ENERGY ALTERNATIVES



Tom Giebultowicz



Energy Alternatives

Tom Giebultowicz

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.

This text was compiled on 05/13/2026

TABLE OF CONTENTS

Licensing

1: Energy Alternatives { what's the meaning?}

- 1.1: Brief semantic considerations
- 1.2: Things to be chosen" What things, what choice?
- 1.3: Back to the definitions- which one? 2a or 2b?

2: General Remarks

- 2.1: Energy and power
  - 2.1.1: Energy Conservation
  - 2.1.2: Energy and Work
  - 2.1.3: Energy Units
  - 2.1.4: Power
  - 2.1.5: Energy expressed in terms of power
- 2.2: Forms of Energy- Physicist's View .
  - 2.2.1: Potential Energy
  - 2.2.2: Kinetic Energy
  - 2.2.3: Energy of an Oscillator
  - 2.2.4: Free Fall, Satellite Motion and the Mechanism of Tides
    - 2.2.4.1: Free Fall
    - 2.2.4.2: Circular Motion
    - 2.2.4.3: Projectile Motion- Very Long Range
    - 2.2.4.4: Satellite Motion
    - 2.2.4.5: Tidal Forces
    - 2.2.4.6: Extreme Tidal Effects
- 2.3: Thermal Energy and Heat
  - 2.3.1: Forms of Energy and Power- an Engineer's Approach

3: Fossil Fuels and Thermal Engines

- 3.1: Fossil Fuel Resources
- 3.2: How much energy there is in fossil fuels?
  - 3.2.1: How much CO2 is released by burning various fuels
- 3.3: Heat Engines
- 3.4: The Efficiency of Thermal Engines
  - 3.4.1: The First Law of Thermodynamics.
  - 3.4.2: Entropy
  - 3.4.3: The Second Law of Thermodynamics
  - 3.4.4: Sad Consequences of the Second Law
  - 3.4.5: The Efficiency of an Ideal Thermal Engine
  - 3.4.6: Even sadder news

4: Nuclear Power

- 4.1: A Note from the Author
- 4.2: Nuclear Physics- Brief Review of Fundamental Facts
- 4.3: Radioactive Decay
  - 4.3.1: Energy Units in Nuclear Physics
  - 4.3.2: The Decay Constants and Half-Lives
- 4.4: Nuclear Cross Sections
- 4.5: Nuclear Fission
  - 4.5.1: Fission and Fusion Energy
  - 4.5.2: How to calculate the mass defect?
  - 4.5.3: Energy Released by the Fission of a Single Nucleus
  - 4.5.4: Nuclear Fusion
- 4.6: Controlling the Fission Chain Reaction- Nuclear Reactors
  - 4.6.1: Essential Cross Sections
  - 4.6.2: No magic tricks are necessary!
  - 4.6.3: Controlling the Chain Reaction and Cooling the Assembly
  - 4.6.4: Power Reactors
  - 4.6.5: Generation IV Reactors and Thorium Fuel
- 4.7: NuScale's Small Modular Reactors – an Oregon-born Project
- 4.8: The Pros and Cons of Nuclear Power

5: Hydropower

- 5.1: Introduction
- 5.2: The Origin of Hydropower
- 5.3: Methods of Harnessing Hydropower
  - 5.3.1: The Kaplan and Francis Turbines
  - 5.3.2: The Pelton Turbine
- 5.4: About One Famous Hydropower Plant Using Pelton Turbines
- 5.5: Pumped Storage Hydroelectric Plants (PSHP)
- 5.6: New Ideas for Hydraulic Energy Storage

6: Windpower

- 6.1: Applications- Winds Over the Seas
- 6.2: Winds Blowing over the Lands
- 6.3: Applications- Land and Off-shore Use
- 6.4: The Physics of a Wind Turbine
  - 6.4.1: Problems with Wind Power

7: Direct Solar Power

- 7.1: Introduction
- 7.2: Harnessing Solar Power
- 7.3: The Concentrated Solar Power (CSP) Technology
  - 7.3.1: Solar Towers Molten Salt Heat Storing Technology
  - 7.3.2: Parabolic Troughs and Fresnel Mirrors
  - 7.3.3: The Linear Fresnel Mirror CSP Technology
  - 7.3.4: CSP – Pros and Contras

- 7.4: Photovoltaic Conversion of Solar Power
  - 7.4.1: Theoretical Background the Nature of Light and the Photoelectric Effect
  - 7.4.2: Semiconductors and Doped Semiconductors
  - 7.4.3: The p-n Junction
  - 7.4.4: Internal Photoelectric Effect in Semiconductors
  - 7.4.5: PV Cells
  - 7.4.6: Solar Cell Efficiency
  - 7.4.7: Inverters
  - 7.4.8: Pros and Cons of PV Solar Power Harnessing

8: Harnessing Power from Oceans and Seas

- 8.1: Introduction
- 8.2: Wavepower
  - 8.2.1: Methods of Harnessing Wavepower
  - 8.2.2: Attenuators
  - 8.2.3: Point Absorbers
  - 8.2.4: Other Wave Power Capturing Devices
- 8.3: Tidal Power
  - 8.3.1: A Bit More of Physics
  - 8.3.2: Tides in Coastal Regions, Tidal Currents
  - 8.3.3: Methods of Harnessing Tidal Power
  - 8.3.4: Tidal Power – Conclusions
- 8.4: Ocean Thermal Energy Conversion (OTEC)
- 8.5: Other Possible Ways of Extracting Power from Oceans
  - 8.5.1: Ocean Currents
  - 8.5.2: Osmotic (Salinity Gradient) Power Generation

9: Geothermal Power

- 9.1: The Non-Uniform Distribution of Geothermal Heat Flux
- 9.2: Geothermal Electricity Generation
  - 9.2.1: The Flash Steam Method
  - 9.2.2: The Binary Cycle Method
- 9.3: The Enhanced Geothermal Systems
- 9.4: Geothermal Heating
  - 9.4.1: Shallow Geothermal Heating

10: Biofuels

- 10.1: The History of Biofuels
- 10.2: New Solid Biofuels
- 10.3: Liquid Biofuels
  - 10.3.1: Bioethanol
  - 10.3.2: Biodiesel Fuel
- 10.4: Other Ways of Harnessing Plant Energy- “Bioelectricity

11: Energy Storage

- 11.1: Introduction
- 11.2: Overcoming the Challenges of Wind Power
- 11.3: Problems with Solar Power- the \Duck Curve"
- 11.4: Energy Storage Technologies
  - 11.4.1: Battery Storage
    - 11.4.1.1: The Voltaic Cell
    - 11.4.1.2: Lead-Acid Batteries
    - 11.4.1.3: Lithium-Ion Batteries
    - 11.4.1.4: The “Families” of Li-ion Batteries Manufactured
    - 11.4.1.5: Batteries for Industry- and Grid-scale Storage
  - 11.4.2: Electricity → potential energy → back to electricity (gravitational energy storage and compressed air energy storage).
  - 11.4.3: Electricity → Kinetic Energy → Electricity Storage Scheme
  - 11.4.4: Electricity → Thermal Energy → Electricity Storage Scheme
    - 11.4.4.1: Molten Silicon
    - 11.4.4.2: Heated Volcanic Rocks
    - 11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage
    - 11.4.4.4: Small and Medium Scale Thermal Energy Storage
  - 11.4.5: Electricity → Chemical Energy
    - 11.4.5.1: The evolution of the hydrogen economy policy in the 21st century
    - 11.4.5.2: Methods of Green Hydrogen Production
    - 11.4.5.3: The recovery of energy stored in the form of hydrogen
    - 11.4.5.4: Hydrogen storage and transportation issues
    - 11.4.5.5: Hydrogen as a “green” fuel suitable for sea- sonal storage

Index

Glossary

Detailed Licensing

Detailed Licensing

Licensing

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

CHAPTER OVERVIEW

1: Energy Alternatives { what's the meaning?}

1.1: Brief semantic considerations

1.2: Things to be chosen" What things, what choice?

1.3: Back to the definitions- which one? 2a or 2b?

1: Energy Alternatives { what's the meaning?} is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

1.1: Brief semantic considerations

Semantics is a branch of linguistics concerned with the meaning of words, phrases, or texts. It seems worthwhile to do a quick ”semantic analysis” of the course title, and explain the underlying motivations of choosing such a title.

The meaning of the term ”energy” has been explained in detail in the preceding section – so now we need to focus our attention on the world ”Alternatives”. Let’s take a highly respected source for word definitions, the *Merriam-Webster Dictionary*, and look up ”Alternative” (we can do that conveniently using the Internet). We find that the word may play the role of either an adjective, or a noun:

Adjective:

1: offering or expressing a choice (e.g., ”several alternative plans”);

2: different from the usual or conventional: such as

> 2A: existing or functioning outside the established cultural, social, or economic system (e.g., ”an alternative newspaper”, ”alternative lifestyles”;
>
> 2B: of, relating to, or being rock music that is regarded as an alternative to conventional rock and is typically influenced by punk rock, hard rock, hip- hop, or folk music;
>
> 2C: of or relating to alternative medicine (e.g., ”alternative therapies”).

Noun:

1:

> 1a: a proposition or situation offering a choice between two or more things only one of which may be chosen;
>
> 1b: an opportunity for deciding between two or more courses or propositions.

2:

> 2a: one of two or more things, courses, or propositions to be chosen;
>
> 2b: something which can be chosen instead (e.g., ”the only alternative to inter- vention”).

3: alternative rock music.

Clearly, the word ”Alternatives” used in our title is not an adjective, because (i) it’s plural, and there is no plural form of adjectives, and (ii) adjectives are normally positioned before the nouns. If ”alternative” were intended to be used as an adjective, the title would be *Alternative Energies*, not *Energy Alternatives*. In the latter, “Alternatives” is obviously a plural noun. And it’s definitely not Item 1a, 1b, or 3 from the noun group quoted above. So, the meaning is either that defined by Item 2a, or that defined by Item 2b. Which one is the meaning preferred by the author of this text, and why? In order to answer this question, let’s for a moment talk about the other word in the title, i.e., the “Energy”.

1.1: Brief semantic considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

1.2: Things to be chosen" What things, what choice?

In this book, the term “energy” will *sometimes* refer to “energy as understood by a physicist” – but primarily, to its meaning in the “engineer’s language”. More specifically, it will refer to forms of energy that can be either used directly, or can be converted in order to obtain:

- Electric power, to be sent to a single user, or to many users;
- Thermal power, to be used for heating homes, cooking, making hot water, and in a range of industrial technologies; and
- Mechanical power, for propelling vehicles, pumping water, lifting loads, etc..

Now, it’s common knowledge that there are old, well-established methods of exploiting available, but depletable, energy resources – we will call them “traditional”. And that there are newer methods, believed (not by everybody, though!) to be “clean” and more friendly to the planet and environment than the “traditional” ones, and based on resources that are not being depleted in spite of extracting power from them. One can think of such sources as “non-depletable” – however, they are commonly referred to as “renewable”, and we will use this term further on[1].

Table \(PageIndex{1}\): Major resources of energy

Traditional			Renewable
Coal and Lignite			
Petroleum derivatives			
Natural Gas Hydropower	→	→	Hydropower
Nuclear power (generations I-III)	→	→	Nuclear power (newer generations)
Wind power	→	→	Wind power
Biofuels	→	→	Biofuels
Geothermal power	→	→	Geothermal power
Solar power	→	→	Solar power
			Wave/Tidal power

The major “traditional” and “renewable” resources are shown in Table \(\PageIndex{1}\). Some resources are shown in both columns. Why? For reasons one may call “historical”. Hydro-power has been used by humans for thousands of years, and, reportedly, the very first hydroelectric power station was built in 1868, even a few years earlier than the first coal-fired station. Oil and natural gas-fired stations emerged only in the XX-th century. Hence, “by seniority” hydro-power has the right to be called “traditional”. On the other hand, hydro-power is very clean, does not emit any atmospheric pollution and it does not deplete any existing resources – actually, the sources of river and stream water the hydro-power plants use are constantly “renewed” by precipitations. Therefore, it is listed in both columns in the table, and the small arrows are meant to show that the “status” of hydropower has changed: when nobody was worried about the CO_2 emission, it was thought of as yet another method of generating electricity – but now, due to its zero-emission property, is considered an important member of the family of technologies of harnessing renewable energy resources.

Nuclear power is also shown in both columns. The first nuclear power plants emerged in the 1950-s, so they are “relatively young” by age, but they quickly became a major component in overall electric power generating systems in many countries (e.g., about 20% in the US, nearly 80% in France). However, in the past there was a number of minor and medium-scale potentially hazardous accidents in nuclear power plants all over the world, and three of real catastrophic scale: Three Mile Island in 1979, Chernobyl in 1986, and Fukushima in 2011. In all three cases the principal reason for the disaster was the malfunctioning of the main cooling systems. Thus, because of safety concerns, the existing facilities, using older Generation I through III reactors of previous generations, have no future and will certainly be shut down when the reactors reach their allowed lifespan. Yet, much effort has

been devoted in recent years to designing new types of reactors (often referred to as “Generation IV” ones), which will be “intrinsically safe”, because the cooling systems will never stop working, even if everything else fails and human operators make all possible errors. One unquestionable advantage of nuclear power plants is that such facilities do not emit CO_2. In order to create a future “zero-emission” electrical utility industry, all power plants using mined combustible fuels (coal, oil and natural gas) need to be gradually eliminated and replaced by facilities using only renewable resources. However, in the opinion of some renowned expert teams, such an ambitious goal may not succeed unless nuclear power generation becomes a meaningful “component” of such industry. It is why nuclear power is listed in the right-hand column[2] in Table 1.

Solar energy, wind energy, geothermal energy and energy from bio-fuels are thought of as the “backbone” of the future zero-emission energy system – but all these energy forms have been used by humanity, at a minor scale, for thousands of years. Therefore, they cannot be totally ignored in the list of “Classical” energy resources – but they are listed in the left column in Table \(\PageIndex{1}\) using a small font.

---

[1]Even though, in the author’s opinion, “non-depletable” would be a better term than “renewable” because the former better describes the actual physical situation.

[2]In fact, the nuclear power sector **does** use mined material as fuels. Currently, it’s the “fissile” ^{235}U isotope which comprises only 0.7% of the natural Uranium. The reminding 99.3% is almost exclusively ^{238}U which is non-fissile. However, there are not yet exploited technologies in which the nonfissile ^{238}U, and a non-fissile Thorium isotope, ^{232}Th, **can be** used as fuels. The combined global resources of ^{238}U and ^{232}Th are not “undepletable”, of course, but they are sufficient to satisfy the energy needs of humanity for many thousands of years, so they may be thought of as “nearly renewable”.

1.2: Things to be chosen" What things, what choice? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

1.3: Back to the definitions- which one? 2a or 2b?

As follows from the preceding subsection, in the present book, the “things” the Merriam-Webster definitions are talking about are the “Traditional” and “Renewable” energy sources.

Suppose that the Author of this text prefers the definition 2b: “Something which can be chosen instead”. What would that mean? Clearly, considering today’s “state of mind” of public opinion, the thing “to be chosen instead” certainly would be the Renewable sources instead of the Traditional ones. In other words, the book would be mostly about the “Renewables”, and, naturally, about their advantages. And since “they are the ones to be chosen”, whenever the “Traditional” are mentioned, a negative, or – at least – an unflattering comment should be added.

Such an approach is definitely not the Author’s intention. On the contrary, he is trying to avoid “hidden messages” – namely, that “the Renewables are ‘the good guys’, and the ‘bad guys’ are the Traditional Methods.” If the text contained such hidden messages, it would **instruct you** what you are supposed to think! And this would be nothing else than **propaganda** (after Wikipedia: *Propaganda is a form of communication to distribute information. ... The information is designed to make people feel a certain way or to believe a certain thing*). Or, one can say that propaganda is “uploading ideas, opinions and ready answers to people’s minds.” An extreme version of propaganda is known as “brainwashing”.

A method that had much in common with today’s propaganda tools was *scholasticism*, a system that dominated Western European learning, teaching and philosophical disputes in Medieval Ages. The foundation of the knowledge passed to students were “dogmas” which had to be accepted without discussion. For example, one such famous (or, rather “infamous”) dogma stated that Earth is the center of Universe. A custodian of the “official dogmas” was the Church, an extremely powerful institution. Openly doubting dogmas was not only discouraged – it could be dangerous: The Church created a special “police” known as the Holy Inquisition, whose task was to pursue philosophers and scientist expressing opinions contradicting the official dogma set. Few philosophers undertook such a risk, as some of them faced dire consequences – the Italian philosopher Giordano Bruno was burnt alive on stake, and Galileo Galilei was given “an offer he could not refuse:” *Revoke your theories, or... .*

The dictatorship of scholasticism was overthrown only in the XVIIIth Century, when the Era of Enlightenment began, preceded by a scientific revolution. Rigid dogmas were rejected, critical thinking was encouraged. The ideas of the Enlightenment Era have evolved into today’s modern scientific thinking: there are no “taboos,” no sacrosanct views. There are no dogmas. No theory should be thought of as “the only right one.” According to today’s standards, nothing is fixed. In any area of science, “the state of understanding things” is described by scientific **laws** and scientific **theories** (in greater detail, the meaning of the words “law” and “theory”, and, additionally, of the word “hypothesis”, is given in Subsection XXX). In contrast to dogmas, the laws and theories of contemporary science should not be thought of as **fixed**. They describe our **current understanding** – but if new facts emerge, we may be forced to change or to reformulate them.

So, in the spirit of contemporary science – no dogmas, no dogma-like conclusions in this book! No opinion imposing on the reader! The author is trying to do his best to be completely honest, and to present the material in an unbiased way enabling you to reach your own conclusions. People may have different opinions about the same things. For instance, former Vice-President Al Gore is a strong proponent of Renewable Energies – while President Trump, as we all know, is definitely not in favor of them.

Then, returning to the question: 2a or 2b? In view of the arguments outlined above, we clearly should chose the 2a definition. We want to present the Traditional Energies and the Renewable Energies as two options – and leave to the Reader the freedom to decide which one of the two she/he will support.

There is still one question remaining. Suppose that this book manages to convince you that Renewable Energies are “the way to go”. Is there much sense for you to learn about all those classical technologies? Or, vice versa: you may conclude that harnessing the Traditional Resources is a wellestablished method, there exists a reliable industry that satisfies all consumer’s needs – changing it at a major scale by unproven technologies of questionable reliability would lead to chaos, jobs would be lost, and so. For those who prefer such thinking, does it make much sense to spend much time to learn about the whole menagerie of methods of harnessing Renewable Resources?

The answer is **yes!**, in either case. Once you have reached your opinion, you would certainly like to be able to defend it in discussions with those who favor the other option. Or not only defend, but even to win in such discussions. And what gives one good chances for winning?

Consider military victories. In one of the most famous battles in the history of warfare, at Cannae at 216 B.C., Hannibal's routed the attacking Roman army, outnumbering his forces by nearly 2:1. How was that possible? One of the decisive factors was the superior Hannibal's intelligence: Before the clash, he had managed to gather much information about the number of enemy troops and of their favorite tactics. It allowed him to prepare a brilliant battle plan. And at Gettysburg in 1863, General Robert Lee's forces suffered a major defeat, after they launched the doomed "Picket Charge" on Union positions. Lee, whose knowledge about the enemy was based on information from a few low-quality sources, badly underestimated the number of opposing troops – while, in contrast, the Union intelligence officers had managed to compile a highly accurate picture of Confederate forces – so that General George G. Meade, the US Commander, knowing of the large numerical superiority, could defend his positions with full confidence.

Verbal disputes are not the same as warfare – but similar rules apply. Think of the following situation: you are a "green" person, and at some social gathering you engage in a discussion with several other people who, as it turns out, all support the views of President Trump on renewable energies. Your goal is to convince them that renewable energies **are** better than those obtained from burning fossil fuels – and to accomplish that, you need not only to know the advantages of renewable energies, but – which is even more important – to know the **disadvantages** of the fossil fuel technologies. You should be able to speak at a tone that is not only persuasive, but which also makes it clear that you have an authentic knowledge of what you are talking about. And the same is true the other way: if you are a supporter of President Trump's energy policy, and you are confronted with several "green activists" – if you don't want to be overwhelmed with their arguments, you should demonstrate that you have a good knowledge of the "Renewable Energies" topic – preferably, even better than that of your opponents – and your choice of the alternative option is a conscious, well thought-out decision, and not a result of "brain washing."

So, in summary – our intention is that this course teaches you about all methods of harnessing power, these based on the traditional resources, as well as these utilizing the renewable resources. And the decision of whether you prefer to support the former vs. the latter, should be yours. We don't want to impose the decision on you – we only want to help you to reach your own.

One more thing to end with. A bit earlier it was said that all modern natural sciences reject dogmas. However, you should be warned that dogmas started "returning with vengeance," n a great measure aided by the Internet. The "new dogmas" are not based on the results of authentic and honest scientific research – although people disseminating them do their best to create an impression that they are highly competent "experts in the field". Some examples of such pseudo-scientific dogmas are: vaccination of children leads to autism; GMOs cause cancer; cancer can be cured by high doses of vitamin D_{17}; anthropogenic global warming is a myth created by a conspiracy of corrupted scientists – and so on. But we are not telling you: "Reject such revelations right away!" – dogmas should not be counter-attacked by "counter-dogmas". Rather, try to perform some research of your own on the subject. The Internet is not only a means of spreading pseudo-scientific sensations – the very same Internet also offers vast opportunities of reaching sources that can be used for checking the trustworthiness of any information.

1.3: Back to the definitions- which one? 2a or 2b? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

2: General Remarks

2.1: Energy and power

2.1.1: Energy Conservation

2.1.2: Energy and Work

2.1.3: Energy Units

2.1.4: Power

2.1.5: Energy expressed in terms of power

2.2: Forms of Energy- Physicist’s View .

2.2.1: Potential Energy

2.2.2: Kinetic Energy

2.2.3: Energy of an Oscillator

2.2.4: Free Fall, Satellite Motion and the Mechanism of Tides

2.2.4.1: Free Fall

2.2.4.2: Circular Motion

2.2.4.3: Projectile Motion- Very Long Range

2.2.4.4: Satellite Motion

2.2.4.5: Tidal Forces

2.2.4.6: Extreme Tidal Effects

2.3: Thermal Energy and Heat

2.3.1: Forms of Energy and Power- an Engineer’s Approach

2: General Remarks is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.1: Energy and power

One may find many different definitions of the notion of energy in various sources. Not exactly ”different”, they are all equivalent – it’s the wording that is different. I think that the best textbook definition is a simple one, and I will combine the wording from an Wikipedia article with some extra comments of my own ”additions”:

In physics, energy is the property that must be transferred to an object in order to perform work on or to heat the object, and can be colorlinks, citecolor=black, converted in form, but not created or destroyed.

OK, for me everything is correct in the above definition – with the exception that I would prefer energy be referred to as an ”entity”, rather than as a ”property”. Entity means ”something that exists”, whereas property, according to the dictionary, is ”an attribute, quality, or characteristic of something”. What I don’t like is this ”of something”. The term ”property” in Wikipedia definition implies therefore that energy is always associated with an **object** – it has to be contained by an object, or transferred from one object to another. In fact, it is ”almost always” the case, but there exist some ”exotic situations” in which it is nor easy to identify ”the object”. For instance, energy – in the form of radiation, such as, e.g., the solar light – is able to travel through empty space. In other words, energy may exist in empty space, or may fill a region of empty space – and is empty space an object? From the viewpoint of advanced modern physics, a firm **no!** is not a good answer to this question. According to modern theories, vacuum is –let me quote a statement I’ve found in Physics Forum – “very much alive and kicking”. Well, but all this “life” is only a **virtual** one – we cannot see it, even with the help of sophisticated modern measuring devices. So, if we appeal to our “conventional wisdom”, calling “empty space” an object would be a bit weird. As will also be calling energy flowing through vacuum “a property” of this vacuum. But if instead of “property” we agree to use “entity”, the problem automatically disappears. Energy travels through vacuum, so it does exist in vacuum it travels through – and “entity”, as said above, is “something that exists”. Now everything is OK!

Actually, whether we should say “entity” or “property” is rather a linguistic dilemma, it’s really not so important. One can even think of it as an example of *tetraphyloctomy* – it’s a Greek term coined by the famous writer Umberto Echo, meaning ”splitting a hair in four”, or paying too much attention in scientific disputes to details that are of little importance, or totally irrelevant. In fact, what is really relevant, is to understand what energy is. Let’s then analyze the Wikipedia definition in closer detail.

2.1: Energy and power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.1.1: Energy Conservation

Let's start with the following statement: *energy... ...can be converted in form, but not created or destroyed*. Very important! Energy cannot "pop out of nothing" (with the exception of Harry Potter movies), it cannot disappear without a trace. This expresses a law, known as the Energy Conservation Law, one of the most fundamental laws of physics. Energy may be passed from one object to another, but the sum of energies in all objects under consideration remains unchanged. In other words, some objects may become "energized", while some other "de-energized", with the total energy unchanged in the process. Also, as was argued above, energy from one object must not necessarily be transferred to "another object", it may be "radiated out", i.e., it may leave the first object in the form of radiation and then travel through empty space. However, the energy of the radiation "sent out" must be exactly equal to the amount of energy lost by the object from which the radiation is emitted.

2.1.1: Energy Conservation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.1.2: Energy and Work

In the Wikipedia definition quoted at the beginning it is stated that: *energy... ...must be transferred to an object in order to perform work on.....the object*. Let's recall what "work" means in physics. Actually, physics recognizes several different types of work – but one of them, the **mechanical work**, is of special importance. Suppose that a body is at rest at a initial position x_i. Then, a force F is exerted to it for some time, and as the result of it the body is displaced to a new "final" position x_f. We will call the difference

$$\Delta x = x_f - x_i$$

the *displacement*, and the product of Δx and the force F:

$$\Delta W = F \cdot \Delta x \tag{2.1.2.1}$$

is the mechanical work performed in the process of shifting the body.

Performed by whom or by what? Well, we don't need to specify, it's enough to assume that there was an *object* capable of exerting the force F on the body. Initially, this object contained certain amount of energy, E_i and after the work was performed, this amount changed to a final lower value \(E_f). So, the difference:

$$\Delta E = E_f - E_i \tag{2.1.2.2}$$

is the energy the object "paid" for performing the work. This is obviously a negative number, right? And now we have reached a very important point – namely, a declaration that is known as the *Work-Energy Theorem*, or the *Work-Energy Principle*, or even by longer name: *The Principle of Equivalence of Work and Energy* – and it simply states that:

$$\Delta W = -\Delta E \tag{2.1.2.3}$$

(note that \Delta E is a negative number, so the minus in the above equation is needed to obtain a positive work). In other words, the above principle simply states that for performing work, the "performer" has to pay an equal price in energy. It also works the other way around: for increasing the energy of an object by ΔE, one has to "pay the same price" in work.

2.1.2: Energy and Work is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.1.3: Energy Units

It was said above that the mechanical energy is of "special importance" in physics. Why? Because the most important energy unit used in physics, the Joule, is defined as - let's again quote Wikipedia - the energy transferred to (or work done on) an object when a force of one newton acts on that object in the direction of its motion through a distance of one meter.

But why is it "the most important"? Well, the thing is that in modern physics there are special regulations, known as the "SI system", stating that the units of all quantities used in physics should be defined in terms of seven **base quantities**. Three of these seven are the units of length, the meter (m), of mass, the kilogram (kg), and of time, the seconds (s). For more details, please look at a NIST Web page (NIST = National Institute of Standards and Technology).

So, the SI unit of force is a Newton:

$$1\ \mathrm{N} = 1\frac{\mathrm{kg}\cdot\mathrm{m}}{\mathrm{s}^2} \tag{2.1.3.1}$$

Why such a combination? In short, it all comes from something known as "the Second Newton's Law of Dynamics". And how it comes? Well, I think that rather than explaining it step-by-step, I should give you a link to a good Internet source. And if you don't yet have an idea "how big" the force of $1\ \mathrm{N}$ is, I can tell you that the weight of something with the mass of **one pound** is approximately $4.5\ \mathrm{N}$; and of 1 kilogram mass, about $9.81\ \mathrm{N}$.

Now, once we know the SI unit of $1\ \mathrm{N}$, we can readily find the SI unit of a Joule:

$$1\ \mathrm{J} = 1\ \mathrm{N}\cdot 1\ \mathrm{m} = 1\frac{\mathrm{kg}\cdot\mathrm{m}^2}{\mathrm{s}^2} \tag{2.1.3.2}$$

2.1.3: Energy Units is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.1.4: Power

If energy is transferred from one object to another (e.g., for performing work), it is not only important to know how much energy is passed, but it may also be essential to know how fast this energy is delivered. So, there is a special unit describing the rate of energy transfer, called **Watt**. Its symbol is "W", and it's defined as the transfer of one Joule per one second:

$$1\ \mathrm{W} = \frac{1\ \mathrm{J}}{1\ \mathrm{s}} = 1\frac{\mathrm{kg}\cdot\mathrm{m}^2}{\mathrm{s}^3} \tag{2.1.4.1}$$

2.1.4: Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.1.5: Energy expressed in terms of power

The Watt and its derivatives $(1\ \mathrm{kW} = 1000\ \mathrm{W}; 1\mathrm{MW} =$ one million Watts; or one miliWatt $= 1\ \mathrm{mW} = 0.001\ \mathrm{W})$ are familiar and often, used units. The notion of power readily appeals to one's imagnination. If you hear someone says: "This is a $60\ \mathrm{W}$ light bulb", or "I have purchased a new $10\ \mathrm{kW}$ air conditioner for my house ", or "The power of the new Conda Civic's engine is $150\ \mathrm{kW}$ ", you usually know immediately what this person is talking about.

In contrast, the Joule is a "much more abstract" unit. If one says: "This light bulb has consumed 216000 Joules of energy", you have to think for a moment or even longer to figure out what the person is talking about. Even though it's simple a 60 W bulb uses 60 Joules every second, so over one hour it used $60\ \mathrm{J/s} \times 3600\ \mathrm{s} = 216000$ Joules.

Therefore, convenient way of describing energy transfer is not to use Joules, but an equivalent unit based on the Watt, called "Watt second", with the symbol Ws, W-s, or W.s. One Watt-second is the amount of energy delivered by a source of $1\ \mathrm{W}$ power over the period of 1 second so, it's indeed the same as 1 Joule. A derivative unit is the Watt-hour, Wh or W-h, equal to 3600 Joules; and perhaps the most often used energy unit in everyday life, the kiloWatthour, $\mathrm{kWh} =$ the energy delivered by a source of $1\ \mathrm{kW}$ power over the period of one hour $= 3,600,000\ \mathrm{J}$.

Look at the bill the power company sends to you you pay for kiloWatthours, not for Joules. From the Web, you may learn that in the year 2015 an average american household consumed $10.8\mathrm{kWh}$ per day, or $901\mathrm{kWh}$ per month. Simple? well, definitely simpler to comprehend than if the same Web source said: "3,243,600,000 Joules per month".

2.1.5: Energy expressed in terms of power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

2.2: Forms of Energy- Physicist's View .

In "pure physics", we essentially recognize two basic types of energy - potential and kinetic.

2.2.1: Potential Energy

2.2.2: Kinetic Energy

2.2.3: Energy of an Oscillator

2.2.4: Free Fall, Satellite Motion and the Mechanism of Tides

2.2.4.1: Free Fall
2.2.4.2: Circular Motion
2.2.4.3: Projectile Motion- Very Long Range
2.2.4.4: Satellite Motion
2.2.4.5: Tidal Forces
2.2.4.6: Extreme Tidal Effects

2.2: Forms of Energy- Physicist's View . is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.1: Potential Energy

Potential Energy is any form of energy that can be "stored" by a physical system for instance, the energy of a stretched spring, a body in a gravitational field, or the energies of some physical fields. The name comes from the fact that such energies, essentially, can be stored for an unlimited time, and then converted to work or other energy form at a chosen moment moment of time. One can then think of such systems "pumped" with energy as of those which "have a potential".

Potential Energy of Stretched Springs

The energy of a stretched or compressed spring can be readily calculated, based on the Hooke's Law that describes the relation between the extension (a.k.a. elongation) of a stretched spring, or the contraction (a.k.a. shortening) of a compressed spring, and the "restoring force" F, i.e., the force tending to bring the spring back to its original length:

$$F = -k \cdot \Delta l \tag{2.2.1.1}$$

where Δl is the change in the spring length (positive for extension, and negative for contraction), and k is the so-called "stiffness constant" of the spring; the minus sign reflects the fact that the restoring force is always oriented towards the equilibrium position, so its sign is always opposite to the sign of Δl.

Knowing the force, we can readily calculate the work that has to be performed for stretching/compressing the spring by Δl. For such an action we need to apply a force which overcomes the "restoring force", i.e., which points in the opposite direction. Since the force needed is not constant, but it changes during the process, we cannot use a simple multiplication as in Equation (1). We have to integrate:

$$\Delta W = \int_0^{\Delta l} F(l)dl = \int_0^{\Delta l} kldl \tag{2.2.1.2}$$

$$= \frac{1}{2}k(\Delta l)^2 \tag{2.2.1.3}$$

Now, invoking the Work-Energy Principle, we can conclude: *The amount of potential energy* [1] *stored in the stretched spring considered is*:

$$\Delta V = \frac{1}{2}k(\Delta l)^2. \tag{2.2.1.4}$$

The discussion on the energy of a stretched spring is an instructive example demonstrating how to use the Work-Energy Principle. It also shows that a spring can be thought of as an elementary h "energy storing" device. In fact, stretched springs were used in the old days for energizing clocks, watches, phonographs, mechanical toys, small movie cameras and many other portable devices. Today, they have been replaced by batteries. Therefore, we will not talk very much about the energy of stretched springs (or "elastic energy") further on with one exception, namely, when we discuss the the energy of a spring oscillator.

Potential Energy due to Gravity

Another good example is the potential energy of a lifted object of mass m. The force of gravity acting on such an object is $F = mg$, where $g = 9.81\ \mathrm{m/s^2}$ is the acceleration due to Earth's gravity. The force is constant, we don't need to integrate so if we lift the mass from the height H_i to the height H_f, we do work $\Delta W = mg\,(H_f - H_i)$. So, again invoking the Work Energy principle, we have shown that by lifting the mass we have increased its potential energy by:

$$\Delta V = mg\,(H_f - H_i) \tag{2.2.1.5}$$

A lifted object can be then used as an "energy reservoir". Actually, such method is used for storing huge amounts of energy in facilities known as "Pumped Storage Hydroelectric Power Plants", about which we will talk later, in a section titled "Hydroelectric

Power". They store energy not by lifting a solid object of mass m, but by pumping water from a lower tank to a higher tank but the general principle is the same.

A lifted object can be then used as an “energy reservoir”. Actually, such method is used for storing huge amounts of energy in facilities known as “Pumped Storage Hydroelectric Power Plants”, about which we will talk later, in a section titled “Hydroelectric Power”. They store energy not by lifting a solid object of mass *m*, but by pumping water from a lower tank to a higher tank – but the general principle is the same.

There is, however, one “tricky issue” with the gravitational potential energy. Suppose that you are standing on a balcony. On the balcony’s floor there is a weight of 1 kg mass. You lift it above your head, so it’s now 2 meters above the balcony’s floor. You say: *Its potential energy is now 1 kg times 2 meters times g, which is 19.62 Joules*. But your friend standing underneath at the ground level disagrees: *The balcony’s floor is three meters above the level, so the weight is now at a height of five meters, hence its potential energy is 49.05 Joules!* Who is right? You both are right and aren’t right. How comes? The thing is that the Eq. 2.8 describes only the **change** of the potential energy, not its absolute value. Giving the potential energy value is a worthless information – one has always specify **relative to what level t**he potential energy is of the value given. So, the correct statements in the above examples should be, respectively: *The potential energy of the weight relative to the balcony floor is 19.62 Joules*, and *The potential energy of the weight relative to the ground level is 49.05 Joules*.

One may insist that if we reach an “universal agreement” – say, that all potential energies should be given relative to the same level, e.g., the sea level, then it would be possible to skip the *relative to* .... part of the statement. Well, such a convention would be possible, but not very practical – everybody would need to know exactly how high about the sea level the object considered is. Using a GPS would not be a good solution, because the altitude information offered by such devices is known to be not too accurate. And the need of always giving information about the “reference level” is not a major inconvenience – therefore, the idea of using an universal reference such as the sea level never caught up.

Which is not 100% true, though. There are situations in which such an universal reference level becomes very useful. Namely, if we consider potential energies of objects at very high altitudes, such as artificial or natural satellites, spacecrafts, or “near-Earth asteroids”. High above the Earth surface the value of *g* can no longer be taken as a constant. It decreases with the altitude *h* as:

$$g(h) = g_0 \frac{R^2}{(R+h)^2} \tag{2.2.1.6}$$

where g_0 = 9.81 m/s^2 is the acceleration due to gravity at the Earth surface, and *R* = 6371, 000 m is the Earth radius.

When considering the potential energy of an object at *h* such that *g*(*h*) is considerably smaller than g_0, we use as the “reference level” *h* = , and we assume that the potential energy at this level is... zero. So, it means that at any distance “closer than infinity” to Earth the potential energy of any object is... negative! Yes, in such convention all potential energies are negative. Weird? At first glance it might seem so. But note that there is nothing wrong with physics: for a lower *h*, the potential energy is “more negative than for a larger *h* – so, with increasing *h* the potential energy indeed increases, as it should be.

We will return to this issue later on, when discussing the energy of the Earth-Moon system, and the ocean tides. Moving the masses of tidal waters needs some energy input, and some part of it is “extracted” from the energy of the Moon’s orbital motion – leading to a small increase (yes!) of the Moon-Earth distance, with the rate of 3.8 cm/year.

---

[1]I use *V* as the symbol of potential energy, but in other books or Web sites you may find different symbols. The same is true for many other quantities in physics. The thing is that there are many more quantities of interest in physics than available Roman and Greek letters. Therefore, the same letter is often used for several different things – e.g. *V* is also used for volume, velocity, voltage and perhaps several other quantities.

2.2.1: Potential Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.2: Kinetic Energy

Kinetic energy is the energy of a moving object, executing a *linear* motion or a *rotational* motion. For a solid object of mass *m* moving with a speed[2] of *v* the formula for kinetic energy is:

$$K = \frac{1}{2}mv^2 \tag{2.2.2.1}$$

Moving fluids or gasses also carry kinetic energy, but the appropriate math formulae are more complicated – actually, about the kinetic energy carried by a moving gas (air) we will talk in the section "Wind Power".

Kinetic energy in linear motion can hardly be used as a practical means of storing energy, because if one tried, the "reservoir would run away" :-). However, this is not the case with kinetic energy of *rotational motion*, i.e., of spinning object.

The kinetic energy of a spinning rigid solid body is given by:

$$K = \frac{1}{2}I\omega^2 \quad \text{with} \quad \omega = 2\pi f \tag{2.2.2.2}$$

where f is the number of revolutions per second executed by the spinning body, and I is a quantity called *the moment of inertia*, which depends on the mass, on the shape, and on the dimensions of the body in question. The formulae for moments of inertia of bodies of many different shapes can be readily found in the Web, for instance, in this Hyperphysics Web site.

In contrast to bodies moving along straight trajectories which are rather useless from the viewpoint of energy storing, spinning bodies can be used and **are widely used** for such storing. Such bodies are commonly known as "flywheels". If a flywheel spins to fast, it may be torn apart by centrifugal forces – therefore, the the amount of energy that can be stored by a flywheel is not spectacular. One interesting idea that emerged in the middle of the last century was to use flywheels in public transportation, for propelling buses. After the flywheel in a such vehicle, called a "gyrobuses", was "charged with energy" by en electric motor, it could run up to 6 km (about 3.5 miles). An advantage of gyrobuses was that they were very "clean", they emitted no fumes, in contrast to buses propelled by Diesel engines. However, there were many problems with maintaining fleets of such vehicles, so the initial enthusiasm gradually died out, and the several existing Gyrobus service systems in Europe and Africa were shut down by the end of the 1950-s.

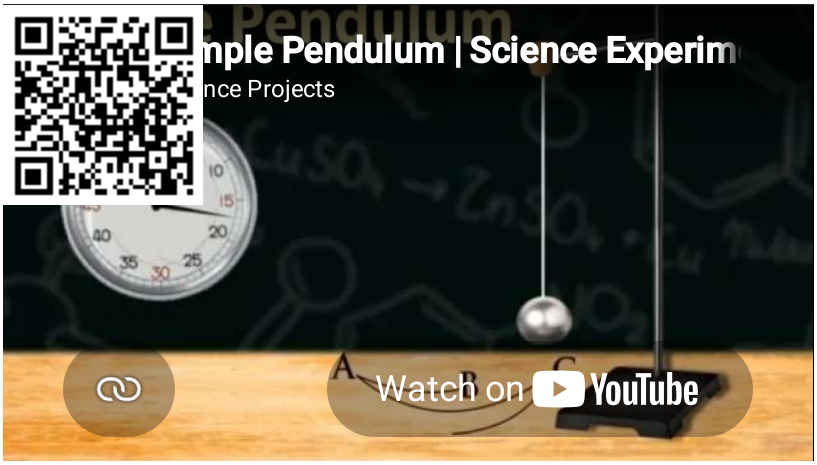
---

[2]Note that as a symbol of speed we use the same letter as we have used a moment ago for potential energy – but now we use **a small letter** for distinction. By the way, in "professional slang" big (capital) letters and small letters are often referred to as, respectively, "upper case letters" and "lower case letters". Why? For historical reasons! You may find an explanation, e.g., in this Web document.

2.2.2: Kinetic Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.3: Energy of an Oscillator

The energy of simple oscillating objects, such as a pendulum, a weight suspended on a spring, or a vibrating guitar string is worth taking a closer look at, even though such things have no direct application in the grand-scale energy usage that is the main subject of this course. Namely, in the oscillating system mentioned above one can clearly see the processes of converting one form of energy to another. Consider a simple pendulum, i.e., a weight of mass m suspended on a piece of a light string. Pull the weight to the right from the equilibrium position, as shown in this short video clip



By doing that, you "charged" the mass m with potential energy ΔV, because it's now slightly higher than it was at the equilibrium position. Now, you let it go - and and the potential energy starts gradually changing to kinetic energy. When the mass passes through the equilibrium point, all the acquired potential energy is gone, it has been entirely converted to kinetic energy. **By the way, here is a small challenge for you:** you may use $K = \Delta V$ and process it further in order to find the speed v of the mass m when it passes through the equilibrium point -- assume that the length of the pendulum string is, say 1 meter, and in the initial displaced position the string made an angle of 20 degrees with a vertical line; please try, it's a good exercise. I am not giving you the value of m because, as you will find out, you don't need to know it. It's a good exercise - and if you have a major problem with obtaining the solution, it will probably mean that you may need some extra studying of very elementary physics in order to do well in this course).

Back to the pendulum: after the initial potential energy is totally lost and totally converted to kinetic energy at the instant of passing through the equilibrium point, the pendulum continues its swing motion, but now the kinetic energy is gradually converted back to potential energy. The process lasts until the pendulum reaches the maximum displacement angle, equal to the initial displacement angle -- but now to the left. At this instance, the kinetic energy is totally gone, the speed of the mass m becomes zero again, and the initial potential energy ΔV is totally "regained". And from this moment on, the process described above restarts, and runs in the opposite direction. A full swing from the right to the lest and back to the right is called "the oscillation cycle", and the time it takes to make such a full cycle is called "the oscillation period".

In the vibration of a guitar string, or in the oscillations of a blob suspended on a spring, you have exactly the same periodic process of conversion of potential energy to kinetic energy and ~back ~to ~potential ~energy. Yet, as mentioned above, the oscillations of such objects do not play any significant role in the grand-scale energy usage, why do we talk about them over here? Well, the reason is simple -- because the energy of oscillations of **atoms** in solids, the sum over all atoms, is what constitutes the **thermal energy** contained in such objects.

Free Fall

If we raise a body above the Earth surface at a distance H, and let it go -- then, no matter what the value of H is, one meter or one million meters, the body starts "falling freely" because it is attracted towards the Earth center by the force of gravity. And, if we forget about the air drag force (which is relevant only for H values of a few tens of kilometers), then we certainly have the right to say: "A free fall is the motion under the influence of the force of gravity, and no other force".

Now, suppose that the in initial velocity of the body considered is not zero, but we give it a "push" in the direction perpendicular to the line drawn between the initial body's position, and the Earth center. Let's call this line "the z direction", and the speed along this line as "the vertical velocity v_z". Let's call the direction perpendicular to the z line as"the x axis", and the speed along the x axis as "the horizontal velocity v_x". So, at the initial moment $v_z = 0$, and $v_x \neq 0$. What happens next? The body starts moving downwards, with constant acceleration g, under the influence of gravity. But there is no horizontal force acting on it, so the initial v_y remains unchanged. The resultant body's trajectory is a parabola, and it hits the Earth surface some distance from the point right below the initial body's position.

In introductory-level physics textbooks the type of motion described above is usually referred to as "projectile motion"}. Three different sub-types are discussed, with the initial "push" directed strictly horizontally, at some angle upwards, or at some angle downwards. Each sub-type has to be treated slightly differently. It's a good approach, instructive and pedagogical: it helps young students to develop an ability called "physical insight", in which a correct "conceptual understanding" of a physical phenomenon considered is combined with a correct choice of "mathematical tools" needed to analyze it in quantitative terms. However, in more advanced textbooks on mechanics one may not find such a distinction: all the types of motion listed above are simply classified as a **free fall**. How comes? Well, as was said, a free fall is the motion of a body under the influence of gravitational forces **only**. And are there forces other than the "gravitational pull" involved in a vertical downward fall, and in the "projectile motion"? No, definitely not!

For obtaining a mathematical description of the motion of a free-falling body, in each case one has to do the same -- namely, to solve the so-called "equation of motion", which always has the familiar form: $F = m \cdot a$, i.e., it's nothing else that what we call the "Newton's Second Law of Dynamics". The only trick is that the force and the acceleration have to be taken in a **vector form**:

$$\vec{F} = m \cdot \vec{a} \tag{2.2.3.1}$$

In the most general case an object moves in a three-dimensional space. Hence, we write both
\)\vec{F}\) and $\vec{a}$ as three-component vectors:

$$(F_x, F_y, F_z) = m \cdot (a_x, a_y, a_z) \tag{2.2.3.2}$$

which can be rewritten as a set of three equations for individual components of the force and acceleration vectors:

$$\begin{aligned} F_x &= m \cdot a_x \\ F_y &= m \cdot a_y \\ F_z &= m \cdot a_z \end{aligned} \tag{2.2.3.3}$$

Now, recall that the acceleration, by definition, is the second time-derivative of the time-dependent position of the object. We can then rewrite the above set as[3]:

$$\begin{aligned} \frac{d^2x}{dt^2} &= \frac{F_x}{m} \\ \frac{d^2y}{dt^2} &= \frac{F_y}{m} \\ \frac{d^2z}{dt^2} &= \frac{F_z}{m} \end{aligned} \tag{2.2.3.4}$$

This is a set of three **differential equations**. You may not know how to solve such equations -- many students, who are interested in "Energy Alternatives", never take math courses at such level, and those who do, take it towards the end of their Sophomore year, or at the Junior year. So, why are the equations mentioned in this text? Well, to advertise how mighty they are! One can use the very same set of equations to solve a vertical fall, all types of projectile motions -- and this is not the end of the list: also, the motion of satellites, on circular orbits as well as on elongated elliptic orbits, the motion of planets, moons, asteroids and comets.

The very same set of equations for all problems listed![1] How can it be? It's simple, because all these types of motion have one thing in common -- the only force acting on the moving object is the force of gravity!

Presenting the solution procedures for specific problems listed above would take too much space in this chapter. However, the Author of this text does not like to make statements, and then ask: "Please accept without proof that the said statement is true". So, several solution procedures will be presented in detail, but in the Appendices. Yet, the Appendices are added to the text with the intention of satisfy the curiosity of students who like math -- but, definitely, reading them is not necessary for students whose goal is only to attain a good \textbf{conceptual} understanding of the course material.

---

1. It’s an average value – Earth is not a perfect sphere, it is somewhat “flattened” at the poles, so that the weight of a body may be slightly different, depending on the geographical region where it is measured; also, the spinning of Earth contributes to a gradual decrease of the body’s weight when it is moved from a pole towards the Equator. Bur the “weight loss” of a body between a pole and the Equator is only about 0.3%, so that with a good approximation one can neglect such effects.

2. \footnote{Earth is not a perfect sphere. The value of the radius we use here, 6371 km, is an average value. But the distance from the Earth center at different geographical regions may be slightly larger or slightly smaller than the average value, so that the actual values may slightly differ from the result obtained from the Eq. 2.15.

3. These equations are a powerful tool -- solving motions resulting from the action of gravity only is just one application, the equations are generally valid, no matter what's the origin of the

2.2.3: Energy of an Oscillator is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

2.2.4: Free Fall, Satellite Motion and the Mechanism of Tides

The formula for the force of gravity acting on a body, commonly referred to as "the body's weight", is:

$$F_{\mathrm{g}} = m \cdot g \tag{2.2.4.1}$$

with g = 9.81 m/s ^{2}, is valid only near the Earth's surface. At higher altitudes the value of g changes. Its dependence on altitude, $g(H)$, where H is the height over the Earth's surface (conventionally, measured from the sea level) can be readily found, if one invokes Newton's Law of Universal Gravitation, which states that two bodies of masses m_1 and m_2, R apart, attract one another with a gravitational force F_{g}, given by the equation:

$$F_{\mathrm{g}} = G\frac{m_1 m_2}{R^2} \tag{2.2.4.2}$$

where G is the universal gravitational constant: G = 6.674\)\times 10^{−11}\) N ·(m/kg)\)^2\). Then, the weight of an object of mass m positioned near the Earth's surface -- i.e., the magnitude of the force attracting it towards the Earth's center is given by:

$$F_{\mathrm{g}} = G\frac{m \cdot M_{\mathrm{e}}}{{R_{\mathrm{e}}}^2} \tag{2.2.4.3}$$

where $M_{\mathrm{e}} = 5.972 \times 10$^{24}\) kilograms is the Earth mass, and $R_{\mathrm{e}} = 6,371 kilometers$[1] is the Earth radius. By equating the right sides of Eqs.(11) and (13), we obtain a formula for the value of g near the earth surface[2]

$$g = G\frac{M_{\mathrm{e}}}{{R_{\mathrm{e}}}^2} = 9.8195 \text{ m/s}^2 \tag{2.2.4.4}$$

At altitude H the distance of an object from the Earth center is $R = R_{\mathrm{e}} + H$. So, for the value of g at this altitude, $g(H)$, we get:

$$g(H) = G\frac{M_{\mathrm{e}}}{R^2} = G\frac{M_{\mathrm{e}}}{(R_{\mathrm{e}} + H)^2} \tag{2.2.4.5}$$

2.2.4.1: Free Fall

2.2.4.2: Circular Motion

2.2.4.3: Projectile Motion- Very Long Range

2.2.4.4: Satellite Motion

2.2.4.5: Tidal Forces

2.2.4.6: Extreme Tidal Effects

2.2.4: Free Fall, Satellite Motion and the Mechanism of Tides is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.4.1: Free Fall

If we raise a body above the Earth surface at a distance H, and let it go -- then, no matter what the value of H is, one meter or one million meters, the body starts "falling freely" because it is attracted towards the Earth center by the force of gravity. And, if we forget about the air drag force (which is relevant only for H values of a few tens of kilometers), then we certainly have the right to say: "A free fall is the motion under the influence of the force of gravity, and no other force".

Now, suppose that the in initial velocity of the body considered is not zero, but we give it a "push" in the direction perpendicular to the line drawn between the initial body's position, and the Earth center. Let's call this line "the z direction", and the speed along this line as "the vertical velocity v_z". Let's call the direction perpendicular to the z line as"the x axis", and the speed along the x axis as "the horizontal velocity v_{x}". So, at the initial moment $v_{\mathrm{z}} = 0$, and $v_{\mathrm{x}} \neq 0$. What happens next? The body starts moving downwards, with constant acceleration g, under the influence of gravity. But there is no horizontal force acting on it, so the initial v_{y} remains unchanged. The resultant body's trajectory is a parabola, and it hits the Earth surface some distance from the point right below the initial body's position.

In introductory-level physics textbooks the type of motion described above is usually referred to as "projectile motion"}. Three different sub-types are discussed, with the initial "push" directed strictly horizontally, at some angle upwards, or at some angle downwards. Each sub-type has to be treated slightly differently. It's a good approach, instructive and pedagogical: it helps young students to develop an ability called "physical insight", in which a correct "conceptual understanding" of a physical phenomenon considered is combined with a correct choice of "mathematical tools" needed to analyze it in quantitative terms. However, in more advanced textbooks on mechanics one may not find such a distinction: all the types of motion listed above are simply classified as a **free fall**. How comes? Well, as was said, a free fall is the motion of a body under the influence of gravitational forces **only**. And are there forces other than the "gravitational pull" involved in a vertical downward fall, and in the "projectile motion"? No, definitely not!

For obtaining a mathematical description of the motion of a free-falling body, in each case one has to do the same -- namely, to solve the so-called "equation of motion", which always has the familiar form: $F = m \cdot a$, i.e., it's nothing else that what we call the "Newton's Second Law of Dynamics". The only trick is that the force and the acceleration have to be taken in a **vector form**:

$$\vec{F} = m \cdot \vec{a} \tag{2.2.4.1.1}$$

In the most general case an object moves in a three-dimensional space. Hence, we write both \)\vec{F}\) and $\vec{a}$ as three-component vectors:

$$(F_{\mathrm{x}}, F_{\mathrm{y}}, F_{\mathrm{z}}) = m \cdot (a_{\mathrm{x}}, a_{\mathrm{y}}, a_{\mathrm{z}}) \tag{2.2.4.1.2}$$

which can be rewritten as a set of three equations for individual components of the force and acceleration vectors:

$$\begin{aligned} F_{\mathrm{x}} &= m \cdot a_{\mathrm{x}} \\ F_{\mathrm{y}} &= m \cdot a_{\mathrm{y}} \\ F_{\mathrm{z}} &= m \cdot a_{\mathrm{z}} \end{aligned} \tag{2.2.4.1.3}$$

Now, recall that the acceleration, by definition, is the second time-derivative of the time-dependent position of the object. We can then rewrite the above set as[3]:

$$\begin{aligned} \frac{d^2x}{dt^2} &= \frac{F_{\mathrm{x}}}{m} \\ \frac{d^2y}{dt^2} &= \frac{F_{\mathrm{y}}}{m} \\ \frac{d^2z}{dt^2} &= \frac{F_{\mathrm{z}}}{m} \end{aligned} \tag{2.2.4.1.4}$$

This is a set of three **differential equations**. You may not know how to solve such equations -- many students, who are interested in "Energy Alternatives", never take math courses at such level, and those who do, take it towards the end of their Sophomore year, or at the Junior
year. So, why are the equations mentioned in this text? Well, to advertise how mighty they are! One can use the very same set of equations to solve a vertical fall, all types of projectile motions -- and this is not the end of the list: also, the motion of satellites, on circular orbits as well as on elongated elliptic orbits, the motion of planets, moons, asteroids and comets. The very same set of equations for all problems listed![1] How can it be? It's simple, because all these types of motion have one thing in common -- the only force acting on the moving object is the force of gravity!

Presenting the solution procedures for specific problems listed above would take too much space in this chapter. However, the Author of this text does not like to make statements, and then ask: "Please accept without proof that the said statement is true". So, several solution procedures will be presented in detail, but in the Appendices. Yet, the Appendices are added to the text with the intention of satisfy the curiosity of students who like math -- but, definitely, reading them is not necessary for students whose goal is only to attain a good \textbf{conceptual} understanding of the course material.

---

1. It’s an average value – Earth is not a perfect sphere, it is somewhat “flattened” at the poles, so that the weight of a body may be slightly different, depending on the geographical region where it is measured; also, the spinning of Earth contributes to a gradual decrease of the body’s weight when it is moved from a pole towards the Equator. Bur the “weight loss” of a body between a pole and the Equator is only about 0.3%, so that with a good approximation one can neglect such effects.

2. \footnote{Earth is not a perfect sphere. The value of the radius we use here, 6371 km, is an average value. But the distance from the Earth center at different geographical regions may be slightly larger or slightly smaller than the average value, so that the actual values may slightly differ from the result obtained from the Eq. 2.15.

3. These equations are a powerful tool -- solving motions resulting from the action of gravity only is just one application, the equations are generally valid, no matter what's the origin of the

2.2.4.1: Free Fall is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.4.2: Circular Motion

Consider an "object" -e.g., a stone or a ball -tied to a string, which is being swung in circles in horizontal plane (e.g., by holding the other end of the string above one's head -probably, everyone once did such an "experiment"). The object moves along a circular path. The string is tensioned, so, apparently, the object pulls it "outwards". Now, recall the Third Newton's Law of Dynamics -it says that if Object A exerts a force $\vec{F}_{A\to B}$ on Object B (called "action force"), then Object B exerts a "reaction force" $\vec{F}_{B\to A}$ on Object A, which is of the same magnitude and of opposite direction as the action force -i.e.,

$$\vec{F}_{B\to A} = -\vec{F}_{A\to B} \tag{2.2.4.2.1}$$

So, if the stone is tensioning the string "outwards", the string is pulling the stone "inwards", always towards the center of the circular path.

The First Newton's Law states that if there is no net force acting on an object, then it either rests or moves along a straight line with constant velocity -meaning that its acceleration is zero, $\vec{a} = 0$. The Second Law states that if there is a non-zero force, the object experiences acceleration that is proportional to the force, and inversely proportional to the object's mass:

$$\vec{a} = \frac{\vec{F}}{m}. \tag{2.2.4.2.2}$$

The "stone on the string", as we concluded above, is always pulled towards its circulat motion center by the force of the tensioned string. So, as follows from the above equation, it must experience an acceleration always directed towards the center. Is there indeed such an acceleration in circular motion? Let's examine such motion in closer detail in mathematical terms. Suppose that a small object -or, even better, a "point-like object" -of mass m is attached to a string of length R, and the other end of the string is fixed at the origin of a Cartesian system with axes X and Y. Let the object rotate in the XY plane. If T is the time it takes the object to make one full revolution, its speed along the circular path is $v = 2\pi R/T$.

Let's choose the instant at which the object intersects the X axis as $t = 0$. Then, at any time t the angle φ (expressed in radians) the string string makes with the X axis is

$$\varphi = \frac{2\pi}{T}t \tag{2.2.4.2.3}$$

The stretched string in the experiment can be thought of as a vector $\vec{R}$ connecting the origin of the XY coordinate frame with the circulating object (usually, it's referred to as the "radius vector"). The projections of $\vec{R}$ on the X and Y axes -call them R_x and R_y, respectively -are:

$$R_x = R\cos\varphi = R\cos\left(\frac{2\pi}{T}t\right) \quad \text{and} \quad R_y = R\sin\varphi = R\sin\left(\frac{2\pi}{T}t\right) \tag{2.2.4.2.4}$$

The expression $(2\pi/T)$ is commonly referred to as the "angular velocity" and is denoted as ω:

$$\omega = \frac{2\pi}{T} \tag{2.2.4.2.5}$$

Using the ω symbol, the two projections can be written in a more compact form, $R_x = R\cos\omega t$ and $R_y = R\sin\omega t$. These two expressions are the Cartesian components of the the vector $\vec{R}$. In vector algebra -in which one can perform various mathematical operations on vectors -a vector can be written as a list of its components, separated by commas, in parentheses\footnote{If $\vec{R}$ were a vector in a three-dimensional space, it would have a three components corresponding to projections on three Cartesian axes, X, Y, and Z, and it would be written then as a three-element list: $\vec{R} = (R_x,\ R_y,\ R_z)$.}

$$\vec{R}(t) = (R_x,\ R_y) = (R\cos\omega t, R\sin\omega t) = R(\cos\omega t, \sin\omega t) \tag{2.2.4.2.6}$$

(constant factors can be factored out). Note that the left-side term is written as $\vec{R}(t)$, because it should be kept in mind that this vector is a function of time. Now, we can calculate the velocity of the object, taking advantage of the fact that the velocity vector is a time derivative of the time-dependent position vector:

$$\vec{v} = \frac{d}{dt}\vec{R}(t) \tag{2.2.4.2.7}$$

For differentiating a vector, one uses a simple receipt -each component is individually differentiated, so we get:

$$\frac{d}{dt}\vec{R}(t) = \left(\frac{d}{dt}R_x(t),\ \frac{d}{dt}R_y(t)\right), \tag{2.2.4.2.8}$$

$$\frac{d}{dt}R_x(t) = \frac{d}{dt}R\cos\omega t = R\omega(-\sin\omega t) = -R\omega\sin\omega t, \tag{2.2.4.2.9}$$

$$\frac{d}{dt}R_y(t) = \frac{d}{dt}R\sin\omega t = R\omega(\cos\omega t) = R\omega\cos\omega t, \tag{2.2.4.2.10}$$

so, finally:

$$\vec{v} = (-R\omega\sin\omega t,\ R\omega\cos\omega t) = R\omega(-\sin\omega t,\ \cos\omega t). \tag{2.2.4.2.11}$$

Now, in a similar fashion, we can calculate the acceleration experienced by the circulating object -by definition, acceleration is a time derivative of the velocity:

$$\vec{a} = \frac{d}{dt}\vec{v}(t). \tag{2.2.4.2.12}$$

Using an analog procedure as the one done step-by-step when differentiating the $\vec{R}(t)$ vector, we obtain:

$$\vec{a} = \frac{d}{dt}\vec{v}(t) = R\omega\frac{d}{dt}(-\sin\omega t,\ \cos\omega t) = \omega^2 R(-\cos\omega t,\ -\sin\omega t), \tag{2.2.4.2.13}$$

which, taking into account the Eq. 2.24, can be rewritten in a final compact form:

$$\vec{a}(t) = \omega^2[-\vec{R}(t)]. \tag{2.2.4.2.14}$$

Note the minus sign -the result means that the object's acceleration vector is exactly parallel to the radius vector $\vec{R}$, but pointing in the **opposite direction** -i.e., towards the center of the object's circular trajectory. It's the famous **centripetal acceleration.**[2]

We have obtained the velocity and the acceleration **vectors**, but we haven't yet mention their **magnitudes**. In general, the magnitude of a "generic" vector, call it $\vec{G}$, is a square root of the sum of the squares of its Cartesian components, and the symbols commonly used for it are $|\vec{G}|$ (also called the vector's "modulus"), or simply the letter symbol alone, without the tiny arrow at the top. So, for a two-dimensional vector:

$$G \equiv |\vec{G}| = \sqrt{G_x^2 + G_y^2} \tag{2.2.4.2.15}$$

where the $\equiv$ symbol means "equivalent to" -and for a three-dimensional one:

$$G \equiv |\vec{G}| = \sqrt{G_x^2 + G_y^2 + G_z^2} \tag{2.2.4.2.16}$$

The radius vector $\vec{R}$ used in our considerations is simply a sector a connecting two points on the XY plane. Its magnitude is, of course, the same as its length:

$$\vec{R} = \sqrt{R_x^2 + R_y^2} = \sqrt{(R\cos\varphi)^2 + (R\sin\varphi)^2} = R\sqrt{\cos^2\varphi + \sin^2\varphi} = R \tag{2.2.4.2.17}$$

By the same token, one can find the magnitude of the velocity vector given by the Eq. 2.25:

$$v = |\vec{v}| = \omega R\sqrt{(-\sin\omega t)^2 + (\cos\omega t)^2} = \omega R\sqrt{\sin^2\omega t + \cos^2\omega t} = \omega R. \tag{2.2.4.2.18}$$

In combination with the Eq. 2.23 it yields:

$$v = \omega R = \frac{2\pi R}{T} \tag{2.2.4.2.19}$$

which is the expected result, considering that over a period of T the object travels the length of the circle, $2\pi R$. Usually, v is called the "linear velocity", to distinguish it from the angular velocity ω, or "tangent velocity", because the $\vec{v}$ vector is tangent to the circular trajectory.

Finally, in a similar way one can find that the magnitude of the centripetal acceleration be equal to $a = \omega^2 R$. Then, by multiplying this expression by another expression, which is usually called "a well-chosen unity" -namely, by $1 = R/R$, and by combining such product with the Eq. 2.27, we obtain an equivalent expression for a in terms of the linear velocity:

$$a = \omega^2 R = \omega^2 R \frac{R}{R} = \frac{(\omega R)^2}{R} = \frac{v^2}{R}. \tag{2.2.4.2.20}$$

The next step is to consider the \textbf{dynamics} of circular motion, i.e., the forces involved. For this purpose, we need to recall the Second Newton's Law of Dynamics, the mathematical form of which was given earlier as the Eq. 2.17 -let's look at it again:

$$\vec{F} = m\vec{a} \tag{2.2.4.2.21}$$

What this equation states, can be expressed by words as follows: "If an object of mass m is experiencing an acceleration $\vec{a}$, there must be a force $\vec{F}$ responsible for it, equal to $\vec{a}$ multiplied by m." Hence, in the circular motion considered, there must be a force acting on the object -considering the Eq. 2.26, we obtain:

$$\vec{F} = m\vec{a} = m\omega^2 (-R) \text{ and } \vec{F} = m\omega^2 R \tag{2.2.4.2.22}$$

The first of the above equations tells us that at any moment there is a force acting on the object, parallel to the radius-vector $\vec{r}$, and pointing towards the motion center. It can be called a \textbf{\textbf{centripetal force}}. Its origin can be easily identified -it's the force with which the stretched string pulls the object towards the motion center.

From the Newton's Third Law of Dynamics it follows that if the string pulls the object inwards, there must be a reaction force the same magnitude and opposite direction, with which the object pulls the string outwards. It is called the \textbf{centrifugal force}.

Students (and not only them) often confuse the centripetal force with the centrifugal force. There is much material related to that confusion in the Web, with many illustrative examples, and it is worth reading it -one can access many such sites simply by typing "centrifugal + centripetal + confusion" in one of the many available search engines. For instance, one can find instructive definitions in the "American Heritage Dictionary of the English Language"-namely:
Centripetal force: *The component of force acting on a body in curvilinear motion that is directed toward the center of curvature or axis of rotation*, and: *Centrifugal force: The apparent force, equal and opposite to the centripetal force, drawing a rotating body away from the center of rotation, caused by the inertia of the body*. These two definitions make it clear that only the centripetal force is a "real force" -i.e., the only force an observer placed in a non-rotating frame will see; whereas the centrifugal force is an apparent force which only an observer placed in the rotating frame would register.

In other words, the only force needed to make a body traveling along a circular route is the centripetal force. Telling the truth, the stone on a string is not the best example in support of that one can think of -because the string not only "transmits" the centripetal force, but it also partially limits the stone's freedom of motion, i.e., it acts as a \textbf{ constraint}.
The centrifugal force emerges from the interaction between the stone and the string.

Much more convincing is to consider a fully unconstrained motion -a good example is the motion of a body in a gravitational field, e.g., of a small artificial satellite of Earth. The only force acting on the satellite is the Earth's gravity, always pointing towards the Earth center. It's the centripetal force. So, does it mean that there is no centrifugal force? Not exactly: there is no such force anywhere nearby the satellite. But the Newton's Third Law is merciless, there always must be a reaction to any force. So, there is a reaction to the said centripetal force -it's the force the satellite exerts on Earth.

We will return to this issue in the "Satellite Motion" section.

1. To get convinced that the equation is true, please check: if, for instance, the object makes one quarter of a full turn, then $\varphi = 90^\circ = \pi/2$, and $t = \frac{1}{4}T$. Now, put the latter into the above equation and you get, indeed, that $\varphi = (2\pi/T) \times (T/4) = \pi/2$; you may make a similar check for any other fraction of a full turn.

2. Centripetal -from the Greek word meaning "tending to move towards a center".

2.2.4.2: Circular Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.4.3: Projectile Motion- Very Long Range

Suppose that a projectile is launched in the horizontal direction (call it X) from a high tower, with the initial velocity of v_x. The height of the tower is H. Find the equation of the projectile trajectory, and the spot it hits the ground. It's a standard problem in introductory level physics. The solution is pretty simple: we call the vertical direction the Z axis, we call the coordinates of the tower top as $x = 0$ and $z = 0$, and assume that the launching takes place a $t = 0$. So, as there is no force in the x direction, v_x does not change, and the x coordinate changes with time as $x(t) = v_x t$. Downwards, there is a motion with acceleration g, so that $z(t) = gt^2$. These two equations together constitute the so-called “parametric equations of the trajectory curve". If one prefers an alternative description of this curve, in the $z = f(x)$ form, one can readily obtain it, solving the first equation for time: $t = x/v_x$, and plugging it into the second, to obtain: $z = -(g/2v_x^2)x^2$, which is the equation of an inverted parabola. And as far as the spot where the projectile hits the ground is concerned, for this spot $z = -H$, so by solving the equation $-H = -(g/2v_x^2)x^2$, we obtain $x = v_x\sqrt{2H/g}$.

Earlier in the text, general equations for the free fall (Eqs. 1.20) were presented. For the case discussed above, we need to use only two of them, for x and z. There is no force in the x direction, and in the z direction the force is $F_z = -mg$. Therefore, we obtain a set of two differential equations:

$$dfracd^2xdt^2 = 0 \quad \text{and} \quad dfracd^2zdt^2 = -g \tag{2.2.4.3.1}$$

It can be readily checked by a pretty simple double differentiating that the two parametric solutions for $x(t)$ and $z(t)$ obtained above by the “introductory physics method" indeed do satisfy both those differential equations.

But will the introductory physics method always work?... Well, note that in such a solving procedure we used an additional “hidden assumption" -namely, that the Earth is flat. As long as the flight range of of the projectile is relatively short -say, not beyond the horizon visible from the ground level -the “flat Earth" is pretty good an approximation because the effects of the Earth curvature are still very small. But if the projectile can fly far beyond the horizon, the curvature can no longer be neglected. In WW II, the battleships' most powerful artillery pieces could fire on targets as far away as about 50 km (\)\tilde{30}\)) miles, and at such distance the target is already about 200 m (1/8 mile) below the imaginary “flat Earth level". And what if we wanted to calculate the trajectory and the range of a projectile with no propulsion, but with the initial speed sufficient to carry it over distances much, much longer than 50 km?

In such a case there is no other way than to use the general equations. If we put the origin of the XZ Carthesian system at the Earth center, then the calculation of the force of gravity at any point above the Earth surface of coordinates x and z is straightforward: the distance from the Earth center is $\sqrt{x^2 + z^2}$.

By replacing R_e in the Eq. 2.13 with this distance we get the total force of gravity acting on mass m at this point:

$$F(x, y) = GdfracM_e m\sqrt{x^2 + z^2}^2. \tag{2.2.4.3.2}$$

In the equations of motion we need not the total force, but its components in the x and z directions. To obtain their values and signs, we have to multiply the total force by, respectively, $-\cos\theta$ and $-\sin\theta$, where θ is the angle defined in Fig. XX. From this Figure, one can readily find that:

$$\cos\theta = dfracx\sqrt{x^2 + z^2} \quad \text{and} \quad \sin\theta = dfracz\sqrt{x^2 + z^2}. \tag{2.2.4.3.3}$$

By combining Eqs. 1.13, 1.30, and 1.31, we obtain generally valid equations of motion for a projectile launched from an arbitrary point above the surface of \textbf{spherical} Earth:

$$dfracd^2xdt^2 = -GM_e dfracx\sqrt{x^2 + z^2}^3 \quad \text{and} \quad dfracd^2zdt^2 = -GM_e dfracz\sqrt{x^2 + z^2}^3 \tag{2.2.4.3.4}$$

To obtain the trajectory of the projectile, one needs to specify the so-called “initial conditions," i.e., the x_0 and z_0 coordinates of the launching point, and the velocity components v_{x0} and v_{z0} at the moment of launching. And then, find the time-dependent coordinates $x(t)$ and $z(t)$, which together constitute a set of parametric equations of the trajectory.

The bad news is that solving these equations using the classical “paper+pencil" method requires applying sophisticated high-level mathematical techniques. Presenting a step-by-step solution procedure in this chapter would make no sense. However, the good news is that a highly accurate solution can be obtained by using a relatively simple numerical algorithm which takes even a simple today's portable computer only a split second to process. This algorithm and a program in the PYTHON language are described in a greater detail in APPENDIX XX. Here, we will only show a few calculated trajectories in a graphic form in a figure.

The calculations are done for realistic data, i.e., the Earth radius is taken as 6371 km, the Earth's mass as 5.9722×10^{24} kg, and $G = 6.6741 \times 10^{-11}$ m\)^3\)kg\)^{-1}\)s\)^{-2}\). If we wanted to make a plot displaying everything in real proportions, the launching point could not be too low, because the plotted trajectory would “blend" with the circle representing the planet's contour. Therefore, the projectile is launched from the altitude of 1000 km -you can either thing of a super-giant tower built at the North Pole, or a spacecraft “hovering" over the North Pole at the altitude of 1000 km, and acting as a platform for a horizontal launch. The value of g at this altitude is markedly lower than at the Earth surface -from the Eq. 2.16, for $H = 1000$ km one gets $g(H) = 7.3362$ m/s\)^2\).

So, the initial coordinates of the launching point are: $x_0 = 0$ and $z_0 = 6371$ km + 1000 km = 7371 000 m. The projectile will always be launched in the horizontal direction, so that always v_{z0} -and we will make calculations for several different values of v_{x0}.

In the Fig. 2.1 one can see a trajectory for the initial projectile velocity $v_{x0} = 6000$ m/s, plotted with black color.[1] The projectile lands at a spot which is about 1/8 of the Earth's circumference from the base of the “launching tower". For $v_{x0} = 6850$ m/s (blue curve) the landing spot is close to the Equator. Now, the range becomes quite sensitive for the initial speed value: for $v_{x0} = 7081$ m/s the projectile lands close to the South Pole. And the most interesting trajectory is definitely that for $v_{x0} = 7353.6$ m/s (red color): the trajectory is a perfect circle, it never hits the ground -the projectile returns precisely to the spot at the top of the launching tower from which it was launched!

In the next section it is explained what is the significance of the “7353.6 m/s" speed. The trajectories for higher launching velocities have one in common -they all return to the launching spot, but their shapes get elongated in the “downwards" direction in the figure (one example plotted with a dashed line is a trajectory for $v_{x0} = 8000$ m/s). They are no longer circles, but ellipses.

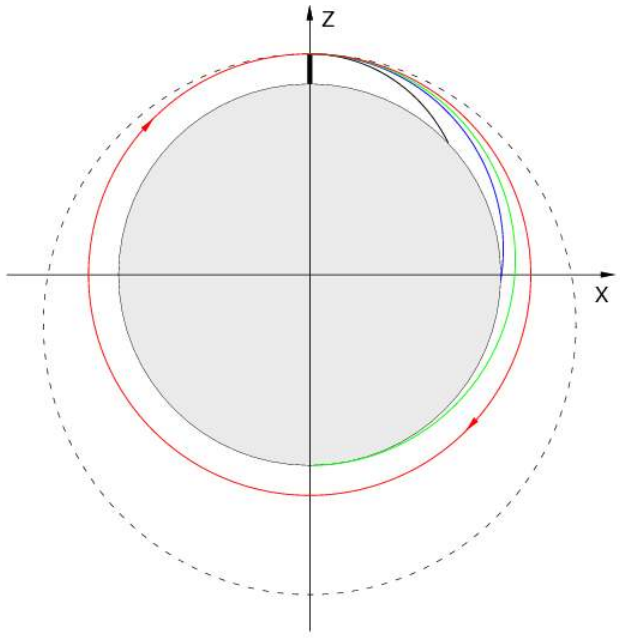


Figure 2.2.4.3.1: The shaded circle symbolizes the Eart, and the sphere bar at its top is a 1000 kilometer high tower erected at the North Pole. The black, blue and green curves are trajectories of projectiles launched horizontally from the tower top with velocities, respectively, 6000, 6850, and 7081 meters per second. They all fall to the ground. But the projectile launched with a velocity of 7353 m/s (red curve) becomes an artificial satellite of Earth -it travels along a trajectory which is a perfect circle and returns to the tower top. A projectile launched with a velocity of 8000 m/s travels along an elliptical trajectory (dashed curve).

However, for $v_{x0} = \sqrt{2} \times 7353.6$ m/s = 10399.6 m/s and higher launching velocities there is yet another change of behavior - the trajectories are no longer closed, but open curves -meaning that the projectiles never return to the launching site, but they “leave Earth for good".

In summary, there are four different types of behavior of the projectile. For zero launching speed, which is obvious, it's just a vertical fall. Then, for $v_{x0} > 0$ the trajectories are first “ballistic curves" ending at the Earth surface. Next, the paths become closed orbits, and the projectiles are now Earth's satellites -and finally, for v_{x0} larger than the so-called “escape velocity", the projectile becomes an “artificial asteroid" and leaves Earth forever.

And it should be stressed that all these four types of behavior have on thing in common - namely, in all cases the \textbf{only} force acting on the projectile is the force of gravity.

Everybody agrees that an appropriate term for a purely vertical drop is “a free fall". Yet, many students react at the first moment th surprise if the same name is used for a ballistic flight, satellite motion, or a spacecraft flight. But physicists insist that as long as there is no engine in the moving object, and it is only the gravity which governs its motion, one should rather invoke the famous “Occam Razor" and use the same term in each case. One more argument in favor of such thinking is that all types of trajectories are obtained as solutions of the very same set of equations of motion that are the mathematical form of the Second Newton's Law of Dynamics.

---

1. 6000 m/s is about 17 times the speed of sound, or about 6 times the speed of a bullet from a powerful military rifle.

2.2.4.3: Projectile Motion- Very Long Range is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.4.4: Satellite Motion

In this subsection special attention will be paid to the perfectly circular trajectory of the projectile discussed in the preceding. Note that the value of acceleration due to gravity at the launching platform 1000 km above the Earth surface, as obtained from Eq. 2.16, is $g = 7.3362\ m/s^2$. It has been found that the projectile flies along a perfectly circular orbit when it's launched with a speed of 7353.6 m/s. The radius of the orbit is 6371 km + 1000 km = 7371,000 m. Let's then ask what is the centripetal acceleration a_{cp} in circular motion which such parameters? From the Eq. 2.28 we get:

$$a_{cp} = \frac{(7353.6\ m/s)^2}{7371,000\ m} = 7.3362\ m/s^2 \tag{2.2.4.4.1}$$

Note: Exactly the same as the g value 7371 km from the Earth center.

