biochemistry and physiology that involve enzymes, are active only within certain pH ranges.

The essential component of a buffer system is a conjugate acid-base pair whose concentration is fairly high in relation to the concentrations of added H^+ or OH^- it is expected to buffer against. A simple buffer system might be a 0.2 *M* solution of sodium acetate; the conjugate pair here is acetic acid HAc and its conjugate base, the acetate ion Ac^-. The idea is that this conjugate pair "pool" will be available to gobble up any small ($\leq 10^{-3}$ *M*) addition of H+ or OH^- that may result from other processes going on in the solution.

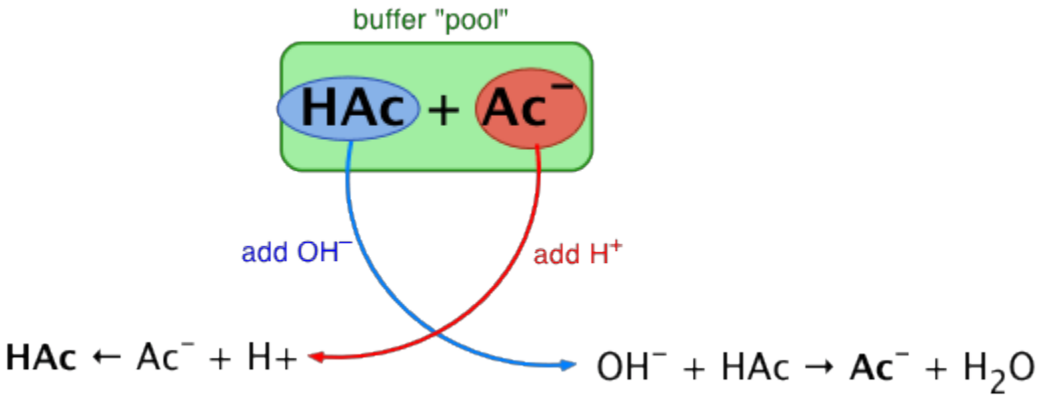


Figure 13.4.1: Buffer

Exercise 13.4.1

Sketch out a similar diagram showing what happens when you *remove* H^+ or OH^-.

Answer available at the end of this section.

You can also think of the process depicted above in terms of the Le Chatelier principle: addition of H^+ to the solution suppresses the dissociation of HAc, partially counteracting the effect of the added acid, as illustrated by the equation at the bottom left of the above diagram.

The Henderson-Hasselbalch Approximation

To develop this more quantitatively, we will consider the general case of a weak acid HA to which a quantity C_b of strong base has been added; think, for example, of acetic acid which has been partially neutralized by sodium hydroxide, yielding the same

conjugate pair described above, although not necessarily identical concentrations. The mass balance for such a system would be

$$[HA] + [A^-] = C_a + C_b = C_T \tag{13.4.6}$$

in which C_t denotes the total concentration of all species in the solution. Because we added a *strong* (completely dissociated) base NaOH to the acid, we also note that

$$C_b = [Na^+] \tag{13.4.7}$$

Recalling the equilibrium expression for a weak acid

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{13.4.8}$$

We can solve this for [H^+]:

$$[H^+] = K_a \frac{[HA]}{[A^-]} \tag{13.4.9}$$

Re-writing this in terms of negative logarithms, this becomes

$$-\log[H_3O^+] = -\log K_a - log[HA] + \log[A^-] \tag{13.4.10}$$

or, since $pK_a = -\log K_a$, we invert the ratio to preserve the positive sign:

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \tag{13.4.11}$$

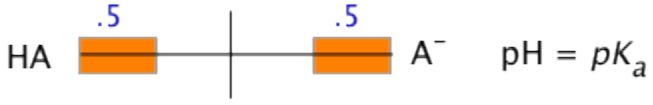
This equation is known as the *Henderson-Hasselbalch Approximation*. It tells us that the pH of a solution containing a weak acid-base system controls the relative concentrations of the acid and base forms of that system.

It is interesting to note that the H-H equation was not developed by chemists!

- Lawrence Henderson (1878-1942) was an American medical doctor who taught at Harvard and who studied the acidity of blood and its relation to respiration. In 1908 he worked out the relation shown in Equation 13.4.8.
- In 1916, K. Hasselbalch, a physiologist at U. of Copenhagen, derived the logarithmic form in Equation 13.4.11.

You may wonder why these two equations, whose derivation we now consider almost trivial, should have immortalized the names of these two scientists. The answer is that the theory of chemical equilibrium was still developing in the early 1900's, and had not yet made its way into chemistry textbooks. Even the concept of pH was unknown until Sørenson's work appeared in 1909. It was the mystery (and medical necessity) of understanding why shortness of breath made the blood more alkaline, and too-rapid breathing made it more acidic, that forced the work of H&H into modern Chemistry.

Of special interest is the case in which the pH of a solution of an acid-base system is set to the value of its pK_a. According to the above equation, when pH = pK_a, the log term becomes zero, so that the ratio [AB^-] / [HA] = 10^0 = 1, meaning that [HA] = [AB^-]. In other words, when the pH of a solution is set to the value of the pK_a of an acid-base pair, the concentrations of the acid- and base forms will be identical. This condition can be represented schematically on a proton-free energy diagram:



This says, in effect, that at when the pH of a solution containing both the acid and it conjugate base is made identical to the acid pK_a, the forms HA and A^- possess identical free energies, and will therefore be present in equal concentrations.

When the pH of a solution is set to the value of the pK_a of an acid-base pair, the concentrations of the acid- and base forms will

be identical.

Making and using Buffer Solutions

Buffers are generally most effective when the buffer system pK_a is not too far from the target pH. Under these conditions, both [HA] and [A^-] are large enough to compensate for the withdrawal or addition, respectively, of hydrogen ions.

Table 13.4.1: *Common buffer Systems*

acid	formulas	pK_a	pH range
phosphoric	H_2PO_4 / HPO_4^-	2.16	1-3
carbonic	H_2CO_3* / HCO_3^-	3.75	3-5
acetic	HCOOH / $HCOO^-$	4.75	4-6
dihydrogen phosphate	$H_2PO_4^-$ / HPO_4^{2-}	7.21	6-8
boric	H_3BO_3 / $H_2BO_3^-$	9.24	8-10
ammonium	NH_4^+ / NH_3	9.25	8-10
bicarbonate	HCO_3^- / CO_3^{2-}	10.3	9-11
monohydrogen phosphate	HPO_4^{2-} / PO_4^{3-}	12.3	11-13

Equation 13.4.9 above and its logarithmic equivalent in Equation 13.4.11 are of limited use in calculations because the exact values [A^-] and [HA] are known only for the special case when the pH of the solution is identical to the pK_a. Most buffer solutions will be adjusted to other pH's, and of course, once the buffered solution begins doing it work by counteracting the effect of additions of H^+ or OH^-, both [A^-] and [HA] will have changed. These two equations are so widely used in practical chemistry (and especially in biochemistry) that they are worth committing to memory.

increases [H^+] and decreases the pH. So you can always arrange the equation to make it come out right. For these reasons, it is useful to rewrite the above two equations by replacing the equilibrium concentrations [A^-] and [HA] with the known initial values C_a and C_b:

$$[H^+] \approx K_a \frac{C_a}{C_b} \tag{13.4.12}$$

$$pH \approx pK_a + \log \frac{C_b}{C_a} \tag{13.4.13}$$

Although these are approximations, they are usually justified because useful buffer systems are always significantly more concentrated than those responsible for adding or removing hydrogen ions. Also, at these higher ionic concentrations, the K_a of the buffer system will seldom be precisely known anyway. Don't expect actual buffer pH's to match calculations to better than 5%.

Example 13.4.2: Buffer Activity

Compare the effects of adding 1.0 mL of 2.5 *M* HCl to

a. 100 mL of pure water
b. 100 mL of 0.20 *M* sodium dihydrogen phosphate solution.

Solution

a)

1. quantity of HCl added: (1.0 mL) × (2.5\ mM/mL) = 2.5 mMol = 0.0025 mol
2. concentration of H^+ in solution: (.0025 mol) ÷ (.101 L) = .025 *M*
3. pH = – log (.025) = 1.60, change in pH = 1.60 – 7.00 = **–5.4**

b) (From Table 13.4.1, the pK_a of $H_2PO_4^-$ is 7.21)

1. Initial C_a ($H_2PO_4^-$) = C_b (HPO_4^{2-}) = 0.20 *M*
2. Initial pH of buffer solution: pH = pK_a + log (C_b/C_a) = 7.21 + 0 = 7.21
3. Added HCl will convert .0025 mol of HPO_4^{2-} to $H_2PO_4^-$, changing C_a to (.20 + .0025) mol ÷ (.101 L) = 2.00 *M*, and reducing C_b to (.20 – .0025) mol ÷ (.101 L) = 1.95 *M*
4. Resulting pH = 7.21 + log (1.95 / 2.00) = 7.21 –.011 = 7.00
5. Change in pH = 7.00 – 7.21 = **–.21**

Note that in both cases the pH is reduced (as it must be if we are adding acid!), but the change is far less in the buffered solution.

Example 13.4.3: Buffer Preparation

How would you prepare 200 mL of a buffer solution whose pH is 9.0?

Solution

The first step would be to select a conjugate acid-base pair whose pK_a is close to the desired pH. Looking at the above table of pK_a values, either boric acid or ammonium ion would be suitable. For this example, we will select the NH_4^+/NH_3 system, K_a = 5.5E–10, pK_a = 9.25.

2) We will use Equation 13.4.12 to determine the ratio C_a/C_b required to set [H^+] to the desired value of 1.00×10^{-9}. To do this, write out the complete equilibrium constant expression for the acid NH_4^+: $NH_4^+ \rightarrow NH_3 + H^+$, and then solve it to find the required concentration ratio:

$$K_a = \frac{[NH_3][H^+]}{[NH_4^+]} = 5.5 \times 10^{-10}$$

$$\frac{C_a}{C_b} = \frac{[NH_3]}{[NH_4^+]} = \frac{5.5 \times 10^{-10}}{[H^+]} = \frac{5.5 \times 10^{-10}}{1.00 \times 10^{-9}} = \frac{5.5 \times 10^{-10}}{10.00 \times 10^{-10}} = 0.55$$

3) The easiest (if not particularly elegant) way to work this out is to initially assume that we will dissolve 1.00 mole of solid NH_4Cl in water, add sufficient OH^- to partially neutralize some of this acid according to the reaction equation $NH_4^+ + OH^- \rightarrow NH_3 + H_2O$, and then add sufficient water to make the volume 1.00 L.

mass of ammonium chloride:

(1 mol) × (53.5 g *M*ol) = 53.8 g

let *x* = moles of NH_4^+ that must be converted to NH_3.
Then C_a/C_b = (1-*x*)/*x* = .55; solving this gives x = .64, 1–x = .36.

4) So to make 1 L of the buffer, dissolve .36 × 53.8 g = 19.3 g of solid NH_4Cl in a small quantity of water. Add .64 mole of NaOH (most easily done from a stock solution), and then sufficient water to make 1.00 L.

5) To make 200 mL of the buffer, just multiply each of the above figures by 0.200.

Pitfalls of the Henderson-Hasselbalch Approximation

The Henderson-Hasselbalch Approximation is widely used in practical calculations. What most books do not tell you is that Eqution 13.4.9 is no more than an “approximation of an approximation” which can give incorrect and misleading results when applied to situations in which the simplifying assumptions are not valid. An exact treatment of conjugate acid-base pairs, including a correct derivation of the Henderson-Hasselbalch equation, is given in the chapter on Exact Calculations.

The Henderson-Hasselbalch Approximation is only valid for fairly high concentrations

The approximations that lead to the H-H equation limit its reliable use to values of C_a and C_b that are within an order of magnitude of each other, and are fairly high. Also, the pK_a of the acid should be moderate.

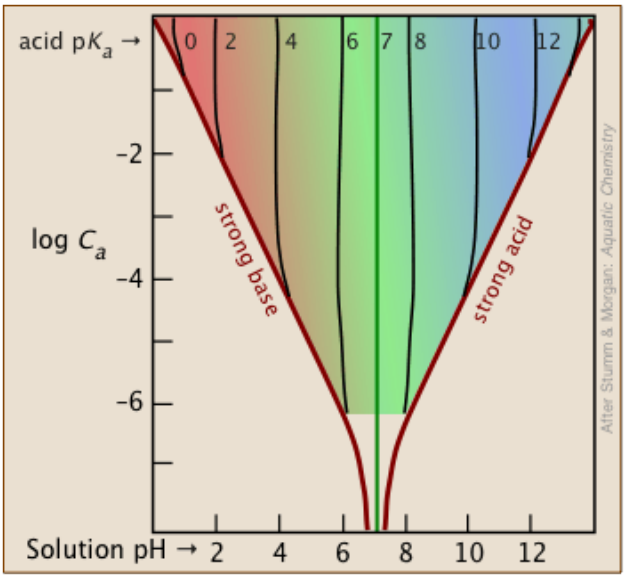


Figure 11.4.X:

The shaded portion of this set of plots indicates the values of C_a and K_a that yield useful results. Clearly, the smaller the buffer concentration, the narrower the range of useable acid pK_as. Most buffer solutions tend to be fairly concentrated, with C_a and C_b typically around 0.01 - 0.1 *M*. Thus a buffer based on a .01 *M* solution of an acid such as chloric ($HClO_3$) with pK_a of 1.9 will fall just outside the "safe" boundary near the upper left part of the diagram.

$$\underbrace{[H^+] = K_a \frac{[HA]}{[A^-]}}_{\text{exact}} \tag{13.4.14}$$

Equation 13.4.14 is simply a re-writing of the equilibrium constant expression, and is therefore always true. Of course, without knowing the actual equilibrium values of [HA] and [A^-], this relation is of little direct use in pH calculations.

$$\underbrace{[H^+] \approx K_a \frac{C_a}{C_b}}_{\text{approximate}} \tag{13.4.15}$$

Equation 13.4.15 is never true, but will yield good results if the acid is sufficiently weak in relation to its concentration to keep the [H^+] from being too high. Otherwise, the high [H^+] will convert a significant fraction of the A^- into the acid form HA, so that the ratio [HA]/[A^-] will differ from C_a /C_b in the above two equations. Consumption of H^+ by the base will also raise the pH above the predicted value as we saw in the preceding problem example.

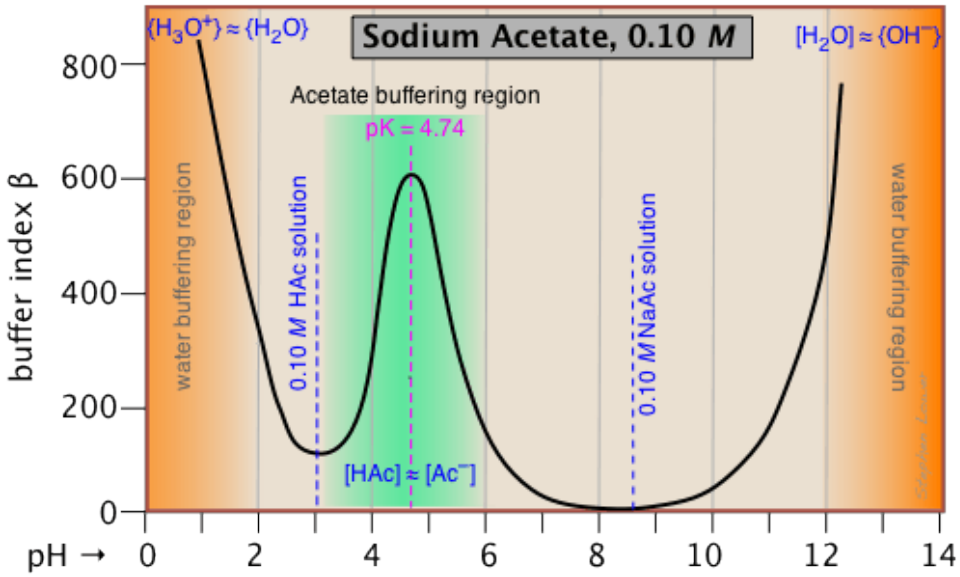
Buffer Index

The terms *buffer intensity* and *buffer capacity* are commonly employed as synonyms for buffer index, but in some contexts, buffer

capacity denotes the quantity of strong acid or strong base which alters the buffer's pH by 1 unit. (see below). How effective is a given buffer system in resisting changes in the pH? The most direct expression of this is the rate of change of the pH as small quantities of strong acid or base are added to the system: $\Delta C/\Delta(\text{pH})$. Expressed in calculus notation, this is the *buffer index*, defined as

$$\beta = \left| \frac{dC}{d(pH)} \right| \tag{13.4.16}$$

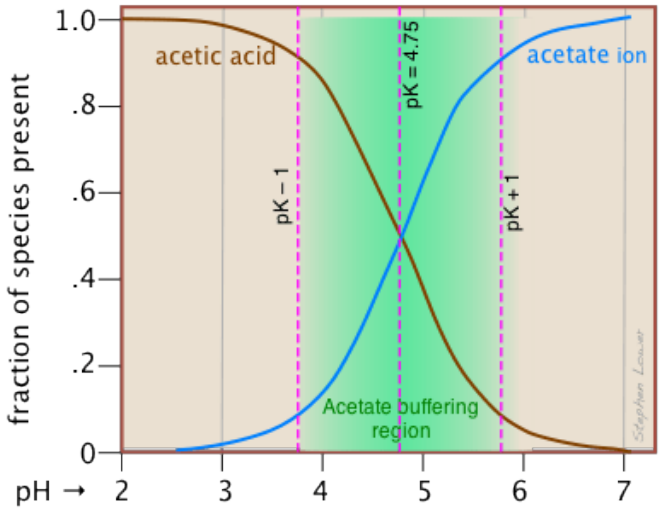
C here refers to the concentration of strong acid or base added to the solution. Because added acid or base affects the pH in opposite ways, we take the absolute value of this function in order to ensure that β is always positive. The value of β can be calculated analytically from C_a, C_b, K_a, K_b and $[H^+]$. By taking the second derivative of β, it can be shown that the buffer index has a maximum value when the pH = pK_a.



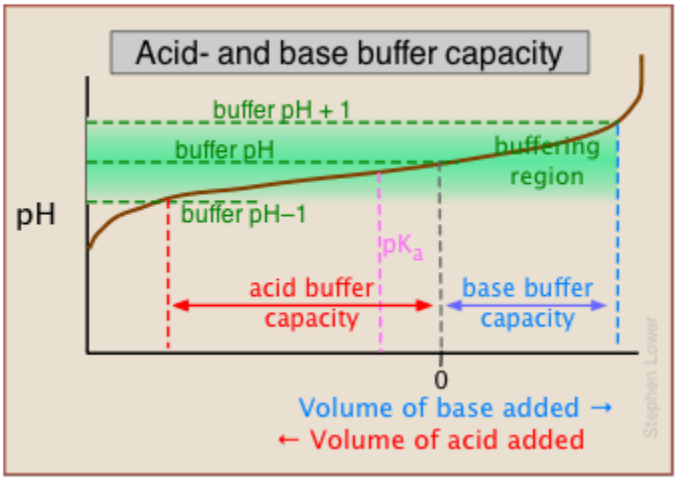
This buffer index plot for a 0.10 M solution of sodium acetate is typical, and confirms that buffering is most efficient within about ±1 pH unit of pK_a. But what about the even greater buffering that apparently occurs at the two extremes of pH (orange shading)? This is due to the buffering associated with the water itself, and will be seen in all aqueous buffer solutions.

Pure water is buffered by the H_3O^+/H_2O conjugate pair at very low pH, and by H_2O/OH- at high pH. This is easily understood if you think about adding some strong acid acid to pure water; even one drop of HCl will send the pH shooting down toward 0. If you continue adding acid, the pH will not drop significantly below 0, because there won't be enough free water molecules remaining to hydrate the HCl to produce H_3O^+ ions — thus the solution is strongly buffered.

Note that if we were to subtract the effect of the NaAc buffering from the above plot, the remaining plot for water itself would exhibit a minimum at pH 7, where both $[H_3O^+]$ and $[OH^-]$ share a common minimum value.



This alternative view shows how the distribution fractions of HAc and Ac^- relate to the effective buffering range of this conjugate pair, which is conventionally defined as ±1 pH unit of the pK_a. The term *buffer capacity* is an alternative means of expressing the ability of a buffer system to absorb the addition of strong acid or base without causing the pH to deviate by more than one unit from that of the pure buffer.



In this example, the buffer capacities for addition of acid and base will differ because the buffer pH has been adjusted to a value that differs from its pK_a.

Buffer systems must be appreciably more concentrated than the concentrations of strong acid or base they are required to absorb while still remaining within the desired pH range. Once the added acid or base has consumed most of one or other of the conjugate species comprising the buffer, the pH will no longer be stabilized. And of course, very small buffer concentrations will approach the pH of pure water.

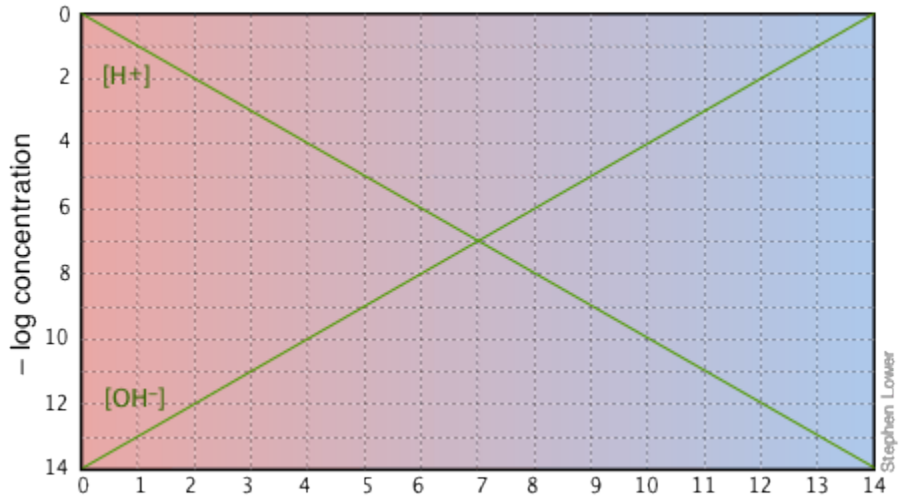
A more complete view: Log-C vs pH plots

These plots are often referred to as *Sillén plots*. **Lars Gunnar Sillén** (1916-1970) was a Swedish chemist who studied the distribution of ionic species in aqueous solutions and especially in the oceans. Because the concentrations of conjugate species can vary over many orders of magnitude, it is far more useful to express them on a logarithmic scale. Since pH is already

logarithmic, one can obtain a "bird's-eye view" of an acid-base system in a compact log-log plot.

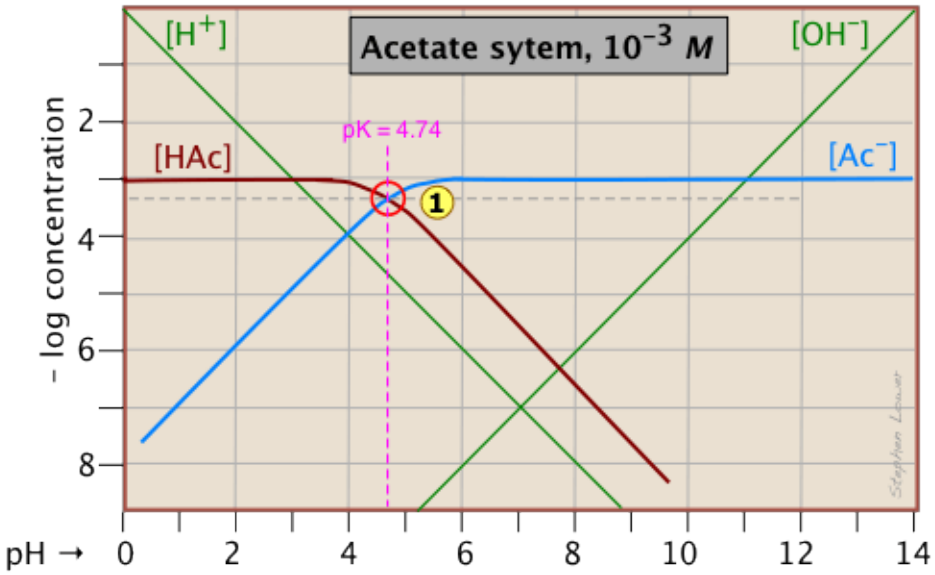
Even better, these plots are easily constructed without any calculations or arithmetic — or even any graph paper — any scrap paper, even the back of an envelope, will be sufficient. A ruler or other straightedge will, however, give more accurate results. In addition, you can use these plots to estimate the pH of a solution of a monoprotic or polyprotic acid or base without contending with quadratic or higher-order equations — or doing any arithmetic at all! The results will not be as precise as you would get from a proper numerical solution, but given the uncertainties of how equilibrium constants are affected by the presence of other ions in the solution, this is rarely a problem.

Aside from these advantages, the use of log-C vs pH plots will afford you an insight into the chemistry of acid-base systems that cannot be obtained simply by doing numerical calculations. The basic form of the plot, and the starting point for any use of such plots, looks like this:



If you examine this plot carefully, you will see that it is nothing more than a definition of pH and pOH, as well as a definition of a neutral solution at pH 7. Notice, for example, that when the pH is 4.0, $[H^+] = 10^{-4}$ *M* and $[OH^-] = 10^{-10}$ *M*.

A simple example: acetic acid solution



Here is a log C plot for a 10^{-3} *M* solution of acetic acid ("HAc") in water. Although it may look a bit complicated at first, it is really very simple. The heavy maroon line on the left plots the concentration of the acid HAc as a function of pH. The blue line on the

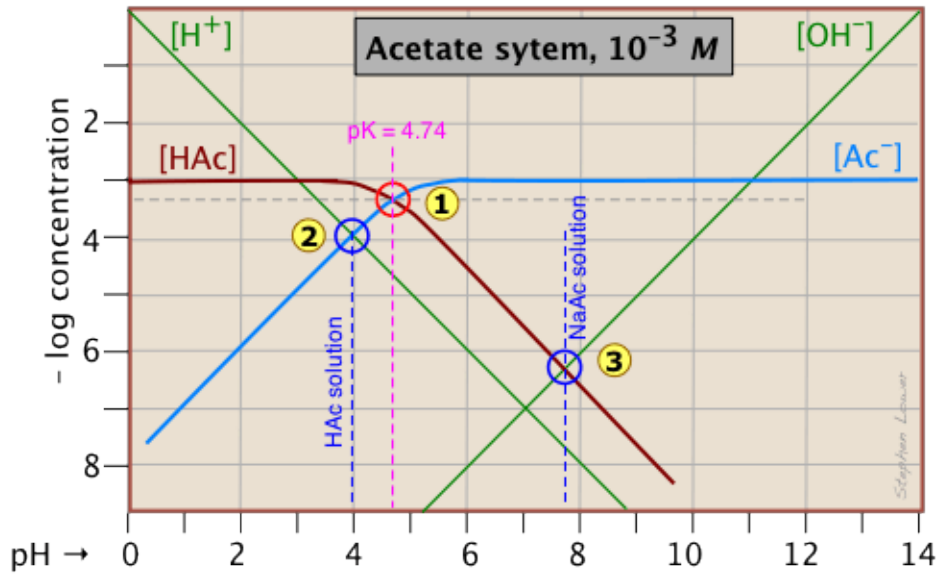
right shows how the concentration of the base Ac^- depends on pH. The horizontal parts of these lines are aligned with "3" on the –log-*C* axis, corresponding to the 10^{-3} *M* nominal concentration (C_a).

How do we know the shapes and placement of the plots for the concentrations of acetic acid [HAc] and the acetate ion [Ac^-] ? Although both of these concentrations will of course vary with the pH, their deviations from 10^{-3} *M* are too small to reveal themselves on a logarithmic plot until the pH approaches the pK_a.

- At very low pH, virtually all of the acetate system will be in its acid form (that is, [HAc] = 10^{-3}), and similarly, at high pH, the base form [Ac^-] = 10^{-3} *M*.
- When pH = pK_a, the concentrations of the conjugate species are identical, as indicated by the crossing of the lines representing the two concentrations. These concentrations will be half of the total concentration given by C_a, so [HAc] = [Ac^-] = ½C_a = 0.5E–3 *M*. Where does this go on the log-C axis? Well, the logarithm of 0.5 is –0.3 (a useful fact to remember!), so the crossing point at (5) is displaced by 0.3 of a log-C unit *below* the –3 level on the *y*-axis.
- Because point (5) defines both the pK_a and concentration of a particular acid-base system, it is known as the *system point*.
- What about the slopes of the plots when they bend down? It turns out that these slopes are +1 for [Ac^-] and –1 for [HAc]. Since the slopes of the lines for [H^+] and [OH^-] are ±1, we can use these as guidelines; just make the plots of [Ac^-] and [HA] parallel to them.
- The curved portions of the plot that joins the the horizontal and diagonal parts on either side of the system point can be drawn freehand without undue error; logarithmic plots are very forgiving!

Estimating the pH

Of special interest in acid-base chemistry are the pH values of a solution of an acid and of its conjugate base in pure water; as you know, these correspond to the beginning and equivalence points in the titration of an acid with a strong base. Continuing with the example of the acetic acid system, we show below another plot that is just like the one above, but with a few more lines and numbers added.



pH of acetic acid in pure water. Suppose you would like to find the pH of a 0.001 M solution of acetic acid in pure water. From the equation

$$HAc + H_2O \rightleftharpoons H_3O^+ + Ac^- \tag{13.4.17}$$

we know that equal numbers of moles of hydronium and acetate ions will be formed, so the concentrations of these species should be about the same:

$$[H^+] \approx [Ac^-] \tag{13.4.18}$$

will hold. The equivalence of these two concentrations corresponds to the point labeled (2) on the log*C*-pH plot; this occurs at a pH of about 4, and this is the pH of a 0.001 M solution of acetic acid in pure water.

pH of sodium acetate in pure water. A 0.001 M solution of NaAc in water corresponds to the composition of a solution of acetic acid that has been titrated to its equivalence point with sodium hydroxide. The acetate ion, being the conjugate base of a weak acid, will undergo hydrolysis according to

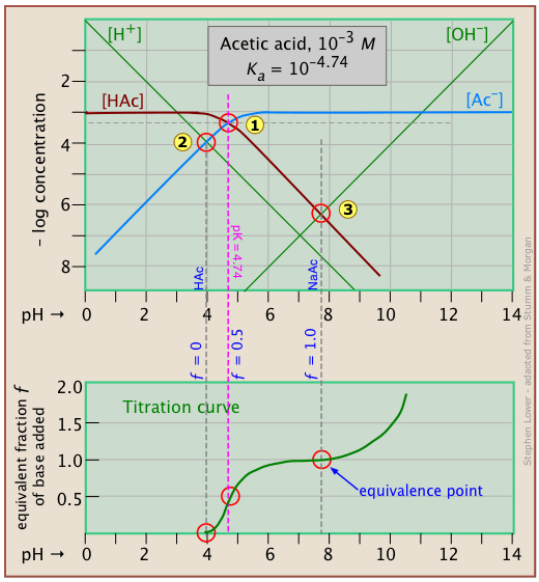
$$Ac^- + H_2O \rightleftharpoons HAc + OH^- \tag{13.4.19}$$

which establishes the approximate relation

$$[HAc] \approx [OH^-] \tag{13.4.20}$$

This condition occurs where the right sloping part of the line representing [HAc] intersects the [OH^-] line at (5) in the plot above. As you would expect for a solution of the conjugate base of a weak acid, this corresponds to an alkaline solution, in this case, at about pH = 7.8.

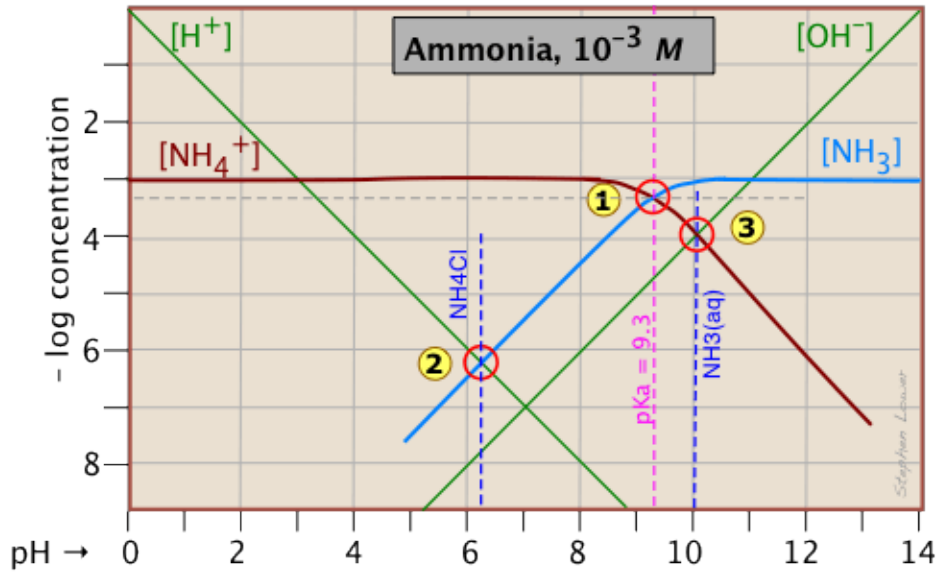
You will have to admit that estimating a pH in this way is far more painless than an ordinary numerical calculation would be!



It is interesting to show the Sillén plot for a weak acid system with its titration curve, both on the same pH scale. In this example, the titration plot has been turned on its side and reversed in order to illustrate the correspondence of its ƒ=0 (initial), ƒ=0.5 and ƒ=1 (equivalence) points with the corresponding points on the upper plot.

Ammonia - a weak base

Ammonia, unlike most gases, is extremely soluble in water. Solutions of NH_3 in water are properly known as *aqueous ammonia*, NH_3(aq); the name "ammonium hydroxide" is still often used, even though there is no evidence to suggest that a species NH_4OH exists. Ammonia is the conjugate base of the *ammonium ion* NH_4^+. The fact that ammonia is a base has no special significance insofar as the construction and interpretation of the log-C vs pH diagram is concerned; such diagrams always refer to conjugate acid-base *systems*, rather than to individual acids and bases.



Thus the system point (5) defines the NH_4^+/NH_3 pair (pK_a 9.3) at 25°C and a total concentration

$$C_T = C_a + C_b = 0.0010\ M. \tag{13.4.21}$$

To estimate the pH of a solution of ammonia in pure water, we make use of the charge balance requirement (known as the *proton condition*)

$$[NH_4^+] + [H^+] = [OH^-] \tag{13.4.22}$$

which we simplify by assuming that, because NH_4^+ is a weak acid, $[NH_4^+] >> [H^+]$ and we can drop the $[H^+]$ term; thus the approximation

$$[NH_4^+] \approx [OH^-] \tag{13.4.23}$$

which corresponds to the crossing point (5) and a pH of 10.0. Similarly, at the crossing point (5), $[NH_3] \approx [H^+]$. This comes from the proton condition

$$[NH_3] + [H^+] - [OH^-] = 0 \tag{13.4.24}$$

by dropping the $[OH^-]$ term on the assumption that its value is negligible compared to $[H^+]$.

Equation 13.4.24 is referred to as the "proton condition". Sorry to sneak it in without warning, but we hope that having piqued your curiosity, you might take the time to look at the following discussion of this important tool for estimating the pH of a solution of a pure acid, base, or a salt.

Mass- and charge balance, and the proton condition

In any aqueous solution of an acid or base, certain conservation conditions are strictly observed. Together with concentration and the K_a or K_b, these put certain constraints on the system that determine the state of the system. The most fundamental of these are conservation of mass and of electric charge, which of course apply to chemical changes of all kinds. We commonly express these as *mass balance* and *charge balance*, respectively. (The latter of these is sometimes referred to as the *electroneutrality condition*.)

Using the above example of the ammonium system as an illustration of this, we are interested in two particular instances of practical importance: what conditions apply to solutions made by dissolving *a)* pure ammonia, or *b)* ammonium chloride, in water?

Sillén diagram

Charge balances

$$C_{T,\,NH_3} = [NH_3] + [NH_4^+] = 10^{-2}\ M$$

$$C_{T,\,Ac} = [HAc] + [Ac^-] = 10^{-2}\ M$$

NH_3: $[H^+] + [NH_4^+] = [OH^-]$

NaAc: $[Na^+] + [H^+] = [Ac^-] + [OH^-]$

OK so far? That's really all we need, but it's usually more convenient to combine these with mass balances on the protons alone, so that we have a single equation for each of the two solutions. The resulting equations are known as the *proton conditions* for the two solutions.

To avoid a lot of algebra, there is a simple short-cut for writing a proton condition equation:

1. Identify the substance you are starting with — the pure acid, base, or salt, and also H_2O itself, that make up the solution whose pH you wish to know at the concentration for which the log C-pH chart is drawn. we call this the *proton reference level* (if that appellation scares you, it is sometimes referred to as the "basis substance".)
2. On the left side of the proton condition, write concentration expressions for all species that possess protons *in excess* of the reference level.
3. On the right side, write concentrations of all species that possess *fewer protons* the reference level.

Let's try it for our ammonia system.

Solution of NH_3 in water:

The proton reference level (PRL) is defined by NH_3(aq) and H_2O.

Proton condition: $[H_3O^+] + [NH_4^+] = [OH^-]$

(Notice that the substances that *define* the PRL (in case, H_2O and NH_3) never appear in the proton condition equation.)

Solution of NH_4Cl (or of any other strong-anion ammonium salt) in water:

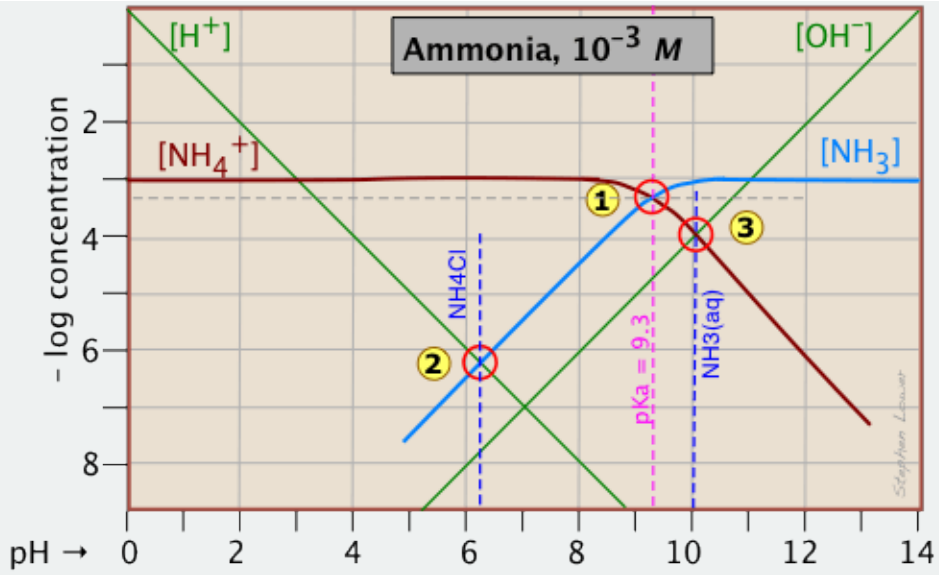
The PRL is defined by NH_4^+ and H_2O.

Proton condition: $[H_3O^+] = [NH_3] + [OH^-]$

(The chloride ion has nothing to do with protons, so it does not appear here.)

This proton condition stuff seems awfully complicated: why do we bother?

Good question: the answer is that what seems complicated at first sometimes turns out to be simpler and easier to understand than any alternative. Think of it this way: acid-base reactions involve the transfer of protons: some protons jump up to higher proton-free energy levels (e.g., $NH_3 + H_2O \rightarrow NH_4^+ + OH^-$), others drop down to proton-vacant levels ($H_3O^+ \rightarrow NH_3 \rightarrow H_2O + NH_4^+$). But no matter what happens, the total number of "available" protons (that is, all "dissociable" protons, including those in H_3O^+) must be conserved — thus the "mass balance on protons". So let's look again at the log C - pH plot for the ammonium system, which we reproduce here for your convenience:



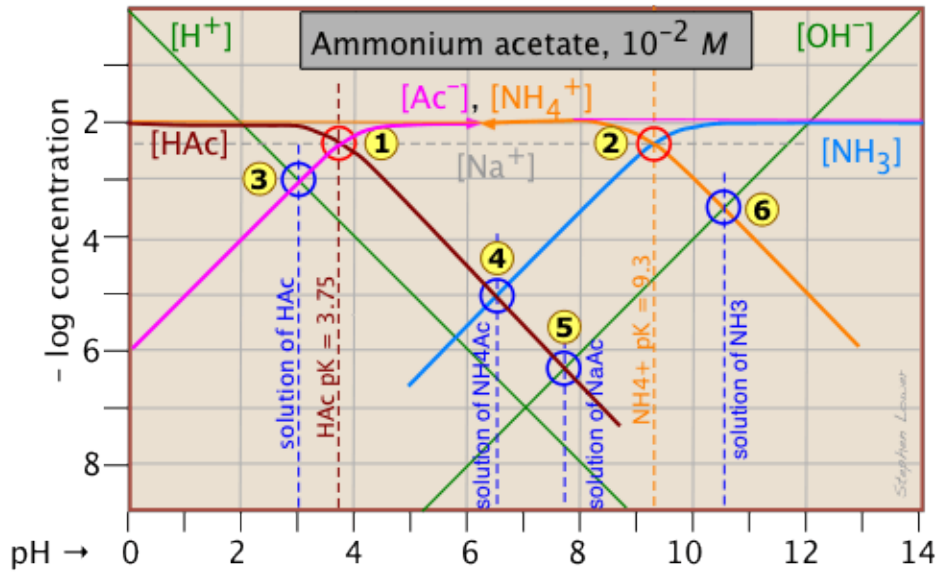
Consider first the solution of ammonium chloride, whose proton condition is given by $[H_3O^+] = [NH_3] + [OH^-]$. We know that NH_4Cl, being the salt of a weak base and a strong acid, will give a solution that is slightly acidic. In such a solution, $[OH^-]$ will be quite small so that we can neglect it without too much error. We can then write the proton condition as $[H^+] \approx [NH_3]$. On the plot above, this corresponds to point (5) where the lines representing these quantities cross. At this pH of around 6.2, $[OH^-]$ is about two orders of magnitude smaller than $[NH_3]$, so we are justified in dropping it.

Similarly, for a solution of ammonia in water, the proton condition $[H_3O^+] + [NH_4^+] = [OH^-]$ that we worked out above can be simplified to $[NH_4^+] \approx [OH^-]$ (point (5)) with hardly any error at all, since $[H^+]$ is here about six orders of magnitude smaller than $[NH_4^+]$.

In our discussion of the plot for the acetic acid system, we arrived at the proton conditions $[H^+] \approx [Ac^-]$ (for HAc in water) and $[H^+] \approx [OH^-]$ (for NaAc) by simply using the stoichiometries of the reactions. This can work in the simplest systems, but not in the more complicated ones involving polyprotic systems. In general, it is far safer to write out the complete proton condition in order to judge what concentrations, if any, can be dropped.

Ammonium acetate solution: solution of a salt

Having looked at the Sillén diagrams for acetic acid and ammonia, let's examine the log C-pH plot for this salt of acetic acid and ammonia.



You will recall that we dealt with solutions of the salts sodium acetate and ammonium chloride in the two examples described above. What is different here is that *both* components of the salt CH_3COONH_4 ("NH_4Ac") are "weak" in the sense that their conjugate species NH_3 and HAc are also present in significant quantities. This means that we are really dealing with *two* acid-base systems — the ones shown previously — on the same plot.

Each system retains its own system point, (1) for the acetate system and (5) for the ammonium system. Note that the two plots previously shown were for 10^{-3} *M* solutions, so the non-system points for this more concentrated solution occur at different pH values. The pH's of 10^{-2} *M* solutions of HAc (5), NaAc (5), and NH_3 are located by using the same proton conditions as before.

The only new thing here is the point (5), which corresponds to the ammonium acetate solution in which we are interested. The proton condition for this solution is $[H^+] + [HAc] = [NH_3(aq)] + [OH^-]$. Inspection of the plot reveals that $[H^+] << [HAc]$ and $[OH^-] << [NH_3(aq)]$, so the proton condition can be simplified to $[HAc] \approx [NH_3(aq)]$, corresponding to the intersection of these two lines at (5).

Polyprotic systems

Now that you are able to find your way around log C-pH plots that encompass two acid-base systems, the polyprotic systems that we describe below should be no trouble at all.

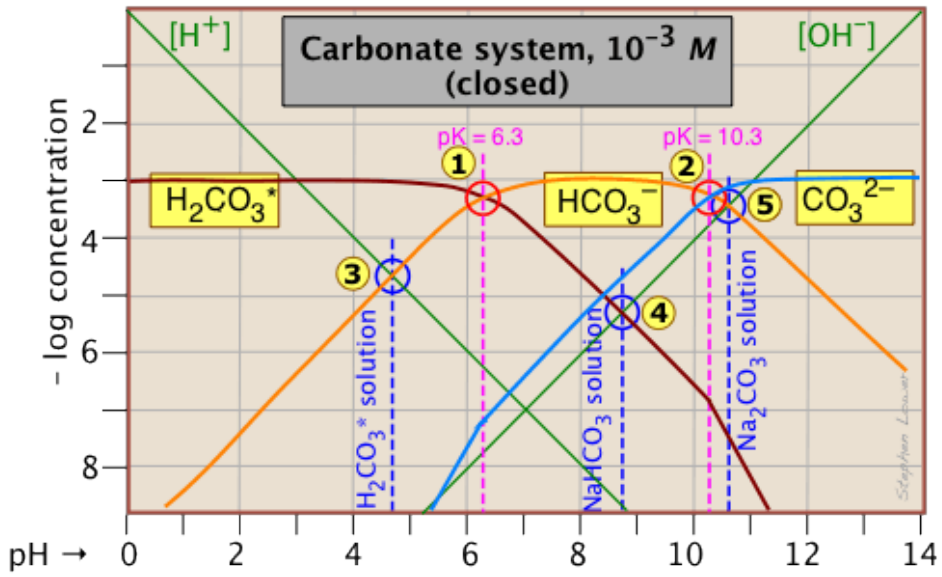
The Carbonate System

A thorough understanding of the carbonic acid-carbonate system is essential for understanding the chemistry of natural waters and the biochemistry of respiration in humans and other animals. There are two special things about this system that you need to know:

- A solution of carbon dioxide in water is mostly in the form of hydrated CO_2, $CO_2(aq)$. But a small fraction of dissolved CO_2 reacts with water to form carbonic acid, H_2CO_3. Because this acid cannot be isolated, common practice is to add an asterisk to the formula to designate the "total" CO_2 concentration
  $[CO_2 + H_2CO_3]$: thus H_2CO_3*.
- Because CO_2 is a gas, solutions containing carbonate species of all kinds can equilibrate with the atmosphere, which normally contains small quantities of CO_2. This means that a solution of sodium bicarbonate that is open to the atmosphere can lose CO_2 to the atmosphere at low pH, and absorb it at high pH. A system of this kind is said to be *open*. Alternatively, a system can be *closed* to exchange with the atmosphere. Sillén plots for open and closed carbonate systems have very

different appearances.

This Sillén plot for the H_2CO_3* closed system is typical of other diprotic acid systems in that there are three conjugate species and two system points. The 10^{-3} *M* concentration plotted here is typical of what is found in groundwaters that are in contact with carbonate-containing sediments.



The pH of a 10^{-3} *M* solution of CO_2, $NaHCO_3$ and Na_2CO_3 are indicted by the blue annotations.

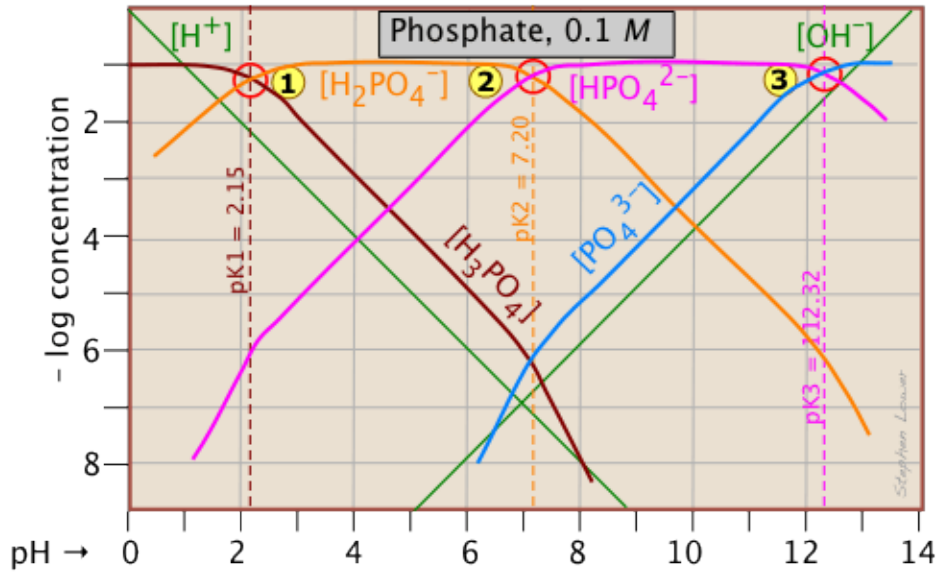
solute, (reference level)	proton condition	approximation
carbon dioxide (H_2CO_3*)	$[H^+] = [HCO_3^-] + 2[CO_3^{2-}] + [OH^-]$	$[H^+] \approx [HCO_3^-]$ (5)
sodium bicarbonate ($NaHCO_3$)	$[H^+] + [H_2CO_3] = [CO_3^{2-}] + [OH^-]$	$[H_2CO_3] \approx [OH^-]$ (5)
sodium carbonate (Na_2CO_3)	$[H^+] + 2[H_2CO_3^*] + [HCO_3^-] = [OH^-]$	$[HCO_3^-] \approx [OH^-]$ (5)

You should take a few moments to verify the approximations in the rightmost column. Notice the factors of 2 that multiply some of the concentrations. For example, the "$2[CO_3^{2-}]$" term in the first line of the table means that the CO_3^{2-} species is two steps away from the proton reference level H_2CO_3*, so it would take two hydrogen ions to balance a single carbonate ion.

Another new feature introduced here is the doubling of the slopes of the lines representing $[H_2CO_3^*]$ and $[CO_3^{2-}]$ where they cross the pH values corresponding to the K_as that are one step removed from their own system points (5) and (5) near –log C = 7.

The Phosphate System

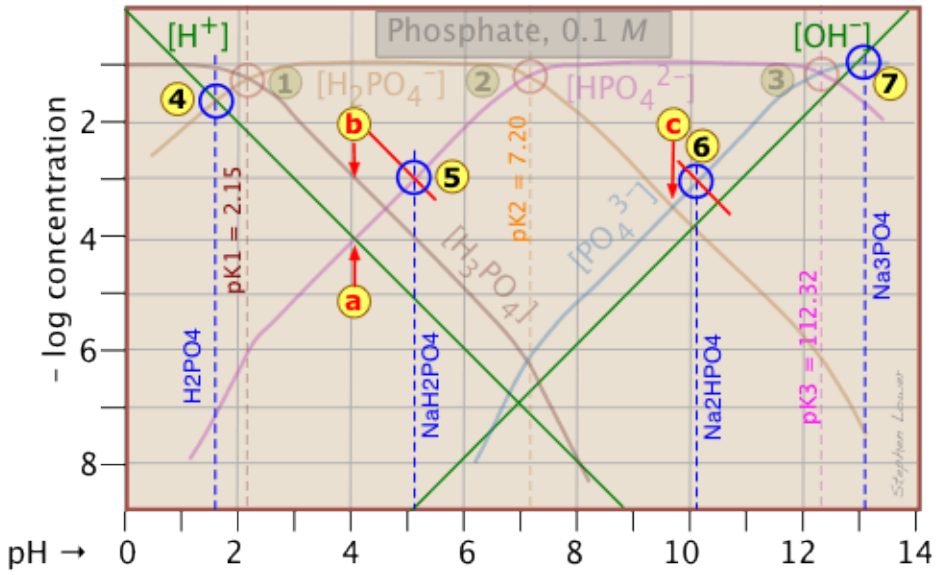
This triprotic system is widely used to prepare buffer solutions in biochemical applications. In its fundamental form, there is nothing new here, other than an additional system point (5) corresponding to the pK_a of $H_2PO_4^-$.



The pH of a 10^{-3} *M* solutions of H_3PO_4, NaH_2PO_4, Na_2HPO_4 and Na_3PO_4can be estimated by setting up the proton conditions as follows. Unfortunately, these estimates become a bit more complicated owing to the proximity of the three system points. But it's far easier than doing a numerical solution which requires solving a firth-degree polynomial!

solute, (reference level)	proton condition	approximation
phosphoric acid (H_3PO_4)	$[H^+] = [H_2PO_4^-] + 2[HPO_4^{2-}] + 3[PO_4^{3-}] + [OH^-]$	$[H^+] \approx [H_2PO_4^-]$ (5)
sodium dihydrogen phosphate (NaH_2PO_4)	$[H^+] + [H_3PO_4] = [HPO_4^{2-}] + 2[PO_4^{3-}] + [OH^-]$	$[H^+] + [H_3PO_4] \approx [HPO_4^{2-}]$ (5)
disodium hydrogen phosphate (Na_2HPO_4)	$[H^+] + 2[H_2PO_4^-] + [H_2PO_4^-] = 3[PO_4^{3-}] + [OH^-]$	$[H_3PO_4\text{-}] = [PO_4^{3-}] + [OH^-]$
trisodium phosphate (Na_3PO_4)	$[H^+] + 3[H_3PO_4] + 2[H_2PO_4^-] + [HPO_4^{2-}] = [OH^-]$	$[HPO_4^{2-}] \approx [OH^-]$

Because the phosphate plot is rather crowded, we show here a modified one in which the details of the pH estimates for the solutions of H_3PO_4 and its salts are emphasized.



Estimation of the pH of 10^{-3} *M* solutions of phosphoric acid (5) and trisodium phosphate , based on the approximations in the above table are straightforward.

Those for NaH_2PO_4 (5) and Na_2HPO_4 , however, are complicated by the presence of two concentration terms on the right side of the respective proton condition approximations.

Unless you are in an advanced course, you may wish to skip the following details. The important thing to know is that graphical estimates of these more problematic systems are fundamentally possible.

Thus for sodium dihydrogen phosphate, if we were to follow the pattern of the other systems, the approximation at the crossing point would be
$[H^+] = [HPO_4^{2-}]$, corresponding to point . But inspection of the plot shows that at that pH (4.00), the concentration of H_3PO_4 is ten times greater than that of H^+ (point). A rough estimate of the pH can be made by constructing a line (shown in red) that is parallel to those for H_3PO_4 and H^+, but is raised up by a factor of 10. This results in the new crossing point (5).

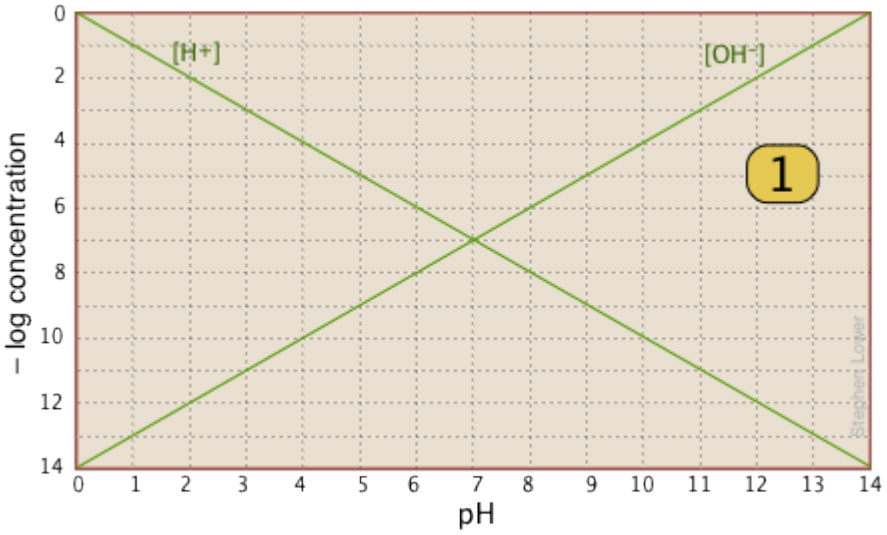
Similarly, the pH of a disodium hydrogen phosphate solution does not correspond to the crossing point because of the significant value of $[HPO_4^{3-}]$ at this pH. Again, we construct a line above this crossing point that predicts a pH corresponding to point .

Although these stratagems may seem rather crude, it should be noted that the uncertainties associated with them tend to be minimized on a logarithmic plot. In addition, it should be borne in mind that in a solution as concentrated as 0.1M, the pK_a values found in tables are not applicable. The alternative of solving for the pH algebraically, taking into consideration activity coefficients for the high ionic concentrations, is rarely worth the effort.

How to draw Sillén diagrams

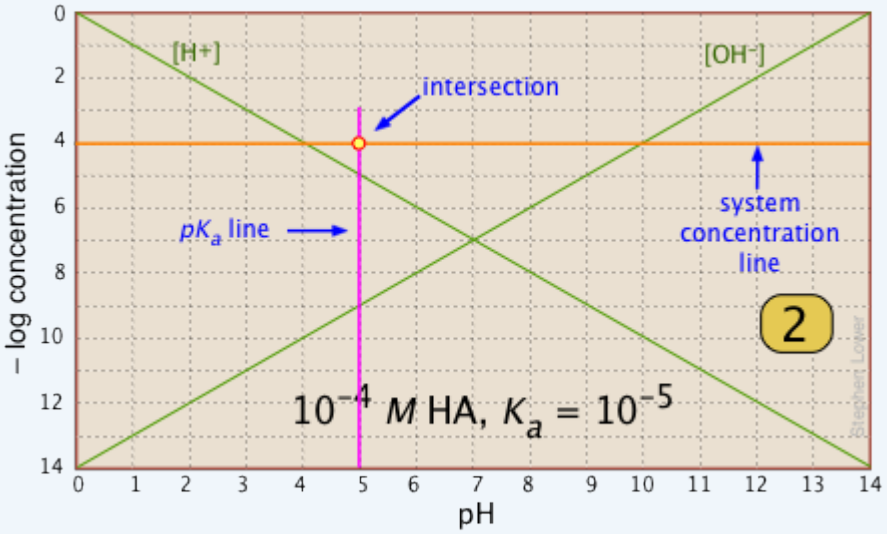
The relation between the pH of a solution and the concentrations of weak acids and their conjugate species is algebraically rather complicated. But over most of the pH range, simplifying assumptions can be made that, when expressed in logarithmic form, plot as straight lines having integral slopes of 0, ±1, ±2, etc. It is only within the narrow pH range near the pK_a that these simplifying assumptions break down, but on a logarithmic plot, one can draw a smooth curve that covers this range without introducing significant error.

Here is your basic stationery — a plot showing the two ions always present in any aqueous solution. In practice, you can usually cut off the bottom part where concentrations smaller than 10^{-10} M.



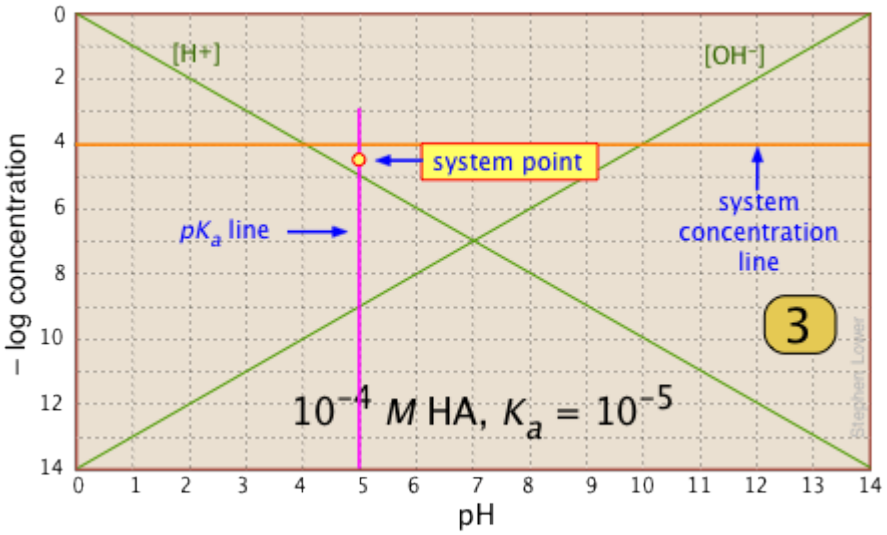
For any given acid-base system, you need to know the pK_a and nominal concentration CT.

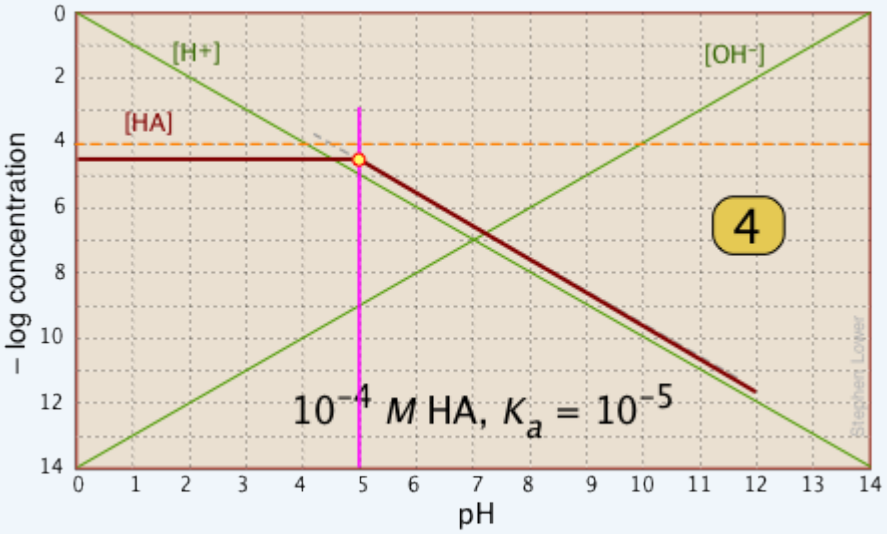
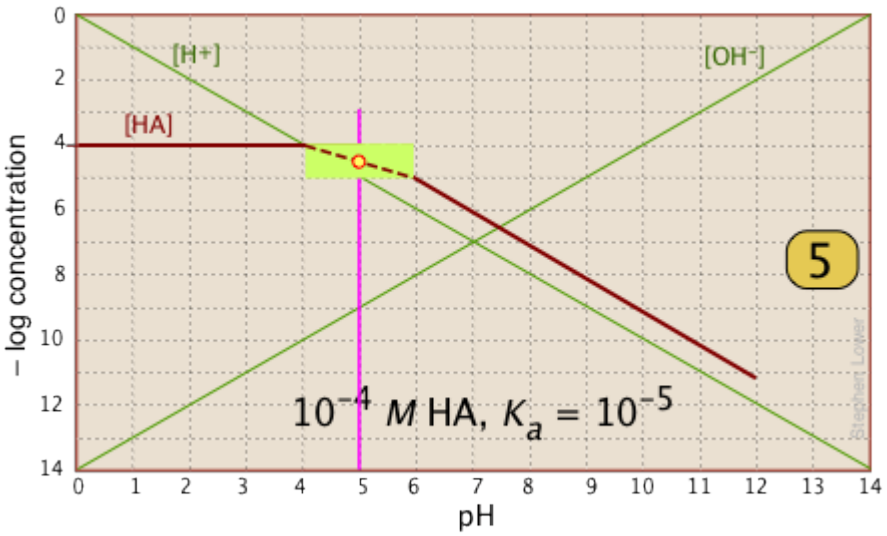
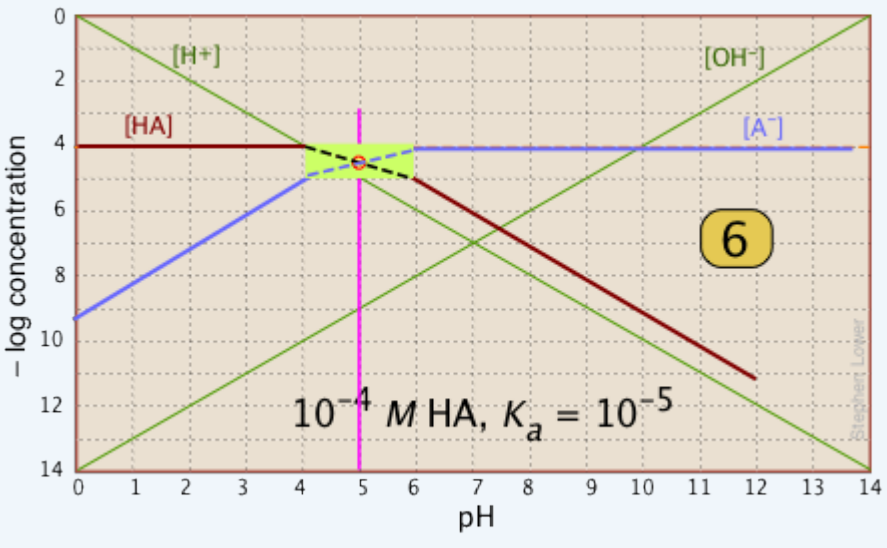
Find the point at which these two lines intersect.



Locate and mark the *system point* on the pK_a line. It will be 0.3 of a log-C unit below the concentration line.

This is where [HA] = [A^-] = ½ C_t. (Recall that log.5 = .3)



For pH < pK_a, [HA] plots as a line through the system point with slope = 0. For pH > pK_a, the line assumes a slope of –1. Use the $[H^+]$ and $[OH^-]$ lines as guides.	
In the pH interval of $\pm pK_a$, the slope changes from 0 to +1. You can approximate this with a sloped line as shown, or "prettify" it with a smooth curve.	
Finally, draw the lines for the conjugate form $[A^-]$, using slopes +1 and 0.	

What's all this good for?

Suppose you have a solution of NaA whose pH is 8. Estimate the concentration of the acid HA in the solution. All you need do is look at the plot. The [HA] line crosses the line marking this pH at – log C = 6, or [HA] = 10^{-6}. This quantity too small to significantly reduce [A^-], which remains at approximately CT = 10^{-6} *M*.

References

1. *Boundless* outlines on buffer solutions. These summarize the main points and problem examples as presented in major textbooks. They are an excellent aid for pre-exam review. *Chemistry: The Central Science* (8 subtopics)
2. A collection of practice exercises (with solutions) on buffers and titrations from Bryn Mawr College.
3. *The Henderson-Hasselbalch equation: its history and limitations* - Henry N. Po and N.M. Senoza, *J.Chem. Ed.* 78 (11) 2001 (1499-1503)
4. A detailed treatment of log-C vs pH diagrams can be found in the book *Acid-base diagrams* by Kahlert and Schotz. It is available online in some college libraries.

Answers

Exercise 13.4.1

When this happens, the ratio [HAc]/[Ac^-], will remain substantially unchanged, as will the pH, as you will see below.

This page titled 13.4: Conjugate Pairs and Buffers is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13.5: Acid/Base Titration

Learning Objectives

Make sure you thoroughly understand the following essential concepts:

- Sketch out a plot representing the *titration of a strong monoprotic acid* by a strong base, or of a strong base titrated by a strong acid. Identify the equivalence point and explain its significance.
- On the plot referred to above, draw a similar plot that would correspond to the same system at a higher or lower *concentration*.
- Sketch out a plot representing the *titration of a weak monoprotic acid* by a strong base, or of a weak base titrated by a strong acid. Identify the equivalence point and half-equivalence points.
- Explain what an acid-base *indicator* is, and how it works.
- When solutions of some *polyprotic acids* are titrated with strong base, not all of the equivalence points can be observed. Explain the most common reasons for this.
- *Calculate the molarity* of a monoprotic acid HA whose titration endpoint occurs after *V* ml of strong base of a given concentration has been added.

The objective of an acid-base titration is to determine C_a, the nominal concentration of acid in the solution. In its simplest form, titration is carried out by measuring the volume of the solution of strong base required to complete the reaction

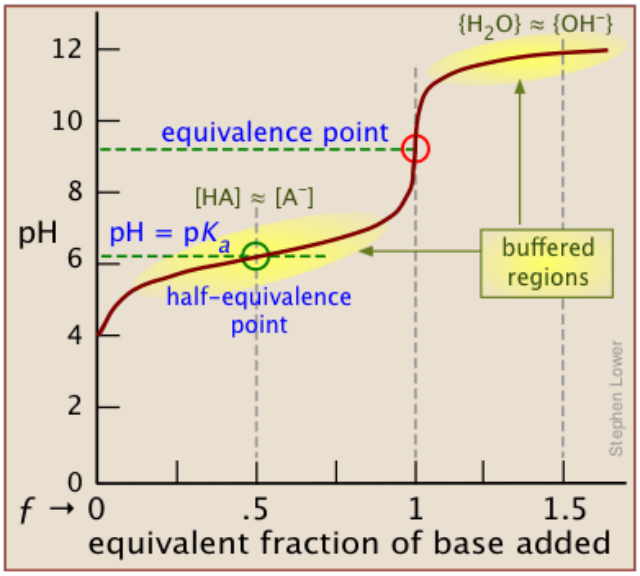
$$\mathrm{H}_n\mathrm{A} + n\,\mathrm{OH}^- \longrightarrow n\,\mathrm{A}^- + n\,\mathrm{H_2O} \tag{13.5.1}$$

in which n is the number of *replaceable hydrogens* in the acid. The point at which this reaction is just complete is known as the *equivalence point*. This is to be distinguished from the *end point*, which is the value we observe experimentally. A *replaceable* hydrogen atom (sometimes called an "acidic" hydrogen) is one that can be donated to a strong base — that is, to an OH^- ion. Thus in acetic acid HCOO**H**, only the hydrogen in the carboxyl group is considered "replaceable".

What we actually measure, of course, is the volume of titrant delivered by the burette. Learning to properly control the stopcock at the bottom of the burette usually requires some instruction and practice, as does the reading of the volume. For highly precise work, the concentration of the titrant itself must be determined in a separate experiment known as "standardization".

Understanding Titration Curves

A plot showing the pH of the solution as a function of the quantity of base added is known as a *titration curve*. These plots can be constructed by plotting the pH as a function of either the volume of base added, or the *equivalent fraction* f which is simply the number of moles of base added per mole of acid present in the solution. In most of the titration curves illustrated in this section, we plot pH as a function of f. It's worth taking some time to thoroughly familiarize yourself with the general form of a titration curve such as the one shown below, in which a weak acid HA is titrated with a strong base, typically sodium hydroxide.



The equivalence point occurs at the pH at which the equivalent fraction ƒ of base added is unity. At this point, the reaction

$$HA + OH^- \rightarrow AB^- + H_2O \tag{13.5.2}$$

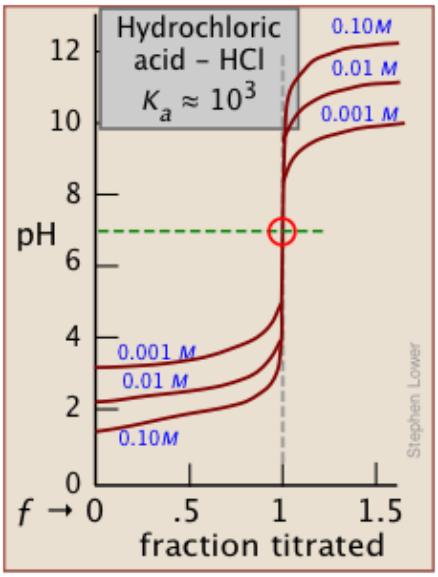
is stoichiometrically complete; a solution initially containing *n* moles of a monoprotic acid HA will now be identical to one containing the same number of moles of the conjugate base A^-. At the *half-equivalence point* ƒ = 0.5, the concentrations of the conjugate species are identical: [HA] = $[A^-]$. This, of course corresponds to a *buffer solution* (hence the relatively flat part of the curve) whose pH is the same as pK_a.

As base is added beyond ƒ = 1, the pH begins to level off, suggesting that another buffered system has come into play. In this case it involves the solvent (water) and hydroxide ion: {H_2O} ≈ {OH-}.

A similar effect is seen at the low-pH side of the curve when a strong acid is titrated, as in the plot for the titration of HCl below. In this case, the buffering is due to {H_3O^+) ≈ {H2O}.

How can this be? Surely, the concentration of OH^-, even when the pH approaches 14, cannot be anything like that of [H_2O] which will be about 55.5 *M* in most solutions! This fine point (along with the mention of H_2O/OH^- buffering) is rarely mentioned in elementary courses because the theory behind it involves some rather esoteric elements of solution thermodynamics. However, in case you are curious, note that the curly brackets in {H_2O} ≈ {OH^-} denote *activities,* not concentrations. And by convention the activity of a pure liquid (H_2O in this case) is unity. At a pH of around 12, pOH = 2, [OH^-] = .01. At this rather high ion concentration, {OH^-} will be somewhat smaller than this, but the two activities will be similar enough to produce the buffering effect we observe.

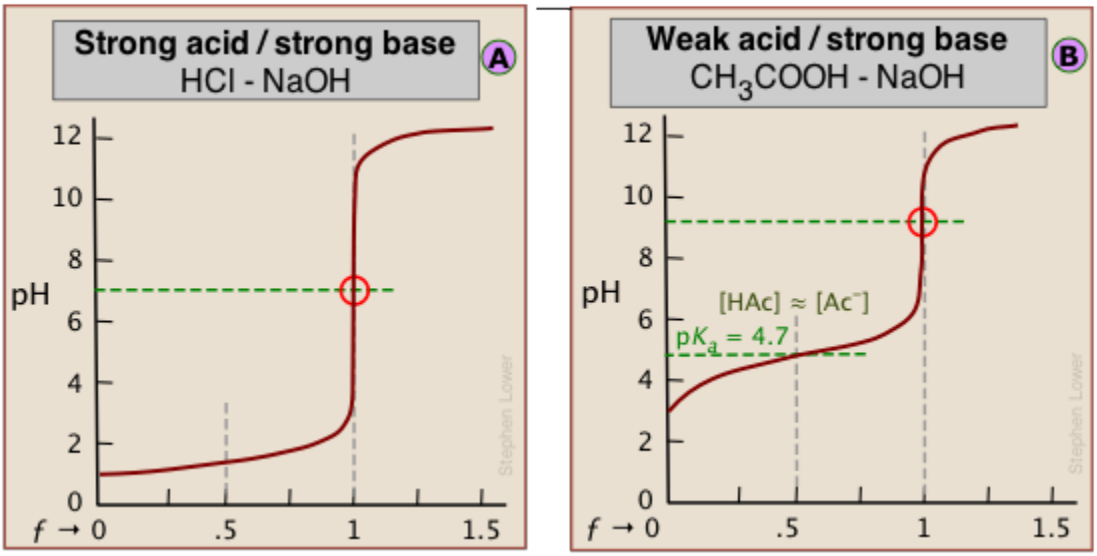
The pH of the solution at its equivalence point will be 7 if we are titrating a strong acid with strong base, as in HCl + NaOH → H_2O + NaCl. However, if the acid is weak, as in the above plot, the solution will be alkaline. This pH can be calculated from C_b and K_b in a manner exactly analogous to that used for calculating the pH of a solution of a weak acid in water.



It is important to understand that the equivalent fraction ƒ of base that must be added to reach the equivalence point is independent of the *strength* of the acid and of its *concentration* in the solution. The whole utility of titration as a means of quantitative analysis rests on this independence; we are in all cases measuring only the *total number of moles of "acidic" hydrogens* in the sample undergoing titration.

Acid and base strengths determine the shape of the curve

Although the strength of an acid has no effect on the location of the equivalence point, it does affect the *shape* of the titration curve and can be estimated on a plot of the curve.



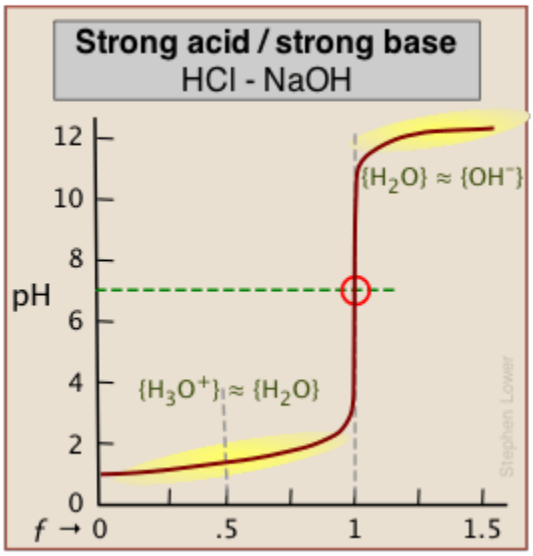
The weaker the acid being titrated, the higher the initial pH (at ƒ=0), and the smaller will be the vertical height of the plot near the equivalence point. As we shall see later, this can make it difficult to locate the equivalence point if the acid is extremely weak.

Estimating the acid strength

As shown in plot B above, the pK_a of a weak acid can be estimated by noting the pH that corresponds to the half-titration point ƒ = 0.5. Recalling that the pH is controlled by the ratio of conjugate species concentrations

$$pH = pK_a + \log \frac{[A]}{[HA]} \tag{13.5.3}$$

it will be apparent that this equation reduces to pH = pK_a when the titration is half complete (that is, when [HA] = [A^-]), the pH of the solution will be identical to the pK_a of the acid. This equation does not work for strong acids owing to the strong buffering that occurs at the very low pH at which ƒ = 0.5.



As indicated here, the buffering has nothing to do with the acid HCl itself (which does not exist as such in water), but rather with its dissociation products H_3O^+ and OH^-, "the strongest acid and base that can exist in water."

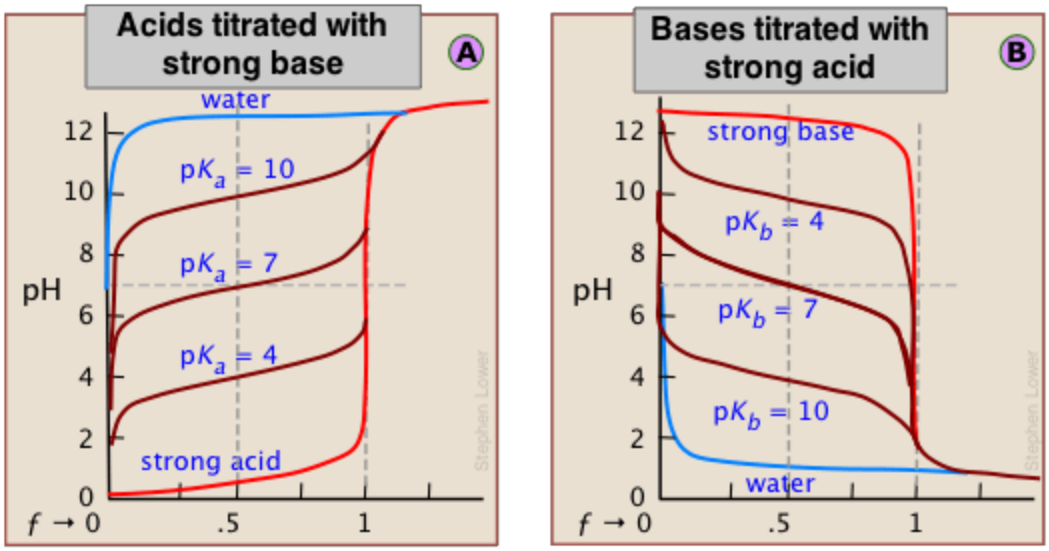
Monoprotic titration curves

The following two principles govern the detailed shape of a titration curve:

- The stronger the acid or base, the greater will be the slope of the curve near the equivalence point;
- The weaker the acid or base, the greater will the deviation of the pH from neutrality at the equivalence point.

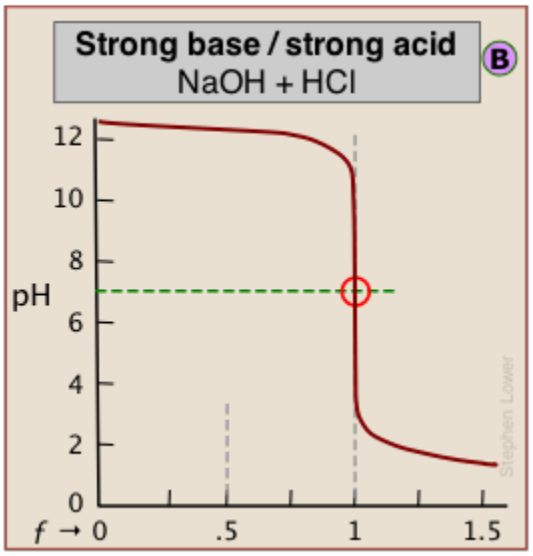
It is important to understand the reasons for these two relations. The second is the simplest to explain. Titration of an acid HA with a base such as NaOH results in a solution of NaA; that is, a solution of the conjugate base A^-. Being a base, it will react with water to yield an excess of hydroxide ions, leaving a slightly alkaline solution. Titration of a weak base with an acid will have the opposite effect.

The extent of the jump in the pH at the equivalence point is determined by a combination of factors. In the case of a weak acid, for example, the initial pH is likely to be higher, so the titration curve starts higher. Further, the weaker the acid, the stronger will be its conjugate base, so the higher will be the pH at the equivalence point. These two factors raise the bottom part of the titration curve. The upper extent of the curve is of course limited by the concentration and strength of the titrant.

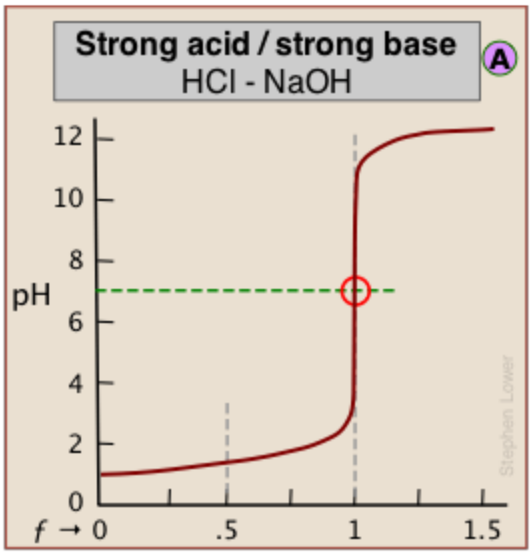


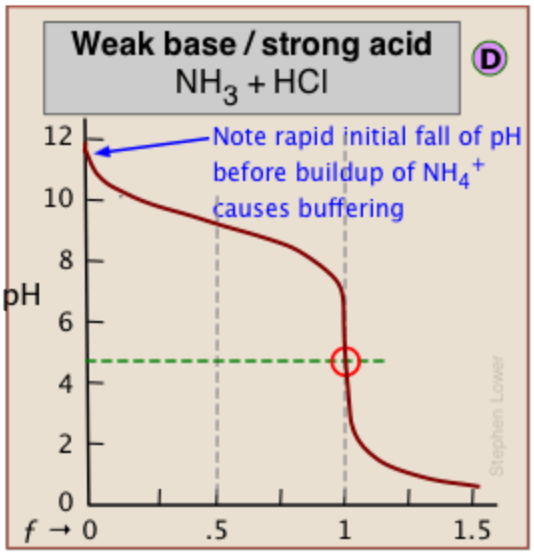
These principles are clearly evident in the above plots for the titrations of acids and bases having various strengths. Notice the blue curves that represent the titration of pure water (a very weak acid) with strong acid or base.

Monoprotic titration curve gallery



When both the titrant and sample are "strong", we get long vertical plots at ƒ = 1. Adding even half a drop of titrant can take us across the equivalence point!

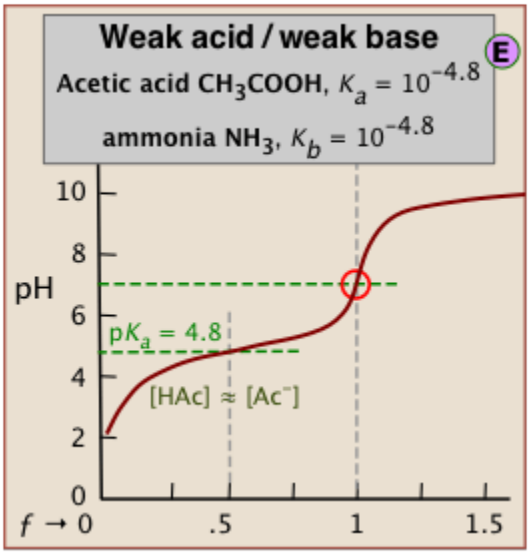




When one of the reactants is weak, the pH changes rapidly at first until buffering sets in.

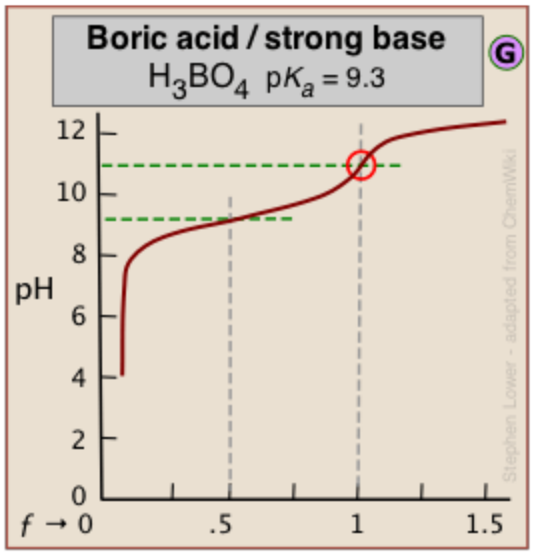
← In (C), the onset of H_2O/OH- buffering near ƒ=1 makes the equivalence point more difficult to locate.

"Weak/weak" titrations tend to be problematic as the buffered regions move closer to ƒ=1. The equivalence point pH of 7 in these examples reflects the near-equality of pK_a and pK_b of the reactants.



Dealing with very weak acids

It can be difficult to reliably detect the equivalence point in the titration of boric acid (pK_a = 9.3) or of other similarly weak acids from the shape of the titration curve*. *An interesting student laboratory experiment that employs an auxiliary reagent (mannitol) to make boric acid stronger and thus more readily titratable was described in *J. Chem Ed.* 2012, 89, 767-770.



The problem here is that aqueous solutions are buffered against pH change at very low and very high pH ranges. An extreme example occurs in the **titration of pure water** with a strong acid or base. At these extremes of pH the concentrations of H_3O^+ and of OH^- are sufficiently great that a competing buffer system (either H_3O^+/H_2O or H_2O/OH^-, depending on whether the solution is highly acidic or highly alkaline) comes into play.

Why we usually use a "strong" titrant

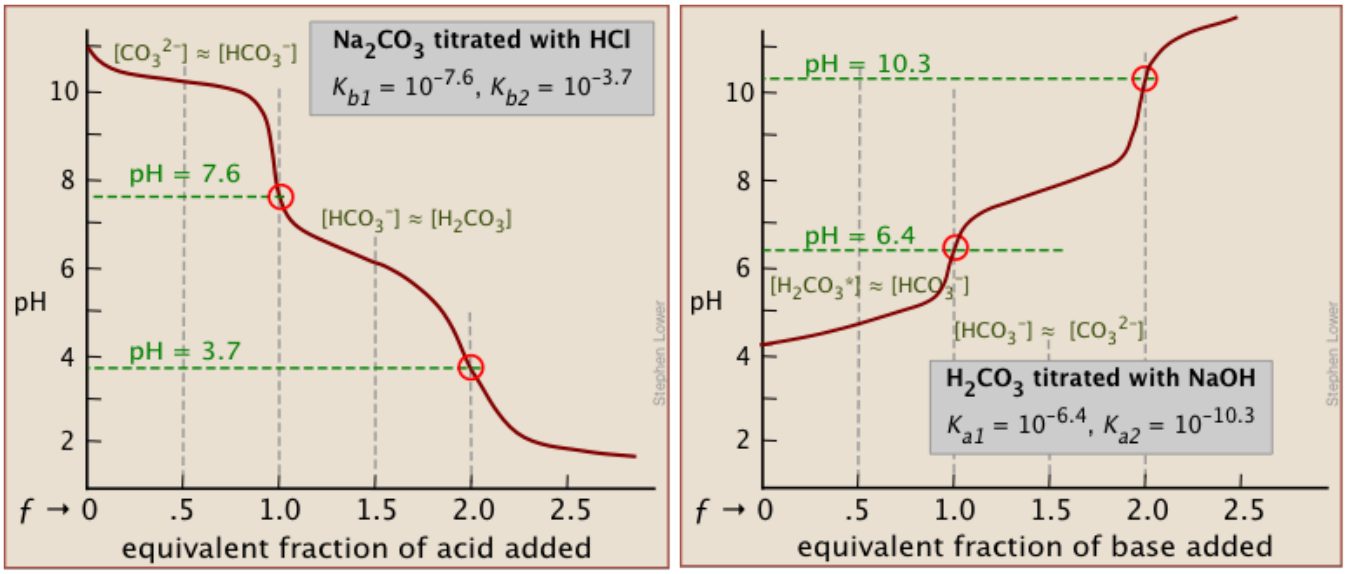
The above plots clearly show that the most easily-detectable equivalence points occur when an acid with is titrated with a strong base such as sodium hydroxide (or a base is titrated with a strong acid.)

In practice, many of the titrations carried out in research, industry, and clinical practice involve *mixture*s of more than one acid. Examples include natural waters, physiological fluids, fruit juices, wine making, brewing, and industrial effluents. For titrating these kinds of samples, the use of anything other than a strong titrant presents the possibility that the titrant may be weaker than one or more of the "stronger" components in the sample, in which case it would be incapable of titrating these components to completion.

In terms of proton-free energies, the proton source (the acidic titrant) would be unable to deliver an equivalent quantity of protons to the (stronger) component of the mixture.

Polyprotic acids

There will be as many equivalence points as there are replaceable hydrogens in an acid. Thus in the extremely important *carbonate system*, equivalence points are seen at both ƒ=1 and ƒ=2:

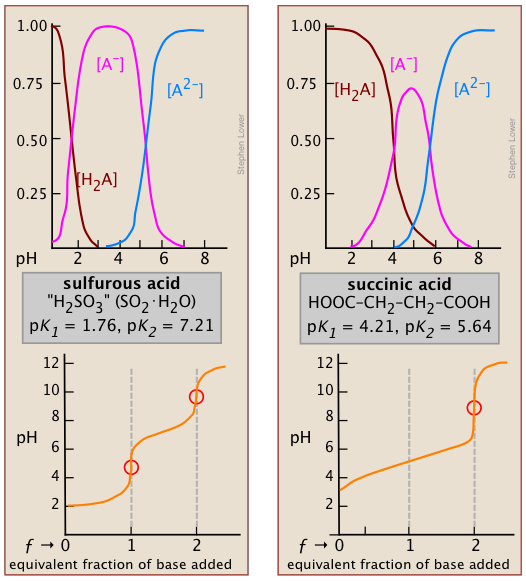


In general, there are two requirements for a clearly discernible jump in the pH to occur in a polyprotic titration:

- The successive K_a's must differ by several orders of magnitude;
- The pH of the equivalence point must not be very high or very low.

Separation of successive equilibrium constants

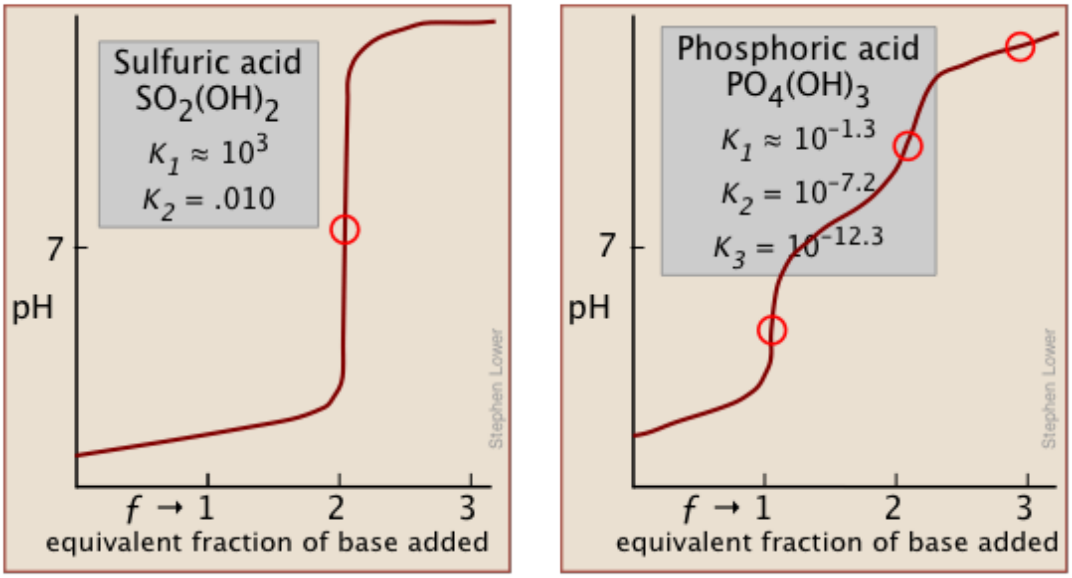
The effect of the first point is seen by comparing the titration curves of two diprotic acids, sulfurous and succinic. The appearance of only one equivalence point in the latter is a consequence of the closeness of the first and second acid dissociation constants. The pK_a's of **sulfurous acid** (below, left) are sufficiently far apart that its titration curve can be regarded as the superposition of those for two independent monoprotic acids having the corresponding K_a's. This reflects the fact that the two acidic –OH groups are connected to the same central atom, so that the local negative charge that remains when HSO_3^- is formed acts to suppress the second dissociation step.



*It can be shown that in the limit of large n, the ratio of K_1/K_2 for a symmetrical dicarboxylic acid HOOC-$(CH_2)_n$- COOH converges to a value of 4.

In **succinic acid**, the two –COOH groups are physically more separated and thus tend to dissociate independently*. Inspection of the species distribution curves for succinic acid (above, right) reveals that the fraction of the ampholyte HA can never exceed 75 percent. That is, there is no pH at which the reaction $H_2A \rightarrow HA^- + H^+$ can be said to be “complete” while at the same time the second step $HA^- \rightarrow A^{2-} + H^+$ has occurred to only a negligible extent. Thus the rise in the pH that would normally be expected as HA is produced will be prevented by consumption of OH^- in the second step which will be well underway at that point; only when all steps are completed and hydroxide ion is no longer being consumed will the pH rise.

Two other examples of polyprotic acids whose titration curves do not reveal all of the equivalence points are **sulfuric** and phosphoric acids. Owing to the leveling effect, the apparent K_{a1} of H_2SO_4 is so close to K_{a2} = 0.01 that the effect is the same as in succinic acid, so only the second equivalence point is detected.



In **phosphoric acid**, the third equivalence point (for HPO_4^{2-}) is obscured by H_2O-OH^- buffering as explained previously.

Detection of the equivalence point

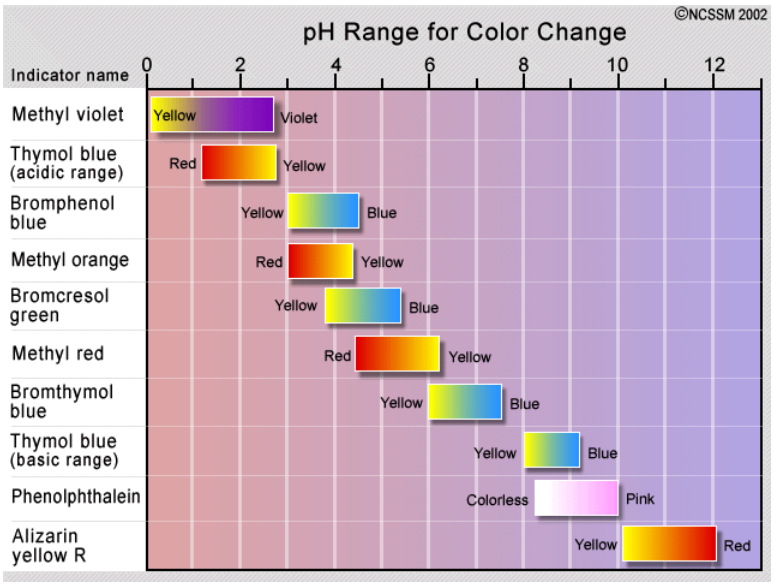
Whether or not the equivalence point is revealed by a distinct "break" in the titration curve, it will correspond to a unique hydrogen ion concentration which can be calculated in advance. There are many ways of determining the equivalence point of an acid-base titration.

Indicators



Don't overshoot the equiv point!

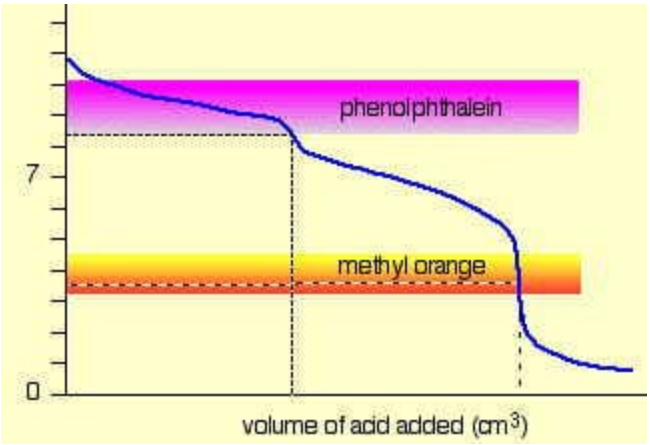
The traditional method of detecting the equivalence point has been to employ an *indicator* dye, which is a second acid-base system in which the protonated and deprotonated forms differ in color, and whose pK_a is close to the pH expected at the equivalence point. If the acid being titrated is not a strong one, it is important to keep the indicator concentration as low as possible in order to prevent its own consumption of OH^- from distorting the titration curve.



The observed color change of an indicator does not take place sharply, but occurs over a range of about 1.5 to 2 pH units. Indicators are therefore only useful in the titration of acids and bases that are sufficiently strong to show a definite break in the titration curve. Some plants contain coloring agents that can act as natural pH indicators. These include cabbage (shown), beets, and hydrangea flowers.



For a strong acid - strong base titration, almost any indicator can be used, although phenolphthalein is most commonly employed. For titrations involving weak acids or bases, as in the acid titration of sodium carbonate solution shown here, the indicator should have a pK close to that of the substance being titrated.



When titrating a polyprotic acid or base, multiple indicators are required if more than one equivalence point is to be seen. The pK_as of phenolphthalein and methyl orange are 9.3 and 3.7, respectively.

Potentiometry: Use a pH meter

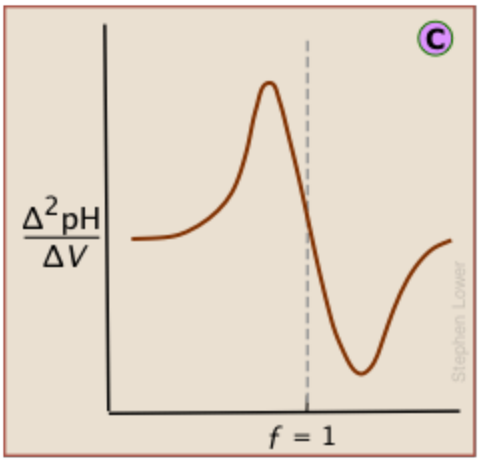
The pH meter detects the voltage produced when the H^+ ions in the solution displace Na^+ ions from a thin glass membrane that is dipped into the solution.



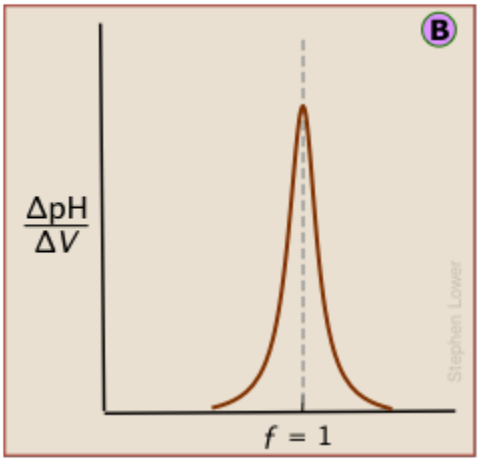
A more modern way of finding an equivalence point is to follow the titration by means of a pH meter. Because it involves measuring the electrical potential difference between two electrodes, this method is known as *potentiometry*. Until around 1980, pH meters were too expensive for regular use in student laboratories, but this has changed; potentiometry is now the standard tool for determining equivalence points.

Improved plotting methods

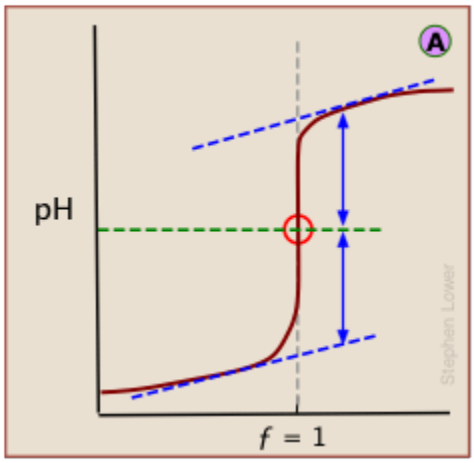
Plotting the pH after each volume increment of titrant has been added can yield a titration curve as detailed as desired, but there are better ways of locating the equivalence point. The most common of these is to take the first or second derivatives of the plot: $d(\text{pH})/dV$ or $d^2(\text{pH})/dV^2$ (of course, for finite increments of pH and volume, these terms would be expressed as $\Delta(\text{pH})/\Delta V$ and $\Delta^2(\text{pH})/\Delta V^2$.)



A **second-derivative** curve locates the inflection point by finding where the *rate* at which the pH changes is zero.



The **differential plot**, showing rate-of-change of pH against titrant volume, locates the inflection point which is also the equivalence point

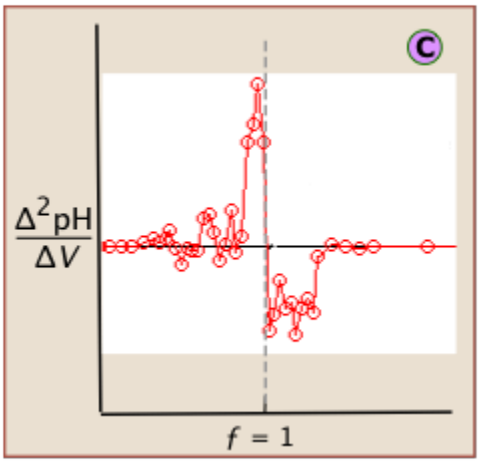


In a **standard plot** of pH-vs-volume of titrant added, the inflection point is located visually as half-way along the steepest part of the curve.

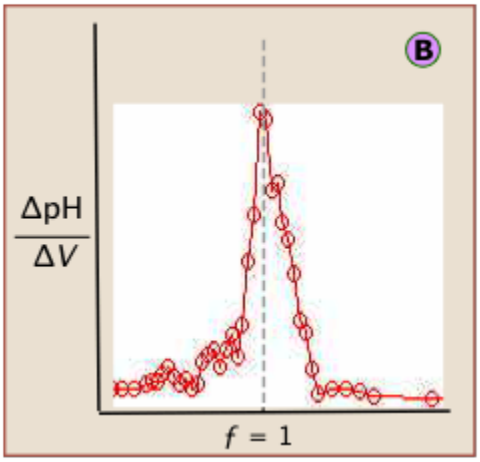
The idealized plots shown above are unlikely to be seen in practice. When the titration is carried out manually, the titrant is added in increments, so even the simple titration curve must be constructed from points subject to uncertainties in volume

measurement and pH (especially if the latter is visually estimated by color change of an indicator.)

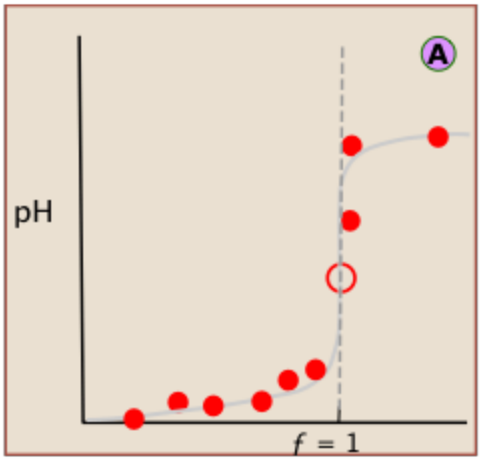
If this data is then converted to differential form, these uncertainties add a certain amount of "noise" to the data.



A second-derivative plot uses pH readings on both sides of the equivalence point, making it easier to locate in the presence of noise.



Locating the equivalence point depends very strongly on correct reading of only one or two pH readings near the top of the plot.



A simple curve, plotted from a small number of pH readings, will not always unambiguously locate the equivalence point.

The "noise" in differential plots can usually be minimized by keeping the titrant and analyte concentrations above 10^{-3} *M*.

Some other ways of following a titration

Monitoring the pH by means of an indicator or by potentiometry as described above are the standard ways of detecting the equivalence point of a titration. However, we have already seen that in certain cases involving polyprotic acids or bases, some of the equivalence points are obscured by their close proximity to others, or by the buffering that occurs near the extremes of the pH range. Similar problems can arise when the solution to be titrated contains several different acids, as often happens when fluids connected with industrial processes must be monitored.

Take the temperature!

Acid-base neutralization reactions HA + B → A^- + BH^+ are always exothermic; when protons fall from their level in the acid to that in the base, most of the free energy drop gets released as heat. If the acid and base are both strong (i.e., totally dissociated), the *enthalpy of neutralization* for the reaction

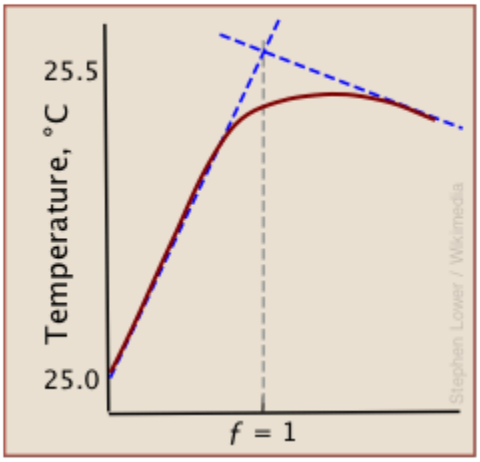
$$\mathrm{H_3O^+ + OH^- \longrightarrow 2\,H_2O} \tag{13.5.4}$$

is -68 kJ mol^{-1}.

See this Wikipedia page for more on thermometric titrations, including many examples. Note also the *video* on this topic in the "Videos" section near the end of this page.

- Thermometric titrations are not limited to acid-base determinations; they can also be used to follow precipitation-, complex formation-, and oxidation-reduction reactions.
- They can be used with polyprotic acids or bases, and with mixtures containing more than one acid or base.
- They are able to follow acid-base titrations that must be carried out in non-aqueous solvents, where other titration methods are not possible.

A typical thermometric titration curve consists of two branches, beginning with a steep rise in temperature as the titrant being added reacts with the analyte, liberating heat. Once the equivalence point is reached, the rise quickly diminishes as heat production stops. Then, as the mixture begins to cool, the plot assumes a negative slope.



Although a rough indication of the equivalence point can be estimated by extrapolating the linear parts of the curve (blue dashed lines), the differential methods described above are generally preferred.

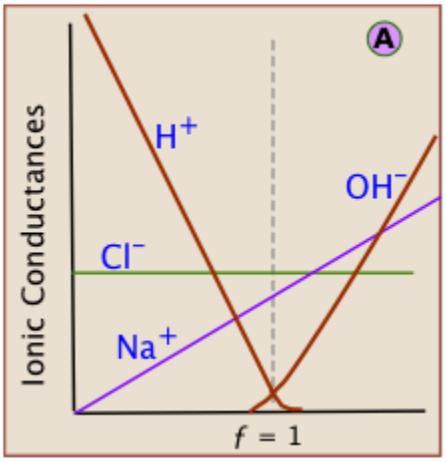
Follow the electrolytic conductance

Acids and bases are electrolytes, meaning that their solutions conduct electric current. The conductivity of such solutions depends

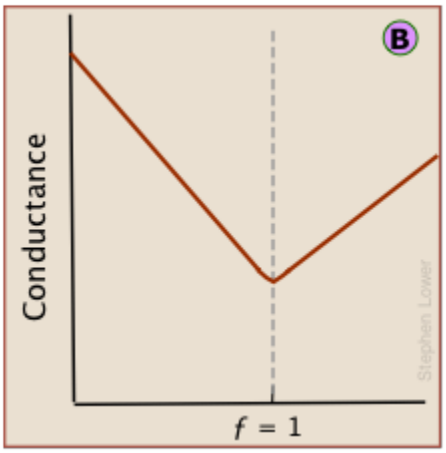
on the concentrations of the ions, and to a lesser extent, on the nature of the particular ions. Any chemical reaction in which there is a change in the total quantity of ions in the solution can usually be followed by monitoring the conductance. Acid-base titrations fall into this category. Consider, for example, the titration of hydrochloric acid with sodium hydroxide. This can be described by the equation

$$\mathrm{H^+ + Cl^- + Na^+ + OH^- \longrightarrow H_2O + Na^+ + Cl^-} \tag{13.5.5}$$

which shows that two of the four species of ions being combined disappear at the equivalence point. During the course of the titration, the conductance of the solution falls as H^+ and Cl^- ions are consumed. At the equivalence point the conductance passes through a minimum, and then rises as continued addition of titrant adds more Na^+ and OH^- ions to the solution.



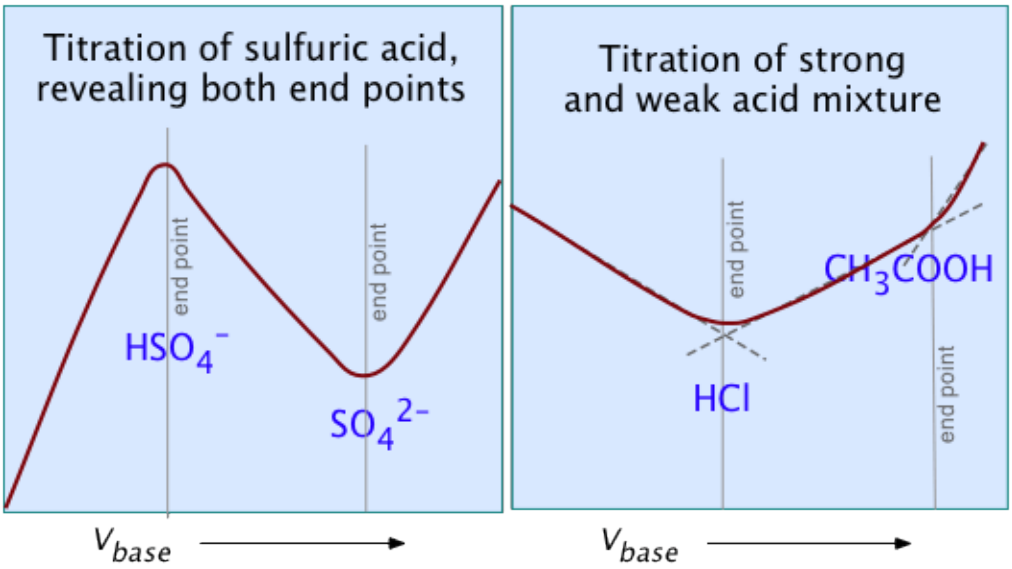
Each kind of ion makes its own contribution to the solution conductivity. If we could observe the contribution of each ion separately, see that the slopes for H^+ and OH^- are much greater. This reflects the much greater conductivities of these ions owing to their uniquely rapid movement through the solution by hopping across water molecules.



However, because the conductances of individual ions cannot be observed directly, conductance measurements always register the total conductances of *all* ions in the solution. The change in conductance that is actually observed during the titration of HCl by sodium hydroxide is the sum of the ionic conductances shown above.

For most ordinary acid-base titrations, conductimetry rarely offers any special advantage over regular volumetric analysis using indicators or potentiometry. This is especially true if the acid being titrated is weak; if the pK_a is much below 2, the rising salt line (Na^+ when titrating with NaOH) will overwhelm the fall in the contribution the small amount of H^+ makes to the conductance, thus preventing any minimum in the total conductance curve from being seen.

However, in some special cases such as those illustrated below, conductimetry is the only method capable of yielding useful results.



These examples illustrate two unique capabilities of conductimetric titrations: (left) Titration of a mixture of two acids and (right) Titration of a strong polyprotic acid →

Because pure H_2SO_2 is a neutral molecule, it makes no contribution to the conductance, which rises to a maximum at the equivalence point.

In four years of college lab sessions, many Chemistry majors will likely carry out fewer than a dozen titrations. However, in the real world, time is money, and long gone are the days when technicians were employed full time just to titrate multiple samples in such enterprises as breweries, food processing (such as blending of canned orange juice), clinical labs, and biochemical research.

Titration calculations

A titration is carried out by adding a sufficient volume \Vo of the titrant solution to a known volume \Vt of the solution being titrated. This addition continues until the *end point* is reached. The end point is our experimental approximation of the *equivalence point* at which the acid-base reaction is stoichiometrically complete (ƒ = 1). The quantity we actually measure at the end point is the volume V_ep of titrant delivered to the solution undergoing titration.

The solution being titrated is often referred to as the *analyte* (the substance being "analyzed") or, less commonly, as the *titrand*. We shall employ the latter term in what follows.

In a simple titration of a monoprotic acid HA by a base B, the equivalence point corresponds to completion of the reaction

$$\mathrm{HA + B \longrightarrow A^- + BH^+} \tag{13.5.6}$$

where equimolar quantities of HA and B have been combined; that is,

$M_{HA} = M_B$ (where M represents the number of moles.)

Recall that the number of moles is given by the product of the volume and concentration:

$$\mathrm{L \times mol\ L^{-1} = mol,}$$

$$\mathrm{mL \times mMol\ ml^{-1} = mmol.}$$

Because we are measuring the *volume* of titrant rather than the number of moles, we need to use its *concentration* to link the two quantities. So if we are titrating the base B with acid HA, the end point is reached when a volume V of HA has been added. The

number of moles of HA we have added at the end point is given by the product of its volume and concentration

$$M_{HA} = V_{HA} \times C_{HA} \tag{13.5.7}$$

And because the reaction 13.5.6 is now complete, $M_{HA} = M_B$; thus, **at the end point of any monoprotic titration,**

$$V_{HA}\, C_{HA} = V_B\, C_B \tag{13.5.8}$$

Equation 13.5.8 is important! In any titration, both the volume and the concentration of the titrant are known, so the unknown concentration is easily calculated.

In titrations carried out in the laboratory, the titrant is delivered by a burette that is usually calibrated in milliliters, so it is more convenient to express M_{HA} in millimoles and C_{HA} in millimoles/mL (mMol ml^{-1}); note that the latter is numerically the same as moles/L.

Example 13.5.1

50.0 mL of 0.100 M hydrochloric acid is titrated with 0.200 M sodium hydroxide. What volume of NaOH solution will have been added at the equivalence point?

Solution

First, find the number of moles of HCl in the titrand:

$$\begin{aligned} M_{\mathrm{HCl}} &= C_{\mathrm{HCl}} \times V_{\mathrm{HCl}} \\ &= (0.100\, mMol\, mL^{-1}) \times (50.0\, mL) \\ &= 5.0\, mMol\, \mathrm{HCl} \end{aligned}$$

This same number of moles of NaOH must be delivered by the burette in order to reach the equivalence point (*i.e.*, M_{NaOH} = 5.0 mMol.)

The volume of NaOH solution M_{NaOH} required is

$$V_{NaOH} = (M_{NaOH} \,/\, C_{NaOH}) = (5.0 \text{ mMol}) \,/\, (0.200 \text{ mMol/mL}) = \mathbf{25\ mL}$$

This page titled 13.5: Acid/Base Titration is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13.6: Applications of Acid-Base Equilibria

Learning Objectives

Make sure you thoroughly understand the following essential concepts:

- Describe the major features of the *global carbon cycle*, and outline the role of acid-base processes in both the slow and fast parts of the cycle.
- Describe the principal acid-base reactions in *aerobic respiration*.
- Define *acid rain*, and explain its origins. Outline its major effects on the environment.

Acid-base reactions pervade every aspect of industrial-, physiological-, and environmental chemistry. In this unit we touch on a few highlights that anyone who studies or practices chemical science should be aware of.

Some applications of Buffers

Buffer solutions and the buffering effect they produce are extremely important in many practical applications of chemistry. The reason for this is that many chemical processes are quite sensitive to the pH; the extent of the reaction, its rate, and even the nature of the products can be altered if the pH is allowed to change. Such a change will tend to occur, for example, when the reaction in question, or an unrelated parallel reaction, consumes or releases hydrogen- or hydroxide ions.

Buffers in biochemistry and physiology

Many reactions that take place in living organisms fall into this category. Most biochemical processes are catalyzed by enzymes whose activities are highly dependent on the pH; if the local pH deviates too far from the optimum value, the enzyme may become permanently deactivated.

- For example, the oxidation of glucose to carbon dioxide and water — the fundamental energy-producing process of aerobic metabolism — releases H^+ ions. The multiple steps of this process are catalyzed by enzymes whose activity is strongly affected by the pH. If the region of the cell where these reactions take place were not buffered, the very act of living would soon reduce the local pH below the range in which the enzymes are active, halting energy production and killing the cell.
- Blood is strongly buffered (mainly by bicarbonate) to maintain its pH at 7.4±0.3; pH values below 7.0 or above 7.8 cause death within minutes.
- The interiors of most cells are buffered near pH 7.0 (by phosphates and proteins)

Buffers in industry

Buffers are employed in a wide variety of laboratory procedures and industrial processes:

- solutions for the calibration of pH meters
- growth of bacteria in culture media
- fermentation processes, including winemaking and the brewing of beer
- controlling the colors of dyes used in coloring fabrics
- shampoos (maintaining pH at or below 7 to conteract the alkalinity of soaps that can cause irritation)
- other personal care products such as baby lotions (keeping the pH around 6 to inhibit the growth of bacteria)
- pH control in printing inks to assure their proper penetration into the paper

Buffers in the environment

- Natural waters (lakes and streams) able to support aquatic life are buffered by the sediments they are in contact with. This buffering is not always able to offset the effects of acid rain.
- Seawater is buffered mainly by bicarbonate and borates. This allows the oceans to absorb about half of the CO_2 emitted to

the atmosphere by human activities.
- Soils, in order to remain productive must maintain a near-neutral pH, not only for plants but also for microorganisms which fix- and recycle nitrogen. Soils rendered highly acidic by acid rain can keep essential nutrients such as phosphates in forms such as insoluble $Ca_3(PO_4)_2$ that prevents their uptake by plants.