So, we can conclude: in a flight along a circular path of radius 7371 km with a speed of 7353.6 m/s an object experiences a centripetal acceleration of $a_{cp} = 7.3362\ m/s^2$ -or, one can say it in a different but equivalent way: if the projectile's mass is m, such a motion has to be maintained by a \textbf{centripetal force} of

$$F_{\mathrm{cp}} = m \times a_{\mathrm{cp}} = m \times 7.3362\ \mathrm{m/s^2} \tag{2.2.4.4.2}$$

And this is exactly the force supplied by gravity on this orbit!

And if we want to have a satellite on circular orbit at a different altitude? The rule is the same for all circular orbits: the centripetal acceleration must be equal to acceleration due to gravity at the orbit's altitude, from which we can find the linear speed needed for maintaining a satellite at any altitude H:

$$\frac{v^2}{R_{\mathrm{e}} + H} = \frac{GM_{\mathrm{e}}}{(R_{\mathrm{e}} + H)^2} \quad \Rightarrow \quad v(H) = \sqrt{\frac{GM_{\mathrm{e}}}{(R_{\mathrm{e}} + H)}} \tag{2.2.4.4.3}$$

If the Earth's surface were perfectly smooth, and there were no atmosphere, a satellite could fly at a nearly-zero altitude. The speed of such satellite should be:

$$v(H = 0) = \sqrt{\frac{GM_{\mathrm{e}}}{R_{\mathrm{e}}}} = \sqrt{g(H = 0) \times R_{\mathrm{e}}} = 7910\ \mathrm{m/s} \tag{2.2.4.4.4}$$

This velocity is commonly referred to as “the first cosmic velocity".

As follows from the discussions in the preceding section and this section, any kind of motion in the gravitational field of Earth, as well as of another planet, moon, or the Sun can be fully described by taking into consideration \textbf{just one} force -the force of gravity -and
solving the equations of motion based on the Second Law of Dynamics, $\vec{F} = m\vec{a}$. There is no need to introduce any other factors. It should be noted that this is exactly the method professional astronomers, including NASA specialists, use for analyzing/predicting the motion of all celestial bodies in the Solar System. The only complication in such calculations is that the total force acting on one object is usually a vector sum of several forces, each of which which corresponds to attractive interaction with another celestial body. For instance, for calculating the motion of Moon -the Earth's natural satellite -relative to Earth, one should take into consideration the attractive forces exerted on Moon both by Earth and Sun. Because its enormous mass, the Sun can be assumed as a “stationary body" -this is the so-called “heliocentric model" -but Earth has to be treated as a body moving under the influence of **two** forces, the attraction by Sun **and** the attraction by Moon. The equations-of-motion for this three-body system are:

$$\begin{aligned} m_m\vec{a}_m &= \vec{F}_{m\to e} + \vec{F}_{m\to s} \\ m_e\vec{a}_e &= \vec{F}_{e\to m} + \vec{F}_{e\to s} \end{aligned} \tag{2.2.4.4.5}$$

where $\vec{F}_{m\to e}$ means: “force attracting Moon towards Earth", and similarly for other symbols. Only such a model is able to reproduce correctly the motion of Moon as seen from Earth.[1]

In the case of small artificial satellites, due to their negligibly small masses compared with that of Earth, a simpler math, such as the Eqs. 1.32, is already sufficient for analyzing their orbits. However, even such math may look scary for students who are at early stages of their education. Therefore, instructors and textbooks often choose to present an explanation of satellite motion that is surely much clearer conceptually, and needs only some pretty simple math. Namely, it says -the motion along a circular orbit produces an outwards-pointing \textbf{centrifugal} force, which exactly compensates the inwards-pointing force of gravity.

Because of such compensation, the explanation goes on, nothing is pulling the satellite out of its orbit. This reasoning is backed by math: the centrifugal force, mv^2/r, is equated with GmM/r^2, which leads to the correct result: $v = \sqrt{GM/r^2}$,essentially the same result as has been obtained in the Eq. 2.33.

However, the fact that a theory yields a correct results does not mean automatically that it is itself correct. There are some elements of doubtful validity in the reasoning -the Newton's First Law of Dynamics states that if all forces acting on a body are compensated, then the body may move, yes, but along a straight line. So how comes that moves along a circular route? Another weakness is that the explanation is based on the concept of centrifugal force. We all are very well familiar with this force, we can experience it in a car taking a turn, on a carousel, or swinging a stone tied to a string overhead. Yet, it is not a real force, it is the force of inertia that emerges in a rotating frame as a response to the centripetal force delivered by the car seat or the carousel seat to our body, or conveyed from our hand to the stone through the string. And, notice, in an empty space where the satellites move there no such thing as a "frame" of any kind.

That simple explanation can be fixed by adding some “extra" elements -which is usually not done by instructors, though. The extra element is abstract, it is an imaginary rotating frame -sort of an invisible string tied one end to the satellite, the other end to the center of the circular orbit -and the tension force in the string equal to the force of gravity acting on the stone. The imaginary rotating frame does not exist, of course -but if it did, the motion of the satellite would be the same as it is without it. So, the model works.

Yet, such a “fixed" model has one major disadvantage: it does work, right, but only in the case of a perfectly circular orbit. But try to use the same reasoning for a satellite on an elliptic orbit. For any point at a continuous curve, there exists an osculating circle -i.e., a circle that touches the curve at this point and has the same tangent and curvature at that point. The radius of such circle is the radius of curvature of the curve at that point. So, for a satellite moving along an elliptic path, if its velocity at a given point is v, and the radius of curvature at this point is r, one can find the “centrifugal force" at this very point equal to mv^2/r. The problem is, however, that the force is directed towards the center of the osculating circle -which, in general, is not the same direction as of the force of gravity vector at the given point. So, the two forces can not compensate one another! So, what happens next to the satellite?

Again, the model can be fixed and used for calculating the parameters of the elliptical orbit -but now the “extras" needed for it are far more complicated than in the case of a circular orbit.

---

1. Even though the equations, as written in this text, may look pretty simple -solving them is definitely not. It took astronomers nearly 200 years to develop sophisticated mathematical tools that enabled them to obtain, using “paper and pencil", satisfactory predictions of the Moon's position for the next tens of years. It should be stressed that the motions of Moon are pretty complicated -one manifestation of that are the highly irregular patterns of the solar and lunar eclipses. But the good news is that today high quality solutions of the equations discussed here can be obtained in minutes using computational methods.}

2.2.4.4: Satellite Motion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.4.5: Tidal Forces

To summarize the considerations in the above four subsection: (i) it was explained that not only a vertical drop in gravitational field, but \textit{any} motion of a body under the influence of the force of gravity alone is classified by physics as a free fall -and (ii), that for the very same reason the satellite motion is nothing else than a special case of the free fall. Why so much attention has been devoted to these issues?, and to showing that satellite motion is just a special case of free fall? Because the next task is to explain the mechanism ocean tides -and it can hardly be done unless the readers has a good understanding of the mechanics of motion of satellites and planets (note that planets are also satellites - of the central star in the planetary system).

The thing is that there are plenty of misconceptions about ocean tides. The explanation of the mechanism responsible for tides given in some college physics textbooks is partially incorrect, and in some it is even altogether incorrect. In Internet there is a number of articles discussing such misconceptions -for example, a thorough discussion presenting several "wrong scenarios" given in various textbooks can be found in this site.

Most high-school students, or even elementary school students already know that the tides are caused by the the Moon gravity -which is correct. Some of them even know that there is also some weaker contribution from Sun, which is also true. But it's not enough to say only that in a college textbook -so the authors feel obliged to present a more detailed physical picture. And as is commonly known, "the devil is in the details". And this is exactly what happens.

Bedeviled are the **two** tidal bulges that form on the oceans. One is on the side of the Earth facing the Moon. Here the explanations are "less bedeviled", saying that the bulge forms because the water in this bulge is closer to the Moon than the Earth's mass center, so it's attracted towards the Moon by a stronger force than that acting on the Earth's bulk. Earth is rocky and stiff, whereas water can flow - therefore it flows towards the stronger gravity and form the bulge.

Quite embarrassing, though, is the formation of a similar bulge on the other side of Earth, the "far side". On the same token as before, here the Moon's gravity acting on things is **weaker** than that acting on the Earth's bulk. So, the fluid water should flow away and one would expect the formation of a "dip" rather than a bulge. It is quite common that for the formation of the other bulge authors blame "the force of inertia". Usually not going far into details -because the devil is in them, as has been said. In fact, it may be really difficult to explain why two different forces produce such a nearly-symmetric effect?

Preface to the mathematical argument

"Ocean tides can be a source of considerable amounts of energy. As is known, tidal phenomena manifest themselves as the formation of two "tidal bulges" -- one of which is under the Moon, the other on the opposite side of the globe. The fact that the latter is formed, although there is no Moon above it, is often explained in a not entirely correct way. In the text below it is shown, by a correct interpretation of Newtonian physics, that both of these "bulges" are formed as a result of the action of one and the same force, known as the "tidal force". However, a thorough explanation of the physics underlying the creation of "the two bulges" requires a series of steps based on mathematical formulas. The author therefore advises the reader, on first reading, to skip the several pages below, to the red "marker" above the text in the box -- and to read only the boxed text and the figures thereafter, which summarize the entire mathematical argument."

The truth is that the two bulges are not produced by two different forces -there is a **single force** responsible for all tidal effects. Astronomers call it the "tidal force". Let's elaborate on the details. Consider a body of mass m and of small size experiencing gravitational attraction from another body of mass M, r apart. The force of attraction, as we know, is given by the Newton's Law of Universal Gravitation:

$$F_{m\to M} = G\frac{mM}{r^2}. \qquad (2.2.4.5.1)$$

Now, suppose that the body of mass m cannot be thought of as a "point-like object", but is of finite size -call it D. A closer analysis of the forces acting on m shows that in addition to $F_{m\to M}$, another force emerges -one which tends to "stretch" the body, along the direction towards the mass M. This force, called the "tidal force", is proportional to D, and inversely proportional to r^3:

$$F_{tidal} \propto \frac{D}{r^3} \quad or \quad F_{tidal} = k\frac{D}{r^3}, \tag{2.2.4.5.2}$$

where k is a certain coefficient depending on the body's shape. Anyway, no matter what the value of k is, because of the force's dependence on R^{-3}, it grows quite rapidly when the distance between m and M decreases. Ocean tides on Earth are not the only known manifestation of the existence of tidal forces. Many other are observed in the Solar System. For instance, powerful tidal forces may occur in comets passing close to Jupiter. Comets are bodies consisting mostly of "dirty ice", which is a relatively soft material -in 1992, a Shoemaker-Levy Comet was ripped apart by tidal forces into several smaller parts during a close approach to Jupiter. The famous rings of Saturn are believed to made of debris of a moon that had been orbiting to close to the planet and was "pulverized" by tidal forces.

As noted, the k coefficient depends on the shape of the body exposed to tidal forces -for a sphere it is different than for the cigar-shaped Oumuamua 750-by-115 foot object, a "visitor" from the interstellar space that zipped through the Solar System at the end of the 2017 year.

In order to explain the origin of the tidal force, it may be instructive to calculate k for a simple object -we choose one whose shape is not necessarily similar to one of the existing celestial bodies, but such that it makes the calculations relatively straightforward. Namely, let's consider a spacecraft orbiting a planet of mass M along a circular trajectory of radius r, measured from the planet's center. On board, there is a small spherical object of mass m. Due to the weightlessness inside an orbiting spacecraft, we can imagine that the spherical object is suspended in mid-air inside the craft, so we can assume it orbits the planet independently, also moving on a circular orbit of radius
r (see Fig. 2.2, plot (a)).

This force -as has been discussed earlier -enforces the satellite motion, i.e., the free fall motion of the spherical object. The acceleration due to the force of gravity F_{gr} acting on the object is:

$$a_{gr} = \frac{F_{gr}}{m} = G\frac{M}{r^2} \tag{2.2.4.5.3}$$

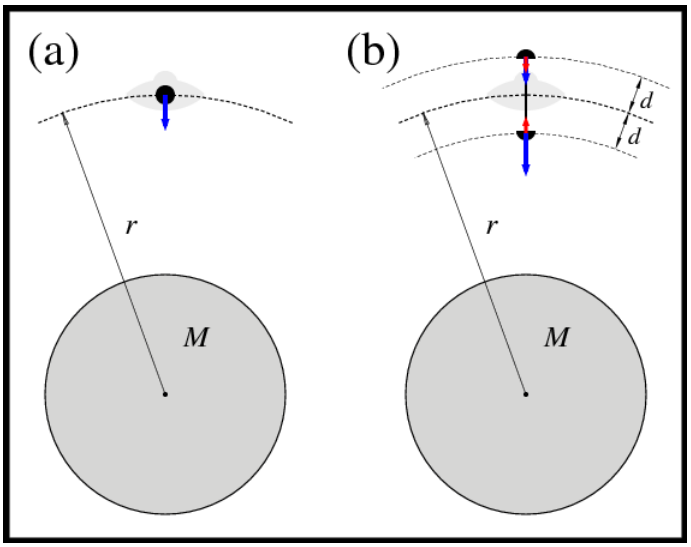


Figure 2.2.4.5.1:(a): A spherical object of mass m is inside a spacecraft orbiting a planet, but one can assume that it orbits the planet independently. (b): The ET-s manning (?) the spacecraft cut the object into two hemispheres of mass $m/2$ each, and attach each one to one end of a long spring. Then they place one hemisphere at a distance of d above the spaceship, and the other at the same distance under it. They connect the other ends of the springs with each other. After that, the spacecraft flies away, leaving the "object" consisting of the two hemispheres connected by a spring of a $2d$ length (shown as a thin line in the figure -the "coils" of the spring can be seen in the Fig. 2.4, which shows the "object" a blown-up scale). The mass center of the "object" is still on the original spacecraft's orbit of radius r So, now each hemisphere orbits the planet with the same angular velocity ω as the craft had, but one moves along a trajectory of radius $r-d$, and the other along a trajectory with radius $r+d$. Because the "lower" hemisphere is closer to the planet than it originally was, the force of gravity acting on it (blue vector) increases -more than needed to hold it on a $r-d$ orbit, tending to pull it down. This pull is compensated by the force of tension of the spring (the red vector). For the "upper" hemisphere, in turn, the force of gravity on the upper orbit is too weak to hold it on this orbit -the hemisphere would "escape upwards", if not pulled down by the spring's tension. So, effectively, one of the $m/2$ masses pulls its spring's end down, the other pulls its spring's end up -producing a tension force in the spring. And this what we call the *tidal force*

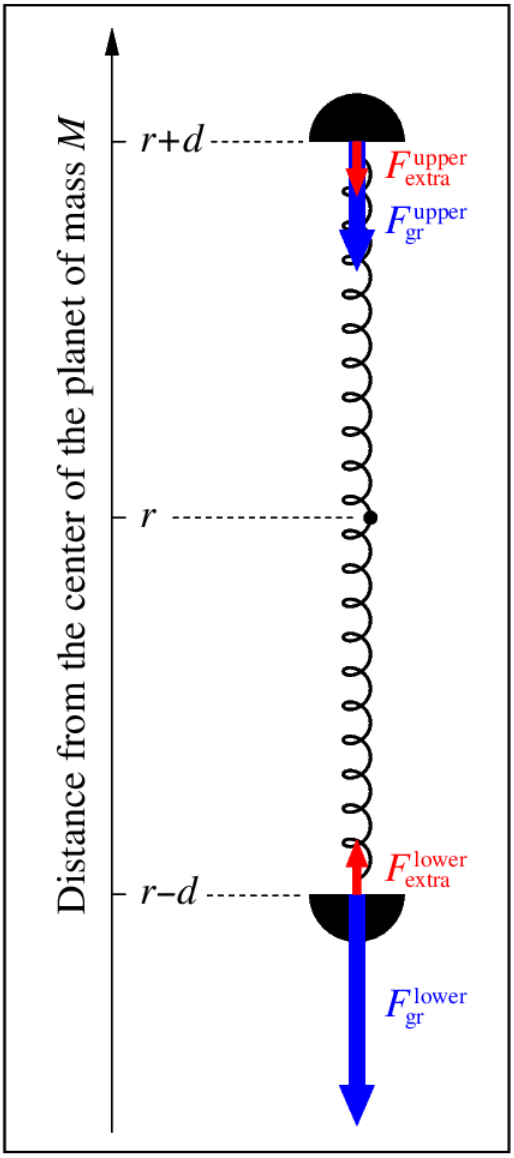


Figure 2.2.4.5.2: The "object" consisting of the two hemispheres and the spring shown here in a larger scale. The black dot in the middle of the spring marks the position of the "object's" mass center.

Now, suppose that the spacecraft crew -not necessarily humans, they may be intelligent ET (extra-terrestrial) beings as well -cut the spherical object into two hemispherical halves of mass $m/2$ each. The one part is attached to a long spring, and lowered from the spacecraft by a distance of d. The other hemispherical part is attached to an identical spring, then raised, and positioned at a distance d above the spacecraft. Suppose that the other ends of the springs are connected, forming a single spring of length $2d$.

The spacecraft is not needed any more, let it fly away. Now the system of two hemispheres, $2d$ apart, connected by the spring -as shown in the Fig. 2.3 - can be thought of a single elongated "object" of mass m (we assume that the mass of the spring is so small that it can be neglected). The object's *mass center* is still at a distance r from the planet center, so that the *net force* of gravity acting on the object is still the F_{gr} given by the Eq. 2.34, and the acceleration of the elongated object due to this force is the a_{gr} given by the Eq. 2.35.

Let's now focus our attention on the "upper" hemisphere. Its mass is $m/2$, and its distance from the planet center is $r + d$, so the force of gravity acting on it -call it F_{gr}^{upper} -is:

$$F_{gr}^{upper} = G\frac{\frac{m}{2}\cdot M}{(r+d)^2} \tag{2.2.4.5.4}$$

The acceleration a_{gr}^{upper} of mass $m/2$ due to such force would be:

$$a_{gr}^{upper} = \frac{F_{gr}^{upper}}{\frac{m}{2}} = G\frac{M}{(r+d)^2} \tag{2.2.4.5.5}$$

But:

$$G\frac{M}{(r+d)^2} < G\frac{M}{r^2}, \quad so\ that \quad a_{gr}^{upper} < a_{gr}, \tag{2.2.4.5.6}$$

which means that the whole "object" is falling with acceleration a_{gr}, yet the force of gravity F_{gr}^{upper} alone acting on the upper hemisphere is not strong enough for producing such acceleration. There must be an "extra" force F_{extra}^{upper} augmenting the downward pull, such that:

$$\frac{F_{gr}^{upper} + F_{extra}^{upper}}{\frac{m}{2}} = a_{gr} \tag{2.2.4.5.7}$$

From which one obtains

$$F_{extra}^{upper} = \frac{m}{2}\cdot a_{gr} F_{gr}^{upper} = G\frac{\frac{m}{2}\cdot M}{r^2} G\frac{\frac{m}{2}\cdot M}{(r+d)^2}, \tag{2.2.4.5.8}$$

which can be written in a more compact form:

$$F_{extra}^{upper} = G\frac{m\cdot M}{2}\left(\frac{1}{r^2}\frac{1}{(r+d)^2}\right) \tag{2.2.4.5.9}$$

Now, let's switch our attention to the "lower" hemisphere. The force of gravity acting on it -call it F_{gr}^{lower} -is:

$$F_{gr}^{lower} = G\frac{\frac{m}{2}\cdot M}{(r-d)^2} \tag{2.2.4.5.10}$$

The acceleration a_{gr}^{lower} of mass $m/2$ due to such force would be:

$$a_{gr}^{lower} = \frac{F_{gr}^{upper}}{\frac{m}{2}} = G\frac{M}{(r-d)^2} \tag{2.2.4.5.11}$$

But:

$$G\frac{M}{(r-d)^2} > G\frac{M}{r^2}, \quad so\ that \quad a_{gr}^{lower} < a_{gr}, \tag{2.2.4.5.12}$$

which means that the force $F^{lower}gr$ acting alone on the "lower" hemisphere would result in a higher downward acceleration than that of the object -therefore, to mitigate the F_{gr}^{lower} force and make the "lower" hemisphere to fall with acceleration a_{gr},
there must be an "extra" force F_{extra}^{lower} pulling the "lower" hemisphere in the upward direction, such that:

$$\frac{F_{gr}^{lower} F_{extra}^{lower}}{\frac{m}{2}} = a_{gr} \tag{2.2.4.5.13}$$

From which we obtain:

$$F_{extra}^{lower} = F_{gr}^{lower}\frac{m}{2}\cdot a_{gr} = G\frac{\frac{m}{2}\cdot M}{(r-d)^2} G\frac{\frac{m}{2}\cdot M}{r^2}, \tag{2.2.4.5.14}$$

which, again, can be written in a more compact form:

$$F_{extra}^{lower} = G\frac{m \cdot M}{2}\left(\frac{1}{(r-d)^2}\frac{1}{r^2}\right) \tag{2.2.4.5.15}$$

The Eqs. 2.40 and 2.45 are not yet the final forms that we need. In order to simplify them even further, let's introduce an auxiliary parameter, $\delta = d/r$. Doing some simple algebra, we get:

$$F_{extra}^{upper} = G\frac{m \cdot M}{2}\left(\frac{1}{r^2}\frac{1}{r^2(1+\frac{d}{r})^2}\right) = G\frac{m \cdot M}{2r^2}\left(1\frac{1}{(1+\delta)^2}\right) \tag{2.2.4.5.16}$$

and

$$F_{extra}^{lower} = G\frac{m \cdot M}{2}\left(\frac{1}{r^2(1-\frac{d}{r})^2} - \frac{1}{r^2}\right) = G\frac{m \cdot M}{2r^2}\left(\frac{1}{(1-\delta)^2}1\right) \tag{2.2.4.5.17}$$

To move further on, we would need to use a small calculus-based trick -in the case the Reader does not remember it, we present the "recipe" as a "Mathematical Detour":

****** Mathematical Detour ******

Any continuous function $f(x)$ can be expanded into a p*ower series*, also known as the *Taylor series* -which is a sum of the form:

$$\begin{aligned} f(x) = f(0) \quad &+ \qquad f'(0)\cdot x + \frac{1}{2}f''(0)\cdot x^2 + \frac{1}{2\cdot 3}f'''(0)\cdot x^3 \\ &+ \quad \frac{1}{2\cdot 3\cdot 4}f''''(0)\cdot x^4 + \frac{1}{2\cdot 3\cdot 4\cdot 5}f'''''(0)\cdot x^5 + ... \end{aligned}$$

where $f'(0)$, $f''(0)$, $f'''(0)$,... are the values of the first, second, third..... , and so on, derivatives of the $f(x)$ function for $x = 0$.

We want to apply such expansions to the functions $1/(1+\delta)^2$ and $1/(1-\delta)^2$ in the Eq. 2.46 and the Eq. 2.47, respectively.

So, for the first of these functions:

$$f(x) = \frac{1}{(1+x)^2} = (1+x)^{-2}; \quad f'(x) = (-2)\cdot(1+x)^{-3}; \tag{2.2.4.5.18}$$

$$f''(x) = (-2)\cdot(-3)\cdot(1+x)^{-4}; \quad f'''(x) = 6\cdot(-4)\cdot(1+x)^{-5}; \tag{2.2.4.5.19}$$

$$f''''(x) = (-5)\cdot(-24)\cdot(1+x)^{-6}; \quad \textit{and so on.} \tag{2.2.4.5.20}$$

Accordingly:

$$\begin{gathered} f(0) = 1; \;\; f'(0) = -2; \;\; f''(0) = 6; \;\; f'''(0) = -24; \\ f''''(0) = 120; \text{ and so on.} \end{gathered} \tag{2.2.4.5.21}$$

Inserting the above numbers into the Eq. 2.48, one obtains:

$$\frac{1}{(1+x)^2} = 12x + 3x^2 4x^3 + 5x^4 - ... \tag{2.2.4.5.22}$$

Next, we use the same recipe for the $1/(1-\delta)^2$ function:

$$f(x) = \frac{1}{(1x)^2} = (1x)^{-2}; \quad f'(x) = (-2)\cdot(-1)\cdot(1-x)^{-3}; \tag{2.2.4.5.23}$$

$$\begin{gathered} f''(x) = (-2)\cdot(-3)\cdot(-1)\cdot(1-x)^{-4}; \\ f'''(x) = 6\cdot(-4)\cdot(-1)\cdot(1-x)^{-5}; \end{gathered} \tag{2.2.4.5.24}$$

$$f''''(x) = (-5)\cdot(-24)\cdot(-1)\cdot(1-x)^{-6}; \quad \textit{and so on.} \tag{2.2.4.5.25}$$

Now, the $f(0)$ value and the derivative values we need are:

$$f(0) = 1; \;\; f'(0) = 2; \;\; f''(0) = -6; \;\; f'''(0) = 24; \\ f''''(0) = -120; \text{ and so on.} \tag{2.2.4.5.26}$$

Hence:

$$\frac{1}{(1-x)^2} = 1 + 2x3x^2 + 4x^3 5x^4 + ... \tag{2.2.4.5.27}$$

****** End of the Mathematical Detour ******,

Based on the Eqs. 2.29 and 2.50, we can write:

$$\frac{1}{(1+\delta)^2} = 12\delta + 3\delta^2 4\delta^3 + 5\delta^4 - ... \tag{2.2.4.5.28}$$

and

$$\frac{1}{(1-\delta)^2} = 1 + 2\delta 3\delta^2 + 4\delta^3 5\delta^4 - ... \tag{2.2.4.5.29}$$

Now, let's think for a moment about the $\delta = d/r$ value. It is natural to assume that r is much larger than d, so that the d/r ratio is much smaller than 1. A symbol for "much smaller" is "\)\ll\)", so that $\delta \ll 1$.

Well, if $\delta \ll 1$, then, obviously, $\delta^2 \ll \delta$, $\delta^3 \ll \delta^2$, and so on. Therefore, the omission of all squares and higher powers of δ will lead only to insignificant changes in the values of the right-hand sums in the two above equations. We can conclude that, to a very good approximation:

$$\frac{1}{(1+\delta)^2} \approx 12\delta \tag{2.2.4.5.30}$$

and

$$\frac{1}{(1-\delta)^2} \approx 1 + 2\delta \tag{2.2.4.5.31}$$

Substitution of the Eqs. 2.53 and 2.54, to the Eqs. 2.46 and 2.47, respectively, yields:

$$F_{extra}^{upper} = G\frac{m\cdot M}{2r^2}\left(1\frac{1}{(1+\delta)^2}\right) = G\frac{m\cdot M}{2r^2}(11+2\delta) = G\frac{m\cdot M}{r^2}\delta \tag{2.2.4.5.32}$$

and

$$F_{extra}^{lower} = G\frac{m\cdot M}{2r^2}\left(\frac{1}{(1-\delta)^2}1\right) = G\frac{m\cdot M}{2r^2}(1+2\delta 1) = G\frac{m\cdot M}{r^2}\delta \tag{2.2.4.5.33}$$

But, as was defined, $\delta = d/r$, and by substituting this to the two above expressions, we obtain the final forms:

$$F_{extra}^{upper} = G\frac{m\cdot M}{r^3}d \quad \textit{and} \quad F_{extra}^{lower} = G\frac{m\cdot M}{r^3}d \tag{2.2.4.5.34}$$

It turns out that both "extra" forces are of the same *magnitude* (we cannot say that they are "equal", because the F_{extra}^{upper} vector is oriented down, and the F_{extra}^{lower} vector is oriented up). What is the origin of the two forces? Clearly, they are delivered by the stretched spring of $2d$ length (if you grab a spring, right hand by one end, left hand by the other end, and you stretch it -each end will pull each hand "inwards", but each hand pulls the spring "outwards", with a force of the same magnitude -which is, of course, a

manifestation of the Newton's Third Law of Dynamics). So, conversely, there is ***a force stretching the object***, originating from the fact that the gravitational field which governs the motion of the object is not uniform -it exerts a stronger pull on one of the object's end than on the other end.

> *In conclusion, the above analysis shows that an object of mass m and of finite size moving in the gravitational field of another object of mass M,* ***always experiences a force, tending to stretch it along the line connecting the mass centers of the two bodies****. The force is proportional to the $m \cdot M$ product, and to the object size - and inversely proportional to the \textit{\textbf{cube}} of the distance between the two objects. It's the famous "tidal force", as it is called by astronomers.*

For a "dumbbell-shaped" object consisting of two $m/2$ masses, connected by a weightless "handle" of $2d$ length, the tidal force is, as has been shown:

$$F_{tidal} = G\frac{m \cdot M}{r^3}d \tag{2.2.4.5.35}$$

Based on this result, by performing a simple integration one can show that for a uniform rod of mass m and length $2d$, the tidal tension force in the middle of the rod is:

$$F_{tidal}^{rod} = G\frac{m \cdot M}{2r^3}d \tag{2.2.4.5.36}$$

As was mentioned earlier, the tidal force is generally of the form $F_{tidal} = k \cdot GmMd/r^3$, where the coefficient k depends on the shape of the body experiencing the force.

People may ask: "OK, but it is a well established fact that it is the Moon which is responsible for the tidal effects occurring on Earth. And everybody knows that the Moon orbits Earth, not the other way. So, it seems to be something wrong with the above explanation."

No, nothing is wrong. Because it is not exactly so that only Moon orbits Earth. In fact, the two bodies form what in astronomy is called a "binary system": both bodies revolve around a common mass center.

So, the tidal force generated by Moon stretches the Earth along the line connecting their move mass centers. The **entire** Earth, not only its fluid part. With some simplification, Earth can be thought of as rocky sphere surrounded by an outer layer of water, i.e., the world oceans. The rocky core is also stretched -yet, since rocks are much less susceptible to stretching forces than fluids, the rocky part "bulges up" less than the fluid part. The average height of bulge on the rocky part is about 20 cm, while at an ocean, far from the shore, water raises, per average, about 60 cm, i.e., 40 cm relative to the ocean bottom.

Why do we say "average"? Well, the tidal effects on Earth are not only due to the Moon, but also due to the Sun. The Moon's mass is $M_M = 7.35 \times 10^{22}$ kg, and its distance from Earth is $r_{M-E} = 3.84 \times 10^8$ m. For the Sun, $M_S = 1.989 \times 10^{30}$ kg, and the Sun-Earth distance $r_{S-E} = 1.496 \times 10^{11}$ m.

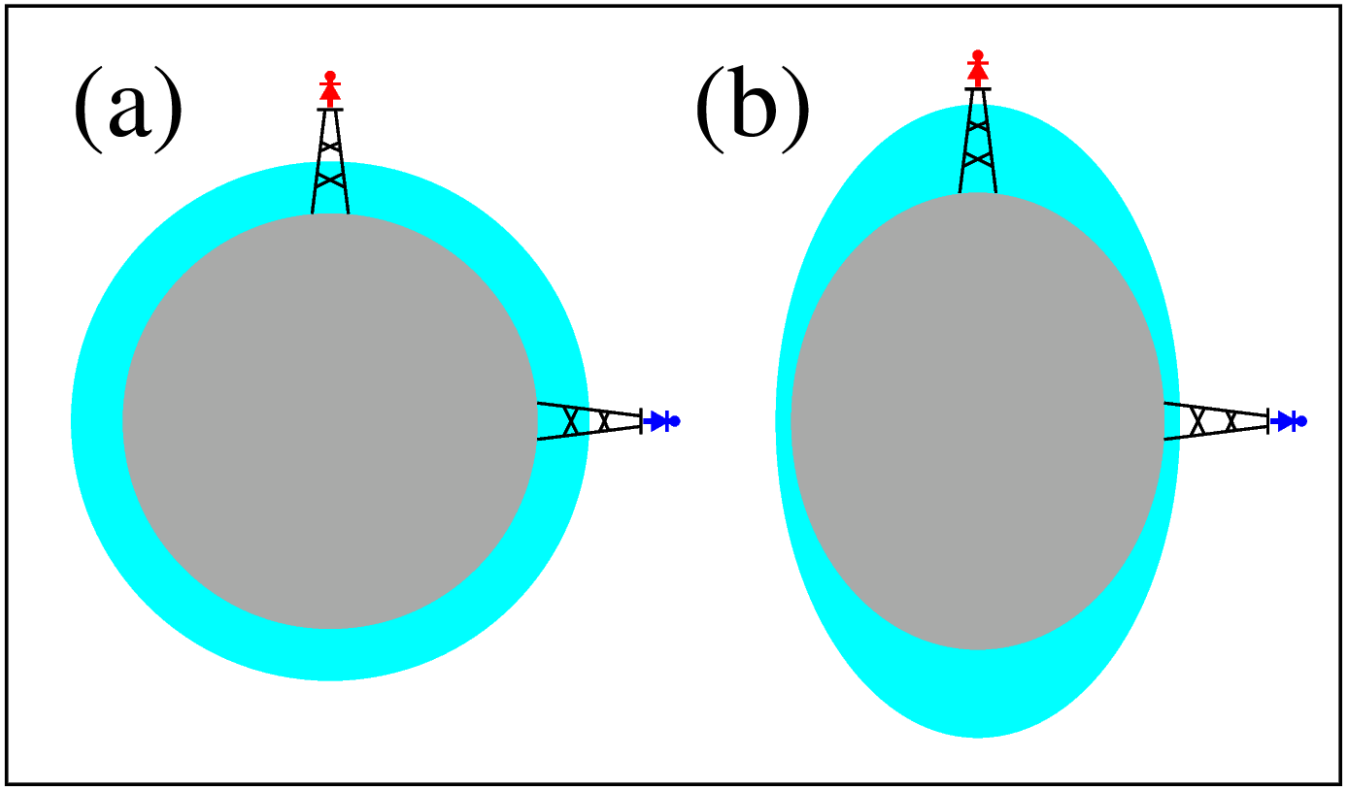


Figure 2.2.4.5.1: The effect of tides force on a planet with a rocky core (gray) covered by ocean (blue). Plot (a): No tidal force. Plot (b): the tidal force stretches both the rocky core and the ocean, but because rocks are stiffer than water, the core bulges up less than the ocean. Observers standing at towers set at the ocean floor (i.e., at the rocky core) do not see the core's motion -they move up and down together with the core's surface, and they cannot see the core's bulging. From their perspective, only the water level rises and falls. The same applies to observers standing on the ocean's shore, they are "connected" to the rocky core and they can only see the motion of waters relatively to the rocky core's surface.

Therefore, if we consider the **ratio** of the tidal forces acting on Earth from the Moon and the Sun, only the above four parameters are important, because all other cancel out:

$$\frac{F_{due\ to\ Moon}}{F_{due\ to\ Sun}} = \frac{M_M \times r_{S-E}^3}{M_S \times r_{M-E}^3} = 2.185, \tag{2.2.4.5.37}$$

meaning that about 68\% of the power of the tidal effects on Earth are due to Moon, and about 3 2\% are due to Sun. When Moon, Sun and Earth are aligned along a straight line (at Full Moon, or at New Moon), their contributions add up -the tide bulges are then the highest. They are called the "spring tides" (it's a tradition; no relation to the time of year). In contrast, when the directions of the Moon and the Sun viewed from Earth make a right angle -i.e., at the First Quarter Moon, or at the Third Quarter Moon -the effects from the Moon and Sun partially cancel out, and the tide bulges are the lowest. Such tides are called the "neap tides".

For an observer on Earth, the two tidal bulges with the "dips" in between can be though of a wave traveling around the globe (actually, it's the bulge which always form below the Moon, and the impression that it "travels" from the East to West comes from the fact that the Earth rotates. Since the Earth spins and the Moon travels on its orbit in the same direction (counter-clockwise, if watched from above the North Pole). A lunar "month" -i.e., is the average period of the Moon's orbit with respect to the line joining the Sun and Earth -is about 29.5 days. So, every day the Moon reaches its highest point in the sky about 48 minutes later than the previous day. Therefore, the "high tides" occur not every 12 hours, but about every 12 hours and 24 minutes.

As noted above, the tide's height at the middle of an ocean is only about 0.5 m. However, it is known that the height of a wave of water may significantly increase if the wave passes from a deep region to a much shallower region, or if it travels along a narrowing bay. Consequently, in some coastal regions on Earth the difference between the high tide and the low tide may be much larger, with the record value of 16 meters in the Bay of Fundy, Canada.

2.2.4.5: Tidal Forces is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.2.4.6: Extreme Tidal Effects

The tidal effects seen on Earth can be thought of as "moderately strong". Note that such effects are inversely proportional to the **cube** of the distance to the body responsible for them. Therefore, if the distance were halved, the tidal effects' power would increase as much as eightfold!

Well, the Earth-Moon and the Earth-Sun average distances are very stable, so the magnitude of the tidal effects on Earth is not going to change in the current epoch. However, there are celestial bodies ibn the Solar System that experience much stronger tidal effects than Earth -sometimes strong enough that they can even lead to “catastrophic" results. It may be instructive to consider such a scenario in somewhat closer detail.

Suppose that there is a satellite of mass m and radius d, orbiting a planet of mass M at a distance of r. Now we will use the same two hemispheres which we considered before for deriving the equation for the magnitude of the tidal forces. With one small change -since the m symbol has been already “taken" for the satellite mass, let's use the Greek equivalent instead -i.e., let the mass of each hemisphere be $\mu/2$.

In the preceding Section, we considered two hemispheres, $2d$ apart, coupled by a spring. Let's now put them on the surface of the satellite, one at the point closest to the planet and the other at the point most distant from it -as is shown in the Figure 2.5. Just put them, but we don't want to attach them to the surface by any constraints.

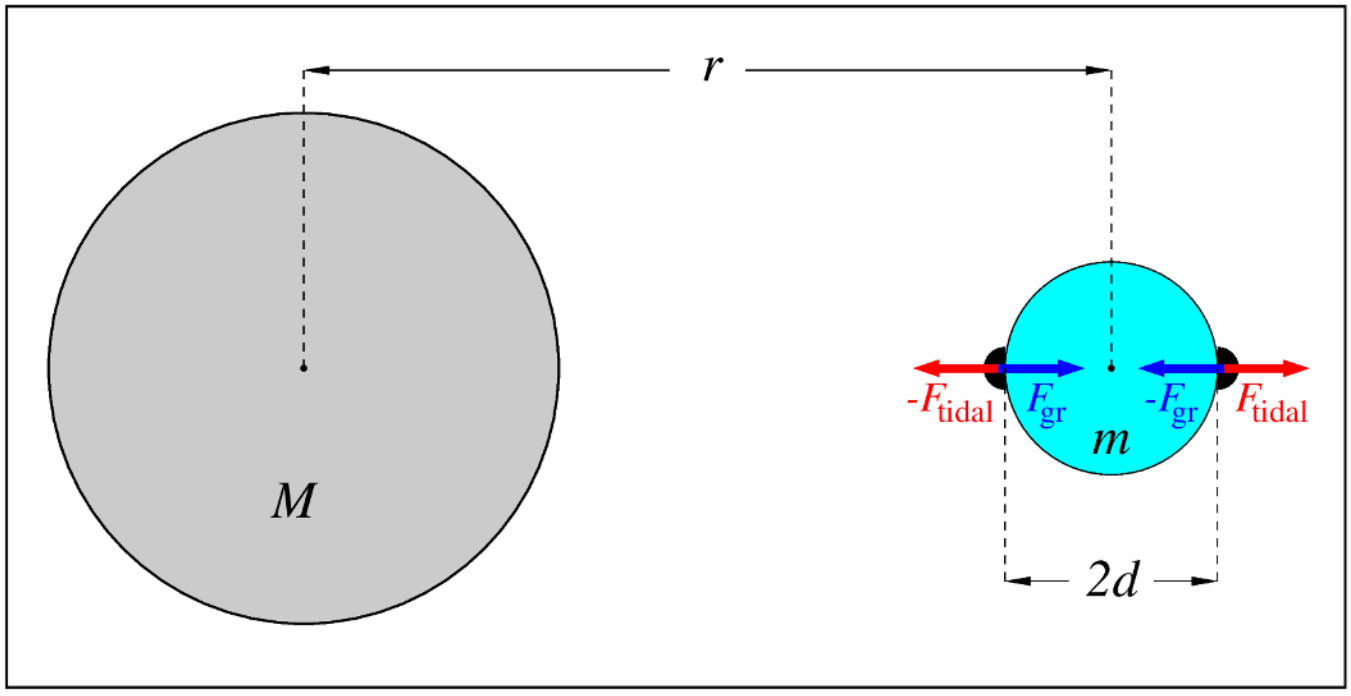


Figure 2.2.4.6.1:

So, we have now the two hemispheres $2d$ apart. Both are pulled “inwards", i.e., towards the middle point,
but not by the stretched spring force as before -now, by the gravity of the satellite, $\pm F_{gr}$ (in the Figure 2.5, the positive direction is, conventionally, from the left to the right). The magnitude of this force is:

$$F_{gr} = G\frac{\frac{\mu}{2} \cdot m}{d^2} \qquad (2.2.4.6.1)$$

This is the effect of the satellite's “own" gravity. But there is also gravity “supplied" by the planet. And as we know from the preceding section, two hemispheres, $2d$ apart, falling freely in a gravitational field of a planet, experience a tidal force tending to “stretch" the pair -i.e., acting upon each of them with an “outwards-directed" force $\pm F_{tidal}$, as shown in the Fig. 2.5. The magnitude of this force is given by the Eq. 2.56:

$$F_{tidal} = G\frac{\mu \cdot M}{r^3}d \tag{2.2.4.6.2}$$

As opposed to F_{gr} which is constant, the F_{tidal} force depends on the satellite-planet distance r -it increases with decreasing distance, and for certain value of r the F_{tidal} force may become equal to F_{gr}. This r value is called “the Roche Limit" in honor of the French astronomer Edouard Albert Roche, who first considere the effects of strong tidal forces on celestial bodies. By equating the Eqs. 2.57 and 2.58 and solving for r, one obtains the formula for the Roche Limit:

$$G\frac{\mu \cdot M}{r^3}d = G\frac{\frac{\mu}{2} \cdot m}{d^2} \quad \Longrightarrow \quad r_{Roche} = d\sqrt[3]{\frac{2M}{m}} \tag{2.2.4.6.3}$$

What happens if the satellite gets even closer to the planet than the Roche Limit? Well, the tidal force will become dominant. The hemisphere closer to the planet will start falling towards the planet. Paradoxically, the other hemisphere will take off from the moon's surface and... ***fly up!***

If the satellite comes even closer to the planet, not just objects loosely lying on the surface will be detached from it, but whole blocks of matter will be ripped off from the satellite -such a process is illustrated by a series of slides in the Wikipedia article titled Roche Limit.

It is commonly accepted that the famous rings of Saturn were created by tidal forces that “pulverized" a satellite that got closer to the planet than the Roche Limit. Anather spectacular manifestation of destructive tidal forces was the fragmentation of the Shoemaker-Levy 9 comet}, "captured" by Jupiter -eventually, in June 1994, the fragments, one after another, collided with the planet, creating spectacular fireballs on its surface.

For Moon whose mean radius $d = 1,738$ km, and mass is 1/81 of the Earth mass, the Eq. 2.48 yields a Roche Limit of about 9,500 km. Note that Roche Limit is the distance between the *centers* of the two bodies in question: accordingly, since the mean Earth's radius is 6,356 km, the distance between the *surfaces* of Earth and of Moon at the Roche Limit would be only (9,500 6,356 1,738) km = 1,406 km -very, very close. The rocks ripped out from Moon would then surely “bombard" Earth. Fortunately, though, the Moon's position is 40 times farther away from Earth than the Roche Limit, and this distance is even slowly *increasing* (about 1.5 inch/year), so there is no danger that such a “bombardment" will ever happen.

2.2.4.6: Extreme Tidal Effects is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.3: Thermal Energy and Heat

The question of what makes objects hot in some circumstances, and cold in some other situations, had intrigued philosophers for millennia.[1] In the XVIII Century, a well-established theory was that the entity responsible for such effects was a mysterious *caloric fluid* -an invisible, weightless and odorless substance, capable of penetrating matter. An object with high content of caloric flu id would get hot, and an object depleted of it would become cold. Only in 1799 a British scientist Humphry Davy performed a famous experiment showing that bodies may be heated up by *performing work on them*, which some time later led to the formulation of modern theory of thermal phenomena, usually referred to as *thermodynamics*. As we know now, it's not the caloric fluid that makes bodies hot -it's the *internal energy*, or *thermal energy*. It's a special category[2] of energy, usually denoted as U for distinction from pure kinetic energy or potential energy.

Only in diluted gases their internal energy can be identified with pure kinetic energy -here U is the sum of the kinetic energy of all individual molecules comprising the gas. In solids, in contrast, all constituent atoms are at fixed positions, they cannot move freely. However, each atom is coupled to several neighboring atoms by psysiological of hot and coldbonds", which act as springs of sub-nanometer lengths. Therefore, each individual atom behaves like a miniature spring oscillator. Then, the U of a given solid is simply the sum of the oscillatory energy of all individual vibrating atoms (more often, we call it *vibrational energy*.

So, it is the energy which decides of whether a body is hot or is cold. The internal energy of a body, call it A, can be changed in two ways. One is, as Humphry Davy discovered, by performing work on the body. For the amount of work performed, we use the symbol ΔW. The other way is by *heat transfer* from another body call it B, brought in contact with body A. If B is psysiological of hot and coldwarmer" than A, some energy flows into A, and if B is psysiological of hot and coldcooler" than A, some energy exits A and flows into B. The portion of energy entering or exiting body A in such a way is called *the heat transfer*, and conventionally denoted as ΔQ. What we have stated above by words, can be expressed in a mathematical form:

$$\Delta U = \Delta W + \Delta Q \tag{2.3.1}$$

which is known as the **First Law of Thermodynamics**.

Two comments concerning the above. One is simply about the terminology. According to the psysiological of hot and coldofficial language of thermodynamics, we cannot say: *It's the heat content that decides of whether the body is warmer or cooler*. The correct way is to say: *It's the amount of internal energy in a body which decides of whether it is warmer or cooler*. The term *heat* should be used only for describing the amount of thermal energy transferred into the body, or out of the body, via contact with another body.

The other comment is about physics. Namely, consider a situation in which there is no heat transfer, $\Delta Q = 0$, only work ΔW \textbf{is performed on the body.} Then, according to the First Law, the content of the internal energy in the body increases by:

$$\Delta U = \Delta W \tag{2.3.2}$$

Seems in perfect agreement with the Work-Energy Principle? Yes, it is!

But now one can think simple mindedly: if so, then the First Law may work in the other way, namely:

$$\Delta W = \Delta U \tag{2.3.3}$$

Meaning that the body can deliver work of ΔW by diminishing its internal energy by ΔU -which also seems to be in perfect agreement with the Work-Energy principle!

But the bad news are that the Fist Law *does not* work the other way ;-(... Thermal energy cannot be converted to mechanical work directly... The troublemaker here is the Entropy which emerges only in the **Second Law of Thermodynamics**.

We will talk about the Second Law somewhat later --because now we have to introduce one more essential thermodynamic parameter -namely, the *temperature*

Temperature

We all have an intuitive concept of temperature, based on -quoting the words of Herbert Callen[3] -*psysiological of hot and cold*. It's also a common knowledge that "temperature is what you read from a thermometer". It is not silly -it's something we call an "operational definition". Sometimes it is really difficult to find something better than such a definition for notions in physics. For instance: "Time is what we read from a clock" -can you think of a better definition of time? It's certainly not easy!

So, in low-level courses of physics instructors or textbooks often decide that such an "operational definition" of temperature is sufficient -with additional information about "standard temperature points" needed for calibrating the thermometer. In the Celsius scale, created in 1742, such points are the water freezing temperature, taken as 0 degrees, and water boiling temperature, taken as 100 degrees. Mr. Fahrenheit, who created a temperature scale even earlier, in 1724, took the temperature of human body as 100 F, and that of the mixture of ice and ammonium chloride as 0 F. Those points were not very precise, so later people started using the same points as in the Celsius scale: water freezes at 32 F, and boils at 212 F. Therefore, the conversion formulae between the two scales are:

$$t_{Fahr.} = 1.8 \times t_{Cels.} + 32F \tag{2.3.4}$$

and

$$t_{Cels.} = \frac{5}{9} \times (t_{Cels.}32F) \tag{2.3.5}$$

However, I can bet that you don't want to think that you are taking only a low-level physics course" -and you'll certainly like to know more, what is the **underlying physics** in the notion of temperature. Well, to satisfy your wish may not be an easy task for an instructor...

The truth is that with the progress in physics, when more and more experimental facts were collected, and more and more understanding of them was achieved, the definition of temperature also had to be consistent with all the accumulated knowledge of thermal phenomena. And presenting the current definition would be difficult, for two reasons. First, because it involves the notion of **Entropy**, which is a highly abstract concept. And second, because the definition in question is given by a mathematical equation including **partial derivatives**. Well, my intention is to make the present text digestible even for students who have not necessarily taken any **calculus classes**, and therefore I try to seldom use even the **simple derivatives**, only when there is really no other option -but the theory of partial derivatives, it's a much more advanced math than the theory of simple derivatives!

Therefore, I will not even try to discuss this super-duper definition here, but I will give you a definition that is perhaps not perfect, but definitely is good enough in the temperature region around room temperature, and for higher temperatures. Which is OK, because in this course we won't be talking of any phenomena that occur in the region of very low temperatures! And one more advantage is that the definition that will follow is primarily a conceptual one", with not much math.

The definition is based on the so-called ”kinetic theory of gases", a theory that matured in the second half of the XIX Century. Please recall what was said above about the internal energy U of diluted gases: it's simply the sum of the kinetic energies of all molecules comprising the gas considered. And what we said about the the energy U in solids? That it is the sum of **vibrational** energies of all atoms in the solid.

Now, remember what happens in vibrational (oscillatory) motion of a spring oscillator: the potential energy of the spring $V_{\text{stretch.}}$ changes to kinetic energy K, K changes to potential energy of the compressed spring $V_{compr.}$, the latter back to K, K back to $V_{stretch.}$, and then the cycle repeats, and so on, and so on. From the above it follows that on average, half of the time the oscillator energy is potential, and half of the time it is kinetic. And averaged over a long time, the potential energy contribution $\langle V \rangle$[4] and the kinetic energy contribution" $\langle K \rangle$ are equal.

So, we can write that the total energy of a spring oscillator is:

$$E_{Oscill.} = \langle V \rangle + \langle K \rangle \tag{2.3.6}$$

But also:

$$\langle V \rangle = \langle K \rangle \tag{2.3.7}$$

So that:

$$E_{Oscill.} = 2\langle K \rangle \tag{2.3.8}$$

In view of the above, since a solid may be thought of as an assembly of miniature atomic oscillators", we can conclude that the thermal energy U of the solid is:

$$U_{solid} = 2\langle K_{all\ atoms} \rangle \tag{2.3.9}$$

In conclusion, we can say that both for a gas and a solid:

$$U \propto \langle K_{all\ atoms} \rangle \tag{2.3.10}$$

where the meaning of the mathematical symbol $\propto$ is proportional to.

Now, it can be further reasoned:

more $U \Rightarrow$ warmer body $\Rightarrow$ higher temperature;

and

less $U \Rightarrow$ cooler body $\Rightarrow$ lower temperature,

which lead to the final conclusion:

The temperature of a body is a measure of the average kinetic energy of constituent atoms. We are nearly done with temperature! But before we finish, I need to tell you about two important conclusions emerging from the considerations in this sub-sub-sub-section. The first conclusion is that the internal energy U is a linear function of the temperature, which we will denote now as T. "Linear function" simply means that the dependent value (\)U\) in the present case) is equal to the independent value (here, the T) multiplied by a constant factor. So:

$$U_{\text{given body}} = \text{constant factor} \cdot T\backslash) \tag{2.3.11}$$

$$= C_{\text{given body}} \cdot T \tag{2.3.12}$$

where the C coefficient is called the "heat capacity" of a given body.

The second conclusion is that that if all internal energy is removed from the body, i.e., when $U = 0$, then it must be $T = 0$. Indeed, zillions of experiments carried out since the birth of modern thermodynamics have made it possible to determine that such a temperature point does exist, at -273.15 °C, or -459.67 °F. This point is called the "absolute zero". The SI system introduces yet another temperature scale, called the "Absolute Temperature Scale", or the "Kelvin Scale", in which the absolute zero is the fundamental thermometric point. It was decided, for simplicity, that one degree in the Kelvin scale, 1 K in short (without the ″°") would be equal to 1 ° C, one degree in the Celsius scale. So, the conversion from the Celsius scale to Kelvins is pretty straightforward:

$$T[K] = t[^\circ C] + 273.15K \tag{2.3.13}$$

([...] means in the units of what is written between the two brackets).

Hence, water freezes at 273.15 K, and boils at 373.15 K.

For distinction, capital T should be used only for expressing temperatures in Kelvins, and lower-case t for temperatures in the Celsius or Fahrenheit scales. Please always pay attention to in which scale the temperature is given, confusing °C with K is one of the most common errors made by students in homeworks and exams in the Ph313 course!

---

1. in the old days, there were no psysiological of hot and cold scientists" of different categories, all people doing any kind research were called psysiological of hot and cold philosophers", what in ancient Greek language meant psysiological of hot and coldthey who love wisdom"

2. Also, a very important category in the context of energy usage, because most of the electric energy consumed today comes from *thermal power plants*, i.e., from big facilities in which thermal energy is converted to electric energy.

3. Herbert Callen, 1919-97, was a distinguished American physicist, author of the textbook T*hermodynamics and an Introduction to Thermostatistics*, the most frequently cited thermodynamic reference in physics research literature.

4. In physics, writing a quantity in between angular brackets: $\langle \mathrm{quantity} \rangle$ is an often used symbol of averaging.

2.3: Thermal Energy and Heat is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

2.3.1: Forms of Energy and Power- an Engineer’s Approach

The “Energy Alternatives" topic depends much more to the realm of engineering sciences, than to the realm of “pure physics". We needed to start with introducing a number of important physical concepts (and we will once in a while return to “pure physics"), but now we will switch a mode that is much closer to an “engineer's approach".

We have to begin with classifying energy forms according to the "sources they come from", rather than their physical nature. And below there is a list of energies and their sources we will discuss throughout the present text. The items are listed in “chronological order", beginning with the energies utilized by the most ancient civilizations, and ending with power sou\-rces that have only been most recently harnessed.

- **Biological energy** – the energy stored in our bodies. Humans started using this form of energy long before they became humans (for moving around, climbing trees, hunting, fighting…).
- **Solar energy** -same as above. E.g., to get warm after a chilly night.
- **Chemical energy** -that stored in the foods. Again, humans started taking advantage of this form long before they became humans.
- **Thermal energy** -that from fire. At the onset of civilization! It was the first form the use of which required human intellect!
- **Animal energy**: here we mean “biological energy", but that stored in the bodies of domesticated animals, not our own bodies.
- **Wind energy** -first ever “hi-tech” devices built by humans used it! In sailing boats, wind mills, simple water pumps for irrigating fields.
- **Hydro-energy** – kinetic/potential energy of water in streams and artificial ponds – probably, started being used about the same time as wind energy. Ancient civilizations, e.g., Egyptians, used both!

Then, there was a long period with no significant progress in using new energy forms...

...Things started moving in fast pace only in the Second Millennium. The new forms that people employed then were:

- **Chemical energy from “synthetic'', i.e. man-made carriers**: gunpowder (a.k.a. “black powder"), first invented in China about 1000 years ago, was perhaps the first?
- **Chemical energy from mined sources: (e.g., coal)**: widespread use begun in the 18^{th} Century; then towards the end of the 18^{th} Century came oil, and natural gas.
- **Energy from heat engines** -when coal became available; the first heat engine was built in England by Newcomen in early 1700-s then a far better machine was built about 50 years later by James Watt. When that happened, the “Industrial Revolution" began!
- **Electrical energy** -19^{th} Century; the proud 19^{th} century people called their times *The Century of Steam and Electricity*.
- **Geothermal energy** -water from hot sources had been used for millennia by people who dwelled nearby, but an utilization at industrial scale began only in the early years of the 20^{th} Century.
- **Atomic (nuclear) energy** – first used for not-so-nice purposes in 1945, for peaceful purposes a few years later.
- **Mass**, as discovered by A. Einstein, is also a form of energy (\)E=mc^2\)).

Is this list complete? Can you think of any other items?

2.3.1: Forms of Energy and Power- an Engineer’s Approach is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

3: Fossil Fuels and Thermal Engines

3.1: Fossil Fuel Resources

3.2: How much energy there is in fossil fuels?

3.2.1: How much CO2 is released by burning various fuels

3.3: Heat Engines

3.4: The Efficiency of Thermal Engines

3.4.1: The First Law of Thermodynamics.

3.4.2: Entropy

3.4.3: The Second Law of Thermodynamics

3.4.4: Sad Consequences of the Second Law

3.4.5: The Efficiency of an Ideal Thermal Engine

3.4.6: Even sadder news

3: Fossil Fuels and Thermal Engines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.1: Fossil Fuel Resources

There are three major types of fossil fuels:

- Those derived from naturally occurring *petroleum* forms, which include liquid oil, tar sands, and oil shales – by refining them, one obtains gasoline, diesel fuel, and aviation fuel (kerosene).Petroleum. Here, we will use the term *crude oil* for all the mined forms, and a general term of *oil-products* for all of the refined forms.
- Natural gas, which is essentially methane CH_4 with small admixtures of other gasses.
- Coal – all its types.

A unit measure used for oil and its liquid derivatives is a *barrel* (37.5 US gallons, or 119.24 liters). Because oil is now the "king" of fossil fuels, it has become a custom to express the amount of the other two fuels in units called *Barrel of Oil Equivalent BOE*. A BOE of coal is the amount of coal that releases, when burned, the same amount of energy as burning a barrel of oil (6.1 GJ)– it's 200 kg, or 0.2 ton of coal. For natural gas at normal pressure, one BOE is 5,800 cubic feet, or 164.23 cubic meters (the US Geological Service uses 6,000 cubic feet, or 170 cubic meters).

Comprehensive data concerning the resources and the consumption of fossil fuels can be found in several different Web documents – we recommend the Fossil Fuels site, a section of a larger Web document maintained by the *Our World in Data*, a project produced by the *Oxford Martin Programme on Global Development* at the University of Oxford, U.K., and is made available in its entirety as a public good.

Resources:

The global resources of recoverable oil (including liquid oil, tar sands, and shale oil) are estimated to be as high as about 1,700 billion barrels. The three countries possessing the largest oil deposits are Venezuela, Saudi Arabia, and Canada (with 300, 266.6, and 172,2 billion barrels, respectively). The US resources are estimated to be about 55 billion barrels.

The estimated global resources of natural gas are 187 trillion cubic meters, equivalent to 1,139 billion BOE.

The estimated coal resources, including all forms, from the high-quality (anthracite) to the lowest-quality lignite (also known as "brown coal") are 1.1 trillion tons, or 5,500 billion BOE.

Consumption:

The data concerning various aspects of fossil fuels are often given in different units (which may be a nightmare for the reader, and even for an author of a text like this one). It's easier to find *annual production* than *annual consumption* data – not a big problem, because most of the oil mined is promptly refined and consumed, so that one can safely assume that the production and consumption data are equal. But to make things worse, the production data in the Internet source quoted above are given in TWh, while we would prefer BOE data. Fortunately, on the Web, using Google or another search engine, one can find right away programs converting one kind of units to another kind. So, one such program informs us that:

1 TWh = 588,235 BOE

Fine! So, let's begin with **oil**. In the *Fossil Fuels* site we read from a chart that in 2016 the annual production of crude was 51,000 TWh – and by multiplying this number by the above conversion factor, we obtain almost exactly 30 billion barrels. Now, we take the global resources, 1,700 barrels – and we get that if oil is mined from now on at the same level as in 2016, the resources will be all gone about 57 years from now.

For the author of this text, who was born the at the same year as the WW II ended, for obvious reasons it's not a threatening perspective – but if you, Dear Reader, are an undergraduate student, it may be a piece of bad news!

Let's then look at **natural gas**. Again, the storage of natural gas is difficult (it may be stored in in a liquefied form, which is the called LNG, but it's being done mostly for transporting the gas, not for storing it for a long period (a much better method of long-time storage is to keep the gas underground!). So, here we can again assume that the annual consumption and the annual production are equal. In the *Fossil Fuels* site we find that the annual production is 37,366 TWH, which converts to 21.98 billion BOE. By dividing the known reserves, 1,139 billion BOE, by the latter number, we obtain a period of about 51.8 years. Again, not very good news!

The last one on our list is **coal**. Since there is a serious danger that all oil and gas resources will be gone before the end of this century, perhaps coal is the "last resource" for future generations? Well, there are no reasons for too much optimism. In 2016, the production of all coal types was about 42,500 TWh, equivalent of 25.0 billion of BOE. Considering that the global resources are 5,500 billion BOE, we get 220 years. Do we? Not really! Because when all oil and natural gas resources are exhausted, coal must substitute for the two other missing fuels! It must be used in much higher annual rate, if we still want to generate as much power from fossil fuels as we did when oil and gas were still available!

Making estimates of how long the coal resources will last if coal indeed will be used to "fill the gap" left by exhausted oil and gas deposits is difficult – let's not try to do that here, but let's rather check what professionals claim. The problem is discussed, for instance, in this Web document. From the graph presented in this site, one can see that the authors of the document don't even expect the coal resources to last beyond the end of this century! There is always some room for optimism, because new deposits of fossil fuels are still being discovered. For instance, the recoverable oil reserves in the United States in 2005 were estimated as 29.9 billion barrels, and ten years later, in 2015, as 50 billion barrels. Such a big change was due to two factors: new oil fields in North Dakota, and the invention of a new "fracking" technology that allows to pump more oil from deposits that were considered fully emptied. Similarly, the estimates of natural gas resources, which were 1200 trillion cubic feet in 2002, increased to over 2,400 trillion cubic feet in 2013. Again, it was due to discovering new gas fields and to developing new technologies of horizontal drilling and fracking.

On the other hand, people who are even strong enthusiasts of using fossil fuels should keep in mind that there are not mined with the sole purpose of burning them in power plants, smelters, heating furnaces, or gasoline, diesel, and jet engines. Oil and natural gas (and coal, to a lesser degree) are the basic raw material for many branches of chemical industry, making products of crucial importance for our lives: plastics, fertilizes, drugs, and dozens of others. If we keep burning fossil fuels as long as they are available, our lives may become miserable... And another important reason to start showing restraint in burning fossil fuels is that we are adding more and more CO_2 to the atmosphere. The concentration of CO_2 in the atmosphere is rapidly growing.

The CO_2 concentration in atmospheric air is monitored by many observation posts around the globe – the best known of which is the one located at Hawaii, at the Mauna Loa volcano at the altitude of 4169 meters above the sea level. Hawaii archipelago is located in the middle of the Pacific Ocean, far away from any big industrial center that might be able to create a local increase above the average concentration. On the graph in Fig. 3.1, one can see a steep curve with characteristic saw-teeth-like oscillations. This is the so-called *Keeilng Curve*, in which the the saw-teeth-like features represent the *seasonal variations* of the carbon dioxide concentration. Why does it change with seasons? The answer is simple: in May, the concentration starts decreasing, which reflects the fact that the vegetation, grasses and trees, start growing new leaves, and they are rapidly absorbing CO_2 from the air. The process lasts for several months, producing a "dip" on the curve, until the leaves fall and start rotting – now, in turn, releasing CO_2 into the air.

Well, one may have some doubts concerning such an explanation. The spring starts in May, yeas, but only at the northern hemisphere – and at the southern, it starts in 6 months later in the same calendar year, in November. So, the southern hemisphere vegetation should give rise to a similar "dip" phase-shifted by 6 month relative to that caused by the northern-hemisphere vegetation. Hence, the two effects should cancel each other!

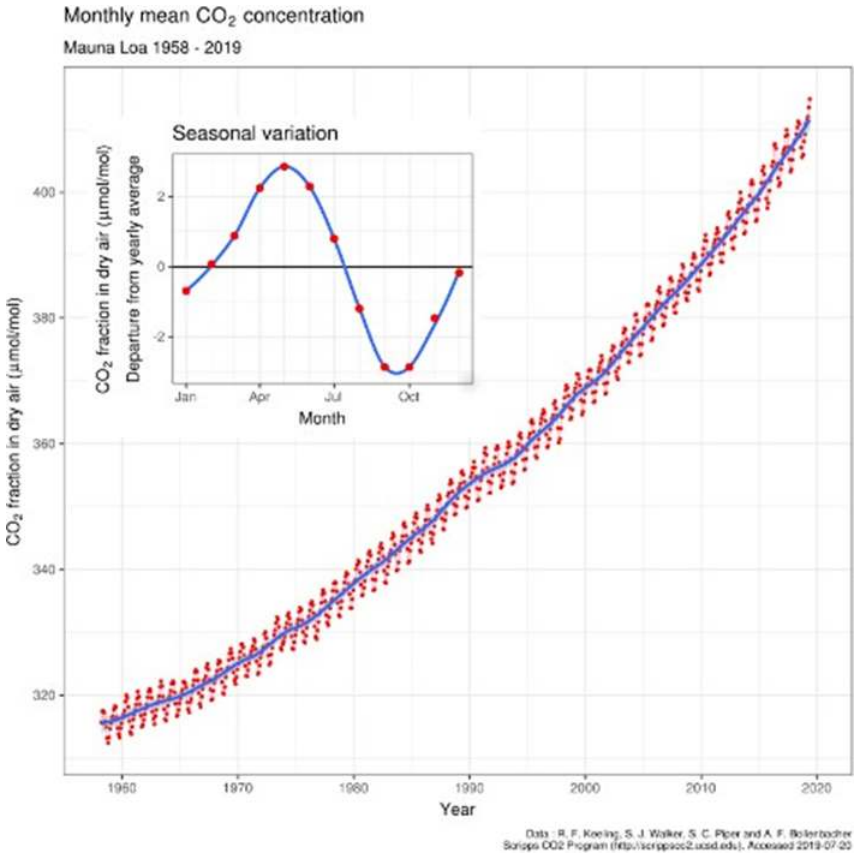


Figure 3.1.1: Atmospheric CO_2 concentration at Mauna Loa Observatory (from Wikimedia.commons)

Yes, it would be so, but the thing is that most landmass, and most vegetation on Earth is on the northern hemisphere. Especially, the boreal forests of northern Russia (the "taiga") and of northern Canada are much stronger CO_2 absorbers/emitters than anything that can be found in the southern hemisphere. For instance, the famous enormous rainforest in the basin of Amazon River is all *evergreen*, so that it does not contribute to the seasonal variation of CO_2. Therefore, the southern hemisphere vegetation plays only a minor role in the process.

However, if there were only seasonal variations, the Keeling Curve, when averaged, should be flat – but in fact, it's rapidly growing. When the Author of this text for the first time learned about the CO_2 concentration in the atmosphere, it was 0.03% = 300 ppm, and now, 60 years later, it has already exceeded 400 ppm.

So what? – some people angrily ask. Where is the problem? Here we have come to one of the major controversies of our times. There is no doubt that the increasing concentration of CO_2 is *antropogenic*, i.e., caused by the activity of us, humans. And in the scientific community, the majority of its members agree that the increasing concentration of CO_2 is responsible for global warming. In 2015, a United Nations Framework Convention on Climate Change was held in Paris. There was a consensus among the 195 participating nations (summarized in the document worked out by the conference, the so-called *Paris Agreement*) that keeping the CO_2 emission at the current (or higher) rate will lead in the near future to a global temperature increase of 2°C over the *pre-industrial level*, i.e., the average global temperature at 1880. The participants declared to voluntarily mitigate the emission of CO_2 by their countries. It is not known exactly by how much the emission should be cut down to meet the goal – there is no well-established theory of how the CO_2 concentration in the atmosphere is correlated with he global temperature. The data on which the Paris Agreement recommendations has been based are published in a 2013 Intergovernmental Panel on Climate Change. The Paris Agreement recommendations are often termed the *carbon budget*. In short, it says that if we want to achieve a 50% probability of limiting warming to 2°, we have to leave the majority (up to 80 percent) of our fossil fuels in the ground (for more details, see the *Unburnable Carbon* Section in the Fossil Fuels Web page. OK, if 80% of fossil fuels should stay in the ground – then, if they are mined at the current rate, they will "run out" much faster than we predicted above – or, the pace of mining them from the underground must be drastically reduced.

On the other hand, there are many groups of people (including some scientists) who argue that man-made global warming is a "hoax". They use different arguments, or no arguments – there make no sense to quote them here, because this would be off-topic. The main point in this Section is: "How long will the fuel resources last?" – so the answer to this question depends not on who is right (those who believe in man-made global warming, or global warming skeptics?), but which one of those two groups will win the dispute. And the skeptics have many supporters – beginning from very-high placed politicians, to big business leaders whose profits are correlated with the rate of fuel extraction, and – last but not least – to people who live from mining fuels, i.e., simple coal miners or operators of gas or oil wells. The combined influence of such people on public opinion may be quite strong, and may decide, for instance, on who wins the elections.

So, it's not possible to predict which group will take over. If the global warming skeptics, then mining of fossil fuels may last longer. Recently, the Trump Administration declared it would withdraw the United States form the Paris Agreement (it cannot happen immediately – the earliest possible effective withdrawal date by the United States cannot be before November 4, 2020, four years after the Agreement came into effect in the United States). It's always possible that other countries may do the same. Regretfully, the only thing that one can do now is "to wait and see".

3.1: Fossil Fuel Resources is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

3.2: How much energy there is in fossil fuels?

A very important characteristic of each fossil fuel type is the *heat of combustion* – in other words, the amount of thermal energy released in the process of burning a mass unit of a given fuel. It is listed in the table below.

Table 3.1. Heat of combustion of fuels

Fuel	MJ/kg	BTU/lb
Hydrogen	141.80	61,000
Methane (Nat. gas)	55.50	23,900
Propane	50.35	21,700
Butane	49.50	20,900
Gasoline	47.30	20,400
Kerosene	46.20	19,860
Diesel	44.80	19,300
Ethanol	29.67	12,750
Methanol	22.65	9,740
Coal (Anthracite)	27.00	11,600
Coal (Lignite)	15.00	6,500
Wood	15.00	6,500
Peat (dry)	15.00	6,500
Peat (dump)	6.00	2,500

In addition, we also give the data for several other substances which are not mined, but are widely used as fuels. The names of fossil fuels or fuels derived from them are shown in bold face, and of other fuels – in regular face.

The heat of combustion is given in the units of MJ/kg, and also of British Thermal Units per pound (BTU/lb). The latter is not a SI unit, but it is still very popular in the US.

Hydrogen H_2 does not occur naturally. Large amounts of hydrogen are used in fertilizer industry for making ammonia NH_3. In the US, 95% of hydrogen used for this purpose is produced by a process called *natural gas reforming*. Hydrogen may be used as fuel in electric cars using *fuel cells* for converting it to electricity. They are real *zero-emission* cars (their exhaust is pure water vapor only), so they are ideal for California. However, because of their high price and a small number of refueling stations, they are not yet very popular: until mid-2018, only slightly over 3,000 of them has been sold, mostly in big CA cities. And it has to be kept in mind that the cars are zero-emission ones, yes, but CO_2 is released to atmosphere when the fuel for them is produced by the natural gas reforming technology.

Propane (C_3H_8) and **Butane** (C_4H_{10}) are gases normally dissolved in crude oil, and released during the rectification process; they are transported and sold in a liquefied form, in pressurized tanks. They are used, e.g., as a popular fuel in gas grills.

Peat is a fossil fuel, yes, but a very “young” one. It formed at the end of the last Ice Age – not tens of millions years ago, like coal. There are few peat-fired power plants in the world, so, essentially, peat is a fossil fuel of only marginal importance.

3.2: How much energy there is in fossil fuels? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.2.1: How much CO2 is released by burning various fuels

In the Table 3.1. we find how much thermal energy is released by burning a unit weight of various fuels – but an important question also is how much CO_2 is released in such processes. In fact, it's easy to find out. The only skill one really needs for doing this is to know how to write the chemical equation of the combustion reaction[1].

How to write the equation of a combustion reaction?

Step One: The left side. On the left side of the equation, you put the chemical formula of a single molecule of the "fuel". For instance: hydrogen occurs in diatomic molecules H_2. Coal, with a not-so-bad approximation, can be though of as pure carbon, C. Natural gas is essentially methane, CH_4. Gasoline is a mixture of a large number of different hydrocarbons, but a "good average" for it may be octane C_8H_{18}. For methanol and ethanol the respective formulas are CH_4O and C_2H_6O. By the way, there are different methods of writing the chemical formulas, some people prefer writing them in a way that is similar to the actual arrangements of atoms in the molecule. Then, methanol is written as CH_3OH, and ethanol as CH_3CH_2OH. But in combustion reaction equations it does not matter which method you prefer: only the total number of atoms of each constituent element must be right. In an ethanol molecule, there are 2 atoms of carbon, 6 atoms of hydrogen, and one atom of oxygen. Note that those numbers are correct, no matter whether you use the C_2H_6O, or the CH_3CH_2OH formula.

The other component necessary for combustion is oxygen, usually taken from the air. Oxygen, like hydrogen, also occurs in diatomic molecules O_2.

So, we can now start writing the left side of the combustion equation. Let's take, for example, ethanol. So, we write down the formulas for ethanol and oxygen:

$$C_2H_6O + xO_2 \longrightarrow \tag{3.2.1.1}$$

We don't know yet how many oxygen molecules will be needed, so that we put x.

Step Two: The right side. It's even easier, because in the reactions of combustion of all fuels listed in Table 3.1. the only products of combustion may be carbon dioxide CO_2 and water H_2O. We don't know yet how many molecules of each of the two are created, so we put, as in "ordinary" equations, the symbols normally used for "unknown", x, y and z :

$$C_2H_6O + xO_2 \longrightarrow yCO_2 + zH_2O \tag{3.2.1.2}$$

Step Three: Determining y and z. It's really easy. On the right side, we

have a total of 2 carbon atom, so that two molecules of CO_2 can be created and a total of 6 hydrogen atoms, so that three H_2O molecules can be created. Therefore:

$$C_2H_6O + xO_2 \longrightarrow 2CO_2 + 3H_2O \tag{3.2.1.3}$$

Step Four: Determining x. We now count the total number of O atoms on the right side: in two CO_2 molecules there are four of them, and in three H_2O molecules there are three of them - seven altogether. And on the left side, one O atom already is in Hhe_2H_6O molecule. To get a total of seven on the left side, we need extra six, so that x must be 3 . So, now we can write the final equation:

$$C_2H_6O + 3O_2 \longrightarrow 2CO_2 + 3H_2O \tag{3.2.1.4}$$

In a similar fashion, we can write the equation of combustion for methanol:

$$CH_4O + 1\frac{1}{2}O_2 \longrightarrow CO_2 + 2H_2O \tag{3.2.1.5}$$

Here the situation is a bit more tricky, because, as you see, on the right side we have four oxygen atoms, and on the left side one is already in the methanol molecule, so we need three extra atoms of oxygen. Some people like to take it this way: three atoms of oxygen are in one-and-one-half oxygen molecule. Other people don't like to use "half-molecules", so they write instead:

$$2CH_4O + 3O_2 \longrightarrow 2CO_2 + 4H_2O \tag{3.2.1.6}$$

Both ways are correct, and which one you want to use depends only on your personal preferences.

For coal, approximated as carbon, the equation is very simple:

$$C + O_2 \longrightarrow CO_2 \tag{3.2.1.7}$$

For gasoline, the equation is:

$$C_8H_{18} + 12\frac{1}{2}O_2 \longrightarrow 8CO_2 + 9H_2O \tag{3.2.1.8}$$

which, if one doesn't like "half-molecules", can also be written as:

$$2C_8H_{18} + 25O_2 \longrightarrow 16CO_2 + 18H_2O \tag{3.2.1.9}$$

How to calculate how much CO_2 is created in a combustion reaction? To do this, one has to first recall what the atomic mass m_a is: it's the mass of an atom of a chemical element expressed in atomic mass units. For the hydrogen atom, $m_a = 1$. A carbon atom is 12 times more massive than a hydrogen atom, so that for C$m_a = 12$, and for an oxygen atom, O, $m_a = 16$. For simple combustion reactions we consider in this Section, it's all we need.

The molecular mass M of a molecule is the sum of the atomic masses of all the constituent elements. For instance, for CO_2, $M = 12 + 2 \times 16 = 44$. For H_2O, one gets $M = 2 \times 1 + 16 = 18$. For O_2, $M = 2 \times 16 = 32$. And for ethanol C_2H_6O, $M = 2 \times 12 + 6 \times 1 + 16 = 46$.

In a chemical reaction equation, the sum of all atomic masses on the left side must be equal to the sum on the right side. It's yet another method of checking whether the equation is correct. So, let's look at the combustion equation of ethanol we have derived a moment ago:

$$C_2H_6O + 3O_2 \longrightarrow 2CO_2 + 3H_2O \tag{3.2.1.10}$$

For the left side, we get: $\Sigma M_{\text{left}} = 46 + 3 \times 32 = 142$ (Σ means "the sum of"). And for the right side, we get: $\Sigma M_{\text{right}} = 2 \times 44 + 3 \times 18 = 142$. So that, $\Sigma M_{\text{left}} = \Sigma M_{\text{right}}$, as expected, showing one more time that our combustion equation is correct.

Now, the calculation of the CO_2 mass becomes very simple. Let's writethe molecular masses instead of the chemical formulas:

$$\underbrace{46}_{\text{fuel}} + \underbrace{96}_{\text{oxygen}} = \underbrace{88}_{CO_2} + \underbrace{54}_{H_2O} \tag{3.2.1.11}$$

Now, one can clearly see: the "input" is 46 mass units of ethanol, and the "output" is 88 mass units of CO_2. This gives us the proportion of the output CO_2 mass to the input ethanol mass units (we are not interested in the input oxygen and the output water). There must be the same proportion of masses, if we use kilograms, not atomic mass units. Denote the mass of CO_2 produced by the combustion of 1 kg of ethanol as X. We can write:

$$\frac{CO_2 \text{ mass (atomic units)}}{\text{Ethanol mass (atomic units)}} = \frac{88}{46} = \frac{X \text{ kg}}{1 \text{ kg}} \tag{3.2.1.12}$$

By solving this proportion, we get:

$$X = (1 \text{ kg}) \times \frac{88}{46} = 1.913 \text{ kg} \tag{3.2.1.13}$$

Ethanol is sometimes used as a fuel - e.g., in tourist stoves - but it's definitely not a fuel used by electric power generating plants. Let's then make similar calculations for two fuels that are used by such plants: coal and natural gas.