Acid-base chemistry in physiology

Acid-base chemistry plays a crucial role in physiology, both at the level of the individual cell and of the total organism. The reasons for this are twofold:

- Many of the major chemical components of an organism can themselves act as acids and/or bases. Thus proteins contain both acidic and basic groups, so that their shapes and their functional activities are highly dependent on the pH.
- Virtually all important metabolic processes involve the uptake or release of hydrogen ions. The very act of being alive tends to change the surrounding pH (usually reducing it); this will eventually kill the organism in the absence of buDering mechanisms.

About two-thirds of the weight of an adult human consists of water. About two-thirds of this water is located within cells, while the remaining third consists of extracellular water, mostly interstial fluid that bathes the cells, and the blood plasma. The latter, amounting to about 5% of body weight (about 5 L in the adult), serves as a supporting fluid for the blood cells and acts as a means of transporting chemicals between cells and the external environment. It is basically a 0.15 M solution of NaCl containing smaller amounts of other electrolytes, the most important of which are HCO_3^- and protein anions.

Respiration, the most important physiological activity of a cell, is an acid-producing process. Carbohydrate substances are broken down into carbon dioxide, and thus carbonic acid:

$$C(H_2O)_n + O_2 \rightarrow CO_2 + nH_2O \tag{13.6.1}$$

Maintenance of acid-base balance

It is remarkable that the pH of most cellular fluids can be kept within such a narrow range, given the large number of processes that tend to upset it. This is due to the exquisite balance between a large number of interlinked processes operating at many different levels. Acid-base balance in the body is maintained by two general mechanisms: selective excretion of acids or bases, and by the buffering action of weak acid-base systems in body fluids.

- Over a 24-hour period, the adult human eliminates the equivalent of about 20-40 moles of H^+ by way of the lungs in the form of CO_2. In addition, the kidneys excrete perhaps 5% of this amount of acid, mostly in the form of $H_2PO_4^-$ and NH_4^+. Owing to their electric charges, these two species are closely linked to salt balance with ions such as Na^+ or K^+ and Cl^-.
- The major buffering system in the body is the carbonate system, which exists mainly in the form of HCO_3^- at normal physiological pH. Secondary buffering action comes from phosphates, from proteins and other weak organic acids, and (within blood cells), the hemoglobin.

There is a huge industry aimed at the notoriously science-challenged "alternative health" market that promotes worthless nostrums such as "ionized water" that are claimed to restore or preserve "the body's acid-base balance". They commonly point out that most foods are "acidic", (i.e., they are metabolized to H_2CO_3), but they never explain that most of this acid is almot immediately exhaled in the form of CO_2. The implication is that our health is being ruined by the resulting "acidification" of the body; some further imply this can be a cause of cancer and other assorted ailments.

People who fall for these expensive scams are in effect paying a tax on scientific ignorance. Those who, like you, have studied Chemistry, should consider that they have a social duty to counter this kind of predatory and deceptive pseudoscience.

Disturbances of acid-base balance

Deviations of the blood plasma pH from its normal value of 7.4 by more than about ±.1 can be very serious. These conditions are known medically as *acidosis* and *alkalosis*. They can be caused by metabolic disturbances such as diabetes and by kidney failure

in which excretion of $H_2PO_4^-$, for example, is inhibited.

Numerous other processes lead to temporary unbalances. Thus hyperventilation, which can result from emotional upset, leads to above-normal loss of CO_2, and thus to alkalosis. Similarly, hypoventilation can act as a compensatory mechanism for acidosis. On the other hand, retention of CO_2 caused by bronchopneumonia, for example, can give rise to acidosis. Acidosis can also result from diarrhea (loss of alkaline fluid from the intestine), while loss of gastric contents by vomiting promotes alkalosis.

Acid rain

The atmosphere is naturally acidic

The natural, unpolluted atmosphere receives acidic, basic, and neutral substances from natural sources (volcanic emissions, salt spray, windblown dust and microbial metabolism) as well as from pollution (Figure 11.6.X). These react in a kind of gigantic acid-base titration to give a solution in which hydrogen ions must predominate to maintain charge balance (indicated by the identical widths of the bar graphs at the bottom labeled "anions" and "cations").

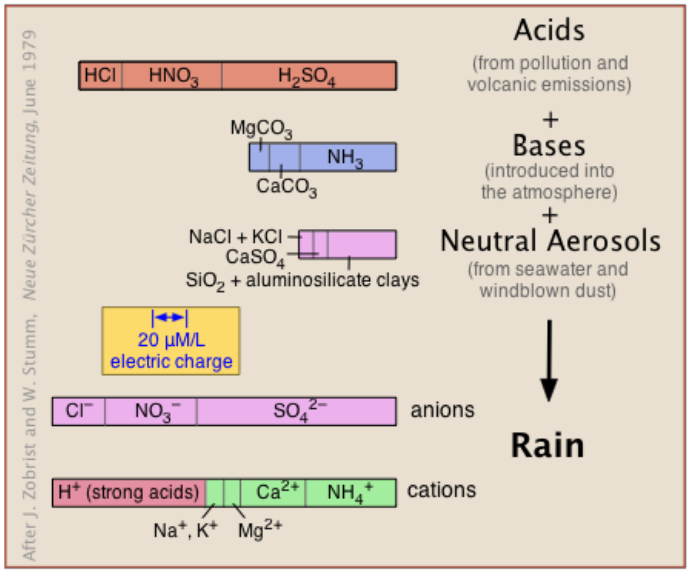


Figure 13.6.1

Carbon dioxide, however, is the major source of acidity in the unpolluted atmosphere. As will be explained farther on, CO_2 makes up 0.032 vol-% of dry air, and dissolves in water to form carbonic acid:

$$CO_{2(g)} + H_2O(l) \rightleftharpoons H_2CO_{3(aq)} \tag{13.6.2}$$

Thus *all* rain is “acidic” in the sense that it contains dissolved CO_2 which will reduce its pH to 5.7. The term *acid rain* is therefore taken to mean rain whose pH is controlled by pollutants which can lower the pH into the range of 3-4, resulting in severe damage to the environment.

Origin of acid rain

Acid rain originates from emissions of SO_2 and various oxides of nitrogen (known collectively as "NO_x") that are formed during combustion processes, especially those associated with the burning of coal. Incineration of wastes, industrial operations such as smelting, and forest fires are other combustion-related sources; these also release particulate material into the atmosphere which plays an important role in concentrating and distributing these acid-forming substances.

Origins of NO_x

Nitrogen oxides are formed naturally by soil bacteria acting on nitrate ions, an essential plant nutrient and itself a product of natural nitrogen fixation. Ammonia, a product of bacterial decomposition of organic matter, is eventually transformed into NO_3^- by bacteria into nitrogen oxides.

The quantity of fixed nitrogen produced industrially to support intensive agriculture now exceeds the amount formed naturally, and has become the major source of anthropogenic nitrogen oxides.

- Atmospheric nitrogen can be thermally decomposed to its various oxides NO_x when subjected to high temperatures; one minor natural source is lightning. More significantly, these oxides are also formed when "clean" fuels such as natural gas (mainly methane, CH_4) is burnt in air.
- NO_x can also react with other atmospheric pollutants to produce *photochemical smog*.

Formation of acids

The gases noted above react with atmospheric oxygen, with each other, and with particulate materials to form sulfuric, nitric, and hydrochloric acids.

- Most of the **H_2SO_4** comes from the photooxidation of SO_2 released from the burning of fossil fuels and from industrial operations such as smelting.
- **Nitric acid** similarly results from photooxidation of nitrogen oxides. Large quantities of these oxides are formed in high-temperature combustion processes in gasoline- and jet engines, and in the turbine engines of fossil-fueled electric power plants.
- **Hydrochloric acid** is formed naturally through the reaction of *sea salt aerosols* with atmospheric H_2SO_4 and HNO_3. The salt particles form when ocean waves break the surface and some of the water evaporates before the spray settles back to the surface. On a much smaller scale, fossil fuel combustion and waste incineration are believed to the principal human-caused sources of HCl.

Wet and dry deposition

There are two basic mechanisms by which acidic material is transported through the environment. The acid that dissolves in the water droplets that form clouds and precipitation, and is eventually incorporated into fog and snowflakes, undergoes *wet deposition*; this is ordinarily what is referred to as "acid rain". Because it is widely dispersed, often high in the atmosphere, it can travel very large distances from the original source.

About 40% of the acidic substances introduced into the atmosphere becomes adsorbed onto particulate matter, which can be soot, smoke, windblown dust, and salt particles formed naturally from sea spray. This spreads less widely (typically around 30 km from the source), and falls onto surfaces by "*dry deposition*". Once the acid-laden particles land on surfaces, ordinary rain, fog, and dew release the acidic components, often in considerably more concentrated form than occurs through dry deposition alone.

Effects on soils

Because soils support the growth of plants and of the soil microorganisms that are essential agents in the recycling of dead plant materials, acid rain has a indirect but profound effect on soil health and plant growth.

Soils containing alkaline components (most commonly limestone $CaCO_3$ and other insoluble carbonates) can neutralize added acid and mitigate its effects. But soils in high mountain regions tend to be thin and unable to provide adequate buffering capacity. The same is true in almost half of Canada, in which granite rock of the Canadian Shield is very close to the surface; the eastern provinces of the country are strongly impacted by acid rain.

- Most soils also contain clay and humic substances that bind and retain ions such as Ca^{2+}, Mg^{2+} and K^+ which are essential plant nutrients. Added H^+ binds even more strongly to these substances, displacing the nutrient ions from the upper layers. The overall effect is to leach these ions to greater depths where they may be inaccessible to plants, or to wash them away into ground water.

- The SO_4^{2-} component of acid rain converts some nutrient cations into insoluble sulfates, reducing their availability to plants.
- Clays, which are complex aluminosilicates, gradually break down, releasing aluminum ions which normally form insoluble $Al(OH)_3$. Addition of H^+ dissolves this hydroxide, raising the concentration of $Al(H_2O)_6^{3+}$ to levels that can be toxic to plants.
- Many toxic heavy metals such as Pb, Cd and Cr are present in trace amounts in soils but are tied up in the form of insoluble salts. Acid rain can mobilize the ions of these metals, much as happens with aluminum.

Effects on plants



↓ Click on image to enlarge
Leaf damaged by acid rain [link]



↓ Click on image to enlarge
Forest in the U.S Andirondacks damaged by acid rain
[link]

The effects on soils noted above affect plants most strongly. However, direct impingment of acidic rain and fog on leaves has other effects which can be especially serious when air pollutants such as SO_2 are present.

- Acid deposition on leaves and needles tends to weaken them by eroding away their protective waxy coatings. This often results in the development of brown spots that interfere with photosynthesis.
- Forests in mountain regions are frequently bathed in clouds and fog which can be even more acidic than normal precipitation, exacerbating the above problem. In addition, the leaves and needles tend to accumulate the fog aerosol into larger droplets. These eventually fall to the forest floor, magnifying the effects on soils described in the previous section.
- Food crops tend to be less affected by acid rain where good farming practices such as the addition of fertilizers to replace depleted nutrients and the addition of limestone to raise the soil pH.

Effects on lakes and aquatic ecosystems

Lakes are subject not only to wet and dry acidic deposition, but also by the water they receive from streams and surface runoff. Any toxic elements released by the action of acid rain on soils and sediments can thus be conveyed to, and concentrated within lakes and the streams that empty them.



↓ Click on image to enlarge
Regions of poorly-buffered lakes in North America in 1980 [link]



↓ Click on image to enlarge
May Lake in the California Sierras - a typical alpine lake
[link]

Lakes in poorly-buffered areas such as are found in alpine regions (western North America, Colorado and much of Switzerland) or on the Canadian Shield and in the Adirondacs and Appalachians) are highly sensitive to acid deposition.

Severely acidified lakes (such as the one depicted at the above right) can be so devoid of life, including algae, that the water appears to be perfectly clear and bright blue in color.

↓ Click on image to enlarge



pH tolerances of some aquatic organisms
[U.S. EPA]

Aquatic organisms are generally adapted to "ordinary" pH conditions of 6-8, but vary greatly in their tolerance of low pH. As pH declines below 6, the diversity of aquatic animals, plants, and microorganisms diminishes.

- Some plant species such as sphagnum mosses and certain filamentous algaes that thrive at very low pH can proliferate in acidified lakes, producing thick mats that seal off oxygen and thus inhibit the decay of litter on the lake floor.
- Acid deposited onto winter snow is released during spring melting and can cause rapid drops of pH in poorly-buffered lakes that receive waters from these sources. This "spring acid shock" can seriously affect viability of organisms such as fish, amphibians, and insects that lay their eggs in the water which hatch in the spring. The hatchlings are often unable to adapt to the rapid change, and end up deformed or killed.

- ↓ Click on image to enlarge



Fish damaged by acid deposition [link]

Even those aquatic species able to survive at in low-pH waters can be indirectly affected if their food supply is pH-limited. For example, mayflies and some other insects, which, which are important food sources for frogs, cannot survive below pH 5.5.

- Fish are severely affected by the aluminum that is released from the action of acid on sediments; it causes a coating of mucus to form on the gills, impeding absorption of oxygen. This has led to the extinction of some species from affected lakes.
- Low pH increases the solubility of calcium salts, making it more difficult for bone to form in developing embryos of fish and amphibians.
- Low pH can also make the eggs of aquatic organism thicker and more difficult for embryos to penetrate, thus delaying hatching. When the embryos continue to grow within the confined space of the egg, their spines can be deformed, interfering with their viability when they finally hatch.

Effects on buildings and statues

Acid rain is not new!

"It has often been observed that the stones and bricks of buildings, especially under projecting parts, crumble more readily in large towns where coal is burnt... I was led to attribute this effect to the slow but constant action of acid rain." Robert Angus Smith, 1856

Acid deposition most strongly affects heritage buildings made of limestone and similar carbonate-containing stone materials.

The principal chemical process involves the reaction of sulfuric acid with calcium carbonate:

$$\mathrm{CO_3\,(s) + H_2SO_4 \longrightarrow CaSO_4\,(s) + CO_2}. \tag{13.6.3}$$

The acid first erodes and breaks up the surface of the stone. As the hydrated calcium sulfate (gypsum) forms, it picks up iron and other components of the stone and forms an unsightly black coating. Some of this gradually blisters off, exposing yet more stone

suface. The gypsum crystals can sometimes grow into the stone, further undermining the surface.

Very old structures such as the Taj Mahal, Notre Dame, the Colosseum and Westminster Abbey have all been affected.



↓ Click on image to enlarge
Stone buildings and monuments are rapidly damaged by acid rain [link]



The exposed upper survaces of the gargoyle are beginning to deteriorate.
[link]

Statues and monuments, including those made from marble, are also susceptible to erosion and damage from acid deposition.

Modern buildings are less affected, although acid rain can erode some painted surfaces.

The Geochemical Carbonate system

The carbonate system, consisting as it does of a soluble gas CO_2, soluble ions HCO_3^- and CO_3^{2-}, and sparingly soluble carbonate salts, spans all four realms of nature: the atmosphere, hydrosphere (mainly the oceans), the lithosphere, and the biosphere. And owing to the acid-base equilibria that govern the transformations between these carbonate species, carbon is readily transported between these geochemical reservoirs. This chapter presents an overview of carbon geochemistry.

Distribution of carbon on Earth

Carbon is the fourth most abundant element in the universe. Within the Earth's crust it ranks 15th, mostly in the form of carbonates in limestones and dolomites. Kerogens, which are fossilized plant-derived carbon mostly in the form of oil shale, constitute another major repository of terrestrial carbon.

Source	relative to atmosphere

carbonate rock	carbonate rock	28,500
fossil carbon	572	10,600
land - organic carbon	0.065	1.22
ocean	3.2	61.8
atmospheric CO_2	0.0535	**1.0**

The geochemical carbon cycle

The carbonate system encompasses virtually all of the environmental compartments– the atmosphere, hydrosphere, biosphere, and major parts of the lithosphere. The complementary processes of photosynthesis and respiration drive a global cycle in which carbon passes slowly between the atmosphere and the lithosphere, and more rapidly between the atmosphere and the oceans. Thus the "carbon cycle" can be divided into "fast" and "slow" parts, operating roughly on annual and geological time scales.

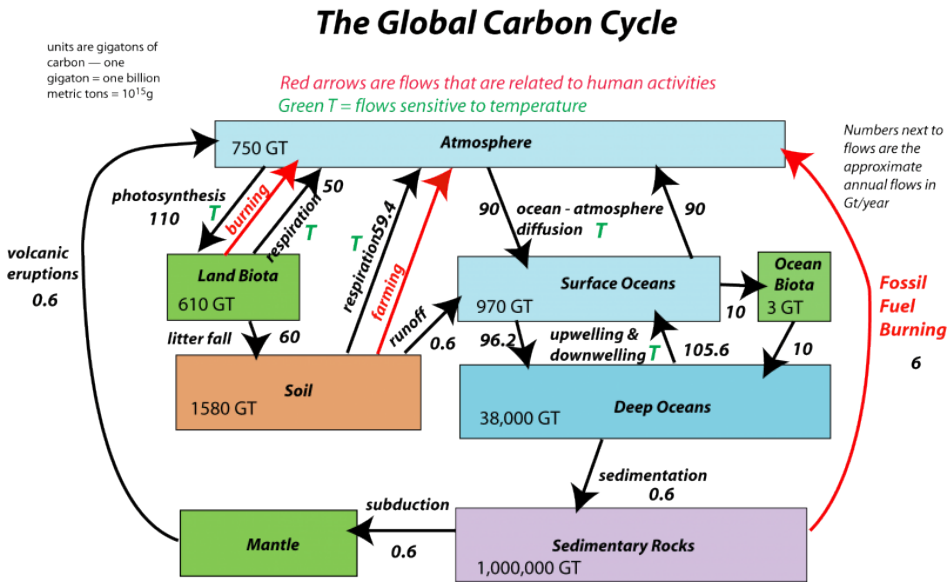


Figure ({13{.}6{.}\eatSpaces{2} \endSpaces}\): *Image by David Bise, Pennsylvania State University; see here for a full discussion of this diagram.*

Another excellent depiction of the carbon cycle, by the U.S. Dept of Energy. (See description):



Carbon Cycle

Carbon dioxide in the atmosphere

CO_2 has probably always been present in the atmosphere in the relatively small absolute amounts now observed. Precambrian limestones, possibly formed by reactions with rock-forming silicates, e.g.

$$\mathrm{CaSiO_3 + CO_2 \longrightarrow CaCO_3 + SiO_2} \tag{13.6.4}$$

have likely had a moderating influence on the CO_2 abundance throughout geological time.

The volume-percent of CO_2 in dry air is 0.032%, leading to a partial pressure of 3×10^{-4} ($10^{3.5}$) atm. In a crowded and poorly-ventilated room, P_{CO2} can be as high as 100×10^{-4} atm. About 54×10^{14} moles per year of CO_2 is taken from the atmosphere by photosynthesis, divided about equally between land and sea. Of this, all except 0.05% is returned by respiration (mostly by microorganisms); the remainder leaks into the slow, sedimentary part of the geochemical cycle where it can remain for thousands to millions of years. Trends in the growth of atmospheric CO_2 for the past five years can be seen on this US NOAA page.

Since the advent of large-scale industrialization around 1860, the amount of CO_2 in the atmosphere has been increasing. Most of this has been due to fossil-fuel combustion; in 1966, about 3.6×10^{15} g of C was released to the atmosphere; this is about 12 times greater than the estimated natural removal of carbon into sediments. The large-scale destruction of tropical forests, which has accelerated greatly in recent years, is believed to exacerbate this effect by removing a temporary sink for CO_2.

About 30-50% of the CO_2 released into the atmosphere by combustion remains there; the remainder enters the hydrosphere and biosphere. The oceans have a large absorptive capacity for CO_2 by virtue of its transformation into bicarbonate and carbonate in a slightly alkaline aqueous medium, and they contain about 60 times as much inorganic carbon as is in the atmosphere. However, effcient transfer takes place only into the topmost
(100 m) wind-mixed layer, which contains only about one atmosphere equivalent of CO_2; mixing time into the deeper parts of the ocean is of the order of 1000 years. For this reason, only about ten percent of added CO_2 is taken up by the oceans.

Most of the carbon in the oceans is in the form of bicarbonate, as would be expected from the pH which ranges between 7.8 and 8.2. In addition to atmospheric CO_2, there is a carbonate input to the ocean from streams. This is mostly in the form of HCO_3^- which derives from the weathering of rocks and terrestrial carbonate sediments, and the acid-base reaction

$$H_2CO_3 + CO_3^{2-} \rightarrow 2HCO_3^- \tag{13.6.5}$$

which can be considered to be the source of bicarbonate in seawater. In this sense, the ocean can be considered the site of a gigantic acid-base titration in which atmospheric acids (mainly CO_2 but also SO_2, HCl, and other acids of volcanic origin) react with bases that originate from rocks and are introduced through carbonate-bearing streams or in windblown dust.

Carbon in the hydrosphere

Dissolution of CO_2 in water

Carbon dioxide is slightly soluble in water:

°C	0	4	10	20
mol L^{-1}	0.077	0.066	0.054	0.039

Henry's law is followed up to a CO_2 pressure of about 5 atm:

$$[\mathrm{CO_2\,(aq)}] = 0.032 P_{\mathrm{CO_2}} \tag{13.6.6}$$

“Dissolved carbon dioxide” consists mostly of the hydrated oxide CO2(aq) together with a small proportion of carbonic acid:

$$[\mathrm{CO_2\,(aq)}] = 650[\mathrm{H_2CO_3}] \tag{13.6.7}$$

The acid dissociation constant K_{a1} that is commonly quoted for "H_2CO_3" is really a *composite* equilibrium constant that includes the above equilibrium. This means that "pure" H_2CO_3 (which cannot be isolated) is a considerably stronger acid than is usually appreciated.

Distribution of carbonate species in aqueous solutions

Water exposed to the atmosphere with P_{CO2} = $10^{3.5}$ atm will take up carbon dioxide, which becomes distributed between the three carbonate species CO_2, HCO_3^- and CO_3^{2-} in proportions that depend on K_1 and K_2 and on the pH. The "total dissolved carbon" is given by the mass balance

$$C_t = [H_2CO_3] + [HCO_3^-] + [CO_3^{2-}] \tag{13.6.8}$$

The distribution of these species as a function of pH can best be seen by constructing a log *C*-pH diagram for C_t = 10^{-5} *M*.

This Sillén plot is drawn for two different concentrations of the carbonate system. The lower one, in heavier lines, is for a 10^{-5} *M* solution, corresponding roughly to atmospheric CO_2 in equilibrium with pure water. The upper plot, for a 10^{-3} *M* solution, is representative of many natural waters such as lakes and streams where the water is in contact with sediments.

It is important to note that this diagram applies only to a system in which C_t is constant. In a solution that is open to the atmosphere, this will not be the case at high pH values where the concentration of CO_2 is appreciable. Under these conditions this ion will react with H_2CO_3 and the solution will absorb CO_2 from the atmosphere, eventually resulting in the formation of a solid carbonate precipitate.

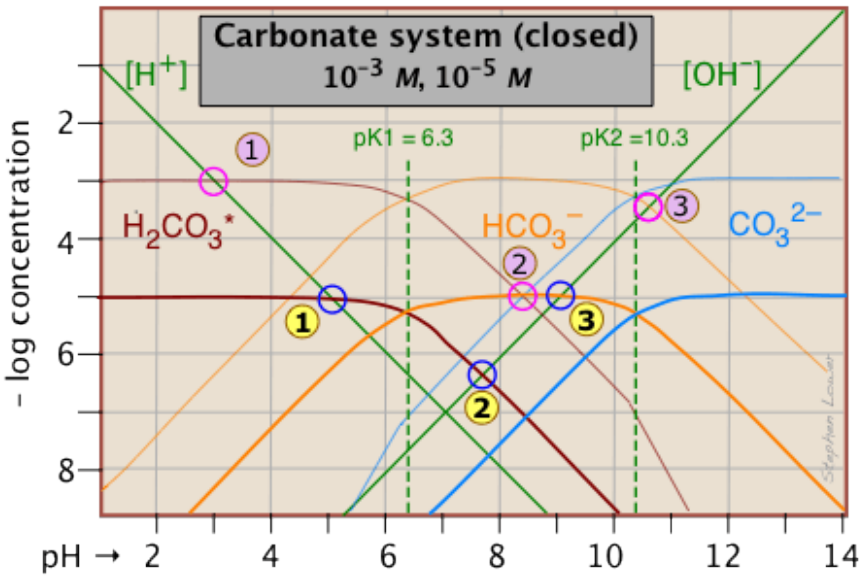


Figure ({13{.}6{.}\eatSpaces{3} \endSpaces}\): Sillén plot for carbonate in water.

The lower of the two above plots can be used to predict the pH of 10^{5} *M* solutions of carbon dioxide, sodium bicarbonate, and sodium carbonate in pure water. The pH values are estimated by using the mass and charge balance conditions for each solution as noted below.

The reasoning leading to these calculations is explained in the discussion of the 10^{-3} *M* carbonate system in a previous chapter.

Solution of CO_2 or H_2CO_3

$[H^+] = [OH^-] + [HCO_3^-] + 2\,[CO_3^{2-}]$(4-6)

which, since the solution will be acidic, can be simplified to

$[H^+] = [HCO_3^-]$ (point 3)(4-7)

Solution of $NaHCO_3$

$[H^+] + [H_2CO_3] = [CO_3^{2-}] + [OH]$(4-8)

or

$[H^+] = [HCO_3^-]$ (point 3) (4-9)

Solution of Na_2CO_3

$[H^+] + 2\,[H_2CO_3] + [CO_3^{2-}] = [OH^-]$(4-10)

or

$[HCO_3^-] = [OH^-]$ (point 3) (4-11)

Carbon in the oceans

Natural waters aquire carbon from sediments they are in contact with, and of course also from the atmosphere. The pH is an important factor here; CO_2 and solid carbonates are more soluble at high pH, and pH also controls the distribution of carbon species, as is seen in the plot just above.

Carbon content and pH of typical natural waters

rain	river/lake water	sea water	
ppm of carbon	1.2	35	140
pH (unpolluted)	5.6	6.5 - 8.5	7.5 - 8.4

At the slightly alkaline pH of most bodies of water (of which the oceans constitute 97% of the earth's surface waters), bicarbonate is the principal dissolved carbon species. The quantity of organic carbon is fairly small.

Carbon content of the oceans , moles × 1018

CO_2(aq)	HCO_3^-	CO_3^{2-}	dead org.	living org.
0.18	2.6	0.33	0.23	0.05

This page titled 13.6: Applications of Acid-Base Equilibria is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

13.7: Exact Calculations and Approximations

Learning Objectives

- Understand the exact equations that are involves in complex acid-base equilibria in aqueous solutions

The methods for dealing with acid-base equilibria that we developed in the earlier units of this series are widely used in ordinary practice. Although many of these involve approximations of various kinds, the results are usually good enough for most purposes. Sometimes, however — for example, in problems involving very dilute solutions, the approximations break down, often because they ignore the small quantities of H^+ and OH^- ions always present in pure water. In this unit, we look at exact, or "comprehensive" treatment of some of the more common kinds of acid-base equilibria problems.

Strong acid-base systems

The usual definition of a “strong” acid or base is one that is completely dissociated in aqueous solution. Hydrochloric acid is a common example of a strong acid. When HCl gas is dissolved in water, the resulting solution contains the ions H_3O^+, OH^-, and Cl^-, but except in very concentrated solutions, the concentration of HCl is negligible; for all practical purposes, molecules of “hydrochloric acid”, HCl, do not exist in dilute aqueous solutions. To specify the concentrations of the three species present in an aqueous solution of HCl, we need three independent relations between them.

These relations are obtained by observing that certain conditions must **always** hold for aqueous solutions:

1. The dissociation equilibrium of water must always be satisfied:
$$[H^+][OH^-] = K_w \tag{13.7.1}$$
2. The undissociated acid and its conjugate base must be in *mass balance*. The actual concentrations of the acid and its conjugate base can depend on a number of factors, but their sum must be constant, and equal to the “nominal concentration”, which we designate here as C_a. For a solution of HCl, this equation would be
$$[HCl] + [Cl^-] = C_a \tag{13.7.2}$$
Because a strong acid is by definition completely dissociated, we can neglect the first term and write the mass balance condition as
$$[Cl^-] = C_a \tag{13.7.3}$$
3. In any ionic solution, the sum of the positive and negative electric charges must be zero; in other words, all solutions are electrically neutral. This is known as the *electroneutrality principle*.
$$[H^+] = [OH^-] + [Cl^-] \tag{13.7.4}$$

The next step is to combine these three limiting conditions into a single expression that relates the hydronium ion concentration to C_a. This is best done by starting with an equation that relates several quantities and substituting the terms that we want to eliminate. Thus we can get rid of the $[Cl^-]$ term by substituting Equation 13.7.3 into Equation 13.7.4 :

$$[H^+] = [OH^-] + C_a \tag{13.7.5}$$

The $[OH^-]$ term can be eliminated by the use of Equation 13.7.1:

$$[H^+] = C_a + \frac{K_w}{[H^+]} \tag{13.7.6}$$

This equation tells us that the hydronium ion concentration will be the same as the nominal concentration of a strong acid as long as the solution is not very dilute. As the acid concentration falls below about 10^{-6} M, however, the second term predominates;

$[H^+]$ approaches $\sqrt{K_w}$ or $10^{-7}M$ at 25 °C. The hydronium ion concentration can of course never fall below this value; no amount of dilution can make the solution alkaline!

> No amount of dilution can make the solution of a strong acid alkaline!

Notice that Equation 13.7.6 is a quadratic equation; in regular polynomial form it would be rewritten as

$$[H^+]^2 - C_a[H^+] - K_w = 0 \tag{13.7.7}$$

Most practical problems involving strong acids are concerned with more concentrated solutions in which the second term of Equation 13.7.7 can be dropped, yielding the simple relation

$$[H^+] \approx [A^-] \tag{13.7.8}$$

In more concentrated solutions, interactions between ions cause their "effective" concentrations, known as their *activities*, to deviate from their "analytical" concentrations. Thus in a solution prepared by adding 0.5 mole of the very strong acid $HClO_4$ to sufficient water to make the volume 1 liter, freezing-point depression measurements indicate that the concentrations of hydronium and perchlorate ions are only about 0.4 *M*. This does not mean that the acid is only 80% dissociated; there is no evidence of $HClO_4$ molecules in the solution. What has happened is that about 20% of the H_3O^+ and ClO_4^- ions have formed ion-pair complexes in which the oppositely-charged species are loosely bound by electrostatic forces. Similarly, in a 0.10 *M* solution of hydrochloric acid, the activity of H^+ is 0.81, or only 81% of its concentration. (See the green box below for more on this.)

Activities are important because only these work properly in equilibrium calculations. The relation between the concentration of a species and its activity is expressed by the *activity coefficient* γ:

$$a = \gamma C \tag{13.7.9}$$

As a solution becomes more dilute, γ approaches unity. At ionic concentrations below about 0.001 *M*, concentrations can generally be used in place of activities with negligible error. Recall that pH is defined as the negative logarithm of the hydrogen ion *activity*, not its concentration.

> Activities of single ions cannot be determined, so activity coefficients in ionic solutions are always the average, or *mean*, of those for all ionic species present. This quantity is denoted as $\gamma_\pm$.

At very high concentrations, activities can depart wildly from concentrations. This is a practical consideration when dealing with strong mineral acids which are available at concentrations of 10 M or greater. In a 12 *M* solution of hydrochloric acid, for example, the mean ionic activity coefficient* is 207. This means that under these conditions with $[H^+]$ = 12, the activity $\{H^+\}$ = 2500, corresponding to a pH of about –3.4, instead of –1.1 as might be predicted if concentrations were being used. These very high activity coefficients also explain another phenomenon: why you can detect the odor of HCl over a concentrated hydrochloric acid solution even though this acid is supposedly "100% dissociated".

At these high concentrations, a pair of "dissociated" ions H^+ and Cl^- will occasionally find themselves so close together that they may momentarily act as an HCl unit; some of these may escape as \(HCl(g)\) before thermal motions break them up again. Under these conditions, "dissociation" begins to lose its meaning so that in effect, dissociation is no longer complete. Although the concentration of \(HCl(aq)\) will always be very small, its own activity coefficient can be as great as 2000, which means that its escaping tendency from the solution is extremely high, so that the presence of even a tiny amount is very noticeable.

Weak monoprotic acids and bases

Most acids are weak; there are hundreds of thousands of them, whereas there are no more than a few dozen strong acids. We can treat weak acid solutions in exactly the same general way as we did for strong acids. The only difference is that we must now include the equilibrium expression for the acid. We will start with the simple case of the pure acid in water, and then go from there to the more general one in which strong cations are present. In this exposition, we will refer to "hydrogen ions" and $[H^+]$ for brevity, and will assume that the acid HA dissociates into H^+ and its conjugate base A^-.

Pure acid in water

In addition to the species H^+, OH^-, and A– which we had in the strong-acid case, we now have the undissociated acid HA; four variables, requiring four equations.

- **Equilibria**

$$[H^+][OH^-] = K_w \tag{13.7.10}$$

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{13.7.11}$$

- **Mass balance**

$$C_a = [HA] + [A^-] \tag{13.7.12}$$

- **Charge balance**

$$[H^+] = [OH^-] + [A^-] \tag{13.7.13}$$

To eliminate [HA] from Equation 13.7.11, we solve Equation 13.7.13 for this term, and substitute the resulting expression into the numerator:

$$K_a = \frac{[H^+]([H^+] - [OH^-])}{C_a - ([H^+] - [OH^-])} \tag{13.7.14}$$

The latter equation is simplified by multiplying out and replacing $[H^+][OH^-]$ with K_w. We then get rid of the $[OH^-]$ term by replacing it with $K_w/[H^+]$

$$[H^+]C_b + [H^+]^2 - [H^+][OH^-] = K_aC_a - K_a[H^+] + K_a[OH^-] \tag{13.7.15}$$

$$[H^+]^2C_b + [H^+]^3 - [H^+]K_w = K_aC_a - K_a[H^+] + \frac{K_aK_w}{[H^+]} \tag{13.7.16}$$

Rearranged into standard polynomial form, this becomes

$$[H^+]^3 + K_a[H^+]^2 - (K_w + C_aK_a)[H^+] - K_a\,K_w = 0 \tag{13.7.17}$$

For most practical applications, we can make approximations that eliminate the need to solve a cubic equation.

Approximation 1: Neglecting Hydroxide Population

Unless the acid is extremely weak or the solution is very dilute, the concentration of OH^- can be neglected in comparison to that of $[H^+]$. If we assume that $[OH^-] \ll [H^+]$, then Equation 13.7.17 can be simplified to

$$K_a \approx \frac{[H^+]^2}{C_a - [H^+]} \tag{13.7.18}$$

which is a quadratic equation:

$$[H^+]^2 + K_a[H^+] - K_aC_a \approx 0 \tag{13.7.19}$$

and thus, from the quadratic formula,

$$[H^+] \approx \frac{K_a + \sqrt{K_a + 4K_aC_a}}{2} \tag{13.7.20}$$

Example 13.7.1: Acetic Acid

Calculate the pH of a 0.0010 M solution of acetic acid, $K_a = 1.74 \times 10^{-5}$.

Solution

First approximation:

$$[H^+] = \sqrt{(1.0 \times 10^{-3}) \times (1.74 \times 10^{-5}} = \sqrt{1.74 \times 10^{-8}} = 1.3 \times 10^{-4}\ M.$$

Applying the "5% test",

$$\frac{1.3 \times 10^{-4}}{1.0 \times 10^{-3}} = 0.13$$

This exceeds 0.05, so we must explicitly solve the quadratic Equation 13.7.19 to obtain two roots: $+1.2 \times 10^{-4}$ and -1.4×10^{-4}. Taking the positive root, we have

$$pH = -\log(1.2 \times 10^{-4}) = 3.9$$

Approximation 2: Very Concentrated Acids

If the acid is fairly concentrated (usually more than 10^{-3} M), a further simplification can frequently be achieved by making the assumption that $[H^+] \ll C_a$. This is justified when most of the acid remains in its protonated form [HA], so that relatively little H^+ is produced. In this event, Equation 13.7.18 reduces to

$$K_a \approx \frac{[H^+]^2}{C_a} \tag{13.7.21}$$

or

$$[H^+] \approx \sqrt{K_a C_a} \tag{13.7.22}$$

Example 13.7.2

Calculate the pH and percent ionization of 0.10 M acetic acid "HAc" (CH_3COOH), $K_a = 1.74 \times 10^{-5}$.

Solution

It is usually best to start by using Equation 13.7.21 as a first approximation:

$$[H^+] = \sqrt{(0.10)(1.74 \times 10^{-5})} = \sqrt{1.74 \times 10^{-6}} = 1.3 \times 10^{-3}\ M$$

This approximation is generally considered valid if $[H^+]$ is less than 5% of C_a; in this case, $[H^+]/C_a$ = 0.013, which is smaller than 0.05 and thus within the limit. This same quantity also corresponds to the ionization fraction, so the percent ionization is 1.3%. The pH of the solution is

$$pH = -\log 1.2 \times 10^{-3} = 2.9$$

Approximation 3: Very Weak and Acidic

If the acid is very weak or its concentration is very low, the H^+ produced by its dissociation may be little greater than that due to the ionization of water. However, if the solution is still acidic, it may still be possible to avoid solving the cubic equation 13.7.17 by assuming that the term $([H^+] - [OH^-]) \ll C_a$ in Equation 13.7.14:

$$K_a = \frac{[H^+]^2}{C_a - [H^+]} \tag{13.7.23}$$

This can be rearranged into standard quadratic form

$$[H^+]^2 + K_a\,[H^+] - K_a\,C_a = 0 \tag{13.7.24}$$

For dilute solutions of weak acids, an exact treatment may be required. With the aid of a computer or graphic calculator, solving a cubic polynomial is now far less formidable than it used to be. However, round-off errors can cause these computerized cubic solvers to blow up; it is generally safer to use a quadratic approximation.

Example 13.7.3: Boric Acid

Boric acid, $B(OH)_3$ ("H_3BO_3") is a weak acid found in the ocean and in some natural waters. As with many boron compounds, there is some question about its true nature, but for most practical purposes it can be considered to be monoprotic with $K_a = 7.3 \times 10^{-10}$:

$$Bi(OH)_3 + 2H_2O \rightleftharpoons Bi(OH)_4^- + H_3O^+$$

Find the [H^+] and pH of a 0.00050 M solution of boric acid in pure water.

Solution

Because this acid is quite weak and its concentration low, we will use the quadratic form Equation 13.7.19, which yields the positive root 6.12×10^{-7}, corresponding to pH = 6.21. Notice that this is only six times the concentration of H^+ present in pure water!

It is instructive to compare this result with what the quadratic approximation would yield, which yield $[H^+] = 6.04 \times 10^{-7}$ so $pH = 6.22$.

Weak bases

The weak bases most commonly encountered are:

- anions A^- of weak acids:

$$A^- + H_2O \rightleftharpoons HA + OH^- \tag{13.7.25}$$

$$CO_3^{2-} + H_2O \rightleftharpoons HCO_3^- + OH^- \tag{13.7.26}$$

- ammonia

$$NH_3 + H_2O \rightleftharpoons NH_4^+ + OH^- \tag{13.7.27}$$

- amines, e.g. methylamine

$$CH_3NH_2 + H_2O \rightleftharpoons CH_3NH_3^+ + H_2O \tag{13.7.28}$$

Solution of an anion of a weak acid

Note that, in order to maintain electroneutrality, anions must be accompanied by sufficient cations to balance their charges. Thus for a C_b *M* solution of the salt NaA in water, we have the following conditions:

- **Species**

Na^+, A^-, HA, H_2O, H^+, OH^-

- **Equilibria**

$$[H^+][OH^-] = K_w \tag{13.7.29}$$

$$K_b = \frac{[HA][OH^-]}{[A^-]} \tag{13.7.30}$$

- **Mass balance**

$$C_b = [Na^+] = [HA] + [A^-] \tag{13.7.31}$$

- **Charge balance**

$$[Na^+] + [H^+] = [OH^-] + [A^-] \tag{13.7.32}$$

Replacing the [Na^+] term in Equation 13.7.31 by C_b and combining with K_w and the mass balance, a relation is obtained that is analogous to that of Equation 13.7.14 for weak acids:

$$K_b = \frac{[OH^-]([OH^-] - [H^+])}{C_b - ([OH^-] - [H^+])} \tag{13.7.33}$$

The approximations

$$K_b \approx \frac{[OH^-]^2}{C_b - [OH^-]} \tag{13.7.34}$$

and

$$[OH^-] \approx \sqrt{K_b\, C_b} \tag{13.7.35}$$

can be derived in a similar manner.

Example 13.7.4: Methylamine

Calculate the pH and the concentrations of all species in a 0.01 M solution of methylamine, CH_3NH_2 ($K_b = 4.2 \times 10^{-4}$).

Solution

We begin by using the simplest approximation Equation 13.7.30:

$$[OH^-] = \sqrt{(K_b C_b -} = \sqrt{(4.2 \times 10^{-4})(10^{-2})} = 2.1 \times 10^{-3}$$

To see if this approximation is justified, we apply a criterion similar to what we used for a weak acid: [OH^-] must not exceed 5% of C_b. In this case,

$$\frac{[OH^-]}{C_b} = \frac{(2.1 \times 10^{-3}}{10^{-2}} = 0.21$$

so we must use the quadratic form Equation 13.7.24 that yields the positive root 1.9×10^{-3} which corresponds to $[OH^-]$

$$[H^+] = \frac{K_w}{[OH^-} = \frac{1 \times 10^{-14}}{1.9 \times 10^{-3}} = 5.3 \times 10^{-12}$$

and

$$pH = -\log 5.3 \times 10^{-12} = 11.3.$$

From the charge balance equation, solve for

$$[CH_3NH_2] = [OH^-]–[H^+] \approx [OH^-] = 5.3 \times 10^{-12}\ M.$$

For the concentration of the acid form (methylaminium ion $CH_3NH_3^+$), use the mass balance equation:

$$[CH_3NH_3^+] = C_b–[CH_3NH_2] = 0.01–0.0019 = 0.0081\ M.$$

Mixtures of Acids

Many practical problems relating to environmental and physiological chemistry involve solutions containing more than one acid. In this section, we will restrict ourselves to a much simpler case of two acids, with a view toward showing the general method of approaching such problems by starting with charge- and mass-balance equations and making simplifying assumptions when justified. In general, the hydrogen ions produced by the stronger acid will tend to suppress dissociation of the weaker one, and both will tend to suppress the dissociation of water, thus reducing the sources of H^+ that must be dealt with.

Consider a mixture of two weak acids HX and HY; their respective nominal concentrations and equilibrium constants are denoted by C_x, C_y, K_x and K_y,

Starting with the charge balance expression

$$[H^+] = [X^-] + [Y^-] + [OH^-] \tag{13.7.36}$$

We use the equilibrium constants to replace the conjugate base concentrations with expressions of the form

$$[X^-] = K_x\frac{[HX]}{[H^+]} \tag{13.7.37}$$

to yield

$$[H^+] = \frac{[HX]}{K_x} + \frac{[HY]}{K_y} + K_w \tag{13.7.38}$$

If neither acid is very strong or very dilute, we can replace equilibrium concentrations with nominal concentrations:

$$[H^+] \approx \sqrt{C_cK_x + C_yK_yK_w} \tag{13.7.39}$$

Example 13.7.5: Acetic Acid and Formic Acid

Estimate the pH of a solution that is 0.10M in acetic acid ($K_a = 1.8 \times 10^{-5}$) and 0.01M in formic acid ($K_a = 1.7 \times 10^{-4}$).

Solution

Because K_w is negligible compared to the C_aK_a products, we can simplify \Equation $ref3-4$:

$$[H^+] = \sqrt{1.8 \times 10^{-6} + 1.7 \times 10^{-6}} = 0.0019$$

Which corresponds to a pH of $-\log 0.0019 = 2.7$

Note that the pH of each acid separately at its specified concentration would be around 2.8. However, if 0.001 M chloroacetic acid (K_a= 0.0014) is used in place of formic acid, the above expression becomes

$$[H^+] \approx \sqrt{1.4 \times 10^{-6} + 1.75 \times 10^{-14}} = 0.00188 \tag{13.7.40}$$

which exceeds the concentration of the stronger acid; because the acetic acid makes a negligible contribution to $[H^+]$ here, the simple approximation given above \Equation 13.7.38 is clearly invalid. We now use the mass balance expression for the stronger acid

$$[HX] + [X^-] = C_x \tag{13.7.41}$$

to solve for [X^-] which is combined with the equilibrium constant K_x to yield

$$[X^-] = C_x - \frac{[H^+][X^]}{K_x} \tag{13.7.42}$$

Solving this for [X^-] gives

$$[X^-] = \frac{C_x K_x}{K_x + [H^+]} \tag{13.7.43}$$

The approximation for the weaker acetic acid (HY) is still valid, so we retain it in the substituted electronegativity expression:

$$[H^+]\frac{C_x K_x}{K_x + [H^+]} + \frac{C_y K_y}{[H^+]} \tag{13.7.44}$$

which is a cubic equation that can be solved by approximation.

Several methods have been published for calculating the hydrogen ion concentration in solutions containing an arbitrary number of acids and bases. These generally involve iterative calculations carried out by a computer. See, for example, *J. Chem. Education* 67(6) 501-503 (1990) and 67(12) 1036-1037 (1990).

Equilibria of Polyprotic Acids

Owing to the large number of species involved, exact solutions of problems involving polyprotic acids can become very complicated. Thus for phosphoric acid H_3PO_4, the three "dissociation" steps yield three conjugate bases:

$$H_3PO_4 \xrightarrow[10^{-2.12}]{K_1} H_2PO_4^- \xrightarrow[10^{-7.2}]{K_2} HPO_4^{2-} \xrightarrow[10^{-12.3}]{K_3} PO_4^{3-}$$

Fortunately, it is usually possible to make simplifying assumptions in most practical applications. In the section that follows, we will show how this is done for the less-complicated case of a diprotic acid. A diprotic acid HA can donate its protons in two steps, yielding first a monoprotonated species HA^- and then the completely deprotonated form A^{2-}.

$$H_2A \xrightarrow{K_1} HA^- \xrightarrow{K_2} A^{2-}$$

Since there are five unknowns (the concentrations of the acid, of the two conjugate bases and of H^+ and OH^-), we need five equations to define the relations between these quantities. These are

- **Equilibria**

$$[H^+][OH^-] = K_w \tag{13.7.45}$$

$$K_1 = \frac{[H^+][HA^-]}{[H_2A]} \tag{13.7.46}$$

$$K_2 = \frac{[H^+][HA^{2-}]}{[HA^-]} \tag{13.7.47}$$

- **Mass balance**

$$C_a = [H_2A] + [HA^-] + [A^{2-}] \tag{13.7.48}$$

- **Charge balance**

$$[H^+] = [OH^-] + [HA^-] + 2[A^{2-}] \tag{13.7.49}$$

(It takes 2 moles of H^+ to balance the charge of 1 mole of A^{2-})

Solving these five equations simultaneously for K_1 yields the rather intimidating expression

$$K_1 = \frac{[H^+]\left([H^+] - [OH^-]\dfrac{2K_2[H^+] - [OH^-]}{[H^+ + 2K_2}\right)}{C_a - \left([H^+] - [OH^-]\dfrac{K_2[H^+] - [OH^-]}{[H^+] + 2K_2}\right)} \tag{13.7.50}$$

which is of little practical use except insofar as it provides the starting point for various simplifying approximations.

If the solution is even slightly acidic, then ([H$^+$] – [OH$^-$]) ≈ [H$^+$] and

$$K_1 = \frac{[H^+]\left([H^+]\dfrac{2K_2[H^+]}{[H^+ + 2K_2}\right)}{C_a - \left([H^+]\dfrac{K_2[H^+]}{[H^+] + 2K_2}\right)} \tag{13.7.51}$$

For any of the common diprotic acids, K_2 is much smaller than K_1. If the solution is sufficiently acidic that $K_2 \ll [H^+]$, then a further simplification can be made that removes K_2 from Equation 13.7.51; this is the starting point for most practical calculations.

$$K_1 \approx \frac{[H^+]^2}{C_a - [H^+]} \tag{13.7.52}$$

Finally, if the solution is sufficiently concentrated and K_1 sufficiently small so that $[H^+] \ll C_a$, then Equation 13.7.52 reduces to:

$$[H^+] \approx \sqrt{K_a C_a} \tag{13.7.53}$$

Acid with conjugate base: Buffer solutions

Solutions containing a weak acid together with a salt of the acid are collectively known as *buffers*. When they are employed to control the pH of a solution (such as in a microbial growth medium), a sodium or potassium salt is commonly used and the concentrations are usually high enough for the Henderson-Hasselbalch equation to yield adequate results.

Exact solution

In this section, we will develop an exact analytical treatment of weak acid-salt solutions, and show how the H$^-$H equation arises as an approximation. A typical buffer system is formed by adding a quantity of strong base such as sodium hydroxide to a solution of a weak acid HA. Alternatively, the same system can be made by combining appropriate amounts of a weak acid and its salt NaA. A system of this kind can be treated in much the same way as a weak acid, but now with the parameter C_b in addition to C_a.

- **Species**

  Na$^+$, A$^-$, HA, H_2O, H$^+$, OH$^-$

- **Equilibria**

$$[H^+][OH^-] = K_w \tag{13.7.54}$$

$$K_a = \frac{[H^+][A^-]}{[HA]} \tag{13.7.55}$$

- **Mass balance**

$$C_a + C - b = [HA] + [A^-] \tag{13.7.56}$$

$$C_b = [Na^+] \tag{13.7.57}$$

- **Charge balance**

$$[Na^+] + [H^+] = [OH^-] + [A^-] \tag{13.7.58}$$

Substituting Equation 13.7.57 into Equation 13.7.58 yields an expression for [A$^-$]:

$$[A^-] = C_b + [H^+] - [OH^-] \tag{13.7.59}$$

Inserting this into Equation 13.7.56 and solving for [HA] yields

$$[HA] = C_b + [H^+] - [OH^-] \tag{13.7.60}$$

Finally, we substitute these last two expressions into the equilibrium constant (Equation 13.7.55):

$$[H^+] = K_a \frac{C_a - [H^+] + [OH^-]}{C_b + [H^+] - [OH^-]} \tag{13.7.61}$$

which becomes cubic in [H$^+$] when [OH$^-$] is replaced by (K_w / [H$^+$]).

$$[H^+]^3 + (C_b + K_a)[H^+]^2 - (K_w + C_a K_a)[H^+] - K_a K_w = 0 \tag{13.7.62}$$

Approximations

In almost all practical cases it is possible to make simplifying assumptions. Thus if the solution is known to be acidic or alkaline, then the [OH$^-$] or [H$^+$] terms in Equation 13.7.61 can be neglected. In acidic solutions, for example, Equation 13.7.61 becomes

$$[H^+] = K_a \frac{C_a - [H^+]}{C_b + [H^+]} \tag{13.7.63}$$

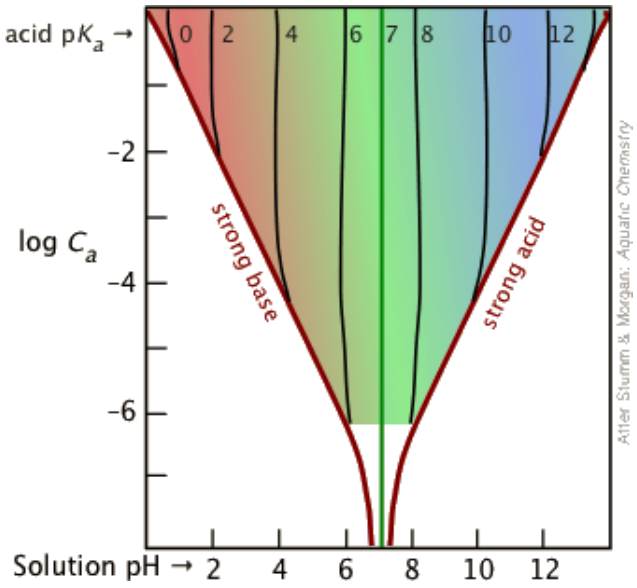
which can be rearranged into a quadratic in standard polynomial form:

$$[H^+]^2 + (C_b + C_a)\,[H^+] - K_a\,C_a = 0 \tag{13.7.64}$$

If the concentrations C_a and C_b are sufficiently large, it may be possible to neglect the [H$^+$] terms entirely, leading to the commonly-seen *Henderson-Hasselbalch Approximation*.

$$[H^+] \approx K_a \frac{C_a}{C_b} \tag{13.7.65}$$

It's important to bear in mind that the Henderson-Hasselbalch Approximation is an "approximation of an approximation" that is generally valid only for combinations of K_a and concentrations that fall within the colored portion of this plot.



Most buffer solutions tend to be fairly concentrated, with C_a and C_b typically around 0.01 - 0.1 *M*. For more dilute buffers and larger K_a's that bring you near the boundary of the colored area, it is safer to start with Equation 13.7.63.

Example 13.7.6: Chlorous Acid Buffer

Chlorous acid $HClO_2$ has a pK_a of 1.94. Calculate the pH of a solution made by adding 0.01 M/L of sodium hydroxide to a -.02 M/L solution of chloric acid.

Solution

In the resulting solution, $C_a = C_b = 0.01M$. On the plots shown above, the intersection of the log C_a = –2 line with the plot for pK_a = 2 falls near the left boundary of the colored area, so we will use the quadratic form 13.7.64.

Substitution in Equation 13.7.64 yields

$$H^+ + 0.02H^+ - (10^{-1.9} x 10^{-2}) = 0$$

which yields a positive root 0.0047 = [H^+] that corresponds to pH = **2.3**.

Note: Using the Henderson-Hassalbach Approximateion (Equation 13.7.65) would give pH = pK_a = 1.9.

This page titled 13.7: Exact Calculations and Approximations is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14: THERMOCHEMISTRY

All chemical changes are accompanied by the absorption or release of heat. The intimate connection between matter and energy has been a source of wonder and speculation from the most primitive times; it is no accident that fire was considered one of the four basic elements (along with earth, air, and water) as early as the fifth century BCE. In this set of lessons, we will review some of the fundamental concepts of energy and heat and the relation between them. We will begin the study of *thermodynamics*, which treats the energetic aspects of change in general, and we will finally apply this specifically to chemical change. Our purpose will be to provide you with the tools to predict the energy changes associated with chemical processes. This will build the groundwork for a more ambitious goal: to predict the direction and extent of change itself.

14.1: Energy, Heat and Work

14.2: The First Law of Thermodynamics

14.3: Molecules as Energy Carriers and Converters

14.4: Thermochemistry and Calorimetry

14.5: Calorimetry

14.6: Applications of Thermochemistry

14.E: Thermochemistry (Exercises)

This page titled 14: Thermochemistry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14.1: Energy, Heat and Work

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- The *potential energy* of an object relates to its location, but there is one additional requirement that must be satisfied for potential energy be present. Explain and give an example.
- Distinguish between the nature of *kinetic energy* that is associated with macroscopic bodies and that is found in microscopic objects such as atoms and molecules.
- Describe the meaning and origins of "*chemical*" *energy*.
- Define the *calorie*.
- *Heat and work* are both expressed in energy units, but they differ from "plain" energy in a fundamental way. Explain.
- ... and state the distinction between heat and work.

All chemical changes are accompanied by the absorption or release of heat. The intimate connection between matter and energy has been a source of wonder and speculation from the most primitive times; it is no accident that fire was considered one of the four basic elements (along with earth, air, and water) as early as the fifth century BCE. In this unit we will review some of the fundamental concepts of energy and heat and the relation between them. We will begin the study of *thermodynamics*, which treats the energetic aspects of change in general, and we will finally apply this specifically to chemical change. Our purpose will be to provide you with the tools to predict the energy changes associated with chemical processes. This will build the groundwork for a more ambitious goal: to predict the direction and extent of change itself.

What is Energy?

Energy is one of the most fundamental and universal concepts of physical science, but one that is remarkably difficult to define in way that is meaningful to most people. This perhaps reflects the fact that energy is not a “thing” that exists by itself, but is rather an attribute of matter (and also of electromagnetic radiation) that can manifest itself in different ways. It can be observed and measured only indirectly through its effects on matter that acquires, loses, or possesses it.

The concept that we call energy was very slow to develop; it took more than a hundred years just to get people to agree on the definitions of many of the terms we use to describe energy and the interconversion between its various forms. But even now, most people have some difficulty in explaining what it is; somehow, the definition we all learned in elementary science ("the capacity to do work") seems less than adequate to convey its meaning. Although the term "energy" was not used in science prior to 1802, it had long been suggested that certain properties related to the motions of objects exhibit an endurance which is incorporated into the modern concept of "conservation of energy". René Descartes (1596-1650) stated it explicitly:

> When God created the world, He "caused some of its parts to push others and to transfer their motions to others..." and thus "He **conserves** motion".*

In the 17th Century, the great mathematician Gottfried Leibniz (1646-1716) suggested the distinction between *vis viva* ("live force") and *vis mortua* ("dead force"), which later became known as kinetic energy (1829) and potential energy (1853).

Kinetic Energy and Potential Energy

Whatever energy may be, there are basically two kinds. Kinetic energy is associated with the **motion** of an object, and its direct consequences are part of everyone's daily experience; the faster the ball you catch in your hand, and the heavier it is, the more you feel it. Quantitatively, a body with a mass m and moving at a velocity v possesses the kinetic energy $mv^2/2$.

Example 14.1.1

A rifle shoots a 4.25 g bullet at a velocity of 965 m s^{-1}. What is its kinetic energy?

Solution

The only additional information you need here is that

$$1\ J = 1\ kg\ m^2\ s^{-2}:$$

$$KE = ½ \times (.00425\ kg)\ (965\ m\ s^{-1})^2 = 1980\ J$$

Potential energy is energy a body has by virtue of its *location*. But there is more: the body must be subject to a "restoring force" of some kind that tends to move it to a location of lower potential energy. Think of an arrow that is subjected to the force from a stretched bowstring; the more tightly the arrow is pulled back against the string, the more potential energy it has. More generally, the restoring force comes from what we call a *force field*— a gravitational, electrostatic, or magnetic field. We observe the consequences of *gravitationa*l potential energy all the time, such as when we walk, but seldom give it any thought.

If an object of mass m is raised off the floor to a height h, its potential energy increases by mgh, where g is a proportionality constant known as the *acceleration of gravity*; its value at the earth's surface is 9.8 m s^{-2}.

Example 14.1.2

Find the change in potential energy of a 2.6 kg textbook that falls from the 66-cm height of a table top onto the floor.

Solution

$$PE = m\,g\,h = (2.6\ kg)(9.8\ m\ s^{-2})(0.66\ m) = 16.8\ kg\ m^2\ s^{-2} = 16.8\ J$$

Similarly, the potential energy of a particle having an electric charge q depends on its location in an electrostatic field.

"Chemical energy"

Electrostatic potential energy plays a major role in chemistry; the potential energies of electrons in the force field created by atomic nuclei lie at the heart of the chemical behavior of atoms and molecules.

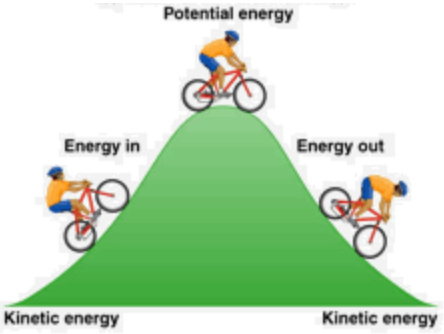
"Chemical energy" usually refers to the energy that is stored in the chemical bonds of molecules. These bonds form when electrons are able to respond to the force fields created by two or more atomic nuclei, so they can be regarded as manifestations of electrostatic potential energy.

In an exothermic chemical reaction, the electrons and nuclei within the reactants undergo rearrangment into products possessing lower energies, and the difference is released to the environment in the form of heat.

Interconversion of potential and kinetic energy

Transitions between potential and kinetic energy are such an intimate part of our daily lives that we hardly give them a thought. It happens in walking as the body moves up and down. Our bodies utilize the chemical energy in glucose to keep us warm and to move our muscles. In fact, life itself depends on the conversion of chemical energy to other forms.

Energy is conserved: it can neither be created nor destroyed. So when you go uphill, your kinetic energy is transformed into potential energy, which gets changed back into kinetic energy as you coast down the other side. And where did the kinetic energy you expended in peddling uphill come from? By conversion of some of the chemical potential energy in your breakfast cereal.



- When drop a book, its potential energy is transformed into kinetic energy. When it strikes the floor, this transformation is complete. What happens to the energy then? The kinetic energy that at the moment of impact was formerly situated exclusively in the moving book, now becomes shared between the book and the floor, and in the form of randomized thermal motions of the molecular units of which they are made; we can observe this effect as a rise in temperature.
- ← Much of the potential energy of falling water can be captured by a water wheel or other device that transforms the kinetic energy of the exit water into kinetic energy. The output of a hydroelectric power is directly proportional to its height above the level of the generator turbines in the valley below. At this point, the kinetic energy of the exit water is transferred to that of the turbine, most of which (up to 90 percent in the largest installations) is then converted into electrical energy.
- Will the temperature of the water at the bottom of a water fall be greater than that at the top? James Joule himself predicted that it would be. It has been calculated that at Niagra falls, that complete conversion of the potential energy of 1 kg of water at the top into kinetic energy when it hits the plunge pool 58 meters below will result in a temperature increase of about 0.14 C°. (But there are lots of complications. For example, some of the water breaks up into tiny droplets as it falls, and water evaporates from droplets quite rapidly, producing a cooling effect.)
- Chemical energy can also be converted, at least partially, into electrical energy: this is what happens in a battery. If a highly exothermic reaction also produces gaseous products, the latter may expand so rapidly that the result is an explosion — a net conversion of chemical energy into kinetic energy (including sound).

Thermal energy

Kinetic energy is associated with motion, but in two different ways. For a macroscopic object such as a book or a ball, or a parcel of flowing water, it is simply given by $\frac{1}{2} mv^2$.

But as we mentioned above, when an object is dropped onto the floor, or when an exothermic chemical reaction heats surrounding matter, the kinetic energy gets dispersed into the molecular units in the environment. This "microscopic" form of kinetic energy, unlike that of a speeding bullet, is completely random in the kinds of motions it exhibits and in its direction. We refer to this as "thermalized" kinetic energy, or more commonly simply as *thermal energy*. We observe the effects of this as a rise in the temperature of the surroundings. The temperature of a body is direct measure of the quantity of thermal energy is contains.

Thermal energy is never completely recoverable

Once kinetic energy is thermalized, only a portion of it can be converted back into potential energy. The remainder simply gets dispersed and diluted into the environment, and is effectively lost.

To summarize, then:

- Potential energy can be converted entirely into kinetic energy..
- Potential energy can also be converted, with varying degrees of efficiency,into electrical energy.
- The kinetic energy of macroscopic objects can be transferred between objects (barring the effects of friction).
- Once kinetic energy becomes thermalized, only a portion of it can be converted back into either potential energy or be concentrated back into the kinetic energy of a macroscopic. This limitation, which has nothing to do with technology but is a fundamental property of nature, is the subject of the second law of thermodynamics.
- A device that is intended to accomplish the partial transformation of thermal energy into organized kinetic energy is known as a *heat engine*.

Energy scales and units

You might at first think that a book sitting on the table has zero kinetic energy since it is not moving. But if you think about it, the earth itself is moving; it is spinning on its axis, it is orbiting the sun, and the sun itself is moving away from the other stars in the

general expansion of the universe. Since these motions are normally of no interest to us, we are free to adopt an arbitrary scale in which the velocity of the book is measured with respect to the table; on this so-called *laboratory coordinate system*, the kinetic energy of the book can be considered zero.

We do the same thing with potential energy. If the book is on the table, its potential energy with respect to the surface of the table will be zero. If we adopt this as our zero of potential energy, and then push the book off the table, its potential energy will be negative after it reaches the floor.

Energy is measured in terms of its ability to perform work or to transfer heat. Mechanical work is done when a force f displaces an object by a distance *d*: $w = f \times d$.

The basic unit of energy is the *joule*. One joule is the amount of work done when a force of 1 newton acts over a distance of 1 m; thus 1 J = 1 N-m. The newton is the amount of force required to accelerate a 1-kg mass by 1 m/sec^2, so the basic dimensions of the joule are kg m^2 s^{-2}. The other two units in wide use. the *calorie* and the *BTU* (British thermal unit) are defined in terms of the heating effect on water. Because of the many forms that energy can take, there are a correspondingly large number of units in which it can be expressed, a few of which are summarized below.

1 calorie will raise the temperature of 1 g of water by 1 C°. The “dietary” calorie is actually 1 kcal. An average young adult expends about 1800 kcal per day just to stay alive. (you should know this definition)	1 cal = 4.184 J
1 BTU (British Thermal Unit) will raise the temperature of 1 lb of water by 1F°.	1 BTU = 1055 J
The **erg** is the c.g.s. unit of energy and a very small one; the work done when a 1-dyne force acts over a distance of 1 cm.	1 J = 10^7 ergs 1 erg = 1 d-cm = 1 g cm^2 s^{-2}
The **electron-volt** is even tinier: 1 e-v is the work required to move a unit electric charge (1 C) through a potential difference of 1 volt.	1 J = 6.24×10^{18} e-v
The watt is a unit of power, which measures the rate of energy flow in J sec^{-1}. Thus the **watt-hour** is a unit of energy. An average human consumes energy at a rate of about 100 watts; the brain alone runs at about 5 watts.	1 J = 2.78×10^{-4} watt-hr 1 w-h = 3.6 kJ
The **liter-atmosphere** is a variant of force-displacement work associated with volume changes in gases.	1 L-atm = 101.325 J
The huge quantities of energy consumed by cities and countries are expressed in **quads**; the **therm** is a similar but smaller unit.	1 quad = 10^{15} Btu = 1.05×10^{18} J
If the object is to obliterate cities or countries with nuclear weapons, the energy unit of choice is the **ton of TNT equivalent**.	1 ton of TNT = 4.184 GJ (by definition)
In terms of **fossil fuels**, we have barrel-of-oil equivalent, cubic-meter-of-natural gas equivalent, and ton-of-coal equivalent.	1 bboe = 6.1 GJ 1 cmge = 37-39 mJ 1 toce = 29 GJ

Heat and work

Heat and work are both measured in energy units, so they must both represent energy. How do they differ from each other, and from just plain “energy” itself? In our daily language, we often say that "this object contains a lot of heat", but this kind of talk

is a no-no in thermodynamics! It's ok to say that the object is "hot", meaning that its temperature is high. The term "heat" has a special meaning in thermodynamics: it is a *process* in which a body (the contents of a tea kettle, for example) acquires or loses energy as a direct consequence of its having a *different temperature* than its surroundings (the rest of the world).

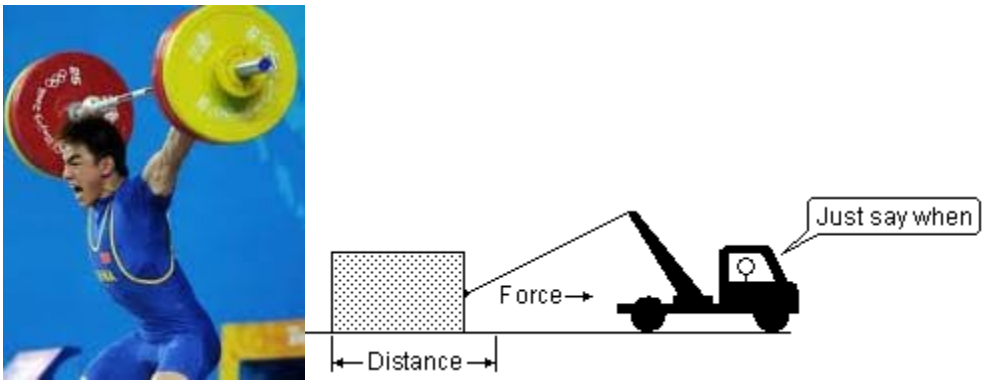
Heat and work are *processes* and cannot be stored

Thermal energy can only flow from a higher temperature to a lower temperature. It is this flow that constitutes "heat". Use of the term "flow" of heat recalls the 18th-century notion that heat is an actual substance called “caloric” that could flow like a liquid.

Transfer of thermal energy can be accomplished by bringing two bodies into physical contact (the kettle on top of the stove, or through an electric heating element inside the kettle). Another mechanism of thermal energy transfer is by radiation; a hot object will convey energy to any body in sight of it via electromagnetic radiation in the infrared part of the spectrum. In many cases, a combination of modes will be active. Thus when you place a can of beer in the refrigerator, both processes are operative: the can radiates heat to the cold surfaces around it, and absorbs it by direct conduction from the ambient air.

So what is *work?*

Work refers to the transfer of energy some means that does not depend on temperature difference.



Work, like energy, can take various forms, the most familiar being mechanical and electrical. *Mechanical work* arises when an object moves a distance Δx against an opposing force f:

\[$w = f\,\Delta x$ N^{-m}\]

with $1\ N^{-m}$ = 1\, J.\)

Electrical work is done when a body having a charge q moves through a potential difference ΔV.

Work, like heat, exists only when energy is being transferred.When two bodies are placed in thermal contact and energy flows from the warmer body to the cooler one,we call the process “heat”. A transfer of energy to or from a system by any means other than heat is called “work”.

Interconvertability of heat and work

Work can be completely converted into heat (by friction, for example), but heat can only be partially converted to work. **Conversion of heat into work** is accomplished by means of a *heat engine*, the most common example of which is an ordinary gasoline engine. The science of thermodynamics developed out of the need to understand the limitations of steam-driven heat engines at the beginning of the Industrial Age. A fundamental law of Nature, the Second Law of Thermodynamics, states that the complete conversion of heat into work is impossible. Something to think about when you purchase fuel for your car!

This page titled 14.1: Energy, Heat and Work is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14.2: The First Law of Thermodynamics

Learning Objectives

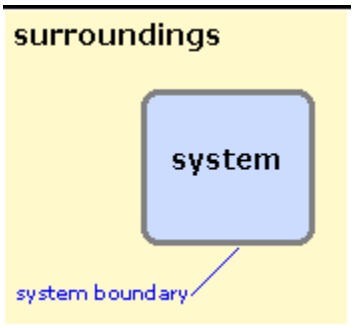
Make sure you thoroughly understand the following essential concepts:

- Define: *system*, *surroundings*, *state properties*, *change of state* (how the latter is calculated).
- Write out the equation that defines the *First Law*, and explain its physical meaning.
- Describe the meaning of a *pathway* in thermodynamics, and how it relates to *state functions*.
- Comment on the meaning of *pressure-volume work*: how is it calculated? What kinds of chemical reactions involve *P-V* work?
- Define *isothermal* and *adiabatic* processes, and give examples of each.
- Using the expansion of a gas as an example, state the fundamental distinction between *reversible* and irreversible changes in terms of the system + surroundings.
- Explain how *enthalpy change* relates to *internal energy*, and how it can be observed experimentally.
- Define *heat capacity*; explain its physical significance, and why there is a difference between C_p and C_v.

"*Energy cannot be created or destroyed*"— this fundamental law of nature, more properly known as **conservation of energy**, is familiar to anyone who has studied science. Under its more formal name of the *First Law of Thermodynamics*, it governs all aspects of energy in science and engineering applications. It's special importance in Chemistry arises from the fact that virtually all chemical reactions are accompanied by the uptake or release of energy. One of the interesting things about thermodynamics is that although it deals with matter, it makes no assumptions about the microscopic nature of that matter. Thermodynamics deals with matter in a macroscopic sense; it would be valid even if the atomic theory of matter were wrong. This is an important quality, because it means that reasoning based on thermodynamics is unlikely to require alteration as new facts about atomic structure and atomic interactions come to light.

A Thermodynamic view of the world

In thermodynamics, we must be very precise in our use of certain words. The two most important of these are *system* and *surroundings*. A *thermodynamic system* is that part of the world to which we are directing our attention. Everything that is not a part of the system constitutes the *surroundings*. The system and surroundings are separated by a *boundary*.



If our system is one mole of a gas in a container, then the boundary is simply the inner wall of the container itself. The boundary need not be a physical barrier; for example, if our system is a factory or a forest, then the boundary can be wherever we wish to define it. We can even focus our attention on the dissolved ions in an aqueous solution of a salt, leaving the water molecules as part of the surroundings. The single property that the boundary must have is that it be clearly defined, so we can unambiguously say whether a given part of the world is in our system or in the surroundings.

> If matter is not able to pass across the boundary, then the system is said to be closed; otherwise, it is open. A closed system may still exchange energy with the surroundings unless the system is an isolated one, in which case neither matter nor energy can pass across the boundary. The tea in a closed Thermos bottle approximates a closed system over a short time interval.

The *properties* of a system are those quantities such as the pressure, volume, temperature, and its composition, which are in principle measurable and capable of assuming definite values. There are of course many properties other than those mentioned above; the density and thermal conductivity are two examples. However, the pressure, volume, and temperature have special significance because they determine the values of all the other properties; they are therefore known as *state properties* because if their values are known then the system is in a definite *state*.

In dealing with thermodynamics, we must be able to unambiguously define the change in the state of a system when it undergoes some process. This is done by specifying changes in the values of the different state properties using the symbol Δ (*delta*) as illustrated here for a change in the volume:

$$\Delta V = V_{final} - V_{initial} \tag{14.2.1}$$

We can compute similar delta-values for changes in *P, V,* n_i (the number of moles of component *i*), and the other state properties we will meet later.

Internal energy is simply the totality of all forms of kinetic and potential energy of the system. Thermodynamics makes no distinction between these two forms of energy and it does not assume the existence of atoms and molecules. But since we are studying thermodynamics in the context of chemistry, we can allow ourselves to depart from "pure" thermodynamics enough to point out that the internal energy is the sum of the kinetic energy of motion of the molecules, and the potential energy represented by the chemical bonds between the atoms and any other intermolecular forces that may be operative.

How can we know how much internal energy a system possesses? The answer is that we cannot, at least not on an absolute basis; all scales of energy are arbitrary. The best we can do is measure changes in energy. However, we are perfectly free to define zero energy as the energy of the system in some arbitrary *reference state*, and then say that the internal energy of the system in any other state is the difference between the energies of the system in these two different states.

The First Law of Thermodynamics

This law is one of the most fundamental principles of the physical world. Also known as the *Law of Conservation of Energy*, it states that energy can not be created or destroyed; it can only be redistributed or changed from one form to another. A way of expressing this law that is generally more useful in Chemistry is that any change in the internal energy of a system is given by the sum of the heat *q* that flows across its boundaries and the work *w* done on the system by the surroundings.

$$\Delta U = q + w \tag{14.2.2}$$

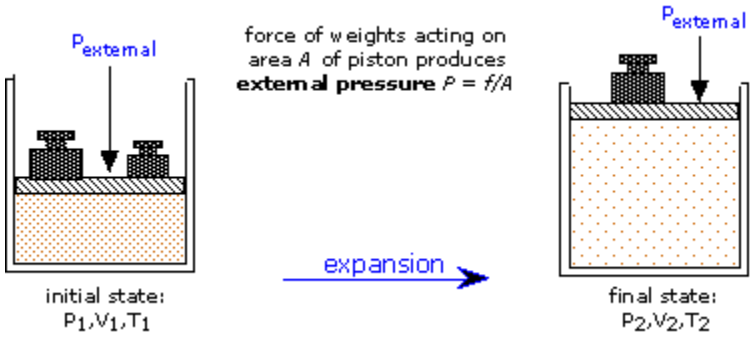
This says that there are two kinds of processes, heat and work, that can lead to a change in the internal energy of a system. Since both heat and work can be measured and quantified, this is the same as saying that any change in the energy of a system must result in a corresponding change in the energy of the world outside the system- in other words, energy cannot be created or destroyed.

There is an important **sign convention** for heat and work that you are expected to know. If heat flows into a system or the surroundings to do work on it, the internal energy increases and the sign of *q* or *w* is positive. Conversely, heat flow out of the system or work done by the system will be at the expense of the internal energy, and will therefore be negative. (Note that this is the opposite of the sign convention that was commonly used in much of the pre-1970 literature.) The full significance of Equation 14.2.2 cannot be grasped without understanding that *U* is a *state function*. This means that a given change in internal energy Δ*U* can follow an infinite variety of *pathways* corresponding to all the possible combinations of *q* and *w* that can add up to a given value of Δ*U*.

As a simple example of how this principle can simplify our understanding of change, consider two identical containers of water initially at the same temperature. We place a flame under one until its temperature has risen by 1°C. The water in the other container is stirred vigorously until its temperature has increased by the same amount. There is now no physical test by which you could determine which sample of water was warmed by performing work on it, by allowing heat to flow into it, or by some combination of the two processes. In other words, there is no basis for saying that one sample of water now contains more "work", and the other more "heat". The only thing we can know for certain is that both samples have undergone identical increases in internal energy, and we can determine the value of simply by measuring the increase in the temperature of the water.

Pressure-volume work

The kind of work most frequently associated with chemical change occurs when the volume of the system changes owing to the disappearance or formation of gaseous substances. This is sometimes called expansion work or *PV*-work, and it can most easily be understood by reference to the simplest form of matter we can deal with, the hypothetical *ideal gas*.



The figure shows a quantity of gas confined in a cylinder by means of a moveable piston. Weights placed on top of the piston exert a force f over the cross-section area *A*, producing a pressure $P = f / A$ which is exactly countered by the pressure of the gas, so that the piston remains stationary. Now suppose that we heat the gas slightly; according to Charles' law, this will cause the gas to expand, so the piston will be forced upward by a distance Δx. Since this motion is opposed by the force *f*, a quantity of work $f\,\Delta x$ will be done by the gas on the piston. By convention, work done by the system (in this case, the gas) on the surroundings is negative, so the work is given by

$$w = -f\Delta x \tag{14.2.3}$$

When dealing with a gas, it is convenient to think in terms of the more relevant quantities pressure and volume rather than force and distance. We can accomplish this by multiplying the second term by *A/A* which of course leaves it unchanged:

$$w = -f\Delta x\frac{A}{A} \tag{14.2.4}$$

By grouping the terms differently, but still not changing anything, we obtain

$$w = -\frac{f}{A}\Delta x A \tag{14.2.5}$$

Since pressure is force per unit area and the product of the length *A* and the area has the dimensions of volume, this expression becomes

$$w = -P\Delta V \tag{14.2.6}$$

It is important to note that although P and V are state functions, the *work* is not (that's why we denote it by a lower-case *w*.) As is shown farther below, the quantity of work done will depend on whether the same net volume change is realized in a single step (by setting the external pressure to the final pressure *P*), or in multiple stages by adjusting the restraining pressure on the gas to successively smaller values approaching the final value of P.

Example 14.2.1

Find the amount of work done on the surroundings when 1 liter of an ideal gas, initially at a pressure of 10 atm, is allowed to expand at constant temperature to 10 liters by:

a. reducing the external pressure to 1 atm in a single step,
b. reducing P first to 5 atm, and then to 1 atm,
c. allowing the gas to expand into an evacuated space so its total volume is 10 liters.

Solution

First, note that ΔV, which is a state function, is the same for each path:

$$V_2 = (10/1) \times (1L) = 10L \tag{14.2.7}$$

so

$$\Delta V = 9\ L \tag{14.2.8}$$

For path *(a)*

$$w = -(1\ atm) \times (9\ L) = \mathbf{-9\ L\text{-}atm}.$$

For path *(b)*, the work is calculated for each stage separately:

$$w = -(5\ atm) \times (2\text{–}1\ L) - (1\ atm) \times (10\text{–}2\ L) = \mathbf{-13\ L\text{-}atm}$$

For path *(c)* the process would be carried out by removing all weights from the piston in Figure $14.2.1$ so that the gas expands to 10 L against zero external pressure. In this case $w = (0\ atm) \times 9\ L = \mathbf{0}$; that is, *no work is done because there is no force to oppose the expansion*.

Adiabatic and Isothermal Processes

When a gas expands, it does work on the surroundings; compression of a gas to a smaller volume similarly requires that the surroundings perform work on the gas. If the gas is thermally isolated from the surroundings, then the process is said to occur *adiabatically*. In an adiabatic change, $q = 0$, so the First Law becomes $\Delta U = 0 + w$. Since the temperature of the gas changes with its internal energy, it follows that adiabatic compression of a gas will cause it to warm up, while adiabatic expansion will result in cooling.

In contrast to this, consider a gas that is allowed to slowly escape from a container immersed in a constant-temperature bath. As the gas expands, it does work on the surroundings and therefore tends to cool, but the thermal gradient that results causes heat to pass into the gas from the surroundings to exactly compensate for this change. This is called an *isothermal* expansion. In an isothermal process the internal energy remains constant and we can write the First Law (Equation $Ref2-1$) as

$$0 = q + w \tag{14.2.9}$$

or

$$q = -w \tag{14.2.10}$$

This illustrates that the heat flow and work done exactly balance each other.

Because no thermal insulation is perfect, truly adiabatic processes do not occur. However, heat flow does take time, so a compression or expansion that occurs more rapidly than thermal equilibration can be considered adiabatic for practical purposes. If you have ever used a hand pump to inflate a bicycle tire, you may have noticed that the bottom of the pump barrel can get quite warm. Although a small part of this warming may be due to friction, it is mostly a result of the work you (the surroundings) are doing on the system (the gas.)

Adiabatic expansion and contractions are especially important in understanding the behavior of the **atmosphere**. Although we commonly think of the atmosphere as homogeneous, it is really not, due largely to uneven heating and cooling over localized areas. Because mixing and heat transfer between adjoining parcels of air does not occur rapidly, many common atmospheric phenomena can be considered at least quasi-adiabatic.

Reversible processes

From Example $14.2.1$ we see that when a gas expands into a vacuum ($P_{external} = 0$ the work done is zero. This is the *minimum work* the gas can do; what is the *maximum work* the gas can perform on the surroundings? To answer this, notice that more work is done when the process is carried out in two stages than in one stage; a simple calculation will show that even more work can be obtained by increasing the number of stages— that is, by allowing the gas to expand against a series of successively lower

external pressures. In order to extract the maximum possible work from the process, the expansion would have to be carried out in an infinite sequence of infinitesimal steps. Each step yields an increment of work $P\,\Delta V$ which can be expressed as $(RT/V)\,dV$ and integrated:

$$w = \int_{V_1}^{V_2} \frac{RT}{V} dv \tag{14.2.11}$$

$$= RT \ln \frac{V_2}{V_1} \tag{14.2.12}$$

Although such a path (which corresponds to what is called a *reversible process*) cannot be realized in practice, it can be approximated as closely as desired.

Even though no real process can take place reversibly (it would take an infinitely long time!), reversible processes play an essential role in thermodynamics. The main reason for this is that q_{rev} and w_{rev} are *state functions* which are important and are easily calculated. Moreover, many real processes take place sufficiently gradually that they can be treated as approximately reversible processes for easier calculation.

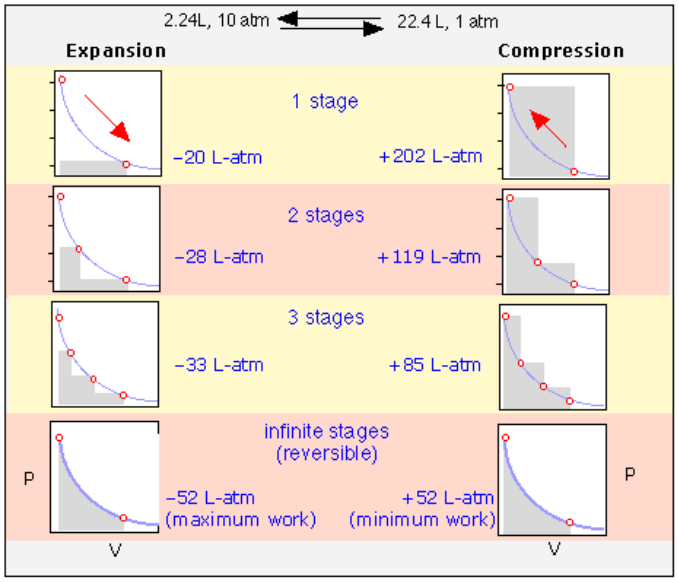


Figure 14.2.1: These plots illustrate the effects of various degrees of reversibility on the amount of work done when a gas expands, and the work that must be done in order to restore it to its initial state by recompressing it. The work, in each case, is proportional to the shaded area on the plot.

Each expansion-compression cycle leaves the gas unchanged, but in all but the one in the bottom row, the surroundings are forever altered, having expended more work in compressing the gas than was performed on it when the gas expanded. Only when the processes are carried out in an ***infinite*** number of steps will the system *and* the surroundings be restored to their initial states— this is the meaning of thermodynamic reversibility.

Heat changes at constant pressure: the Enthalpy

For a chemical reaction that performs no work on the surroundings, the heat absorbed is the same as the change in internal energy: $q = \Delta U$. But many chemical processes do involve work in one form or another:

- If the total volume of the reaction products exceeds that of the reactants, then the process performs work on the surroundings in the amount $P\Delta V$, in which P is the pressure exerted by the surroundings (usually the atmosphere) on the system.
- A reaction that drives an electrical current through an external circuit performs *electrical work* on the surroundings.

For an isothermal process, pressure-volume work affects the heat *q*

We will consider only pressure-volume work in this lesson. If the process takes place at a constant pressure, then the work is given by $P\Delta V$ and the change in internal energy will be

$$\Delta U = q – P\Delta V \tag{14.2.13}$$

Bear in mind why q is so important: the heat flow into or out of the system is directly measurable. ΔU, being "internal" to the system, is not directly observable. Thus the amount of heat that passes between the system and the surroundings is given by

$$q = \Delta U + P\Delta V \tag{14.2.14}$$

This means that if an exothermic reaction is accompanied by a net increase in volume under conditons of constant pressure, some heat additional to ΔU must be absorbed in order to supply the energy expended as work done on the surroundings if the temperature is to remain unchanged *(isothermal* process.)

For most practical purposes, changes in the volume of the system are only significant if the reaction is accompanied by a difference in the moles of gaseous reactants and products.

For example, in the reaction $H_2(g) + O_2(g) \rightarrow ½H_2O(l)$, the total volume of the system decreases from that correponding to 2 moles of gaseous reactants to 0.5 mol of liquid water which occupies only 9 mL — a volume so small in comparison to that of the reactants that it can be neglected without significant error. So all we are really concerned with is the difference in the number of moles of gas Δn_g:

Δn_g = (0 – 2) mol = –2 mol

This corresponds to a net contraction (negative expansion) of the system, meaning that the surroundings perform work on the system.

The molar volume of an ideal gas at 25° C and 1 atm is

(298/273) × (22.4 L mol^{-1}) = 24.5 L mol^{-1}

Remember the sign convention: a flow of heat or performance of work that supplies energy is positive; if it consumes energy, it is negative. Thus work performed by the surroundings diminishes the energy of the surroundings ($w_{surr} < 0$) and increases the energy of the system ($w_{sys} > 0$).

and the work done (by the surroundings on the system) is

(1 atm) (–2 mol)(24.5 L mol^{-1}) = –49.0 L-atm.

Using the conversion factor 1 J = 101.3 J, and bearing in mind that work performed *on* the system supplies energy *to* the system, the work associated solely with the volume change of the system increases its energy by

(101.3 J/L-atm)(–49.0 L-atm) = 4964 J = 4.06 kJ

Example 14.2.2

The above reaction $H_2(g) + ½\ O_2(g) \rightarrow H_2O(l)$ is carried out at a constant pressure of 1 atm and a constant temperature of 25° C. What quantity of heat *q* will cross the system boundary (and in which direction?) For this reaction, the change in internal energy is ΔU = –281.8 kJ/mol.

Solution

281.8 kJ of heat. The work performed by the surroundings supplies an additional energy of $P\ \Delta V$ = 4.06 kJ to the system. The total energy change of the system is thus

q = (–281.8 + 4.06) k J = –285.8 kJ

(Eq. 4.2 above.) In order to maintain the constant 25° temperature, an equivalent quantity of heat must pass from the system to the surroundings.

This terminology can be somewhat misleading unless you bear in mind that the conditions ΔP and ΔT refer to the differences between the *inital* and *final* states of the system — that is, *before* and *after* the reaction. During the time the reaction is in progress, the temperature of the mixture will rise or fall, depending on whether the process is exothermic or endothermic. But because ΔT is a state function, its value is independent of what happens "in between" the initial state (reactants) and final state (products). The same is true of ΔV.

Enthalpy hides work and saves it too!

Because most chemical changes we deal with take place at constant pressure, it would be tedious to have to explicitly deal with the pressure-volume work details that were described above. Fortunately, chemists have found a way around this; they have simply defined a new state function that incorporates and thus hides within itself any terms relating to incidental kinds of work (*P-V*, electrical, etc.). Since both ΔP and ΔV in Equation 14.2.14 are state functions, then q_P, the heat that is absorbed or released when a process takes place at constant pressure, must also be a state function and is known as the *enthalpy change* ΔH

$$\Delta H \equiv q_P = \Delta U + P\Delta V \tag{14.2.15}$$

Since most processes that occur in the laboratory, on the surface of the earth, and in organisms do so under a constant pressure of one atmosphere, Equation 14.2.15 is the form of the First Law that is of greatest interest to most of us most of the time.

Example 14.2.3

Hydrogen chloride gas readily dissolves in water, releasing 75.3 kJ/mol of heat in the process. If one mole of HCl at 298 K and 1 atm pressure occupies 24.5 liters, find ΔU for the system when one mole of HCl dissolves in water under these conditions.

Solution

In this process the volume of liquid remains practically unchanged, so ΔV = –24.5 L. The work done is

$$\begin{aligned} w &= -P\Delta V \\ &= -(1\ atm)(-24.5\ L) \\ &= 24.6\ L-atm \end{aligned}$$

(The work is positive because it is being done on the system as its volume decreases due to the dissolution of the gas into the much smaller volume of the solution.) Using the conversion factor 1 L-atm = 101.33 J mol^{-1} and substituting in Equation 14.2.15 we obtain

$$\begin{aligned} \Delta U &= q + P\Delta V \\ &= -(75,300\ J) + [101.33\ J/L-atm)(24.5\ L-atm)] \\ &= -72.82\ kJ \end{aligned}$$

In other words, if the gaseous HCl could dissolve without volume change, the heat released by the process (75.3 kJ) would cause the system's internal energy to diminish by 75.3 kJ. But the disappearance of the gaseous phase reduces the volume of the system. This is equivalent to compression of the system by the pressure of the atmosphere performing work on it and consuming part of the energy that would otherwise be liberated, reducing the net value of ΔU to **–72.82 kJ**.

The Heat Capacity

For systems in which no change in composition (chemical reaction) occurs, things are even simpler: to a very good approximation, the enthalpy depends only on the temperature. This means that the temperature of such a system can serve as a direct measure of its enthalpy. The functional relation between the internal energy and the temperature is given by the *heat capacity* measured at constant pressure:

$$c_p = \frac{dH}{dT} \tag{14.2.16}$$

(or $\Delta H/\Delta T$ over a finite duration) An analogous quantity relates the heat capacity at constant *volume* to the internal energy:

$$c_v = \frac{dU}{dT} \tag{14.2.17}$$

The greater the heat capacity of a substance, the smaller will be the effect of a given absorption or loss of heat on its temperature. Heat capacity can be expressed in joules or calories per mole per degree (molar heat capacity), or in joules or calories per gram per degree; the latter is called the specific heat capacity or just the specific heat. The difference between c_p and c_v is of importance only when the volume of the system changes significantly— that is, when different numbers of moles of gases appear on either side of the chemical equation. For reactions involving only liquids and solids, c_p and c_v are for all practical purposes identical.

This page titled 14.2: The First Law of Thermodynamics is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14.3: Molecules as Energy Carriers and Converters

Learning Objectives

Make sure you thoroughly understand the following essential concepts:

- Describe the sources of potential energy and kinetic energy contained in a molecule.
- Describe the nature of "*thermal*" *energy*, and how it relates to other forms of kinetic energy and to *temperature*.
- Explain why the simplest molecules (monatomic and diatomic) have smaller heat capacities than polyatomic molecules.
- Similarly, explain why the dependence of heat capacity on the temperature is different for monatomic and polyatomic molecules.

All molecules at temperatures above absolue zero possess *thermal energy*— the randomized kinetic energy associated with the various motions the molecules as a whole, and also the atoms within them, can undergo. Polyatomic molecules also possess *potential energy* in the form of chemical bonds. Molecules are thus both vehicles for **storing and transporting energy**, and the means of **converting it from one form to another** when the formation, breaking, or rearrangement of the chemical bonds within them is accompanied by the uptake or release of *heat*.

Chemical Energy: Potential + Kinetic

When you buy a liter of gasoline for your car, a cubic meter of natural gas to heat your home, or a small battery for your flashlight, you are purchasing energy in a chemical form. In each case, some kind of a chemical change will have to occur before this energy can be released and utilized: the fuel must be burned in the presence of oxygen, or the two poles of the battery must be connected through an external circuit (thereby initiating a chemical reaction inside the battery.) And eventually, when each of these reactions is complete, our source of energy will be exhausted; the fuel will be used up, or the battery will be “dead”.

Where did the energy go? It could have gone to raise the temperature of the products, to perform work in expanding any gaseous products or to push electrons through a circuit. The remainder will reside in the chemical potential energy associated with the products of the reaction.

Chemical substances are made of atoms, or more generally, of positively charged nuclei surrounded by negatively charged electrons. A molecule such as dihydrogen, H_2, is held together by electrostatic attractions mediated by the electrons shared between the two nuclei. The total potential energy of the molecule is the sum of the repulsions between like charges and the attractions between electrons and nuclei:

$$PE_{total} = PE_{electron-electron} + PE_{nucleus-nucleus} + PE_{nucleus-electron} \tag{14.3.1}$$

In other words, the potential energy of a molecule depends on the time-averaged relative locations of its constituent nuclei and electrons. This dependence is expressed by the familiar potential energy curve which serves as an important description of the chemical bond between two atoms.

Translation refers to movement of an object as a complete unit. Translational motions of molecules in solids or liquids are restricted to very short distances comparable to the dimensions of the molecules themselves, whereas in gases the molecules typically travel hundreds of molecular diameters between collisions.

In gaseous hydrogen, for example, the molecules will be moving freely from one location to another; this is called *translational motion*, and the molecules therefore possess *translational kinetic energy*

$$KE_{trans} = \frac{mv^2}{2} \tag{14.3.2}$$

in which v stands for the average velocity of the molecules; you may recall from your study of gases that v, and therefore KE_{trans}, depends on the temperature.

In addition to translation, molecules composed of two or more atoms can possess other kinds of motion. Because a chemical

bond acts as a kind of spring, the two atoms in H_2 will have a natural ***vibrational frequency***. In more complicated molecules, many different modes of vibration become possible, and these all contribute a vibrational term KE_{vib} to the total kinetic energy. Finally, a molecule can undergo ***rotational*** motions which give rise to a third term KE_{rot}. Thus the total kinetic energy of a molecule is the sum

$$KE_{total} = KE_{trans} + KE_{vib} + KE_{rot} \tag{14.3.3}$$

The total energy of the molecule (its *internal energy U*) is just the sum

$$U = KE_{total} + PE_{total} \tag{14.3.4}$$

Although this formula is simple and straightforward, it cannot take us very far in understanding and predicting the behavior of even one molecule, let alone a large number of them. The reason, of course, is the chaotic and unpredictable nature of molecular motion. Fortunately, the behavior of a large collection of molecules, like that of a large population of people, can be described by statistical methods.

How Molecules Store Thermal Energy

As noted above, the *heat capacity* of a substance is a measure of how sensitively its temperature is affected by a change in heat content; the greater the heat capacity, the less effect a given flow of heat *q* will have on the temperature.

Thermal energy is randomized kinetic energy

We also pointed out that *temperature* is a measure of the average kinetic energy due to *translational motions* of molecules. If vibrational or rotational motions are also active, these will also accept thermal energy and reduce the amount that goes into translational motions. Because the temperature depends only on the latter, the effect of the other kinds of motions will be to reduce the dependence of the internal energy on the temperature, thus raising the *heat capacity* of a substance.

Table 14.3.1*: Molar heat capacities (kJ mol^{-1} K^{-1}) of some gaseous substances at constant pressure.*

monatomic		diatomic		triatomic	
He	20.5	CO	29.3	H_2O	33.5
Ne	20.5	N_2	29.5	D_2O	34.3
Ar	20.5	F_2	31.4	CO_2	37.2
Kr	20.5	Cl_2	33.9	CS_2	45.6

Whereas monatomic molecules can only possess translational thermal energy, two additional kinds of motions become possible in polyatomic molecules.

A linear molecule has an axis that defines two perpendicular directions in which rotations can occur; each represents an additional degree of freedom, so the two together contribute a total of ½ *R* to the heat capacity.

Vibrational and rotational motions are not possible for monatomic species such as the noble gas elements, so these substances have the lowest heat capacities. Moreover, as you can see in the leftmost column of Table 1, their heat capacities are all the same. This reflects the fact that translational motions are the same for all particles; all such motions can be resolved into three directions in space, each contributing one *degree of freedom* to the molecule and ½ *R* to its heat capacity. (*R* is the gas constant, 8.314 J K^{-1}).

Think of a "degree of freedom" as a kind of motion that adds kinetic energy to a molecule.

For a non-linear molecule, rotations are possible along all three directions of space, so these molecules have a rotational heat capacity of 3/2 *R*. Finally, the individual atoms within a molecule can move relative to each other, producing a vibrational motion. A molecule consisting of *N* atoms can vibrate in 3*N* –6 different ways or modes. Each vibrational mode contributes *R* (rather than ½ *R*) to the total heat capacity. (These results come from advanced mechanics and will not be proven here.)

Table 14.3.2*: Contribution of molecular motions to heat capacity*

type of motion →	translation	rotation	vibration
monatomic	3/2*R*	0	0
diatomic	3/2 *R*	*R*	*R*
polyatomic	3/2 *R*	3/2 *R*	3*N* – 6
separation between adjacent levels, (kJ mol^{-1})	6.0×10^{-17} J (O_2)	373 J (HCl)	373 J (HCl)

Monatomic molecules have the smallest heat capacities

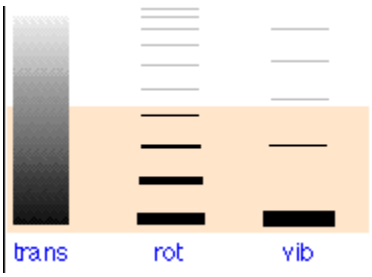
Now we are in a position to understand why more complicated molecules have higher heat capacities. The total kinetic energy of a molecule is the sum of those due to the various kinds of motions:

$$KE_{total} = KE_{trans} + KE_{rot} + KE_{vib} \tag{14.3.5}$$

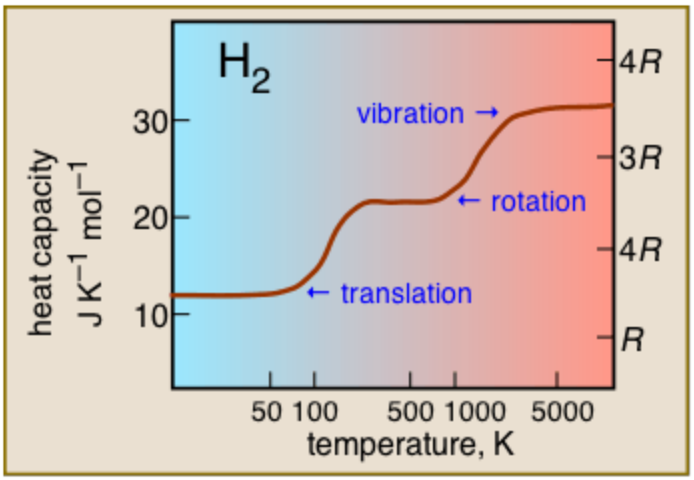
When a monatomic gas absorbs heat, all of the energy ends up in translational motion, and thus goes to increase its temperature. In a polyatomic gas, by contrast, the absorbed energy is partitioned among the other kinds of motions; since only the translational motions contribute to the temperature, the temperature rise is smaller, and thus the heat capacity is larger. There is one very significant complication, however: classical mechanics predicts that the energy is always partitioned equally between all degrees of freedom. Experiments, however, show that this is observed only at quite high temperatures. The reason is that these motions are all *quantized*. This means that only certain increments of energy are possible for each mode of motion, and unless a certain minimum amount of energy is available, a given mode will not be active at all and will contribute nothing to the heat capacity.

Translational energy levels are effectively a continuum

The shading indicates the average thermal energy available at 300 K. Only those levels within this range will have significant occupancy as indicated by the thickness of the lines in the two rightmost columns. At 300 K, only the lowest vibrational state and the first few rotational states will be active. Most of the thermal energy will be confined to the translational levels whose minute spacing (10^{-17} J) causes them to appear as a continuum.



Heat capacity of dihydrogen as a function of temperature. This plot is typical of those for other polyatomic molecules, and shows the practical consequences of the spacings of the various forms of thermal energy. Thus translational motions are available at virtually all temperatures, but contributions to heat acapacity by rotational or vibrational motions can only develop at temperatures sufficiently large to excite these motions.



It turns out that translational energy levels are spaced so closely that they these motions are active almost down to absolute zero, so all gases possess a heat capacity of at least 3/2 *R* at all temperatures. Rotational motions do not get started until intermediate temperatures, typically 300-500K, so within this range heat capacities begin to increase with temperature. Finally, at very high temperatures, vibrations begin to make a significant contribution to the heat capacity

The strong intermolecular forces of liquids and many solids allow heat to be channeled into vibrational motions involving more than a single molecule, further increasing heat capacities. One of the well known “anomalous” properties of liquid water is its high heat capacity (75 J mol^{-1} K^{-1}) due to intermolecular hydrogen bonding, which is directly responsible for the moderating influence of large bodies of water on coastal climates.

Heat capacities of metals

Metallic solids are a rather special case. In metals, the atoms oscillate about their equilibrium positions in a rather uniform way which is essentially the same for all metals, so they should all have about the same heat capacity. That this is indeed the case is embodied in the *Law of Dulong and Petit*. In the 19th century these workers discovered that the molar heat capacities of all the metallic elements they studied were around to 25 J mol^{-1} K^{-1}, which is close to what classical physics predicts for crystalline metals. This observation played an important role in characterizing new elements, for it provided a means of estimating their molar masses by a simple heat capacity measurement.

Standard enthalpy change

Under the special conditions in which the pressure is 1 atm and the reactants and products are at a temperature of 298 K, ΔH becomes the *standard enthalpy change* $\Delta H°$. Chemists usually refer to the "enthalpy change of a reaction" as simply the "enthalpy of reaction", or even more simply as the "*heat of reaction*". But students are allowed to employ this latter shortcut only if they are able to prove that they know the meaning of enthalpy.

Since most changes that occur in the laboratory, on the surface of the earth, and in organisms are subjected to an approximately constant pressure of "one atmosphere" and reasonably salubrious temperatures, most reaction heats quoted in the literature refer to $\Delta H°$. But the high pressures and extreme temperatures frequently encountered by chemical engineers, geochemists, and practicioners of chemical oceanography, often preclude the convenience of the "standard" values.

The rearrangement of atoms that occurs in a chemical reaction is virtually always accompanied by the liberation or absorption of heat. If the purpose of the reaction is to serve as a source of heat, such as in the combustion of a fuel, then these heat effects are of direct and obvious interest. We will soon see, however, that a study of the energetics of chemical reactions in general can lead us to a deeper understanding of chemical equilibrium and the basis of chemical change itself.

In chemical thermodynamics, we define the zero of the enthalpy and internal energy as that of the elements as they exist in their stable forms at 298K and 1 atm pressure. Thus the enthalpies *H* of Xe(g), O_2(g) and C(diamond) are all zero, as are those of H_2 and Cl_2 in the reaction

$$H_{2(g)} + Cl_{2(g)} \rightarrow 2HCl_{(g)} \tag{14.3.6}$$

The enthalpy of two moles of HCl is smaller than that of the reactants, so the difference is released as heat. Such a reaction is said to be *exothermic*. The reverse of this reaction would absorb the same quantity of heat from the surroundings and be *endothermic*. In comparing the internal energies and enthalpies of different substances as we have been doing here, it is important to compare equal numbers of moles, because energy is an *extensive property* of matter. However, heats of reaction are commonly expressed on a molar basis and treated as *intensive properties*.

Changes in enthalpy and internal energy

We can characterize any chemical reaction by the change in the internal energy or enthalpy:

$$\Delta H = H_{final} - H_{initial} \tag{14.3.7}$$

The significance of this can hardly be exaggerated because ΔH, being a state function, is entirely independent of how the system gets from the initial state to the final state. In other words, the value of ΔH or ΔU for a given change in state is independent of the *pathway* of the process.

Consider, for example, the oxidation of a lump of sugar to carbon dioxide and water:

$$\mathrm{C_{12}H_{22}O_{11} + 12\,O_2\,(g) \longrightarrow 12\,CO_2\,(g) + 11\,H_2O\,(l)} \tag{14.3.8}$$

This process can be carried out in many ways, for example by burning the sugar in air, or by eating the sugar and letting your body carry out the oxidation. Although the mechanisms of the transformation are completely different for these two pathways, the overall change in the enthalpy of the system (the atoms of carbon, hydrogen and oxygen that were originally in the sugar) will be identical, and can be calculated simply by looking up the standard enthalpies of the reactants and products and calculating the difference

$$\Delta H = [12 \times H(\mathrm{CO_2})] + [11 \times H(\mathrm{H_2O})] - H(\mathrm{C_{12}H_{22}O_{11}}) = -5606\,kJ \tag{14.3.9}$$

The same quantity of heat is released whether the sugar is burnt in the air or oxidized in a series of enzyme-catalyzed steps in your body.

Enthalpy increases with temperature

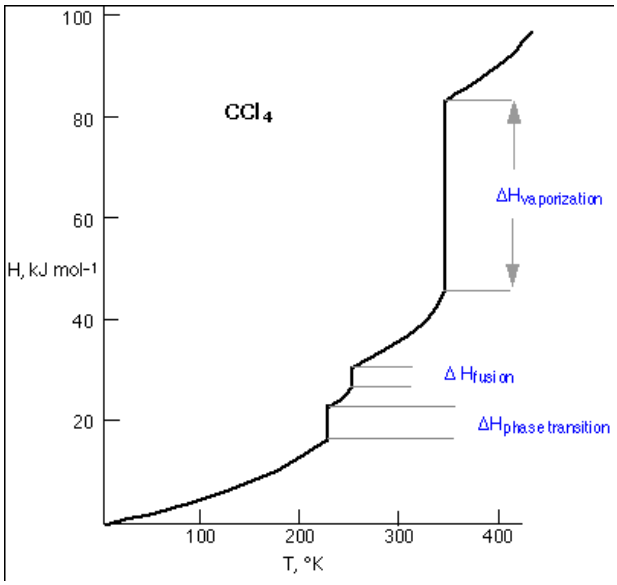
When the temperature of a substance is raised, it absorbs heat. The enthalpy of a system increases with the temperature by the amount $\Delta H = C_p\,\Delta T$. The defining relation

$$\Delta H = \Delta U + P\,\Delta V \tag{14.3.10}$$

tells us that this change is dominated by the internal energy, subject to a slight correction for the work associated with volume change. Heating a substance causes it to expand, making ΔV positive and causing the enthalpy to increase slightly more than the internal energy. Physically, what this means is that if the temperature is increased while holding the pressure constant, some extra energy must be expended to push back the external atmosphere while the system expands. The difference between the dependence of *U* and *H* on temperature is only really significant for gases, since the coefficients of thermal expansion of liquids and solids are very small.

Enthalpy of phase changes

A plot of the enthalpy of a system as a function of its temperature is called an *enthalpy diagram*. The slope of the line is given by C_p. The enthalpy diagram of a pure substance such as water shows that this plot is not uniform, but is interrupted by sharp breaks at which the value of C_p is apparently infinite, meaning that the substance can absorb or lose heat without undergoing any change in temperature at all. This, of course, is exactly what happens when a substance undergoes a phase change; you already know that the temperature the water boiling in a kettle can never exceed 100 until all the liquid has evaporated, at which point the temperature of the steam will rise as more heat flows into the system.



A plot of the enthalpy of carbon tetrachloride as a function of its temperature provides a concise view of its thermal behavior. The slope of the line is given by the heat capacity C_p. All *H*-vs.-*C* plots show sharp breaks at which the value of C_p is apparently infinite, meaning that the substance can absorb or lose heat without undergoing any change in temperature at all. This, of course, is exactly what happens when a substance undergoes a **phase change**; you already know that the temperature of the water boiling in a kettle can never exceed 100°C until all the liquid has evaporated, at which point the temperature (of the steam) will rise as more heat flows into the system.

The lowest-temperature discontinuity on the CCl_4 diagram corresponds to a solid-solid phase transition associated with a rearrangement of molecules in the crystalline solid.

This page titled 14.3: Molecules as Energy Carriers and Converters is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14.4: Thermochemistry and Calorimetry

Learning Objectives

Make sure you thoroughly understand the following essential concepts:

- Write a *balanced thermochemical equation* that expresses the enthalpy change for a given chemical reaction. Be sure to correctly specificy the *physical state* (and, if necessary, the *concentration*) of each component.
- Write an equation that defines the *standard enthalpy of formation* of a given chemical species.
- Define the *standard enthalpy change* of a chemical reaction. Use a table of standard enthalpies of formation to evaluate $\Delta H_f°$.
- State the principle on which *Hess' law* depends, and explain why this law is so useful.
- Describe a simple *calorimeter* and explain how it is employed and how its heat capacity is determined.

The heat that flows across the boundaries of a system undergoing a change is a fundamental property that characterizes the process. It is easily measured, and if the process is a chemical reaction carried out at constant pressure, it can also be predicted from the difference between the enthalpies of the products and reactants. The quantitative study and measurement of heat and enthalpy changes is known as *thermochemistry*.

Thermochemical Equations and Standard States

In order to define the thermochemical properties of a process, it is first necessary to write a *thermochemical equation* that defines the actual change taking place, both in terms of the formulas of the substances involved and their physical states (temperature, pressure, and whether solid, liquid, or gaseous.

To take a very simple example, here is the complete thermochemical equation for the vaporization of water at its normal boiling point:

$$\mathrm{H_2O}(\ell,\ 373\,\mathrm{K},\,1\,\mathrm{atm}) \longrightarrow \mathrm{H_2O}\,(\mathrm{g},\,373\,\mathrm{K},\,1\,\mathrm{atm})\ \ \Delta H = 40.7\,kJ\,mol^{-1}$$

The quantity 40.7 is known as the *enthalpy of vaporization* (often referred to as “heat of vaporization”) of liquid water. The following points should be kept in mind when writing thermochemical equations:

Any thermodynamic quantity such as ΔH that is associated with a thermochemical equation **always** refers to the number of moles of substances explicitly shown in the equation. Thus for the synthesis of water we can write

$$2\,\mathrm{H_2}\,(\mathrm{g}) + \mathrm{O_2}(\mathrm{g}) \longrightarrow 2\,\mathrm{H_2O}\,(\mathrm{l})\ \ \Delta H = -572\,kJ$$

or

$$\mathrm{H_2}\,(\mathrm{g}) + \tfrac{1}{2}\,\mathrm{O_2}(\mathrm{g}) \longrightarrow \mathrm{H_2O}\,(\mathrm{l})\ \ \Delta H = -286\,kJ$$

The thermochemical equations for reactions taking place in solution **must** also specify the concentrations of the dissolved species. For example, the enthalpy of neutralization of a strong acid by a strong base is given by

$$\mathrm{H^+}\,(\mathrm{aq},\,1\mathrm{M},\,298\,\mathrm{K},\,1\,\mathrm{atm}) + \mathrm{OH^-}\,(\mathrm{aq},\,1\mathrm{M},\,298\,\mathrm{K},\,1\,\mathrm{atm}) \longrightarrow \mathrm{H_2O}(\ell, 373\ \mathrm{K},\ 1\,\mathrm{atm})\ \ \Delta H = -56.9\,kJ\,mol^{-1}$$

in which the abbreviation *aq* refers to the hydrated ions as they exist in aqueous solution. Since most thermochemical equations are written for the standard conditions of 298 K and 1 atm pressure, we can leave these quantities out if these conditions apply both before and after the reaction. If, under these same conditions, the substance is in its preferred (most stable) physical state, then the substance is said to be in its **standard state**.

Thus the standard state of water at 1 atm is the solid below 0°C, and the gas above 100°C. A thermochemical quantity such as ΔH that refers to reactants and products in their standard states is denoted by $\Delta H°$.

Concentrations vs. Effective Concentrations (Activities)

In the case of dissolved substances, the standard state of a solute is that in which the "effective concentration", known as the activity, is unity. For non-ionic solutes the activity and molarity are usually about the same for concentrations up to about 1M, but for an ionic solute this approximation is generally valid only for solutions more dilute than 0.001-0.01M, depending on electric charge and size of the particular ion.

Standard Enthalpy of Formation

The enthalpy change for a chemical reaction is the difference

$$\Delta H = H_{products} - H_{reactants} \tag{14.4.1}$$

If the reaction in question represents the formation of one mole of the compound from its elements in their standard states, as in

$$\mathrm{H_2\,(g)} + \tfrac{1}{2}\,\mathrm{O_2\,(g)} \longrightarrow \mathrm{H_2O\,(l)} \quad \Delta H = -286\ kJ \tag{14.4.2}$$

then we can arbitrarily set the enthalpy of the **elements** to zero and write

$$\begin{aligned} H_f^o &= \sum H_f^o(products) - \sum H_f^o(reactants) \\ &= -286\ kJ - 0 \\ &= -268\ kJ\,mol^{-1} \end{aligned}$$

which defines the *standard enthalpy of formation* of water at 298 K. The value H_f ° = -268 kJ tells us that when hydrogen and oxygen, each at a pressure of 1 atm and at 298 K (25° C) react to form 1 mole of liquid water also at 25°C and 1 atm pressure, 268 kJ will have passed from the system (the reaction mixture) into the surroundings. The negative sign indicates that the reaction is exothermic: the enthalpy of the product is smaller than that of the reactants. The standard enthalpy of formation is a fundamental property of any compound. Table T1 list H_f ° values (usually alongside values of other thermodynamic properties) in their appendices.

> The standard enthalpy of formation of a compound is defined as the heat associated with the formation of one mole of the compound from its elements in their standard states.

In general, the *standard enthalpy change for a reaction* is given by the expression

$$\Delta H_f^o = \sum H_f^o(products) - \sum H_f^o(reactants) \tag{14.4.3}$$

in which the $\sum H_f^o$ terms indicate the sums of the standard enthalpies of formations of all products and reactants. The above definition is one of the most important in chemistry because it allows us to predict the enthalpy change of any reaction without knowing any more than the standard enthalpies of formation of the products and reactants, which are widely available in tables.

The following examples illustrate some important aspects of the standard enthalpy of formation of substances.

Molarity

The thermochemical equation defining H_f^o is always written in terms of one mole of the substance in question> For example, the relavante thermodynamic equation for the heat of formation of ammonia ($\mathrm{NH_3}$) is:

$$\tfrac{1}{2}\,\mathrm{N_2\,(g)} + \tfrac{3}{2}\,\mathrm{H_2(g)} \longrightarrow \mathrm{NH_3\,(g)} \;\; \Delta H^o = -46.1\,kJ\,(\text{per mole of } \mathrm{NH_3})$$

Allotropes

The standard heat of formation of a compound is always taken in reference to the forms of the elements that are most stable at 25°C and 1 atm pressure.

A number of elements, of which sulfur and carbon are common examples, can exist in more then one solid crystalline form (called allotropes). For carbon dixoide ($ceCO2$, one can construct two thermodynamics equations:

$$\mathrm{C(graphite) + O_2\,(g) \longrightarrow CO_2\,(g)} \;\; \Delta H^o \equiv H_f^o = -393.5\,kJ\,mol^{-1}$$

$$\mathrm{C(diamond) + O_2\,(g) \longrightarrow CO_2\,(g)} \;\; \Delta H^o = -395.8\,kJ\,mol^{-1}$$

However for carbon, the graphite form is the more stable form and the correct thermodynamic equations for the heat of formation.

Heats of Vaporization

The physical state of the product of the formation reaction must be indicated explicitly if it is not the most stable one at 25°C and 1 atm pressure:

$$\mathrm{H_2\,(g) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow H_2O\,(aq)} \;\; \Delta H^o \equiv H_f^o = -285.8\,kJ\,mol^{-1}$$

$$\mathrm{H_2\,(g) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow H_2O\,(g)} \;\; \Delta H^o = -241.8\,kJ\,mol^{-1}$$

Notice that the difference between these two ΔH^o values is just the *heat of vaporization* of water.

Endothermic Heats of Formation

Although the formation of most molecules from their elements is an exothermic process, the formation of some compounds is mildly endothermic:

$$\mathrm{\tfrac{1}{2}\,N_2\,(g) + O_2\,(g) \longrightarrow NO_2\,(g)} \;\; \Delta H^o \equiv H_f^\circ = +33.2\,kJ\,mol^{-1}$$

A positive heat of formation is frequently associated with instability— the tendency of a molecule to decompose into its elements, although it is not in itself a sufficient cause. In many cases, however, the rate of this decomposition is essentially zero, so it is still possible for the substance to exist. In this connection, it is worth noting that all molecules will become unstable at higher temperatures.

Non-Existing Compounds

The thermochemical reactions that define the heats of formation of most compounds cannot actually take place. For example, the direct synthesis of methane from its elements

$$\mathrm{C(graphite) + 2\,H_2\,(g) \longrightarrow CH_4\,(g)}$$

cannot be observed directly owing to the large number of other possible reactions between these two elements. However, the standard enthalpy change for such a reaction be found indirectly from other data, as explained in the next section.

Heats of Atomization

The standard enthalpy of formation of gaseous atoms from the element is known as the *heat of atomization*.

Heats of atomization are always positive, and are important in the calculation of bond energies.

$$\mathrm{Fe\,(s)} \longrightarrow \mathrm{Fe\,(g)}\ \Delta H°= 417\,kJ\,mol^{-1}$$

Ions in Solution

The standard enthalpy of formation of an ion dissolved in water is expressed on a separate scale in which that of $\mathrm{H^+\,(aq)}$ is **defined as zero**.