Let's take anthracite, the highest-quality mined coal (over 95% of pure carbon), and ignore the 5% that may not be carbon. The equation of combustion is then pretty simple:

$$C + O_2 \longrightarrow CO_2 \tag{3.2.1.14}$$

The fuel input is 12 atomic mass units, the CO_2 output is 44 atomic mass units. Hence, the combustion of 1 kg of anthracite produces:

$$X = (1 \text{ kg}) \times \frac{44}{12} = 3.67 \text{ kg of } CO_2 \tag{3.2.1.15}$$

Natural gas, as mentioned, is almost pure methane CH_4. The equation of combustion is then:

$$CH_4 + 2O_2 \longrightarrow CO_2 + H_2O \tag{3.2.1.16}$$

The molecular mass of methane is 16 atomic mass units, so the mass X of CO_2 created in the reaction is:

$$X = (1\ \text{kg}) \times \frac{44}{16} = 2.75\ \text{kg of } CO_2 \tag{3.2.1.17}$$

It's only $3/4$ of the CO_2 released by burning the same mass of coal. But one often hears that natural gas is a "much greener" fuel than coal, so perhaps you have expected a more significant difference?

But the difference is significant! The thing is that we should not look at the amount of CO_2 released by burning a mass unit of a given fuel, but at the amount of CO_2 that needs to be released to the atmosphere for obtaining a unit measure - say, $1GJ$ - of thermal energy by burning a givenfuel. So, let's see how such figures look like for coal and natural gas. Let's take the data from Table 3.1 and from the calculations we have made, but for 1000 kg (a ton) of fuel, not a single kilogram:

- The burning of $1000\ \text{kg}$ of coal releases 27.0 GJ of heat, and $3670\ \text{kg}$ of CO_2. Accordingly, for obtaining $1GJ$ of heat from coal one has to "dump" $3670/27 = 136\ \text{kg}$ of CO_2 to the atmosphere.
- The burning of $1000\ \text{kg}$ of natural gas releases $55GJ$ of heat, and 2750 kg of CO_2. Therefore, for producing $1GW$ of heat from natural gas one has to send $2440/55 = 50.0\ \text{kg}$ of CO_2 to the atmosphere.

As follows from the above, natural gas is "nearly three times greener a fuel" than coal! In fact, in practice this ratio may be even higher, because of two reasons. First, we have made our calculations for anthracite, the highest quality coal. It's the so-called smokeless coal [2]. Only about 1% of all global coal resources are anthracite, the rest are lesser quality forms with lower heat of combustion than anthracite. A really "bad guy" is lignite, the socalled brown coal. With its heat of combustion as low as $15MJ/kg$, it may produce as much as $250\ \text{kg}$ of CO_2 for $1GJ$ of heat, a score about five times worse that natural gas. Nonetheless, "brown coal" is willingly used by some European countries because huge rich deposits of this fuel are placed shallowly underground. Like the coal deposits in Montana – however, in Montana it's a good quality coal, not the highly polluting lignite.

The other reason that the natural gas is much "greener" than coal is the difference in the efficiency of thermal engines using coal and those using natural gas as fuel. Coal is the standard fuel in power plants using steam turbines – currently, it's the most popular type of thermal power plants. Here, coal is used to heat steam-generating boilers, from which steam is directed to turbines. The combined efficiency of such a system may be only somewhat higher than 30%. It means that the amount of CO_2 created for producing 1 GW of *electric energy* may be as high as (136 kg)/0.3 ~ 450 kg of CO_2. On the other hand, the efficiency of *gas turbines*, converting thermal energy from gas burning directly to mechanical energy, may be as high as45% – and in some newest models of turbines, the record-high efficiency is 62%. Then, the production of 1 GJ of electric power releases only (50.0 kg)/0.62 = 81 kg of CO_2. 81 kg as opposed to 450 kg, natural gas, if such gas turbines are widely used, may be 5.56 times "greener" than coal!

---

1 From his past experience in teaching the *Energy Alternatives* course at OSU, the Author knows that not all students taking this course have taken a basic chemistry course, either during their high school education, or at OSU. The term *chemical reaction equation* may sound scary to them – which is a *wrong reaction*(!), such equations are pretty simple indeed, and we here explain step-by-step how to write them.

2 During the Civil War, the "blockade runners" – i.e., the Confederacy merchant ships trying to sneak through the blockade imposed by the Union – used anthracite as a fuel. Other types of coal were much cheaper, but they produced a lot of smoke, revealing the position of the "blockade runner" to the Union Navy ships enforcing the blockade.

3.2.1: How much CO2 is released by burning various fuels is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.3: Heat Engines

Heat engines are devices that convert thermal energy to mechanical work. They are all around us. They may be quite annoying if you live in the center of a big city. There are many different types. Chronologically, the first types used in industry, used for pulling trains on railroads, and used for propelling commercial and navy vessels were piston steam engines. At the end of the 19th century internal combustion engines were invented – gasoline and diesel engines – and they started gradually replacing steam engines. Some of the "steamers", though, were die-hards: steam automobiles, the famous *Stanley Steamers*, were manufactured in the US until the mid-1920-s. Steam railway locomotives were still used in many countries even after the year 1950. And during the World War Two years, between 1941 and 1945, the American shipyards mass-manufactured, at an unprecedented scale, as many as 2710 cargo steamers, known as the *U.S. Liberty class* ships.

Steam turbines. However, the decline of the piston steam engines coincided with the emergence of a new steam engine type – namely, the *steam turbine*. They turned out to be particularly useful in devices where huge power was needed: for propelling the largest ships – and in power plants generating hundreds or thousands of megawatts of electric power. The first modern steam turbine was built in 1884 by Sir Charles Parsons (1854-1931). An generator connected to his turbine produced 7.5 kW of electric power. During his lifetime, the power of a single turbine units increased to 50 MW. Today, the two largest existing steam turbines are capable of generating 1,600 MW each.

In today's global electric power generating system steam turbines are extremely important, indeed. Their main application is in conventional thermal power plants – those fueled by coal, natural gas, or oil. Also, all existing nuclear power plants utilize them: nuclear reactors are only generating heat, later used for making steam sent to turbines. And they are also vital components in some renewable energy systems – in solar power plants based on the technology of Concentrated Solar Power (CSP). Altogether, power plants using steam turbines generate more than 50% of electric power currently consumed worldwide.

Because of the leading role of turbines in power generation, we will not discuss the design or functioning of piston engines here, no matter what type they are steam, petrol or diesel. For students who are strongly interested in such engines, we can recommend an instructive Web site in which 20 animated schemes of piston engines (plus one of a jet engine) are shown. It's also recommended to pay special attention to the Newcomen Atmospheric engine in this Web page. You may also want to see this Web page. The Newcomen Engine, designed in 1712, was the very first thermal engine that found a wide application in mining industry (several hundred of them were built throughout the 18th Century. Their construction and the operation principle were very simple – not to say "primitive". So why it's worth paying attention to such an engine? Well, because of the simplicity of its operation, it's pretty straightforward to calculate its thermal efficiency (pretty low!), and to discuss where most of the input thermal energy is wasted. It's a good homework problem!

Let's go back to the steam turbine. Its spinning part, called *rotor*, is a sturdy shaft, on which a number of discs are mounted. Each discs has a number of blades attached to its perimeter, as shown in Fig. 3.3.1.

There are additional stationary elements of the turbine assembly, called *stator discs*, or *diaphragms*, inserted in between each pair of rotor discs.



Figure 3.3.1: The rotor of a steam turbine. Actually, here there are rather two rotors mounted on a single common shaft – which is an often used design. There are two high-pressure steam inlets in the middle, sending steam in opposite directions. Note that there are spaces between successive discs, into which *stators* (a.k.a. *diaphragms*) are inserted when the rotor is mounted in its casing.

A very instructive figures explaining how a rotor and the stator "halfdisc" are mounted in the turbine casing are shown in this Web page – see Figs. 6.19 and 6.21 in this document (essentially, the page is about gas turbines – but the arrangement of the rotor and stator elements is the same in both turbine types). The role of the stators is explained pretty well by an animation in this *Technology Partnership* Web page. In the animated graph seen in this page right after it opens up, there is a picture of a steam turbine with two rotor disc, and one stator disc in between. The steam that leaves the nozzles, moving (if looked at from the top of the picture) "to the right". It enters the spaces between the blades of the first rotor disc and passes some of its momentum to the blades. But in the process the steam is "redirected": now, looking from the top of the picture, it moves "to the left". But the next rotor disc again "expects" the incoming steam to "arrive from the right". And this is why the stator disc is for: the steam enters the spaces between the blades moving "to the left", but due to the curvature of the blades, it's redirected and leaves the stator disc "moving to the right".

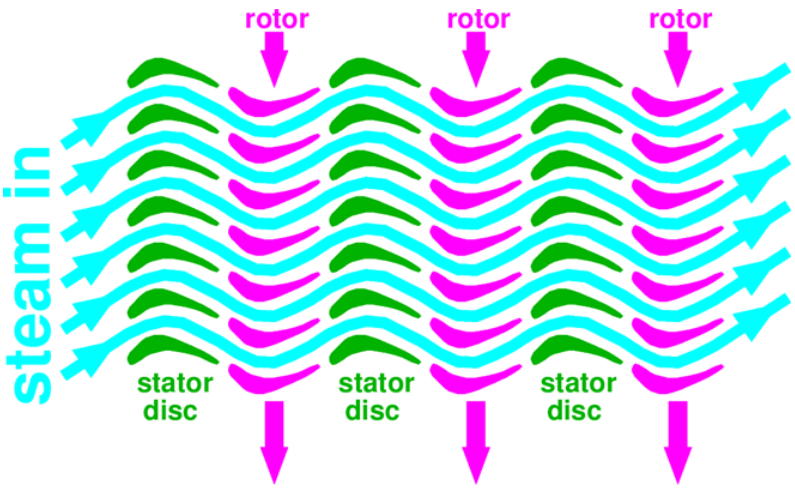


Figure 3.3.2: A graph explaining the role of the stator blades in a steam turbine (see the text).

Another graph explaining the role of the stator blades is shown in Fig. 3.3.2. The rotor blades are the purple ones, and the stator blades are green. The steam arriving from the left is first "redirected" up by the green static blades – so that they can "push" the purple rotor blades "downwards". But after exiting the space between the rotor blades, the steam moves "upwards". So, another passage between static blades redirects it "downwards" again – and so on.

In the Fig. 3.3.1 one can see that the turbine rotor has a sophisticated shape – in the region where the steam enters it has a small diameter, which gradually increases, and becomes much larger where the steam exits. Why is it so? As the steam gradually passes

its energy to the turbine rotor, its temperature and its pressure decreases. Decreasing pressure of a gas causes it to expand in volume; decreasing temperature has an opposite effect. But in a steam turbine the pressure drop has a far stronger effect than the temperature drop – therefore, the net effect is that the volume of the steam strongly increases in the course of its passage through the turbine – and therefore the last rotor disc is so much larger that the first one.

The turbine is only one part of a complete thermal power plant. A simplified scheme of a complete installation is shown in the Fig. 3.3.3. It must include, of course, a "steam source". By tradition, it's called *boiler*. In the old days, the boilers were huge water tanks mounted on the top of coal furnaces. Inside the tank, water was boiling – usually at pressure much higher than the atmospheric pressure. After "doing its work" in the engine, the exhaust steam was simply released into the air. When most water was used up, the tank had to be opened and refilled with fresh water. It was an inconvenient and time-consuming procedure. A significant improvement was the invention of a *condenser*. Now, the system became a "closed-loop" one: the exhaust steam was cooled down so it condensed back to liquid water, which was then re-injected by a special pump back into the boiler.

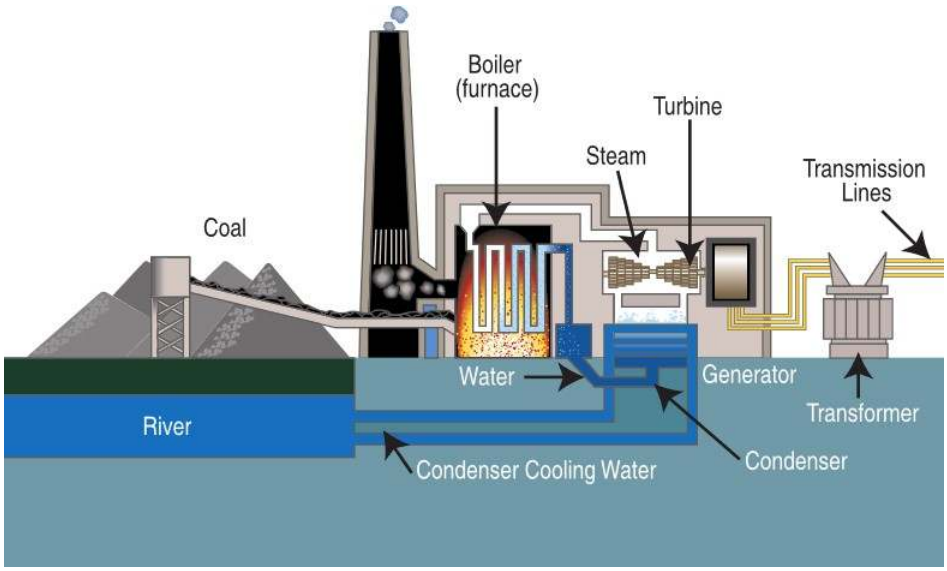


Figure 3.3.3: A simplified scheme of a modern coal-fired thermal power plant (from Wikimedia Commons).

Next, the engineers realized that a big-volume boiler was not necessary. Once the water circulated, a tank could be replaced by a system of coiled tubes. It had an additional advantage that higher pressure could be used. So, water could be heated up to higher and higher temperatures – up to the so called *critical point*, which for water occurs at 374°C (647 K), at pressure 218 times the atmospheric pressure (below the critical point water may still exist in liquid form; above it, only in gas form). And it was good, considering that the efficiency of a turbine increases with the increasing steam temperature (it will be discussed in closer detail in the next Section).

Finally, the engineers realized that not only liquid water could be heated. *If increasing the steam temperature improves the efficiency, why don't we heat up the steam even more?* – was their thinking. What was needed for that, was just adding more tubes to the "tank-less" boiler. And so the newest generation boilers were born. They produce the so-called *superheated steam*, of temperature about 550°C (823 K).

Also, the furnaces underwent a gradual evolution: in the old days, the coal was shuffled manually into them by workers called *stokers* – you may see them, e.g., in this YouTube clip about the *Titanic* steam engines, starting at 1 minute 38 seconds (by the way, in the same YouTube one can hear that the steam pressure in the boilers supplying the *Titanic's* giant piston-engines was 215 pounds per square inch (PSI), which is 14.6 atmospheres – and today the inlet steam pressure in state-of-the art steam turbines may be more than 10 times higher).



Gradually, the stokers were replaced by conveyor belts. And in the most advanced installations, coal is first the coal is first pulverized into fine dust, and then injected into the furnace using a jet of compressed air.

The alternative of using coal is to fuel the boilers with oil and natural gas. Especially, the latter option is attractive, because – as was shown in the preceding section – natural gas produces at least 3 times less CO_2 than coal for the same amount of power generated. Another reason is the ease of fuel transportation, gas can be delivered to the plant by a pipeline. Therefore, more and more coal-fired power plants in the US are being converted to natural gas fuel.

Gas turbines. The conversion of a coal-fired power plant to a gas-fired one involves changing the coal-burning furnace to a system of gas burners. The other parts of the boiler may stay. However, in principle it's possible to make a "deep conversion" in which the entire boiler and even the steam turbine is removed. The steam turbine may be replaced by a gas turbine that does not need any boiler!

In a gas turbine, air and natural gas of ambient temperature are pumped into a combustion chamber, where they are ignited. Natural gas is essentially a pure methane CH_4, and air is a mixture of 21% of oxygen (O_2), 78% of nitrogen (M_2), and 1% of argon ((Ar). Nitrogen and argon do not take part in the combustion reaction, and methane and oxygen change in the process to CO_2 and water vapor. Due to the high heat of combustion of natural gas, the temperature of the mixture of gases after the reaction is as high as 1600°C, and its pressure increases considerably. From the combustion chamber, it passes to the turbine, which has a very similar design to a steam turbine. The main difference is the temperature. In the latter, the inlet steam temperature is 550°C, or sometimes a bit higher – but it's still a temperature tolerated by steel. In a gas turbine, the blades must tolerate a much higher temperature of inlet gases, so they have to be made of special alloys. But besides, the design and operation of both types of turbines are very similar. As mentioned, the largest existing steam turbines can produce as much power as 1600 MW. Currently, single-unit gas turbines with record-high performance made by Siemens yield less, nearly 600 MW – but as far as their efficiency is concerned, they outperform steam turbines. The latter, when running on extremely high-temperature steam, may reach up to 35% efficiency, while from a single-unit gas turbine one can get even 45%.

And this is not the end, because gas turbines offer the possibility of working "in team" *in combination* with steam turbines, and then the efficiency of such a co-generation assembly may exceed 60%. How such a combination is set up, is explained in the Fig. 3.5.

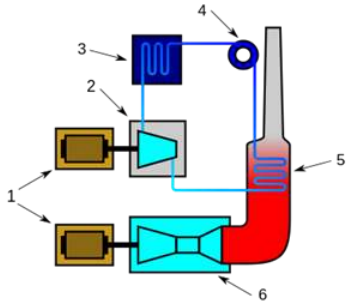


Figure 3.3.3: A scheme of co-generation power plant: 1 – electric generators, 2 – steam turbine, 3 – condenser, 4 – pump, 5 – steam generator, 6 – gas turbine.(from Wikimedia Commons).

The idea is simple: the exhaust gases from a gas turbine are still pretty hot (450 - 650°C), which is an ideal steam temperature for a modern steam turbine. Therefore, the exhaust gases are used for generating steam – much like in a boiler in a conventional power plant using steam turbines only. In practice, the power ratio of the two turbines making a “tandem” is close to 2:1. For instance, a typical combination is a 400 MW set consisting of a 270 MW primary gas turbine and a 130 MW secondary steam turbine. On April 28, 2016, a plant in in Bouchain, France, was certified by Guinness World Records as the worlds most efficient combined cycle power plant at 62.22%. However, General Electric is expecting to break the 65% level soon. Note that 65% means generating twice as much electricity from the same amount of gas fuel than a plant using steam turbines only would do. As you see, heat (or, rather thermal energy) from steam can be transformed to mechanical energy. And there is a range of other heat engine types that can be employed to perform many useful tasks (e.g., powering your car).

3.3: Heat Engines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

3.4: The Efficiency of Thermal Engines

3.4.1: The First Law of Thermodynamics.

3.4.2: Entropy

3.4.3: The Second Law of Thermodynamics

3.4.4: Sad Consequences of the Second Law

3.4.5: The Efficiency of an Ideal Thermal Engine

3.4.6: Even sadder news

3.4: The Efficiency of Thermal Engines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.4.1: The First Law of Thermodynamics.

As we have seen, heat (or, rather thermal energy) from fuels can be trans- formed to mechanical energy. And, let us say it one more time, there is a range of thermal engine types that can be “employed” to perform many useful tasks.

It is essentially what *The First Law of Thermodynamics* states. It’s the *good guy*, everybody is happy with it, But it’s not the only law that governs the transformation of heat to work. The other one is the *Second Law of Thermodynamics*, and this is the *bad guy*.

This Section is mainly about the Second Law and its consequences – but before we start talking about it, we should briefly recapitulate what the First Law says in a more exact manner. We need some definitions:

- **A system:** A system: a single body, or more bodies that are in contact with one another.
- **Internal Energy:** There is a physical quantity called the *internal ther- mal energy* of a system or “internal energy” in short. Conventionally, it is denoted as U .
- Energy may be added to the system, thus increasing its U . (we call such a process “heating”) – either through a contact with another system, or by performing work on the system (for instance: if the system is a portion of gas, one way of performing work on it is *compressing it*).
- Energy may be taken away from a system, thus lowering its U (we call such a process “cooling”) – either through a contact with another system, or by causing the system to perform work (for instance: if the system were a gas, it would perform work by *expanding*). However, the process of taking energy away can not continue forever, when all energy is taken away, the system reaches the *absolute zero* point, at which its temperature is zero Kelvins (0 K).

Now, we can write the First Law in mathematical terms: essentially, it’s the Energy Conservation Law, but expressed in a way specifically applying to thermal phenomena:

$$\Delta U = \Delta Q + \Delta W \tag{3.4.1.1}$$

where:

ΔU : The total change in the system internal energy.

ΔQ : The change of U due to transfer of heat (heat flowing in or out from another system, in contact with the system considered).

ΔW : The change of U due to mechanical work done on the system, or the work delivered by the system (then, negative).

Important: A common misconception is to confuse heat with the internal energy. Internal energy is the amount of energy contained by the system. Heat is the energy that flows in or out from/to a warmer/cooler body which is in contact with the system.

3.4.1: The First Law of Thermodynamics. is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.4.2: Entropy

The First Law was an easy part. But in order to explain what the Second Law talks about, we have to introduce the notion of *Entropy*.

Entropy is widely regarded as one of the most difficult concepts in academic physics curriculum. It's a parameter that characterizes the state of a thermal system. Other state parameters are the internal energy U, volume V, the amount of substance (usually expressed as the number of moles N a mole consists of 6.022×10^{23} molecules [1] of a given substance). Furthermore, the temperature T, and the pressure p. They are all "intuitively clear", right? In contrast, entropy, conventionally denoted as S, is an abstract function.

It's mathematically defined, and the definition is not very complicated. Namely, if a portion of heat ΔQ is transferred in or out of a body of temperature T (either an individual body, or a body being a part of a larger system, otherwise isolated) the change ΔS in the body's/system's entropy is:

$$\Delta S = \frac{\Delta Q}{T} \tag{3.4.2.1}$$

So, by adding heat one can increase the entropy of a system, and by removing heat e.g., through contact with a cooler body one can lower the systems entropy.

An Example. Perhaps a good thing to do now would be to discuss a simple example. OK, consider a $1\ \mathrm{kg}$ sample of ordinary water, H_2O. At very low temperature it's a solid piece of ice. If we start adding heat to it by small portions ΔQ, each time its entropy will grow by $\Delta S = \Delta Q/T$, according to the Equation 3.7. But when we add heat, we also add energy, right? So, after each step, the temperature will also grow a bit.

As the process continues, the temperature will gradually increase, as illustrated in the Fig. 3.6 - until it reaches the ice melting point $T_f = 273\ \mathrm{K}$. In order to melt a mass unit of ice, one has to add an amount of heat called the heat of fusion, equal $334\ \mathrm{kJ/kg}$, or $6.01\ \mathrm{kJ}$ per one mole of water (the mass of one mole of H2O is $18.01528\ \mathrm{g}$).

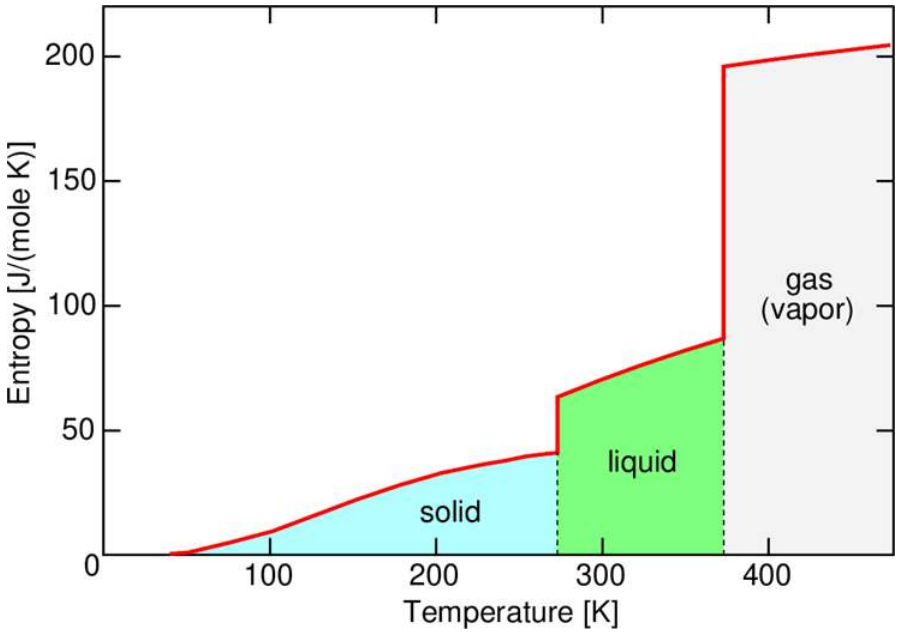


Figure 3.4.2.1: The change of entropy of one mole of water, when heat is grad- ually added by small portions, starting from a very low temperature. When the ice melting point is reached (273 K), the temperature does not change until all ice melts down – which produces the vertical “step” on the curve. Another such “step” occurs at water boiling temperature – again, the tem- perature of 373 K does not change as long as all water “boils off” – this transition require adding 6.75 times more heat than in the case of ice melt- ing, therefore the “step” is considerably higher.

Now, the temperature stops growing, until the small portions ΔQ add up to $334\ \mathrm{kJ}$ (for a $1\ \mathrm{kg}$ sample) or $6.01\ \mathrm{kJ}$ for a single mole. But the entropy keeps growing during the process, as shown by the vertical part of the curve in the Fig. 3.6. During the entire

melting process, it increases by $334\ \mathrm{kJ/273\ K} = 1.22\ \mathrm{kJ/K}$ (or by $6.01\ \mathrm{kJ/273\ K} = 22.01\ \mathrm{J/K}$ for one mole). After all ice changes to water, the temperature starts increasing again, until it reaches the boiling temperature $T_{\mathrm{b}} = 373\ \mathrm{K}$. Again, the temperature stops growing, because in order to change one kg of liquid water to water vapor, the energy of $2256\ \mathrm{kJ}$ is needed (or $40.68\ \mathrm{kJ}$ per one mole). So, again the temperature stops growing, until all small portions ΔQ add up to this value - but the entropy in the course of this process increases by $2256\ \mathrm{kJ/373\ K} = 6.05\ \mathrm{kJ/K}$ per one kilogram (or by $40.68\ \mathrm{kJ/373\ K} = 109.06\ \mathrm{J/K}$ for one mole), creating another "vertical jump" on the curve, much higher than that at $273\ \mathrm{K}$. After that, the temperature of water, now in gas (vapor) form, will again start increasing.

The purpose of the above example is to show that the value of entropy can be obtained from calculations. In the Web one can find tables and online calculators.

Another example The process of water heating described above is reversible. If we started to take away small portions of heat, we would return to the original state along the same curve. There are, however, processes that cannot be reversed - called irreversible. Consider the following example: there are two $1\ \mathrm{kg}$ blocks of aluminum, placed into a container that is totally impermeable to heat - no heat can flow in or flow out. The blocks initially do not touch one another. The temperature of one of them (Block 1) is T_1, and of the other (Block 2) is T_2, with $T_1 > T_2$.
The relation between the amount of heat ΔQ that flows in or out of aluminum, and the change ΔT of its temperature is caused by the process is:

$$\Delta Q = C_{\mathrm{Al}} \Delta T \text{ or } \Delta T = \frac{\Delta Q}{C_{\mathrm{Al}}} \tag{3.4.2.2}$$

where C_{Al} is a "material constant" called heat capacity. For Aluminum, its value is $900\ \mathrm{J/kgK}$.

If we now bring the two blocks together to establish thermal contact between them, heat will start flowing from Block 1 into Block 2 . The process will last until the temperature of both gets equal. Call this temperature T_{12}. An obvious guess is that it is equal to the arithmetic average of T_1 and T_2, i.e.,

$$T_{12} = \frac{T_1 + T_2}{2} \tag{3.4.2.3}$$

To check whether it's a right guess, lets calculate the heat transferred out of Block 1 and heat transferred into Block2. The energy conservation law states that they must be equal! OK, let's then see. First, we find the change of the temperature of Block 1. An algorithm for a temperature change ΔT is always: the final temperature minus the initial one. So, we get:

$$\Delta T_1 = T_{12} - T_1 = \frac{T_1 + T_2}{2} - T_1 = -\frac{T_1 - T_2}{2} \tag{3.4.2.4}$$

So that the heat transfer out of Block 1 is, according to the Eq. 3.8 : $\Delta Q_{\text{out of Block 1}} = -C_{\mathrm{Al}} \frac{T_1 - T_2}{2}$

In a similar fashion, we get:

$$\Delta T_2 = T_{12} - T_2 = \frac{T_1 + T_2}{2} - T_2 = \frac{T_1 - T_2}{2} \tag{3.4.2.5}$$

and $\Delta Q_{\text{into Block 2}} = C_{\mathrm{Al}} \frac{T_1 - T_2}{2}$

So, everything is correct - the same amount flows out of Block 1 (therefore the minus sign) as it flows in Block 2.

However, let's now see what happens with the entropy. Let's focus on the very beginning of the process, when only a tiny amount of heat dQ managed to get out of Block 1, and into Block 2. So, the change of entropy of Block 1 is $dS_1 = -dQ/T_1$ (minus, because dQ flows out of Block 1), and of Block 2 is $dS_2 = dQ/T_2$. The total change of entropy of the system of two block is the sum of the two:

$$dS_{\mathrm{tot}} = dS_1 + dS_2 = -\frac{dQ}{T_1} + \frac{dQ}{T_2} = dQ\left[-\frac{1}{T_1} + \frac{1}{T_2}\right] = dQ\left[\frac{T_1 - T_2}{T_1 T_2}\right] \tag{3.4.2.6}$$

Note that the last right-hand expression in square brackets is a non-zero positive number. What does it mean? It means that the entropy of thetwoblock system increases in the course of the heat transfer process. But not only during the initial short period - the same math remains valid for any moment until the heat stops flowing. So, the conclusion is that this process of heat exchange between the blocks increases the entropy of the two-block system.

The exchange of heat between two bodies is always an irreversible process: the heat flows spontaneously from the warmer to the cooler body, but the opposite may never happen. There are many other types of irreversible thermodynamic processes, and all have one in common: if such processes occur in thermally isolated system (often called adiabatic by professionals), the entropy of the system always increases. Entropy may stay constant only if all processes in an adiabatic system are fully reversible. And the entropy in an adiabatic system can never ever decrease!

---

1 Why such an "exotic" number? Well, once it has been decided that a mole of hydrogen, the simplest of all elements, would be the number of hydrogen atoms in a mass unit, a *gram* (g). And there are $6.022 \times 10_{23}$ atoms in a gram of hydrogen – it's called the *Avogadro Number*. Then it has been decided that a mole of any other pure substance is the Avogadro Number of molecules of this substance.

3.4.2: Entropy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.4.3: The Second Law of Thermodynamics

Why do we pay so much attention to entropy? Why is it so important? Well, it's because the entropy is – let's use such a metaphor – the "leading character" in the Second law of Thermodynamics.

One scientist once counted all non-identical formulations of the Second Law he was able to find – and his result was about one hundred. But non- identical does not mean that they are different: no, no, they are all equivalent! They all convey exactly the same information, only the differently (sometimes, very differently!) formulated. In some of them entropy is explicitly mentioned, and in some of them it is not. For instance, one of this author's favorite formulation of the Second Law is the following: *It's not possible to build a Perpetual Motion Machine of the Second Kind*.

In fact, what the second law states, is expressed in the last paragraph of the preceding Section. In a more compact form, one can say: *In a closed thermodynamic (adiabatic) system, the entropy may either remain constant, or increase*. Or, even in a compacter form:

> *In a closed thermodynamic (adiabatic) system,* $\frac{dS(t)}{dt} \geqslant 0$

If the time derivative of a function of time is zero, it means that the function does not change in time; and if the derivative is greater than zero, it means that its value is increasing in time.

Why does the Author so stubbornly stress each time that the Second Law applies only to thermally isolated, i.e., adiabatic systems? Because he knows from his long teaching practice that students often forget to add this clause. The Author himself forgot about it at the thermodynamics exam at his senior year. Fortunately, other answers were correct, so the exam ended up with a B. Some instructors treat the omission of this clause as a cardinal error (not the Author, remembering his own fault – he believes that giving the student "another chance" is a better way of dealing with the problem).

In a system which is not thermally isolated, the entropy may decrease. Take a glass of water and on a cold winter day leave it outside. Water will freeze, its entropy will lower considerably. It's all OK.

3.4.3: The Second Law of Thermodynamics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.4.4: Sad Consequences of the Second Law

After the First Law has been discovered, some people started thinking: *if heat is equivalent to mechanical work, we should be able to take a big reservoir of thermal energy – such as, e.g., an ocean – and extract work from it. There is so much thermal energy in any ocean that there should be enough mechanical work for everybody for thousands of years!*

A machine that would extract thermal energy from oceans and convert it to usable work was given the name of *Perpetual Motion Machine of the Second Kind (PMMoSK)*. The first kind was a machine that would produce work out of nothing.

Some inventors seriously believed that making such a machine would be possible. They spent much time working on projects and prototypes. The failure of all such efforts was certainly one of the factors that helped to discover the second law of thermodynamics.

Who has “derived” the Second Law? Not a human scientist. It was Mother Nature who decided that things should be that way. The role of human scientist was only that by analyzing a wide variety of facts they reached a firm conclusion that there is such a law of nature[1].

A simple explanation why it is not possible to build a PMMoSK is presented in Fig. 3.4.4.1. Consider an isolated homogeneous system. An amount of work ΔW can be added to it to increase its internal energy U - the result is the same as if one added an equivalent amount of heat ΔQ, so that the entropy of the system increases by $\Delta S = \Delta W/T$.

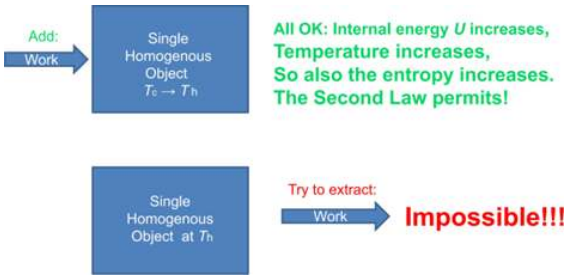


Figure 3.4.4.1: Adding work to a homogeneous isolated system is possible - but extracting energy at the expense of its energy U is not.

The increase of entropy is permitted by the Second Law, so everything is OK with the process. But not with an opposite one! Extracting a portion of work ΔW from the system would be equivalent with taking a portion of entropy $\Delta S = \Delta W/T$ away from it - so, in consequence, to lowering the system entropy, which is strictly forbidden by the Second Law.

---

1. In fact, other laws of physics were discovered in the same way. For instance, Newton did not “derive” his famous second law of dynamics, $F = ma$. By analyzing observations, he concluded that there is such a law of nature. And many generations of physicist later strengthened his conclusion by performing zillions of tests. Something *derived* by scientists may only be a *theorem*, not a law.

3.4.4: Sad Consequences of the Second Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.4.5: The Efficiency of an Ideal Thermal Engine

But there are devices converting heat to mechanical work! Yes, there are. There are ways of outsmarting the Second Law. But for a price!

Consider an isolated system (Fig. 3.4.5.1) containing a heated body. The only way of extracting work from a heated body of temperature T_{h} is to build a more complicated system that would make it possible to keep the entropy at a constant level. The “trick” is as follow: the system should contain, in addition, a “heat engine” that draws a “portion” of heat ΔQ_{h} from the heated body, and converts some part of it - very important, some part of it only! - to mechanical work. And there must be a cold body or a "heat sink" of temperature T_{c} necessarily lower than T_{h} where the remaining heat can be absorbed ("dumped").

The entropy taken from the heated body is:

$$\Delta S_{\mathrm{h}} = \frac{\Delta Q_{\mathrm{h}}}{T_{\mathrm{h}}} \tag{3.4.5.1}$$

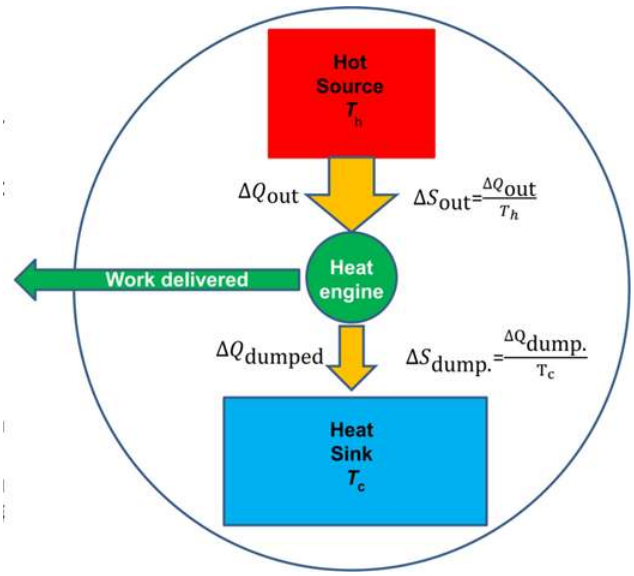


Figure 3.4.5.1: An isolated system containing hot body, an ideal thermal engine, and a cold body, used for deriving the equation for the maximum possible thermal engine efficiency.

The engine is using some part of ΔQ_{h} to do work ΔW, so that the remaining heat "to be dumped" to the cold body is:

$$\Delta Q_{\mathrm{c}} = \Delta Q_{\mathrm{h}} - \Delta W \tag{3.4.5.2}$$

So the entropy ΔS_{c} passed to the cold body is:

$$\Delta S_{\mathrm{c}} = \frac{\Delta Q_{\mathrm{c}}}{T_{\mathrm{c}}} = \frac{\Delta Q_{\mathrm{h}} - \Delta W}{T_{\mathrm{c}}} \tag{3.4.5.3}$$

The total entropy cannot decrease, so it must be:

$$\Delta S_{\mathrm{c}} = \Delta S_{\mathrm{h}} \tag{3.4.5.4}$$

By substituting for ΔS_{c} and ΔS_{h} the results from the Eqs. 3.9 and 3.10, we get:

$$\frac{\Delta Q_{\mathrm{h}}}{T_{\mathrm{h}}} = \frac{\Delta Q_{\mathrm{h}} - \Delta W}{T_{\mathrm{c}}} \tag{3.4.5.5}$$

By performing simple algebraic operations, one can convert the above to:

$$\frac{\Delta W}{\Delta Q_{\mathrm{h}}} = 1 - \frac{T_{\mathrm{c}}}{T_{\mathrm{h}}} \tag{3.4.5.6}$$

Note that the left side of the above equation is the ratio of the mechanical energy delivered by the heat engine to the thermal energy taken from the hot body - in other words, it is the efficiency of converting the input heat to the output mechanical energy. We can

write then:

$$\epsilon_{\text{conv}} = 1 - \frac{T_{\text{c}}}{T_{\text{h}}} \tag{3.4.5.7}$$

Or, we often prefer to express the efficiency in percents, then this equation takes the form:

$$\epsilon_{\text{conv}}[\%] = \left(1 - \frac{T_{\text{c}}}{T_{\text{h}}}\right) \times 100\% \tag{3.4.5.8}$$

This result is known as the Carnot Law[1], in honor of Nicolas Leonard Sadi Carnot (1796-1832), a French engineer whose works published in 1824 contributed greatly to discovering the Second Law of Thermodynamics and to understand the role of entropy in thermal phenomena.

What has been said by now, may raise one question: is it possible to build an engine consistent with all the assumptions used in the considerations above? It's a good question! But, fortunately, the answer is yes. A "prototypical" engine of such kind, known as the Carnot Engine, uses a gas as the "working fluid". Later, many other types of Carnot Engines were conceived, not necessarily using a gas working fluid.

"Sadi" was perhaps a good middle name for the discoverer of this "law", because, regretfully, it brings us a sad message... Namely, no heat engine can attain a higher efficiency of converting thermal energy to work than that permitted by the Carnot Law. The consequence are not so pleasant... Lets consider a modern power plant, which uses steam turbines. The highest temperature of steam from "state of the arts" flame-heated boilers is $t = 650^{\circ}\text{C}$, which translates to $650 + 273 = 923\ \text{K}$. And the exhaust steam leaves the turbine at atmospheric pressure, so that its temperature cannot belower than 100°C - therefore, the "heat sink" temperature should be taken as $\sim 100^{\circ}\text{C} = 373\ \text{K}$. We get:

$$\epsilon = 1 - \frac{373\ \text{K}}{923\ \text{K}} = 0.596 \tag{3.4.5.9}$$

It means that only about then 68% of the thermal energy "invested" is converted to work, and about 32% "goes down the drain", i.e., is dumped in the heat sink. Not a brilliant performance, you may think, but one can survive with such an efficiency...

---

1. As noted before, from an orthodox viewpoint it should not be called a law, but rather a theorem. But it is difficult to overcome a tradition. The Author will not try, and therefore the formulation Carnot Law will be used further on.

3.4.5: The Efficiency of an Ideal Thermal Engine is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

3.4.6: Even sadder news

But there are more bad news for you: namely, the “Carnot efficiency” is correct only in highly idealized situations. One can build engines which would obey the Carnot Law, yes look at the following Web sites: e.g., a NASA Web site, the renowned “Hyperphysics” site, or this site in Electropaedia (I like the British “Electropaedia”, one can find good “non-nonsense” articles over there).

However, such laboratory-built engines have to work extremely slowly in order to deliver output work consistent with the Carnot Law. “Extremely slowly” means that they, yes, deliver work – but not **power**. And power is what we really need! We need engines that produce **maximum power** from a given amount of thermal energy! To make the long story short: one can make power-maximizing heat engines, there is even a special theory of such engines in thermodynamics, they are called “endoreversible heat engines”. The thing is that their operation involves processes which the science of thermodynamics recognizes as **irreversible** and their nasty effect is that they produce an additional portion of en- tropy. This extra entropy also has to be removed from the engine, so that even more heat has to be “dumped” into the heat sink. The result is that even less heat can be converted to output work. In short, the efficiency of a power-maximizing heat engine is given by the Chambadal-Novikov formula:

$$\epsilon_{\mathrm{C-N}} = 1 - \sqrt{\frac{T_{\mathrm{c}}}{T_{\mathrm{h}}}} \tag{3.4.6.1}$$

Novikov and Chambadal are the two gentlemen who in 1957 independently did pioneering theoretical studies on power-maximizing engines. The theory is quite complicated, it will not make sense to discuss its details over here if you are interested, you may find more in the following Web sources: Endoreversible thermodynamics, or in this article - as well as in references listed in these two sources. The bad news is that, as noted, the above theory is "pretty complicated" but the good news is that the final theoretical formula is pretty similar to the Carnot Equation note that there is only an extra square root symbol! So, it's not the original Carnot's equation, but the Chambadal-Novikov formula we should use for estimating the efficiency of practical heat engines. In the example we have considered above, we should use:

$$\epsilon_{\mathrm{C-N}} = 1 - \sqrt{\frac{373\ \mathrm{K}}{923\ \mathrm{K}}} = 0.364 \tag{3.4.6.2}$$

It means that not 40% of the energy released from burning fuel, but as much as 64% of this energy "goes down the drain"!

Table 3.4.6.1. Comparison of the real efficiency of thermal plants with calculations using the Carnot and the Chambdal-Novikov formulae (the symbol η means the same as E in the text) (from the Wikipedia article quoted above).

Power Plant	T_L (°C)	T_H (°C)	η(Carnot)	η (Endoreversible)	η (Observed)
West Thurrock (UK) coal-fired power plant	25	565	0.64	0.40	0.36
CANDU (Canada) nuclear power plant	25	300	0.48	0.28	0.30
Larderello (Italy) geothermal power plant	80	250	0.33	0.178	0.16

As follows from the data in the table (from the Wikipedia site linked above), the Chambadal-Novikow formula yields results that are pretty close to the real thermal efficiencies attained in real power plants. But the 36% efficiency appears to be even too high for most existing thermal power plants, due to extra losses of heat in not-too-well engineered installations the real efficiency in such plants is seldom higher than 30%. 70% of thermal energy released goes down the drain!



3.4.6: Even sadder news is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

4: Nuclear Power

4.1: A Note from the Author
4.2: Nuclear Physics- Brief Review of Fundamental Facts
4.3: Radioactive Decay
4.3.1: Energy Units in Nuclear Physics
4.3.2: The Decay Constants and Half-Lives
4.4: Nuclear Cross Sections
4.5: Nuclear Fission
4.5.1: Fission and Fusion Energy
4.5.2: How to calculate the mass defect?
4.5.3: Energy Released by the Fission of a Single Nucleus
4.5.4: Nuclear Fusion
4.6: Controlling the Fission Chain Reaction- Nuclear Reactors
4.6.1: Essential Cross Sections
4.6.2: No magic tricks are necessary!
4.6.3: Controlling the Chain Reaction and Cooling the Assembly
4.6.4: Power Reactors
4.6.5: Generation IV Reactors and Thorium Fuel
4.7: NuScale's Small Modular Reactors – an Oregon-born Project
4.8: The Pros and Cons of Nuclear Power

4: Nuclear Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.1: A Note from the Author

Of all members of the whole “menagerie of powers”, i.e., of various methods of harnessing energy resources, nuclear power is surely the most controversial one, and it evokes the strongest emotions.

Therefore, the Author of this text believes that it may be appropriate to begin the present Chapter with **Personal Statement**. Here it comes: When I was an active researcher in the area of experimental physics yes, I was, for more than 40 years of my professional life the topic of my research was to investigate different phenomena in solid state physics by means of neutron scattering. And the principal sources of neutron radiation that can be used for such studies are nuclear reactors. So, over the period of 40 years I often spent hours and hours in close proximity of a nuclear reactor (actually, of eight reactors in five different countries). And nothing harmful ever happened to me. So, I am “positively biased”: when I think about nuclear reactors, I don’t think of them as of bloodthirsty wild animals, but rather as of “domestic animals”. Therefore, I know that when I lecture about nuclear power, I have to be extremely cautious. I told you at the beginning that the “battle cry” in this course if ANP, it is, “Absolutely no propaganda!”. I will do my best, then, to avoid sneaking in any propaganda when talking about nuclear power I will present facts only, facts that can be easily verified in widely available sources. OK., this is the end of my personal statement!

4.1: A Note from the Author is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.2: Nuclear Physics- Brief Review of Fundamental Facts

We have to start with a brief review of fundamental facts in nuclear physics. So, atoms consist of “shells” of electrons, which surround a nucleus a tiny col- lection of particles known as “nucleons”: there are two kinds of them, namely, protons (positively charged particles, with a charge of the same magnitude as the negative electron charge) and neutrons (particles only slightly more massive then protons, but with no electric charge). The size of an atom is of the order of 10^{-10} m – one tenth of one billionth of a meter. The size of a nucleus is about one hundred thousand times smaller than the size of an atom.

So, in summary:

- An atom consists of a positive nucleus, and a “shell” of negative elec- trons;
- A nucleus consists of positive protons and neutrons that have no charge. The proton and electron charges are of the same magnitude, but of opposite signs;
- The number of protons in the nucleus is the same as the number of electrons in the electronic shell, so that the atom has a zero net charge.
- There is one common name for the protons and neutrons comprising an atomic nucleus – the *nucleons*. As, for example, in a kindergarten class: one may say: *There are 10 girls and 8 boys in the class*, or: *There are 18 children in the class*.

Particles with positive electric charge repel one another, and protons in the nucleus are so closely packed that these repulsive forces are very strong. So, in atomic nuclei there is an extra **attractive** force that holds all the constituent protons and neutrons together. It’s known as the “strong force”. It would be too long a story to talk about this force in our course but if you are interested, much material about strong nuclear forces can be found in the Web, e.g., in this Wikipedia article. The only important fact that we need to point out here is that neutrons play an important role in holding the nucleus together. With the exception of the hydrogen nucleus, consisting of a single proton, there are no other nuclei with no neutrons. And the number of neutrons cannot be too low or too high. The rules are not simple, it’s better to use a graph to explain what is going on.

Figure 4.2.1 shows the so-called “Chart of known nuclei”. There are stable nuclei, and radioactive (unstable) nuclei, that decay after some time. The “lifetimes” of known unstable nuclei span from microseconds to billions of years.

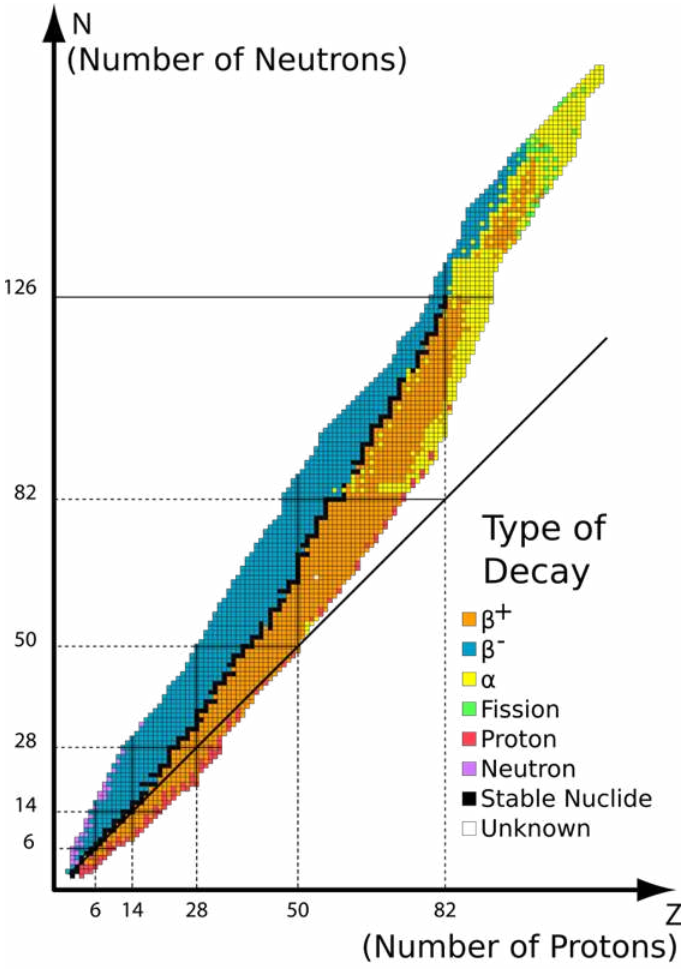


Figure 4.2.1:: Chart of known nuclei showing stable and unstable ones, with the stable ones forming the so-called “stability path” or “stability valley”(black data points). The color of the other data points indicate the type decay of unstable nuclei. The types listed are described in the text, with the exception of the decay by emission of a proton, or a neutron – those two modes of decay are extremely rare (from Wikimedia Commons).

The stable nuclei – we know 253 of them – form the so-called “stability path” (black squares in the graph). As you can see, the ratio of the number of protons (Z) to the number of neutrons (N) is about 1 for the light stable nuclei, and about 2/3 for the heavy ones. There are over 5000 known radioactive (unstable) nuclei, about 350 of them occurring naturally, and the rest produced artificially, most often by bombarding stable nuclei with high-energy protons or ions.

The graph shows the properties of nuclei depending on the number of constituent neutrons (on the abscissa) and protons (on the ordinate). Black dots indicate stable nuclei, blue dots unstable nuclei, i.e. such that have a limited lifetime only and eventually undergo a radioactive decay.

Notation: By convention, Z is the number of protons in a nucleus. Because each proton has a positive electric charge of the same magnitude as the negative charge of an electron, Z can also be thought of as the total charge of a given nucleus expressed in the units of electron charge. N is the number of neutrons, and A is the total number of all nucleons therefore, $A = Z + N$. Also, the convention is that for describing nuclei we use the chemical symbol and put Z as the left subscript and A as the left superscript: ${}^{A}\mathrm{X}$. One small comment: Z, the number of protons in the nucleus is equal to the number of electrons in a given atom, because atoms are electrically neutral. So, in fact, writing Z in the nucleus’ symbol is redundant. Therefore, Z is often omitted for instance, instead of ${}^{12}\mathrm{C}$, we more often write simply ${}^{12}\mathrm{C}$. Another small comment: sometimes people instead of using the left sub- and super- scripts (its the official way, sanctioned by international organizations) put Z and A as **the right** sub- and superscripts, e.g., C^{12}, or simply C^{12}. Such “unorthodox” notation seldom leads to misunderstandings, though. By the way, there is an often-used alternative way of describing nuclei – one uses not the chemical symbol, but the full name of the element, starting with a capital, and then a dash and the A number: e.g., Carbon-12, Carbon-14, or Uranium-235.

The subscript + superscript convention is also used for single particles. The proton and the neutron are both nucleons, so for both A = 1. For protons Z = 1, and for neutral neutrons Z = 0, so for them we use the symbols:

$$ {}_{1}^{1}p \text{ and } {}_{0}^{1}n \tag{4.2.1} $$

respectively. The electron is not a nucleon, but since electrons take part in several types of nuclear processes, one can use the same notation for them, with $A = 0$ (meaning: "not a nucleon"), and $Z = -1$:

$$ {}_{-1}^{0}e \text{ or } {}_{-1}^{0}\beta \tag{4.2.2} $$

are two equivalent symbols. Finally, there are two more particle types created in nuclear processes: gamma radiation quantum, and neutrinos. Neither is a nucleus nor has is an electric charge, so their symbols are:

$$ {}_{0}^{0}\gamma \text{ and } {}_{0}^{0}\nu \tag{4.2.3} $$

respectively.

Note that the number of neutrons in stable nuclei is either equal to the number of protons ($N = Z$, as in some light nuclei, such as, e.g., in carbon ${}_{6}^{12}\mathrm{C}$), or greater , $N > Z$. There is only one known exception from this rule, a stable nucleus with two protons and only one neutron: $Z = 2$ identifies it as helium, so that the full symbol is ${}_{2}^{3}\mathrm{He}$ (it's an extremely rare isotope of helium, in ordinary helium nucleus there are two neutrons). For heavier nuclei, the number of neutrons grows faster, than the number of protons. In the heaviest of all stable nuclei, an isotope of lead, $Z = 82$ and $A = 208$ $\left({}_{82}^{208}\mathrm{Pb}\right)$, so that he N/Z ratio in it is $(208 - 82)/82 = 126/82 = 1.54$.

Isotopes: Atoms with nuclei in which the number of protons (Z) is the same, but with different numbers of neutrons, are called isotopes of the same chemical element. The chemical properties of atoms depend on the number of electrons in their shells or, equivalently, on the number of protons Z in their nuclei. Atoms with different number of electrons are different chemical elements:

$$ \begin{array}{l} Z = 1: \text{ Hydrogen;} \\ Z = 2: \text{ Helium;} \\ Z = 3: \text{ Lithium;} \\ Z = 4: \text{Beryllium;} \\ Z = 5: \text{ Boron;} \\ Z = 6: \text{ Carbon;} \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ Z = 92: \text{ Uranium} \end{array} \tag{4.2.4} $$

However, there may be several nuclei with the same number of protons Z, but a different number of neutrons N. Atoms with such nuclei have all the same number of electrons, which determines their chemical properties. They are all atoms of the same chemical element, but they are different isotopes of that element. Some even have special names: $Z = 1, N = 0$ is the "ordinary" Hydrogen; $Z = 1, N = 1$ is Deuterium (stable); $Z = 1, N = 3$ is Tritium (radioactive). But for Z greater or equal 2 different names are no longer used. Instead, they have the same name and are distinguished by the mass number A - e.g.: Carbon-12, Carbon-13, and carbon-14; or Uranium-233, Uranium-235, and Uranium-238.

4.2: Nuclear Physics- Brief Review of Fundamental Facts is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.3: Radioactive Decay

Learning Objectives

- To learn how to edit a libretext
- To get credit

Decay processes

Radioactive decay is a process in which a nucleus of an unstable isotope, called the “parent nucleus”, loses energy by emitting radiation, and trans- forms itself into a nucleus of a “daughter isotope”. There are several types of radioactive decay, listed below.

One important case, not shown in the Fig. 4.3.1, is the decay of a *neutron*. The neutron is a single particle, not a nucleus, but an extremely important component of all atomic nuclei, with the exception of the simplest one, the nucleus of ordinary hydrogen, which is a single proton. Protons repel each other due to their electric charge $+e$. All nuclei with two or more protons do contain neutrons, so that – in a naive picture – they may be thought of as “glue particles” that neutralize the repulsive forces between protons and hold the nucleus together. In fact, the nucleus is held together by the “strong force”, an interaction much stronger than the repulsive Coulomb force, but of a very short range.

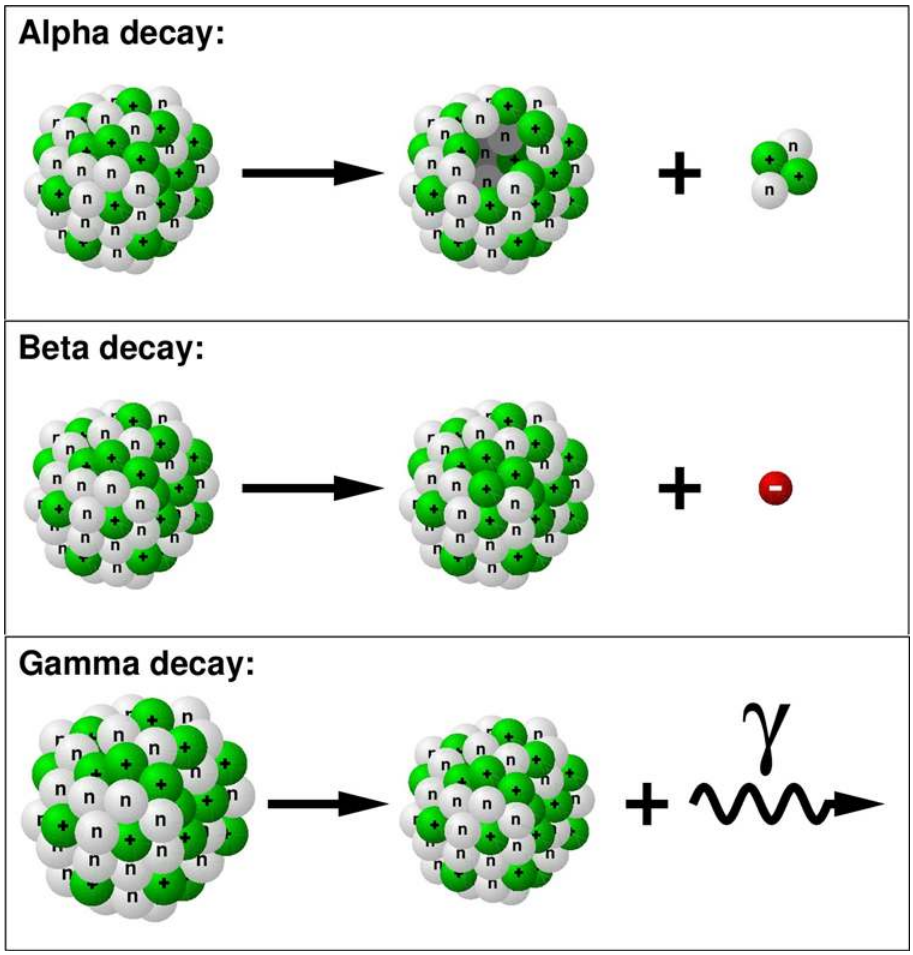


Figure 4.3.1: The three most often occuring modes of nuclear decay: α, β minus and γ. In the case of the first two the decay produces a new nucleus (see the text for more details), but in γ decay the number of neutrons and protons is the same before and after the act – the transition is from an excited state with higher energy to a state with lower energy. However, a question arises, how to depict an “excited state” of the atomic nucleus in the figure? We decided to use a metaphorical comparison with a balloon: by inflating a balloon, we increase its energy and size; and if we release the air, the energy decreases and the size shrinks. but it remains the same balloon! So the larger nucleus in the figure symbolizes one in the state of a higher energy – and the smaller is the same nucleus in a lower energy state, after a portion of energy is “released” – or jettisoned – in the form of a quantum of electromagnetic radiation, called a “gamma quantum”. But in real gamma decay it doesn’t have to be that the nucleus “shrinks”!.

A free neutron is not stable, it decays into a proton, an electron and a neutral particle called *neutrino*[1] (symbol: *v*), with an average lifetime of about 15 minutes (it's called a "beta-minus", or a β^-decay). However, in stable nuclei the neutrons never decay. In contrast, in some unstable nuclei they do decay - see the list below - but their average lifetime in the nucleus may be either shorter, or longer than in the free state (sometimes, as long as many years).

As far as nuclei are concerned, there is a number of possible decay modes:

- The Alpha (or a) decay - mostly in heavy nuclei. The nucleus emits an a-particle, which is the same as the nucleus of the light element Helium4 $\left({}_2^4\mathrm{He}\right)$. It means that the parent nucleus transforms into a daughter nucleus with 4 nucleons less, and 2 protons less. It can be written in the form of an equation, with X and Y symbolizing, respectively, the parent and the daughter nucleus:

$$ {}_Z^A\mathrm{X} \rightarrow {}_{Z-2}^{A-4}\mathrm{Y} + {}_2^4\mathrm{He} + \text{ kinetic energy} \tag{4.3.1} $$

  In the Fig. 4.3.1 we drew the "daughter nucleus" as if the detached alpha particle left a "pothole" on its surface. Of course, in a real alpha decay such a pothole is unlikely to be created - and we only drew it so that it could be clearly seen that the "daughter nucleus" is the "parent" one minus four nucleons.
- The beta decays (there are more than one possible modes:

  * The most common is the beta-minus (β^-)decay: The nucleus emits an electron, ${}^0e$, and a neutrino v. The number of nucleons (A) does not change, but the number of protons increases by 1 , and the number of neutrons (N) decreases by 1 :

$$ {}_Z^A\mathrm{X} \rightarrow {}_{Z+1}^{A}\mathrm{Y} + {}_{-1}^0e + {}_0^0\overline{\nu} + \text{ kinetic energy} \tag{4.3.2} $$

  It means that one of the constituent neutrons undergoes a β^- decay the same way as a free neutron does. Note that in the illustration of a beta minus decay in Fig. 4.3.1 a nucleon that is a neutron in the "parent" nucleus (right in its center) is a proton in the "daughter" nucleus. By the way, the neutrino is not shown in Fig. 4.3.1.

  *The less common beta-plus (β^+)decay. Now the nucleus emits a positron - a particle which is the "twin brother" of the electron: it has the same mass, but a positive charge of the same magnitude as that of the electron. It means that one of the constituent protons undergoes a change to a neutron:

$$ {}_Z^A\mathrm{X} \rightarrow {}_{Z-1}^{A}\mathrm{Y} + {}_{+1}^0e + {}_0^0\nu + \text{ kinetic energy} \tag{4.3.3} $$

  * A pretty exotic β process is the electron capture: the nucleus "captures" one electron from its electronic shell, and uses it to convert one
  of the constituent protons into a neutron:

$$ {}_Z^A\mathrm{X} + {}_{-1}^0e \rightarrow {}_{Z-1}^{A}\mathrm{Y} + {}^{0}\overline{\nu}_0 \tag{4.3.4} $$

  All energy released in the process is carried away by the antineutrino - there is no practical way to detect this elusive particle, so that the process can be detected only by indirect methods.

Gamma Decay

- The gamma decay. Let's talk first about the well-known Bohr Model of Hydrogen Atom - because it's good aid for explaining what the gamma-decay is.

In the Bohr Model an electron travels around a single proton in circular "orbits" corresponding to specific "allowed" energies $E_1, E_2, E_3, ...$, where $E_1 < E_2, E_2 < E_3$, and so on. When on the orbit with the lowest energy E_1, such a state may last forever, and it is referred to as the "ground state". On any orbit with higher energy the electron can stay only for a very short time, and then "jumps" to an orbit with lower energy - and such jump is associated with an emission of a photon of light of energy equal to to the difference between the energies of the two orbits. The frequency f of a photon is related to its energy by the Planck Constant: $E_{\rm ph} = hf$, so the energy and the frequency of the emitted photon is:

$$E_{\rm ph} = hf = E_{\rm n} - E_{\rm m} \quad \text{where} \quad n > m \geq 1 \tag{4.3.5}$$

The Bohr Model is now considered obsolete - it has been replaced by a much better theory, in which there are no longer orbits, but discrete states. Anyway, the rules are still the same: there is the ground state E_1 and a number of states with higher energies - called excited states. The electron "jumps" from an excited state to another excited state with lower energy, or at once to the ground state, and the Eq. 1 is still valid in the modern theory.

Let's go back to nuclear Y decay. There is one important difference between this decay and the decays discussed before: namely, thereis no such thing as an "isolated" γ decay in a parent nucleus. There is always a preceding a or β decay which leaves the "daughter" nucleus in an excited state. Nuclei do have excited states, in analogy to the excited states of electrons in atoms. Likewise, nuclei cannot remain in their excited states for a long time - so they "jump down" to the ground state, "jettisoning" the excitation energy - again, in analogy to electronic "jumps" in atoms - in the form of one or more photons of electromagnetic radiations. Yes, gamma radiation, X-rays, ultraviolet light, visible light, infrared light, microwaves, and radio waves areall forms of electromagnetic radiation. The only difference between photons of visible light and γ photons is in their energies: typically, a Y photon carries $10^4 - 10^6$ times more energy than a visible light photon.

Nuclear Fission

- An exotic mode of decay, occurring only in some very heavy nuclei is fission: the nucleus "splits" into two, sometimes releasing a few free neutrons. Fission may occur spontaneously, but it is very rare phenomenon-however, it may be stimulated by an absorption of a free neutron by the nucleus. We will discuss fission in greater detail later on, because essentially all electric power generated in nuclear power plants comes from fission reactions in uranium isotopes and some other heavy elements.

Beta Decya

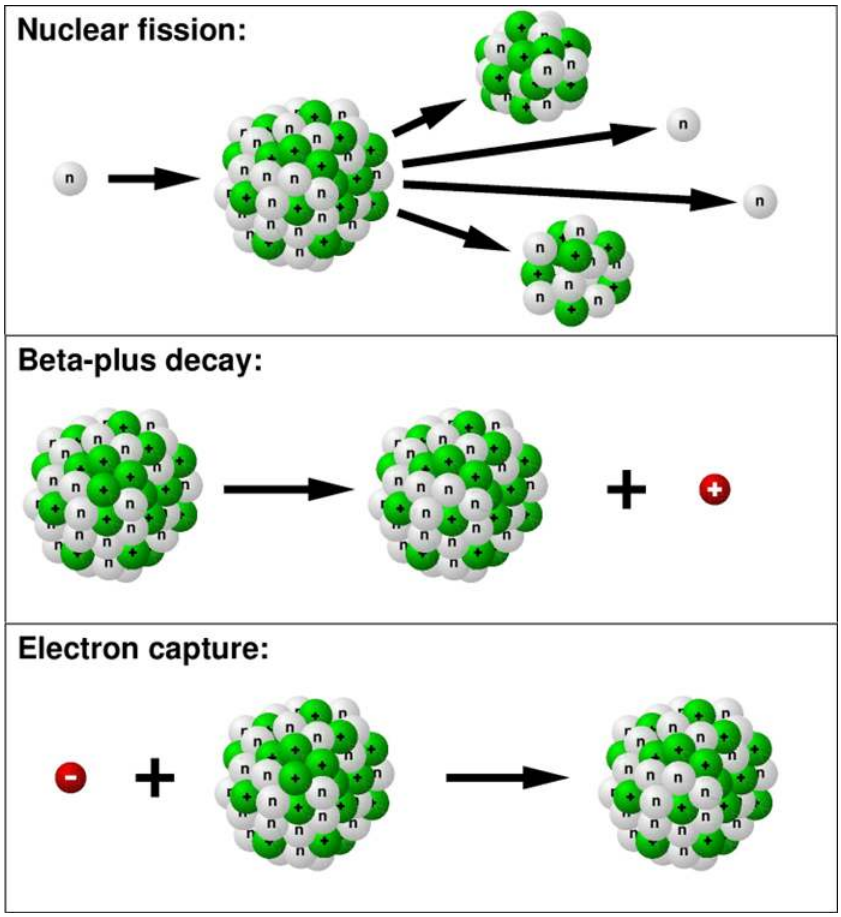


Figure 4.3.2: A fission of a heavy nucleus stimulated by absorption of a neutron. In the 235 isotope of Uranium such a process produces, in addition to the two “daughter” nuclei, two or more free neutrons which may trigger a “chain reaction” in other Uranium nuclei. A heavier Uranium isotope 238 also undergoes fission triggered by impinging neutrons, but it produces the “daughters” only, no free neutrons. The next panel show the less common beta decay types: the so-called “beta plus”, in which not en electron, but a **positron** is emitted, and one of the constituent protons in the nucleus changes into a neutron. And the third panel depicts the “electron capture” – a process in which the nucleus “captures” one of the electrons orbiting it. Neutrinos are emitted in both those processes, but are not shown in the plots.

---

1. Neutrinos are elementary particles of extremely small mass. There are three kinds ("flavors") of neutrinos, in each kind there is a "partner" and "anti-partner", making six different neutrinos altogether. They interact extremely weakly with matter, therefore they are very hard to detect. The neutrinos created in β decay processes are either the "electron neutrinos" (V_e) or "electron anti-neutrinos" $\left(\bar{V}_e\right)$. In this text, we skip the "e" subscript because neutrinos created in β decay processes are always of the "electron" flavor.

4.3: Radioactive Decay is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.3.1: Energy Units in Nuclear Physics

The relevant energies in nuclear physics, even the largest ones, are only tiny fractions of a Joule, the standard unit in the SI system. For instance, the products of the most energetic a, β and γ decays carry energies of the order of 10^{-13} J. The record-high energies, those observed in fission reactions, are still of the order of one hundredth of one billionth $\left(10^{-11}\right)$ of one Joule. Working with such small numbers is certainly not too convenient. Therefore, nuclear physicists created a more "user-friendly" energy unit, the electronVolt, with a symbol "eV". By definition, it is the amount of energy gained (or lost) by the charge of a single electron moving across an electric potential difference of one volt: $1\mathrm{eV} = 1.602^{*}10^{-19}$.

Practical units based on the electron-Volt are: the kilo-electron-Volt, 1 keV = one thousand eV; the mega-electron-Volt, $1\mathrm{MeV}$ = one million eV; as well as the milli-electronvolt, $1\mathrm{meV}$ = one thousandth of $1\mathrm{eV}$, and the micro-electron-Volt, $1\mu\mathrm{eV}$ = one millionth of $1\mathrm{eV}$.

4.3.1: Energy Units in Nuclear Physics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.3.2: The Decay Constants and Half-Lives

Radioactive decay is a stochastic, a.k.a. a random process. Suppose that we have just one radioactive nucleus: there is no theory that would allow to predict the moment at which the decay would occur. Or, suppose that we have two such nuclei. Is there a way to say which one will decay first? No, there is no such way.

Trustworthy prediction can be made, yes, but only concerning a large number of nuclei of a given isotope. Suppose that the number of nuclei in a specimen at an initial moment of time $t_0 = 0$ is N, and that we monitor the number of decays dN occurring in a short period of time, dt. Alternatively, dN can also be thought of as the "decrease in the population of radioactive nuclei in the specimen occurring over a time period dt " - and then it should be a negative number. Naturally, one can expect that the number of decays is proportional to length of the time period considered, and to the number of nuclei in the specimen - if there twice as many nuclei in the specimen, there should be twice as many of decays, right? - so we can write:

$$dN \propto -Ndt, \quad \text{or} \quad dN = -\lambda N dt \tag{4.3.2.1}$$

where λ is just the coefficient of proportionality.

When the decay process continues, the nuclei population gradually decreases. In order to find out how many nuclei will remain in the specimen after a longer elapse of time - say, from t_0 to t_1- one can treat the N in Eq. 4.2 as a function of time: $N = N(t)$. Then, after rewriting the equation as:

$$\frac{dN}{N} = -\lambda dt \tag{4.3.2.2}$$

one can perform an integration:

$$\int_{N(t_0)}^{N(t_1)} \frac{dN}{N} = \int_{t_0}^{t_1} -\lambda dt \tag{4.3.2.3}$$

Here one can use the known general integration formulas, in which x is a variable, and a, b and c are constant numbers:

$$\int_a^b \frac{dx}{x} = [\ln x]_a^b = \ln b - \ln a = \ln \frac{b}{a} \tag{4.3.2.4}$$

and

$$\int_a^b c dx = c \int_a^b dx = c[x]_a^b = c(b-a) \tag{4.3.2.5}$$

By applying them to the Eq. 4.3, we obtain:

$$\ln \frac{N(t_1)}{N(t_0)} = -\lambda (t_1 - t_0) \tag{4.3.2.6}$$

Recall the definition of the natural logarithm function: if $y = \ln x$, it means that $x = e^y$ (where $e = 2.718281\ldots$ is the base of the natural logarithm). One can write then: T

$$\frac{N(t_1)}{N(t_0)} = e^{-\lambda(t_1 - t_0)} \Rightarrow N(t_1) = N(t_0) \times e^{-\lambda(t_1 - t_0)} \tag{4.3.2.7}$$

This formula can be simplified: t_1, the moment "the stopwatch is started", can be taken as zero; the symbol for the initial population of radioactive nuclei, $N(t_0)$, can be replaced a simpler one, N_0, and the subscript at t_1 can be dropped, yielding a simple elegant final formula:

$$N(t) = N_0 e^{-\lambda t} \tag{4.3.2.8}$$

This is an important equation, expressing what is usually called "the exponential law of radioactive decay". The λ parameter in the equation is usually referred to as "the decay constant". It's an individual characteristic of the isotope described by the equation. There is no theory making it possible to calculate the λ value for any radioisotope with given A and Z numbers: it has to be determined by experimental measurements.

There is, however, a more "user friendly" and conceptually clearer parameter to be used in the Eq. 4.4 instead of λ. Let's ask the following question: if the initial population of radioactive nuclei in a given specimen is N_0, how much time will elapse until the population drops to one-half of its initial vale, i.e., to $N(t) = \frac{1}{2}N_0$? Let's call it "the half-life time", and introduce for it the $\tau_{1/2}$ symbol. The relation between λ and $\tau_{1/2}$ can be readily found from the Eq. 4.4 if we do the following:

$$\frac{1}{2}N_0 = N_0 e^{-\lambda\tau_{1/2}} \tag{4.3.2.9}$$

Here N_0 cancels out, and if we then take the logarithms of both sides, we obtain:

$$\ln\frac{1}{2} = \ln\left(e^{-\lambda\tau_{1/2}}\right) \Rightarrow -\ln 2 = -\lambda\tau_{1/2} \Rightarrow \lambda = \frac{\ln 2}{\tau_{1/2}} \tag{4.3.2.10}$$

By inserting this result to Eq. 4.4, we obtain a new form of this equation:

$$N(t) = N_0 e^{-\frac{\ln 2}{\tau_{1/2}}t} \text{ or } \quad N(t) = N_0 e^{-\frac{0.693}{\tau_{1/2}}t} \tag{4.3.2.11}$$

The physical meaning of the half-lifetime parameter $\tau_{1/2}$ is "intuitionally clear", and therefore it is commonly used as a radioisotope characteristic. It should be stressed that radioactive decay is a phenomenon completely insensitive to any extraneous influence: one may apply ultra-low or ultra-high temperature, ultra-high pressure, high magnetic or high electric field to an isotope sample - and measurements of $\tau_{1/2}$ in such drastic conditions will always yield the same result. And no matter whether the investigated sample has a solid, a liquid, or a gas form. It has been confirmed by "zillions" of experiments carried out on a wide variety of radioactive materials.

The half-lifetimes of known radioisotopes vary from milliseconds to billions of years. On Earth, the Naturally Occurring Radioactive Materials (NORM) are:

- Either "primordial" radioisotopes with their $\tau_{1/2}$-s longer, or at least not considerably shorter than the age of Earth (about 5 billion years): Uranium-238 ($\tau_{1/2} = 4.5 \times 10^9$ years), Thorium-232 $(\tau_{1/2} = 1.4^{*}10^{10}$ years), Potassium-40 $\tau_{1/2} = 1.25 \times 10^9$ years), Uranium-235 $(t_{1/2} = 7^{*}10^{8}$ years). For U-235 about seven half-life periods have elapsed since the Earth was formed; it means that there had been over 100 times more U-235 on Earth than there is now. At present, in natural uranium mix, there is only 0.7% of $\mathrm{U}-235$, and the rest is $\mathrm{U}-238$.
- Any radioisotope with a $\tau_{1/2}$ considerably shorter than 5 billion years, if present on Earth at the time of its formation, would long be gone. However, such elements may emerge as products of the decay of Uranium or Thorium which do not transform at once to a stable daughter isotope, but their decay initiate a chain of successive decay events, consisting of morethan 10 different transformations untilit finally reaches a stable nucleus. Some of the elements in such chains may have a relatively long lifetimes. For instance, Polonium $-210\,(\tau_{1/2} = 138$ days) and Radium-226 ($\tau_{1/2} = 1601$ years), the first radioactive elements other than Uranium and Thorium - discovered by Maria and PierreCurie, for which they were awarded two Nobel Prizes.
- Any NORM which relatively short $\tau_{1/2}$ period - if it is not an element of the decay chain - must have emerged as the result of the bombardment of Earth by high energy radiation from outer space. This is the case, for instance, with the famous ${}^{14}_{6}C$ isotope used for "radiocarbon dating" (196o Nobel Prize in Chemistry). The C-14 isotope is produced in the atmosphere, when a high-energy neutron penetrates a nitrogen nucleus and knocks a proton out of it:

$${}^{14}_{7}\mathrm{N} + {}^{1}_{0}n \rightarrow {}^{14}_{6}\mathrm{N} + {}^{1}_{1}p \tag{4.3.2.12}$$

There is one more comment worth making, about a misconception that is not so uncommon. Namely, after learning that there is such a thing as the half-life time, some students may have a tendency of thinking: OK, if after $\tau_{1/2}$ period one-half of the initial number of radioactive nuclei remains, then after another $\tau_{1/2}$ period all will be gone. This is not true, of course. After another $\tau_{1/2}$ period $1/4$ of the initial population will remain, and after yet another one, $1/8$ of it – and so on.

4.3.2: The Decay Constants and Half-Lives is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.4: Nuclear Cross Sections

Atomic nuclei are so small that they cannot be seen even through the most powerful existing microscopes. In order to investigate them, one has to use a number of other methods. One is to bombard them subatomic particles such as, e.g., protons, neutrons, or a-particles - and observe what happens.

In the available literature one can find the approximate formula for the radius R of an atomic nucleus of an atomic number A - assuming that it is a spherical object:

$$R = r_0 \ \mathrm{A}^{\frac{1}{3}} \tag{4.4.1}$$

where r_0 is usually given as $1.2\mathrm{fm}$ or $1.25\mathrm{fm}$ $\left(1\mathrm{fm} = 10^{-15}\ \mathrm{m} = 10^{-13}\ \mathrm{cm}\right)$.

Suppose that a particle, e.g., a neutron, is flying towards the nucleus. For a moment, let's assume that the particle can see like a one-eyed creature. What would it see? Obviously, a sphere, of radius R watched from a distance by one eye looks like a disk of radius R. The surface area of such disk is, of course, the same as the area of the cross section of the sphere.

The symbol used for the cross section area in nuclear physics is σ (Greek sigma). Consider a nucleus of an element from the middle of the Periodic Table - say, with A - and find the value of σ for it:

$$\sigma = \pi R^2 = \pi \times \left(1.25 \times 10^{-13}\ \mathrm{cm}\right)^2 \times 100^{\frac{2}{3}} = 1.058 \times 10^{-24}\ \mathrm{cm}^2 \tag{4.4.2}$$

For nuclei with other A values the results will be all of the order of $10^{-24}\ \mathrm{cm}^2$. Therefore nuclear physicists specializing in bombarding nuclei with different particles introduced, for convenience, a special unit of the name "barn": 1 barn $= 10^{-24}\ \mathrm{cm}^2$. Also, physicists have discarded the word "area", now σ is referred to simply as the "cross section".

One may wonder: What is so special in the cross section that a special unit has been created for it? And what is the cross section good for in nuclear physics? We show an example below, and later on, it will become clear that understanding the notion of cross section is essential for understanding fully how a nuclear reactor works.

An example In lead metal (Pb) the packing of atoms is such that there are $3.296^{*}1\mathrm{O}^{22}$ atoms $/\mathrm{cm}^3$. If a low-energy neutron passing through the metal hits a nucleus, it's bounced off its path (in professional terminology, one says "scattered"). The cross section of a single nucleus of the metal for such process is 11.2 barns.

Suppose that a beam of neutrons is incident at a right angle on a lead plate of 2 cmthickness. Find what fraction of neutrons pass the plate without changing their flight direction, and what fraction of neutrons are bounced off their original flight direction?

Answer. Draw (mentally) a $1\ \mathrm{cm} \times 1\ \mathrm{cm}$ square on the slab's surface. How many atoms there are behind the square's face? The volume of the material behind is $1\ \mathrm{cm}^2 2\ \mathrm{cm} = 2\ \mathrm{cm}^3$, so that behind the face there are 6.592×10^{22} atoms. What is the total cross section area of all nuclei of these atoms? The answer is: $6.592 \times 10^{22} \times 11.2 \times 10^{-24}\ \mathrm{cm}^2 = 0.751\ \mathrm{cm}^2$. So, an incoming neutron "sees" a $1\ \mathrm{cm}^2$ square on the slab surface, and behind it it sees an enormous number of tiny discs, the total surface area of which obscure obscure $0.751\ \mathrm{cm}^2$ of the square's surface area - it means, 75.1% of it. Hence, a single neutron has 75.1% chance of hitting one of the discs and being bounced out of its original flight path, while it has 24.9% chance of passing through the plate without any collision. In other words, if there are many neutrons in the impinging beam, 24.9% of them will emerge from the other side of the slab, as if there was no obstacle of any kind put across the beam's path.

So, as the example shows, working with cross sections maybe pretty straightforward - but now we have to reveal a secret concerning what has been said above: namely, the model in which the cross section is taken as $\sigma = \pi R^2$, would be correct, yes, but if our "nuclei" were billiard balls, and our projectile-particles were, for instance, balls from a BB gun. In other words, in our "macro-world". But this model seldom describes correctly collisions in the micro-world, which is the realm of quantum phenomena.

The cross sections in the micro-world may be very different even for nuclei with the same A. So, the cross section σ is not the same as the cross section area of a simple sphere. For instance, the Xenon- 135 nucleus, a real neutron devourer - an accomplice of the Chernobyl disaster, and a source of strong headaches for nuclear reactor operators - has the geometric cross section of about 1.5 barn, and the cross-section for neutron capture (absorption) as incredibly high as 2,600,000 barns! It means that for an incoming

neutron the Xe-125 nucleus "looks like" a disc with a radius one thousand times as large as the radius of the nucleus' geometrical cross section!

Another strange property of nuclear cross sections - especially, for neutron capture - is their dependence on the incident particle energy E_i. For some nuclei, the σ vs. E_i plots exhibit sharp peaks, or even a large number of such peaks. Wewill return to discussing such characteristics shortly, because they are essential for understanding how nuclear reactors work.

One more example (we will need the result when discussing the design of a nuclear reactor): The total cross section of Uranium-238 for neutrons of $1.5\mathrm{MeV}$ energy ("total" means: for any kind of interaction) is $\sigma = 1.1$ barn. The atomic number density (i.e., number of atoms per 1 cubic cm) for Uranium metal is $\rho_{\mathrm{A}} = 4.8 \times 10^{22}$ atoms $/\mathrm{cm}^3$. What is the probability thata $1.5\mathrm{MeV}$ neutron passes "untouched" trough a $1\ \mathrm{cm}$ thick plate of Uranium metal?

Answer: Consider a $1\ \mathrm{cm} \times 1\ \mathrm{cm}$ square drawn on the surface of the plate. The plate is $1\ \mathrm{cm}$ thick, so that the volume of the material behind the square is $1\ \mathrm{cm}^3$, and there are 4.8×10^{22} atoms behind it. The total surface area of that many 1.1 barn discs "seen" by a neutron is $4.8 \times 10^{22} \times 0.3 \times 10^{-24}\ \mathrm{cm}^2 = 0.0528\ \mathrm{cm}^2$. Hence, the probability that a neutron interacts with a nucleus is 5.28% - and it is 94.72% that it will "fly untouched all the way through".

4.4: Nuclear Cross Sections is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.5: Nuclear Fission

After the discovery of the neutron by James Chadwick in 1932, and after learning, shortly afterwards, how to prepare strong neutron sources, physicists rushed to carry out experiments in which they "bombarded" with neutrons any nuclei that were available. Literally, all elements from the periodic table were investigated in such way. Typically, the bombardment led to the so-called (n, y) reaction:

$$ {}^{A}_{Z}\mathrm{X} + {}^{1}_{0}n \rightarrow {}^{A+1}_{Z}\mathrm{Y} + {}^{0}_{0}\gamma \tag{4.5.1} $$

In most cases, the daughter nucleus X was radioactive - so, a large number of new previously unknown radioisotopes were discovered.

In December 1938 such an experiment was performed on uranium by two German radiochemists, Otto Hahn and Fritz Strassmann. In the irradiated material they found Barium, an element with $Z = 56$. Since for Uranium $Z = 92$, Barium could be created by a typical (n, γ) reaction. The riddle was solved shortly afterwards by Lise Meitner, a long-time Hahn collaborator, who was Jewish and therefore some time earlier she had had to leave then Nazi Germany and to seek safety in Sweden. Meitner deduced that a reaction product that much lighter than the parent nucleus could only mean that a completely new process had occurred - namely, that the uranium nucleus had been split by the incident neutron. Soon afterwards, Meitner's explanation was confirmed by a large number of experiments performed independently by several labs. A new term has been coined for the newly discovered process namely, "nuclear fission".

Natural uranium consists of 99.3% of the U-238 isotope, and only 0.7\% of the U-235 isotope. But the fission process in the Hahn and Strasmann occurred in U-235 nuclei only. It turned out that the production of Barium was only one of the many other possible decay schemes. The two daughter nuclei produced by the fission are called "fragments": their Z numbers must always add up to 92 , but there are altogether about 150 possible such combinations. Two more properties of the fission reaction are of great importance:

- The total energy released by the fission process is unusually high about $200\mathrm{MeV}$ per one event, 100 times more than in the most energetic known a and β decays; and
- Each fission reaction, in addition to the two fragment, releases a few (2.4 per average) free "daughter neutrons".

Especially, the discovery of the latter property created a great excitement among physicist, because it meant that a "chain reaction" might occur. Namely, one one neutron incident on a block of U-235 might fission one nucleus and release 2-3 daughter neutrons; which in turn might fission 2-3 more U-235 nuclei, producing some 5-6 "second-generation daughter neutrons", these might trigger up to six more fission reaction, and so on - an avalanche might develop and last up to the moment when there would be no more U-235 nuclei left in the block. Such process might release an enormous amount of energy: in one pound of pure U-235, if all nuclei underwent fission, would release nearly 4×10^{14} J of energy. It's about the same amount of energy that would be released by burning 10^7 kilograms, or three million gallons of gasoline.

As follows from the above, pure Uranium-235 isotope would be a very dangerous material: a single neutron - and there are always such particles around us, created by cosmic radiation-might trigger a powerful explosion. In fact, the very first atomic bomb ever used in warfare was nothing else than a big block of nearly-pure U-235 isotope. Good news, however, is that natural uranium contains only 0.7% of the dangerous isotope and extracting the latter from the former is an incredible difficult process, requiring a huge industrial-scale facility to be carried out. And a small amount of U-235 say, a pound or so - is not dangerous at all. The thing is that the "daughter neutrons", necessary to maintain the chain reaction, are more likely to escape to the outside than to induce another fission event. In order to "keep enough daughter neutrons inside", one has to diminish the "escape route" - in other words, to make attain a small-enough ratio of the surface area to the block volume. ForU-235 a chain reaction may develop in a sphere of radius nearly 9 cm, and a mass of about 52 kg. It's the infamous "critical mass" of Uranium 235

Interestingly, the other Uranium isotope in the natural composition may also undergo a fission reaction [2] - however, the essential difference between the U-235 and U-238 fission is that the latter reaction does not yield any daughter neutrons. Hence, a chain reaction can't occur in U-238. So,there are two categories of nuclei that exhibit the fission reaction. The ones that release daughter neutrons and are thus capable to maintain chain reaction are called fissile (a term derived by merging "fissionable" and "fertile"). Of naturally occurring heavy nuclei only U-235 is fissile, and only U-238 fissionable. Yet, there are "human-made" fissile nuclei,

Plutonium-239 (obtained by irradiating U-238 with neutrons) and U-233 (by irradiating Thorium-232). There are more man-made fissile materials, but only Pu-239 and U-233 are of practical importance from the viewpoint of electric power generation.

---

1. In U-238 fission can be triggered only by neutrons of relatively high energy, $> 1\mathrm{MeV}$ -in contrast, quite surprisingly, for U-235 it's the low-energy neutrons which are the most efficient in triggering the fission reaction.

4.5: Nuclear Fission is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.5.1: Fission and Fusion Energy

As noted, in atomic nuclei the constituent nucleons, protons and neutrons, are held together by an attractive interaction called the "strong force". If we want to remove a nucleon from the nucleus, we need to do some work against the strong force. This work is equal to the energy with which the nucleon was "bound" to the nucleus, so we call it the binding energy. If we measure the mass of the nucleus before removing the nucleon, and after removing it, we finf that the mass of the original nucleus is lower than the sum of masses of the nucleus after the operation plus the mass of the removed nucleon. How comes? Well, from the famous Einstein's relation between mass and energy, $E = mc^2$. Note that in the process of removing the nucleus we did work W, so we added energy $E = W$ to the nucleus + nucleon system. If we convert the work done to mass, $\Delta m = W/c^2$, then we obtain the difference:

$$\Delta m = [(M_{\text{after removal}} + M_{\text{nucleon}}) - M_{\text{original}}] = \frac{W}{c^2} \tag{4.5.1.1}$$

If we had a magic tool enabling us to remove nucleons one by one, and measuring the work done at each step, and then added up the work done at each step, we would get the total binding energy E_{tbe} for the original nucleus. As we say,

$$E_{\text{tbe}} = \sum_{i=1}^{N-1} W_i \tag{4.5.1.2}$$

where Wi is the work we did in each i^{h} step of removing nuclei (Note that if there are N nucleons, we need only N - 1 steps to completely separate them from one another).

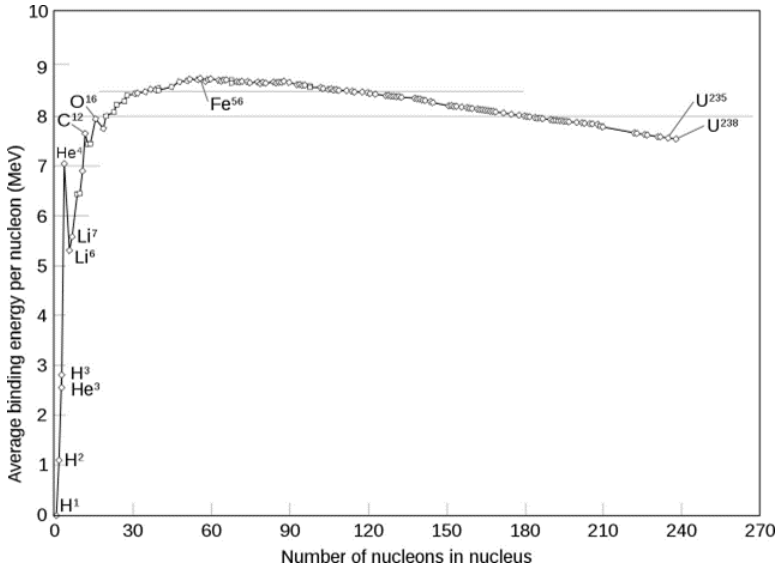


Figure 4.5.1.1: : Binding energy per nucleon versus the *A* number (from Wikimedia Commons).

So, by the same token as before, we may conclude:

$$\Delta M = \sum m_i - M_{\text{original}} = \frac{E_{\text{tbe}}}{c^2} \tag{4.5.1.3}$$

where the sum is over the masses of all constituent nuclei. The ΔM is called "the mass defect" of the given nucleus. The mass defect of all stable nuclei and important radioisotopes are listed in many publications and Web documents - but seldom in mass units, usually as mass defect per nucleon $(\Delta M/N)$ and converted to energy units by multiplying by c^2 – then its called "Binding Energy" and given in the units of MeV.

A graph displaying the binding energy per nucleon is shown in Fig. 4.5.1.1.

4.5.1: Fission and Fusion Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.5.2: How to calculate the mass defect?

The atomic masses of all stable and long-lived elements have been determined with a very high precision, and they can be readily found on the Web -usually, given in a.m.u. (often shortened to single "u"), which means "atomic mass units". The most useful units for the discussion in this section are not a.m.u., but the e.m.u. energy equivalent expressed in MeV. A highly precise value of the a.m.u. $\mathrm{MeV/c^2}$ conversion factor can be found in the NIST Web page: 1 a.m.u. $= 9 \cdot 314940954 * 0^5 \mathrm{MeV/c^2}$.

Let's, for example, consider the U-235. It's atomic mass (taken from the Web page of the Nuclear Data Center, Japan Atomic Energy Agency) is 235.043931368 a.m.u., and after performing the multiplication one gets the value of $2.189420329 * 10^5 \mathrm{MeV/c^2}$. There is one "trick" that has to be done here: namely, the tables almost always list the values of atomic masses, whereas we need the mass of the **nucleus** - so we have to subtract the masses of all electrons in the atom. The energy equivalent of an electron is $0.511\ \mathrm{MeV/c^2}$ and there are 92 of then in an Uranium atom, so after subtracting the total mass of the electrons, we get the value of $2.188950209^* 10^5 \mathrm{MeV}/c^2$ for the mass of the nucleus.

Now, we need the energy-equivalents for the masses of a single proton and a singleneutron - again, from the NIST Web page we find: $m_p = 938.2720813\ \mathrm{MeV/c^2}$ and $m_n = 939.5654133 \mathrm{MeV}/c^2$. In the U-235 nucleus there are 92 protons and 143 neutrons, and the total mass of such combination is:

$$92 \times m_p + 143 \times m_n = 2.206788856 \times 10^5 \mathrm{MeV}/c^2 \tag{4.5.2.1}$$

Subtraction the m ass of the nucleus from the above result yields the mass defect value:

$$\Delta M = (2.206788856 - 2.188950209) \times 10^5 \mathrm{MeV}/c^2 = 1783.864682 \mathrm{MeV}/c^2 \tag{4.5.2.2}$$

This method is straightforward, but tedious - but trustworthy values can be obtained more conveniently from the "Nuclear Binding Energy Finder" provided on the Web by Wolfram Alpha. For U-235 this service returns the value of $7.590907 \mathrm{MeV}/c^2$ for a single nucleon. Multiplication by 235 yields $1783.863145 \mathrm{MeV}/c^2$ for the entire nucleus, in excellent agreement with the result obtained by the more laborious method.

A "**technical" remark**: In the above calculations, all decimal numbers of the values provided by the Web sources were taken - which was an unnecessary "overkill", the calculations in the next subsection demonstrate that taking just six decimals is a sufficient precision. But if the final result is to be a difference between two similar very large numbers, it never hurts to be cautious.

4.5.2: How to calculate the mass defect? is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.5.3: Energy Released by the Fission of a Single Nucleus

Let's use as an example the reaction that led to the discovery of nuclear fission. The "key" to the discovery was the presence of element Barium with $Z = 56$. In a fission reaction the Z numbers of the daughter nuclei must add up to Z of the parent nucleus, so that the other element must have been $92 - 56 = 36$, which is the atomic number of the element Krypton. As noted earlier, the fission of U-235 nuclei may produce as many as 150 different combinations of daughter nuclei - so for a single it is not possible to predict how many neutrons there will be in each fragment. Here is one likely "scenario":

$$n + {}^{235}_{92}\mathrm{U} \rightarrow {}^{144}_{56}\mathrm{Ba} + {}^{89}_{36}\mathrm{Kr} + 3n \tag{4.5.3.1}$$

Note that the number of protons on both sides are the same: 92 in U, and 56 $+36 = 92$ in Ba and Kr. The same is true for neutrons: left side, 1 incident plus (235-92) in U is 144 ; right side: (144-56) in Ba plus (89-36) in Kr plus 3 released also $= 144$

The calculation of the energy release in the process is pretty straightforward. We need to compare the masses of all objects at both sides of the equation. For that, we need to know the mass defects of all nuclei - we have already got it for U-235, and using the aforementioned Web tool, we readily find: $1190.23\mathrm{MeV/c^2}$ for Ba-144 and $766.91\mathrm{MeV}$ for $\mathrm{Kr} - 89$. Keeping in mind that the mass of a nucleus is equal to the mass of all protons plus the mass of all neutrons minus the mass defect.

In Table 1, all items needed for calculating the total masses of the elements represented on the left side, and the right side of the Eq. 4.7 are collected. Our objective is to find the difference between the total masses "before" and "after" the fission event.

Table 1. All relevant components arranged in a way making easy "balancing the mass book" in the reaction described by Eq. 4.7.

Left side: $1m_n$
Left side: $143m_n + 92m_p - 1783.36\mathrm{MeV}/c^2$
Right side: $88m_n + 56m_p - 1190.23\mathrm{MeV}/c^2$
Right side: $53m_n + 36m_p - 766.91\mathrm{MeV}/c^2$
Right side: $3m_n$

The way things are arranged in Table 1 facilitates such calculations: it makes it clear immediately that when the "right side" (final) masses are subtracted from the "left side" (initial) ones, all proton and neutron masses cancel out, and the result is:

$$\begin{gathered} M_{\text{total initial}} - M_{\text{total final}} = \\ = [(-1783.36) - (-1190.23) - (-766.91)]\tfrac{\mathrm{MeV}}{c^2} = +173.78\tfrac{\mathrm{MeV}}{c^2} \end{gathered} \tag{4.5.3.2}$$

The "plus" sign before the final result was added intentionally, in order to emphasize that a significant mass component is missing in the right side of the equation. How comes that the "mass book doesn't balance"? Here the Einstein's famous formula $E = mc^2$ steps in handy, the missing mass has been converted to $173.78\mathrm{MeV}$ of "pure energy": its main component is the kinetic energy of the two daughter nuclei flying apart (like fragments of the bombshell in an explosion of a bomb), several MeV is the kinetic energy of the fast neutrons released, and some energy may be released in the form of one or more Y photons.

Altogether, the total energy released in a single fission event is even higher. Note that the N/Z ratio for U-235, 1.554, is considerable higher than that for stable isotopes from the Z range close to 46 , i.e., one-half of the Uranium's Z. Let's check: the element 46 is Palladium, a mixture of stable isotopes with an average A of 106.42, meaning an average $N = (A - Z)$ of 60.42, and average N/Z ratio of 1.313, much less than that of Uranium. Consequently, the fission product nuclei havea "strong surplus" of neutrons. Let's take a look at the nuclei in the Eq. 4.7: for the heavier fragment $N/Z = 1.57$, while for natural Barium it is 1.452; and for the lighter fragment it is $N/Z = 1.472$, while for natural Krypton gas it is 1.328. It means that both fragments have a huge "surplus" of neutrons, and thus they should be strongly radioactive with very short lifetimes. In fact, sometimes the fission products in real nuclear reactors have lifetimes as short as a small fraction of one second. In some of them the N/Z ratio may be so high that they decay through the emission of a neutron - a decay process is not seen in any naturally occurring radioactive material. By the way, such neutrons are called "prompt neutrons" and they play an important role in controlling the chain reaction in nuclear reactors.

The decay of the fission fragments yields additional energy. A detailed list of all processes participating in the overall energy yield in nuclear fission can be found, e.g., in this Hyperphysics Web page. In addition to the immediate energy yield coming from the

loss of mass of U-235, the decay of the fission products may contributs extra $30 - 40\mathrm{MeV}$, so that a reasonable estimate of the average energy release per single fission reaction can be taken as 210-215 MeV[1].

---

1. In fact, the net energy released depends of the timescale because not all decay products of fission reaction fragments are short-lived. For instance, the infamous Caesium-137 that contaminated huge land areas after the Chernobyl disaster has the half-life of over 30 years

4.5.3: Energy Released by the Fission of a Single Nucleus is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.5.4: Nuclear Fusion

In Fig. 4.3 one can see that the binding energy for ^{4}He is unusually high. If we look at the binding energies of several nuclei from this Z region ($^2_1$H is also known as "Deuterium", and$_1$ $^3_1$H as "Tritium"):

Nucleus	E_{\text {Bind
$^1_1$H	0
$^2_1$H	2.2245
$^3_1$H	8.4820
$^4_2$He	28.2970
$^{11}_5$ B	76.2060

In contrast to fission reactions, in which energy is released due to the split of a heavy nucleus into two lighter nuclei, energy can also be released in fusion reactions, in which two light nuclei merge into a single heavier one.

Consider the reaction:

$$^2_1\text{H} + ^2_1\text{H} \rightarrow ^4_2\text{He} + \text{ Energy} \tag{4.5.4.1}$$

Using the data from the table provided above and the same method as is used in the preceding section for calculating the energy released in U-235 fission, it can be readily found that the energy released in a single reaction is $28.297\text{MeV} - 2\chi 2.2245\text{MeV} = 23.848\text{MeV}$. It's about 10 times less from a single fission event, but the mass of the "fuel" is about 60 times less, so by the fusion of one pound of Deuterium one should obtain about 6 times more energy than from the fission of all nuclei in one pound of U-235. And Deuterium is available commercially for $\$138$ per $20\ \text{g}$ (in $100\ \text{g}$ of "heavy water" D_2O), while $\text{U}-235$ is offered for $\$1500$ per $o.1\text{g}$, or $\$300,000$ per $20\ \text{g}$, over two thousand times more expensive than Deuterium.

So, a fusion reaction might be a much, much cheaper way for generating power than the currently used methods based on Uranium fission. The fuel for generating $1\ \text{kW}$ power might be even much, much cheaper than the oil or natural gas needed for generating $1\ \text{kW}$ in a conventional power plant. The picture would be really glamorous, if not the fact that currently there exists no technology making it possible to conduct such reaction in a controlled manner. The only time humans succeeded to initiate a Deuterium-Deuterium (D-D)fusion reaction was during the test of the very first American thermonuclear charge "Ivy-Mike", detonated on Nov. 1, 1952. It was definitely not a deliverable bomb, but a device of a total weight of nearly 75 metric tons.

A much easier to ignite than the D-D fusion is the Deutrium-Tritium (D-T)reaction:

$$^2_1\text{H} + ^3_1\text{H} \rightarrow ^4_2\text{He} + ^1_0n + 17.59\text{MeV} \tag{4.5.4.2}$$

This reaction is used in all currently existing deliverable "hydrogen bombs". Many of them have been successfully tested - the largest one was the Soviet test of a monster, which later became known in the West as the "Tsar Bomb", which was detonated in October 1961, with a yield about 5000 times more powerful than that of the bomb which on August 6, 1945 had wiped out the city of Hiroshima.

Until now, however, all attempts to realize a controlled D-T fusion process lasting for a prolonged time have failed. But countless experiments carried out so far (since the 1950-ties) have offered much hint of how to design a giant machine promising a good chance for achieving a "breakthrough" the construction of such machine, a huge effort code-named ITER jointly sponsored by 35 countries, is currently under way in France. The latest schedule revision foresees the first plasma runs in 2025 . The schedule was revised several times before the last revision - but let's hope there will be no more revisions...

It is perhaps worth mentioning about other possible fusion schemes that are now being researched. A big inconvenience of the D-T "fuel cycle" is the fact that this reaction fast neutrons are created. Prolonged exposure to them may make the parts of "fusion

reactors" highly radioactive. The plans for ITER are that it will only be a research facility. But dealing with intense high-energy neutron radiation may become a major challenge for the designers of future power generators utilizing the D-T reaction. In view of that, much interest is currently focused on the so-called "aneutronic fusion" - i.e., reaction, in which no neutrons are produced. One such reaction is the "Hydrogen-Boron" fusion:

$$ {}_{1}^{1}p + {}_{5}^{11}\mathrm{B} \rightarrow 3\,{}_{2}^{4}\mathrm{He} + 8.7\mathrm{MeV} \tag{4.5.4.3} $$

Another advantage of this "cycle" is that both fuel components, hydrogen and boron, are inexpensive and easily available. A new interesting idea of using powerful laser beams for achieving a sustained fusion process in the H-B fuel mix has recently emerged. It is certainly worth paying attention to any new developments in this research project.

4.5.4: Nuclear Fusion is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

4.6: Controlling the Fission Chain Reaction- Nuclear Reactors

As has been discussed before, a chain reaction may occur in a block of fissile material, big enough to assure that not too many fission neutrons manage to reach the surface from the inside and escape. This is the so-called “critical mass”.

But if a chain reaction occurred in a block of pure or nearly-pure fissile material, the consequences might be catastrophic (well, sometimes this is the goal it how nuclear weapon works). In this text, however, we want to talk only about **peaceful** applications of nuclear energy. A chain reaction may release much energy that can be converted to electricity – but only if there are means for controlling the process. There is no way of controlling the process in a bulk piece of pure fissile isotopes. The “avalanche” develops very fast, the chain reaction simply causes the material to explode. For making the chain reaction “controllable”, it must be slowed down. But how to achieve such a goal?

Let’s begin with a discussion of what means are at our disposal. The first and the most important question is: what kind of“nuclear fuel” can we use? Now, there are several different types available. But let’s talk about the situation at the very beginning of the Atomic Age: then, the pioneers were left by Mother Nature with just one choice: natural uranium.has already left us with little choice: natural uranium. Its is a mixture of two isotopes: it contains 99.3% Uranium-238, which is not fissile, and only 0.7% of Uranium-235, which is. Only one atom out of 140 not-fissile ones.

But it’s not the only problem: U-238 is not merely a “neutral” component, it is *fissionable*, meaning that it gladly absorbs fission neutrons from U-235, undergoes a fission, contributes 200 MeV of energy but releases no neutrons. No neutrons needed for maintaining the chain reaction! In other words, U-238 is a “kidnapper” of fission neutrons. There is no way that a chain reaction, controlled or uncontrolled, can ever occur in a piece of natural uranium

Nevertheless, Enrico Fermi, Leo Szilard, and a few other pioneers man- aged to successfully start a self-sustained and controlled chain reaction on Dec. 2, 1942, in a device they called “the pile”, which used nothing else but natural uranium as a fuel. After that success, a number of other reac- tors were built, all using natural uranium which remained the only available nuclear fuel for several more years. Then, many new types of fuel emerged (e.g., “enriched” Uranium with increased U-235 content, or uranium mixed with the “man-made” fissile element Plutonium). Most currently existing reactors in civilian nuclear power plants worldwide use those new types of fuel. But for explaining how a nuclear reactor works it still make sense to begin with an analysis of the design and the physical processes occurring in a simple reactor prototypes using natural uranium fuel[4]. And the basic principles of operation of modern reactors remain essentially the same as in those old prototypes

4.6.1: Essential Cross Sections

4.6.2: No magic tricks are necessary!

4.6.3: Controlling the Chain Reaction and Cooling the Assembly

4.6.4: Power Reactors

4.6.5: Generation IV Reactors and Thorium Fuel

4In fact, the progress in developing new more sophisticated fuel types has not made natural uranium fuel completely obsolete: there is a CANDU family of Canadian-designed power reactors running on natural uranium – 29 operational in several different countries, and a few more under construction

4.6: Controlling the Fission Chain Reaction- Nuclear Reactors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.6.1: Essential Cross Sections

The key for understanding the physical foundations of the operation of a nuclear reactor is the energy dependence of relevant cross sections of U-235 and $U-238$. The plots are shown in Fig. 4.6.1.1.

In the process of U-235 fission, high energy neutrons are created - with average energy of about $1.6\mathrm{MeV}$. For such energies, the U-238 cross section for fission is somewhat lower (about 50\%) than that of U-235 (see the red and the green curves on the right-side part of Fig. 4.6.1.1). Yet, there are 140 atoms of U-238 per one atom of U-235, so the neutron has a very small chance, one-out-of-seventy to hit another U-235 nucleus and thus to support the chain reaction - and if it hits a U-238 nucleus, it may trigger a fission, but - as was stated before - the U-238 fission does not release any neutrons, so that it does not support the chain reaction.

The situation drastically changes in the region of very low energies, of $0.1\mathrm{eV}$ and less. Here, the fission cross section for U-238 is of the order of micro-barns only. The dominant is the cross section for "radiative capture of neutron", i.e., the U-238 nucleus emits a gamma photon and changes into U-239 . But now the cross section for U-235 fission is about 3 timeshigher, so that the 0.7% of U-235 nuclei have a five-to-one chance over the 99.3% of U-238 to capture the neutron. So, the chain reaction may be saved - if the fission neutron could be removed from the uranium fuel, forced to give away its high energy - or lowering it about ten million times (from $1\mathrm{MeV}$ down to $0.1\mathrm{eV}$) - and only then returned to the fuel.

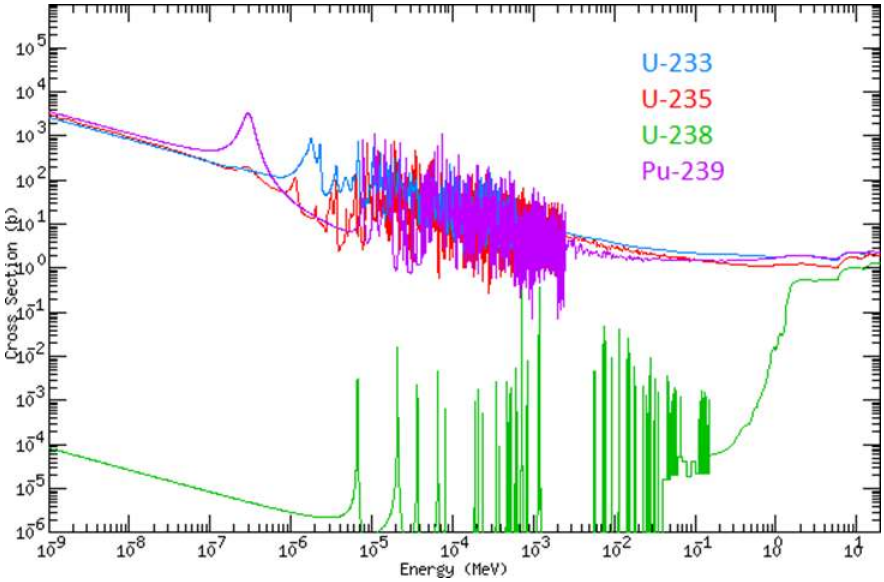


Figure 4.6.1.1: Upper curves: The cross sections for fission vs. neutron energy for U-235 and two other fissile isotopes (U-233 and plutonium Pu-239) as a function of incident neutron energy – note the plot is in a “log-log” scale. Lower curve: the same for the non-fissile U-238. Note that the vertical scale in the lower plot is more compressed than in the upper plot.Note that in certain energy regions the cross sections “oscillate” in a very irregular way. But below the neutron energy of about 1 eV (= 10^{-6} MeV the cross section of the three fissile nuclei are more than a million times higher than that for U-238.

[1] $U-239$ after half-life of only 2.35 minutes changes via a β decay to Neptune, and after another β decay with half-life of 2.4 days changes to Plutonium-239, with half-life over 20,000 years - not a nice element, as it is the essential component in the A-bombs and H bombs.

4.6.1: Essential Cross Sections is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.6.2: No magic tricks are necessary!

The “recipe” given above may sound a bit like one of those prescriptions from Harry Potter books – but, in fact, no supernatural powers are needed. As it follows from the result we have obtained in the second Example in Section 1.4, a fission neutron can travel a relatively long way in pure U-238. Natural uranium has only 0.7% of the U-235 isotope, which only marginally changes the cross section we used in the calculations. Accordingly, there is no need to use any tricks for removing the fission neutron from the fuel element: if the fuel is in the form of long rods, half inch or one inch in diameter, most of the fission neutrons will fly out “all by themselves”.

So, the first step of the “recipe” is completed. Now comes the step number two: take away the energy from the fission neutrons. How do we take away kinetic energy from neutrons?

The answer is simple: the only way of doing it without destroying the neutron is by colliding it with atomic nuclei. Some atomic nuclei would gladly “swallow” an impinging neutron, but some have no appetite for them whatsoever and the neutron simply “bounces off” the nucleus. Here the behavior of the two objects shows much analogy to the collision of two billiard balls of different masses.

It may be helpful now to recall the introductory physics problem: a ball of mass *m* moving with speed *v* is approaching, on a head-on collision course, another ball of mass *M* which is at rest (*V* = 0). Assuming that the collision is perfectly elastic (i.e., both energy and momentum is conserved in the process), find the kinetic energy of each ball after the collision. We will not solve the problem here, but only quote the results. The velocities after the collision are:

$$v' = v\frac{m-M}{m+M} \quad \text{and} \quad V' = v\frac{2m}{m+M} \tag{4.6.2.1}$$

and the kinetic energies:

$$K'_m = \frac{mv^2}{2}\left(\frac{m-M}{m+M}\right)^2 \text{ and } \quad K'_M = \frac{Mv^2}{2}\left(\frac{2m}{m+M}\right)^2 \tag{4.6.2.2}$$

The thing of real interest for us is the ratio of the energy K^t_m of the m ball after the collision to that before the collision (\left(K_{m}=\frac{1}{2} m v^{2}\right)\) :

$$\frac{K'_m}{K_m} = \left(\frac{M-m}{M+m}\right)^2 \tag{4.6.2.3}$$

There is one reason this simple result from introductory classical mechanics is shown here: namely, from his long experience in teaching the basic reactor physic, the Author knows that the problem of "slowing down" neutrons through collisions can be "counterintuitive". Many times a student, when given the question: What do you thing? What is the more efficient way of slowing down neutrons - in other words, of lowering their kinetic energy - by colliding them with light nuclei, or with heavy nuclei? responded with confidence: With heavy nuclei, I believe! - and was a bit surprised when told that "with light nuclei" was the right answer. Evidently, it's an example of a simple problem which is misleading if one applies physical thinking based exclusively on the so-called "conceptual understanding".

But it is easy to fix such incorrect thinking. Suppose that you throw a golf ball against a bowling ball. The golf ball bounces back, the bowling ball hardly even budges. Now, take a toy called the "Newton's Cradle" in which there are several steel balls suspended from a frame, so that one can study collisions. Use only two balls and see what happens, if one ball hits the other which is at rest: the ball that moved stops, the other takes over its entire kinetic energy.

Indeed, this is exactly what the Eq. 4.8 states. Let's check: the golf ball mass is 1.62 ounces, and the mass of the bowling ball is 16 pounds, i.e., the latter is 158 times heavier than the former. So, for such a mass proportion we get from the equation we get that the energy of the small ball after the collision is still $(157/159)^2 = 97.5\%$ of that before the collision, so it loses only 2.5% of its initial energy.

Now, consider a fission neutron. We want it to get slowed from energy 1.6 MeV down to $0.1\mathrm{eV}$ - i.e., $1.6 \times 1\mathrm{O}^7$ times, through a chain of multiple collisions. If in each collision the neutron loses 2.5% of its energy, then it needs to collide nearly 700 times [1]

In contrast, if the neutron collides with a deuteron, ${}^{2}_{1}\mathrm{H}$, the mass ratio M/m is close to 2 , then after a single collision only $(1/3)^2 = 1/9$ remains, and only 9 successive collisions are needed to reduce the neutron energy from $1.6\mathrm{MeV}$ to $0.1\mathrm{eV}$. And for Carbon, ${}^{12}_{6}\mathrm{C}$, it is, respectively $(11/13)^2 = 0.715$, and 50 collisions.

In fact, it is not exactly so simple - the Eq. \(4.6.2.3\) was derived for the assumption that the collision is central, whereas in the real world not all collisions are central, not always the impinging particle moves exactly along the line connecting the centers of both particles. In non-central collisions neutrons loseless energy than central ones - therefore, the actual number of collisions needed to slow fission neutrons down to $0.1\mathrm{eV}$ may be higher by a factor 2-3 than the values obtained above. Anyway, it is true that Deuterium and Carbon are very efficient as "neutron energy reducers". In professional language, the process of slowing neutrons down is called moderation, and the materials used for this purpose are called *moderators*. Pure Deuterium would be difficult to use because it's a gas; but an excellent moderator is Deuterium oxide, D_2O, which is a liquid known as "heavy water". The element of Carbon may occur in many different forms – e.g., black sooth, graphene, diamond, and graphite are all pure forms of carbon. But only graphite is used as a neutron moderator.

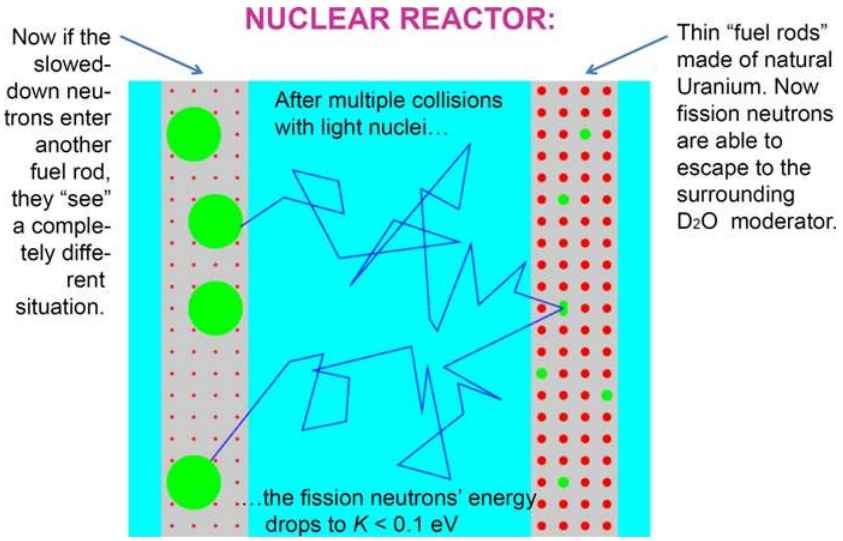


Figure 4.6.2.1:A fission neutron with energy of 1.6 MeV has little chance of triggering another fission event in the same fuel rod because the U-235 nuclei are a small minority. But after escaping the rod, a neutron experiences multiple collisions with the moderator nuclei, loosing its kinetic at each collision. With the energy of 0.1 eV or lower, when re-entering a fuel rod, neutrons "see" the U-235 nuclei as objects of several hundred times larger sizes than the U-238 nuclei – most of them will therefore trigger new fission events.

As follows from the above, the uranium fuel rod should be immersed in a tank of heavy water D_2O, or put into a block of graphite. Then, the fission neutrons escaping from the rod get slowed down to energies below 0.1 eV.

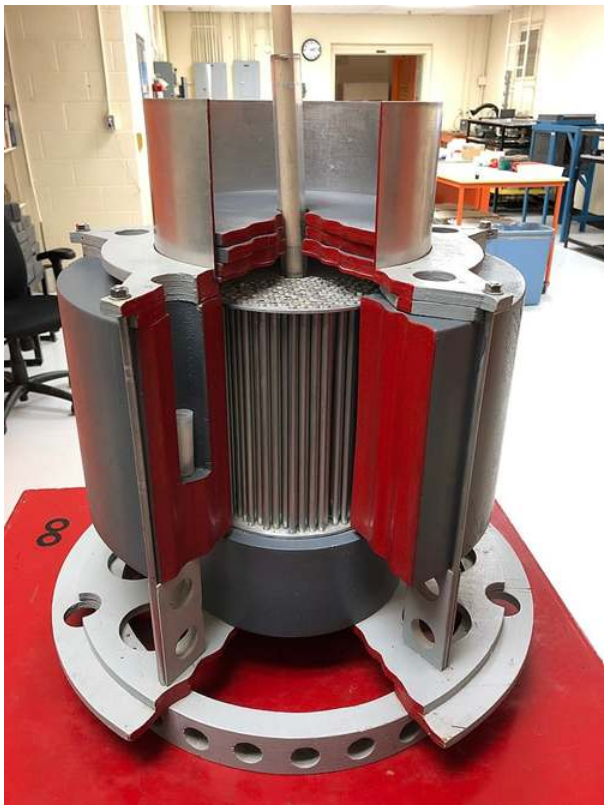


Figure 4.6.2.2: Typical assembly of fuel rods in a reactor core – here shown in a “demo-model” of a small research reactor (from Wikimedia Commons).

What should be done next to obtain a controlled chain reaction? If a slow neutron now reenters a fuel rod, it “sees” a completely different situation. The cross section of the fissile U-235 nuclei is now several hundred times larger than that of U-238 nuclei. So, even though the U-235 nuclei are still a 1:140 minority, it is more likely that the neutron will interact with one of them, than with a U-238 nucleus. And trigger another fission, releasing 2-3 new fission neutrons –and the chain reaction will continue!

There is still one small problem – after zig-zagging in the moderator, the neutron may simply escape. How to make sure it re-enters the fuel rod? Well, should it necessarily be the same fuel rod? Of course not, another fuel rod will be as good as the one in which the fission neutron has originated. To increase the probability of re-entering, the uranium fuel in reactors is arranged in the form of a lattice consisting of several hundred, or even of several thousand fuel rods.

---

1. If after the first collision the kinetic energy is 0.975 of the initial value, then after the second it is $0.975 \times 0.975 = 0.9752$. After the third collision it is $0.9752 \times 0.975 = 0.975$s, and so on. In order to find the number of collisions that will reduce the energy 1.6×10**7** times - in other words, so it is only $6.25 \times 10^{-}$of the initial energy - one has to solve the equation $0.975^{N} = 6.25 \times 10^{-}$, which can be readily done using a calculator.

4.6.2: No magic tricks are necessary! is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.6.3: Controlling the Chain Reaction and Cooling the Assembly

If a chain reaction starts in an assembly of fuel rod, it may very rapidly develop into a process running out of control. Therefore, special rods are placed in between the fuel rods in order to control the reaction intensity. Such rods are made of materials strongly absorbing neutrons with energies in the range of a few eV. The most often used material are Cadmium metal, or compounds of Boron. The control rods can be moved up or down in guiding tubes.

In a normal procedure of starting up the reactor the control rods at first are fully lowered. Then, a small neutron source is placed in the middle of the fuel rods – nothing will start unless there are some “seed” neutrons in the system. Then, the control rods are slowly being raised, while the flux of neutrons in the moderator is all the time monitored. At some moment, the neutron flux starts increasing even if the control rods are no longer moving – it means that the reactor “has reached criticality”, it doesn’t need “seed neutrons” any more. The extra source may be thus removed. Raising the control rods may continue, until the rate of the fission reactions reaches a desired level. Then, a special mechanism move the rods slightly up the rate slows down, or slightly down, if the rate starts speeding up. Such an action is called a “negative feedback”; other examples of negative feedback you must know is the cruise control in a car: if the car’s speed starts decreasing because, for instance, the is going slightly uphill – the cruise control electronics slightly increases the pressure on the gas pedal. And vice versa, if the car starts accelerating because the road now goes slightly downhill, the automaton will release the pressure on the gas pedal.

The system of controlling the reactor power by moving the control rods is an “electro-mechanical feedback loop”. However, there is another negative feedback mechanism, we can call it “intrinsic”, because it does not need any electronics and electric motors. The thing is that the efficiency of the reactor moderator is usually slightly temperature-dependent: if the temperature goes up, the flux of neutrons with the lowest energies slightly decreases. So a small increase of the reactor power causes a small increase in the moderator’s temperature – and a small drop in the reactor power – which in turn lowers the moderator temperature – and so on.

But as long as the reactors works, thermal energy is generated inside, about 200 MeV for each fission event. One could argue: well, since energy is continuously released, the temperature can only increase – hence, the “power stabilizing thermal feedback effect will never work! This is certainly true. However, any reactor, if supposed to operate for a prolonged time, **must** be equipped with a cooling system removing the heat generated at a constant rate. And then the power stabilizing thermal feedback effect will indeed work.

4.6.3: Controlling the Chain Reaction and Cooling the Assembly is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.6.4: Power Reactors

It is possible to build high-power reactors using natural uranium as the fuel, and heavy water D_2O as the moderator – as discussed in the preceding section. This design is used in Canadian CANDU reactors – there are some such units in power plants generating as much as nearly 1000 MW of electric power. Natural uranium is the cheapest reactor fuel – but, regretfully, D_2O is pretty expensive and therefore the CANDU-type reactors have never become very popular – there are only about 30 of them inactive service worldwide, out of the total number of 449 ((data for the end of the year 2017).

Another option for natural uranium rectors is to use high-purity graphite as the moderator, much less expensive than D_2O. In fact, all reactors built in the US during the first few years of the "Atomic Age" were graphite moderated. Such reactors are perhaps the best choice for the production of Plutonium, the material needed for making nuclear weapon. But definitely they are a poorer choice for civilian electric power generating plants. Such reactors are known to be unstable in certain circumstances. An instability resulting from human errors led to the catastrophic explosion of the huge RBMK-1000 graphite power reactor at Chernobyl in 1986. In 2013, there were still 11 such reactors in service in Russia, but new units have not been built after 1986.

The most common type now is a reactor moderated by ordinary water, H_2O. From the discussion following the derivation of the Eq. 4.8 one might conclude that ordinary hydrogen, with its nucleus consisting of a single proton, should be even a better moderator than Deuterium whose nucleus is approximately twice as heavy as a single proton. This is true – a neutron loses more energy in a collision with a single proton than with a deuteron. But protons not only slow down neutrons – quite often a neutron hitting a proton "decides" not to get bounced off, but to stay coupled with it instead. Such a neutron, of course, is lost and cannot support the chain reaction. In other words, ordinary water has too strong an absorption coefficient: there is no way of building a natural-uranium fueled reactor with a H_2O moderator. Yes, natural-uranium fueled. But once "enriched" fuel became available – i.e., a mixture of uranium isotopes with the U-235 content significantly higher than 0.7% – building reactors moderated with ordinary water became feasible. It turns out that enriched uranium with U-235 content of 3-5% already makes a good fuel for such reactors.

Today, over 400 power reactors operating worldwide are water-moderated. There are two basic types used, "Boiling Water Reactor" (BWR) and "Pres- surized Water Reactor" (PWR). The basic design scheme of each type is shown in Figs. Figure 4.6.4.1 and Figure 4.6.4.2. In both types water is used both as a moderator and as a coolant. In both types the fuel rod assembly (the reactor's "core") is placed inside a pressurized water tank, usually referred to as a "vessel". The BWR design is simpler. It operates at a pressure of about 7.5 MPa (75 times the atmospheric pressure) at which the water boiling point is close to 300°C (the dependence of the water boiling point on pressure is shown in Fig. Figure 4.6.4.1). The thermal power generated in the core heats the water to the boiling point, and the steam generated by the process is sent directly to a turbine – where its thermal power is converted to mechanical power, which is used next to operate an electric generator. The exhaust steam is next sent to a "condenser", where it's cooled down and changed back to water – and then injected back into the pressure vessel by special pumps.

One disadvantage of the BWR system is that water/steam which inside the reactor is in direct contact with high radiation is sent directly to the turbine – the water becomes radioactive and even though its radioactivity decays in days, a leak in the steam/water loop may cause a serious safety hazard.

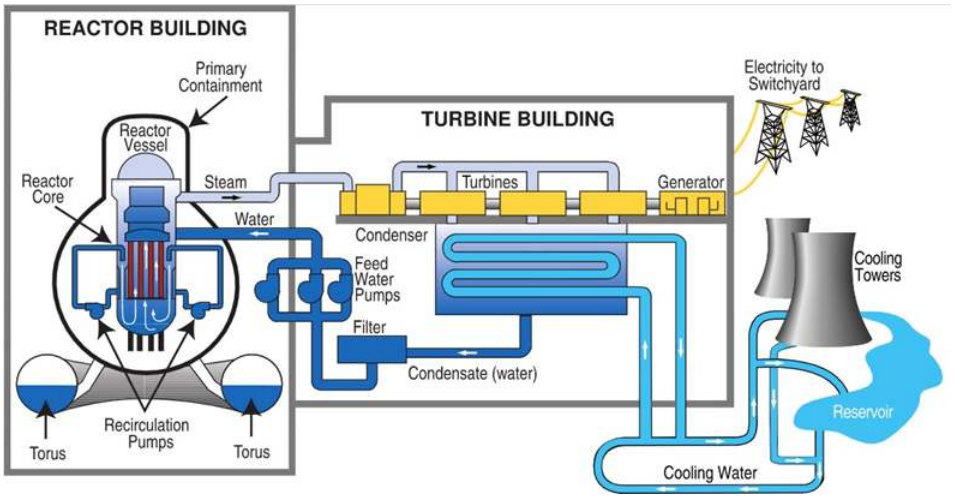


Figure 4.6.4.1: Basic design scheme of a BWR reactor (from Wikimedia Commons).

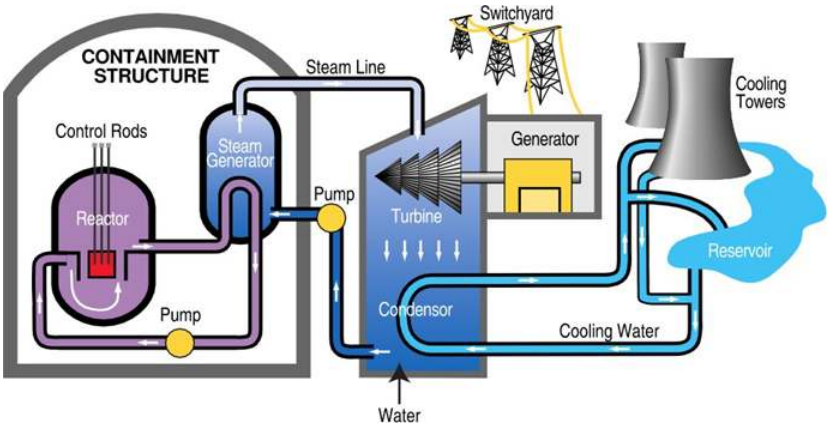


Figure 4.6.4.2: Basic design schemes of a PWR reactor (from Wikimedia Commons).

A much safer design is the PWR reactor in which water that is in direct contact with the core travels only in a closed “primary loop”. The pressure in the reactor vessel is kept higher than the water boiling point, so that it always stays liquid. An essential component of the primary loop is a “steam generator” – essentially, it is a special type of a heat exchanger, a device used to transfer heat between two streams of gas or liquid.

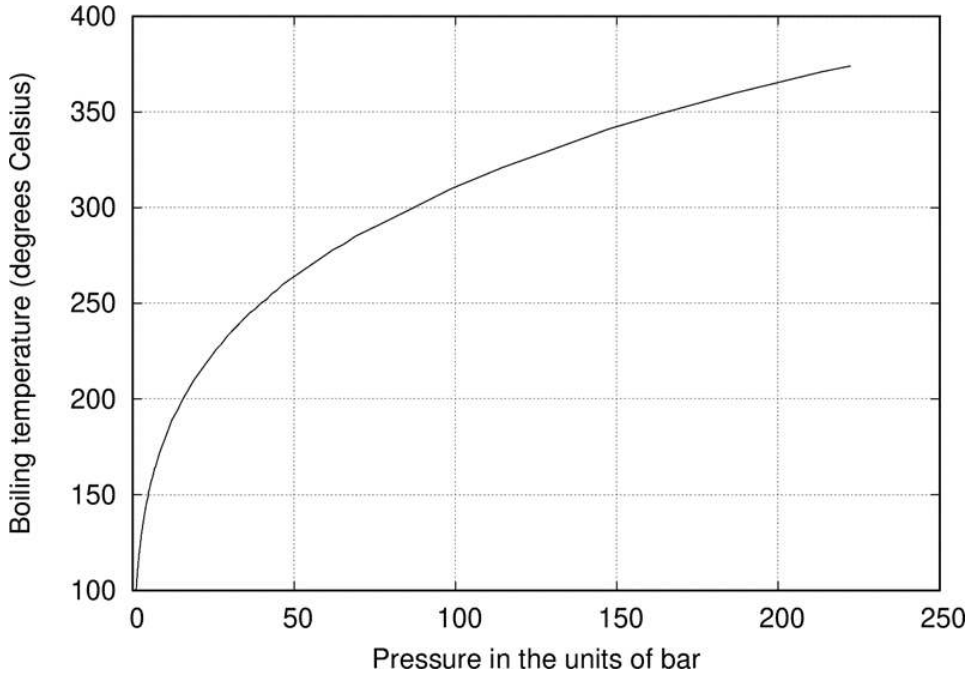


Figure 4.6.4.23: Boiling temperature of water vs. pressure. The curve does not go any farther beyond what is plotted in the figure – it ends at the so-called "critical point", above which water does not exist any more in liquid form. The "bar" is defined as 100 000 Pa – i.e., it's only slightly lower than the standard atmospheric pressure at the sea level, equal to 101300 Pa (the plot is based on the boiling temperature vs. pressure data listed in The Engineering ToolBox Web page.

The primary loop water is pumped through the generator, where it passes its thermal energy to a stream of water from the secondary loop. The water from the primary loop and that from the secondary loop are not in direct contact – all heat is passed from one stream to the other through the walls of metal tubes. Before exiting the generator, the water in the secondary loop is hot enough to change to steam. This steam is sent to a turbine, then the exhaust steam is condensed to water and sent back to the steam generator. So, the steam that is operating the turbine never gets in contact with high radioactivity. The PWR design is thus safer that the BWR one – on the other hand, it's more complicated and its efficiency of converting heat to mechanical power is slightly lower than in the case of the BWR reactors.

It should be noted that both BWR and PWR reactors have an "intrinsic safety mechanism" built in: if the cooling system fails and the water leaks out, the chain reaction in the core simply stops. It's not the end of the problems because the decay of the short-lived fission fragments still generated heat and after a short time the core may start melting and shortly afterwards may form a "pool" of molten metal at the bottom of the vessel. But at least there is no such an explosion as the one that happened in Chernobyl: in a graphite reactor, cooling water is circulated through a network of pipes inside the graphite core – but it is the graphite which is the moderator. So, If the water leaks out, the chain reaction does not stop – in a moment the reactor may get catastrophically overheated and explode.

Over the years, there has been a steady progress in developing new de- signs of BWR and PWR reactors with improved parameters and enhanced safety systems. The reactors built before the year 1990 are referred to as the "Second Generation" devices, and those built later, as the "Third Genera- tion" ones. Now, most attention is attracted by "Generation IV" reactors which have not yet been built, but "are on the horizon".

4.6.4: Power Reactors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.6.5: Generation IV Reactors and Thorium Fuel

Generation IV reactors are not a reality – not yet – they exist mostly as concepts, some of them in advanced phase of planning, and only in a few cases there exist functioning prototypes. In view of that, it does not make sense to discuss details of those new concepts in the present e-text. Rather, it seems to be a better idea to provide links to Web sites where much information about those innovative ideas can be found.

There exists an international organization, the Generation IV Interna-tional Forum (GIF). In its Web page, it declares: *GIF is a co-operative international endeavour which was set up to carry out the research and de- velopment needed to establish the feasibility and performance capabilities of the next generation nuclear energy systems.*

GIF has selected six reactor technologies for further research and development. They are listed in this Web page. By clicking on each item one can get much more information about the given concept.

It should be noted that out of those six goals, four are fast-neutron reactors. We haven't talked yet about fast-neutron reactor technologies in this Chapter. Until now, we discussed only the principles of operation of reactors using natural or low-enrichment (3-5%) uranium as a fuel. Such reactors **do need** moderators. However, a chain reaction may also be supported by **fast neutrons**: this is the case in nuclear weapon. Atomic bombs are made of highly enriched uranium (80% or more of U-235) or Plutonium-239[1] which is 100% a fissile material. Well, but atomic bombs are expected to explode in a most violent way. This is achieved by making the mass of U-235 or Pu-239 *critical* (e.g., by combining two or more blocks of sub-critical mass into a single block). In a fast-neutron reactor the mass of nuclear fuel must reach the critical level, too, to start the chain reaction – but the goal is to keep the rate at a steady controlled level.

Out of the four fast-neutron reactor types on the GIF goal list, only one type has underwent extensive testing in several large-scale prototypes – namely, the liquid sodium cooled type. Two large-scale French devices, Phoenix and Super Phoenix were tested for about 25 years before the year 2009, demonstrating that there were capable of generating power at the level of hundreds of Mega Watts for prolonged periods. But the unquestionable leader in research on such reactors is Russia, where the testing program was crowned by commissioning the first-in-the-world regular power plant BN-800 which has started supplying the grid with 880 MW in November, 2016 – with an even larger reactor, BN-1200, under construction and expected to be commissioned in the near future.

The fast-neutron reactors are especially interesting due to their "fuel breeding capability". Namely, when operating, they release more fission neu- trons than are needed to maintain the chain reaction. The "extra" neutrons may be absorbed by U-238 nuclei – either those in the fuel, or those in a spe- cial "blanket" surrounding the reactor core – to create Plutonium-239, which can be later used as the fuel for the same reactor. It turns out that amount of Pu-239 produced may be **even higher** than the amount of fissile material spent for maintaining the chain reaction. In other words, the reactor may "self-supply" itself with nuclear fuel. In practice, it means that the fertile[8] U-238 can be converted to valuable fuel, so that not only 0.7% of the nuclei present in natural uranium, but 100% of them can be used for generating power in future nuclear power plants.

Thorium Fuel

One more fascinating opportunity is that fast-neutron reactors can use not only U-238, but also Thorium-232 for "breeding". Thorium-232 is a non- fissionable element, but irradiated with neutrons, it changes to Th-233, which decays with a half-life of 22.3 minutes to protactinium Pa-233, which in turn decays with a half-life of 27 days to Uranium-233:

$$ {}^{232}_{90}\mathrm{Th} + {}^{1}_{0}\mathrm{n} \longrightarrow {}^{233}_{90}\mathrm{Th} \xrightarrow[22.3\ \mathrm{min}]{\beta^-} {}^{233}_{91}\mathrm{Pa} \xrightarrow[27\ \mathrm{days}]{\beta^-} {}^{233}_{92}\mathrm{U} \qquad (4.6.5.1) $$

Uranium-233, like U-235, is a fissile isotope. There is one problem with such "breeding" reaction because the Protactinium-233 has a high cross section for neutron absorption. If it captures a neutron before it decays to U-233, it leads to the production of a lighter Protactinium-232 isotope in an exotic reaction:

$$ {}^{233}_{91}\mathrm{Pa} + {}^{1}_{0}\mathrm{n} \longrightarrow {}^{232}_{91}\mathrm{Pa} + 2{}^{1}_{0}\mathrm{n} \qquad (4.6.5.2) $$

which decays to U-232, which is an unwanted contamination. Extra means are needed to deal with it; nonetheless, using Thorium as a nuclear fuel is definitely feasible. Because thorium is three to four times more abundant than uranium, thorium fuel may

become an important player in the future system of satisfying the growing appetite for power on the planet of Earth.

---

1. Plutonium ^{239}Pu is a “man-made” element which is created during a normal operation of an uranium reactor by absorption of neutrons by U-238 nuclei. The resulting ^{239}U undergoes two β decays and becomes ^{239}Pu, the half-lifetime of which is over 20 000 years. Plutonium is chemically separated from spent fuel rods. U-238 is not fissile, but if it captures a neutron, then a decay chain starts, in which after a short time a **fissile** nucleus emerges – namely, Plutonium-239. Therefore, U-238 is called “fertile” – as is another non-fissile material, Thorium-232, which can also be transmuted to a fissile Uranium-233 through irradiation with neutrons.

4.6.5: Generation IV Reactors and Thorium Fuel is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.7: NuScale's Small Modular Reactors – an Oregon-born Project

Nuclear power has not been the "darling" of public opinion. The Three Mile Island Accident of March 28, 1979, and the Chernobyl disaster of April 26, 1986, greatly contributed to the emergence of such negative attitude. The first of those events had created widespread fear, but not really serious consequences. The other was an authentic catastrophe. After that, for a quarter of century there were no serious accidents in the global nuclear power sector and its bad reputation began to slowly improve – until a new blow that hit a Japanese nuclear power complex in Fukushima Daiichi on March 11, 2011. The enthusiasm for nuclear energy almost completely disappeared due to this disaster.

Nonetheless, in the opinion of many experts issuing a "death penalty" on nuclear power would be an unforgivable mistake. As is well known, a goal supported by more and more people is to attain a total "decarbonization" in the 21st Century. It's a colossal challenge. To make the Reader aware of the scale of the problem, let's do some numbers.

Current global consumption of fossil fuels is around 14,000 MTOE per year. MTOE is an energy unit, acronym of "Million Tonnes of Oil Equiv- alent". It's instructive to convert MTOE-s to units more commonly used: there are several Web calculators (e.g., this one) for quick conversion to other units. For instance, one can find that 1 MTOE is equivalent to 11,630 GWh = 11.63 TWh. 1 GW is the power generated by a medium-size conventional power plant; there are 24,365 = 8,760 hours in a year, so that the total energy output of such a plant is 8,760 GWh = 8.76 TWh – or 0.753 MTOE. Now, let's take a look at the data published by the International Energy Agency. One can find that the total global generation of wind energy in 2019 was 1430 TWh, and 900 TWh of solar energy – corresponding, respectively, to 123 and 77 MTOE, or 200 MTOE together. Well, we see solar panels and wind turbines all around, but the data are bit disappointing – in order to replace all CO_2-emitting energy sources, we would need to add 60 times more solar panels and 60 times more windmill than those currently installed! But the aforementioned 14,000 MTOE is the total consumption of en- ergy, which includes all sectors, such as transportation, metallurgy, chemical industry, etc. – the gross electric power generation in 2019 is about 27,000 TWh, corresponding to 2320 MTOE, which is about 16.6% MTOE consumed. Let's keep in mind that the wind and solar sources generate almost exclu- sively electric power – so that the 200 MTOE of the two sources combined contributes about 8.6% to the gross electricity production.

And now, let's take a look at the nuclear sector, which generates power that is 100% emission-free. In 2019, an the total global output of those plans is about 2560 TWh = 220 MTOE, or about 9.5% of the gross electric energy production. More than the solar and wind energy combined.

In developed countries the dominance of nuclear energy may be even higher: in the US, nuclear generation, and wind + solar generation were, respectively, 19% and 8.2% of the total, and in France, respectively, 74.5% and less than 1% (2016 data).

In view of the above, many experts in the field and not only them agree that – once we face a "Herculean challenge" of total "decarbonization" of energy acquisition well before the end of the 21st – it does not make much sense to eliminate the electricity generating branch that is "patently zero- emitting". Rather, it may be a good idea to build new nuclear power sources in parallel to new wind and solar power sources.

Guidelines for New Reactor Safety Standards

The specter of the Three-Mile Island, Chernobyl and Fukushima Daiichi disasters still haunts our thinking of nuclear power. It is instructive to consider what went wrong in those catastrophic events. In the TMI case, faulty operation of some devices and sensors triggered a chain of malfunctions which, when combined with human errors, led eventually to a partial meltdown of the reactor core. In the case of Chernobyl catastrophe, what contributed were serious errors in the design of the reactor (of a different type than the reactors used in most Western countries), combined with a series of human errors (the whole story cannot be summarized in a few sentences, for details the Reader is referred to one the many reports published by appropriate Agencies – e.g., to this, a relatively compact one – and many other can be found in the Web). Finally, in the Fukushima Daiichi there was no obvious human error made during the disaster itself. The error in question had been made years earlier when it had been decided to build a power reactor complex at the ocean shore, in an area where catastrophic earthquakes and tsunamis may happen.

In all these three cases it was the failure of the core cooling system that initiated the critical phase of the catastrophe, from which"there was no way out". And the "hearths" of the cooling systems – as in almost all other power reactors in the world – were colossal pumps, driven by electric motors supplied from the power grid.

Another common denominator in all of them was that the malfunction- ing reactors were large units, with net generating capacity close to, or even larger than 1 GW. So, all the reactors contained a large amount of radioactive elements. Only in the case of TMI the release of radioactive materials was relatively modest – in the other two cases large areas of land were contaminated.

The lessons that can be drawn from the facts presented can be summa- rized as follows:

- Future reactors should be designed in such a way that both in normal operation and in emergency situations human intervention would not be necessary.
- Make the core cooling system completely autonomous, preferably by complete elimination of pumps supplied by power from external sources, prone to power outages. it is entirely conceivable to build a deep shaft cove
- Future generations of reactors should be much smaller than those being currently operated – not units of power of the order of 1 GW (1000 MW), but of 100 MW or even smaller. Why? Even if a reactor is 100% immune to the mechanical failures and human error, there are still unpredictable cataclysmic events such fall of a meteorite, or 9/11. The amount of radiation released in such an event would depend on the mass of the uranium fuel in the reactor. In a typical reactor of 1 GW power there is about 100 tons of fuel, and in a 50 MW reactor there will be only some 5 tons. Small size of a reactor also gives a much higher guarantee that in the case of a cataclysmic event the fuel will not be released and spread out over a large area.

  All power reactors are enclosed in a reinforced “containment shells” designed to contain the escape of radioactive substances in the case of an emergency. However, placing such a shell of a 1 GW reactor under- ground would pose an extremely difficult technical problem. Therefore, all such facilities (the existing and the planned ones) are placed above the ground level, thus making possible targets for a meteorite strike , or for a 9/11-style attack. In contrast, for small reactors with power of several dozen MW it is entirely conceivable to build a containment structure in the form of a deep shaft underground shaft.
- Finally, the location of the reactor must be very thoughtfully consid- ered. Areas where tsunamis or floods may happen, or where earth- quakes may occur, must be excluded from such considerations.

The NuScale Module

The design of the module created by NuScale Power is consistent with the guidelines outlined in the preceding subsection (only the last item, related to the location, must remain the concern of those who will buy the reactor, and not of them who have designed the reactor).

In fact, what is called “a module” by the company, is a compact nuclear reactor capable of operating all by itself, and of producing steam that can be used (by a separate facility) to generate about 60 MW of electric power. The reactor is slim, only about 9 feet (2.7 m) in diameter and of the height of 65 feet (20 m). The module contains only about 5 percent of the nuclear fuel of a conventional 1,000 MW_e nuclear reactor, which is one factor that contributes to its safety. Another factor is the module’s slimness – it can be housed in an underground cylindrical “well” made of reinforced concrete. Needless to say, such an underground tube is a much sturdier “containment structure” that anything built above the ground level.

One can say that 60 MW of power is not much. Therefore, the design is such that up to twelve individual modules can be teamed up, yielding the output power up to 720 MW, not much less than than the typical power of about 1 GW of a typical conventional power reactor. There are many advantages of such an arrangement. Defects cannot be ruled out entirely, but in a “12-pack arrangement” it is quite unlikely that more that one unit will be malfunctioning – that module can be turned off and taken care off. In the worst scenario, it can be replaced by a “fresh” module. So, the system will be working for a while as an “eleven-pack” with power reduced by less than 10%. Another advantage is there may be no need of shutting down the system for refueling. Most conventional reactors in the 1 GW output power range typically refuel every 18 to 24 months, and during such an operation the reactors needs to be shut down for three weeks, or even longer. In a “12- pack” SMR power plant, the modules will not be refueled *in situ* – instead, they will be removed and transported to the factory, where the refueling operation will take place. And the removed unit can be replaced by a “fresh one”, brought from the factory. While a single unit is removed and replaced, the remaining eleven may keep working. Accordingly, the power plant will keep working with the power reduced by less than 10% for a relatively short period – but its operation will never need to cease completely.

But the most remarkable technological novelty in the NuScale Power’s SMR design is the core cooling system. The pumps are completely eliminated – instead, the circulation of the cooling agent (water) is maintained by a natural physical phenomenon,

known as *convection*. It's a heat transfer mechanism, in which heated gas or liquid travels away from the heat source, carrying thermal energy along.

An instructive example – the hot air balloon. A spectacular demon- stration of a convection-like effect is a hot air balloon. In such an aircraft, a large volume (typically 2 - 5 thousand cubic meters) is contained within a "bag", called the *envelope*. The density of air decreases with increasing temperature, at constant pressure it is inversely proportional to the absolut temperature (i.e., the one expressed in Kelvins). The air temperature in a balloon is usually kept at about 120˚C = 393 K. The ambient air temper- ature at hot air balloon events is usually close to 20˚C = 293 K, and the ambient air density not too high above the sea level (e.g., in Willamette Valley) is about ρ_A = 1.2 kg/m^3. Then, the hot air density inside the envelope is ρ_H = 1.2 (293/393) kg/m^3 = 0.895 kg/m^3.

The *buoyant force B* acting on the balloon – i.e., on a body of volume V "immersed" in ambient fluid (air) of density ρ_A – can be readily found from the Archimedes Law: $B = V \rho_A g$, where g = 9.81 m/s^2 is the acceleration due to gravity. But the hot gas inside the balloon is not weightless, its weight (i.e., the force with which Earth attracts it) is $W = V \cdot \rho_H \cdot g$. By subtracting it from the buoyant force B we get the *lifting force* F_L of the balloon:

$$F_L = V \cdot (\rho_A - \rho_H) \cdot g \tag{4.7.1}$$

One can readily find that for the conditions outlined above, each cubic meter of hot air contributes a lifting force of (1.2 0.895) 9.81 N = 2.99 N. 2.99 N is the weight of a mass of 0.305 kg. Then, a typical balloon of 2500 m^3 volume can lift a total mass of 2500 · 0.305 kg = 763 kg.

Natural convection in gases and liquids. An "envelope" is not needed for a portion of hot air to rise, of course. It happens whenever air at a low altitude is heated – for instance, as is discussed in Chapter 6, it is the global- scale convection that gives rise to the so-called *trade winds*. Air convection over lands produces "up-blowing winds", taken advantage of by many bird species, as well as by pilots flying gliders.

An artificial device utilizing convection – known to humans for thousands of years – is a chimney. A vertical pipe in which the convection creates an up-moving air current that captures smoke and combustion products from the inside of a stove or a fireplace and blows them out.

Convection also occurs in liquids. It plays an important role in the heat transportation in oceans and other deep bodies of water (by the way, vertical convection currents in the Oregon's Crater Lake are studied and monitored by OSU oceanographers, as the Lake can be thought of as a "microscale model" of an ocean).

In a teaching lab scale, convection in liquids can be visualized in simple experiments – e.g., by injecting a "blob" of dyed hot water near the bottom of a beaker filled with clear cold water, and observing how the "blob" raises towards the surface. Many such experiments are presented at YouTube, for instance – in this movie clip, in another one, and in yet another one. The Reader is encouraged to look for other similar movies, there are many other demonstration of convection in the YouTube.

The NuScale module unique convection-based core cooling system. In a reactor using water as a moderator/coolant, the core temperature should be maintained at 300˚C or slightly higher. But why not at even higher a temperature? Well, we have to keep in mind that substances exist only at liquid state at temperatures lower than the so-called *critical point*, specific for a given liquid (for a list, see the table at the end of this Wikipedia article. For water, the critical temperature is 374˚C. But not only the temperature is relevant – another essential parameter is the *pressure*. As we all know, at normal atmospheric pressure liquid water exists only up to 100˚C (212 F). At higher temperatures, in order to keep water liquid the pressure has to be increased – for a given temperature, it has to be kept above the so-called *boiling point pressure*. The boiling point parameters for water up to the critical point can be found in in this Web page (from the graph, or by using the on-line tool). For the temperature of 300˚C, one gets the boiling point pressure of 86 bara (*bara*, or *bar* is a unit of pressure liked by engineers; it's equal to 100,000 N/m^2 – which is slightly less than the normal atmospheric pressure, equal to 1.013 bara). Consequently, inside all boiling water reactors (BWRs) and pressurized water reactors (PWRs) the pressure during normal operation must be of the order of 100 bara.

The NuScale modular reactor is essentially a small PWR, so that the core temperature and the pressure inside will be similar as in the "big" reactors of this type. But there is one more temperature dependent parameter of water that is essential for the module operation: namely, ***the density*** . As all other gases, liquids and solids, water experiences ***thermal expansion*** : the volume of a given mass of water *expands* in response to increasing temperature – hence, its density decreases. *decreases*. For a temperature as high as

300°C the change of water density is already substantial. At 4°C the density of water is exactly 1000 kg/m^3. At 100°C – as can be found from the graph in Fig. 4.7.1 – it's 950 kg/m^3, and at 300°C it is already as low as 720 kg/m^3.

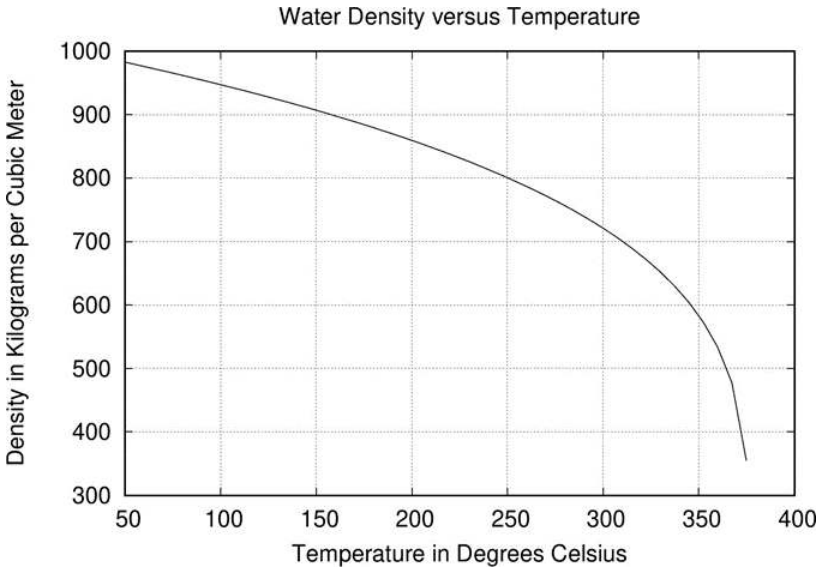


Figure 4.7.1: The dependence of water density on temperature (plotted based on the data listed in this Dortmund Data Bank Web page).

Suppose that 100°C water enters the core of a NuScale SMR from below and inside the core it is heated up to 300°C. Its density – as follows from the numbers given above – dramatically decreases. It gives rise to a strong convection, leading to an upward flow of the heated water and escape through the core's top – whereby a new portion of cool water is sucked into the core from below. In other words, the convection acts as a pump circulating water through the core and carrying out thermal energy from it.

In order to assure that a large amount of thermal energy is removed from the core every second, the circulation must be made really "vigorous". Is there any way to speed up the convection? Yes, there is, and people have discovered a proper "trick" centuries ago when experimenting with chimneys. They found out that the higher the chimney was, the stronger was the chim- ney "draft" and the faster was the flow of hot gases in the chimney duct. This is often referred to as the stack effect. If we look at old photos of factory complexes from the "industrial revolution" of the 19th century, there are tall chimneys everywhere. Why? Because before the age of electricity even in large factories there was typically only a single huge steam engine, delivering power to a large number of individual machines through a "cat's cradle" of shafts and pulleys. Such steam engines consumed tons of coal, the burning of which required an intense intake of air into the boiler furnaces. The rate of air influx is equal to the rate of combustion products (the "flue gas") removal from the furnace by the chimney. In order to speed up the removal, a strong "draft" was needed and this is why very tall chimneys were built. For the same reason – as one can see on old photos (e.g., that of "Titanic") – there were very tall "smokestacks" on merchant and navy steamships.



Figure 4.7.1: Photo of factory chimneys from the age of the 19th century industrial revolution (source: Wikimedia Commons).

The very same “stack effect” is taken advantage of in the NuScale re- actor module to enhance the core cooling. In the Figure 4.11 it’s explained schematically how this system works. The water heated up in the core passes next to long vertical channel acting as a “chimney” for producing an upward convection “drag”, thus speeding up the passage of water through the core. At the top of this “stack” is a “steam generator” – essentially, it’s a heat exchanger in which heat is transferred from water to steam circulating in a separate loop – this steam passes next to the turbines in the electricity- generating unit.

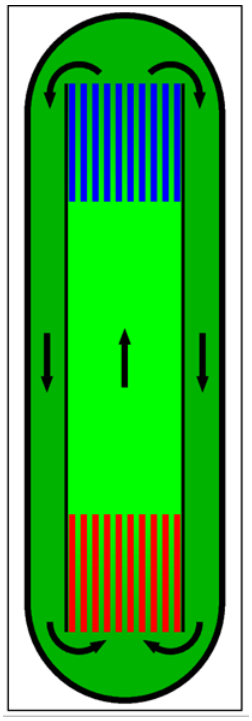


Figure 4.7.3: A scheme explaining how the convection-driven core cooling system works. The red bars are the fuel rods in the core, and the blue ones are the pipes in the “steam generator”. The low-temperature (100°C) water with higher density is colored dark-green, and the high-temperature (300°C) water with lower density is colored light-green. The cooler water enters the core from below, where it is heated up and its density drops. When it emerges from the core’s top, it is driven up by convention enhanced by a long “stack” between the core and the steam generator. After giving away its heat in the steam generator, the cooled-down water with higher density “sinks” towards the bottom, where it may reenter the core and go through another heating-up/cooling-down cycle. In the steam generator water injected into it is heated up and it leaves the generator as a high-pressure hot steam, which is sent to a steam turbine driving an electric generator. However, the drawing of the steam generator in the figure is very simplified and the pipes needed to inject water and to lead steam away are not shown.

The cooled-down water emerging from the “steam generator” again has a higher density and therefore it “sinks” tothe bottom of the system from which it may reenter the core, and undergo another heating-up/cooling-down cycle.