The standard heat of formation of a **dissolved ion** such as $\mathrm{Cl^-\,(aq)}$ (that is, formation of the ion from the element) cannot be measured because it is impossible to have a solution containing a single kind of ion. For this reason, ionic enthalpies are expressed on a separate scale on which $\mathrm{H}_f{}^o$ of the hydrogen ion at *unit activity* (1 M effective concentration) is defined as zero:

$$\tfrac{1}{2}\,\mathrm{H_2\,(g)} \longrightarrow \mathrm{H^+\,(aq)}\ \ \Delta H^o \equiv H_f^o = 0\,kJ\,mol^{-1}$$

Other ionic enthalpies (as they are commonly known) are found by combining appropriate thermochemical equations (as explained in Section 3 below). For example, $H_f°$ of HCl(aq) is found from the enthalpies of formation and solution of HCl(g), yielding

$$\tfrac{1}{2}\,\mathrm{H_2\,(g)} + \tfrac{1}{2}\,\mathrm{Cl_2\,(g)} \longrightarrow \mathrm{HCl\,(aq)}\quad \Delta H^o \equiv H_f^o = -167\,kJ\,mol^{-1}$$

Because $H_f°$ for H^+(aq) is zero, this value establishes the standard enthalpy of the chloride ion. The standard enthalpy of formation of Ca^{2+}(aq), given by

$$\mathrm{Ca\,(s) + Cl_2\,(g)} \longrightarrow \mathrm{CaCl_2\,(aq)}$$

could then be calculated by combining other measurable quantities such as the enthalpies of formation and solution of $\mathrm{CaCl_2\,(s)}$ to find $H_f°$ for $CaCl_2$(aq), from which $H_f°$ of $\mathrm{Ca^{2+}\,(aq)}$ is found by difference from that of $\mathrm{Cl^-\,(aq)}$. Tables of the resulting ionic enthalpies are widely available (see here).

Hess' Law and Thermochemical Calculations

Two or more chemical equations can be combined algebraically to give a new equation. Even before the science of thermodynamics developed in the late nineteenth century, it was observed that the heats associated with chemical reactions can be combined in the same way to yield the heat of another reaction. For example, the standard enthalpy changes for the oxidation of graphite and diamond can be combined to obtain ΔH° for the transformation between these two forms of solid carbon, a reaction that cannot be studied experimentally.

$$\mathrm{C(graphite) + O_2(g)} \longrightarrow \mathrm{CO_2\,(g)}\ \ \Delta H^o = -393.51\,kJ\,mol^{-1} \tag{14.4.4}$$

$$\mathrm{C(diamond) + O_2(g)} \longrightarrow \mathrm{CO_2\,(g)}\ \ \Delta H^o = -395.40\,kJ\,mol^{-1} \tag{14.4.5}$$

Subtraction of the second reaction from the first (i.e., writing the second equation in reverse and adding it to the first one) yields

$$\mathrm{C(graphite)} \longrightarrow \mathrm{C(diamond)}\ \ \Delta H^\circ = 1.89\,kJ\,mol^{-1} \tag{14.4.6}$$

This principle, known as *Hess' law of independent heat summation* is a direct consequence of the enthalpy being a state function. Hess' law is one of the most powerful tools of chemistry, for it allows the change in the enthalpy (and in other thermodynamic functions) of huge numbers of chemical reactions to be predicted from a relatively small base of experimental data.

Germain Henri Hess

Germain Henri Hess (1802-1850) was a Swiss-born professor of chemistry at St. Petersburg, Russia. He formulated his famous law, which he discovered empirically, in 1840. Very little appears to be known about his other work in chemistry.

Standard Enthalpies of Combustion

Because most substances cannot be prepared directly from their elements, heats of formation of compounds are seldom determined by direct measurement. Instead, Hess' law is employed to calculate enthalpies of formation from more accessible data. The most important of these are the standard *enthalpies of combustion*. Most elements and compounds combine with oxygen, and many of these oxidations are highly exothermic, making the measurement of their heats relatively easy. For example, by combining the heats of combustion of carbon, hydrogen, and methane, we can obtain the standard enthalpy of formation of methane, which as we noted above, cannot be determined directly (Example 14.4.1).

Example 14.4.1: Combustion of Methane

Use the following heat of formation/combustion information to estimate the standard heat of formation of methane $\mathrm{CH_4}$.

Multiple \label

Solution

The standard heat of formation of methane is defined by the reaction

$$\mathrm{C(graphite) + 2\,H_2\,(g) \longrightarrow CH_4\,(g)} \quad \Delta H^o = ? \tag{14.4.7}$$

Our task is thus to combine the top three equations in such a way that they add up to (4).

Step 1

Begin by noting that (3), the combustion of methane, is the only equation that contains the CH_4 term, so we need to write it in reverse (not forgetting to reverse the sign of $\Delta H°$!) so that CH_4 appears as the product.

$$\mathrm{CO_2\,(g) + 2\,H_2O\,(g) \longrightarrow CH_4\,(g) + 2\,O_2\,(g)} \quad \Delta H^o = +890\,kJ\,mol^{-1} \tag{14.4.8}$$

Step 2

Since H_2O does not appear in the net reaction (4), add two times (2) to cancel these out. Notice that this also cancels one of the oxygens in (3Rev):

$CO_2(g)$ + $2H_2O(g)$ → **$CH_4(g)$** + $2O_2(g)$ $\Delta H°$ = +890 kJ mol^{-1}(P1-3Rev)
→ $2H_2O(g)$ $\Delta H°$ = -484 kJ mol^{-1}(P1-2)

Step 3

Finally, get rid of the remaining O_2 and CO_2 by adding (1); this also adds a needed C:

→ **$CH_4(g)$** + $2O_2(g)$ $\Delta H°$ = +890 kJ mol^{-1}(P1-3Rev)
→ $2H_2O(g)$ $\Delta H°$ = -572 kJ mol^{-1}(P1-2)
→ $CO_2(g)$ $\Delta H°$ = -393 kJ mol^{-1}(P1-1)

Step 4

So our creative cancelling has eliminated all except the substances that appear in (4). Just add up the enthalpy changes and we are done:

$$\mathrm{C(graphite) + 2\,H_2\,(g) \longrightarrow CH_4\,(g)} \tag{14.4.9}$$

(The tablulated value is -74.6 kJ mol^{-1})

Calorimetry: Measuring ΔH in the laboratory

How are enthalpy changes determined experimentally? First, you must understand that the only thermal quantity that can be observed directly is the heat q that flows into or out of a reaction vessel, and that q is numerically equal to $\Delta H°$ only under the special condition of constant pressure. Moreover, q is equal to the standard enthalpy change only when the reactants and products are both at the same temperature, normally 25°C. The measurement of q is generally known as *calorimetry*.

The most common types of calorimeters contain a known quantity of water which absorbs the heat released by the reaction. Because the specific heat capacity of water (4.184 J g^{-1} K^{-1}) is known to high precision, a measurement of its temperature rise due to the reaction enables one to calculate the quantity of heat released.

In all but the very simplest calorimeters, some of the heat released by the reaction is absorbed by the components of the calorimeter itself. It is therefore necessary to "calibrate" the calorimeter by measuring the temperature change that results from the introduction of a known quantity of heat. The resulting *calorimeter constant*, expressed in J K^{-1}, can be regarded as the “heat capacity of the calorimeter”. The known source of heat is usually produced by passing a known quantity of electric current through a resistor within the calorimeter, but it can be measured by other means as described in the following problem example.

Example 14.4.2: Heat Capacity of a Calorimeter

In determining the heat capacity of a calorimeter, a student mixes 100.0 g of water at 57.0 °C with 100.0 g of water, already in the calorimeter, at 24.2°C. (yhe specific heat of water is 4.184 J g^{-1} K^{-1}). After mixing and thermal equilibration with the calorimeter, the temperature of the water stabilizes at 38.7°C. Calculate the heat capacity of the calorimeter in J/K.

Solution

The hot water loses heat, the cold water gains heat, and the calorimter itself gains heat, so this is essentially a thermal balance problem. Conservation of energy requires that

$$q_{hot} + q_{cold} + q_{cal} = 0$$

We can evaluate the first two terms from the observed temperature changes:

$$q_{hot} = (100\ g)(38.7\ K - 57.0\ K)(4.184\ J\ g^{-1}K^{-1}) = -7,657\ J$$

$$q_{cold} = (100\ g)(38.7\ K - 24.2\ K)(4.184\ J\ g^{-1}K^{-1}) = 6,067\ J$$

So

$$q_{cal} = 7,657\ J - , 6067\ J = 1,590;\ J$$

The ***calorimeter constant*** *is*

$$\frac{1590\ J}{38.7\ K - 24.2\ K} = 110\ J\ K^{-1}$$

Note: Strictly speaking, there is a fourth thermal balance term that must be considered in a highly accurate calculation: the water in the calorimeter expands as it is heated, performing work on the atmosphere.

For reactions that can be initiated by combining two solutions, the temperature rise of the solution itself can provide an approximate value of the reaction enthalpy if we assume that the heat capacity of the solution is close to that of the pure water — which will be nearly true if the solutions are dilute.

For example, a very simple calorimetric determination of the standard enthalpy of the reaction

$$\mathrm{H^+(aq) + OH^-(aq) \longrightarrow H_2O(\ell)}$$

could be carried out by combining equal volumes of 0.1M solutions of HCl and of NaOH initially at 25°C. Since this reaction is exothermic, a quantity of heat q will be released into the solution. From the temperature rise and the specific heat of water, we

obtain the number of joules of heat released into each gram of the solution, and q can then be calculated from the mass of the solution. Since the entire process is carried out at constant pressure, we have $\Delta H° = q$.

For reactions that cannot be carried out in dilute aqueous solution, the reaction vessel is commonly placed within a larger insulated container of water. During the reaction, heat passes between the inner and outer containers until their temperatures become identical. Again, the temperature change of the water is observed, but in this case we need to know the value of the calorimeter constant described above.

The Bomb Calorimeter

$\Delta H_{combustion}$, since these are essential to the determination of standard enthalpies of formation of the thousands of new compounds that are prepared and characterized each month. In a constant volume calorimeter, the system is sealed or isolated from its surroundings, and this accounts for why its volume is fixed and there is no volume-pressure work done. A bomb calorimeter structure consists of the following:

- Steel bomb which contains the reactants
- Water bath in which the bomb is submerged
- Thermometer
- A motorized stirrer
- Wire for ignition

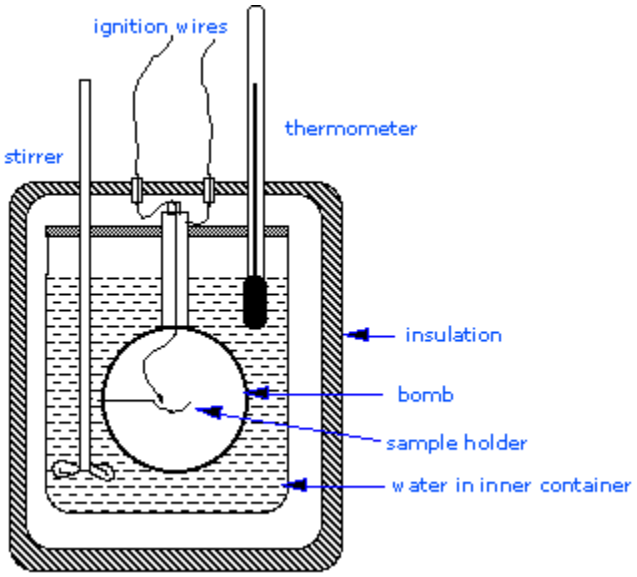


Figure 1: A simplified bomb calorimeter

Since the process takes place at constant volume, the reaction vessel must be constructed to withstand the high pressure resulting from the combustion process, which amounts to a confined explosion. The vessel is usually called a “bomb”, and the technique is known as *bomb calorimetry*. The reaction is initiated by discharging a capacitor through a thin wire which ignites the mixture.

Another consequence of the constant-volume condition is that the heat released corresponds to q_V, and thus to the internal energy change ΔU rather than to ΔH. The enthalpy change is calculated according to the formula

$$\Delta H = q_v + \Delta n_g RT \tag{14.4.10}$$

- Δn_g is the change in the number of moles of gases in the reaction.

Example 14.4.3: Combustion of Biphenyl

A sample of biphenyl $(C_6H_5)_2$ weighing 0.526 g was ignited in a bomb calorimeter initially at 25°C, producing a temperature rise of 1.91 K. In a separate calibration experiment, a sample of benzoic acid C_6H_5COOH weighing 0.825 g was ignited under identical conditions and produced a temperature rise of 1.94 K. For benzoic acid, the heat of combustion at constant pressure is known to be 3226 kJ mol^{-1} (that is, ΔU = -3226 kJ mol^{-1}.) Use this information to determine the standard enthalpy of combustion of biphenyl.

Solution

Begin by working out the calorimeter constant:

- Moles of benzoic acid:

$$\frac{(0.825g}{122.1\ g/mol} = 0.00676\ mol$$

- Heat released to calorimeter:

$$(0.00676\ mol) \times (3226\ kJ/mol) = 21.80\ kJ$$

- Calorimeter constant:

$$\frac{21.80\ kJ}{1.94\ K} = 11.24\ kJ/K$$

Now determine $\Delta U_{combustion}$ of the biphenyl ("BP"):

- moles of biphenyl:

$$\frac{0.526\ g}{154.12\ g/mol} = 0.00341\ mol$$

- heat released to calorimeter:

$$(1.91\ K) \times (11.24\ kJ/K) = 21.46\ kJ$$

- heat released per mole of biphenyl:

$$\frac{21.46\ kJ}{0.00341\ mol} = 6,293\ kJ/mol$$

$$\Delta U_{combustion}(BP) = -6,293\ kJ/mol$$

(This is the heat change at constant volume, q_v; the negative sign indicates that the reaction is exothermic, as all combustion reactions are.)

From the reaction equation

$$(C_6H_5)_{2(s)} + \frac{19}{2}O_{2(g)} \rightarrow 12CO_{2(g)} + 5H_2O_{(l)}$$

we have

$$\Delta n_g = 12 - \frac{19}{2} = \frac{-5}{2}$$

Thus the volume of the system *decreases* when the reaction takes place. Converting to ΔH, we can write the following equation. Additionally, recall that at constant volume, $\Delta U = q_V$.

$$\begin{aligned}\Delta H &= q_V + \Delta n_g RT \\ &= \Delta U - \left(\frac{5}{2}\right)(8.314\ J\ mol^{-1}\ K^{-1})(298\ K) \\ &= (-6,293\ kJ/mol) - (6,194\ J/mol) = (-6,293 - 6.2)\ kJ/mol = -6299\ kJ/mol\end{aligned}$$

A common mistake here is to forget that the subtracted term is in J, not kJ. Note that the additional 6.2 kJ in ΔH compared to ΔU reflects the work that the surroundings do on the system as the volume of gases decreases according to the reaction equation.

Determining the Heat of Reaction

The amount of heat that the system gives up to its surroundings so that it can return to its initial temperature is the heat of reaction. The heat of reaction is just the negative of the thermal energy gained by the calorimeter and its contents ($q_{calorimeter}$) through the combustion reaction.

$$q_{rxn} = -q_{calorimeter} \tag{14.4.11}$$

where

$$q_{calorimeter} = q_{bomb} + q_{water} \tag{14.4.12}$$

If the constant volume calorimeter is set up the same way as before, (same steel bomb, same amount of water, etc.) then the heat capacity of the calorimeter can be measured using the following formula:

$$q_{calorimeter} = (\text{ heat capacity of calorimeter}) \times \Delta T \tag{14.4.13}$$

Heat capacity is defined as the amount of heat needed to increase the temperature of the entire calorimeter by 1 °C. The equation above can also be used to calculate q_{rxn} from $q_{calorimeter}$ calculated by Equation 2. The heat capacity of the calorimeter can be determined by conducting an experiment.

Example 14.4.4: Heat of Combustion

1.150 g of sucrose goes through combustion in a bomb calorimeter. If the temperature rose from 23.42°C to 27.64°C and the heat capacity of the calorimeter is 4.90 kJ/°C, then determine the heat of combustion of sucrose, $C_{12}H_{22}O_{11}$, in kJ per mole of $C_{12}H_{22}O_{11}$.

Given:

- mass of $C_{12}H_{22}O_{11}$: 1.150 g
- $T_{initial}$: 23.42°C
- T_{final}:27.64°C
- Heat Capacity of Calorimeter: 4.90 kJ/°C

Solution

Using Equation 4 to calculate $q_{calorimeter}$:

$$q_{calorimeter} = (4.90\ kJ/°C) \times (27.64 - 23.42)°C = (4.90 \times 4.22)\ kJ = 20.7\ kJ \tag{14.4.14}$$

Plug into Equation 2:

$$q_{rxn} = -q_{calorimeter} = -20.7\ \ kJ$$

But the question asks for kJ/mol $C_{12}H_{22}O_{11}$, so this needs to be converted:

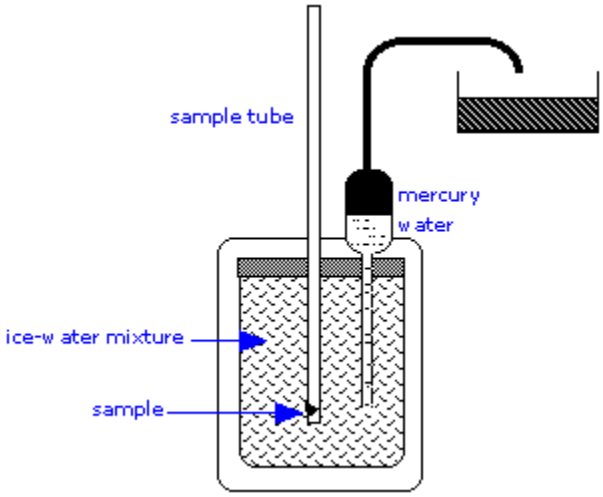
$$q_{rxn} = \frac{-20.7\ kJ}{1.150\ g\ C_{12}H_{22}O_{11}} = \frac{-18.0\ kJ}{g\ C_{12}H_{22}O_{11}}$$

Per Mole $C_{12}H_{22}O_{11}$:

$$q_{rxn} = \frac{-18.0\ kJ}{g\ C_{12}H_{22}O_{11}} \times \frac{342.3\ g\ C_{12}H_{22}O_{11}}{1\ mol\ C_{12}H_{22}O_{11}} = \frac{-6.16 \times 10^3\ kJ}{mol\ C_{12}H_{22}O_{11}}$$

Calorimeters

Although calorimetry is simple in principle, its practice is a highly exacting art, especially when applied to processes that take place slowly or involve very small heat changes, such as the germination of seeds. Calorimeters can be as simple as a foam plastic coffee cup, which is often used in student laboratories. Research-grade calorimeters, able to detect minute temperature changes, are more likely to occupy table tops, or even entire rooms:



The **ice calorimeter** is an important tool for measuring the heat capacities of liquids and solids, as well as the heats of certain reactions. This simple yet ingenious apparatus is essentially a device for measuring the change in volume due to melting of ice. To measure a heat capacity, a warm sample is placed in the inner compartment, which is surrounded by a mixture of ice and water.

The heat withdrawn from the sample as it cools causes some of the ice to melt. Since ice is less dense than water, the volume of water in the insulated chamber decreases. This causes an equivalent volume of mercury to be sucked into the inner reservoir from the outside container. The loss in weight of this container gives the decrease in volume of the water, and thus the mass of ice melted. This, combined with the heat of fusion of ice, gives the quantity of heat lost by the sample as it cools to 0°C.

This page titled 14.4: Thermochemistry and Calorimetry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14.5: Calorimetry

Learning Objectives

Make sure you thoroughly understand the following essential concept:

- Describe a simple *calorimeter* and explain how it is employed and how its heat capacity is determined.

Constant Volume Calorimetry, also know as bomb calorimetry, is used to measure the heat of a reaction while holding volume constant and resisting large amounts of pressure. Although these two aspects of bomb calorimetry make for accurate results, they also contribute to the difficulty of bomb calorimetry. In this module, the basic assembly of a bomb calorimeter will be addressed, as well as how bomb calorimetry relates to the heat of reaction and heat capacity and the calculations involved in regards to these two topics.

Introduction

Calorimetry is used to measure quantities of heat, and can be used to determine the heat of a reaction through experiments. Usually a coffee-cup calorimeter is used since it is simpler than a bomb calorimeter, but to measure the heat evolved in a combustion reaction, constant volume or bomb calorimetry is ideal. A constant volume calorimeter is also more accurate than a coffee-cup calorimeter, but it is more difficult to use since it requires a well-built reaction container that is able to withstand large amounts of pressure changes that happen in many chemical reactions.

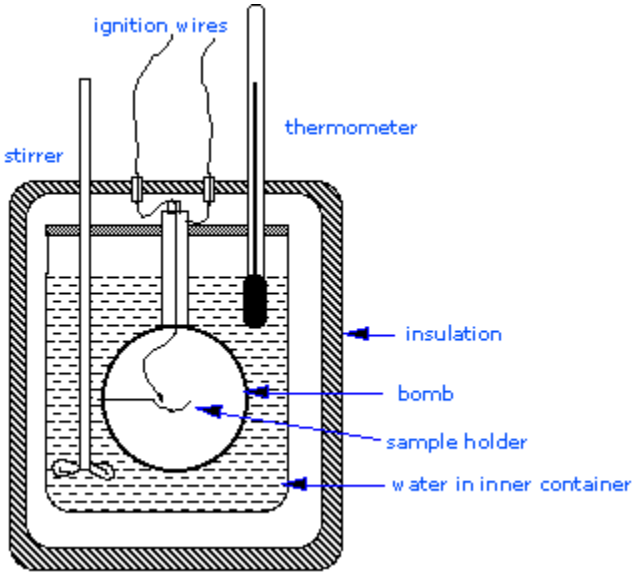


Figure 14.5.1: A simplified bomb calorimeter.

Most serious calorimetry carried out in research laboratories involves the determination of heats of combustion $\Delta H_{combustion}$, since these are essential to the determination of standard enthalpies of formation of the thousands of new compounds that are prepared and characterized each month. In a constant volume calorimeter, the system is sealed or isolated from its surroundings, which accounts for why its volume is fixed and there is no volume-pressure work done. A bomb calorimeter structure consists of the following:

- Steel bomb which contains the reactants
- Water bath in which the bomb is submerged
- Thermometer
- A motorized stirrer
- Wire for ignition

Since the process takes place at constant volume, the reaction vessel must be constructed to withstand the high pressure resulting from the combustion process, which amounts to a confined explosion. The vessel is usually called a “bomb”, and the technique is known as *bomb calorimetry*. The reaction is initiated by discharging a capacitor through a thin wire which ignites the mixture.

Another consequence of the constant-volume condition is that the heat released corresponds to q_v, and thus to the internal energy change ΔU rather than to ΔH. The enthalpy change is calculated according to the formula

$$\Delta H = q_v + \Delta n_g RT \tag{14.5.1}$$

where Δn_g is the change in the number of moles of gases in the reaction.

Example 14.5.1: Combustion of Biphenyl

A sample of biphenyl ($(C_6H_5)_2$) weighing 0.526 g was ignited in a bomb calorimeter initially at 25°C, producing a temperature rise of 1.91 K. In a separate calibration experiment, a sample of benzoic acid (C_6H_5COOH) weighing 0.825 g was ignited under identical conditions and produced a temperature rise of 1.94 K. For benzoic acid, the heat of combustion at constant volume is known to be 3,226 kJ mol^{-1} (that is, ΔU = –3,226 kJ mol^{-1}.) Use this information to determine the standard enthalpy of combustion of biphenyl.

Solution

Begin by working out the calorimeter constant:

- Moles of benzoic acid:

$$\frac{0.825g}{122.1\ g/mol} = 0.00676\ mol$$

- Heat released to calorimeter:

$$(0.00676\ mol) \times (3226\ kJ/mol) = 21.80\ kJ$$

- Calorimeter constant:

$$\frac{21.80\ kJ}{1.94\ K} = 11.24\ kJ/K$$

Now determine $\Delta U_{combustion}$ of the biphenyl ("BP"):

- moles of biphenyl:

$$\frac{0.526\ g}{154.12\ g/mol} = 0.00341\ mol$$

- heat released to calorimeter:

$$(1.91\ K) \times (11.24\ kJ/K) = 21.46\ kJ$$

- heat released per mole of biphenyl:

$$\frac{21.46\ kJ}{0.00341\ mol} = 6,293\ kJ/mol$$

$$\Delta U_{combustion}(BP) = -6,293\ kJ/mol$$

This is the heat change at constant volume, q_v; the negative sign indicates that the reaction is exothermic, as all combustion reactions are.

From the balanced reaction equation

$$(C_6H_5)_2\,(s) + \tfrac{29}{2}\,O_2\,(g) \rightarrow 12\,CO_2\,(g) + 5\,H_2O\,(l)$$

we can calculate the change in the moles of gasses for this reaction

$$\Delta n_g = 12 - \frac{29}{2} = \frac{-5}{2}$$

Thus the volume of the system *decreases* when the reaction takes place. Converting to ΔH, we can write the following equation. Additionally, recall that at constant volume, $\Delta U = q_V$.

$$\begin{aligned}\Delta H &= q_V + \Delta n_g RT \\ &= \Delta U - \left(\frac{5}{2}\right)(8.314\ J\ mol^{-1}\ K^{-1})(298\ K) \\ &= (-6,293\ kJ/mol) - (6,194\ J/mol) \\ &= (-6,293 - 6.2)\ kJ/mol \\ &= -6299\ kJ/mol\end{aligned}$$

A common mistake here is to forget that the subtracted term is in J, not kJ. Note that the additional 6.2 kJ in ΔH compared to ΔU reflects the work that the surroundings do on the system as the volume of gases decreases according to the reaction equation.

Determining the Heat of Reaction

The amount of heat that the system gives up to its surroundings so that it can return to its initial temperature is the heat of reaction. The heat of reaction is just the negative of the thermal energy gained by the calorimeter and its contents ($q_{calorimeter}$) through the combustion reaction.

$$q_{rxn} = -q_{calorimeter} \tag{14.5.2}$$

where

$$q_{calorimeter} = q_{bomb} + q_{water} \tag{14.5.3}$$

If the constant volume calorimeter is set up the same way as before, (same steel bomb, same amount of water, etc.) then the heat capacity of the calorimeter can be measured using the following formula:

$$q_{calorimeter} = (\text{ heat capacity of calorimeter}) \times \Delta T \tag{14.5.4}$$

Heat capacity is defined as the amount of heat needed to increase the temperature of the entire calorimeter by 1 °C. The equation above can also be used to calculate q_{rxn} from $q_{calorimeter}$ calculated by Equation 14.5.2. The heat capacity of the calorimeter can be determined by conducting an experiment.

Example 14.5.4: Heat of Combustion

1.150 g of sucrose goes through combustion in a bomb calorimeter. If the temperature rose from 23.42 °C to 27.64 °C and the heat capacity of the calorimeter is 4.90 kJ/°C, then determine the heat of combustion of sucrose, $C_{12}H_{22}O_{11}$ (in kJ per mole of $C_{12}H_{22}O_{11}$).

Solution

Given:

- mass of $C_{12}H_{22}O_{11}$: 1.150 g
- $T_{initial}$: 23.42°C
- T_{final}:27.64°C
- Heat Capacity of Calorimeter: 4.90 kJ/°C

Using Equation 14.5.4 to calculate $q_{calorimeter}$:

$$\begin{aligned} q_{calorimeter} &= (4.90\ kJ/°C) \times (27.64 - 23.42)°C \\ &= (4.90 \times 4.22)\ kJ = 20.7\ kJ \end{aligned}$$

Plug into Equation 14.5.2:

$$\begin{aligned} q_{rxn} &= -q_{calorimeter} \\ &= -20.7\ kJ \end{aligned}$$

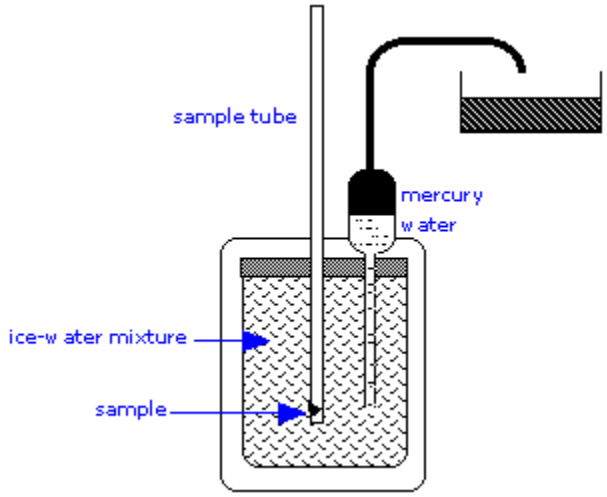
But the question asks for kJ/mol $\mathrm{C_{12}H_{22}O_{11}}$, so this needs to be converted:

$$\begin{aligned} q_{rxn} &= \frac{-20.7\ kJ}{1.150\ g\ C_{12}H_{22}O_{11}} \\ &= \frac{-18.0\ kJ}{g\ C_{12}H_{22}O_{11}} \end{aligned}$$

Convert to per Mole $\mathrm{C_{12}H_{22}O_{11}}$:

$$\begin{aligned} q_{rxn} &= \frac{-18.0\ kJ}{\cancel{g\ \mathrm{C_{12}H_{22}O_{11}}}} \times \frac{342.3\ \cancel{g\ \mathrm{C_{12}H_{22}O_{11}}}}{1\ mol\ \mathrm{C_{12}H_{22}O_{11}}} \\ &= \frac{-6.16 \times 10^3\ kJ}{mol\ \mathrm{C_{12}H_{22}O_{11}}} \end{aligned}$$

Although calorimetry is simple in principle, its practice is a highly exacting art, especially when applied to processes that take place slowly or involve very small heat changes, such as the germination of seeds. Calorimeters can be as simple as a foam plastic coffee cup, which is often used in student laboratories. Research-grade calorimeters, able to detect minute temperature changes, are more likely to occupy table tops, or even entire rooms:



The **ice calorimeter** is an important tool for measuring the heat capacities of liquids and solids, as well as the heats of certain reactions. This simple yet ingenious apparatus is essentially a device for measuring the change in volume due to melting of ice. To measure a heat capacity, a warm sample is placed in the inner compartment, which is surrounded by a mixture of ice and water.

The heat withdrawn from the sample as it cools causes some of the ice to melt. Since ice is less dense than water, the volume of water in the insulated chamber decreases. This causes an equivalent volume of mercury to be sucked into the inner reservoir from the outside container. The loss in weight of this container gives the decrease in volume of the water, and thus the mass of

ice melted. This, combined with the heat of fusion of ice, gives the quantity of heat lost by the sample as it cools to 0°C.

This page titled 14.5: Calorimetry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14.6: Applications of Thermochemistry

Virtually all chemical processes involve the absorption or release of heat, and thus changes in the internal energy of the system. In this section, we survey some of the more common chemistry-related applications of enthalpy and the First Law. While the first two sections relate mainly to chemistry, the remaining ones impact the everyday lives of everyone.

Enthalpy diagrams and their uses

Comparison and interpretation of enthalpy changes is materially aided by a graphical construction in which the relative enthalpies of various substances are represented by horizontal lines on a vertical energy scale. The zero of the scale can be placed anywhere, since energies are always arbitrary; it is generally most useful to locate the elements at zero energy, which reflects the convention that their standard enthlapies of formation are zero.

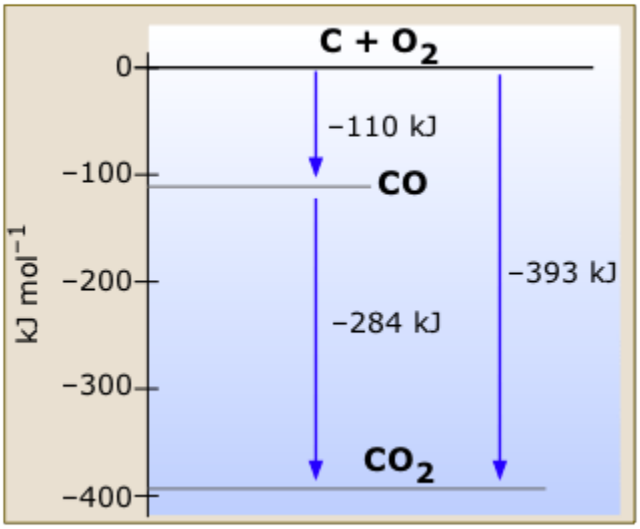


Figure 14.6.1: Enthalpy Diagram of CO

This very simple *enthalpy diagram* for carbon and oxygen and its two stable oxides (Figure 14.6.1) shows the changes in enthalpy associated with the various reactions this system can undergo. Notice how Hess's law is implicit in this diagram; we can calculate the enthalpy change for the combustion of carbon monoxide to carbon dioxide, for example, by subtraction of the appropriate arrow lengths without writing out the thermochemical equations in a formal way.

The zero-enthalpy reference states refer to graphite, the most stable form of carbon, and gaseous oxygen. All temperatures are 298 K.

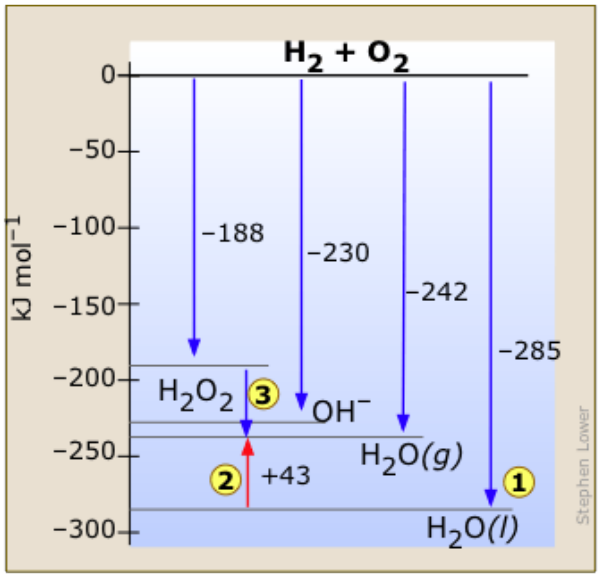


Figure 14.6.2: Enthalpy Diagram of H2O

This enthalpy diagram for the hydrogen-oxygen system (Figure 14.6.2) shows the known stable configurations of these two elements. Reaction of gaseous H_2 and O_2 to yield one mole of liquid water releases 285 kJ of heat (1). If the H_2O is formed in the gaseous state, the energy release will be smaller. Notice also that...

- The heat of vaporization of water (2) (an endothermic process) is clearly found from the diagram.
- Hydrogen peroxide H_2O_2, which spontaneously decomposes into O_2 and H_2O, releases some heat (3) in this process. Ordinarily this reaction is so slow that the heat is not noticed. But the use of an appropriate catalyst can make the reaction so fast that it has been used to fuel a racing car.

You may have heard the venerable urban legend, probably by now over 80 years old, that some obscure inventor discovered a process to do this, but the invention was secretly bought up by the oil companies in order to preserve their monopoly. The enthalpy diagram for the hydrogen-oxygen system shows why this cannot be true — there is simply no known compound of H and O that resides at a lower enthalpy level.

Enthalpy diagrams are especially useful for comparing groups of substances having some common feature. This one shows the molar enthalpies of species relating to two hydrogen halides, with respect to those of the elements. From this diagram we can see at a glance that the formation of HF from the elements is considerably more exothermic than the corresponding formation of HCl. The upper part of this diagram shows the gaseous atoms at positive enthalpies with respect to the elements. The endothermic processes in which the H_2 and the dihalogen are dissociated into atoms can be imagined as taking place in two stages, also shown. From the enthalpy change associated with the dissociation of H_2 (218 kJ mol^{-1}), the dissociation enthalpies of F_2 and Cl_2 can be calculated and placed on the diagram.

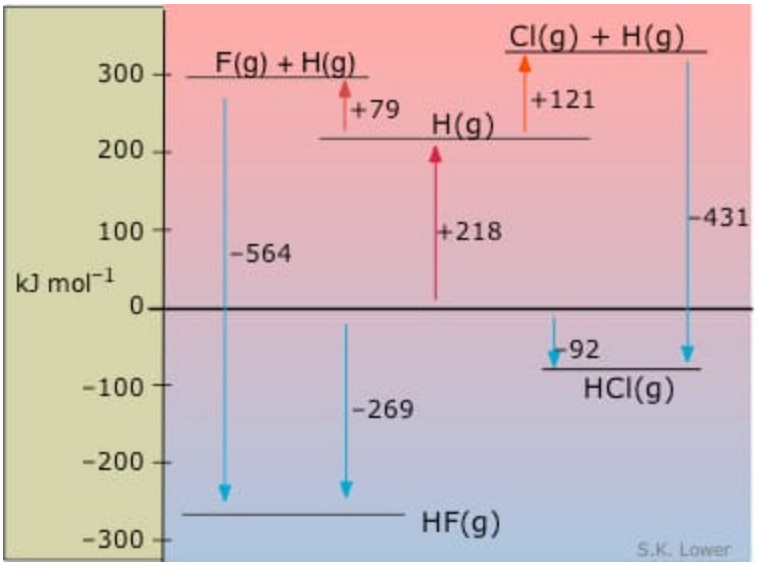


Figure 14.6.3: Enthalpy Diagram

Bond Enthalpies vs. Bond Energies

The enthalpy change associated with the reaction

$$\mathrm{HI\,(g) \longrightarrow H\,(g) + I\,(g)} \tag{14.6.1}$$

is the *enthalpy of dissociation* of the HI molecule; it is also the *bond energy* of the hydrogen-iodine bond in this molecule. Under the usual standard conditions, it would be expressed either as the bond enthalpy *H*°(HI,298K) or internal energy *U*°(HI,298); in this case the two quantities differ from each other by Δ*PV* = *RT*. Since this reaction cannot be studied directly, the H–I *bond enthalpy* is calculated from the appropriate standard enthalpies of formation:

½ H_2*(g)* → H*(g)*	+ 218 kJ
½ I_2*(g)* → I*(g)*	+107 kJ
½ H_2*(g)* + 1/2 I_2*(g)* → HI*(g)*	–36 kJ
HI*(g)* → H*(g)* + I*(g)*	**+299 kJ**

Bond energies and enthalpies are important properties of chemical bonds, and it is very important to be able to estimate their values from other thermochemical data. The total bond enthalpy of a more complex molecule such as ethane can be found from the following combination of reactions:

C_2H_6*(g)* → 2C*(graphite)* + 3 H_2*(g)*	84.7 kJ
3 H_2*(g)* → 6 H*(g)*	1308 kJ
2 C*(graphite)* → 2 C*(g)*	1430 kJ
C_2H_6*(g)* → 2 C*(g)* + 6H*(g)*	**2823 kJ**

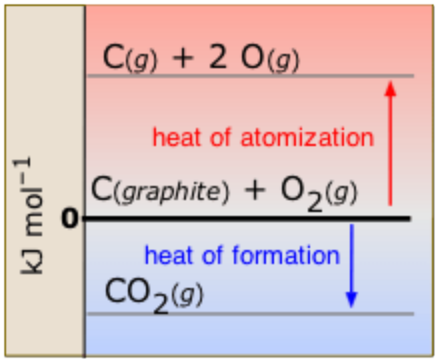


Figure 14.6.4: Comparing Standard Heats of Atomization (red) and Formation (blue) for carbon

When a molecule in its ordinary state is broken up into gaseous atoms, the process is known as *atomization* (Figure 14.6.4). The standard *enthalpy of atomization* refers to the transformation of an element into gaseous atoms:

$$\mathrm{C_{(graphite)} \longrightarrow C\,(g)} \quad \Delta H^o = 716.7\ kJ \tag{14.6.2}$$

Atomization is *always* an endothermic process. Heats of atomization are most commonly used for calculating bond energies. They are usually measured spectroscopically.

Pauling's rule and Average Bond Energy

The total bond energy of a molecule can be thought of as the sum of the energies of the individual bonds.

Pauling's Rule is only an approximation, because the energy of a given type of bond is not really a constant, but depends somewhat on the particular chemical environment of the two atoms. In other words, all we can really talk about is the average energy of a particular kind of bond, such as C–O, for example, the average being taken over a representative sample of compounds containing this type of bond, such as CO, CO_2, $COCl_2$, $(CH_3)_2CO$, CH_3COOH, etc.

Despite the lack of strict additivity of bond energies, Pauling's Rule is extremely useful because it allows one to estimate the heats of formation of compounds that have not been studied, or have not even been prepared. Thus in the foregoing example, if we know the enthalpies of the C–C and C–H bonds from other data, we could estimate the total bond enthalpy of ethane, and then work back to get some other quantity of interest, such as ethane's enthalpy of formation. By assembling a large amount of experimental information of this kind, a consistent set of average bond energies can be obtained. The energies of double bonds are greater than those of single bonds, and those of triple bonds are higher still (Table 14.6.1).

Table 14.6.1: Average energies of some single bonds (KJ/mol)

H	C	N	O	F	Cl	Br	I	Si	
H	436	415	390	464	569	432	370	295	395
C	345	290	350	439	330	275	240	360	
N	160	200	270	200	270				
O	140	185	205	185	200	370			
F	160	255	160	280	540				
Cl	243	220	210	359					
Br	190	180	290						
I	150	210							
Si	230								

Energy Content of Fuels

A *fuel* is any substance capable of providing useful amounts of energy through a process that can be carried out in a controlled manner at economical cost. For most practical fuels, the process is combustion in air (in which the oxidizing agent O_2 is available at zero cost.) The *enthalpy of combustion* is obviously an important criterion for a substance's suitability as a fuel, but it is not the only one; a useful fuel must also be easily ignited, and in the case of a fuel intended for self-powered vehicles, its *energy density* in terms of both mass (kJ kg^{-1}) and volume (kJ m^{-3}) must be reasonably large. Thus substances such as methane and propane which are gases at 1 atm must be stored as pressurized liquids for transportation and portable applications.

Table 14.6.2: Energy densities of some common fuels

fuel	MJ kg^{-1}
wood (dry)	15
coal (poor)	15
coal (premium)	27
ethanol[a]	30
petroleum-derived products	45
methane, liquified natural gas	54
hydrogen[b]	140

Notes on the above table

[a] **Ethanol** is being strongly promoted as a motor fuel by the U.S. agricultural industry. Note, however, that according to some estimates, it takes 46 MJ of energy to produce 1 kg of ethanol from corn. Some other analyses, which take into account optimal farming practices and the use of by-products arrive at different conclusions; see, for example, this summary with links to several reports.

[b] Owing to its low molar mass and high heat of combustion, **hydrogen** possesses an extraordinarily high energy density, and would be an ideal fuel if its critical temperature (33 K, the temperature above which it cannot exist as a liquid) were not so low. The potential benefits of using hydrogen as a fuel have motivated a great deal of research into other methods of getting a large amount of H_2 into a small volume of space. Simply compressing the gas to a very high pressure is not practical because the weight of the heavy-walled steel vessel required to withstand the pressure would increase the effective weight of the fuel to an unacceptably large value. One scheme that has shown some promise exploits the ability of H_2 to "dissolve" in certain transition metals. The hydrogen can be recovered from the resulting solid solution (actually a loosely-bound compound) by heating.

Energy content of foods

What, exactly, is meant by the statement that a particular food "contains 1200 calories" per serving? This simply refers to the standard enthalpy of combustion of the foodstuff, as measured in a *bomb calorimeter*. Note, however, that in nutritional usage, the calorie is really a kilocalorie (sometimes called "large calorie"), that is, 4184 J. Although this unit is still employed in the popular literature, the SI unit is now commonly used in the scientific and clinical literature, in which energy contents of foods are usually quoted in kJ per unit of weight.

Although the mechanisms of oxidation of a carbohydrate such as glucose to carbon dioxide and water in a bomb calorimeter and in the body are complex and entirely different, the net reaction involves the same initial and final states, and must be the same for any possible pathway:

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O \quad \Delta H^o = -20.8\ kJ\ mol^{-1} \tag{14.6.3}$$

Glucose is a *sugar*, a breakdown product of *starch*, and is the most important energy source at the cellular level; fats, proteins, and other sugars are readily converted to glucose. By writing balanced equations for the combustion of sugars, fats, and proteins, a comparison of their relative energy contents can be made. The stoichiometry of each reaction gives the amounts of oxygen taken

up and released when a given amount of each kind of food is oxidized; these gas volumes are often taken as indirect measures of energy consumption and metabolic activity; a commonly-accepted value that seems to apply to a variety of food sources is 20.1 J (4.8 kcal) per liter of O_2 consumed.

For some components of food, particularly proteins, oxidation may not always be complete in the body, so the energy that is actually available will be smaller than that given by the heat of combustion. Mammals, for example, are unable to break down cellulose (a polymer of sugar) at all; animals that derive a major part of their nutrition from grass and leaves must rely on the action of symbiotic bacteria which colonize their digestive tracts. The amount of energy available from a food can be found by measuring the heat of combustion of the waste products excreted by an organism that has been restricted to a controlled diet, and subtracting this from the heat of combustion of the food (Table 14.6.3).

Table 14.6.3: Energy content and availability of the major food components

type of food	food	$\Delta H°$ (kJ g^{-1})	percent availability
Protein	meat	22.4	92
	egg	23.4	
Fat	butter	38.2	
	animal fat	39.2	95
Carbohydrate	starch	17.2	
	glucose (sugar)	15.5	99
	ethanol	29.7	100

The amount of energy an animal requires depends on the age, sex, surface area of the body, and of course on the amount of physical activity. The rate at which energy is expended is expressed in watts: 1 W = 1 J sec^{-1}. For humans, this value varies from about 200-800 W. This translates into daily food intakes having energy equivalents of about 10-15 MJ for most working adults. In order to just maintain weight in the absence of any physical activity, about 6 MJ per day is required.

Table 14.6.4: Metabolic rates of some animals

animal	kJ hr^{-1}	kJ $kg^{-1}hr^{-1}$
mouse	82	17
cat	34	6.8
dog	78	3.3
sheep	193	2.2
human	300	2.1
horse	1430	1.1
elephant	5380	0.7

The above table is instructive in that although larger animals consume more energy, the energy consumption *per unit of body weight* decreases with size. This reflects the fact the rate of heat loss to the environment depends largely on the surface area of an animal, which increases with mass at a greater rate than does an animal's volume ("size").

Thermodynamics and the weather

Hydrogen bonds at work

It is common knowledge that large bodies of water have a “moderating” effect on the local weather, reducing the extremes of

temperature that occur in other areas. Water temperatures change much more slowly than do those of soil, rock, and vegetation, and this effect tends to affect nearby land masses. This is largely due to the high heat capacity of water in relation to that of land surfaces— and thus ultimately to the effects of *hydrogen bonding*. The lower efficiency of water as an absorber and radiator of infrared energy also plays a role.

The specific heat capacity of water is about four times greater than that of soil. This has a direct consequence to anyone who lives near the ocean and is familiar with the daily variations in the direction of the winds between the land and the water. Even large lakes can exert a moderating influence on the local weather due to water's relative insensitivity to temperature change.

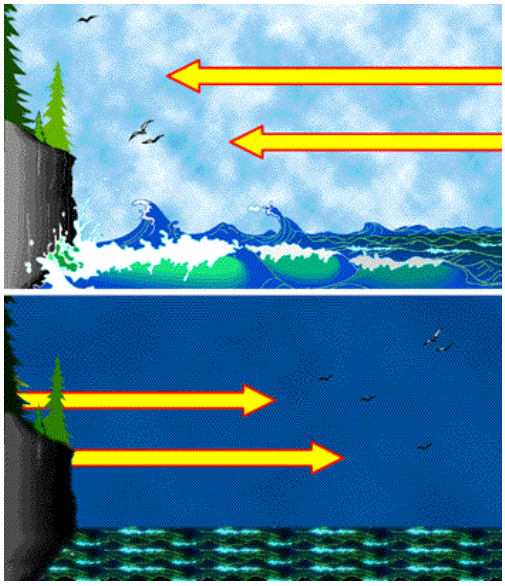


Figure 14.6.5: Coastal Winds. The above images are from the NDBC site.

During the daytime the land and sea receive approximately equal amounts of heat from the Sun, but the much smaller heat capacity of the land causes its temperature to rise more rapidly. This causes the air above the land to heat, reducing its density and causing it to rise. Cooler oceanic air is drawn in to fill the void, thus giving rise to the daytime **sea breeze**. In the evening, both land and ocean lose heat by radiation to the sky, but the temperature of the water drops less than that of the land, continuing to supply heat to the oceanic air and causing it to rise, thus reversing the direction of air flow and producing the evening **land breeze**.

Why it gets colder as you go higher: the adiabatic lapse rate

The air receives its heat by absorbing far-infrared radiation from the earth, which of course receives its heat from the sun. The amount of heat radiated to the air immediately above the surface varies with what's on it (forest, fields, water, buildings) and of course on the time and season. When a parcel of air above a particular location happens to be warmed more than the air immediately surrounding it, this air expands and becomes less dense. It therefore rises up through the surrounding air and undergoes further expansion as it encounters lower pressures at greater altitudes.

Whenever a gas expands against an opposing pressure, it does work on the surroundings. According to the First Law $\Delta U = q + w$, if this work is not accompanied by a compensating flow of heat into the system, its internal energy will fall, and so, therefore, will its temperature. It turns out that heat flow and mixing are rather slow processes in the atmosphere in comparison to the convective motion we are describing, so the First Law can be written as $\Delta U = w$ (recall that w is negative when a gas expands.) Thus as air rises above the surface of the earth it undergoes adiabatic expansion and cools. The actual rate of temperature decrease with altitude depends on the composition of the air (the main variable being its moisture content) and on its heat capacity. For dry air, this results in an *adiabatic lapse rate* of 9.8 C° per km of altitude.

Santa Anas and Chinooks

Just the opposite happens when winds develop in high-altitude areas and head downhill. As the air descends, it undergoes compression from the pressure of the air above it. The surroundings are now doing work on the system, and because the process

occurs too rapidly for the increased internal energy to be removed as heat, the compression is approximately adiabatic.

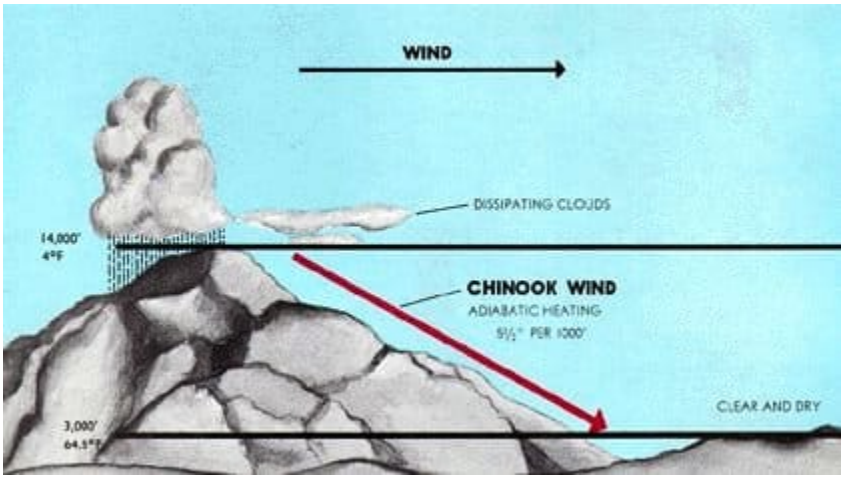


Figure 14.6.6: Adiabatic warming of downward moving air produces the warm Chinook wind. Image from U.S. FAA AC 00-61, Chapter 6, Figure 41. Adiabatic warming of downward moving air produces the warm Chinook wind.

The resulting winds are warm (and therefore dry) and are often very irritating to mucous membranes. These are known generically as **Föhn winds** (which is the name given to those that originate in the Alps). In North America they are often called **chinooks** (or, in winter, "snow melters") when they originate along the Rocky Mountains. Among the most notorious are the **Santa Ana** winds of Southern California which pick up extra heat (and dust) as they pass over the Mohave Desert before plunging down into the Los Angeles basin (Figure 13.5.X). Their dryness and high velocities feed many of the disastrous wildfires that afflict the region.

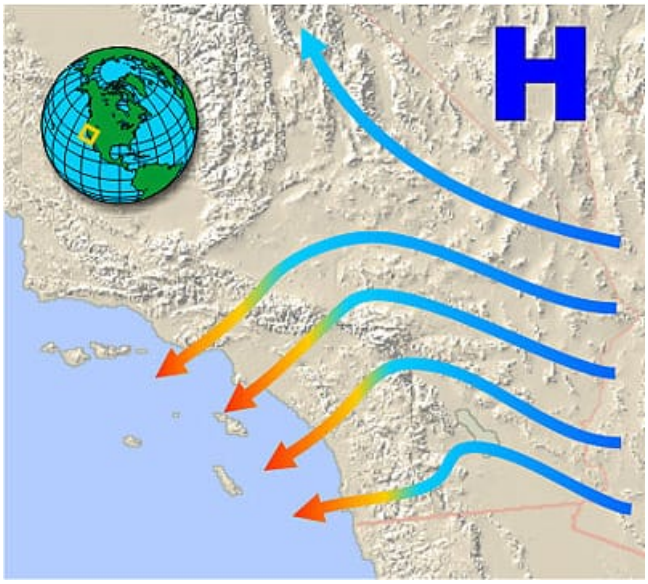


Figure 14.6.7: Map illustration showing propagation of a Santa Ana wind event in southern California, USA (U.S. National Oceanic and Atmospheric Administration)

This page titled 14.6: Applications of Thermochemistry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

14.E: Thermochemistry (Exercises)

13.1: Energy, heat and work

13.2: The First Law of Thermodynamics

13.3: Molecules as energy carriers and converters

13.4: Thermochemistry

13.5: Calorimetry

Q13.5.1

After going through combustion in a bomb calorimeter a sample gives off 5,435 cal. The calorimeter experiences an increase of 4.27°C in its temperature. Using this information, determine the heat capacity of the calorimeter in kJ/°C.

Q13.5.2

Referring to the example given above about the heat of combustion, calculate the temperature change that would occur in the combustion of 1.732 g $C_{12}H_{22}O_{12}$ in a bomb calorimeter that had the heat capacity of 3.87 kJ/°C.

Q13.5.3

Given the following data calculate the heat of combustion in kJ/mol of xylose,$C_5H_{10}O_5$(s), used in a bomb calorimetry experiment: mass of $C_5H_{10}O_5$(s) = 1.250 g, heat capacity of calorimeter = 4.728 kJ/°C, Initial Temperature of the calorimeter = 24.37°C, Final Temperature of calorimeter = 28.29°C.

Q13.5.4

Determine the heat capacity of the bomb calorimeter if 1.714 g of naphthalene, $C_{10}H_8$(s), experiences an 8.44°C increase in temperature after going through combustion. The heat of combustion of naphthalene is -5156 kJ/mol $C_{10}H_8$.

Q13.5.5

What is the heat capacity of the bomb calorimeter if a 1.232 g sample of benzoic acid causes the temperature to increase by 5.14°C? The heat of combustion of benzoic acid is -26.42 kJ/g.

S13.5.1

Use equation **4** to calculate the heat of capacity:

$$q_{calorimeter} = \ heat\ capicity\ of\ calorimeter\ x\ \Delta T$$

5435 cal = heat capacity of calorimeter x 4.27°C

Heat capacity of calorimeter = (5435 cal/ 4.27°C) x (4.184 J/1 cal) x (1kJ/1000J) = 5.32 kJ/°C

S13.5.2

The temperature should increase since bomb calorimetry releases heat in an exothermic combustion reaction.

Change in Temp = (1.732 g $C_{12}H_{22}O_{11}$) x (1 mol $C_{12}H_{22}O_{11}$/342.3 g $C_{12}H_{22}O_{11}$) x (6.61 x 10^3 kJ/ 1 mol $C_{12}H_{22}O_{11}$) x (1°C/3.87kJ) = 8.64°C

S13.5.3

[(Heat Capacity x Change in Temperature)/mass] =[((4.728 kJ/°C) x(28.29 °C – 24.37 °C))/1.250 g] = 14.8 kJ/g xylose

$$q_{rxn}$$ = (-14.8 kJ/g xylose) x (150.13 g xylose/ 1 mol xylose) = -2.22x10^3 kJ/mol xylose

S13.5.4

Heat Capacity = [(1.714 g $C_{10}H_8$) x (1 mol $C_{10}H_8$/128.2 g $C_{10}H_8$) x (5.156x10^3 kJ/1 mol $C_{10}H_8$)]/8.44°C = 8.17 kJ/ °C

S13.5.5

Heat Capacity = [(1.232 g benzoic acid) x (26.42 kJ/1 g benzoic acid)]/5.14°C = 6.31 kJ/ °C

13.6: Applications of Thermochemistry

This page titled 14.E: Thermochemistry (Exercises) is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15: THERMODYNAMICS OF CHEMICAL EQUILIBRIA

Trying to introduce chemical thermodynamics to beginning students is always problematic; to do it "properly" requires a degree of rigor that rarely succeeds for more than a small fraction of the class. Although a full formal development is rarely appropriate at this level, I believe that the value of developing students' understanding of the fundamental concepts is generally underappreciated. This requires some understanding of the ways thermal energy is dispersed in matter— something that is not a part of classical thermodynamics and is not supported by most textbooks, but which is in keeping with the molecular focus of modern chemical science.

The equilibrium value for a reversible reaction is an important quantity that characterizes a chemical reaction, but what factors govern its value? In particular, is there any way that we can predict the value of the equilibrium constant of a reaction solely from information about the products and reactants themselves, without any knowledge at all about the mechanism or other details of the reaction? The answer is yes, and this turns out to be the central purpose of chemical thermodynamics:

The purpose of thermodynamics is to predict the equilibrium composition of a system from the properties of its components.

Don't let the significance of this pass you by; it means that we can say with complete certainty whether or not a given change is possible, and if it is possible, to what extent it will occur— without the need to study the particular reaction in question. To a large extent, this is what makes chemistry a science, rather than a mere cataloging of facts.

15.1: Energy Spreading Drives Spontaneous Change

15.2: Entropy Rules

15.3: The Second Law of Thermodynamics

15.4: Free Energy and the Gibbs Function

15.5: Thermodynamics of Mixing and Dilution

15.6: Free Energy and Equilibrium

15.7: Some Applications of Entropy and Free Energy

15.8: Quantum states, Microstates, and Energy spreading in Reactions

This page titled 15: Thermodynamics of Chemical Equilibria is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.1: Energy Spreading Drives Spontaneous Change

Learning Objectives

You are expected to be able to define and explain the significance of terms identified in ***bold***.

- S*pontaneous change* is that which, once initiated, proceeds on its own until some *state of* ***equilibrium*** (mechanical, thermal, chemical, etc.) is attained.
- All ***natural processes*** (those that occur in the world) ultimately involve spontaneous change.
- ***Thermal energy*** is ***kinetic energy*** associated with the random motions of atoms and molecules. These states of motion are quantized into energy states which can be realized in huge numbers of different ways that we call ***microstates***. will be ***thermally accessible*** and likely to be populated.
- Spontaneous change is driven by the tendency of thermal energy to spread into as many microstates of the system and surroundings as are thermally accessible.

Drop a teabag into a pot of hot water, and you will see the tea diffuse into the water until it is uniformly distributed throughout the water. What you will *never* see is the reverse of this process, in which the tea would be sucked up and re-absorbed by the teabag. The making of tea, like all changes that take place in the world, possesses a “natural” direction.

What is a spontaneous process?

Processes that proceed in a definite direction when left to themselves and in the absence of any attempt to drive them in reverse — are known as *natural process*or *spontaneous* changes. In many cases our everyday life experiences teach us the direction in which spontaneous change can occur, and anything that runs counter to these expectations is immediately sensed as weird. In other cases, including that of most chemical change, we often have no obvious guidelines, and must learn how to apply the laws of thermodynamics which ultimately govern all spontaneous change.

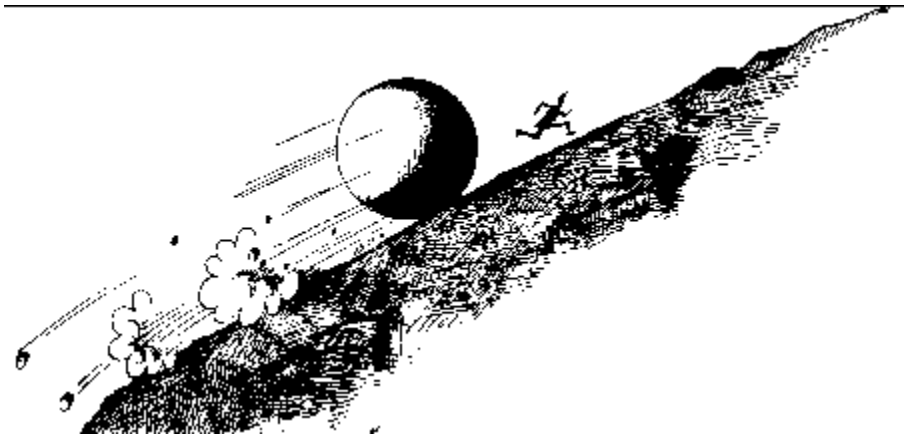


Figure 15.1.1: A non-natural process: Steinberg's famous New Yorker cartoon. Saul Steinberg (1914-1999) was one of America's most admired artists whose unique and imaginative cartoon style graced the pages of the New Yorker magazine for more than 60 years.

In order to answer the above question, let's begin by thinking about the outcomes of the following four experiments, each of which illustrates a **natural process that proceeds spontaneously only in a single direction.**

Experiment 15.1.1

A stack of one hundred coins is thrown into the air. After they have come to rest on the floor, the numbers that land “heads up” and “tails up” are noted.

Net change: ordered coins → randomized coins (roughly equal numbers of heads and tails.)

Energetics: no relevant net change in energy

Why it does not go in reverse: Simple statistics shows that the probability that the coins will land in one particular arrangement of the huge number that are possible is vanishingly small.

Example 15.1.2

Two identical blocks of copper, one at 200°C and the other at 100°C, are brought into contact in a thermally-insulated environment. Eventually the temperatures of both blocks reach 150°C.

Net change: block1 (200°) + block2 (100°) → combined blocks (150°)

Energetics: Heat (randomized molecular kinetic energy) flows from the warmer block to the cooler one until their temperatures are identical.

Why it does not go in reverse: Dispersal of kinetic energy amongst the copper atoms is a random process; the chances that such a process would lead to a non-uniform sharing of the energy are even smaller than in the case of the 100 coins because of the much greater number (around 10^{22}) of particles involved.

Experiment 15.1.3

A book or some other solid object is held above a table top, and is then allowed to fall.

Net change: book in air  book on table top; potential energy → organized kinetic energy  thermal energy.

Energetics: At the instant just before the end of its fall, the potential energy the object acquired when it was raised will exist entirely as kinetic energy $mv^2/2$ in which m is the mass of the object and v is its velocity. Each atom of which the object is composed will of course possess a proportionate fraction of this energy, again with its principal velocity component pointing down. Superimposed on this, however, will be minute thermal displacements that vary randomly in magnitude and direction from one instant to the next. The sum total of these constitutes the thermal energy contained in the object.

When the object strikes the table top, its motion ceases and we say its kinetic energy is zero. Energy is supposed to be conserved, so where did it disappear to? The shock of impact has resulted in its dispersal into greatly augmented thermal motions of the atoms, both of the object itself and of the area of the table top where the impact occurred. In other words, the kinetic energy of organized motion the object had just before its motion stopped has been transformed into kinetic energy of random or disorganized motion (thermal energy) which spreads rapidly away from the point of impact.

Why it does not go in reverse: Once the kinetic energy of the book has been dispersed amongst the molecules of the book and the table top, the probability of these randomized motions reappearing at the surface where the two objects are in contact and then acting in concert to propel the object back into the air) is negligible.

Experiment 15.1.4

One mole of gas, initially at 300 K and 2 atm pressure, is allowed to expand to double its volume, keeping the temperature constant.

Net change: Increase in volume of gas.

Energetics: No change in energy if the gas behaves ideally.

Why it does not go in reverse: Simple statistics: the probability that N randomly moving objects (flies in a bottle, for example,) will at any time all be located in one half of the container is $(1/2)^N$. For chemically-significant values of N (10^{20}, say) this probability is indistinguishable from zero.

All of the changes described above take place spontaneously, meaning that

- Once they are allowed to commence, they will proceed to the finish without any outside intervention.
- It would be inconceivable that any of these changes could occur in the reverse direction (that is, be undone) without changing the conditions or actively disturbing the system in some way.

What determines the direction in which spontaneous change will occur? **It is clearly not a fall in the energy**, since in most cases cited above the energy of the system did not change. Even in the case of the falling book, in which the potential energy of the system (the book) falls, energy is conserved overall; if there is no net loss of energy when these processes operate in the forward or natural direction, it would not require any expenditure of energy for them to operate in reverse. In other words, contrary to what all too many people appear to believe, the First Law of Thermodynamics cannot predict the direction of a natural process.

> The direction of a spontaneous process is not governed by the energy change and thus the **First Law of Thermodynamics** cannot predict the direction of a natural process

Direction through disorder

In our examination of the processes described above, we saw that although the total energy of the system and the surroundings (and thus, of the world) is unchanged, **there is something about the world that has changed, and this is its degree of disorder**.

After coins have been tossed or cards shuffled, the final state is invariably one of greater disorder. Similarly, the molecules of a gas can occupy a larger number of possible positions in space if the volume is larger, so the expansion of a gas is similarly accompanied by an increase in randomness

How can we express disorder quantitatively? From the example of the coins, you can probably see that simple statistics plays a role: the probability of obtaining three heads and seven tails after tossing ten coins is just the ratio of the number of ways that ten different coins can be arranged in this way, to the number of all possible arrangements of ten coins.



Using the language of molecular statistics, we say that a collection of coins in which a given fraction of its members are heads-up constitutes a *macroscopic state* of the system. Since we do not care which coins are heads-up, there are clearly numerous configurations of the individual coins which can result in this "macrostate". Each of these configurations specifies a *microscopic state* of the system. The greater the number of microstates that correspond to a given macrostate, the greater the probability of that macrostate. To see what this means, consider the possible outcomes of a toss of four coins (Table 15.1.1):

Table 15.1.1: Coin Toss Results

macrostate	ways	probability	microstates
0 heads	1	1/16	TTTT
1 head	4	4/16 = 1/4	HTTT THTT TTHT TTTH
2 heads	6	6/16 = 3/8	HHTT HTHT HTTH THHT TTHH THTH
3 heads	4	4/16 = 1/4	HHHT HTHH HHTH THHH
4 heads	1	1/16	HHHH

A toss of four coins will yield one of the five outcomes (macrostates) listed in the leftmost column of the table. The second column gives the number of "ways"— that is, the number of head/tail configurations of the set of coins (the number of microstates)—that can result in the macrostate. The probability of a toss resulting in a particular macrostate is proportional to the number of microstates corresponding to the macrostate, and is equal to this number, divided by the total number of possible microstates (in this example, 2^4 =16). An important assumption here is that all microstates are equally probable; that is, the toss is a "fair" one in which the many factors that determine the trajectory of each coin operate in an entirely random way.

> The greater the number of **microstates** that correspond to a given **macrostate**, the greater the probability of that **macrostate**.

Disorder is more probable than order because there are so many more ways of achieving it. Thus coins and cards tend to assume random configurations when tossed or shuffled, and socks and books tend to become more scattered about a teenager's room during the course of daily living. But there are some important differences between these large-scale mechanical, or macro systems, and the collections of sub-microscopic particles that constitute the stuff of chemistry, and which we will refer to here generically as **molecules**. Molecules, unlike macro objects, are capable of accepting, storing, and giving up energy in tiny amounts (quanta), and act as highly efficient carriers and spreaders of thermal energy as they move around. Thus, **in chemical systems**,

1. **We are dealing with huge numbers of particles**. This is important because statistical predictions are always more accurate for larger samples. Thus although for the four tosses there is a good chance (62%) that the H/T ratio will fall outside the range of 0.45 - 0.55, this probability becomes almost zero for 1000 tosses. To express this in a different way, the chances that 1000 gas molecules moving about randomly in a container would at any instant be distributed in a sufficiently non-uniform manner to produce a detectable pressure difference between the two halves of a container will be extremely small. If we increase the number of molecules to a chemically significant number (around 10^{20}, say), then the same probability becomes

indistinguishable from zero.

2. **Once the change begins, it proceeds spontaneously.** That is, no external agent (a tosser, shuffler, or teenager) is needed to keep the process going. Gases will spontaneously expand if they are allowed to, and reactions, one started, will proceed toward equilibrium.
3. **Thermal energy is continually being exchanged between the particles of the system, and between the system and the surroundings.** Collisions between molecules result in exchanges of momentum (and thus of kinetic energy) amongst the particles of the system, and (through collisions with the walls of a container, for example) with the surroundings.
4. **Thermal energy spreads rapidly and randomly throughout the various energetically accessible microstates of the system.** The **direction** of spontaneous change is that which results in the maximum possible spreading and sharing of thermal energy.

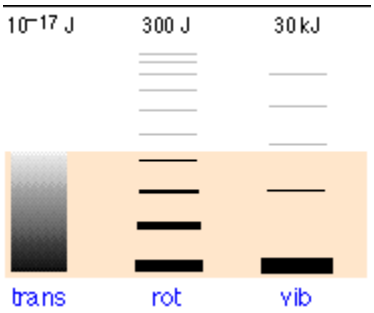
The importance of these last two points is far greater than you might at first think, but to fully appreciate this, you must recall the various ways in which thermal energy is stored in molecules— hence the following brief review.

The Spreading of Energy

Thermal energy is the portion of a molecule's energy that is proportional to its *temperature*, and thus relates to motion at the molecular scale. What kinds of molecular motions are possible? For monatomic molecules, there is only one: actual movement from one location to another, which we call *translation*. Since there are three directions in space, all molecules possess three modes of translational motion.

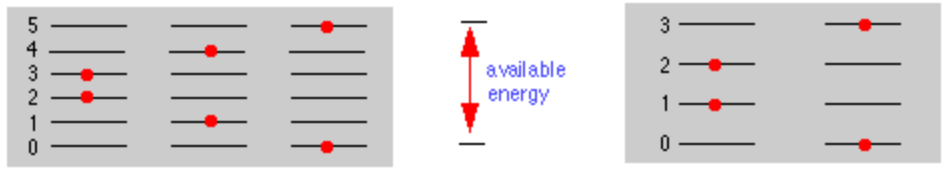
For polyatomic molecules, two additional kinds of motions are possible. One of these is *rotation*; a linear molecule such as CO_2 in which the atoms are all laid out along the *x*-axis can rotate along the *y*- and *z*-axes, while molecules having less symmetry can rotate about all three axes. Thus linear molecules possess two modes of rotational motion, while non-linear ones have three rotational modes.

Finally, molecules consisting of two or more atoms can undergo internal *vibrations*. For freely moving molecules in a gas, the number of vibrational modes or patterns depends on both the number of atoms and the shape of the molecule, and it increases rapidly as the molecule becomes more complicated.



The relative populations of the translational, rotational and vibrational energy states of a typical diatomic molecule are depicted by the thickness of the lines in this schematic (not-to-scale!) diagram. The colored shading indicates the total thermal energy available at a given temperature. The numbers at the top show order-of-magnitude spacings between adjacent levels. It is readily apparent that virtually all the thermal energy resides in translational states.

Notice the greatly different spacing of the three kinds of energy levels. This is extremely important because it determines the number of energy quanta that a molecule can accept, and, as the following illustration shows, the number of different ways this energy can be distributed amongst the molecules.



The quantity of energy indicated by the vertical arrow can be distributed in three ways in a collection of molecules whose allowed energy levels are spaced as shown here.

In molecules with more widely-spaced energy levels, there are fewer ways of distributing the same total energy.

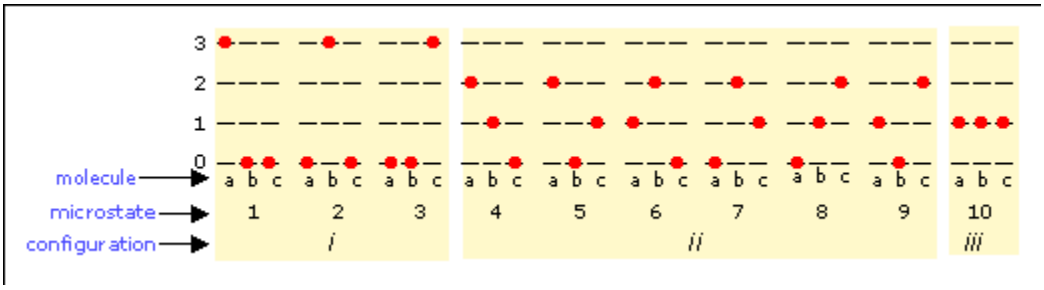
Accessible energy states. The more closely spaced the quantized energy states of a molecule, the greater will be the number of ways in which a given quantity of thermal energy can be shared amongst a collection of these molecules.

The spacing of molecular energy states becomes closer as the mass and number of bonds in the molecule increases, so we can generally say that the more complex the molecule, the greater the *density* of its energy states. This is a feature of the particle-in-a-box model, which predicts that the separation of the energy states of a gas confined within a box depends on the inverse square of the box length (and on the inverse of the particle mass as well.)

Quantum states, microstates, and energy spreading

At the atomic and molecular level, all energy is *quantized*; each particle possesses discrete states of kinetic energy and is able to accept thermal energy only in packets whose values correspond to the energies of one or more of these states. Polyatomic molecules can store energy in rotational and vibrational motions, and all molecules (even monatomic ones) will possess translational kinetic energy (thermal energy) at all temperatures above absolute zero. The energy difference between adjacent translational states is so minute (roughly 10^{-30} J) that translational kinetic energy can be regarded as *continuous* (non-quantized) for most practical purposes.

The number of ways in which thermal energy can be distributed amongst the allowed states within a collection of molecules is easily calculated from simple statistics, but we will confine ourselves to an example here. Suppose that we have a system consisting of three molecules and three quanta of energy to share among them. We can give all the kinetic energy to any one molecule, leaving the others with none, we can give two units to one molecule and one unit to another, or we can share out the energy equally and give one unit to each molecule. All told, there are ten possible ways of distributing three units of energy among three identical molecules as shown here:



Each of these ten possibilities represents a distinct *microstate* that will describe the system at any instant in time. Those microstates that possess identical distributions of energy among the accessible quantum levels (and differ only in which particular molecules occupy the levels) are known as **configurations**. Because all microstates are equally probable, the probability of any one configuration is proportional to the number of microstates that can produce it. Thus in the system shown above, the configuration labeled *ii* will be observed 60% of the time, while *iii* will occur only 10% of the time.

As the number of molecules and the number of quanta increases, the number of accessible microstates grows explosively; if 1000 quanta of energy are shared by 1000 molecules, the number of available microstates will be around 10^{600}— a number that greatly exceeds the number of atoms in the observable universe! The number of possible configurations (as defined above) also increases, but in such a way as to greatly reduce the probability of all but the most probable configurations. Thus for a sample of a gas large enough to be observable under normal conditions, only a single configuration (energy distribution amongst the

quantum states) need be considered; even the second-most-probable configuration can be neglected.

The bottom line: any collection of molecules large enough in numbers to have chemical significance will have its therrmal energy distributed over an unimaginably large number of microstates. The number of microstates increases exponentially as more energy states ("configurations" as defined above) become accessible owing to

- Addition of energy quanta (higher temperature),
- Increase in the number of molecules (resulting from dissociation, for example).
- the volume of the system increases (which decreases the spacing between energy states, allowing more of them to be populated at a given temperature.)

Increasing temperature increases the number of microstates in a state and hence the entropy of the state.

Energy-spreading changes the world

Energy is conserved; if you lift a book off the table, and let it fall, the total amount of energy in the world remains unchanged. All you have done is transferred it from the form in which it was stored within the glucose in your body to your muscles, and then to the book (that is, you did work on the book by moving it up against the earth's gravitational field). After the book has fallen, this same quantity of energy exists as thermal energy (heat) in the book and table top.

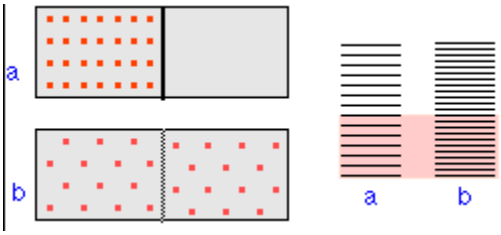
What has changed, however, is the availability of this energy. Once the energy has spread into the huge number of thermal microstates in the warmed objects, the probability of its spontaneously (that is, by chance) becoming un-dispersed is essentially zero. Thus although the energy is still "there", **it is forever beyond utilization or recovery.**

The profundity of this conclusion was recognized around 1900, when it was first described as the "heat death" of the world. This refers to the fact that every spontaneous process (essentially every change that occurs) is accompanied by the "dilution" of energy. The obvious implication is that all of the molecular-level kinetic energy will be spread out completely, and nothing more will ever change. Not a happy thought!

Gases Spontaneously Expand and never contract

Everybody knows that a gas, if left to itself, will tend to expand so as to fill the volume within which it is confined completely and uniformly. What "drives" this expansion? At the simplest level it is clear that with more space available, random motions of the individual molecules will inevitably disperse them throughout the space. But as we mentioned above, the allowed energy states that molecules can occupy are spaced more closely in a larger volume than in a smaller one. The larger the volume available to the gas, the greater the number of energy states (and thus microstates) its thermal energy can occupy. Since all such microstates within the thermally accessible range of energies are equally probable, the expansion of the gas can viewed as a consequence of the tendency of energy to be spread and shared as widely as possible. Once this has happened, the probability that this sharing of energy will reverse itself (that is, that the gas will spontaneously contract) is so minute as to be unthinkable.

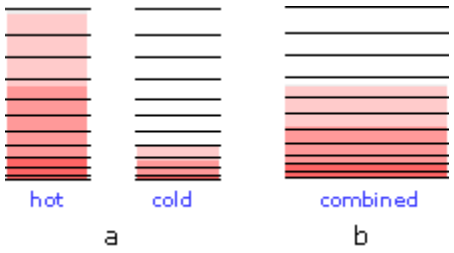
The same can in fact be said for even other highly probable distributions, such as having 49.999% of the molecules in the left half of the container and 50.001% in the right half. Even though the number of possible configurations that would yield this distribution of molecules is inconceivably great, it is essentially negligible compared to the number that would correspond to an exact 50-percent distribution.



The illustration represents the allowed thermal energy states of an ideal gas. The larger the volume in which the gas is confined, the more closely-spaced are these states, resulting in a huge increase in the number of microstates into which the available thermal energy can reside; this can be considered the origin of the thermodynamic "driving force" for the spontaneous expansion of a gas.

Heat Spontaneously flows from Hot to Cold

Just as gases spontaneously change their volumes from “smaller-to-larger”, the flow of heat from a warmer body to a cooler always operates in the direction “warmer-to-cooler” because this allows thermal energy to occupy a larger number of energy states as new ones are made available by bringing the cooler body into contact with the warmer one; in effect, the thermal energy becomes more “diluted”.



When two bodies at different temperatures are placed in thermal contact, thermal energy flow from the warmer into the cooler one until they reach the same temperature.

a. Schematic depiction of the thermal energy states in two separated identical bodies at different temperatures (indicated by shading.)
b. When the bodies are brought into thermal contact, thermal energy flows from the higher occupied levels in the warmer object into the unoccupied levels of the cooler one until equal numbers are occupied .

As you might expect, the increase in the amount of energy spreading and sharing is proportional to amount of heat transferred q, but there is one other factor involved, and that is the temperature at which the transfer occurs. When a quantity of heat q passes into a system at temperature T, the degree of dilution of the thermal energy is given by q/T

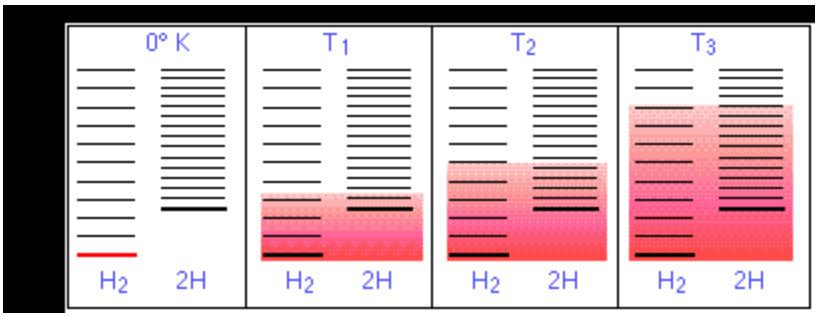
To understand why we have to divide by the temperature, consider the effect of very large and very small values of T in the denominator. If the body receiving the heat is initially at a very low temperature, relatively few thermal energy states are occupied, so the amount of energy spreading can be very great. Conversely, if the temperature is initially large, the number of new thermal energy states that become occupied will be negligible compared to the number already active.

Energy spreading and sharing in chemical reactions

When a chemical reaction takes place, two kinds of changes relating to thermal energy are involved:

1. The ways that thermal energy can be stored within the reactants will generally be different from those for the products. For example, in the reaction $H_2 \rightarrow 2\ H$, the reactant dihydrogen possesses vibrational and rotational energy states, while the atomic hydrogen in the product has translational states only— but the total number of translational states in two moles of H is twice as great as in one mole of H_2. Because of their extremely close spacing, translational states are the only ones that really count at ordinary temperatures, so we can say that thermal energy can become twice as diluted (“spread out”) in the product than in the reactant. If this were the only factor to consider, then dissociation of dihydrogen would always be spontaneous and H_2 would not exist.
   must be taken up from the surroundings in order to break the H–H bond. In other words, the *ground state* (the energy at which the manifold of energy states begins) is higher in H, as indicated by the vertical displacement of the right half in each of the four panels below.

Energy levels in H_2 and 2 H. The number of thermally accessible energy states (indicated by the shading in the diagram below) increases with temperature, but because 2 moles of H possess twice as many translational states as one mole of H_2, dissociation becomes increasingly favored at higher temperatures as more of these H states become thermally accessible.



This argument can be generalized to all molecules, illustrating an important principle:

> All molecules spontaneously absorb heat **and** dissociate at high temperatures.

The ability of energy to spread into the product molecules is constrained by the availability of sufficient thermal energy to produce these molecules. This is where the temperature comes in. At absolute zero the situation is very simple; no thermal energy is available to bring about dissociation, so the only component present will be dihydrogen.

- As the temperature increases, the number of populated energy states rises, as indicated by the shading in the diagram. At temperature T_1, the number of populated states of H_2 is greater than that of 2H, so some of the latter will be present in the equilibrium mixture, but only as the minority component.
- At some temperature T_2 the numbers of populated states in the two components of the reaction system will be identical, so the equilibrium mixture will contain H_2 and "2H" in equal amounts; that is, the mole ratio of H_2/H will be 1:2.
- As the temperature rises to T_3 and above, we see that the number of energy states that are thermally accessible in the product begins to exceed that for the reactant.

The result is exactly what the *Le Chatelier Principle* predicts: the equilibrium state for an endothermic reaction is shifted to the right at higher temperatures. This is all very well for helping you understand the direct connection between energy spreading when a chemical reaction occurs, but it is of little help in achieving our goal of predicting the direction and extent of chemical change. For this, we need to incorporate the concept of **energy spreading** into thermodynamics.

This page titled 15.1: Energy Spreading Drives Spontaneous Change is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.2: Entropy Rules

Learning Objectives

You are expected to be able to define and explain the significance of terms identified in ***bold***.

- A ***reversible*** *process* is one carried out in infinitessimal steps after which, when undone, both the system and surroundings (that is, the world) remain unchanged (see the example of gas expansion-compression below). Although true reversible change cannot be realized in practice, it can always be approximated.
- ((in which a process is carried out.
- As a process is carried out in a more reversible manner, the value of w approaches its maximum possible value, and q approaches its minimum possible value.
- Although q is not a state function, the quotient q_{rev}/T is, and is known as the *entropy*.
- ***energy*** within a system.
- The entropy of a substance increases with its molecular weight and complexity and with temperature. The entropy also increases as the pressure or concentration becomes smaller. Entropies of gases are much larger than those of condensed phases.
- The ***absolute entropy*** of a pure substance at a given temperature is the sum of all the entropy it would acquire on warming from absolute zero (where S=0) to the particular temperature.

Entropy is one of the most fundamental concepts of physical science, with far-reaching consequences ranging from cosmology to chemistry. It is also widely mis-represented as a measure of "disorder", as we discuss below. The German physicist Rudolf Clausius originated the concept as "energy gone to waste" in the early 1850s, and its definition went through a number of more precise definitions over the next 15 years.

Previously, we explained how the tendency of thermal energy to disperse as widely as possible is what drives all spontaneous processes, including, of course **chemical reactions**. We now need to understand how the direction and extent of the spreading and sharing of energy can be related to measurable thermodynamic properties of substances— that is, of **reactants** and **products**.

You will recall that when a quantity of heat q flows from a warmer body to a cooler one, permitting the available thermal energy to spread into and populate more microstates, that the ratio q/T measures the extent of this energy spreading. It turns out that we can generalize this to other processes as well, but there is a difficulty with using q because it is not a state function; that is, its value is dependent on the pathway or manner in which a process is carried out. This means, of course, that the quotient q/T cannot be a state function either, so we are unable to use it to get differences between reactants and products as we do with the other state functions. The way around this is to restrict our consideration to a special class of pathways that are described as *reversible*.

Reversible and irreversible changes

A change is said to occur reversibly when it can be carried out in a series of infinitesimal steps, each one of which can be undone by making a similarly minute change to the conditions that bring the change about. For example, the reversible expansion of a gas can be achieved by reducing the external pressure in a series of infinitesimal steps; reversing any step will restore the system and the surroundings to their previous state. Similarly, heat can be transferred reversibly between two bodies by changing the temperature difference between them in infinitesimal steps each of which can be undone by reversing the temperature difference.

The most widely cited example of an **irreversible** change is the free expansion of a gas into a vacuum. Although the system can always be restored to its original state by recompressing the gas, this would require that the surroundings perform work on the gas. Since the gas does no work on the surrounding in a free expansion (the external pressure is zero, so $P\Delta V$ = 0,) there will be a permanent change in the surroundings. Another example of irreversible change is the conversion of mechanical work into frictional heat; there is no way, by reversing the motion of a weight along a surface, that the heat released due to friction can be restored to the system.

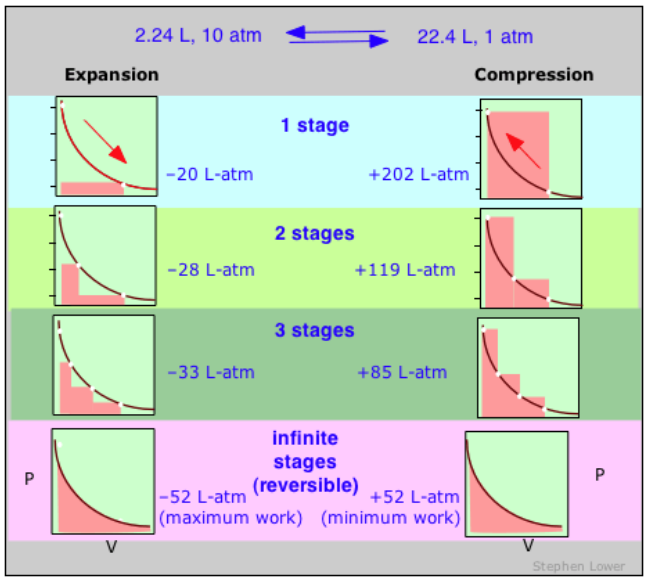


Figure 15.2.1: Reversible vs. Irreversible Expansions and Compressions

These diagrams show the same expansion and compression ±ΔV carried out in different numbers of steps ranging from a single step at the top to an "infinite" number of steps at the bottom. As the number of steps increases, the processes become less irreversible; that is, the difference between the work done in expansion and that required to re-compress the gas diminishes. In the limit of an ”infinite” number of steps (bottom), these work terms are identical, and both the system and surroundings (the “world”) are unchanged by the expansion-compression cycle. In all other cases the system (the gas) is restored to its initial state, **but the *surroundings* are forever changed.**

Definition: Reversible Changes

A reversible change is one carried out in such as way that, when undone, both the system and surroundings (that is, the world) remain unchanged.

It should go without saying, of course, that any process that proceeds in infinitesimal steps would take infinitely long to occur, so thermodynamic reversibility is an idealization that is never achieved in real processes, except when the system is already at equilibrium, in which case no change will occur anyway! So why is the concept of a reversible process so important?

The answer can be seen by recalling that the change in the internal energy that characterizes any process can be distributed in an infinity of ways between heat flow across the boundaries of the system and work done on or by the system, as expressed by the *First Law* $\Delta U = q + w$. Each combination of q and w represents a different pathway between the initial and final states. It can be shown that as a process such as the expansion of a gas is carried out in successively longer series of smaller steps, the absolute value of q approaches a minimum, and that of w approaches a maximum that is characteristic of the particular process.

Thus when a process is carried out reversibly, the w-term in the First Law expression has its greatest possible value, and the q-term is at its smallest. These special quantities w_{max} and q_{min} (which we denote as q_{rev} and pronounce “q-reversible”) have unique values for any given process and are therefore state functions.

Work and reversibility

For a process that reversibly exchanges a quantity of heat q_{rev} with the surroundings, the entropy change is defined as

$$\Delta S = \frac{q_{rev}}{T} \tag{15.2.1}$$

This is the basic way of evaluating ΔS for constant-temperature processes such as phase changes, or the isothermal expansion

of a gas. For processes in which the temperature is not constant such as heating or cooling of a substance, the equation must be integrated over the required temperature range, as discussed below.

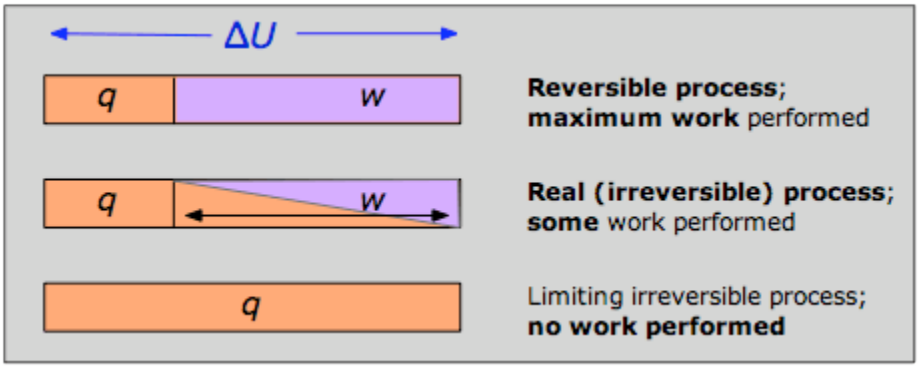


Figure 15.2.2: Note that the reversible condition implies w_{max} and q_{min}. The impossibility of extracting all of the internal energy as work is essentially a statement of the Second Law.

If no real process can take place reversibly, what use is an expression involving q_{rev}? This is a rather fine point that you should understand: although transfer of heat between the system and surroundings is impossible to achieve in a truly reversible manner, this idealized pathway is only crucial for the definition of ΔS; by virtue of its being a state function, the same value of ΔS will apply when the system undergoes the same net change via any pathway. For example, the entropy change a gas undergoes when its volume is doubled at constant temperature will be the same regardless of whether the expansion is carried out in 1000 tiny steps (as reversible as patience is likely to allow) or by a single-step (as irreversible a pathway as you can get!) expansion into a vacuum.

The physical meaning of entropy

Entropy is a measure of the degree of spreading and sharing of thermal energy within a system. This “spreading and sharing” can be spreading of the thermal energy into a larger volume of *space* or its sharing amongst previously inaccessible microstates of the system. The following table shows how this concept applies to a number of common processes.

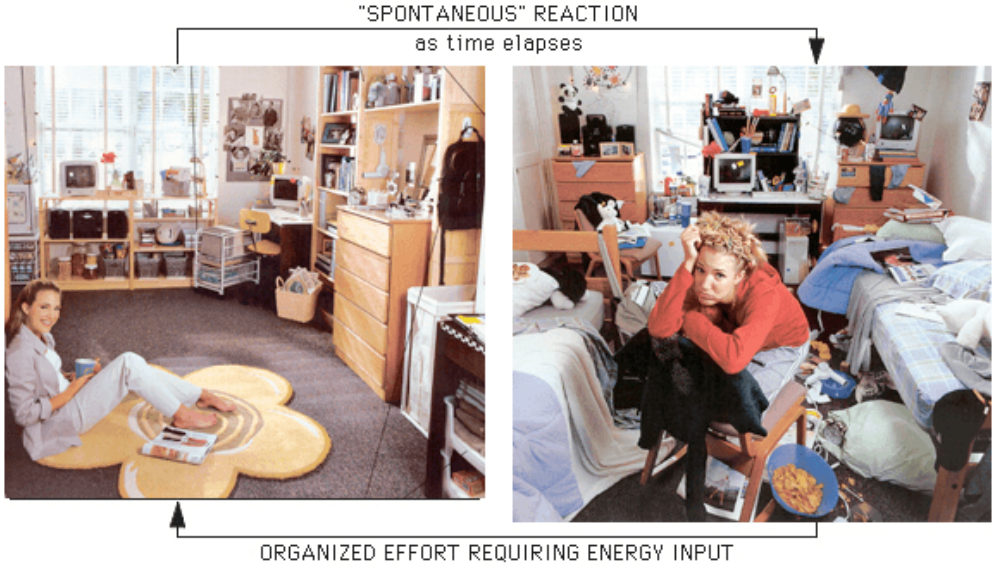
Table 15.2.1: Observations and Explanations in terms of Entropy changes

system and process	source of entropy increase of system
A deck of cards is shuffled, or 100 coins, initially heads up, are randomly tossed.	**This has nothing to do with entropy** because macro objects are unable to exchange thermal energy with the surroundings within the time scale of the process
Two identical blocks of copper, one at 20°C and the other at 40°C, are placed in contact.	The cooler block contains more unoccupied microstates, so heat flows from the warmer block until equal numbers of microstates are populated in the two blocks.
A gas expands isothermally to twice its initial volume.	A constant amount of thermal energy spreads over a larger volume of space
1 mole of water is heated by 1C°.	The increased thermal energy makes additional microstates accessible. (The increase is by a factor of about $10^{20,000,000,000,000,\,000,000,000}$.)
Equal volumes of two gases are allowed to mix.	The effect is the same as allowing each gas to expand to twice its volume; the thermal energy in each is now spread over a larger volume.
One mole of dihydrogen, H_2, is placed in a container and heated to 3000K.	Some of the H_2 dissociates to H because at this temperature there are more thermally accessible microstates in the 2 moles of H.
The above reaction mixture is cooled to 300K.	The composition shifts back to virtually all H_2because this molecule contains more thermally accessible microstates at low temperatures.

Entropy is an *extensive* quantity; that is, it is proportional to the quantity of matter in a system; thus 100 g of metallic copper has twice the entropy of 50 g at the same temperature. This makes sense because the larger piece of copper contains twice as many

quantized energy levels able to contain the thermal energy.

Entropy is still described, particularly in older textbooks, as a measure of disorder. In a narrow technical sense this is correct, since the spreading and sharing of thermal energy does have the effect of randomizing the disposition of thermal energy within a system. But to simply equate entropy with “disorder” without further qualification is extremely misleading because it is far too easy to forget that entropy (and thermodynamics in general) applies only to molecular-level systems capable of exchanging thermal energy with the surroundings. Carrying these concepts over to macro systems may yield compelling analogies, but it is no longer science. it is far better to avoid the term “disorder” altogether in discussing entropy.



Entropy and Probability

The distribution of thermal energy in a system is characterized by the number of quantized microstates that are accessible (i.e., among which energy can be shared); the more of these there are, the greater the entropy of the system. This is the basis of an alternative (and more fundamental) definition of entropy

$$S = k \ln \Omega \tag{15.2.2}$$

in which *k* is the Boltzmann constant (the gas constant per molecule, 1.38×10^{-23} J K^{-1}) and Ω (omega) is the number of microstates that correspond to a given macrostate of the system. The more such microstates, the greater is the probability of the system being in the corresponding macrostate. For any physically realizable macrostate, the quantity Ω is an unimaginably large number, typically around $10^{10^{25}}$ for one mole. By comparison, the number of atoms that make up the earth is about 10^{50}. But even though it is beyond human comprehension to compare numbers that seem to verge on infinity, the thermal energy contained in actual physical systems manages to discover the largest of these quantities with no difficulty at all, quickly settling in to the most probable macrostate for a given set of conditions.

The reason S depends on the *logarithm* of Ω is easy to understand. Suppose we have two systems (containers of gas, say) with S_1, Ω_1 and S_2, Ω_2. If we now redefine this as a single system (without actually mixing the two gases), then the entropy of the new system will be

$$S = S_1 + S_2 \tag{15.2.3}$$

but the number of microstates will be the product $\Omega_1\Omega_2$because for each state of system 1, system 2 can be in any of Ω_2 states. Because

$$\ln(\Omega_1\Omega_2) = \ln \Omega_1 + \ln \Omega_2 \tag{15.2.4}$$

Hence, the additivity of the entropy is preserved.

If someone could make a movie showing the motions of individual atoms of a gas or of a chemical reaction system in its equilibrium state, there is no way you could determine, on watching it, whether the film is playing in the forward or reverse direction. Physicists describe this by saying that such systems possess *time-reversal symmetry*; neither classical nor quantum mechanics offers any clue to the direction of time.

However, when a movie showing changes at the *macro*scopic level is being played backward, the weirdness is starkly apparent to anyone; if you see books flying off of a table top or tea being sucked back up into a tea bag (or a chemical reaction running in reverse), you will immediately know that something is wrong. At this level, time clearly has a direction, and it is often noted that because the entropy of the world as a whole always increases and never decreases, it is entropy that gives time its direction. It is for this reason that entropy is sometimes referred to as "time's arrow".

But there is a problem here: conventional thermodynamics is able to define entropy change only for reversible processes which, as we know, take infinitely long to perform. So we are faced with the apparent paradox that thermodynamics, which deals only with differences between states and not the journeys between them, is unable to describe the very process of change by which we are aware of the flow of time.

The direction of time is revealed to the chemist by the progress of a reaction toward its state of equilibrium; once equilibrium is reached, the net change that leads to it ceases, and from the standpoint of that particular system, the flow of time stops. If we extend the same idea to the much larger system of the world as a whole, this leads to the concept of the "heat death of the universe" that was mentioned briefly in the previous lesson.

Absolute Entropies

Energy values, as you know, are all relative, and must be defined on a scale that is completely arbitrary; there is no such thing as the absolute energy of a substance, so we can arbitrarily define the enthalpy or internal energy of an element in its most stable form at 298K and 1 atm pressure as zero. The same is *not* true of the entropy; since entropy is a measure of the “dilution” of thermal energy, it follows that the less thermal energy available to spread through a system (that is, the lower the temperature), the smaller will be its entropy. In other words, as the absolute temperature of a substance approaches zero, so does its entropy. This principle is the basis of the *Third law of thermodynamic*s, which states that the entropy of a perfectly-ordered solid at 0 K is zero.

Third law of thermodynamics

The entropy of a perfectly-ordered solid at 0 K is zero.

The *absolute entropy* of a substance at any temperature above 0 K must be determined by calculating the increments of heat *q* required to bring the substance from 0 K to the temperature of interest, and then summing the ratios *q*/*T*. Two kinds of experimental measurements are needed:

1. The enthalpies associated with any **phase changes** the substance may undergo within the temperature range of interest. Melting of a solid and vaporization of a liquid correspond to sizeable increases in the number of microstates available to accept thermal energy, so as these processes occur, energy will flow into a system, filling these new microstates to the extent required to maintain a constant temperature (the freezing or boiling point); these inflows of thermal energy correspond to the heats of fusion and vaporization. The entropy increase associated with melting, for example, is just $\Delta H_{fusion}/T_m$.
2. The **heat capacity** *C* of a phase expresses the quantity of heat required to change the temperature by a small amount ΔT , or more precisely, by an infinitesimal amount dT . Thus the entropy increase brought about by warming a substance over a range of temperatures that does not encompass a phase transition is given by the sum of the quantities $C\ dT/T$ for each increment of temperature dT . This is of course just the integral

$$S_{0^o \rightarrow T^o} = \int_{o^o}^{T^o} \frac{C_p}{T} dt \tag{15.2.5}$$

Because the heat capacity is itself slightly temperature dependent, the most precise determinations of absolute entropies require that the functional dependence of *C* on *T* be used in the above integral in place of a constant *C*.

$$S_{0^o \rightarrow T^o} = \int_{o^o}^{T^o} \frac{C_p(T)}{T} dt \tag{15.2.6}$$

When this is not known, one can take a series of heat capacity measurements over narrow temperature increments ΔT and measure the area under each section of the curve in Figure 15.2.3.

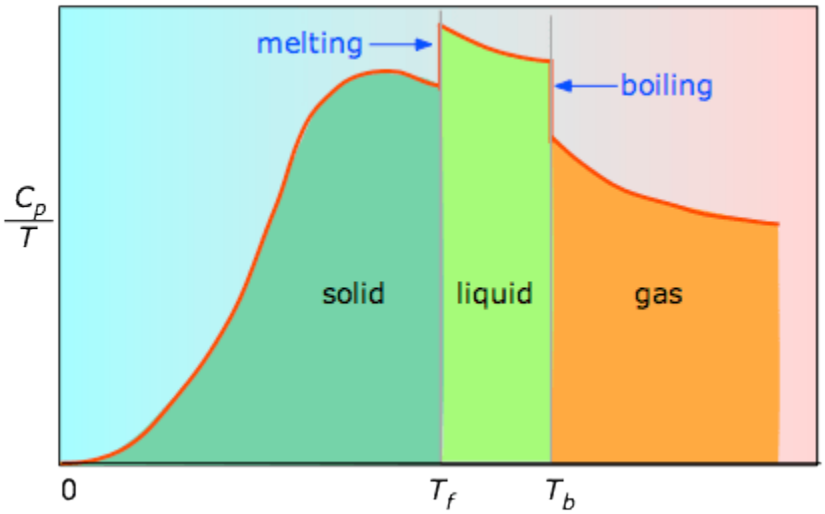


Figure 15.2.3: Heat capacity/temperature as a function of temperature

The area under each section of the plot represents the entropy change associated with heating the substance through an interval ΔT. To this must be added the enthalpies of melting, vaporization, and of any solid-solid phase changes. Values of C_p for temperatures near zero are not measured directly, but can be estimated from quantum theory.

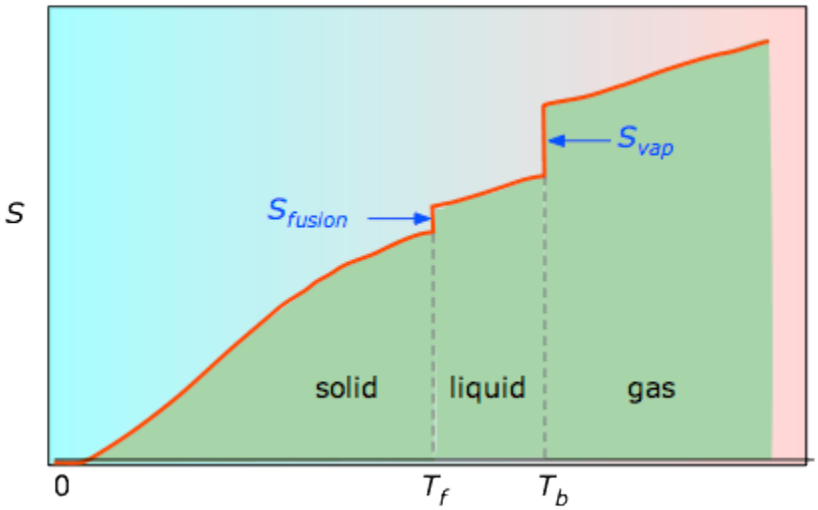


Figure 15.2.4: Molar entropy as a function of temperature

/ T_b are added to obtain the absolute entropy at temperature T. As shown in Figure 15.2.4 above, the entropy of a substance increases with temperature, and it does so for two reasons:

- As the temperature rises, more microstates become accessible, allowing thermal energy to be more widely dispersed. This is reflected in the gradual increase of entropy with temperature.
- The molecules of solids, liquids, and gases have increasingly greater freedom to move around, facilitating the spreading and sharing of thermal energy. Phase changes are therefore accompanied by massive and discontinuous increase in the entropy.

Standard Entropies of substances

The *standard entropy* of a substance is its entropy at 1 atm pressure. The values found in tables are normally those for 298K, and are expressed in units of J K^{-1} mol^{-1}. The table below shows some typical values for gaseous substances.

Table 15.2.2: Standard entropies of some gases at 298 K, J K^{-1} mol^{-1}

He	126	H_2	131	CH_4	186
Ne	146	N_2	192	$H_2O(g)$	187
Ar	155	CO	197	CO_2	213
Kr	164	F_2	203	C_2H_6	229
Xe	170	O_2	205	*n* -C_3H_8	270
Cl_2	223	*n* -C_4H_{10}	310		

Note especially how the values given in this Table 15.2.2:illustrate these important points:

- Although the standard internal energies and enthalpies of these substances would be zero, the entropies are not. This is because there is no absolute scale of energy, so we conventionally set the “energies of formation” of elements in their standard states to zero. Entropy, however, measures not energy itself, but its dispersal amongst the various quantum states available to accept it, and these exist even in pure elements.
- It is apparent that entropies generally increase with molecular weight. For the noble gases, this is of course a direct reflection of the principle that translational quantum states are more closely packed in heavier molecules, allowing of them to be occupied.
- The entropies of the diatomic and polyatomic molecules show the additional effects of rotational quantum levels.

Table 15.2.3: Standard entropies of some solid elements at 298 K, J K^{-1} mol^{-1}

C(diamond)	C(graphite)	Fe	Pb	Na	S(rhombic)	Si	W
2.5	5.7	27.1	51.0	64.9	32.0	18.9	33.5

The entropies of the solid elements are strongly influenced by the manner in which the atoms are bound to one another. The contrast between diamond and graphite is particularly striking; graphite, which is built up of loosely-bound stacks of hexagonal sheets, appears to be more than twice as good at soaking up thermal energy as diamond, in which the carbon atoms are tightly locked into a three-dimensional lattice, thus affording them less opportunity to vibrate around their equilibrium positions. Looking at all the examples in the above table, you will note a general inverse correlation between the hardness of a solid and its entropy. Thus sodium, which can be cut with a knife, has almost twice the entropy of iron; the much greater entropy of lead reflects both its high atomic weight and the relative softness of this metal. These trends are consistent with the oft-expressed principle that the more “disordered” a substance, the greater its entropy.

Table 15.2.4: Standard entropy of water at 298 K, J K^{-1} mol^{-1}

solid	liquid	gas
41	70	186

Gases, which serve as efficient vehicles for spreading thermal energy over a large volume of space, have much higher entropies than condensed phases. Similarly, liquids have higher entropies than solids owing to the multiplicity of ways in which the molecules can interact (that is, store energy.)

How Entropy depends on Concentration

As a substance becomes more dispersed in space, the thermal energy it carries is also spread over a larger volume, leading to an increase in its entropy. Because entropy, like energy, is an extensive property, a dilute solution of a given substance may well possess a smaller entropy than the same volume of a more concentrated solution, but the entropy *per mole* of solute (the *molar*

entropy) will of course always increase as the solution becomes more dilute.

For gaseous substances, the volume and pressure are respectively direct and inverse measures of concentration. For an ideal gas that expands at a constant temperature (meaning that it absorbs heat from the surroundings to compensate for the work it does during the expansion), the increase in entropy is given by

$$\Delta S = R \ln \left(\frac{V_2}{V_1} \right) \tag{15.2.7}$$

Note: If the gas is allowed to cool during the expansion, the relation becomes more complicated and will best be discussed in a more advanced course.

Because the pressure of a gas is inversely proportional to its volume, we can easily alter the above relation to express the entropy change associated with a change in the pressure of a perfect gas:

$$\Delta S = R \ln \left(\frac{P_1}{P_2} \right) \tag{15.2.8}$$

Expressing the entropy change directly in concentrations, we have the similar relation

$$\Delta S = R \ln \left(\frac{c_1}{c_2} \right) \tag{15.2.9}$$

Although these equations strictly apply only to perfect gases and cannot be used at all for liquids and solids, it turns out that in a dilute solution, the solute can often be treated as a gas dispersed in the volume of the solution, so the last equation can actually give a fairly accurate value for the entropy of dilution of a solution. We will see later that this has important consequences in determining the equilibrium concentrations in a homogeneous reaction mixture.

How thermal energy is stored in molecules

Thermal energy is the portion of a molecule's energy that is proportional to its *temperature*, and thus relates to motion at the molecular scale. What kinds of molecular motions are possible? For monatomic molecules, there is only one: actual movement from one location to another, which we call *translation*. Since there are three directions in space, all molecules possess three modes of *translational motion*.

For polyatomic molecules, two additional kinds of motions are possible. One of these is *rotation*; a linear molecule such as CO_2 in which the atoms are all laid out along the x-axis can rotate along the y- and z-axes, while molecules having less symmetry can rotate about all three axes. Thus linear molecules possess two modes of rotational motion, while non-linear ones have three rotational modes. Finally, molecules consisting of two or more atoms can undergo internal *vibrations*. For freely moving molecules in a gas, the number of vibrational modes or patterns depends on both the number of atoms and the shape of the molecule, and it increases rapidly as the molecule becomes more complicated.

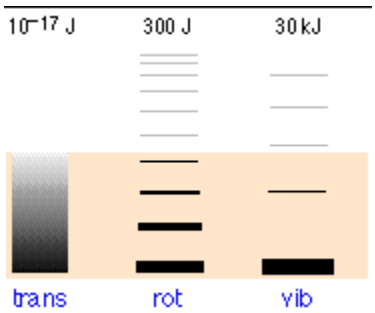


Figure 15.2.5

The relative populations of the quantized translational, rotational and vibrational energy states of a typical diatomic molecule are depicted by the thickness of the lines in this schematic (not-to-scale!) diagram. The colored shading indicates the total thermal energy available at a given temperature. The numbers at the top show order-of-magnitude spacings between adjacent levels. It is readily apparent that virtually all the thermal energy resides in translational states.

Notice the greatly different spacing of the three kinds of energy levels. This is extremely important because it determines the number of energy quanta that a molecule can accept, and, as the following illustration shows, the number of different ways this energy can be distributed amongst the molecules.

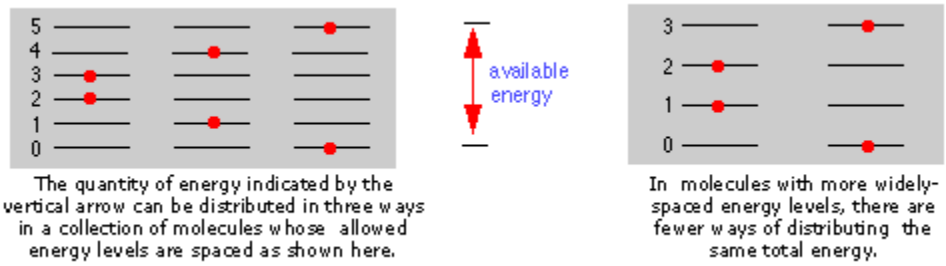


Figure 15.2.6

The more closely spaced the quantized energy states of a molecule, the greater will be the number of ways in which a given quantity of thermal energy can be shared amongst a collection of these molecules.

Note

The spacing of molecular energy states becomes closer as the mass and number of bonds in the molecule increases, so we can generally say that the more complex the molecule, the greater the density of its energy states.

Quantum states, microstates, and energy spreading

At the atomic and molecular level, all energy is *quantized*; each particle possesses discrete states of kinetic energy and is able to accept thermal energy only in packets whose values correspond to the energies of one or more of these states. Polyatomic molecules can store energy in rotational and vibrational motions, and all molecules (even monatomic ones) will possess translational kinetic energy (thermal energy) at all temperatures above absolute zero. The energy difference between adjacent translational states is so minute that translational kinetic energy can be regarded as *continuous* (non-quantized) for most practical purposes.

The number of ways in which thermal energy can be distributed amongst the allowed states within a collection of molecules is easily calculated from simple statistics, but we will confine ourselves to an example here. Suppose that we have a system consisting of three molecules and three quanta of energy to share among them. We can give all the kinetic energy to any one molecule, leaving the others with none, we can give two units to one molecule and one unit to another, or we can share out the energy equally and give one unit to each molecule. All told, there are ten possible ways of distributing three units of energy among three identical molecules as shown here:

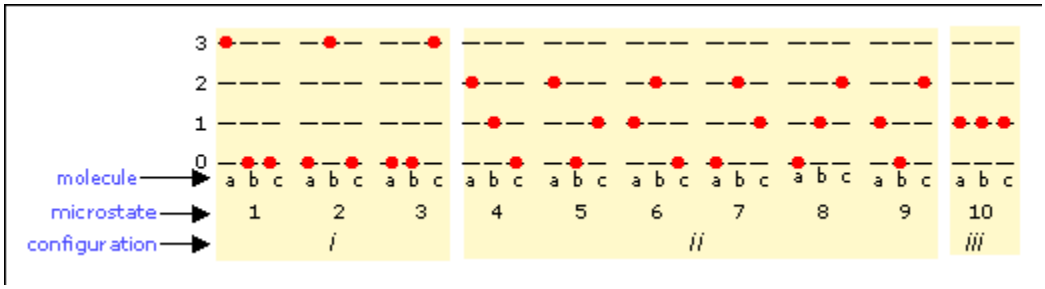


Figure 15.2.7

Each of these ten possibilities represents a distinct *microstate* that will describe the system at any instant in time. Those microstates that possess identical distributions of energy among the accessible quantum levels (and differ only in which particular molecules occupy the levels) are known as **configurations**. Because all microstates are equally probable, the probability of any one configuration is proportional to the number of microstates that can produce it. Thus in the system shown above, the configuration labeled *ii* will be observed 60% of the time, while *iii* will occur only 10% of the time.

As the number of molecules and the number of quanta increases, the number of accessible microstates grows explosively; if 1000 quanta of energy are shared by 1000 molecules, the number of available microstates will be around 10^{600}— a number that greatly exceeds the number of atoms in the observable universe! The number of possible configurations (as defined above) also increases, but in such a way as to greatly reduce the probability of all but the most probable configurations. Thus for a sample of a gas large enough to be observable under normal conditions, only a single configuration (energy distribution amongst the quantum states) need be considered; even the second-most-probable configuration can be neglected.

The bottom line: any collection of molecules large enough in numbers to have chemical significance will have its therrmal energy distributed over an unimaginably large number of microstates. The number of microstates increases exponentially as more energy states ("configurations" as defined above) become accessible owing to

- Addition of energy quanta (higher temperature),
- Increase in the number of molecules (resulting from dissociation, for example).
- the volume of the system increases (which decreases the spacing between energy states, allowing more of them to be populated at a given temperature.)

Heat Death: Energy-spreading changes the world

Energy is *conserved*; if you lift a book off the table, and let it fall, the total amount of energy in the world remains unchanged. All you have done is transferred it from the form in which it was stored within the glucose in your body to your muscles, and then to the book (that is, you did work on the book by moving it up against the earth's gravitational field). After the book has fallen, this same quantity of energy exists as thermal energy (heat) in the book and table top.

What *has* changed, however, is the availability of this energy. Once the energy has spread into the huge number of thermal microstates in the warmed objects, the probability of its spontaneously (that is, by chance) becoming un-dispersed is essentially zero. Thus although the energy is still "there", it is forever beyond utilization or recovery. The profundity of this conclusion was recognized around 1900, when it was first described at the "heat death" of the world. This refers to the fact that every spontaneous process (essentially every change that occurs) is accompanied by the "dilution" of energy. The obvious implication is that all of the molecular-level kinetic energy will be spread out completely, and nothing more will ever happen.

Why do gases tend to expand, but never contract?

Everybody knows that a gas, if left to itself, will tend to expand and fill the volume within which it is confined completely and uniformly. What "drives" this expansion? At the simplest level it is clear that with more space available, random motions of the individual molecules will inevitably disperse them throughout the space. But as we mentioned above, the allowed energy states that molecules can occupy are spaced more closely in a larger volume than in a smaller one. The larger the volume available to the gas, the greater the number of microstates its thermal energy can occupy. Since all such states within the thermally accessible range of energies are equally probable, the expansion of the gas can be viewed as a consequence of the tendency of thermal energy to be spread and shared as widely as possible. Once this has happened, the probability that this sharing of energy will reverse itself (that is, that the gas will spontaneously contract) is so minute as to be unthinkable.

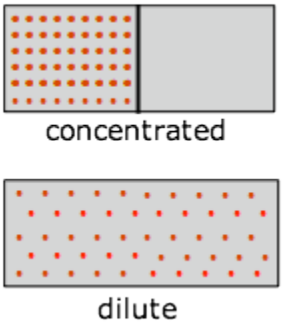


Figure 15.2.7

Imagine a gas initially confined to one half of a box (Figure 15.2.7). The barrier is then removed so that it can expand into the full volume of the container. We know that the entropy of the gas will increase as the thermal energy of its molecules spreads into the enlarged space. In terms of the spreading of thermal energy, Figure 23.2.X may be helpful. The tendency of a gas to expand is

due to the more closely-spaced thermal energy states in the larger volume (2).

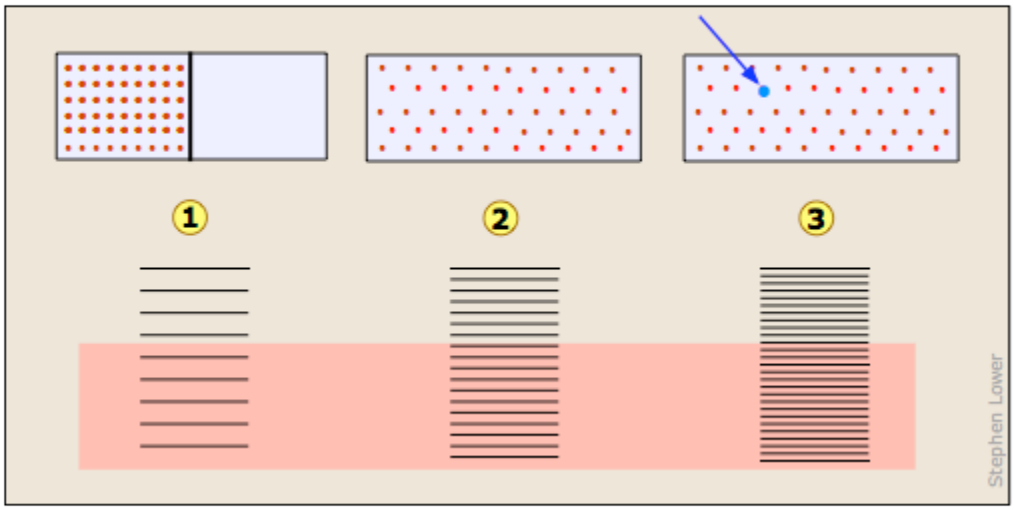


Figure 15.2.8: The pink shading represents the thermally-accessible range of microstates for a given temperature.

Entropy of mixing and dilution

Mixing and dilution really amount to the same thing, especially for idea gases. Replace the pair of containers shown above with one containing two kinds of molecules in the separate sections (Figure 15.2.9). When we remove the barrier, the "red" and "blue" molecules will each expand into the space of the other. (Recall Dalton's Law that "each gas is a vacuum to the other gas".) However, notice that although each gas underwent an expansion, the overall process amounts to what we call "mixing".

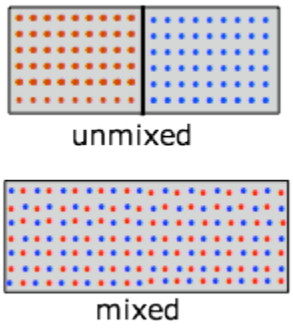


Figure 15.2.9: Mixing of a gas

What is true for gaseous molecules can, in principle, apply also to solute molecules dissolved in a solvent. But bear in mind that whereas the enthalpy associated with the expansion of a perfect gas is by definition zero, ΔH's of mixing of two liquids or of dissolving a solute in a solvent have finite values which may limit the miscibility of liquids or the solubility of a solute. But what's really dramatic is that when just *one molecule* of a second gas is introduced into the container ((3) in Figure 15.2.8), an unimaginably huge number of new configurations become possible, greatly increasing the number of microstates that are thermally accessible (as indicated by the pink shading above).

Why heat flows from hot to cold

Just as gases spontaneously change their volumes from “smaller-to-larger”, the flow of heat from a warmer body to a cooler one always operates in the direction “warmer-to-cooler” because this allows thermal energy to populate a larger number of energy microstates as new ones are made available by bringing the cooler body into contact with the warmer one; in effect, the thermal energy becomes more “diluted”.

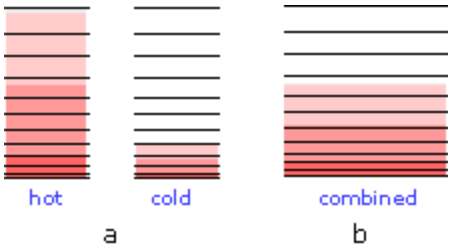


Figure 15.2.10: Part a of the figure is a schematic depiction of the thermal energy states in two separated identical bodies at different temperatures (indicated by shading.)

When the bodies are brought into thermal contact (*b*), thermal energy flows from the higher occupied levels in the warmer object into the unoccupied levels of the cooler one until equal numbers are occupied in both bodies, bringing them to the same temperature. As you might expect, the increase in the amount of energy spreading and sharing is proportional to the amount of heat transferred *q*, but there is one other factor involved, and that is the *temperature* at which the transfer occurs. When a quantity of heat *q* passes into a system at temperature *T*, the degree of dilution of the thermal energy is given by

$$\frac{q}{T} \tag{15.2.10}$$

To understand why we have to divide by the temperature, consider the effect of very large and very small values of *T*in the denominator. If the body receiving the heat is initially at a very low temperature, relatively few thermal energy states are initially occupied, so the amount of energy spreading into vacant states can be very great. Conversely, if the temperature is initially large, more thermal energy is already spread around within it, and absorption of the additional energy will have a relatively small effect on the degree of thermal disorder within the body.

Chemical reactions: why the equilibrium constant depends on the temperature

When a chemical reaction takes place, two kinds of changes relating to thermal energy are involved:

1. The ways that thermal energy can be stored within the reactants will generally be different from those for the products. For example, in the reaction $H_2 \rightarrow$ 2 H, the reactant dihydrogen possesses vibrational and rotational energy states, while the atomic hydrogen in the product has translational states only— but the total number of translational states in two moles of H is twice as great as in one mole of H_2. Because of their extremely close spacing, translational states are the only ones that really count at ordinary temperatures, so we can say that thermal energy can become twice as diluted ("spread out") in the product than in the reactant. If this were the only factor to consider, then dissociation of dihydrogen would always be spontaneous and this molecule would not exist.
2. In order for this dissociation to occur, however, a quantity of thermal energy (heat) $q = \Delta U$ must be taken up from the surroundings in order to break the H–H bond. In other words, the ground state (the energy at which the manifold of energy states begins) is higher in H, as indicated by the vertical displacement of the right half in each of the four panels below.

In Figure 15.2.11 are schematic representations of the translational energy levels of the two components H and H_2 of the hydrogen dissociation reaction. The shading shows how the relative populations of occupied microstates vary with the temperature, causing the equilibrium composition to change in favor of the dissociation product.

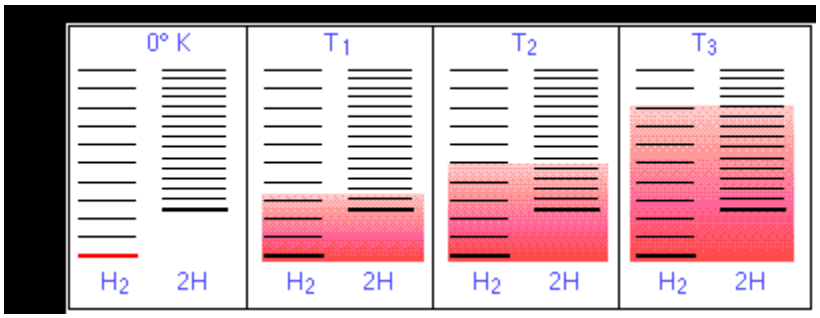


Figure 15.2.11:

The ability of energy to spread into the product molecules is constrained by the availability of sufficient thermal energy to produce

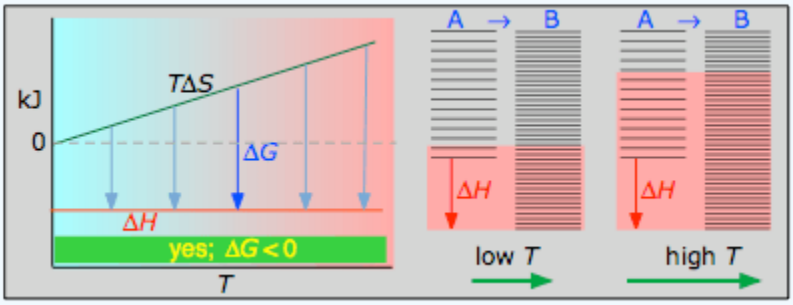
these molecules. This is where the temperature comes in. At absolute zero the situation is very simple; no thermal energy is available to bring about dissociation, so the only component present will be dihydrogen.

- As the temperature increases, the number of populated energy states rises, as indicated by the shading in the diagram. At temperature T_1, the number of populated states of H_2 is greater than that of 2H, so some of the latter will be present in the equilibrium mixture, but only as the minority component.
- At some temperature T_2 the numbers of populated states in the two components of the reaction system will be identical, so the equilibrium mixture will contain H_2 and "2H" in equal amounts; that is, the mole ratio of H_2/H will be 1:2.
- As the temperature rises to T_3 and above, we see that the number of energy states that are thermally accessible in the product begins to exceed that for the reactant, thus favoring dissociation.

The result is exactly what the LeChatelier Principle predicts: the equilibrium state for an endothermic reaction is shifted to the right at higher temperatures.

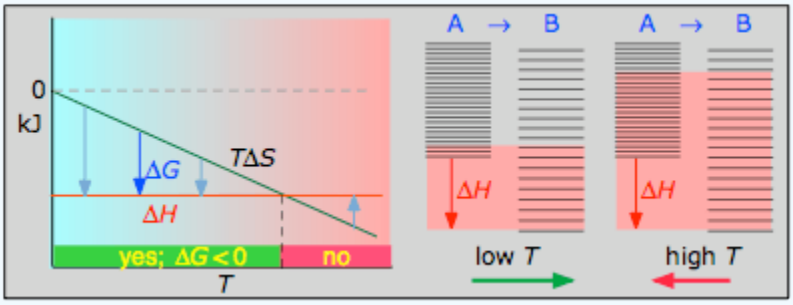
The following table generalizes these relations for the four sign-combinations of ΔH and ΔS. (Note that use of the standard $\Delta H°$ and $\Delta S°$ values in the example reactions is not strictly correct here, and can yield misleading results when used generally.)

Exothermic reaction, $\Delta S > 0$



This *combustion reaction*, like most such reactions, is **spontaneous at all temperatures**. The positive entropy change is due mainly to the greater mass of CO_2 molecules compared to those of O_2.

< 0

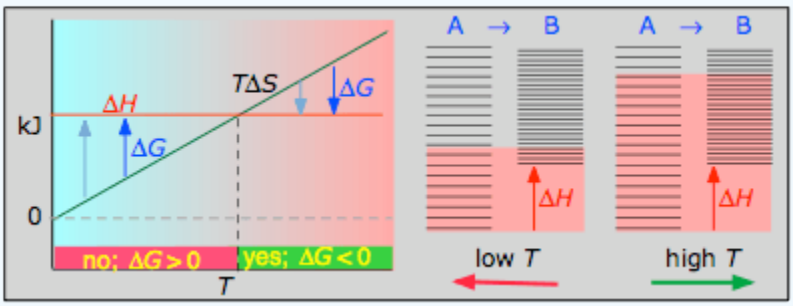


$3\ H_2 + N_2 \rightarrow 2\ NH_3(g)$

- $\Delta H° = -46.2$ kJ
- $\Delta S° = -389$ J K^{-1}
- $\Delta G° = -16.4$ kJ at 298 K

The decrease in moles of gas in the **Haber ammonia synthesis** drives the entropy change negative, making the reaction **spontaneous only at low temperatures**. Thus higher *T*, which speeds up the reaction, also reduces its extent.

> 0

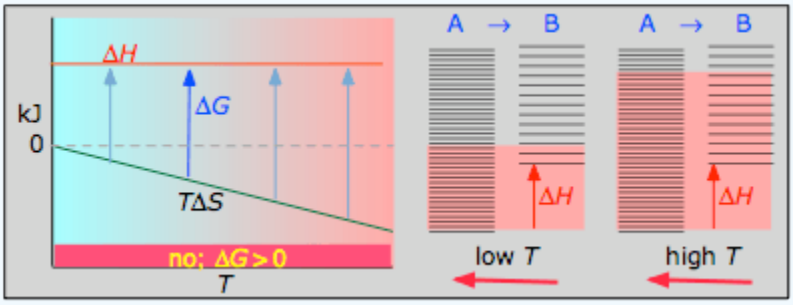


$N_2O_4(g) \rightarrow 2\ NO_2(g)$

- $\Delta H°$ = 55.3 kJ
- $\Delta S°$ = +176 J K^{-1}
- $\Delta G°$ = +2.8 kJ at 298 K

Dissociation reactions are typically endothermic with positive entropy change, and are therefore **spontaneous at high temperatures**.Ultimately, all molecules decompose to their atoms at sufficiently high temperatures.

< 0



$½\ N_2 + O_2 \rightarrow NO_2(g)$

- $\Delta H°$ = 33.2 kJ
- $\Delta S°$ = –249 J $K^{-}1$
- $\Delta G°$ = +51.3 kJ at 298 K

This reaction is **not spontaneous at any temperature**, meaning that *its reverse is always spontaneous*. But because the reverse reaction is kinetically inhibited, NO_2 can exist indefinitely at ordinary temperatures even though it is thermodynamically unstable.

Phase changes

Everybody knows that the **solid** is the stable form of a substance at low temperatures, while the gaseous state prevails at high temperatures. Why should this be? The diagram in Figure 15.2.12 shows that

1. the density of energy states is smallest in the solid and greatest (much, much greater) in the gas, and
2. the ground states of the liquid and gas are offset from that of the previous state by the heats of fusion and vaporization, respectively.

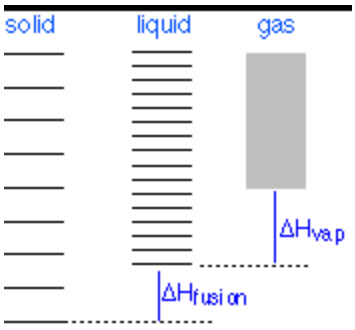


Figure 15.2.12:

Changes of phase involve exchange of energy with the surroundings (whose energy content relative to the system is indicated (with much exaggeration!) by the height of the yellow vertical bars in Figure 15.2.13. When solid and liquid are in equilibrium (middle section of diagram below), there is sufficient thermal energy (indicated by pink shading) to populate the energy states of both phases. If heat is allowed to flow into the surroundings, it is withdrawn selectively from the more abundantly populated levels of the liquid phase, causing the quantity of this phase to decrease in favor of the solid. The temperature remains constant as the heat of fusion is returned to the system in exact compensation for the heat lost to the surroundings. Finally, after the last trace of liquid has disappeared, the only states remaining are those of the solid. Any further withdrawal of heat results in a temperature drop as the states of the solid become depopulated.

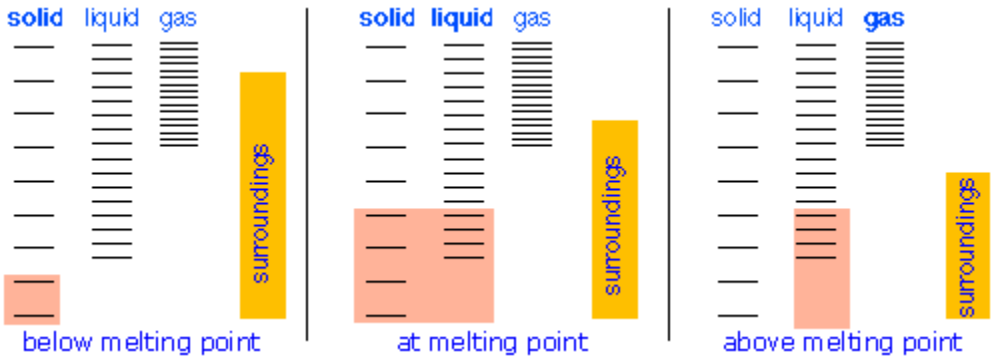


Figure 15.2.13:

Colligative Properties of Solutions

Vapor pressure lowering, boiling point elevation, freezing point depression and osmosis are well-known phenomena that occur when a non-volatile solute such as sugar or a salt is dissolved in a volatile solvent such as water. All these effects result from "dilution" of the solvent by the added solute, and because of this commonality they are referred to as *colligative properties* (Lat. *co ligare*, connected to.) The key role of the solvent concentration is obscured by the greatly-simplified expressions used to calculate the magnitude of these effects, in which only the solute concentration appears. The details of how to carry out these calculations and the many important applications of colligative properties are covered elsewhere. Our purpose here is to offer a more complete explanation of *why* these phenomena occur.

Basically, these all result from the effect of **dilution of the solvent** on its entropy, and thus in the increase in the density of energy states of the system in the solution compared to that in the pure liquid. Equilibrium between two phases (liquid-gas for boiling and solid-liquid for freezing) occurs when the energy states in each phase can be populated at **equal densities**. The temperatures at which this occurs are depicted by the shading.

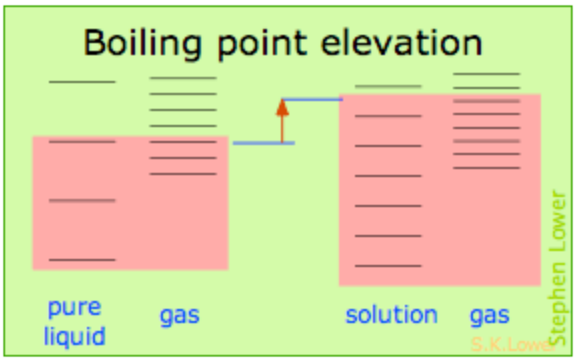


Figure 15.2.14: Boiling Point Elevation from an entropy perspective

Dilution of the solvent adds new energy states to the liquid, but does not affect the vapor phase. This raises the temperature required to make equal numbers of microstates accessible in the two phases.

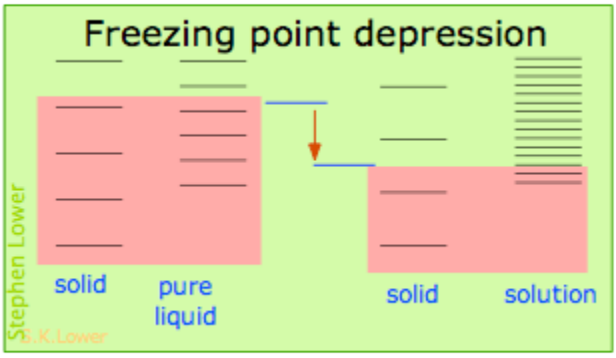


Figure 15.2.15: Freezing Point Depression from an entropy perspective

Dilution of the solvent adds new energy states to the liquid, but does not affect the solid phase. This reduces the temperature required to make equal numbers of states accessible in the two phases.

Effects of pressure on the entropy: Osmotic pressure

When a liquid is subjected to hydrostatic pressure— for example, by an inert, non-dissolving gas that occupies the vapor space above the surface, the vapor pressure of the liquid is raised (Figure 15.2.16). The pressure acts to compress the liquid very slightly, effectively narrowing the potential energy well in which the individual molecules reside and thus increasing their tendency to escape from the liquid phase. (Because liquids are not very compressible, the effect is quite small; a 100-atm applied pressure will raise the vapor pressure of water at 25°C by only about 2 torr.) In terms of the entropy, we can say that the applied pressure reduces the dimensions of the "box" within which the principal translational motions of the molecules are confined within the liquid, thus reducing the density of energy states in the liquid phase.

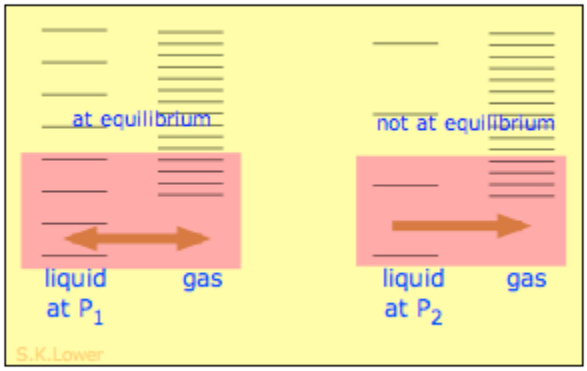


Figure 15.2.16:

Applying hydrostatic pressure to a liquid increases the spacing of its microstates, so that the number of energetically accessible states in the gas, although unchanged, is relatively greater— thus increasing the tendency of molecules to escape into the vapor phase. In terms of free energy, the higher pressure raises the free energy of the liquid, but does not affect that of the gas phase.

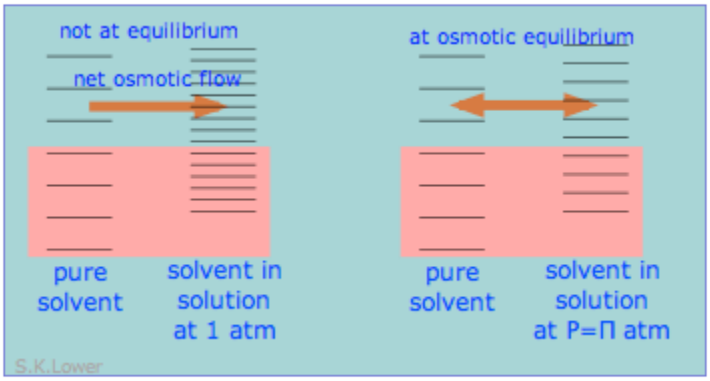


Figure 15.2.17: Osmotic Pressure from an entropy perspective

This phenomenon can explain **osmotic pressure**. Osmotic pressure, students must be reminded, is not what drives osmosis, but is rather the hydrostatic pressure that must be applied to the more concentrated solution (more dilute solvent) in order to *stop* osmotic flow of solvent into the solution. The effect of this pressure Π is to slightly increase the spacing of solvent energy states on the high-pressure (dilute-solvent) side of the membrane to match that of the pure solvent, restoring osmotic equilibrium.

This page titled 15.2: Entropy Rules is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.3: The Second Law of Thermodynamics

Learning Objectives

You are expected to be able to define and explain the significance of terms identified in *green type*.

- In any *macroscopic change*, the *entropy of the world* (that is, system + surroundings) always increases; it never decreases.
- Processes that do not exchange heat with the surroundings (such as the free expansion of a gas into a vacuum) involve entropy change of the system alone, and are always spontaneous.
- A *heat engine* is a device that converts heat into work. The fraction of heat that can be converted into work is limited by the fall in temperature between the input to the engine and the exhaust.
- According to the *Second Law of Thermodynamics*, complete conversion of heat into work by a *spontaneous cyclic process* is impossible.

The *First Law of thermodynamics*, expressed as $\Delta U = q + w$, is essentially a statement of the law of conservation of energy. The significance of this law is that it tells us that any proposed process that would violate this condition can be dismissed as impossible, without even inquiring further into the details of the process.

Why is the First Law not enough?

For simple mechanical operations on macroscopic objects, the First Law, conservation of energy, is all we usually need to determine such things as how many joules of energy is required to lift a weight or to boil some water, how many grams of glucose you must metabolize in order to climb a hill, or how much fuel your car needs to drive a given distance.

But if you think about it, there are a number of "simple mechanical operations" that never occur, even though they would not violate energy conservation.

- Suppose you drop a book onto a table top. The kinetic energy contained in the falling book is dispersed as thermal energy, slightly warming the book and the table top. According to the First Law, there is no reason why placing pre-warmed book on a warmed table top should not be able to propel the book back into the air. Similarly, why can't the energy imparted to the nail (and to the wood) by a hammer not pop the nail back out?
- One might propose a scheme to propel a ship by means of a machine that takes in seawater, extracts part of its thermal energy which is used to rotate the propeller, and then tosses the resulting ice cubes overboard. As long as the work done to turn the propeller is no greater than the heat required to melt the ice, the First Law is satisfied.
- Because motion of the air molecules is completely random, there is no reason why all of the molecules in one half of a room cannot suddenly "decide" to move into the other half, asphyxiating the unfortunate occupants of that side. (To the extent that air behaves as a perfect gas, this doesn't involve the First Law at all.)

What do all these scenarios that conform to the First Law but are nevertheless never seen to occur have in common? In every case, energy becomes less spread out, less "diluted". In the first two examples, thermal energy (dispersed) gets concentrated into organized kinetic energy of a macroscopic object— a book, a propeller. In the third case, the thermal energy gets concentrated into a smaller volume as the gas contracts.

The second law of thermodynamics says in effect, that the extent to which any natural process can occur is limited by the dilution of thermal energy (increase in entropy) that accompanies it, and once the change has occurred, it can never be un-done without spreading even more energy around. This is one of the most profound laws of nature, and should be a part of every educated person's world view. It is unfortunate that this law is so widely misrepresented as simply ordaining the increase in "disorder". A more brief statement of the Second Law (for those who know the meaning of "entropy") is

The more formal and historical ways of stating the Second Law will be presented farther below after we introduce the topic of

heat engines. It is also worth knowing this important consequence of the Second Law: Just because the energy is “there” does not mean it will be available to do anything useful.

Entropy and Spontaneous Change

We explained how processes that take place spontaneously always proceed in a direction that leads to the spreading and sharing of thermal energy.

- A book falls to the tabletop (rather than absorbing heat and jumping up from it) because its kinetic energy changes into thermal energy which is widely dispersed into the molecules of the book and the table.
- A gas expands and solutions mix because the thermal energy their molecules possess get spread over a larger volume of space.
- Hydrogen gas dissociates into H atoms which share thermal energy amongst more particles and a greater volume of space. (But only if the temperature is high enough to make the huge number of new microstates energetically accessible.)

Because all natural processes lead to the spreading and sharing of thermal energy, and because entropy is a measure of the extent to which energy is dispersed in the world, it follows that:

> In any spontaneous macroscopic change, the entropy of the world increases.

All natural processes that allow the free exchange of thermal energy amongst chemically-significant numbers of particles are accompanied by a spreading or “dilution” of energy that leaves the world forever changed. In other words, **all spontaneous change leads to an increase in the entropy of the world**. At first sight, this might seem to be inconsistent with our observations of very common instances in which there is a clear decrease in entropy, such as the freezing of a liquid, the formation of a precipitate, or the growth of an organism.

System + surroundings = the world!

... but **it’s the entropy of the system *plus* surroundings that counts!** It is important to understand that the criterion for spontaneous change is the entropy change of the system and the surroundings— that is, of the “world”, which we denote by ΔS_{total}:

$$\Delta S_{total} = \Delta S_{system} + \Delta S_{surroundings} \tag{15.3.1}$$

The only way the entropy of the *surroundings* can be affected is by exchange of heat with the system:

$$\Delta S_{surroundings} = \frac{q_{surr}}{T} \tag{15.3.2}$$

Thus the freezing of water is accompanied by a flow of heat (the heat of fusion) into the surroundings, causing ΔS_{surr} to increase. At temperatures below the freezing point, this increase more than offsets the decrease in the entropy of the water itself, so ΔS_{world} exceeds zero and the process is spontaneous. The problem example below works this out in detail for a specific example.

Note that it does not matter whether the change in the system occurs reversibly or irreversibly; as mentioned previously, it is always possible to define an alternative (irreversible) pathway in which the amount of heat exchanged with the surroundings is the same as q_{rev} ; because ΔS is a state function, the entropy change of the surroundings will have the same value as for the unrealizable reversible pathway.

If there is no flow of heat into or out of the surroundings, the entropy change of the system and that of the world are identical. Examples of such processes, **which are always spontaneous**, are the free expansion of an ideal gas into a vacuum, and the mixing of two ideal gases. In practice, almost all processes involving mixing and diffusion can be regarded as driven exclusively by the entropy increase of the system.

Most processes involving chemical and phase changes involve the exchange of heat with the surroundings, so their tendency to occur cannot always be predicted by focusing attention on the system alone. Further, owing to the $-q/T$ term in $\Delta S_{surroundings}$, the spontaneity of all such processes will depend on the temperature, as we illustrated for the dissociation of H_2 previously.

As a quantitative example, let us consider the **freezing of water**. We know that liquid water will spontaneously change into ice

when the temperature drops below 0°C at 1 atm pressure. Since the entropy of the solid is less than that of the liquid, we know the entropy of the water (the system here) will decrease on freezing. The amount of decrease is found by dividing the heat of fusion of ice by the temperature for the reversible pathway, which occurs at the normal freezing point:

$$\Delta S_{system} = \frac{-6000\ J/mol}{273\ K} = -21.978\ J/mol \tag{15.3.3}$$

If the process is actually carried at 0°C, then the heat of fusion is transferred to the surroundings at the same temperature, and the entropy of the surroundings increases by

$$\Delta S_{surroundings} = \frac{6000\ J/mol}{273\ K} = 21.979\ J/mol \tag{15.3.4}$$

so that $\Delta S_{total} = 0$. Under these conditions the process can proceed in either direction (freezing or melting) without affecting the entropy of the world; this means that both ice and liquid water can be present simultaneously without any change occurring; the system is said to be in *equilibrium*.

Suppose now that the water is supercooled to –1°C before it freezes. The entropy change of the water still corresponds to the reversible value q_{rev}/T = (–6000J)/(273K). The entropy change of the surroundings, however, is now given by

$$\Delta S_{surroundings} = \frac{6000\ J/mol}{273\ K} = 22.059\ J/mol \tag{15.3.5}$$

The total entropy change is now

$$\Delta S_{total} = (-21.978 + 22.059)J;\ K^{-1}mol^{-1} = +0.081\ J\ K^{-1}mol^{-1} \tag{15.3.6}$$

indicating that the process can now occur ("is spontaneous") only in the one direction.

Why did we use 273 K when evaluating ΔS_{system} and 272 K for calculating $\Delta S_{surroundings}$? In the latter case it is possible to formulate a reversible pathway by which heat can be transferred to the surroundings at any temperature. ΔS_{system}, however, is a state function of water, and will vary with temperature only slightly.

Note that in order to actually freeze water, it must be cooled to very slightly below its normal freezing point, a condition known as supercooling. Freezing of supercooled water is of course an irreversible process (once it starts, it cannot be stopped except by raising the temperature by a finite amount), and the positive value of ΔS_{total} tells us that this process will occur spontaneously at temperatures below 273 K. Under these conditions, the process is driven by the entropy increase of the surroundings resulting from flow of the heat of fusion of water into the surroundings.

Does the entropy of the world ever decrease?

The principle that thermal energy (and the molecules carrying it) tends to spread out is based on simple statistics. It must be remembered, however, that the laws of probability have meaningful application only to systems made up of large numbers of independent actors. If you trap a hundred flies in a bottle, they will generally distribute themselves more or less uniformly throughout the container; if there are only four flies, however, it is quite likely that all of them will occasionally be located in one particular half of the bottle.

Similarly, you can trust with complete certainty that the spontaneous movement of half the molecules of the air to one side of the room you now occupy will not occur, even though the molecules are moving randomly and independently. On the other hand, if we consider a box whose dimensions are only a few molecular diameters, then we would expect that the random and short-term displacement of the small number of particles it contains to one side of the box would occur quite frequently. This is, in fact, the cause of the blueness of the sky: random fluctuations in the air density over tiny volumes of space whose dimensions are comparable with the wavelength of light results in selective scattering of the shorter wavelengths, so that blue light is scattered out, leaving the red light for the enjoyment of sunset-watchers to the east.

This refers to the irregular zig-zag-like movement of extremely small particles such as plant pollen when they are suspended in a drop of liquid. Any such particle is continually being buffeted by the thermal motions of the surrounding liquid molecules. If size of the particle is very large compared to that the the liquid molecules, the forces that result from collisions of these molecules with the particle will cancel out and the particle remains undisturbed. If the particle is very small, however (perhaps only a thousand times larger than a molecule of the liquid), then the chances that it will undergo sufficiently more hits from one direction than from another during a brief interval of time become significant.

In these two examples, the entropy of the system decreases without any compensating flow of heat into the surroundings, leading to a net (but only temporary) decrease in the entropy of the world. This does not represent a failure of the Second Law, however, because no one has ever devised a way to extract useful work from these processes.

Heat Engines

The Industrial Revolution of the 19th century was largely driven by the invention of the steam engine. The first major use of such engines was to pump water out of mines, whose flooding from natural seepage seriously limited the depths to which they could be driven, and thus the availability of the metal ores that were essential to the expansion of industrial activities. The steam engine is a type of heat engine, a device that converts heat, provided by burning a fuel, into mechanical work, typically delivered through the motion of a piston in opposition to an opposing force. An engine is therefore an energy conversion device in which, ideally, every joule of heat released by combustion of the fuel could be extracted as work at the output shaft; such an engine would operate at 100 percent efficiency.



However, engineers of the time were perplexed to find that the efficiencies of steam engines were rather low (usually around 20%), with most of the heat being exhausted uselessly to the environment. Everyone understood that an efficiency exceeding 100% would be impossible (that would violate conservation of energy, and thus the First Law), but it was not clear why efficiencies could not rise significantly beyond the small values observed even as mechanical designs improved



The answer was found by a young French engineer, Sadi Carnot, who in 1824 published an analysis of an idealized heat engine that is generally considered to be the foundation of the science of thermodynamics— notwithstanding the fact that Carnot still accepted the belief that heat is a fluid-like substance called “caloric”. We will not replicate his analysis here (this is normally done in more advanced courses in physical chemistry), but will simply state his conclusion in his own [translated] words:

"The production of motive power is then due in steam-engines not to an actual consumption of caloric, but to its **transportation** from a warm body to a cold body...the production of heat alone is not sufficient to give birth to the impelling power: it is necessary that there should also be cold; without it, the heat would be useless. The ultimate attainable efficiency of any heat engine will depend on the temperatures at which heat is supplied to and removed from it."

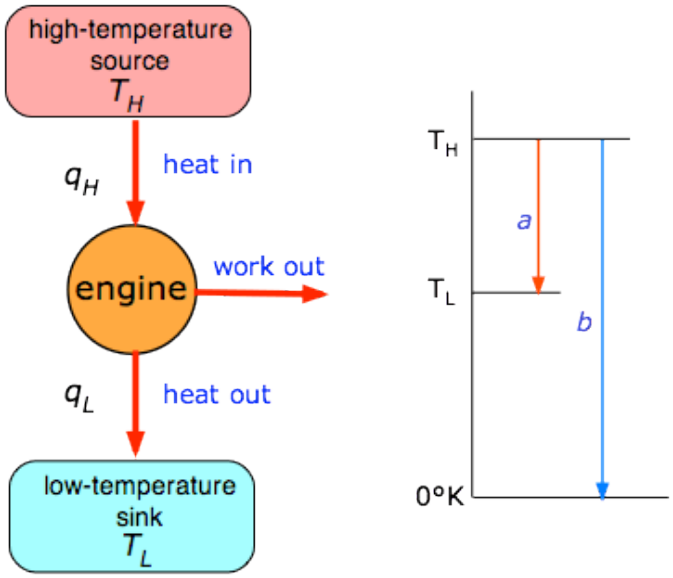


Figure 15.3.1: Efficiency of a heat engine. Left: common schematic representation of a heat engine. Right: diagramatic representation of Eq. 14 (below); the efficiency is the ratio of the temperature intervals a :b.

The left side of the figure represents a generalized heat engine into which a quantity of heat q_H, extracted from a source or “reservoir” at temperature T_H is partly converted into work w. The remainder of the heat q_L is exhausted to a reservoir at a lower temperature T_L. In practice, T_H would be the temperature of the steam in a steam engine, or the temperature of the combustion mixture in an internal combustion or turbine engine. The low temperature reservoir is ordinarily that of the local environment. The efficiency ε (*epsilon*) of a heat engine is the fraction of the heat abstracted from the high temperature reservoir that can be converted into work:

$$\varepsilon = \frac{w}{q_H} \tag{15.3.7}$$

Carnot's crucial finding (for which he would certainly have deserved a Nobel prize if these had existed at the time) is that the efficiency is proportional to the "distance" in temperature that the heat can "fall" as it passes through the engine:

$$\varepsilon = 1 - \frac{T_L}{T_H} \tag{15.3.8}$$

This is illustrated graphically in the right half of the figure just above, in which the efficiency is simply the fraction of the "complete" fall (in temperature) to absolute zero (arrow *b*) that the heat undergoes in the engine (arrow *a*.) Clearly, the only way to attain 100% efficiency would be to set the temperature of the exhaust reservoir to 0°K, which would be impossible. For most terrestrial heat engines, T_L is just the temperature of the environment, normally around 300 K, so the only practical way to improve the efficiency is to make T_H as high as possible. This is the reason that high pressure (superheated) steam is favored in commercial thermal power plants. The highest temperatures (and the greatest operating efficiencies) are obtained in gas turbine engines. However, as operating temperatures rise, the costs of dealing with higher steam pressures and the ability of materials such as turbine blades to withstand high temperatures become significant factors, placing an upper limit of around 600K on T_H, thus imposing a maximum of around 50 percent efficiency on thermal power generation.

For **nuclear plants**, in which safety considerations require lower steam pressures, the efficiency is lower. One consequence of this is that a larger fraction of the heat is exhausted to the environment, which may result in greater harm to aquatic organisms when the cooling water is returned to a stream or estuary.

Example 15.3.1

Several proposals have been made for building a heat engine that makes use of the temperature differential between the surface waters of the ocean and cooler waters that, being more dense, reside at greater depth. If the exhaust temperature is 5°C, what is the maximum amount of work that could be extracted from 1000 L of surface water at 10°C? (The specific heat capacity of water is 4.184 J $g^{-1}K^{-1}$.)

Solution

The amount of heat (q_H) that must be extracted to cool the water by 5 K is (4.184 J $g^{-1}K^{-1}$)(10^6 g)(5 K) = 2.09 × 10^7 J. The ideal thermodynamic efficiency is given by

$$1 - \frac{278\ K}{283\ K} = 0.018 \tag{15.3.9}$$

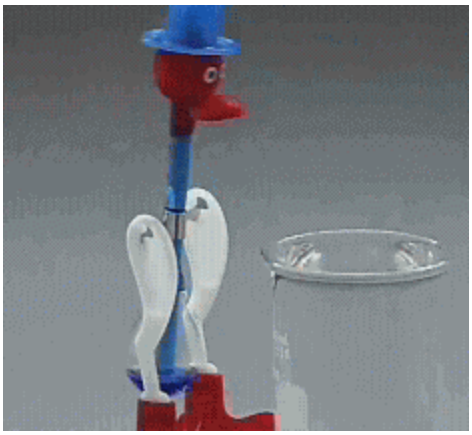
The amount of work that could be done would be

$$(0.018)(2.09 \times 10^7\ J) = 3.7 \times 10^6\ J \tag{15.3.10}$$

Comment: It may be only 1.8% efficient, but it's free!

The drinking bird as a heat engine

Few toys illustrate as many principles of physical science as this popular device that has been around for many years. At first glance it might appear to be a perpetual motion machine, but it's really just a simple heat engine. Modern "dippy birds" (as they are sometimes called) utilize dichloromethane as the working fluid.



This liquid boils at 39° C, and therefore has a rather high vapor pressure at room temperature. The liquid (to which a dye is often added for dramatic effect) is stored in a reservoir at the bottom of the bird. The bird's beak is covered with felt which, when momentarily dipped in water, creates a cooling effect as the water evaporates. This causes some of the CH_2Cl_2 vapor to condense in the head, reducing the pressure inside the device, causing more liquid to boil off and re-condense in the head. The redistribution of fluid upsets the balance, causing the bird to dip its beak back into the water. Once the head fills with liquid, it drains back into the bottom, tipping the bird upright to repeat the cycle.

We will leave it to you to relate this to the heat engine diagram above by identifying the heat source and sink, and estimate the thermodynamic efficiency of the engine.

Heat Pumps

If a heat engine is run “in reverse” by performing work on it (that is, changing “work out” to “work in” in Fig 8), it becomes a device for transporting heat against a thermal gradient. Refrigerators and air conditioners are the most commonly-encountered heat pumps. A heat pump can also be used to heat the interior of a building. In this application, the low temperature reservoir can be a heat exchanger buried in the earth or immersed in a well. In this application heat pumps are more efficient than furnaces or electric heating, but the capital cost is rather high.

The Second Law: what it means

It was the above observation by Carnot that eventually led to the formulation of the *Second Law of Thermodynamics* near the end of the 19th Century. One statement of this law (by Kelvin and Planck) is as follows:

Definition: Second Law of Thermodynamics (Kelvin Definition)

It is impossible for a *cyclic process* connected to a reservoir *at one temperature* to produce a *positive amount of work* in the surroundings.

To help you understand this statement and how it applies to heat engines, consider the schematic heat engine in the figure in which a working fluid (combustion gases or steam) expands against the restraining force of a weight that is mechanically linked to the piston. From a thermodynamic perspective, the working fluid is the system and everything else is surroundings. Expansion of the fluid occurs when it absorbs heat from the surroundings; return of the system to its initial state requires that the surrounding do work on the system. Now re-read the above statement of the Second Law, paying special attention to the italicized phrases which are explained below:

- A *cyclic process* is one in which the system returns to its initial state. A simple steam engine undergoes an expansion step (the power stroke), followed by a compression (exhaust stroke) in which the piston, and thus the engine, returns to its initial state before the process repeats.
- “**At one temperature**” means that the expansion and compression steps operate isothermally. This means that $\Delta U = 0$; just enough heat is absorbed by the system to perform the work required to raise the weight, so for this step $q = -w$.

- "**A positive amount of work in the surroundings**" means that the engine does more work on the surroundings than the surroundings do on the engine. Without this condition the engine would be useless.

Note carefully that the Second Law applies only to a *cyclic process*— isothermal expansion of a gas against a non-zero pressure always does work on the surroundings, but an engine must repeat this process continually; to do so it must be returned to its initial state at the end of every cycle. When operating isothermally, the work $-w$ it does on the surroundings in the expansion step (power stroke) is nullified by the work $+w$ the surroundings must do on the system in order to complete the cycle. The Second Law can also be stated in an alternative way:

Definition: Second Law of Thermodynamics (Planck Definition)

It is impossible to construct a machine operating in cycles that will convert heat into work without producing any other changes.

Thus the Second Law does allow an engine to convert heat into work, but only if "other changes" (transfer of a portion of the heat directly to the surroundings) are allowed. And since heat can only flow spontaneously from a source at a higher temperature to a sink at a lower temperature, the impossibility of isothermal conversion of heat into work is implied.

This page titled 15.3: The Second Law of Thermodynamics is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.4: Free Energy and the Gibbs Function

Learning Objectives

- Gibbs Energy is a state function defined as $G = H–TS$.
- The practical utility of the Gibbs function is that ΔG for any process is negative if it leads to an increase in the entropy of the world. Thus spontaneous change at a given temperature and pressure can only occur when it would lead to a decrease in G.
- The sign of the standard free energy change ΔG^o of a chemical reaction determines whether the reaction will tend to proceed in the forward or reverse direction.
- Similarly, the relative signs of ΔH^o and ΔH^o determine whether the spontaneity of a chemical reaction will be affected by the temperature, and if so, in what way.
- The existence of sharp melting and boiling points reflects the differing temperature dependancies of the free energies of the solid, liquid, and vapor phases of a pure substance, which are in turn reflect their differing entropies.

Previously, we saw that it is the **sum** of the entropy changes of the system and surroundings that determines whether a process will occur spontaneously. In chemical thermodynamics we prefer to focus our attention on the system rather than the surroundings, and would like to avoid having to calculate the entropy change of the surroundings explicitly.

In this unit we introduce a new thermodynamic function, the *free energy*, which turns out to be the single most useful criterion for predicting the direction of a chemical reaction and the composition of the system at equilibrium. However, the term "free energy", although still widely used, is rather misleading, so we will often refer to it as "Gibbs energy." The free energy enables us to do this for changes that occur at a constant temperature and pressure (the *Gibbs energy*) or constant temperature and volume (the *Helmholtz energy*.)

Free energy: the Gibbs function

The Gibbs energy (also known as the *Gibbs function* or *Gibbs Potential*) is defined as

$$G = H–TS \tag{15.4.1}$$

in which S refers to the entropy of the *system*. Since H, T and S are all state functions, so is G. Thus for any change in state (under constant temperature), we can write the extremely important relation

$$\Delta G = \Delta H–T\Delta S \tag{15.4.2}$$

How does this simple equation encompass the entropy change of the world ΔS_{total}, which we already know is the sole criterion for spontaneous change from the second law of thermodynamics? Starting with the definition

$$\Delta S_{total} = \Delta S_{surr} + \Delta S_{sys} \tag{15.4.3}$$

we would first like to get rid of ΔS_{surr}. How can a chemical reaction (a change in the *system*) affect the entropy of the *surroundings*? Because most reactions are either exothermic or endothermic, they are accompanied by a flow of heat q_p across the system boundary. The enthalpy change of the reaction ΔH is defined as the flow of heat into the system from the surroundings when the reaction is carried out at constant pressure, so the heat withdrawn from the surroundings will be $–q_p$ which will cause the entropy of the surroundings to change by $–q_p/T = –\Delta H/T$. We can therefore rewrite Equation 15.4.3 as

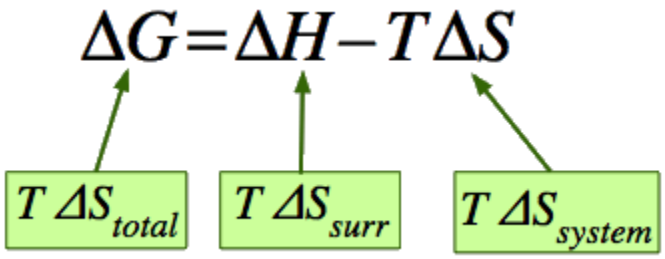
$$\Delta S_{total} = \frac{-\Delta H}{T} + \Delta S_{sys} \tag{15.4.4}$$

Multiplying each side by $-T$, we obtain

$$-T\Delta S_{total} = \Delta H - T\Delta S_{sys} \tag{15.4.5}$$

which expresses the entropy change of the world in terms of thermodynamic properties of the *system* exclusively. If $-T\Delta S_{total}$ is

denoted by ΔG, then we have Equation 15.4.2 which defines the *Gibbs energy* change for the process.



From the foregoing, you should convince yourself that G will decrease in any process occurring at constant temperature and pressure which is accompanied by an overall increase in the entropy. The constant temperature is a consequence of the temperature and the enthalpy appearing in the preceding Equation 15.4.5. Since most chemical and phase changes of interest to chemists take place under such conditions, the Gibbs energy is the most useful of all the thermodynamic properties of a substance, and (as we shall see in the lesson that follows this one) it is closely linked to the equilibrium constant.

Some textbooks and teachers say that the free energy, and thus the spontaneity of a reaction, depends on both the enthalpy and entropy changes of a reaction, and they sometimes even refer to reactions as "energy driven" or "entropy driven" depending on whether ΔH or the $T\Delta S$ term dominates. This is technically correct, but misleading because it disguises the important fact that ΔS_{total}, which this equation expresses in an indirect way, is the *only* criterion of spontaneous change.

Helmholtz Energy is also a "Free Energy"

We will deal only with the Gibbs energy in this course. The Helmholtz free energy is of interest mainly to chemical engineers (whose industrial-scale processes are often confined to tanks and reactors of fixed volume) and some geochemists whose interest is centered on the chemistry that occurs deep within the earth's surface.

Gibbs Energy and Chemical Change

Remember that ΔG is meaningful only for changes in which the *temperature and pressure remain constant*. These are the conditions under which most reactions are carried out in the laboratory; the system is usually open to the atmosphere (constant pressure) and we begin and end the process at room temperature (after any heat we have added or which is liberated by the reaction has dissipated.) The importance of the Gibbs function can hardly be over-stated: it serves as the single master variable that determines whether a given chemical change is thermodynamically possible. Thus if the free energy of the reactants is greater than that of the products, the entropy of the world will increase when the reaction takes place as written, and so the reaction will tend to take place spontaneously. Conversely, if the free energy of the products exceeds that of the reactants, then the reaction will not take place in the direction written, but it will tend to proceed in the reverse direction.

> ΔG serves as the single master variable that determines whether a given chemical change is thermodynamically possible. Moreover, it determines the direction and extent of chemical change.

In a spontaneous change, Gibbs energy always decreases and never increases. This of course reflects the fact that the entropy of the world behaves in the exact opposite way (owing to the negative sign in the $T\Delta S$ term).

$$\mathrm{H_2O\,(l)} \rightarrow \mathrm{H_2O\,(s)} \tag{15.4.6}$$

water below its freezing point undergoes a decrease in its entropy, but the heat released into the surroundings more than compensates for this, so the entropy of the world increases, the free energy of the H_2O diminishes, and the process proceeds spontaneously.

> In a spontaneous change, Gibbs energy **always** decreases and never increases.

An important consequence of the one-way downward path of the free energy is that once it reaches its minimum possible value, all net change comes to a halt. This, of course, represents the state of chemical equilibrium. These relations are nicely summarized as follows:

- $\Delta G < 0$: reaction can spontaneously proceed to the right:

$$A \rightarrow B$$

- $\Delta G > 0$: reaction can spontaneously proceed to the left:

$$A \leftarrow B$$

- $\Delta G = 0$: the reaction is at **equilibrium** and both $[A]$ and $[B]$ will not change:

$$A \rightleftharpoons B.$$

No need to find the value of ΔG for a Specific Reaction!

This might seem strange, given the key importance ΔG in determining whether or not a reaction will take place in a given direction. It turns out, however, that it is almost never necessary to explicitly evaluate ΔG. As we will show in the lesson that follows this one, it is far more convenient to work with the equilibrium constant of a reaction, within which ΔG is "hidden". This is just as well, because for most reactions (those that take place in solutions or gas mixtures) the value of ΔG depends on the **proportions** of the various reaction components in the mixture; it is not a simple sum of the "products minus reactants" type, as is the case with ΔH.

Recalling the condition for spontaneous change

$$\Delta G = \Delta H – T\Delta S < 0 \tag{15.4.7}$$

it is apparent that the temperature dependence of ΔG depends almost entirely on the entropy change associated with the process. (We say "almost" because the values of ΔH and ΔS are themselves slightly temperature dependent; both gradually increase with temperature). In particular, notice that in the above equation **the sign of the entropy change determines whether the reaction becomes more or less spontaneous as the temperature is raised.** For any given reaction, the sign of ΔH can also be positive or negative. This means that there are four possibilities for the influence that temperature can have on the spontaneity of a process:

Case 1: ΔH < 0 and ΔS > 0

Both enthalpic ΔH and entropic $-T\Delta S$ terms will be negative, so ΔG will be negative regardless of the temperature. An exothermic reaction whose entropy increases will be spontaneous at all temperatures.

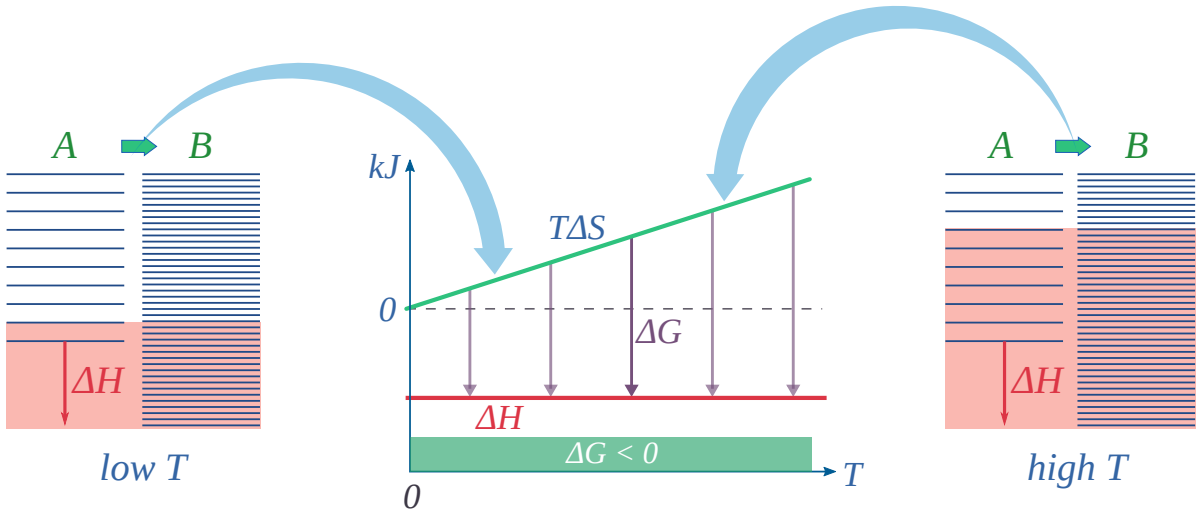


Figure 15.4.1: A hypothetical A ⟶ B reaction that is favored enthalpically and entropically (spontaneous at all temperatures). (CC BY 4.0; Ümit Kaya via LibreTexts)

Case 2: ΔH < 0 and ΔS < 0

If the reaction is sufficiently exothermic it can force ΔG negative only at temperatures below which $|T\Delta S| < |\Delta H|$. This means that there is a temperature $T = \Delta H/\Delta S$ at which the reaction is at equilibrium; the reaction will only proceed spontaneously below this temperature. The freezing of a liquid or the condensation of a gas are the most common examples of this condition.

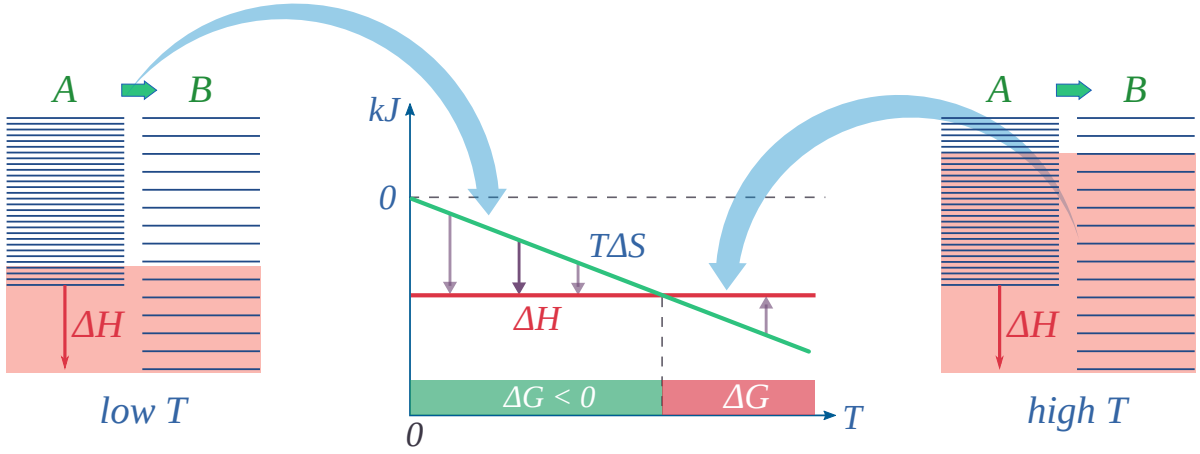


Figure 15.4.2: A hypothetical A ⟶ B reaction that is favored enthalpically and is not favored entropically (spontaneous at low temperatures). (CC BY 4.0; Ümit Kaya via LibreTexts)

Case 3: ΔH > 0 and ΔS > 0

This is the reverse of the previous case; the entropy increase must overcome the handicap of an endothermic process so that $T\Delta S > \Delta H$. Since the effect of the temperature is to "magnify" the influence of a positive ΔS, the process will be spontaneous at temperatures above $T = \Delta H/\Delta S$. (Think of melting and boiling.)

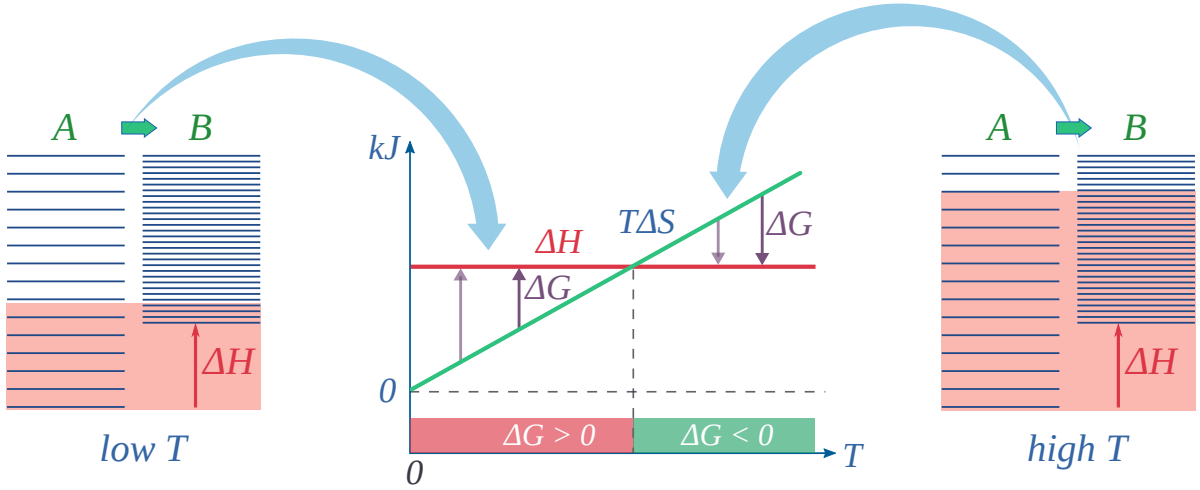


Figure 15.4.3: A hypothetical A ⟶ B reaction that is not favored enthalpically and is favored entropically (spontaneous at high temperatures). (CC BY 4.0; Ümit Kaya via LibreTexts)

Case 4: $\Delta H > 0$ and $\Delta S < 0$

With both ΔH and ΔS working against it, this kind of process will not proceed spontaneously at any temperature. Substance A always has a greater number of accessible energy states, and is therefore always the preferred form.

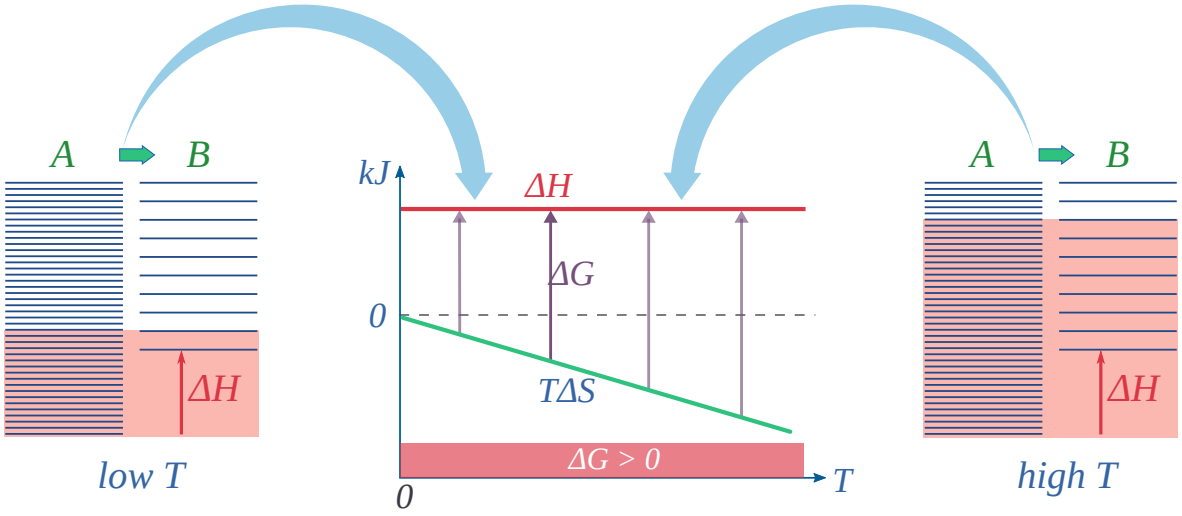


Figure 15.4.4: A hypothetical $\mathrm{A} \longrightarrow \mathrm{B}$ reaction that is not favored enthalpically and is not is favored entropically (not spontaneous at temperatures). (CC BY 4.0; Ümit Kaya via LibreTexts)

The plots above are the important ones; do not try to memorize them, but make sure you understand and can explain or reproduce them for a given set of ΔH and ΔS.

- Their most important differentiating features are the position of the ΔH line (above or below the is $T\Delta S$ line), and the slope of the latter, which of course depends on the sign of ΔS.
- The reaction A → B will occur spontaneously only when ΔG is negative (blue arrows pointing down.)
- $T\Delta S$ plots are not quite straight lines as shown here. Similarly, the lines representing ΔH are even more curved.

The other two plots on each diagram are only for the chemistry-committed.

- Each pair of energy-level diagrams depicts the relative spacing of the microscopic energy levels in the reactants and products as reflected by the value of $\Delta S°$. (The greater the entropy, the more closely-spaced are the quantized microstates.)
- The red shading indicates the range of energy levels that are accessible to the system at each temperature. The spontaneous direction of the reaction will always be in the direction in which the red shading overlaps the greater number of energy levels, resulting in the maximum dispersal of thermal energy.
- Note that the vertical offsets correspond to $\Delta H°$ for the reaction.
- Never forget that it is the ability of thermal energy to spread into as many of these states as possible that determines the tendency of the process to take place. None of this is to scale, of course!

The Standard Gibbs Energy

You have already been introduced to the terms such as ΔU^o and ΔH^o in which the ^{o} sign indicates that all components (reactants and products) are in their *standard states*. This concept of standard states is especially important in the case of the free energy, so let's take a few moments to review it. More exact definitions of the conventional standard states can be found in most physical chemistry textbooks. In specialized fields such as biochemistry and oceanography, alternative definitions may apply. For example, the "standard pH" of zero (corresponding to $[H^+] = 1\ M$) is impractical in biochemistry, so pH = 7 is commonly employed. For most practical purposes, the following definitions are good enough:

- **gases:** 1 atmosphere partial pressure
- **pure liquids:** the liquid under a total (hydrostatic) pressure of 1 atm.
- **solutes:** an effective concentration of 1 mol L^{-1} (1 mol dm^{-3}). ("Effective" concentrations approach real concentrations as the

latter approach zero; for practical purposes, these can be considered identical at real concentrations smaller than about 10^{-4} molar.)
- **solids:** the pure solid under 1 atm pressure

Reminder on Standard States

- There is actually no "standard temperature", but because most thermodynamics tables list values for 298.15 K (25° C), this temperature is usually implied.
- These same definitions apply to standard enthalpies and internal energies.
- Do not confuse these thermodynamic standard states with the "standard temperature and pressure" (STP) widely employed in gas law calculations.

To make use of Gibbs energies to predict chemical changes, we need to know the free energies of the individual components of the reaction. For this purpose we can combine the standard enthalpy of formation and the standard entropy of a substance to get its **standard free energy of formation**

$$\Delta G_f^o = \Delta H_f^o – T\Delta S_f^o \tag{15.4.8}$$

Recall that the symbol ° refers to the *standard state* of a substance measured under the conditions of 1 atm pressure or an effective concentration of 1 mol L^{-1} and a temperature of 298 K. Then determine the standard Gibbs energy of the reaction according to

$$\Delta G^o = \sum \Delta G_f^o \text{ (products)} – \sum \Delta G_f^o \text{ (reactants)} \tag{15.4.9}$$

As with standard heats of formation, the standard free energy of a substance represents the free energy change associated with the formation of the substance from the elements in their most stable forms as they exist under the standard conditions of 1 atm pressure and 298 K. Standard Gibbs free energies of formation are normally found directly from tables. Once the values for all the reactants and products are known, the standard Gibbs energy change for the reaction is found by Equation 15.4.8. Most tables of thermodynamic values list ΔG_f^o values for common substances (e.g., Table T2), which can, of course, always be found from values of ΔH_f° and ΔS_f°.

Example 15.4.1

Find the standard Gibbs energy change for the reaction

$$\mathrm{CaCO_3\,(s)} \rightarrow \mathrm{CaO\,(s) + CO_2\,(g)}$$

The ΔG_f^o ° values for the three components of this reaction system are $\mathrm{CaCO_3\,(s)}$: –1128 kJ mol^{-1}, CaO*(s)*: –603.5 kJ mol^{-1}, CO_2*(g)*: –137.2 kJ mol^{-1}.

Solution

Substituting into Equation 15.4.8, we have

$$\Delta G^o = (–603.5–137.2)–(–1128)kJ\,mol^{-1} = +130.9\,kJ\,mol^{-1}$$

This indicates that the process is not spontaneous under standard conditions (i.e., solid calcium carbone will not form solid calcium oxide and CO_2 at 1 atm partial pressure at 25° C).

Comment: This reaction is carried out on a huge scale to manufacture cement, so it is obvious that the process can be spontaneous under different conditions.

The practical importance of the Gibbs energy is that it allows us to make predictions based on the properties (ΔG° values) of the reactants and products themselves, eliminating the need to experiment. But bear in mind that while thermodynamics always correctly predicts whether a given process *can* take place (is spontaneous in the thermodynamic sense), it is unable to tell us if it

will take place at an observable rate.

When thermodynamics says "no", it means exactly that. When it says "yes", it means "maybe".

Example 15.4.2

The reaction

$$\tfrac{1}{2}\,\mathrm{O_2\,(g)} + \mathrm{H_2\,(g)} \longrightarrow \mathrm{H_2O\,(l)}$$

is used in fuel cells to produce an electrical current. The reaction can also be carried out by direct combustion.

Thermodynamic data: molar entropies in J mol^{-1} K^{-1}: O_2(g) 205.0; H_2(g)130.6; H_2O(l) 70.0; H_2O(l) $\Delta H°_f$ = –285.9 kJ mol^{-1}.

Use this information to find

a. The amount of heat released when the reaction takes place by direct combustion;
b. The amount of electrical work the same reaction can perform when carried out in a fuel cell at 298 K under reversible conditions;
c. The amount of heat released under the same conditions.

Solution

First, we need to find ΔH^o and ΔS^o for the process. Recalling that the standard enthalpy of formation of the elements is zero,

$$\begin{aligned}\Delta H^o &= \Delta H^p_f(\text{products}) – \Delta H^\circ_f(\text{reactants}) \\ &= –285.9\,kJ\,mol^{–1} – 0 \\ &= –285.9\,kJ\,mol^{–1}\end{aligned}.$$

Similarly,

$$\begin{aligned}\Delta S^o &= S^o_f(\text{products}) – S^o_f(\text{reactants}) \\ &= (70.0) – (½ \times 205.0 + 130.6) \\ &= –163\,J\,K^{–1}mol^{–1}\end{aligned}$$

a. When the hydrogen and oxygen are combined directly, the heat released will be $\Delta H^o = –285.9\ kJ\ mol^{–1}$.
b. The maximum electrical work the fuel cell can perform is given by

$$\begin{aligned}\Delta G^o &= \Delta H^o – T\Delta S^o \\ &= –285.9\,kJ\,mol^{–1} – (298\,K)(–163\,JK^{–1}mol^{–1}). \\ &= –237.2\,kJ\,mol^{–1}.\end{aligned}$$

c. The heat released in the fuel cell reaction is the difference between the enthalpy change (the total energy available) and the reversible work that was expended:

$$\begin{aligned}\Delta H^o – \Delta G^o &= T\Delta S^o \\ &= (298\,K)(–163\,JK^{–1}mol^{–1}) \\ &= –48,800\,J\,mol^{–1} \\ &= –48.8\,kJ\,mol^{–1}.\end{aligned}$$

The foregoing example illustrates an important advantage of fuel cells. Although direct combustion of a mole of hydrogen gas yields more energy than is produced by the same net reaction within the fuel cell, the latter, in the form of electrical energy, can be utilized at nearly 100-percent energy efficiency by a motor or some other electrical device. If the thermal energy released by direct combustion were supplied to a heat engine, second-law considerations would require that at least half of this energy be "wasted" to the surroundings.

ΔG vs. ΔG°: what's the difference?

ΔG° refer to *single, specific chemical changes* in which all components (reactants and products) are in their *standard states*.

The ΔG_f^o of a substance, like ΔH_f^o, refers to the reaction in which that substance is formed from the elements as they exist in their most stable forms at 1 atm pressure and (usually) 298 K. Both of these terms are by definition zero for the elements in their standard states. There are only a few common cases in which this might create some ambiguity:

Table 15.4.1*: Standard Gibbs Energies of Select Substances*

Stable Form	ΔG_f^o (kJ mol^{-1})	Unstable Form	ΔG_f^o (kJ mol^{-1})
$O_2\,(g)$	0	$O_3\,(g)$	163.2
C(graphite)	0	C(diamond)	2.9
S(rhombic)	0	S(monoclinic)	0.1
P(white)	0	$P_4\,(g)$	24.4

Ions in aqueous solution are a special case; their standard free energies are relative to the hydrated hydrogen ion $H^+\,(aq)$ which is assigned $\Delta G_f^o = 0$.

ΔG is very different from ΔG°. The distinction is nicely illustrated in Figure 15.4.5 in which ΔG is plotted on a vertical axis for two hypothetical reactions having opposite signs of ΔG°. The horizontal axis schematically expresses the relative concentrations of reactants and products at any point of the process. Note that the origin corresponds to the composition at which half of the reactants have been converted into products.

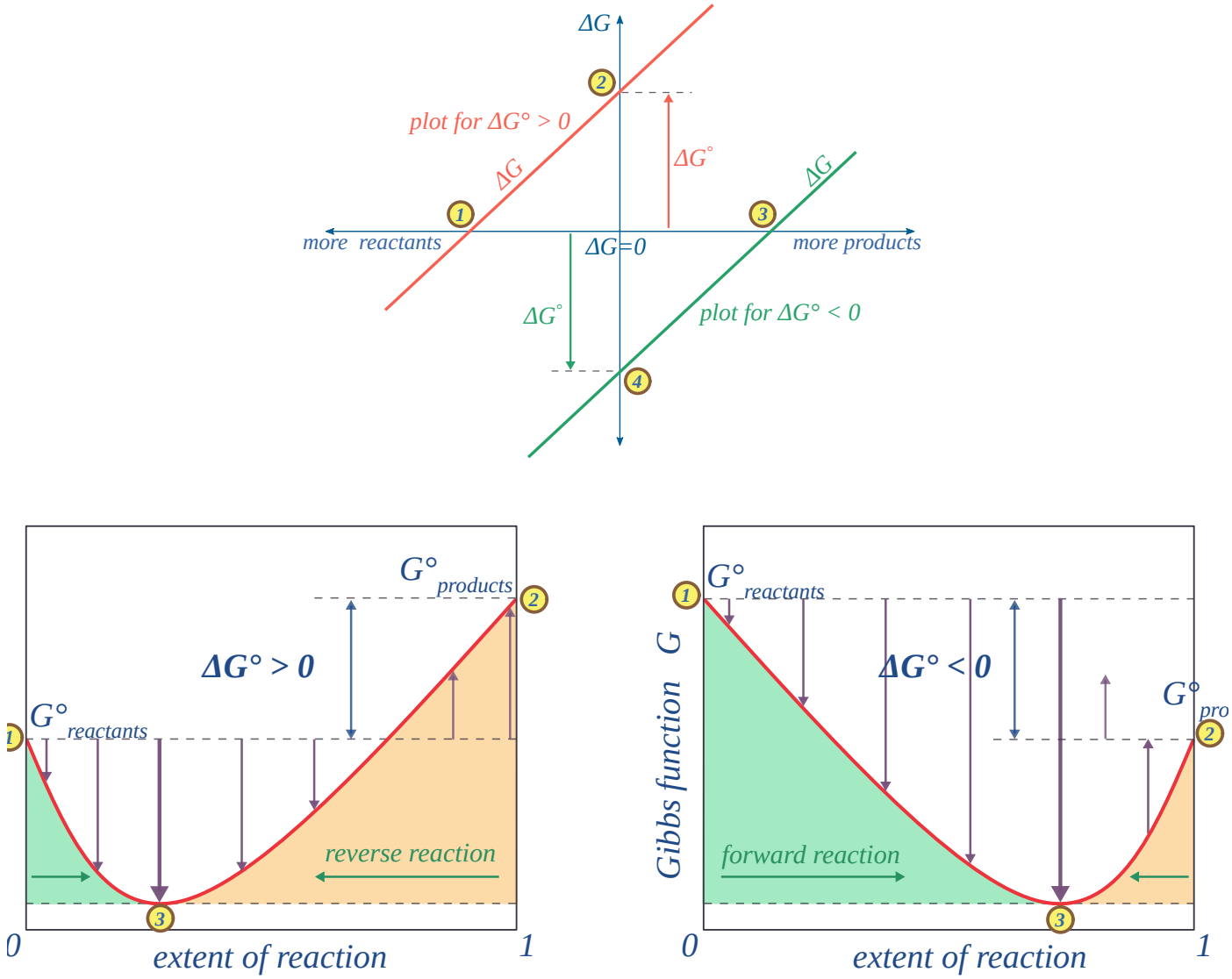


Figure 15.4.5: (CC BY 4.0; Ümit Kaya via LibreTexts)

Take careful note of the following:

- for the $\Delta G° > 0$ reaction. Notice that there are an infinite number of these values, depending on the progress of the reaction. In contrast there is only a single value (2) of $\Delta G°$, corresponding to the composition at which $\Delta G = 0$ ((1)).
- At this point, some products have been formed, but the composition is still dominated by reactants.
- If we begin at a composition to the left of (1), ΔG will be negative and the composition will move to the right. Similarly if we begin with a composition to the right of (1), ΔG will be positive and the composition will move to the left.
- The plot on the right is for the $\Delta G° < 0$ reaction, for which $\Delta G°$ is shown at (4). At its equilibrium point (3), there are more products than reactants. If we start at a composition to the right of (3), the composition will tend to move to the left. If the initial composition is to the left of (3), the reaction will tend to proceed to the right.
- What would happen if $\Delta G°$ were 0? The equilibrium point of such a reaction would be at the origin, corresponding to half the reactants being converted to products.

The important principle you should understand from this is that a negative $\Delta G°$ does not mean that the reactants will be

completely transformed into products. By the same token, a positive $\Delta G°$ does not mean that no products are formed at all.

It should now be clear from the discussion above that a given reaction carried out under standard conditions is characterized by a *single value* of $\Delta G°$.

The reason for the Gibbs energy minimum at equilibrium relates to the increase in entropy when products and reactants coexist in the same phase. As seen in the plot, even a minute amount of "contamination" of products by reactants reduces the free energy below that of the pure products. In contrast, composition of a chemical reaction system undergoes continual change until the equilibrium state is reached. So the a single reaction can have an infinite number of ΔG values, reflecting the infinite possible compositions between the extremes of pure reactants (zero extent of reaction) and pure products (unity extent of reaction).

In the example of a reaction A → B, depicted in the above diagram, the standard free energy of the products (2) is smaller than that of the reactants (1), so the reaction will take place spontaneously. This does not mean that each mole of pure A will be converted into one mole of pure B. For reactions in which products and reactants occupy a single phase (gas or solution), the meaning of "spontaneous" is that the equilibrium composition (3) will correspond to an extent of reaction greater than 0.5 but smaller than unity.Note, however, that for $\Delta G°$ values in excess of about ±50 kJ mol^{-1}, the equilibrium composition will be negligibly different from zero or unity extent-of-reaction. The physical meaning of ΔG is that it tells us how far the free energy of the system has changed from $G°$ of the pure reactants (1). As the reaction proceeds to the right, the composition changes, and ΔG begins to fall. When the composition reaches (3), ΔG reaches its minimum value and further reaction would cause it to rise. But because free energy can only decrease but never increase, this does not happen. The composition of the system remains permanently at its equilibrium value.

A *G vs.* extent-of-reaction diagram for a non-spontaneous reaction can be interpreted in a similar way; the equilibrium composition (3) will correspond to an extent of reaction greater than zero but less than 0.5. In this case, the minimum at (3) reflects the increase in entropy when the reactants are "contaminated" by a small quantity of products.

If all this detail about ΔG seems a bit overwhelming, do not worry: it all gets hidden in the equilibrium constant and reaction quotient that we discuss in the next lesson!

Interpretation of Standard Gibbs energy changes

Although it is ΔG rather than ΔG^o that serves as a criterion for spontaneous change at constant temperature and pressure, ΔG^o values are so readily available that they are often used to get a rough idea of whether a given chemical change is possible. This is practical to do in some cases, but not in others:

Example

It generally works for reactions such as

$$4\,\mathrm{NH_3\,(g)} + 5\,\mathrm{O_2\,(g)} \longrightarrow 4\,\mathrm{NO\,(g)} + 6\,\mathrm{H_2O\,(g)}$$

with $\Delta G^o = -1,010\ kJ$.

(industrially important for the manufacture of nitric acid) because ΔG^o is so negative that the reaction will be spontaneous and virtually complete under just about any reasonable set of conditions.

Example

The following reaction expresses the fact that the water molecule is thermodynamically stable:

$$2\,\mathrm{H_2\,(g)} + \tfrac{1}{2}\,\mathrm{O_2(g)} \longrightarrow \mathrm{H_2O\,(l)}$$

with $\Delta G^o = -237.2\ kJ$.

Note that this refers to *liquid* water (the standard state of H_2O at 25°). If you think about it, a negative standard Gibbs energy of formation (of which this is an example) can in fact be considered a definition of molecular stability.

Example

Similarly, dissociation of dihydrogen into its atoms is highly unlikely under standard conditions:

$$\mathrm{H_2O\,(g)} \longrightarrow \mathrm{2\,H\,(g) + O\,(g)}$$

with $\Delta G^o = +406.6\ kJ$.

Again, an analogous situation would apply to any stable molecule.

Example

Now consider the dissociation of dinitrogen tetroxide

$$\mathrm{N_2O_4\,(g)} \longrightarrow \mathrm{2\,NO_2\,(g)}$$

with $\Delta G^o = +2.8kJ$.

in which the positive value of $\Delta G°$ tells us that N_2O_4 at 1 atm pressure will not change into two moles of NO_2 at the same pressure, but owing to the small absolute value of $\Delta G°$, we can expect the spontaneity of the process to be quite sensitive to both the temperature (as shown in the table below) and to the pressure in exactly the way the Le Chatelier principle predicts.

Example

For reactions involving dissolved ions, one has to be quite careful. Thus for the dissociation of the weak hydrofluoric acid

$$\mathrm{HF\,(aq)} \longrightarrow \mathrm{H^+\,(aq) + F^-\,(aq)}$$

with $\Delta G^o = -317\ kJ$.

it is clear that a 1 mol/L solution of HF will not dissociate into 1M ions, but this fact is not very useful because if the HF is added to water, the initial concentration of the fluoride ion will be zero (and that of H^+ very close to zero), and the Le Chatelier principle again predicts that *some* dissociation will be spontaneous.

Example

It is common knowledge that dissociation of water into hydrogen- and hydroxyl ions occurs only very sparingly:

$$\mathrm{H_2O\,(l)} \longrightarrow \mathrm{H^+\,(aq) + OH^-\,(aq)}$$

with $\Delta G^o = 79.9\ kJ$.

which correctly predicts that the water will not form 1M (effective concentration) of the ions, but this is hardly news if you already know that the product of these ion concentrations can never exceed 10^{-14} at 298K.

Example

Finally, consider this most familiar of all phase change processes, the vaporization of liquid water:

$$H_2O\,(l) \longrightarrow H_2O\,(g)$$

with $\Delta G^o = 8.58\ kJ$.

Conversion of liquid water to its vapor at 1 atm partial pressure does not take place at 25° C, at which temperature the equilibrium partial pressure of the vapor (the "vapor pressure") is only 0.031 atm (23.8 torr.) Gaseous H_2O at a pressure of 1 atm can only exist at 100° C. Of course, water left in an open container at room temperature will spontaneously evaporate if the partial pressure of water vapor in the air is less than 0.031 atm, corresponding to a relative humidity of under 100%

Finding the Equilibrium Temperature

A reaction is in its equilibrium state when

$$\Delta G = \Delta H – T\Delta S = 0 \tag{15.4.10}$$

The temperature at which this occurs is given by

$$T = \frac{\Delta H}{\Delta S} \tag{15.4.11}$$

If we approximate ΔH by ΔH^o and ΔS by ΔS^o, so Equation 15.4.10 would be

$$\Delta G \approx \Delta H^o – T\Delta S^o = 0 \tag{15.4.12}$$

We can then estimate the normal boiling point of a liquid. From the following thermodynamic data for water:

$H_2O(l)$	→	$H_2O(g)$	
–285.9		–241.8	$\Delta H°$ = +44.1 kJ
69.9		188.7	$\Delta S°$ = +118.7 J K^{-1}

Caution!

Because ΔH° values are normally expressed in kilojoules while ΔS° is given in joules, a very common student error is to overlook the need to express both in the same units.

We find that liquid water is in equilibrium with water vapor at a partial pressure of 1 atm when the temperature is

$$T = \frac{44,100\ J}{118.7\ J\ K^{-1}} = 371.5\ K \tag{15.4.13}$$

But "*the normal boiling point of water is 373 K*", you say? Very true. The reason we are off here is that both $\Delta H°$ and $\Delta S°$ have their own temperature dependencies; we are using the "standard" 25° values without correcting them to 100° C. Nevertheless, if you think about it, the fact that we can estimate the boiling point of a liquid from a table of thermodynamic data should be rather impressive! Of course, the farther one gets from 298 K, the more unreliable will be the result. Thus for the dissociation of dihydrogen into its atoms,

$H_2(g)$	→	2 H(g)	
0		2 x (218.0)	$\Delta H°$ = +436.0 kJ
130.6		2 x (114.6)	$\Delta S°$ = +98.6 J K^{-1}

$$T = \frac{436,000\ J}{98.6\ JK^{-1}} = 4,420\ K \tag{15.4.14}$$

All one can say here is that H_2 will break down at something over 3000 K or so. (You may already know that *all* molecules will dissociate into their atoms at high temperatures.) We tend to think of high temperatures as somehow "forcing" molecules to dissociate into their atoms, but this is wrong. In order to get the H–H bond to vibrate so violently through purely thermal excitation that the atoms would fly apart, a temperature more like 30,000 K would be required. The proper interpretation is at the temperature corresponding to $\Delta H/T\Delta S$, the molecule *spontaneously* absorbs energy from the surroundings sufficient to overcome the H-H bond strength.

Predicting the Effects of Temperature

The $T\Delta S$ term interacts with the ΔH term in ΔG to determine whether the reaction can take place at a given temperature. This can be more clearly understood by examining plots of $T\Delta S^o$ and ΔH^o as functions of the temperature for some actual reactions. Of course these parameters refer to standard states that generally do not correspond to the temperatures, pressures, or concentrations that might be of interest in an actual case. Nevertheless, these quantities are easily found and they can usefully predict the way that temperature affects these systems.

Case 1: Exothermic reaction, ΔS° > 0

$$\mathrm{C(graphite) + O_2\,(g) \longrightarrow CO_2\,(g)}$$

- $\Delta H^o = -393\ kJ$
- $\Delta G^o = -394\ kJ$ at $298\ K$

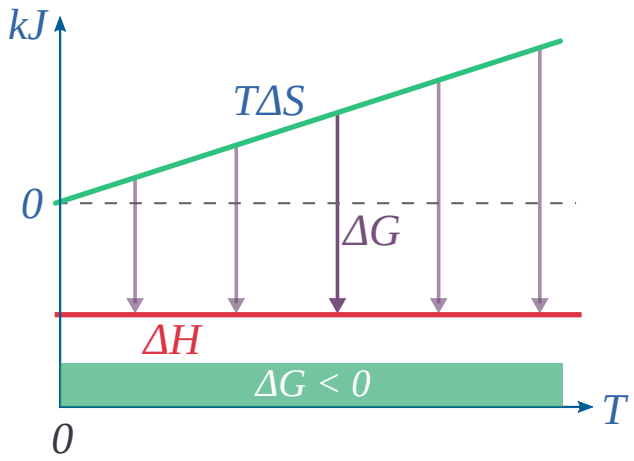


Figure 15.4.6: (CC BY 4.0; Ümit Kaya via LibreTexts)

This *combustion reaction*, like most such reactions, is **spontaneous at all temperatures**. The positive entropy change is due mainly to the greater mass of CO_2 molecules compared to those of O_2.

Case 2: Exothermic reaction, ΔS° < 0

$$3\,H_2 + N_2 \longrightarrow 2\,NH_3\,(g)$$

- $\Delta H^o = -46.2\ kJ$
- $\Delta G^o = -16.4\ kJ$ at $298\ K$

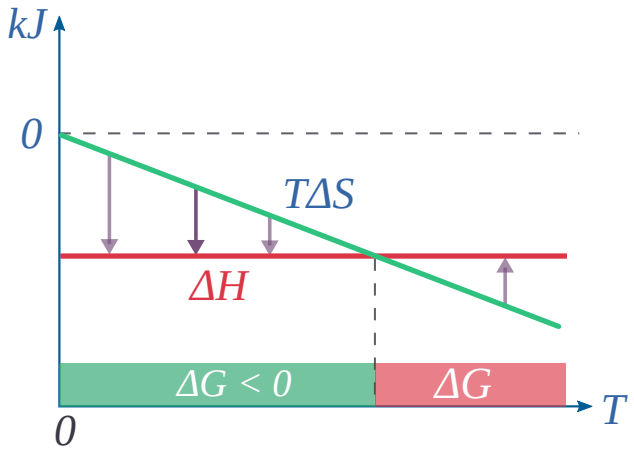


Figure 15.4.7: (CC BY 4.0; Ümit Kaya via LibreTexts)

The decrease in moles of gas in the **Haber ammonia synthesis** drives the entropy change negative, making the reaction **spontaneous only at low temperatures**. Thus higher T, which speeds up the reaction, also reduces its extent.

Case 3: Endothermic reaction, ΔS° > 0

$$N_2O_4\,(g) \longrightarrow 2\,NO_2\,(g)$$

- $\Delta H^o = 55.3\ kJ$
- $\Delta G^o = +2.8\ kJ$ at $298\ K$

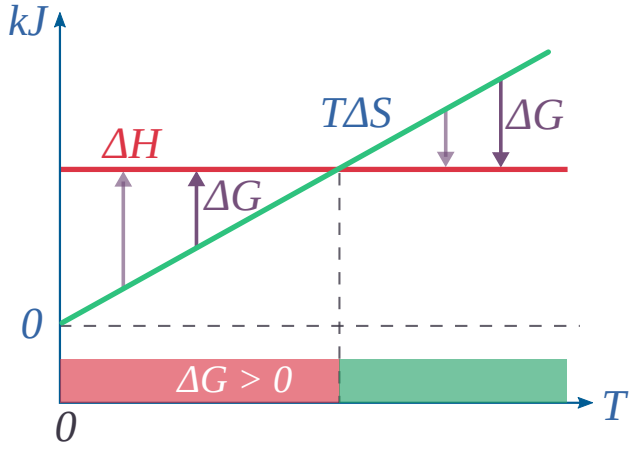


Figure 15.4.8: (CC BY 4.0; Ümit Kaya via LibreTexts)

Dissociation reactions are typically endothermic with positive entropy change, and are therefore **spontaneous at high temperatures**. Ultimately, all molecules decompose to their atoms at sufficiently high temperatures.

Case 4: Endothermic Reaction, ΔS° < 0

$$\tfrac{1}{2}\,\mathrm{N_2\,(g)} + \mathrm{O_2\ (g)} \longrightarrow \mathrm{NO_2\,(g)}$$

- $\Delta H^\circ = 33.2\ kJ$
- $\Delta S^o = -249\ J\ K^{-1}$
- $\Delta G^o = +51.3\ kJ$ at $298\ K$

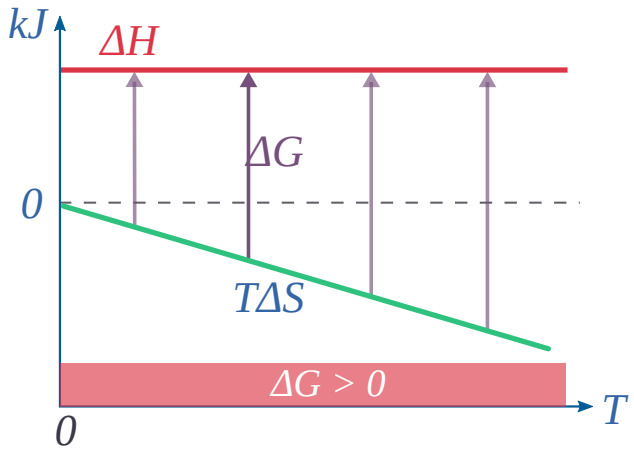


Figure 15.4.9: (CC BY 4.0; Ümit Kaya via LibreTexts)

This reaction is **not spontaneous at any temperature**, meaning that *its reverse is always spontaneous*. But because the reverse reaction is kinetically inhibited, NO_2 can exist indefinitely at ordinary temperatures even though it is thermodynamically unstable.

Concluding remarks on Gibbs Energy

The appellation “free energy” for G has led to so much confusion that many scientists now refer to it simply as the *Gibbs energy*. The “free” part of the older name reflects the steam-engine origins of thermodynamics with its interest in converting heat into work: ΔG is the maximum amount of energy, which can be “freed” from the system to perform useful work. By "useful", we mean work other than that which is associated with the expansion of the system. This is most commonly in the form of electrical work (moving electric charge through a potential difference), but other forms of work (osmotic work, increase in surface area) are also possible.

A much more serious difficulty with the Gibbs function, particularly in the context of chemistry, is that although G has the *units* of energy (joules, or in its intensive form, J mol^{-1}), it lacks one of the most important attributes of energy in that it is not *conserved*. Thus, although the free energy always falls when a gas expands or a chemical reaction takes place spontaneously, there need be no compensating *increase* in energy anywhere else. Referring to G as an energy also reinforces the false but widespread notion that a fall in energy must accompany any change. But if we accept that energy is conserved, it is apparent that the only necessary condition for change (whether the dropping of a weight, expansion of a gas, or a chemical reaction) is the *redistribution* of energy. The quantity \(–ΔG\) associated with a process represents the quantity of energy that is “shared and spread”, which as we have already explained is the meaning of the increase in the entropy. The quotient $-\Delta G/T$ is in fact identical with ΔS_{total}, the entropy change of the world, whose increase is the primary criterion for any kind of change.

G differs from the thermodynamic quantities H and S in another significant way: it has no physical reality as a property of matter, whereas H and S can be related to the quantity and distribution of energy in a collection of molecules. The free energy is simply a useful construct that serves as a criterion for change and makes calculations easier.

What Gibbs Energy is not...

- Gibbs Energy is not free energy
- Gibbs Energy is not energy
- Gibbs Energy is not even "real"

This page titled 15.4: Free Energy and the Gibbs Function is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.5: Thermodynamics of Mixing and Dilution

The free energy of a pure liquid or solid at 1 atm pressure is just its molar free energy of formation ΔG multiplied by the number of moles present. For gases and substances in solution, we have to take into account the *concentration* (which, in the case of gases, is normally expressed in terms of the pressure). We know that the lower the concentration, the greater the entropy, and thus the smaller the free energy. The following excerpt from this lesson serves as the starting point for the rest of the present lesson.

Entropy Depends on Concentration

As a substance becomes more dispersed in space, the thermal energy it carries is also spread over a larger volume, leading to an increase in its entropy. Because entropy, like energy, is an extensive property, a dilute solution of a given substance may well possess a smaller entropy than the same volume of a more concentrated solution, but the entropy *per mole* of solute (the *molar* entropy) will of course always increase as the solution becomes more dilute.

For gaseous substances, the volume and pressure are respectively direct and inverse measures of concentration. For an ideal gas that expands at a constant temperature (meaning that it absorbs heat from the surroundings to compensate for the work it does during the expansion), the increase in entropy is given by

$$\Delta S = R \ln \left(\frac{V_2}{V_1} \right) \tag{15.5.1}$$

(If the gas is allowed to cool during the expansion, the relation becomes more complicated and will best be discussed in a more advanced course.)

Because the pressure of a gas is inversely proportional to its volume, we can easily alter the above relation to express the entropy change associated with a change in the pressure of a perfect gas:

$$\Delta S = R \ln \left(\frac{P_1}{P_2} \right) \tag{15.5.2}$$

Expressing the entropy change directly in concentrations c, we have the similar relation

$$\Delta S = R \ln \left(\frac{C_1}{C_2} \right) \tag{15.5.3}$$

Although these equations strictly apply only to perfect gases and cannot be used at all for liquids and solids, it turns out that in a dilute solution, the solute can often be treated as a gas dispersed in the volume of the solution, so the last equation can actually give a fairly accurate value for the entropy of dilution of a solution. We will see later that this has important consequences in determining the equilibrium concentrations in a homogeneous reaction mixture.

The Gibbs energy of a Gas: standard states

The pressure of a perfect gas does not affect its enthalpy, but it does affect the entropy (box at left), and thus, through the $-T\Delta S$ term, the free energy. When the pressure of such a gas changes from P_1 to P_2, the Gibbs energy change is

$$\Delta G = \Delta H - T\Delta S = 0 - RT \ln \left(\frac{P_1}{P_2} \right) \tag{15.5.4}$$

How can we evaluate the free energy of a specific sample of a gas at some arbitrary pressure? First, recall that the standard molar free energy $G°$ that you would look up in a table refers to a pressure of 1 atm. The free energy per mole of our sample is just the sum of this value and any change in free energy that would occur if the pressure were changed from 1 atm to the pressure of interest

$$G = G^o + RT \ln \left(\frac{P_1}{1\ atm} \right) \tag{15.5.5}$$

which we normally write in abbreviated form

$$G = G^o + RT \ln P \tag{15.5.6}$$

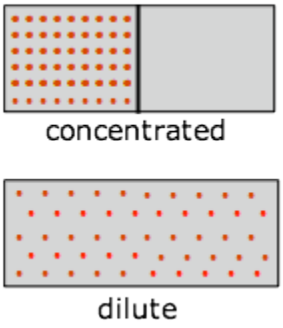
Escaping Tendency

The term *escaping tendency* is not commonly used in traditional thermodynamics because it is essentially synonymous with the free energy, but it is worth knowing because it helps us appreciate the physical significance of free energy in certain contexts. The higher the pressure of a gas, the greater will be the tendency of its molecules to leave the confines of the container; we will call this the *escaping tendency*. The above equation tells us that the pressure of a gas is a directly observable measure of its free energy (*G*, not *G*°!). Combining these two ideas, we can say that the free energy of a gas is also a measure of its escaping tendency.

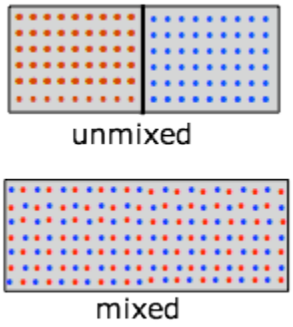
From Gases to Solutions: Mixing and Dilution

All substances, given the opportunity to form a homogeneous mixture with other substances, will tend to become more dilute. This can be rationalized simply from elementary statistics; there are more equally probable ways of arranging one hundred black marbles and one hundred white marbles, than two hundred marbles of a single color. For massive objects like marbles this has nothing to do with entropy, of course. However, when we are dealing with huge numbers of molecules capable of storing, exchanging and spreading thermal energy, mixing and expansion are definitely entropy-driven processes.

It can be argued, in fact, that mixing and expansion are really very similar; after all, when we mix two gases, each is expanding into the space formerly occupied exclusively by the other.



Suppose, for example, that we have a gas initially confined to one half of a box, and we then remove the barrier so that it can expand into the full volume of the container. We know that the entropy of the gas will increase as the thermal energy of its molecules spreads into the enlarged space; the actual increase, according to Equations 15.5.1 and 15.5.2 above, is $R \ln 2$. And from Equation 15.5.6, the change in G will be $-RT \ln 2$.



Now let us repeat the experiment, but starting this time with "red" molecules in one half of the container and "blue" ones in the right half. Because we now have two gases expanding into twice their initial volumes, the changes in *S* and *G* will be twice as great:

$$\Delta S = 2R \ln 2 \tag{15.5.7}$$

$$\Delta G = -2RT \ln 2 \tag{15.5.8}$$

However, notice that although each gas underwent an expansion, the overall process in this case is equivalent to *mixing*.

What is true for gaseous molecules can, in principle, apply also to solute molecules dissolved in a solvent. An important qualification here is that the solution must be an *ideal* one, meaning that the strength of interactions between all type of molecules (solutes A and B, and the solvent) must be the same. Remember that the enthalpy associated with the expansion of a perfect gas is by definition zero. In contrast, the ΔH_{mixing} of two liquids or of dissolving a solute in a solvent have finite values which may limit the miscibility of liquids or the solubility of a solute.

Given this proviso, we can define the Gibbs energy of dilution or mixing by substituting this equation into the definition of Δ*G*:

$$\Delta G_{dilution} = \Delta H_{dilution} - RT \ln \left(\frac{C_1}{C_2} \right) \tag{15.5.9}$$

If the substance in question forms an **ideal solution** with the other components, then $\Delta H_{diution}$ is zero, and we can write

$$\Delta G_{dilution} = RT \ln \left(\frac{C_2}{C_1} \right) \tag{15.5.10}$$

These relations tell us that the dilution of a substance from an initial concentration C_1 to a more dilute concentration C_2 is accompanied by a **decrease in the free energy**, and thus will occur spontaneously. By the same token, the spontaneous "un-dilution" of a solution will not occur (we do not expect the tea to diffuse back into the tea bag!) However, un-dilution can be forced to occur if some means can be found to supply to the system an amount of energy (in the form of work) equal to $\Delta G_{dilution}$.

An important practical example of this is the metabolic work performed by the kidneys in concentrating substances from the blood for excretion in the urine.

To find the Gibbs energy of a solute at some arbitrary concentration, we proceed in very much the same way as we did for a gas: we take the sum of the standard free energy plus any change in the free energy that would accompany a change in concentration from the standard state to the actual state of the solution. From Equation 15.5.10 it is easy to derive an expression analogous to Equation 15.5.6:

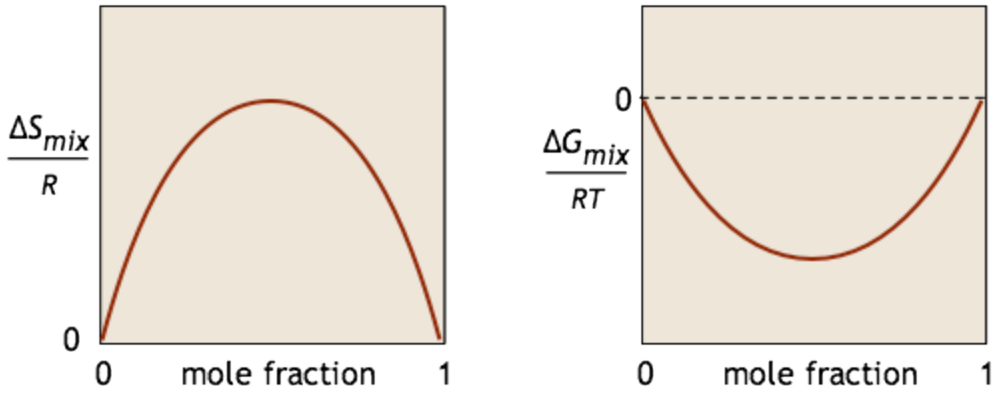
$$G = G^o + RT \ln C \tag{15.5.11}$$

which gives the free energy of a solute at some arbitrary concentration C in terms of its value G^o in its standard state.

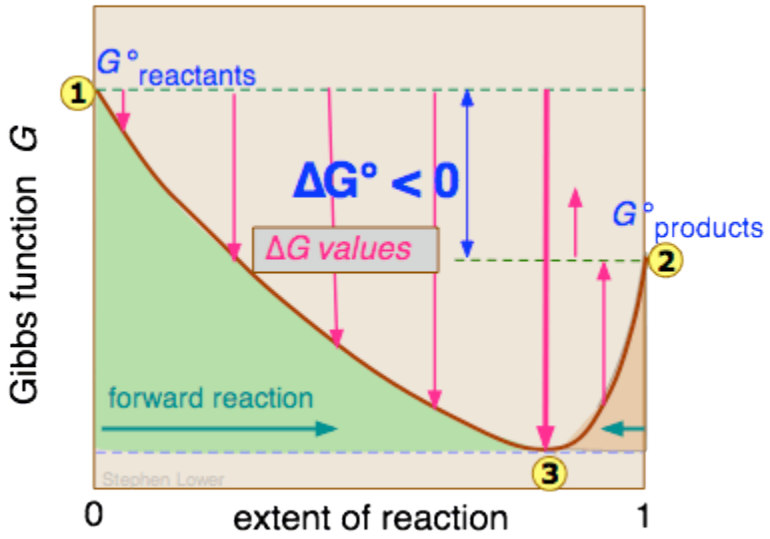
Although Equation 15.5.11 has the same simple form as Equation 15.5.6, its practical application is fraught with difficulties, the major one being that it does not usually give values of G that are consistent with experiment, especially for solutes that are ionic or are slightly soluble. This is because most solutions (especially those containing dissolved ions) are far from *ideal*; intermolecular interactions between solute molecules and between solute and solvent bring back the enthalpy term that we left out in deriving Equation 15.5.10. In addition, the structural organization of the solution becomes concentration dependent, so that the entropy depends on concentration in a more complicated way than is implied by the concentration analog of Equation 15.5.10 .

Chemical Reactions and Mixing

We characterize the tendency for a chemical reaction A → B to occur at constant temperature and pressure by the value of its standard Gibbs energy change Δ*G*°. If this quantity is negative, we know that the reaction will take place spontaneously. However, have you ever wondered why it is that substance A is not completely transformed into B if the latter is thermodynamically more stable? The answer is that if the reaction takes place in a single phase (gas or liquid), something else is going on: A and B are mixing together, and this process creates its own Gibbs energy change ΔG_{mixing}. For a simple binary mixture of A and B (without any reaction), the changes in S and G can be represented by these simple plots:



We will not try to prove it here, but it turns out that no matter how much lower the Gibbs energy of the products compared to that of the reactants, the free energy of the system can always be reduced even more if some of the reactants remain in the solution to contribute a ΔG_{mixing} term. This is the reason that a plot of G as a function of the composition of such a system has a minimum at some point short of complete conversion.



Diffusion refers to the transport of a substance across a concentration gradient. The direction is always toward the region of lower concentration. You should now see that from a thermodynamic standpoint, these processes are identical in that they both represent the spontaneous "escape" of molecules from a region of higher concentration (lower entropy, higher Gibbs energy) to a region of lower concentration.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

Video 15.5.1: Tea diffusing away from a tea bag.

Activity and Standard State of the Solute

Instead of complicating G° by trying to correct for all of these effects, chemists have chosen to retain its simple form by making a single small change in the form of 15.5.11:

$$G = G^o + RT \ln a \tag{15.5.12}$$

This equation is guaranteed to work, because *a*, the *activity* of the solute, is its *thermodynamically effective* concentration. The relation between the activity and the concentration is given by

$$a = \gamma c \tag{15.5.13}$$

where γ is the *activity coefficient*. As the solution becomes more dilute, the activity coefficient approaches unity:

$$\lim_{c \to 0} \gamma = 1 \tag{15.5.14}$$

The price we pay for this simplicity is that the relation between the concentration and the activity at higher concentrations can be quite complicated, and must be determined experimentally for every different solution.

The question of what standard state we choose for the solute (that is, at what concentration is *G*° defined, and in what units is it expressed?) is one that you will wish you had never asked. We might be tempted to use a concentration of 1 molar, but a solution this concentrated would be subject to all kinds of intermolecular interaction effects, and would not make a very practical standard state. These effects could be eliminated by going to the opposite extreme of an "infinitely dilute" solution, but by Equation 15.5.10 this would imply a free energy of minus infinity for the solute, which would be awkward. Chemists have therefore agreed to define the standard state of a solute as one in which the concentration is 1 molar, but all solute-solute interactions are magically switched off, so that γ is effectively unity. Since this is impossible, no solution corresponding to this standard state can actually exist, but this turns out to be only a small drawback, and seems to be the best compromise between convenience, utility, and reality.

This page titled 15.5: Thermodynamics of Mixing and Dilution is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.6: Free Energy and Equilibrium

Learning Objectives

You are expected to be able to define and explain the significance of terms identified in *italic type*.

- As a *homogeneous chemical reaction* proceeds, the Gibbs energies of the reactants become more negative and those of the products more positive as the composition of the system changes.
- The *total Gibbs energy of the system* (reactants + products) always becomes more negative as the reaction proceeds. Eventually it reaches a minimum value at a system composition that defines the *equilibrium composition* of the system, after which time no further net change will occur.
- The *equilibrium constant* for the reaction is determined the standard Gibbs energy change: $\Delta G° = -RT \ln K_p$
- The sign of the temperature dependence of the equilibrium constant is governed by the sign of $\Delta H°$. This is the basis of the *Le Chatelier Principle*.
- The Gibbs energies of solid and liquid components are constants that do not change with composition. Thus in *heterogeneous reactions* such as phase changes, the total Gibbs energy does not pass through a minimum and when the system is not at equilibrium only all-products or all-reactants will be stable.
- Two reactions are *coupled* when the product of one reaction is consumed in the other. If $\Delta G°$ for the first reaction is positive, the overall process can still be spontaneous if $\Delta G°$ for the second reaction is sufficiently negative— in which case the second reaction is said to "drive" the first reaction.

Under conditions of constant temperature and pressure, chemical change will tend to occur in whatever direction leads to a decrease in the value of the *Gibbs energy*. In this lesson we will see how G varies with the *composition* of the system as reactants change into products. When G falls as far as it can, all net change comes to a stop. The *equilibrium composition* of the mixture is determined by $\Delta G°$ which also defines the *equilibrium constant* K.

The Road to Equilibrium is Down the Gibbs Energy Hill

This means, of course, that if the total Gibbs energy G of a mixture of reactants and products goes through a minimum value as the composition changes, then all net change will cease— the reaction system will be in a state of *chemical equilibrium*. You will recall that the relative concentrations of reactants and products in the equilibrium state is expressed by the *equilibrium constant*. In this lesson we will examine the relation between the Gibbs energy change for a reaction and the equilibrium constant.

To keep things as simple as possible, we will consider a homogeneous chemical reaction of the form

$$A + B \rightleftharpoons C + D \tag{15.6.1}$$

in which all components are gases at the temperature of interest. If the sum of the standard Gibbs energies of the products is less than that of the reactants, $\Delta G°$ for the reaction will be negative and the reaction will proceed to the right. But how far? If the reactants are *completely* transformed into products, the equilibrium constant would be infinity. The equilibrium constants we actually observe all have finite values, implying that even if the products have a lower Gibbs energy than the reactants, some of the latter will always remain when the process comes to equilibrium.

> A *homogeneous* reaction is one in which everything takes place in a single gas or liquid phase.

To understand how equilibrium constants relate to $\Delta G°$ values, assume that all of the reactants are gases, so that the Gibbs energy of gas A, for example, is given at all times by

$$G_A = G_A^\circ + RT \ln P_A \tag{15.6.2}$$

The Gibbs energy change for the reaction is *sum of the Gibbs energies* of the products, minus *the sum of Gibbs energies* of the reactants:

$$\Delta G = \underbrace{G_C + G_D}_{\text{products}} \underbrace{-G_A - G_B}_{\text{reactants}} \tag{15.6.3}$$

Using Equation 15.6.2 to expand each term on the right of Equation 15.6.3, we have

$$\Delta G = (G^\circ_C + RT\ln P_C) + (G^\circ_D + RT\ln P_D) - (G^\circ_B + RT\ln P_B) - (G^\circ_A + RT\ln P + A) \tag{15.6.4}$$

We can now express the G° terms collectively as ΔG°, and combine the logarithmic pressure terms into a single fraction

$$\Delta G = \Delta G^\circ + RT\ln\left(\frac{P_C P_D}{P_A P_B}\right) \tag{15.6.5}$$

which is more conveniently expressed in terms of the *reaction quotient* Q.

$$\Delta G = \Delta G^\circ + RT\ln Q \tag{15.6.6}$$

The Gibbs energy G is a quantity that becomes more negative during the course of any natural process. Thus as a chemical reaction takes place, G only falls and will never become more positive. Eventually a point is reached where any further transformation of reactants into products would cause G to increase. At this point G is at a minimum (see below), and no further net change can take place; the reaction is then at equilibrium.

> Although Equations 15.6.2-15.6.6 are strictly correct only for perfect gases, we will see later that equations of similar form can be applied to many liquid solutions by substituting concentrations for pressures.

Example 15.6.1: Dissociation of Dinitrogen Tetroxide

Consider the gas-phase dissociation reaction

$$\mathrm{N_2O_4} \rightarrow 2\,\mathrm{NO_2}$$

which is a simple example of the Gibbs energy relationships in a homogeneous reaction.

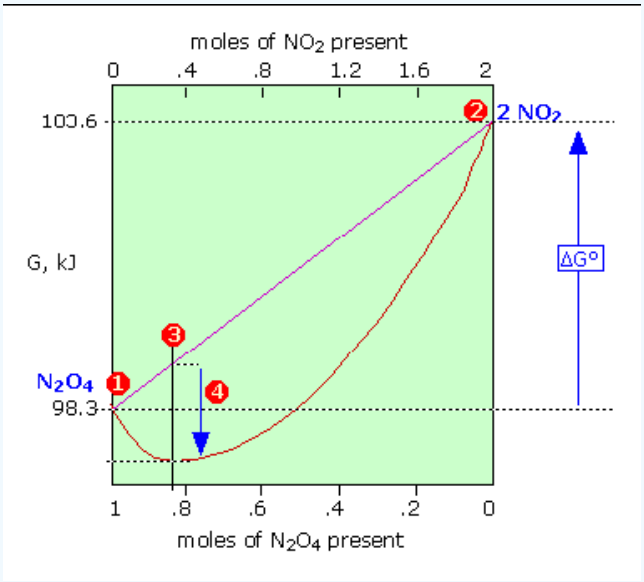


Figure 15.6.1: Gibbs energy for the thermal dissociation reaction of dinitrogen tetroxide

The Gibbs energy of 1 mole of N_2O_4 **(1)** is smaller than that of 2 moles of NO_2 **(2)** by 5.3 kJ; thus $\Delta G^o = +5.3\ \mathrm{kJ}$ for the complete transformation of reactants into products. The straight diagonal line shows the Gibbs energy of all possible compositions if the two gases were prevented from mixing. The red curved line show the Gibbs energy of the actual reaction mixture. This passes through a minimum at **(3)** where 0.814 mol of N_2O_4 are in equilibrium with 0.372 mol of NO_2. The

difference **(4)** corresponds to the Gibbs energy of mixing of reactants and products which always results in an equilibrium mixture whose Gibbs energy is lower than that of either pure reactants or pure products. Thus some amount of reaction will occur even if $\Delta G°$ for the process is positive.

What's the difference between ΔG and $\Delta G°$?

It's very important to be aware of this distinction; that little ° symbol makes a world of difference! First, the standard Gibbs energy change **$\Delta G°$** has a single value for a particular reaction at a given temperature and pressure; this is the difference

$$\sum G_f^\circ(\text{products}) - \sum G_f^\circ(\text{reactants}) \tag{15.6.7}$$

that are tabulated in thermodynamic tables. It corresponds to the Gibbs energy change for a process that never really happens: the *complete transformation* of pure N_2O_4 into pure NO_2 at a constant pressure of 1 atm.

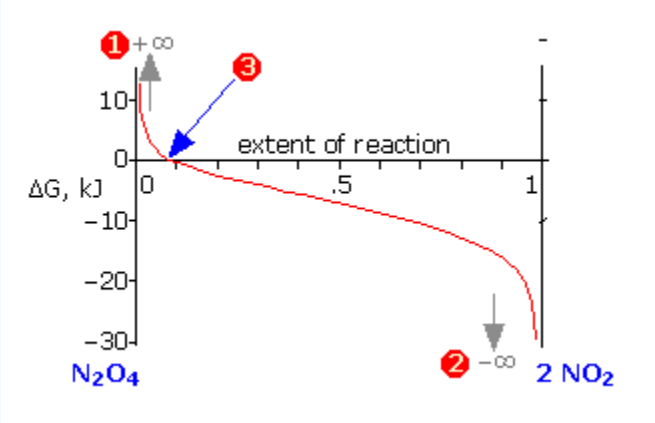


Figure 15.6.2: Gibbs energy change for the thermal dissociation reaction of dinitrogen tetroxide

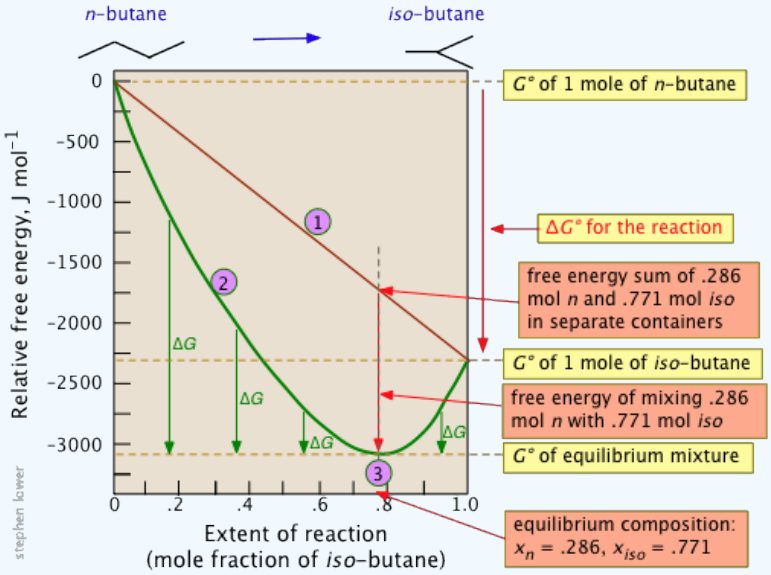
The other quantity ΔG, defined by Equation 15.6.6, represents the total Gibbs energies of all substances in the reaction mixture at any particular system composition. In contrast to ΔG° which is a *constant* for a given reaction, ΔG varies continuously as the composition changes, finally reaching zero at equilibrium. ΔG is the "distance" (in Gibbs energy) from the equilibrium state of a given reaction. Thus for the limiting cases of pure N_2O_4 or NO_2 (as far from the equilibrium state as the system can be!),

$$Q = \frac{[NO_2]^2}{[N_2O_4]} = \pm\infty \tag{15.6.8}$$

which makes the logarithm in Equation 15.6.6, and thus the value of ΔG, approach the same asymptotic limits **(1)** or **(2)**. As the reaction proceeds in the appropriate direction ΔG approaches zero; once there **(3)**, the system is at its equilibrium composition and no further net change will occur.

Example 15.6.2: Isomerization of Butane

The standard molar Gibbs energy change for this very simple reaction is –2.26 kJ, but mixing of the unreacted butane with the product brings the Gibbs energy of the equilibrium mixture down to about –3.1 kJ mol^{-1} at the equilibrium composition corresponding to 77 percent conversion.



Notice particularly that

- The sum of the Gibbs energies of the two gases (*n*-butane and *iso*-butane) separately varies linearly with the composition of the mixture (red line 1).
- The green curve 2 adds the Gibbs energy of mixing to the above sum; its minimum defines the equilibrium composition.
- As the composition approaches the equilibrium value 3, ΔG (which denotes how much farther the Gibbs energy of the system can fall) approaches zero.

The detailed calculations that lead to the values shown above can be found here.

Why reactions lead to *mixtures* of reactants and products

We are now in a position to answer the question posed earlier: if $\Delta G°$ for a reaction is negative, meaning that the Gibbs energies of the products are more negative than those of the reactants, why will some of the latter remain after equilibrium is reached? The answer is that no matter how low the Gibbs energy of the products, the Gibbs energy of the system can be reduced even more by allowing some of the products to be "contaminated" (i.e., diluted) by some reactants. Owing to the entropy associated with mixing of reactants and products, no homogeneous reaction will be 100% complete. An interesting corollary of this is that any reaction for which a balanced chemical equation can be written can in principle take place to some extent, however minute that might be.

Gibbs energies of mixing of products with reactants tend to be rather small, so for reactions having $\Delta G°$ values that are highly negative or positive (±20 kJ mol^{-1}, say), the equilibrium mixture will, for all practical purposes, be either [almost] "pure" reactants or products.

The Equilibrium Constant

Now let us return to Equation 15.6.6 which we reproduce here:

$$\Delta G = \Delta G^\circ + RT \ln Q \tag{15.6.9}$$

As the reaction approaches equilibrium, ΔG becomes less negative and finally reaches zero. At equilibrium $\Delta G = 0$ and $Q = K$, so we can write (must know this!)

$$\Delta G^\circ = -RT \ln K_p \tag{15.6.10}$$

in which K_p, the equilibrium constant expressed in pressure units, is the special value of Q that corresponds to the equilibrium composition.

This equation is one of the most important in chemistry because it relates the equilibrium composition of a chemical reaction system to measurable physical properties of the reactants and products. If you know the entropies and the enthalpies of formation of a set of substances, you can predict the equilibrium constant of any reaction involving these substances without the need to know anything about the mechanism of the reaction.

> Instead of writing Equation 15.6.10 in terms of K_p, we can use any of the other forms of the equilibrium constant such as K_c (concentrations), K_x (mole fractions), K_n(numbers of moles), etc. Remember, however, that for ionic solutions especially, only the K_a, in which *activities* are used, will be strictly valid.

It is often useful to solve Equation 15.6.10 for the equilibrium constant, yielding

$$K = \exp\left(\frac{-\Delta G}{RT}\right) \tag{15.6.11}$$

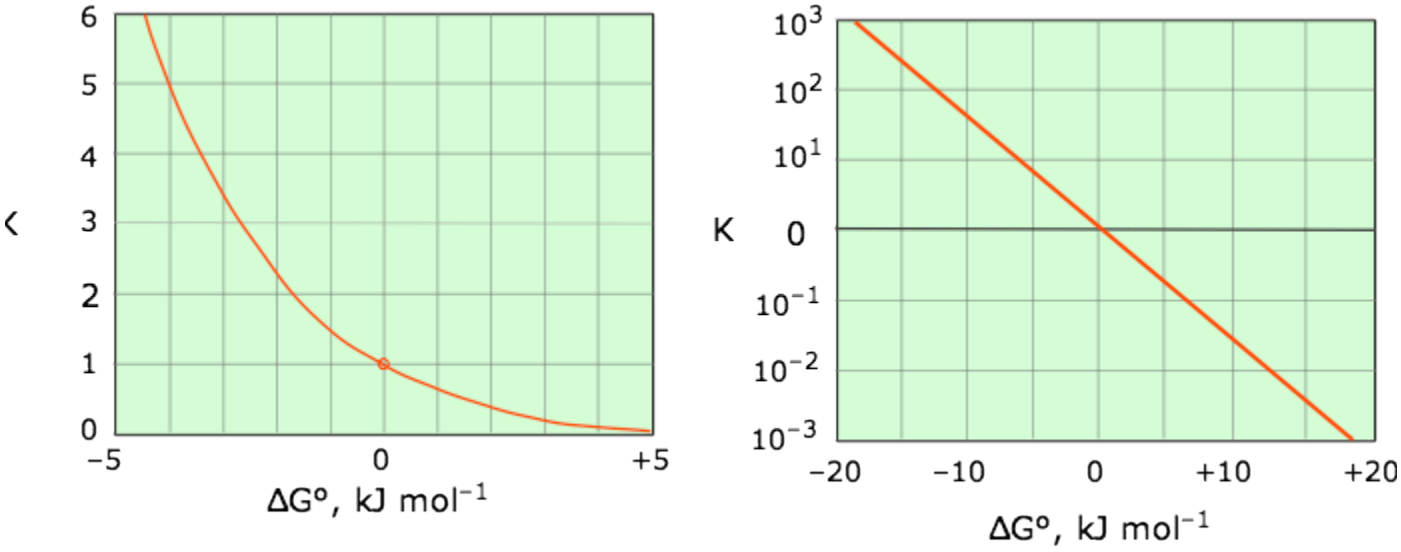


Figure 15.6.3

This relation is most conveniently plotted against the logarithm of K as shown in Figure 15.6.3, where it can be represented as a straight line that passes through the point (0,0).

Example 15.6.3

Calculate the equilibrium constant for the reaction from the following thermodynamic data:

$$\mathrm{H^+\,(aq) + OH^-\,(aq) \rightleftharpoons H_2O\,(l)}$$

$H^+(aq)$	$OH^-(aq)$	$H_2O(l)$
$\Delta H_f°$, kJ mol^{-1}	0	−230.0	−285.8

$S°$, J K^{-1} mol^{-1}
0*
–10.9
70.0

* Note that the standard entropy of the hydrogen ion is zero by definition. This reflects the fact that it is impossible to carry out thermodynamic studies on a single charged species. All ionic entropies are relative to that of $\mathrm{H^+\,(aq)}$, which explains why some values (as for aqueous hydroxide ion) are negative.