Fig. 4.7.3 also explains why such a reactor has to be “tall and slim”

- it’s because the “stack” has to be long enough to produce a strong “draft”
- strong enough to guarantee, in turn, that the water circulation will be fast enough to remove from the core the entire 60 MW of thermal power generated in it.

As follows from above, the water circulation in the system is driven only by convection and is not aided by any pump powered by electricity from an external source. Nate that this is crucial for the system safety! In the Three Mile Island and Fukushima Daiichi disasters it was the failure of the core cooling systems that led to the reactor core meltdown – an irreversible catastrophic event. Even though the chain reactions in the cores had been stopped, the heat was still generated due to the decay of fission products the lifetime of which can be as long as days or weeks. To prevent the core meltdown, it is necessary to keep the circulation of the coolant through the core for a long time. In the NuScale SMR design, the coolant circulation is driven by the heat generated in the core, no matter whether it’s due to the chain reaction, or only due to the decay of the fission products. In other words, the core cooling system is “intrinsically safe”.

It’s of crucial importance, but it still isn’t enough. The coolant circu- lation, yes, removes heat from the core – but, in turn, something has to removed the heat from the coolant. Otherwise, all the water in the SMR would eventually reach the same

temperature and the convection-driven cool- ing mechanism would stop. Normally, it's the steam generator that removes heat from the coolant, after it passes through core. So, suppose that there is a major disaster that knocks down the power generating system: water does not circulate any longer through the steam generator, its cooling action is completely stopped. The temperature inside the closed SMR "capsule" keeps raising, as also does the pressure – eventually, the capsule bursts open, nothing is cooling the core anymore and everything is heading towards a meltdown.

However, the NuScale designers have also taken such a scenario into con- sideration. They added another safety mechanism into the system. How it works is explained in the Fig. 4.7.4. The reactor capsule is installed into another larger "outer" shell and the space between the two vessels is evac- uated during the normal operation. The entire assemble is installed in an underground water-filled "well". Normally, the vacuum jacket around the inner capsule provides an efficient thermal insulation, so that any heat gen-erated in the reactor is removed by the steam generator only. Yet, if the chain reaction is stopped and, in addition, the steam generator becomes in- active and no longer removes any heat from the reactor, then valves in the outer capsule are opened and the space between the two shells is flooded with water. Now, the water inside the reactor circulates due to convection and heat is transferred out of it across the capsule wall, to the water that replaced the vacuum jacket.

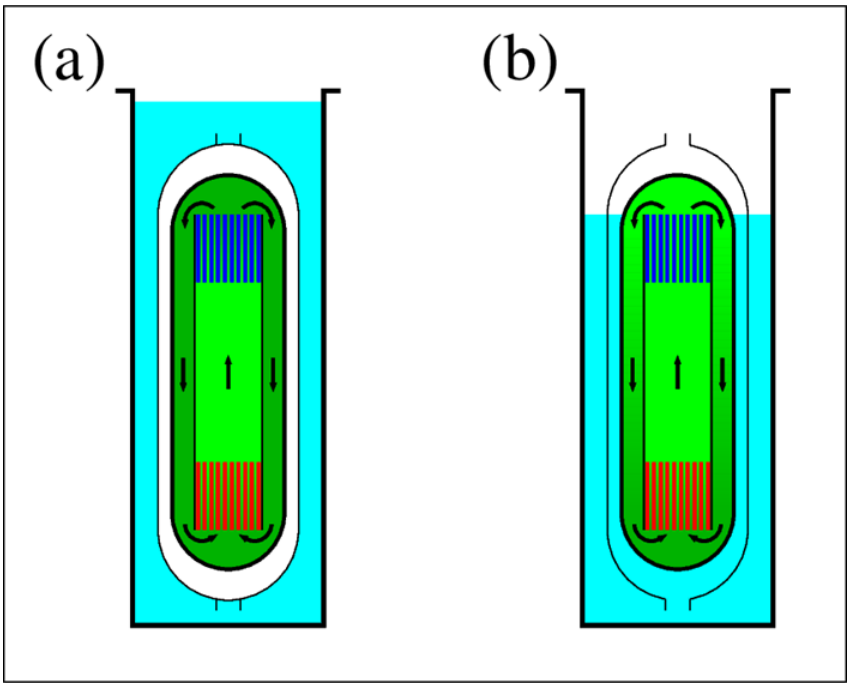


Figure 4.7.4: The reactor unit in the NuScale design is housed inside another "outer" capsuke. The whole device is placed in a deep well-like shaft filled with water. During normal operation, as shown in part (a), the space between the "inner" and "outer" capsule is evacuated, i.e., there is no thermal contact between them. In emergency situations, when the steam generator does not remove heat emerging from the core any longer, an emergency cooling system is activated, as shown in part (b): water from the "well" is let into the space between the two capsules. Now, the hot water flowing down inside the "inner" capsule passes its heat across the capsule wall to the "well water".

This heat makes the water boil, so that steam escapes from the well and the water level in it gradually lowers. There is enough water in the well to cool the reactor in this way for a month. It is assumed that the system will not stay abandoned for such a long period and some intervention will begin sooner. In fact, the simplest action of such kind is simply to keep adding water to the well. So, if necessary, the emergency cooling of the reactor can be extended by periodically refilling the well for a time period much longer than a month, until other more radical repairing actions become possible.

The diagrams in Figures 4.7.3 and 4.7.4 are pretty simplified – in order to make it easier to explain the general principle on which the reactor cooling works. For much more detailed drawings (like, e.g., this one) and for a much more detailed description of the safety features of the reactors and their functioning, the Reader is referred to the Home Page of the NuScale Power Company, and to many "daughter pages" that can be accessed from this Home Page.

4.7: NuScale's Small Modular Reactors – an Oregon-born Project is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

4.8: The Pros and Cons of Nuclear Power

This is an excellent topic for a classroom discussion! So, instead of presenting a list of “pros” and “contras”, the Author asks the readers to ponder and consider arguments they would use in such a discussion.

4.8: The Pros and Cons of Nuclear Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

5: Hydropower

5.1: Introduction

5.2: The Origin of Hydropower

5.3: Methods of Harnessing Hydropower

5.3.1: The Kaplan and Francis Turbines

5.3.2: The Pelton Turbine

5.4: About One Famous Hydropower Plant Using Pelton Turbines

5.5: Pumped Storage Hydroelectric Plants (PSHP)

5.6: New Ideas for Hydraulic Energy Storage

5: Hydropower is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

5.1: Introduction

The oldest “green power” harnessed of by humans (even before they had become humans!) is direct solar power. They used it to get warm in the morning, to make dried berries, meat, or fish... The second oldest is biofuel power, until modern times all fires needed wood, or fuels extracted from animals, live or dead:: burning animal fat in primitive lamps was a method of making light, and dried cow’s dung burned in simple furnaces was used for cooking food, or for keeping people’s dwellings warm at nights.

In chronological order, Hydro-power and wind power probably share the position of the third-oldest methods of extracting energy from natural sources. Which one should we discuss first? Well, as patriotic Oregonians, let’s give the priority to hydro-, because most of electric power generated in our State comes from hydro-power plants.

We don’t know when exactly our ancestors learned how to harness the energy of flowing waters for doing usable work – but certainly, it was a long time ago. Archaeological findings reveal that quite sophisticated water wheel machines were already used in ancient Egypt for lifting river water to irrigation systems in the 4th century BC. Ancient Romans at the end of the BC Era and the beginning of modern Era used waterwheels for powering various types of machinery, such as, e.g., sawmills. An excellent overview of the history of the waterwheel up to the modern days is given in this Wikipedia article.

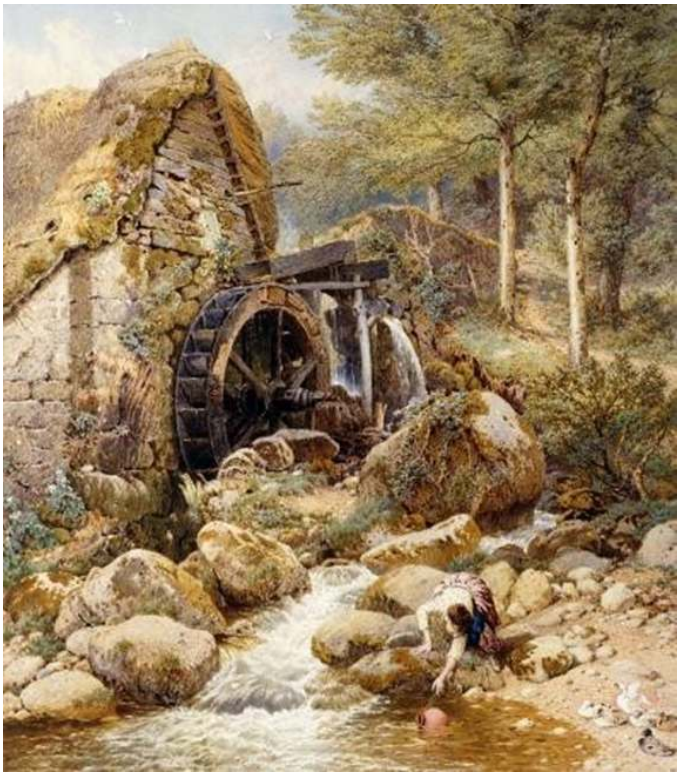


Figure 5.1.1:An old waterwheel, probably from the XIX Century (source: Wikimedia Commons).

The old charming waterwheel shown in the Fig. 5.1.1. could generate perhaps 1 or 2 kiloWatts of power, at the most. Things has changed dramatically when people started using hydro-power for generating electricity – which happened shortly after the discovery of electromagnetic induction by Michael Faraday, shortly afterwards followed by the invention of electric generator (also known as “dynamo”) based on the Faraday Law. Combining a dynamo with a waterwheel was a “no-brainer”. After the first hydro-power plant was built – reportedly, in 1878 – new hydropower plants started rapidly emerging. Due to the ever-growing worldwide demand for electric power, the hydropower plants were also growing, so that more and more Mega-, and later Giga-Watts of power could be supplied by them. Today the record-keeper, the “Dam of Three Gorges” hydro-power plant, generates 24 GigaWatts of power, some 20 milion times more than an ancient waterwheel. The power of the largest American hydro-power plants are not as spectacular as that of the Chinese and Brazilian supermonsters but they are by no means small. The largest American – the Grand Coule Dam – is, by the way (or by the FREEway, you may see it from I-86) on the Columbia River, not very far from us. Note that the famous Hoover Dam is not the largest American hydro-power plant as far as the power generated is concerned contrary to what many people think, it’s perhaps “the best known”, because of the height of the dam and the spectacular surroundings.



Figure 5.1.2: The dam of Three Gorges in China, the largest hydropower plant in the world, delivering 24 000 MW of power (source: Wikimedia Commons).



Figure 5.1.3: The 14 000 MW Itaipu Power Plant, until recently the largest in the world. It's owned jointly by Brazil and Paraguay (the Itaipu Riveris the borderline between the two countries; source:Wikimedia Commons).

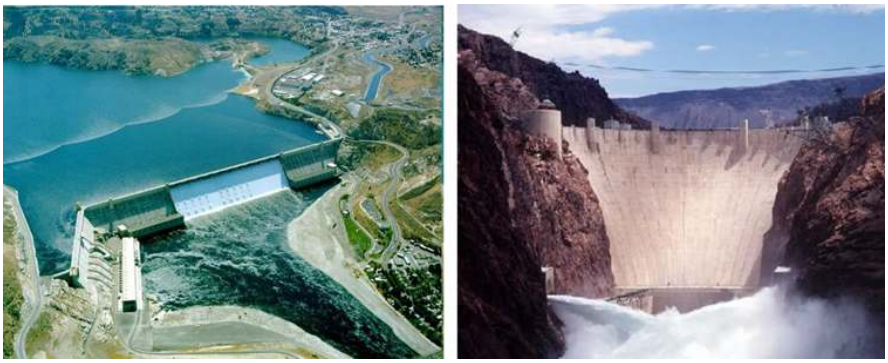


Figure 5.1.4: Left image: the largest US 6800 MW hydro-power plant, at the Grand Coule Dam on Columbia River. Right Image: the 2080 MW Hoover Dam hydro-power plant at Colorado River (source: Wikimedia Commons).

The power of the largest US hydropower plant, the Grand Coule, is listed as 6800 MW, and its energy output per one year as 22.6 Tera-Watt-hours. But if you divide this number by the number of hours in a year (365 24), we get the average power of only 2580 MW.

How comes? Well, this is something normal for most hydropower plants in the world. Usually, the power listed is the **maximum** power they can generate. But there is not enough water in the river to allow them to run at maximum power all the time. So why to install machinery with the output power $2\frac{1}{2}$ times larger than the power of the water flow in the river? The answer is simple: namely, the demand for electricity is not constant, it significantly varies over the 24 hour period, with two huge peaks during the morning and the evening rush-hours. Therefore, at other times the plants run at power lower than the average, using less water than flows into the reservoir – in other words, water is being stored behind the dam. And then, at rush hours, the water stored may be used for running at full power to help the grid to deal with the huge demand. For Hoover Dam (2080 MW maximum power, and 4 TW annually) the difference is even more pronounced:

$$\frac{4\ TW}{365 \times 24\ hours} = 457\ MW$$

it is, 4.5 times less than the maximum power.

The ability to store energy is an important advantage of the hydropower plants. In addition, they can react very quickly for sudden increase in the demand for power – they can switch from a low-power output to high-power output in a few minutes. Much faster than the “number two” on the list, the turbines fueled by natural gas, for which the “start-up time” is usually given as 25-30 minutes (for plants using steam turbines – which provide most of the electric power in the US – the start-up time may be as long as several hours).

5.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

5.2: The Origin of Hydropower

Please keep in mind that Hydro-power is nothing else than converted solar power. There are many good things spanning indirectly from the energy coming from Sun. Water from oceans, seas, lakes, rivers, and even from moist soil, when heated up by solar radiation, yields water vapor that rises up and forms clouds the clouds then condensate and the water returns to the Earth surface in various forms of precipitation: as rain, snow, hail... The water that has returned to mountains and other high areas then begins its "trek" toward the oceans. On the way, it loses its potential energy. Harnessing hydro-power by us is nothing else than "capturing" this potential energy before it all dissipates.

The "water cycle" is illustrated in Fig. 5.2.1, in which it is stressed that all energy that can be extracted from flowing or falling waters is actually solar power converted by Mother Nature. In the next Figure 5.2.2, it is shown how much radiative power our planet receives from the Sun – and how much of it is converted to hydropower that can be harnessed by us. Its almost one-half of the total global consumption of power. We still use only a fraction of those available resources – but even if we managed to harness them all, they would not satisfy all our needs. However, since there are still available hydropower resources not yet taken advantage of, it would be a wise thing to start using them. In contrast to fossil fuel resources, hydro-power resources cannot be exhausted!

In the upper graph in Fig. 5.2.3 it is shown how the global population "appetite" for electric power is growing over time – and what is the contribution of different sectors of power industry to satisfying these needs.

According to the World Energy Council, in the year 2016 the share of hydropower in global power generation was 16.4%. As indicated by predictions, this gure is not expected to change signicantly over the next 20-30 years.

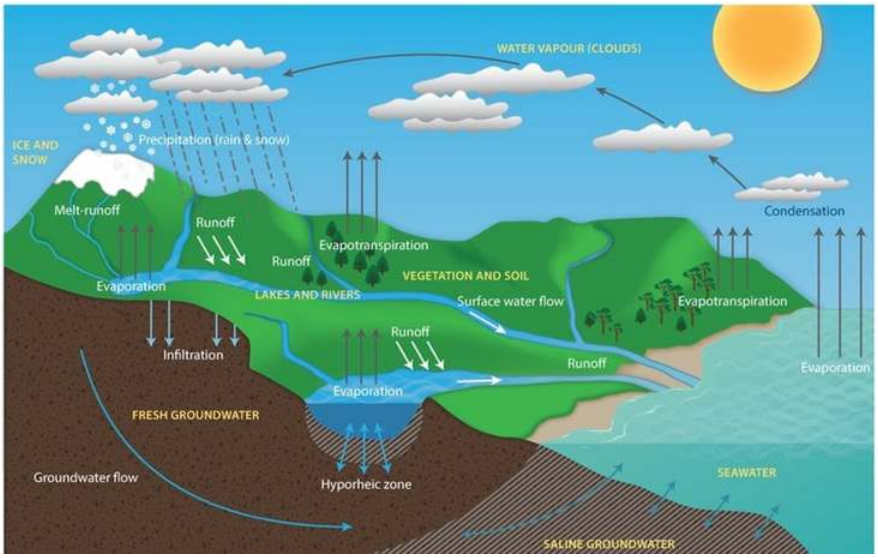


Figure 5.2.1: The water cycle (source: Wikimedia Commons).

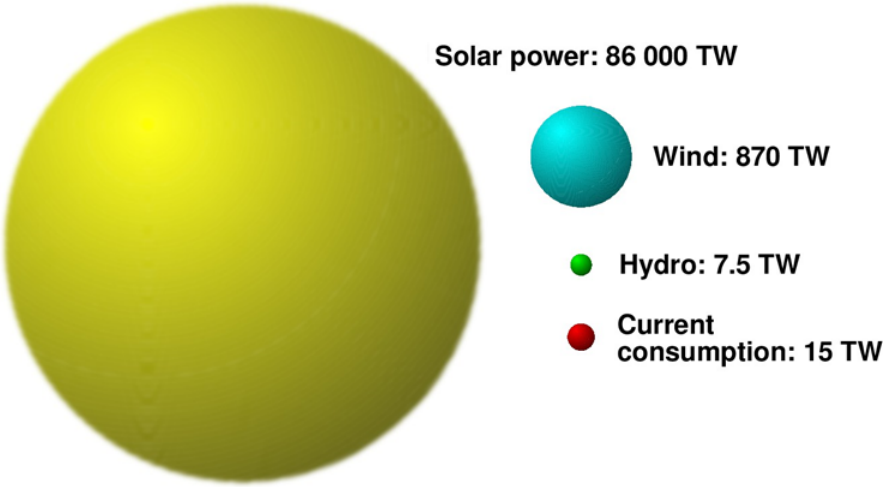


Figure 5.2.2: The total radiative power delivered to the Earth surface by Sun, and the available resources of two other forms of power, compared with the current total global consumption all all energy forms. The volume of each sphere is proportional to the corresponding power listed next to it (aop).

In the lower graph in Fig. 5.2.3 it is shown how much hydropower is generated by different countries. China generates most, but if one considers how big the populations in each country is and calculates the hydropower generation per one million population, then Brazil (current population slightly over 200 million) is the unquestionable winner, and the US is also ahead of China.

It is also remarkable that Brazil satisfies over 90% of its power needs from hydro. Norway, in which the hydro-sector satisfies 100% or even more of the country's needs (more than 100% – how is it possible? Yes, it is, Norway exports its surplus of electric power to other countries, primarily to Denmark through an undersea transmission line capable of transmitting up to 2 GW of power).

In the US, according to the US Energy Information Administration, the share of hydropower in the total usage is perhaps less impressible – it is about 6.5% of the total.

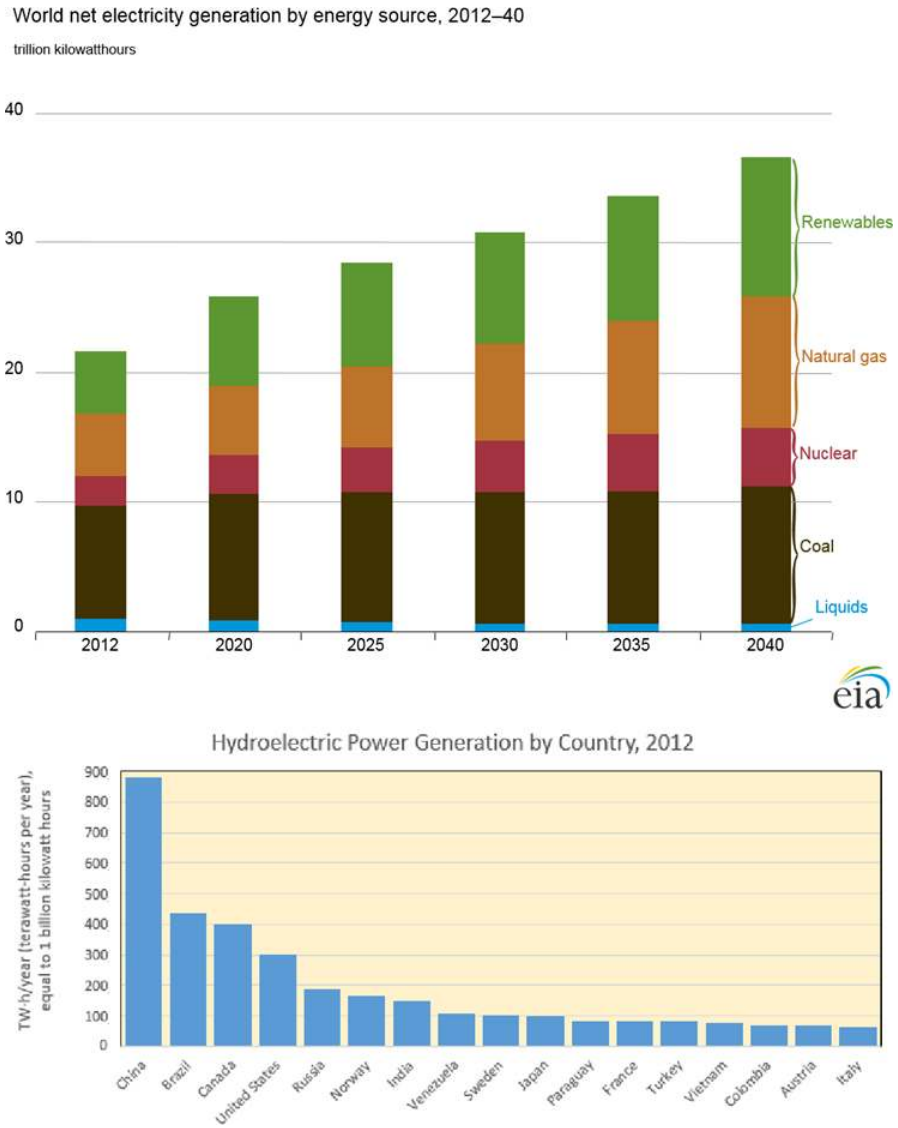


Figure 5.2.3: Upper graph: The global production of electric power showing the share of each major technique – prediction running up to the year 2040. Lower graph: generation of electricity in hydropower plants, by countries.

Figure 5.2.3: Upper graph: The global production of electric power showing the share of each major technique – prediction running up to the year 2040. Lower graph: generation of electricity in hydropower plants, by countries.

However, in the US the share of hydropower in individual states is very the different. The graph in the page below is from the year 1995 and is no longer very accurate. Current data from individual states can be found in this Wikipedia article, or in this US Energy Information Administration Web site – in the latter, there is a map with all states and if you click on one, you get info not only about hydropower generation in it, but also a bounty of info about all other power sources in this state.

Table 5.2.3 – Hydroelectric generaration in 2018 in TWh: Total USA and the leading states (all other states generate less tan 5 TWh each)

USA total	292.52
Washington	80.86
Oregon	35.44
New York	29.63
California	26.33
Montana	11.40
Alabama	11.14
Idaho	11.02
Tennessee	10.29
Arizona	6.98
North Carolina	6.61
South Dakota	6.27

Source: U.S. Energy Information Administration Detailed State Data 1990-2018 Net Generation by State by Type of Producer by Energy Source

5.2: The Origin of Hydropower is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

5.3: Methods of Harnessing Hydropower

How exactly is electric power generated at hydropower facilities? There are two major techniques. One is by exploiting rivers in which the stream gradient is not very high, but high is the volume of water flowing downriver. Along the river bed, the flowing water gradually loses its potential energy. If there is a dam on the river giving rise to an artificial lake, however, then the potential energy of water flowing into the lake is “fixed” at a value cor- responding to the altitude of the lake’s surface. Now, if there is an opening in the dam at the lake’s bottom, the water squirts out, its potential energy at the surface level is converted to kinetic energy.

The speed of the “jet” flowing out is given by the solution of the famous tank draining problem, formulated by Daniel Bernoulli nearly 300 years ago – and is:

$$v = \sqrt{2g\,(H_{\text{high}} - H_{\text{low}})}$$

where g is the acceleration due to the Earth’s gravity. For details of the formula derivation, look, e.g., at this Web site.

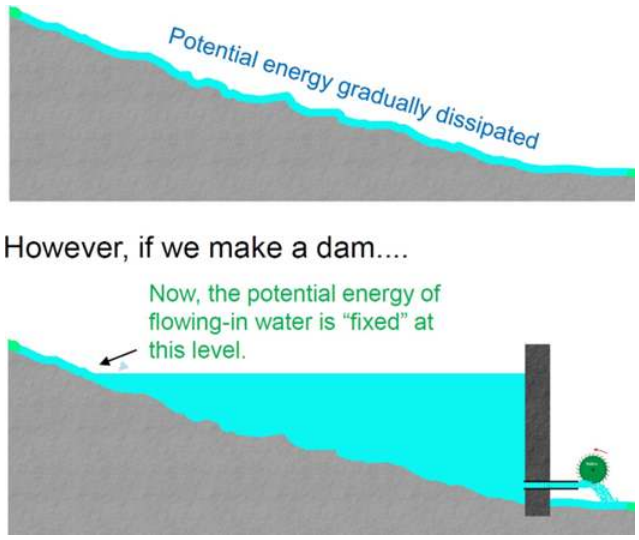


Figure \(\PageIndex{1}\): Making a dam saves the energy of flowing water – otherwise, the energy is gradually dissipated along the river bed.

5.3.1: The Kaplan and Francis Turbines

5.3.2: The Pelton Turbine

5.3: Methods of Harnessing Hydropower is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

5.3.1: The Kaplan and Francis Turbines

The kinetic energy of the “water jet” is harnessed and converted to mechanical work (needed for turning the electricity generator). A widely used turbine type is the Kaplan Turbine, looking very much like a giant propeller. Its efficiency is very high, it can “capture” over 90% of the kinetic energy of the outlet stream. Today, however, another turbine type is taking over – namely, the *Francis Turbine* (Fig. 5.11). It was invented 170 years ago by a Massachusetts engineer James B. Francis. Mechanically, it’s design is more

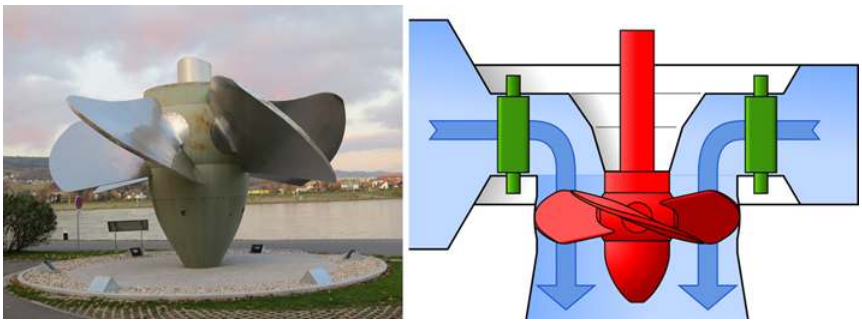


Figure 5.3.1.1: On the left: the rotor of a Kaplan turbine. On the right, there is a scheme showing how the rorors are most often mounted: at the lower end of a verical shaft, the other end of which turns an electric generator (source: Wikipedia Commons).

complicated than that of the Kaplan turbine. Also, its efficiency is slightly lower than the Kaplan’s turbine efficiency. Yet, it has one considerable advantage over the Kaplan type – name, it is a **reversible** machine,

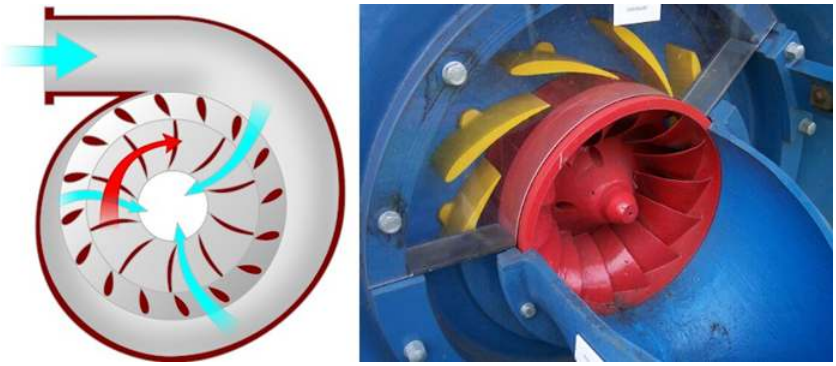


Figure 5.3.1.1: The flow of water in the Francis turbine (left diagram – see the text for more explanation) and, in the right, a photograph of a real turbine with partially exposed elements, the yellow vanes and the red runner (source: Wikimedia Commonas).

Figure 5.10: The flow of water in the Francis turbine (left diagram – see the text for more explanation) and, in the right, a photograph of a real turbine with partially exposed elements, the yellow vanes and the red runner (source: Wikimedia Commonas).

it can work in **two directions**: (i) as a turbine harnessing the energy from down-flowing water, and (ii) as an electric motor-powered **pump**, capable of lifting water from one reservoir to another one, located much higher (even 2000 ft. higher). Such capability makes the Francis turbines especially valuable in *Pumped Storage Hydroelectic Plants* (PSHP) which are becoming more and more popular – we will return to the discussion about the advantages of pumped-storage plants later in this Chapter.

While the operation of the Kaplan turbine seems quite simple and does not require any special explanation (everybody on many occasions have had some experience with propellers – those in toys, in motor boats, in portabe fans, etc., right?) – the operating principle of the Francis turbine may not be immediately clear, if we look only at photographs and diagrams.So it’s a good idea to explain the role of individual components in the device. A pretty good description if given in a Wikipedia article – here is, in short, how it explains the things: as shown in the diagram on the left in Fig. 5.10, water enters a spiral casing (known as the “volute casig”, or the “scroll case”), inside which there is a rotor, usually called a “runner”. Throughout the casing’s length there are numerous openings (known as wicket gates or simply “wickets”) through which water flows inwards and impinge on the blades of the rototr. As can bee seen in the scheme in Fig. 5.10, the cross section area of the spiral casing gradually decreases along its circumference in order to maintain constant speed of water along the casing despite the fact that its volume gradually decreases. In

addition, in each wicket gate there is a “stay vane”, i.e., a profiled blade (the yellow blades in the right photograph in Fig. 5.10) that allows to send the water streams impinging on the rotor at different angles, appropriately to the speed at which it rotates.

However, an even better way of explaining the Francis’ turbine operatio n principle is definitely to use a dynamic video presentation – for instance, one such YouTube piece worth recommending is this one.



In addition to the “reversibility”, another important advantage of the Francis turbines is that they allow to maintain the a constant speed of rotation even when the flow of the input water varies in a wide range. Today, almost all turbines installed in new hydropower plants are of the Francis type, and about 60% of the hydropower generated worldwide comes from such turbines.

It is also worth looking at the animated picture in this website which shows the “block scheme” and the operation of a typical hydropower plant (note that the tube through which water from the reservoir reaches the turbnbe is called “the penstock”). There are also several similar movie clips in YouTube, e.g., this nice short video from New Zealand. In some of the static or animated pictures in the Web (not in the two linked above!).



One thing is shown incorrectly –namely, the penstock’s inlet is positioned close to the reservoir bottom. In fact, it’s typically located close to the water surface. But no matter from where the water is taken, the kinetic energy of the flow reaching the turbine is the same – can you explain why?

5.3.1: The Kaplan and Francis Turbines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

5.3.2: The Pelton Turbine

Sometimes there is much less water in a stream than in a big river – but the stream flows at high altitude, and at some point cascades down a steep mountain side, or even forms a waterfall. It's a good situation for building a power plant of another type. Namely, the water is "sent downhill" through a long penstock. For each 10 meters (33 feet) of vertical drop the pressure in the pipe increases by one Atmosphere[1], and if the overall vertical fall is large, e.g., several hundred meters, then the pressure at the bottom of the pipeline becomes really high, and water squirts from the nozzle at the pipeline's end with a very high speed. The turbine type appropriate for using in such a situation is the Pelton Turbine.

The Pelton Turbine has one considerable advantage for an instructor. Namely, when talking about the Kaplan and the Francis turbine, the need instructor tells you that both have high efficiency, they can "intercept" over 90% of the energy carried by flowing water. But if you start asking indepth questions, e.g., *What specifically makes the efficiency so high?*, then the answer may be either: *It's what the measurements show* – in other words, the instructor honestly admits that he/she doesn't know much more – or, if the instructor **does know**, you may expect an answer: *Well, I can tell you, but it may help if you first read a book about fluid mechanics*. And fluid mechanics, you must know, is a highly sophisticated branch of Newtonian mechanics, loaded with pretty high level calculus.

The Pelton Turbine also has a very high efficiency – even higher that the Kaplan and Francis – and the explanation is pretty simple. No need to delve into the intricacies of Newtonian fluid mechanics! Everything will become clear if we first consider two really simple problems: – so that if a pressure is expressed in the units of atm, one quickly gets the idea of what is being talked about.

Problem 5.3.2.1

A big U-HAUL truck is parked on a street. A boy throws a basketball at the truck's back door with the speed of v, a. How much of the ball's kinetic energy will be transferred to the truck? Assume that the ball-door collision is perfectly elastic.

Solution

if the collision is perfectly elastic, the ball is bounced back with the same speed v, so it's kinetic energy remains unchanged – *ergo*, no energy is transferred to the truck.

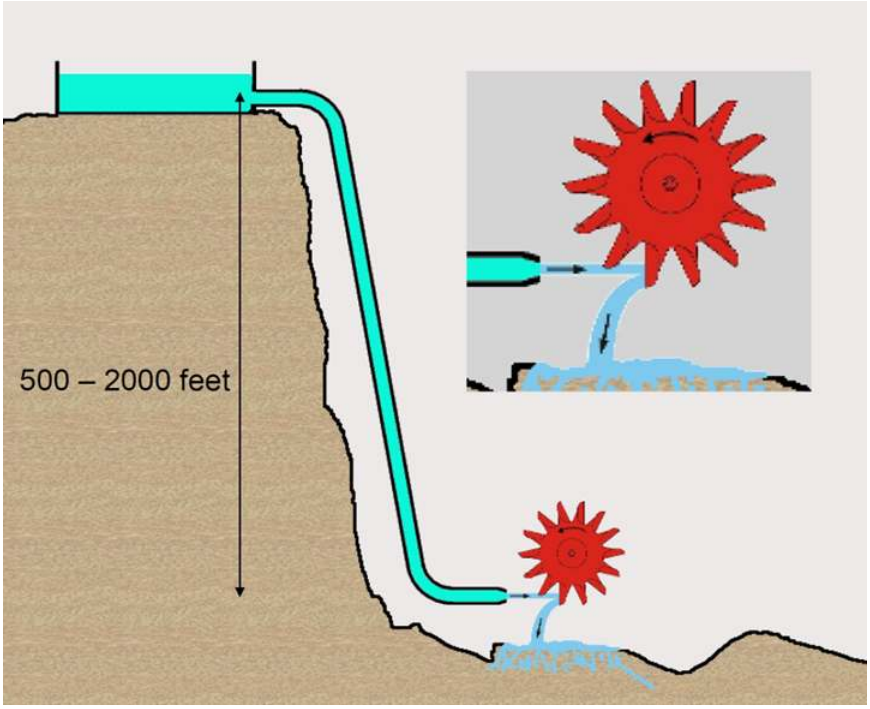


Figure 5.3.2.1: Typical arrangement of a power plant using Pelton turbines: the reservoir is located high above the turbine house. Water is brought down through a pipeline ("penstock"). Due to the specific geography of Norway, such powerplants are popular in this country. Pipelines running down on mountain slopes are a common view in the Norwegian landscape (aod).

Problem 5.3.2.2

Now, assume that the truck started moving, and when the ball hits the door, its speed is exactly $v/2$. How much of the ball's kinetic energy will be transferred now?

Solution

For an observer moving together with the truck, the ball hits the door with a $v - v/2 = v/2$ speed. After the collision, it moves with the speed of $v/2$ relative to the truck, meaning that its speed relative to the street is now zero – *ergo*, it has no kinetic energy, meaning that **all** kinetic energy it had before the collision has been transferred to the truck

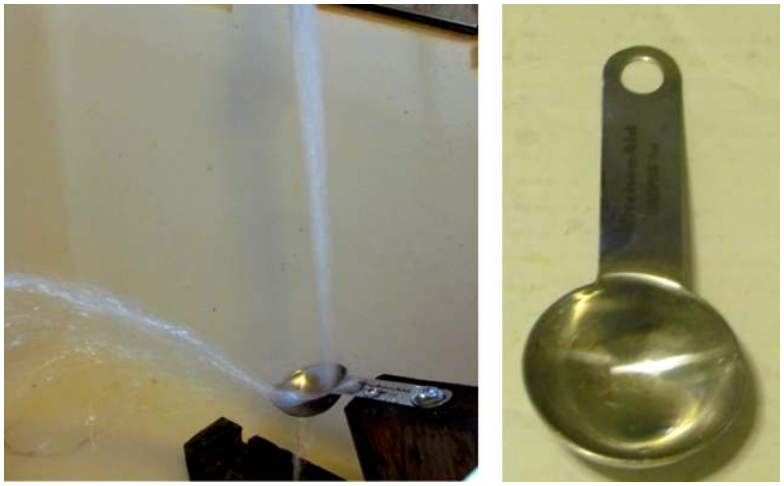


Figure 5.3.2.2: Left side: a measuring spoon acting as a "water jet reverser". The same effect sometimes gives a person washing dishes an unexpected and unpleasant "shower" (aoph)



Figure 5.3.2.3: The Pelton turbine also uses elements that act as “turnarounders” of water streams hitting them. They are called “buckets” and a number of them is mounted at the perimeter of the runner. The water jet hits the “ridge” in the bucket’s center and is split into two streams that are both turned around (source: Wikimedia commons).

Now, let’s think: is it possible to do the same not to a ball, but to a jet of water? Is it possible to find a gizmo that would “backreflect” the jet so that it would retain its kinetic energy? The answer is yes and, most probably, everyone who was ever washing dishes in a kitchen sink did have a “bad experience” with such a gizmo: a spoon, if carelessly placed under a running faucet, can give one an unexpected unpleasant “shower”. In Fig. 5.3.2.2 one can see the results of a “kitchen sink experiment” carried out by the author: a stream of water from the faucet falls on a measuring spoon of spherical shape (a blown-up picture of this spoon is in the right-side picture) and is almost completly reversed and runs upwards.

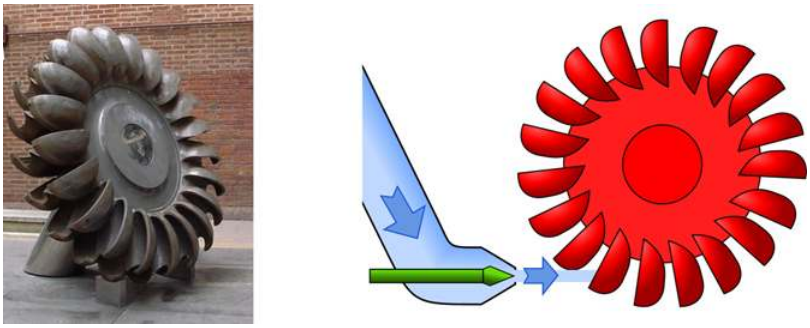


Figure 5.3.2.4: Left side: a complete “runner” with a full set of “buckets”. atteched. The Pelton turbine runners do have a lot of charm, don’t they? Therefore those from decomissioned turbines are often used as decorations in parks or outdor art galleries. The one in the photo is such a decoration in the front of a building in Barcelona, Spain. Right side: a scheme of a Pelton turbine assembly with a single nozzle; usually, more one nozzle is used in order to increase the power output from the runner (source: Wikimedia Commons).

A Pelton turbine uses a number of “buckets” attached to a wheel called a “runner”(Fig. 5.3.2.4, left). In a working assembly (Fig 5.3.2.4, right), the water jet is sent to a bucket center. The “reversal” of the water jet is equivalent to the elastic collision of the ball with the truck door that was discussed earlier. So, if the turbine rotates with such a speed that the bucket moves away from the nozzle with a speed equal to one-half of the water jet speed, then the water flows out of the recessing bucket with zero speed – meaning that all its kinetic energy has been passed to the turbine.

The reasoning that has been used in the problem with the basketball hitting a receding truck remains perfectly valid for the water jet reversal in a Pelton bucket – but we can redo the problem in a more formal way. Let’s use the symbols from Fig. 5.3.2.5: velocities observed in the static reference frame are marked with the subscript “S”, and those observed by a “virtual person” sitting in the bucket are marked with the subscript “B”.

Water is squirting from the nozzle with velocity V_S, but because the bucket is receding from the nozzle with velocity U_S, the virtual observer sitting in the bucket sees a jet flowing in with velocity V_B:

$$V_B = V_S - U_S$$

The jet is reversed inside the bucket, and the virtual observer sees it flowing out with velocity $V_B^t = -V_B$, so that

$$V_B^t = -V_S + U_S$$

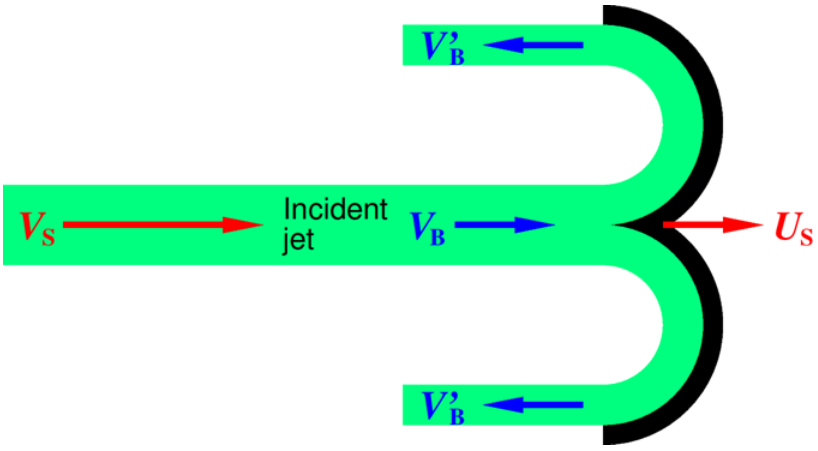


Figure 5.3.2.5: An idealized analysis of a water jet velocity reversal in the Pelton bucket. All "red" velocity vectors are those observed in the static reference frame and the "blue" ones are those seen by a virtual observer moving along with the bucket (aop).

To get the speed V_S^t of the flowing out water seen in the static system, we have now to **add** the recession velocity to V_B^t :

$$V_S^t = V_B^t + U_S = -V_S + 2U_S$$

From the above, one gets immediately that V_S is zero when the bucket recedes with a speed equal to one-half the speed of the water jet squirting out from the nozzle. And $V_S^t = 0$ means that the water flowing out of the bucket carries zero energy –*ergo*, all energy carried by the water jet entering the bucket is "left inside the bucket", i.e., is transferred to the turbine wheel. It implies that with the condition $U_S = 0.5V_S$ the efficiency of the Pelton turbine id 100% – but one has to remember that "nothing is perfect", there are always some deviations from idealized assumption, so in practice the efficiency will be always lower than 100%.

---

[1]An Atmosphere (symbol: atm) is a pressure unit that was widely used in the old days, but now is barely tolerated by the official SI system. But sometimes it is convenient to use it, because the pressure of 1 atm is approximately equal to the pressure of ambient air.

5.3.2: The Pelton Turbine is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

5.4: About One Famous Hydropower Plant Using Pelton Turbines

In Fig. 5.4.1 you can see a photo of at real plant, which is perhaps the most famous of all Pelton-type installations in the world. Its located in Rjukan, Norway, in a narrow valley. During World War II, when Norway got under Nazi occupation, the occupants started using the electric power generated in the plant for extracting the so-called “heavy water” from ordinary water. In heavy water, normal hydrogen is replaced by its heavier isotope, Deuterium. In ordinary water, there is roughly one Deuterium atom per 5500 atoms of normal hydrogen. The process of extracting D_2O from ordinary water, H_2O, consumed much electric power. However, D_2O was crucial for the Nazi efforts of building an atomic bomb. The Allies tried more than once to destroy the plant from the air, but, because of its location in a narrow valley – or, a canyon rather – it was impossible to hit it with bombs. Finally, the plant was twice sabotaged by a British commando teams who parachute-landed in nearby mountain area. This authentic story was used in 1965 for making a movie “The Heroes of Telemark” in which one of the commandos was played by a famous American actor, Kirk Douglas (“Telemark” is the name of a mountain area in Norway, where Rjukan is located).



Figure 5.4.1: Perhaps the most famous hydroelectric power plant in the world using the Pelton turbine technology: in Rjukan, Norway, located in a narrow canyon. Water is brought down from a high altitude by ten parallel penstocks seen in the upper left corner of the photograph (contemporary view; source: Wikimedia Commons).

5.4: About One Famous Hydropower Plant Using Pelton Turbines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

5.5: Pumped Storage Hydroelectric Plants (PSHP)

One great advantage of hydropower technology is that it makes it possible to build plants in which large amount of energy can be stored and used later “on demand”. Such complexes are called “pumped storage plants”. In the area of energy storage, they are definitely the record-keepers. Energy can be stored in other ways, in electric batteries, or thermally in huge reservoirs of molten salts or as compressed air, (the Chapter 11 in this text is devoted specifically to energy storage methods). However, the largest existing hydroelectric storage complex (in the US, in Bath County, Virginia– and here is a 7-minute video) can store about 50 times more energy than the largest currently existing electric battery systems.

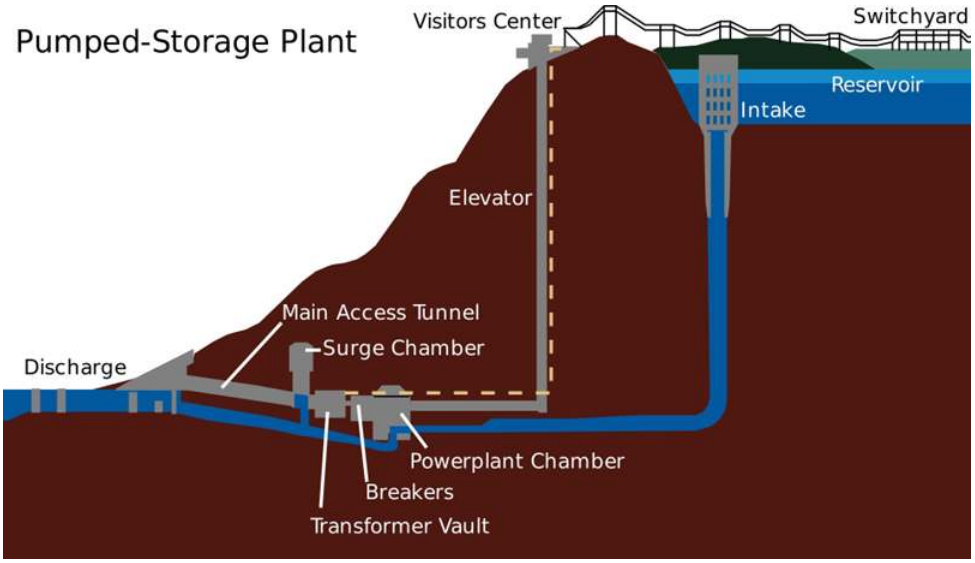


Figure 5.5.1: A general scheme of the Raccoon Mountain Pumped Storage Hydroelectric Plant. It uses dual-action Francis turbines. Details of the turbines and the motors/generators are not shown in the figure, we have to understand that they are all hidden in the unit marked as the “Powerplant Chamber” (source: Wikimedia Commons).

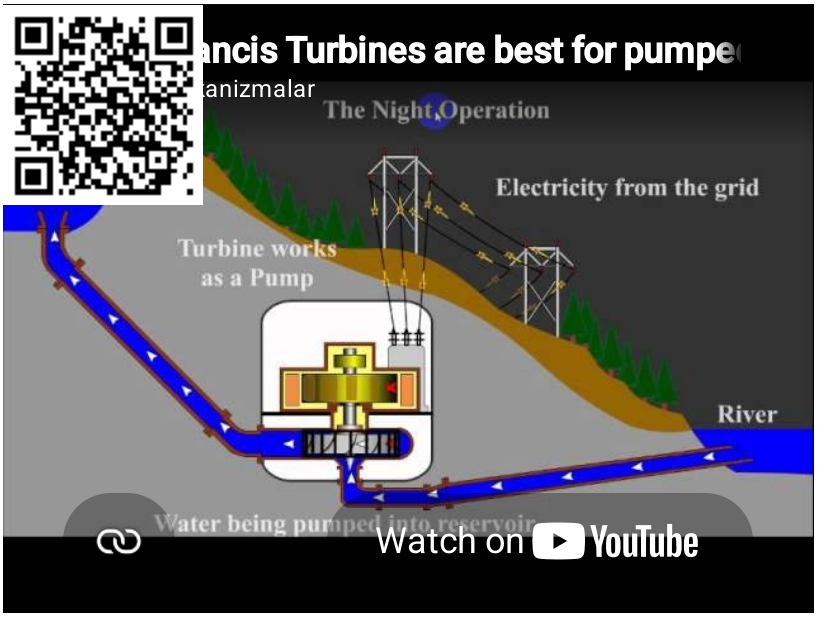


The idea of hydropower storage is very simple one needs two reservoirs, called the “lower” and the “upper”. When there is surplus of electric power (e.g., in the night hours), water is pumped from the lower pool to the upper one – this is the “storage mode”. Then, when the utility system uses maximum power (e.g., during the “peek hours”, the water from the upper pool is sent to turbines this part of the operation, called the “generating mode”, is exactly the same as in an “ordinary” hydropower plant.

There are some problems with this technology, though. If the complex is expected to store a large amount of energy, then the upper pool should be able to contain a huge amount of water, and it should be located considerably higher than the lower pool. The thing is that there are not too many places in the world that offer favorable conditions: first, there must be a mountain with a flat top and enough space in the summit region enabling one to build a sizable “upper pool”. Second, there must be a long steep slope, at the bottom of which the “lower pool” has to be installed. Third, there must be plenty of water available in the area. From the quiz about

the Bath County storage complex, you will certainly conclude that finding other locations good for building complexes of similar magnitude may be a difficult task – and in many geographic areas, just a hopeless task!

There is a nice animation on Youtube showing the operation of a pumped storage system accompanied by an interesting comment.



There are several possible ways of building PSHP installations. One possible variant is to make the pumping unit and the electricity generating unit completely separate. It is how the first PSHPs were built. Yet, a smarter solution is to use the generator as an electric motor. It should be noted that electric power generators usually can work “the other way”, as motors. When the “input” to a generator is work, it converts it to an “output” in the form of electric power; and vice-versa, if the “in-put” is electric current sent to the generator, the “output is mechanical work. Accordingly, the next step was to build installations in which one generator/motor services both the turbine and the pump. And the newest simplification of the system is to use the Francis Turbine which is, as was mentioned earlier, a double-action device that can operate both ways: as a turbine extracting power from downhillflowing water, or as a pump sending water uphill. Essentially, all pumped storage installations built in the recent past use the Francis turbine/pump technology.

If you would like to find a more “in-depth” description of the Francis turbine technology, this article is certainly worth reading – because of excellent graphic material, beautiful photos and instructive technical drawings.

In Fig. 5.5.2 there is a graph showing a typical daily activity of a pumpedstorage plant: Pumping (mostly, during night hours), and Power Generation the power generated may sharply go up and down, depending on how much “backup” the nation’s power system needs at a given moment. Older installations are not capable of reacting so quickly to the demand – in the generation mode they “prefer” to deliver power at a constant rate. However, the ability of quickly changing the output power and to react promptly to changing power demand becomes crucial in times when the utility systems relay more and more renewable power sources. Windpower and solarpower, as we all know, may fluctuate quite rapidly, especially, when weather fronts are passing over a given region. Therefore, the energy storing facilities should be ready to smooth out the “ups and downs” in the power delivered by photovoltaic panels and wind turbines. Indeed, some new PSHPs are able to raise the output from zero to maximum power in about a minute.

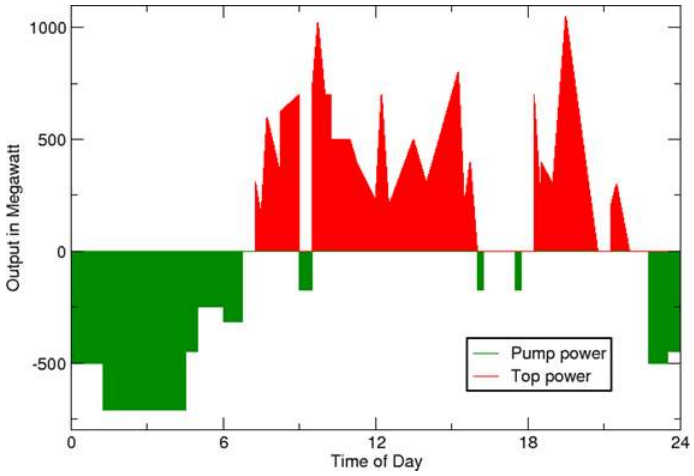


Figure 5.5.2: An example of typical power distribution over a 24 hour period of a PSHP. Green color is the power used for pumping; red is the power delivered back to the grid (source: Wikimedia Commons).

5.5: Pumped Storage Hydroelectric Plants (PSHP) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

5.6: New Ideas for Hydraulic Energy Storage

All existing permanent PSHP installation use freshwater (from rivers or lakes). In principle, seawater could be used as well, with the sea acting as the lower reservoir. In many places in the world there are cliffs in the coastal areas, so only upper reservoirs need to be built. One concern is that salty water is corrosive, but this is a problem modern metallurgy and material science could surely deal with. So far, one prototypical 30 MW seawater installation was tested at the Japanese Okinawa Island. More information about new ideas of building seawater PSHPs can be found in this Wikipedia article.

One fascinating idea, also discussed in the Wiki article quoted, is to build PSHP installations in which a sea or an ocean would act as the **upper reservoir**. But what about the lower reservoir? – for this role, the authors of the idea say, one can employ a huge tank which would be sunk and anchored at the ocean bottom at the depth of several hundred meters, and connected to the shore with a penstock.

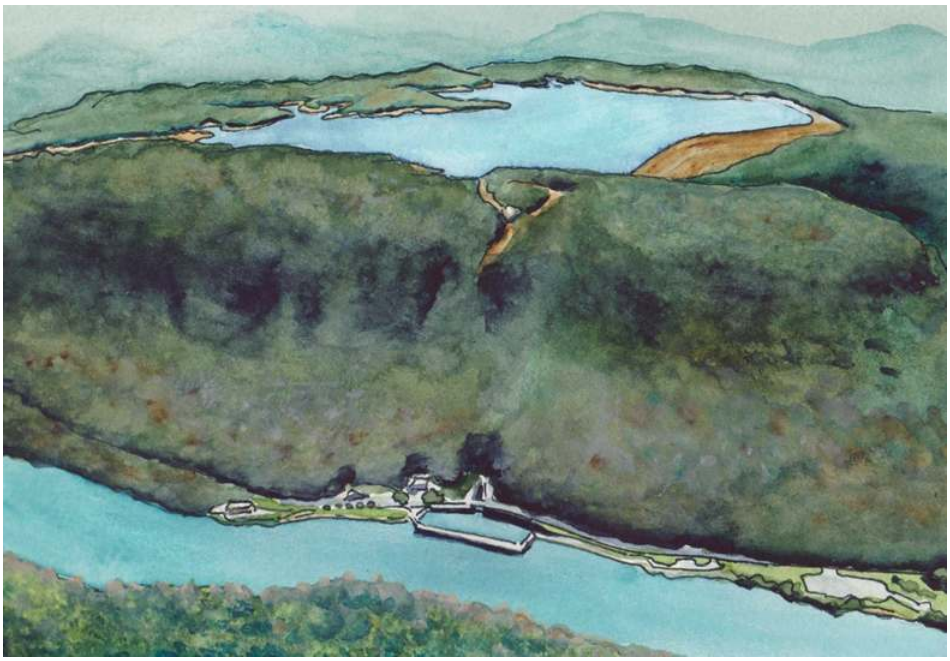


Figure 5.6.1: The Raccoon Mountain PSHP, in this author's opinion, is beautifully composed into the landscape. So beautifully that it inspires artists who like to paint it from different perspectives: here is one such example, an aquarel painted by Ms. Julia Giebultowicz. The facility, owned by the Tennessee Valley Authority, can deliver 1650 MW for 22 hours. More info in this Web page. Some interesting math and discussion related to this PSHP and to energy storage problem in general can be found in this Web site.

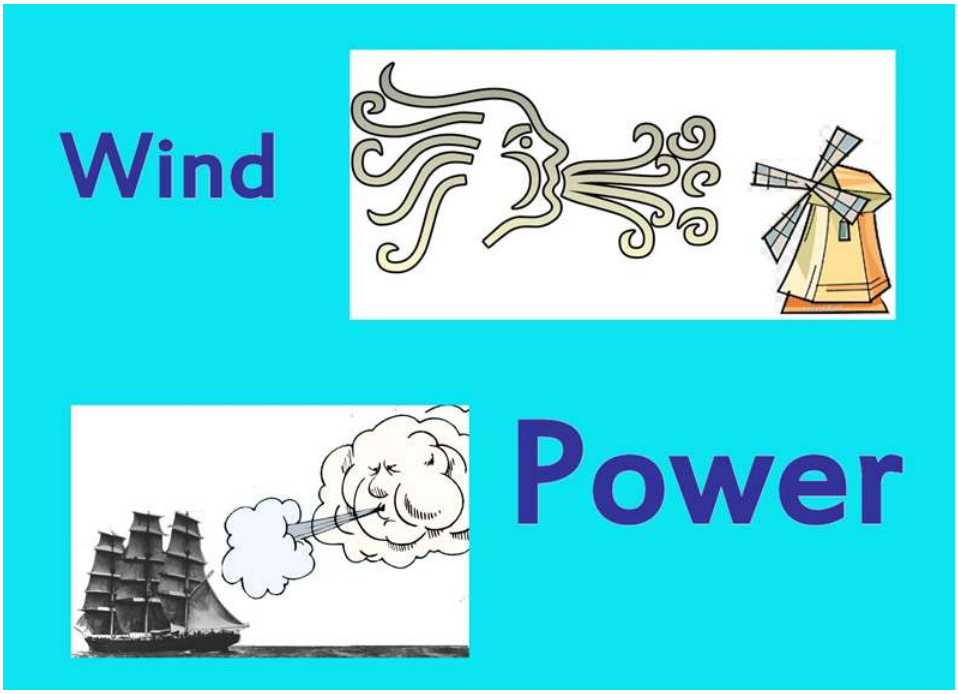
The turbine/pump and the motor/generator assembly, according to the plan, would be located inside the lower reservoir and connected to the shore with a powerline. In the pumping mode, the pump would empty the bottom reservoir; and in the generation mode, water from the ocean surface would flow down through the penstock and drive the turbine before refilling the reservoir.

A reservoir in the form of a sphere 30 meters in diameter has the capacity of 10600 cubic meters. If it were sunk to the depth of 700 meters, the water flowing in from the surface and passing through the turbine would do a work of 75 GigaJoules. If the process takes one hour, this work corresponds to a power of (75 GJ)/(3600 s) = 20.6 MW; or to 5.15 MW of power – enough to satisfy the needs of a community of over 1000 households over that time period. What is needed is an ocean floor that is rapidly descending, so that the lower reservoir can be installed not too far from the coast. It's often the case with islands such as, e.g., Hawaii, Canaries, or New Zealand.

5.6: New Ideas for Hydraulic Energy Storage is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

6: Windpower



Wind power is nothing else than converted solar power (as is, e.g., hydropower, windpower, wavepower, or the power generated by burning biofuels).

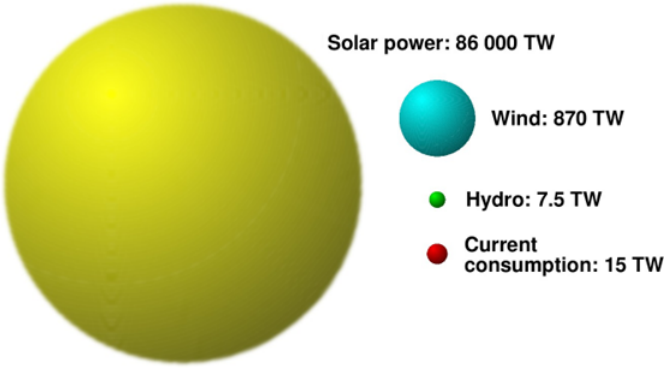


Figure 6.1: The total power of all winds blowing over the globe is slightly more than 1% of all solar power reaching the Earth surface – but is over 50 times larger than the total global consumption of power.

In general, wind is caused by the fact that different areas on the Earth surface get heated up at different rate by sunshine (or the rate of cooling after the sunset is different).

One well-known example a winds which are always blowing over the oceans in the same direction are the *trade winds* (a.k.a. *passats*) One is blowing between the Equator and the 30th parallel north[1] from the North- East and is called the *Northeasterly Trade Wind*. The other is blowing and between the Equator and the 30th parallel from the South-East and is called the *Southeasterly Trade Wind*.

Both these trade winds are caused by hot air rising in the vicinity of Equator. The air “escaping up” leaves behind a “void”, i.e., an area of low pressure. When reaching the troposphere, the rising-up stream of air splits into two, one moving north, and the other moving south. At high altitude, the air cools down, gets heavier, and eventually falls back towards the the Earth surface in the

vicinity of the 30th parallel north and the 30th parallel south, causing an increase of the air pressure in both these areas. So, the air starts moving from them towards the area of lower pressure, i.e., towards the Equator.

The above description may suggest that both winds blow directly towards the Equator from the North ans from the South. And, as was stated earlier, actually they blow from the North-East and from the South-East. How come? Well, there is yet another factor involved, called the *Coriolis Force*.

The Earth, as we all know, is a rotating body. The forces acting on objects in rotating systems are described by Newtonian Mechanics. There are two such forces – the *centrifugal force* acting on objects which are **fixed** in the rotating system (we all know from this force from rides in a merry-go-round), and the Coriolis force which acts on objects moving in the rotating system. The winds are moving objects, so it is the Coriolis force they experience. The general theory of the Coriolis force is a bit complicated – but it is not necessary to study it in detail, a simplified explanation is all we need.

First, consider the speed of a fixed point at the Equator. Its circumference is about 40 000 km. The Earth makes one revolution in 24 hours, or in 24 3600 s = 86400 s. So, the speed of a point at the Equator is (40000 km)/(86400 s) = 463 m/s. The Earth is rotating from the West to the East, so it is 463 m/s in the eastern direction.

Now, consider a point at the 30th parallel. This parallel is a circle of circumference equal to (40000 km) cos 30° = 34641 km. By dividing it by 86400 s, we get the speed of 401 m/s. The air starting its trip towards the Equator moves then with a speed of 401 m/s eastbound – but the point at the Equator is moving faster by 62 m/s in the same direction. So, the wind is "lagging behind" and when it finally reaches the Equator, it is at a point that are pretty more to the West than the point from which it started its trek. Consequently, an observer at the Equator will see a wind coming from the North-East, not straight from the North –or coming from the South-East, not directly from the South.

But this is not the end. Not all air descending from above near the 30th parallel would attempt to travel towards the Equator – some will floe in the opposite direction, to the North or to the South. The Coriolis Force will then push the wind in westbound. So, above and under thea areas where the North-East and South-East trade wind blow, there are winds blowing from the South-West and North-West. Such winds were called *westerlies* by anglo-saxon sailors, and *anti-passates* by German and Dutch sailors.

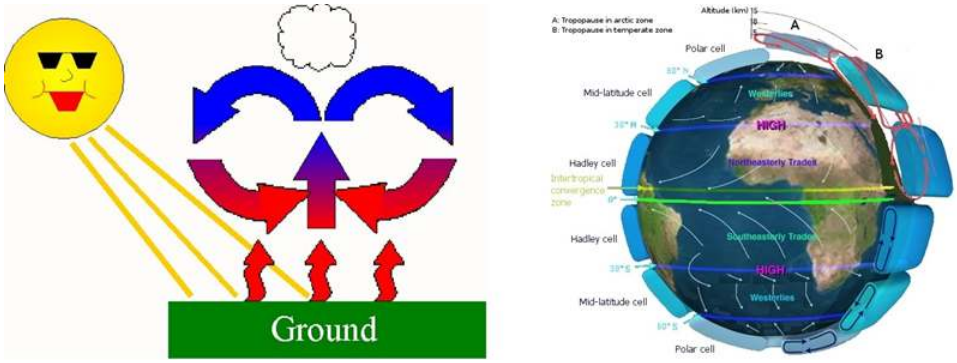


Figure 6.2: : Left side: the mechanism creating the tradewinds. Hot air raises at the Equator zone, leaving behind an area of low pressure. The air cools down and descends close to the 30th parallel, building up the pressure; then the air flows towards the area of low pressure. Right side: the effect of the Coriolis force on the air returning to the Equator, giving rise to the Trade Winds (a.k.a. "Easterlies") and on the "Westerlies"(sources: left – free clipart; right – NASA educational resources

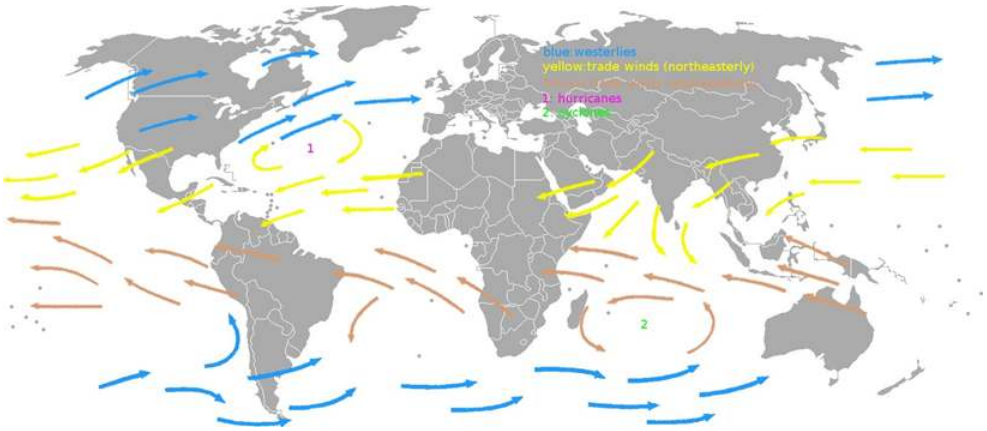


Figure 6.3:The distribution of the tradewinds (yellow & brown arrows) and the westerlies (blue arrows) over the globe – note that they blow above the waters, but do not reach far inland (source: NOAA).

1The 30th parallel north is a circle of latitude that is 30 degrees north of the Earth's equatorial plane.

6.1: Applications- Winds Over the Seas
6.2: Winds Blowing over the Lands
6.3: Applications- Land and Off-shore Use
6.4: The Physics of a Wind Turbine
6.4.1: Problems with Wind Power

6: Windpower is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

6.1: Applications- Winds Over the Seas

One of the first practical applications of wind power by humans was to use it as a “propellants” for sailing boats. In the Mediterranean area, extensive trade already existed one thousand years before the beginning of the modern era. Many nations and tribes were scattered on numerous islands and peninsulas in that area. Transporting goods by boats was a convenient (and sometimes the only possible) way of delivering them to customers. The traders quickly learned to take advantage of winds – if they were blowing in the right direction; otherwise, the merchants had to use oars (driven by slaves), or, if their boats were not equipped with such “propulsion system”, they had simply to for a wind blowing in a favorable direction.

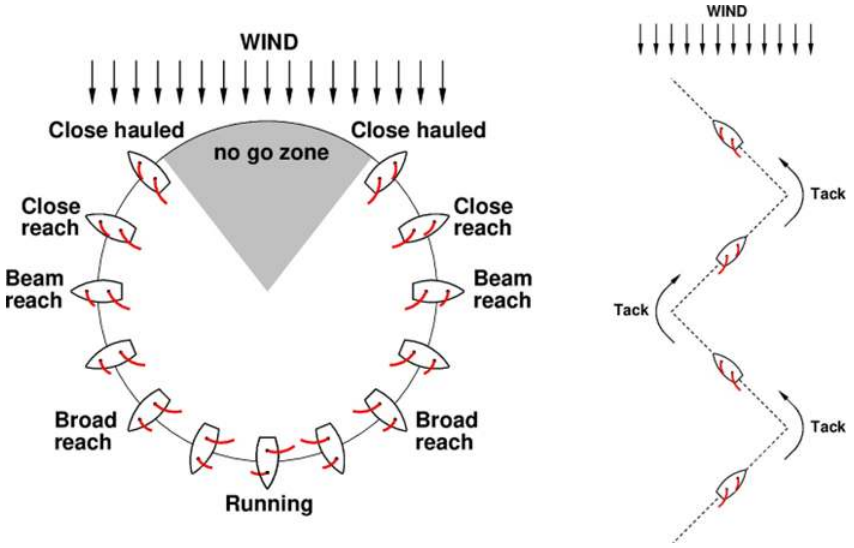


Figure 6.1.1:: Left: possible sailing directions relative to the wind. Only the directions closer than about 45◦ to up-wind direction are excluded. Until the mid-16th century Europeans used sails that allowed only sailing downwind and in the “beam reach” direction. The art of sailing “upwind”, as close as 45◦ relative to the wind was known to Arabs long before the Europeans learned it by observing Arab boats at the Mediterranean. Right: the tacking maneuver, by zigzagging the boat can effectively sail upwind, or “beat to windward” according to the old nautical terminology.

Over a period of two thousand years, the sails did not change very much – they were just big fabric rectangles. Seamen call such sailing *square-rigged* . The upper edge of the sail is carried by a spar, i.e., a horizontal bar with its center attached to the mast. A boat with such a sail can only travel downwind. Going far into the Atlantic in such sailboats was too risky – how would one return home, if there was no guarantee that there would be a wind blowing in the right direction? Even the Vikings, the Scandinavian warrior tribes who were perhaps the best sailors of all times, in the X and XI century had enough courage to sail thousands of miles away from their bases only because they always carried oars to be used if the wind was blowing in their faces.

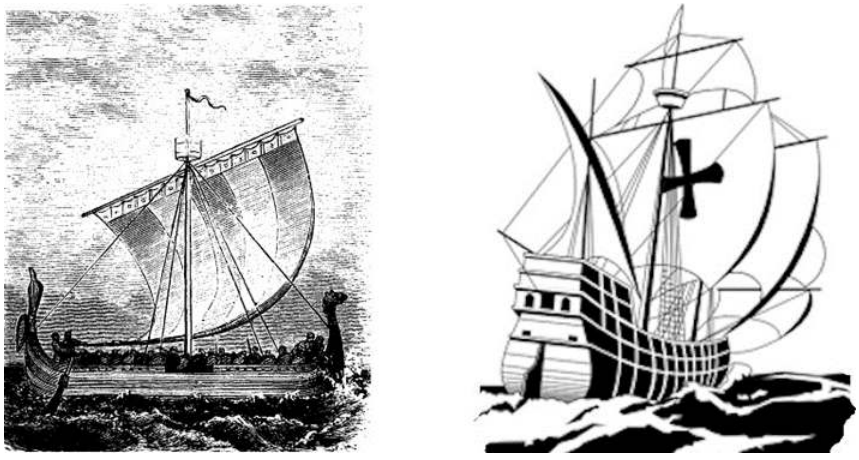


Figure 6.1.2: Left: a Viking boat with the oars hidden. Right: a galeon with several square-rigged sails, but still not able to sail up-wind.

In Portugal and in Spain, where the shipbuilding technology was pretty well developed, in the XIV century the shipyards were able to build large galleons with the capacity up to 200 tons, but still square rigged, so that they were not yet worthy ocean-going crafts.

A real technological revolution took place only in the XV century. Technically, what happened, was the beginning of a widespread use of the so-called *lateen sail* or *latin-rig* which is a **triangular** sail with one side carried by a *lateen*, i.e., a long rod mounted at an angle on the mast.

It's not exactly known who and when had invented the lateen sail, but it is clear that such a rig was used long before the XV century in the Middle East and at the Indian Ocean. Then it was used by Arabs and the Moors at the Mediterranean Sea at their characteristic boats they called dhows and feluccas. The Western Europeans did not pay much attention to them – until a Portuguese prince Henry the Navigator realized that they had the ability the square-rigged ships lacked – namely, to "beat to windward", i.e., to go closer than 90° to the up-wind direction. Under Henrys sponsorship, the Portuguese shipyards developed a highly successful new vessel type, the Caravel.

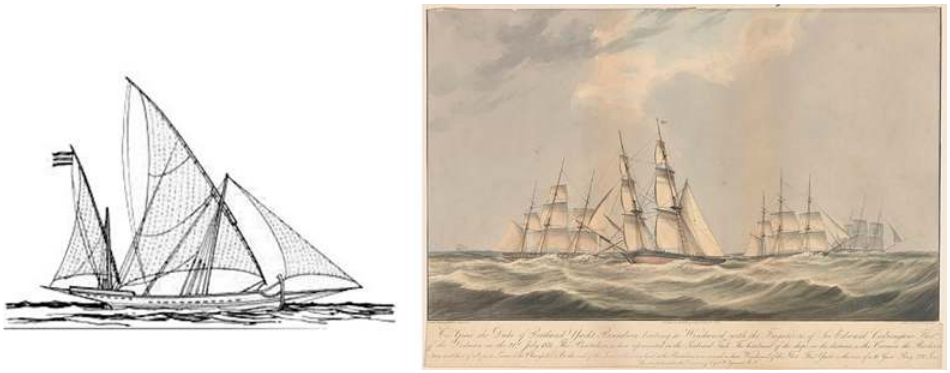


Figure 6.1.3: Left: a felucca, a boat sailed by Arabs, capable of "beating to windward" due to its lateen-rigged triangular sails. As far as beating to windward is concerned, the caption below the picture on the right (hard to read) states: "His Grace the Duke of Portland Yacht Pantaloon beating to Windward with the Frigates &c. of Sir Edward Codrington's Fleet off the Dodman on the 21st July 1831". Portland? An Oregonian accent? No, apparently no relation to our Portland.



Figure 6.1.4: : Back side of the *Monument of Discoverers* in Lisbon, Portugal. The "discoveries" made by Portuguese navigators who had reached and had taken "into the possession of Portuguese Crown" lands located thousands of miles from their homeland brought an enormous prosperity to that small European nation. And it's clear who paid for that wealth. Perhaps, it's the reason why the name of the monument emphasizes that those brave sailors were "discoverers", but it omits the fact that they were often brutal colonizers.



Figure 6.1.5:Front side of the monument, with the statue of Prince Henry the Navigator, holding a model of a caravel – note the two lateen sails it carries (both pictures taken by the Author).

One of the first task of the caravel captains was to explore the western coast of Africa. As can be seen in Fig. 3, after passing the "Bulge of Africa" – the westernmost part of the continent – the ships encountered a steady south- eastern headwind. For the square-rigged boats it was not possible to proceed any further along the coast. The caravels changed everything, by the end of the XV century the Portuguese reached the southern tip of Africa and proceeded further, eventually reaching India. Christopher Columbus also used caravels in his historic trip in 1492.

New "discoveries" of more and more remote lands followed. The caravels were modified, favors were returned to square-rigged sails because with the wind blowing from the back the square-rigged boats could run even faster than lateen-rigged ones. Therefore, new designs carried both types of rigging. Those types of vessels could reach any region on the globe. New trade routes were opened, regular communication between all continents was established. Until mid-XIX century sailing boats "ruled the waves". In the mid-1800s incredibly fast sailing vessels, called "clippers", or "tall ships", carrying a pyramid of sails, maintained a regular communication across Atlantic and Pacific and between Great Britain and Australia (see the image of a "clipper" at the front page of this Chapter). However, in the mid-1800-s a new type of propulsion for boats emerged – namely, the steam engine – and was quickly taking over. By the end of XIX century the era of big sailing boats was over, it was steam that ruled the waves since then. And then, steam engines had to give way to giant Diesel motors. It's them who rule the waves today.

Anyway, today there are probably more sailing boats than ever before. But not the big ones with pyramids of sails. Several tens of "tall ships" still exist over the world, maintained by merchant fleets, or navies, for training young people to become hardened seamen. Those "tall ships" converge every 8 or 12 years at galas called *Operation Sail* or *OpSail* . But the "center of gravity" has moved to small sailing boats, used mostly for tourism, recreational sailing or racing. Such boats carry only triangular sails, the so-called "Bermuda rig" (with time, the lateen rig evolved to gaff rig, and eventually to Bermuda rig).

But who knows? One day the tall ships may come back to the oceans. The wind is a "fuel" that costs nothing and produces no CO_2 emission at all. The progress in material science has made it possible to produce new ultralight but sturdy masts and sails. In the old days, the tall ships had to carry crews up to several tens of sailors because all operations on sails were manual and required climbing the masts. For example, the famous clipper Cutty Sark with displacement of 2100 tons needed a crew of more than 30 men. It is expected that a new-generation merchant ship with ten times or even larger displacement and with fully automated operation of the sails, will need less than ten crew members.

6.1: Applications- Winds Over the Seas is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

6.2: Winds Blowing over the Lands

The patterns of winds blowing over the lands are much more complicated than those of winds over the oceans and seas. The speed and the direction of winds depend on many more factors. In Fig.6.2.1 there is a map of the average annual speed of wind over the continental United States at the height of 80 meters over the ground level, from the Web page of the US Department of Energy. Maps for other heights can be found on the same Web page.

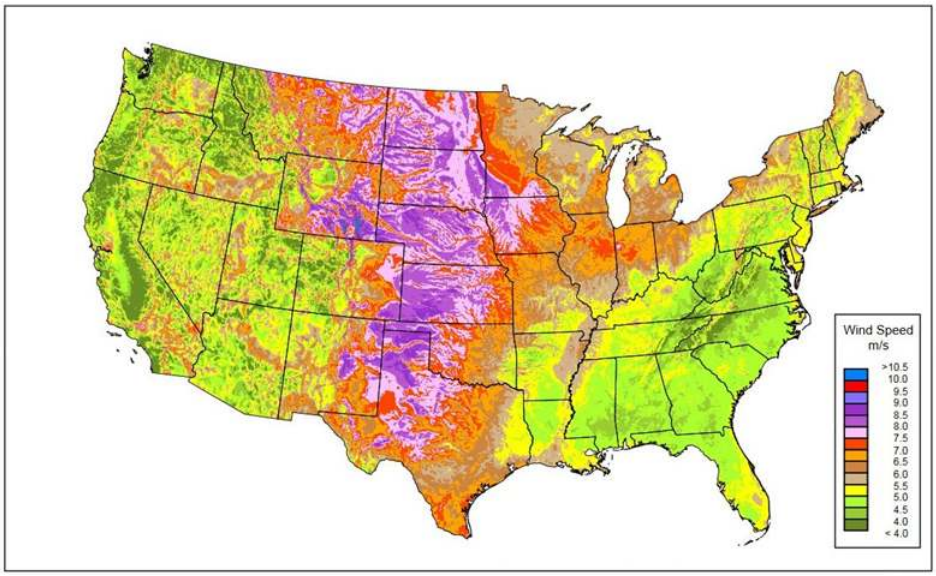


Figure 6.2.1: The average annual speed of wind at the height of 80 m over the ground lever for the continental United States (source: Wikipedia).