Solution

From the above data, we can evaluate the following quantities:

$$\begin{aligned}\Delta H^o &= \sum \Delta H_f^o(\text{products}) - \sum \Delta H_f^o(\text{reactants}) \\ &= (-285.8) - (-230) \\ &= -55.8\, kJ\ mol^{-1}\end{aligned}$$

$$\begin{aligned}\Delta S^o &= \sum \Delta S^o(\text{products}) - \sum \Delta S°(\text{reactants}) \\ &= (70.0) - (-10.9) \\ &= +80.8\, J\ K^{-1}\ mol^{-1}\end{aligned}$$

The value of $\Delta G\,°$ at 298 K is

$$\begin{aligned}\Delta H^o - T\Delta S^o &= (-55800) - (298)(80.8) \\ &= -79900\, J\, mol^{-1}\end{aligned}$$

From Equation 15.6.11 we have

$$\begin{aligned}K &= \exp\left(\frac{-79900}{8.314 \times 298}\right) \\ &= e^{32.2} = 1.01 \times 10^{-14}\end{aligned}$$

Equilibrium and Temperature

We have already discussed how changing the temperature will increase or decrease the tendency for a process to take place, depending on the sign of ΔS°. This relation can be developed formally by differentiating the relation

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \tag{15.6.12}$$

with respect to the temperature:

$$\frac{d(-\Delta G^\circ)}{dT} = -\Delta S^\circ \tag{15.6.13}$$

Hence, the sign of the entropy change determines whether the reaction becomes more or less allowed as the temperature increases.

We often want to know how a change in the temperature will affect the value of an equilibrium constant whose value is known at some fixed temperature. Suppose that the equilibrium constant has the value K_1 at temperature T_1 and we wish to estimate K_2 at temperature T_2. Expanding Equation 15.6.11 in terms of ΔH° and ΔS°, we obtain

$$-RT_1 \ln K_1 = \Delta H^\circ - T_1\Delta S^\circ \tag{15.6.14}$$

and

$$-RT_2 \ln K_2 = \Delta H^\circ - T_2\Delta S^\circ \tag{15.6.15}$$

Dividing both sides by *RT* and subtracting, we obtain

$$\ln K_1 - \ln K_2 = -\left(\frac{\Delta H^\circ}{RT_1} - \frac{\Delta H^\circ}{RT_2}\right) \tag{15.6.16}$$

Which is most conveniently expressed as the ratio

$$\ln \frac{K_1}{K_2} = -\frac{\Delta H^\circ}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right) \tag{15.6.17}$$

This is its theoretical foundation of Le Chatelier's Principle with respect to the effect of the temperature on equilibrium:

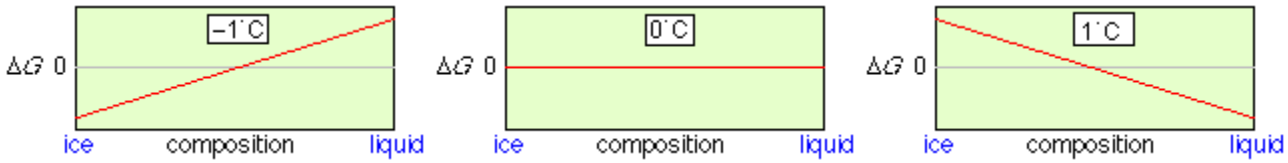
- if the reaction is exothermic $\Delta H^\circ < 0$, then increasing temperature will make the second exponential term smaller and K will decrease. The equilibrium will then "shift to the left".
- If $\Delta H^\circ > 0$, then increasing *T* will make the exponent less negative and K will increase and the equilibrium will "shift to the right".

This is an extremely important relationship, but not just because of its use in calculating the temperature dependence of an equilibrium constant. Even more important is its application in the "reverse" direction to experimentally determine Δ*H*° from two values of the equilibrium constant measured at different temperatures. Direct calorimetric determinations of heats of reaction are not easy to make; relatively few chemists have the equipment and experience required for this rather exacting task. Measurement of an equilibrium constant is generally much easier, and often well within the capabilities of anyone who has had an introductory Chemistry course. Once the value of ΔH° is determined it can be combined with the Gibbs energy change (from a single observation of *K*, through Equation 15.6.11) to allow ΔS° to be calculated through Equation 15.6.13.

Equilibrium Without Mixing: it's all or nothing

You should now understand that for *homogeneous reactions* (those that take place entirely in the gas phase or in solution) the equilibrium composition will never be 100% products, no matter how much lower their Gibbs energy relative to the reactants. As was summarized in the N_2O_4-dissociation example discussed previously. This is due to "dilution" of the products by the reactants. In *heterogeneous reactions* (those which involve more than one phase) this dilution, and the effects that flow from it, may not be possible.

A particularly simple but important type of a heterogeneous process is *phase change*. Consider, for example, an equilibrium mixture of ice and liquid water. The concentration of H_2O in each phase is dependent only on the density of the phase; there is no way that ice can be "diluted" with water, or vice versa. This means that at all temperatures other than the freezing point, the lowest Gibbs energy state will be that corresponding to pure ice or pure liquid. Only at the freezing point, where the Gibbs energies of water and ice are identical, can both phases coexist, and they may do so in any proportion.



Gibbs energy of the ice-water system

Only at 0°C can ice and liquid water coexist in any proportion. Note that in contrast to the homogeneous N_2O_4 example, there is no Gibbs energy minimum at intermediate compositions.

Coupled Reactions

Two reactions are said to be *coupled* when the product of one of them is the reactant in the other:

$$A \rightarrow B$$

and

$$B \rightarrow C$$

If the standard Gibbs energy of the first reaction is positive but that of the second reaction is sufficiently negative, then for the overall process will be negative and we say that the first reaction is "driven" by the second one. This, of course, is just another way of describing an effect that you already know as the Le Chatelier principle: the removal of substance B by the second reaction causes the equilibrium of the first to "shift to the right". Similarly, the equilibrium constant of the overall reaction is the product of the equilibrium constants of the two steps.

1 $Cu_2S(s) \rightarrow 2\ Cu(s) + S(s)$	$\Delta G° = +\ 86.2$ kJ	$\Delta H° = +\ 76.3$ kJ
2 $S(s) + O_2(g) \rightarrow SO_2(g)$	$\Delta G° = -300.1$ kJ	$\Delta H° = +\ 296.8$ kJ
3 $Cu_2S(s) \rightarrow 2\ Cu(s) + SO_2(g)$	$\Delta G° = -213.9$ kJ	$\Delta H° = -\ 217.3$ kJ

In the above example, reaction **1** is the first step in obtaining metallic copper from one of its principal ores. This reaction is endothermic and it has a positive Gibbs energy change, so it will not proceed spontaneously at any temperature. If Cu_2S is heated in the air, however, the sulfur is removed as rapidly as it is formed by oxidation in the highly spontaneous reaction **2**, which supplies the Gibbs energy required to drive **1**. The combined process, known as *roasting*, is of considerable industrial importance and is one of a large class of processes employed for winning metals from their ores.

This page titled 15.6: Free Energy and Equilibrium is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.7: Some Applications of Entropy and Free Energy

Learning Objectives

You are expected to be able to define and explain the significance of terms identified in *bold*

- What do all of the ***colligative properties*** of solutions have in common in terms of entropy?
- At the melting and boiling points, the two phases in equilibrium have equal numbers of energetically accessible microstates. How does the addition of a nonvolatile solute upset this balance in the two cases?
- Describe the effect of increasing the hydrostatic pressure on a liquid has on its vapor pressure, and suggest a reason for this.
- Under what conditions does ***osmotic flow*** occur? What is the fundamental driving force? What is the definition of *osmotic pressure*?
- Explain the critical role of temperature in the extraction of metals from their oxide ores.
- Describe the bioenergetic cycle of free energy in terms of the roles of glucose, electron-acceptors, ATP, and photosynthesis.
- Define aerobic and non-aerobic oxidation, and explain why they differ in efficiency.
- Sketch a simple ***proton-free energy*** *diagram* showing why an acid reacts with a base.
- Sketch a simple ***electron-free energy*** *diagram* showing why a metal reacts with an oxidizing agent that is below it in the electromotive series.
- What thermodynamic factors are involved in the spontaneous contraction of an elastic material such as a rubber band?

Thermodynamics may appear at first to be a rather esoteric subject, but when you think about it, almost every chemical (and biological) process is governed by changes in entropy and free energy. Examples such as those given below should help you connect these concepts with the real world.

Colligative properties of solutions

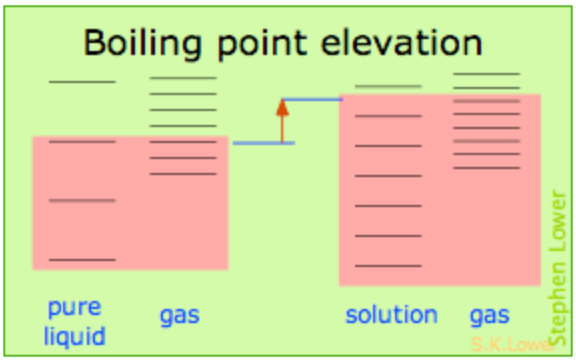
Vapor pressure lowering, boiling point elevation, freezing point depression and osmosis are well-known phenomena that occur when a non-volatile solute such as sugar or a salt is dissolved in a volatile solvent such as water. All these effects result from “dilution” of the solvent by the added solute, and because of this commonality they are referred to as *colligative properties* (Lat. *co ligare*, connected to.) The key role of the solvent concentration is obscured by the expressions used to calculate the magnitude of these effects, in which only the solute concentration appears. The details of how to carry out these calculations and the many important applications of colligative properties are covered in the unit on solutions. Our purpose here is to offer a more complete explanation of why these phenomena occur.

The conventional explanation is that dilution of a liquid by a non-volatile solute reduces the vapor pressure or "escaping tendency" of the liquid in that phase, leading to a net transport of material into that phase. Equilibrium can then be re-established by adjusting the temperature of applying a hydrostatic pressure to the solution. The stable phase at any temperature will be the one having the lowest free energy, and from which molecules have the smallest escaping tendency.

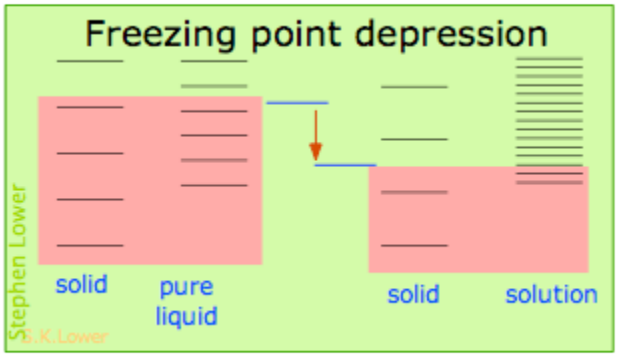
A more fundamental approach is to recall that dilution of a liquid creates uncountable numbers of new microstates, increasing the density of quantum states in the solution compared to that in the pure liquid. To the extent that these new states are thermally accessible, they will become populated at the expense of some of the microstates of the other phase. Equilibrium between two phases can be restored by adjusting the temperature or pressure so that equal numbers of microstates are occupied in each phase.

Boiling point elevation and freezing point depression

These effects are readily understood in terms of the schematic diagrams shown below. The red shading indicates the temperatures required to make equal numbers of microstates thermally accessible in the two phases.

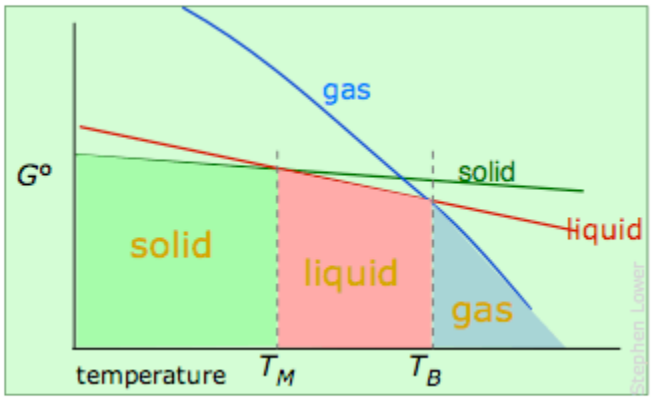


Dilution of the solvent adds new energy states to the liquid, but does not affect the vapor phase. This raises the temperature required to make equal numbers of microstates accessible in the two phases.

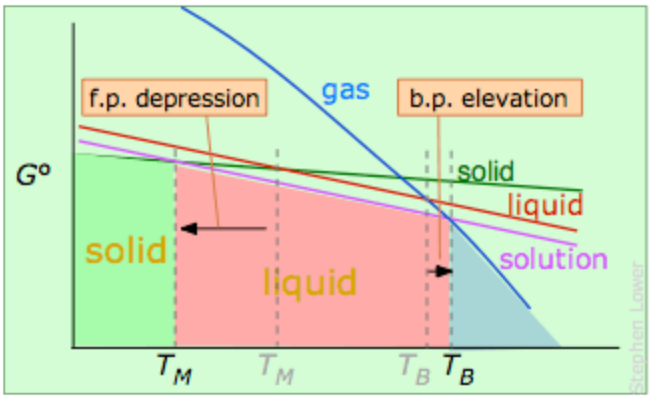


Dilution of the solvent adds new energy states to the liquid, but does not affect the solid phase. This reduces the temperature required to make equal numbers of states accessible in the two phases. The shifts of the boiling- and and freezing points in opposite directions reflects the higher energies of vapor microstates and lower energies of solid microstates in relation to those of the liquid. Note that these diagrams are purely schematic and nowhere near to scale!

As mentioned previously, the more conventional explanation of bp elevation and fp depression is given in terms of vapor pressures. The latter are measures of the free energies of molecules in a phase, and the relationships are best understood by showing plots of $-T\Delta S°$ for each phase in terms of the temperature:



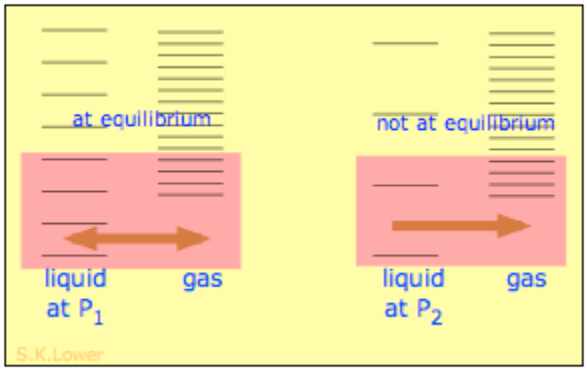
Melting and boiling points of a pure substance.
The stable phase at any temperature will be the one having the lowest free energy, and from which molecules have the smallest escaping tendency.



Melting- and boiling points of a solvent containing a non-volatile solute.
The solute "dilutes" the solvent, reducing its free energy (purple line) and shifting the transition temperatures in opposite directions.

The pressure on a liquid affects its volatility

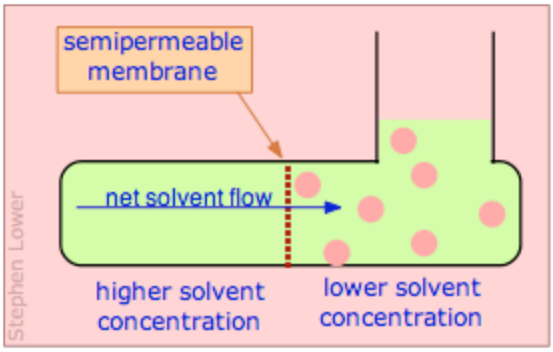
When a liquid is subjected to hydrostatic pressure— for example, by an inert, non-dissolving gas that occupies the vapor space above the surface, the vapor pressure of the liquid is raised. The pressure acts to compress the liquid very slightly, effectively narrowing the potential energy well in which the individual molecules reside and thus increasing their tendency to escape from the liquid phase. (Because liquids are not very compressible, the effect is quite small; a 100-atm applied pressure will raise the vapor pressure of water at 25° C by only about 2 torr.) In terms of the entropy, we can say that the applied pressure reduces the dimensions of the "box" within which the principal translational motions of the molecules are confined within the liquid, thus reducing the density of the quantized energy states in the liquid phase.



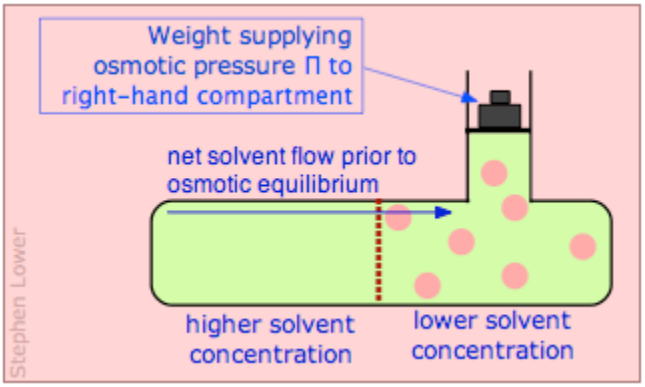
Why pressure affects the volatility of a liquid. Applying hydrostatic pressure to a liquid increases the spacing of its microstates, so that the number of energetically accessible states in the gas, although unchanged, is relatively greater— thus increasing the tendency of molecules to escape into the vapor phase. In terms of free energy, the higher pressure raises the free energy of the liquid, but does not affect that of the gas phase.

Osmotic flow and osmotic pressure

Molecules in a fluid are always subject to random thermal motions. If a given kind of molecule is present in greater concentration in one region of a fluid, the effect of these random motions will be to produce a net migration into the region of lower concentration; this migration is known as *thermal diffusion*.

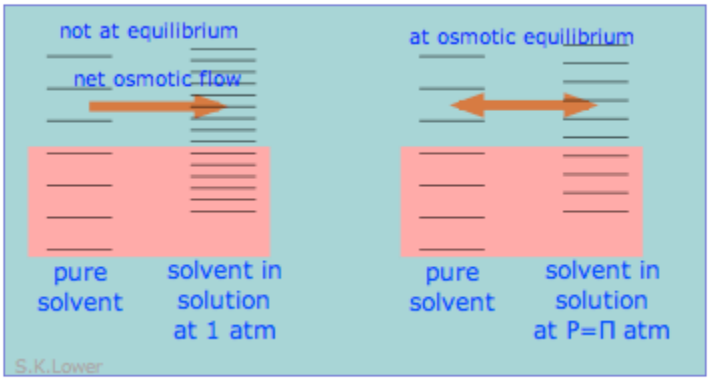


Osmotic flow (commonly known simply as *osmosis*) occurs when molecules of a solvent diffuse through a membrane that is permeable only to those molecules, and into a solution in which the solvent is "diluted" by the presence of solute molecules. In the simplest example, there is pure solvent on the left side of the membrane; on the other side the same solvent contains a solute whose molecules are too large to pass through the membrane. Because the solvent concentration on the right side will always be smaller than that on the left, osmotic flow will continue indefinitely if the right side can accommodate the increased volume.



If, however, the liquid on the right side of the membrane is placed under hydrostatic pressure, the driving force for osmotic flow will diminish. If the pressure is raised sufficiently high, osmotic flow will come to a halt; the system is then said to be in osmotic equilibrium. The pressure required to bring about osmotic equilibrium (and thus to stop osmotic flow) is known as the *osmotic pressure*, commonly denoted by Π (Greek capital *pi*).

Osmotic equilibrium, like any kind of equilibrium, occurs when the free energies (that is, the escaping tendencies) of the diffusible molecules are the same on the two sides of the membrane. The free energy on the right side, initially lower due to the lower solvent concentration, is raised by the hydrostatic pressure, making it identical with that of the pure liquid on the left.



From the standpoint of microscopic energy states, the effect of the applied pressure is to slightly increase the spacing of solvent energy states on right side of the membrane so as to equalize the number of accessible states on the two sides, as shown here in a very schematic way.

Extraction of Metals from their Oxides

Since ancient times, the recovery of metals from their ores has been one of the most important applications of chemistry to civilization and culture. The oldest, and still the most common smelting process for oxide ores involves heating them in the presence of carbon. Originally, charcoal was used, but industrial-scale smelting uses coke, a crude form of carbon prepared by pyrolysis (heating) of coal. The basic reactions are:

$$MO + C \rightleftharpoons M + CO \tag{15.7.1}$$

$$MO + ½O_2 \rightleftharpoons M + ½CO_2 \tag{15.7.2}$$

$$MO + CO \rightleftharpoons M + CO_2 \tag{15.7.3}$$

Each of these can be regarded as a pair of *coupled reactions* in which the metal M and the carbon are effectively competing for the oxygen atom. Using reaction 15.7.1 as an example, it can be broken down into the following two parts:

$$MO \rightleftharpoons M + ½O_2 \quad \Delta G^o > 0 \tag{15.7.4}$$

$$C + ½O_2 \rightleftharpoons CO \quad \Delta G^o < 0 \tag{15.7.5}$$

At ordinary environmental temperatures, reaction 15.7.4 is always spontaneous in the reverse direction (that is why ores form in the first place!), so $\Delta G°$ of Reaction 15.7.4 will be positive. $\Delta G°$ for reaction 15.7.5 is always negative, but at low temperatures it will not be sufficiently negative to drive 15.7.4.

The *smelting* process depends on the different ways in which the free energies of reactions like 15.7.4 and 15.7.4 vary with the temperature. This temperature dependence is almost entirely dominated by the $T\Delta S°$ term in the Gibbs function, and thus by the entropy change. The latter depends mainly on Δn_g, the change in the number of moles of gas in the reaction. Removal of oxygen from the ore is always accompanied by a large increase in the system volume so ΔS for this step is always positive and the reaction becomes more spontaneous at higher temperatures. The temperature dependences of the reactions that take up oxygen vary, however (Table 23.6.X).

Table 23.6.X: The temperature dependences of the reactions

Reaction	Δn_g	$d(\Delta G°)/dT$
C + ½ O_2 → CO	0.5	<0>
C + O_2 → CO_2	0	0
CO + ½ O_2 → CO_2	-0.5	>0

A plot of the temperature dependences of the free energies of these reactions, superimposed on similar plots for the oxygen removal reactions 15.7.4 is called an *Ellingham diagram*. For a given oxide MO to be smeltable, the temperature must be high enough that reaction 15.7.4 falls below that of at least one of the oxygen-consuming reactions. The slopes of the lines on this diagram are determined by the sign of the entropy change.

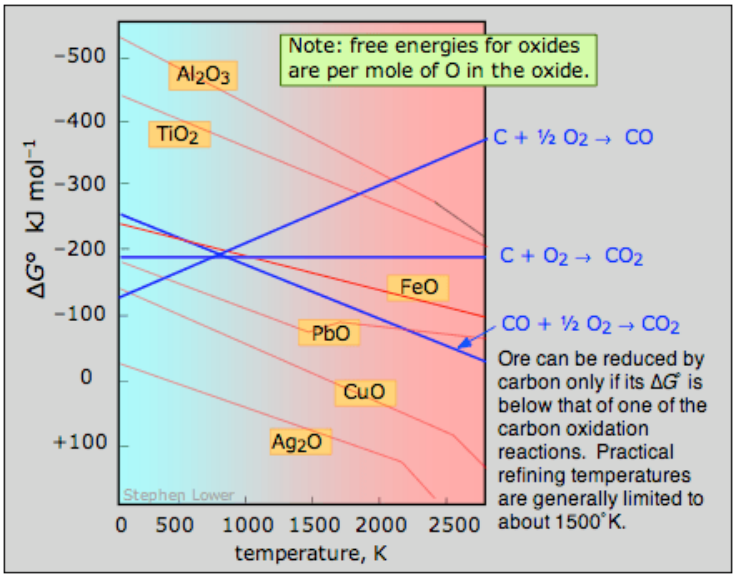


Figure 15.7.1: Ellingham diagram. An ore can be reduced by carbon only if its Gibbs free energy of formation falls below that of one of the carbon reduction reactions (blue lines.) Practical refining temperatures are generally limited to about 1500°K.

Examination of the Ellingham diagram shown above illustrates why the metals known to the ancients were mainly those such as copper and lead, which can be obtained by smelting at the relatively low temperatures that were obtainable by the methods available at the time in which a charcoal fire supplied both the heat and the carbon. Thus the bronze age preceded the iron age; the latter had to await the development of technology capable of producing higher temperatures, such as the blast furnace. Smelting of aluminum oxide by carbon requires temperatures too high to be practical; commercial production of aluminum is accomplished by electrolysis of the molten ore.

Bioenergetics

Many of the reactions that take place in living organisms require a source of free energy to drive them. The immediate source of this energy in heterotrophic organisms, which include animals, fungi, and most bacteria, is the sugar **glucose**. Oxidation of glucose to carbon dioxide and water is accompanied by a large negative free energy change

$$\mathrm{C_6H_{12}O_6 + 6\,O_2 \;\rightarrow 6\,CO_2 + 6\,H_2O} \quad \Delta G^o = -2,880\,kJ\,mol^{-1} \tag{15.7.6}$$

Of course it would not do to simply “burn” the glucose in the normal way; the energy change would be wasted as heat, and rather too quickly for the well-being of the organism! Effective utilization of this free energy requires a means of capturing it from the glucose and then releasing it in small amounts when and where it is needed. This is accomplished by breaking down the glucose in a series of a dozen or more steps in which the energy liberated in each stage is captured by an “energy carrier” molecule, of which the most important is **adenosine diphosphate**, known as **ADP**. At each step in the breakdown of glucose, an ADP molecule reacts with inorganic phosphate (denoted by P_i) and changes into **ATP**:

$$ADP + P_i \rightarrow ATP \tag{15.7.7}$$

$$\Delta G^\circ = +30\ kJ\ mol^{-1} \tag{15.7.8}$$

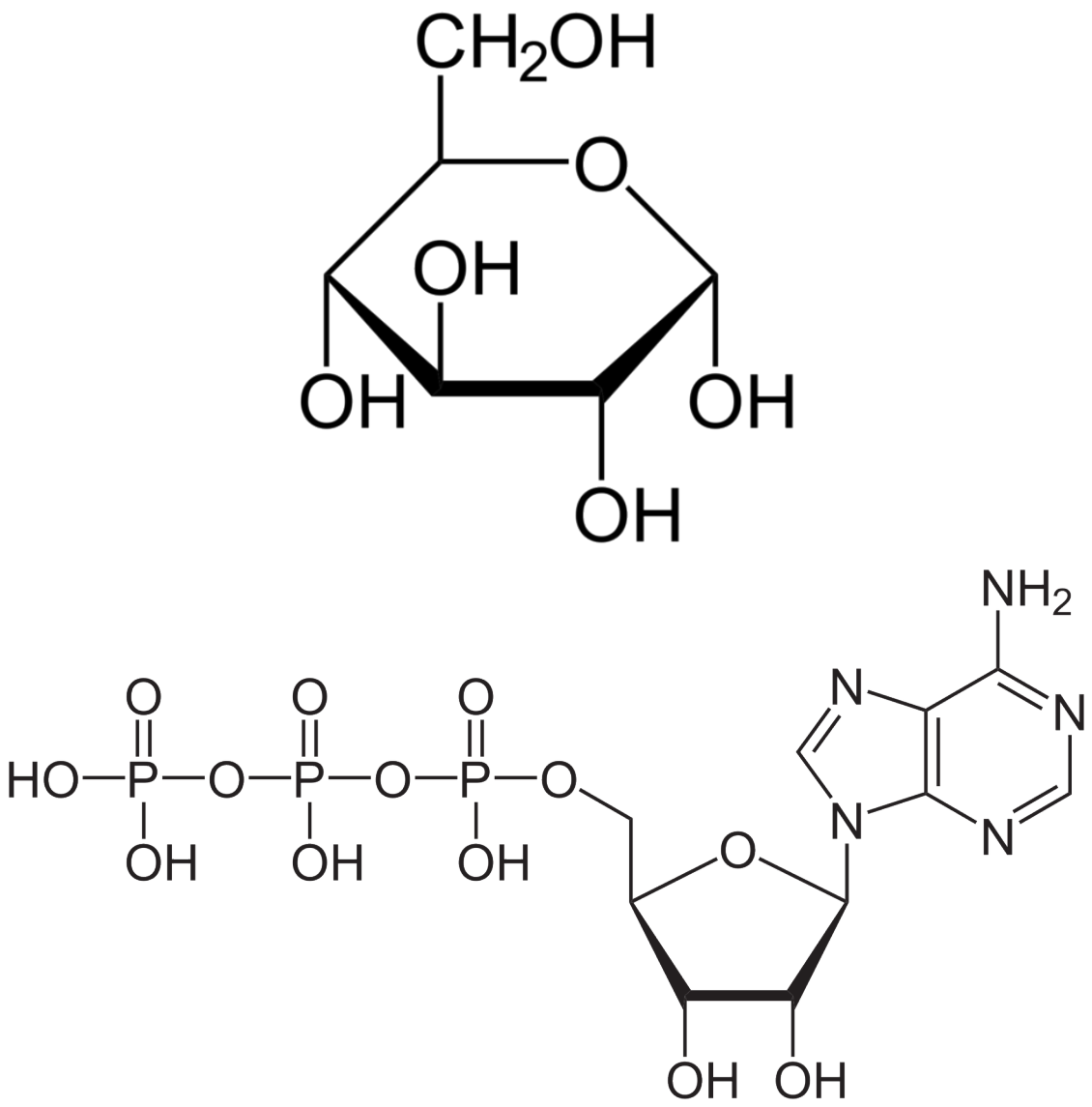


Figure 23.6.X: Structures of glucose and ATP

The 30 kJ mol^{-1} of free energy stored in each ATP molecule is released when the molecule travels to a site where it is needed and loses one of its phosphate groups, yielding inorganic phosphate and ADP, which eventually finds its way back the site of glucose metabolism for recycling back into ATP. The complete breakdown of one molecule of glucose is coupled with the production of 38 molecules of ATP according to the overall reaction

$$\ce{C6H12O6 + 6 O2 + 38 Pi + 38 ADP \rightarrow 38 ATP + 6CO2 + 44 H2O} \label{6-3}$$

For each mole of glucose metabolized, 38 × (30 kJ) = 1140 kJ of free energy is captured as ATP, representing an energy efficiency of 1140/2880 = 0.4. That is, 40% of the free energy obtainable from the oxidation of glucose is made available to drive other metabolic processes. The rest is liberated as heat.

Where does the glucose come from? Animals obtain their glucose from their food, especially cellulose and starches that, like glucose, have the empirical formula $\{CH_2O\}$. Animals obtain this food by eating plants or other animals. Ultimately, all food comes from plants, most of which are able to make their own glucose from $CO_2$and H_2O through the process of photosynthesis. This is just the reverse of Eq 40 in which the free energy is supplied by the quanta of light absorbed by chlorophyll and other photosynthetic pigments.

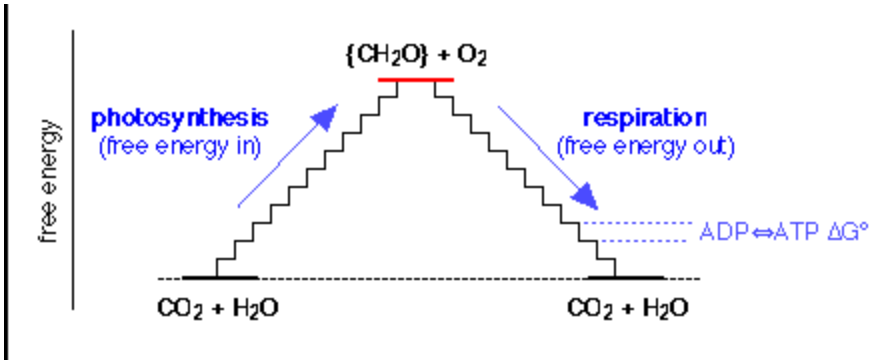


Figure 23.6.X: The photosynthesis-respiration free energy cycle. The zigzags represent the individual steps of energy capture by ADP and its delivery to metabolic processes by ATP.

This describes *aerobic respiration*, which evolved after the development of photosynthetic life on Earth began to raise the concentration of atmospheric oxygen. Oxygen is a poison to most life processes at the cellular level, and it is believed that aerobic respiration developed as a means to protect organisms from this peril. Those that did not adapt have literally "gone underground" and constitute the more primitive anaerobic bacteria.

The function of oxygen in respiration is to serve as an acceptor of the electrons that glucose loses when it undergoes oxidation. Other electron acceptors can fulfill the same function when oxygen is not available, but none yields nearly as much free energy. For example, if oxygen cannot be supplied to mammalian muscle cells as rapidly as it is needed, they switch over to an anaerobic process yielding lactic acid instead of CO_2:

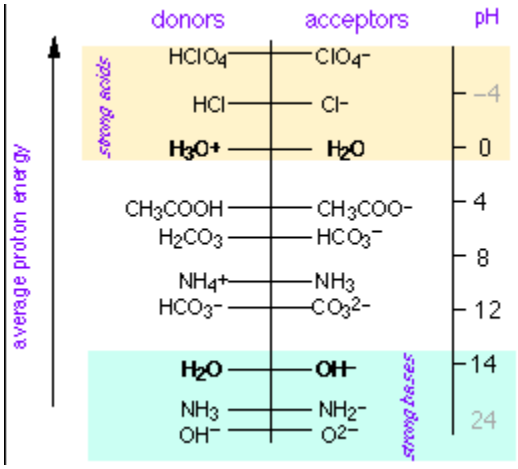
$$\mathrm{C_6H_{12}O_6 + 2\,ADP \longrightarrow 2\,CH_3CH(OH)COOH} \quad \Delta G^o = -218\,kJ\,mol^{-1} \tag{15.7.9}$$

In this process, only (2 × 30 kJ) = 60 kJ of free energy is captured, so the efficiency is only 28% on the basis of this reaction, and it is even lower in relation to glucose. In "aerobic" exercising, one tries to maintain sufficient lung capacity and cardiac output to supply oxygen to muscle cells at a rate that promotes the aerobic pathway.

Respiration and many other metabolic processes involve electron-transfer reactions.

Acid-base reactions: the fall of the proton

According to the widely useful *Brønsted-Lowry concept*, an acid is a proton donor and a base is a proton acceptor. In 1953, Gurney showed how this idea could be made even more useful by placing acid-base conjugate pairs on a *proton-free energy scale*.



In this view, acids are *proton sources* and bases are *proton sinks*. Protons fall spontaneously from acids to fill sinks in which the proton free energy levels are lower. The pH is a measure of the average proton free energy in the solution; when this quantity is the same as the proton free energy level of a conjugate pair, the two species are present in equal concentrations (this

corresponds, of course to the equality of pH and pK_a in the conventional theory.)

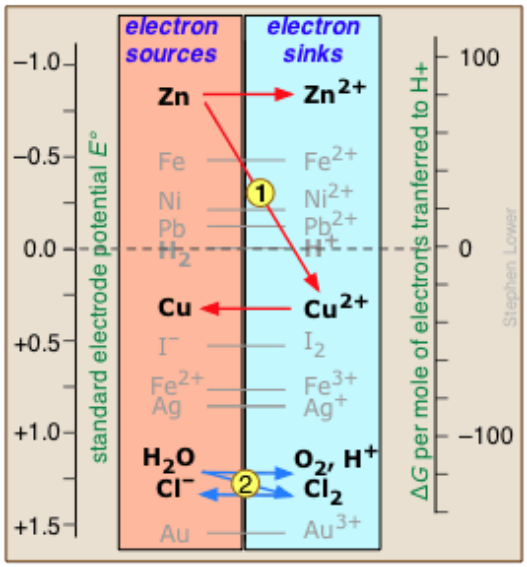
The proton-free energy concept is commonly employed in aquatic environmental chemistry in which multiple acid-base systems must be dealt with on a semi-quantitative bases. It is, however, admirably adapted to any presentation of acid-base chemistry, even at the first-year college level, and it seems a shame that it never seems to have made its way into the ordinary curriculum.

Oxidation-reduction: the fall of the electron

Oxidation-reduction reactions proceed in a direction that allows the electron to “fall” (in free energy) from a “source” to a “sink”. Later on when you study electrochemistry you will see how this free energy can manifest itself as an electrical voltage and be extracted from the system as electrical work.

Electron-free energy levels and the electromotive series

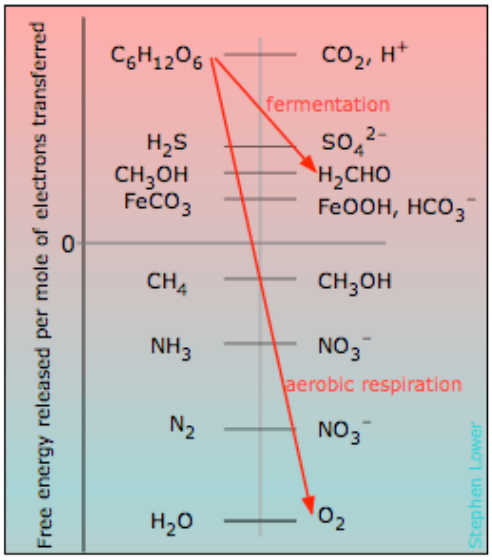
You may already have seen an electromotive force table that shows the relative tendencies of different reducing agents to donate electrons. The same information can be presented in another way that relates electron-donating (reducing) power to the "fall" in free energy that electrons undergo when they are transferred to an oxidizing agent.



Any available sink on the right side will tend to drain electrons from a source above it. For example, immersion of metallic zinc in a solution of $CuSO_4$ allow electrons to "fall" (1) from the high free energy they possess in Zn down to the lower free energy level in the newly-introduced Cu^{2+} ions. This will result in the reduction of Cu^{2+} to metallic copper and the oxidation of the zinc (red arrows.)

Similarly, addition of chlorine to water will introduce a new electron sink (Cl_2) that lies below the free energy of the accessible electrons in H_2O, draining them away (1) and producing O_2 and Cl^- (blue arrows.) Note especially the positions of the H_2/ H^+ and H_2O/O_2,H^+ couples on this chart, as they define the range of E°s for substances that will not decompose water. The zero for G° is arbitrarily set at the electron activity at which the H_2/H^+ couple is at equilibrium; this corresponds to E° = 0 volts on the ordinary electromotive scale.

Coupled redox reactions are central to bioenergetics



Figure, numerous electron acceptors can serve this function, but none yields nearly as much free energy as aerobic respiration whose net reaction is

$$\{CH_2O\} \rightarrow CO_2 + H_2O \tag{15.7.10}$$

In the above chart, the zero for *G*° is arbitrarily set at the electron free energy required to maintain the H_2/H^+ couple is at equilibrium; this corresponds to E° = 0 volts on the ordinary electromotive scale.

Organisms that live in environments where oxygen is lacking, such as marshes, muddy soils, and the intestinal tracts of animals, must utilize other electron acceptors to extract free energy from carbohydrate. A wide variety of inorganic ions such as sulfate and nitrate, as well as other carbon compounds can serve as electron acceptors, yielding the gaseous products like H_2S, NH_3 and CH_4 which are commonly noticed in such locations. From the location of these acceptors on the scale, it is apparent that the amount of energy they can extract from a given quantity of carbohydrate is much less than for O_2. One reason that aerobic organisms have dominated the earth is believed to be the much greater energy-efficiency of oxygen as an electron acceptor.

What did organisms use for food before there was a widespread supply of carbohydrate in the world? Any of the electron sources near the upper left of the table can in theory serve this function, although at reduced energy efficiency. As a matter of fact, there are still a number of these autotrophic bacteria around whose “food” is CH_4 , CH_3OH, $FeCO_3$, and even H_2!

Thermodynamics of Rubber Bands

Rubber is composed of random-length chains of polymerized *isoprene* molecules. The poly(isoprene) chains are held together partly by weak intermolecular forces, but are joined at irregular intervals by covalent disulfide bonds so as to form a network. The intermolecular forces between the chain fragments tend to curl them up, but application of a tensile force can cause them to elongate. The disulfide cross-links prevent the chains from slipping apart from one another, thus maintaining the physical integrity of the material. Without this cross-linking, the polymer chains would behave more like a pile of spaghetti.

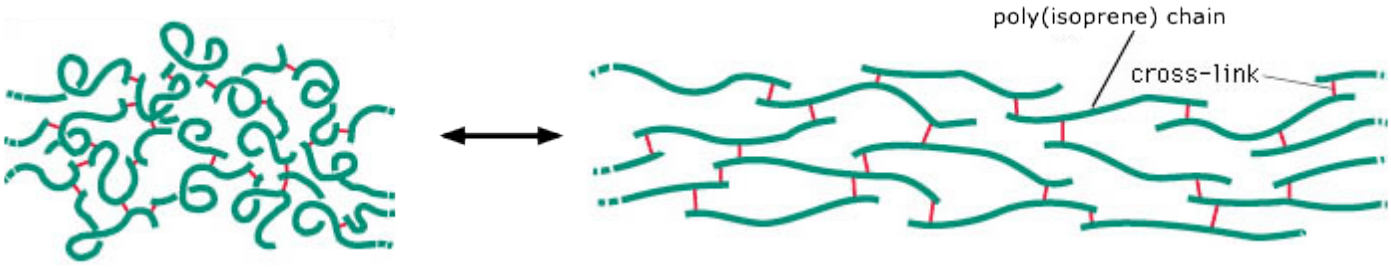


Figure 23.6.X: This image is adapted from one intended to show the structure of the rubber-like protein elastin, which is found in many organisms, including humans.

The ability of rubber bands and other elastic substances to undergo a change in physical dimensions in response to a change in the applied stretching force is subject to the same laws of thermodynamics as any other physical process. You can investigate this for yourself.

Example 15.7.1

Hold a rubber band (the thicker the better) against your upper lip, and notice how the temperature changes when the band is stretched, and then again when it is allowed to contract.

a. Use the results of this observation to determine the signs of ΔH, ΔG and ΔS for the process

$$rubber_{stretched} \rightarrow rubber_{unstretched} \tag{15.7.11}$$

b. How will the tendency of the stretched rubber to contract be changed if the temperature is raised?

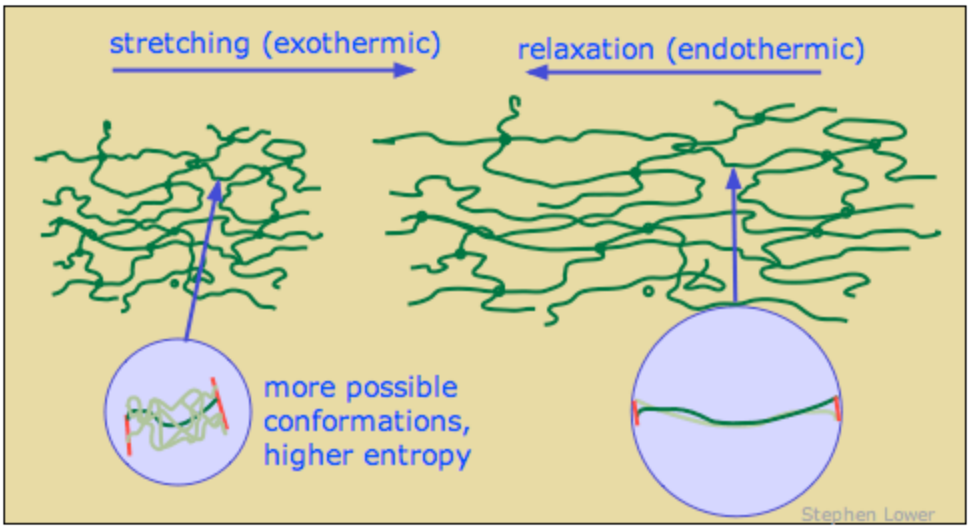
Solution

a. Contraction is obviously a spontaneous process, so ΔG is negative. You will have observed that heat is given off when the band is stretched, meaning that contraction is endothermic, so $\Delta H > 0$. Thus according to $\Delta G = \Delta H - T\Delta S$, ΔS for the contraction process must be positive.
b. Because $\Delta S > 0$, contraction of the rubber becomes more spontaneous as the temperature is raised.

Why is Rubber Elastic?

When an ordinary material is placed under tension, the strain energy is taken up by bond distortions and is stored as electrostatic potential energy which rises very rapidly so as to greatly inhibit further elongation. In rubber-like polymers, this does not happen; the strain energy is instead stored as thermal (kinetic) energy.

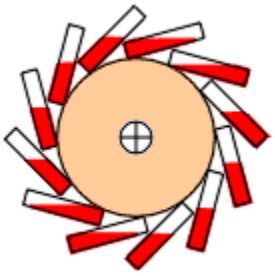
Free polymer chains naturally tend to curl up in complex and ever-changing ways as thermal energy allows random bond rotations to take place. In a rubber-like material in its relaxed state, the portions of the polymer chains between cross-links are continually jumping between different randomly-coiled configurations.



When the rubber is stretched, the polymer segments straighten out as the applied force overcomes the weak dispersion force interactions that caused the strands to curl. Each chain segment is pulled into an almost-straight conformation, thus greatly reducing the quantity of thermal energy it can store. The excess thermal energy spreads into the material and is lost in the form of heat. When the rubber relaxes, the polymer strands curl up again and soak up thermal energy.

The spontaneous contraction of rubber is largely an entropy-driven process. The number of energetically-equivalent ways of distributing thermal energy amongst the nearly-linear polymer chains of the stretched state of rubber is insignificant compared to those available when the chains are curled up in random ways, so the un-stretched form of rubber is statistically the most likely one by overwhelming odds. As noted in part (b) of the above problem example, the gain in entropy when the rubber contracts drives ΔG more negative at higher temperatures. This means that a rubber band, held at constant tension in stretched state, will contract when it is heated.





This fact can be put to use in an interesting way. Replace the spokes of a bicycle wheel with rubber bands, and shine a heat lamp on one side of the wheel. The contraction of the heated bands will shift the wheel off-center, causing it to rotate. This rotation will continue indefinitely as long as the heat source is present. The device has become a heat engine whose working "fluid" is rubber!

This recalls the classic perpetual motion machine design in which a wheel is caused to rotate by continually-shifting unbalanced weights (the one depicted here uses hinged vials of mercury). That, as we saw, would violate the Second Law by producing work in a cyclic process without degrading heat to a lower temperature. The rubber-band heat engine avoids this pitfall by absorbing heat from an external source on one side of the wheel, and releasing it at a lower temperature on the unheated side.

This page titled 15.7: Some Applications of Entropy and Free Energy is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

15.8: Quantum states, Microstates, and Energy spreading in Reactions

Learning Objectives

You are expected to be able to define and explain the significance of terms identified in ***bold***.

energy within a system.

- The entropy of a substance increases with its molecular weight and complexity and with temperature. The entropy also increases as the pressure or concentration becomes smaller. Entropies of gases are much larger than those of condensed phases.

The Physical Meaning of Entropy

Entropy is a measure of the degree of spreading and sharing of thermal energy within a system. This “spreading and sharing” can be spreading of the thermal energy into a larger volume of *space* or its sharing amongst previously inaccessible microstates of the system. The following table shows how this concept applies to a number of common processes.

Table 15.8.1*: Observations and Explanations in terms of Entropy changes*

system and process	source of entropy increase of system
A deck of cards is shuffled, or 100 coins, initially heads up, are randomly tossed.	**This has nothing to do with entropy** because macro objects are unable to exchange thermal energy with the surroundings within the time scale of the process
Two identical blocks of copper, one at 20°C and the other at 40°C, are placed in contact.	The cooler block contains more unoccupied microstates, so heat flows from the warmer block until equal numbers of microstates are populated in the two blocks.
A gas expands isothermally to twice its initial volume.	A constant amount of thermal energy spreads over a larger volume of space
1 mole of water is heated by 1C°.	The increased thermal energy makes additional microstates accessible. (The increase is by a factor of about $10^{20,000,000,000,000,\,000,000,000,000}$.)
Equal volumes of two gases are allowed to mix.	The effect is the same as allowing each gas to expand to twice its volume; the thermal energy in each is now spread over a larger volume.
One mole of dihydrogen, H_2, is placed in a container and heated to 3000K.	Some of the H_2 dissociates to H because at this temperature there are more thermally accessible microstates in the 2 moles of H.
The above reaction mixture is cooled to 300K.	The composition shifts back to virtually all H_2because this molecule contains more thermally accessible microstates at low temperatures.

Entropy is an *extensive* quantity; that is, it is proportional to the quantity of matter in a system; thus 100 g of metallic copper has twice the entropy of 50 g at the same temperature. This makes sense because the larger piece of copper contains twice as many quantized energy levels able to contain the thermal energy.

Thermal energy is the portion of a molecule's energy that is proportional to its *temperature*, and thus relates to motion at the molecular scale. What kinds of molecular motions are possible? For monatomic molecules, there is only one: actual movement from one location to another, which we call *translation*. Since there are three directions in space, all molecules possess three modes of *translational motion*.

For polyatomic molecules, two additional kinds of motions are possible. One of these is *rotation*; a linear molecule such as CO_2 in which the atoms are all laid out along the x-axis can rotate along the y- and z-axes, while molecules having less symmetry can rotate about all three axes. Thus linear molecules possess two modes of rotational motion, while non-linear ones have three

rotational modes. Finally, molecules consisting of two or more atoms can undergo internal *vibrations*. For freely moving molecules in a gas, the number of vibrational modes or patterns depends on both the number of atoms and the shape of the molecule, and it increases rapidly as the molecule becomes more complicated.

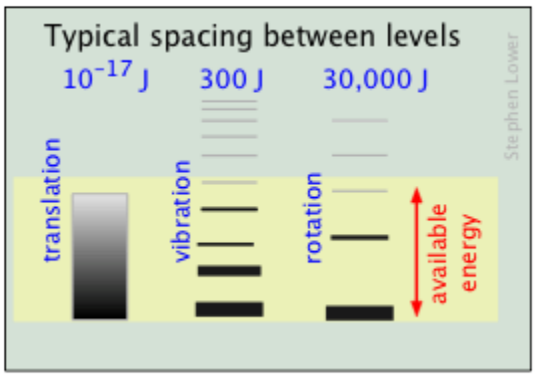


Figure 15.8.1: Relative Energy levels for translation (left), rotation (middle) and vibration (right)

The relative populations of the quantized translational, rotational and vibrational energy states of a typical diatomic molecule are depicted by the thickness of the lines in this schematic (not-to-scale!) diagram. The colored shading indicates the total thermal energy available at an arbitrary temperature. The numbers at the top show order-of-magnitude spacings between adjacent levels.

Notice that the spacing between the quantized translational levels is so minute that they can be considered nearly continuous. This means that at all temperatures, the thermal energy of a collection of molecules resides almost exclusively in translational microstates. At ordinary temperatures (around 25° C), most of the molecules are in their zero-level vibrational and rotational states (corresponding to the bottom-most bars in the diagram.) The prevalence of vibrational states is so overwhelming that we can effectively equate the thermal energy of molecules with translational motions alone.

The number of ways in which thermal energy can be distributed amongst the allowed states within a collection of molecules is easily calculated from simple statistics. A very important point to bear in mind is that the number of discrete microstates that can be populated by an arbitrary quantity of energy depends on the spacing of the states. As a very simple example, suppose that we have two molecules (depicted by the orange dots) in a system total available thermal energy is indicated by the yellow shading.

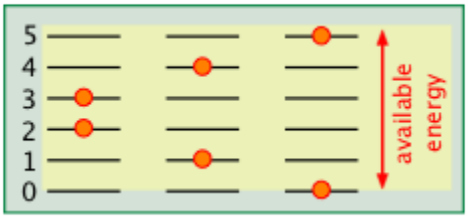


Figure 15.8.2*A*

In the system with the more closely-spaced energy levels, there are three possible microstates, while in the one with the more widely-spaced levels, only two possibilities are available.

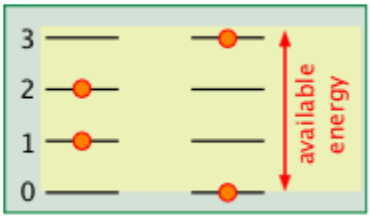


Figure 15.8.2*B*

The spacing of molecular energy states becomes closer as the mass and number of bonds in the molecule increases, so we can generally say that the more complex the molecule, the greater the density of its energy states.

At the atomic and molecular level, all energy is *quantized*; each particle possesses discrete states of kinetic energy and is

able to accept thermal energy only in packets whose values correspond to the energies of one or more of these states. Polyatomic molecules can store energy in rotational and vibrational motions, and all molecules (even monatomic ones) will possess translational kinetic energy (thermal energy) at all temperatures above absolute zero. The energy difference between adjacent translational states is so minute that translational kinetic energy can be regarded as *continuous* (non-quantized) for most practical purposes.

The number of ways in which thermal energy can be distributed amongst the allowed states within a collection of molecules is easily calculated from simple statistics, but we will confine ourselves to an example here. Suppose that we have a system consisting of three molecules and three quanta of energy to share among them. We can give all the kinetic energy to any one molecule, leaving the others with none, we can give two units to one molecule and one unit to another, or we can share out the energy equally and give one unit to each molecule. All told, there are ten possible ways of distributing three units of energy among three identical molecules as shown here:

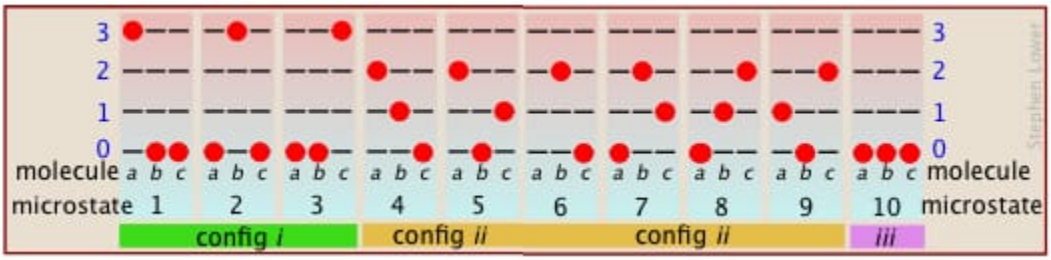


Figure 15.8.3

Each of these ten possibilities represents a distinct microstate that will describe the system at any instant in time. Those microstates that possess identical distributions of energy among the accessible quantum levels (and differ only in which particular molecules occupy the levels) are known as **configurations**. Because all microstates are equally probable, the probability of any one configuration is proportional to the number of microstates that can produce it. Thus in the system shown above, the configuration labeled *ii* will be observed 60% of the time, while *iii* will occur only 10% of the time.

As the number of molecules and the number of quanta increases, the number of accessible microstates grows explosively; if 1000 quanta of energy are shared by 1000 molecules, the number of available microstates will be around 10^{600}— a number that greatly exceeds the number of atoms in the observable universe! The number of possible configurations (as defined above) also increases, but in such a way as to greatly reduce the probability of all but the most probable configurations. Thus for a sample of a gas large enough to be observable under normal conditions, only a single configuration (energy distribution amongst the quantum states) need be considered; even the second-most-probable configuration can be neglected.

The bottom line: any collection of molecules large enough in numbers to have chemical significance will have its therrmal energy distributed over an unimaginably large number of microstates. The number of microstates increases exponentially as more energy states ("configurations" as defined above) become accessible owing to

- Addition of energy quanta (higher temperature),
- Increase in the number of molecules (resulting from dissociation, for example).
- the volume of the system increases (which decreases the spacing between energy states, allowing more of them to be populated at a given temperature.)

Using Entropy to Understand Spontaneous Processes

Heat Death

Energy is *conserved*; if you lift a book off the table, and let it fall, the total amount of energy in the world remains unchanged. All you have done is transferred it from the form in which it was stored within the glucose in your body to your muscles, and then to the book (that is, you did work on the book by moving it up against the earth's gravitational field). After the book has fallen, this same quantity of energy exists as thermal energy (heat) in the book and table top.

What *has* changed, however, is the availability of this energy. Once the energy has spread into the huge number of thermal microstates in the warmed objects, the probability of its spontaneously (that is, by chance) becoming un-dispersed is essentially zero. Thus although the energy is still "there", it is forever beyond utilization or recovery. The profundity of this conclusion was

recognized around 1900, when it was first described at the “heat death” of the world. This refers to the fact that every spontaneous process (essentially every change that occurs) is accompanied by the “dilution” of energy. The obvious implication is that all of the molecular-level kinetic energy will be spread out completely, and nothing more will ever happen.

Gases Expansions

Everybody knows that a gas, if left to itself, will tend to expand and fill the volume within which it is confined completely and uniformly. What “drives” this expansion? At the simplest level it is clear that with more space available, random motions of the individual molecules will inevitably disperse them throughout the space. But as we mentioned above, the allowed energy states that molecules can occupy are spaced more closely in a larger volume than in a smaller one. The larger the volume available to the gas, the greater the number of microstates its thermal energy can occupy. Since all such states within the thermally accessible range of energies are equally probable, the expansion of the gas can be viewed as a consequence of the tendency of thermal energy to be spread and shared as widely as possible. Once this has happened, the probability that this sharing of energy will reverse itself (that is, that the gas will spontaneously contract) is so minute as to be unthinkable.

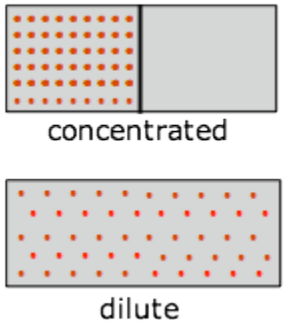


Figure 15.8.4

Imagine a gas initially confined to one half of a box (Figure 15.8.4). The barrier is then removed so that it can expand into the full volume of the container. We know that the entropy of the gas will increase as the thermal energy of its molecules spreads into the enlarged space. In terms of the spreading of thermal energy, Figure 15.8.5 may be helpful. The tendency of a gas to expand is due to the more closely-spaced thermal energy states in the larger volume (2).

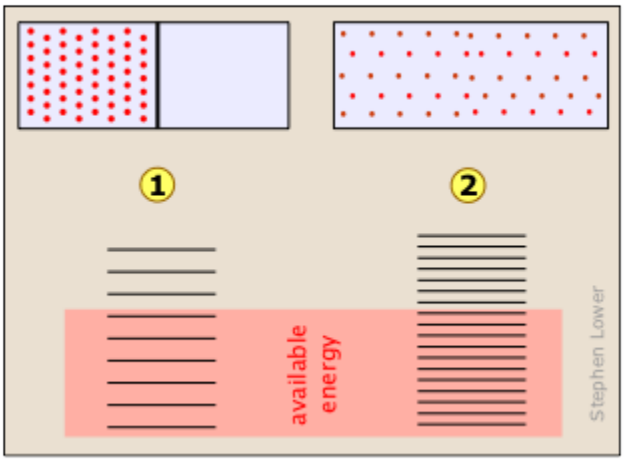


Figure 15.8.5: The pink shading represents the thermally-accessible range of microstates for a given temperature.

Mixing and dilution

Mixing and dilution really amount to the same thing, especially for idea gases. Replace the pair of containers shown above with one containing two kinds of molecules in the separate sections (Figure 15.8.6). When we remove the barrier, the "red" and "blue" molecules will each expand into the space of the other. (Recall Dalton's Law that "each gas is a vacuum to the other gas".) However, notice that although each gas underwent an expansion, the overall process amounts to what we call "mixing".

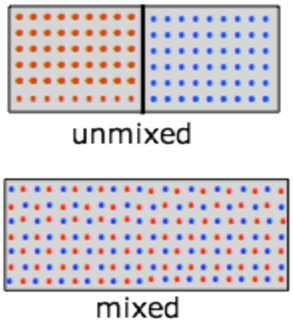
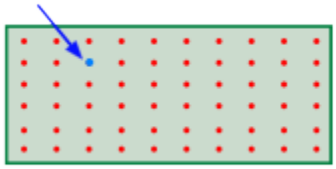


Figure 15.8.6: Mixing of a gas

What is true for gaseous molecules can, in principle, apply also to solute molecules dissolved in a solvent. But bear in mind that whereas the enthalpy associated with the expansion of a perfect gas is by definition zero, ΔH's of mixing of two liquids or of dissolving a solute in a solvent have finite values which may limit the miscibility of liquids or the solubility of a solute.

It's unfortunate the the simplified diagrams we are using to illustrate the greater numbers of energetically accessible microstates in an expanded gas or a mixture of gases fail to convey the immensity of this increase. Only by working through the statistical mathematics of these processes (beyond the scope of first-year Chemistry!) can one gain an appreciation of the magnitude of the probabilities of these spontaneous processes.



It turns out that when just *one molecule* of a second gas is inroduced into the container of another gas, an unimaginably huge number of new configurations becom available. This happens because the added molecule (indicated by the blue arrow in the diagram) can in principle replace any one of the old (red) ones, each case giving rise to a new microstate.

Heat Flow

ust as gases spontaneously change their volumes from “smaller-to-larger”, the flow of heat from a warmer body to a cooler one always operates in the direction “warmer-to-cooler” because this allows thermal energy to populate a larger number of energy microstates as new ones are made available by bringing the cooler body into contact with the warmer one; in effect, the thermal energy becomes more “diluted”. In this simplified schematic diagram, the "cold" and "hot" bodies differ in the numbers of translational microstates that are occupied, as indicated by the shading. When they are brought into thermal contact, a hugely greater number of microstates are created, as is indicated by their closer spacing in the rightmost section of the diagram, which represents the combined bodies in thermal equilibrium. The thermal energy in the initial two bodies fills these new microstates to a level (and thus, temperature) that is somewhere between those of the two original bodies.

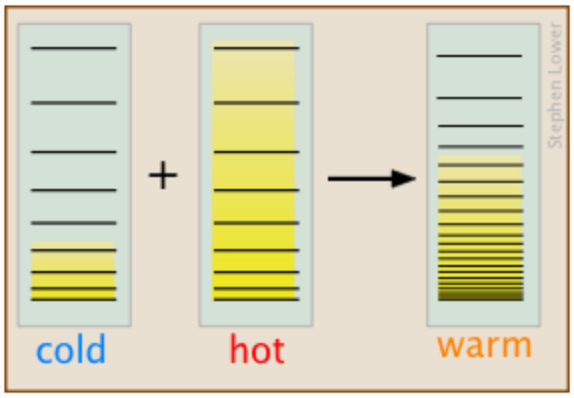


Figure 15.8.7: Part a of the figure is a schematic depiction of the thermal energy states in two separated identical bodies at different temperatures (indicated by shading.)

Note that this explanation applies equally well to the case of two solids brought into thermal contact, or two the mixing of two fluids having different temperatures.

As you might expect, the increase in the amount of energy spreading and sharing is proportional to the amount of heat transferred *q*, but there is one other factor involved, and that is the *temperature* at which the transfer occurs. When a quantity of heat *q* passes into a system at temperature *T*, the degree of dilution of the thermal energy is given by

$$\frac{q}{T} \tag{15.8.1}$$

To understand why we have to divide by the temperature, consider the effect of very large and very small values of T in the denominator. If the body receiving the heat is initially at a very low temperature, relatively few thermal energy states are initially occupied, so the amount of energy spreading into vacant states can be very great. Conversely, if the temperature is initially large, more thermal energy is already spread around within it, and absorption of the additional energy will have a relatively small effect on the degree of thermal disorder within the body.

Chemical Reactions and Equilibrium

When a chemical reaction takes place, two kinds of changes relating to thermal energy are involved:

1. The ways that thermal energy can be stored within the reactants will generally be different from those for the products. For example, in the reaction $H_2 \rightarrow$ 2 H, the reactant dihydrogen possesses vibrational and rotational energy states, while the atomic hydrogen in the product has translational states only— but the total number of translational states in two moles of H is twice as great as in one mole of H_2. Because of their extremely close spacing, translational states are the only ones that really count at ordinary temperatures, so we can say that thermal energy can become twice as diluted ("spread out") in the product than in the reactant. If this were the only factor to consider, then dissociation of dihydrogen would always be spontaneous and this molecule would not exist.
2. In order for this dissociation to occur, however, a quantity of thermal energy (heat) \($q = \Delta U$\) must be taken up from the surroundings in order to break the H–H bond. In other words, the ground state (the energy at which the manifold of energy states begins) is higher in H, as indicated by the vertical displacement of the right half in each of the four panels below.

In Figure 15.8.8 are schematic representations of the translational energy levels of the two components H and H_2 of the hydrogen dissociation reaction. The shading shows how the relative populations of occupied microstates vary with the temperature, causing the equilibrium composition to change in favor of the dissociation product.

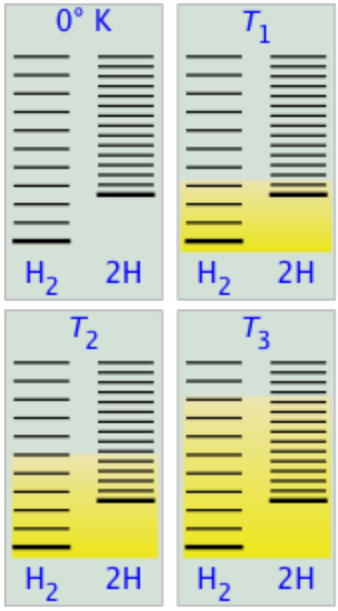


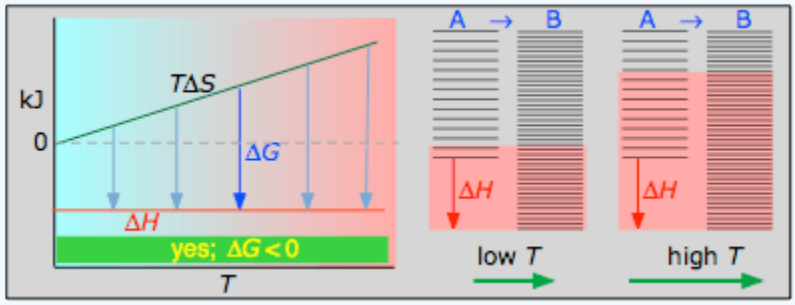
Figure 15.8.8:

The ability of energy to spread into the product molecules is constrained by the availability of sufficient thermal energy to produce these molecules. This is where the temperature comes in. At absolute zero the situation is very simple; no thermal energy is available to bring about dissociation, so the only component present will be dihydrogen.

- As the temperature increases, the number of populated energy states rises, as indicated by the shading in the diagram. At temperature T_1, the number of populated states of H_2 is greater than that of 2H, so some of the latter will be present in the equilibrium mixture, but only as the minority component.
- At some temperature T_2 the numbers of populated states in the two components of the reaction system will be identical, so the equilibrium mixture will contain H_2 and "2H" in equal amounts; that is, the mole ratio of H_2/H will be 1:2.
- As the temperature rises to T_3 and above, we see that the number of energy states that are thermally accessible in the product begins to exceed that for the reactant, thus favoring dissociation.

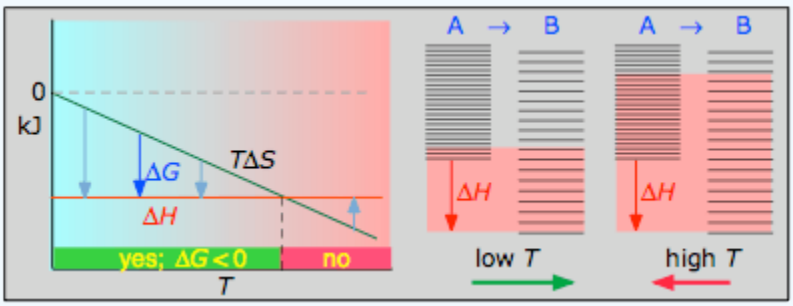
The result is exactly what the Le Chatelier Principle predicts: the equilibrium state for an endothermic reaction is shifted to the right at higher temperatures.

The following table generalizes these relations for the four sign-combinations of ΔH and ΔS. (Note that use of the standard ΔH° and ΔS° values in the example reactions is not strictly correct here, and can yield misleading results when used generally.)

> 0	
	

This *combustion reaction*, like most such reactions, is **spontaneous at all temperatures**. The positive entropy change is due mainly to the greater mass of CO_2 molecules compared to those of O_2.

< 0

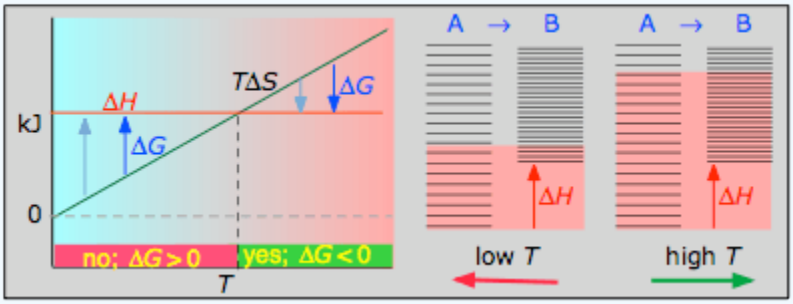


$$3\,H_2 + N_2 \longrightarrow 2\,NH_3\,(g) \tag{15.8.2}$$

- $\Delta H°$ = –46.2 kJ
- $\Delta S°$ = –389 J K^{-1}
- $\Delta G°$ = –16.4 kJ at 298 K

The decrease in moles of gas in the **Haber ammonia synthesis** drives the entropy change negative, making the reaction **spontaneous only at low temperatures**. Thus higher *T*, which speeds up the reaction, also reduces its extent.

> 0

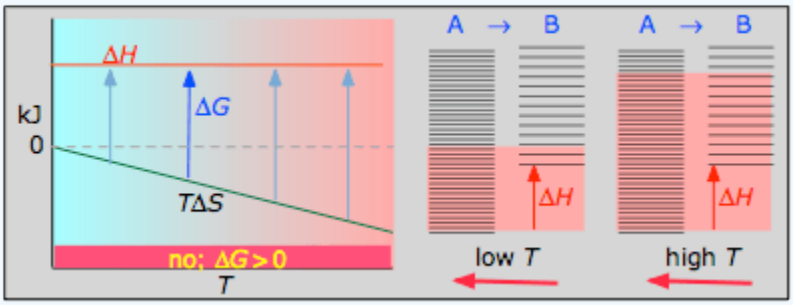


$$N_2O_4\,(g) \longrightarrow 2\,NO_2\,(g) \tag{15.8.3}$$

- $\Delta H°$ = 55.3 kJ
- $\Delta S°$ = +176 J K^{-1}
- $\Delta G°$ = +2.8 kJ at 298 K

Dissociation reactions are typically endothermic with positive entropy change, and are therefore **spontaneous at high temperatures**.Ultimately, all molecules decompose to their atoms at sufficiently high temperatures.

< 0



$$\tfrac{1}{2}\,N_2 + O_2 \longrightarrow NO_2\,(g) \tag{15.8.4}$$

- $\Delta H°$ = 33.2 kJ
- $\Delta S°$ = –249 J K^{-1}
- $\Delta G°$ = +51.3 kJ at 298 K

This reaction is **not spontaneous at any temperature**, meaning that *its reverse is always spontaneous*. But because the reverse reaction is kinetically inhibited, NO_2 can exist indefinitely at ordinary temperatures even though it is thermodynamically unstable.

Phase Changes

Everybody knows that the **solid** is the stable form of a substance at low temperatures, while the gaseous state prevails at high temperatures. Why should this be? The diagram in Figure 15.8.9 shows that

1. the density of energy states is smallest in the solid and greatest (much, much greater) in the gas, and
2. the ground states of the liquid and gas are offset from that of the previous state by the heats of fusion and vaporization, respectively.

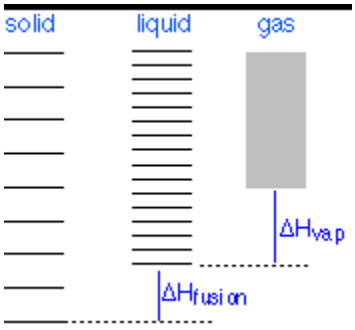


Figure 15.8.9:

Changes of phase involve exchange of energy with the surroundings (whose energy content relative to the system is indicated (with much exaggeration!) by the height of the yellow vertical bars in Figure 15.8.13. When solid and liquid are in equilibrium (middle section of diagram below), there is sufficient thermal energy (indicated by pink shading) to populate the energy states of both phases. If heat is allowed to flow into the surroundings, it is withdrawn selectively from the more abundantly populated levels of the liquid phase, causing the quantity of this phase to decrease in favor of the solid. The temperature remains constant as the heat of fusion is returned to the system in exact compensation for the heat lost to the surroundings. Finally, after the last trace of liquid has disappeared, the only states remaining are those of the solid. Any further withdrawal of heat results in a temperature drop as the states of the solid become depopulated.

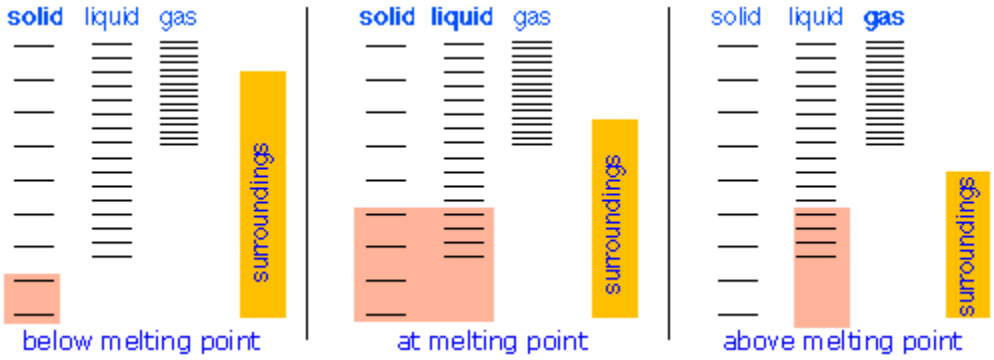


Figure 15.8.10:

Colligative Properties of Solutions

Vapor pressure lowering, boiling point elevation, freezing point depression and osmosis are well-known phenomena that occur when a non-volatile solute such as sugar or a salt is dissolved in a volatile solvent such as water. All these effects result from

"dilution" of the solvent by the added solute, and because of this commonality they are referred to as *colligative properties* (Lat. *co ligare*, connected to.) The key role of the solvent concentration is obscured by the greatly-simplified expressions used to calculate the magnitude of these effects, in which only the solute concentration appears. The details of how to carry out these calculations and the many important applications of colligative properties are covered elsewhere. Our purpose here is to offer a more complete explanation of *why* these phenomena occur.

Basically, these all result from the effect of **dilution of the solvent** on its entropy, and thus in the increase in the density of energy states of the system in the solution compared to that in the pure liquid. Equilibrium between two phases (liquid-gas for boiling and solid-liquid for freezing) occurs when the energy states in each phase can be populated at **equal densities**. The temperatures at which this occurs are depicted by the shading.

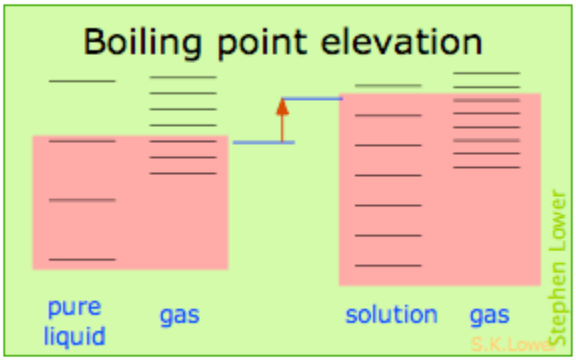


Figure 15.8.11: Boiling Point Elevation from an entropy perspective

Dilution of the solvent adds new energy states to the liquid, but does not affect the vapor phase. This raises the temperature required to make equal numbers of microstates accessible in the two phases.

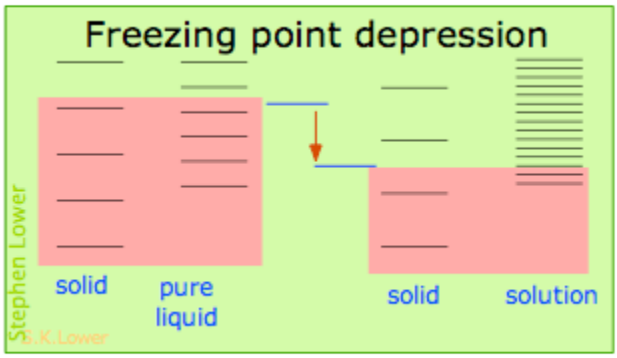


Figure 15.8.12: Freezing Point Depression from an entropy perspective

Dilution of the solvent adds new energy states to the liquid, but does not affect the solid phase. This reduces the temperature required to make equal numbers of states accessible in the two phases.

Osmotic Pressure: Effects of Pressure on the Entropy

When a liquid is subjected to hydrostatic pressure— for example, by an inert, non-dissolving gas that occupies the vapor space above the surface, the vapor pressure of the liquid is raised (Figure 15.8.13). The pressure acts to compress the liquid very slightly, effectively narrowing the potential energy well in which the individual molecules reside and thus increasing their tendency to escape from the liquid phase. (Because liquids are not very compressible, the effect is quite small; a 100-atm applied pressure will raise the vapor pressure of water at 25°C by only about 2 torr.) In terms of the entropy, we can say that the applied pressure reduces the dimensions of the "box" within which the principal translational motions of the molecules are confined within the liquid, thus reducing the density of energy states in the liquid phase.

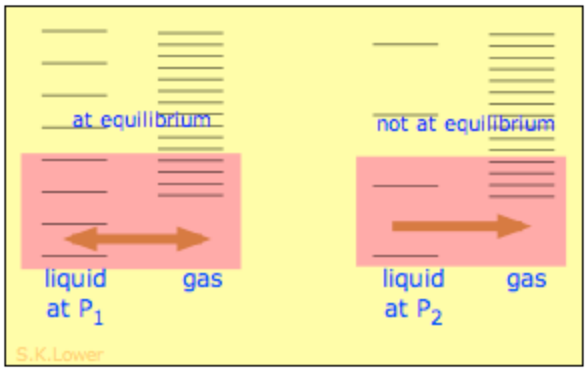


Figure 15.8.13:

Applying hydrostatic pressure to a liquid increases the spacing of its microstates, so that the number of energetically accessible states in the gas, although unchanged, is relatively greater— thus increasing the tendency of molecules to escape into the vapor phase. In terms of free energy, the higher pressure raises the free energy of the liquid, but does not affect that of the gas phase.

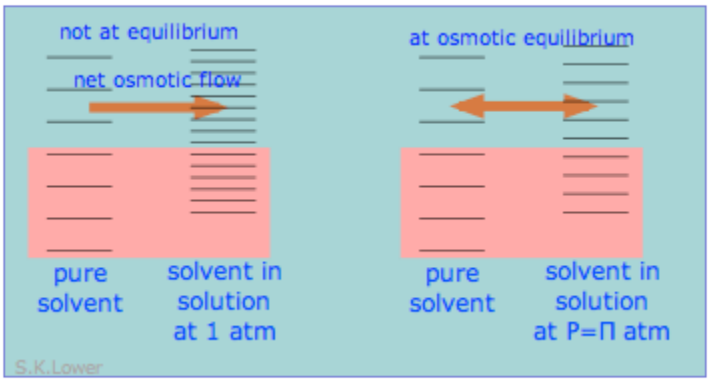


Figure 15.8.14: Osmotic Pressure from an entropy perspective

This phenomenon can explain **osmotic pressure**. Osmotic pressure, students must be reminded, is not what drives osmosis, but is rather the hydrostatic pressure that must be applied to the more concentrated solution (more dilute solvent) in order to *stop* osmotic flow of solvent into the solution. The effect of this pressure Π is to slightly increase the spacing of solvent energy states on the high-pressure (dilute-solvent) side of the membrane to match that of the pure solvent, restoring osmotic equilibrium.

This page titled 15.8: Quantum states, Microstates, and Energy spreading in Reactions is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by LibreTexts.

16: ELECTROCHEMISTRY

Electrochemistry is the study of electricity and how it relates to chemical reactions. In electrochemistry, electricity can be generated by movements of electrons from one element to another in a reaction known as redox reaction, or oxidation-reduction reaction.

16.1: Chemistry and Electricity

16.2: Galvanic cells and Electrodes

16.3: Cell Potentials and Thermodynamics

16.4: The Nernst Equation

16.5: Applications of the Nernst Equation

16.6: Batteries and Fuel Cells

16.7: Timeline of Battery Development

16.8: Electrochemical Corrosion

16.9: Corrosion Gallery

16.10: Electrolytic Cells and Electrolysis

This page titled 16: Electrochemistry is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.1: Chemistry and Electricity

Learning Objectives

Make sure you thoroughly understand the following essential ideas. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- **Electroneutrality principle** - Bulk matter cannot have a chemically-significant unbalance of positive and negative ions.
- Dissolution of a metal in water can proceed to a measurable extent only if some means is provided for removing the excess negative charge that remains. This can be by electron-acceptor ions in solution, or by drawing electrons out of the metal through an external circuit.
- Interfacial potentials - these exist at all phase boundaries. In the case of a metal in contact with an electrolyte solution, the interfacial region consists of an electric double layer.
- The potential difference between a metal and the solution is almost entirely located across the very thin double layer, leading to extremely large potential gradients in this region.

The connection between chemistry and electricity is a very old one, going back to ALESSANDRO VOLTA'S discovery, in 1793, that electricity could be produced by placing two dissimilar metals on opposite sides of a moistened paper.

Electroneutrality

Nature seems to strongly discourage any process that would lead to an excess of positive or negative charge in matter. Suppose, for example, that we immerse a piece of zinc metal in pure water. A small number of zinc atoms go into solution as Zn^{2+} ions, leaving their electrons behind in the metal:

$$Zn(s) \rightarrow Zn^{2+} + 2e^{-} \tag{16.1.1}$$

As this process goes on, the electrons which remain in the zinc cause a negative charge to build up within the metal which makes it increasingly difficult for additional positive ions to leave the metallic phase. A similar buildup of positive charge in the liquid phase adds to this inhibition. Very soon, therefore, the process comes to a halt, resulting in a solution in which the concentration of Zn^{2+} is still too low (around 10^{-10} *M*) to be detected by ordinary chemical means.

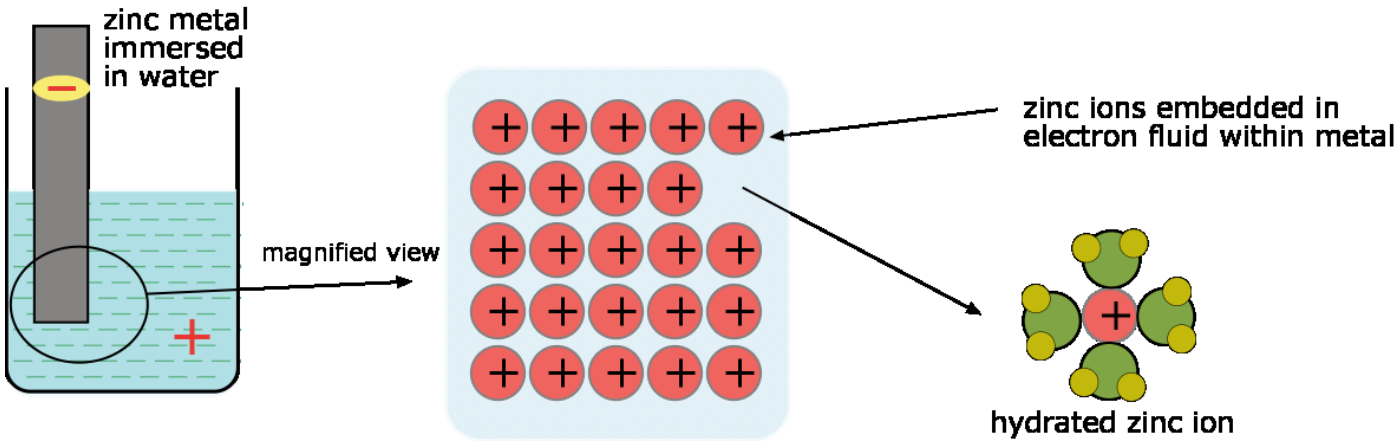


Figure 16.1.1. Transport of zinc ions from the metal to water; the build-up of negative charge in the metal (and positive charge in the solution) soon brings the process to a halt.

There would be no build-up of this opposing charge in the two phases if the excess electrons could be removed from the metal or the positive ions consumed as the electrode reaction proceeds. For example, we could drain off the electrons left behind in the zinc through an external circuit that forms part of a complete *electrochemical cell*; this we will describe later. Another way to remove these same electrons is to bring a good electron acceptor (that is, an *oxidizing agent*) into contact with the electrode. A

suitable acceptor would be hydrogen ions; this is why acids attack many metals. For the very active metals such as sodium, water itself is a sufficiently good electron acceptor.

The degree of charge unbalance that is allowed produces differences in electric potential of no more than a few volts, and corresponds to unbalances in the concentrations of oppositely charged particles that are not chemically significant. There is nothing mysterious about this prohibition, known as the electroneutrality principle; it is a simple consequence of the thermodynamic work required to separate opposite charges, or to bring like charges into closer contact. The additional work raises the free energy change of the process, making it less spontaneous.

The only way we can get the oxidation of the metal to continue is to couple it with some other process that restores electroneutrality to the two phases. A simple way to accomplish this would be to immerse the zinc in a solution of copper sulfate instead of pure water. The zinc metal quickly becomes covered with a black coating of finely-divided metallic copper. The reaction is a simple **reduction-oxidation reaction** involving a transfer of two electrons from the zinc to the copper:

$$Zn(s) \rightarrow Zn^{2+} + 2e^{-} \tag{16.1.2}$$

and

$$Cu^{2+} + 2e^{-} \rightarrow Cu(s) \tag{16.1.3}$$

The dissolution of the zinc is no longer inhibited by a buildup of negative charge in the metal, because the excess electrons are removed from the zinc by copper ions that come into contact with it. At the same time, the solution remains electrically neutral, since for each Zn ion introduced to the solution, one Cu ion is removed. The net reaction

$$Zn(s) + Cu^{2+} \rightarrow Zn^{2+} + Cu(s) \tag{16.1.4}$$

quickly goes to completion.

Potential differences at interfaces

The transition region between two phases consists of a region of charge unbalance known as the *electric double layer*. As its name implies, this consists of an inner monomolecular layer of adsorbed water molecules and ions, and an outer diffuse region that compensates for any local charge unbalance that gradually merges into the completely random arrangement of the bulk solution. In the case of a metal immersed in pure water, the electron fluid within the metal causes the polar water molecules to adsorb to the surface and orient themselves so as to create two thin planes of positive and negative charge. If the water contains dissolved ions, some of the larger (and more polarizable) anions will loosely bond (*chemisorb*) to the metal, creating a negative inner layer which is compensated by an excess of cations in the outer layer.

Electrochemistry is the study of reactions in which charged particles (ions or electrons) cross the interface between two phases of matter, typically a metallic phase (the *electrode*) and a conductive solution, or *electrolyte*. A process of this kind can always be represented as a chemical reaction and is known generally as an *electrode process*. Electrode processes (also called *electrode reactions*) take place within the double layer and produce a slight unbalance in the electric charges of the electrode and the solution. Much of the importance of electrochemistry lies in the ways that these potential differences can be related to the thermodynamics and kinetics of electrode reactions. In particular, manipulation of the interfacial potential difference affords an important way of exerting external control on an electrode reaction.

The interfacial potential differences which develop in electrode-solution systems are limited to only a few volts at most. This may not seem like very much until you consider that this potential difference spans a very small distance. In the case of an electrode immersed in a solution, this distance corresponds to the thin layer of water molecules and ions that attach themselves to the electrode surface, normally only a few atomic diameters. Thus a very small voltage can produce a very large potential gradient. For example, a potential difference of one volt across a typical 10^{-8} cm interfacial boundary amounts to a potential gradient of 100 million volts per centimeter— a very significant value indeed!

Interfacial potentials are not confined to metallic electrodes immersed in solutions; they can in fact exist between any two phases in contact, even in the absence of chemical reactions. In many forms of matter, they are the result of adsorption or ordered alignment of molecules caused by non-uniform forces in the interfacial region. Thus colloidal particles in aqueous suspensions selectively adsorb a given kind of ion, positive for some colloids, and negative for others. The resulting net electric charge prevents the particles from coming together and coalescing, which they would otherwise tend to do under the influence of

ordinary van der Waals attractions.

Interfacial potential differences are not directly observable. The usual way of measuring a potential difference between two points is to bring the two leads of a voltmeter into contact with them. It's simple enough to touch one lead of the meter to a metallic electrode, but there is no way you can connect the other lead to the solution side of the interfacial region without introducing a second electrode with its own interfacial potential, so you would be measuring the sum of two potential differences. Thus s*ingle electrode potentials*, as they are commonly known, are not directly observable.

What we *can* observe, and make much use of, are potential differences between *pair*s of electrodes in *electrochemical cells*. This is the topic of the next page in this series.

Summary

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

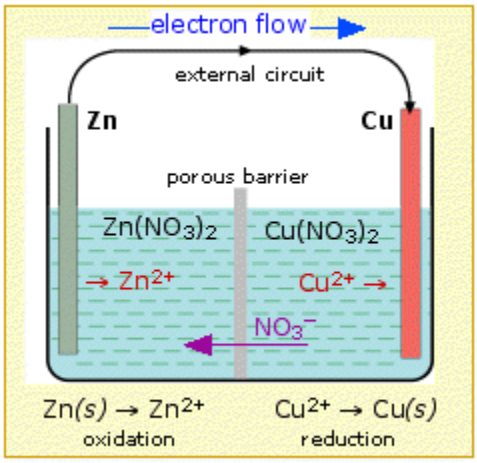
- **Electroneutrality principle** - Bulk matter cannot have a chemically-significant unbalance of positive and negative ions.
- Dissolution of a metal in water can proceed to a measurable extent only if some means is provided for removing the excess negative charge that remains. This can be by electron-acceptor ions in solution, or by drawing electrons out of the metal through an external circuit.
- Interfacial potentials - these exist at all phase boundaries. In the case of a metal in contact with an electrolyte solution, the interfacial region consists of an electric double layer.
- The potential difference between a metal and the solution is almost entirely located across the very thin double layer, leading to extremely large potential gradients in this region.

This page titled 16.1: Chemistry and Electricity is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.2: Galvanic cells and Electrodes

It is physically impossible to measure the potential difference between a piece of metal and the solution in which it is immersed. We can, however, measure the *difference* between the potentials of two electrodes that dip into the same solution, or more usefully, are in two different solutions. In the latter case, each electrode-solution pair constitutes an oxidation-reduction *half cell*, and we are measuring the sum of the two *half-cell potentials*.

This arrangement is called a **galvanic cell**. A typical cell might consist of two pieces of metal, one zinc and the other copper, each immersed each in a solution containing a dissolved salt of the corresponding metal. The two solutions are separated by a porous barrier that prevents them from rapidly mixing but allows ions to diffuse through.



If we connect the zinc and copper by means of a metallic conductor, the excess electrons that remain when Zn^{2+} ions emerge from the zinc in the left cell would be able to flow through the external circuit and into the right electrode, where they could be delivered to the Cu^{2+} ions which become "discharged", that is, converted into Cu atoms at the surface of the copper electrode. The net reaction is the oxidation of zinc by copper(II) ions:

$$Zn_{(s)} + Cu^{2+} \rightarrow Zn^{2+} + Cu_{(s)} \tag{16.2.1}$$

but this time, the oxidation and reduction steps (half reactions) take place in separate locations:

left electrode:	$Zn(s) \rightarrow Zn^{2+} + 2e^-$	*oxidation*
right electrode:	$Cu^{2+} + 2e^- \rightarrow Cu(s)$	*reduction*

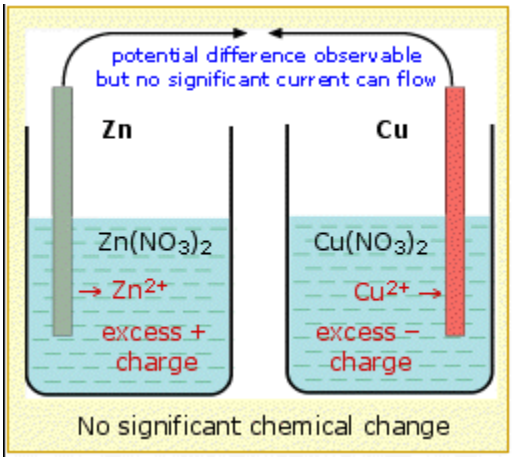
Electrochemical cells allow measurement and control of a redox reaction

The reaction can be started and stopped by connecting or disconnecting the two electrodes. If we place a variable resistance in the circuit, we can even control the rate of the net cell reaction by simply turning a knob. By connecting a battery or other source of current to the two electrodes, we can force the reaction to proceed in its non-spontaneous, or reverse direction. By placing an ammeter in the external circuit, we can measure the amount of electric charge that passes through the electrodes, and thus the number of moles of reactants that get transformed into products in the cell reaction.

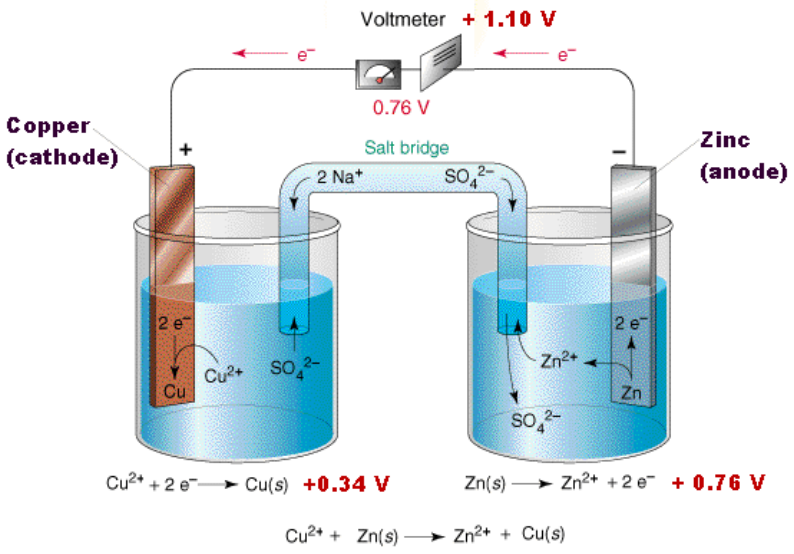
Electric charge q is measured in coulombs. The amount of charge carried by one mole of electrons is known as the **Faraday**, which we denote by F. Careful experiments have determined that 1 F = 96467 C. For most purposes, you can simply use 96,500 Coulombs as the value of the faraday. When we measure electric current, we are measuring the rate at which electric charge is transported through the circuit. A current of one ampere corresponds to the flow of one coulomb per second.

Charge Transport within the Cell

For the cell to operate, not only must there be an external electrical circuit between the two electrodes, but the two electrolytes (the solutions) must be in contact. The need for this can be understood by considering what would happen if the two solutions were physically separated. Positive charge (in the form of Zn^{2+}) is added to the electrolyte in the left compartment, and removed (as Cu^{2+}) from the right side, causing the solution in contact with the zinc to acquire a net positive charge, while a net negative charge would build up in the solution on the copper side of the cell. These violations of *electroneutrality* would make it more difficult (require more work) to introduce additional Zn^{2+} ions into the positively-charged electrolyte or for electrons to flow into right compartment where they are needed to reduce the Cu^{2+} ions, thus effectively stopping the reaction after only a chemically insignificant amount has taken place.



In order to sustain the cell reaction, the charge carried by the electrons through the external circuit must be accompanied by a compensating transport of ions between the two cells. This means that we must provide a path for ions to move directly from one cell to the other. This ionic transport involves not only the electroactive species Cu^{2+} and Zn^{2+}, but also the *counterions*, which in this example are nitrate, NO_3^-. Thus an excess of Cu^{2+} in the left compartment could be alleviated by the drift of these ions into the right side, or equally well by diffusion of nitrate ions to the left. More detailed studies reveal that both processes occur, and that the relative amounts of charge carried through the solution by positive and negative ions depends on their relative *mobilities*, which express the velocity with which the ions are able to make their way through the solution. Since negative ions tend to be larger than positive ions, the latter tend to have higher mobilities and carry the larger fraction of charge.



In the simplest cells, the barrier between the two solutions can be a porous membrane, but for precise measurements, a more complicated arrangement, known as a *salt bridge*, is used. The salt bridge consists of an intermediate compartment filled with a concentrated solution of KCl and fitted with porous barriers at each end. The purpose of the salt bridge is to minimize the natural potential difference, known as the *junction potential*, that develops (as mentioned in the previous section) when any two phases (such as the two solutions) are in contact. This potential difference would combine with the two half-cell potentials so as introduce a degree of uncertainty into any measurement of the cell potential. With the salt bridge, we have two liquid junction potentials instead of one, but they tend to cancel each other out.

Cell description conventions

In order to make it easier to describe a given electrochemical cell, a special symbolic notation has been adopted. In this notation the cell we described above would be

$$Zn(s) \mid Zn^{2+}(aq) \parallel Cu^{2+}(aq) \mid Cu(s)$$

There are several other conventions relating to cell notation and nomenclature that you are expected to know:

- The **anode** is where oxidation occurs, and the **cathode** is the site of reduction. In an actual cell, the identity of the electrodes depends on the direction in which the net cell reaction is occurring.
- If electrons flow from the left electrode to the right electrode (as depicted in the above cell notation) when the cell operates in its spontaneous direction, the potential of the right electrode will be higher than that of the left, and the cell potential will be positive.
- "Conventional current flow" is from positive to negative, which is opposite to the direction of the electron flow. This means that if the electrons are flowing from the left electrode to the right, a galvanometer placed in the external circuit would indicate a current flow from right to left.

Electrodes and Electrode Reactions

An *electrode reaction* refers to the net oxidation or reduction process that takes place at an electrode. This reaction may take place in a single electron-transfer step, or as a succession of two or more steps. The substances that receive and lose electrons are called the electroactive species.

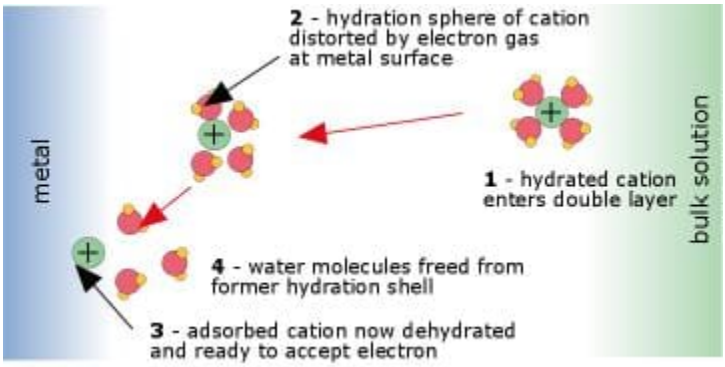


Fig. 4: Electron transfer at an anode

This process takes place within the very thin interfacial region at the electrode surface, and involves quantum-mechanical tunneling of electrons between the electrode and the electroactive species. The work required to displace the H_2O molecules in the hydration spheres of the ions constitutes part of the *activation energy* of the process.

In the example of the Zn/Cu cell we have been using, the electrode reaction involves a metal and its hydrated cation; we call such electrodes metal-metal ion electrodes. There are a number of other kinds of electrodes which are widely encountered in electrochemistry and analytical chemistry.

Ion-ion Electrodes

Many electrode reactions involve only ionic species, such as Fe^{2+} and Fe^{3+}. If neither of the electroactive species is a metal, some other metal must serve as a conduit for the supply or removal of electrons from the system. In order to avoid complications that would arise from electrode reactions involving this metal, a relatively inert substance such as platinum is commonly used. Such a half cell would be represented as

$$\text{Pt(s) | } Fe^{3+}(aq), Fe^{2+}(aq) \text{ || ...}$$

and the half-cell reaction would be

$$Fe^{2+}(aq) \rightarrow Fe^{3+}(aq) + e^- \tag{16.2.2}$$

The reaction occurs at the surface of the electrode (Fig 4 above). The electroactive ion diffuses to the electrode surface and adsorbs (attaches) to it by van der Waals and Coulombic forces. In doing so, the waters of hydration that are normally attached to any ionic species must be displaced. This process is always endothermic, sometimes to such an extent that only a small fraction of the ions be able to contact the surface closely enough to undergo electron transfer, and the reaction will be slow. The actual electron-transfer occurs by quantum-mechanical tunnelling.

Gas Electrodes

Some electrode reactions involve a gaseous species such as H_2, O_2, or Cl_2. Such reactions must also be carried out on the surface of an electrochemically inert conductor such as platinum. A typical reaction of considerable commercial importance is

$$Cl^-(aq) \rightarrow ½Cl_2(g) + e^- \tag{16.2.3}$$

Similar reactions involving the oxidation of Br_2 or I_2 also take place at platinum surfaces.

Insoluble–salt Electrodes

A typical electrode of this kind consists of a silver wire covered with a thin coating of silver chloride, which is insoluble in water. The electrode reaction consists in the oxidation and reduction of the silver:

$$AgCl(s) + e^- \rightarrow Ag(s) + Cl^-(aq) \tag{16.2.4}$$

The half cell would be represented as

$$...||Cl^-(aq)|AgCl(s)|Ag(s) \tag{16.2.5}$$

Although the usefulness of such an electrode may not be immediately apparent, this kind of electrode finds very wide application in electrochemical measurements, as we shall see later.

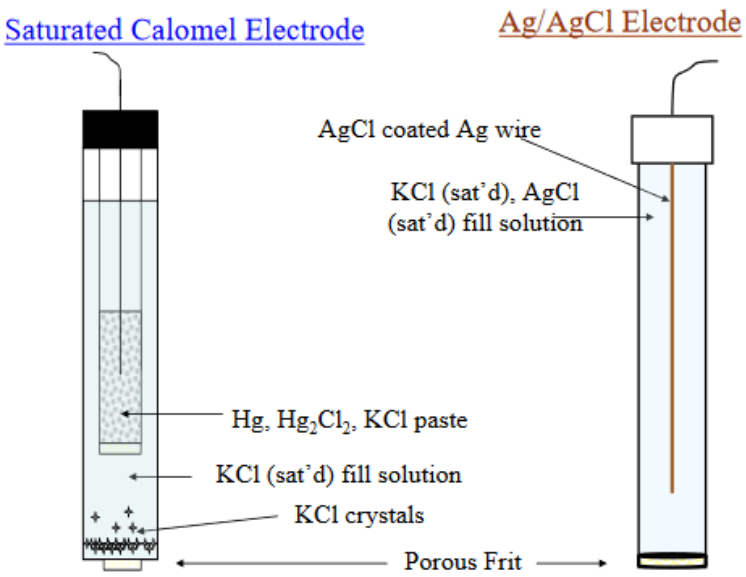
Reference Electrodes

In most electrochemical experiments our interest is concentrated on only one of the electrode reactions. Since all measurements must be on a complete cell involving two electrode systems, it is common practice to employ a *reference electrode* as the other half of the cell. The major requirements of a reference electrode are that it be easy to prepare and maintain, and that its potential be stable. The last requirement essentially means that the concentration of any ionic species involved in the electrode reaction must be held at a fixed value. The most common way of accomplishing this is to use an electrode reaction involving a saturated solution of an insoluble salt of the ion. One such system, the silver-silver chloride electrode has already been mentioned:

$$Ag|AgCl(s)|Cl^-(aq)||... \tag{16.2.6}$$

$$Ag(s) + Cl^-(aq) \rightarrow AgCl(s) + e^- \tag{16.2.7}$$

This electrode usually takes the form of a piece of silver wire coated with AgCl. The coating is done by making the silver the anode in an electrolytic cell containing HCl; the Ag^+ ions combine with Cl^- ions as fast as they are formed at the silver surface.



The other common reference electrode is the *calomel electrode*; calomel is the common name for mercury(I) chloride. Such a half cell would be represented as

$$Hg|Hg^{2+}(aq)|KCl||... \tag{16.2.8}$$

and the half-cell reaction would be

$$Hg(l) + Cl^{-} \rightarrow ½HgCl2(s) + e^{-} \tag{16.2.9}$$

The potentials of both of these electrodes have been very accurately determined against the hydrogen electrode. The latter is seldom used in routine electrochemical measurements because it is more difficult to prepare; the platinum surface has to be specially treated by preliminary electrolysis. Also, there is need for a supply of hydrogen gas which makes it somewhat cumbersome and hazardous.

Summary and additional notes

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- A *galvanic cell* (sometimes more appropriately called a *voltaic cell*) consists of two *half-cells* joined by a *salt bridge* or some other path that allows ions to pass between the two sides in order to maintain *electroneutrality*.
- The conventional way of representing an electrochemical cell of any kind is to write the oxidation half reaction on the left and the reduction on the right. Thus for the reaction

$$Zn(s) + Cu^{2+} \rightarrow Zn^{2+} + Cu(s)$$

we write

$$Zn(s) \mid Zn^{2+}(aq) \parallel Cu^{2+}(aq) \mid Cu(s)$$

in which the single vertical bars represent *phase boundaries*. The double bar denotes a *liquid-liquid boundary* which in laboratory cells consists of a salt bridge or in ion-permeable barrier. If the net cell reaction were written in reverse, the cell notation would become

$$Cu(s) \mid Cu^{2+}(aq) \parallel Zn^{2+}(aq) \mid Zn(s)$$

Remember: the ***R**eduction* process is always shown on the ***R**ight*.

at the electrode surface. The energy required to displace water molecules from the hydration shell of an ion as it approaches the electrode surface constitutes an *activation energy* which can slow down the process. Even larger activation energies (and slower reactions) occur when a molecule such as O_2 is formed or consumed.

This page titled 16.2: Galvanic cells and Electrodes is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.3: Cell Potentials and Thermodynamics

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially imortant that you know the precise meanings of all the highlighted terms in the context of this topic.

- When we refer to the "standard potential of a half-cell" or "couple" M^{2+}/M, we mean the potential difference $E_{right} - E_{left}$ of the cell
- If the potential difference of this cell is positive ($E_{right} - E_{left} > 0$), electrons will flow through an external circuit from the Pt/H_2 electrode to the M electrode and the cell reaction will spontaneously proceed in the direction written. The more positive the cell potential, the greater the tendency of this reaction to occur and the stronger the oxidizing agent M^{2+}.
- Through the relation $E° = - \Delta G°/nF$ it is apparent that a standard half-cell reduction potential is simply the decrease in the free energy per mole of electrons transferred to H^+ ions under the conditions that define the SHE. Strong reducing agents (good electron donors) have more negative $E°$s, while strong oxidizing agents (good acceptors) have more positive $E°$s.
- For a more general cell $X(s) \mid X^+ \parallel M^{2+} \mid M(s)$, $E°$ is similarly the fall in free energy per electron-mole when M^{2+} is reduced by X. This reaction can proceed spontaneously only if the cell potential is positive ($\Delta G°$ negative.)
- An *electron free energy diagram* that displays various redox couples on a vertical scale of free energies relative to H^+ serves as a convenient means of visualizing the possible reactions when two or more redox-active pairs are present in a solution. The position of a redox couple in relation to those of the H_2/ H^+ and $H_2O/O_2,H^+$ couples is especially significant because it indicates whether a given species will be thermodynamically stable in water.
- *Latimer diagrams* provide a convenient means of correlating the various oxidation states of a particular element.

It has long been known that some metals are more "active" than others in the sense that a more active metal can "displace" a less active one from a solution of its salt. The classic example is the one we have already mentioned on the preceding page:

$$\mathrm{Zn\,(s) + Cu^{2+} \rightarrow Zn^{2+} + Cu\,(s)}$$

Here zinc is more active because it can displace (precipitate) copper from solution. If you immerse a piece of metallic zinc in a solution of copper sulfate, the surface of the zinc quickly becomes covered with a black coating of finely-divided copper, and the blue color of the hydrated copper(II) ion diminishes. Similar comparisons of other metals made it possible to arrange them in the order of their increasing electron-donating (reducing) power. This sequence became known as the electromotive or activity series of the metals.

The most active (most strongly reducing) metals appear on top, and least active metals appear on the bottom. A more active metal (such as Zn) will donate electrons to the cation of a less active metal (Cu2+, for example.) Notice the special role of hydrogen here; although H2 does not have the physical properties of a metal, it is capable of being "displaced" (a rather archaic term seldom used in modern chemistry) from H2O or H+-containing (acidic) solutions. Note that the "active" metals are all "attacked by acids"; what this really means is that they are capable of donating electrons to H+.

displace H_2 from water, steam, or acids	**Li**	$2\ Li(s) + 2\ H_2O(l) \rightarrow 2\ LiOH(aq) + H_2(g)$
	K	$2\ K(s) + 2\ H_2O(l) \rightarrow 2\ KOH(aq) + H_2(g)$
	Ca	$Ca(s) + 2\ H_2O(l) \rightarrow Ca(OH)_2(s) + H_2(g)$
	Na	$2\ Na(s) + 2\ H_2O(l) \rightarrow 2\ NaOH(aq) + H_2(g)$
displace H_2 from steam or acids	**Mg**	$Mg(s) + 2\ H_2O(g) \rightarrow Mg(OH)_2(s) + H_2(g)$
	Al	$2\ Al(s) + 6\ H_2O(g) \rightarrow 2\ Al(OH)_3(s) + 3\ H_2(g)$
	Mn	$Mn(s) + 2\ H_2O(g) \rightarrow Mn(OH)_2(s) + H_2(g)$
	Zn	$Zn(s) + 2\ H_2O(g) \rightarrow Zn(OH)_2(s) + H_2(g)$
	Fe	$Fe(s) + 2\ H_2O(g) \rightarrow Fe(OH)_2(s) + H_2(g)$
displace H_2 from acids only	**Ni**	$Ni(s) + 2\ H^+(aq) \rightarrow Ni^{2+}(aq) + H_2(g)$
	Sn	$Sn(s) + 2\ H^+(aq) \rightarrow Sn^{2+}(aq) + H_2(g)$
	Pb	$Pb(s) + 2\ H^+(aq) \rightarrow Pb^{2+}(aq) + H_2(g)$
H_2		
can't displace H_2	**Cu**	
	Ag	
	Pt	
	Au	

The activity series has long been used to predict the direction of oxidation-reduction reactions. Consider, for example, the oxidation of Cu by metallic zinc that we have mentioned previously. The fact that zinc is near the top of the activity series means that this metal has a strong tendency to lose electrons. By the same token, the tendency of Zn to accept electrons is relatively small. Copper, on the other hand, is a poorer electron donor, and thus its oxidized form, Cu, is a fairly good electron acceptor. We would therefore expect the reaction

$$Zn(s) + Cu^{2+} \rightarrow Zn^{2+} + Cu(s) \tag{16.3.1}$$

to proceed in the direction indicated, rather than in the reverse direction. An old-fashioned way of expressing this is to say that "zinc will displace copper from solution". The above table is of limited practical use because it does not take into account the concentrations of the dissolved species. In order to treat these reactions quantitatively, it is convenient to consider the oxidation and reduction steps separately.

Standard half-cell potentials

When a net reaction proceeds in an electrochemical cell, oxidation occurs at one electrode (the anode) and reduction takes place at the other electrode (the cathode.) We can think of the cell as consisting of two half-cells joined together by an external circuit through which electrons flow and an internal pathway that allows ions to migrate between them so as to preserve electroneutrality.

Reduction potentials

Each half-cell has associated with it an electrode-solution potential difference whose magnitude depends on the nature of the particular electrode reaction and on the concentrations of the dissolved electroactive species. The sign of this potential difference depends on the direction (oxidation or reduction) in which the electrode reaction proceeds. To express them in a uniform way, we adopt the convention that half-cell potentials are always defined for the reduction direction. Thus the half-cell potential for the Zn/Zn^{2+} electrode (or *couple* as it is sometimes called) always refers to the reduction reaction

$$Zn^{2+} + 2e^- \rightarrow Zn(s) \tag{16.3.2}$$

In the cell $Zn(s)$ | $Zn^{2+}(aq)$ || $Cu^{2+}(aq)$ | $Cu(s)$ the zinc appears on the left side, indicating that it is being oxidized, not reduced. For this reason, the potential difference contributed by the left half-cell has the opposite sign to its conventional half-cell potential. More generally, we can define the cell potential (or cell EMF) as

$$E_{cell} = \Delta V = E_{right} – E_{left} \tag{16.3.3}$$

in which "right" and "left" refer to the cell notation convention ("**r**eduction on the **r**ight") and not, of course, to the physical orientation of a real cell in the laboratory. If we expand the above expression we see that the cell potential

$$E_{cell} = V_{Cu} – V_{solution} + V_{solution} – V_{Zn} \tag{16.3.4}$$

is just the difference between the two half-cell potentials E_{right} and E_{left}.

Reference half-cells

The fact that individual half-cell potentials are not directly measurable does not prevent us from defining and working with them. Although we cannot determine the absolute value of a half-cell potential, we can still measure its value in relation to the potentials of other half cells. In particular, if we adopt a reference half-cell whose potential is arbitrarily defined as zero, and measure the potentials of various other electrode systems against this reference cell, we are in effect measuring the half-cell potentials on a scale that is relative to the potential of the reference cell.

The reference cell that has universally been adopted for this purpose is the *hydrogen half-cell*

$$Pt|½H_{2(g)}|H^+_{(aq)}|| \ldots \tag{16.3.5}$$

in which hydrogen gas is allowed to bubble over a platinum electrode having a specially treated surface which catalyzes the reaction

$$½H_{2(g)} \rightarrow H^+ + e^- \tag{16.3.6}$$

When this electrode is operated under standard conditions of 1 atm H_2 pressure, 25°C, and pH = 0, it becomes the standard hydrogen electrode, sometimes abbreviated SHE.

To measure the relative potential of some other electrode couple M^{2+}/M, we can set up a cell

$$Pt|H_{2(g)}|H^+||M^{2+}_{(aq)}|M_{(s)} \tag{16.3.7}$$

whose net reaction is

$$H_2(g) + M^{2+}(aq) \rightarrow 2H^+ + M(s)$$

the potential difference between the platinum and M electrodes will be

$$E_{cell} = V_M – V_{solution} + V_{solution} – V_{Pt}$$

but since the difference $V_{solution} – V_{Pt}$ is by definition zero for the hydrogen half-cell, the cell potential we measure corresponds to

$$E_{cell} = V_M - V_{solution}$$

which is just the potential (relative to that of the SHE) of the half-cell whose reaction is

$$M^{2+} + 2e^- \rightarrow M_{(s)} \tag{16.3.8}$$



Measurement of a standard reduction potential.

The M^{2+}/M half-cell is on the left, and the standard hydrogen cell is on the right. The two half-cells are joined through the salt bridge in the middle. The more "active" the metal M (the greater its tendency to donate electrons to H^+), the more negative will be $E_{cell} = \Delta V = E_{right} - E_{left}$

Standard [reduction] potentials are commonly denoted by the symbol *E*°. *E*° values for hundreds of electrodes have been determined (mostly in the period 1925-45, during which time they were referred to as "oxidation potentials") and are usually tabulated in order of increasing tendency to accept electrons (increasing oxidizing power.)

Table 2: some standard reduction potentials

oxidant (electron acceptor)	reductant (elecron donor)	E°, volts
Na^+	Na*(s)*	-2.71
Zn^{2+}	Zn*(s)*	-.76
Fe^{2+}	Fe*(s)*	-.44
Cd^{2+}	Cd*(s)*	-.40
Pb^{2+}	Pb*(s)*	-.126

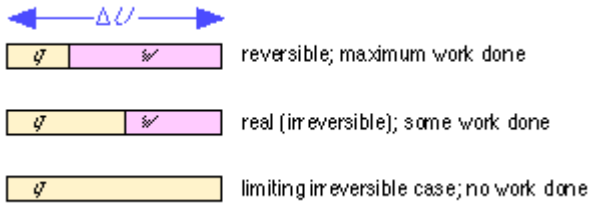
H_2*(g)*

0.000

AgCl*(s)*
Ag*(s)* + Cl^-*(aq)*

+.222

Hg_2Cl_2*(s)*



2Cl^-*(aq)* + 2Hg( )

+.268

Cu^{2+}
Cu*(s)*

+.337

$I_2(s)$
$2\ I^-$

+.535

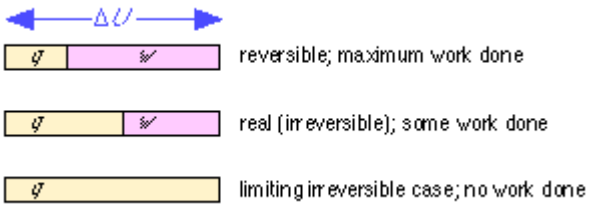
Fe^{3+}
Fe^{2+}

+.771

Ag^+
$Ag(s)$

+.799

$O_2(g) + 4H^+$



$2\ H_2O($ $)$

+1.23

$Cl_2(g)$
$2\ Cl^-$

+1.36

Note particularly that

- Unlike the activity series mentioned above, this table includes *non-metallic substances*, and it is *quantitative*.
- The more positive the half-cell EMF, the greater the tendency of the reductant to donate electrons, and the smaller the tendency of the oxidant to accept electrons.
- A species in the leftmost column can act as an oxidizing agent to any species above it in the reductant column.
- Oxidants such as Cl_2 that are below H_2O will tend to decompose water.

Given the $E°$ values for two half reactions, you can easily predict the potential difference of the corresponding cell: simply add the reduction potential of the reduction half-cell to the *negative* of the reduction potential (that is, to the oxidation potential) of the oxidation reaction.

Example 16.3.1

Find the standard potential of the cell

$$Cu(s)\ |\ Cu^{2+}\ ||\ Cl^-\ |\ AgCl(s)\ |\ Ag(s)$$

and predict the direction of electron flow when the two electrodes are connected.

Solution

The net reaction corresponding to this cell will be:

$$Cu(s) + 2AgCl(s) \rightarrow 2\ Ag(s) + 2\ Cl^{-}(aq) + Cu^{2+}(aq)$$

Where $Cu(s)/Cu^{2+}$ is being Oxidized and $Ag(s)/Ag^{+}$ is being Reduced

$$E_{cell} = E_{reduction} + E_{oxidation}$$ Or Written another way $E_{cell} = E_{right} - E_{left}$ =(.222 + (– .337)) v = **-0.115 v**

Since E_{cell} is negative, the reaction will run in the **opposite** direction. The correct net reaction will be:

$$2\ Ag(s) + 2\ Cl^{-}(aq) + Cu^{2+}(aq) \rightarrow 2AgCl(s) + Cu(s)$$

Where $Ag(s)/Ag^{+}$ is being Oxidized and $Cu(s)/Cu^{2+}$ is being Reduced

$$E_{cell} = E_{reduction} + E_{oxidation}$$ Or Written another way $E_{cell} = E_{right} - E_{left}$ =(.337 + (– .222)) v = **+0.115 v**

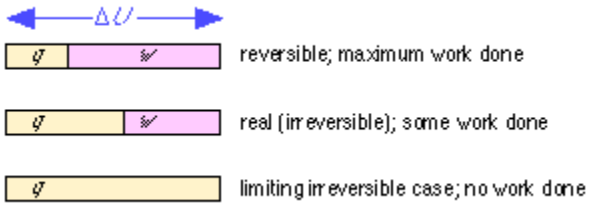
Since this potential is positive, the reaction will proceed to the right; electrons will be withdrawn from the silver electrode and flow through the external circuit into the copper electrode. Note carefully that in combining these half-cell potentials, we did not multiply *E*° for the Cu^{2+}/Cu couple by two. The reason for this will be explained later.

Cell potentials and Free (Gibbs) energy

From the above, it should be apparent that the potential difference between the electrodes of a cell is a measure of the tendency for the cell reaction to take place: the more positive the cell potential, the greater the tendency for the reaction to proceed to the right. But we already know that the standard free energy change expresses the tendency for any kind of process to occur under the conditions of constant temperature and pressure. Thus $\Delta G°$ and $E°$ measure the same thing, and are related in a simple way:

$$\Delta G° = -nFE° \tag{16.3.9}$$

... or in more detail (see below for explanations of the units given for voltage)



A few remarks are in order about this very fundamental and important relation:

- The negative sign on the right indicates that a *positive cell potential* (according to the sign convention discussed previously) implies a *negative free energy change*, and thus that the cell reaction will spontaneously proceed to the *right*.
- Electrical work is done when an electric charge *q* moves through a potential difference ΔV. The right side of Eq. 2 refers to the movement of *n* moles of charge across the cell potential *E*°, and thus has the dimensions of *work*.
- The value of $\Delta G°$ expresses the maximum useful work that a system can do on the surroundings. "Useful" work is that which can be extracted from the cell by electrical means to operate a lamp or some other external device. This excludes any *P-V* work that is simply a consequence of volume change (which could conceivably be put to some use!) and which would be performed in any case, even if the reactants were combined directly. This quantity of work $-\Delta G°$ can only be extracted from the system under the limiting conditions of a *thermodynamically reversible* change, which for an electrochemical cell implies zero current. The more rapidly the cell operates, the less electrical work it can supply.
- If *F* is expressed in coulombs per mole of electrons, the electrical work is in joules per mole. To relate these units to electrical units, recall that the coulomb is one amp-sec, and that *power*, which is the rate at which work is done, is measured in watts, which is the product of amps and volts

1 J = 1 watt-sec = 1 (amp-sec) × volts

Thus the *volt* has the dimensions of joules/coulomb– the energy produced per quantity of charge passing through the cell. Because voltage is the quotient of two extensive quantities, it is itself *intensive*. When we multiply the anodic and cathodic half-reactions by the stoichiometric factors required to ensure that each involves the same quantity of charge, the free energy change and the number of coulombs both increase by the same factor, leaving the potential (voltage) unchanged. This explains why we do not have to multiply the *E*°s of the anode and cathode reactions by stoichiometric factors when we are finding the potential of a complete cell.

If Eq. 2 is solved for *E*°, we have

$$E^\circ = \frac{\Delta G^\circ}{nF} \tag{16.3.10}$$

This states explicitly that the cell potential is a measure of the free energy change *per mole of electrons transferred*, which is a brief re-statement of the principle explained immediately above. To see this more clearly, consider the cell

Cu*(s)* | Cu^{2+} || Cl^- | AgCl*(s)* | Ag*(s)*

for which we list the standard reduction potentials and Δ*G*°s of the half-reactions:

reaction	***E*°**	***-nFE*°= Δ*G*°**
cathode: 2 × [AgCl(s) + e^- → Ag(s) + Cl^-] *anode:* Cu(s) → Cu^{2+} + 2 e^-	+.222 v −(+.337) V	−42800 J +65000 J
net: 2 Ag*(s)* + 2 Cl^-*(aq)* + Cu^{2+}*(aq)* → AgCl*(s)* + Cu*(s)* *cell:* Cu*(s)* | Cu^{2+}*(aq)* || AgCl*(s)* | Cl^-*(aq)* | Ag*(s)*	−.115 v	+22200 J

Here we multiply the cathodic reaction by two in order to balance the charge. Because the anodic reaction is written as an oxidation, we reverse the sign of its *E*° and obtain $E_{cell} = E_{right} - E_{left}$ = **−.115 volt** for the cell potential. The negative cell potential tells us that this reaction will not proceed spontaneously.

When the electrons don't cancel out

Note, however, that if we are combining two half reactions to obtain a third half reaction, the *E*° values are not additive, since this third half-reaction is not accompanied by another half reaction that causes the charges to cancel. Free energies are always additive, so we combine them, and use Δ*G*° = −*nFE*° to find the cell potential.

Example 16.3.2

Calculate *E*° for the electrode Fe^{3+}/Fe*(s)* from the standard potential of the couples Fe^{3+}/Fe^{2+} and Fe^{2+}/Fe*(s)*

Solution

Tabulate the values and calculate the Δ*G*°s as follows:

(i) Fe^{3+} + e^- → Fe^{2+}	E°_1 = .771 v , ΔG°_1 = −.771 *F*

(ii) $Fe^{2+} + 2\ e^- \rightarrow Fe(s)$	$E°_2$= –.440 v , $\Delta G°_2$ = +.880 *F*
(iii) $\mathbf{Fe^{3+} + 3\ e^- \rightarrow Fe(s)}$	$E°_3$ = ? , $\Delta G°_3$ = +.109 *F*

The Gibbs energy for half-reaction *(iii)* is 0.109*nF*, so $E°_3$ = –.109/3 = **–.036 v**

The fall of the electron

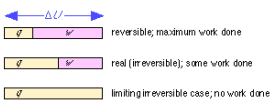
A table of standard half-cell potentials summarizes a large amount of chemistry, for it expresses the relative powers of various substances to donate and accept electrons by listing reduction half-reactions in order of increasing E° values, and thus of increasing spontaneity. The greater the value of E°, the greater the tendency of the substance on the left to acquire electrons, and thus the stronger this substance is as an oxidizing agent.

If you have studied elementary chemical thermodynamics, you will have learned about the role that a quantity called the *Gibbs free energy*, usually referred to as simply the "*free energy*", plays in determining the direction of any chemical change. The rule is that all spontaneous change (that is, all reactions that proceed to the "right") is associated with a fall in the free energy, and the greater the degree of that fall ($\Delta G°$), the greater will be the tendency for the reaction to take place.

If you are not familiar with the concept of free energy, just think of it as something like potential energy, which similarly decreases when spontaneous mechanical events occur, such as the dropping of a weight.

Since oxidation-reduction processes involve the transfer of an electron from a donor to an acceptor, it makes sense to focus on the electron and to consider that it falls from a higher-free energy environment (the reductant, or "source") to a lower-free energy one (the oxidant, or "sink".)

As can be seen from the diagram below, this model makes it far easier to predict what will happen when two or more oxidants and reducants are combined; the electron "falls" as far as it can, filling up oxidizing agents (sinks) from the bottom up, very much in the same way as electrons fill atomic orbitals as we build up larger atoms.

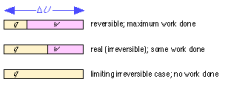


Electron-free energy diagram of redox couples

This chart is essentially an abbreviated form of a table of standard potentials in which the various couples are displayed on a vertical scale corresponding to $E° = -\Delta G°/nF$. Any available sink on the right side will tend to drain electrons from a source above it. For example, immersion of metallic zinc in a solution of $CuSO_4$ will result in the reduction of Cu^{2+} to metallic copper (red arrows.) Similarly, addition of chlorine to water will tend to oxidize the water, producing O_2 and Cl^- (blue arrows.) Note especially the positions of the H_2/H^+ and $H_2O/O_2,H^+$ couples on this chart, as they define the range of E°s for substances that will not decompose water (green region.)

A more detailed table with a more complete explanation can be seen on the "Fall of the electron" tutorial page; it is strongly recommended that you take the time to acquire a thorough understanding of this concept.

At this point, it might be worth calling your attention to the similar way of depicting acid-base reactions as representing the "fall of the proton" as shown below and described much more thoroughly here.



Proton-free energy diagram of acid-base systems

Acids are *proton sources* (donors), bases are *proton sinks*. Protons "fall" (in free energy) whenever a base is present that presents proton-empty free energy levels. The red arrows show what happens when acetic acid is titrated with a strong base; the results are acetate ion and water. Note here again the crucial role of water, both as a proton acceptor (forming hydronium ion) and as a proton donor (forming hydroxide ion.) Note also that the pH of a solution is a direct measure of the average free energy of protons in the solution (relative to H_3O^+.)

An important difference between proton transfer and electron transfer reactions is that the latter can vary greatly in speed, from almost instantaneous to so slow as to be unobservable. Acid-base reactions are among the fastest known.

Considerable insight into the chemistry of a single element can be had by comparing the standard electrode potentials (and thus the relative free energies) of the various oxidation states of the element. The most convenient means of doing this is the *Latimer diagram*. As examples, diagrams for iron and chlorine are shown below.

The formulas of the species that represent each oxidation state of the element are written from left to right in order of decreasing oxidation number, and the standard potential for the reduction of each species to the next on the right is written in between the formulas. Potentials for reactions involving hydrogen ions will be pH dependent, so separate diagrams are usually provided for acidic and alkaline solutions (effective hydrogen ion concentrations of 1M and 10^{-14} M, respectively).

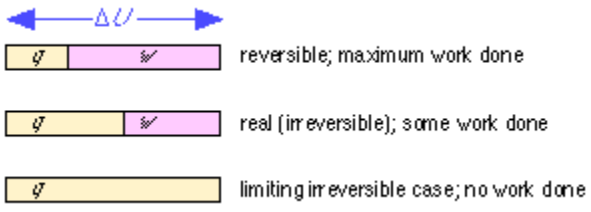
The more positive the reduction potential, the greater will be the tendency of the species on the left to be reduced to the one on the right. To see how Latimer diagrams are used, look first at the one for iron in acid solution. The line connecting Fe^{3+} and Fe^{2+} represents the reaction

$$Fe^{3+} + e^- \rightarrow Fe^{2+}$$

whose positive E° (.440 v) indicates that metallic iron will dissolve in acidic solution to form Fe^{2+}. Because the oxidation of this species to the +3 state has a negative potential (-0.771v; moving to the left on the diagram reverses the sign), the +2 state will be the stable oxidation state of iron under these conditions.

Disproportionation

An important condition to recognize in a Latimer diagram is when the potential on the left of a species is less positive than that on the right. This indicates that the species can oxidize and reduce itself, a process known as *disproportionation*. As an example, consider Cl_2 in alkaline solution. The potential for its reduction to Cl^- is sufficiently positive (+1.35 v) to supply the free energy necessary for the oxidation of one atom of chlorine to hypochlorite. Thus elemental chlorine is thermodynamically unstable with respect to disproportionation in alkaline solution, and the same it true of the oxidation product, ClO^- (hypochlorite ion).



Behavior of chlorine in water

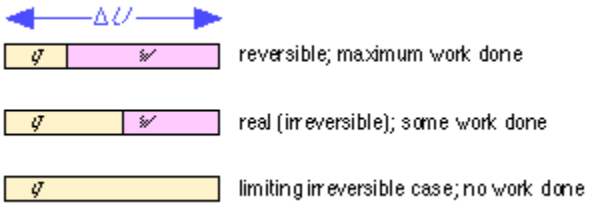
Cl_2 can oxidize water (green arrows, top) and also undergo disproportionation (purple arrows, bottom). In the latter process, one Cl_2 molecule donates electrons to another. Bear in mind that many oxidation-reduction reactions, unlike most acid-base

reactions, tend to be very slow, so the fact that a species is thermodynamically unstable does not always mean that it will quickly decompose. Thus the two reactions shown in the figure are normally very slow.

Thermodnamics of Galvanic cells

The free energy change for a process represents the maximum amount of non-*PV* work that can be extracted from it. In the case of an electrochemical cell, this work is due to the flow of electrons through the potential difference between the two electrodes. Note, however, that as the rate of electron flow (i.e., the current) increases, the potential difference must decrease; if we short-circuit the cell by connecting the two electrodes with a conductor having negligible resistance, the potential difference is zero and no work will be done. The full amount of work can be realized only if the cell operates at an infinitessimal rate; that is, reversibly.

You should recall that this is exactly analogous to the expansion of an ideal gas. The full amount of work $w = PdV$ is extracted only under the special condition that the external pressure P opposing expansion is only infinitessimally smaller than the pressure of the gas itself. If the gas is allowed to expand into a vacuum ($P = 0$), no work will be done.



The total amount of energy a reaction can supply under standard conditions at constant pressure and temperature is given by $\Delta H°$. If the reaction takes place by combining the reactants directly (no cell) or in a short-circuited cell, no work is done and the heat released is ΔH. If the reaction takes place in a cell that performs electrical work, then the heat released is diminished by the amount of electrical work done. In the limit of reversible operation, the heat released becomes

$$\Delta H = \Delta G° + T\,\Delta S$$

$$\mathrm{Pt} \mid \mathrm{H_2}(g) \mid \mathrm{H^+} \parallel \mathrm{M^{2+}}\,(aq) \mid \mathrm{M}(s)$$

whose left half consists of a *standard hydrogen electrode* (SHE) and whose net reaction is

$$\mathrm{H_2}(g) + \mathrm{M^{2+}}(aq) \rightarrow 2\mathrm{H^+} + \mathrm{M}(s)$$

This page titled 16.3: Cell Potentials and Thermodynamics is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.4: The Nernst Equation

Learning Objectives

Make sure you thoroughly understand the following essential ideas. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- The Nernst equation relates the effective concentrations (*activities*) of the components of a cell reaction to the standard cell potential. For a simple reduction of the form $M^{n+} + ne^- \rightarrow M$, it tells us that a half-cell potential will change by $59/n$ mV per 10-fold change in the activity of the ion.
- Ionic concentrations can usually be used in place of activities when the total concentration of ions in the solution does not exceed about about $0.001 M$.
- In those reactions in which H^+ or OH^- ions take part, the cell potential will also depend on the pH. Plots of *E vs.* pH showing the stability regions of related species are very useful means of summarizing the redox chemistry of an element.

The *standard cell potentials* we discussed in a previous section refer to cells in which all dissolved substances are at unit *activity*, which essentially means an "effective concentration" of 1 M. Similarly, any gases that take part in an electrode reaction are at an effective pressure (known as the *fugacity*) of 1 atm. If these concentrations or pressures have other values, the cell potential will change in a manner that can be predicted from the principles you already know.

Cell Potentials Depend on Concentrations

Suppose, for example, that we reduce the concentration of Zn^{2+} in the Zn/Cu cell from its standard effective value of 1 M to an to a much smaller value:

$$Zn(s)|Zn^{2+}(aq, 0.001\ M)||Cu^{2+}(aq)|Cu(s) \tag{16.4.1}$$

This will reduce the value of Q for the cell reaction

$$Zn(s) + Cu^{2+} \rightarrow Zn^{2+} + Cu(s) \tag{16.4.2}$$

thus making it more spontaneous, or "driving it to the right" as the *Le Chatelier principle* would predict, and making its free energy change ΔG more negative than $\Delta G\,°$, so that E would be more *positive* than $E°$. The relation between the actual cell potential E and the standard potential $E°$ is developed in the following way. We begin with the equation derived previously which relates the standard free energy change (for the complete conversion of products into reactants) to the standard potential

$$\Delta G° = -nFE° \tag{16.4.3}$$

By analogy we can write the more general equation

$$\Delta G = -nFE \tag{16.4.4}$$

which expresses the change in free energy for any extent of reaction— that is, for any value of the reaction quotient Q. We now substitute these into the expression that relates ΔG and $\Delta G\,°$ which you will recall from the chapter on chemical equilibrium:

$$\Delta G = \Delta G° + RT \ln Q \tag{16.4.5}$$

which gives

$$-nFE = -nFE° + RT \ln Q \tag{16.4.6}$$

which can be rearranged to

$$\underbrace{E = E° - \frac{RT}{nF} \ln Q}_{\text{applicable at all temperatures}} \tag{16.4.7}$$

This is the *Nernst equation* that relates the cell potential to the standard potential and to the activities of the electroactive species. Notice that the cell potential will be the same as $E°$ only if Q is unity. The Nernst equation is more commonly written in base-10 log form and for 25 °C:

$$\underbrace{E = E° - \frac{0.059}{n} \log_{10} Q}_{\text{Applicable at only 298K}} \tag{16.4.8}$$

Significance of the Nernst Equation

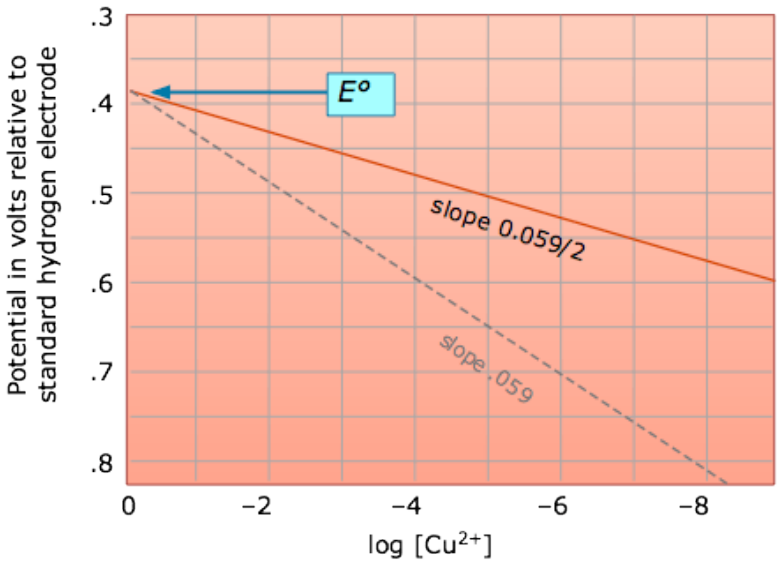
The Nernst equation tells us that a half-cell potential will change by 59 millivolts per 10-fold change in the concentration of a substance involved in a one-electron oxidation or reduction; for two-electron processes, the variation will be 28 millivolts per decade concentration change. Thus for the dissolution of metallic copper

$$Cu_{(s)} \rightarrow Cu^{2+} + 2e^- \tag{16.4.9}$$

the potential

$$E = (-0.337) - 0.0295 \log_{10}[Cu^{2+}] \tag{16.4.10}$$

becomes more positive (the reaction has a greater tendency to take place) as the cupric ion concentration decreases. This, of course, is exactly what the Le Chatelier Principle predicts; the more dilute the product, the greater the extent of the reaction.

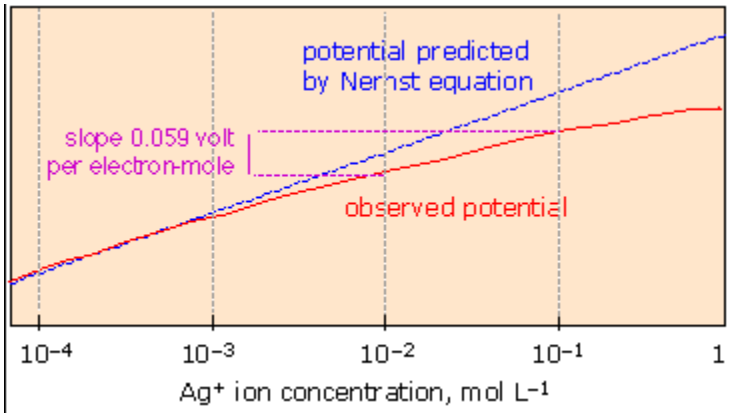


Electrodes with poise

The equation just above for the Cu/Cu^{2+} half-cell raises an interesting question: suppose you immerse a piece of copper in a solution of pure water. With Q = $[Cu^{2+}]$ = 0, the potential difference between the electrode and the solution should be infinite! Are you in danger of being electrocuted? You need not worry; without any electron transfer, there is no charge to zap you with. Of course it won't be very long before some Cu^{2+} ions appear in the solution, and if there are only a few such ions per liter, the potential reduces to only about 20 volts. More to the point, however, the system is so far from equilibrium (for example, there are not enough ions to populate the electric double layer) that the Nernst equation doesn't really give meaningful results. Such an electrode is said to be un*poised*. What ionic concentration is needed to poise an electrode? I don't really know, but I would be suspicious of anything much below 10^{-6} M.

The Nernst Equation works only in Dilute Ionic Solutions

Ions of opposite charge tend to associate into loosely-bound ion pairs in more concentrated solutions, thus reducing the number of ions that are free to donate or accept electrons at an electrode. For this reason, the Nernst equation cannot accurately predict half-cell potentials for solutions in which the total ionic concentration exceeds about 10^{-3} M.



How the cell potential really depends on concentration! The Nernst equation accurately predicts cell potentials only when the equilibrium quotient term Q is expressed in activities. Ionic activities depart increasingly from concentrations when the latter exceed 10^{-4} to 10^{-5} M, depending on the sizes and charges of the ions.

If the Nernst equation is applied to more concentrated solutions, the terms in the reaction quotient Q must be expressed in "effective concentrations" or activities of the electroactive ionic species. The activity coefficient γ) relates the concentration of an ion to its activity a in a given solution through the relation

$$a = \gamma c \tag{16.4.11}$$

Since electrode potentials measure activities directly, activity coefficients can be determined by carrying out appropriate EMF measurements on cells in which the concentration of the ion of interest is known. The resulting Es can then be used to convert concentrations into activities for use in other calculations involving equilibrium constants.

Cell potentials and pH: stability diagrams

As most of us recall from our struggles with balancing redox equations in beginning chemistry courses, many electron-transfer reactions involve hydrogen ions and hydroxide ions. The standard potentials for these reactions therefore refer to the pH, either 0 or 14, at which the appropriate ion has unit activity. Because multiple numbers of H^+ or OH^- ions are often involved, the potentials given by the Nernst equation can vary greatly with the pH. It is frequently useful to look at the situation in another way by considering what combinations of potential and pH allow the stable existence of a particular species. This information is most usefully expressed by means of a *E*-vs.-pH diagram, also known as a *Pourbaix diagram*.

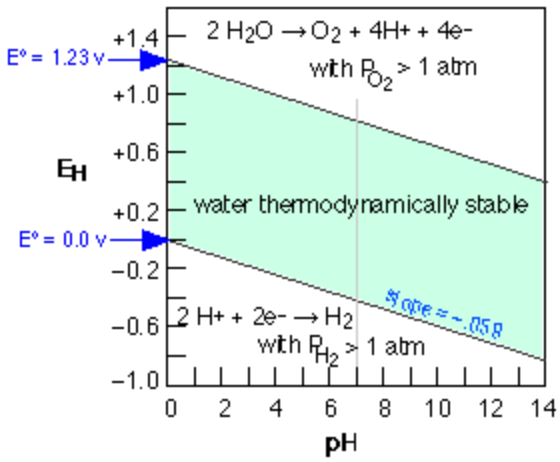


Figure 16.4.2: Stability (Pourbaix) diagram for water. The two E° values shown at the left refer to "standard" conditions of unit H^+ activity (pH=0) and gas pressures of 1 atm. At combinations of pH and E that lie outside the shaded area, the partial pressures of O_2 or H_2 exceed 1 atm, signifying the decomposition of water. The unity partial pressures are of course arbitrary criteria; in a system open to the atmosphere, water can decompose even at much lower H_2 partial pressures, and at oxygen pressures below 0.2 atm. Fortunately, these processes are in most cases quite slow.

As was noted in connection with the shaded region in the figure below, water is subject to decomposition by strong oxidizing agents such as Cl_2 and by reducing agents stronger than H_2. The reduction reaction can be written either as

$$2H^+ + 2e^- \rightarrow H_2(g) \tag{16.4.12}$$

or, in neutral or alkaline solutions as

$$2H_2O + 2e^- \rightarrow H_2(g) + 2OH^- \tag{16.4.13}$$

These two reactions are equivalent and follow the same Nernst equation

$$E_{\mathrm{H^+/H_2}} = E^\circ_{\mathrm{H^+/H_2}} + \frac{RT}{nF}\ln\left(\frac{[\mathrm{H^+}]^2}{P_{\mathrm{H_2}}}\right)$$

$$E_{H^+/H_2} = E^o_{H^+/H_2} + \frac{RT}{nF}\ln\left(\frac{[H^+]^2}{P_{H_2}}\right) \tag{16.4.14}$$

which, at 25°C and unit H_2 partial pressure reduces to

$$E = E^\circ - \frac{0.059}{2} \times 2pH = -0.059\ pH \tag{16.4.15}$$

Similarly, the oxidation of water

$$2H_2O \rightarrow O_2(g) + 4H^+ + 2e^- \tag{16.4.16}$$

is governed by the Nernst equation

$$E_{O_2/H_2O} = E^o_{O_2/H_2O} + \frac{RT}{nF}\ln\left(P_{O_2}[H^+]^4\right) \tag{16.4.17}$$

which similarly becomes E = 1.23 – 0.059 pH, so the E-*vs*-pH plots for both processes have identical slopes and yield the stability diagram for water in Figure 16.4.2. This Pourbaix diagram has special relevance to electrochemical corrosion of metals. Thus metals above hydrogen in the activity series will tend to undergo oxidation (corrosion) by reducing H^+ ions or water.

Chlorine in water

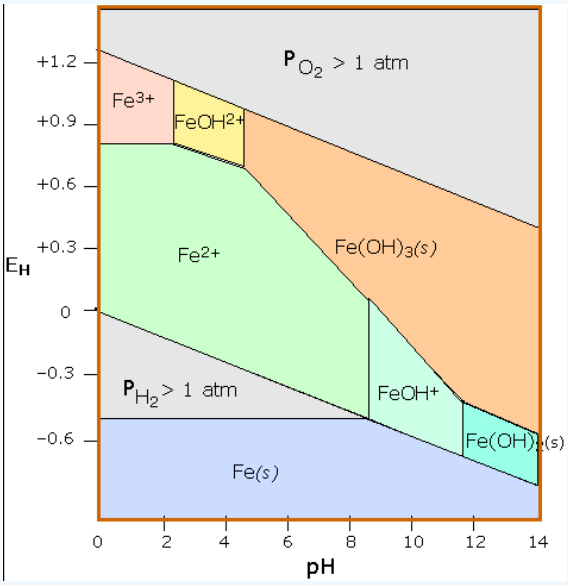
Because chlorine is widely used as a disinfectant for drinking water, swimming pools, and sewage treatment, it is worth looking at its stability diagram. Note that the effective bactericidal agent is not Cl_2 itself, but its oxidation product hypochlorous acid HOCl which predominates at pH values below its pK_a of 7.3. Note also that

- Cl_2 is unstable in water except at very low pH; it decomposes into HOCl and Cl^-.
- Hypochlorous acid and its anion are stronger oxidants than O_2 and thus subject to decomposition in water. The only stable chlorine species in water is Cl^-.
- Decomposition of HOCl occurs very slowly in the dark, but is catalyzed by sunlight. For this reason the chlorine in outside swimming pools must be frequently renewed.
- Decomposition of Cl_2 and HOCl by reaction with organic material in municipal water supply systems sometimes makes it necessary to inject additional chlorine at outlying locations.

Each solid line represents a combination of E and pH at which the two species on either side of it can coexist; at all other points, only a single species is stable. Note that equilibria between species separated by diagonal lines are dependent on both E and pH, while those separated by horizontal or vertical lines are affected by pH only or E only, respectively.

Iron

Stability diagrams are able to condense a great amount of information into a compact representation, and are widely employed in geochemistry and corrosion engineering. The Pourbaix diagram for iron is one of the more commonly seen examples.



Pourbaix diagram for iron. *Three oxidation states of iron (0, +2 and +3) are represented on this diagram. The stability regions for the oxidized iron states are shown only within the stability region of H_2O. Equilibria between species separated by vertical lines are dependent on pH only.*

The +3 oxidation state is the only stable one in environments in which the oxidation level is controlled by atmospheric O_2. This is the reason the Earth's crust contains iron oxides, which developed only after the appearance of green plants which are the source of O_2. Iron is attacked by H^+ to form H_2 and Fe(II); the latter then reacts with O_2 to form the various colored Fe(III) oxides that constitute "rust". Numerous other species such as oxides and hydrous oxides are not shown. A really "complete" diagram for iron would need to have at least two additional dimensions showing the partial pressures of O_2 and CO_2.

Concentration Cells

From your study of thermodynamics you may recall that the process

$$\text{solute } \textit{(concentrated)} \rightarrow \text{solute } \textit{(dilute)}$$

is accompanied by a fall in free energy, and therefore is capable of doing work on the surroundings; all that is required is some practical way of capturing this work. One way of doing this is by means of a concentration cell such as

$$\text{Cu}\textit{(s)} \mid \text{CuNO}_3\textit{(.1 M)} \parallel \text{CuNO}_3\textit{(.01 M)} \mid \text{Cu}\textit{(s)}$$

$$\text{cathode: } \text{Cu}^{2+}\textit{(.1 M)} + 2e^- \rightarrow \text{Cu}\textit{(s)}$$

$$\text{anode: Cu}\textit{(s)} \rightarrow \text{Cu}^{2+}\textit{(.01 M)} + 2e^-$$

$$\text{net: } \text{Cu}^{2+}\textit{(.1 M)} \rightarrow \text{Cu}^{2+}\textit{(.01 M)}$$

which represents the transport of cupric ion from a region of higher concentration to one of lower concentration.

The driving force for this process is the free energy change ΔG associated with the concentration gradient (C_2 – C_1), sometimes known as the free energy of dilution:

$$\Delta G_{dilution} = RT \ln(C_2 – C_1) \tag{16.4.18}$$

Note, however, that Cu^{2+} ions need not physically move between the two compartments; electron flow through the external circuit creates a "virtual" flow as copper ions are created in the low-concentration side and discharged at the opposite electrode. Nitrate ions must also pass between the cells to maintain *electroneutrality*. The Nernst equation for this cell is

$$E = E^\circ - \left(\frac{0.059}{n}\right) \log_{10} Q = 0 - 0.29 \log_{10} 0.1 = +0.285\, V \tag{16.4.19}$$

Note that E° for a concentration cell is always zero, since this would be the potential of a cell in which the electroactive species are at unit activity in both compartments.

> E° for a concentration cell is **always** zero,

This page titled 16.4: The Nernst Equation is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.5: Applications of the Nernst Equation

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Atmospheric oxygen is a strong oxidizing agent, but in the absence of a suitable catalyst this reaction is ordinarily too slow to be observable.
- The pE is a useful index of electron availability, especially in natural waters and biological systems in which multiple redox systems are usually present.
- Solubility products can often be measured by constructing a cell in which one of the electrodes involves the sparingly soluble salt, and the net cell equation corresponds to the solubility reaction.
- Potentiometric titrations are widely used to measure concentrations of species that are readily oxidized or reduced.
- Measurement of pH is accomplished with an electrode consisting of a thin membrane of glass in which Na^+ ions are exchanged with H^+.
- Membrane (Donnan) potentials arise when passage of an ion across a semipermeable membrane is selectively facilitated or inhibited.
- Nerve conduction occurs not by a flow of charge through the nerve, but by a wave of depolarization resulting from the concerted action of active channels that govern the passage of ions (mainly K^+ and Na^+) across the membrane enclosing the nerve.

We ordinarily think of the oxidation potential being controlled by the concentrations of the oxidized and reduced forms of a redox couple, as given by the Nernst equation. Under certain circumstances it becomes more useful to think of *E* as an independent variable that can be used to control the value of Q in the Nernst equation. This usually occurs when two redox systems are present, one being much more concentrated or kinetically active than the other. By far the most important example of this is the way atmospheric oxygen governs the composition of the many redox systems connected with biological activity.

Oxygen and the Aquatic Environment

The presence of oxygen in the atmosphere has a profound effect on the redox properties of the aquatic environment— that is, on natural waters exposed directly or indirectly to the atmosphere, and by extension, on organisms that live in an aerobic environment.This is due, of course, to its being an exceptionally strong oxidizing agent and thus a low-lying sink for electrons from most of the elements and all organic compounds. Those parts of the environment that are protected from atmospheric oxygen are equally important because it is only here that electrons are sufficiently available to produce the "reducing" conditions that are essential for processes varying from photosynthesis to nitrogen fixation.

Example 16.5.1

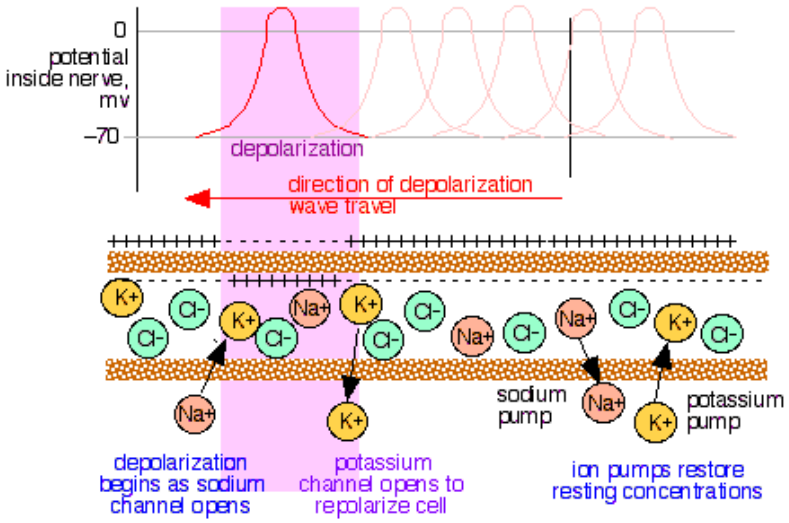
Estimate the redox potential of a natural water that is in equilibrium with the atmosphere at pH 7 and 298 K. What fraction of a dilute solution Fe^{2+} will be in its oxidized form Fe^{3+} in such a water?

Solution

The relevant *E*°s are

- 1.23 v for $O_2(g) + 4H^+ + 4e^- \rightarrow 2H_2O$ and
- 0.77 V for the Fe^{3+}/Fe^{2+} couple.

(a) The potential (with respect to the SHE, of course) is given by the Nernst equation



which works out to E = 0.82 volt. As the Le Chatelier principle predicts, the higher pH (lower $[H^+]$ compared to that at the "standard" pH of zero) reduces the electron-accepting tendency of oxygen.

(b) The Nernst equation for the reduction of Fe^{3+} is E = .77–.059 log Q, in which Q is the ratio $[Fe^{2+}]/[Fe^{3+}]$. With E set by the O_2/H_2O couple, this becomes

$$0.82 = 0.77 - 0.059 \log Q$$

which gives $Q = 10^{-0.85}$ or $[Fe^{2+}]/[Fe^{3+}]$ = 0.14/1, so the fraction of the iron in its oxidized form is 1/1.14 = 0.88.

If we can have pH, why not pE?

As you will recall from your study of acid-base chemistry, the pH of a solution (defined as –log $\{H^+\}$) is a measure of availablity (technically, the *activity*) of protons in the solution. As is explained in more detail here, protons tend to "fall" (in free energy) from filled donor levels (acids) to lower acceptor levels (bases.) Through the relation

$$[H^+] \approx K_a \frac{C_a}{C_b} \tag{16.5.1}$$

which can be rewritten as

$$\frac{C_a}{C_b} \approx \frac{[H^+]}{K_a} \tag{16.5.2}$$

in which the pH is treated as an independent variable that controls the ratio of the conjugate forms of any acid-base pairs in the solution:

$$\log\left(\frac{C_a}{C_b}\right) \approx pH–pK_a \tag{16.5.3}$$

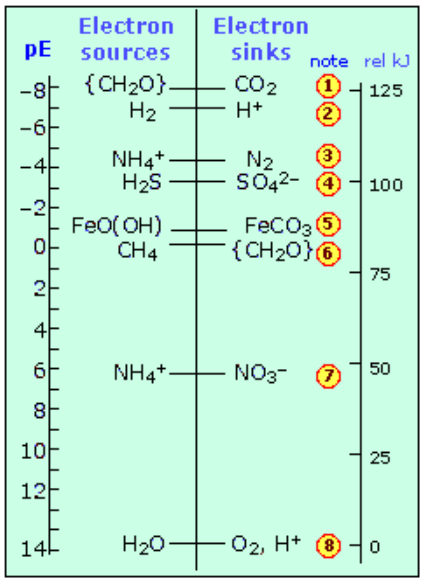
In the same way, we can define the pE as the negative log of the *electron activity* in the solution:

$$pE = -\log e^- \tag{16.5.4}$$

of these particles (but not to their "concentrations") when we are considering their availability to donors and acceptors. We will not get into the details of how pE is actually calculated (it is of course related to the ordinary standard electrode potential). To get an idea of its significance, consider the following chart that shows the pE° values of some redox systems that are of immense

importance in the aquatic environment.

- These pE values refer to typical environmental conditions with pH=7 and oxygen partial pressure of 0.21 atm. The scale below shows the free energy of a mole of electrons relative to their level in H_2O.
- The two conjugate forms of any redox pair are present in equal concentrations when the pE is at the level at which the pair is shown. At pE's above or below this level, the reduced or oxidized form will predominate.
- The sugar *glucose*, denoted by the general formula for carbohydrates $\{CH_2O\}$, is the source of chemical energy for most organisms. Note that it is thermodynamically stable (and thus capable of being formed by photosynthesis) only under highly reducing conditions.
- Organisms derive their metabolic free energy when electrons fall from glucose to a lower-lying acceptor on the right.
- Delivery of electrons from glucose to O_2 (**8**) is the source of metabolic free energy for all aerobic organisms, yielding 125 kJ per mole of electrons transferred.



A few other points about this plot are worth noting:

- Anaerobic organisms must make do with electron-acceptors above oxygen. The poor bacteria that depend on reducing hydrogen ions (**2**) have it worst of all, gaining only a tiny amount of metabolic energy to produce a tiny puff of hydrogen gas.
- Reaction (**3**) is not much more efficient, but it is the vital first link in the process of natural *nitrogen fixation* that gets carried out in the protected electron-rich environment of the organisms that live in the root nodules of legumes.
- The consequences of (**4**) are readily apparent if you have ever noticed the rotten-egg odor of some poorly-aerated muddy soils.
- Reaction (**6**) is known as *fermentation*; it takes place in the anarobic soils of marshes and bogs (hence the nickname "marsh gas" for methane) and in the insides of animals from termites to cows, and sometimes to our embarrassment, in ourselves. Notice that glucose plays the double role of electron donor and acceptor here; this is a *disproportionation* reaction.
- Finally, if you have ever had much to do with babys' diapers, you have likely noticed the smell of ammonia produced as the anaerobic bacteria from within deliver electrons to nitrate ions (**7**). It's important to bear in mind that the reactions discussed above are mediated by living organisms; without the necessary enzymes to catalyze them, their rates are essentially zero.

For a more detailed chart, see Falling through the respiratory chain.

Analytical chemistry applications

A very large part of Chemistry is concerned, either directly or indirectly, with determining the concentrations of ions in solution. Any method that can accomplish such measurements using relatively simple physical techniques is bound to be widely exploited. Cell potentials are fairly easy to measure, and although the Nernst equation relates them to ionic activities rather than to concentrations, the difference between them becomes negligible in solutions where the total ionic concentration is less than

about 10^{-3} *M*.

The concentrations of ions in equilibrium with a sparingly soluble salt are sufficiently low that their direct determination can be quite difficult. A far simpler and common procedure is to set up a cell in which one of the electrode reactions involves the insoluble salt, and whose net cell reaction corresponds to the dissolution of the salt. For example, to determine the K_{sp} for silver chloride, we could use the cell

$$Ag_{(s)}|Ag^{+}(?\ M)||Ag^{+},Cl^{-}|AgCl_{(s)}|Ag_{(s)} \tag{16.5.5}$$

whose net equation corresponds to the dissolution of silver chloride:

cathode:	AgCl*(s)* + e^- → Ag*(s)* + Cl^-*(aq)*	*E*° = +0.222 v
anode:	Ag*(s)* → Ag^+*(aq)* + e^-	*E*° = –(+0.799) v
net:	AgCl*(s)* → Ag^+ + Cl^-	*E*° = -0.577 v

The standard potential for the net reaction refers to a hypothetical solution in which the activities of the two ions are unity. The cell potential we actually observe corresponds to *E* in the Nernst equation, which is then solved for *Q* which gives K_{sp} directly.

Potentiometric titrations

In many situations, accurate determination of an ion concentration by direct measurement of a cell potential is impossible due to the presence of other ions and a lack of information about activity coefficients. In such cases it is often possible to determine the ion indirectly by titration with some other ion. For example, the initial concentration of an ion such as Fe^{2+} can be found by titrating with a strong oxidizing agent such as Ce^{4+}. The titration is carried out in one side of a cell whose other half is a reference electrode:

$$\text{Pt}(s)\ |\ Fe^{2+},\ Fe^{3+}\ ||\ \textit{reference electrode}$$

Initially the left cell contains only Fe^{2+}. As the titrant is added, the ferrous ion is oxidized to Fe^{3+} in a reaction that is virtually complete:

$$Fe^{2+} + Ce^{4+} \rightarrow Fe^{3+} + Ce^{3+} \tag{16.5.6}$$

The cell potential is followed as the Fe^{2+} is added in small increments. Once the first drop of ceric ion titrant has been added, the potential of the left cell is controlled by the ratio of oxidized and reduced iron according to the Nernst equation

$$E = 0.68 - 0.059\ \log \frac{]Fe^{3+}]}{[Fe^{2+}]} \tag{16.5.7}$$

which causes the potential to rise as more iron becomes oxidized.

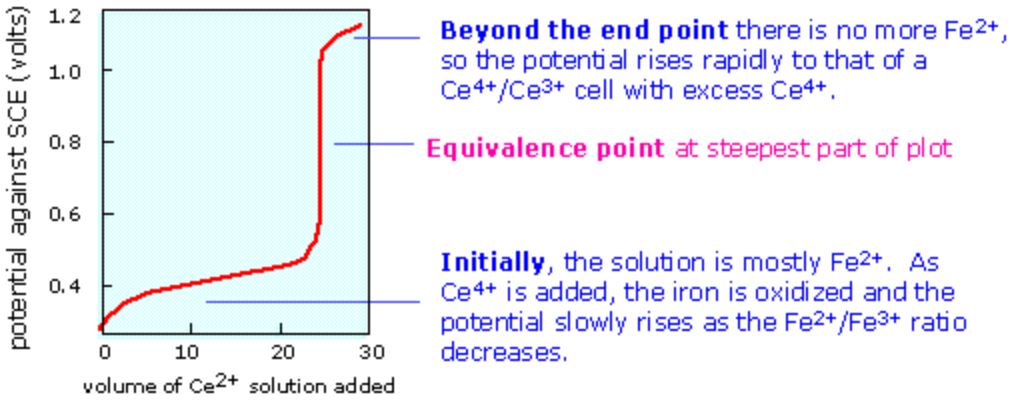


Figure 16.5.1: potentiometric titration curve

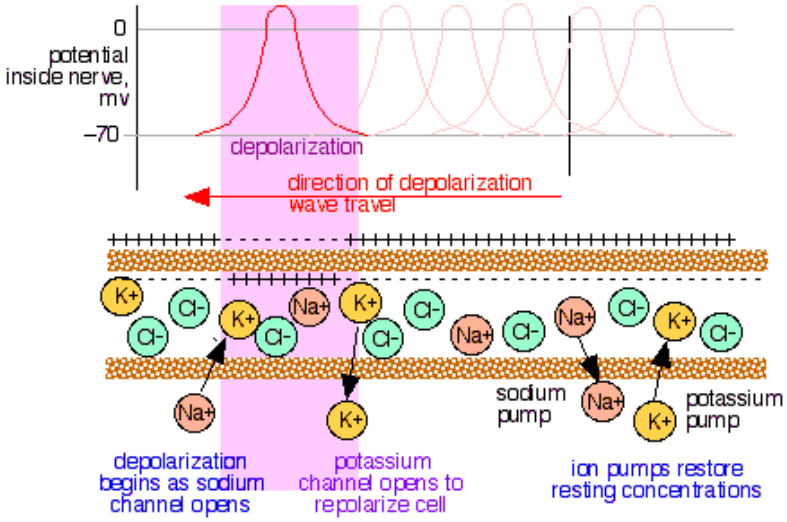
When the equivalence point is reached, the Fe^{2+} will have been totally consumed (the large equilibrium constant ensures that this will be so), and the potential will then be controlled by the concentration ratio of Ce^{3+}/Ce^{4+}. The idea is that both species of a redox couple must be present in reasonable concentrations poise an electrode (that is, to control its potential according to the Nernst equation.) If one works out the actual cell potentials for various concentrations of all these species, the resulting titration curve looks much like the familiar acid-base titration curve. The end point is found not by measuring a particular cell voltage, but by finding what volume of titrant gives the steepest part of the curve.

Measurement of pH

Since pH is actually defined in terms of hydrogen ion *activity* and not its concentration, a *hydrogen electrode* allows a direct measure of $\{H^+\}$ and thus of –log $\{H^+\}$, which is the pH. All you need is to measure the voltage of a cell

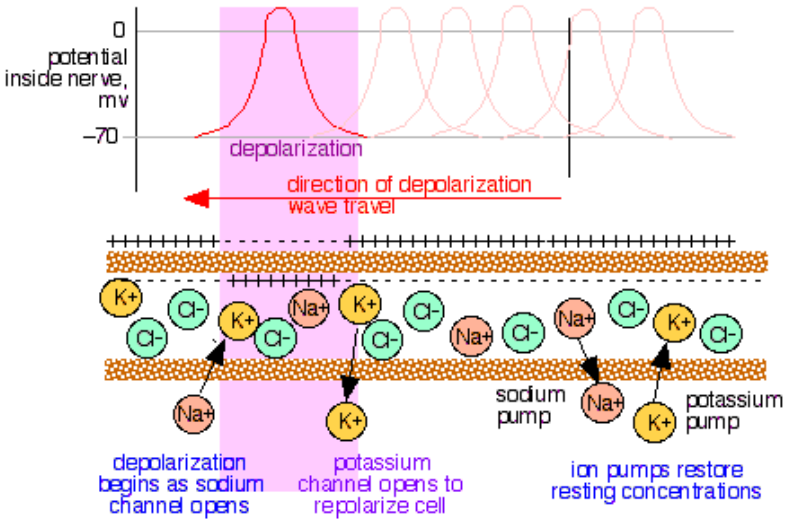
$$H_2(g, 1\ atm) \mid Pt \mid H^+(?\ M) \parallel \textit{reference electrode}$$

In theory this is quite simple, but when it was first employed in the pre-electronics era, it required some rather formidable-looking apparatus (such as the L&N vibrating-reed electrometer setup from the 1920's shown here) and the use of explosive hydrogen gas. Although this arrangement (in which the reference electrode could be a standard hydrogen electrode) has been used for high-precision determinations since that time, it would be impractical for routine pH measurements of the kinds that are widely done, especially outside the research laboratory.



The glass electrode for pH measurements

In 1914 it was discovered that a thin glass membrane enclosing a solution of HCl can produce a potential that varies with the hydrogen ion activity $\{H^+\}$ in about the same way as that of the hydrogen electrode. Glass electrodes are manufactured in huge numbers for both laboratory and field measurements. They contain a built-in Ag-AgCl reference electrode in contact with the HCl solution enclosed by the membrane.

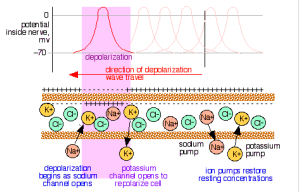


The potential of a glass electrode is given by a form of the Nernst equation very similar to that of an ordinary hydrogen electrode, but of course without the H_2:

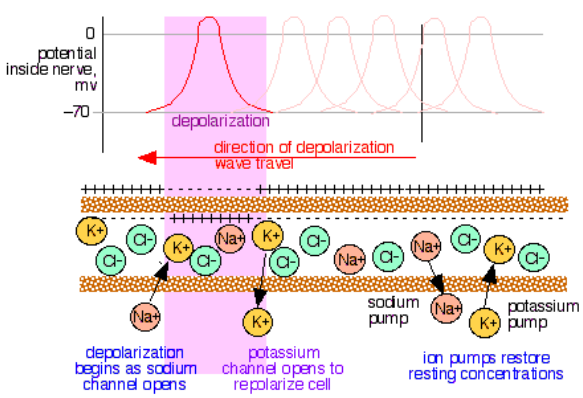
$$E_{membrane} = A + (RT/F) \ln (\{H^+\} + B)$$

in which *A* and *B* are constants that depend on the particular glass membrane.

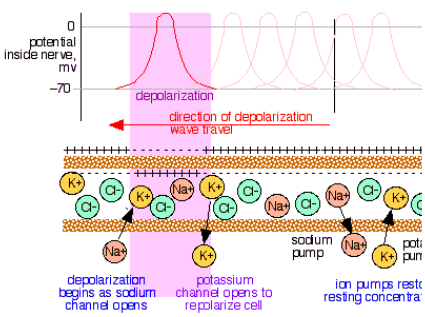
The reason a glass membrane would behave in this way was not understood until around 1970. It now appears that hydrogen ions in the external solution diffuse through the glass and push out a corresponding number of the Na^+ ions which are normally present in most glasses. These sodium ions diffuse to whichever side of the membrane has the lower concentration, where they remain mostly confined to the surface of the glass, which has a porous, gelatinous nature. It is the excess charge produced by these positive ions that gives rise to the pH-dependent potential. The first commercial **pH meter** was developed by Arnold Beckman (1900-2004) while he was a Chemistry professor at CalTech. He was unable to interest any of the instrumentation companies in marketing it, so he founded his own company and eventually became a multi-millionaire philanthropist.



Arnold Beckman



An early Beckman pH meter



A typical modern pH meter

Ion-selective electrodes

The function of the membrane in the glass electrode is to allow hydrogen ions to pass through and thus change its potential, while preventing other cations from doing the same thing (this selectivity is never perfect; most glass electrodes will respond to moderate concentrations of sodium ions, and to high concentrations of some others.) A glass electrode is thus one form of ion-

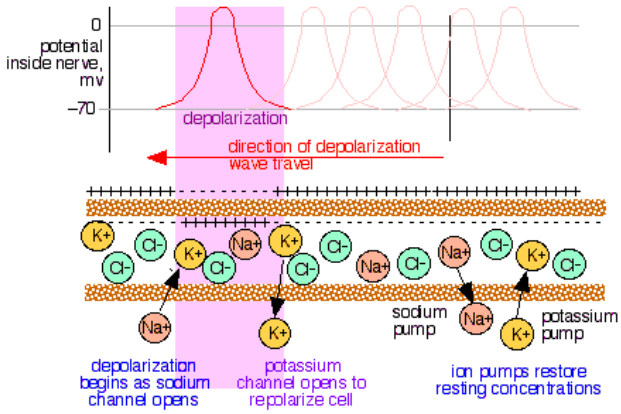
selective electrode. Since about 1970, various other membranes have been developed which show similar selectivities to certain other ions. These are widely used in industrial, biochemical, and environmental applications.

Membrane potentials and nerve conduction

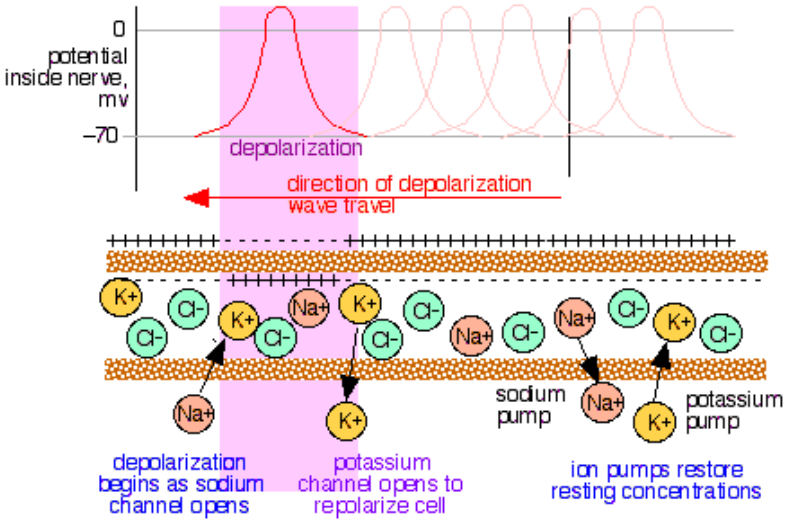
You may recall the phenomena of osmosis and osmotic pressure that are observed when two solutions having different solute concentrations are separated by a thin film or membrane whose porosity allows small ions and molecules to diffuse through, but which holds back larger particles. If one solution contains a pair of oppositely-charged ionic species whose sizes are very different, the smaller ions may pass through the semipermeable membrane while the larger ones are retained. This will produce a charge imbalance between the two solutions, with the original solution having the charge sign of the larger ion. Eventually the electrical work required to bring about further separation of charges becomes too large to allow any further net diffusion to take place, and the system settles into an equilibrium state in which a constant potential difference (usually around a volt or less) is maintained. This potential difference is usually called a membrane potential or *Donnan potential* after the English chemist who first described this phenomenon around 1930.

Origin of a membrane potential

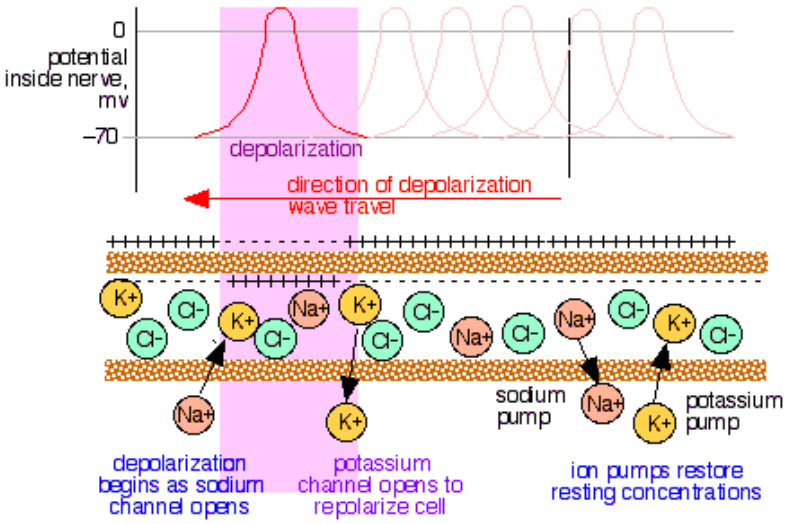
If the smaller ions are able to diffuse through the membrane but the larger ions cannot, a potential difference will develop between the two solutions. This membrane potential can be observed by introducing a pair of platinum electrodes.



The figure shows a simple system containing the potassium salt of a protein on one side of a membrane, and potassium chloride on the other. The proteinate anion, being too large to diffuse through the membrane, gives rise to the potential difference. The value of this potential difference can be expressed by a relation that is essentially the same as the Nernst equation, although its derivation is different. The membrane potential can be expressed in terms of the ratio of either the K^+ or Cl^- ion activities:



The membrane surrounding most living cells contains sites or "channels" through which K^+ ions are selectively transported so that the concentration of K^+ inside the cell is 10-30 times that of the intracellular fluid. Taking the activity ratio as about 20, the above equation predicts that the potential difference $\theta_{inside} - \theta_{outside}$ will be



which is consistent with observed values. Transport of an ion such as K^+ from a region of low concentration into the more concentrated intercellular fluid requires a source of free energy, which is supplied by ATP under enzymatic control. The metabolic processes governing this action are often referred to as "ion pumps".

Nerve conduction

Transmission of signals through the nervous system occurs not by the movement of a charge carrier through the nerve, but by waves of differential ion concentrations that travel along the length of the nerve. These concentration gradients are reduced by protein-based ion channels and ATP-activated (and energy-consuming) ion pumps specific to K^+ and Ca^{2+} ions.

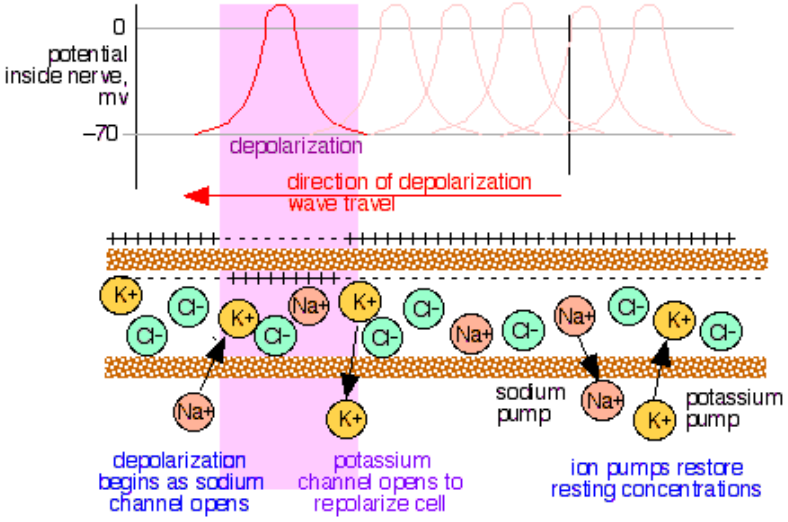


Figure 16.5.2: Conduction of nerve impulses

We sometimes think of our nerves as the body's wiring, but the "electricity" that they transmit is not a flow of electrons, but a rapidly-traveling wave of depolarization involving the transport of ions through the nerve membrane. The normal potential

difference between the inner and outer parts of nerve cells is about –70 mv as estimated above. Transmissin of a nerve impulse is initiated by a reduction of this potential difference to about –20 mv. This has the effect of temporarily opening the Na^+ channel; the influx of these ions causes the membrane potential of the adjacent portion of the nerve to collapse, leading to an effect that is transmitted along the length of the nerve. As this pulse passes, K^+ and Na^+ pumps restore the nerve to its resting condition.

Summary

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

- Atmospheric oxygen is a strong oxidizing agent, but in the absence of a suitable catalyst this reaction is ordinarily too slow to be observable. In organisms that possess the appropriate enzymes, the reaction is sufficiently fast to control the availability of electrons to other redox systems.
- The pE is a useful index of electron availability, especially in natural waters and biological systems in which multiple redox systems are usually present. Although electrons tend to "fall" (in free energy) from sources to sinks, the outcome depends very strongly on the rates of the various reactions. This is quite different from proton-exchange (acid-base) reactions which are uniformly fast.
- Solubility products can often be measured by constructing a cell in which one of the electrodes involves the sparingly soluble salt, and the net cell equation corresponds to the solubility reaction.
- Potentiometric titrations are widely used to measure concentrations of species that are readily oxidized or reduced. The reaction must be fast and have a very large equilibrium constant. The equivalance point is detected by the rapid change in potential that occurs when control of the cell potential passes from the redox system of the analyte to that of the titrant.
- Measurement of pH is accomplished with an electrode consisting of a thin membrane of glass in which Na^+ ions are exchanged with H^+.
- Membrane (Donnan) potentials arise when passage of an ion across a semipermeable membrane is selectively facilitated or inhibited. The classic example is a sodium proteinate solution in which the protein anion is too large to pass through the membrane. In organisms, ion-specific channels or "pumps" have a similar effect.
- Nerve conduction occurs not by a flow of charge through the nerve, but by a wave of depolarization resulting from the concerted action of active channels that govern the passage of ions (mainly K^+ and Na^+) across the membrane enclosing the nerve.

This page titled 16.5: Applications of the Nernst Equation is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.6: Batteries and Fuel Cells

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented below.

- A battery is a galvanic cell in which some of the free energy change associated with a spontaneous electron-transfer reaction is captured in the form of electrical energy.
- A secondary or storage battery is one in which the electron-transfer reaction can be reversed by applying a charging current from an external source.
- A fuel cell is a special type of battery in which the reactants are supplied from an external source as power is produced. In most practical fuel cells, H^+ions are produced at the anode (either from H_2 or a hydrocarbon) and oxygen from the air is reduced to H2O at the cathode.
- The cathodic reduction of O_2 is kinetically limited, necessitating the use of electrode surfaces having high catalytic activity.
- The electrodes in batteries must have very high effective surface areas, and thus be highly porous. This requirement may conflict with the other important one of efficient diffusion of reactants and products in the narrow channels within the pores.
- Batteries and fuel cells designed to power vehicles and portable devices need to have high charge-to-weight and charge-tovolume ratios.

One of the oldest and most important applications of electrochemistry is to the storage and conversion of energy. You already know that a *galvanic cell* converts chemical energy to work; similarly, an *electrolytic cell* converts electrical work into *chemical free energy*. Devices that carry out these conversions are called *batteries*. In ordinary batteries the chemical components are contained within the device itself. If the reactants are supplied from an external source as they are consumed, the device is called a *fuel cell*.

Introduction

The term *battery* derives from the older use of this word to describe physical attack or "beating"; Benjamin Franklin first applied the term to the electrical shocks that could be produced by an array of charged glass plates. In common usage, the term "call" is often used in place of battery. For portable and transportation applications especially, a battery or fuel cell should store (and be able to deliver) the maximum amount of energy at the desired rate (power level) from a device that has the smallest possible weight and volume. The following parameters are commonly used to express these attributes:

- Storage capacity or *charge density*, coulombs/liter or coulombs/kg;
- *Energy density*, J/kg or watt-hour/lb
- *Power density*, watts/kg
- *Voltage efficiency*, ratio of output voltage to E°
- Lifetime: *shelf-life* (resistance to self-discharge) or *charge/recharge cycles*

Primary and Secondary Batteries

A *secondary* or *storage battery* is capable of being recharged; its electrode reactions can proceed in either direction. During charging, electrical work is done on the cell to provide the free energy needed to force the reaction in the non-spontaneous direction. A *primary cell*, as expemplified by an ordinary flashlight battery, cannot be recharged with any efficiency, so the amount of energy it can deliver is limited to that obtainable from the reactants that were placed in it at the time of manufacture.

The lead-acid storage cell

The most well-known storage cell is the lead-acid cell, which was invented by Gaston Planté in 1859 and is still the most widely used device of its type. The cell is represented by

$$Pb(s)|PbSO_4(s)|H_2SO_4(aq)||PbSO_4(s), PbO_2(s)|Pb(s) \tag{16.6.1}$$

and the net cell reaction is

$$Pb(s) + PbO_2(s) + 2H_2SO_4(aq) \rightarrow 2PbSO_4(s) + 2H_2O \tag{16.6.2}$$

The reaction proceeds to the right during discharge and to the left during charging. The state of charge can be estimated by measuring the density of the electrolyte; sulfuric acid is about twice as dense as water, so as the cell is discharged, the density of the electrolyte decreases.

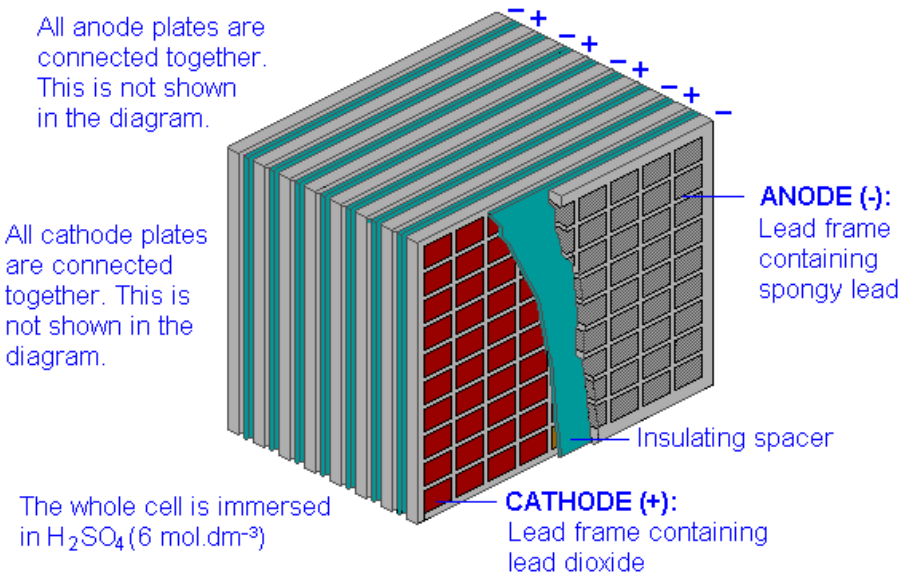


Figure 16.6.1: Lead-acid storage cell and detail of its plate construction

The technology of lead-acid storage batteries has undergone remarkably little change since the late 19th century. Their main drawback as power sources for electric vehicles is the weight of the lead; the maximum energy density is only about 35 Ah/kg, and actual values may be only half as much. There are also a few other problems:

- The sulfuric acid electrolyte becomes quite viscous when the temperature is low, inhibiting the flow of ions between the plates and reducing the current that can be delivered. This effect is well-known to anyone who has had difficulty starting a car in cold weather.
- These batteries tend to slowly self-discharge, so a car left idle for several weeks might be unable to start.
- Over time, $PbSO_4$ that does not get converted to PbO_2 due to lack of complete discharge gradually changes to an inert form which limits the battery capacity. Also, "fast" charging causes rapid evolution of hydrogen from the water in the electrolyte; the bubbles form on the lead surface and can tear PbO_2 off the positive plate. Eventually enough solid material accumulates at the bottom of the electrolyte to short-circuit the battery, leading to its permanent demise.

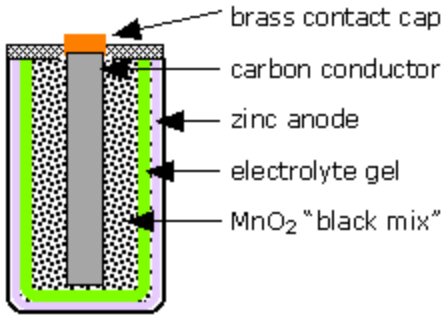
The LeClanché "dry cell"

The most well-known primary battery has long been the common "dry cell" that is widely used to power flashlights and similar devices. The modern dry cell is based on the one invented by Georges Leclanché in 1866. The electrode reactions are

$$Zn \rightarrow Zn^{2+} + 2e^- \tag{16.6.3}$$

$$2MnO_2 + 2H^+ + 2e^- \rightarrow Mn_2O_3 + H_2O \tag{16.6.4}$$

Despite its name, this cell is not really "dry"; the electrolyte is a wet paste containing NH_4Cl to supply the hydrogen ions. The chemistry of this cell is more complicated than it would appear from these equations, and there are many side reactions and these cells have limited shelf-lifes due to self discharge. (In some of the older ones, attack by the acidic ammonium ion on the zinc would release hydrogen gas, causing the battery to swell and rupture, often ruining an unused flashlight or other device.) A more modern version, introduced in 1949, is the *alkaline cell* which employs a KOH electrolyte and a zinc-powder anode which permits the cell to deliver higher currents and avoids the corrosive effects of the acidic ammonium ion on the zinc.



Exercise 16.6.1

What is the cell notation for the LeClanché dry cell?

The most important of these are:

- *Effective surface area of the electrode*. A 1-cm^2 sheet of polished metal presents far less active surface than does one that contains numerous surface projections or pores. All useful batteries and fuel cells employ highly porous electrodes. Recent advances in nanotechnology are likely to greatly improve this parameter.
- *Current density of electrode surface*. Expressed in amperes m^{-2}, this is essentially a measure of the catalytic ability of the electrode, that is, its ability to reduce the activation energy of the electron transfer process.
- *Rate at which electroactive components can be delivered to or depart from the active electrode surface*. These processes are controlled by thermal diffusion and are inhibited by the very narrow pores that are needed to produce the large active surface area.
- *Side reactions and irreversible processes*. The products of the discharge reaction may tend to react with the charge-storing components. Thermal diffusion can also cause self-discharge, limiting the shelf life of the battery. Recharging of some storage batteries may lead to formation of less active modifications of solid phases, thus reducing the number of charge/discharge cycles possible.

Clearly, these are all primarily kinetic and mechanistic factors which require a great deal of experimentation to understand and optimize.

The Fuel Cell

Conventional batteries supply electrical energy from the chemical reactants stored within them; when these reactants are consumed, the battery is "dead". An alternative approach would be to feed the reactants into the cell as they are required, so as to permit the cell to operate continuously. In this case the reactants can be thought of as "fuel" to drive the cell, hence the term fuel cell.

Although fuel cells were not employed for practical purposes until space exploration began in the 1960's, the principle was first demonstrated in 1839 by Sir William Grove, a Welsh lawyer and amateur chemist. At the time, it was already known that water could be decomposed into hydrogen and oxygen by electrolysis; Grove tried recombining the two gases in a simple apparatus, and discovered what he called "reverse electrolysis"— that is, the recombination of H_2 and O_2 into water— causing a potential difference to be generated between the two electrodes:

	$H_2(g) \rightarrow 2\ H^+ + 2e^-$	$E° = 0$ v
	$½\ O_2 + 2\ H^+ + 2e^- \rightarrow H_2O(l)$	$E° = +1.23$ v
	$H_2(g) + ½\ O_2(g) \rightarrow H_2O(l)$	$E° = +1.23$ v

It was not until 1959 that the first working hydrogen-oxygen fuel cell was developed by Francis Thomas Bacon in England. Modern cells employ an alkaline electrolyte, so the electrode reactions differ from the one shown above by the addition of OH^- to both sides of the equations (note that the net reaction is the same):

	$H_2(g) + 2\ OH^- \rightarrow 2\ H_2O + 2\ e^-$	$E° = 0$ v
	$½\ O_2\ (g) + 2\ H_2O + 2\ e^- \rightarrow 2\ OH^-$	$E° = +1.23$ v
	$H_2(g) + ½\ O_2(g) \rightarrow H_2O$	$E° = +1.23$ v

Although hydrogen has the largest energy-to-mass ratio of any fuel, it cannot be compressed to a liquid at ordinary temperatures. If it is stored as a gas, the very high pressures require heavy storage containers, greatly reducing its effective energy density. Some solid materials capable of absorbing large amount of H_2 can reduce the required pressure. Other fuels such as alcohols, hydrocarbon liquids, and even coal slurries have been used; methanol appears to be an especially promising fuel.

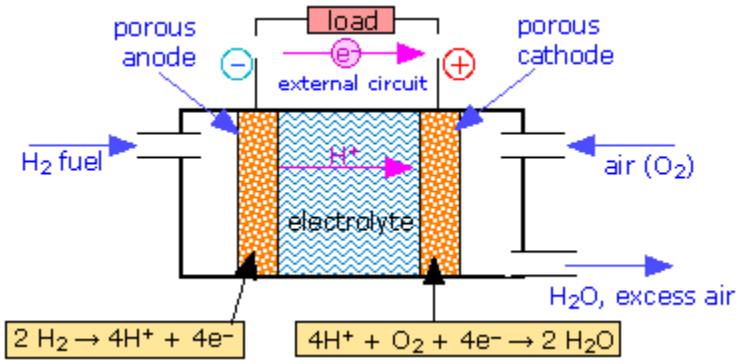


Figure 16.6.2: Schematic diagram of a modern hydrogen-oxygen fuel cell. Commonly used electrolytes are NaOH solution, phosphoric acid, or solid oxides. A major limitation of any oxygen-consuming fuel cell is the slow rate of the reduction of this element at a cathode. The best cathode surfaces are usually made of platinum, which is a major cost factor in fuel cell design.

One reason for the interest in fuel cells is that they offer a far more efficient way of utilizing chemical energy than does conventional thermal conversion. The work obtainable in the limit of reversible operation of a fuel cell is 229 kJ per mole of H_2O formed. If the hydrogen were simply burned in oxygen, the heat obtainable would be $\Delta H = 242$ kJ mol^{-1}, but no more than about half of this heat can be converted into work so the output would not exceed 121 kJ mol^{-1}. This limit is a consequence of the Second Law of Thermodynamics. The fraction of heat that can be converted into work (η) is a function of how far (in temperature) the heat falls as it flows through the engine and into the surroundings; this fraction is given by

$$\eta = \frac{1 - T_{high}}{T_{low}} \tag{16.6.5}$$

At normal environmental temperatures of around 300 K, this would have to be at least 600 K for 50% thermal efficiency.

The major limitation of present fuel cells is that the rates of the electrode reactions, especially the one in which oxygen is reduced, tend to be very small, and thus so is the output current per unit of electrode surface. Coating the electrode with a suitable catalytic material is almost always necessary to obtain usable output currents, but good catalysts are mostly very expensive substances such as platinum, so that the resulting cells are too costly for most practical uses. There is no doubt that if an efficient, low-cost catalytic electrode surface is ever developed, the fuel cell would become a mainstay of the energy economy.

Certain types of bacteria are able to oxidize organic compounds to carbon dioxide while directly transferring electrons to electrodes. These so-called *electricigen* organisms may make it possible to convert renewable biomass and organic waste directly into electricity without the wasted energy and pollution produced by direct combustion. In one experiment, a graphite electrode immersed in ordinary mud (containing humic materials) was able to produce measurable amounts of electricity.

Summary

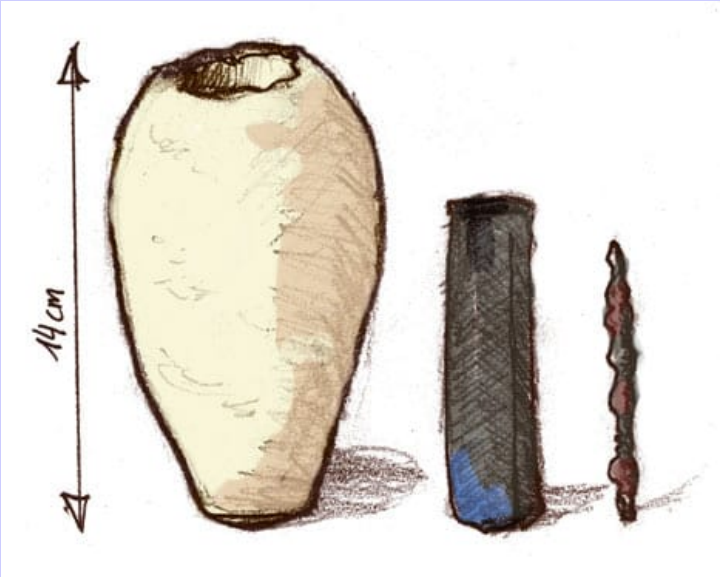
Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially imortant that you know the precise meanings of all the highlighted terms in the context of this topic.

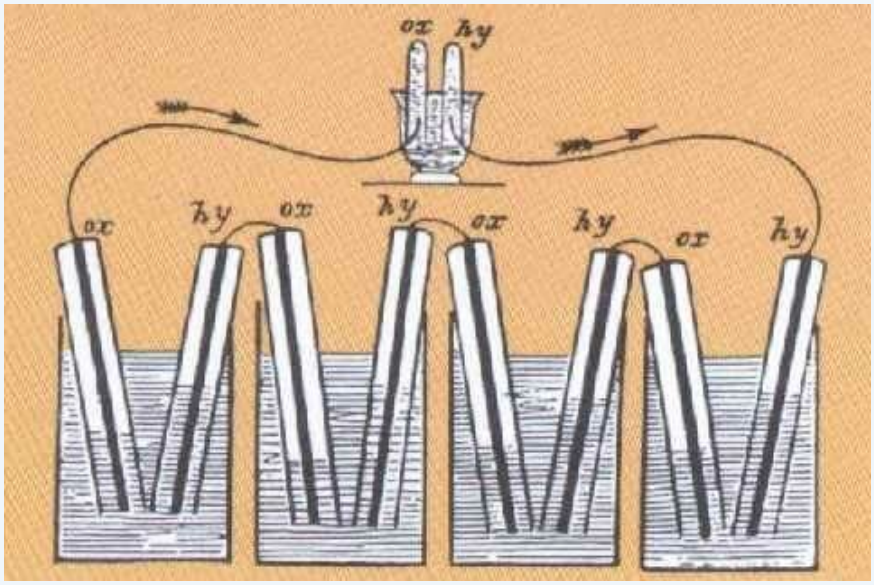
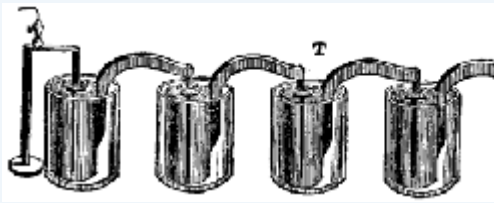
- A battery is a galvanic cell in which some of the free energy change associated with a spontaneous electron-transfer reaction is captured in the form of electrical energy.
- A secondary or storage battery is one in which the electron-transfer reaction can be reversed by applying a charging current from an external source.
- A fuel cell is a special type of battery in which the reactants are supplied from an external source as power is produced. In most practical fuel cells, H^+ ions are produced at the anode (either from H_2 or a hydrocarbon) and oxygen from the air is reduced to H2O at the cathode.
- The cathodic reduction of O_2 is kinetically limited, necessitating the use of electrode surfaces having high catalytic activity.
- The electrodes in batteries must have very high effective surface areas, and thus be highly porous. This requirement may conflict with the other important one of efficient diffusion of reactants and products in the narrow channels within the pores.
- Batteries and fuel cells designed to power vehicles and portable devices need to have high charge-to-weight and charge-tovolume ratios.

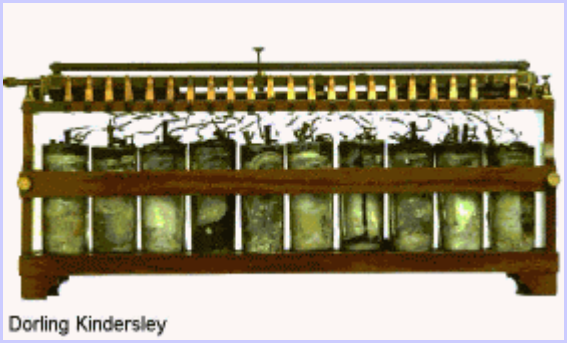
This page titled 16.6: Batteries and Fuel Cells is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.7: Timeline of Battery Development

Although the development practical batteries largely paralelled the expansion of electrical technology from about the mid-19th century on, it is now thought that a very primitive kind of battery was apparently in use more than 2000 years ago. The brief popularity of electrically powered automobiles in the 1920's encouraged storage battery development. The widespread use of portable "personal" electrical devices has kept the search for better batteries very much alive.

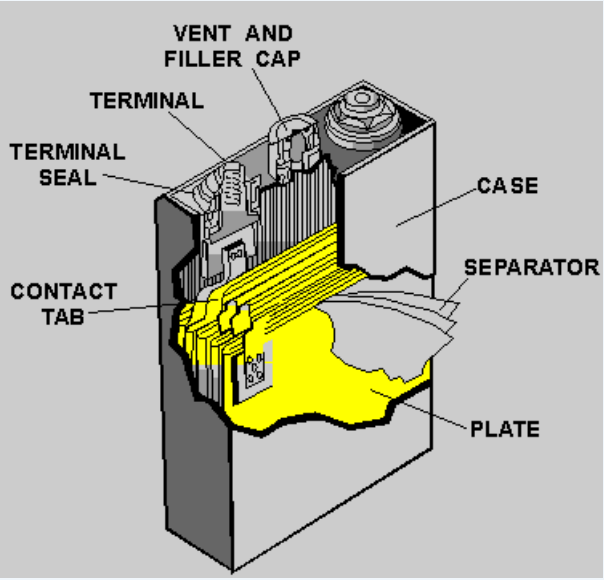
"Baghdad Battery" - 1000 BCE?  *Drawing of the three pieces. (CC-BY-SA 2.5; Ironie)*	Earthenware jars containing an iron rod surro copper cylinder were discovered near Baghda They are believed to have been used by the P civilization that occupied the region about 20 as a source of electricity to plate gold onto si
Allesandro Volta 1782 	His "Voltaic pile", a stack of zinc and silver disk by a wet cloth containing a salt or a weak acid the first battery known to Western civilization.

Sir Humphry Davy 1813	Davy builds a 2000-plate battery that occupies feet in the basement of Britain's Royal Society. batteries provided power for the first public de of electric lighting (carbon arc).
Michael Faraday, 1830's	Faraday discovered the fundamentals of galvar electrolysis that put electrochemistry on a firm basis.
1836 - Daniell cell (also known as a Crow's Foot or Gravity cell.) 	John Daniell (English chemist and meterologis the **first modern storage cell** based on Farad principles. This consists of a large glass jar wit star-shaped electrode in the bottom and a zinc shaped electrode suspended near the top. The jar was filled with a concentrated copper sulfat top of this was poured dilute sulfuric acid, who density kept it on top. This was the first practic find wide use to power telegraphs and railway s systems and home doorbells.
1839 - William Grove (Welsh) 	Grove was best known in the 19th century for acid battery" which came into wide use in ear  Now, however, he is most famous for his "gas battery" in which discovered "reverse electroly recombination of H_2 and O_2 following electrol at platinum electrodes. This was the first dem what we now know as the hydrogen-oxygen **fu** below.)

1859 - Gaston Planté (French)  	Invents the first **lead-acid storage cell** which two sheets of lead separated by a rubber sheet spiral and immersed in dilute sulfuric acid.
1866 - Georges Leclanché (French) 	By 1868 twenty thousand Leclanché cells were telegraph systems. The original Leclanché cells porous pots which were heavy and subject to b Within twenty years other inventors had modifi into what we now know as "dry cells" which bec used in the first flashlights (1909) and in batte radios of the 1920s.
1881 - Faure and others 	Development of the **first practical lead-acid** The major improvement over Planté's design w addition of a paste of $PbSO_4$ to the positive pl

1905 Nickel-iron cell **(Thomas Edison)** 	Edison, who was as much a chemist as an all- inventor, thought that the lead in Planté-type them too heavy, and that having acid in conta metal was an inherently bad idea. After much experimentation, he developed a successful a battery. The **Edison cell** uses an iron anode, cathode, and KOH electrolyte. This cell is extr and is still used in certain industrial applicati was never able to displace the lead-acid cell hoped.
1950s	A similar cell, employing a nickel anode instea the first rechargeable cell that was small enou in portable consumer devices. Its main disadv it is ruined by complete discharge.
1949 - Alkaline dry cell - Lew Urry (Eveready Battery Co.)	First commercial alkaline dry cell. These substi the corrosive NH_4Cl used in the older dry cells times longer.
1947 - Mercury cell (Ruben and Mallory, 1950's)	This was one of the first "button"-type cells wh widely used in cameras and hearing aids. The the 1.34 v output made them popular for use i instruments and cardiac pacemakers. The net is $Zn(s) + HgO(s) \rightarrow ZnO(s) + Hg(l)$ Most countries have outlawed sales of these to reduce mercury contamination of the envir

Nickel-Cadmium (NiCad) cells

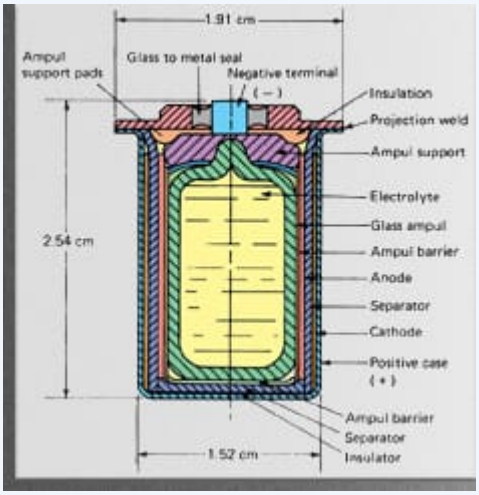


The NiCad cell quickly become one of the mos
rechargeable batteries for small consumer dev
deliver high current and undergo hundreds of c
discharge cycles. Because cadmium is an envi
toxin, their use is being discouraged.

1959 - Fuel cell - Francis Bacon (UK)

The first practical fuel cell was developed by Br
Francis Bacon (1904-1992). This hydrogen-oxy
an alkaline electrolyte and inexpensive nickel e

Late 1960's - Nickel-metal hydride cells

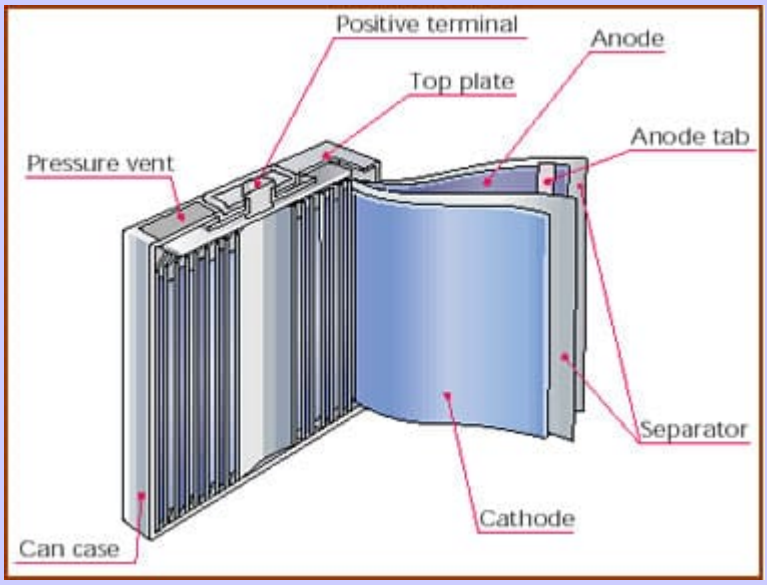


The hydride ion H^- would be an ideal cathode
except for the fact that its oxidation product H
The discovery that certain compounds such a
$ZrNi_2$ can act as "hydrogen sponges" made it
employ metal hydrides as a cathode material.
peculiarity of Ni-MH cells is that recharging th
exothermic process, so that proper dissipatio
must be allowed for. These batteries are wide
phones, computers, and portable power tools.
electrode reactions take place in a concentra
electrolyte:

Cathode (+): $NiOOH + H_2O + e^- \rightarrow Ni(OH)_2 +$

Anode (-): $(1/x)\ MH_x + OH^- \rightarrow (1/x)\ M + H_2O +$

1990s - Lithium cells (Sony Corp.)



Lithium is an ideal anode material owing to it and high reduction potential, making Li-base most compact ways of storing electrical ener cells are used in wristwatches, cardiac pacem digital cameras. Both primary (non-rechargeb rechargeable types have been available for sc More recent applications are in portable powe perhaps most importantly, in electric-powere automobiles.

Modern lithium cells operate by transporting between electrodes into which the ions can b intercalated. Cathodes are lithium transition-m such as $LiCoO_3$, while anodes are lithium-con carbon, LiC_6. The species that undergoes oxi reduction is not lithium, but the transition me Co(III)-Co(IV).

Lithium batteries as incendiary devices

There have been numerous reports of fires and explosions associated with lithium batteries. In 2006, the Dell Corporation had to recall 4.1 million Sony batteries that had been shipped with Dell's laptop computers and were judged to be at risk owing to a manufacturing defect. This illustrates the difficulty of concentrating a large amount of chemical energy into a small package, which is of course the goal of all battery developers eager to meet commercial demands ranging from consumer personal electronics to electrically-powered cars. The fully-charged Li^+-deficient lithium cobalt oxide cathodes are inherently unstable, held in check only by a thin insulating membrane which, if accidentally breached, can lead to thermal runaway involving gaseous oxygen, carbon, organic solvents, and (in some cases) lithium chlorate— all the components necessary for a fierce fire.

Much research has gone into the development of fail-safe membranes. In one type, made by ExxonMobil and targeted at the automotive market, the pores are designed to close up and thus inhibit the passage of lithium ions when the temperature rises above a safe level.

Biological Batteries

Finally, we should mention the biological batteries that are found in a number of **electric fish**. The "electric organs" of these fish are modified muscle cells known as *electrocytes* which are arranged in long stacks. A neural signal from the brain causes all the electrocytes in a stack to become polarized simultaneously, in effect creating a battery made of series-connected cells. Most electric fish produce only a small voltage which they use for navigation, much in the way that bats use sound for echo-location of prey. The famous electric eel, however, is able to produce a 600-volt jolt that it employs to stun nearby prey.

This page titled 16.7: Timeline of Battery Development is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.8: Electrochemical Corrosion

Learning Objectives

Make sure you thoroughly understand the following essential ideas. It is especially important that you know the precise meanings of all the highlighted terms in the context of this topic.

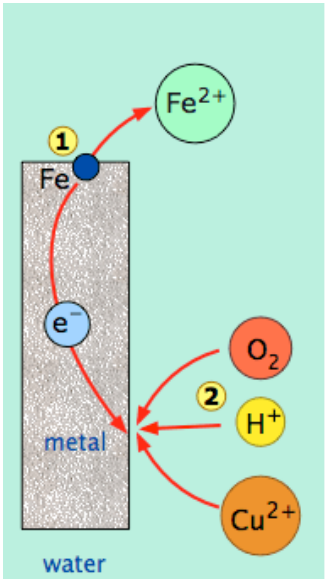
- Electrochemical corrosion of metals occurs when electrons from atoms at the surface of the metal are transferred to a suitable electron acceptor or *depolarizer*. Water must be present to serve as a medium for the transport of ions.
- The most common depolarizers are oxygen, acids, and the cations of less active metals.
- Because the electrons flow through the metallic object itself, the anodic and cathodic regions (the two halves of the electrochemical cell) can be at widely separated locations.
- Anodic regions tend to develop at locations where the metal is stressed or is protected from oxygen.
- Contact with a different kind of metal, either direct or indirect, can lead to corrosion of the more active one.
- Corrosion of steel can be inhibited by *galvanizing*, that is, by coating it with zinc, a more active metal whose dissolution leaves a negative charge on the metal which inhibits the further dissolution of Fe^{2+}.
- *Cathodic protection* using an external voltage source is widely used to protect underground structures such as tanks, pipelines and piers. The source can be a sacrificial anode of zinc or aluminum, or a line-operated or photovoltaic power supply.

Corrosion can be defined as the deterioration of materials by chemical processes. Of these, the most important by far is electrochemical corrosion of metals, in which the oxidation process $M \rightarrow M^{+} + e^{-}$ is facilitated by the presence of a suitable electron acceptor, sometimes referred to in corrosion science as a *depolarizer*.

In a sense, corrosion can be viewed as the spontaneous return of metals to their ores; the huge quantities of energy that were consumed in mining, refining, and manufacturing metals into useful objects is dissipated by a variety of different routes. The economic aspects of corrosion are far greater than most people realize; the estimated cost of corrosion in the U.S. alone was $276 billion per year. Of this, about $121 billion was spent to control corrosion, leaving the difference of $155 billion as the net loss to the economy. Utilities, especially drinking water and sewer systems, suffer the largest economic impact, with motor vehicles and transportation being a close second.

Corrosion Cells and Reactions

The special characteristic of most corrosion processes is that the oxidation and reduction steps occur at separate locations on the metal. This is possible because metals are conductive, so the electrons can flow through the metal from the anodic to the cathodic regions (Figure 16.8.1). The presence of water is necessary in order to transport ions to and from the metal, but a thin film of adsorbed moisture can be sufficient.



1 $Fe \rightarrow Fe^{2+} + 2e^-$

Fe atom at metal surface dissolves into moisture film, leaving negative charge in metal.

2 Corrosion continues as a *depolarizer* removes electrons from metal. Common depolarizers are:

oxygen:
$O_2 + 4H_2O \rightarrow 4e^- + 4OH^-$

acid: $2H^+ + 2e^- \rightarrow H_2$

cation of more-noble metal:
$Cu^{2+} + 2e^- \rightarrow Cu$

Figure 16.8.1: Corrosion is a two-step process. Figure 16.8.1: Electrochemical corrosion of iron. Corrosion often begins at a location (1) where the metal is under stress (at a bend or weld) or is isolated from the air (where two pieces of metal are joined or under a loosely-adhering paint film.) The metal ions dissolve in the moisture film and the electrons migrate to another location (2) where they are taken up by a *depolarizer*. Oxygen is the most common depolarizer; the resulting hydroxide ions react with the Fe^{2+} to form the mixture of hydrous iron oxides known as *rust*. (CC BY 3.0 Unported; Stephen Lower)

A corrosion system can be regarded as a short-circuited electrochemical cell in which the anodic process is something like

$$\mathrm{Fe\,(s)} \rightarrow \mathrm{Fe^{2+}\,(aq)} + 2\,\mathrm{e^-} \tag{16.8.1}$$

and the cathodic steps may invove the reduction of oxygen gas

$$\mathrm{O_2} + 2\,\mathrm{H_2O} + 4\,\mathrm{e^-} \rightarrow 4\,\mathrm{OH^-} \tag{16.8.2}$$

or the reduction of protons

$$\mathrm{H^+} + \mathrm{e^-} \rightarrow \tfrac{1}{2}\,\mathrm{H_2\,(g)} \tag{16.8.3}$$

or the reduction of a metal ion

$$\mathrm{M^{2+}} + 2\,\mathrm{e^-} \rightarrow \mathrm{M\,(s)} \tag{16.8.4}$$

where M is a metal.

Which parts of the metal serve as anodes and cathodes can depend on many factors, as can be seen from the irregular corrosion patterns that are commonly observed. Atoms in regions that have undergone stress, as might be produced by forming or machining, often tend to have higher free energies, and thus tend to become anodic.

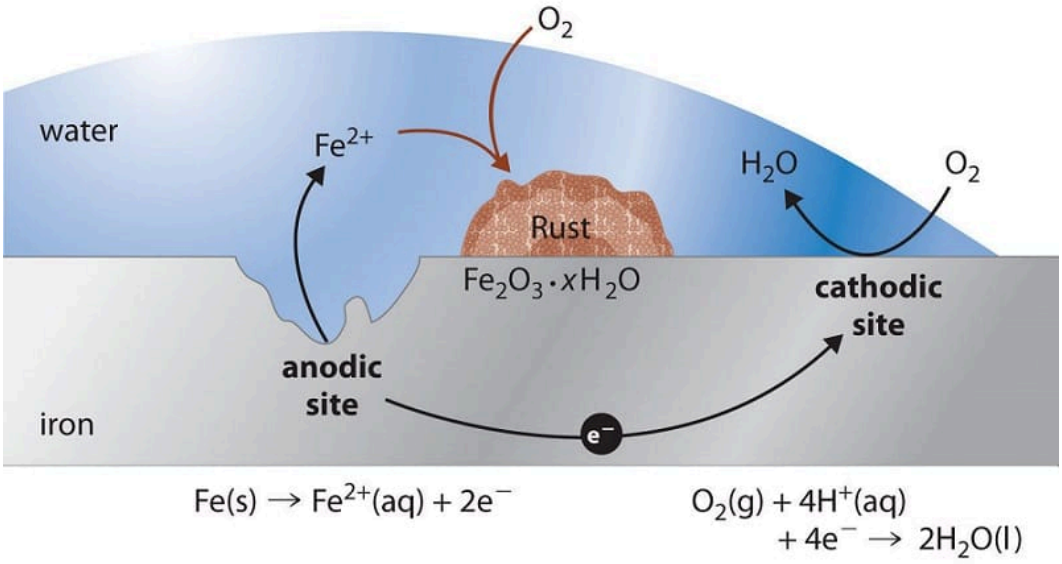


Figure 16.8.2: Schematic diagram of corrosion cells on iron. (CC BY-NSA-NC; Anonymous by request)

If one part of a metallic object is protected from the atmosphere so that there is insufficient O_2 to build or maintain the oxide film, this "protected" region will often be the site at which corrosion is most active. The fact that such sites are usually hidden from view accounts for much of the difficulty in detecting and controlling corrosion.

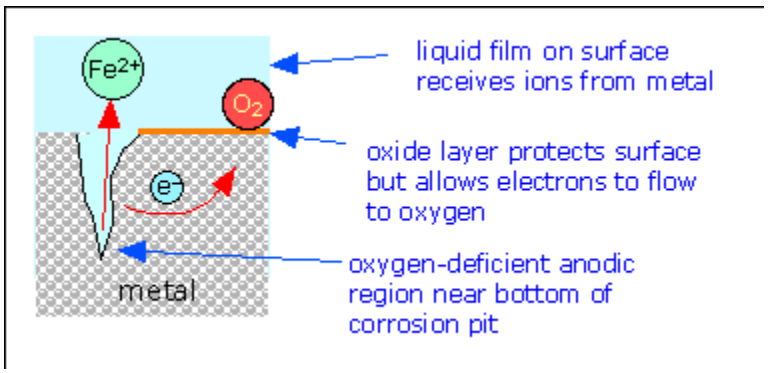


Figure 16.8.3: Pitting corrosion Most metals are covered with a thin oxide film which inhibits anodic dissolution. When corrosion does occur, it sometimes hollows out a narrow hole or pit in the metal. The bottoms of these pits tend to be deprived of oxygen, thus promoting further growth of the pit into the metal. (CC BY 3.0 Unported; Stephen Lower)

In contrast to anodic sites, which tend to be localized to specific regions of the surface, the cathodic part of the process can occur almost anywhere. Because metallic oxides are usually semiconductors, most oxide coatings do not inhibit the flow of electrons to the surface, so almost any region that is exposed to O_2 or to some other electron acceptor can act as a cathode. The tendency of oxygen-deprived locations to become anodic is the cause of many commonly-observed patterns of corrosion.

Anyone who has owned an older car has seen corrosion occur at joints between body parts and under paint films. You will also have noticed that once corrosion starts, it tends to feed on itself. One reason for this is that one of the products of the O_2 reduction reaction is hydroxide ion. The high pH produced in these cathodic regions tends to destroy the protective oxide film, and may even soften or weaken paint films, so that these sites can become anodic. The greater supply of electrons promotes more intense cathodic action, which spawns even more anodic sites, and so on.



Figure 16.8.4: Rusted out car. Severely rusted floorboards of 1990 Chrysler New Yorker. Car was damaged to the point where it became unsafe to drive. Under the rusting section between the two holes were the two brake lines (Public Domain; Bige1977 via Wikipedia).

A very common cause of corrosion is having two dissimilar metals in contact, as might occur near a fastener or at a weld joint. Moisture collects at the junction point, acting as an electrolyte and forming a cell in which the two metals serve as electrodes. Moisture and conductive salts on the outside surfaces provide an external conductive path, effectively short-circuiting the cell and producing very rapid corrosion; this is why cars rust out so quickly in places where salt is placed on roads to melt ice.

Dissimilar-metal corrosion can occur even if the two metals are not initially in direct contact. For example, in homes where copper tubing is used for plumbing, there is always a small amount of dissolved Cu^{2+} in the water. When this water encounters steel piping or a chrome-plated bathroom sink drain, the more-noble copper will plate out on the other metal, producing a new metals-in-contact corrosion cell. In the case of chrome bathroom sink fittings, this leads to the formation of Cr^{3+} salts which precipitate as greenish stains.

Control of Corrosion

Since both the cathodic and anodic steps must take place for corrosion to occur, prevention of either one will stop corrosion. The most obvious strategy is to stop both processes by coating the object with a paint or other protective coating. Even if this is done, there are likely to be places where the coating is broken or does not penetrate, particularly if there are holes or screw threads. A more sophisticated approach is to apply a slight negative charge to the metal, thus making it more difficult for the reaction to take place:

$$\mathrm{M} \longrightarrow \mathrm{M}^{2+} + 2\,\mathrm{e}^{-}. \tag{16.8.5}$$

Protection Method 1: Sacrificial Coatings

One way of supplying this negative charge is to apply a coating of a more active metal. Thus a very common way of protecting steel from corrosion is to coat it with a thin layer of zinc; this process is known as galvanizing.The zinc coating, being less noble than iron, tends to corrode selectively. Dissolution of this sacrificial coating leaves behind electrons which concentrate in the iron, making it cathodic and thus inhibiting its dissolution.

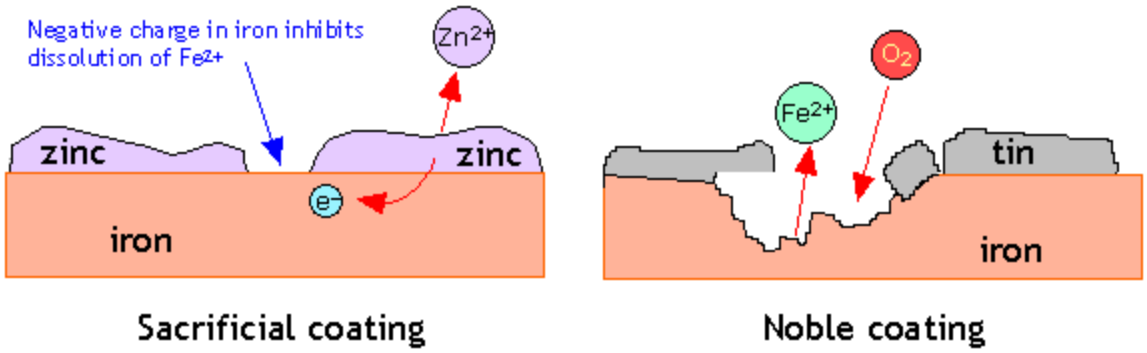


Figure 16.8.5: Sacrificial Coatings (CC BY 3.0 Unported; Stephen Lower)

The effect of plating iron with a less active metal provides an interesting contrast. The common tin-plated can (on the right) is a good example. As long as the tin coating remains intact, all is well, but exposure of even a tiny part of the underlying iron to the moist atmosphere initiates corrosion. The electrons released from the iron flow into the tin, making the iron more anodic so now the tin is actively promoting corrosion of the iron! You have probably observed how tin cans disintegrate very rapidly when left outdoors.

Protection Method 2: Cathodic Protection

A more sophisticated strategy is to maintain a continual negative electrical charge on a metal, so that its dissolution as positive ions is inhibited. Since the entire surface is forced into the cathodic condition, this method is known as *cathodic protection*. The source of electrons can be an external direct current power supply (commonly used to protect oil pipelines and other buried structures), or it can be the corrosion of another, more active metal such as a piece of zinc or aluminum buried in the ground nearby, as is shown in the illustration of the buried propane storage tank below.

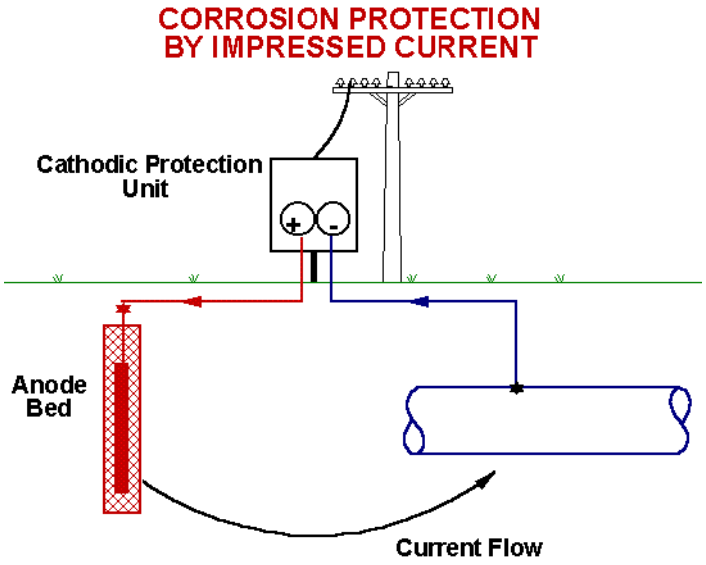
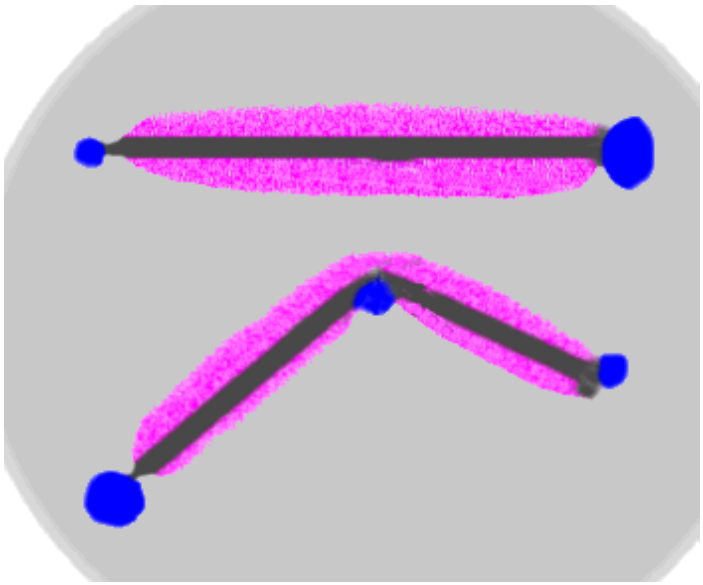


Figure 16.8.6: Cathodic Protection (CC BY 3.0 Unported; Stephen Lower)

This page titled 16.8: Electrochemical Corrosion is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.9: Corrosion Gallery

Corrosion gallery



Corrosion of a nail

The nails are immersed in agar which forms a moist so gel. The agar also contains phenolphthalein and hexacyanoiron(III) $Fe(CN_6)^-$ which forms a deep blue c ("prussian blue") in the presence of Fe^{2+}. The blue colo are clearly associated with those parts of the nail that h been stressed, thus faciliting the anodic release of Fe^2 from the metal. The the pink color shows the cathodic regions that have been made alkaline by the reaction

$$O_2 + 2\ H_2O + 4e^- \rightarrow 4\ OH^-$$

This clearly shows the separation between the anodic a cathodic processes in corrosion.

[Illustration from U of West Indies:

Water distribution main

If you live in the older part of a city where the mains are 50-100 years old, the water you drink may well have passed through a pipe in this condition! Severe corrosion like this is more common in areas where the water is acidic. Such water comes from mountain snowmelt and runoff, and usually acquires its acidity from dissolved atmospheric carbon dioxide.

Waters from rivers, lakes, and especially groundwaters from wells have usually been in sufficiently long contact with carbonate-containing sediments to have been neutralized. Water-works administrators like to make the water slightly alkaline and slightly supersaturated in calcium carbonate in order to maintain a thin coating of solid carbonate on the interior of the pipe which acts to protect it from corrosion.



Corrosion of reinforcing bars in concrete

All large concrete structures contain steel reinforcing bars ("rebars") that help ensure structural integrity under varying load conditions and especially during earthquakes. Intrusion of water, even in the form of fog or mists, can lead to serious corrosion damage, as seen in this picture of this column which supports a highway overpass.



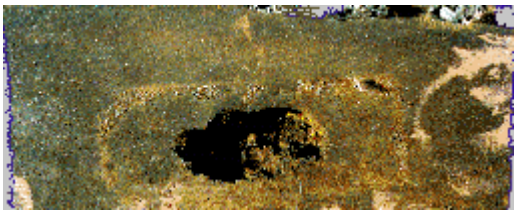


Corrosion at metallic joints

The picture shows two steel structural members joined cast iron flanges which have been bolted together. For some reason, one of the pieces has become more anoc than the other, leading to extensive corrosion of the up part.

Bacterial-assisted corrosion

This gas pipe was buried in a red soil that contained iron pyrites (FeS.) The bacterium *thiobacillus ferrooxidans* derives its energy by oxidizing Fe^{2+} to the more soluble Fe^{3+}, transferring the electrons to O_2. It also oxidizes the sulfur, producing sulfuric acid. The resulting chemical cocktail has eaten a hole into the pipe.



lin

These galvanized bolts were used to join wooden beams together. Subsequent movement of the beams due to varying load conditions abraded the zinc coating. A lack of oxygen near the centers of the bolts also likely contributed to the corrosion by preventing the formation of a protective oxide film.



Pitting corrosion



This page titled 16.9: Corrosion Gallery is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

16.10: Electrolytic Cells and Electrolysis

Learning Objectives

Make sure you thoroughly understand the following essential ideas.

- In electrolysis, an external power source supplies the free energy required to drive a cell reaction in its non-spontaneous direction. An electrolytic cell is in this sense the opposite of a galvanic cell. In practice, the products of electrolysis are usually simpler than the reactants, hence the term electro-*lysis*.
- Transport of ions in the electrolyte in response to the potential difference between the electrodes ("drift") is largely restricted to the regions very close to the electrodes. Ionic transport in the greater part of the electrolyte is by ordinary thermal diffusion— the statistical tendency of concentrations to become uniform.
- When an aqueous solution is subjected to electrolysis, the oxidation or reduction of water can be a competing process and may dominate if the applied voltage is sufficiently great. Thus an attempt to electrolyze a solution of $NaNO_3$ will produce only H_2 and O_2.
- A large number of electrolysis processes are employed by industry to refine metals and to produce both inorganic and organic products. The largest of these are the chloralkali industry (chlorine and "caustic"), and the refining of aluminum; the latter consumes approximately 5% of the electrical power generated in North America.

Electrolysis refers to the decomposition of a substance by an electric current. The electrolysis of sodium and potassium hydroxides, first carried out in 1808 by Sir Humphrey Davey, led to the discovery of these two metallic elements and showed that these two hydroxides which had previously been considered un-decomposable and thus elements, were in fact compounds.

Electrolysis of molten alkali halides is the usual industrial method of preparing the alkali metals:

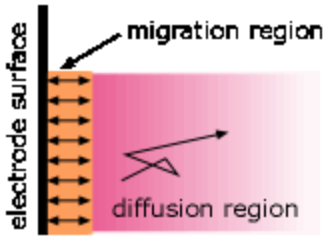
$Na^+ + e^- \rightarrow Na(l)$	$E° = -2.71$ V
$Cl^- \rightarrow ½\ Cl_2(g) + e^-$	$E° = -1.36$ V
$Na^+ + Cl^- \rightarrow Na(l) + ½\ Cl_2(g)$	$E° = -4.1$ V

Ions in aqueous solutions can undergo similar reactions. Thus if a solution of nickel chloride undergoes electrolysis at platinum electrodes, the reactions are

$Ni^{2+} + 2\ e^- \rightarrow Ni(s)$	$E° = -0.24$ V
$2\ Cl^- \rightarrow Cl_2(g) + 2\ e^-$	$E° = -1.36$ V
$Ni^{2+} + 2\ Cl^- \rightarrow Ni(s) + Cl_2(g)$	$E° = -1.60$ v

Both of these processes are carried out in electrochemical cells which are forced to operate in the "reverse", or non-spontaneous direction, as indicated by the negative for the above cell reaction. The free energy is supplied in the form of electrical work done on the system by the outside world (the surroundings). This is the only fundamental difference between an *electrolytic cell* and the *galvanic cell* in which the free energy supplied by the cell reaction is extracted as work done on the surroundings.

A common misconception about electrolysis is that "ions are attracted to the oppositely-charged electrode." This is true only in the very thin interfacial region near the electrode surface. Ionic motion throughout the bulk of the solution occurs mostly by diffusion, which is the transport of molecules in response to a concentration gradient. Migration— the motion of a charged particle due to an applied electric field, is only a minor player, producing only about one non-random jump out of around 100,000 random ones for a 1 volt cm^{-1} electric field. Only those ions that are near the interfacial region are likely to undergo migration.



Electrolysis in aqueous solutions

Water is capable of undergoing both *oxidation*

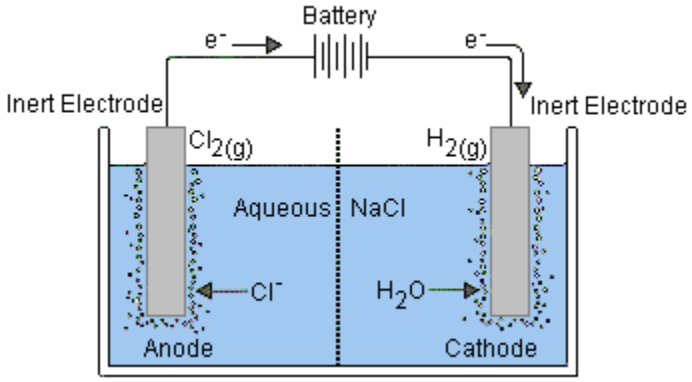
$$H_2O \rightarrow O_{2(g)} + 4H^+ + 2e^- \quad E^o = -1.23V \tag{16.10.1}$$

and *reduction*

$$2H_2O + 2e^- \rightarrow H_{2(g)} + 2OH^- \quad E^o = -0.83\ V \tag{16.10.2}$$

Thus if an aqueous solution is subjected to electrolysis, one or both of the above reactions may be able to compete with the electrolysis of the solute. For example, if we try to electrolyze a solution of sodium chloride, hydrogen is produced at the cathode instead of sodium:

	$H_2O + 2\ e^- \rightarrow H_2(g) + 2\ OH^-$	E =+0.41 V ($[OH^-] = 10^{-7}$ M)
	$Cl^- \rightarrow ½\ Cl_2(g) + e^-$	*E*° = –1.36 V
	$Cl^- + H_2O \rightarrow 2\ H_2(g) + ½\ Cl_2(g) + 2\ OH^-$	*E* = –0.95 V



[This illustration is taken from the excellent Purdue University Chemistry site]Reduction of Na+ (E° = –2.7 v) is energetically more difficult than the reduction of water (–1.23 V), so in aqueous solution the latter will prevail.

Electrolysis of salt ("brine") is carried out on a huge scale and is the basis of the *chloralkali* industry.

Electrolysis of water

Pure water is an insulator and cannot undergo significant electrolysis without adding an electrolyte. If the object is to produce hydrogen and oxygen, the electrolyte must be energetically more difficult to oxidize or reduce than water itself. Electrolysis of a solution of sulfuric acid or of a salt such as $NaNO_3$ results in the decomposition of water at both electrodes:

	$H_2O + 2\ e^- \rightarrow H_2(g) + 2\ OH^-$	E =+0.41 V ($[OH^-] = 10^{-7}$ M)

	$2\ H_2O \rightarrow O_2(g) + 4\ H+ + 2\ e^-$	E° = -0.82 V
	$2\ H_2O(l) \rightarrow 2\ H_2(g) + O_2(g)$	E = -1.23 V

Electrolytic production of hydrogen is usually carried out with a dilute solution of sulfuric acid. This process is generally too expensive for industrial production unless highly pure hydrogen is required. However, it becomes more efficient at higher temperatures, where thermal energy reduces the amount of electrical energy required, so there is now some interest in developing high-temperature electrolytic processes. Most hydrogen gas is manufactured by the steam reforming of natural gas.

Faraday's laws of electrolysis

One mole of electric charge (96,500 coulombs), when passed through a cell, will discharge half a mole of a divalent metal ion such as Cu^{2+}. This relation was first formulated by Faraday in 1832 in the form of two *laws of electrolysis*:

1. The weights of substances formed at an electrode during electrolysis are directly proportional to the quantity of electricity that passes through the electrolyte.
2. The weights of different substances formed by the passage of the same quantity of electricity are proportional to the equivalent weight of each substance.

The equivalent weight of a substance is defined as the molar mass, divided by the number of electrons required to oxidize or reduce each unit of the substance. Thus one mole of V^{3+} corresponds to three equivalents of this species, and will require three faradays of charge to deposit it as metallic vanadium.

Most stoichiometric problems involving electrolysis can be solved without explicit use of Faraday's laws. The "chemistry" in these problems is usually very elementary; the major difficulties usually stem from unfamiliarity with the basic electrical units:

- *current* (amperes) is the rate of *charge transport*; 1 amp = 1 C/sec.
- *power* (watts) is the rate of *energy* production or consumption;

$$1W = 1J/sec = 1volt - amp; 1watt - sec = 1J, 1kW - h = 3600J. \tag{16.10.3}$$

Example 16.10.1

A metallic object to be plated with copper is placed in a solution of $CuSO_4$.

a. To which electrode of a direct current power supply should the object be connected?
b. What mass of copper will be deposited if a current of 0.22 amp flows through the cell for 1.5 hours?

Solution

a. Since Cu^{2+} ions are being reduced, the object acts as a cathode and must be connected to the negative terminal (where the electrons come from!)
b. The amount of charge passing through the cell is

(0.22 amp) × (5400 sec) = 1200 C

or

(1200 C) ÷ (96500 c F^{-1}) = 0.012 F

Since the reduction of one mole of Cu^{2+} ion requires the addition of two moles of electrons, the mass of Cu deposited will be

(63.54 g mol^{-1}) (0.5 mol Cu/F) (.012 F) = 0.39 g of copper

Example 16.10.2

How much electric power is required to produce 1 metric ton (1000 kg) of chlorine from brine, assuming the cells operate at 2.0 volts and assuming 100 % efficiency?

Solution

- moles of Cl_2 produced: (10^6 g) ÷ 70 g mol^{-1} = 14300 mol Cl_2
- faradays of charge: (2 F/mol) × (14300 mol) = 28600 F
- charge in coulombs: (96500 C/F) × (28600 F) = 2.76 × 10^9 C
- duration of electrolysis: (3600 s/h) x (24 h) = 86400 s
- current (rate of charge delivery): (2.76 × 10^9 amp-sec) ÷ (86400 sec) = 32300 amps
- power (volt-amps): (2.0 V) × (32300 amp) = 64.6 kW
- energy in kW-h: (64.6 kW) × (24 h) = 1550 kW-h
- energy in joules: (1550 kW-h) × (3.6 MJ/kW-h) = 5580 MJ (megajoules)

(In the last step, recall that 1 W = 1 J/s, so 1 kW-h = 3.6 MJ)

Industrial Electrolytic Processes

For many industrial-scale operations involving the oxidation or reduction of both inorganic and organic substances, and especially for the production of the more active metals such as sodium, calcium, magnesium, and aluminum, the most cost-effective reducing agent is electrons supplied by an external power source. The two most economically important of these processes are described below.

The chloralkali industry

The electrolysis of brine is carried out on a huge scale for the industrial production of chlorine and caustic soda (sodium hydroxide). Because the reduction potential of Na^+ is much higher than that of water, the latter substance undergoes decomposition at the cathode, yielding hydrogen gas and OH^-.

anode reactions	$2\ Cl^- \rightarrow Cl_2(g) + 2\ e^-$	-1.36 V	i
	$4\ OH^- \rightarrow O_2(g) + 2\ H_2O + 4\ e^-$	-0.40 V	ii
cathode reactions	Na+ + e^- → Na(s)	-2.7 V	iii
	$H_2O + 2\ e^- \rightarrow H_2(g) + 2\ OH^-$	+.41 V	iv

A comparison of the *E*°s would lead us to predict that the reduction (*ii*) would be favored over that of (*i*). This is certainly the case from a purely energetic standpoint, but as was mentioned in the section on fuel cells, electrode reactions involving O_2 are notoriously slow (that is, they are kinetically hindered), so the anodic process here is under kinetic rather than thermodynamic control. The reduction of water (*iv*) is energetically favored over that of Na^+ (*iii*), so the net result of the electrolysis of brine is the production of Cl_2 and NaOH ("caustic"), both of which are of immense industrial importance:

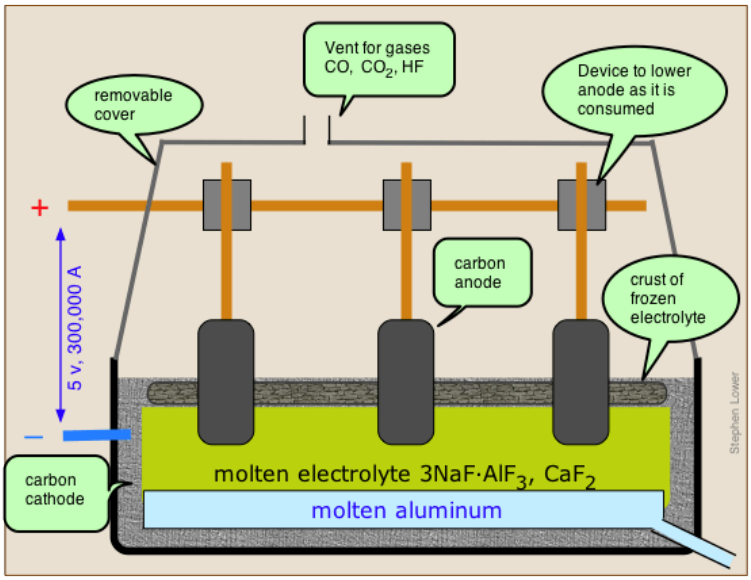
$$2\,\mathrm{NaCl} + 2\,\mathrm{H_2O} \longrightarrow 2\,\mathrm{NaOH} + \mathrm{Cl_2\,(g)} + \mathrm{H_2\,(g)}$$

Since chlorine reacts with both OH^- and H_2, it is necessary to physically separate the anode and cathode compartments. In modern plants this is accomplished by means of an ion-selective polymer membrane, but prior to 1970 a more complicated cell was used that employed a pool of mercury as the cathode. A small amount of this mercury would normally find its way into the

plant's waste stream, and this has resulted in serious pollution of many major river systems and estuaries and devastation of their fisheries.