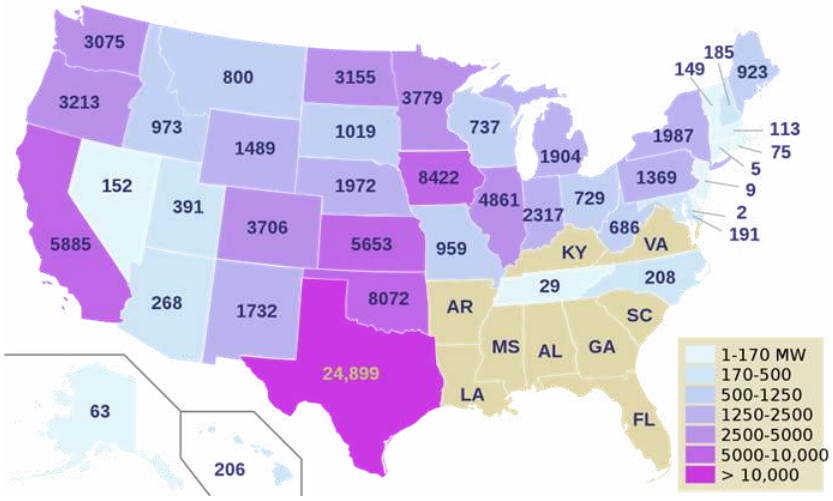


Figure 6.2.2: US map showing the total wind power installed in each State (source: Wikipedia).

But the direction and the speed of wind in a given location may change even many times over the same day and a big-scale map like the one in Fig. 6.2.1 cannot include such characteristics. In a region with many hills and valleys – and especially in a mountain region – much more detailed characteristics of the local wind patters are needed, if one plans to build wind turbines in a given area.

6.2: Winds Blowing over the Lands is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

6.3: Applications- Land and Off-shore Use

It is not known exactly when people started using the power of wind to do work on land – work which previously needed engaging human or ani- mal muscles. Probably it was not as early as harnessing wind for propelling riverboats. The latter requires using a big piece of cloth and a few poles and pieces of rope. Performing land works such as grain grinding or water pumping – which were the first task wind was used for – already required building a **machinery**. There are indications that Hammurabi, the king of Mesopotamia, tried to use windpower for irrigation purposes as early as 1800 B.C.. But harnessing wind power at a larger scale has began only in the Modern Era (see the document *History of wind power development* . The first known wind machines were **vertical-axis** windmills built in Persia (today's Iran) and Afghanistan. It seems that the art of building windmill was brought to Europe from Middle East only towards the end of the Medieval Age. The Persian-type vertical axis windmills were soon replaced by more efficient horizontal-axis models, the precursors of modern wind turbines. The progress in windmill technology was the fastest in the Netherlands where such machines were crucial for pumping water from fields which otherwise would be flooded by seawater. In other countries the principal task of windmills was rather grain grinding. Windmills competed with mills using waterwheels and were certainly victorious in areas where creeks were scarce but there was much wind. Windmills, for instance, were popular in Spain – but not only for grain grinding, as is described in the literary masterpiece *Don Quixote* novel (see the article in Encyclopedia Britannica) written by Miguel de Cervantes in 1605.



Figure 6.3.1:A typical Dutch water-pumping windmill. There are still about 1000 such beautiful wind-machines in Holland today (source: Wikimedia Commons).

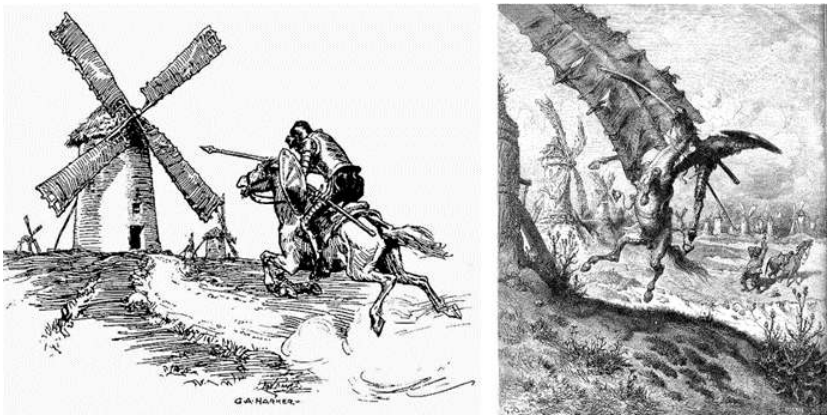


Figure 6.3.2: Left: Don Quixote charging at a windmill. He thought that the windmill was an evil giant waving his arms (by G. A. Harker, a contemporary artist-illustrator) *Tilting at windmills* or *jousting windmills* is an idiomatic expression in most European languages which means attacking imaginary enemies. Right: this action did not end well for the unfortunate Don Quixote, as seen in the illustration bu Gustave Dore, a famous 19th-century French artist (source: Wikimedia Commons, both figures).

A highly successful windmill type was the *American multi- blade wind turbine*. The manufacturing of early models with wooden blades started around 1850. No long afterwards they were replaced by steel profiled blades. The main application of the turbines was water pumping. The turbines were manufactured for over one hundred years. About six million of them were installed all over the United States, mostly at farms. A large American turbine-type rotor was used in 1888 in a pioneer installation for generating electricity built in Cleveland, Ohio by Charles F. Brush (see Fig. 6.13). It generated 12 kW of electric power and worked for the next twenty years.

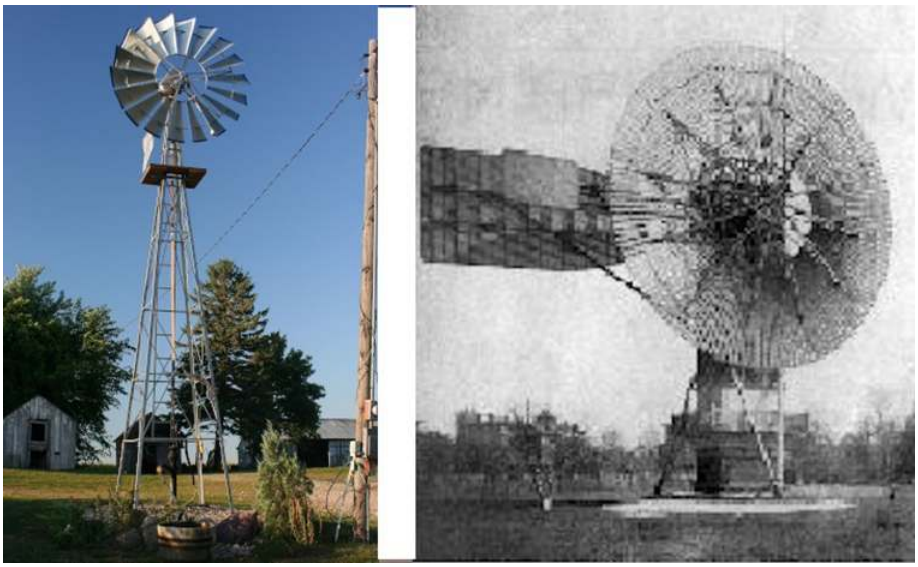


Figure 6.3.3:Left: a multiblade American water-pumping wind turbine at a farm, with a water tank next to it. Right: the *Brush post*, the first electricity generating windmill built by C. F. Brush in 1888 (picture from a public downloading site.

However, in general towards the end of the XIX century the use of wind- power started dwindling because of the emergence of steam engine and later, electric motors and Diesel engines. Research on possible new applications of windpower did continue, but not at a large scale.

A renaissance of interest in windpower was triggered by the Oil Crisis of October 1973, when the oil prices surged overnight from $3 to $12 for a barrel due to embargo for selling petroleum products to the US imposed by OPEC countries. From today's perspective $12 may seem as a bargain, but in 1973 it was a tremendous shock. All of a sudden methods of generating power **not involving** combustive materials started attracting considerable attention. It coincided with environmental concerns that scientists had started raising about a decade earlier. At the time of the Oil Crisis scientists' concerns were already shared with a broader public opinion. The atmosphere was therefore conducive for launching new research programs on "clean energy". In particular, research sites where scientist and engineers developed and tested new types of wind turbines capable of generating electric power mushroomed all over the US and Western Europe. The goal was to find designs that would combine several advantages, such as: high efficiency and high output, sturdiness warrantying long periods of reliable action with minimum maintenance servicing, easiness of mass production and of setting up "wind farms" grouping large numbers of individual units.

There were many competing ideas. For example, some engineers opted for turbines with vertical axes – like in the old Persian windmills – but with highly sophisticated rotors. Such machines, it was argued, offer two considerable advantages: first, the electric generator may be then placet on the ground level. It makes servicing it much easier – and the column, carrying only the rotor, may be much lighter. Second, a vertical-axis rotor works well no matter from where the wind blows, there is no need for adjusting anything if the wind direction suddenly changes.

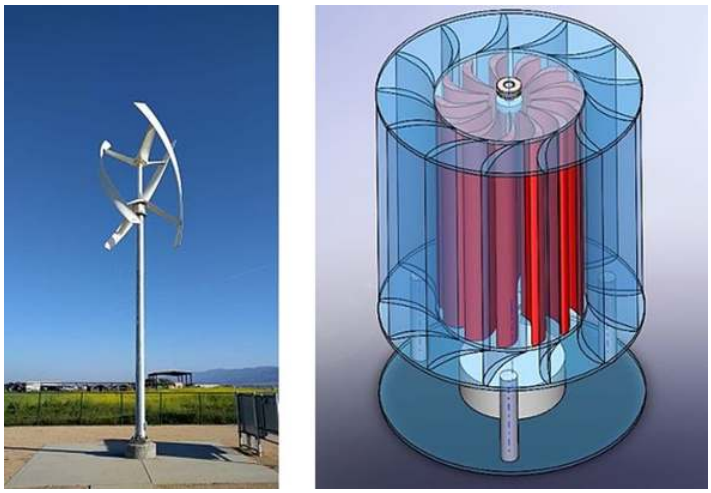


Figure 6.3.4: Left: a vertical axis Darrieus turbine. Right: the rotor of a vertical axis Vortexis turbine.(source: Wikipedia, both figures).

The alternative is a turbine with a horizontal axis, reminiscent of the old Dutch windmills. Here, there is a dilemma: the most straightforward option is to couple the electric generator to the same shaft to which the rotor is attached – meaning that the tower should carry both the rotor's and the generator's weight. Or, one can "transmit" the torque from the rotor to the ground level by a long shaft running inside the column by applying a 90° gear at the top of the tower (which, in fact, is the case in the Dutch windmills). Another disadvantage of the "Dutch style" is that the turbine rotor must always face the incoming wind – which requires installing an extra wind direction sensor, and a mechanism that would turn the entire "tower head" if the wind direction changes.

It should be noted that analogous situations happened many times before in the history of technology. Take, for example, the automobiles. At 1910 there were three types of cars available in the US. The largest group were electric cars. There were also gasoline cars, and there were steam cars (yes![1]). But finally, after about 1925, the gasoline cars were the winner, and the other two types simply disappeared. What decided about the gasoline engine victory? Well, a whole bunch of entangled factors – not only of technical nature, "marker forces" also played a significant role.

It is difficult to say what exactly decided about the winner in the competition between the many wind turbine designs. There were too many quarrels, failed tests, successful tests, bankruptcies, etc. – too many even for trying to present a very brief account of what happened during the two tumultuous decades in the windpower R&D and marketing sectors after the Oil Crisis of 1973 – but a good review of important facts is given in Parts 3 to 5 of the Illustrated History of Wind Power Development.

In short: it is clear who is the winner. Take a ride over rural regions of the United States where there are many windfarms. Or a trip over rural Germany (over there, one doesn't need to look for a windfarm – it is difficult to find a spot in that country from which it is **not possible** to see any wind turbine). Everywhere you go, you see the same standard model: a horizontal- axis three-blade turbine with the electric generator positioned right behind it, in a "nacelle" situated at the top of a tower. The turbine should always face the wind, i.e., the turbine axis should always be parallel to the wind velocity vector. The misalignment between these two directions is called the *yaw angle* – for ideal turbine work, this angle should be zero. If the wind direction varies, the nacelle can be rotated in a controlled manner in order to always keep the yaw angle possibly close to zero.

In YouTube, one can find good videos explaining how different elements and mechanism work:

- How the blades are profiled and how the turbine "pitch" is controlled:
- How the wind direction and speed are determined by the on-board electronics, and how the nacelle yaw and the pitch of the blades are varied by small electric motors with appropriate gears depending on the wind direction and speed;
- How the electric generator is coupled with the turbine.



Figure 6.3.5:Before the electricity-generating wind turbines have emerged, the most powerful wind machine ever built was probably Cutty Sark (left picture), the legendary “wool and tea” clipper launched in 1846, one of the fastest sailing ships of all times. It was estimated that she harnessed 10,000 horsepower of wind, equivalent to about 7.4 MW. The first wind turbines of comparable power were built only around 2010. But now (2020) the record-keepers (mostly offshore turbines, as the one shown in the right photo, are capable of delivering about 15 MW – and this may not be the end (sources: Wikimedia Commons ad Wikipedia, respectively).

The rotor of large “wind power converters” – as the manufacturers prefer to call their products – do not turn very fast: typically, 15-20 revolutions per minute (rpm). Most often, the electric generators used have to turn much faster, 1500-2000 rpm. In such designs the rotor and the generator are connected via a gear that increases as much as needed the “output” speed of rotation. The “anatomy” of such a model is shown in this YouTube clip.



Another possible solution is to use a generator turning with the same speed as the rotor – it’s called a “direct drive”. The generators used in such arrangements, called *annular* (meaning “ring-shaped”) are of large diameters. A process of manufacturing a generator of such type is shown (starting 7 min. 45 s from the beginning) in this video made by Enercon, a major manufacturer of high-power wind energy converters.



The most powerful wind power converter – as of the end of 2017 – is the 9.5 MW offshore windmill shown in Fig. 6.3.5. The list of 10 most powerful wind machines, periodically updated, is given by the Windpower Monthly magazine. The largest on-land wind machine is the 7.5 MW Enercon E-126.

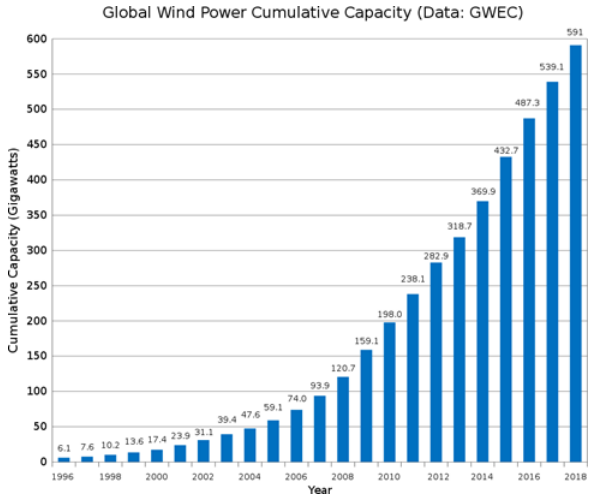


Figure 6.3.1: Cumulative installed windpower capacity since 2001 in MegaWatts (source: Wikipedia).

The *Windpower Monthly* magazine provides a list of top ten wind turbine makers in the world. The list of all current manufacturers worldwide can be found at the Web page of *The Windpower - Wind Energy Market Inteligence*[2]. By clicking on the manufacturers name, one can get for free the list of wind turbine models ever made by this company. There is much more interesting information i n the latter Web site, but – unfortunately – one has to pay for it. The total installed windpower nametag capacity[3] worldwide approaches 500 GW (as for the beginning of 2018). As can be seen in Fig. 6.13, the growth is exponential-like, with the total capacity doubling over a period of about 5 years. One can expect that the 1000 GW level will be reached about the year 2023 (for comparison, the total installed nametag electricity generation capacity in the United States, according to the U.S. Energy Information Administration, was 1177 GW in 2016, and the total windpower nametag generation capacity was in the same year was 82 GW).

---

1. The famous *Stanley Steamer* vehicles did not look like a railway steam engine: no smokestack, no towed coal-carrying tender. They were elegant cars, using gasoline as the fuel, looking even better than other car types of the same era.

2. Not the same organization as the *Windpower Monthly*.

3. Nametag capacity – a.k.a. the *nominal capacity, rated capacity, installed capacity, maximum effect*, is the maximum possible sustained output from a facility, e.g., from a power generating plant.

6.3: Applications- Land and Off-shore Use is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

6.4: The Physics of a Wind Turbine

Let's consider the question: how much energy does wind carry? It turns out that finding the answer is a pretty straightforward task. Suppose that the wind blows with a speed of V . Now, let's put an "imaginary tube" with cross section of A parallel to the wind's velocity direction. Let δt be an arbitrarily chosen time period. Over Δt the air particles the wind carries travel the distance of \(V *times* \delta t \), right? So let the length of our tube be \(L = V *times* \delta t \)

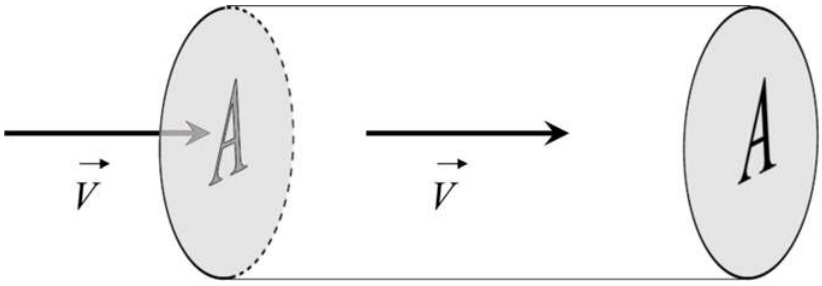


Figure 6.4.1: An "imaginary pipe" the wind of velocity V blows through. A is the cross section area of the tube, and its length\(L = V *times* \delta t \), where δt is an arbitrarily chosen time period.

$$volume = A \times L = A \times V \times \delta t \tag{6.4.1}$$

What we need is the mass *m* of the air inside the tube. The general relation between the mass and the volume is: *mass* = *volume* × *density*. Denoting the density of the air as ρ, we get then:

$$m = A \times L = A \times V \times \Delta t \times \rho \tag{6.4.2}$$

The next step is to find the kinetic energy K of the air portion inside the tube. All we need is to use the known formula: $K = mV^2/2$. Combining it with the expression for m we have gotten, we obtain:

$$K = \frac{A \times \rho \times V^3}{2} \Delta t \tag{6.4.3}$$

But it is not exactly the energy which is of interest to us - we want to find the power the wind carries. The power, as we know, is:

$$P = \frac{\text{the amount of energy delivered}}{\text{the time it took to deliver it}} \tag{6.4.4}$$

We have found the energy K which the "portion" of the air inside the tube carries. How much time it takes it to leave the pipe through its outlet? The length of the pipe is L, and the air inside travels with speed V, so thetime the "portion" in question needs to get completely out through the outlet is:

$$\frac{L}{V} = \frac{V \times \Delta t}{V} = \Delta t \tag{6.4.5}$$

So, energy K passes through the tube outlet in time Δt. Using then K from the Eq. 6.1, we the "almost-final" result:

$$P = \frac{K}{\Delta t} = \frac{A \times \rho \times V^3 \times \Delta t}{2 \Delta t} = \frac{A \times \rho \times V^3}{2} \tag{6.4.6}$$

Let's check if the unit of P is correct. There is a convention: [X] = the unit of X. Therefore, $[A] = \mathrm{m}^2; \rho] = \mathrm{kg/m}^3$; and $[V] = \mathrm{m/s}$. Hence:

$$[P] = \mathrm{m}^2 * \ \mathrm{kg/m}^3 * \ \mathrm{m}^3/\mathrm{s}^3 = \mathrm{kg} * \ \mathrm{m}^2/\mathrm{s}^3 \tag{6.4.7}$$

which is indeed the unit of power, the Watt, right? ($1\ \mathrm{W} = 1\ \mathrm{J/s} = 1\ \mathrm{N}^*\ \mathrm{ms} = 1\ (\ \mathrm{kg} * \ \mathrm{m/s}^2)^*\ \mathrm{m/s} = 1\ \mathrm{kg}^*\ \mathrm{m}^2/\mathrm{s}^3$ – right!)

The Eq. 6.2 is already a useful formula - if we know how big is the area A to which the wind "delivers" its power. For example, is the rotor of a wind turbine is R, then the area in question is $A = \pi R^2$. Sometimes, however, we want to know only how much power the wind carries per a unit surface area - denote it as p. The formula gets even simpler:

$$p = \frac{P}{A} = \frac{A \times \rho \times V^3}{2A} = \frac{\rho \times V^3}{2} \tag{6.4.8}$$

The air density ρ is a function of temperature and altitude. At sea level and at temperature 15° (59° F) air has a density of approximately 1.225 kg/m^3. The recipe for calculating ρ at other circumstances are given in this Wikipedia article.

If we know the density of air, the speed of wind, and the radius R of a given turbine, is it enough to find out how much power the turbine deliver, using the Eq. 2? Well... almost. There are still some extra factors. One of them is purely of a physical nature. The problem is that a wind turbine cannot "extract" **all kinetic energy** from the incoming wind. If it did, what would happen? Well, the kinetic energy of the air **after passing the turbine** would be zero, meaning also that its velocity would be zero – this is clearly not possible, because the air would start "accumulating" behind the rotor and would start blocking the incoming wind! The air behind the rotor **must keep moving!** So, what happens to the "downstream" wind? In order to accommodate the slower moving air, the stream behind the rotor has to widen considerably.

Here again is a link to a video explaining the "anatomy" of a wind energy converter – let's watch it, starting from the 4th minute: the reasons why not all energy carried by the "upstream" wind cannot be "captured" by the turbine is very clearly explained.



Then, how much power can be captured from the wind? This question has been answered in a paper published in 1919 by a German physicist Albert Betz who proved that the maximum fraction of the upstream kinetic energy K that can be "absorbed" by an ideal "actuator" – not necessarily a turbine, but any device capable of converting wind energy to another energy form– is $\frac{16}{27}$ K, or 59.3% of K. It is known as the **Betz Law** – and how it can be derived based on the energy, momentum and mass conservation laws is described, for instance, in the Wikipedia article linked.

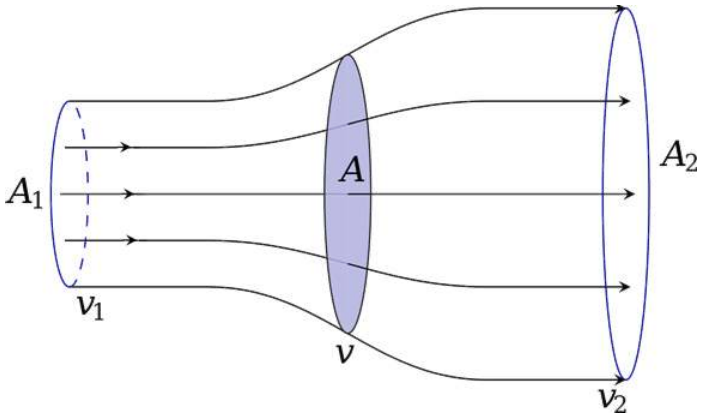


Figure 6.4.2: Retaining some speed enables the downstream air to escape and not to block the upstream air – but the downstream has to get wider than the upstream (source: Wikipedia). Detaled derivation of the Betz coefficient of performance can be found in this Wikipedia article or in this blog.

Let us repeat - 59.3% is the maximum fraction, assuming that the "actuator" is an ideal device. In reality, even the best turbines can capture only a fraction of the "Betz maximum". So, in math language - for a turbine rotor of radius R the maximum theoretical power allowed by the Betz Law is:

$$P_{\text{Betz lim.}} = \frac{16}{27} \times \frac{\pi R^2 \rho V^3}{2} \tag{6.4.9}$$

And in practice, a turbine delivers only a fraction ε of this power, so that the net power obtainable is:

$$P_{\text{obtainable}} = \frac{16}{27} \times \frac{\pi R^2 \rho V^3}{2} \times \epsilon \tag{6.4.10}$$

where E, to make things worse, is not constant, but is the function of something called the "tip-speed ratio" - we will create a TSR symbol for it:

$$\text{tip speed ratio } = TSR = \frac{\text{Linear speed of the blade tip}}{\text{Wind speed}} \tag{6.4.11}$$

We can therefore write the final result for the net windpower harnessed by a turbine:

$$P_{\text{net}} = \frac{8\pi R^2 \rho V^3}{27} \times \epsilon \quad \text{where } \epsilon = \epsilon(TSR). \tag{6.4.12}$$

The bad news is that that there is no simple mathematical expression describing the $\mathcal{E}(TSR)$ function. The shape of this function is different for different rotor types in for each type it has to be determined by experimental measurements.

The results of measurements of the net efficiency \(varepsilon_{\text {net }}=P_{\text {net }}/P \) plotted versus the Tip Speed Ratio for different types of rotors are presented graphically in a figure uploaded to the ResearchGate.net (a networking site for scientists) by Moshe Zilberman. From the figure it can be seen that the overall efficiency for a modern three-blade rotor has the maximum value of about 46% = 0.46 for the TSR value of about 5.5. By comparing it to the Betz Limit value, 0.593, we find that the value of E in Eq. 4 for this STR value is 0.46/0.593 = 0.776 = 77.6%, meaning that the turbine converts nearly $\frac{3}{4}$ of the wind power available by the Betz Law to mechanical power. For the American Multiblade farm windmill the maximum net efficiency is as low as 0.31, corresponding to E = 52%[1]. As noted, there is no simple algorithm for calculating the *E*(*TSR*) coefficient. However, there is a simple way of dealing with this problem – namely, the power output from a given type of turbine for different wind velocities can be measured experimentally and the results can be stored in an easily accessible database. A known Internet tool of this kind is a Swiss Wind Turbine Power Calculator. It contains the data for more than 50 types of the most popular turbines. After selecting the type, one gets the measured values of the output power of the turbine for speeds of wind from 1 to 30 m/s, with a 1 m/s increment. Such results constitute what is usually referred to as the “power curve” of the given turbine. For instance, for a popular (over 15,000 installed worldwide!) Vestas-80 with “nametag” output of 2,000 kW, the calculator yields:

Table 6.2: The performance of the Vestas-80 turbine

Speed of wind (m/s)	Power output (kW)	Speed of wind (m/s)	Power output (kW)	Speed of wind (m/s)	Power output (k)W
1	0	11	1652	21	2000
2	0	12	1892	22	2000
3	0	13	1982	23	2000
4	66	14	2000	24	2000
5	156	15	2000	25	2000
6	285	16	2000	26	0
7	467	17	2000	27	0
8	706	18	2000	28	0
8	706	18	2000	28	0
9	1001	19	2000	29	0
10	1329	20	2000	30	0

As follows from the above discussion, the three-blade turbine, which is dominant type used today, can deliver power as high as over 75% of the Betz Limit. It is instructive to confront that with the Power Calculator figures. Let's take, for instant, $V = 10\ \mathrm{m/s}$, for which the Calculator gives us $P = 1,329\ \mathrm{kW}$. The rotor diameter of Vestas V8o is 80 meters, so that the radius $R = 40\ \mathrm{m}$, and $A = \pi R^2$ is $5026.5\ \mathrm{m}^2$. The Power Calculator data are for the air density of $1.225\ \mathrm{kg/m}^3$. For these numbers from the Eq. 6.2 one gets $P = 3,078,761\ \mathrm{W} = 3,078.8\ \mathrm{kW}$. Correction for the Betz Limit, i.e., multiplication by $16/27$, yields $1,824.5\ \mathrm{kW}$. It means that the efficiency of the V-8o turbine is $1.329/1,824.5 = 0.7284 = 72.8\%$ - close indeed to the predicted 75%. But the Reader is encouraged to make similar checks for several other speed of wind values. It turn out that an efficiency close to 75% is obtained only at favorable wind conditions: not too weak and not too strong.

One reason for the V8o popularity is that it starts generating electric power at wind velocity as low as only $4\ \mathrm{m/s}$. At V exceeding $6\ \mathrm{m/s}$ it's output already is above 70% of the Betz Limit (see the rightmost column in the table), but next, for V exceeding $10\ \mathrm{m/s}$, the efficiency starts gradually falling. Why? The answer is simple, the maximum output power the generator in the V-8o turbine is capable to deliver is $2000\ \mathrm{kW} = 2\mathrm{MW}$. Any electric device has a limit power it can tolerate, otherwise it may overheat or short-circuit. And the power an electric generator delivers depends onhow fast it rotates. Apparently, at wind's velocity over 13 m/s the generator reaches its maximum allowed speed of rotation. Now, if *V* keeps increasing, the efficiency of the rotor is artificially lowered, in order not to allow the ro- tor to turn any faster. It can be done by changing the "pitch" of the blades

– again, we link here the YouTube video where it is shown, 3 minutes and 45 seconds after the beginning, how it is done by rotating the entire blades relative to the rotor hub.



But for wind speed \(\gt 25 \mathrm{~m} / \mathrm{s}\) it is no longer safe to let the rotor turn – so the blades are set to a neutral position in which they generate no torque and a special electromagnetic brake is engaged to completely immobilize the rotor.

---

1. It should be noted, however, that for millions of farmers who installed American Multiblade turbines not their power, but the **torque** offered by those windmills was the most important factor. In contrast to two- and three-bladed turbines, the multiblade rotors produce a high torque right from the moment the wind starts blowing – it's called the "start-up" torque. And the torque is crucial if the turbine is used, for operating a reciprocating water pump – what was their main task at American farms.

6.4: The Physics of a Wind Turbine is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

6.4.1: Problems with Wind Power

Wind power offers great advantages indeed: no CO_2 emission, not toxic emissions of any other kind, and the wind gives away it's energy completely for free. So, why don't we shut down all the polluting power plants and install instead enough wind turbines to generate as much power as those polluting plant do?

Well, it surely looks as a very good scenario for everyone who is concerned about the environment, the dwindling resources of mined fuels, and the devastating effects of mining. Yet, there are certain problems with wind power that make such a "grand plan" not fully realistic.

First of all, we should take into consideration something that is called the Capacity Factor. It is defined as follows. Consider a wind farm with a number of turbines. Each turbine has a "nameplate power", i.e., the maximum electric power it can generate (for instance, for the Vestas V8o turbine discussed in the preceding section it is $2MW$). Now, add up the nameplate powers of all individual turbines comprising the farm. It will be the total maximum power output of the entire farm. Then, think of a prolonged time period: a month, a quarter, or a full year - call it Δt. Multiply the maximum power output by this time period, and you will get the maximum energy output from the farm over Δt.

Then, find out what is the actual amount of energy that the farm has generated over the the time period considered - and calculate the capacity factor, defined as:

$$\text{capacity factor } = \frac{\text{actual energy output over } \Delta t}{\text{maximum energy output over } \Delta t} \tag{6.4.1.1}$$

The bad news is that the capacity factors for wind power facilities are not very high. An unquestionable leader in wind power technology is Denmark – it's a very windy country, and most of its wind power comes from off-shore wind farms. The best of those attain the capacity factor as high as 50%, and for all installations it reaches about 40%. But Mother Nature is not as generous for the United States – here, the average for all states, according to the Energy Information Administration, is 27%.

Suppose that one day we decide to eliminate all fossil fuel power plants and to replace them with wind turbines. In 2016, according to the U.S. EIA, 33.8% of electric power generated in the US came from natural gas power plants, and 30.4% from coal-burning plants. And the total installed nameplate power was 1177 GW. So, 64.2% out of 1177 GW, = 726.7 GW was the nameplate capacity of all fossil fuel facilities. And the nameplate capacity of all windpower facilities was 82 GW. But due to the capacity factor, these facilities generated only 27% of 82 GW = 22.1 GW. What's the conclusion? Well, it is that if we wanted to eliminate all 726.7 GW obtained from fuel burning and replace by 726.7 GW generated by wind, we would need to increase the existing potential of windfarms 726.7/22.1 = 33 times! It means that next to each existing wind farm we would need to build 32 wind farms of the same size!

It would be a grandiose project – unfortunately, it still won't be enough. The problem is the natural intermittency of wind – especially, over lands, there may be pronounced changes in the the wind velocity – sometimes ir- regular variations in the wind speed and direction occurring over relatively short periods of time, so they should be termed as "fluctuations". In Fig 6.4.1.1 it is shown how such fluctuations look like in Poland, a country located in central Europe[1]. The total consumption of electric power in Poland may reach about 25 GW in the peak hours, so that the utility sector needs to have at least 5 GW of reserve power in "peakers" ready to be launched on a short notice in order to compensate for a possible sudden hushing of the wind.

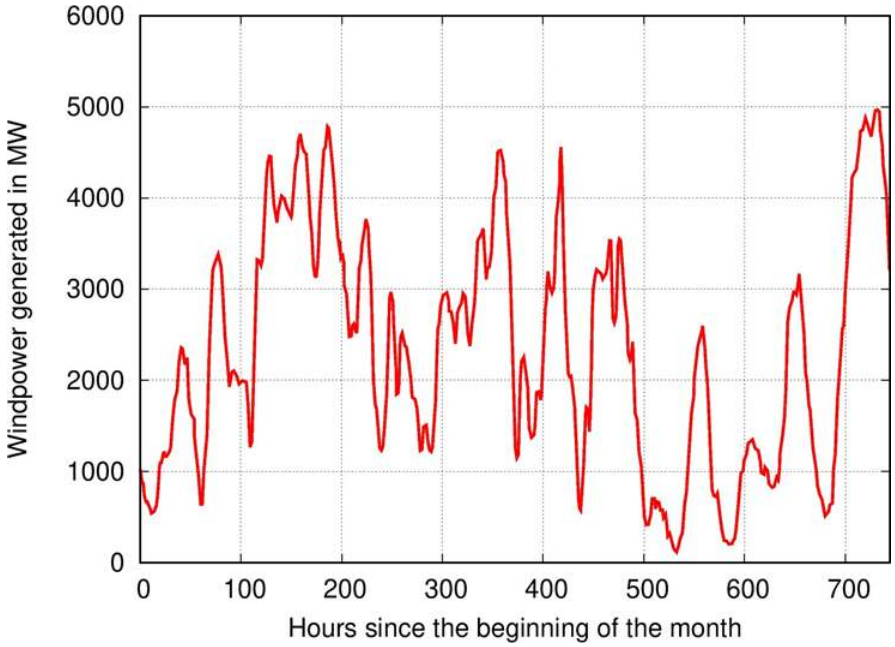


Figure 6.4.1.1: Typical pattern of windpower electric generation in Poland. The data are from December 2019, recorded every hour over 31 times 24 = 744 hours. Source: hourly data published by the operator of the national grid in Poland "Polskie Sieci Elektroenergetyczne (PSE)" and systematically monitored on Mrs. Jozefa Sokolowska's blog – chart based on the data collected in her blog, with her permission.

And now, a crucial question – how much power **should be** generated? From any worker of an utility company at any point of the globe you will get the same answer: **As much as the customers demands!** It cannot be less, because the grid which is not able to satisfy the needs of all customers may react with "black-outing" some regions of the area it services. And a surplus is not better, a grid into which too much power is "pumpeded" may react irresponsibly. Utility companies know how to handle such situations

– normally, in the system they administer there are several different types of power plants, some of them able to quickly change their output power to respond to the fluctuations in the windpower generation. For instance, hydroelectric facilities may increase or decrease their output power in a few minutes. And if there are pumped storage hydroelectric plants in system, they may be quite happy if there is a sudden surge in the windpower. Another "backup power sources" widely used are natural gas turbines, whose start-up times from zero to full power is typically about 30 minutes.

And if it's not enough, the utilities may still rely on import from or ex- port to other regions. In Europe, there are special transmission lines between several countries that can be activated in such situations – one example is the 2 GW line connecting Denmark with its "sister county" Norway. Denmark relays heavily on windpower – and in Norway, power is generated almost ex- clusively in hydroelectric plants. If in Denmark there is a surge in windpower generation, the surplus is sent to Norway, which then curtails the output from its hydro plants – and if there is a "dip" in windpower in Denmark, Norway sends more water to its turbines and sends the surplus power to Denmark. In the US individual states can help each other in the same way, or simi- lar export-import may be arranged between American states and Canadian provinces.

There may be, however, even more serious situations which the utility workers call "ramp events". Normally, the utility systems in which a sig- nificant percent of power comes from wind farms observe the forecasts with attention. If the wind are expected to weaken, they make plans for many hours ahead how to use their backup sources. A ramp event is a situation when the dip in windpower appears to be much more serious than that fore- casted. An event of such kind that received much publicity had occurred in Texas on February 26, 2008. In short, on that day in the afternoon the wind died out completely in a western part of the state where there were several farms with large turbines. Over three hours their contribution to the state's grid fell by 1500 MW. It coincided with a spike in demand. The grid opera- tors started heavily importing power from other regions – but so much power had to be sent that the transmission lines started overheating. The system was nearly collapsing – but fortunately, the operators had a "last resource" they decided to use. Namely, being aware that such situation may happen, some time before they had struck an agreement with several large industrial customers – allowing to cut off power for them in emergency situations after a 10-minute warning period. 1200 MW was taken off from the grid, and the collapse was avoided. With the exception of the aforementioned

industrial plans, other power consumers were totally unaware that the situation really dramatic and if drastic measures were not taken at the last moment, the collapse of the grid was imminent.

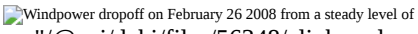


Figure 6.4.1.1: Total Texas windpower profile on February 26, 2008 (from the National Renewable Energy Laboratory Technical Report NREL/TP-5500- 49218).

The Texas 02/26/2008 ramp event was described and analyzed in several professional reports and even in periodics of wider circulations[2]. What was learned from it certainly helped grid operators to be better prepared for such possibly-catastrophic events in the future. It should be noted that the collapse of the grid in one region may trigger an avalanche effect causing that eventually a whole macro regions are blacked out. Such “cascading” was responsible for two major power outages in the Northeast in 1965 and in 2003. In each tens of millions of people were left without electricity for many hours.

In conclusion – there is no doubt that winds as very friendly and clean source of electric power – and may also be used for other purposes, e.g., wind may again start helping commercial vessels to cross the oceans. The available wind power is some 40 times larger than all electric power currently consumed by humanity. And no matter how much windpower we use, there is no fear that the wind resources will be ever depleted – our Sun is taking care of that.

On the other hand, the facts that have been presented above should be taken as a warning against “overenthusiasm”. The capacity factors of 30% or 40% may look good if one looks only at them – in such narrow perspective one may conclude that the wind sector alone may be able to satisfy the needs of a country as big as the U.S.. Well, the capacity factor becomes highly misleading if one “forgets” that it is an **average** based on data collected over a long period, usually over the entire year, or over a season. The factor’s value alone doesn’t tell anything about the sudden windpower surges, die-outs and ramp events. Here a positive example may be the Danish-Norwegian partnership. Denmark is a leading nation as far as the windpower is concerned, with the capacity factor as high as 42%. It deserves praise, but it also means that the Danish utility grid may be knocked down in the case of a sudden wind hush. In the Danish-Norwegian partnership a recipe has been worked out that fixes such problems: namely, Norway, who is able to considerably increase the output of it’s hydroelectric facilities “on call”, simply acts a gigantic “back-up generator” for the entire Danish national utility system (which relies, as noted, in 42% on windpower – in professional terminology, one would say: *the windpower **penetration** in Danish electric power generation system is 42%*.).

Unfortunately, there is no “another Norway” – or, rather a “Super-Norway” next to the U.S. borders which would be able to come to rescue of the U.S. wind sector in case of a huge developing ramp event. Such a backup system should exist **inside** the U.S.. Is it possible to get one? Yes, definitely. One way towards that is *diversity* : the entire system should consist of not only of one major component, but of several renewable energy “branches”, each of which should have some “reserve capacity”. Also, the success in develop- ing a comprehensive renewable energy sector, capable of eliminating power sources using fossil fuels, depends very much on developing in parallel an en- ergy storage system mighty enough to balance out sudden substantial losses of power due to an unexpected ramp event – let us note that such events are not limited to windpower. Solar power may also exhibit ramping effects. In general, solar and wind power are not easy to manage due to their variable and stochastic nature – especially, when penetration of one of those sources or both attains a high level. This discussion will be continued in the chapters devoted to solar power and to the storage of energy.

---

1. Poland is often referred to as an “Eastern European” country – which is perhaps correct from the viewpoint of political sciences. But geographically, Poland is located in the **very center** of the continent of Europe. No matter how this center is defined, either as the center of gravity of a figure obtained by clipping Europe out of the World map, or as the center of a circle circumscribed about such figure.

2. See, e.g., *Renewables Test IQ of the Grid* by Dan Charles, in *SCIENCE*, 10 Apr. 2009, vol. 324, www.sciencemag.org)

6.4.1: Problems with Wind Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

7: Direct Solar Power

7.1: Introduction
7.2: Harnessing Solar Power
7.3: The Concentrated Solar Power (CSP) Technology
7.3.1: Solar Towers Molten Salt Heat Storing Technology
7.3.2: Parabolic Troughs and Fresnel Mirrors
7.3.3: The Linear Fresnel Mirror CSP Technology
7.3.4: CSP – Pros and Contras
7.4: Photovoltaic Conversion of Solar Power
7.4.1: Theoretical Background the Nature of Light and the Photoelectric Effect
7.4.2: Semiconductors and Doped Semiconductors
7.4.3: The p-n Junction
7.4.4: Internal Photoelectric Effect in Semiconductors
7.4.5: PV Cells
7.4.6: Solar Cell Efficiency
7.4.7: Inverters
7.4.8: Pros and Cons of PV Solar Power Harnessing

7: Direct Solar Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.1: Introduction

Direct solar power – the solar radiation arriving directly from Sun – is the "greenest" of all energy sources, one that dwarfs all other available energy sources combined (including those who are "transformed" forms of solar energy).

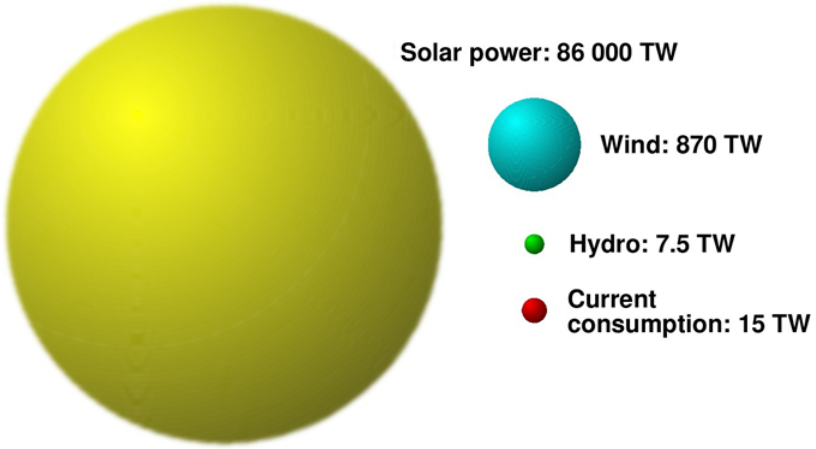


Figure 7.1.1: Graphic comparison of the total power of the incoming solar radiation with maximum power obtainable from some other renewable energy resources.

Some important definitions: *Solar Constant* , or the *Total Solar Irradiance (TSI)*

- the total power of solar electromagnetic radiation that falls on a unit surface area at a vertical angle above the Earth atmosphere, at a distance of 1 astronomic unit[1] from the Sun. It's value is Extra close brace or missing open brace it's the average value over a 11-year solar cycle, with an amplitude of about 0.1%, and short-period fluctuations – see the TSI variation graph for the years 1980-90).

Knowing the TSI value, one can calculate the total solar power P_{tot} reaching the top of the Earth'a atmosphere. The average Earth's radius is $R = 6371\ km$. But from the Author's 10 year perspective of teaching the *Energy Alternatives* course it follows that the task is not "100% conceptually clear".

** **A Digression: the Instructor's Adventure with** P_{tot} ** Every year the above task was given as a classroom test, together with a "hint": *The surface area of a sphere is* $4\pi R_2$*, and the surface area of a disc (one side) is* πR^2*. Which formula shall you multiply by the TSI value to get the right value?*. The answers showed that the problem was **not** conceptually clear for many students. Few students chose the $4\pi R^2$ formula, but a surprisingly large number of them insisted that one should take $2\pi R^2$. When asked why, they usually gave an explanation like this: $4\pi R^2$ *is not the correct formula, because it's the* ***entire*** *surface area – but only one-half of the Earth's surface is illuminated by Sun, there is night on the other half, so the right formula to take is* $2\pi R^2$.

The Author tried to explain why that was wrong: *Look*, he persuaded, *the TSI definition talks about the* ***radiation that falls on a unit surface area at a vertical angle****. But this is not the case for all Earth, the radiation arrives from one direction only, so for most spots on Earth it does not fall at a vertical angle!*. So, the students reluctantly agreed that indeed, $2\pi R^2$ would not be the right formula – but such conclusion, regretfully, did not bring them any closer to the right formula.

Not surprising! Because if the radiation impinges a unit area not at right angle, but at an angle $\theta < 90_\circ$, than the power that falls on this area is not TSI, but $TSI * sin\theta$. And if a parallel beam of radiation falls on a sphere, then the θ angle is different for different spots on the sphere's surface... How should one deal with such a problem?

The students were right. For calculating the total power impinging the illuminated side of a sphere one would need to take into account the θangle and perform an integration over the illuminated part, which would not be a trivial task. Not too many students taking the Energy Alternatives course have enough background in calculus for setting up and solving the integral needed.

But the integral is not necessary! The instructor's fault was to give an "insufficient hint". A good hint would be like this: *suppose that you have magic powers, and you want to block all sunlight coming to Earth by placing a gigantic black disc in front of it. What is the radius of the disc you would need, and how much solar power it would absorb?*

Now the problem is easy! To block all sunlight from Earth, you need to put a disc of radius $R = 6371\ km$ in front of the planet, on the path of the incoming sunlight and perpendicular to the rays. So, the power deposited now on the disc is $TSI\pi R^2$. And this is the

solar power blocked from reaching Earth. So, if we remove the disc, the power deposited on Earth would be the same, $P_{tot} = TSI\pi R^2$. And if we now insert the numbers, we get:

$$P_{tot} = 1361.5\ W/m^2 \times \pi \times (6,371,000m)^2 = 1.736 \times 10^{17}\ W = 174\ PW \quad (7.1.1)$$

**** End of the Digression. ****

Insolation. $TSI = 1361.5 W/m^2$ only at the top of the Earth's atmosphere. Of the arriving 174 PW of power, only about 89 PW reaches the Earth's surface. In Fig. 7.1.3 it is shown what is weakening the power when sunlight travels through the atmosphere. First of all, some radiation (about 10 PW) is back-reflected by the atmosphere, and even more by the clouds (about 35 PW – that much is back-reflected from the clouds in the famous picture of Earth taken by the Apollo 17 crew in 1972.



Figure 7.1.2: The famous photograph of Earth, taken by the Apollo 17 crew in 1972 (source: NASA, Wikimedia Commons).

One can see the white clouds, and the whole photograph is bluish in color. It comes from the fact that the efficiency of "back-reflecting" and scattering light by atmospheric gases depends on the light color: for visible radiation, the effect is the strongest for for the shortest wavelengths, i.e., for the color blue[2]. In the same photo, near the top one can also see what's undoubtedly the contour of Africa. The color of the landmass is brownish

– Earth's surface also back-reflects some impinging sunlight (about 7 PW), and the back-reflected color obviously depends on whether the light reflected from water surface or from a landmass. And a big portion (some 33 PW) of the impinging radiation is simply absorbed by the atmosphere.

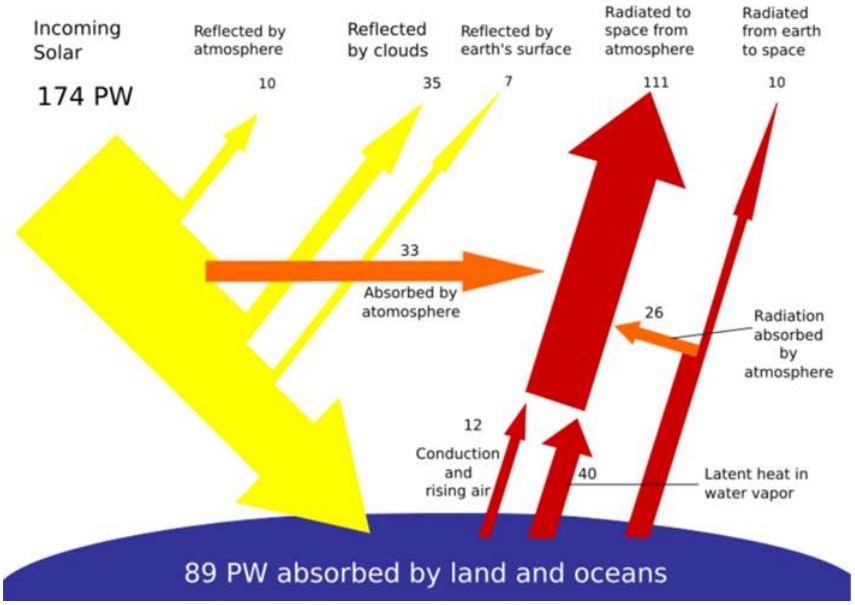


Figure 7.1.3: Breakdown of incoming solar power (source: Wikimedia Commons).

So, effectively, of the impinging 174 PW only about 89 PW reaches the Earth surface and is therefore available for being converted for other forms of power by humans. It means that the "effective TSI value" at the Earth surface on a sunny day may be between 0.8 and 1.0 kW/m^2, depending on the geographic location. This brings us to the notion of insolation.

Insolation (not to be confused with *insulation*, or with *shoe insoles*): a measure of solar radiation energy received on a given surface area in a given time. It is commonly expressed as a 24-hour average irradiance in Watts per square meter (W/m^2). Note: a square meter of *actual terrain* at a given location, not a square meter perpendicular to the incoming radiation, as in the definition of STI.

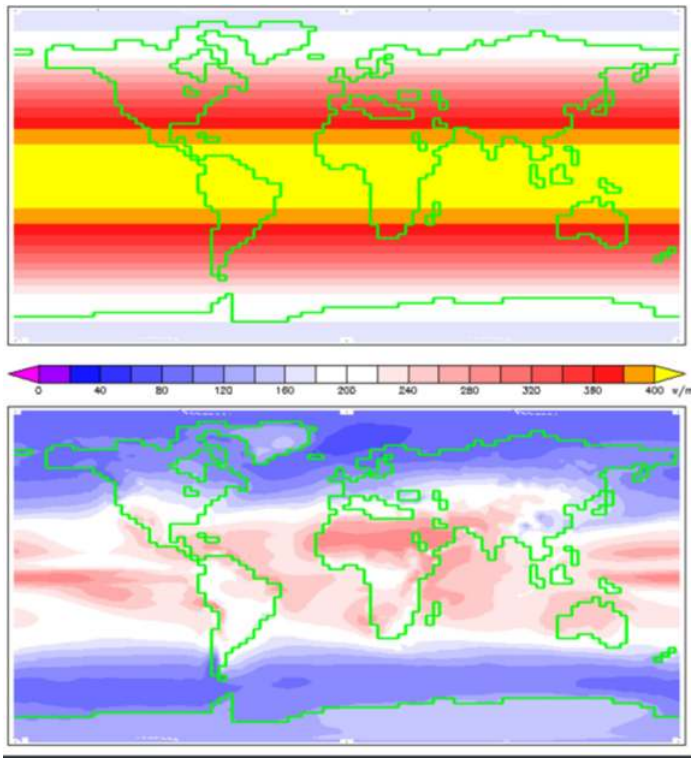


Figure 7.1.4: Annual insolation. Upper panel: at the top of the atmosphere. Lower panel: at the Earth surface. The scale is in W/m^2 units (source: Wikimedia Commons).

The geographic distribution of the insolation is shown in Fig. 7.1.4. At the top of the atmosphere the maximum insolation occurs in the equatorial regions, where the value, as indicated by the yellow color, may exceed \400 W/m^{2} \) 24h. It's easy to understand where such number comes from. Consider a 1 m^2 "imaginary unit surface area" (call it IUSA[3]) at the top of the atmosphere at it's equatorial region. The atmosphere rotates together with our planet, so that for 12 hours IUSA is in darkness. Then, it crosses what astronomers call the *terminator* , which is the line that divides the daylit side and the dark night side of a planetary body. At this moment the solar rays fall on the IUSA at a θ angle equal zero – and for the next six hours, θ gradually increases, finally reaching $\theta = 90_{\circ}$ when IUSA crosses the line drawn through the centers of Sun and Earth. And from that moment on, θ keeps decreasing and falls to zero when IUSA again crosses the Terminator, now re-entering the dark night side of Earth.

And from that moment on, θ keeps decreasing and falls to zero when IUSA again crosses the Terminator, now re-entering the dark night side of Earth.

The sunlight power deposited at a unit surface area, when it's not perpendicular to the incoming rays, is $P = \sin\theta \cdot 1362\ \mathrm{W/m^2}$. When θ gradually changes from o to 90°, or to $\pi/2$ in radians, $\sin\theta$ increases from o to 1 , and the average value of the $\sin\theta$ function is then:

$$(\sin\theta)_{\text{aver.}} = \frac{\int_0^{\pi/2}\sin\theta d\theta}{\pi/2} = \frac{1}{\pi/2} = 2/\pi = 0.6366 \tag{7.1.2}$$

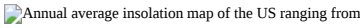
So, during the first 6-hour period the average power delivered to the IUSA is $0.63661362\ \mathrm{W/m^2} = 867\ \mathrm{W/m^2}$. For the next 6-hour period it's also 867 $\mathrm{W/m^2}$, and then, for the 12 nighttime hours, its zero. So, the average for a 24-hour period is $867/2 = 433.5\ \mathrm{W/m^2}$, which indeed agrees well withthe color-scale in Fig. 7.4.

If we put our IUSA not at the Equator, but more to the north or more to the south at a Φ latitude, the $433.5\ \mathrm{W/m^2}$ insolation has to be additionally multiplied by $\cos\Phi$. For instance, for Corvallis Φ is close to 45° North, so that the insolation for an IUSA flying high over our city would be 434:0.7071 $= 307\ \mathrm{W/m^2}$. More about calculating insolation at different latitudes can be found in a dedicated Wikipedia article.

In the solar power industry, they usually prefer to use not insolation, but solar irradiance instead, expressed not as the 24-hour average of sunlight power deposited to an area of a square meter, but as the total radiation energy deposited to such an area. The unit used then is $\mathrm{kWh/m^2}$. The conversion from insolation expressed in $\mathrm{W/m^2}$ to solar irradiance expressed in $\mathrm{kWh/m^2}$ is easy, one simply needs to multiply the former by 24 . For instance, the $307\ \mathrm{W/m^2}$ insolation result we have obtained for an IUSA flying high over Corvallis translates into a solar irradiance of $307\ \mathrm{W/m^2} \times 24\ \mathrm{h} = 7368\mathrm{Wh/m^2} = 7.37\mathrm{kWh/m^2}$.

But as far as the insolation at the ground level is considered, it may be much lower than that measured at the top of the atmosphere. For instance, for Corvallis the highest is in July: $5.94\mathrm{kWh/m^2}$ day, which is $5.94/7 \cdot 37 = 0.806 = 80.6\%$ of the annual average at the top of the atmosphere. In December it's the lowest, $1.00\mathrm{kWh/m^2}$ day, which is 13.6% of that annual average. What's the reason of such a big difference? Everybody knows, the clouds! So, not only the geographic latitude is important, the number of sunshine hours in the region also matters.

There are 8766 hours in a year, daytime hours are slightly more than one half of that number (because of the size of the Sun disc – the daytime lasts from the moment the disc starts emerging from under the horizon, to the moment the disc completely disappears). In the sunniest regions of Earth, such as the Sahara Desert, there are more that 4000 sunshine hours in a year, which translates into average annual insolation of 6.3 kWh/m^2/day. The map of average insolation in the USA is shown in Fig. 7.1.5.



8 kWh/m_{2} day in the southwest to 3 in Alaska and the northeast"
src="/@api/deki/files/56353/clipboard_e2f51b0d53caad88cdbbae2703d8c4ff3.png">

Figure 7.1.5: Annual insolation map of the US (source: Natl. Renewable Energy Laboratory via Wikimedia Commons).

---

1. An astronomical unit, a.u., is the average Earth-Sun distance, 149 597 871 kilometers.

2. For the same reason the setting sun is red – at the sunset, the light has to travel a long way through the atmosphere, so that most blue light is back-reflected and "ejected" from the beam of sunlight, while these effects are much weaker for the red light.

3. It's a name created *ad hoc* by the Author, you'll not find it in "official literature".

7.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.2: Harnessing Solar Power

As shown in the Fig. 7.3, out of the total of 174 PW of solar power reaching the Earth, only 89 PW (51%) reach the Earth surface. Yet, we don't need to care about the 49% "losses", because the remaining 51% still represent a tremendous amount of power, if we compare it with the current global consumption of all forms of power, which net about 21 TW. The latter is only a tiny fraction of the 89 PW = 89 000 TW, let's check: 21 TW/89 PW = 21/89000 = 0.00024 = 0.024%.

This is an impressive figure – but numbers written on paper or on com- puter screens still make a much weaker impression than the same information presented in a *graphic form*. Such a piece of graphics, definitely worth look- ing at, was created by a German scientist, Dr. Matthias Loster – please click on this link to watch it.

Dr. Loster's graph shows a map of Earth, with a global distribution of insolation shown using a color scale. On this map, there are six black dots. Each dot corresponds to a circular area of radius 200 km, so it's total surface area is about 125 000 square kilometers. Suppose that each such area is totally covered by devices converting solar radiation to other forms of usable power – i.e., each dot is a giant "solar power plant". And it turns out that the sum of output power from those six "power plants" would be 18 TW – which was the total global consumption of power in the year 2006, when Dr. Loster made the graph.

The total area of all lands on Earth is about 150 million square kilometers. The total area of all "dots" correspond to about 750 thousand square kilometers – about 0.5% of the total land area.

Is it feasible to use 0.5% of the land area of Earth to build solar power plants? Well, let's take a closer look at Dr. Loster's graph, and then on a graph showing the largest desert areas on Earth.

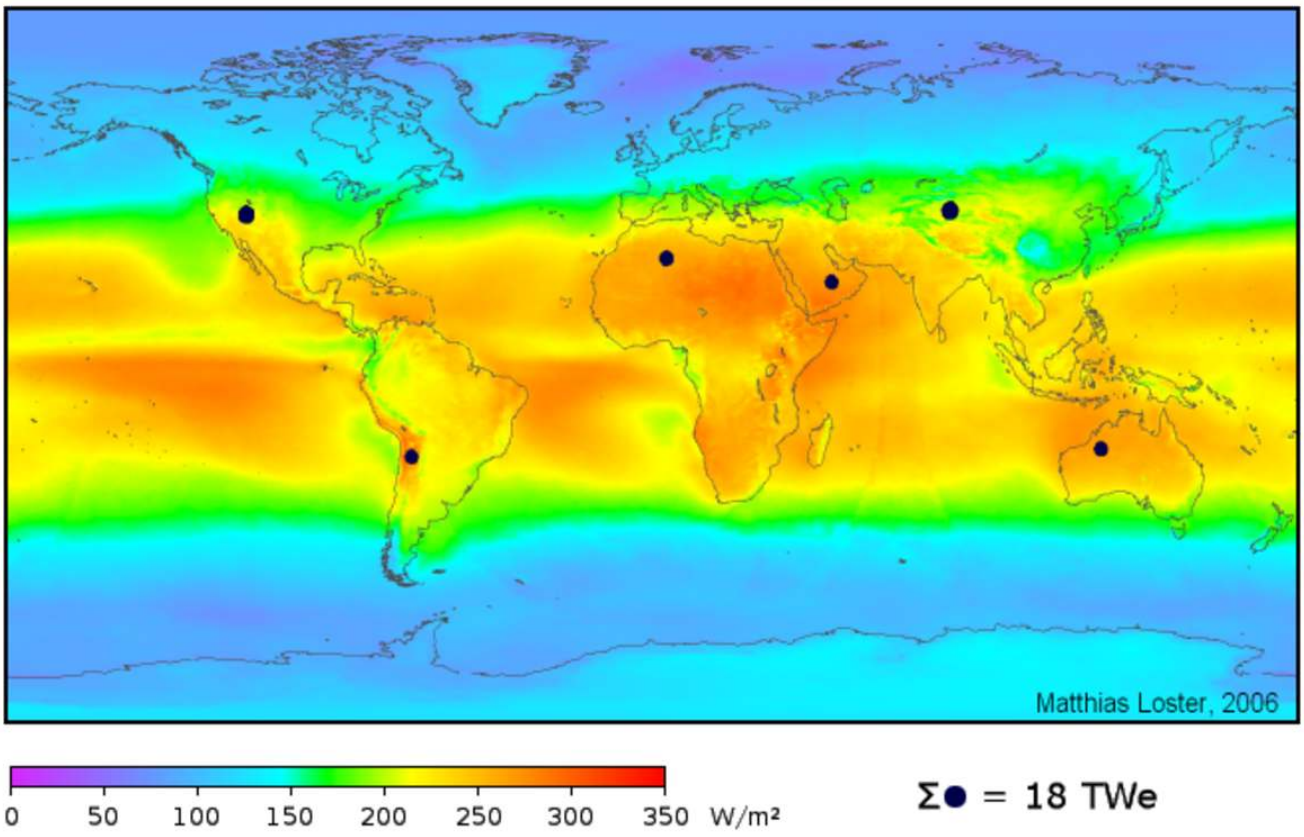


Figure 7.2.1: Mattias Loster's idea that six large solar arrays could power the world

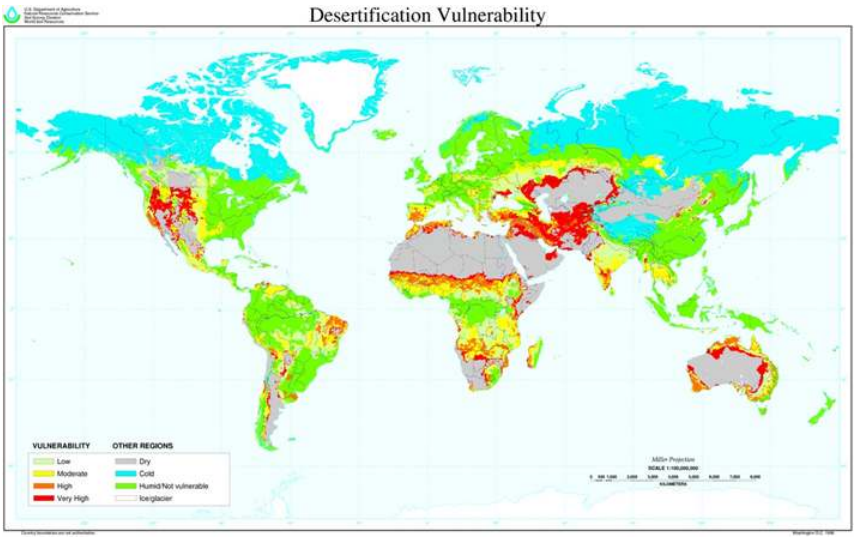


Figure 7.2.2: A map of desertification, showing the largest desert areas on Earth (source: Wikimedia Commons).

It's clear that Dr. Loster put his six dots in the middle of large deserts area. So, those giant solar plants won't "steal" land from agriculture. Or occupy land that can be used for building cities – a perspective of living in an area where the temperature may reach 120°F, and there is now water, is not very attractive for most people.

In Dr. Loster's graphics, there are then two "messages". One is simple: *See!* There is "more than enough" solar radiation to satisfy all humanity's needs for power – and it seems likely that even if the demands double or triple, there still will be enough!

The other "message" is a bit more "hidden" because the graph does not show that the dots are all located in desert areas which are not very suitable for use for other purposes. What is more visible, however: they are arranged around the globe in such a way that some of them are always illuminated. In other words, the message is: *It is possible to build a global solar power system which will provide power 24 hours a day because some of the constituent power plant will always be active*. It's clearly a great idea – yet, one must realize that the creators of such a global system would have to deal with two huge challenges: one is of technical nature, and the other of political nature.

The technical challenge is that the power from a real global system must be available to all dwellers of the globe. So, it would be necessary not only to build the plants, but also to create a "world wide web" of transmission lines. Would it be feasible at all? Certainly, such a global system cannot be created overnight. But there are reasons for optimism. Currently, the world's longest existing transmission line is in Brazil. It uses the High-Voltage Direct Current (HVDC) technology, is 2,375 km (1,476 miles) long, and transmits over 6 GW of power. China intends to build an even longer (3,000 km) Ultra-HVDC line, capable of transmitting 12 GW of power[1]. So, if a large number of such transmission lines are built, reaching to all countries, the world-wide power transmission network may become a reality. Well, perhaps in principle...

In principle, because the other challenge is really serious. A global solar power system can work only in a **politically united world**. Currently, the world is divided into a number of political blocs, which compete with one another, or even wage wars with one another. If a global power system were created, then, naturally, the regions located closer to the power generating centers would have more control over the transmission lines than the regions located far away from the centers. And if a conflict emerges between two regions, one of them having more control over the system than the other – "blackouting" the adversary may be used as a powerful weapon. In today's world, where new and new political conflicts erupt all the time, no country would be willing to risk its energy security. Therefore, the creation of a global solar power system "must wait for better times".

But a global system, of course, is not the only option. Almost all coun- tries in the world are now investing in solar power. The power installed by the end of 2016 is listed by country in this Wikipedia article. The world- wide cumulative power in 2016 was at least 308 GW. As far as the "new renewable power sources" are concerned, the cumulative solar power in 2016 was second only to cumulative worldwide wind power which in 2016 was 487 GW. However, the growth of installed solar power capacity in 2017 was 98 GW, almost twice as high as the growth of the installed wind power capacity in the same year (52.6 GW). Therefore, as far as the cumulative installed power is concerned, solar power may soon become the leader. Which is not surprising, considering that – as follows from the data shown in Fig. 7.1 – there is about 100 times more available solar power on Earth than wind power.

So, solar radiation may offer us a real "El Dorado[2] usable energy". The only problem is how to convert it to electric power. Well, not exactly a problem, because there exist two mature (or almost-mature) technologies for performing such a conversion. One is the *solar thermal conversion*, or the *concentrated solar power (CSP)* method. The CPS power plants are large

facilities, with power of tens, or even a few hundreds of MW. They can be installed only in areas of very high insolation, which is a drawback – but they can keep generating power for several hours after the sunset. The other technology is the *photovoltaic (PV) conversion*. There neither low nor high limit for power of PV installations – rooftop installations at single family houses are of a few kW power, while giant arrays of PV panels may deliver hundreds of MW. The PV installations are also "more tolerant" to weather caprices. We will discuss the CSP and the PV technologies in greater details in the following two Sections.

---

1. To give the Reader an idea of what 12 GW is: it's more than the average power consumption in small industrialized European countries like Denmark or Switzerland, and about the same as the average power consumption in the Netherlands.

2. El Dorado – a city of fabulous riches believed to exist in South America by 16th century explorers.

7.2: Harnessing Solar Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

7.3: The Concentrated Solar Power (CSP) Technology

The simplest method of heating a boiler of in such a way is to put the boiler at the top of a high tower surrounded by a number of mirrors, called *heliostats*. A heliostat is a mirror, usually a flat one, which can be rotated and the tilt of which can be changed – so that, when Sun moves across the sky, it always sends the reflected beam of sunlight to the predetermined target.

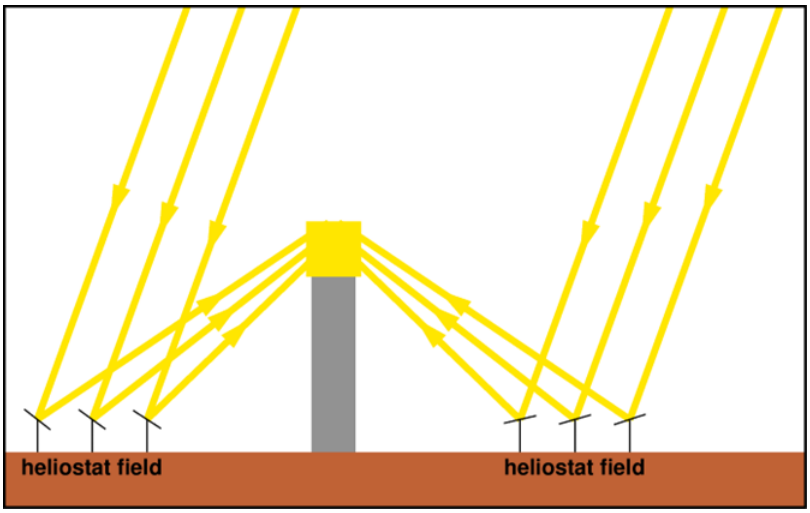


Figure 7.3.1: A scheme of a CSP power plant. A tower is surrounded by fields of heliostats – e.g., of large mirrors, the individual orientation of each is computer-controlled so that the reflected beams is always sent towards the "receiver" at the tower's top – no matter what's the position of Sun in the sky (aop).

An example: In order to make a quick estimate of how many heliostats are needed in such a power plant. Suppose it's noon on a sunny day. Suppose that the effective Total Solar Irradiance in the area where the plant is located is 0.8 kW/m^2, and each heliostat carries a mirror of the $3 \times 3\ m$ size, i.e., with the surface area of $9\ m^2$. But a single heliostat does not send a beam of \(9 \; m^2 \times \;0.8 \;kW/m^{2} = 7.2 kW \) towards the tower, because the sunlight does not fall on the heliostats at a right angle. Actually, the incident angle is different for each individual heliostat. In order to make our life simpler, let's assume that the average power sent by an individual heliostat is just one-half of the power we have calculated above, i.e., 3.6 kW (perhaps it's too conservative a value, but, remember – we want to make an *estimate* only). So, for depositing the power of 1 MW = 1000 kW at the boiler, we need to set up a field of 1000/3.6 = 278 heliostats.

Let's take a look at an existing CSP plant – for instance, at the Crescent Dunes Solar Energy Project, a facility capable of generating 110 MW of electric power. After clicking on the link provided, one will see the aerial photograph of the heliostat field. A big one, isn't it? It's difficult to count the heliostats, but their number is given in the article: 10,347. OK, in the example we worked on above we found that 278 heliostats are needed for 1 MW. So, 10,347 of them should yield 10347/278 = 37.2 MegaWatts. Why not 110 MW? Well, if we keep reading the article, we find that they use much larger heliostats than in our example, not of $9\ m^2$, but of $115.7\ m^2$. 115.7/9 = 12.9 times larger! So, we have to multiply the 37.2 MW by this number and we get... 37.2 MW 12.9 = 480 MW. Now, much to large a power! What's going on?

Well, we have to keep in mind that 480 MW we have gotten is the *thermal power delivered to the boiler*. And the steam from the boiler is sent to a turbine + generator system, like in a conventional thermal power plant.

Typically, as was discussed in Chapter X – the efficiency of converting steam thermal power to electric power in such a system is 30%. So, 30% of 480 MW is 144 MW – closer to the figure of 110 MW. Considering that what we got from our example was an **estimate** only, we can conclude that the 144 MW result is "reasonably close" to the real value of 110 MW.

But there is another possible explanation for the difference. There is one more piece of important information in the article about he Crescent Dunes facility: namely, it can generate power for several hours **after** the sunset. How? During the daytime hours some of the sunshine energy "captured" is not sent to the turbine, but is stored and later reused for electricity generation, when there is no

sunshine. So, we probably had to use not 480 MW in our calculations, but 480 MW *minus* the amount of power sent to a reservoir to be used later. And how such a reservoir works is explained in the Subsection that follows.

Molten Salt Heat Storing Technology

In the Crescent Dunes power plant there is huge dual-chamber reservoir containing M = 32 000 000 kg of molten salt – a mixture of 54% of KNO_3 and 46% of $NaNO_3$ (sometimes referred to as the "solar salt"). Both components are inexpensive fertilizers. In the 54/46 weight proportion they form a so-called Eutectic system with the lowest melting temperature (131°C). The reservoir consists of two thermally-insulated tanks: the "cold one", which stores molten salt of 288°C temperature, and the "hot one", to where the molten salt is transferred after being heated during the "storing" phase – and where it's stored at temperature of 566°C.

During the "generating" phase, the salt travels the other way: it's pumped to a "steam generator", which is nothing else than a *heat exchanger* [1] in which the molten salt is cooled from 566°C down to 288°, in the process heating up water and making it a 566°C steam.

The heat capacity [2] of the molten salt mixture is about $c = 1.5$ J/kg1°C. In the steam generator the molten salt is cooled down from $566^{\circ}\mathrm{C}$ to $288^{\circ}\mathrm{C}$, i.e., by $\delta t = 566^{\circ}\mathrm{C} - 288^{\circ}\mathrm{C} = 278^{\circ}\mathrm{C}$. Hence, the total amount of heat ΔQ given away by the $M = 32 \cdot 10^6$ kg of the "solar salt" is:

$$\Delta Q = M \cdot c \cdot \Delta t = 32 \cdot 10^6 \text{ kg} \times 1.5 \cdot 10^3 \text{ J/kg} \cdot 1^{\circ}\mathrm{C} \times 278^{\circ}\mathrm{C} = 1.33 \cdot 10^{13} \text{ J} \tag{7.3.1}$$

If this thermal energy is released over the period of 10 hours $= 36000$ s, the thermal power released is 1.3310^{13} J/36000 s $= 36910^6$ W $= 369$MW. Of thermal power - but if we assume 30% efficiency in conversion to electric power, we obtain... 0.3$69 MW = 110.7 MW!

So, it seems that we got too low a result (144 MW) because it was based on estimates that had been "too conservative". Probably, the real story is such that the thermal power deposited at the tower top in Crescent Dunes is not 480MW, but rather 740MW - of which one half, 370MW is converted to electricity right away, and the other 370MW is stored in the molten salt tanks.

The example of the Crescent Dunes solar power plant points out to a considerable advantage of the CSP technology - the capability of generating power after sunset. With the molten saltstorage reservoir, a power plant may be even able to generate electricity at the same power level for full 24 hours. However, the largest tower-type CSP power plant in the US, the Ivanpah Solar Power Facility with three heliostat fields a net power of 392MW, does not use thermal storage. Other US solar power plants do, although they are not tower-type plants (for the list of the largest CPS plants in the world, please see this Web document).

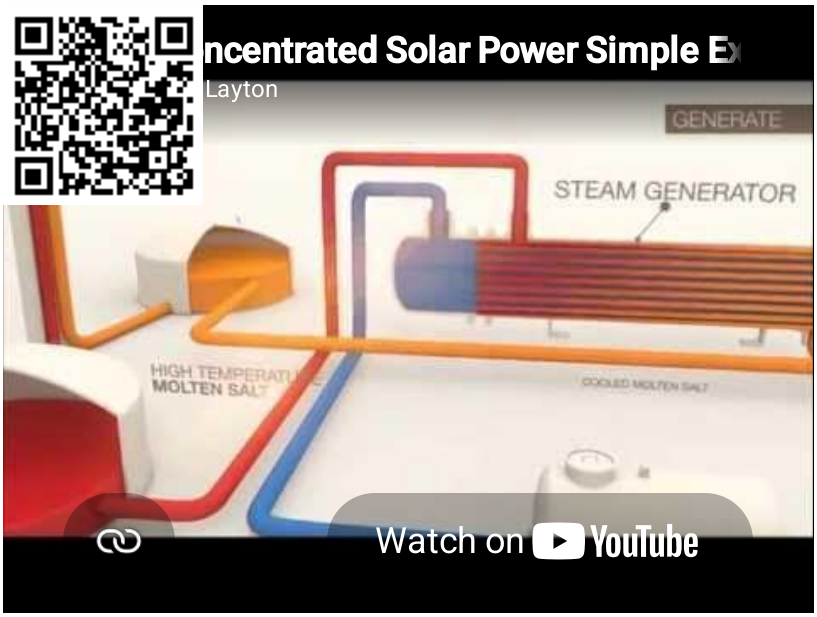
It's worth watching a short but highly instructive YouTube video explaining how the molten salt energy storage system works.

If this thermal energy is released over the period of 10 hours = 36 000 s, the thermal **power** released is 1.33 10^{13}J/36000s = 369 10^6 W = 369 MW. Of **thermal** power – but if we assume 30% efficiency in conversion to electric power, we obtain... 0.3 369 MW = 110.7 MW!

So, it seems that we got too low a result (144 MW) because it was based on estimates that had been "too conservative". Probably, the real story is such that the thermal power deposited at the tower top in Crescent Dunes is not 480 MW, but rather 740 MW – of which one half, 370 MW is converted to electricity right away, and the other 370 MW is stored in the molten salt tanks.

The example of the Crescent Dunes solar power plant points out to a considerable advantage of the CSP technology – the capability of generating power after sunset. With the molten salt storage reservoir, a power plant may be even able to generate electricity at the same power level for full 24 hours. However, the largest tower-type CSP power plant in the US, the Ivanpah Solar Power Facility with three heliostat fields a net power of 392 MW, does not use thermal storage. Other US solar power plants do, although they are not tower-type plants (for the list of the largest CPS plants in the world, please see this Web document).

It's worth watching a short but highly instructive YouTube video explain- ing how the molten salt energy storage system works.



Now, the question is, why do we want to use liquid salts, and not, for instance, metals? The answer is simple – metals which are liquids at the same temperature range as the "solar salt" have much lower specific heat. For instance, for lead (Pb) in its liquid state c = 0.14 J/kg 1˚C, over ten times less than the c of "solar salt", which explains why it's not a good heat storage agent. But the popularity "solar salt" comes from the fact that there exist a huge fertilizer industry producing millions of tons of nitrates per year – therefore, KNO_3 and $NaNO_3$ are widely available. But the "solar salt" is not the only possible mixture of salts with good heat storage properties – there are many other possible combinations, and their properties are reported in many Web publications (e.g., in this one).