Electrolytic refining of aluminum

Aluminum is present in most rocks and is the most abundant metallic element in the earth's crust (eight percent by weight.) However, its isolation is very difficult and expensive to accomplish by purely chemical means, as evidenced by the high E° (–1.66 V) of the Al^{3+}/Al couple. For the same reason, aluminum cannot be isolated by electrolysis of aqueous solutions of its compounds, since the water would be electrolyzed preferentially. And if you have ever tried to melt a rock, you will appreciate the difficulty of electrolyzing a molten aluminum ore! Aluminum was in fact considered an exotic and costly metal until 1886, when Charles Hall (U.S.A) and Paul Hérault (France) independently developed a practical electrolytic reduction process.



The **Hall-Hérault process** takes advantage of the principle that the melting point of a substance is reduced by admixture with another substance with which it forms a homogeneous phase. Instead of using the pure alumina ore Al_2O_3 which melts at 2050°C, it is mixed with cryolite, which is a natural mixture of NaF and AlF_3, thus reducing the temperature required to a more manageable 1000°C. The anodes of the cell are made of carbon (actually a mixture of pitch and coal), and this plays a direct role in the process; the carbon gets oxidized (by the oxide ions left over from the reduction of Al^{3+} to CO, and the free energy of this reaction helps drive the aluminum reduction, lowering the voltage that must be applied and thus reducing the power consumption. This is important, because aluminum refining is the largest consumer of industrial electricity, accounting for about 5% of all electricity generated in North America. Since aluminum cells commonly operated at about 100,000 amperes, even a slight reduction in voltage can result in a large saving of power.

The net reaction is

$$2\,Al_2O_3 + 3\,C \rightarrow 4\,Al + 3\,CO_2$$

However, large quantities of CO and of HF (from the cryolite), and hydrocarbons (from the electrodes) are formed in various side reactions, and these can be serious sources of environmental pollution.

This page titled 16.10: Electrolytic Cells and Electrolysis is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17: CHEMICAL KINETICS AND DYNAMICS

The chapters in this section deal with the rates and mechanisms of chemical change. These topics stand in contrast to the subjects of **equilibrium** and **thermodynamics** that control the *direction* of chemical change. Chemical change is guided and driven by energetics (thermodynamics), but the actual route it takes and the speed with which it occurs is the subject of "dynamics". Dynamics is itself divided into two general areas: kinetics, which deals with the rate of change and is the subject of this lesson. Mechanistics, introduced in a later lesson, is an exploration of the "road map" that links reactants to products.

17.1: Rates of reactions and rate laws

17.2: Reaction Rates Typically Change with Time

17.3: Collision and activation- the Arrhenius law

17.4: Reaction Mechanisms

17.5: Kinetics of Reactions in Solution

17.6: Catalysts and Catalysis

17.7: Experimental methods of chemical kinetics

This page titled 17: Chemical Kinetics and Dynamics is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17.1: Rates of reactions and rate laws

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Describe the contrasting roles of *thermodynamics* and *kinetics* in understanding chemical change.
- Given a balanced net equation, write an expression for the *rate of a reaction*.
- Sketch a curve showing how the *instantaneous rate* of a reaction might change with time.
- Determine the *order* of a reaction of the form A → B + C from experimental data for the concentrations of its products at successive times.
- Describe the *initial rate* and *isolation* methods of determining the orders of the individual reactants in a reaction involving multiple reactants.
- Explain the difference between *differential* and *integral rate laws*.
- Sketch out a plot showing how the concentration of a component ([A] or ln [A]) that follows *first-order kinetics* will change with time. Indicate how the magnitude of the rate constant affects this plot
- Define the *half-life* of a reaction.
- Given the half-life for a first-order reaction A → products along with the initial value of $[A]_o$, find $[A]_t$ at a subsequent time an integral number of half-lifes later.
- Describe the conditions under which a reaction can appear to have an order of zero.

Chemical change is guided and driven by energetics, but the actual route it takes and the speed with which it occurs is the subject of "dynamics". Dynamics is itself divided into two general areas: kinetics, which deals with the rate of change and is the subject of this lesson. Mechanistics, introduced in a later lesson, is an exploration of the "road map" that links reactants to products.

The energetic aspects of change are governed by the laws of thermodynamics (the "dynamics" part of this word is related to the historical origins of the field and is not a part of dynamics in the sense of these lessons.)

Energetics + dynamics = chemical change

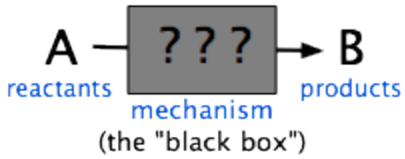
Chemical change is driven by the tendency of atoms and molecules to rearrange themselves in a way that results in the maximum possible dispersion of thermal energy into the world. The observable quantity that measures this spreading and sharing of energy is the *free energy* of the system. As a chemical change takes place, the quantities of reactants and products change in a way that leads to a more negative free energy. When the free energy reaches its minimum possible value, there is no more *net change* and the system is said to be in *equilibrium*.

The beauty of thermodynamics is that it enables us to unfailingly predict the net direction of a reaction and the composition of the equilibrium state even without conducting the experiment; the standard free energies of the reactants and products, which can be independently measured or obtained from tables, are all we need.

Thermodynamics points the way and makes it possible...but it says nothing about how long it will take to get there!

It is worth noting that the concept of "time" plays no role whatsover in thermodynamics. But kinetics is all about time. The "speed" of a reaction — how long it takes to reach equilibrium — bears no relation at all to how *spontaneous* it is (as given by the sign and value of $\Delta G°$) or whether it is exothermic or endothermic (given by the sign of $\Delta H°$). Moreover, there is no way that reaction rates can be predicted in advance; each reaction must be studied individually. One reason for this is that

The stoichiometric equation for the reaction says nothing about its *mechanism*. By mechanism, we mean, basically, "who does what to whom". Think of a reaction mechanism as something that goes on in a "black box" that joins reactants to products:



The inner workings of the black box are ordinarily hidden from us, are highly unpredictable and can only be inferred by indirect means. Consider, for example, the gas-phase formation reactions of the hydrogen halides from the elements.

The thermodynamics of these reactions are all similar (they are all highly exothermic), but their *dynamics* (their kinetics and mechanisms) could not be more different.

$H_{2(g)} + I_{2(g)} \rightarrow 2HI_{(g)}$	$H_{2(g)} + Br_{2(g)} \rightarrow 2HBr_{(g)}$	$H_{2(g)} + Cl_{2(g)} \rightarrow 2HCl_{(g)}$
Careful experiments, carried out over many years, are consistent with the simplest mechanism: a single collision between the two reactant molecules results in a rearrangement of the bonds:	One might be tempted to suppose that this would proceed in a similar way, but experiments reveal that the mechanism of this reaction is far more complex. The reaction takes place in a succession of steps, some of which involve atomic H and Br.	The mechanism of this reaction is different again. Although the first two reactions reach equilibrium in minutes to an hour or so at temperatures of 300 to 600 K, a mixture of hydrogen and chlorine will not react at all in the dark, but if you shine a light on the mixture, it goes off with a bang as the instantaneous reaction releases heat and expands the gas explosively.

What is particularly noteworthy is that these striking differences cannot be reliably predicted from theory; they were revealed only by experimentation.

The rates of chemical reactions

Chemical reactions vary greatly in the speed at which they occur. Some are essentially instantaneous, while others may take years to reach equilibrium. The speed of a chemical reaction may be defined as the change in concentration of a substance divided by the time interval during which this change is observed:

$$\text{rate} = \frac{\Delta\text{concentration}}{\Delta\text{time}} \tag{17.1.1}$$

For a reaction of the form A + B → C, the rate can be expressed in terms of the change in concentration of any of its components:

$$\text{rate} = \frac{-\Delta[A]}{\Delta t} \tag{17.1.2}$$

$$\text{rate} = \frac{-\Delta[B]}{\Delta t} \tag{17.1.3}$$

$$\text{rate} = \frac{\Delta[C]}{\Delta t} \tag{17.1.4}$$

in which Δ[A] is the difference between the concentration of A over the time interval $\Delta t = t_2 – t_1$:

$$\Delta[A] = [A]_2 – [A]_1 \tag{17.1.5}$$

Notice the minus signs in Equations 17.1.2 and 17.1.2; the concentration of a *reactant* always *decreases* with time, so Δ[A] and Δ[B] are both negative. Since negative rates does not make much sense, rates expressed in terms of a reactant concentration are always preceded by a minus sign in order to make the rate come out positive.

Consider now a reaction in which the coefficients are different:

$$A + 3B \rightarrow 2D \tag{17.1.6}$$

It is clear that [B] decreases three times as rapidly as [A], so in order to avoid ambiguity when expressing the rate in terms of different components, it is customary to divide each change in concentration by the appropriate coefficient:

$$\text{rate} = \frac{-\Delta[A]}{\Delta t} = \frac{-\Delta[B]}{3\Delta t} = \frac{+\Delta[D]}{2\Delta t} \tag{17.1.7}$$

Each of the above quotients is a legitimate expression of the rate of this particular reaction; they all yield the same number. Which one you employ when doing a calculation is largely a matter of convenience.

Example 17.1.1: Oxidation of Ammonia

For the oxidation of ammonia

$$4NH_3 + 3O_2 \rightarrow 2N_2 + 6H_2O$$

it was found that the rate of formation of N_2 was 0.27 mol L^{-1} s^{-1}.

a. At what rate was water being formed?
b. At what rate was ammonia being consumed?

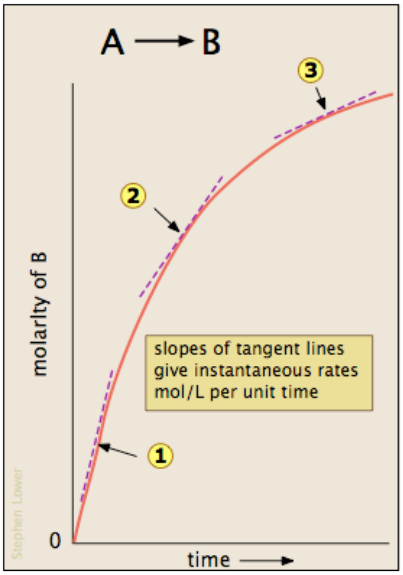
Solution:

a. From the equation stoichiometry, $\Delta[H_2O] = 6/2\ \Delta[N_2]$, so the rate of formation of H_2O is 3 × (0.27 mol L^{-1} s^{-1}) = 0.81 mol L^{-1} s^{-1}.
b. 4 moles of NH_3 are consumed for every 2 moles of N_2 formed, so the rate of disappearance of ammonia is 2 × (0.27 mol L^{-1} s^{-1}) = 0.54 mol L^{-1} s^{-1}.

Comment: Because of the way this question is formulated, it would be acceptable to express this last value as a negative number.

Instantaneous rates

Most reactions slow down as the reactants are consumed. Consequently, the rates given by the expressions shown above tend to lose their meaning when measured over longer time intervals Δt. Thus for the reaction whose progress is plotted here, the actual rate (as measured by the increasing concentration of product) varies continuously, being greatest at time zero. The *instantaneous rate* of a reaction is given by the slope of a tangent to the concentration-vs.-time curve. Three such rates have been identified in this plot.



An instantaneous rate taken near the beginning of the reaction (t = 0) is known as an *initial rate* (label **(1)** here). As we shall soon see, initial rates play an important role in the study of reaction kinetics. If you have studied differential calculus, you will know that these tangent slopes are *derivatives* whose values can very at each point on the curve, so that these instantaneous rates are really *limiting rates* defined as

$$rate = \lim_{\Delta t \to 0} \frac{-\Delta[A]}{\Delta t} \tag{17.1.8}$$

However, if you does not know calculus, just bear in mind that the larger the time interval Δt, the smaller will be the precision of the instantaneous rate.

Instantaneous rates are also known as differential rates.

Rate Laws and Reaction Order

The relation between the rate of a reaction and the concentrations of reactants is expressed by its *rate law*. For example, the rate of the gas-phase decomposition of dinitrogen pentoxide

$$2N_2O_5 \rightarrow 4NO_2 + O_2 \tag{17.1.9}$$

has been found to be directly proportional to the concentration of N_2O_5:

$$\text{rate} = k[N_2O_5] \tag{17.1.10}$$

Be very careful about confusing equilibrium constant expressions with those for rate laws. The expression for K_{eq} can always be written by inspecting the reaction equation, and it contains a term for each component (raised to the appropriate power) whose concentration changes during the reaction. For this reaction it is given by

$$K_{eq} = \frac{[NO_2]^2[O_2]}{[N_2O_5]^2} \tag{17.1.11}$$

In contrast, the expression for the rate law generally bears no necessary relation to the reaction equation, and must be determined **experimentally**. More generally, for a reaction of the form

$$n_A A + n_B B + ... \rightarrow products \tag{17.1.12}$$

the rate law will be

$$rate = [A]^a[B]^b... \tag{17.1.13}$$

, n_B.

Since the rate of a reaction has the dimensions of (concentration/time), the dimensions of the rate constant *k* will depend on the exponents of the concentration terms in the rate law. To make this work out properly, if we let p be the sum of the exponents of the concentration terms in the rate law

$$p = a + b + \ldots \tag{17.1.14}$$

then *k* will have the dimensions (concentration^{1-p}/time).

Reaction order

The *order* of a rate law is the sum of the exponents in its concentration terms. For the N_2O_5 decomposition with the rate law $k[N_2O_5]$, this exponent is 1 (and thus is not explicitly shown); this reaction is therefore a first order reaction. We can also say that the reaction is "first order in N_2O_5". For more complicated rate laws, we can speak of the overall reaction order and also the orders with respect to each component. As an example, consider a reaction

$$A + 3B + 2C \rightarrow \text{products} \tag{17.1.15}$$

whose experimental rate law is

$$rate = k[A][B]^2 \tag{17.1.16}$$

We would describe this reaction as

- third-order overall,
- first-order in A,
- second-order in B, and
- zero-order in C.

Zero-order means that the rate is independent of the concentration of a particular reactant. However, enough C must be present to allow the equilibrium mixture to form.

Example 17.1.2

The rate of oxidation of bromide ions by bromate in an acidic aqueous solution

$$6H^+ + BrO_3^- + 5Br^- \rightarrow 3Br_2 + 3H_2O$$

is found to follow the rate law

$$rate = k[Br^-][BrO_3^-][H^+]^2 \tag{17.1.17}$$

What happens to the rate if, in separate experiments,

a. [BrO_3^-] is doubled;
b. the pH is increased by one unit;
c. the solution is diluted to twice its volume, with the pH kept constant by use of a buffer?

Solution

a. Since the rate is first-order in bromate, doubling its concentration will double the reaction rate.
b. Increasing the pH by one unit will *decrease* the [H^+] by a factor of 10. Since the reaction is second-order in [H^+], this will decrease the rate by a factor of 100.
c. Dilution reduces the concentrations of both Br_2 and BrO_3^- to half their original values. Doing this to each concentration alone would reduce the rate by a factor of 2, so reducing both concentration will reduce the rate by a factor of 4, to (½)×(½) = ¼ of its initial value.

How reaction orders are observed

Observing rate-vs.-concentration proportionality

In order to determine the value of the exponent in a rate equation term, we need to see how the rate varies with the concentration of the substance. For a single-reactant decomposition reaction of the form

A → products

in which the rate is –*d*[A]/*dt*, we simply plot [A] as function of time, draw tangents at various intervals, and see how the slopes of these tangents (the instantaneous rates) depend on [A].

- If doubling the concentration of A doubles the rate, then the reaction is first-order in A.
- If doubling the concentration results in a four-fold rate increase, the reaction is second-order in A.

Example 17.1.3

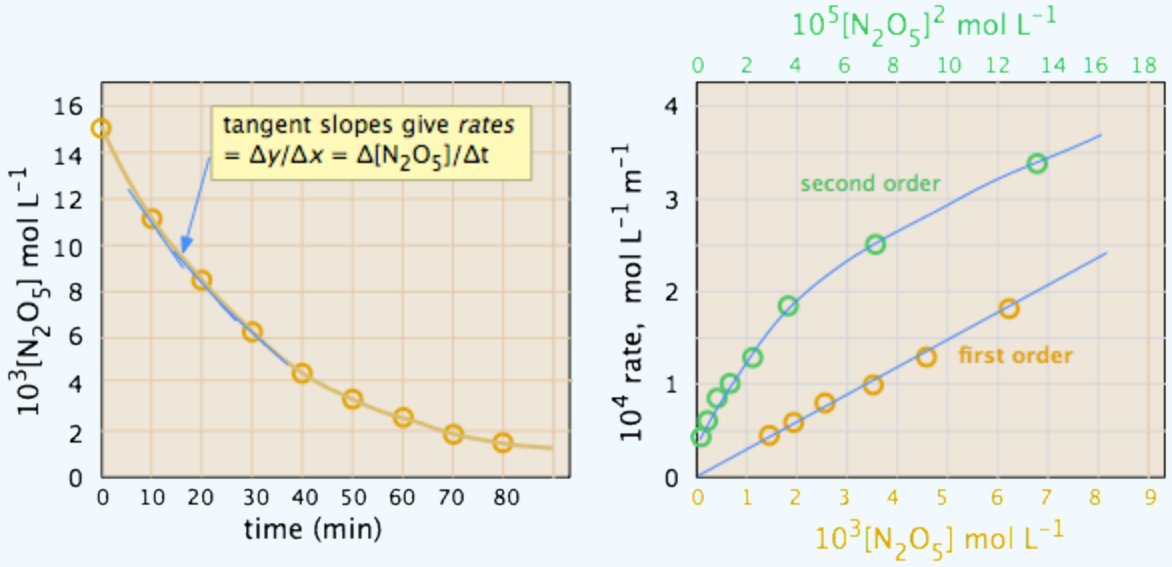
Use the tabulated experimental data to determine the order of the reaction

$$2N_2O_5 \rightarrow 4NO_2 + O_2$$

Time (min)	p(N_2O_5)	[N_2O_5] mol L^{-1}	Rate (mol L^{-1} min^{-1})
0	301.6	0.0152	
10	224.8	0.0113	3.4×10^{-4}
20	166.7	0.0084	2.5
30	123.2	0.0062	1.8
40	92.2	0.0046	1.3
69.1	69.1	0.0035	1.0

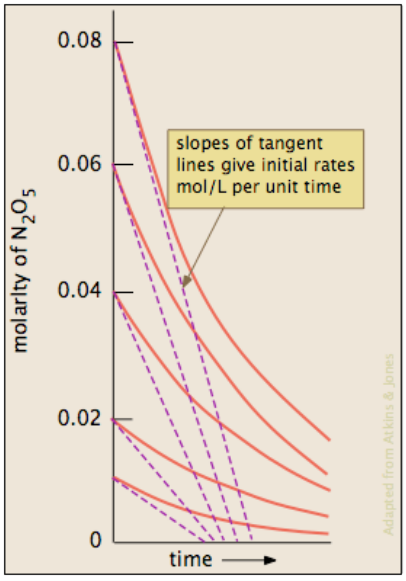
Solution

The ideal gas law can be used to convert the partial pressures of N_2O_5 to molar concentrations. These are then plotted (left) to follow their decrease with time. The rates are computed from the slopes of the tangents (blue lines) and their values plotted as a function of $[N_2O_5]$ and $[N)2O)5]^2$. It is apparent that the rates are directly proportional to $[N)2O)5]^1$, indicating that this is a first-order reaction.

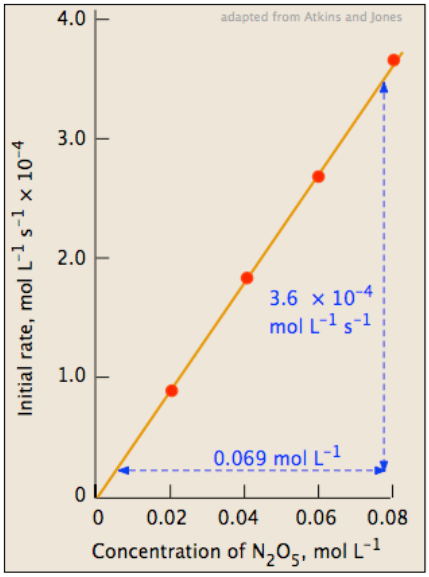


Initial rate method

When there is more than one reactant, the method described above is rarely practical, since the concentrations of the different reactants will generally fall at different rates, depending on the stoichiometry. Instead, we measure only the rate near the beginning of the reaction, before the concentrations have had time to change significantly. The experiment is then repeated with a different starting concentration of the reactant in question, but keeping the concentrations of any others the same. After the order with respect to one component is found, another series of trials is conducted in which the order of another component is found.



Using the same dinitrogen pentoxide decomposition as in the previous example, we conduct a series of runs using five different initial concentrations of N_2O_5. Note that we use only the tangents at the beginning of each plot — that is, at times t=0.



We then plot the five initial rates of consumption of N_2O_5 as a function of its molar concentration. As before, we see that these rates are directly proportional to $[N_2O_5]$. The slope of this plot gives the value of the rate constant.

$$k = \frac{3.6 \times 10^{-4}\ \text{mol L}^{-1}\ \text{s}^{-1}}{0.069\ \text{mol L}^{-1}} = 5.2 \times 10^{-3}\ \text{s}^{-1}$$

rate = (5.2 × 10^{-3}) [N2O5] mol L^{-1} s^{-1}

Example 17.1.4

A study of the gas-phase reduction of nitric oxide by hydrogen

$$2NO + 2H_2 \rightarrow N_2 + 2H_2O$$

yielded the following initial-rate data (all pressures in torr):

experiment	P(NO)	P(H_2)	initial rate (torr s–1)
1	359	300	1.50
2	300	300	1.03
3	152	300	0.25
4	300	289	1.00
5	300	205	0.71
6	200	147	0.51

Find the order of the reaction with respect to each component. In looking over this data, take note of the following:

- The six runs recorded here fall into two groups, in which the initial pressures of H_2 and of NO, respectively, are held constant.

- All the data are expressed in pressures, rather than in concentrations. We can do this because the reactants are gases, whose concentrations are directly proportional to their partial pressures when *T* and *V* are held constant. And since we are only interested in comparing the ratios of pressures and rates, the units cancel out and does notmatter. It is far easier experimentally to adjust and measure pressures than concentrations.

Solution

Experiments 2 and 3: Reduction of the initial partial pressure of NO by a factor of about 2 (300/152) results in a reduction of the initial rate by a factor of about 4, so the reaction is second-order in nitric oxide.

Experiments 4 and 6: Reducing the initial partial pressure of hydrogen by a factor of approximately 2 (289/147) causes a similar reduction in the initial rate, so the reaction is first-order in hydrogen.

The rate law is thus

$$rate = k[NO]^2[H_2] \tag{17.1.18}$$

Dealing with multiple reactants: the isolation method

It is not always practical to determine orders of two or more reactants by the method illustrated in the preceding example. Fortunately, there is another way to accomplish the same task: we can use excess concentrations of all the reactants except the one we wish to investigate. For example, suppose the reaction is

$$A + B + C \rightarrow \text{products} \tag{17.1.19}$$

and we need to find the order with respect to [B] in the rate law. If we set $[B]_o$ to 0.020 M and let $[A]_o = [C]_o$ = 2.00M, then if the reaction goes to completion, the change in [A] and [C] will also be 0.020 M which is only 1 percent of their original values. This will often be smaller than the experimental error in determining the rates, so it can be neglected. By "flooding" the reaction mixture with one or more reactants, we are effectively *isolating* the one in which we are interested.

This page titled 17.1: Rates of reactions and rate laws is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17.2: Reaction Rates Typically Change with Time

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially imortant that you know the precise meanings of all the green-highlighted terms in the context of this topic.

- Describe the contrasting roles of *thermodynamics* and *kinetics* in understanding chemical change.
- Given a balanced net equation, write an expression for the *rate of a reaction*.
- Sketch a curve showing how the *instantaneous rate* of a reaction might change with time.
- Determine the *order* of a reaction of the form A → B + C from experimental data for the concentrations of its products at successive times.
- Describe the *initial rate* and *isolation* methods of determining the orders of the individual reactants in a reaction involving multiple reactants.
- Explain the difference between *differential* and *integral rate laws*.
- Sketch out a plot showing how the concentration of a component ([A] or ln [A]) that follows *first-order kinetics* will change with time. Indicate how the magnitude of the rate constant affects this plot
- Define the *half-life* of a reaction.
- Given the half-life for a first-order reaction A → products along with the initial value of $[A]_o$, find $[A]_t$ at a subsequent time an integral number of half-lifes later.
- Describe the conditions under which a reaction can appear to have an order of zero.

On this page we extend the concept of differential rate laws introduced on the previous page to integral rate laws and reaction *half-life* that are of great importance in most practical applications of kinetics.

Differential and integral rate laws

Measuring instantaneous rates as we have described on the previous page of this unit is the most direct way of determining the rate law of a reaction, but is not always convenient, and it may not even be possible to do so with any precision.

- If the reaction is very fast, its rate may change more rapidly than the time required to measure it; the reaction may be finished before even an initial rate can be observed.
- In the case of very slow reactions, observable changes in concentratons occur so slowly that the observation of a truly "instantaneous" rate becomes impractical.

The ordinary rate law (more precisely known as the *instantaneous* or *differential* rate law) tells us how the rate of a reaction depends on the concentrations of the reactants. But for many practical purposes, it is more important to know how the concentrations of reactants (*and* of *products*) change with time. For example, if you are carrying out a reaction on an industrial scale, you would want to know how long it will take for, say, 95% of the reactants to be converted into products.

This is the purpose of an *integrated rate law*.

Integrating the rate law

This is easy to do, but only some courses expect you to know how to do it. For a quick run-through, click here. If you have had even a bit of calculus, here is an opportunity to put it to use!

Expressing the "speed" of a reaction: the half-life

How long does it take for a chemical reaction to occur under a given set of conditions? As with many "simple" questions, no meaningful answer can be given without being more precise. In this case,

How do we define the point at which the reaction is "completed"?

A reaction is "completed" when it has reached equilibrium — that is, when concentrations of the reactants and products are no longer changing. If the equilibrium constant is quite large, then the answer reduces to a simpler form: the reaction is completed when the concentration of a reactant falls to zero. In the interest of simplicity, we will assume that this is the case in the remainder of this discussion.

"How long?" may be too long

If the reaction takes place very slowly, the time it takes for every last reactant molecule to disappear may be too long for the answer to be practical. In this case, it might make more sense to define "completed" when a reactant concentration has fallen to some arbitrary fraction of its initial value — 90%, 70%, or even only 20%.

The particular fraction one selects depends on the cost of the reactants in relation to the value of the products, balanced against the cost of operating the process for a longer time or the inconvenience of waiting for more product. This kind of consideration is especially imporant in industrial processes in which the balances of these costs affect the profitability of the operation.

The half-life of a reaction

Instead of trying to identify the time required for the reaction to become completed, it is far more practical to specify the the time required for the concentration of a reactant to fall to half of its initial value. This is known as the *half-life* (or half-time) of the reaction.

First-order reactions

The law of exponential change

The rate at which a reactant is consumed in a first-order process is proportional to its concentration at that time. This general relationship, in which a quantity changes at a rate that depends on its instantaneous value, is said to follow an *exponential law*.

Exponential relations are widespread in science and in many other fields. Consumption of a chemical reactant or the decay of a radioactive isotope follow the exponential decay law. Its inverse, the law of exponential growth, describes the manner in which the money in a continuously-compounding bank account grows with time, or the population growth of a colony of reproducing organisms.

The reason that the exponential function $y=e^x$ so efficiently describes such changes stems from the remarkable property that $dy/dx = e^x$; that is, e^x is its own derivative, so the rate of change of y is identical to its value at any point. A nice discussion of this property can be found here .

The first-order integrated rate law

The integrated rate law for a first-order reaction

A → products

is a common example of the law of exponential change. For a reactant A, its concentration $[A]_t$ at time t is given by

$$[A]_t = [A]_o \times e^{-kt}$$

in which $[A]_o$ is its initial concentration and k is the first-order rate constant.

The "e" in the exponential term is of course the base of the natural logarithms, and the negative sign in its exponent means that the value of this term diminishes as t increases, as we would expect for any kind of a decay process.

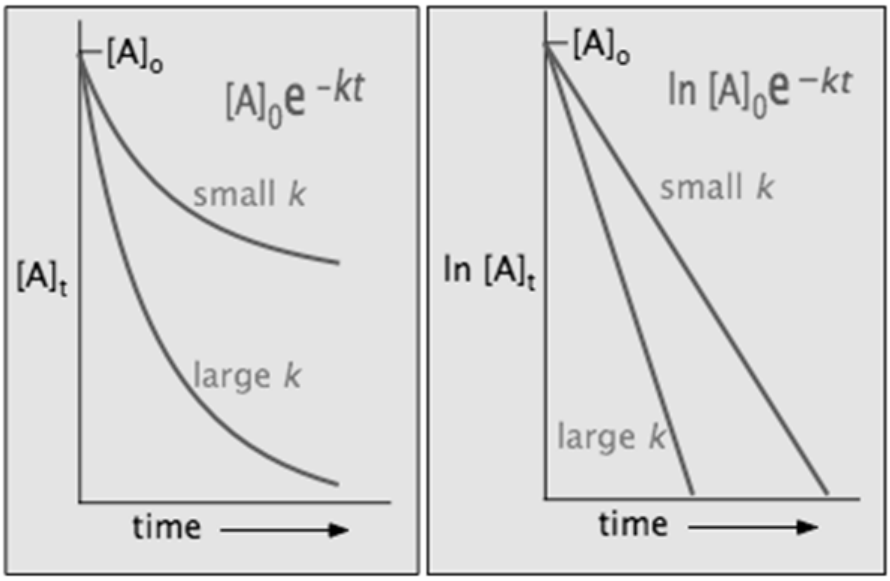
A more convenient form of the integrated rate law is obtained by taking the natural logarithm of both sides:

$\ln [A] = -kt + \ln [A]_o$(4-1)

This has the form of an equation for a straight line

$y = mx + b$

in which the slope m corresponds to the rate constant k. This means that, for a first-order reaction, a plot of ln [A] as a function of time gives a straight line with a slope of $-k$.



Half-life

After a period of one half-life, $t = t_{½}$ and we can write

$$\frac{[A]_{1/2}}{[A]_0} = \frac{1}{2} = e^{-kt_{1/2}}$$

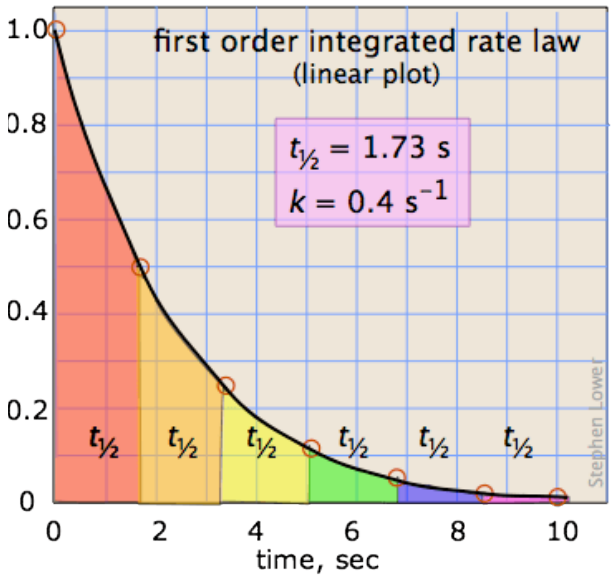
(in which we express the exponential as a function in order to make it stand out more prominently.) Taking logarithms of both sides (remember that
$\ln e^x = x$) yields

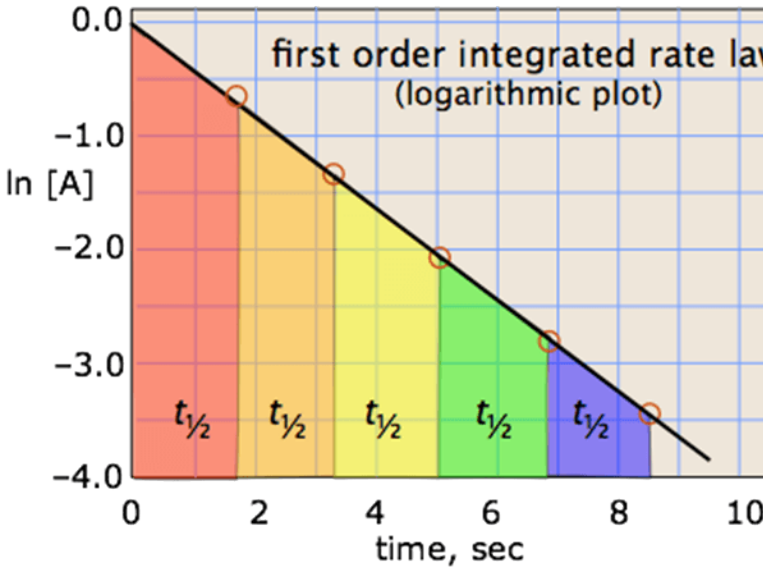
$$\ln 0.5 = -kt$$

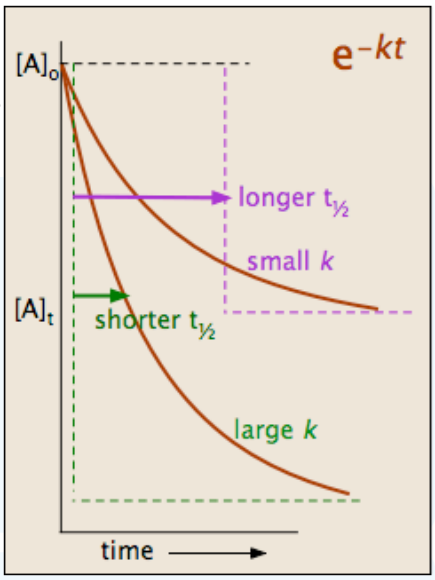
Solving for the half-life, we obtain the simple relation

$$t_{1/2} = \frac{0.693}{k}$$

which tells us that the half-life of a first-order reaction is a constant. This means that 100,000 molecules of a reactant will be reduced to 50,000 in the same time interval needed for ten molcules to be reduced to five.







It should be clear that the rate constant and the half life of a first-order process are inversely related.

Example 17.2.1

The half-life of a first-order reaction was found to be 10 min at a certain temperature. What is its rate constant in reciprocal seconds?

Solution:

From the above equation, k = −0.693/(600 s) = 0.00115 s^{-1}

The decay of radioactive nuclei is always a first-order process.

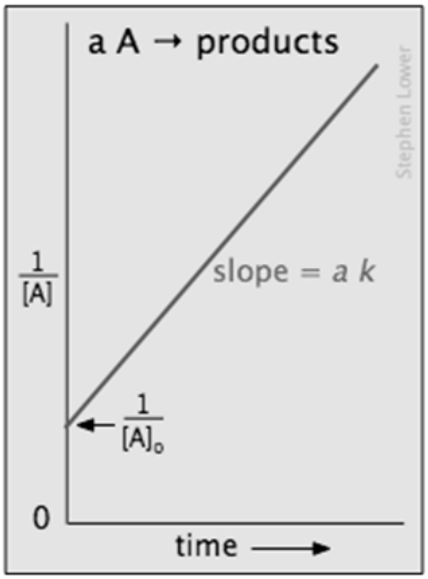
Example 17.2.2

The mass-241 isotope of americium, widely used as an ionizing source in smoke detectors, has a half-life of 432 years.

a) What fraction of the Am^{241} in a smoke detector will have decayed after 50 years?

b) How long will it take for the activity to decline to 80% of its initial value?

c) What would be the "seventh-life" of Am^{241}?



Second-order reactions

Integration of the second-order rate law

$$-\frac{d[\ \mathrm{A}]}{dt} = -k|\ \mathrm{A}|^2$$

yields

$$-\frac{1}{|\ \mathrm{A}]} = \frac{1}{|\ \mathrm{A_o}|} + kt$$

which is easily rearranged into a form of the equation for a straight line (try showing this yourself!) and yields plots similar to the one shown on the left below.

The half-life is given by

$$t_{1/2} = \frac{1}{k|\ \mathrm{A}_0|}$$

(see here for details)

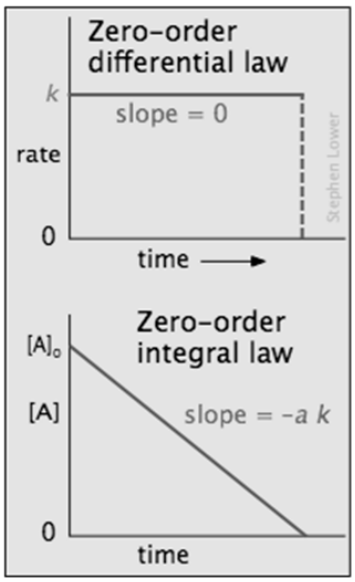
Notice that the half-life of a second-order reaction depends on the initial concentration, in contrast to its constancy for a first-order reaction. For this reason, the concept of half-life for a second-order reaction is far less useful.

Zero-order processes

In some reactions, the rate is apparently independent of the reactant concentraton, in which case

> Integration of the differential rate law yields $[A]_t = [A]_0 - k$

$$rate = k[\ \mathrm{A}]^0 = k$$



Note the word "apparently" in the preceding sentence; zero-order kinetics is always an artifact of the conditions under which the reaction is carried out. For this reason, reactions that follow zero-order kinetics are often referred to as *pseudo-zero-order* reactions.

Clearly, a zero-order process cannot continue after a reactant has been exhausted. Just before this point is reached, the reaction will revert to another rate law instead of falling directly to zero as depicted at the upper left.

There are two general conditions that can give rise to zero-order rates:

This situation commonly occurs when a reaction is catalyzed by attachment to a solid surface (*heterogeneous catalysis*) or to an enzyme.

For example, the decomposition of nitrous oxide

$$N_2O\ (g) \rightarrow N_2(g) + \tfrac{1}{2}\, O_2(g)$$

in the presence of a hot platinum wire (which acts as a catalyst) is zero-order, but it follows more conventional kinetics when carried out entirely in the gas phase.

In this case, the N_2O molecules that react are limited to those that have attached themselves to the surface of the solid catalyst. Once all of the sites on the limited surface of the catalyst have been occupied, additional gas-phase molecules must wait until the decomposition of one of the adsorbed molecules frees up a surface site.

Enzyme-catalyzed reactions in organisms begin with the attachment of the substrate to the active site on the enzyme, leading to the formation of an *enzyme-substrate complex*. If the number of enzyme molecules is limited in relation to substrate molecules, then the reaction may appear to be zero-order.

This is most often seen when two or more reactants are involved. Thus if the reaction

A + B → products

is first-order in both reactants so that

rate = k [A][B]

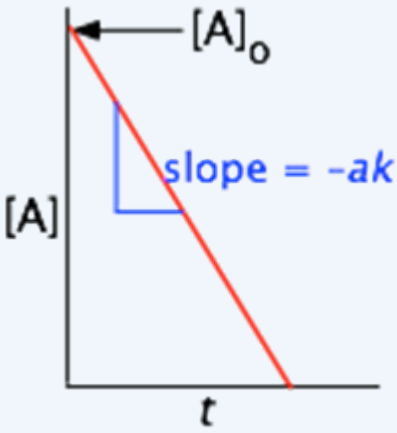
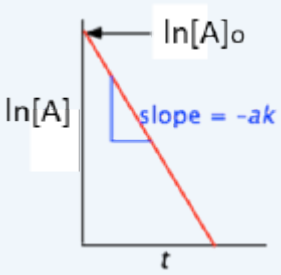
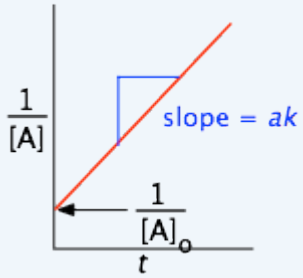
then if B is present in great excess, the reaction will appear to be zero order in B (and first order overall). This commonly happens when B is H_2O and the reaction is carried out in aqueous solution.

1. Only a small fraction of the reactant molecules are in a location or state in which they are able to react, and this fraction is

continually replenished from the larger pool.
2. When two or more reactants are involved, the concentrations of some are much greater than those of others

Summary

The following table compares the rate parameters of zero-, first-, and second-order reactions of the form ***a*A → products**. See also this excellent page from Purdue University that shows plots of [A], ln [A], and 1/[A] for each reaction order.

zero order	first order	second order	
differential rate law	rate = k	rate = k[A]	rate = k[A]^2
integrated rate law	$[A] = [A]_0 = -akt$	$\ln \frac{[\mathrm{A}]_0}{[\mathrm{A}]} = akt$	$\frac{1}{[\mathrm{A}]} - \frac{1}{[\mathrm{A}]_0} = akt$
half-life	$\frac{[\mathrm{A}]_0}{2ak}$	$\frac{\ln 2}{ak} = \frac{.693}{ak}$	$\frac{1}{ak[\mathrm{A}]_0}$
plot that gives straight line	[A] vs. t	ln [A] vs. t	1/[A] vs. t
plot, showing interpretation of slope and intercept			

This page titled 17.2: Reaction Rates Typically Change with Time is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17.3: Collision and activation- the Arrhenius law

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially imortant that you know the precise meanings of all the green-highlighted terms in the context of this topic.

- Explain the meaning of a reaction mechanism and define elementary step and intermediate.
- Describe the role of collisions in reaction mechanisms, and explain why not all collisions lead to the formation of products.
- Sketch out activation energy diagrams for simple reactions that are endothermic or exothermic,
- Explain how an activated complex differs from an intermediate.
- Define catalyst, and sketch out an activation energy diagram that illustrates how catalysts work.
- Explain the significance of the various terms that appear in the Arrhenius Law.
- Sketch out a typical Arrhenius Law plot for a hypothetical reaction at higher and lower temperatures.
- Explain how the activation energy of a reaction can be determined experimentally.
- Explain the significance of the various terms that appear in the pre-exponential factor of the Arrhenius equation.

Why are some reactions so much faster than others, and why are reaction rates independent of the thermodynamic tendency of the reaction to take place? These are the central questions we address in this unit. In doing so, we open the door to the important topic of *reaction mechanisms*: what happens at the microscopic level when chemical reactions take place? We can thank Prof. **Svante Arrhenius** for unlocking this door! To keep things as simple as possible, we will restrict ourselves to reactions that take place in the **gas phase.**The same principles will apply to reactions in liquids and solids, but with added complications that we will discuss in a later unit.

Reaction mechanisms

The *mechanism* of a chemical reaction is the sequence of actual events that take place as reactant molecules are converted into products. Each of these events constitutes an *elementary step* that can be represented as a coming-together of discrete particles ("collison") or as the breaking-up of a molecule ("dissociation") into simpler units. The molecular entity that emerges from each step may be a final product of the reaction, or it might be an *intermediate* — a species that is created in one elementary step and destroyed in a subsequent step, and therefore does not appear in the net reaction equation.

A reaction mechanism must ultimately be understood as a "blow-by-blow" description of the molecular-level events whose sequence leads from reactants to products. These elementary steps (also called *elementary reactions*) are almost always very simple ones involving one, two, or [rarely] three chemical species which are classified, respectively, as

unimolecular	A →	by far the most common
bimolecular	A + B →	
termolecular	A + B + C →	very rare

Collision theory of chemical change: Molecules must collide before they can react

This fundamental rule must guide any analysis of an ordinary chemical reaction mechanism.



This explains why termolecular processes are so uncommon. The kinetic theory of gases tells us that for every 1000 binary collisions, there will be only one event in which three molecules simultaneously come together. Four-way collisions are so improbable that this process has never been demonstrated in an elementary reaction. Consider a simple bimolecular step

$$\mathrm{A} + \mathrm{B} \longrightarrow \text{products}$$

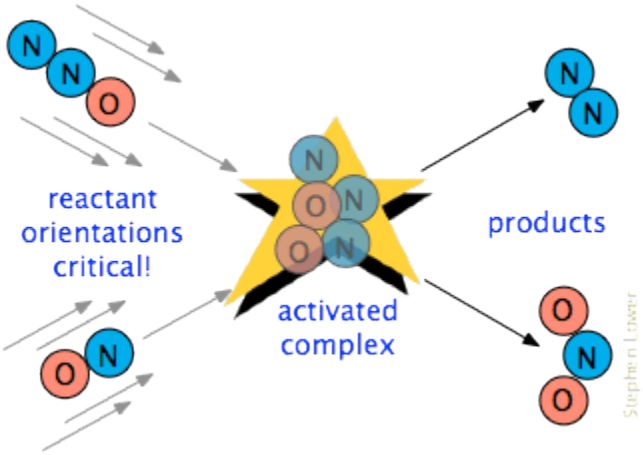
Clearly, if two molecules A and B are to react, they must approach closely enough to disrupt some of their existing bonds and to permit the creation of any new ones that are needed in the products. We call such an encounter a *collision*.

The frequency of collisions between A and B in a gas will be proportional to the concentration of each; if we double [A], the frequency of A-B collisions will double, and doubling [B] will have the same effect. So if all collisions lead to products, than the rate of a bimolecular process will be first-order in A and B, or second-order overall:

$$\text{rate} = k[\mathrm{A}][\mathrm{B}]$$

However, not all collisions are equal. In a gas at room temperature and normal atmospheric pressure, there will be about 10^{33} collisions in each cubic centimetre every second. If every collision between two reactant molecules yielded products, all reactions would be complete in a fraction of a second.

When two billiard balls collide, they simply bounce off of each other. This is also the most likely outcome if the reaction between A and B requires a significant disruption or rearrangement of the bonds between their atoms. In order to effectively initiate a reaction, collisions must be sufficiently *energetic* (kinetic energy) to bring about this bond disruption. More about this further on.



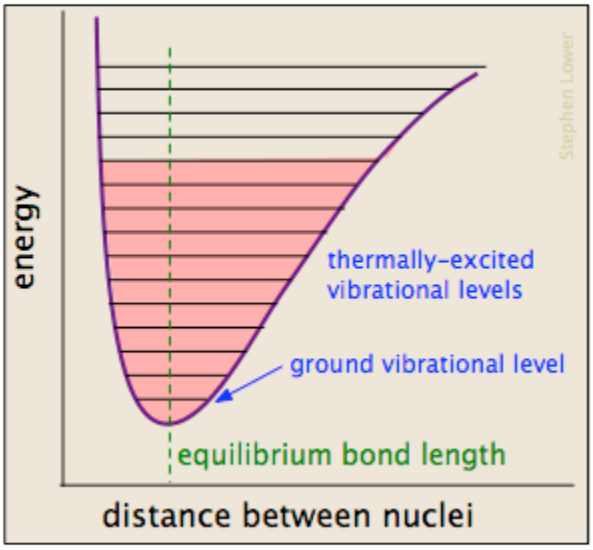
And there is often one additional requirement. In many reactions, especially those involving more complex molecules, the reacting species must be oriented in a manner that is appropriate for the particular process. For example, in the gas-phase reaction of dinitrogen oxide with nitric oxide, the oxygen end of N_2O must hit the nitrogen end of NO; reversing the orientation of either

molecule prevents the reaction.

Owing to the extensive randomization of molecular motions in a gas or liquid, there are always enough correctly-oriented molecules for some of the molecules to react. But of course, the more critical this orientational requirement is, the fewer collisions will be effective.

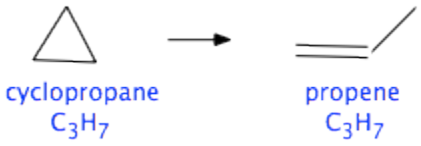
Anatomy of a collision

Energetic collisions between molecules cause interatomic bonds to stretch and bend farther, temporarily weakening them so that they become more susceptible to cleavage. Distortion of the bonds can expose their associated electron clouds to interactions with other reactants that might lead to the formation of new bonds. Chemical bonds have some of the properties of mechanical springs, whose potential energy depends on the extent to which they are stretched or compressed. Each atom-to-atom bond can be described by a potential energy diagram that shows how its energy changes with its length. When the bond absorbs energy (either from heating or through a collision), it is elevated to a higher quantized vibrational state (indicated by the horizontal lines) that weakens the bond as its length oscillates between the extended limits corresponding to the curve.

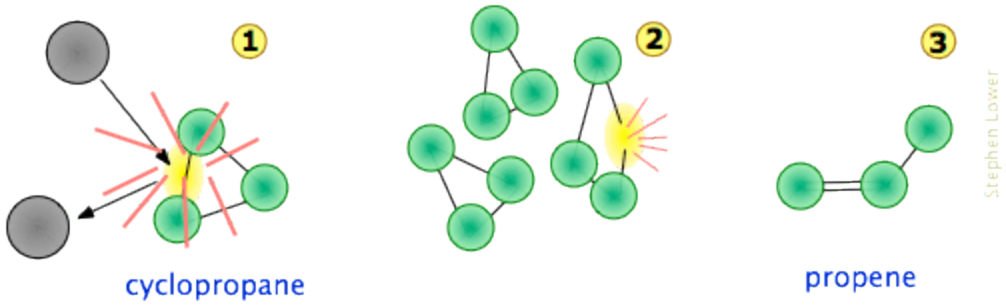


A particular collision will typically excite a number of bonds in this way. Within about 10^{-13} second this excitation gets distributed among the other bonds in the molecule in rather complex and unpredictable ways that can concentrate the added energy at a particularly vulnerable point. The affected bond can stretch and bend farther, making it more susceptible to cleavage. Even if the bond does not break by pure stretching, it can become distorted or twisted so as to expose nearby electron clouds to interactions with other reactants that might encourage a reaction.

Consider, for example, the isomerization of cyclopropane to propene which takes place at fairly high temperatures in the gas phase.



We can imagine the collision-to-product sequence in the following [grossly oversimplified] way:



Note that

- To keep things simple, we do not show the hydrogen atoms here. This is reasonable because C–C bonds are weaker then C–H bonds and thus less likely to be affected.
- The collision at 2 will usually be with another cyclopropane molecule, but because no part of the colliding molecule gets incorporated into the product, it can in principle be a noble gas or some other non-reacting species;
- Although the C–C bonds in cyclopropane are all identicial, the instantaneous localization of the collisional energy can distort the molecule in various ways (2), leading to a configuration sufficiently unstable to initiate the rearrangement to the product.

Unimolecular processes also begin with a collision

The cyclopropane isomerization described above is typical of many decomposition reactions that are found to follow first-order kinetics, implying that the process is unimolecular. Until about 1921, chemists did not understand the role of collisions in unimolecular processes. It turns out that the mechanisms of such reactions are really rather complicated, and that at very low pressures they do follow second-order kinetics. Such reactions are more properly described as *pseudounimolecular*.

Activation energy

Higher temperatures, faster reactions

The chemical reactions associated with most food spoilage are catalyzed by enzymes produced by the bacteria which mediate these processes. It is common knowledge that chemical reactions occur more rapidly at higher temperatures. Everyone knows that milk turns sour much more rapidly if stored at room temperature rather than in a refrigerator, butter goes rancid more quickly in the summer than in the winter, and eggs hard-boil more quickly at sea level than in the mountains. For the same reason, cold-blooded animals such as reptiles and insects tend to be noticeably more lethargic on cold days.

It is not hard to understand why this should be. Thermal energy relates direction to motion at the molecular level. As the temperature rises, molecules move faster and collide more vigorously, greatly increasing the likelyhood of bond cleavages and rearrangements as described above.

Activation energy diagrams

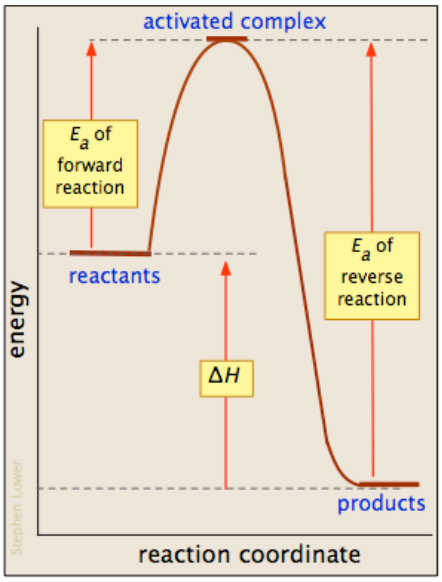
Most reactions involving neutral molecules cannot take place at all until they have acquired the energy needed to stretch, bend, or otherwise distort one or more bonds. This critical energy is known as the *activation energy* of the reaction. *Activation energy diagrams* of the kind shown below plot the total energy input to a reaction system as it proceeds from reactants to products.

In examining such diagrams, take special note of the following:

- The "*reaction coordinate*" plotted along the abscissa represents the changes in atomic coordinates as the system progresses from reactants to products. In the very simplest elementary reactions it might correspond to the stretching or twisting of a particular bond, and be shown to a scale. In general, however, the reaction coordinate is a rather abstract concept that

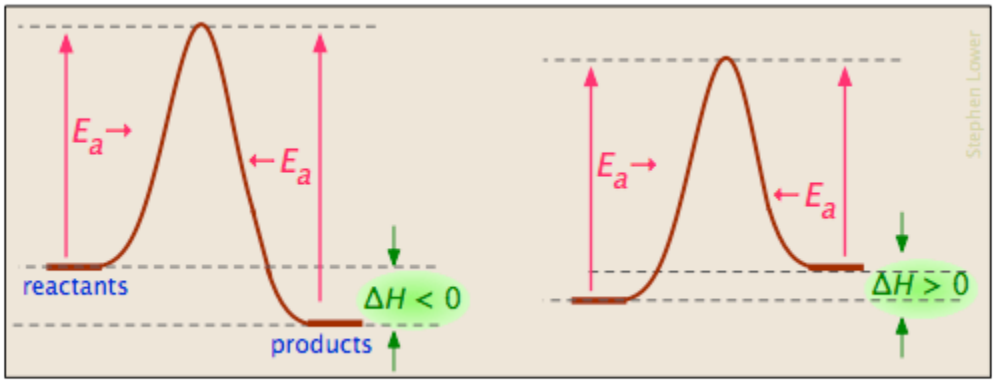
cannot be tied to any single measurable and scaleable quantity.

- The *activated complex* (also known as the *transition state*) represents the structure of the system as it exists at the peak of the activation energy curve. It does *not* correpond to an identifiable intermediate structure (which would more properly be considered the product of a separate elementary process), but rather to whatever configuration of atoms exists during the collision, which lasts for only about 0.1 picosecond.
- Activation energy diagrams always incorporate the energetics (ΔU or ΔH) of the net reaction, but it is important to understand that the latter quantities depend solely on the thermodynamics of the process which are always independent of the reaction pathway. This means that the same reaction can exhibit different activation energies if it can follow alternative pathways.
- With a few exceptions for very simple processes, activation energy diagrams are largely conceptual constructs based on our standard *collision model* for chemical reactions. It would be unwise to read too much into them.

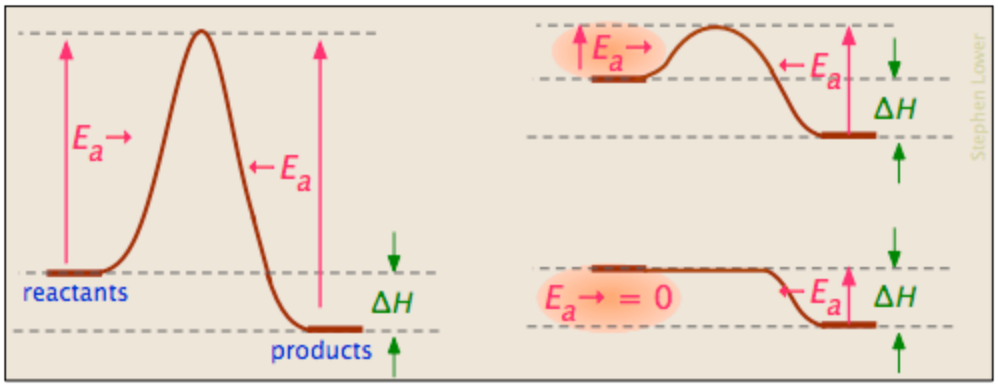


Gallery of activation energy plots

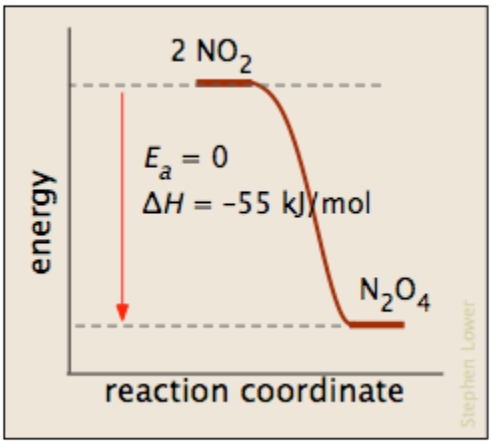
Activation energy diagrams can describe both exothermic and endothermic reactions:



... and the activation energies of the forward reaction can be large, small, or zero (independently, of course, of the value of ΔH):



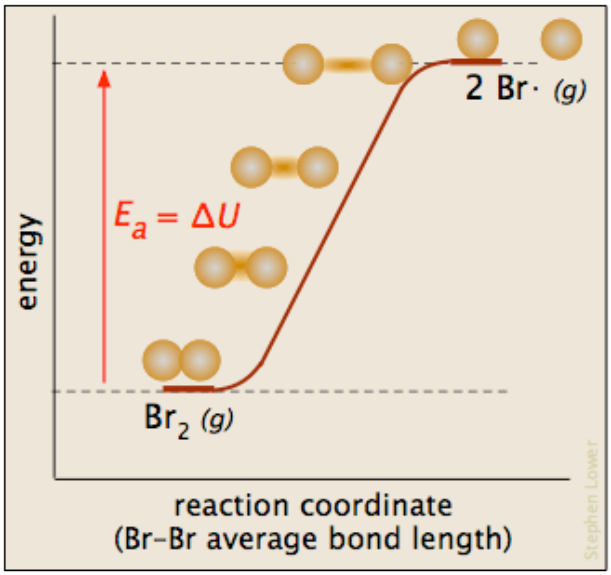
Processes with zero activation energy most commonly involve the combination of oppositely-charged ions or the pairing up of electrons in free radicals, as in the dimerization of nitric oxide (which is an odd-electron molecule).



In this plot for the dissociation of bromine, the E_a is just the enthalpy of atomization

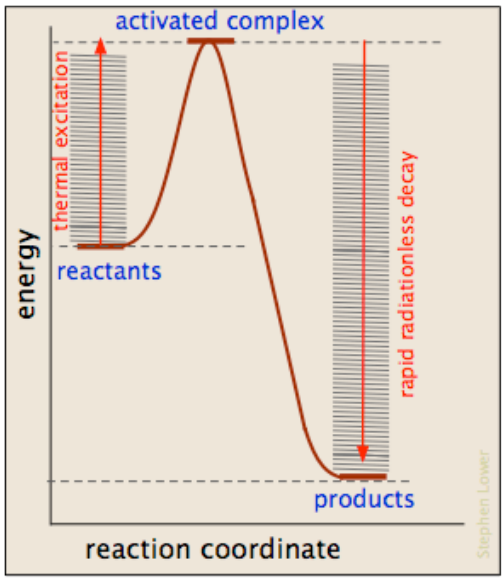
$$Br_2(g) \rightarrow 2\ Br\cdot\ (g)$$

and the reaction coordinate corresponds roughly to the stretching of the vibrationally-excited bond. The "activated complex", if it is considered to exist, is just the last, longest "stretch". The reverse reaction, being the recombination of two radicals, occurs immediately on contact.



Where does the activation energy come from?

In most cases, the activation energy is supplied by thermal energy, either through intermolecular collisions or (in the case of thermal dissocation) by thermal excitation of a bond-stretching vibration to a sufficiently high quantum level.

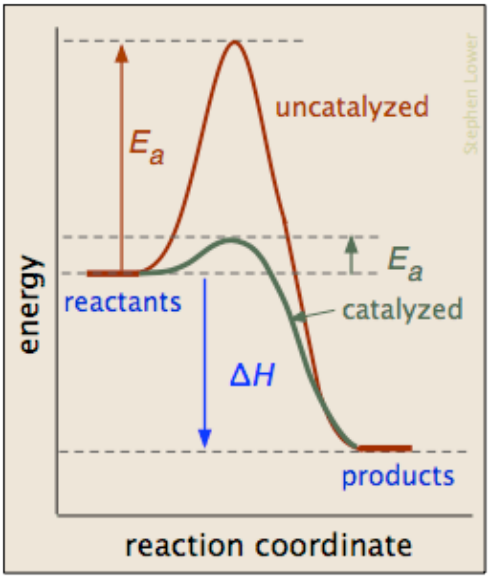


As products are formed, the activation energy is returned in the form of vibrational energy which is quickly degraded to heat. It's worth noting, however, that other sources of activation energy are sometimes applicable:

- Absorption of light by a molecule (photoexcitation) can be a very clean and efficient, but it doesn't always work. It's not enough that the wavelength of the light correspond to the activation energy; it must also fall within the absorption spectrum of the molecule, and (in a complex molecule) enough of it must end up in the right part of the molecule, such as in a particular bond.
- Electrochemical activation. Molecules capable of losing or gaining electrons at the surface of an electrode can undergo activation from an extra potential (known as the *overvoltage*) between the electrode and the solution. The electrode surface often plays an active role, so the process is also known as *electrocatalysis*.

Catalysts can reduce activation energy

A *catalyst* is usually defined as a substance that speeds up a reaction without being consumed by it. More specifically, a catalyst provides an alternative, lower activation energy pathway between reactants and products. As such, they are vitally important to chemical technology; approximately 95% of industrial chemical processes involve catalysts of various kind. In addition, most biochemical processes that occur in living organisms are mediated by *enzymes*, which are catalysts made of proteins.

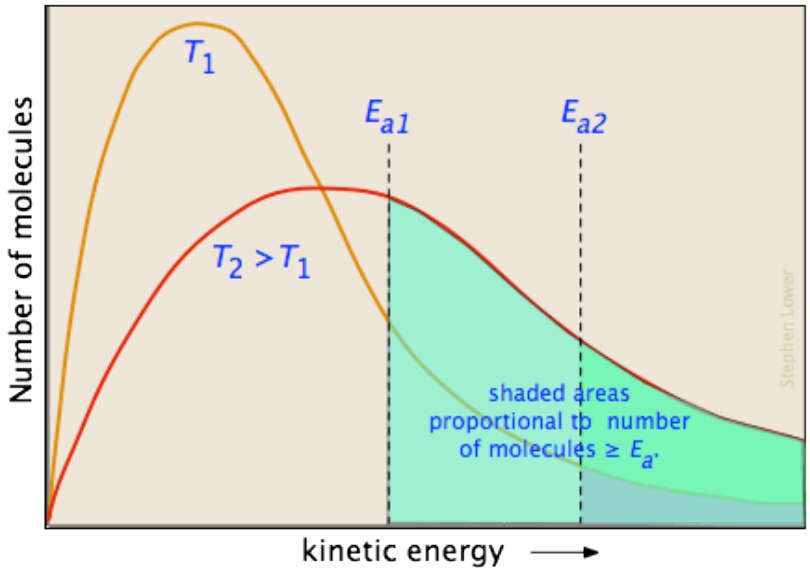


It is important to understand that a catalyst affects only the *kinetics* of a reaction; it does *not* alter the thermodynamic tendency for the reaction to occur. Thus there is a single value of ΔH for the two pathways depicted in the plot on the right

Temperature and kinetic energy

A review of the principles of gas molecular velocities and the Boltzmann distribution can be found on the "KMT-classic" page.

In the vast majority of cases, we depend on thermal actvation, so the major factor we need to consider is what fraction of the molecules possess enough kinetic energy to react at a given temperature. According to kinetic molecular theory, a opulation of molecules at a given temperature is distributed over a variety of kinetic energies that is described by the Maxwell-Boltzman distribution law.



The two distribution plots shown here are for a lower temperature T_1 and a higher temperature T_2. The area under each curve represents the total number of molecules whose energies fall within particular range. The shaded regions indicate the number of molecules which are sufficiently energetic to meet the requirements dictated by the two values of E_a that are shown.

It is clear from these plots that the fraction of molecules whose kinetic energy exceeds the activation energy increases quite rapidly as the temperature is raised. This the reason that virtually all chemical reactions (and all elementary reactions) are more rapid at higher temperatures.

2 The Arrhenius law

By 1890 it was common knowledge that higher temperatures speed up reactions, often doubling the rate for a 10-degree rise, but the reasons for this were not clear. Finally, in 1899, the Swedish chemist Svante Arrhenius (1859-1927) combined the concepts of activation energy and the Boltzmann disribution law into one of the most important relationships in physical chemistry:



$$k = A\, e^{-\frac{E_a}{RT}}$$

activation energy

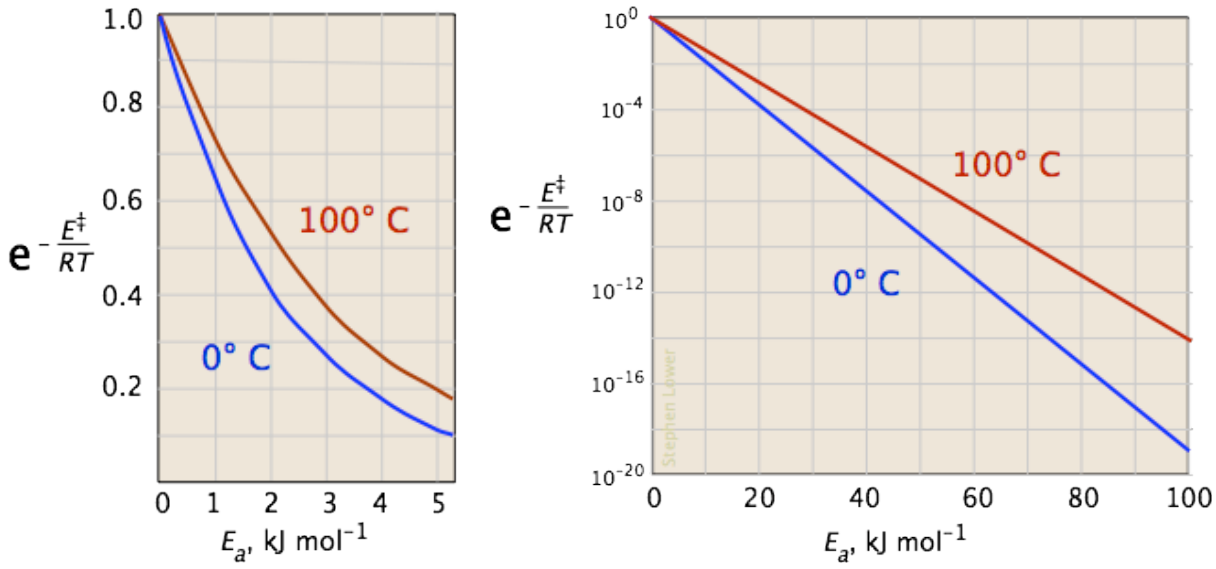
pre-exponential factor

average kinetic energy

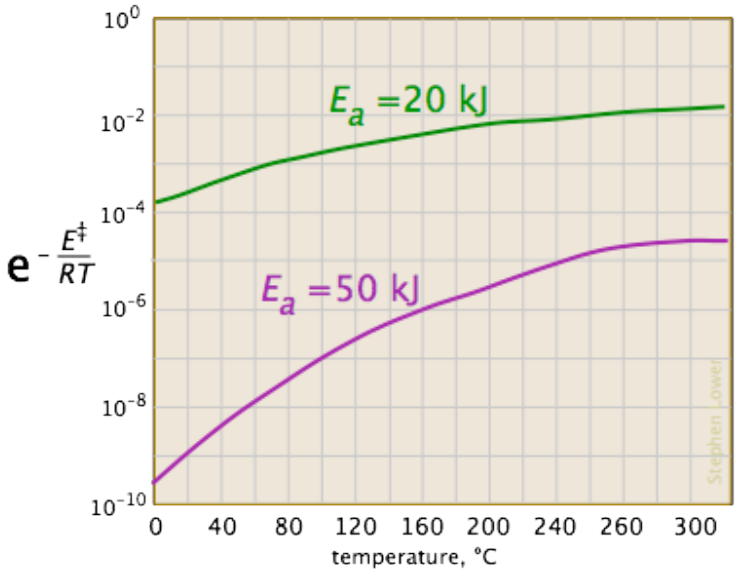
Take a moment to focus on the meaning of this equation, neglecting the A factor for the time being.

First, note that this is another form of the exponential decay law we discussed in the previous section of this series. What is "decaying" here is not the concentration of a reactant as a function of time, but the magnitude of the rate constant as a function of the exponent $-E_a/RT$. And what is the significance of this quantity? If you recall that RT is the *average kinetic energy*, it will be apparent that the exponent is just the ratio of the activation energy E^a to the average kinetic energy. The larger this ratio, the smaller the rate (hence the negative sign.) This means that high temperature and low activation energy favor larger rate constants, and thus speed up the reaction. And because these terms occur in an exponent, their effects on the rate are quite substantial.

The two plots below show the effects of the activation energy (denoted here by $E^{\ddagger}$) on the rate constant. Even a modest activation energy of 50 kJ/mol reduces the rate by a factor of 10^8.



The logarithmic scale in the right-hand plot leads to nice straight lines, as described under the next heading below.



Looking at the role of temperature, we see a similar effect. (If the *x*-axis were in "kilodegrees" the slopes would be more comparable in magnitude with those of the kilojoule plot at the above right.)

Determining the activation energy

The Arrhenius equation

$$k = A\mathrm{e}^{-E_a/RT} \tag{17.3.1}$$

can be written in a non-exponential form which is often more convenient to use and to interpret graphically. Taking the logarithms of both sides and separating the exponential and pre-exponential terms yields

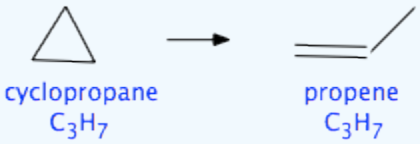
$$\ln k = \ln\left(A\mathrm{e}^{-E_a/RT}\right) = \ln A + \ln\left(\mathrm{e}^{-E_a/RT}\right) \tag{17.3.2}$$

$$\ln k = \ln A - \frac{E_a}{RT} \tag{17.3.3}$$

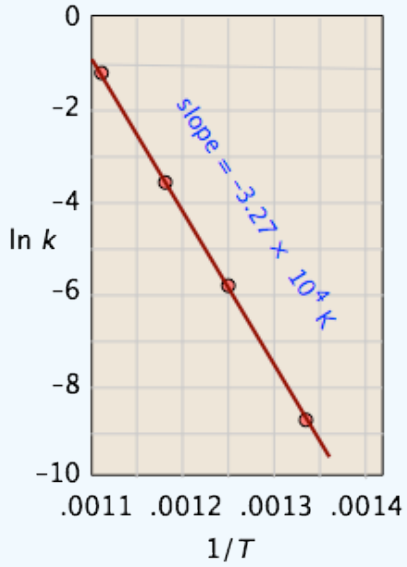
which is the equation of a straight line whose slope is $-E_a/R$. This affords a simple way of determining the activation energy from values of k observed at different temperatures; we just plot $\ln k$ as a function of $1/T$.

Example 17.3.1: Isomerization of cyclopropane

Thus for the isomerization of cyclopropane to propene



the following data were obtained (calculated values shaded in pink):



T, °C	477	523	577	623
1/T, K^{-1} × 10^3	1.33	1.25	1.18	1.11
***k*, s^{-1}**	0.00018	0.0027	0.030	0.26
ln *k*	−8.62	−5.92	−3.51	−1.35

From the calculated slope, we have

$$-(E_a/R) = -3.27 \times 10^4 \text{ K}$$

$$E_a = -(8.314 \text{ J mol}^{-1} \text{ K}^{-1})(-3.27 \times 10^4 \text{ K}) = 273 \text{ kJ mol}^{-1}$$

Comment: This activation energy is rather high, which is not surprising because a carbon-carbon bond must be broken in order to open the cyclopropane ring. (C–C bond energies are typically around 350 kJ/mol.) This is why the reaction must be carried out at high temperature.

You don't always need a plot

(... if you are willing to live a bit dangerously!) Since the ln *k*-vs.-1/*T* plot yields a straight line, it is often convenient to estimate the activation energy from experiments at only two temperatures. To see how this is done, consider that

$$\ln k_2 - \ln k_1 = \left(\ln A - \frac{E_a}{RT_2}\right) - \left(\ln A - \frac{E_a}{RT_1}\right) = \frac{E_a}{R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$$

(... in which we have made the ln-*A* term disappear by subtracting the expressions for the two ln-*k* terms.) Solving the expression on the right for the activation energy yields

$$E_a = \frac{R \ln \frac{k_2}{k_1}}{\frac{1}{T_1} - \frac{1}{T_2}}$$

Example 17.3.2

A widely used rule-of-thumb for the temperature dependence of a reaction rate is that a ten-C° rise in the temperature approximately doubles the rate. (This is obviously not generally true, especially when a strong covalent bond must be broken.) But for a reaction that does show this behavior, what would the activation energy be?

Solution

We will center our ten-degree interval at 300 K. Substituting into the above expression yields

$$E_a = \frac{8.314 \ln \frac{2}{1}}{\frac{1}{295} - \frac{1}{305}}$$ = (8.314)(0.693) / (.00339 - 0.00328)

= (5.76 J mol^{-1} K^{-1}) / (0.00011 K^{-1}) = 52400 J mol^{-1} = **52.4 kJ mol^{-1}**

Example 17.3.3

It takes about 3.0 minutes to cook a hard-boiled egg in Los Angeles, but at the higher altitude of Denver, where water boils at 92°C, the cooking time is 4.5 minutes. Use this information to estimate the activation energy for the coagulation of egg albumin protein.

Solution:

The ratio of the rate constants at the elevations of LA and Denver is 4.5/3.0 = 1.5, and the respective temperatures are 373K and 365K. With the subscripts 2 and 1 referring to LA and Denver respectively, we have

E_a = (8.314)(ln 1.5) / (1/365 – 1/273) = (8.314)(.405) / (0.00274 – 0.00366)

= (3.37 J mol^{-1} K^{-1}) / (0.000923 K^{-1}) = 3650 J mol^{-1} = **3.65 kJ mol^{-1}**

Comment: This rather low value seems reasonable because protein denaturation involves the disruption of relatively weak hydrogen bonds; no covalent bonds are broken

- Many biological processes exhibit a temperature dependence that follows the Arrhenius law, and can thus be characterized by an activation energy. See this interesting Dartmouth U. page that looks at the kinetics of cricket chirps.
- In an article on the *Kinetics of Popping of Popcorn* (Cereal Chemisty 82(1): 53-59), J. Byrd and M. Perona found that popping follows a first-order rate law with an activation energy of 53.8 kJ/mol.

The pre-exponential factor

It is now time to focus in on the pre-exponential term *A* in the Arrhenius equation. We have been neglecting it because it is not directly involved in relating temperature and activation energy, which is the main practical use of the equation. But since *A* multiplies the exponential term, its value clearly contributes to the value of the rate constant and thus of the rate.

$$k = Ae^{-E_a/RT}$$

and of the temperature.

If this fraction were unity, the Arrhenius law would reduce to

$$k = A \tag{17.3.4}$$

In other words, *A* is the fraction of molecules that would react if either the activation energy were zero, or if the kinetic energy of all molecules exceeded E_a — admittedly, an uncommon scenario.

It's all about collisions

So what would limit the rate constant if there were no activation energy requirements? The most obvious factor would be the rate at which reactant molecules come into contact. This can be calculated from kinetic molecular theory and is known as the *frequency-* or *collision factor Z*. In some reactions, the relative orientation of the molecules at the point of collision is important, so we can also define a geometrical or *steric factor* (commonly denoted by ρ (Greek lower case *rho*).

$$k = A e^{-\frac{E_a}{RT}}$$

$A = Z \times \rho$

frequency factor — steric factor

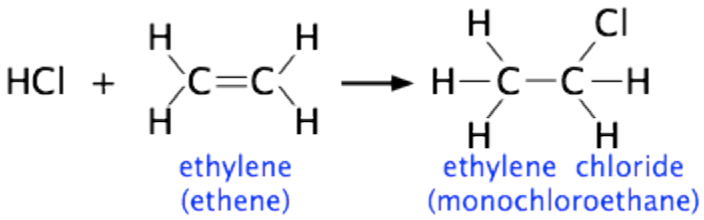
In general, we can express *A* as the product of these two factors:

$$A = Z\rho \tag{17.3.5}$$

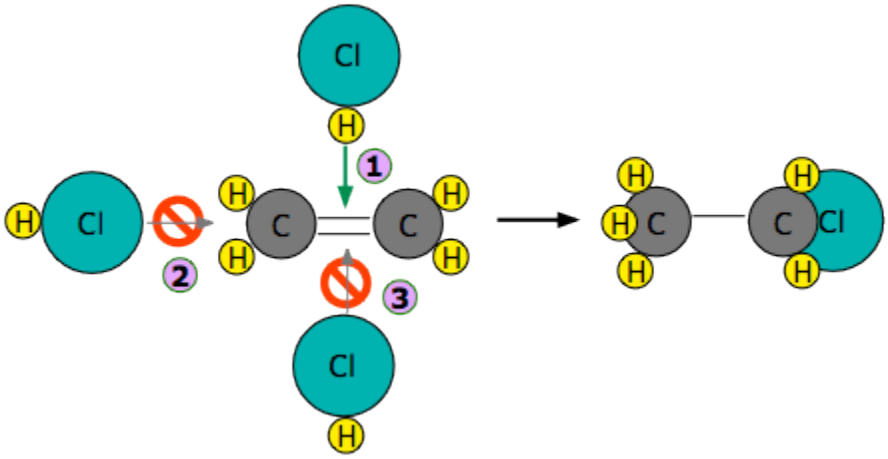
Values of ρ are generally very difficult to assess; they are sometime estimated by comparing the observed rate constant with the one in which *A* is assumed to be the same as *Z*.

Direction makes a difference

The more complicated the structures of the reactants, the more likely that the value of the rate constant will depend on the trajectories at which the reactants approach each other. We showed one example of this near the top of the page, but for another, consider the addition of a hydrogen halide such as HCl to the double bond of an alkene, converting it to a chloroalkane. This kind of *electrophilic addition reaction* is well-known to all students of organic chemistry.

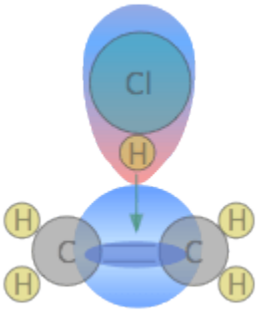


Experiments have shown that the reaction only takes place when the HCl molecule approaches the alkene with its hydrogen-end, and in a direction that is approximately perpendicular to the double bond, as shown at 1 below.



The reason for this becomes apparent when we recall that HCl is highly polar owing to the high electronegativity of chlorine, so that the hydrogen end of the molecule is slightly positive. The double bond of ethene consists of two clouds of negative charge corresponding to the σ (*sigma*) and π (*pi*) molecular orbitals. The latter, which extends above and below the plane of the C_2H_4 molecule, interacts with and attracts the HCl molecule.

If, instead, the HCl approaches with its chlorine end leading as in 3, electrostatic repulsion between the like charges causes the two molecules to bounce away from each other before any reaction can take place. The same thing happens in 2; the electronegativity difference between carbon and hydrogen is too small to make the C–H bond sufficiently polar to attract the incoming chlorine atom.



The lesson you should take from this example is that once you start combining a variety of chemical principles, you gradually develop what might be called "chemical intuition" which you can apply to a wide variety of problems. This is far more important than memorizing specific examples. Now that you know what it takes to get a reaction started, you are ready for the next lesson that describes their actual mechanisms.

This page titled 17.3: Collision and activation- the Arrhenius law is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17.4: Reaction Mechanisms

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially important that you know the precise meanings of all the green-highlighted terms in the context of this topic.

- Explain what is meant by the *mechanism* of a reaction.
- Define an *elementary reaction*, and state how it differs from an ordinary *net chemical reaction*.
- Sketch out an activation energy diagram for a multistep mechanism involving a *rate-determining step*, and relate this to the activation energy of the overall reaction.
- Write the rate law expression for a two-step mechanism in which the rate constants have significantly different magnitudes.
- Write the rate law expression for a three-step reaction in which one step is a *rapid equilibrium*, and the other two steps have significantly different magnitudes.
- Define a *chain reaction*, and list some of the different kinds of steps such a reaction will involve.
- Define a *branching chain reaction*, and explain how such reactions can lead to explosions.

We are now ready to open up the "black box" that lies between the reactants and products of a net chemical reaction. What we find inside may not be very pretty, but it is always interesting because it provides us with a blow-by-blow description of ***how*** chemical reactions take place.

The *mechanism* of a chemical reaction is the sequence of actual events that take place as reactant molecules are converted into products. Each of these events constitutes an *elementary step* that can be represented as a coming-together of discrete particles ("collision") or as the breaking-up of a molecule ("dissociation") into simpler units. The molecular entity that emerges from each step may be a final product of the reaction, or it might be an *intermediate* — a species that is created in one elementary step and destroyed in a subsequent step, and therefore does not appear in the net reaction equation.

For an example of a mechanism, consider the decomposition of nitrogen dioxide into nitric oxide and oxygen. The net balanced equation is

$$2\,\mathrm{NO_2\,(g)} \longrightarrow 2\,\mathrm{NO\,(g)} + \mathrm{O_2\,(g)}$$

The mechanism of this reaction is believed to involve the following two elementary steps:

Multiple \label

Note that the intermediate species $\mathrm{NO_3}$ has only a transient existence and does not appear in the net equation.

Properties of a Mechanism

A useful reaction mechanism

- consists of a series of *elementary steps* (defined below) that can be written as chemical equations, and whose sum gives the net balanced reaction equation;
- must agree with the experimental rate law;
- can rarely if ever be proved absolutely.

It is important to understand that the mechanism of a given net reaction may be different under different conditions. For example, the dissociation of hydrogen bromide

$$2\,\mathrm{HBr\,(g)} \longrightarrow \mathrm{H_2\,(g)} + \mathrm{Br_2\,(g)}$$

proceeds by different mechanisms (and follows different rate laws) when carried out in the dark (*thermal decomposition*) and in

the light (*photochemical decomposition*). Similarly, the presence of a *catalyst* can enable an alternative mechanism that greatly speeds up the rate of a reaction.

Elementary Steps

A reaction mechanism must ultimately be understood as a "blow-by-blow" description of the molecular-level events whose sequence leads from reactants to products. These elementary steps (also called *elementary reactions*) are almost always very simple ones involving one, two, or [rarely] three chemical species which are classified, respectively, as

unimolecular	$A \rightarrow$	by far the most common
bimolecular	$A + B \rightarrow$	common
termolecular	$A + B + C \rightarrow$	very rare

Elementary reactions differ from ordinary net chemical reactions in two important ways:

- The rate law of an elementary reaction can be written by inspection. For example, a bimolecular process always follows the second-order rate law $rate = k[A][B]$.
- Elementary steps often involve unstable or reactive species that do not appear in the net reaction equation.

Some net reactions do proceed in a single elementary step, at least under certain conditions. However, without careful experimentation, one can never be sure. The gas-phase formation of HI from its elements was long thought to be a simple bimolecular combination of $\mathrm{H_2}$ and $\mathrm{I_2}$, but it was later found that under certain conditions, it follows a more complicated rate law.

Multi-step (Consecutive) Reactions

Mechanisms in which one elementary step is followed by another are very common.

$$\mathrm{A + B \longrightarrow \not{Q}} \qquad \text{(step 1)}$$

$$\mathrm{B + \not{Q} \longrightarrow C} \qquad \text{(step 2)}$$

$$\mathrm{A + 2\,B \longrightarrow C} \qquad \text{(net reaction)}$$

(As must always be the case, the net reaction is just the sum of its elementary steps.)

In this example, the species Q is an **intermediate**, usually an unstable or highly reactive species. If both steps proceed at similar rates, rate law experiments on the net reaction would not reveal that two separate steps are involved here. The rate law for the reaction would be

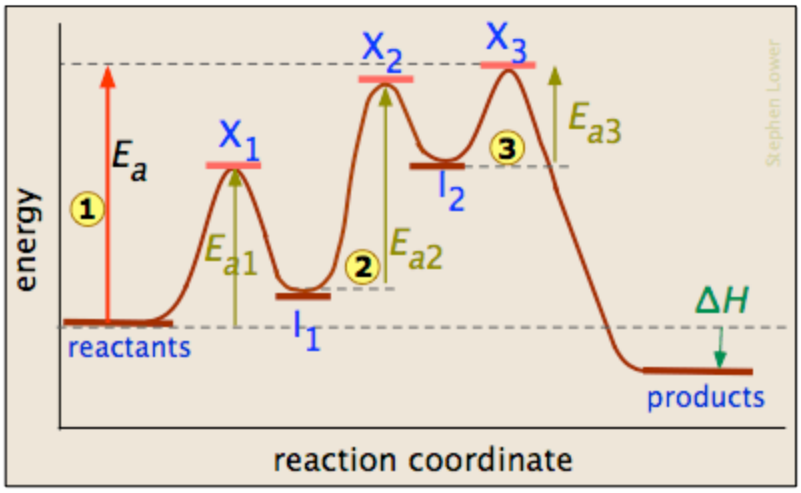
$$rate = k[A][B]^2$$

(Bear in mind that intermediates such as Q **cannot** appear in the rate law of a net reaction.)

When the rates are quite different, things can get interesting and lead to quite varied kinetics as well as some simplifying approximations. When the rate constants of a series of consecutive reactions are quite different, a number of relationships can come into play that greatly simplify our understanding of the observed reaction kinetics.

Approximation 1: The Rate-Determining Step Approximation

The rate-determining step is also known as the **rate-limiting step**. We can generally expect that one of the elementary reactions in a sequence of consecutive steps will have a rate constant that is smaller than the others. The effect is to slow the rates of all the reactions — very much in the way that a line of cars creeps slowly up a hill behind a slow truck.



The three-step reaction depicted here involves two intermediate species **I_1** and **I_2**, and three activated complexes numbered **$X_{1\text{-}3}$**. Although the step (3) $I_2 \rightarrow$ products has the smallest individual activation energy E_{a3}, the energy of X_3 with respect to the *reactants* determines the activation energy of the overall reaction, denoted by the leftmost vertical arrow (1). Thus the rate-determining step is (3)

$$X_1 \rightarrow X_2.$$

Chemists often refer to elementary reactions whose forward rate constants have large magnitudes as "fast", and those with forward small rate constants as "slow". Always bear in mind, however, that as long as the steps proceed in single file (no short-cuts!), **all of them will proceed at the same rate**. So even the "fastest" members of a consecutive series of reactions will proceed as slowly as the "slowest" one.

Approximation 2: The Rapid Equilibrium Approximation

In many multi-step processes, the forward and reverse rate constants for the formation of an intermediate Q are of similar magnitudes and sufficiently large to make the reaction in each direction quite rapid. Decomposition of the intermediate to product is a slower process:

$$\mathrm{A} \underset{\blacksquare}{\overset{k_1}{\rightleftharpoons}} \mathrm{Q} \xrightarrow{k_2} \mathrm{B}$$

This is often described as a rapid equilibrium in which the concentration of Q can be related to the equilibrium constant

$$K = \frac{k_1}{k_{-1}}$$

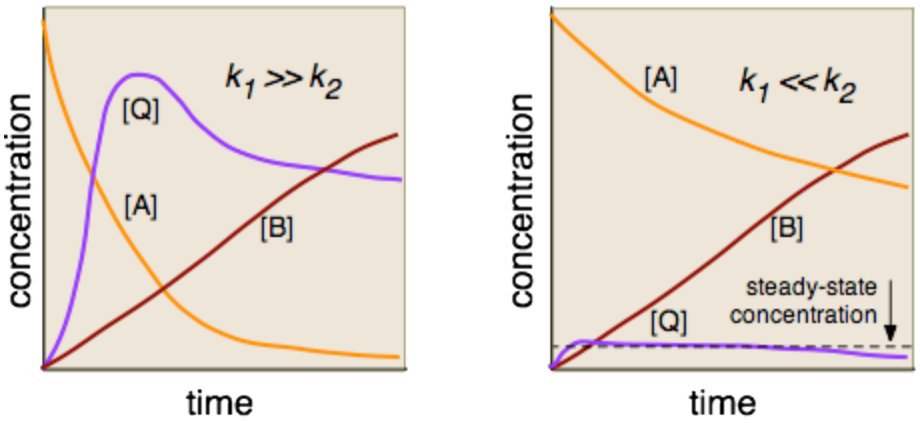
This is just the Law of Mass Action. It should be understood, however, that true equilibrium is never achieved because Q is continually being consumed; that is, the rate of formation of Q always exceeds its rate of decomposition. For this reason, the steady-state approximation described below is generally preferred to treat processes of this kind.

Approximation 3: The Steady-State Approximation

Consider a mechanism consisting of two sequential reactions

$$\mathrm{A} \xrightarrow{k_1} \mathrm{Q} \xrightarrow{k_2} \mathrm{B}$$

in which Q is an intermediate. The time-vs-concentration profiles of these three substances will depend on the relative magnitudes of k_1 and k_2, as shown in the following diagrams. Construction of these diagrams requires the solution of sets of simultaneous differential equations, which is [fortunately!] beyond the scope of this course.



The steady-state approximation is usually not covered in introductory courses, although it is not particularly complicated mathematically.

In the left-hand diagram, the rate-determining step is clearly the conversion of the rapidly-formed intermediate into the product, so there is no need to formulate a rate law that involves Q. But on the right side, the formation of Q is rate-determining, but its conversion into B is so rapid that $[Q]$ never builds up to a substantial value. (Notice how the plots for $[A]$ and $[B]$ are almost mutually inverse.) The effect is to maintain the concentration of Q at an approximately constant value. This *steady-state approximation* can greatly simplify the analysis of many reaction mechanisms, especially those that are mediated by enzymes in organisms.

Converting Mechanisms into Rate Laws

We are unable to look directly at the elementary steps hidden within the "black box" of the reaction mechanism, we are limited to proposing a sequence that would be consistent with the reaction order which we can observe. Chemical intuition can guide us in this, for example by guessing the magnitudes of some of the activation energies. In the end, however, the best we can do is to work out a mechanism that is *plausible*; we can never "prove" that what we come up with is the actual mechanism.

Example $17.4.1A$: A + B → C

Consider the following reaction:

$$A + B \rightarrow C$$

One possible mechanism might involve two intermediates Q and R:

Step 1	$\mathrm{A + B \xrightarrow{k_1} Q}$	(slow, rate-determining)
Step 2	$\mathrm{Q + A \xrightarrow{k_2} R}$	(fast)
Step 3	$\mathrm{R + B \xrightarrow{k_3} C}$	(fast)

The rate law corresponding to this mechanism would be that of the rate-determining step:

$$\text{rate} = k_1[A][B].$$

If the first step in a mechanism is rate-determining, it is easy to find the rate law for the overall expression from the mechanism. If the second or a later step is rate-determining, determining the rate law is slightly more complicated and often requires either of the two approximation above to identify.

Example 17.4.1B: A + B → C

An alternative mechanism for the following reaction:

$$A + B \rightarrow C$$

in which the rate-determining step involves one of the intermediates would display third-order kinetics:

Step 1	$\mathrm{A + B \overset{k_1}{\rightleftharpoons} [redacted]}$	(rapid equilibrium)
Step 2	$\mathrm{Q + A \xrightarrow{k_2} R}$	(slow, rate-determining)
Step 3	$\mathrm{R + B \xrightarrow{k_3} C}$	(fast)

Since intermediates cannot appear in rate law expressions, we must express $[Q]$ in the rate-determining step in terms of the other reactants. To do this, we make use of the fact that Step **1** involves an equilibrium constant K_1:

$$K_1 = \frac{k_1}{k_{-1}} = \frac{[Q]}{[A][B]}$$

Solving this for $[Q]$, we obtain

$$[Q] = K_1[A][B].$$

We can now express the rate law for **2** as

$$\begin{aligned} \text{rate} &= k_2 K_1 [A][B][A] \\ &= k[A]^2[B] \end{aligned}$$

in which the constants k_2 and K_1 have been combined into a single constant k.

Example 17.4.2: F_2 + 2 NO_2 → 2 NO_2F

Consider the following reaction:

$$\mathrm{F_2 + 2\,NO_2 \longrightarrow 2\,NO_2F}$$

Application of the "chemical intuition" mentioned in the above box would lead us to suspect that any process that involves breaking of the strong F–F bond would likely be slow enough to be rate limiting, and that the resulting atomic fluorines would be very fast to react with another odd-electron species:

Step 1	$\mathrm{F_2 + 2\,NO_2 \xrightarrow{k_1} NO_2F + F + NO_2}$	(slow, rate-determining)
Step 2	$\mathrm{F + NO_2 \xrightarrow{k_2} NO_2F}$	(very fast)

If this mechanism is correct, then the rate law of the net reaction would be that of the rate-determining step:

$$\text{rate} = k_1[F_2][NO_2]$$

Example 17.4.3: Decomposition of ozone

Ozone is an unstable allotrope of oxygen that decomposes back into ordinary dioxygen according to the net reaction

$$2\,O_3 \longrightarrow 3\,O_2$$

A possible mechanism would be the simple one-step bimolecular collision suggested by the reaction equation, but this would lead to a second-order rate law which is not observed. Instead, experiment reveals a more complicated rate law:

$$\text{rate} = [O_3]^2[O_2]^{-1}$$

What's this? It looks as if O_2 actually *inhibits* the reaction in some way. The generally-accepted mechanism for this reaction is:

Step 1	$O_3 \underset{[redacted]}{\overset{k_1}{\rightleftharpoons}} O_2 + O$	(rapid equilibrium)
Step 2	$O + O_3 \xrightarrow{k_2} 2\,O_2$	(rate-determining)

Does this seem reasonable? Note that

- The equilibrium in Step **1** should be quite rapid because ozone's instability predicts a low activation energy for the forward process. The same can be said for the reverse process which involves the highly reactive oxygen atom which we would expect to rapidly gobble up one of the nonbonding electron pairs on the O_2 molecule.
- The inhibitory effect of O_2 can be explained by the Le Chatelier effect — being a product of the equilbrium in **1**, buildup of oxygen forces it back to the left.
- In Step **2**, one might expect the O atom to react with ozone as rapidly as it does with O_2 in the reverse of Step 1, but other studies show that this is not the case. The rate constant k_2 is probably fairly large. Given the strong covalent bond in O_2, we not expect any significant reverse reaction.

To translate this mechanism into a rate law, we first write the equilibrium constant for Step **1** and solve it for the concentration of the intermediate:

$$K = \frac{[O_2][O]}{[O_3]}$$

$$[O] = \frac{k_1[O_3]}{k_{-1}[O_2]}$$

We substitute this value of $[\mathrm{O}]$ into the rate expression $[\mathrm{O}][\mathrm{O_3}]$ for Step 2, which yields the experimentally-obtained rate law

$$rate = k_1 K \frac{[O_3]^2}{[O_2]}$$

Example 17.4.4: A reaction having an apparently Negative E_a

Consider the gas-phase oxidation of nitric oxide:

$$2\,NO + O_2 \longrightarrow 2\,NO_2$$

This reaction, like most third-order reactions, is not termolecular but rather a combination of an equilibrium followed by a subsequent bimolecular step:

Step 1	$NO + NO \underset{[redacted]}{\overset{k_1}{\rightleftharpoons}} N_2O_2$	(equilibrium, eq. const. *K*)
Step 2	$N_2O_2 + O_2 \xrightarrow{k_2} 2\,NO_2$	(rate-limiting)

Since the intermediate N_2O_2 may not appear in the rate equation, we need to express its concentration in terms of the reactant NO. As in the previous example, we do this through the equilibrium constant of Step **1**:

$$K = \frac{[N_2O_2]}{[NO]^2}$$

$$[N_2O_2] = K[NO]^2$$

$$\text{rate} = k_2[N_2O_2][O_2]$$
$$= k_2K[NO]^2[O_2]$$

The unusual feature of this net reaction is that its rate diminishes as the temperature increases, suggesting that the activation energy is negative. Reaction **1** involves bond formation and is exothermic, so as the temperature rises, K decreases (Le Chatelier effect). At the same time, k_2 increases, but not sufficiently to overcome the decrease in K. So the apparently negative activation energy of the overall process is simply an artifact of the magnitudes of the opposing temperature coefficients of k_2 and K.

Chain Reactions (Positive Feedback Mechanisms)

Many important reaction mechanisms, particularly in the gas phase, involve intermediates having unpaired electrons, commonly known as *free radicals*. Free radicals are often fairly stable thermodynamically and may be quite long-lived by themselves, but they are highly reactive, and hence kinetically labile. The dot ·, representing the unpaired electron, is not really a part of the formula, and is usually shown only when we want to emphasize the radical character of a species. The "atomic" forms of many elements that normally form diatomic molecules are free radicals; H·, O·, and Br· are common examples. The simplest and most stable (ΔG = +87 kJ/mol) molecular free radical or "*odd-electron molecule*" is nitric oxide, NO·.

The most important chemical property of a free radical is its ability to pass the odd electron along to another species with which it reacts. This *chain-propagation* process creates a new radical which becomes capable of initiating another reaction. Radicals can, of course, also react with each other, destroying both ("*chain termination*") while creating a new covalent-bonded species.

Max Bodenstein

Much of the pioneering work in this field, of which the HBr synthesis is a classic example, was done by the German chemist Max Bodenstein (1871-1942)



The synthesis of hydrogen bromide from its elements illustrates the major features of a chain reaction. The figures in the right-hand column are the activation energies per mole.

Step 1	$Br_2 + M \rightarrow 2\ Br\cdot$	chain initiation ($\Delta H°$ = +188 kJ)
Step 2	$Br\cdot + H_2 \rightarrow H\cdot + HBr$	chain propagaton (+75 kJ)
Step 3	$H\cdot + Br_2 \rightarrow HBr + Br\cdot$	chain propagation (–176 kJ)
Step 4	$H\cdot + HBr \rightarrow H_2 + Br\cdot$	chain inhibition (–75 kJ)
Step 5	$2\ Br\cdot + M \rightarrow M^* + Br_2$	chain termination (–188 kJ)

Note the following points:

- To start the reaction, a free radical must be formed (**1**). If the temperature of the gas is fairly high, then Br· can be formed from the more energetic collisions of Br_2 molecules with some other molecule M (most likely a second Br_2). This is known as *thermal activation*; Another way of creating free radicals is *photochemical activation*.
- Reactions **2** and **3** consume a free radical but form another, thus "propagating" the chain.
- In reaction **4**, a molecule of the product is destroyed, thus partially un-doing the net process.
- If the only the first four reactions were active, then the cycle would continue indefinitely. But reaction **5** consumes the Br· *chain carrier* without producing new radicals, thus *terminating* the chain. The function of the molecule M is to absorb some of the kinetic energy of the collision so that the two bromine atoms do not simply bounce apart. The product M* represents a thermally-excited M which quickly dissipates its energy to other molecules.
- Other reactions, such as the recombination of hydrogen atoms, also take place, but their contribution to the overall kinetics is usually very small.

The rate laws for chain reactions tend to be very complex, and often have non-integral orders.

Branching Chain Reactions: Explosions

The gas-phase oxidation of hydrogen has been extensively studied over a wide range of temperatures and pressures.

$$\mathrm{H_2\,(g) + \tfrac{1}{2}\,O_2\,(g) \longrightarrow H_2O\,(g)} \quad \Delta H^o = -242\,kJ/mol$$

This reaction does not take place at all when the two gases are simply mixed at room temperature. At temperatures around 500-600°C it proceeds quite smoothly, but when heated above 700° or ignited with a spark, the mixture explodes. As with all combustion reactions, the mechanism of this reaction is extremely complex (you do not want to see the rate law!) and varies somewhat with the conditions. Some of the major radical formation steps are

Step 1	$H_2 + O_2 \rightarrow HO_2\cdot + H\cdot$	chain initiation
Step 2	$H_2 + HO_2\cdot \rightarrow HO\cdot + H_2O$	chain propagaton
Step 3	$H\cdot + O_2 \rightarrow HO\cdot + O\cdot$	chain propagaton + branching

Step 4	$O\cdot + H_2 \rightarrow HO\cdot + H\cdot$	chain propagation + branching
Step 5	$HO\cdot + H_2 \rightarrow H_2O + H\cdot$	chain propagation
The species $HO_2\cdot$ is known as *perhydroxyl*; $HO\cdot$ is *hydroxyl* (not to be confused with the *hydroxyl ion* $HO:^-$)		

Reactions **3** and **4** give birth to more radicals than they consume, so when these are active, each one effectively initiates a new chain process causing the rate overall rate to increase exponentially, producing an explosion.

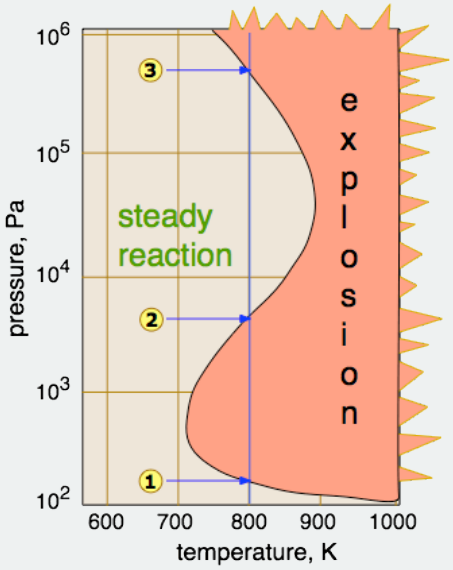
Explosion Limits

An explosive reaction is a highly exothermic process that once initiated, goes to completion very rapidly and cannot be stopped. The destructive force of an explosion arises from the rapid expansion of the gaseous products as they absorb the heat of the reaction. There are two basic kinds of chemical explosions:

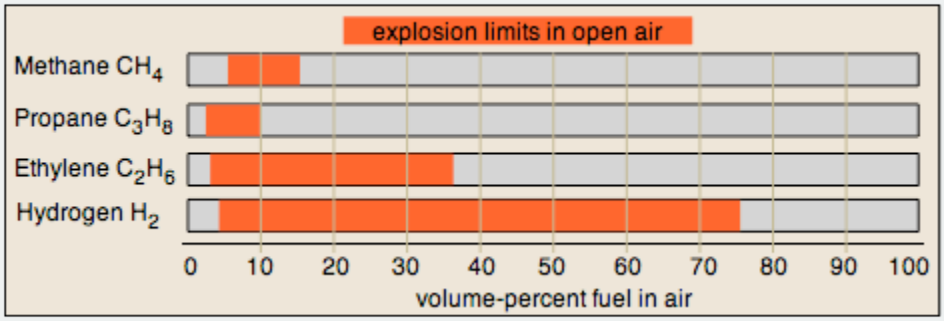
- Thermal explosions occur when heat is released by a reaction more rapidly than it can escape from the reaction space. This causes the reaction to proceed more rapidly, releasing even more heat, resulting in an ever-accelerating runaway rate.
- Chain-branching explosions occur when the number of chain carriers increases exponentially, effectively seeding new reaction centers in the mixture. The process then transforms into a thermal explosion.

Whether or not a reaction proceeds explosively depends on the balance between formation and destruction of the chain-carrying species. This balance depends on the temperature and pressure, as illustrated here for the hydrogen-oxygen reaction.

- Direct recombination of chain carriers generally requires a three-body collision with another molecule to absorb some of the kinetic energy; such ternary processes are unlikely at very low pressures. Thus below the lower explosion limit (1), the radicals (including those produced by a spark) are usually able to reach the walls of the container and combine there — or in the case of an open-air explosion, simply combine with other molecules as they exit the active volume of the reaction.
- Within the explosion zone between (1) and (1) (which, for most gases, are known as the *lower-* and *upper explosion limits*), the propagation and branching processes operate efficiently and explosively, even when the mixture is heated homogeneously.
- Above (2), the concentration of gas molecules is sufficient to enable ternary collisions which allow chain-termination processes to operate efficiently, thus suppressing branching. Above this upper limit, reactions involving most gases proceed smoothly.
- Hydrogen is unusual in that it exhibits a third explosion limit (3). The cause of this was something of a mystery for some time, but it is now believed that explosions in this region do not involve branching, but are thermally induced by the reaction
  $HO_2\cdot + H_2 \rightarrow H_2O_2 + H\cdot$.



The lower explosion limit of gas mixtures varies with the size, shape, and composition of the enclosing container. Needless to say, experimental determination of explosion limits requires some care and creativity. Upper and lower explosion limits for several common fuel gases are shown below.



This page titled 17.4: Reaction Mechanisms is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17.5: Kinetics of Reactions in Solution

Learning Objectives

Make sure you thoroughly understand the following essential ideas:

- Describe some of the major differences between the kinetics of reactions in the gas phase, compared with those in liquid solutions.
- What role do solvent cages play in solution kinetics?
- Explain the distinction between diffusion-control and activation-control of reaction rates in solutions.
- How can the polarity of a solvent affect the energetics of a reaction mechanism?

The kinetics fundamentals we covered in the earlier sections of this lesson group relate to processes that take place in the gas phase. But chemists and biochemists are generally much more concerned with solutions. This lesson will take you through some of the extensions of basic kinetics that you need in order to understand the major changes that occur when reactions take place in liquid solutions.

What's different about kinetics in liquid solutions?

Most of the added complications of kinetics and rate processes in liquid solutions arise from the *much higher density* of the liquid phase. In a typical gas at atmospheric pressure, the molecules occupy only about 0.2 per cent of the volume; the other 99.8 percent is empty space. In a liquid, molecules may take up more than half the volume, and the "empty" spaces are irregular and ever-changing as the solvent molecules undergo thermal motions of their own.

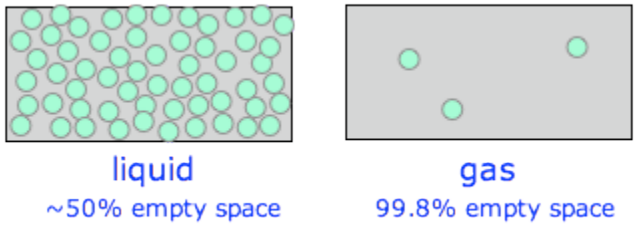


Figure 17.5.1: Liquid vs. Gases.

In a typical liquid solution, the solvent molecules massively outnumber the reactant solute molecules, which tend to find themselves momentarily (~10^{-11} sec) confined to a "hole" within the liquid. This trapping becomes especially important when the solvent is strongly hydrogen-bonded as is the case with water or alcohol.

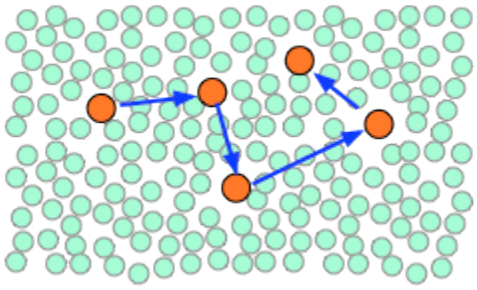
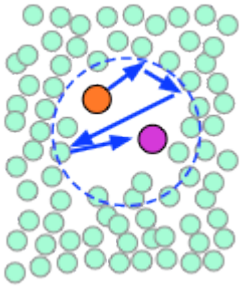
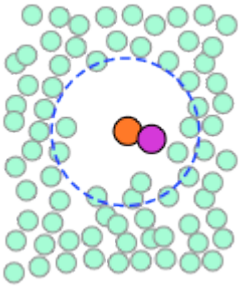
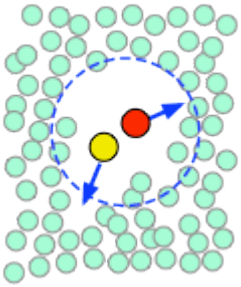
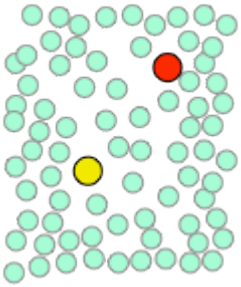


Figure 17.5.2: Brownian motion of a particle in solution

When thermal motions occasionally release a solute molecule from this trap, it will jump to a new location. The jumps are very fast (10^{-12} - 10^{-13} sec) and short (usually a few solvent-molecule diameters), and follow an entirely random pattern, very much as in **Brownian motion**. Consider a simple bimolecular process A + B → products. The reactant molecules will generally be jumping from hole to hole in the solvent matrix, only occasionally finding themselves in the same *solvent cage* where thermal motions are

likely to bring them into contact.

Table 17.5.1*: Solvent cages and encounter pairs*

			
A pair of reactants end up in the same solvent cage, where they bounce around randomly and exchange kinetic energy with the solvent molecules.	Eventually the two reactants form an *encounter pair*. If they fail to react the first time, they have many more opportunities during the lifetime of the cage.	The products form and begin to move away from each other.	Finally, after about 10^{-11} sec, the solvent cage breaks up and the products diffuse away.

The process can be represented as

$$A + B \rightarrow \{AB\} \rightarrow \text{products} \tag{17.5.1}$$

in which the $\{AB\}$ term represents the caged reactants including the *encounter pair* and the activated complex.

Contrast this scenario with a similar reaction taking place in the gas phase; the molecules involved in the reaction will often be the only ones present, so a significant proportion of the collisions will be A-B encounters. However, if the collision should fail to be energetically or geometrically viable, the reactant molecules fly apart and are unlikely to meet again anytime soon. In a liquid, however, the solute molecules are effectively in a constant state of collision — if not with other reactants, then with solvent molecules which can exchange kinetic energy with the reactants. So once an A-B encounter pair forms, the two reactants get multiple whacks at each other, greatly increasing the probability that they will obtain the kinetic energy needed to kick them over the activation hump before the encounter pair disintegrates.

Limiting Cases: Diffusion-Controlled and Activation-Controlled Reactions

The encounter pair model introduces some new rate parameters:

$$\mathrm{A + B} \underset{\blacksquare}{\overset{k_1}{\rightleftharpoons}} \mathrm{AB} \longrightarrow \text{products} \tag{17.5.2}$$

The first step is an equilibrium between the reactants outside and inside the solvent cage. The rate constants k_1 and k_2 reflect those relating to diffusion of molecules through the solvent; their values are strongly dependent on the viscosity (and thus the temperature) of the solvent. (Note that k_1 is a second-order rate constant, while k_2 is first-order.)

Diffusion is the transport of a substance through a concentration gradient; that is, from a region of higher concentration to one of lower concentration. Think of the way the color of tea spreads out when a tea bag is immersed in hot water. Diffusion occurs because random thermal motions are statistically more likely to move molecules out of a region of higher concentration than in the reverse direction, simply because in the latter case fewer molecules are available to make the reverse trip. Eventually the concentrations become uniform and equilibrium is attained.

As molecules diffuse through a liquid, they must nudge neighboring molecules out of the way. The work required to do this amounts to an activation energy, so diffusion can be thought of as a kinetic process with its own rate constant k_d and activation energy. These parameters depend on the sizes of the solute and solvent molecules and on how strongly the latter interact with each other. This suggests two important *limiting cases* for reactions in solution.

For water at room temperature, k_1 is typically 10^9-10^{10} dm^{-3} mol^{-1} s^{-1} and k_2 is around 10^{-9}-10^{-10} dm^{-3} mol^{-1} s^{-1}. Given these values, $k_3 > 10^{12}$ s^{-1} implies diffusion control, while values $< 10^9$ s^{-1} are indicative of activation control.

- **Diffusion Controlled** ($k_3 \gg k_2$): If the activation energy of the A+B reaction is very small or if escape of molecules from the {AB} cage is difficult, the kinetics will be dominated by k_1, and thus by the activation energy of diffusion. Such a process is said to be *diffusion controlled*. Reactions in aqueous solution in which E_a > 20 kJ/mol are likely to fall into this category.
- **Activation Controlled** ($k_3 \ll k_2$): Alternatively, if the activation energy of the A+B reaction dominates the kinetics, and the reaction is *activation-controlled*.

Several general kinds of reactions are consistently very "fast" and thus are commonly found to be diffusion-controlled in most solvents:

Gas-phase rate constants are normally expressed in units of mol s^{-1}, but rate constants of reactions in solution are conventionally given in mol/L units, or dm^3 mol^{-1} s^{-1}. Conversion between them depends on a number of assumptions and is non-trivial.

- **Recombination of atoms and radicals**

  For example the formation of I_2 from I atoms in hexane at 298 K has k_3 = 1.3×10^{12} dm^3 mol^{-1} s^{-1}.

- **Acid-base reactions** which involve the transport of H^+ and OH^- ions tend to be very fast.

  The most famous of these is one of the fastest reactions known:

  $$H^+ + OH^- \rightarrow H_2O$$

  for which k_3 = 1.4×10^{11} dm^3 mol^{-1} s^{-1} at 298 K.

Polar solvents such as water and alcohols interact with ions and polar molecules through attractive dipole-dipole and ion-dipole interactions, leading to lower-energy solvated forms which stabilize these species. In this way, a polar solvent can alter both the thermodynamics and kinetics (rate) of a reaction.

Solvent Thermodynamic Effect

If the products of the reaction are markedly more or less polar than the reactants, solvent polarity can change the overal thermodynamics (equilibrium constant) of the reaction. Nowhere is this more apparent than when an ionic solid such as salt dissolves in water. The Na^+ and Cl^- ions are bound together in the solid through strong coulombic forces; pulling the solid apart in a vacuum or in a nonpolar solvent is a highly endothermic process. In contrast, dissolution of NaCl in water is slightly exothermic and proceeds spontaneously.

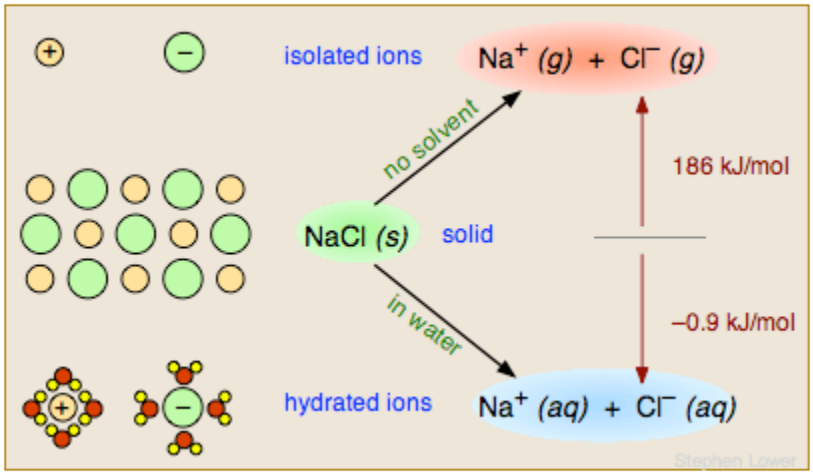


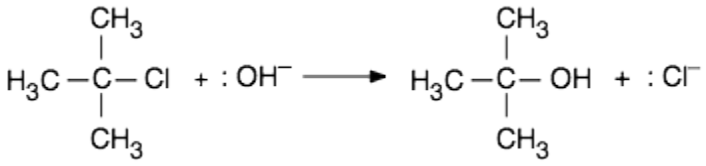
Figure 17.5.3: The solvent thermodynamics effect for table salt in water.

The water facilitates this process in two important ways. First, its high dielectric constant of 80 reduces the force between the

separated ions to 1/80 of its normal value. Secondly, the water molecules form a *solvation shell* around the ions (lower left), rendering them energetically (thermodynamically) more stable than they were in the NaCl solid.

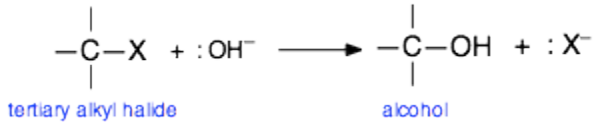
Solvent Kinetic Effect

In the same way, a reaction whose mechanism involves the formation of an intermediate or activated complex having a polar or ionic character will have its activation energy, and thus its rate, subject to change as the solvent polarity is altered. As an example we will consider an important class of reactions that you will hear much about if you take a course in organic chemistry. When an aqueous solution of a strong base such as KOH is added to a solution of *tertiary*-butyl chloride in ethanol, the chlorine is replaced by a hydroxyl group, leaving *t*-butyl alcohol as a product:

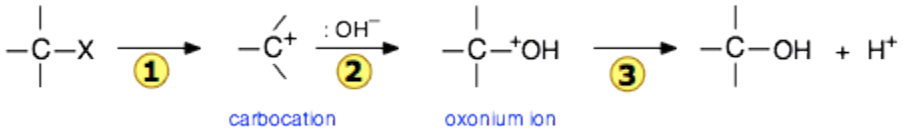


This reaction is one of a large and important class known as *S_N1 nucleophilic substitution processes* that are discussed in most organic chemistry courses. In these reactions, a species that possesses a pair of non-bonding electrons (also called a *nucleophile* or *Lewis base*) uses them to form a new bond with an *electrophile* — a compound in which a carbon atom has a partial positive charge owing to its bonds to electron-withdrawing groups. In the example here, other nucleophiles such as NH_3 or even H_2O would serve as well.

In order to reflect the generality of this process and to focus on the major changes that take place, we will represent this reaction as



Extensive studies of this class of reactions in the 1930's revealed that it proceeds in two activation energy-controlled steps, followed by a simple dissociation into the products:



In step (1), which is rate-determining, the chlorine leaves the alkyl chloride which becomes an intermediate known as a *carbocation* ("cat-ion"). These ions, in which the central carbon atom lacks a complete octet, are highly reactive, and in step (2) the carbocation is attacked by the hydroxide ion which supplies the missing electron. The immediate product is another cation in which the positive charge is on the oxygen atom. This *oxonium ion* is unstable and rapidly dissociates ((3))into the alcohol and a hydrogen ion.