7.3.1: Solar Towers Molten Salt Heat Storing Technology

7.3.2: Parabolic Troughs and Fresnel Mirrors

7.3.3: The Linear Fresnel Mirror CSP Technology

7.3.4: CSP – Pros and Contras

---

1. Heat exchangers are described in greater detail in the *Geothermal Energy* Chapter.

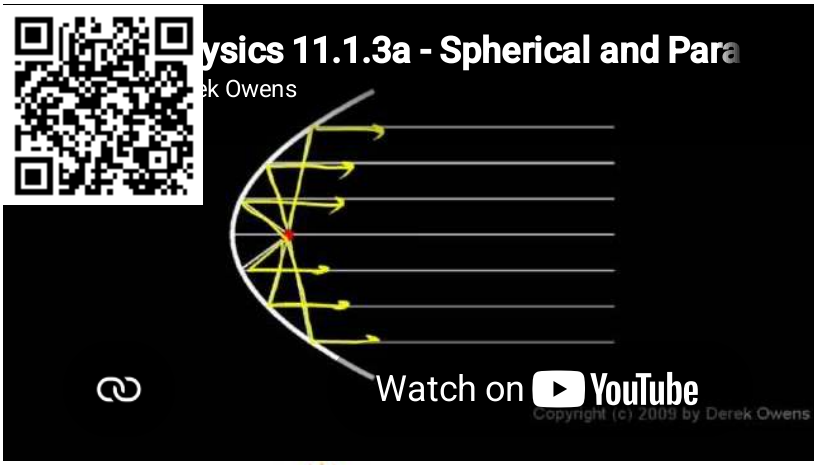
2. Heat capacity, commonly denoted as c, is the amount of heat needed to raise the temperature of a mass unit of a given substance by 1˚C (or, equivalently, by 1 K), or the amount of heat given away by the mass unit of this substance when it's cooled down by 1˚C/1 K.

7.3: The Concentrated Solar Power (CSP) Technology is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.3.1: Solar Towers Molten Salt Heat Storing Technology

In the solar tower CSP technology, all sunlight is focused on a single bulk absorber. An alternative method is to use *linear* absorbers in the form of a long pipes running over a light-reflecting troughs. The geometry of such system is depicted in the Fig. 7.3.1.1.

From introductory physics courses one should remember that a *spherical concave mirror* (one that you get by "cutting a slice of a spherical shell" does not have a perfect focal point. If one needs a mirror that really focuses light rays into a single point, one has to use a *parabolic mirror*. For them who do not remember that material too well, a short video will surely help to refresh their memories.



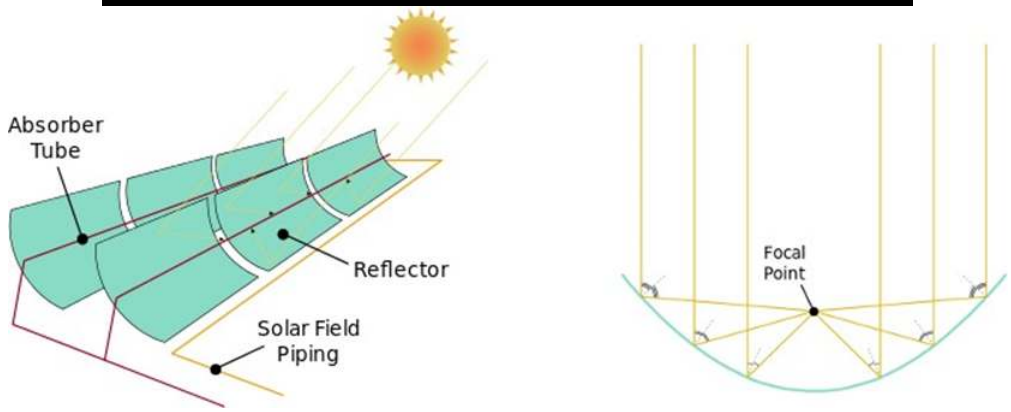


Figure 7.3.1.1: Left: The geometry of a parabolic trough system, with linear absorbers. Right: A cross- section of a trough, with the absorber running along the "focal line" of the parabolic reflector (source: Wikipedia).

There is a similar story with trough-shaped mirrors. To make a simple trough, one may take a thin-walled pipe and cut a "slice" out of it (but not by cutting in the direction perpendicular to the pipe axis – then one gets a ring! – the cutting must be *parallel* to the pipe's axis. From such a operation, one gets a "cylindrical trough". Which is not good, because a concave mirror made of a cylindrical trough suffers the same problem as a spherical mirror: its focus is not perfect. In order to get a trough mirror focusing all incident rays along a perfect line, the trough profile must not be a circle, but a *parabola*. Fortunately, it's not a big deal to make such a trough

– in YouTube one can find many video-instruction how to make a parabolic trough reflector in one's garage (for instance, this one – it's combined with a good "theoretical background"). People use backyard parabolic trough mirrors for a variety of purposes – e.g., for "solar cooking", water heating, or even for running miniature steam turbines to make their own electricity.

In real parabolic trough power plants the absorbers have the form of two concentric pipes. The outer pipe is made of glass. The inner pipe is the actual absorber. There is a small gap between the outer and the inner pipe – it's evacuated so that the absorber tube is thermally insulated by the vacuum layer. Inside the absorber tube a heat-collecting fluid is circulated

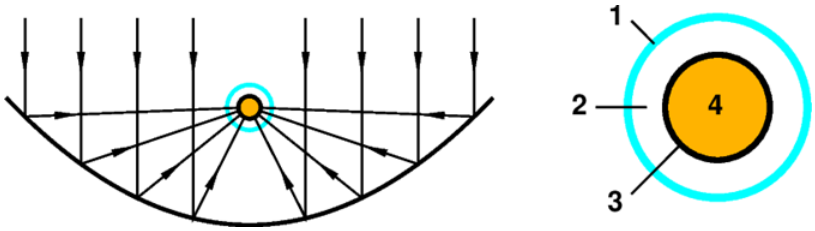


Figure 7.3.1.2: The structure of the absorber tube (a.k.a. the “receiver”). The outer glass tube (1) and the “vacuum jacket” (2) in between prevents contact of the absorber tube with air. If there were not such shielding and the absorber tube had a direct contact with surrounding air, most of the heat deposited to the absorber would be “captured” by air molecules and taken away by convection. There is no way of completely eliminating losses of heat from the absorbers, but the vacuum insulation keeps them at a reasonably low level. Inside the glass tube there is the actual absorber – a metal tube (3) with black outer surface. A cooling liquid (4 – e.g., silicon oil) circulates in the absorber tube and next passes through a “steam generator” where it transmits the heat it has collected to steam which then drives a turbine (aop).

Individual troughs can be combined to make large arrays, covering many acres. Such power plants may have the same heat-storing capability as the tower-type Crescent Dunes plant discussed in the preceding Section.

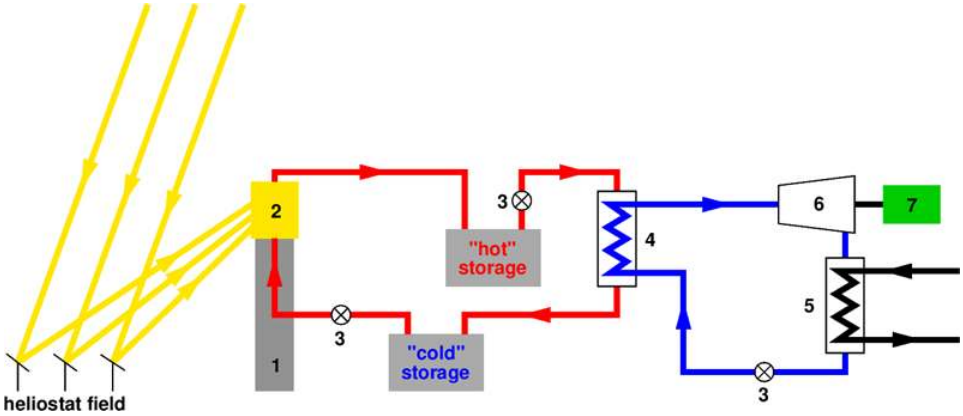


Figure 7.3.1.3: A scheme of a tower-type solar power plant with molten-salt energy storing capability. 1 – the tower, 2 – the receiver, 3 – pumps, 4 – steam generator (i.e., a heat exchanger), 5 – exhaust steam condenser with cooling water runing through it, 6 – steam turbine, and 7 – a generator of “green” electric power (aop).

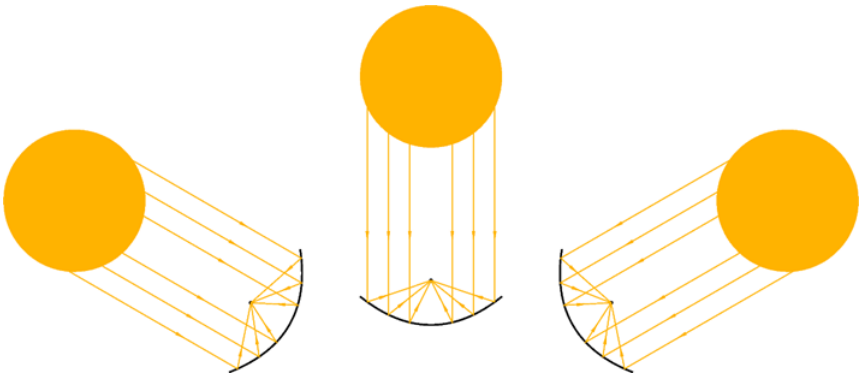


Figure 7.3.1.4: Parabolic troughs may be oriented parallel to the north-south direction – as the sun moves across the sky, the trough symmetry plane’s must remain parallel to the impinging rays, so that the trough inclination varies accordingly (the situation illustrated by the image). Another possible trough orientation is parallel to the east-west direction – then the inclination of the trough’s symmetry plane is set to the angle at which sunlight impinges at noon at the given day – and it remains constant constant throughout the day (aop).

When Sun moves across the sky, in tower-type CSP plants each heliostat has to be individually moved. In the trough-type plants the troughs also have to be moved, but the motion is much simpler. If the long axis of the trough is perfectly aligned with the north-south direction, the only motion needed to always keep the reflected sun rays precisely focused on the absorbers is a “rocking motion”, as shown in Fig. 7.3.1.4. It requires a simpler mechanism than that needed in a heliostat for following the Sun position in the sky. It is also a great convenience that **all troughs in a solar field** must assume the very same position with respect to Sun – so that a single electronic device may control the motion of all troughs. Controlling the motion of ten thousand heliostats in the solar

field of a tower-type CSP plant, where each heliostat has to be individually aligned, must be much more difficult a task! The list of the largest CSP plants in the world can be found in this Web page. As can be seen, in the US the parabolic trough technology is the dominant one, with 1350 MW nameplate capacity installed, whereas the combined power of the tower-type facilities is 517 MW. The proportion may change, though, if the announced plans of building a “mammoth” 1600 MW Sandstone Solar Energy Project are implemented. So, it’s difficult to say about the two technologies, which one is “better”. Probably, as is often the case, the competition between the two technologies will be resolved by market economy mechanisms.

7.3.1: Solar Towers Molten Salt Heat Storing Technology is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.3.2: Parabolic Troughs and Fresnel Mirrors

It's the least widespread CSP technology. As in the parabolic trough CSP technology, a long "linear" absorber is used in it. However, instead of a parabolic trough made of a single piece of appropriately profiled reflecting sheet, it uses an array of long flat light-reflecting "stripes". Each stripe is inclined at a different angle, such that the beam of solar light reflected reflected by it is incident on the absorber.

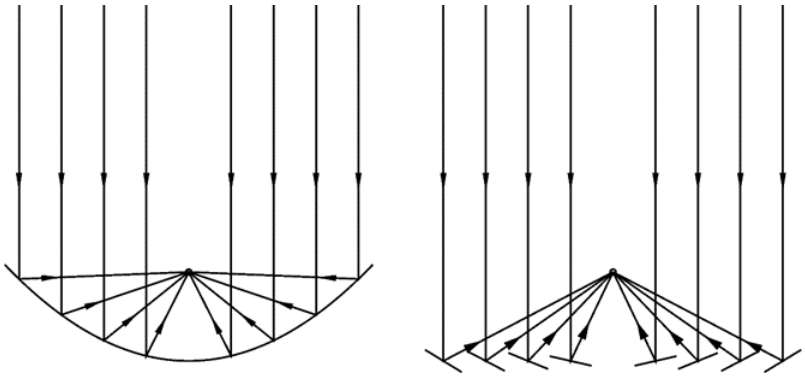


Figure 7.3.2.1: Solar light focused at en absorber in a parabolic trough (left image). If the trough shape is described by a $y = ax^2$ parabola, all the reflected rays intersect at a single line located $1/4a$ above the trough bottom. Another way of focusing reflected sunrays on a linear a bsorber is to use the so-called "linear Fresnel mirror" consisting of a number of flat light-reflecting "stripes" (right image). Fresnel mirrors are less expensive than parabolic troughs, but the latter are more convenient to work with and therefore they are more often used (aop).

A photograph of a large solar field in a CSP plant using linear Fresnel mirrors is shown in this Web page.

When the Sun moves across the sky, each reflecting "stripe" should be rotated at a different angle. So, controlling the rotation is more complicated than in the case of a parabolic trough, but perhaps still much simpler than controlling the alignment of many thousands of heliostats in tower-type CSP plants. Anyway, the Fresnel mirror technology is not very popular – in the list of the largest CSP plants one can find only one such 125 MW plant operating in India, and one under construction in South Africa

7.3.2: Parabolic Troughs and Fresnel Mirrors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.3.3: The Linear Fresnel Mirror CSP Technology

It's the least widespread CSP technology. As in the parabolic trough CSP technology, a long "linear" absorber is used in it. However, instead of a parabolic trough made of a single piece of appropriately profiled reflecting sheet, it uses an array of long flat light-reflecting "stripes". Each stripe is inclined at a different angle, such that the beam of solar light reflected reflected by it is incident on the absorber.

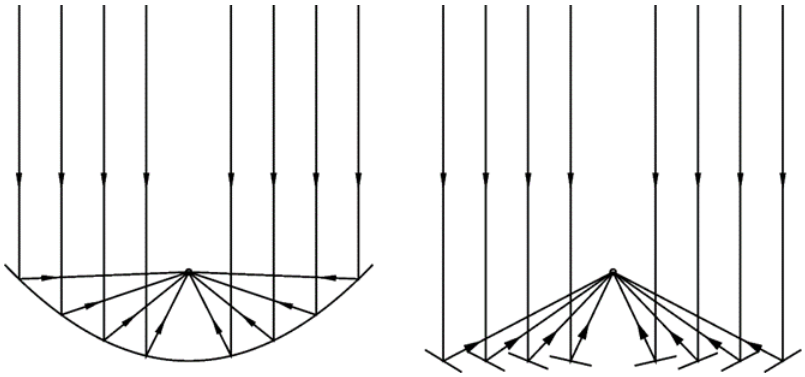


Figure 7.3.3.1:Solar light focused at en absorber in a parabolic trough (left image). If the trough shape is described by a $y = ax^2$ parabola, all the reflected rays intersect at a single line located $1/4a$ above the trough bottom. Another way of focusing reflected sunrays on a linear a bsorber is to use the so-called "linear Fresnel mirror" consisting of a number of flat light-reflecting "stripes" (right image). Fresnel mirrors are less expensive than parabolic troughs, but the latter are more convenient to work with and therefore they are more often used (aop).

A photograph of a large solar field in a CSP plant using linear Fresnel mirrors is shown in this Web page.

When the Sun moves across the sky, each reflecting "stripe" should be rotated at a different angle. So, controlling the rotation is more complicated than in the case of a parabolic trough, but perhaps still much simpler than controlling the alignment of many thousands of heliostats in tower-type CSP plants. Anyway, the Fresnel mirror technology is not very popular – in the list of the largest CSP plants one can find only one such 125 MW plant operating in India, and one under construction in South Africa.

7.3.3: The Linear Fresnel Mirror CSP Technology is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.3.4: CSP – Pros and Contras

Pros:

- Fully “renewable”, carbon-free energy
- Operating costs are low – the “fuel”, the solar light, is essentially for free. The only costs are service and maintenance costs.
- Can utilize thermal storage to better match supply with demand (not possible in PV power plants).
- High efficiency of converting solar power to electric power – about twice the efficiency of the PV installations.

Contras:

- Intermittent – therefore, CSP plants can be located only at regions of exceptionally high number of sunshine hours/year: e.g., in Nevada, California, at major desert areas such as Sahara, or Gobi; in Europe, essentially only in Spain an in Portugal.
- Heavily location dependent (see the preceding item).
- Construction/installation costs are high.
- A single installation requires a considerable amount of space.
- Manufacturing processes often create pollution.
- Environmental concerns in the case of tower-type installations: birds flying to close to the tower (the beams aimed at the absorber are in- visible!) get burned alive.

7.3.4: CSP – Pros and Contras is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

7.4: Photovoltaic Conversion of Solar Power

Photovoltaic (PV) conversion is a direct conversion of solar radiation to elec- tric power. Currently (mid-2018), the installed capacity of PV solar power station is over 400 GW, versus slightly over 5 GW of installed power in CSP solar plants. Hence, the PV method is clearly the dominant one.

7.4.1: Theoretical Background the Nature of Light and the Photoelectric Effect

7.4.2: Semiconductors and Doped Semiconductors

7.4.3: The p-n Junction

7.4.4: Internal Photoelectric Effect in Semiconductors

7.4.5: PV Cells

7.4.6: Solar Cell Efficiency

7.4.7: Inverters

7.4.8: Pros and Cons of PV Solar Power Harnessing

7.4: Photovoltaic Conversion of Solar Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.1: Theoretical Background the Nature of Light and the Photoelectric Effect

How does the PV conversion work? Well, the underlying physics is much more complicated than in the case of CSP conversion. In the latter, things are easy to understand: it's our everyday experience that solar light carries energy, and when it impinges objects, it heats them up. With PV conversion things are more complicated. But if we wanted to outline a detailed theory of PV conversion, not a single Section would be enough – an entire book would be needed. Therefore, out of necessity, we can present only a "simple-minded theory" of PV conversion in this chapter. By a "simple-minded theory" the Author understand an outline in which everything is true, but a number of things have to be accepted without a proof or based only on a simplified reasoning.

We have to start with the microscopic nature of light. The truth is that this nature seems mysterious to many, because in some phenomena light behaves like a typical wave, and in some other phenomena it behaves like a beam of particle-like objects. We call it a "dual nature". Some people riot against such interpretation, insisting that a micro-world object must be either a wave or a particle, because these two "natures" are mutually exclusive. Indeed, they are – but in the *macro-world* we live in. In the *micro-world*, which we cannot observe directly with our senses, because the scale is too small, the rules may be different – and a variety of experiments show that they are different, indeed.

The PV conversion is one of those phenomena in which light exhibits its particle-like nature. The physical effect taken advantage of in PV conversion is a version of the *photoelectric effect* (PE), called *internal*. The other version, the *external* PE, was first discovered by Heinrich Hertz in 1887: he observed that light incident on a metallic plate ejects electrons out of it. It caused a big consternation in the community of physicists, because in 1887 the wave-like nature of light seemed to be firmly established – and a wave **cannot** knock electrons out of a metal! (PAWAP – please accept without a proof). Only after 18 years an explanation of the PE was given by Albert Einstein[1]. Einstein's explanation was based on a courageous hypothesis – namely, that in the PE light behaves like a stream of miniature particle-like objects he called "light quanta" (only some time later people started calling them *photons*, as we still do today). Some "big names in physics" at that time insisted that *Einstein had lost contact with reality*. However, with time, new and new facts supporting Einstein's theory had been discovered, so that eventually it was universally accepted.

The Einstein's theory of photons, or "light particles", still "borrows" something from the wave theory of light. Namely, any wave is characterized by a *frequency*, usually denoted as f. The two other parameters characterizing a wave are the *wavelength* λ and the *speed of propagation* v. These three parameters are related as:

$$f = \frac{v}{\lambda} \tag{7.4.1.1}$$

For light, the speed of propagation is always the same, denoted as c, and it's value is very close to 300,000 km/s. So, the frequency of a light wave is:

$$f = \frac{c}{\lambda} \tag{7.4.1.2}$$

The wavelength λ is different for each light color. And the good news is that λ for a given light color can be readily measured using a simple device called *optical grating* or *diffraction grating* – it's a standard exercise in introductory physics laboratory. And the speed of light c has been measured with a great precision in zillions experiments, so it's value is very well known. Therefore, having measured λ for a given light color, we can readily find the characteristic frequency f of this color.

So, in Einstein's theory what the photons "share" with the light theory is the frequency f . But the entirely new element in the theory is that each "light particle", or each photon carries an amount of energy E_{photon} equal:

$$E_{photon} = h \cdot f \tag{7.4.1.3}$$

where h is an important physical constant, commonly referred to as the *Planck Constant*.

Accordingly, each individual photon carries some energy. It's not kinetic energy, because a photon has no mass – but energy it has.

Now, we have to say something about electrons in metals. In 1905, when Einstein published his revolutionary PE paper, the theory of electrons in metals was not even in "its infancy". However, Einstein correctly assumed that an electron is held inside the metal by some attractive force – therefore, for pulling the electron outside some work has to be done. Einstein called it the *work function*, and attributed a symbol of Φ to it.

What happens, then, if a photon enters the metal[2]? It may interact with an electron and pass its energy to it – i.e., in the process the photon disappears and its energy hf is converted to electron's kinetic energy. After "swallowing" the photon energy, the electron becomes a *photoelectron*. If the photoelectron energy is lower than the "work function" Φ $the extra energy acquired by the electron won't be sufficient to get it out of the metal. But if \(hf > \Phi$, the photoelectron **can** get out and it still retains some of the acquired energy.

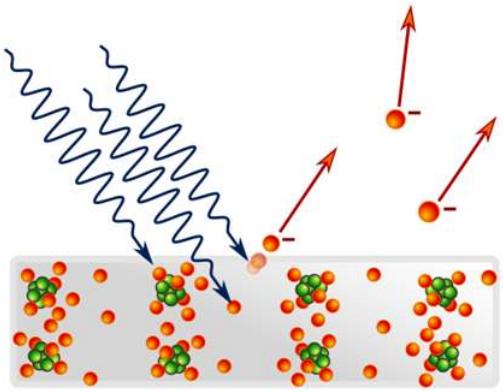


Figure 7.4.1.1: Photoelectric effect in a metal. In metals, the outrmost atomic electrons leave their "parent" atoms and form an "electron gas" (a.k.a. "Fermi gas"). Such electrons may be ejected from the metal by impinging photons of light from the visible or ultraviolet energy range. Atomic electrons from deeper levels are too strongly bound to their parent atoms and cannot be ejected by such photons (source:Wikipedia).

As follow from the above reasoning, the kinetic energy $K_{el.}$ of the photoelectron after it gets out of the metal is:

$$K_{el} = h \cdot f - \Phi \tag{7.4.1.4}$$

It's the famous Einstein's equation for the photoelectric effect. For the *external* photoelectric effect, to be exact – external, because the electron *exits* the metal.

In the Eq. 7.4.1.4, the energy of the photoelectrons should be denoted as $K_{max.}$ – why that "max"? Well, when Einstein derived the Eq. 7.4, he probably assumed that the value of Φ is the same for all electrons in the metal. However, as it turned out later, there is a distribution of ψ values for different electrons. What we now call Φ, is the *minimum value of the work function* for a given metal. Hence, in the case of Potassium, the work needed to "pull out" an electron from the metal may be *2.0 electron-Volts, or more*. Therefore, the $K_{el.}$ parameter in the Eq. 7.4 should be thought of as the *maximum photoelectron energy*, and the photoelectron velocity corresponding to this energy should thought of as *the maximum speed of the photoelectron*.

How can the energy of photoelectrons be investigated? The apparatus needed for that is very simple.

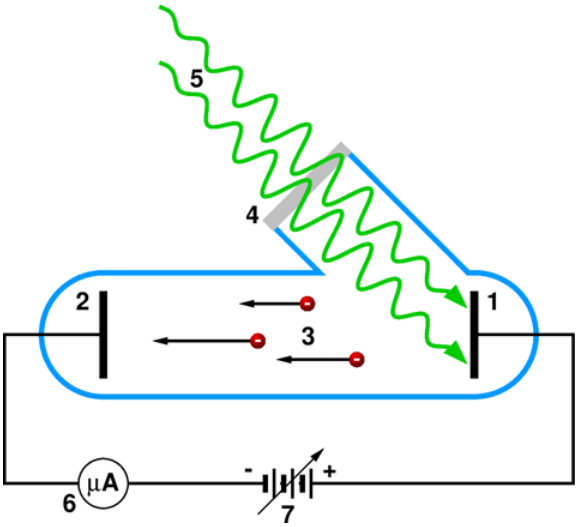


Figure 7.4.1.2: The apparatus used for investigating the photoelectric effect. Inside an evacuated glass tube (blue line) there are: 1 – cathode made of the investigated metal; 2 – anode, the electrode collecting the photoelectrons; 3– photoelectrons; even if all incident photons have the same energy, the kinetic energy of all photoelectrons is not the same – the famous Einstein's equation describes only the **maximum** energy. 4 – a quartz window; in contrast to glass, quartz is transparent for ultraviolet (UV) light, so that such window makes it possible to study the effect for UV light. 5 – the photons incident on the catode; 6 – a sensitive current meter; 7 – a variable voltage source (aop).

The electrode made of the metal investigated, called *cathode*, or *photocathode*, is placed in an evacuated glass tube, together with another electrode called the *anode*. The two electrodes are connected to a variable voltage source, enabling one to change the potential between them. Also, there is an ammeter in the circuit for measuring the current of photoelectrons, usually referred to as the *photocurrent*.

If we want to determine the value of Φ, we connect the *negative* terminal of the voltage source to the anode, and the positive terminal to the photocathode, and we start with a zero voltage. After the photocathode is illuminated, photoelectrons are ejected from it. They have certain kinetic energy, so they reach the anode – and so, a current is flowing. Now, we start gradually increase the voltage – i.e., making the anode *more and more negative*. A negative electrode repels the electrons. So, as the voltage is increased, less and less photoelectrons have their energy high enough to reach the anode, and the current gradually decreases. At some negative voltage U_{max}, the current completely stops flowing.

Now, we only need to recall the definition of the unit of *electron-Volt* : *One electron-Volt, the eV, is equal to the gain or to the loss of kinetic energy of an electron passing trough a potential difference of 1 Volt*. Accordingly, if the fastest photoelectrons are stopped by a potential difference of U_{max}, it means that their energy in the units of eV is equal to theU_{max} in the units of Volts. Very simple, indeed – we get the K_{el} in Eq. 7.4 just by reading U_{max} from a voltmeter. And the value of hf for the light we use we know from other measurement. Then, we get the value of Φ by solving the Eq. 7.4: $\Phi = h \cdot f - K_{el}$.

Internal Photoelectric Effect

All the discussion above was about the *external* PE. But, as mentioned, there is also an *internal* PE. What's the difference? Well, if the energy $h \cdot f$ of photons is lower than Φ,, any photoelectrons created by such photons have no chance to exit the material. Such is the situation in the case of the red light in the Fig. 7.12. If the red light produces any photoelectrons in Potassium metal, they will stay inside and they will blend into the "crowd" of electrons that are are already there. So, since there is no chance to observe such photoelectrons, there is no much interest among physicists in the internal photoelectric effect in metal.

But there is another family of materials in which such internal photoelectrons do produce conspicuous effects – namely, in semiconductors. What's important for us, semiconductors are the materials from which PV panels are made. And the internal PE plays a crucial role in the electricity generating process in such panels.

But before we start explaining **how** the internal PE is taken advantage of in the generation process, we have to explain what semiconductors are, what is the "doping" of semiconductors, and how doped semiconductors conduct electric current.

1. It was what for Einstein was awarded with a Nobel Prize in 1921 – not for the relativity theory, as many people wrongly believe.

2. Light is reflected from a metal surface, but the process involves penetration at a depth of the order of nanometers.

7.4.1: Theoretical Background the Nature of Light and the Photoelectric Effect is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.2: Semiconductors and Doped Semiconductors

There are many different semiconductor materials. But most semiconductor devices used today are based on silicon, and almost all PV solar panels in the world are made of silicon[1]. Therefore, we will focus our attention on silicon only.

Silicon (Si) is the fourteenth element in the Periodic Table, which means that its nucleus contains 14 protons. So, to stay neutral, an Si atom also contains 14 electrons. Four of them are the *valence electrons*, i.e., those that form the outermost electron shell of the atoms. The valence electrons are used by atoms to make chemical bonds with other atoms (of different elements, or with atoms of the same element). There are several types of chemical bonding. One of the is the so-called *covalent bonding*, in which pairs of electrons are shared by two atoms. In silicon crystals, each Si atom has four other Si atoms as its nearest-neighbors. It's coupled to each nearest- neighbor by a covalent bond, to which each of the two contributes one valence electron. A simplified scheme of the arrangement of atoms and the valence electrons in a Si crystal is shown in the Fig. 7.4.2.1. The scheme in the figure is largely simplified, because in a Si crystal an atom and its four nearest neighbors *do not* lie in the same plane: an atom is positioned in the center of a *tetrahedron* formed by it's four nearest neighbors. Yet, there is no way of graphically presenting a 3-dimensional (3-D) arrangement of atoms on a 2-D piece of paper, or a computer screen – so, out of necessity, in the figure the 3-D actual arrangement is replaced by a 2-D square lattice, in which there is the same number of nearest neighbors as in the real 3-D crystal.

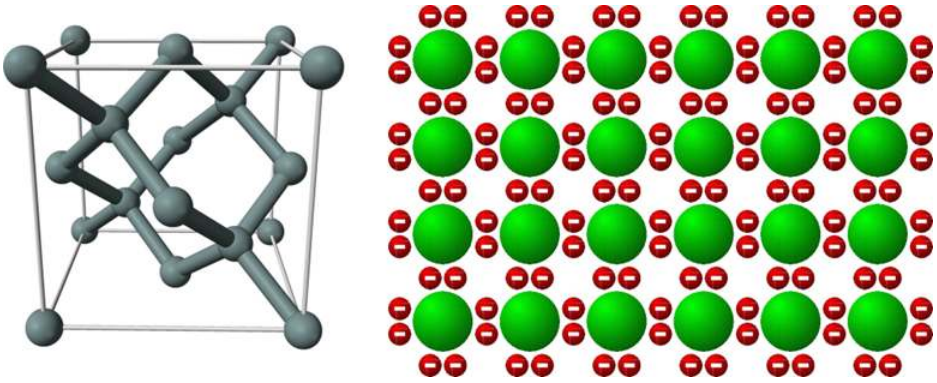


Figure 7.4.2.1: Left image: In solid silicon the atoms are arranged into a so-called "three-dimmensional tetrahedral lattice", in which each atom is coupled with four nearest neighbors that form the vertices of a tetrahedron. The atoms are coupled by **covalent bonds** – i.e., every pair of nearest neighbors shares a pair of electrons. If one attempted to present such a situation graphically by adding electrons to the left-side image, it would become really messy – therefore, it's better to use an approximation and plot instead a two-dimensional square array to show a square array as this in the right-side image: here the green spheres represent the "cores", i.e., atoms "stripped" from their four outer electrons the four outer electrons – which are shown as small red spheres marked with a minus sign. The pairs of electrons in between Si atoms represent symbolically the covalent bonds (sources: left image – Wikimedia Commons, the right one plotted by the author).

By the way, the spatial arrangement of atoms in a Si crystal is identical as the arrangement of carbon (C) atoms in a diamond crystal – therefore one can often find the following statement in literature: *silicon crystallizes in the diamond structure*.

In a pure Si crystals the electrons forming the covalent bonds are "firmly anchored" to their "parent atoms" and are very reluctant to move. Therefore, pure Si is a very poor conductor of electric current. The same is true for other semiconductors. They are not insulators, but their electric conductivity is orders of magnitude lower that the conductivity of metals. In logarithmic scale the conductivity of semiconductors is located more or less *half way* between the conductivity of insulators and of metals – it's where their name comes from, *semi-* meaning "half" in Latin.

However, there are ways of greatly enhancing the conductivity of silicon (and other semiconductors, too) by an artificial method known as *doping*. There are methods of substituting a small number of Si atoms in a crystals by other elements, then called *dopants*, or *doping agents*. A doping agent often used in silicon is phosphorus (P), the 15th element in the Periodic Table. It has then a total of 15 electrons, of which *five* are valence electrons. If a P atoms is placed in between four Si neighbors, four of its electrons are paired with the neighbor electrons to form covalent bonds – and the fifth phosphorus electron has no electron to get paired with, so it remains only weakly coupled with its "parent" P atom. Because of the thermal vibrations of atoms in the crystal, such electron gets easily "de-coupled" from its "parent" and can migrate away, becoming an *mobile electron*. And it's mobile electrons which are needed in a material to conduct electric current through it. So, silicon doped with phosphorus becomes a conductor – not as good a conductor as a typical metal (e.g., Cu, or Al), but much-much better than a pure silicon.

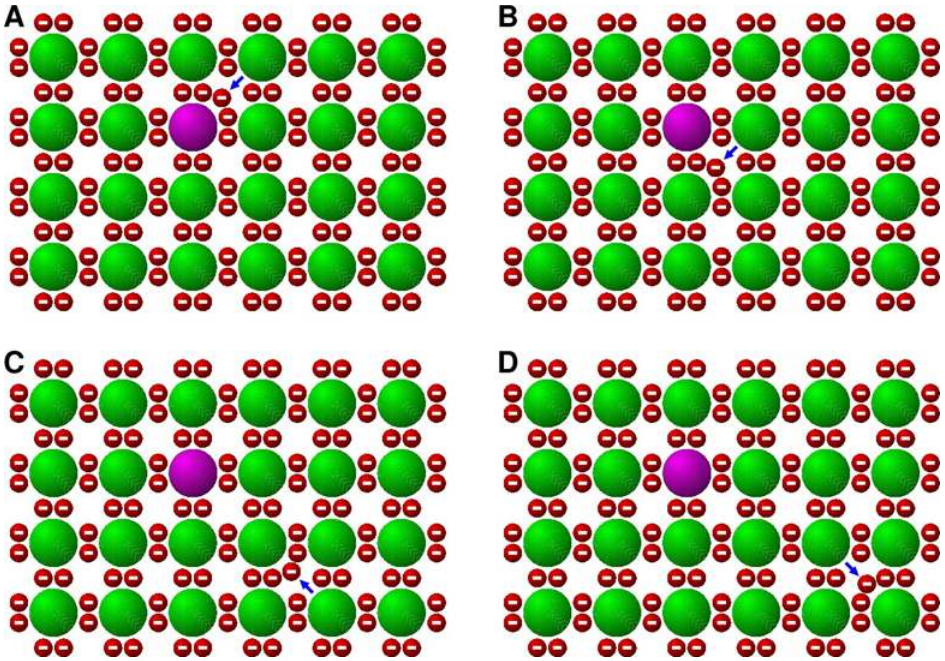


Figure 7.4.2.2: A model of n-type silicon, i.e., one in which a small number of silicon atoms are replaced by atoms of an element with five external electrons (e.g., by atoms of Phosphorus, P, shown as the magenta- color sphere). Of the five, one electron (indicated by the small blue arrow in image **A**) cannot participate in covalent bonding. It's only weakly coupled with its "parent" atom and may gradually migrate away from it, as shown in the images marked as **B**, **C** and **D** (source: aop).

The phosphorus atoms in silicon are called *donors*, because they donate extra electrons to the crystal. Silicon doped with phosphorus is referred to as *n- type* – because it contains extra *negative* current-carrying particles (mobile electrons, it means). However, let's keep in mind that the crystal as a whole remains neutral, because the charge of those "extra electrons" is balanced by the extra positive charge of the "parent" P atoms.

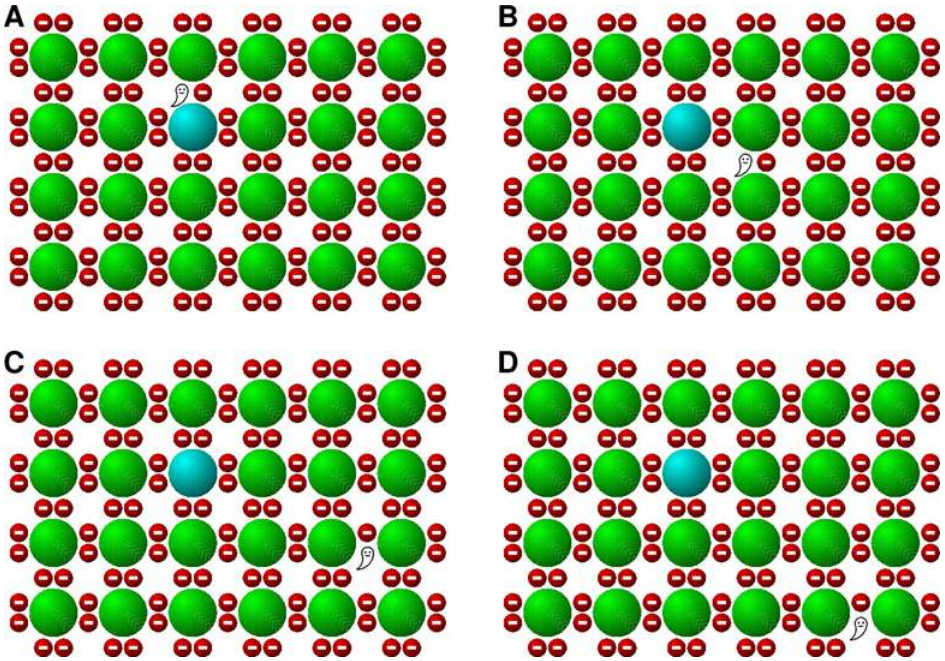


Figure 7.4.2.3: A model of p-type Si, in which a small number of silicon atoms are replaced by atoms of an element with only **three** outer electrons (e.g., by Boron, B, shown as the cyan-colored sphere). Then, in one covalent bond there is only a single electron – where the other should be, there is a "hole". It is not easy to draw emptiness, something that does exist and at the same time does not exist – therefore, we used a small ghost as a symbol of such hole (ghosts do not exist, by they also do exist, right?). Like an extra electron in n-type Si, a hole may also migrate from it's "parent atom". Thus, it can act as a positive current carrier (source: aop)

Interestingly, this is not the only way of making Si conducting. One can dope the Si crystal with atoms which have not 5, but 3 valence electrons. For instance, with boron (B), the fifth element in Periodic Table. It has a total of five electrons, and three of them are valence electrons. So, when a B atom sits in between four Si neighbor atoms, only three covalent bonds can be fully formed – in the fourth, one electron is missing. Such an incomplete bond is called a *hole*. But such hole is not firmly "anchored" to the B atom– it may start migrating from its "parent": it may exchange places with an electron from a nearby covalent bond, then with an electron form another bond, and so on – as is illustrated in Fig. 7.4.2.3. The effect is such as if a *positive charge* were wandering around the crystal lattice. As electrons in an n-type crystal, such holes can also carry electric current across the crystal. Because the current carriers are positive, such crystals are referred to as *p- type* ones. And because the B atoms effectively "capture" electrons, they are called *acceptors*.

Again, as in the case of n-type silicon, it should be stressed that the crystal as a whole remains neutral, even though it contains positive holes – because their positive charge is neutralized by the negative charge of the electrons "trapped" by the boron acceptors.

---

1. with the exception, e.g., of the PV panels supplying power to spacecrafts like the Skylab – in order to achieve the highest possible efficiency, they are made of combinations of several semiconducting materials; but they are only a small fraction of all PV panels manufactured globally

7.4.2: Semiconductors and Doped Semiconductors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.3: The p-n Junction

Suppose that a Si crystal plate is p-type on one side, and n-type on the other, and at some depth below the surface the two types get in contact – what's created then is a so-called *p-n junction* (some people prefer calling it *n-p junction*, which is essentially the same). The interface area should be very thin, with thickness of the order of 1 micrometer or less. Such p-n junctions play extremely important roles in modern semiconductor electronics. Any piece of modern electronics, such as a cellphone, a PC, or the GPS in your car, they all contain millions of p-n junctions. The p-n junction is also the "heart" of every PV solar power converter.

Let's first discuss what happens to the loose electrons and holes roaming around in the n-type and p-type areas on both sides of the p-n junction. They exhibit a tendency of doing the same as any particles suspended in any kind of medium normally do – namely, of diffusing, or spreading over a larger and larger area. This tendency is limited by the crystal boundaries – the only possible direction is *across the p-n interface*. And so they do!

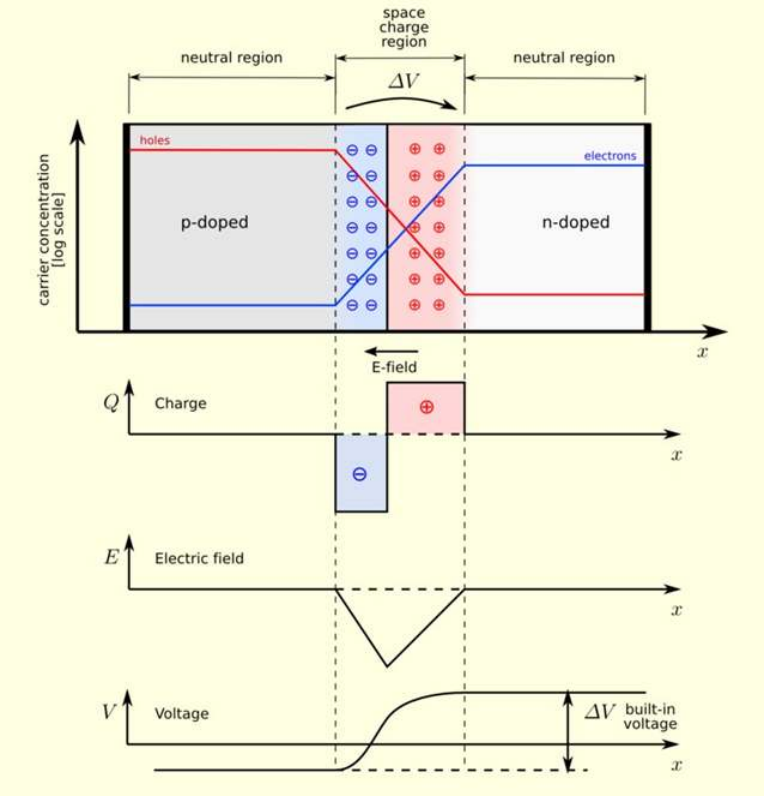


Figure 7.4.3.1: The creation of a space-charge region in a *p n* junction due to the diffusion of electrons and holes across the *p n* interface. The existence of the space-charge gives rise to an electric field, and creates a potential steps, commonly referred to as the "built-in voltage" (more details in the text; source: Wikipedia).

To explain the effects occurring in p-n junction we will use a concise version of the explanation outlined in a Wikipedia article, as well as an instructive graph from this article, copied into the Fig. 7.4.3.1. The physical process that determines the properties of a *p-n* junction is diffusion: some electrons from the n-type block penetrate some distance into the p-type block – and vice versa: some holes from the p-type block penetrate some distance into the n-type block. As noted, the *p*-type and the *n*-type materials are normally electrically neutral – so, due to the hole diffusion, a positive electric charge is accumulated on the right side of the *p n* interface; and due to the diffusion of electrons, a negative charge accumulates on the left side of the interface, as shown in the *Q*-plot in the Fig. 7.4.3.1. The existence of regions of opposite electric charges, some distance apart, gives rise to an electric field, pointing in the direction from the positive region towards the negative region. The magnitude of the *E* field as a function of *x* is shown in the *E*-plot in the figure. Note it's always negative – it's because the field vector is pointing to the left, which is the "negative" direction in the graph.

The last plot in the Fig. 7.4.3.1 shows the electric potential as the function of x. If there is an electric field parallel to the x direction, there is always electric potential difference between two different points, x_1 and x_2. The basic relation between the magnitude of

electric field parallel to the x axis and the x-dependent potential $V(x)$ is:

$$E = -\frac{dV}{dx} \tag{7.4.3.1}$$

Hence, the potential difference - commonly referred to as the voltage - can be obtained by integrating:

$$V(x_2) - V(x_1) = -\int_{x_1}^{x_2} E(x)dx \tag{7.4.3.2}$$

The $E(x)$ function in the present case is a triangle, and it's easy to show that an integral over such triangle function is of the shape of a "rounded step", as shown in the last plot in the Fig. 7.4.3.1.

The potential step existing in the $p \to n$ junction is commonly referred to as the "built-in voltage". And this voltage step is exactly what allows a pn junction to act as a converter of solar energy to electric power. But to explain how exactly such conversion happens, we should return for a moment to the internal photoelectric effect.

7.4.3: The p-n Junction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.4: Internal Photoelectric Effect in Semiconductors

As noted, there is not chance to observe internal photoelectric effects in metals. The concentration of mobile electrons in metals is very high, of the order of $10^{21} - 10^{22}$ per one cubic centimeter. A relatively small number of photoelectrons generated inside the metal would not change it properties in an observable way.

But in semiconductors the situation may be quite different. Think of a pure semiconductor material, such as that shown in the Fig. 7.15. Essentially, all electrons in such crystals are "employed" in forming the covalent interatomic bonds. In order to "rip out" an electron from a bond some energy Phi is needed (in analogy to the energy needed to eject an electron from a metal in the external photoelectric effect). But no such energy is available – so, effectively, the electrons are "immobilized", none are available to carry electric current across the crystal.

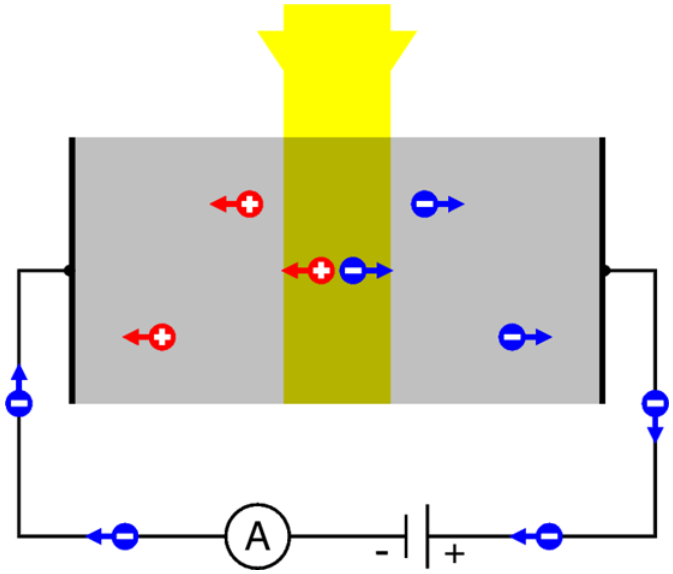


Figure 7.4.4.1: Internal photoelectric effect in a semiconductor: light, penetrating the material, creates *electron-hole pairs*. If a voltage is applied across the crystal, the electric field separates the pair – negative electrons drift towards the positive electrode, while the positive holes towards the negative electrode. Electrons can enter the metallic electrode and they continue flowing in the wires making the circuit. The holes cannot move in metals, so at the negative electrode they *recombine* with incoming electrons. The current in the circuit can be measured by an ammeter (source: aop).

But the situation may change if the crystal is illuminated by photons of energy hf higher than Φ. Then, the *internal photoelectric effect* may occur. And if an electron is "knocked out" of a covalent bond, it leaves behind an incomplete binding – which is nothing else that a *hole*. In other words, what happens is that a photon disappears, but in the process it creates a pair – an electron, and a hole. And both can migrate away from the spot where the event took place.

The effect can be detected, if we apply voltage across the crystal using two metal electrodes attached to opposite crystal sites. It's shown schematically in the Fig. 7.4.4.1. The electrons and the holes are pulled in opposite directions: the negative electrons migrate towards the positive electrode, and the positive holes towards the negative electrode. The electrons can enter the electrode and then can flow down the wire to the voltage source (the battery) – and then[1], through the ammeter to the negative electrode. Here the electrons meet the holes flowing from inside the crystal. Holes cannot flow down a metal wire – so they *recombine* with the electrons. A *recombination* is an opposite process as compared with the *pair creation*[2].

The current flowing in the circuit (referred to as the *photocurrent*) can be measured by a sensitive ammeter. It is proportional to the intensity of the incident light – and therefore such a method of internal photoelectric effect detection is used – e.g., by photographers – for measuring light intensity in simple devices called "photometers". Most often, The semiconducting material used in inexpensive photometers of such type is Cadmium Sulfide (CdS) – such CdS "photoelements" could be purchased in *Radio Shack* stored when they still existed, and people used them to build their own amateur light intensity detectors. Why CdS? Because the Φ value in it is exceptionally low, so the internal photoelectric effect may occur in this material for low-energy photons, such as those of red light or of infrared light.

1.Let’s keep in mind that electric current never can *enter a battery and to “stay inside”*. The battery acts like a pump, it only “pulls in” electron through one terminal, and makes the same number of electrons to leave through the other terminal. A “voltage source” is not a source of electron, it only “energizes” electrons passing through it.

2. A photon creates an electron-hole pair, and disappears. An opposite process may be a “mirror image”: a hole and an electron meet and “re-recreate” the covalent bond, and the energy released in the process “re-creates” a photon. This is called a *radiative recombination*, and such a process occurs in the well-known *light emitting diodes (LEDs)*. However, more often the recombination is *non-radiative*, the energy released by the recombining pair is “utilized” in a different way by the crystal.

7.4.4: Internal Photoelectric Effect in Semiconductors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.5: PV Cells

A silicon PV cell is a thin (0.5 - 1 mm) wafer of *p*-type Si, on the top of which there is a thin layer of *n*-type Si. So, a short distance below the illuminated surface there is an *np* junction. The photoelectrons generated leave the cell through the surface, and return through the surface of the "dark side". There must be an electrode on the top surface to collect the photoelectrons – but such electrode would obscure the incident sunlight. Therefore, it's made as a grid of parallel thin wires. There is no such problem with the bottom electrode through which the photocurrent returns to the cell – it may cover the entire surface.

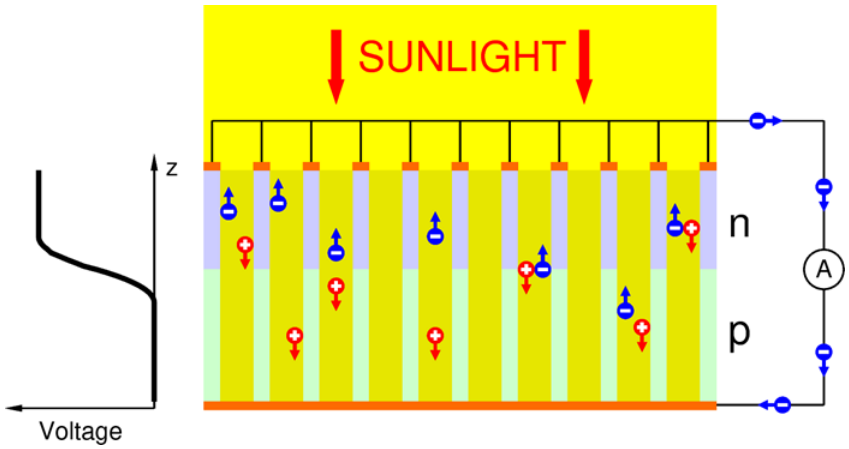


Figure 7.4.5.1: A scheme explaining how a photocell works. Sunlight enters through the spaces between the wires forming the top electrode (their width is much exaggerated in the plot). The *n p* junction is a short distance below the surface. The sunlight photons create electron-hole pairs, which are driven apart by voltage – electron towards the surface, holes towards the bottom Where does the voltage comes from? This is all the ingenuity of the design! The voltage employed for separating electrons from holes is the "built-in" voltage of the *n p* junction! (as shown by the graph on the left). There is no need to apply an external voltage source, like in the Fig. 7.19 – the current flows "all by itself". So, an illuminated PV cell becomes a current source. The output voltage is close to the "built-in" voltage step, typically 0.6 Volt (source: aop).

The voltage from a single cell is far too low for applying it to some practi- cal purposes. Therefore, in PV panels several tens of single cells are connected in series to deliver a higher voltage. For instance, a typical panel of about 25 inches by 54 inches size contains 36 cells connected in series to deliver about 21.5 Volts – when no current is taken from it. As in all other types of current sources, the output voltage drops when current is taken from them. An important question is *how much the voltage drops* in dependence of the current (Amps) flowing through an external load. Conventionally, for pre- senting such information about a given current source one uses a plot called a with the volt scale on the abscissa (horizontal) axis, and the amp scale on the ordinate (vertical) axis. A typical shape of the *I-V characteristic* of a single silicon solar cell is shown in the Fig. 7.4.5.2. It's interesting to notice that for most of the characteristic the I-V curve is almost flat (professionals say: *there is a* ***plateau*** *region in the characteristic*) – which is easy to under- stand: the output current is proportional to the number of photoelectrons leaving the cell in one second – and the number of photoelectrons generated every second is proportional to the number of photons reaching the cell every second.

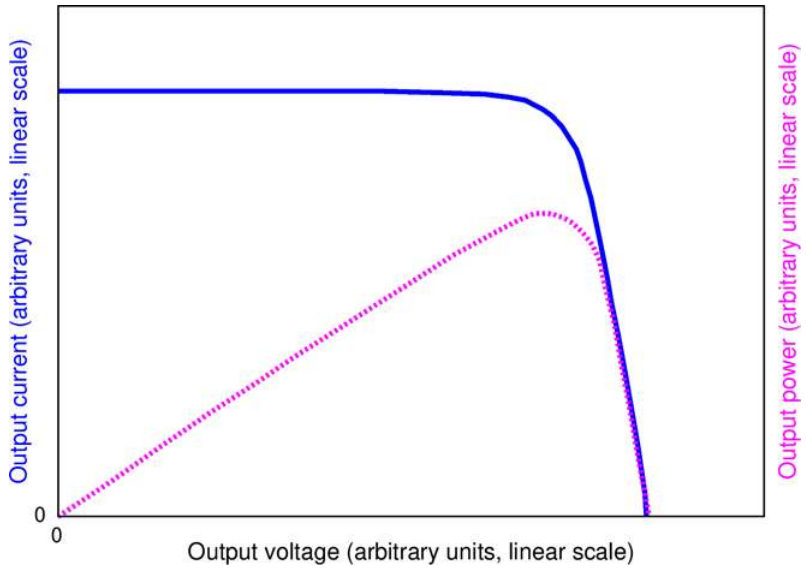


Figure 7.4.5.2: A typical shape of a *constant illumination* I-V characteristic of a single solar cell (the blue curve). The output voltage V when the cell delivers no current (called Voc, where "oc" stands for "open circuit") is 0.6 V. The magenta-color curve is the output power as a function of the output voltage. The power P delivered to an "external load" (e.g., such as a bulb) by electric current is simply $P = V\,I$, the voltage across the load times the current flowing through it. So, for each point of the blue curve, the corresponding power is calculated by multiplying the abscissa (volts) by the ordinate (amps). The voltage corresponding to the maximum of the power curve, V_{MP} is the output voltage at which a **maximum power** is delivered to an external load (source: aop).

The power curve at the I-V characteristic of a power cell has a distinct maximum, usually for output voltage equal to about 80% of the open circuit voltage V_{oc}.

In Fig. 7.4.5.3 typical I-V characteristics of PV panels for several different incident sunlight intensities are shown.

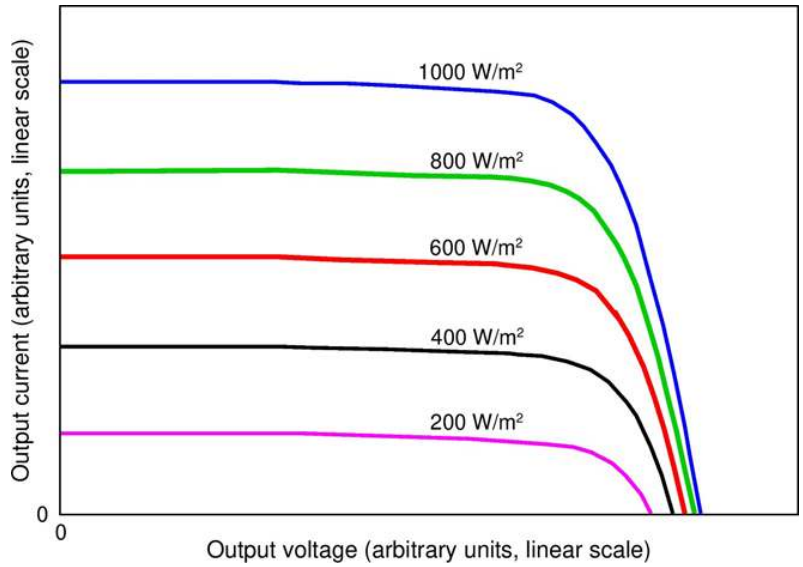


Figure 7.4.5.3: Typical IV characteristics for a PV panel for different intensities of the incident solar light (source: aop).

From the figure one can see that the output current always exhibits a *plateau*, with the current magnitude roughly proportional to the light intensity. The V_{oc} slightly drop with decreasing sunlight intensity as also the maximum power voltage does, but the maximum power essentially is proportional to the incident light intensity. So, a solar panel works not only on sunny days, it can also deliver power on cloudy days. Obviously, much less, in handbooks with instructions for users of PV installation they talk about 10% - 25% of their rated capacity.

7.4.5: PV Cells is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.6: Solar Cell Efficiency

Solar cell efficiency is the portion of sunlight energy that can be converted into electricity.

Let's begin our discussion with a description of the spectrum of solar light. It can be divided into three regions:

- *infrared (IR)* – light with wavelength λ longer than 700 nm, or photons with energy lower than 1.77 eV;
- *visible* – light with wavelength between 700 nm (red color) and 400 nm (violet color) – or with photon energy between 1.77 eV and 3.10 eV;
- *ultraviolet (UV)* – with wavelength shorter than 400 nm, and photon energy higher than 3.1 eV.

In solar light, 46% of power is carried by IR photons, 47% by visible light photons, and 7% by UV photons.

Different light wavelengths may contribute in a different way to the output power of a silicon cell. In Si, the minimum energy needed to produce an electron-hole pair (we called that energy as Ψ in Section 3.4.4) is 1.14 eV. It means that all photons with energies lower than 1.14 eV, and wavelength longer than 1100 nm are not absorbed by silicon, and they take no part in electricity generation in a Si PV cell.

But high photon energy offers little advantage in current generation. In the process of pair creation the photon disappears. 1.14 eV of its energy has gone to "knocking out" the electron from the covalent bond – and what happens to the remaining energy? Well, it's converted into kinetic energy of the electron and the hole. But what determines the output power of the PV cell is the **number** of electron-hole pairs produced, not their energy. The extra energy of a pair is quickly lost in collisions with atoms. So, of the 3.1 eV of a violet light photon energy only 1.14 eV does "useful work", whereas the rest, 1.96 eV or 63% of the the photon energy, is wasted.

What's described above is not the only factor contributing to lowering the efficiency of PV cell. But we will not discuss the topic any further because the calculation of the maximum theoretical efficiency of a PV cell is a complicated task. The Author would rather refer the the Readers who are specifically interested in the topic to other sources, such as, e.g., this Wikipedia article, and references therein. The Wiki article presents in greater detail the results reported by Shockley and Queisser in a 1961 paper. These authors, trough thorough calculations, determined the theoretical efficiency limit for single *np* junction[1] solar cells made of different materials. For instance, for a silicon cell such as the one discussed in the preceding Section, the Shockley- Queisser theory predicts a maximum possible efficiency of 33.2%.

Practical Efficiency

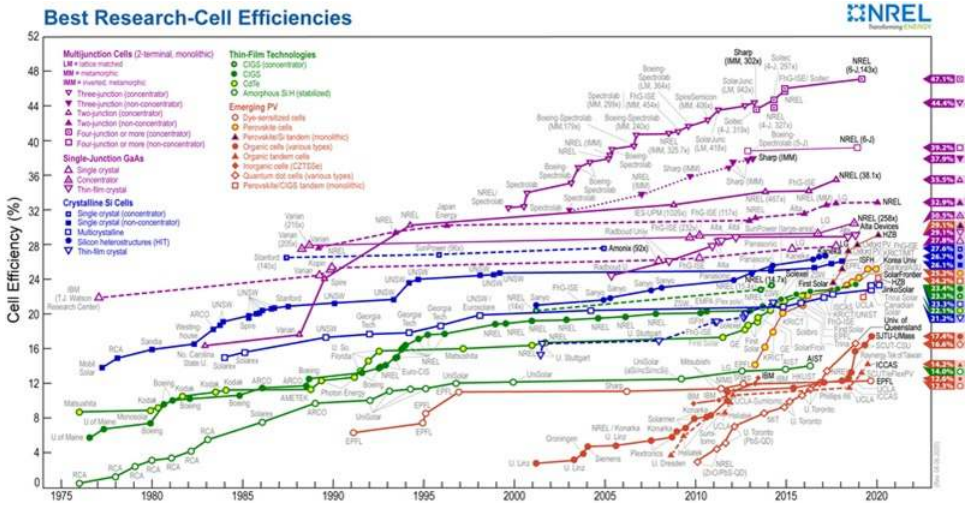


Figure 7.4.6.1:The best research cell efficiencies since 1976 – a chart compiled by the National Renewable Energy Lab (source: NREL, via Wikimedia Commons). The chart is periodically updated, the version shown is from April 2020. The image reproduced here is, for obvious reasons, only a low-resolution copy. But it is highly recommended to take a look at the high-resolution original, which can be downloaded from this NREL Web page.

But for ordinary users of PV cells what really matters is the *practical efficiency*. For the mass-produced Si cells this figure is about 15%. But there is a continuing progress in the field. Competing manufacturers are devoting much effort to get closer to the

Shockley-Queisser limit. For more money one can now purchase panels with efficiency over 20%. A list of the most efficient products on the 2018 market is given, e.g., in this Web page. It's also interesting to watch a graph showing, for the period since 1976, how the efficiency of various types of PV cells improves with the passage of time. The graph, originally published by the National Renewable Energy Laboratory, is presented above as the Fig. 7.4.6.1

The plot looks like the proverbial "spaghetti plate", but it's worth spending a few minutes to examine different curves. It can be found in many different Web pages which enable the reader to study a blown-up version on the screen – for instance, in this Wiki article (which has already been quoted here) in the section *Comparison*.

It should be noted that the NREL chart shows the best results obtained in research labs – the efficiency of mass-manufactured products is always lower. However, one can see that the efficiency of single crystal Si cells – the most popular type today – in some research was as high as over 27%. One can therefore expect that an efficiency over 20% may soon become a standard for mass-produced PV panels.

In this Chapter, we have not mention yet some new emerging technologies, such as organic PV cells, or cells made of semiconductors of *perovskite* structure. Such materials may offer a substantial reduction of manufacturing costs. The new emerging technologies have not yet entered the market – but, in the future, who knows? So, for those of you who are interested in solar power and think of becoming a solar power user in the future, it is recommended to observe the news related to news related to new technologies of harnessing solar energy.

Going back to efficiency: it should be noted that efficiency is not always a critical factor. If a PV panel generates an amount of power that does not satisfy its owner, one possible solution of the problem may be replacing the panel with one of a higher efficiency. But a simpler (and probably a cheaper) solution may be adding more panels of the same type.

There is a clear practical difference between the efficiency of a thermal engine, and the efficiency of a solar PV panel. Consider a conventional coal- burning power plant. The efficiency of an average facility of such kind is about 30%. It means that if the plant uses, say, 100 tons of coal, only the energy released by burning 30 tons of coal is converted to electric energy – and the energy released by burning 70 tons of coal is wasted (it is spent mostly for heating river water or ambient air in the cooling system, as required by the Second Law of Thermodynamics). However, one needs to pay not only for those "good" 30 tons, but also for the 70 tons from which the energy is wasted, and nothing can be done to prevent it.

In the case of solar power things look slightly different. If, say, 15% of solar power is converted to usable energy, 85% is wasted, right? But almost all solar power reaching the Earth surface is "wasted" in a similar way! Nobody needs to pay for that 85% of solar power which is "wasted" due to the not-very-high conversion efficiency of a solar cell! It's all for free! Not enough power? Just add more panels! Well, this is not for free, but it is a one-time investment. After the period of amortization – which is not very long at the current prices – there will be no need to pay for the power, the "usable" as well for the "wasted".

---

1. By using sophisticate techniques, it is possible to prepare PV cells in which solar light passes through more than one *np* junction – each made of a different semiconductor material. Such "multilayered" PV cells may attain practical efficiency as high as 46%. They are very expensive, so they are used mostly when there is a "dramatic need" for power – e.g., in the Skylab, or in spacecrafts sent to distant planets in the Solar System.

7.4.6: Solar Cell Efficiency is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.7: Inverters

The PV panels generate electric power – and what next?...

A homeowner can use the power at home only, or sell it to the utility company. But the electricity, how it flows out from the panels, is not good for any of the two purposes. The thing is that all American home installations deliver 110 V (essentially, all wall outlets, and all lightening delivers/uses such voltage) or 240 V (used by kitchen ovens, or, e.g., well pumps) – and, which is also important, of **alternating current**. If, say, at some moment the right slot in the wall outlet is *plus*, and the left slot is *minus*, then after 1/120 second the left is *plus* and the right is *minus* – and after another 1/120 second again the right is *plus* and the left is *minus*. So, the voltage oscillates with a period of 1/60 second, or with the frequency of 60 Hertz.

On the other hand, if the output voltage from a solar panel is delivered by two wires, then one of them is always *plus*, and the other is always *minus*. In other words, the solar panels generate *direct current*. It's completely incompatible with a household installation – and there is absolutely no chance to sell direct current to the utility company.

The solution of the problem is an electronic device called a *solar inverter* (or *PV inverter*, or *solar converter*). Such device converts the output DC power from the PV panels to AC power with the same voltage and frequency as the power delivered by the utility company – so that the output can be used at home for powering household devices without changing anything. Also, the surplus power can be sent to the utility company (if it's ready to purchase it) – it's then sent out of the home by the same power line through which the company delivers power to the house.

There are many types of inverters, depending of the specific needs of the owner of the PV installation. Some people may opt to cut the connection with the grid whatsoever – they will need an "off-grid-type" converter, most likely with storage batteries that can deliver power after the sunset and be recharged during the sunlight hours. An intelligent off-grid converter will take care of all such needs. People who want to sell the surplus power to the utility company need some other type of specialized inverter. And so on.

The usage of PV solar installation at a practical scale has begin about 40 years ago. The "heart" of an inverter at that time was a transformer with an iron core, a bulky and very heavy device. The weight of an inverter capable of converting, say, 5 kW of power, could be as much as 200 pounds. Fortunately, the progress in semiconductor technology – and specifically, the development of special transistors called "power MOS FET" – made it possible to eliminate the bulky iron-core transistors and to reduce the weight considerably – as well as the price.

The efficiency of modern solar inverters is very high – typically, in the product specification cards it's given as 98% or even 98.5%.

7.4.7: Inverters is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

7.4.8: Pros and Cons of PV Solar Power Harnessing

The PV conversion of solar power is much less controversial than the CSP technology. A considerable advantage is that after the installation, a PV system is essentially maintenance free – what is only needed is removing the dust from the glass surfaces of the panels once in a while. The panels occupy some space – but in small household installations the space occupied is a rooftop. And large industrial-scale installations can be located at terrains of low or no value for agriculture.

One drawback of PV facilities is that they cannot deliver power during nighttime hours, as the the CSP plants using the molten salt technology do. Currently existing storage batteries may solve the problem at a single household scale – but industrial-scale PV installations of tens or hundres of MV capacity would need storage batteries capable to deliver a similar power for many hours – such batteries have not yet been developed. But let's be optimistic, much R&D work with the aim of developing such "monster batteries" is going on, so the problem may be eventually solved.

An argument that can be often heard from people who are adversaries of renewable energy (yes, there are such people – even among the most influential American politicians) is that a *payback time* of PV panels is very long, even longer than their the average lifetime. The *payback time* is defined as the time needed for a panel to generate the amount of electric energy equal to the energy that has been used for manufacturing the panel. It's a demagogic and unfair argument, though – responsible research quoted by this Wikipedia article – see the *Energy Payback section* – shows that the average payback time for the most widely used silicon crystal PV devices is about 2 years, as compared with the expected average lifetime of about 30 years of such devices.

7.4.8: Pros and Cons of PV Solar Power Harnessing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

8: Harnessing Power from Oceans and Seas

8.1: Introduction

8.2: Wavepower

8.2.1: Methods of Harnessing Wavepower

8.2.2: Attenuators

8.2.3: Point Absorbers

8.2.4: Other Wave Power Capturing Devices

8.3: Tidal Power

8.3.1: A Bit More of Physics

8.3.2: Tides in Coastal Regions, Tidal Currents

8.3.3: Methods of Harnessing Tidal Power

8.3.4: Tidal Power – Conclusions

8.4: Ocean Thermal Energy Conversion (OTEC)

8.5: Other Possible Ways of Extracting Power from Oceans

8.5.1: Ocean Currents

8.5.2: Osmotic (Salinity Gradient) Power Generation

8: Harnessing Power from Oceans and Seas is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.1: Introduction

There are several ways of extracting energy from the oceans and seas. We will list them in the order of *the most realistic* $-\rightarrow$ *the least realistic*:

- Wave energy (the most realistic option);
- Tidal energy (already being used, but at a small scale);
- Thermal methods (based on the difference of surface water tempera- ture, and deep water temperature (prototypical installations);
- Energy of oceanic currents (possible, but not yet harnessed – and maybe it's better to leave it like that?).
- Osmotic (salinity gradient) power generation

A good auxiliary text to the introduction to this chapter may be this European Commission document which provides a brief but comprehensive review of what energy resources the world has in oceans. Current research on renewable ocean energy resources is focused primarily on wave energy. If one assumes that all current research projects sum up to 100%, then – according to a very rough estite by the author – 50% are projects in this area; 25% are those on the energy potential of tidal currents; 10% each are those on tidal barrages and on Ocean Thermal Energy Conversion (OTEC), and 5% are dedicated to the exploitation of water salinity gradient using methods based on the phenomenon of osmosis.

As far as wave power is concerned, it should be stressed that it receives a strong interest from OSU researchers. OSU is a member of the Pacific Ma- rine Energy Center (PMEC), the other two university partners of which are University of Washington (UW) and University of Alaska Fairbanks (UAF). The OSU's research in this partnership is focused primarily on wave energy. UW's research on tidal energy, and UAF's research generally on *hydrokinetic energy*. OSU and UW are additionally linked in the Northwest National Marine Renewable Energy Center (NNMREC) partnership.

Hydrokinetic energy harnessing, by the way, is a general term meaning "extracting energy from flowing waters". It comprises technologies that are already close to their maturity (wave energy harnessing) as well as technolo- gies that are just emerging (extracting energy from river and tidal currents). For reviews, please see, for example: one review, or another review.

8.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

8.2: Wavepower

Ocean waves carry very much power – many things convince us of this, for instance: the erosion of the coastal rocks, the noise the waves attacking the Devil's Churn at the Oregon Coast make, and last but not least, how a wave can accelerate a wave surfer. The power of the waves on oceans, seas, lakes and rivers is all converted solar power. This is a two-step conversion: first, solar power gives rise to winds – and the the winds pass their energy to waves. What is the exact physical mechanism of generating waves by wind is explained, e.g., in this Web article by Ahmed Sawan, or in this Wikipedia article. In Fig. 8.2.1 there is a plot by S. J. Yang form this latter source that is used to explain the role of pressure fluctuations and the so-called "shear instability" in the formation of waves.

We will not try to get very deep into the physics of waves of water. We will only limit our attention to the power a wave carries. The power from a wave is conventionally expressed in W/m units, i.e., the power in Watts over one meter of crest length (distance along an individual crest).

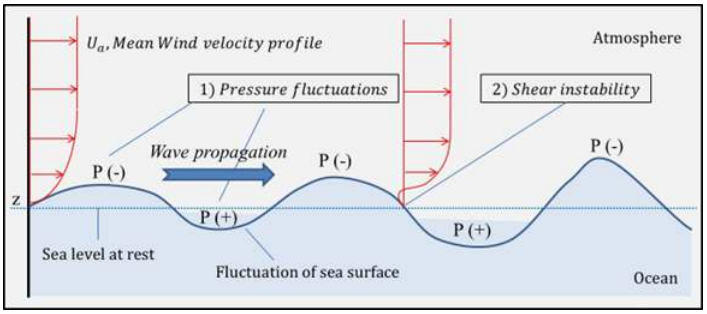


Figure 8.2.1: A plot by S. J. Yang from the Wikipedia article used to explain the role of pressure fluctuations and the so-called "shear instability" in the creation of waves. By the way, physicists – the author is one of them – like to say that waves are "excited", rather than "created".

Please accept without derivation the formulae for the wave power and energy density. The power, per unit length of the wave's crest, is:

$$P = \frac{\rho g^2 T H^2}{64\pi} \mathrm{W/m} \tag{8.2.1}$$

where:

ρ = the density of sea water $= 1025\ \mathrm{kg/m^3}$;

g = acceleration due to gravity $= 9.81\ \mathrm{m/s^2}$;

T = period of wave (in seconds; the averages are: winter $8\ \mathrm{s}$, summer $6\ \mathrm{s}$);

H = wave height (in meters; the averages are: winter $3.5\ \mathrm{m}$, summer $1.5\ \mathrm{m}$).Here H is the significant height of the wave, i.e., its height measured from the "trough to the crest".

The energy density per a unit area is:

$$E = \frac{\rho g H^2}{16}\ \mathrm{J/m^2} \tag{8.2.2}$$

The dependence of the power of waves reaching Oregon Coast on the time of year is shown in Fig. 8.2.2.

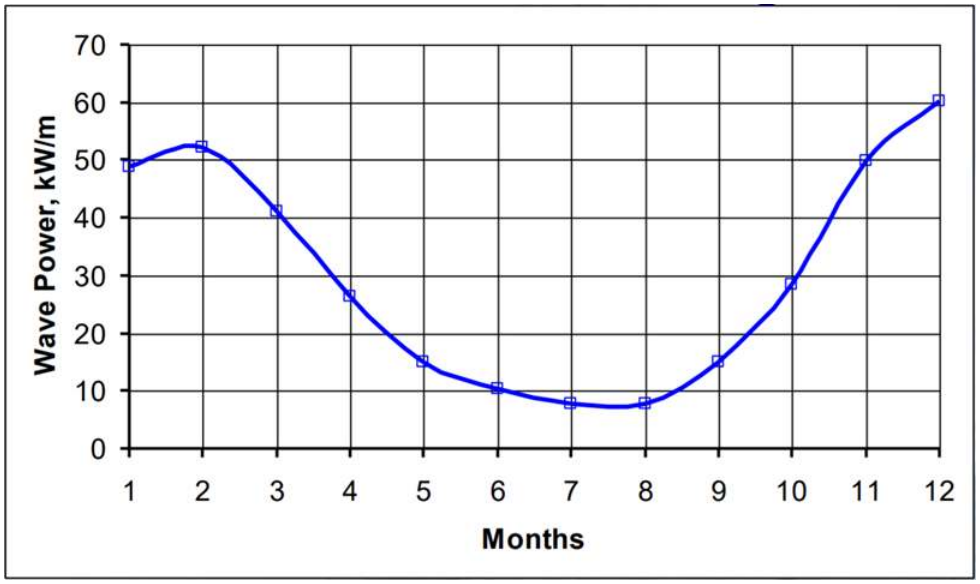


Figure 8.2.2: The monthly average power of waves reaching Oregon Coast, measured over 12 consecutive months (from a PMEC Web site).

If we use the Eq. 8.1 to calculate the power of waves at Oregon Coast in wintertime $(H = 3.5\ \mathrm{m}, T = 8\ \mathrm{s})$, we get:

$$P = \frac{1025 \times 9.81^2 \times 3.5^2 \times 8}{64\pi} = 48079\ \mathrm{W/m} = 48.1\ \mathrm{kW/m} \tag{8.2.3}$$

quite consistent with the graph in Fig. 8.2.2.

Oregon has about 300 miles of coastal line, or 480 kilometers. If we multiply it by $50\ \mathrm{kW/m}$, we get $24\mathrm{GW}$. The electric energy consumption per year in Oregon, according to US Energy Information Administration, is $60.2\mathrm{TWh}$, which corresponds to average power consumption of $6.9\mathrm{GW}-$ which means that in wintertime the poor Oregon Coast is beaten mercilessly by waves with a power three and a half times greater than the electric power consumed by the entire state!

In another US EIA Web page we find that annual energy potential of wave off the coasts of the United States is estimated to be as much as 2.64 *trillion kilowatt-hours, or the equivalent of about 65% of U.S. electricity generation in 2015*.

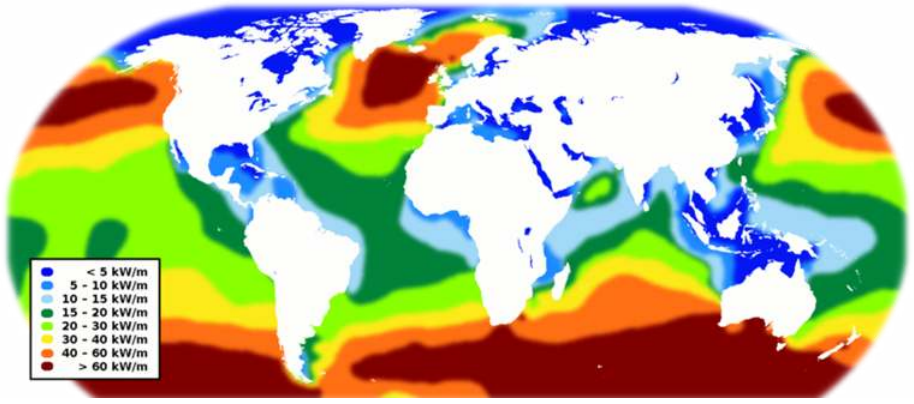


Figure 8.2.3: The global distribution of wavepower (source: Wkipedia Commons).

The data in Fig. 8.2.3, taken from on OSU NMMREC Web page, show the distribution of wavepower not only at the coastal regions, but all over the global oceans. In order to find the *total energy* of all ocean waves worldwide, one has to integrate such data. It's difficult to find such information in the available sources, though. If fact, the total energy of all waves at all oceans is not particularly relevant. Why? Well, nobody will ever try to harness wave energy from waters located far away from the coasts. Obviously, only the coastal zones are really important, for two reasons – first, devices capturing wave power have to be anchored,

which is not possible too far from the shore. And second, the electric power power generated has to be sent back to the coast by cables that cannot be too long.