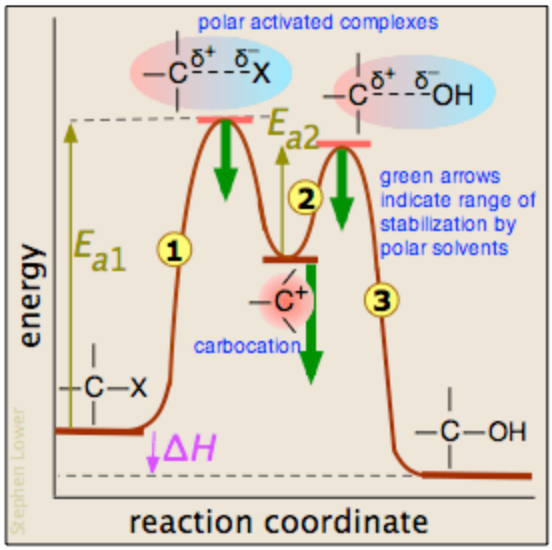


Figure 17.5.4: reaction Coordinate

The reaction coordinate diagram helps us understand the effect of solvent polarity on this reaction. Polar solvent molecules interact most strongly with species in which the electric charge is concentrated in one spot. Thus the carbocation is stabilized to a greater degree than are the activated complexes in which the charge is spread out between the positive and negative ends. As the heavy green arrows indicate, a more polar solvent will stabilize the carbocation more than it will either of the activated complexes; the effect is to materially reduce the activation energy of the rate-determining step, and thus speed up the reaction. Because neither the alkyl chloride nor the alcohol is charged, the change in solvent polarity has no effect on the equilibrium constant of the reaction. This is dramatically illustrated by observing the rate of the reaction in solvents composed of ethanol and water in varying amounts:

Table 17.5.2: Data

% water	10	20	30	40	50	60
$k_1 \times 10^6$	1.7	9.1	40.3	126	367	1294

This page titled 17.5: Kinetics of Reactions in Solution is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17.6: Catalysts and Catalysis

Learning Objectives

Make sure you thoroughly understand the following essential ideas which have been presented above. It is especially imortant that you know the precise meanings of all the green-highlighted terms in the context of this topic.

- What are *catalysts*, and how do they work in terms altering the parameters of a reaction?
- Describe the similarities and differences between the three principal classes of catalysts.
- Define *physisorption* and *chemisorption*, and explain the role of the latter in initiating a catalytic event.
- What is the meaning and significance of the *Sabatier Principle*?
- What are the principal differences between the *Langmuir-Hinshelwood* and *Eley-Rideal* mechanisms of heterogeneous catalysis?
- Describe the role of the *enzyme-substrate complex*.
- How might the particular properties of the *amino-acids* surrounding the active site of an enzyme contribute to its catalytic effect?
- Describe the *lock-and-key* and *induced-fit* models of enzyme action.
- Explain the function and significance of *allosteric sites* on an enzyme.

It almost seems like magic! A mixture of gaseous H_2 and O_2 can coexist indefinitely without any detectable reaction, but if a platinum wire is heated and then immersed in the gaseous mixture, the reaction proceeds gently as the heat liberated by the reaction makes the wire glow red-hot. Catalysts play an essential role in our modern industrial economy, in our stewardship of the environment, and in all biological processes. This lesson will give you a glimpse into the wonderful world of catalysts, helping you to understand what they are and how they work.

What are Catalysts?

Catalysts have no effect on the equilibrium constant and thus on the equilibrium composition. Catalysts are substances that speed up a reaction but which are not consumed by it and do not appear in the net reaction equation. Also — and this is very important — *catalysts affect the forward and reverse rates equally*; this means that catalysts have no effect on the equilibrium constant and thus on the composition of the equilibrium state. Thus a catalyst (in this case, sulfuric acid) can be used to speed up a reversible reaction such as ester formation or its reverse, ester hydrolysis:

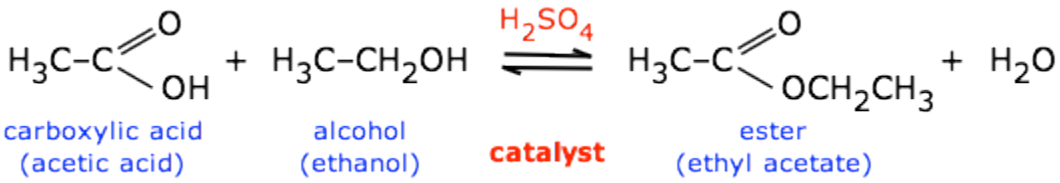


Figure 17.6.1: An acid catalyzed reactions

The catalyst has no effect on the equilibrium constant or the direction of the reaction. The direction can be controlled by adding or removing water (Le Chatelier principle).

Catalysts function by allowing the reaction to take place through an alternative mechanism that requires a smaller activation energy. This change is brought about by a specific interaction between the catalyst and the reaction components. You will recall that the rate constant of a reaction is an exponential function of the activation energy, so even a modest reduction of E_a can yield an impressive increase in the rate.

Catalysts provide alternative reaction pathways

Catalysts are conventionally divided into two categories: *homogeneous* and *heterogeneous*. *Enzymes*, natural biological catalysts, are often included in the former group, but because they share some properties of both but exhibit some very special properties of their own, we will treat them here as a third category.

Some common examples of catalysis

How to burn a Sugar Cube

When heated by itself, a sugar cube (sucrose) melts at 185°C but does not burn. But if the cube is rubbed in cigarette ashes, it burns before melting owing to the catalytic action of trace metal compounds in the ashes.

Platinum as an Oxidation Catalyst

The surface of metallic platinum is an efficient catalyst for the oxidation of many fuel vapors. This property is exploited in flameless camping stoves (left). The image at the right shows a glowing platinum wire heated by the slow combustion of ammonia on its surface. However, if you dip a heated Pt wire into *liquid* ammonia, you get a miniature explosion: see video below.

YouTube Video

This interactive element is not available in this format. Visit the link or scan the QR code to access it online.

View this YouTube Video online

This is oxidation of ammonia by oxygen catalysed by warmed platinum wire. Oxygen is bubbled through ammonia solution, in which it mixes with ammonia gas present. The reaction causes the platinum wire to glow, and the hot wire ignites a mixture of ammonia and oxygen.

Decomposition of Hydrogen Peroxide

Hydrogen peroxide is thermodynamically unstable according to the reaction

$$2\,H_2O_2 \longrightarrow 2\,H_2O + O_2 \qquad \Delta G^o = -210\,kJ\,mol^{-1} \tag{17.6.1}$$

In the absence of contaminants this reaction is very slow, but a variety of substances, ranging from iodine, metal oxides, trace amount of metals, greatly accelerate the reaction, in some cases almost explosively owing to the rapid release of heat. The most effective catalyst of all is the enzyme *catalase*, present in blood and intracellular fluids; adding a drop of blood to a solution of 30% hydrogen peroxide induces a vigorous reaction.

Rapid liberation of O_2 can result in a spectacular bubble bath if some soap is added. [This same reaction has been used to power a racing car!

Potassium iodide efficiently catalyzes H_2O_2 deposition. This short video shows what happens when some colored soap, H_2O_2, and KI are combined. But "don't try this at [your] home"!

Each kind of catalyst facilitates a different pathway with its own activation energy. Because the rate is an exponential function of E_a (Arrhenius equation), even relatively small differences in E_a's can have dramatic effects on reaction rates. Note especially the values for catalase; the chemist is still a rank amateur compared to what Nature can accomplish through natural selection!

catalyst	Ea kJ/mol	relative rate
no catalyst	75	1
iodide ion	56	2140

colloidal platinum	50	24,000
catalase (enzyme)	21	2,900,000,000

How catalytic activity is expressed

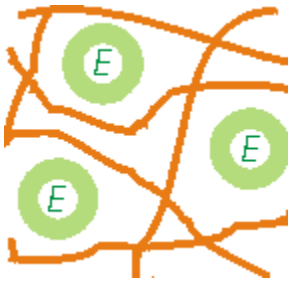
Changes in the rate constant or of the *activation energy* are obvious ways of measuring the efficacy of a catalyst. But two other terms have come into use that have special relevance in industrial applications.

Turnover number

The turnover number (TON) is an average number of cycles a catalyst can undergo before its performance deteriorates (see below). Reported TONs for common industrial catalysts span a very wide range from perhaps 10 to well over 10^5, which approaches the limits of diffusion transport.

Turnover Frequency

This term, which was originally applied to enzyme-catalyzed reactions, has come into more general use. It is simply the number of times the overall catalyzed reaction takes place per catalyst (or per active site on an enzyme or heterogeneous catalyst) per unit time:is defined as



The number of active sites S on a heterogeneous catalyst is often difficult to estimate, so it is often replaced by the total area of the exposed catalyst, which is usually experimentally measurable. TOFs for heterogeneous reactions generally fall between 10^{-2} to 10^2 s^{-1}.

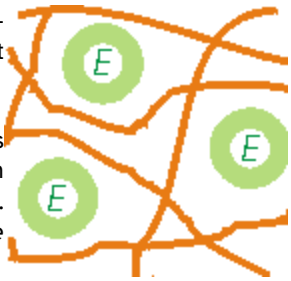
Homogeneous Catalysis

As the name implies, homogeneous catalysts are present in the same phase (gas or liquid solution) as the reactants. Homogeneous catalysts generally enter directly into the chemical reaction (by forming a new compound or complex with a reactant), but are released in their initial form after the reaction is complete, so that they do not appear in the net reaction equation.

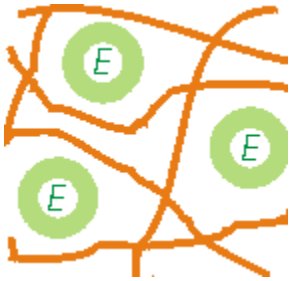
Iodine-catalyzed *cis-trans* isomerization

Unless you are taking an organic chemistry course in which your instructor indicates otherwise, **don't try to memorize these mechanisms**. They are presented here for the purpose of convincing you that catalysis is not black magic, and to familiarize you with some of the features of catalyzed mechanisms. It should be sufficient for you to merely convince yourself that the individual steps make chemical sense.

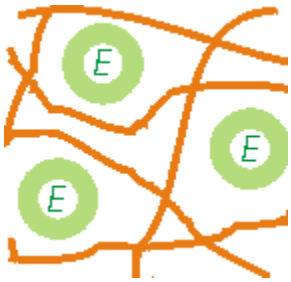
You will recall that *cis-trans isomerism* is possible when atoms connected to each of two doubly-bonded carbons can be on the same (*cis*) or opposite (*trans*) sides of the bond. This reflects the fact that rotation about a double bond is not possible.

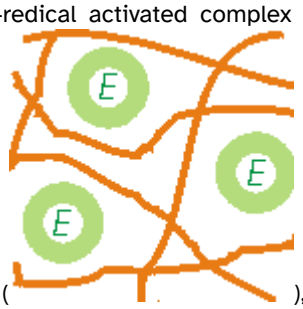


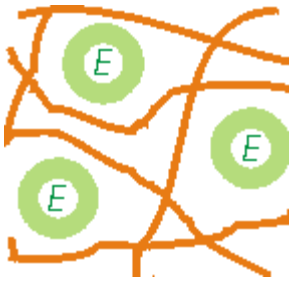
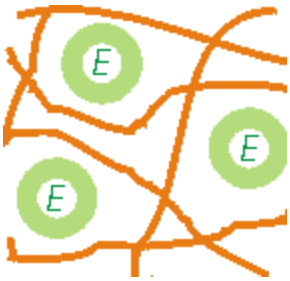
Conversion of an alkene between its *cis*- and *trans* forms can only occur if the double bond is temporarily broken, thus freeing the two ends to rotate. Processes that cleave covalent bonds have high activation energies, so cis-trans isomerization reactions tend to be slow even at high temperatures. Iodine is one of several catalysts that greatly accelerate this process, so the isomerization of *butene* serves as a good introductory example of homogeneous catalysis.

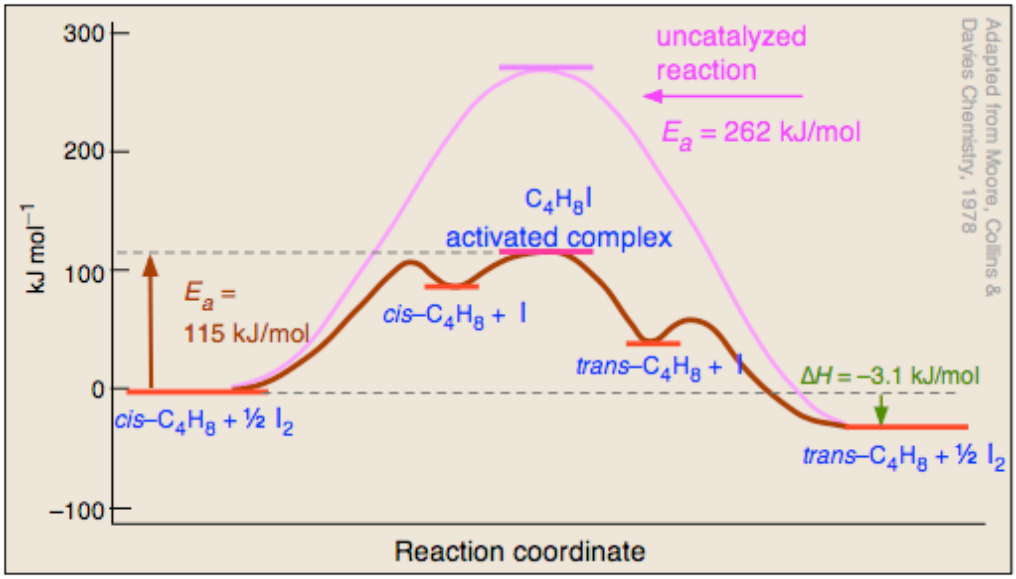


The mechanism of the iodine-catalyzed reaction is believed to involve the attack of iodine atoms (formed by the dissociation equilibrium  on one of the carbons in Step 2:



During its brief existence, the free-radical activated complex can undergo rotation about the C—C bond, so that when it decomposes by releasing the iodine (), a portion of the reconstituted butene will be in its *trans* form.

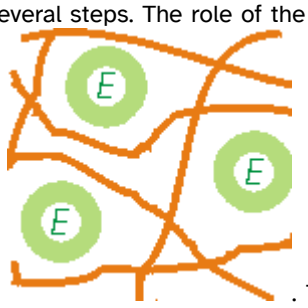
Finally, the iodine atom recombine into diiodine. Since processes  and  cancel out, iodine does not appear in the net reaction equation — a requirement for a true catalyst.

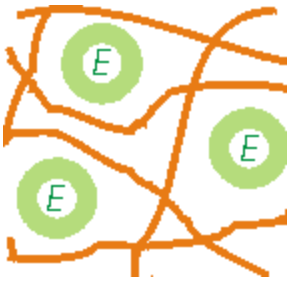


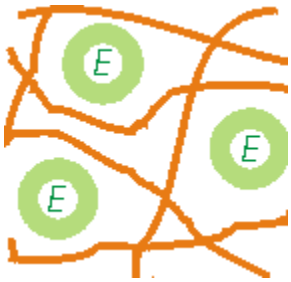
Acid-base catalysis

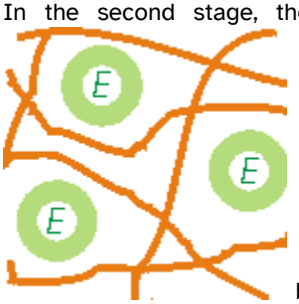
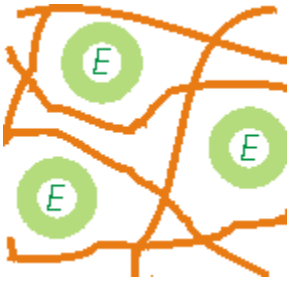
Many reactions are catalyzed by the presence of an acid or a base; in many cases, both acids and bases will catalyze the same reaction. As one might expect, the mechanism involves the addition or removal of a proton, changing the reactant into a more kinetically labile form. A simple example is the addition of iodine to propanone

$I_2 + (CH_3)_2–C=O \rightarrow (CH_2I)(CH_3)–C=O$

The mechanism for the acid-catalyzed process involves several steps. The role of the acid is to provide a proton that attaches to the carbonyl oxygen, forming an unstable oxonium ion  . The latter rapidly rearranges into an enol


(i.e., a carbon connected to both a double bond (*ene*) and a hydroxyl (*ol*) group.) This completes the catalytic part of the process, which is basically an acid-base (proton-transfer) reaction in which the role of the proton is to withdraw an electron from the ketone oxygen.



In the second stage, the enol reacts with the iodine. The curved arrows indicate shifts in electron locations. In 
below, an electron is withdrawn from the π orbital of the double bond by one of the atoms of the I_2 molecule. This induces a shift of electrons in the latter, causing half of this molecule to be expelled as an iodide ion. The other half of the iodine is now an iodonium ion I^+ which displaces a proton from one of the methyl groups. The resultant carbonium ion 
then expels the -OH proton to yield the final neutral product.

Perhaps the most well-known acid-catalyzed reaction is the hydrolysis (or formation) of an ester — a reaction that most students encounter in an organic chemistry laboratory course. This is a more complicated process involving five steps; its mechanism is discussed here. See also this U. Calgary site, which describes both the acid- and base-catalyzed reaction.

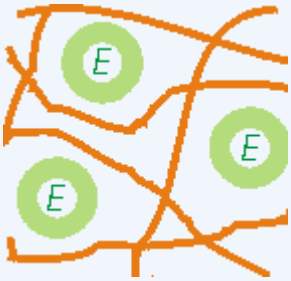
Oxidation-reduction catalysis

Many oxidation-reduction (electron-transfer) reactions, including direct oxidation by molecular oxygen, are quite slow. Ions of transition metals capable of existing in two oxidation states can often materially increase the rate. An example would be the reduction of Fe^{3+} by the vanadium ion V^{3+}:

$$V^{3+} + Fe^{3+} \longrightarrow V^{4+} + Fe^{2+} \tag{17.6.2}$$

This reaction is catalyzed by either Cu^+or Cu^{3+}, and the rate is proportional to the concentration of $V^{3+}$and of the copper ion, but

independent of the Fe^{3+} concentration. The mechanism is believed to involve two steps:

1 $V^{3+} + Cu^{2+} \xrightarrow{k_1} V^{4+} + Cu^{+}$	(rate-determining)
2 $Fe^{3+} + Cu^{+}$  $Fe^{2+} + Cu^{2+}$	(very fast)

(If Cu^{+} is used as the catalyst, it is first oxidized to Cu^{2+} by step 2.)

Hydrogen Peroxide, again

Ions capable of being oxidized by an oxidizing agent such as H_2O_2 can serve as catalysts for its decomposition. Thus H_2O_2 oxidizes iodide ion to iodate, when then reduces another H_2O_2 molecule, returning an I^- ion to start the cycle over again:

$$H_2O_2 + I^- \rightarrow H_2O + IO^- \tag{17.6.3}$$

$$H_2O_2 + IO^- \rightarrow H_2O + O_2 + I^- \tag{17.6.4}$$

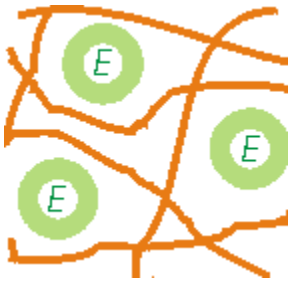
Iron(II) can do the same thing. Even traces of metallic iron can yield enough Fe^{2+} to decompose solutions of hydrogen peroxide.

$$H_2O_2 + Fe^{2+} \rightarrow H_2O + Fe^{3+} \tag{17.6.5}$$

$$H_2O_2 + Fe^{3+} + 2H^+ \rightarrow H_2O + O_2 + Fe^{2+} + 2H^+ \tag{17.6.6}$$

Heterogeneous catalysts

As its name implies, a heterogeneous catalyst exists as a separate phase (almost always a solid) from the one (most commonly a gas) in which the reaction takes place. The catalytic affect arises from disruption (often leading to dissociation) of the reactant molecules brought about by their interaction with the surface of the catalyst.



← Model of a catalyst consisting of clusters of 8-10 platinum atoms (blue) deposited on an aluminum oxide surface. This catalyst efficiently removes hydrogen atoms from propane, converting it into the industrially-important propylene. [source]

Unbalanced forces at surfaces

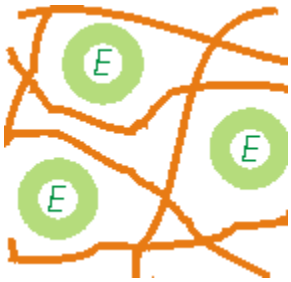
You will recall that one universal property of matter is the weak attractive forces that arise when two particles closely approach each other. (See here for a quick review.) When the particles have opposite electric charges or enter into covalent bonding, these

far stronger attraction dominate and define the "chemistry" of the interacting species.

The molecular units within the bulk of a solid are bound to their neighbors through these forces which act in opposing directions to keep each under a kind of "tension" that restricts its movement and contributes to the cohesiveness and rigidity of the solid.

At the surface of any kind of condensed matter, things are quite different. The atoms or molecules that reside on the surface experience unbalanced forces which prevents them from assuming the same low potential energies that characterize the interior units. (The same thing happens in liquids, and gives rise to a variety of interfacial effects such as surface tension.)

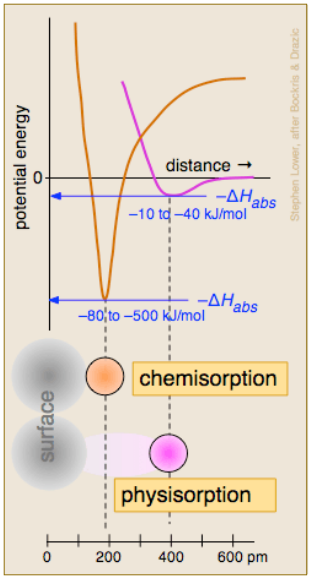
But in the case of a solid, in which the attractive forces tend to be stronger, something much more significant happens. The molecular units that reside on the surface can be thought of as partially buried in it, with their protruding parts (and the intermolecular attractions that emerge from them) exposed to the outer world. The strength of the attractive force field which emanates from a solid surface varies in strength depending on the nature of the atoms or molecules that make up the solid.



Don't confuse a**d**sorption with a**b**sorption; the latter refers to the bulk uptake of a substance into the interior of a porous material. At the microscopic level, of course, absorption also involves adsorption. The process in which molecules in a gas or a liquid come into contact with and attach themselves to a solid surface is known as *adsorption*. Adsorption is almost always an exothermic process and its strength is conventionally expressed by the enthalpy or "heat" of adsorption ΔH_{ads}.

Chemisorption and Physisorption

Two general categories of adsorption are commonly recognized, depending on the extent to which the electronic- or bonding structure of the attached molecule is affected. When the attractive forces arise from relatively weak van der Waals interactions, there is little such effect and ΔH_{ads} tends to be small. This condition is described as *physical adsorption* (*physisorption*). Physisorption of a gas to a surface is energetically similar to the condensation of the gas to a liquid, it usually builds up multiple layers of adsorbed molecules, and it proceeds with zero activation energy.

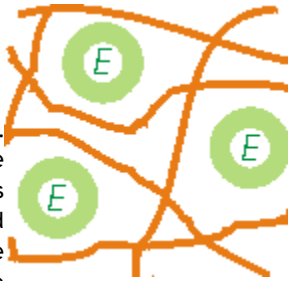


Of more relevance to catalytic phenomena is *chemisorption*, in which the adsorbate is bound to the surface by what amounts to a

chemical bond. The resulting disruption of the electron structure of the adsorbed species "activates" it and makes it amenable to a chemical reaction (often dissociation) that could not be readily achieved through thermal activation in the gas or liquid phase. In contrast to physisorption, chemisorption generally involves an activation energy (supplied by ΔH_{ads}) and the adorbed species is always a monolayer.

Mechanisms of reactions on surfaces

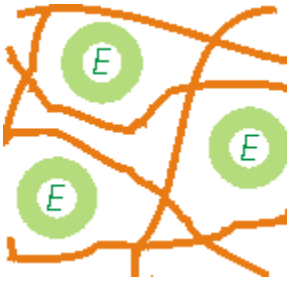
Dissociative adsorption

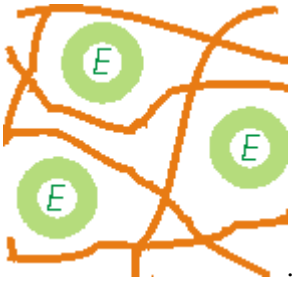


The simplest heterogeneous process is chemisorption followed by bond-breaking as described above. The most common and thoroughly-studied of these is the dissociation of hydrogen which takes place on the surface of most transition metals. The single 1s electron of each hydrogen atom coordinates with the *d* orbitals of the metal, forming a pair of chemisorption bonds (indicated by the red dashed lines). Although these new bonds are more stable than the single covalent bond they replace, the resulting hydrogen atoms are able to migrate along the surface owing to the continuous extent of the *d*-orbital conduction band.

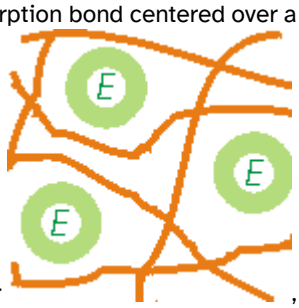
The Langmuir-Hinshelwood mechanism

Although the adsorbed atoms ("*adatoms*") are not free radicals, they are nevertheless highly reactive, so if a second, different molecular species adsorbs onto the same surface, an interchange of atoms may be possible. Thus carbon monoxide can be oxidized to CO_2 by the process illustrated below:

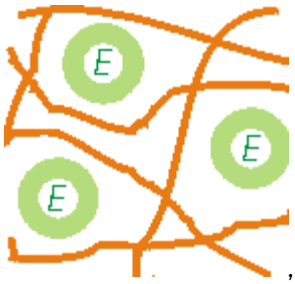




In this example, only the O_2 molecule undergoes dissociation . The CO molecule adsorbs without dissociation (2), configured perpendicular to the surface with the chemisorption bond centered over a hollow space between the



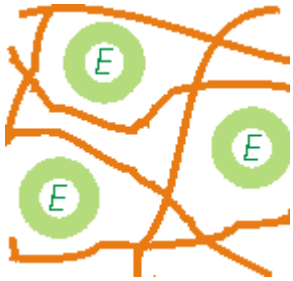
metal atoms. After the two adsorbed species have migrated near each other , the oxygen atom switches



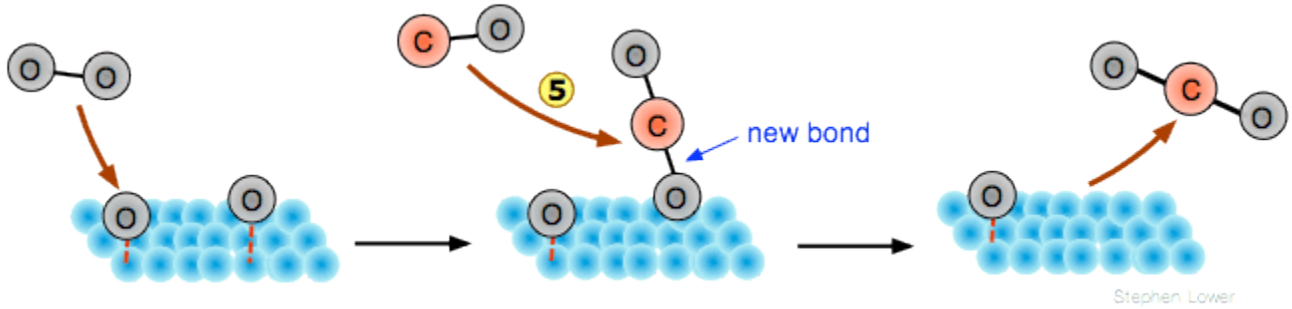
its attachment from the metal surface to form a more stable C=O bond with the carbon , followed by release of the product molecule.

The Eley-Rideal mechanism

An alternative mechanism eliminates the second chemisorption step; the oxygen adatoms react directly with the gaseous CO



molecules by replacing the chemisorption bond with a new C–O bond as they swoop over the surface:



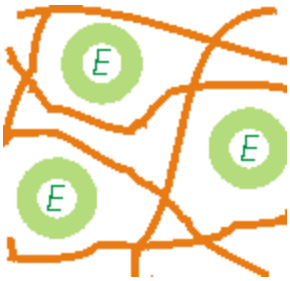
Examples of both mechanisms are known, but the Langmuir-Hinshelwood mechanism is more importantin that it exploits the activation of the adsorbed reactant. In the case of carbon monoxide oxdation, studies involving molecular beam experiments support this scheme. A key piece of evidence is the observation of a short time lag between contact of a CO molecule with the surface and release of the CO_2, suggesting that CO remains chemisorbed during the interval.

The Sabatier Principle

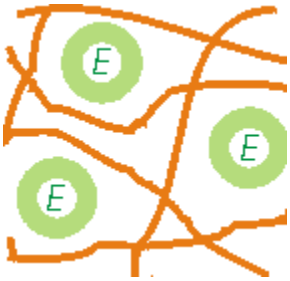
To be effective, these processes of adsorption, reaction, and desorption must be orchestrated in a way that depends critically on the properties of the catalyst in relation to the chemisorption properties (ΔH_{ads}) of the reactants and products.

- Adsorption of the reactant onto the catalytic surface (**2**) must be strong enough to perturb the bonding within the species to dissociate or activate it;
- If the resulting fragments must migrate to other locations on the surface (**3-4**), their chemisorption must be weak enough to allow this movement but not so small that they escape before they have a chance to react;
- The product species must have sufficiently small ΔH_{ads} values to ensure their rapid desorption from the catalyst (**5**) so that surface is freed up to repeat the cycle

The importance of choosing a catalyst that achieves the proper balance of the heats of adsorption of the various reaction components is known as the *Sabatier Principle*, but is sometimes referred to as the "just-right" or "Goldilocks principle". Remember the story of Goldilocks and the Three Bears? ... or see this UTube video.



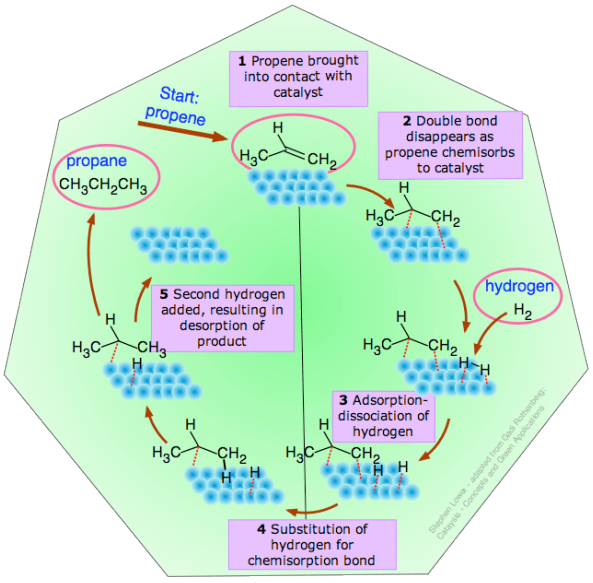
In its application to catalysis, this principle is frequently illustrated by a "volcano diagram" in which the rate of the catalyzed reaction is plotted as a function of ΔH_{ads} of a substrate such as H_2 on a transition metal surface.



The plot at the left shows the relative effectiveness of various metals in catalyzing the decomposition of formic acid HCOOH. The vertical axis is plotted as temperature, the idea being that the better the catalyst, the lower the temperature required to maintain a given rate.

The catalytic cycle

This term refers to the idealized sequence of steps between the adsorption of a reactant onto the catalyst and the desorption of the product, culminating in restoration of the catalyst to its original condition. A typical catalytic cycle for the hydrogenation of propene is illustrated below.



This particular reaction

$H_3C–CH=CH_2 + H_2 \rightarrow H_3C–CH_2–CH_3$

takes place spontaneously only in the reverse direction, but it is representative of the process used to hydrogenate the carbon-

carbon double bonds in vegetable oils to produce solid saturated fats such as margarine.

Catalyst poisoning and breakdown

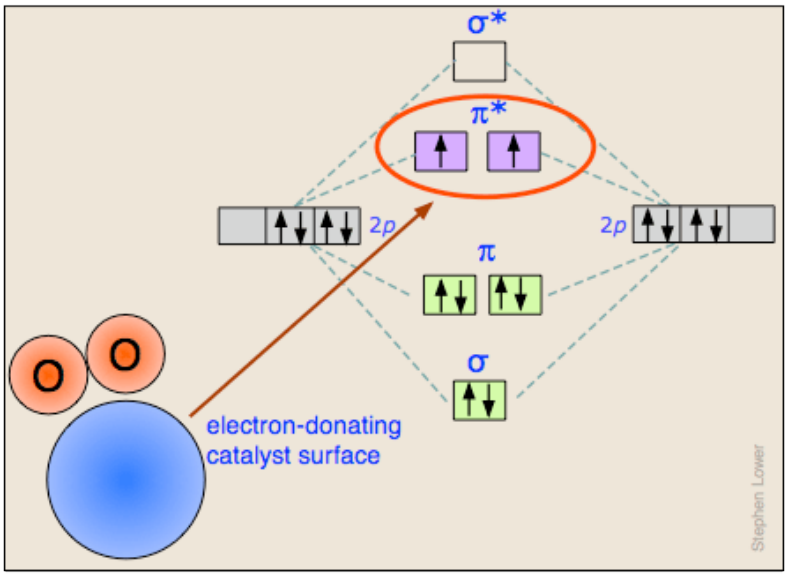
Catalyst poisoning, brought about by irreversible binding of a substance to its surface, can be permanent or temporary. In the latter case the catalyst can be regenerated, usually by heating to a high temperature. In organisms, many of the substances we know as "poisons" act as catalytic poisons on enzymes. If catalysts truly remain unchanged, they should last forever, but in actual practice, various events can occur that limit the useful lifetime of many catalysts.

- Impurities in the feedstock or the products of side reactions can bind permanently to a sufficient number of active sites to reduce catalytic efficiency over time, "poisoning" the catalyst.
- Physical deterioration of the catalyst or of its support, often brought about by the high temperatures sometimes used in industrial processes, can reduce the effective surface area or the accessibility of reactants to the active sites.

Catalysts tend to be rather expensive, so it is advantageous if they can be reprocessed or regenerated to restore their activity. It is a common industrial practice to periodically shut down process units to replace spent catalysts.

How heterogeneous catalysts work

The actual mechanisms by which adsorption of a molecule onto a catalytic surface facilitates the cleavage of a bond vary greatly from case to case.

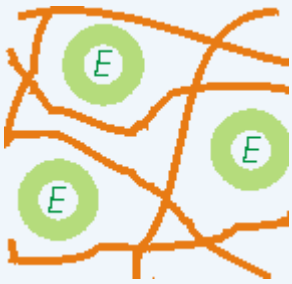
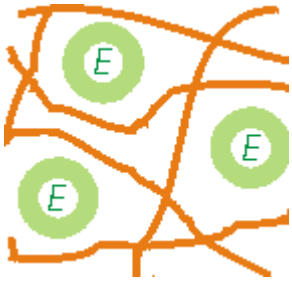
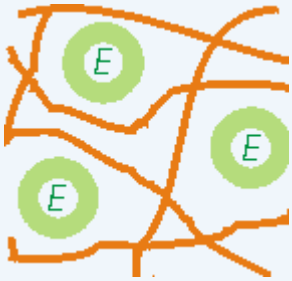
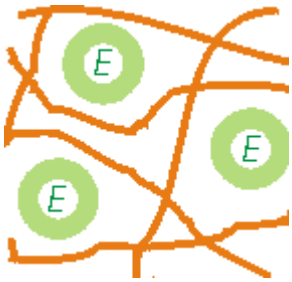
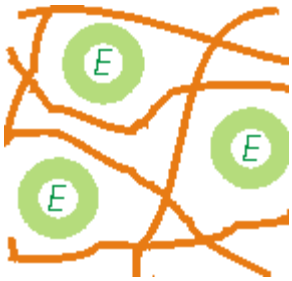


We give here only one example, that of the dissociation of dixoygen O_2 on the surface of a catalyst capable of temporarily donating an electron which enters an oxygen antibonding molecular orbital that will clearly destabilize the O–O bond. (Once the bond has been broken, the electron is given back to the catalyst.)

Types of catalytically active surfaces

Heterogeneous catalysts mostly depend on one or more of the followng kinds of surfaces:

Active surface type	Remarks

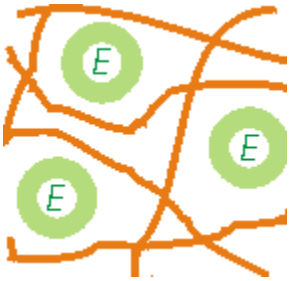
	Atoms at surfaces or crystal edges in macromolecular 2- and 3-D networks such as graphite or quartz may have free valences
	Surfaces and edges are sites of intense electric fields able to interact with ions and polar molecules
	Oxides can acquire H^+ and/or OH^- groups able to act as acid- or base catalyts
	"Electron gas" at metal surfaces can perturb bonding in substrate molecules
	Vacant *d* orbitals can provide a variety of coordination sites for activation.
	Semiconductors (including many oxides) can supply electrons, thermally excited through reasonably small (<50 kJ) band gaps.

Some factors affecting catalyst efficacy

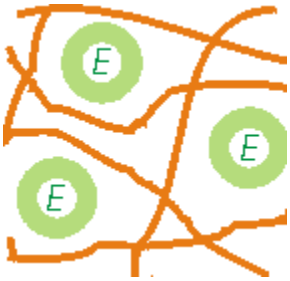
Since heterogeneous catalysis requires direct contact between the reactants and the catalytic surface, the area of active surface

goes at the top of the list. In the case of a metallic film, this is not the same as the nominal area of the film as measured by a ruler; at the microscopic level, even apparently smooth surfaces are highly irregular, and some cavities may be too small to accommodate reactant molecules.

Consider, for example, that a 1-cm cube of platinum (costing roughly $1000) has a nominal surface area of only 6 cm^2. If this is broken up into 10^{12} smaller cubes whose sides are 10^{-6} m, the total surface area would be 60,000 cm^2, capable in principle of increasing the rate of a Pt-catalyzed reaction by a factor of 10^4. These very finely-divided (and often very expensive) metals are typically attached to an inert *supporting surface* to maximize their exposure to the reactants.



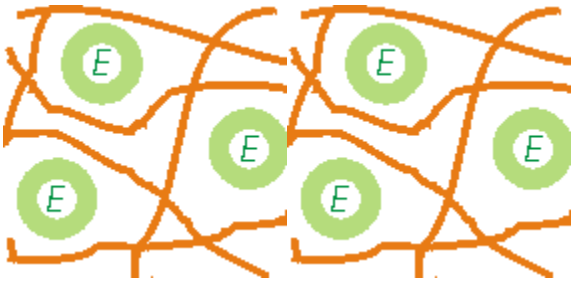
Surface topography. At the microscopic level, even an apparently smooth surface is pitted and uneven, and some sites will be more active than others. Penetration of molecules into and out of some of the smaller channels of a porous surface may become rate-limiting.



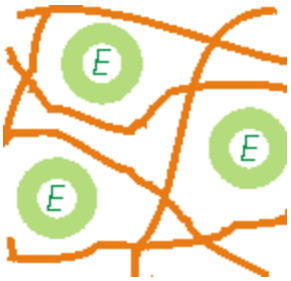
An otherwise smooth surface will always possess a variety of defects such as steps and corners which offer greater exposure and may be either the only active sites on the surface, or overly active so as to permanently bind to a reactant, reducing the active area of the surface. In one study, it was determined that kink defects constituting just 5 percent of platinum surface were responsible for over 90% of the catalytic activity in a certain reaction.

Steric factors

When chemisorbtion occurs at two or more locations on the reactant, efficient catalysis requires that the spacing of the active centers on the catalytic surface be such that surface bonds can be formed without significant angular distortion.



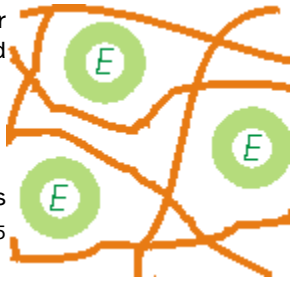
Thus activation of the ethylene double bond on a nickel surface proceeds efficiently because the angle between the C—Ni bonds and the C—C is close to the tetrahedral value of 109.5° required for carbon sp^3 hybrid orbital formation. Similarly, we can expect that the hydrogenation of benzene should proceed efficiently on a surface in which the active sites are spaced in the range of 150 and 250 pm.



This is one reason why many metallic catalysts exhibit different catalytic activity on different crystal faces.

Some special types of heterogeneous catalysts

Metal-cluster catalysts



As the particle size of a catalyst is reduced, the fraction of more highly exposed step, edge, and corner atoms increases. An extreme case occurs with nano-sized (1-2 nm) metal cluster structures composed typically of 10-50 atoms.

[link] →

Metallic gold, well known for its chemical inertness, exhibits very high catalytic activity when it is deposited as metallic clusters on an oxide support. For example, O_2 dissociates readily on Au_{55} clusters which have been found to efficiently catalyze the oxidation of hydrocarbons [article].

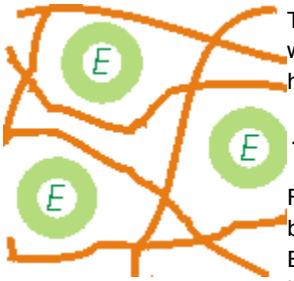
Zeolite catalysts

Zeolites are clay-like aluminosilicate solids that form open-framework microporous structures that may contain linked cages, cavities or channels whose dimensions can be tailored to the sizes of the reactants and products. To those molecules able to diffuse through these spaces, zeolites are in effect "all surface", making them highly efficient. This size-selectivity makes them important for adsorption, separation, ion-exchange, and catalytic applications. Many zeolites occur as minerals, but others are made synthetically in order to optimize their properties.

As catalyts, zeolites offer a number of advantages that has made them especially important in "green chemistry" operations in which the number of processing steps, unwanted byproducts, and waste stream volumes are minimized.

Enzymes and biocatalysis

"A Jug of Wine, a Loaf of Bread, and . . ." Enzymes!



This distortion of Robert FitzGerald's already-distorted translation of the famous quatrain from the wonderful *Rubaiyat of Omar Khayyam* underlines the central role that enzymes and their technology have played in civilization since ancient times.

← Illustration by Edmund Dulac (1882-1953)

Fermentation and wine-making have been a part of human history and culture for at least 8000 years, but recognition of the role of catalysis in these processes had to wait until the late nineteenth century. By the 1830's, numerous similar agents, such as those that facilitate protein digestion in the stomach, had been discovered. The term "enzyme", meaning "from yeast", was coined by the German physiologist Wilhelm Kühne in 1876. In 1900, Eduard Buchner (1860-1917, 1907 Nobel Prize in Chemistry) showed that fermentation, previously believed to depend on a mysterious "life force" contained in living organisms such as yeast, could be achieved by a cell-free "press juice" that he squeezed out of yeast.

By this time it was recognized that enzymes are a form of catalyst (a term introduced by Berzelius in 1835), but their exact

chemical nature remained in question. They appeared to be associated with proteins, but the general realization that enzymes *are* proteins began only in the 1930s when the first pure enzyme was crystallized, and did not become generally accepted until the 1950s. It is now clear that nearly all enzymes are proteins, the major exception being a small but important class of RNA-based enzymes known as ribozymes.

Proteins are composed of long sequences of amino acids strung together by amide bonds; this sequence defines the *primary structure* of the protein.. Their huge size (typically 200-2000 amino acid units, with total molecular weights 20,000 - 200,000) allows them to fold in complicated ways (known as secondary and tertiary structures) whose configurations are essential to their catalytic function.

Because enzymes are generally very much larger than the reactant molecules they act upon (known in biochemistry as *substrates*), enzymatic catalysis is in some ways similar to heterogeneous catalysis. The main difference is that the binding of a subtrate to the enzyme is much more selective.

Precursors and cofactors

Most enzymes come into being as inactive precursors (*zymogens*) which are converted to their active forms at the time and place they are needed.

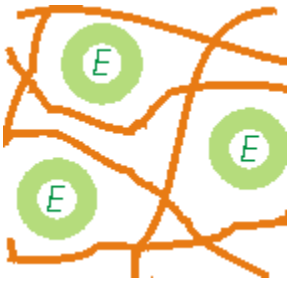
- For example, the enzymes that lead to the clotting of blood are supposed to remain inactive until bleeding actually begins; a major activating factor is exposure of the blood to proteins in the damaged vessel wall.
- The enzyme that catalyzes the formation of lactose (milk sugar) in the mammary gland is formed during pregnancy, but it remains inactive until the time of birth, when hormonal changes cause a modifier unit to bind to and activate the enzyme.

Conversion to the active form may involve a simple breaking up of the protein by hydrolysis of an appropriate peptide bond or the addition of a phosphate or similar group to one of the amino acid residues.

Many enzyme proteins also require "helper" molecules, known as *cofactors*, to make them catalytically active. These may be simple metal ions (many of the trace nutrient ions of Cu, Mn, Mo, V, etc.) or they may be more complex organic molecules which are called *coenzymes*. Many of the latter are what we commonly refer to as *vitamins*. Other molecules, known as *inhibitors*, decrease enzyme activity; many drugs and poisons act in this way.

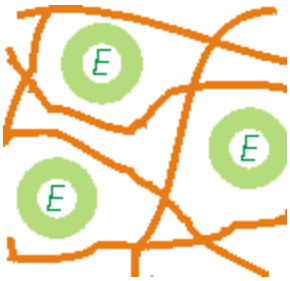
The enzyme-substrate complex

The standard model of enzyme kinetics consists of a two-step process in which an enzyme binds reversibly to its substrate S (the reactant) to form an *enzyme-substrate complex* ES:



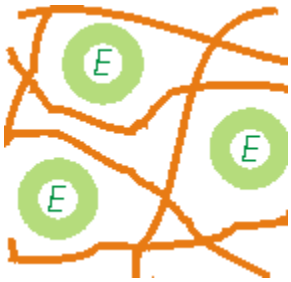
The enzyme-substrate complex plays a role similar to that of the activated complex in conventional kinetics, but the main function of the enzyme is to stabilize the transition state.

In the second, essentially irreversible step, the product and the enzyme are released:



> The overall kinetics of the process were worked out in 1913 by the German biochemist Leonor Michaelis and his Canadian student Maud Menten.

The basic kinetic treatment of this process involves the assumption that the concentrations [E] and [ES] reach steady-state values which do not change with time. (The detailed treatment, which is beyond the scope of this course, can be found here.)

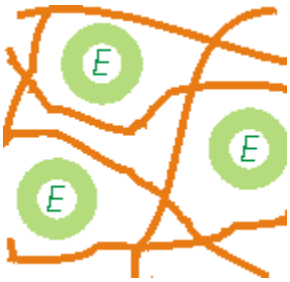


The overall process is described by the *Michaelis-Menten equation* which is plotted here. The *Michaelis constant* k_M is defined as shown, but can be simplified to the ES dissociation constant k_{-1}/k_1 in cases when dissociation of the complex is the rate-limiting step. The quantity v_{max} is not observed directly, but can be determined from k_M as shown here.

Enzymes are proteins

In order to understand enzymes and how they catalyze reactions, it is first necessary to review a few basic concepts relating to proteins and the amino acids of which they are composed.

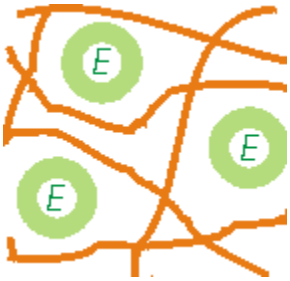
Amino acids: polar, nonpolar, positive, negative.



The 21 amino acids that make up proteins all possess the basic structure shown here, where R represents either hydrogen or a side chain which may itself contain additional $-NH_2$ or –COOH groups. Both kinds of groups can hydrogen-bond with water and with the polar parts of substrates, and therefore contribute to the amino acid's polarity and hydophilic nature. Side chains that contain longer carbon chains and especially benzene rings have the opposite effect, and tend to render the amino acid non-polar and hydrophobic.

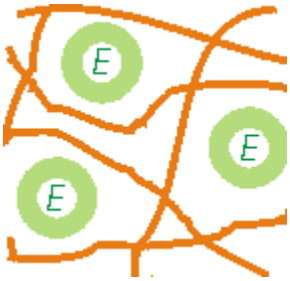
> The pK_a of an ionizable group is the pH at which half of these groups will be ionized. For example, if a –COOH group has a pK_a of 2.5, it will exist almost entirely in this form at a pH of about 1.5 or less, and as the negative ion $-COO^-$ at pH above about 3.5.

Both the $-NH_2$ and –COOH groups are ionizable (i.e., they can act as proton donors or acceptors) and when they are in their ionic forms, they will have an electric charge. The –COOH groups have pKa's in the range 1.8-2.8, and will therefore be in their ionized forms $-COO^-$ at ordinary cellular pH values of around 7.4. The amino group pKa's are around 8.8-10.6, so these will also normally be in their ionized forms NH_3^+.



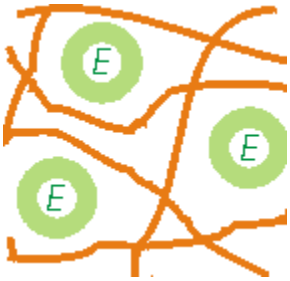
This means that at ordinary cellular pH, both the carboxyl and amino groups will be ionized. But because the charges have opposite signs, an amino acid that has no extra ionizable groups in its side chain will have a net charge of zero. But if the side chain contains an extra amino or carboxyl group, the amino acid can carry a net electric charge. The following diagram illustrates typical amino acids that fall into each of the groups described above.

> For clarity, only the charges in the side chains are shown here, but of course the carboxyl and amino groups at the head end of each amino acid will also be ionized. The small green numbers are pK_a values. See here for a complete list of amino acids that occur in proteins.



Proteins

Proteins are made up of one or more chains of amino acids linked to each other through peptide bonds by elimination of a water molecule.

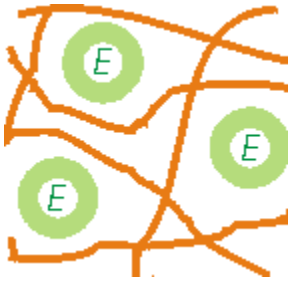


The product shown above is called a *peptide*, specifically it is a *dipeptide* because it contains two *amino acid residue*s (what's left after the water has been removed.) Proteins are simply very long polypeptide chains, or combinations of them. (The distinction between a long polypeptide and a small protein is rather fuzzy!)

Globular proteins

Most enzymes fall into the category of *globular proteins*. In contrast to the *fibrous proteins* that form the structural components

of tissues, globular proteins are soluble in water and rarely have any systematic tertiary structures. They are made up of one or more amino-acid ("*peptide*") chains which fold into various shapes that can roughly be described as spherical — hence the term "globular", and the suffix "globin" that is frequently appended to their names, as in "hemoglobin".



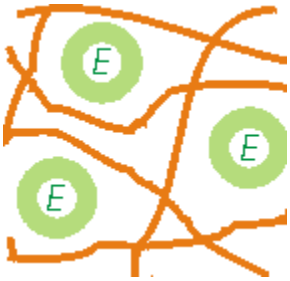
Protein folding is a spontaneous process that is influenced by a number of factors. One of these is obviously their primary amino-acid sequence that facilitates formation of intramolecular bonds between amino acids in different parts of the chain. These consist mostly of hydrogen bonds, although disulfide bonds S—S between sulfur-containing amino acids are not uncommon.

In addition to these intramolecular forces, interactions with the surroundings play an important role. The most import of these is the *hydrophobic effect*, which favors folding conformations in which polar amino acids (which form hydrogen bonds with water) are on the outside, while the so-called *hydrophobic* amino acids remain in protected locations within the folds.

How enzymes work: the active site

The catalytic process mediated by an enzyme takes place in a depression or cleft that exposes the substrate to only a few of the hundreds-to-thousands of amino acid residues in the protein chain. The high specificity and activity of enzyme catalysis is sensitively dependent on the shape of this cavity and on the properties of the surrounding amino acids

In 1894, long before it was clear that enzymes are proteins, the German chemist Emil Fischer suggested the so-called lock-and-key model as a way of understanding how a given enzyme can act specificilly on only one kind of substrate molecule. This model is essentially an elaboration of the one we still use for explaining heterogeneous catalysis.

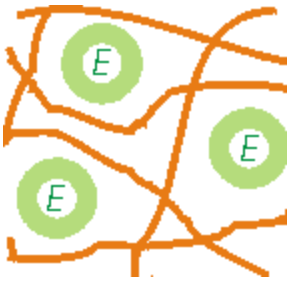


Although the basic lock-and-key model continues to be useful, it has been modified into what is now called the *induced-fit model*. This assumes that when the substrate enters the active site and interacts with the surrounding parts of the amino acid chain, it reshapes the active site (and perhaps other parts of the enzyme) so that it can engage more fully with the substrate.

One important step in this process is to squeeze out any water molecules that are bound to the substrate and which would interfere with its optimal positioning.

Within the active site, specific interactions between the substrate and appropriately charged, hydrophlic and hydrophobic amino acids of the active site then stabilize the transition state by distorting the substrate molecule in such a way as to lead to a transition state having a substantially lower activation energy than can be achieved by ordinay non-enzymatic catalysis. Beyond this point, the basic catalytic steps are fairly conventional, with acid/base and nucleophilic catalysis being the most common.

For a very clear and instructive illustration of a typical multi-step sequence of a typical enzyme-catalyzed reaction, see this page from Mark Bishop's online textbook, from which this illustration is taken.



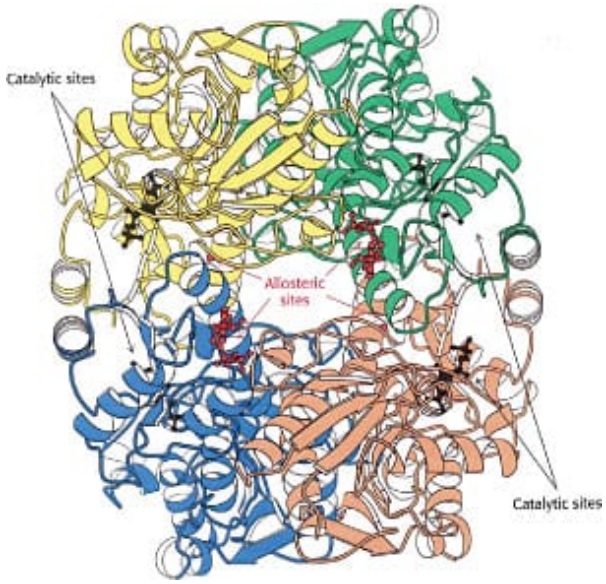
Enzyme regulation and inhibition

If all the enzymes in an organism were active all the time, the result would be runaway chaos. Most cellular processes such as the production and utilization of energy, cell division, and the breakdown of metabolic products must operate in an exquisitely choreographed, finely-tuned manner, much like a large symphony orchestra; no place for jazz-improv here!

Nature has devised various ways of achieving this; we described the action of precursors and coenzymes above. Here we focus on one of the most important (and chemically-interesting) regulatory mechanisms.

Allosteric regulation: tweaking the active site

There is an important class of enzymes that possess special sites (distinct from the catalytically active sites) to which certain external molecules can reversibly bind. Although these *allosteric sites*, as they are called, may be quite far removed from the catalytic sites, the effect of binding or release of these molecules is to trigger a rapid change in the folding pattern of the enzyme that alters the shape of the active site. The effect is to enable a signalling or regulatory molecule (often a very small one such as NO) to modulate the catalytic activity of the active site, effectively turning the enzyme on or off.



In some instances, the product of an enzyme-catalyzed reaction can itself bind to an allosteric site, decreasing the activity of the enzyme and thus providing negative feedback that helps keep the product at the desired concentration. It is believed that concentrations of plasma ions such as calcium, and of energy-supplying ATP are, are regulated in this way.

Allosteric enzymes are more than catalysts: they act as control points in metabolic and cellular signalling networks.Allosteric enzymes frequently stand at the beginning of a sequence of enzymes in a metabolic chain, or at branch points where two such chains diverge, acting very much like traffic signals at congested intersections.

Enzyme inhibition

As is the case with heterogeneous catalysts, certain molecules other than the normal substrate may be able to enter and be retained in the active site so as to competitively inhibit an enzyme's activity. This is how penicillin and related antibiotics work; these molecules covalently bond to amino acid residues in the active site of an enzyme that catalyzes the formation of an essential component of bacterial cell walls. When the cell divides, the newly-formed progeny essentially fall apart.

Enzymes outside the cell



Artist Mike Perkins' conception of an immobilized enzyme within a functionalized nanoporous silica pore. The enzyme is shown as green with positively charged regions shown in red. The blue structures inside the pores represent the negatively charged functional groups added to the mesoporous silica to create a favorable chemical environment for the enzyme. The favorable environment in each pore attracts the enzyme molecule to move into the unoccupied pore. This environment stabilizes and increases the chemical reactivity of the enzyme, allowing it to convert harmful substrate materials (purple particles) into useful or harmless products (yellow and red particles). [link]

Enzymes have been widely employed in the food, pulp-and-paper, and detergent industries for a very long time, but mostly as impure whole-cell extracts.

In recent years, developments in biotechnology and the gradual move of industry from reliance on petroleum-based feedstocks and solvents to so-called "green" chemistry have made enzymes more attractive as industrial catalysts. Compared to the latter, purified enzymes tend to be expensive, difficult to recycle, and unstable outside of rather narrow ranges of temperature, pH, and solvent composition.

Immobilized enzymes

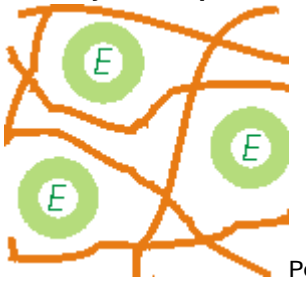
Many of the problems connected with the use of free enzymes can be overcome by immobilizing the enzyme. This can be accomplished in several ways:

Binding the enzyme to an inert solid supporting material



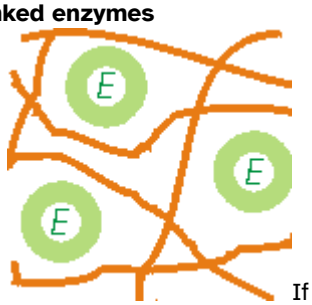
This was first done in 1916, using charcoal or aluminum hydroxide. since the 1960's when this method became more popular, synthetic polymers, often engineered for a specific application, have come into wide use. Some natural biopolymers such as cellulose or starch, and inorganic solids such as silica and alumina have also been employed.

Trapping the enzyme in a porous material



Polymerization of acrylamide in the presence of an enzyme yields a porous gel which alllows free diffusion of substrate and products, Other materils used for this purpose include natural gels such as alginate (obtained from seaweed), porous silica or ceramic solids, and zeolites.

Cross-linked enzymes



If the enzyme is chemically linked to another substance (usually a very small molecule), the resulting solid, known as a cross-linked enzyme aggregate, can retain high catalytic efficiency over a very wide range of conditions. A more recent related development involves precipitating a less-purified form of the enzyme (by addition of ethylene glycol or ammonium sulfate, for example) and then cross-linking the resulting aggregates.

[images↑]
References: more to explore

Catalysis

1. Gadi Rothenberg- *Catalysis: Concepts and Green Applications* (Wiley-VCH, 2008)
2. R.A. van Santan, M. Neurock - *Molecular Heterogeneous Catalysis: A conceptual and computational approach* (Wiley-VCH, 2006)
3. History of Catalysis: List of Web pages, including Fifty Years of Catalysis - A list of major advances in the field 1949-1999. More detailed History pages from UK
4. Catalysis and its applications - a very readable and interesting account of the applications of catalysis to various industrially-important fields.

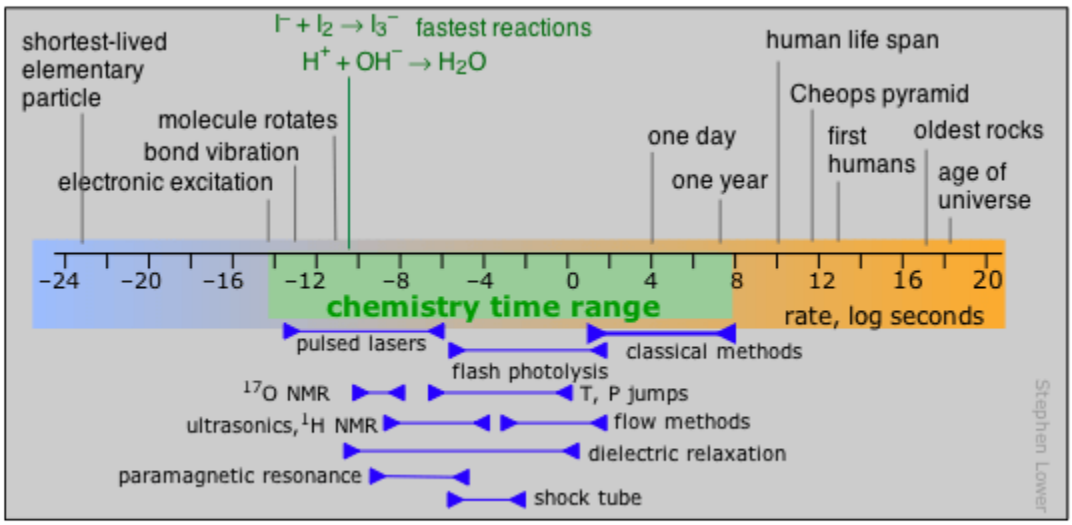
This page titled 17.6: Catalysts and Catalysis is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

17.7: Experimental methods of chemical kinetics

Studies of the dynamics of chemical processes impinge on almost every area of chemistry and biochemistry. It is useful for students even at the general chemistry level to have some understanding of the experimental techniques that have informed what you have already learned about kinetics.

The time domain of chemical kinetics

Kinetics deals with rates of change, so as a preface to this section, it is useful to consider the range of times within which the study of chemical processes takes place. Prior to about 1945, the practical range for kinetic investigations began at about 10^3 sec and extended out to around 10^8 sec, this upper limit being governed by the length of time that a graduate student (who usually does the actual work) can expect to get a Ph.D. degree. The development of flash photolysis and flow methods in the mid-1940s moved the lower time limit down to the millisecond range. The most dramatic changes began to occur in the 1970s as the availablity of lasers capable of very short-lived pulses and fast electronics gradually opened up the study of reactions that are completed in nanoseconds, picoseconds, and even femtoseconds (10^{-15} sec).



The basic way of obtaining the information needed to determine rate constants and reaction orders is to bring the reactants together and then measure successive changes in concentration of one of the components as a function of time. Two important requirements are

- The time required to take a measurement must be very short compared to the time the reaction takes to run to completion;
- The temperature must be held constant — something than can pose a problem if the reaction is highly exothermic.

Measuring the concentration of a reactant or product directly — that is by chemical analysis, is awkward and seldom necessary. When it cannot be avoided, the reaction sample must usually be *quenched* in some way in order to stop any further change until its composition can be analyzed. This may be accomplished in various ways, depending on the particular reaction. For reactions carried out in solution, especially enzyme-catalyzed ones, it is sometimes practical to add a known quantity of acid or base to change the pH, or to add some other inhibitory agent. More commonly, however, the preferred approach is to observe some physical property whose magnitude is *proportional* to the extent of the reaction.

Classical and laboratory-class methods

These methods are applicable to reactions that are not excessively fast, typically requiring a few minutes or hours to run to completion. They were about the only methods available before 1950. Most first-year laboratory courses students will include at least one experiment based on one of these methods.

Optical methods

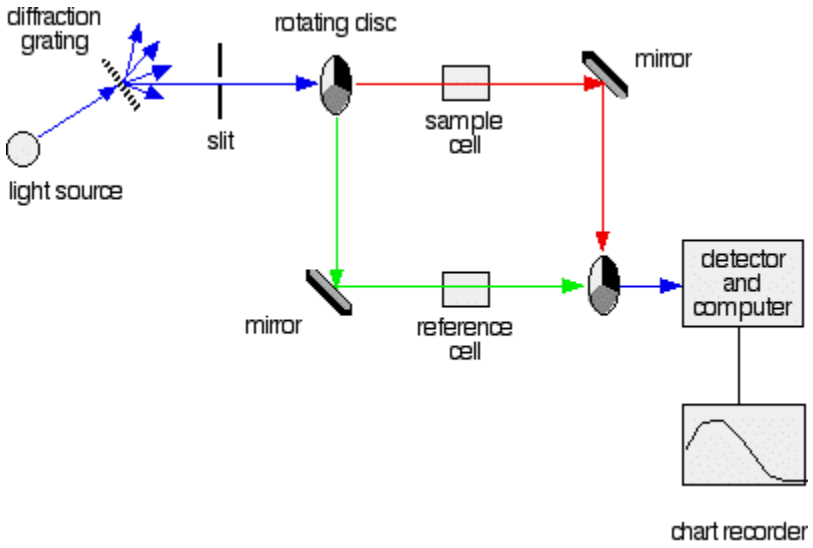
Light absorption is perhaps the most widely-used technique. If either a reactant or a product is colored, the reaction is easily followed by recording the change in transmission of an appropriate wavelength after the reaction is started.

When a beam of light passes through a solution containing a colored substance, the fraction that is absorbed is directly proportional to the concentration of that substance and to the length of the light's path through the solution. The latter can be controlled by employing a cell or cuvette having a fixed path length.

If I_0 is the intensity of the light incident on the cell and I is the intensity that emerges on the other side, then the *percent absorption* is just $100 \times I / I_0$. Because a limited 1-100 scale of light absorption is often inadequate to express the many orders of magnitude frequently encountered, a logarithmic term optical density is often employed. The relation between this, the cell path length, concentration, and innate absorption ability of the colored substance is expressed by Beer's law.

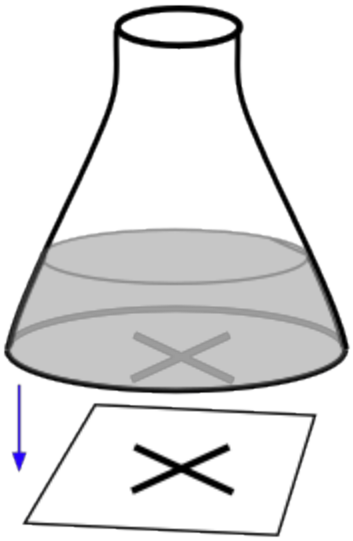
The simplest absorbance measuring device is a *colorimeter* in which a beam of white light from an incandescent lamp is passed through a cell (often just a test tube) and onto a photodetector whose electrical output is directly proportional to the light intensity. Before beginning the experiment, a *zeroing control* sets the meter to zero when the light path is blocked off, and a *sensitivity control* sets it to 100 when a cell containing an uncolored solution (a "blank") is inserted.

The sensitivity and selectivity of such an arrangement is greatly enhanced by adjusting the wavelength of the light to match the absorption spectrum of the substance being measured. Thus if the substance has a yellow color, it is because blue light is being absorbed, so a blue color filter is placed in the light path.



More sophisticated *absorption spectrophotometers* employ two cells (one for the sample and another *reference cell* for the blank. They also allow one to select the particular wavelength range that is most strongly absorbed by the substance under investigation, which can often extend into the near-ultraviolet region.

Light scattering measurements (made with a *nephthelometer*) can be useful for reactions that lead to the formation a fine precipitate/ A very simple student laboratory experiment of this kind can be carried out by placing a conical flask containing the reaction mixture on top of a marked piece of paper. The effects of changing the temperature or reactant concentrations can be made by observing how long it takes for precipitate formation to obscure the mark on the paper.



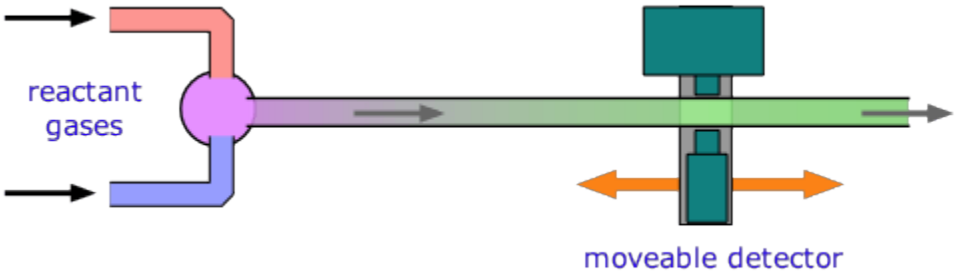
Other optical methods such a fluorescence and polarimetry (measurement of the degree to which a solution rotates the plane of polarized light) are also employed when applicable.

- **Measurement of gases**: The classical method of following reactions that produce changes in the number of moles of gases is to observe changes in pressure or volume. An alternative method for following the loss of a gas is to place the reaction container on an electronic balance and monitor the loss in weight. This is generally less accurate than pressure measurements, but is sometimes used in student experiments.
- **pH measurements**: Many reactions yield or consume hydrogen ions and are conveniently followed by means of a pH meter.
- **Electrical Conductance:** Reactions that yield or consume ionic substances are often studied by measuring the electrical conductance of the solution. *Conductimetry* is usually carried out by balancing the solution conductance with a known resistance in a bridge arrangement. An audio-frequency alternating current is used in order to avoid electrolysis

The study of rapid reactions

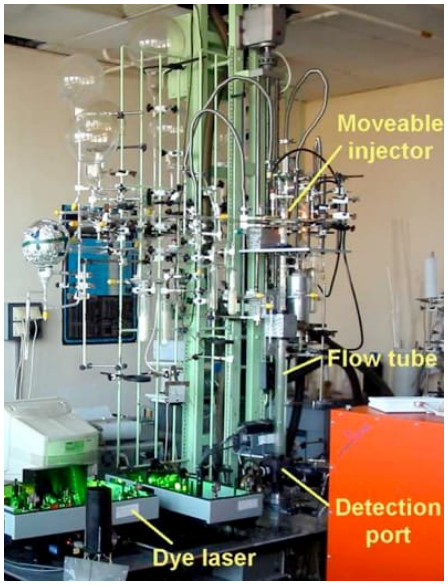
Rapid mixing

The traditional experimental methods described above all assume that we can follow the reaction after its components have combined into a homogeneous mixture of known concentrations. But what can we do if the time required to complete the mixing process is comparable to or greater than the time needed for the reaction to run to completion? For reactions that take place in milliseconds, the standard approach since the 1950s has been to employ a flow technique of some kind. An early example was used to study fast gas-phase reactions in which one of the reactants is a free radical such as OH that can be produced by an intense microwave discharge acting on a suitable source gas mixture. This gas, along with the other reactant being investigated, is made to flow through a narrow tube at a known velocity.



If the distance between the point at which the reaction is initiated and the product detector is known, then the time interval can

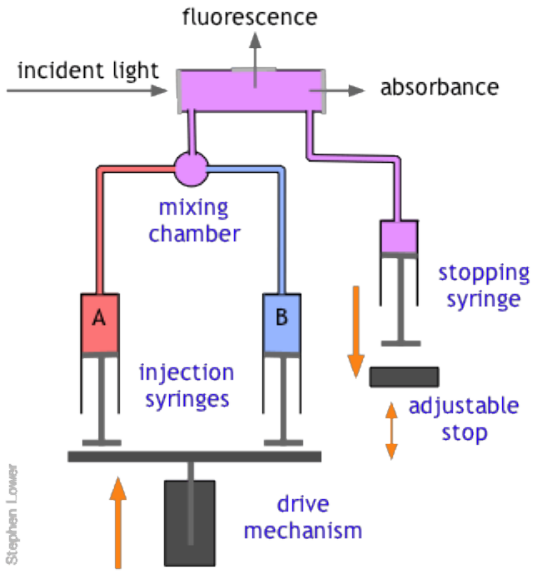
be found from the flow rate. By varying this distance, the time required to obtain the maximum yield can then be determined.



Although this method is very simple in principle, it can be rather complicated in practice, as the illustration shows. Owing to the rather large volumes required, his method is more practical for the study of gas-phase reactions than for solutions, for which the stopped-flow method described below is generally preferred.

Stopped-flow and Quenched-flow methods

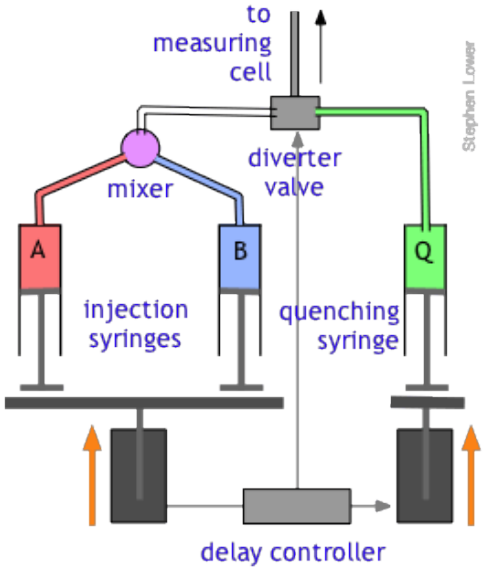
These are by far the most common means of studying fast solution-phase reactions over time intervals of down to a fraction of a millisecond. The use of reasonably simple devices is now practical even in student laboratory experiments. These techniques make it possible to follow not only changes in the concentrations of reactants and products, but also the buildup and decay of reaction intermediates.





The basic stopped-flow apparatus consists of two or more coupled syringes that rapidly inject the reactants into a small mixing chamber and then through an observation cell that can be coupled to instruments that measure absorption, fluorescence, light scattering, or other optical or electrical properties of the solution. As the solution flows through the cell, it empties into a stopping syringe that, when filled, strikes a backstop that abruptly stops the flow. The volume that the stopping syringe can accept is adjusted so that the mixture in the cell has just become uniform and has reached a steady state; at this point, recording of the cell measurement begins and its change is followed.

Of course, there are many reactions that cannot be followed by changes in light absorption or other physical properties that are conveniently monitored. In such cases, it is often practical to *quench* (stop) the reaction after a desired interval by adding an appropriate quenching agent. For example, an enzyme-catalyzed reaction can be stopped by adding an acid, base, or salt solution that denatures (destroys the actvity of) the protein enzyme. Once the reaction has been stopped, the mixture is withdrawn and analyzed in an appropriate manner.



The quenched-flow technique works something like the stopped-flow method described above, with a slightly altered plumbing arrangement. The reactants A and B are mixed and fed directly through the diverter valve to the measuring cell, which is not shown in this diagram. After a set interval that can vary from a few milliseconds to 200 sec or more, the controller activates the quenching syringe and diverter valve, flooding the cell with the quenching solution.

Too fast to mix? No problem!

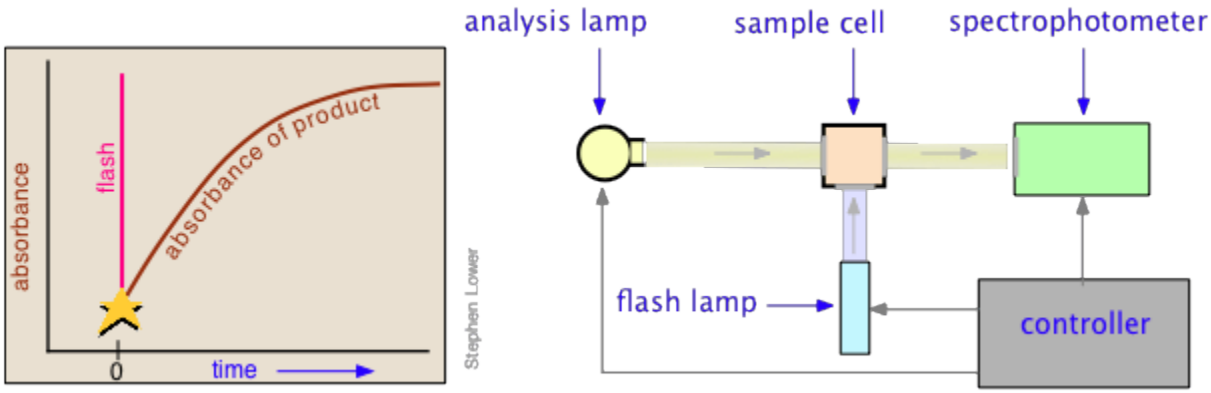
In order to investigate reactions that are complete in less than a millisecond, one can start with a pre-mixed sample in which one of active reactants is generated *in situ*. Alternatively, a rapid change in pressure or temperature can alter the composition of a reaction that has already achieved equilibrium.

Flash photolysis

Many reactions are known which do not take place in the absence of light whose wavelength is sufficiently short to supply the activation energy needed to break a bond, often leading to the creation of a highly reactive radical. A good example is the combination of gaseous Cl_2 with H_2, which procedes explosively when the system is illuminated with visible light. In *flash photolysis*, a short pulse of light is used to initiate a reaction whose progress can be observed by optical or other means.

Photolysis refers to the use of light to decompose a molecule into simpler units, often ions or free radicals. In contrast to *thermolysis* (decomposition induced by high temperature), photolysis is able to inject energy into a molecule almost instantaneously and can be much "cleaner", meaning that there are fewer side reactions that often lead to complex mixtures of products. Photolysis can also be highly *specific*; the wavelength of the light that triggers the reaction can often be adjusted to activate one particular kind of molecule without affecting others that might be present.

All this had been known for a very long time, but until the mid-1940's there was no practical way of studying the kinetics of the reactions involving the highly reactive species producd by photolysis. In 1945, Ronald Norrish of Cambridge University and his graduate student George Porter conceived the idea of using a short-duration flash lamp to generate gas-phase CH_2 radicals, and then following the progress of the reaction of these radicals with other species by means of absorption spectroscopy.



Norrish and Porter shared the 1967 Nobel Prize in Chemistry for this work.Nanosecond flash photolysis setup.

In a flash photolysis experiment, recording of the absorbance of the sample cell contents is timed to follow the flash by an interval that can be varied in order to capture the effects produced by the product or intermediate as it is formed or decays. Flash durations of around 1 millisecond permitted one to follow processes having lifetimes in the microsecond range, but the advent of fast lasers gradually extended this to picoseconds and femtoseconds.

Flash photolysis revolutionized the study of organic photochemistry, especially that relating to the chemistry of free radicals and other reactive species that cannot be isolated or stored, but which can easily be produced by photolysis of a suitable precursor. It has proven invaluable for understanding the complicated kinetics relating to atmospheric chemistry and smog formation. More recently, flash photolysis has become an important tool in biochemistry and cellular physiology.

Perturbation-relaxation methods

Many reactions, especially those that take place in solution, occur too rapidly to follow by flow techniques, and can therefore only be observed when they are already at equilibrium. The classical examples of such reactions are two of the fastest ones ever observed, the dissociation of water

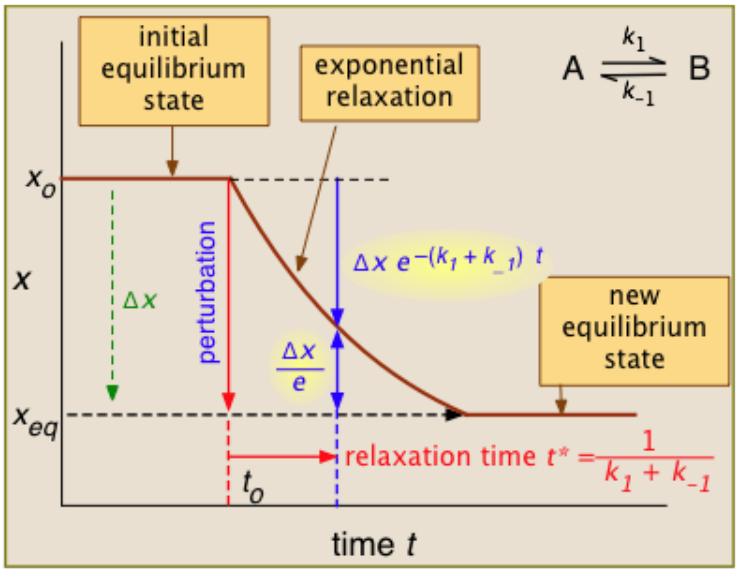
$$2H_2O \rightarrow H_3O^+ + OH^- \tag{17.7.1}$$

and the formation of the triiodide ion in aqueous solution

$$I^- + I_2 \rightarrow I_3^- \tag{17.7.2}$$

Reactions of these kinds could not be studied until the mid-1950s when techniques were developed to shift the equilibrium by imposing an abrupt physical change on the system.

For example, if the reaction A $\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}}$ B is endothermic, then according to the Le Chatelier principle, subjecting the system to a rapid jump in temperature will shift the equilibrium state to one in which the product B has a higher concentration. The composition of the system will than begin to shift toward the new equilibrium composition at a rate determined by the kinetics of the process.



For the general case illustrated here, the quantity "x" being plotted is a measurable quantity such as light absorption or electrical conductivity that varies linearly with the composition of the system. In a first-order process, x will vary with time according to

$x_t = x_o\, e^{-kt}$

After the abrupt perturbation at time t_o, the relaxation time t^* is defined as the half-time for the return to equilibrium — that is, as the time required for xo to decrease by $\Delta x/e = \Delta x/2.718$. The derivation of t^* and the relations highlighted in yellow can be found in most standard kinetics textbooks. Temperature jumps are probably most commonly used. They can be brought about in various ways:

This is the method that Manfred Eigen pioneered when, in the early 1960's, he measured the rate constant of what was then the fastest reaction ever observed:

$H^+ + OH^- \rightarrow H_2O$ $k = 1.3 \times 10^{11}\ M^{-1}\ sec^{-1}$

A detailed description of the kinetics of this process can be seen here. Eigen shared the 1967 Nobel Prize in Chemistry with Porter and Norrish, who developed flash photolysis. Eigen's Nobel lecture, in keeping with his legendary sense of humor, was titled "immeasurably fast reactions". His later work has focused on self-organizing systems and the origin of life, molecular genetics, and neurology.

- **high-voltage electric discharge**: A capacitor, charged to 5-10 kV, is discharged through a solution to which an electrolyte has been added to provide a conductive path.
- **laser irradiation**: The sample is irradiated with a laser whose wavelength corresponds to an absorption peak in the sample. Infrared lasers are often used for this purpose.

- **mixing of two pre-equilibrated solutions**: Two solutions, otherwise identical but at different temperatures, are rapidly mixed in a stopped-flow type of apparatus. Although this method is not as fast, it has the advantage of allowing both negative and positive T-jumps. The device shown here uses 0.1-mL samples and provides jumps of up to ±40 C° over a few microseconds. Observation times, however, are limited to 1-2 milliseconds owing to thermal dissipation.

Pressure jumps

According to the Le Chatelier principle, a change in the applied pressure will shift the equilibrium state of any reaction which involves a change in the volume of a system. Aside from the obvious examples associated with changes in the number of moles of gases, there are many more subtle cases involving formation of complexes, hydration shells and surface adsorption, and phase changes. One area of considerable interest is the study of protein folding, which has implications in diseases such as Parkinson's and Alzheimer's.

The pressure-jump is applied to the cell through a flexible membrane that is activated by a high-pressure gas supply, or through an electrically-actuated piezoelectric crystal. The latter method is employed in the device shown here, which can produce P-jumps of around 1GPa over sub-millisecond time intervals.

More on P-jumps: this 1999 review covers many non-biochemical applications. See also this article on protein folding and aggregation.

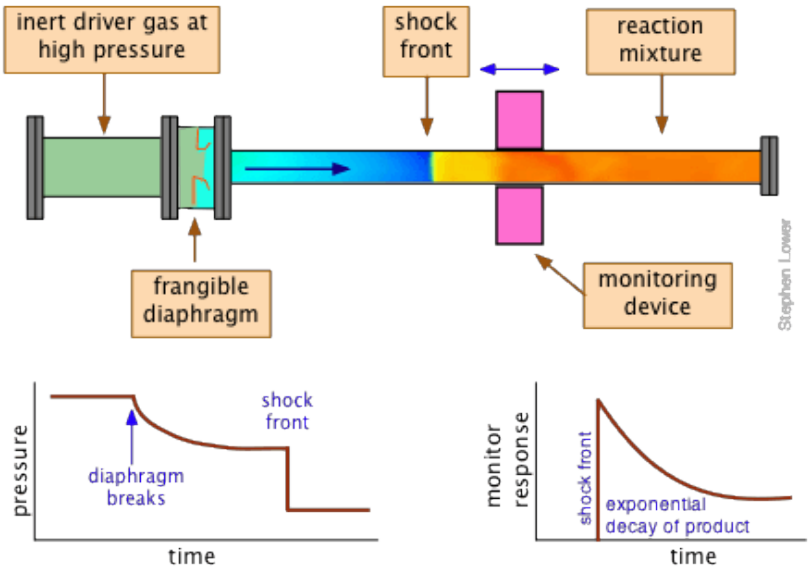
Shock tubes: extreme jumping

When a change in pressure propagates through a gas at a rate greater than the ordinary compressions and rarefactions associated with the travel of sound, a moving front (a *shock wave*) of very high pressure forms. This in turn generates an almost instantaneous rise in the temperature that can approach several thousand degrees in magnitude.

A shock tube is an apparatus in which shock waves can be generated and used to study the kinetics of gas-phase reactions that are otherwise inaccessible to kinetic measurements. Since all molecules tend to dissociate at high temperatures, shock tubes are widely used to study dissociation processes and the chemistry of the resulting fragments. For example, the shock-induced decomposition of carbon suboxide provides an efficient means of investigating carbon atom reactions:

$$C_3O_2 \rightarrow C + CO \tag{17.7.3}$$

Shock tube techniques are also useful for studying combustion reactions, including those that proceed explosively.



The shock tube itself consists of two sections separated by a breakable diaphragm of metal or plastic. One section is filled with a "driver" gas at a very high pressure, commonly helium, but often mixed with other inert gases to adjust the properties of the shock. The other, longer section of the tube contains the "driven" gas — the reactants — at a low pressure, usually less than 1

atmosphere.

The reaction is initiated by causing the diaphragm to rupture, either by means of a mechanical plunger or by raising the pressure beyond its bursting point. The kinetics of the reaction are monitored by means of an absorption or other optical monitoring device that is positioned at a location along the reaction tube that is appropriate to the time course of the reaction.

This page titled 17.7: Experimental methods of chemical kinetics is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.

INDEX

A

Activation controlled
17.5: Kinetics of Reactions in Solution
activity
13.2: Strong Monoprotic Acids and Bases
activity coefficient
13.2: Strong Monoprotic Acids and Bases
allotropes
4.3: Formulas and Their Meaning
anhydrous
9.9: Bonding in Coordination Complexes
antibonding
9.8: Molecular Orbital Theory
autonumheader:yes2
9.8: Molecular Orbital Theory
Avogadro's constant
4.2: Avogadro's Number and the Mole
avogadro's law
6.2: Ideal Gas Model - The Basic Gas Laws
Avogadro's number
4.2: Avogadro's Number and the Mole
azeotropes
8.9: Distillation

B

band gap
9.11: Bonding in Semiconductors
barometer
2.3: The Measure of Matter
3.1: Units and Dimensions
battery
16.6: Batteries and Fuel Cells
biomagnification
11.6: Phase Distribution Equilibria
body centered cubic
7.8: Cubic Lattices and Close Packing
boiling point
7.5: Changes of State
bomb calorimeter
14.5: Calorimetry
bond dissociation energy
14.6: Applications of Thermochemistry
bond enthalpy
14.6: Applications of Thermochemistry
bond order
9.8: Molecular Orbital Theory
bonding molecular orbital
9.8: Molecular Orbital Theory
Boyle temperature
6.6: Real Gases and Critical Phenomena
Boyle's law
6.2: Ideal Gas Model - The Basic Gas Laws
Brownian motion
17.5: Kinetics of Reactions in Solution

C

calorimetry
14.4: Thermochemistry and Calorimetry
14.5: Calorimetry
catalysis
17.6: Catalysts and Catalysis
catalysts
17.6: Catalysts and Catalysis
cathodic protection
16.8: Electrochemical Corrosion
cell potential
16.3: Cell Potentials and Thermodynamics
chain reaction
17.4: Reaction Mechanisms
Charles' Law
6.2: Ideal Gas Model - The Basic Gas Laws
chelate
9.9: Bonding in Coordination Complexes
chemical energy
2.2: Energy, Heat, and Temperature
chemisorption
17.6: Catalysts and Catalysis
chromatography
2.1: Classification and Properties of Matter
closed packing
7.8: Cubic Lattices and Close Packing
colligative properties
15.7: Some Applications of Entropy and Free Energy
colloid
7.10: Colloids and their Uses
conduction band
9.11: Bonding in Semiconductors
conjugated double bonds
9.7: The Hybrid Orbital Model II
conservation of energy
14.2: The First Law of Thermodynamics
coordination compound
9.9: Bonding in Coordination Complexes
coordination number
9.5: Molecular Geometry
corrosion
16.8: Electrochemical Corrosion
corrosion resistance
16.8: Electrochemical Corrosion
coupled reactions
15.6: Free Energy and Equilibrium
covalent bond
9.1: Three Views of Chemical Bonding
critical point
7.5: Changes of State
crystallization
2.1: Classification and Properties of Matter
cubic lattices
7.8: Cubic Lattices and Close Packing
cumulene
9.7: The Hybrid Orbital Model II

D

Dalton's Law of Partial Pressure
6.3: Dalton's Law
Daniell cell
16.7: Timeline of Battery Development
dative bond
9.4: Polar Covalence
degree of dissociation
13.4: Conjugate Pairs and Buffers
derived units
3.1: Units and Dimensions
diamagnetic
9.8: Molecular Orbital Theory
Diffusion Controlled
17.5: Kinetics of Reactions in Solution
dipole moment
9.4: Polar Covalence
distillation
2.1: Classification and Properties of Matter
8.9: Distillation
distribution coefficient
11.6: Phase Distribution Equilibria
distribution ratio
11.6: Phase Distribution Equilibria

E

E.V.E.N. principle
6.2: Ideal Gas Model - The Basic Gas Laws
electrolysis
16.10: Electrolytic Cells and Electrolysis
electrolytic cell
16.10: Electrolytic Cells and Electrolysis
electronegativity
9.4: Polar Covalence
elementary reaction
17.4: Reaction Mechanisms
empirical formula
4.3: Formulas and Their Meaning
encounter pair
17.5: Kinetics of Reactions in Solution
enthalpy of vaporization
14.4: Thermochemistry and Calorimetry
entropy of mixing
15.5: Thermodynamics of Mixing and Dilution
entropy of solution
8.2: Thermodynamics of Solutions
equilibrium constant
11.3: Reaction Quotient
equivalence point
13.5: Acid/Base Titration
equivalent fraction
13.5: Acid/Base Titration
extensive property
2.1: Classification and Properties of Matter

F

field:genchem
Chem1 (Lower)
first law of thermodynamics
14.2: The First Law of Thermodynamics
formal charge
9.4: Polar Covalence
formula weight
4.3: Formulas and Their Meaning

fractionating column
8.9: Distillation
fuel cell
16.6: Batteries and Fuel Cells

H

hardness
7.7: Ionic and Ion-Derived Solids
heat
14.4: Thermochemistry and Calorimetry
heat capacity
2.2: Energy, Heat, and Temperature
heat of atomization
14.4: Thermochemistry and Calorimetry
heat of vaporization
14.4: Thermochemistry and Calorimetry
heat transfer
14.1: Energy, Heat and Work
Hess' Law
14.4: Thermochemistry and Calorimetry
Hexagonal Closest Packed
7.8: Cubic Lattices and Close Packing
high spin
9.9: Bonding in Coordination Complexes
hybrid orbital
9.6: The Hybrid Orbital Model
hybridization
9.6: The Hybrid Orbital Model
hydrogen bonding
7.3: Hydrogen-Bonding and Water
hydrophbic effect
8.2: Thermodynamics of Solutions

I

ice calorimeter
14.5: Calorimetry
ideal gas law
6.2: Ideal Gas Model - The Basic Gas Laws
indicators
13.5: Acid/Base Titration
insulator
9.11: Bonding in Semiconductors
intensive property
2.1: Classification and Properties of Matter
intermolecular forces
7.2: Intermolecular Interactions
interstitial sites
7.8: Cubic Lattices and Close Packing
ion pair
13.2: Strong Monoprotic Acids and Bases
ionic bond
9.1: Three Views of Chemical Bonding
ionic solid
7.7: Ionic and Ion-Derived Solids
ionization fraction
13.4: Conjugate Pairs and Buffers

K

Kelvin
2.2: Energy, Heat, and Temperature
Kelvin: Units
2.2: Energy, Heat, and Temperature

L

Laing tetrahedron
7.1: Matter under the Microscope
Leveling Effect
10.3: Acid-base reactions à la Brønsted
lever rule
8.9: Distillation
ligand
9.9: Bonding in Coordination Complexes
liter (unit)
3.1: Units and Dimensions
low spin
9.9: Bonding in Coordination Complexes
lyophilic colloids
7.10: Colloids and their Uses

M

mean free path
6.5: More on Kinetic Molecular Theory
mechanism
17.4: Reaction Mechanisms
metal
9.1: Three Views of Chemical Bonding
metallic bonds
9.10: Bonding in Metals
metric ton (unit)
3.1: Units and Dimensions
molar mass
4.3: Formulas and Their Meaning
molar volume
4.2: Avogadro's Number and the Mole
molecular orbital
9.8: Molecular Orbital Theory
molecular orbital diagram
9.8: Molecular Orbital Theory
molecular weight
4.3: Formulas and Their Meaning

N

Nernst Equation
16.4: The Nernst Equation
normal melting
7.5: Changes of State

O

octahedral hole
7.8: Cubic Lattices and Close Packing
osmosis
8.5: Colligative Properties - Osmotic Pressure
osmotic pressure
8.5: Colligative Properties - Osmotic Pressure
oxidation numbers
9.4: Polar Covalence

P

paramagnetic
9.8: Molecular Orbital Theory
Pauli exclusion principle
5.5: The Quantum Atom
Pauling's rule
14.6: Applications of Thermochemistry
petroleum refining
8.9: Distillation
phase diagram
7.5: Changes of State
Phase Distribution Equilibria
11.6: Phase Distribution Equilibria
physisorption
17.6: Catalysts and Catalysis
PN Junction
9.11: Bonding in Semiconductors
Polar Covalence
9.4: Polar Covalence
polydentate ligand
9.9: Bonding in Coordination Complexes
Propagation of Error
3.2: The Meaning of Measure
Pseudoscience
1.2: Pseudoscience

R

Rapid Equilibrium Approximation
17.4: Reaction Mechanisms
reaction quotient
11.3: Reaction Quotient
refractories
7.7: Ionic and Ion-Derived Solids
reverse osmosis
8.6: Reverse Osmosis
rounding
3.3: Significant Figures and Rounding off
rust
16.8: Electrochemical Corrosion

S

Sabatier Principle
17.6: Catalysts and Catalysis
Sacrificial Coatings
16.8: Electrochemical Corrosion
semiconductor
9.11: Bonding in Semiconductors
SI units
2.3: The Measure of Matter
3.1: Units and Dimensions
significant figures
3.3: Significant Figures and Rounding off
simple cubic structure
7.8: Cubic Lattices and Close Packing
solvent cage
17.5: Kinetics of Reactions in Solution
solvent kinetic effect
17.5: Kinetics of Reactions in Solution
sp2 hybrid orbital
9.6: The Hybrid Orbital Model
sp3 hybrid orbital
9.6: The Hybrid Orbital Model
sp3d
9.7: The Hybrid Orbital Model II
sp3d2 hybrid orbital
9.7: The Hybrid Orbital Model II
specific heat
2.2: Energy, Heat, and Temperature
spectrochemical series
9.9: Bonding in Coordination Complexes
standard state
14.4: Thermochemistry and Calorimetry

strong acid
13.2: Strong Monoprotic Acids and Bases
strong base
13.2: Strong Monoprotic Acids and Bases
super acid
13.2: Strong Monoprotic Acids and Bases
surface tension
7.3: Hydrogen-Bonding and Water

T

tetrahedral hole
7.8: Cubic Lattices and Close Packing
thermal energy
2.2: Energy, Heat, and Temperature
thermosetting polymers
7.9: Polymers and Plastics
titration
13.5: Acid/Base Titration
titration analysis
13.5: Acid/Base Titration
titration curve
13.5: Acid/Base Titration
triple point
7.5: Changes of State
turnover number
17.6: Catalysts and Catalysis

U

ultramicroscope
7.10: Colloids and their Uses

V

vapor pressure
7.5: Changes of State
VSEPR
9.5: Molecular Geometry

W

water of hydration
9.9: Bonding in Coordination Complexes
work
14.1: Energy, Heat and Work

This page titled Index is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by LibreTexts.

INDEX

A

Activation controlled
17.5: Kinetics of Reactions in Solution
activity
13.2: Strong Monoprotic Acids and Bases
activity coefficient
13.2: Strong Monoprotic Acids and Bases
allotropes
4.3: Formulas and Their Meaning
anhydrous
9.9: Bonding in Coordination Complexes
antibonding
9.8: Molecular Orbital Theory
autonumheader:yes2
9.8: Molecular Orbital Theory
Avogadro's constant
4.2: Avogadro's Number and the Mole
avogadro's law
6.2: Ideal Gas Model - The Basic Gas Laws
Avogadro's number
4.2: Avogadro's Number and the Mole
azeotropes
8.9: Distillation

B

band gap
9.11: Bonding in Semiconductors
barometer
2.3: The Measure of Matter
3.1: Units and Dimensions
battery
16.6: Batteries and Fuel Cells
biomagnification
11.6: Phase Distribution Equilibria
body centered cubic
7.8: Cubic Lattices and Close Packing
boiling point
7.5: Changes of State
bomb calorimeter
14.5: Calorimetry
bond dissociation energy
14.6: Applications of Thermochemistry
bond enthalpy
14.6: Applications of Thermochemistry
bond order
9.8: Molecular Orbital Theory
bonding molecular orbital
9.8: Molecular Orbital Theory
Boyle temperature
6.6: Real Gases and Critical Phenomena
Boyle's law
6.2: Ideal Gas Model - The Basic Gas Laws
Brownian motion
17.5: Kinetics of Reactions in Solution

C

calorimetry
14.4: Thermochemistry and Calorimetry
14.5: Calorimetry
catalysis
17.6: Catalysts and Catalysis
catalysts
17.6: Catalysts and Catalysis
cathodic protection
16.8: Electrochemical Corrosion
cell potential
16.3: Cell Potentials and Thermodynamics
chain reaction
17.4: Reaction Mechanisms
Charles' Law
6.2: Ideal Gas Model - The Basic Gas Laws
chelate
9.9: Bonding in Coordination Complexes
chemical energy
2.2: Energy, Heat, and Temperature
chemisorption
17.6: Catalysts and Catalysis
chromatography
2.1: Classification and Properties of Matter
closed packing
7.8: Cubic Lattices and Close Packing
colligative properties
15.7: Some Applications of Entropy and Free Energy
colloid
7.10: Colloids and their Uses
conduction band
9.11: Bonding in Semiconductors
conjugated double bonds
9.7: The Hybrid Orbital Model II
conservation of energy
14.2: The First Law of Thermodynamics
coordination compound
9.9: Bonding in Coordination Complexes
coordination number
9.5: Molecular Geometry
corrosion
16.8: Electrochemical Corrosion
corrosion resistance
16.8: Electrochemical Corrosion
coupled reactions
15.6: Free Energy and Equilibrium
covalent bond
9.1: Three Views of Chemical Bonding
critical point
7.5: Changes of State
crystallization
2.1: Classification and Properties of Matter
cubic lattices
7.8: Cubic Lattices and Close Packing
cumulene
9.7: The Hybrid Orbital Model II

D

Dalton's Law of Partial Pressure
6.3: Dalton's Law
Daniell cell
16.7: Timeline of Battery Development
dative bond
9.4: Polar Covalence
degree of dissociation
13.4: Conjugate Pairs and Buffers
derived units
3.1: Units and Dimensions
diamagnetic
9.8: Molecular Orbital Theory
Diffusion Controlled
17.5: Kinetics of Reactions in Solution
dipole moment
9.4: Polar Covalence
distillation
2.1: Classification and Properties of Matter
8.9: Distillation
distribution coefficient
11.6: Phase Distribution Equilibria
distribution ratio
11.6: Phase Distribution Equilibria

E

E.V.E.N. principle
6.2: Ideal Gas Model - The Basic Gas Laws
electrolysis
16.10: Electrolytic Cells and Electrolysis
electrolytic cell
16.10: Electrolytic Cells and Electrolysis
electronegativity
9.4: Polar Covalence
elementary reaction
17.4: Reaction Mechanisms
empirical formula
4.3: Formulas and Their Meaning
encounter pair
17.5: Kinetics of Reactions in Solution
enthalpy of vaporization
14.4: Thermochemistry and Calorimetry
entropy of mixing
15.5: Thermodynamics of Mixing and Dilution
entropy of solution
8.2: Thermodynamics of Solutions
equilibrium constant
11.3: Reaction Quotient
equivalence point
13.5: Acid/Base Titration
equivalent fraction
13.5: Acid/Base Titration
extensive property
2.1: Classification and Properties of Matter

F

field:genchem
Chem1 (Lower)
first law of thermodynamics
14.2: The First Law of Thermodynamics
formal charge
9.4: Polar Covalence
formula weight
4.3: Formulas and Their Meaning

fractionating column
8.9: Distillation
fuel cell
16.6: Batteries and Fuel Cells

H

hardness
7.7: Ionic and Ion-Derived Solids
heat
14.4: Thermochemistry and Calorimetry
heat capacity
2.2: Energy, Heat, and Temperature
heat of atomization
14.4: Thermochemistry and Calorimetry
heat of vaporization
14.4: Thermochemistry and Calorimetry
heat transfer
14.1: Energy, Heat and Work
Hess' Law
14.4: Thermochemistry and Calorimetry
Hexagonal Closest Packed
7.8: Cubic Lattices and Close Packing
high spin
9.9: Bonding in Coordination Complexes
hybrid orbital
9.6: The Hybrid Orbital Model
hybridization
9.6: The Hybrid Orbital Model
hydrogen bonding
7.3: Hydrogen-Bonding and Water
hydrophbic effect
8.2: Thermodynamics of Solutions

I

ice calorimeter
14.5: Calorimetry
ideal gas law
6.2: Ideal Gas Model - The Basic Gas Laws
indicators
13.5: Acid/Base Titration
insulator
9.11: Bonding in Semiconductors
intensive property
2.1: Classification and Properties of Matter
intermolecular forces
7.2: Intermolecular Interactions
interstitial sites
7.8: Cubic Lattices and Close Packing
ion pair
13.2: Strong Monoprotic Acids and Bases
ionic bond
9.1: Three Views of Chemical Bonding
ionic solid
7.7: Ionic and Ion-Derived Solids
ionization fraction
13.4: Conjugate Pairs and Buffers

K

Kelvin
2.2: Energy, Heat, and Temperature
Kelvin: Units
2.2: Energy, Heat, and Temperature

L

Laing tetrahedron
7.1: Matter under the Microscope
Leveling Effect
10.3: Acid-base reactions à la Brønsted
lever rule
8.9: Distillation
ligand
9.9: Bonding in Coordination Complexes
liter (unit)
3.1: Units and Dimensions
low spin
9.9: Bonding in Coordination Complexes
lyophilic colloids
7.10: Colloids and their Uses

M

mean free path
6.5: More on Kinetic Molecular Theory
mechanism
17.4: Reaction Mechanisms
metal
9.1: Three Views of Chemical Bonding
metallic bonds
9.10: Bonding in Metals
metric ton (unit)
3.1: Units and Dimensions
molar mass
4.3: Formulas and Their Meaning
molar volume
4.2: Avogadro's Number and the Mole
molecular orbital
9.8: Molecular Orbital Theory
molecular orbital diagram
9.8: Molecular Orbital Theory
molecular weight
4.3: Formulas and Their Meaning

N

Nernst Equation
16.4: The Nernst Equation
normal melting
7.5: Changes of State

O

octahedral hole
7.8: Cubic Lattices and Close Packing
osmosis
8.5: Colligative Properties - Osmotic Pressure
osmotic pressure
8.5: Colligative Properties - Osmotic Pressure
oxidation numbers
9.4: Polar Covalence

P

paramagnetic
9.8: Molecular Orbital Theory
Pauli exclusion principle
5.5: The Quantum Atom
Pauling's rule
14.6: Applications of Thermochemistry
petroleum refining
8.9: Distillation
phase diagram
7.5: Changes of State
Phase Distribution Equilibria
11.6: Phase Distribution Equilibria
physisorption
17.6: Catalysts and Catalysis
PN Junction
9.11: Bonding in Semiconductors
Polar Covalence
9.4: Polar Covalence
polydentate ligand
9.9: Bonding in Coordination Complexes
Propagation of Error
3.2: The Meaning of Measure
Pseudoscience
1.2: Pseudoscience

R

Rapid Equilibrium Approximation
17.4: Reaction Mechanisms
reaction quotient
11.3: Reaction Quotient
refractories
7.7: Ionic and Ion-Derived Solids
reverse osmosis
8.6: Reverse Osmosis
rounding
3.3: Significant Figures and Rounding off
rust
16.8: Electrochemical Corrosion

S

Sabatier Principle
17.6: Catalysts and Catalysis
Sacrificial Coatings
16.8: Electrochemical Corrosion
semiconductor
9.11: Bonding in Semiconductors
SI units
2.3: The Measure of Matter
3.1: Units and Dimensions
significant figures
3.3: Significant Figures and Rounding off
simple cubic structure
7.8: Cubic Lattices and Close Packing
solvent cage
17.5: Kinetics of Reactions in Solution
solvent kinetic effect
17.5: Kinetics of Reactions in Solution
sp2 hybrid orbital
9.6: The Hybrid Orbital Model
sp3 hybrid orbital
9.6: The Hybrid Orbital Model
sp3d
9.7: The Hybrid Orbital Model II
sp3d2 hybrid orbital
9.7: The Hybrid Orbital Model II
specific heat
2.2: Energy, Heat, and Temperature
spectrochemical series
9.9: Bonding in Coordination Complexes
standard state
14.4: Thermochemistry and Calorimetry

strong acid
13.2: Strong Monoprotic Acids and Bases
strong base
13.2: Strong Monoprotic Acids and Bases
super acid
13.2: Strong Monoprotic Acids and Bases
surface tension
7.3: Hydrogen-Bonding and Water

T

tetrahedral hole
7.8: Cubic Lattices and Close Packing
thermal energy
2.2: Energy, Heat, and Temperature
thermosetting polymers
7.9: Polymers and Plastics
titration
13.5: Acid/Base Titration
titration analysis
13.5: Acid/Base Titration
titration curve
13.5: Acid/Base Titration
triple point
7.5: Changes of State
turnover number
17.6: Catalysts and Catalysis

U

ultramicroscope
7.10: Colloids and their Uses

V

vapor pressure
7.5: Changes of State
VSEPR
9.5: Molecular Geometry

W

water of hydration
9.9: Bonding in Coordination Complexes
work
14.1: Energy, Heat and Work

This page titled Index is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by LibreTexts.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by LibreTexts.

DETAILED LICENSING

Overview

Title: Chem1 (Lower)

Webpages: 156

All licenses found:

- CC BY 3.0: 98.1% (153 pages)
- Undeclared: 1.9% (3 pages)

By Page

- Chem1 (Lower) - *CC BY 3.0*
  - Front Matter - *CC BY 3.0*
    - TitlePage - *CC BY 3.0*
    - InfoPage - *CC BY 3.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Fundamentals of Science and Chemistry - *CC BY 3.0*
    - 1.1: Introduction to Chemistry - *CC BY 3.0*
    - 1.2: Pseudoscience - *CC BY 3.0*
  - 2: Essential Background - *CC BY 3.0*
    - 2.1: Classification and Properties of Matter - *CC BY 3.0*
    - 2.2: Energy, Heat, and Temperature - *CC BY 3.0*
    - 2.3: The Measure of Matter - *CC BY 3.0*
    - 2.4: The Meaning of Measure - *CC BY 3.0*
  - 3: Measuring Matter - *CC BY 3.0*
    - 3.1: Units and Dimensions - *CC BY 3.0*
    - 3.2: The Meaning of Measure - *CC BY 3.0*
    - 3.3: Significant Figures and Rounding off - *CC BY 3.0*
    - 3.4: Reliability of a measurement - *CC BY 3.0*
    - 3.5: Drawing Conclusions from Data - *CC BY 3.0*
  - 4: The Basics of Chemistry - *CC BY 3.0*
    - 4.1: Atoms, Elements, and the Nucleus - *CC BY 3.0*
    - 4.2: Avogadro's Number and the Mole - *CC BY 3.0*
    - 4.3: Formulas and Their Meaning - *CC BY 3.0*
    - 4.4: Chemical Equations and Stoichiometry - *CC BY 3.0*
    - 4.5: Introduction to Chemical Nomenclature - *CC BY 3.0*
    - 4.6: Significant Figures and Rounding - *CC BY 3.0*
  - 5: Atoms and the Periodic Table - *CC BY 3.0*
    - 5.1: Primer on Quantum Theory - *CC BY 3.0*
    - 5.2: Quanta - A New View of the World - *CC BY 3.0*
    - 5.3: Light, Particles, and Waves - *CC BY 3.0*
    - 5.4: The Bohr Atom - *CC BY 3.0*
    - 5.5: The Quantum Atom - *CC BY 3.0*
    - 5.6: Atomic Electron Configurations - *CC BY 3.0*
    - 5.7: Periodic Properties of the Elements - *CC BY 3.0*
    - 5.8: Why Don't Electrons Fall into the Nucleus? - *CC BY 3.0*
  - 6: Properties of Gases - *CC BY 3.0*
    - 6.1: Observable Properties of Gas - *CC BY 3.0*
    - 6.2: Ideal Gas Model - The Basic Gas Laws - *CC BY 3.0*
    - 6.3: Dalton's Law - *CC BY 3.0*
    - 6.4: Kinetic Molecular Theory (Overview) - *CC BY 3.0*
    - 6.5: More on Kinetic Molecular Theory - *CC BY 3.0*
    - 6.6: Real Gases and Critical Phenomena - *CC BY 3.0*
  - 7: Solids and Liquids - *CC BY 3.0*
    - Front Matter - *CC BY 3.0*
      - TitlePage - *CC BY 3.0*
      - InfoPage - *CC BY 3.0*
    - 7.1: Matter under the Microscope - *CC BY 3.0*
    - 7.2: Intermolecular Interactions - *CC BY 3.0*
    - 7.3: Hydrogen-Bonding and Water - *CC BY 3.0*
    - 7.4: Liquids and their Interfaces - *CC BY 3.0*
    - 7.5: Changes of State - *CC BY 3.0*
    - 7.6: Introduction to Crystals - *CC BY 3.0*
    - 7.7: Ionic and Ion-Derived Solids - *CC BY 3.0*
    - 7.8: Cubic Lattices and Close Packing - *CC BY 3.0*
    - 7.9: Polymers and Plastics - *CC BY 3.0*
    - 7.10: Colloids and their Uses - *CC BY 3.0*
    - Back Matter - *CC BY 3.0*
      - Index - *CC BY 3.0*
  - 8: Solutions - *CC BY 3.0*
    - 8.1: Solutions and their Concentrations - *CC BY 3.0*
    - 8.2: Thermodynamics of Solutions - *CC BY 3.0*
      - 8.2.2A: Solutions of Gaseous Solutes in

Gaseous Solvents - *CC BY 3.0*
- 8.2.2B: Solutions of Gaseous Solutes in Liquid Solvents - *CC BY 3.0*
- 8.2.2C: Solutions of Liquid Solutes in Liquid Solvents - *CC BY 3.0*
- 8.2.2D: Solutions of Solid Solutes in Liquid Solvents - *CC BY 3.0*

- 8.3: Colligative Properties- Raoult's Law - *CC BY 3.0*
- 8.4: Colligative Properties- Boiling Point Elevation and Freezing Point Depression - *CC BY 3.0*
- 8.5: Colligative Properties - Osmotic Pressure - *CC BY 3.0*
- 8.6: Reverse Osmosis - *CC BY 3.0*
- 8.7: Colligative Properties and Entropy - *CC BY 3.0*
- 8.8: Ideal vs. Real Solutions - *CC BY 3.0*
- 8.9: Distillation - *CC BY 3.0*
- 8.10: Ions and Electrolytes - *CC BY 3.0*
  - 8.10.9A: Electrolytes and Electrolytic Solutions - *CC BY 3.0*
  - 8.10.9B: The nature of ions in aqueous solution - *CC BY 3.0*
  - 8.10.9C: Weak and Strong Electrolytes - *CC BY 3.0*
  - 8.10.9D: Ionic migration - *CC BY 3.0*
  - 8.10.9E: Some applications of electrolytic conduction - *CC BY 3.0*

◦ 9: Chemical Bonding and Molecular Structure - *CC BY 3.0*
- 9.1: Three Views of Chemical Bonding - *CC BY 3.0*
- 9.2: Molecules - Properties of Bonded Atoms - *CC BY 3.0*
- 9.3: Models of Chemical Bonding - *CC BY 3.0*
- 9.4: Polar Covalence - *CC BY 3.0*
- 9.5: Molecular Geometry - *CC BY 3.0*
- 9.6: The Hybrid Orbital Model - *CC BY 3.0*
- 9.7: The Hybrid Orbital Model II - *CC BY 3.0*
- 9.8: Molecular Orbital Theory - *CC BY 3.0*
- 9.9: Bonding in Coordination Complexes - *CC BY 3.0*
- 9.10: Bonding in Metals - *CC BY 3.0*
- 9.11: Bonding in Semiconductors - *CC BY 3.0*
- 9.12: The Shared-Electron Covalent Bond - *CC BY 3.0*

◦ 10: Fundamentals of Acids and Bases - *CC BY 3.0*
- 10.1: Introduction to Acids and Bases - *CC BY 3.0*
- 10.2: Aqueous Solutions- pH and Titrations - *CC BY 3.0*
- 10.3: Acid-base reactions à la Brønsted - *CC BY 3.0*
- 10.4: Acid-Base Reactions - *CC BY 3.0*
- 10.5: Lewis Acids and Bases - *CC BY 3.0*
- 10.6: Types of Acids and Bases - *CC BY 3.0*
- 10.7: Acid-Base Gallery - *CC BY 3.0*

◦ 11: Chemical Equilibrium - *CC BY 3.0*
- 11.1: Introduction to Chemical Equilibrium - *CC BY 3.0*
- 11.2: Le Chatelier's Principle - *CC BY 3.0*
- 11.3: Reaction Quotient - *CC BY 3.0*
- 11.4: Equilibrium Expressions - *CC BY 3.0*
- 11.5: Equilibrium Calculations - *CC BY 3.0*
- 11.6: Phase Distribution Equilibria - *CC BY 3.0*

◦ 12: Solubility Equilibria - *CC BY 3.0*

◦ 13: Acid-Base Equilibria - *CC BY 3.0*
- 13.1: Introduction to Acid/Base Equilibria - *CC BY 3.0*
- 13.2: Strong Monoprotic Acids and Bases - *CC BY 3.0*
- 13.3: Finding the pH of weak Acids, Bases, and Salts - *CC BY 3.0*
- 13.4: Conjugate Pairs and Buffers - *CC BY 3.0*
- 13.5: Acid/Base Titration - *CC BY 3.0*
- 13.6: Applications of Acid-Base Equilibria - *CC BY 3.0*
- 13.7: Exact Calculations and Approximations - *CC BY 3.0*

◦ 14: Thermochemistry - *CC BY 3.0*
- 14.E: Thermochemistry (Exercises) - *CC BY 3.0*
- 14.1: Energy, Heat and Work - *CC BY 3.0*
- 14.2: The First Law of Thermodynamics - *CC BY 3.0*
- 14.3: Molecules as Energy Carriers and Converters - *CC BY 3.0*
- 14.4: Thermochemistry and Calorimetry - *CC BY 3.0*
- 14.5: Calorimetry - *CC BY 3.0*
- 14.6: Applications of Thermochemistry - *CC BY 3.0*

◦ 15: Thermodynamics of Chemical Equilibria - *CC BY 3.0*
- 15.1: Energy Spreading Drives Spontaneous Change - *CC BY 3.0*
- 15.2: Entropy Rules - *CC BY 3.0*
- 15.3: The Second Law of Thermodynamics - *CC BY 3.0*
- 15.4: Free Energy and the Gibbs Function - *CC BY 3.0*
- 15.5: Thermodynamics of Mixing and Dilution - *CC BY 3.0*
- 15.6: Free Energy and Equilibrium - *CC BY 3.0*
- 15.7: Some Applications of Entropy and Free Energy - *CC BY 3.0*
- 15.8: Quantum states, Microstates, and Energy spreading in Reactions - *CC BY 3.0*

◦ 16: Electrochemistry - *CC BY 3.0*
- 16.1: Chemistry and Electricity - *CC BY 3.0*
- 16.2: Galvanic cells and Electrodes - *CC BY 3.0*
- 16.3: Cell Potentials and Thermodynamics - *CC BY 3.0*
- 16.4: The Nernst Equation - *CC BY 3.0*
- 16.5: Applications of the Nernst Equation - *CC BY 3.0*

- 16.6: Batteries and Fuel Cells - *CC BY 3.0*
- 16.7: Timeline of Battery Development - *CC BY 3.0*
- 16.8: Electrochemical Corrosion - *CC BY 3.0*
- 16.9: Corrosion Gallery - *CC BY 3.0*
- 16.10: Electrolytic Cells and Electrolysis - *CC BY 3.0*

◦ 17: Chemical Kinetics and Dynamics - *CC BY 3.0*
- 17.1: Rates of reactions and rate laws - *CC BY 3.0*
- 17.2: Reaction Rates Typically Change with Time - *CC BY 3.0*
- 17.3: Collision and activation- the Arrhenius law - *CC BY 3.0*
- 17.4: Reaction Mechanisms - *CC BY 3.0*
- 17.5: Kinetics of Reactions in Solution - *CC BY 3.0*
- 17.6: Catalysts and Catalysis - *CC BY 3.0*
- 17.7: Experimental methods of chemical kinetics - *CC BY 3.0*

◦ Back Matter - *CC BY 3.0*
- Index - *CC BY 3.0*
- Glossary - *CC BY 3.0*
- Detailed Licensing - *Undeclared*

This page titled Chem1 (Lower) is shared under a CC BY 3.0 license and was authored, remixed, and/or curated by Stephen Lower via source content that was edited to the style and standards of the LibreTexts platform.