But even making estimates of how much wave power can be potentially retrieve from coastal regions is not a straightforward task – the best illustration of this is that in a 2011 report of the International Energy Agency, a highly respected autonomous Paris-based intergovernmental organization, this estimate is given with a very high uncertainly margin: namely, from 8,000 to 80,000 TWh per year. In order to set these numbers in a right perspective, let's say that the total global electric energy consumption in 2015 was 20,900 TWh – in other words, the realistic assessment of the available wavepower resources is that it can satisfy between 0.4 to nearly 4 times the global demand for electric power. Even if only the former of the two numbers were realistic, it's definitely worth going for it!

8.2.1: Methods of Harnessing Wavepower

8.2.2: Attenuators

8.2.3: Point Absorbers

8.2.4: Other Wave Power Capturing Devices

8.2: Wavepower is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.2.1: Methods of Harnessing Wavepower

There is a large number of devices for converting wave power to electric power: from those that have already been working for several years, through prototypes at various stages of development, to those that are only concepts in their inventors' minds. It's far too many of them to describe them all in this e-text. We will present only a "small representation" of the entire "menagerie".

In the Author's opinion, a reasonable approach has been used in a Wikipedia article on wave power: namely, of identifying six main generic concepts un- derlying the construction of wave energy converters:

1. Point absorber;

2. Attenuator;

3. Oscillating wave surge converter;

4. Oscillating water column;

5. Overtopping device;

6. Submerged pressure differential.

These six concepts are graphically illustrated in Fig. 8.4.

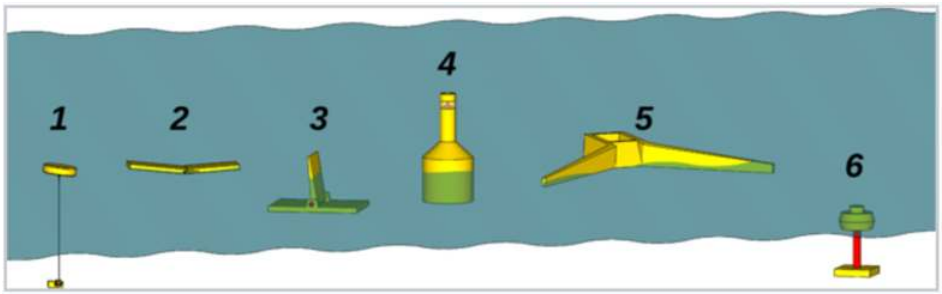


Figure 8.2.1.1: The six generic concepts of wave energy converters – as illustrated graphically in the linked Wikipedia article.

The overtopping device is probably historically the oldest concept, and its functioning is perhaps the most straightforward. The incoming wave runs over a ramp and into a reservoir, the bottom of which is at a higher higher than average water level. From the reservoir, the water return to the ocean through an opening in which a turbine is installed, connected to an electric generator.

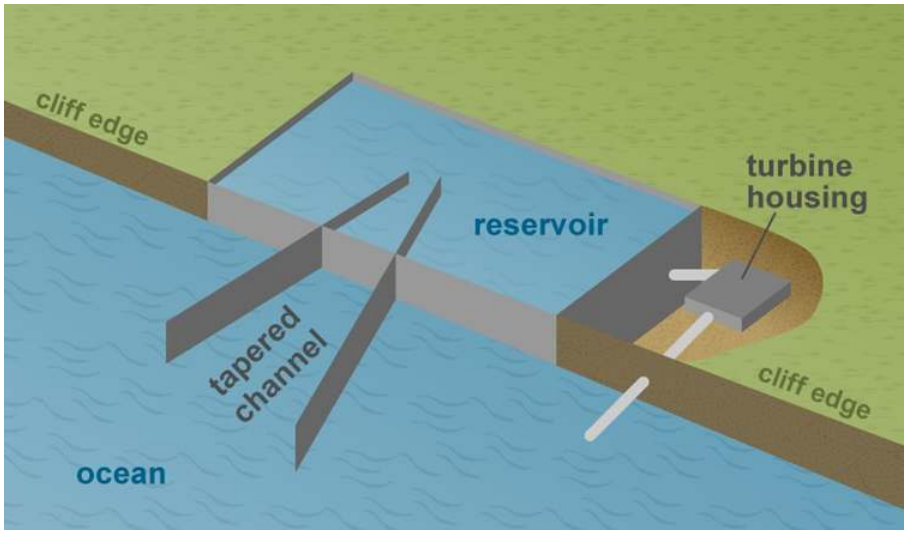
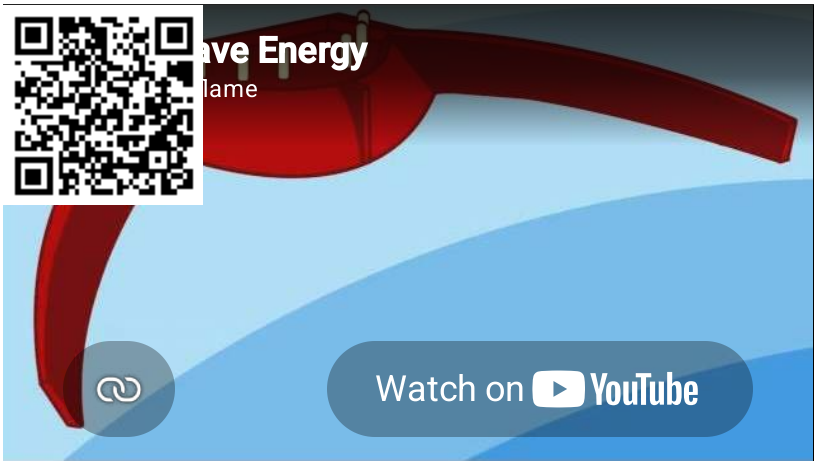


Figure 8.2.1.1: The idea of an overtopping device (source: EIA Web page).

8.2.1: Methods of Harnessing Wavepower is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.2.2: Attenuators

However, we would like to focus our attention on two other more technically advanced concepts. A good idea is probably to show them “in action”, and there is a simple YouTube video clip that explains pretty well how they work. OK, once you have watched the video, we can continue with a more detailed explanation.



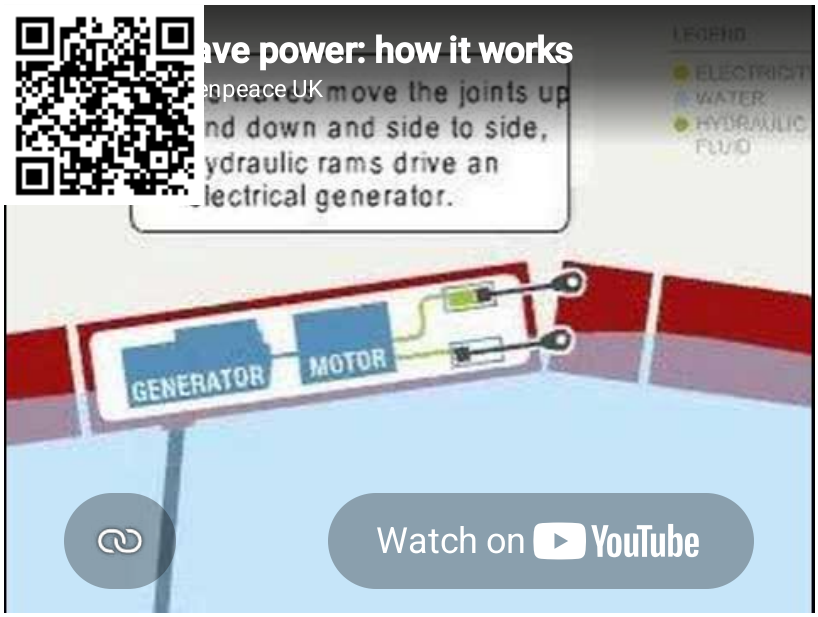
One highly interesting type of wave power converter is known as the *attenuator*. Many of such devices have been already installed in Europe – most R&D work and prototype testing has been done on Pelamis-type attenuators, named after *Pelamis Wave Power*, a company

which developed them.

The Pelamis-type attenuator has the form of a long “caterpillar”, consist- nig of 4-5 floating segments connected to one another with hinged joints. The length of each segment is approximately equal to one half of the crest-crest distance of the incoming waves. The caterpillar is moored at one end to the ocean bottom, so that it gets aligned perpendicular to the incoming wave crests.

When the waves pass by, the hinged segments form a “broken line shape” approximating the shape of the waveform. The angle between two neigh- boring segments varies thus in an oscillatory manner. This motion is used to drive pistons in two pumps. The pistons move in an alternating fashion, as shown in Fig. 8.10: when the hinged joint moves up, the piston in the lower pump is moving in, and the one in the upper pump is withdrawing. And when the join moves down, this motion pattern is reversed. So, the hydraulic fluid in the pumps is always pushed through the hydraulic motor, one way or another. The hydraulic motor always turns in the same direction, no matter which direction the fluid circulates – and it drives the electric generator.

It will surely be instructive to watch another YouTube clip, illustrating specifically the action of the Pelamis-type attenuators.



And here is yet an- other one, showing not animated graphics, but a real attenuator floating in the waves off the coasts of Portugal.



From the pictures, it's difficult to get the right idea of what's the size of a Pelamis-type attenuator. They are really big! The first model, P1, consisting of four sections, is 3.5 meter (12 feet) in diameter, and 120 m (400 ft.) long. It can generate 750 kW of power. The second-generation machine P2 has 5 sections, its 4 m in diameter, and 180 m long, and it weights 1350 tons.

8.2.2: Attenuators is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.2.3: Point Absorbers

The other important category of wavepower-to-electric-power converters we would like to describe in somewhat closer detail are **Point Absorbers**. In general, they are buoys anchored to the bottom by a cable. Usually, they con- sist of a floating cylinder that is attached to the cable, so it stays at a constant distance from the bottom – and of surrounding it another doughnut-shaped floating element that can move up and down. When waves are passing, the outer element moves up and down relative to he inner cylinder.

The relative motion of these two parts should be translated to an electric signal. There are several conceivable methods of doing this, e.g., with first to use a rack-and-pinon gear to change the reciprocating linear motion to rotary motion, and then to use the latter to turn a conventional electric generator. However, a most ingenious and elegant of the existing solution is certainly the one invented at OSU. Namely, it's based on a **direct** conversion of linear reciprocating motion into an electric signal. As any electric generator, it's based on the Faraday's Law of Electromagnetic Induction. It states, in short, that a variable magnetic field passing through a solenoid generates electric current in it. More accurately, it needs to be the **magnetic flux** Φ in the solenoid, the product of the magnetic field's magnitude (commonly denotes as *B*)in the solenoid and the cross-section area of it, $\Phi = \pi r^2 B$, where *r* is the solenoid's radius. If something in physics is *variable*, it means that it *time derivative* is not zero. Accordingly, for generating current in a solenoid it must be:

$$\frac{d\Phi}{dt} \neq 0 \tag{8.2.3.1}$$

The mathematical form of the Faraday's Law is very simple, it just says that voltage E generated in a single ring of wire is:

$$\mathcal{E} = -\frac{d\Phi}{dt} \tag{8.2.3.2}$$

The minus sign in the formula simply tells one what's the direction of the current induces, but it's not so important for our considerations, because we are mostly interested in the magnitude of the induced current, while it's direction is not particularly relevant. So, if a solenoid is not a single ring, but it consists of N turns of wire, the voltage is N times larger:

$$\mathcal{E} = -N\frac{d\Phi}{dt} \tag{8.2.3.3}$$

But Φ, as we say, is the product of B and the cross section area πr^2 of the solenoid. A solenoid has a fixed size, so in the $\pi r^2 B$ expression only B may vary in time. Therefore, one can factor out πr^2 from the derivative, and the equation becomes:

$$\mathcal{E} = -N\pi r^2 \frac{dB}{dt} \tag{8.2.3.4}$$

The simplest way of varying the field *B* in a solenoid is to stick a bar magnet forth and back into it. It indeed induces voltage in a solenoid, as is demonstrated, for instance in this instructive YouTube clip, or in another Youtube clip. In both it is shown that the strength of the induced signal is proportional to *N* , and to how fast the magnetic field penetrating the solenoid is changing.

The design of the point-absorber OSU buoy is a bit more complicated. The coils is mounted in the moving part of the device, but instead of a single bar magnets, in the inner cylinder moored to the bottom there is a number of disc magnets, as shown in Fig. 8.6.

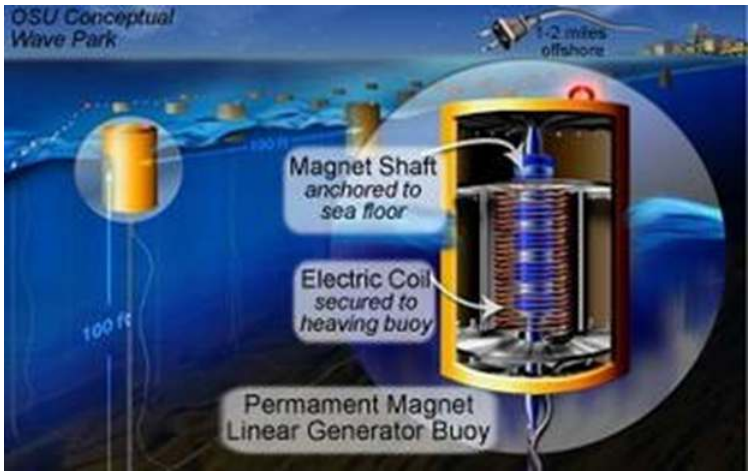


Figure 8.2.3.1: The arrangement of the coil and multiple magnets forming a linear current generator in the OSU point absorber buoy.

Everybody knows the devices called *alternators* – they produce electricity in our cars, and if they break down, there is a drama. Such alternators con- vert **rotary motion** to an electric signal. In contrast, the coil plus the rod including multiple magnets in the buoys constitute a device known as a *linear alternator*. More information about the design of such alternators specially designed to work in wave power converters can be found, e.g., in this document from Louisiana State University authored by Rajkumar Parthasarathy. We will not discuss any further the parameters of the electric signal generated, and how the signal is furtjer processed – the information of importance for us is that a single OSU buoy can generate as much as 50 kW of electric power. And at the end, it's worth watching a short YouTube piece showing an OSU L10 converter and the way of how it is installed.



8.2.3: Point Absorbers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.2.4: Other Wave Power Capturing Devices

There are too many wave power converter prototypes to present them all in this text. New designs are still emerging. For instance, there is a highly interesting Danish prototype of a very large machine with a numbers of floaters, shown in this YouTube clip.



One can expect that the wave power technology has a bright future due to its unquestionable advantages. As noted earlier, the existing resources are such that they may satisfy a substantial part of humanity's needs for electric power. The wave power is "environmentally friendly". The performance of a"wave power farm" does depend on the weather, yes – as in the case of solar power and wind power farms – but it's by far more predictable. In particular, no such "ramp effects" can take place as those occurring in the case of windpower (see the *Windpower* Chapter). No abrupt drops in power generated by photovoltaic installation due to clouds obscuring the sun. The swell of the ocean, the height of the waves can be pretty accurately forecasted for several days ahead.

8.2.4: Other Wave Power Capturing Devices is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

8.3: Tidal Power

8.3.1: A Bit More of Physics

8.3.2: Tides in Coastal Regions, Tidal Currents

8.3.3: Methods of Harnessing Tidal Power

8.3.4: Tidal Power – Conclusions

8.3: Tidal Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.3.1: A Bit More of Physics

The physical mechanism of ocean tides has been described in detail in Chap- ter Two. Let's briefly recall: two tidal bulges are created on opposite sides of the Earth, one on the side facing the Moon, the other on the "far side" with respect to Moon. A common misconception is to attribute each bulge to different mechanisms. The truth is that **both** are manifestations of the same force, referred to by physicists and astronomers as the *tidal force*. Any celestial body experiencing gravitational pull from another celestial body **does experience** the tidal force: it tends to stretch the body (actually, both bodies) along the line connecting the centers of mass of both bodies. For instance, if one of the bodies were a giant water balloon, perfectly spherical if no force were acting on it – then, subjected to a tidal force, this water balloon would change its spherical as if it tended to assume a football shape.

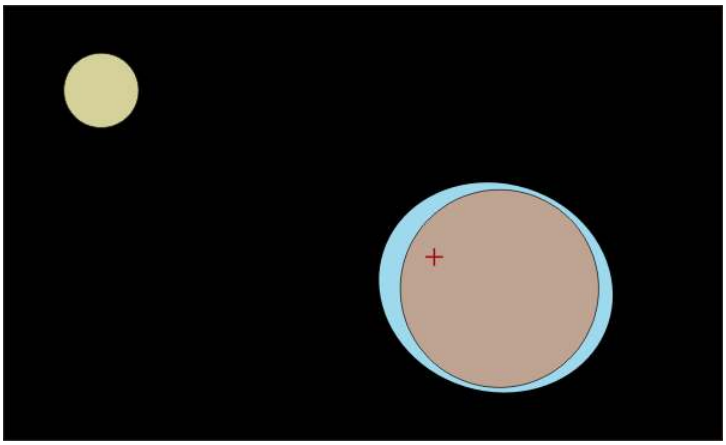


Figure 8.3.1.1: Bulging under influence of tidal forces. A stretching tidal force parallel to the line connecting the centers of Earth and Moon act both on ocean waters and on the rocky bulk of Earth – so both are stretched towards Moon, but the rocky bulk is stiffer than the waters, so that the bulging of Earth's rocky core is weaker than the bulging of ocean waters (the magnitude of the bulging in the figure is greatly exaggerated, of course).The cross shows the exact location of the mass center of the Earth-Moon system (source: Wikimedia Commons).

Well, one could say, if so, then the whole Earth should be stretched, and not only the waters. The answer is that not only the waters are experiencing the tidal force, the rest of the Earth is experiencing them as well. So how comes that only the water bulges up? This is not true, the rocky core of Earth also bulges up – yet, because its greater stiffness it bulges up less than the water does. Therefore, what we observe is the actual bulging up of water minus the bulging up Earth.

Because of the apparent motion of Moon across the sky, the tidal bulges also move, following the Moon's position. Accordingly, for an observer at a fixed location on Earth the two bulges with the two low tides in between look like a "tidal wave" with two "crests" (one at the Earth's near side and one at its far side, with respect to Moon) and two troughs. It takes the crests (and the troughs as well) 24 hours and 50 minutes to travel all the way around the globe (the same time period that separates successive Moon rises – note that if you observe a Moon rise one evening, next evening it will occur 50 minutes later). Or, the tide bulges arrive every 12 hours and 25 minutes.

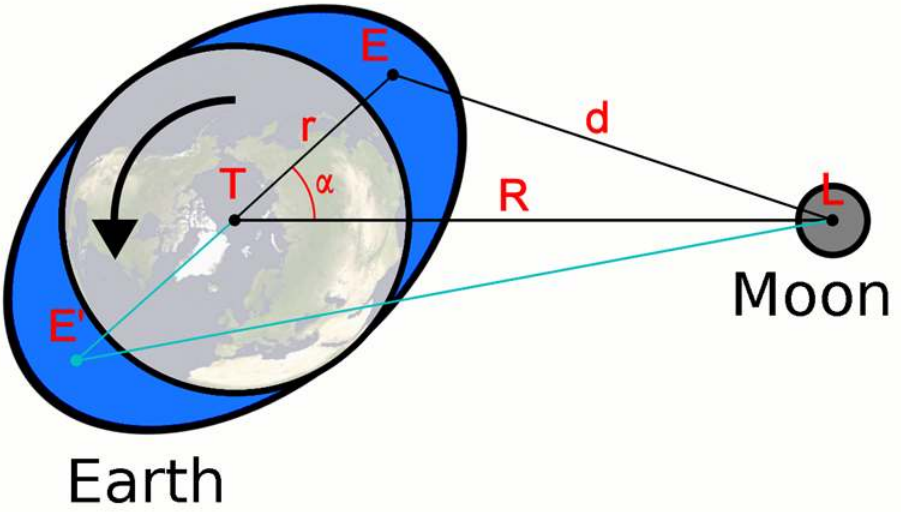


Figure 8.3.1.1: The offset of the tidal bulges – for clarity, the offset angle in the figure is increased about 10 times (source: Wikimedia Commons).

There is one more interesting effect, known as the *tidal bulge offset*. Namely, the highest tide occurs not at the moment the moon is directly overhead. There is friction between the tidal waters and the Earth's surface. The Earth rotation is much faster than the orbital rotation of Moon – Earth makes one revolution in one day = 24 hours, while it takes Moon 29 days to complete a single orbit. Due to the faster Earth's rotation, the friction "captures" tidal bulge and moves it about 3° ahead of the line connecting the centers of the two bodies, as shown if Fig. 8.12. It means that for an ob- server on Earth the maximum bulging occurs not when the Moon is directly overhead, but when Moon is 3° behind its overhead position. To turn by an angle of 3° takes Earth about 12 minutes. Accordingly, for the observer the maximum tide "lags behind" the Moon's overhead position by about 12 minutes. Well, one could say, if so, then the whole Earth should be stretched, and not only the waters. The answer is that not only the waters are experi- encing the tidal force, the rest of the Earth is experiencing them as well. So

how comes that only the water bulges up? This is not true, the rocky core of Earth also bulges up – yet, because its greater stiffness it bulges up less than the water does. Therefore, what we observe is the actual bulging up of water minus the bulging up Earth.

The existence of the tidal bulge offset has important consequences. Be- cause of the asymmetric position of the bulges relative to Moon, the Earth's gravitational field at the Moon's position is not perfectly symmetric – there appears in it an additional weak component tangent to the Moon's orbit, oriented in the same direction as the Moon's orbital velocity vector. This extra force continuously "pushing" Moon tends to increase its kinetic energy, which results in slow increase of the Earth-Moon distance – about $1\frac{1}{2}$ inch per year. A reaction to this force acting on Moon is a force of the same magni- tude, and opposite orientation (due to Newton's 3rd Law of Dynamics), thus tending to **slow down** the Earth's rotation. Such a process indeed occurs, 100 years from now the day will be longer than the present one by about 2 milliseconds (or will be longer than today by one second about 50,000 years from now). So, the process is very slow. However, it also provides an answer to the question of what is the source of the energy that drives all the tidal effects on Earth. The answer is:*the resource from which energy is taken to move tidal waters is the kinetic energy of Earth's rotation. Because energy is taken from this reservoir, the kinetic energy of Earth's rotation decreases, resulting in slowing the rotation down.*

We have not discussed the role of Sun in tidal phenomena on Earth. The contribution of Sun is 44 percent that of the Moon, weaker than that of the Moon. The Sun's impact on the tides i.e., slightly less than half. The effects can add up, which produces the strongest tides – or partially subtract, producing the weakest one. Much information about the Sun's role is given in this Wikipedia article.

8.3.1: A Bit More of Physics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.3.2: Tides in Coastal Regions, Tidal Currents

The tide amplitude depends on several factors, such as the relative positions of Moon and Sun, the geographic location, and a few others. In the middle of big Oceans the amplitude of the tide (the difference between the highest level and the lowest level) may reach almost 1 meter. However, near the continental coastal areas the situation may change, sometimes even dramatically.

As noted before, the tidal waters can be thought as a wave traveling around the world, with the two bulges acting as “crests”. The propagation of this wave would be smooth, if there was an ocean all around the globe. Which is the case only in the southern latitudes, south from the southern tips of South America and Africa, and the coasts of Antarctica. More to the north, the “global tidal wave” encounters continental coasts.

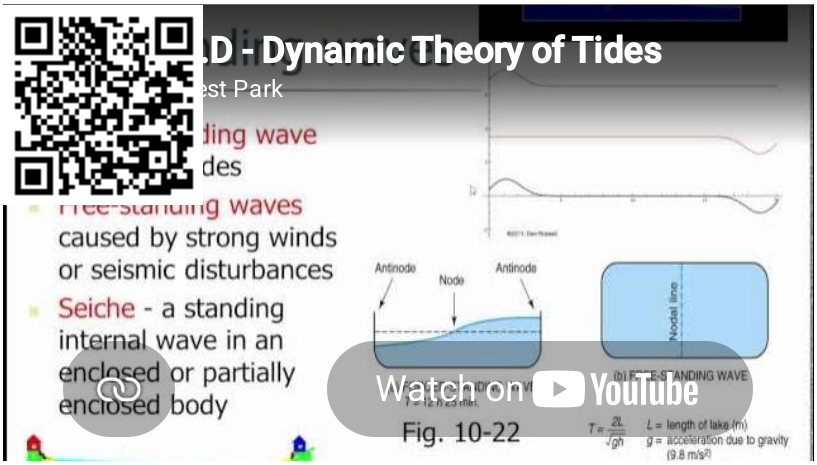
Let’s first think of an ordinary ocean wave on a day with average weather. Far away from the coast the wave doesn’t look very impressive – it quietly “rolls” towards the shoreline. But when such waves get really close to the shore, to the region where the water is getting shallow, the waves start “mis- behaving”, they grow in height, and if they finally encounter not a beach, but a vertical rocky obstacle (there are many such locations along Oregon Coast), then the waves end their lives with spectacular “explosions”, with thunder- ous noises and fountains of foam. The much less common (fortunately!) tsunami waves exhibit a somewhat similar behavior. The wavelength of ordinary ocean waves is only 30 - 50 meters. Tsunami waves, in contrasts, have an extremely long wavelength, form 10 to several hundred kilometers. But their amplitude at deep waters is surprisingly small – one foot or even less. However, when they encounter shallow water, they grow in height and nay produce a surge which is ten meters, or over several tens of meter high.

There is not a close analogy (fortunately!), but at least a distant analogy between the behavior of ordinary waves and tsunami waves, and of the behavior of the tidal bulge when it arrives from deep water areas and approaches a coast. The forces responsible for the creation of ordinary ocean waves and tsunami waves are not the same as those underlying the creation of a tide bulge. Nonetheless, the latter shows similar tendencies as the former in the shallow-water regions. But the tidal bulge is “decisively better behaved”, it grows in height gradually, over the period of hours, and only by a factor or two or so. But, well, it’s not only so simple. If the tidal wave encounters a funnel-shaped geographical objects, such as, e.g., bays, straights, or river estuaries, they may have a “focusing effect” and drive the tide much higher. For instance, the highest tide in the world (15 meters) occurs in the Bay of Fundy at the eastern coast of Canada.

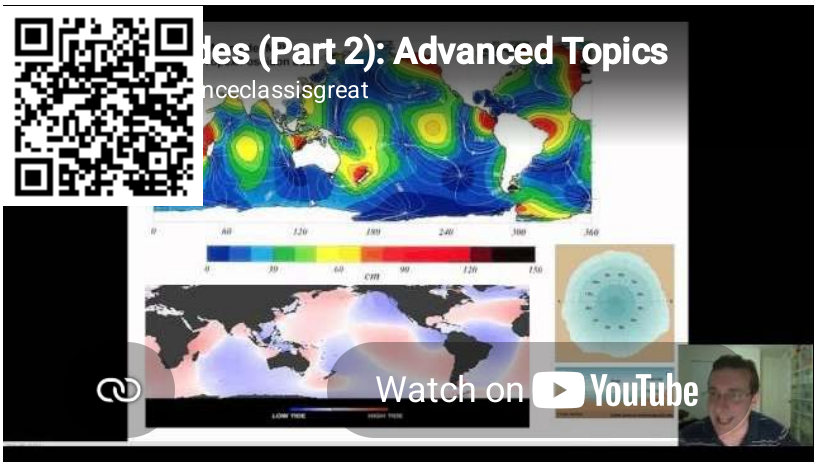


Figure 8.3.2.1: The high tide and the low tide in the Bay of Fundy (source: Wikimedia Commons).

In fact, the pattern of tidal effects over the globe is influenced not only by shapes of coastal lines, but by a number of other factors. In addition to areas of extra-high tide, there are also many “no-tide” areas, called *amphidromic points*. One can find out about other peculiar tide phenomena from this YouTube clip



or from another YouTube in which there is more about the peculiarities of the Bay of Fundy: the tide over there is record high, but there are no two, but **only one** high tide per day. may be



The map showing the tide heights is pretty complicated. Since it continually changes, it's most instructive to watch a "dynamic map", shown, e.g., in this YouTube clip.

Whenever water moves, there are always currents. The simplest tides that alternately raise and lower the water level at your favorite beach also pro- duce currents, perpendicular to the shore line, called *rectilinear*, or *reversing* currents. The current flowing towards the shore is called the *flood current*, and that carrying water away from the coast is called the *ebb current*. Those currents are not very strong (a tidal ebb currents should not be confused with offshore *ripe currents* which are produced by a different mechanism).

But if tidal waters flow into a closed body of water such as a bay, a fjord, a lagoon or a natural harbor, through a relatively narrow inlet or canal. In such circumstances the tidal current in the inlet/outlet area may become really strong. One known such example located in the Oregon's nearest- neighbor, the State of Washington, is *Puget Sound* , a complex system of interconnected waterways and basins with three outlets to the Pacific Ocean via the *Strait of Juan de Fuca* – one major exit (*Admiralty Inlet*) and two minor, *Swinomish Channel* and the *Deception Pass*. It's interesting that the water flow through the Deception Pass is only about 2% of the total tidal exchange between Puget Sound and the Strait of Juan de Fuca, but the speed of the current in the narrowest part of the outlet, under the bridges, may be as high as 8 knots (about 15 kilometers/hour).



Figure 8.3.2.2: Deception Pass – the tidal currents are the strongest under the bridge (source: Wikimedia Commons).

Another famous example of an extraordinarily strong tidal current is the *Saltstraumen Maelstrom* in Norway. It forms in a 3 km long, and only 150 m wide channel connecting a large *Skjerstadfjorden* fjord with a smaller one, *Saltfjorden* – the latter being widely open to the sea. The current in channel attains a speed as high as 37 kilometers/hour, nearly 2.5 times higher than the current speed in Deception Pass. Norwegian people insist that it's fastest tidal current in the world!

The fact that tidal phenomena can be very diverse in areas not too far apart can promote the creation of currents. For instance, one of the fastest tidal currents in the world exists in the Cook Strait, the narrow see passage separating the two island of New Zealand. Not far from the country's capital, Wellington, is the strait's narrowest part in which the current attains a speed up to 4 knots. It may not look impressive at the first glance, compared with the 8 knot speed of the current in Deception Pass, or the 20 knots in the Saltstraumen Maelstrom. But let's keep in mind that the total width of the two channels at Deception Pass is about 200 meters, and the width of the Saltstraumen Maelstrom channel is 150 metes – while the width of the Cook Strait in its narrowest part is 20 **kilometers** wide, about 100 times wider than the Deception Pass channels.



Figure 8.3.2.1: New Zealand, the Cook Strait

There are, of course, many other places all over the globe where strong tidal currents occur. Especially, in areas where there are many bays, peninsulas, and islands.

8.3.2: Tides in Coastal Regions, Tidal Currents is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.3.3: Methods of Harnessing Tidal Power

According to an Encyclopedia Britannica article, the total power contained in tides worldwide is estimated to be 3000 GW – it's three times the total electricity generating power installed in the US, and about 20% of the global power consumption. Of those 3000 GW, it is estimated that between 120 and 400 GW can be generated by *tidal barrages*, and the remainder by underwater turbines capturing power from tidal currents.

Tidal Barrages. This technology is now well established. The idea is very simple: one needs a "tidal basin" – a relatively narrow bay, a river estuary, or a fjord, and to build a "barrage" – a dam that would capture the high tide water. Then, during the low tide period, this water can be send to turbines. Essentially, it's the same operating scheme as that used by conventional hydropower plants with dams built on rivers. However, those power plants work only "one-way" – naturally, the high water level is on the "upriver" side of the dam, and water, after flowing through the turbines, continues flowing "downriver". In contrast, a barrage plant can use a "two-way cycle", in which the same water flows **twice** through the turbines. Suppose that the cycle begins when the "tidal basin" – the reservoir behind the barrage is empty, and there is a high-tide water level on the sea-side of the barrage. Then valves (called "sluice gates") are set open, letting water to flow in through the turbines, until the reservoir gets filled. Now, the valves are closed, and the power plant has to pause and wait for the low tide. When it happens, the valves are again opened and water from the filled reservoir flows back to the sea, again passing through the turbines, until the reservoir is emptied. The valves are now closed and the plant takes another pause, waiting for the next high tide – when it comes, a new cycle begins. So, during a half-day cycle, there are two generating sessions, each followed by a pause.

In France, the Rance Tidal Power Station built at the mouth of the La Rance river estuary has been operating since 1966. Its location is very favorable because the tidal range in the estuary is as high as 8 meters. The plant's maximum power is 240 MW (but due to the natural intermittency of tidal the average power is only 57 MW, corresponding to a capacity factor of about 24%. Another slightly larger facility (254 MW)is the Sihwa Lake Tidal Power Station in South Korea, commissioned in 2011.

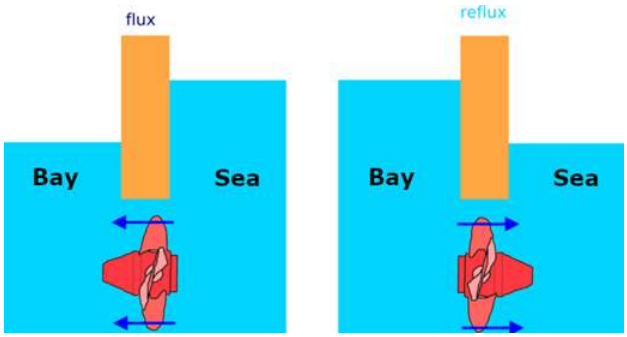


Figure 8.3.3.1: Tidal barrage power plant cycle. Left image: at the high tide, the sluice gates are set open, and the reservoir filling process starts. Right image: at the low tide, the sluice gates are set open again, and reservoir emptying process starts (source: Wikimedia Commons).



Figure 8.3.3.2: The French tidal barrage power plant on the Rance (source: Wikimedia Commons).

As of the beginning of the 2018 year, there are no other existing tidal barrage power plants with power comparable to the two listed above. Nonetheless, there are several ongoing projects of building more such facilities at different parts of the world.

One of the reasons for the slow progress in developing new major barrage power plant is certainly the high cost of building large dams. After the construction is completed, the energy is “essentially for free”, the only costs are due to maintenance and service. But in the case of the La Rance plan it took 20 years to pay back the initial investment.

Tidal Stream Turbines. A much less expensive way is to harness the energy of *tidal currents*, often referred to as *tidal streams*. No huge construc- tions, no earthmoving, no many thousand tons of reinforced concrete are needed.

The “tidal stream generators”, as they are called, are pretty similar to wind turbines. One can use the same math as that used for wind turbines to calculate the expected power output from a single unit – namely, the Eq. 6.2 from the *Windpower* chapter, with $A = \pi R^2$:

$$P = \frac{\pi R^2 \rho V^3}{2} \tag{8.3.3.1}$$

where R, as before, is the radius of the turbine, and now V is the speed not of wind, but of the current, and ρ is the density not of the air, but now of the salty seawater, approximately $1025\ \mathrm{kg/m^3}$.

Let's do an example: a turbine of $10\mathrm{m}$ diameter $(R = 5\ \mathrm{m})$ is placed in the current in Cook Strait. The maximum current speed over there, as we remember, is 4 knots $= 4 \times 1852\ \mathrm{m/}$ hour $= 7408\ \mathrm{m/}$ hour $= 7408/3600\ \mathrm{m/s} = 2.06\ \mathrm{m/s}$. So, $P = 1/2 \times (\pi \times 5^2 \times 1025 \times 2.06^3)\ \mathrm{W} = 352\ \mathrm{kW}$. Well, but the Betz Limit does apply, and the efficiency factor, so let's take only one half of this. $170\ \mathrm{kW}$, it's still a very decent power, it's more than what one hundred average American households consume.

Another interesting problem: let's make a rough estimate [1], what's the total energy of the water comprising the tidal current in the Cooks Strait? Why only a rough estimate? Well, we don't know all essential parameters, we have to base our calculation on a few "guesses". We know that the maximum current speed in the Strait is $2.06\ \mathrm{m/s}$, that is $20\ \mathrm{km}$ wide, and that its average depth is slightly above 100 fathoms. One fathom is 6 feet. Well, but

we don't know whether the current speed is uniform all the way from the surface to the bottom, or perhaps it isn't? Let's try to make our estimates "conservative", it's always safer to "undershoot" than to "overshoot". So, let's assume that the speed $V = 2.06\ \mathrm{m/s}$ applies only to the depth of $100\ \mathrm{m}$. Now, in order to obtain the total energy, we need to take into account the entire water in the strait. And how long it is? Again we have to make a guess: let it be $20\ \mathrm{km}$. So, in our "model", it's like a river $20000 = 20^4$ meters wide, $2 \times 10^4\ \mathrm{m}$ long, and $10^2\ \mathrm{m}$ deep, and all that seawater is flowing with a speed of $2.06\ \mathrm{m/s}$. The total volume of the water flowing is:

$$v = \left(2 \times 10^4\right) \times \left(2 \times 10^4\right) \times 10^2 \text{ m}^3 = 4 \times 10^{10} \text{ m}^3. \tag{8.3.3.2}$$

For calculating the total kinetic energy we need not the volume, but the mass of the moving object - so we multiply the volume by the density of seawater, 1025 kg/m^3, to obtain $m = 4.1 \times 10^{13}$ kilograms. The last step is to use the well-known formula for kinetic energy:

$$K = mV^2/2 = 4.1 \times 10^{13} \times 2.06^2/2 \text{ kg} \cdot \text{m}^2/\text{s}^2 = 8.7 \times 10^{13} \text{ J} \tag{8.3.3.3}$$

Would the fact matter that the turbine from the preceding example is "stealing" energy from the tidal flow with a rate of 170 kW $= 170$ ooo W $= 1.7 \times 10^5$ J/s ? Suppose that the "theft" is continued for 3 hours 10^4 seconds, approximately the time period the high tide lasts. So, the total "stolen" amount is 1.7×10^9 J. Is it much? Let's check:

$$\frac{1.7 \times 10^9}{8.7 \times 10^{13}} = 1.95 \times 10^{-5} = 0.00195\% \tag{8.3.3.4}$$

It's only a tiny part of the total energy of the current. It shows that one can install as many as 500 such turbines as we have considered in the first example, and they will extract only about one percent of the total power of the tidal current. It does not seem likely that the natural flow of tide water in the Cook Strait would be significantly affected by such turbine system.

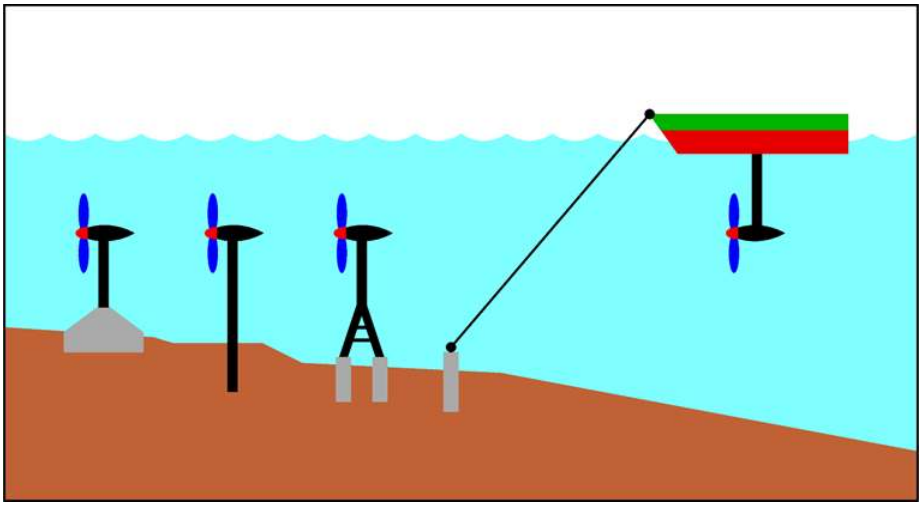


Figure 8.3.3.3: Possible ways of installing tidal turbines. From the left: (i) gravity base, (ii) mono-pile, (iii) piled jacket, and (iv) below the bottom of a floating barge, moored to the sea botton (source: drawn by the author).

One more interesting fact about tidal stream turbines. Installing them on the waterway bottom in a similar way as installing wind turbines on dry land – it is, at the top of a tower erected from a solid foundation – is not the only option. There is even a simpler way: to use a flat bottom of a barge as the base of a “pointing down tower”, and mount the turbine at the lower end of such “tower”. The barge should be moored, of course, to sturdy posts at the bottom (or, if the tidal stream flows in a narrow canal, to rocks on both sides of such passage).

Harnessing power of tidal streams has several advantages. The machinery is relatively simple. There is no need to permanently change the landscape, because all components of the system may be either hidden underwater, or – if a floating system is used – a moored barge is never thought of as a *permanent* object distorting the landscape. In contrast to wind turbines that may kill birds, the rotation speed of underwater turbines is much slower, so they may not pose a serious threat to aquatic life (but there, nevertheless, some other concerns related to the ecosystems).

---

1. By an *estimate*, a physicist usually means *a result that differs no more than an* ***order of magnitude*** *from the actual value*. And *an order of magnitude* means: ten times larger, or ten times smaller.

8.3.3: Methods of Harnessing Tidal Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.3.4: Tidal Power – Conclusions

The tidal motions of ocean and sea waters offer an renewable energy source with vast potential. This energy is absolutely clean, has a minimal envi- ronmental impact, and, despite its intermittent nature, is highly predictable (centuries ago, sailors already knew the art of forecasting tidal height and tidal current strength). According to US governmental sources, tidal streams in the United States may provide power equal to nearly 10% of the country's total electric power consumption (3911 TWh in 2017). Another official institution, the presents more optimistic estimates – that even more than one third of all needs for electricity used in the U.S. could be satisfied by power drawn from the seas.

It is estimated that harnessing local tidal currents can provide about 20% of Great Britain's needs for electricity. These expectations are consistent with the estimates of the total global resources of tidal power given in the article of *Encyclopedia Britannica*, quoted at the beginning of this Section (3000 GW, which is nearly 20% of the current global consumption of electric power).

But so far, ocean resources (including wave power) are expected to pro- duce only 1 GW of electric power worldwide by 2020. Half of that power will come from the two large existing tidal barrage facilities in France and in South Korea. The remainder may come from a large number of small installations, none of which being a real commercial facility – they are all small-scale prototypes, or demonstration installations. 1 GW is only a tiny fraction of the power generated worldwide by two other major renewable sources: wind (over 500 GW in 2017) and solar (over 300 GW in 2016).

Why is the marine energy lagging so badly behind? There are several conceivable reasons. In Author's opinion, the present low price of mined fuels plays its role. But it should affect wind power and solar power as well, but it doesn't. But wind and power certainly gets support from people regardless of where they live, while for inland-dwellers the term "marine technologies" may sound like an abstraction. Well, but it would hardly explain the proportion of one to one thousand.

Some more insight into the problem is given in a recent article *Will Tidal and Wave Energy Ever Live Up to Their Potential?* by Sophia Schweitzer[1]. She lists several other possible reasons, such as environmental concerns and the challenges of underwater technology. But, more optimistically, she also points out that to signs indicating that a strong growth of interest in marine power generating may happen in near future. Well, let's then be optimistic and hope that it will really happen!

---

1. Sophia V. Schweitzer is an independent science writer focusing on climate change and the environment. Her work has appeared in numerous publications including Pacific Standard, Wired.com, and American Forests.

8.3.4: Tidal Power – Conclusions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.4: Ocean Thermal Energy Conversion (OTEC)

All the methods of extracting energy from oceans and seas we have so far discussed in this Chapter have one thing in common: they are all based on harnessing energy from different forms of water **motion**. More accurately, they convert kinetic and potential energy of flowing waters to electric energy. In OTEC, however, the kinetic and potential energy of water is irrelevant – it's the **thermal energy** that is converted to electricity.

OTEC is in certain respect similar to electricity generation from low temperature geothermal waters, described in the Geothermal Energy Chapter. Both method use a heat engine. A heat engine, let's recall, needs a "hot source" of higher temperature $T_{\rm hot}$, and a "cold source" of lower temperature $T_{\rm cold}$ (let's always remember, these temperatures should be given in absolute scale, i.e., in Kelvins). In geothermal facilities, the "hot source" is water pumped from a deep well reaching a geothermal reservoir. And the most often used "cold source" is surface water, e.g., from a stream or a lake.

In an OTEC installation, in contrast, the scheme is "reversed": the "cold source" is ocean water pumped from a depth of 1000 meters, and the "hot source" is surface water. One interesting property of all oceans is that at a $1000\ \rm m$ depth the water temperature is always close to $4-5^{\circ}\rm C$, i.e. to $T=277278\ \rm K$. No matter whether it's close to the Equator, or close to the North Pole. And the "hot source" is the water at the ocean's surface.

Let's now recall that the highest theoretical efficiency of a thermal engine is given by the Carnot Law:

$$\epsilon = 1 - \frac{T_{\rm cold}}{T_{\rm hot}} \tag{8.4.1}$$

and the maximum efficiency of a power-maximizing thermal engine is:

$$\epsilon' = 1 - \sqrt{\frac{T_{\rm cold}}{T_{\rm hot}}} \tag{8.4.2}$$

So, for obtaining a good conversion efficiency $T_{\rm hot}$ should be considerably the difference will be to high. But at tropical lattitudes, the temperature of the surface waters may be $25^{\circ}\rm C = 298\ K$ or even higher. One can therefore obtain a realistic efficiency $\varepsilon' = 1 - (278/298)^{1/2} = 0.034 = 3.4\%$. Well? Not particularly impressive?... In conventional coal or natural gas burning thermal power plants the efficiency may be event ten times higher... But let's keep in mind that coal and natural gas cost money, and here the thermal energy used is for free![1] So, it may be worth going for it. And this is exactly the idea of an OTEC power plant.

In all practical implementations of OTEC installations, the thermal engines used are steam turbines. Of course, one should use not water as the working fluid, but something with the boiling temperature much lower than $T_{\rm hot.}$ As in the case of low-temperature geothermal power plants, a good choice may be ammonia $\rm NH_3$, circulating in a closed system. The basic scheme of such a system is shown in Fig. 8.18. Liquid ammonia is pumped through an "evaporator" - i.e., a heat exchanger, through which also "hot" surface water is run. The heat from water is passed to ammonia, which boils and changes to gas, or "ammonia steam". It's then passed to the turbine, where its thermal energy is converted to mechanical work. In the process, the ammonia gas cols down, but it's still a gas. So, it has to be converted back to a liquid, which requires taking away another portion of heat from it. It's therefore sent to another heat exchanger, called "condesor", through which cold water pumped from the depth is passing. The ammonia gas gives away its remaining heat to the cold water and condensates back to a liquid. Finally, it passes through a pump, which injects it back to the evaporator.

As noted, there are some cost involved. Some power is needed to pump surface water through the evaporator. Some more power is needed to run the pump (labeled as 4 in Fig. 8.18). And much more power is needed for pumping cold water from the $1000\ \rm m$ dept and through the condenser. In practice, all that power needed consumes about 50% of the electric power generated.

In 1974, The U.S. established the Natural Energy Laboratory of Hawaii Authority (NELHA) at Keahole Point on the Kona coast of Hawaii. Hawaii is an ideal site for OTEC research: the temperature of surface waters is high, 25° or more. From the coasts, there is a relatively easy access to deep cold waters: the Hawaiian Islands are essentially steep-slope mountains, rising from the bottom of the Pacific, more than 3 miles deep at that area. So, access to cold water can be obtained by lowering a pipeline along the steep underwater slope.

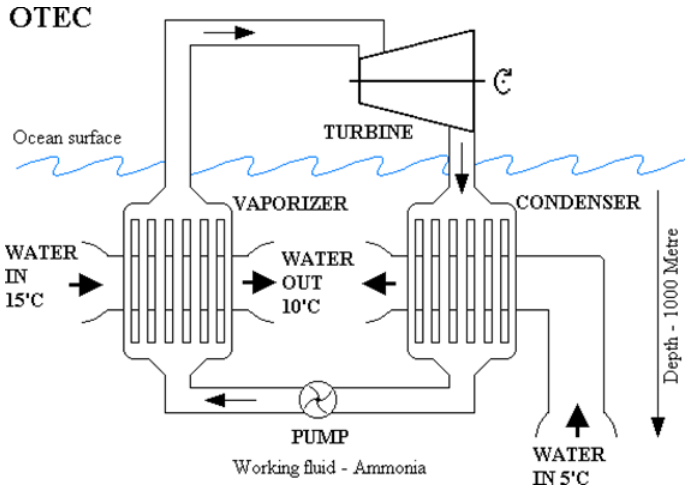


Figure 8.4.1: A diagram of an OTEC installation using ammonia as the working fluid. In the evaporator, the ammonia boils and its vapor proceeds to the turbine. In the condenser, it changes back to a liquid, and the pump injects it back to the evaporator. Such a system can work even if the difference between the temperature of the surface water and that from the depth is as small as 10° – but more energy can be harnessed, of course, by using surface water of temperature not 15°C, but 25-30°C. (source: Wikipedia Commons).

In 1993 a 255 kW[4] prototype OTEC power plant started operating at the Keahole Point, and continued running for six years. The experience gained over that period was used to design a new permanent OTEC installation. And, indeed, on August 25, 2013 a new 105 kW installation started operating, when the Hawaii Governor, David Ige, symbolically ”flipped the switch” to activate it. The power is not particularly impressive, but, in contrast to the installation tested in the 1990-s, the new one **is connected to the grid**. And it’s still not the end of the long-range project, which is a part of the plan of making Hawaii totally energy-independent by 2045. The next OTEC-related step in the plan is to build a 10-20 MW size installation.

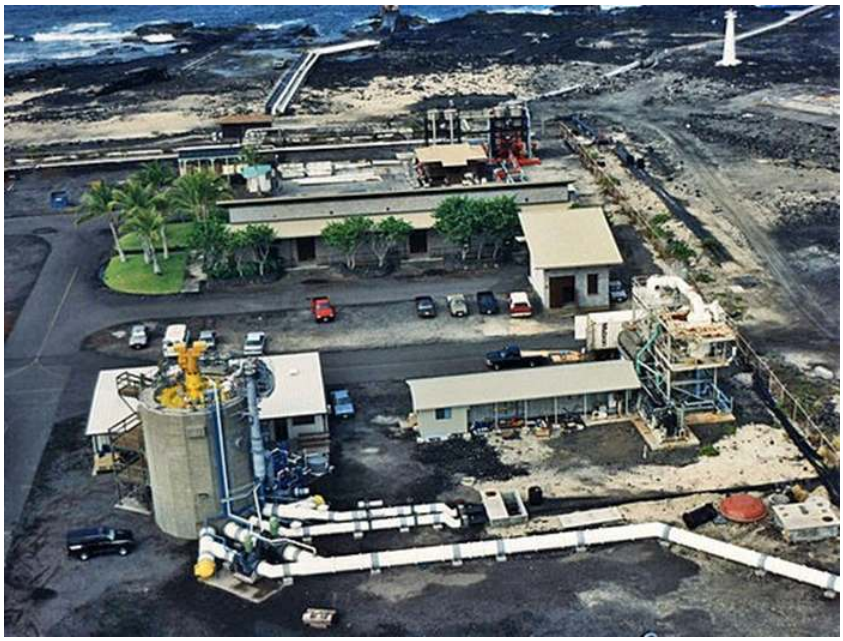


Figure 8.4.1: The first OTEC installation at Keahole Point at the Hawaii Big Island (a photograph from the 1990-s). A photo of the new 105 kW installation is shown in this Web page of Makai Ocean Engineering Company, together with much interesting info about the future OTEC plans for Hawaii (source: Wikimedia Commons).
)

The OTEC technique isn’t not a new idea. It was first proposed almost 140 years ago by a French physicist Jacques-Arsene d’Arsonval. His student, Georges Claude, built the first-ever working 22 kW OTEC installation at Cuba in 1930. More about the OTEC history and the status of this tech- nology can be found in a Wikipedia article, or in this Way2Science Web page.

OTEC can be used only where the surface water temperature is high enough. But there are many such places. The installations should not neces- sarily be located at the coasts. Other possible locations are off-shore platform on continental shelves, or *floating OTEC power plants*. The latter may be located at places where the oceans is 1000 m deep and then use a vertical

cold-water pipeline, which is the easiest to install. Different assessments re- ported estimate the available global OTEC resources (i.e., as much power can be extracted, without disrupting the thermal equilibrium of oceans) to be from 30% to as much as 2-3 times the current global consumption of electric power.

The latter figures would mean that we would be able to supply the en- tire world with OTEC electricity, and still much extra capacity would be available, if needed. A considerable advantage of OTEC power is that – like geothermal power – it is available at a constant level for 24 hours per 7 days in a week. Well, it is available only in geographical areas lying not so far from the Equator (10° 15°). But this is not an unsolvable problem, electric power can be sent over distances of thousands of miles! So, why don't we rush to shut down all the dirty coal burning power plants and start building large OTEC facilities? Well, for sure the major culprit for the lack of action is the prohibitively high initial cost. It's sad. But at least one thing is optimistic: if one day we run out of fossil fuels, then even **OTEC alone** may be able to supply us with no less electric power than we obtain now by burning those dirty CO_2 emitters. And add tide power, geothermal power, all other renewable sources – there is absolutely no fear that there would be a catastrophic never-ending power outage when all the fossil fuels resources are completely emptied. But what about the prohibitively high costs of replacing the fossil fuel plants with renewable power sources? Don't worry! When brown-outs and black-outs become more often and last longer and longer, the governments will gladly revise the budgets and find money to deal with those "prohibitively high costs".

An example: at the end, let's work on an example which will make it easier perhaps to understand why the costs are "prohibitively high". Consider an OTEC installation in which the 25°C (298 K) surface water is used. The thermal engine in the installation extracts the heat from it, and in the process it's cooled down to 15°C (288 K) in the "evaporator", and is returned to the ocean as "wastewater" (as shown in Fig. 8.4.1).

Suppose that one ton (1000 kg) of the surface water runs through the evaporator. Specific heat of seawater (the amount of heat that has to be added to a mass unit to heat it up by 1 K, or taken away from a mass unit to cool it down by 1 K) is 3.85 kJ/kg K. OK, so if 1000 kg of seawater is passed through the evaporator and cooled by 10K, the amount of heat taken away from it is:

$$\Delta Q = 1000 \text{ kg} \times 10 \text{ K} \times 3.85 \text{ kJ/kg} \cdot \text{K} = 38500 \text{ kJ} = 38.5\text{MJ} \tag{8.4.3}$$

Now, suppose that the thermal efficiency of the heat engine used in the installation is $E' = 3.4\% = 0.034$, as we obtained earlier from the Eq. 8.5. It means that 3.4% of ΔQ is converted to work W :

$$w = \Delta Q \times \epsilon' = 38.5\text{MJ} \times 0.034 = 1.309\text{MJ}. \tag{8.4.4}$$

Suppose that the installation "consumes" 1000 kg of surface water every second. So, it delivers 1.309 MJ of power every second, i.e., 1.309 MW of power.

OK, it follows from the above that to get 1MW of output power from the turbine, we need to use 1000 kg$/1.309 = 764$ kg of water every second. Since the density of ocean water is 1025 kg$/\text{m}^3$, the volume of 764 kg of water is 0.775 cubic meter. OK, it still doesn't look scary, does it?

But 1MW is not so much of power. The average consumption of electric power is about 6GW. So, if we wanted to build an OTEC of 1GW power, we have to multiply 0.775 m^3 by one thousand, which is 775 cubic meters per second. Is it much or isn't it? Well, do you know what's the discharge of our Willamette River at it mouth? The long time average, the Web says, is 933 m^3/s. So, for 1GW output power one would to run through the evaporator a stream of water equal to about 80% of the water discharge of Willamette. And for 6GW, it would be 4650 m^3/s. The discharge of Columbia River is $7,500$ m^3/s. So, to get our 6GW, we would need to run through the evaporator a "river" of water as big as 62% of the Columbia River!

I think the above is a good illustration of the scale of the challenges that future builders of large-scale OTEC power plants would face...

---

1. Not exactly... One has to pay, too, but not with money. In a moment, it will be explained.

2. 250 kW was the output power of the generator. But the net power generated was only about one-half of it, after correcting for the power needed to run all pumps.

8.4: Ocean Thermal Energy Conversion (OTEC) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

8.5: Other Possible Ways of Extracting Power from Oceans

8.5.1: Ocean Currents

8.5.2: Osmotic (Salinity Gradient) Power Generation

8.5: Other Possible Ways of Extracting Power from Oceans is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.5.1: Ocean Currents

At the beginning of this Chapter, we listed two more ocean sources from which energy can be harvested. One of them are ocean currents. Not tidal currents we have already talked about – those flowing forth and back in the same rhythm as the tides around the globe. But there are also permanent current always flowing in the same direction. The most famous of them is perhaps the warm Gulf Current, which emerges from the Gulf of Mexico, flows along the coast of North America, then traverses the Atlantic and flows along the coasts of Europe.

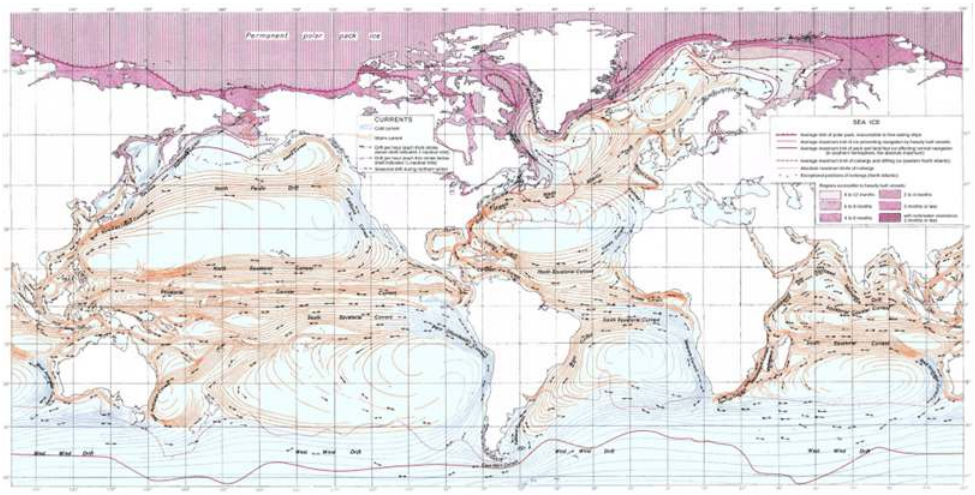


Figure 8.5.1.1: An interesting map of global currents, prepared in 1943 for the US Navy (source: Wikimedia Commons).

Another current we Oregonians should all know is the cold Alaska Cur- rent, flowing along the US West Coast. Because of this current, the Oregon beaches are not crowded even at the hottest summer days. Compare Oregon beaches with those in Croatia, a Mediterranean country at about the same latitude as Oregon, and with its coast famous for its beauty, at places looking like "Oregon's twin coast". And please click on this and see what happens at Croatian beaches where there is easy access to water!

So, one can see that a cold current can significantly influence the climate at the coast. But at the coast only? Think of the Corvallis and the whole Willamette Valley: the summers are mild, with not too many really hot days. Think of what's going on at the same latitude East Coast.

Then, in Pacific Northwest we experience the cooling effect of Alaska Cur- rent. Then we can expect that a warm current should have an opposite effect, right? And it's indeed what happens. In Europe, countries located close to the Atlantic Coast benefit from the warming effect of the Gulf Stream. Especially, Scandinavian countries. Take, for example cities at the Norwegian ocean coast, and cities (well, settlements...) at the western coast of Greenland, where there is no Gulf Stream. One can find "pairs" which have a very similar latitude: for instance, Narvik, Norway, and Rodebay, Greenland – the average winter temperature in the latter is about 6°C lower. Well, six degrees Celsius (over 10 degrees Fahrenheit) makes a big difference.

The Gulf Stream, because of its width (100 km), depth (800 to 1200 m) and speed (about 2,5 m/s at the surface) carries an enormous amount of energy. So, it may be tempting idea to extract some of it. For instance, good conditions for extracting energy from the Gulf Stream do exist at the coasts of Florida. The ocean over there is not very deep, one can install turbines at the bottom, exactly the same way as can be done with tidal stream turbines, discussed in the preceding Section. And in contrast to tidal streams, which flow one way, than pause, then the other way, than pause, and so on – the Gulf Stream flows with a constant power 24 hours/day. So, much more convenient than with tidal streams, isn't it? According to some estimates, the power harnessed from the Gulf Stream could easily supply 50% of the power consumed in all Florida State...

Nice and tempting? Perhaps. But the Author of this text categorically refuses to discuss it any further. In the Author's opinion, playing with permanent ocean currents may have far-reaching and dangerous consequences. Why? Because we still don't know exactly what is the "motor" propelling those currents. We know that it's principally a "thermal engine", fueled by solar power and wind power (which, in turn, also is converted solar power). Water warmed up by Sun expands in volume, so some part of it has to flow away. Well, it can not flow away permanently, it must be substituted by cold water from the depths. And, indeed, in addition to surface currents, there are also deep water currents flowing in opposite directions.

Presently, at the best, we have only what physicists call a “qualitative understanding” (and students usually call a “conceptual understanding”) of the motors setting all those currents in motion. We are not able to say what may happen if we start extracting energy from the Gulf Stream near the coasts of North America. Are we sure that nothing wrong is going to happen to the path of the Stream several thousand miles away? To the part flowing along the coasts of Scandinavia? As long as we don’t have **comprehensive quantitative** model of global ocean currents, enabling one to calculate the effects of any “human intervention” into the power of the currents, it certainly better not to try *meddling into the forces of nature that keep the current running*. How do we know that they are not sensitive to “stealing” some power from them? Europe would never forgive America if some irresponsible experimenting “switched off” the heat the Gulf Stream has been delivering to it for many thousands of years.

Well, one can ask, if the Author is so strongly against harnessing the power of Gulf Stream and other major global ocean currents, why doesn’t he oppose exploiting tidal currents? Here, the answer is easy. The mechanism propelling the tidal streams are pretty well understood. And every tidal stream is not a global, but **a local** phenomenon. Such currents never flow very far: with the speed of 2 m/s, like the stream in the Cook Strait, the water flows perhaps 10 miles during the high tide, then 10 miles the other way during the low tide, and so on. And during each such cycle the streams dissipate their energy, due to friction, erosion, creation of waves, etc. If some part of the energy were not extracted by turbines, it would be lost anyway.

8.5.1: Ocean Currents is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

8.5.2: Osmotic (Salinity Gradient) Power Generation

Every day over the globe a huge volume of river water flows into oceans and “disappears” – not exactly, but it gets mixed with sea water. Is the process of mixing freshwater with seawater energetically completely neutral? No, it isn’t, the mixing process does release some thermal energy, about 2.2 kJ per liter. But it’s definitely not enough to run any thermal engine.

There is, however, a futuristic concept of capturing that “energy of mix- ing”, using a phenomenon called “osmosis”. It’s a very real phenomenon playing a major role in all living organisms.

Let’s first explain what osmosis is. Suppose that you take an empty rectangular aquarium of an $A´B´H$ volume, where A is the longer side, and H is the height. Inside, in the middle, you install a vertical semipermeable membrane. So now there are two open containers of $½´A´B´H$ volume each.

The semipermeable membrane is such that it has micro-pores wide enough so that a water molecule can pass through, but too tight for a molecule of salt, NaCl, to pass[1]. Now, to our divided aquarium, let’s pour: on one side, pure water, to the height of H/2. And to the other half, to the same level, sea water – i.e., 3.5% water solution of table salt, NaCl.

Then wait. And soon a strange thing starts to happen: relative to the original level, the seawater level starts increasing, and the freshwater level – decreasing. Apparently, the water molecules start sneaking through the pores. But note, water molecules are present on both sides of the membrane! They should also pass in the other direction! Perhaps our membrane has one- way micro-holes? Let’s check: let’s empty the aquarium and pour fresh water into the part where the seawater has been, and vice-versa. Well, everything happens as it has happened before. So what’s going on? Why are the water molecules from the freshwater side “privileged”, and those from the seawater side “underprivileged”?

Well, because osmosis plays a crucial role in life processes, and not only for this reason, the phenomenon of osmosis has been thoroughly investigated for about two hundred years – and now our understanding of it, practical as well as theoretical, is really good. However, the “cornerstone” in the theo- retical explanation is the notion of *chemical potential*. It’s name is perhaps a bit misleading, because it may suggest that it’s something that has an ap- plication only in chemistry. Which is not true! In many areas of physics it has, too!

The chemical potential, commonly denoted as μ, is in certain sense analogous to temperature. Risking a slight simplification, one can say: *if one adds/removes heat to/from a given object, the change in its temperature is the measure of how much energy has ben added/removed from it*. And for μ, similarly: *if you add/remove constituent particles to a given object, the change in its μ is the measure of how much energy has been added/removed from it*. Simple? Surely! Yet, in the introductory level academic physics course there is very little time for thermodynamics, so that the chemical potential is not included – in fact, more and more often thermal phenomena are altogether skipped. So, even though the μ-based theory makes it possible to obtain a thorough explanation of all the facts observed in an osmosis process, we cannot present it here.

Instead, we can relay on a conceptual description (the Author hopes that not too many voices of protest). A nice and concise explanation of such kind – one that the Author of this text likes – is presented at the beginning of the Web document posted by Dr. Howard S. Neufeld, a Professor of phys- iological plant ecology at the Appalachian University, who also has strong ties with Corvallis. So, let me quote Dr. Neufeld: *When substances are dis- solved in the water, such as salt or sugar, they cause water molecules to lose kinetic energy, because they are attracted to charged sites on these ions and molecules, effectively immobilizing them, and keeping them from evaporating. This lowers the overall energy state of the water....* And then, after a mo- ment, the text continues: *Now, if we were to take two containers of water, and separate them by a semi-permeable membrane (one that allows water to go through its pores, but not solutes like salt or sugar), and add sugar to one side, this would result in a lowering of the kinetic energy of the water-sugar solution. Thus, from a statistical-probability point of view, we would expect the molecules of pure water to encounter the membrane more often than the lower energy water molecules on the solution side, and thus, over time, water will move from the pure water to the solution. Of course some water molecules do go the other way, but the net exchange favors movement into the solution.* ***This is known as osmosis***.

Beautiful and simple. This is all we need for understanding the *Odsmosis Power* idea (but if you are not satisfied with the above explanation and you would like to review the more advanced version based on μ, please continue reading Dr. Neufeld document, there is such version in the page, too!

The Fig. 8.5.2.1 explains what the *osmotic pressure* is. When extra pressure is exerted to the solution, the water molecules in it get extra "vigor", and a process of "reverse osmosis" begins.

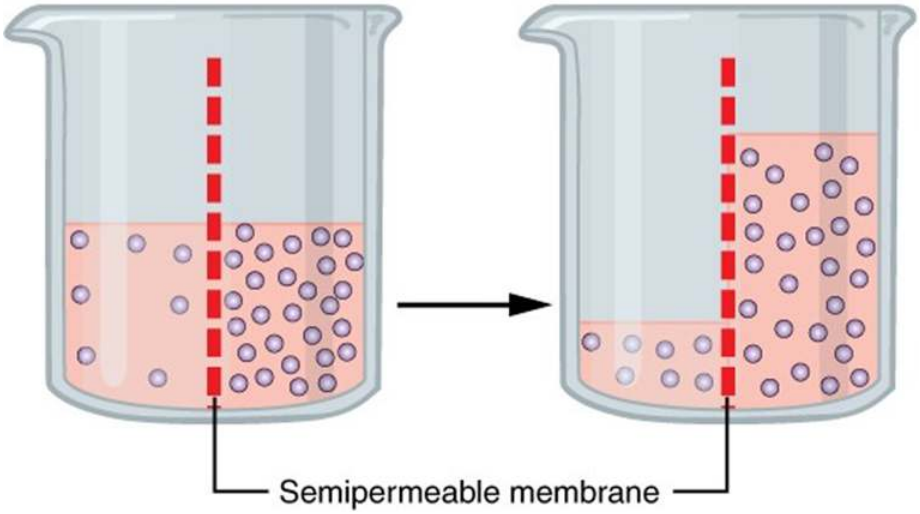


Figure 8.5.2.1: A beaker with two compartments separated by a semipermeable membrane. In one of them (left) there is pure water, and in the other there is NaCl solution. Initially, the levels in both compartments are the same. When osmosis starts and continues, the liquid level in the left container gradually drops, and in the left one it gradually raises. Because of the level difference, the fluid in the right compartment exerts extra hydrostatic pressure on the membrane. This pressure causes that water molecules are "pressed stronger" against the membrane, which pushes them back through the pores. At some moment, the rate of the back-diffusion becomes equal to the rate of the forward-diffusion, and then *osmotic equilibrium* sets in. The pressure needed to establish such equilibrium is called the *osmotic pressure*.

In the experiment shown in Fig. 8.5.2.1, when the osmosis process proceeds, extra hydrostatic pressure $p_h = h\,\rho\,g$ builds up, where h is the difference of fluid levels in the two vessels, ρ is the density of the solution, and g is the acceleration due to gravity. The rate of the reverse diffusion is proportional to the pressure p_h. At certain moment, the pressure attains such a value that the rates of the forward-osmosis and reverse-osmosis become equal – then the system reaches *osmotic equilibrium (source: Wikipedia).*

There is a theoretical formula for calculating the osmotic pressure for any combinations of pure solvent and a solution (solvent + solute). The set of parameters needed for calculations is pretty simple: the molecular mass of the solute, the solution concentration (i.e., the mass of the solute per one kg of solvent), and the temperature at which the process occur. The formula has been tested in zillion experiments, so it can be trusted. The formula, as well and a convenient *osmotic pressure calculator* is provided at HyperPhysics Web Page.

Let's use the HyperPhysics calculator to find the osmotic pressure for a pure water - sea water combination. The concentration of salt in seawater is about 3.5%, so in the boxes "solute mass" and the "solution volume" we put, respectively, 35 grams and $1000\ \mathrm{cm}^3$. Next, the calculator needs the molecular weight of the solute, so we add the atomic weights of Na (22.99) and $\mathrm{Cl}(35.45)$, so we type in 58.44. And finally, it needs the temperature in absolute scale - so let put in $293\ \mathrm{K}$, which is a pleasant room temperature. And we get a prompt answer from the calculator, $1459\mathrm{kPa}$.

Well, kiloPacal means $1000\ \mathrm{N/m}^2$, but it is not a very popular pressure unit. In America everybody knows what is a pound per square inch pressure, but a kPa? $1459\mathrm{kPa}$ may not look unusual until we do some checking. OK, then let's check what's the pressure exerted by a 1 meter column of seawater. $p = h \times \rho \times g$, right? For seawater $\rho = 1025\ \mathrm{kg/m}^3$, so:

$$p = 1\ \mathrm{m} \times 1025\ \mathrm{kg/m}^3 \times 9.81\ \mathrm{m/s}^2 = 10055\left(\ \mathrm{kg \cdot m/s}^2\right)/\mathrm{m}^2 \tag{8.5.2.1}$$

$$= 10055\ \mathrm{N/m}^2 = 10055\ \mathrm{Pa} = 10.055\mathrm{kPa} \tag{8.5.2.2}$$

Only now the osmotic pressure provided by the calculator becomes really impressive: it turns out it's equal to the hydrostatic pressure of a column of seawater 145 meters high, almost one-half of the height of the famous Eiffel Tower in Paris! Or, one can say, the osmotic pressure of seawater in apure water-seawater combination is 14.4 times the atmospheric pressure.

Such a very high water column could never be obtained in the apparatus shown in the Fig. 8.22. Why? Because when the osmosis goes on, more and more pure water enters the seawater compartment, gradually diluting the seawater. And the osmotic pressure is

proportional to the solute concentration in the solvent. Therefore, even if the "neck" of the bottle-like vessel in the Fig. 8.22 were much longer, the column height would stop increasing after reaching a much smaller value than $h = 145\ \mathrm{m}$.

But you don't need columns as high as $145\ \mathrm{m}$ to operate water turbines! Much lower h values could also do. And this is exactly the idea underlying the osmotic power generation, or salinity gradient power generation method. Use water river as the "pure water", and seawater as the "solution" - and by osmosis, raise the level of the fluid in the seawater compartment high enough so that the hydrostatic pressure which builds up is sufficient to drive a turbine.

In a Web article a team of Norwegian researchers presents a thorough explanation of how osmoticenergy production can beimplemented in practice. Diagrams of such power plants are shown and the issue of manufacturing semi-permeable warranting a high efficiency of the process membranes is dis- cussed in detail.

The method discussed in the Norwegian paper is not the only one possible – more sophisticated methods have been conceived. More details about such ideas can be found in this Web document, in which there is also a link to a short YouTube video explaining the principles of the osmotic power generation method. Other instructive articles can be found in Wikipedia, in Sciencedirect.com, or in this report from the Science Advances journal.

According to the data discussed in the Web documents linked, the world resources of osmotic power generation are such that they may satisfy more that 10% of the current global electric power consumption. It’s therefore a bit disappointing that until now only installations with sub-kiloWatt power has been tested. Nothing has yet emerged over the level of “early R&D projects”. The relatively low prices of fossil fuels are probably a “strong slowing-down factor” for such projects. But it’s important to continue the R&D work, because – who knows? – such times may come that the world will need to quickly activate all existing reserves.

---

1. Many such membranes are available. In the old days, researchers used membranes taken from slaughtered animals. Fortunately, now we can make synthetic semipermeable membranes using different methods. For instance, by irradiating thin plastic sheets by beams of heavy high-energy ions, which punch “microholes” of desired width. Such mem- branes are widely used, e.g., in the apparatus for artificial dialysis – they save life of people whose kidneys do not function normally. Or, in the Big 5 store in Corvallis for about 20 bucks you may buy a “magic straw” for backpackers: one may drink water from very con- taminated creeks or waterholes, and 99.99% of germs have no chance of sneaking through the micropores in the straw.

8.5.2: Osmotic (Salinity Gradient) Power Generation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

9: Geothermal Power

The Earth's interior is hot – incredibly hot! According to estimates made by learned scientists, the temperature at the Earth's center is roughly the same as at the Sun's surface, nearly 6000°. And because the Earth's surface, as you know, is much cooler, heat continuously drifts from the Earth's interior to the surface and out. The total flux of the thermal energy emerging from the Earth's interior is about 45 TW – more that twice as much as the total global consumption of power by all humanity.

What is the source of that energy coming out "from down under"? Well, in part it's the primordial energy. As the science of cosmology teaches us, the Solar System formed about 5 billion years ago, from a spinning cloud of dust, called the "solar nebula". Over 90% of the nebula's mass formed the Sun, and the remainder coalesced into smaller objects, one of which was our Earth. The coalescence happened primarily due to gravity – in other words, the gravity performed a colossal work for merging all those dust particles together – and, as we knot from the First Law of Thermodynamics, all that work ended up as thermal energy. So, the Earth at the beginning of its existence was red-hot – then, it stated radiating out its thermal energy. But for a sphere that big loosing energy is not a uniform process: first the areas close to the surface area cool down. It creates a temperature gradient that enables the heat residing closet to the center to start flowing out. But this is not the whole story, because more thermal continues to be generated within the entire Earth's volume: it's coming from the decay of long-lived radioactive isotopes such as Uranium 238, Thorium 232, and Potassium 40 (with half-life times, respectively, of 4.5, 14.1, and 1.25 billion years). It's not exactly known how much of the energy presently emerging from the Earth's interior is of primordial origin, and how much is due to radioactive decay of minerals – in some sources one can find a 50-50 proportion, in some other 20% primordial, 80% from radioactive decay. But, in fact, it's not so important for us – we know that the total power is about 45 TW, and this is all we need to know.

45 TW is an impressive number – but keep in mind that the surface area of Earth is also pretty big, about 126 billion acres, or 0.126 Tera-acre – so the average flux of heat emerging from "the netherworld" is only about 360 W/acre, or 0.088 W/m^2. For comparison, the heat flux emerging from the body of an adult human at ambient temperature of 20° is of the order of 20 W/m^2. And we know that the heat our body produces is enough for keeping us warm in a sleeping bag at nighttime at a backpack hiking trip – and that's all. So, 0.088 W/m^2 can hardly be thought of as an impressive figure...

Suppose that we have a "magic technology" enabling us to collect 100% of the geothermal heat emerging from the ground, and to convert it all to electric power. It only requires covering the ground with a thermally insulating "blanket". Suppose that using this magic technology we want to produce free electric power for all people and all institutions in Benton County. The population is 90,000 people. The average electric power generation per capita in Oregon is 1.9 kW, so the power we need is 90,000 1.9 kW = 171 MW. And the geothermal power flux is 0.088 MW per square kilometer. So, we need to cover 171/0.088 = 1943 square kilometers. But the whole land area of Benton County 1759 square kilometers! We cover all Benton County with our magic "Geothermal Power Capturing Blanket", no leaving a single square inch for people, homes, animals, farmland, forests, OSU campus... And what?! Even then we are not able to generate enough electricity!

9.1: The Non-Uniform Distribution of Geothermal Heat Flux
9.2: Geothermal Electricity Generation
9.2.1: The Flash Steam Method
9.2.2: The Binary Cycle Method
9.3: The Enhanced Geothermal Systems
9.4: Geothermal Heating
9.4.1: Shallow Geothermal Heating

9: Geothermal Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

9.1: The Non-Uniform Distribution of Geothermal Heat Flux

The above example clearly shows that if the total geothermal power were evenly distributed over the Earth surface, there would be no chance of making a ny practical use of it. Fortunately, the geothermal power flux is not evenly distributed over the planets surface and there are areas where the flux density is orders of magnitude higher than the average value of 0.088 W/m^2.

To explain what's going on, for a moment we have to focus our attention on the Earth's inner structure. With the current advanced drilling technol- ogy, we can probe directly the composition of the rocks to the depth of about 10 km below the Earth surface. But the Earth's radius is 6371 kilometers. How can we get an insight into the areas at greater depths than 10 km? Well, there is something that can be used for "X-raying" those regions – namely, the seismic waves. Such waves are created by earthquakes. Not only those major catastrophic ones, but also by smaller ones which cannot be felt, but can be detected by instruments called *seismographs* (there are about 500,000 detectable earthquakes every year). The seismic waves can travel long distances, even across the globe. When passing from one layer to another with different properties, they exhibit the effects of refraction and reflection, in analogy to the known behavior of light. The branch of science focused on investigating the origin and the propagation of seismic waves is referred to as *seismology* . The foundations of modern instrumental seismology were laid 170 years ago by Robert Mallet. Systematic research on seismic waves continued since then have revealed a clear image of the inner structure of Earth. In short, it consists of several spherical "layers" of different properties, as shown in Fig. 9.1.1.

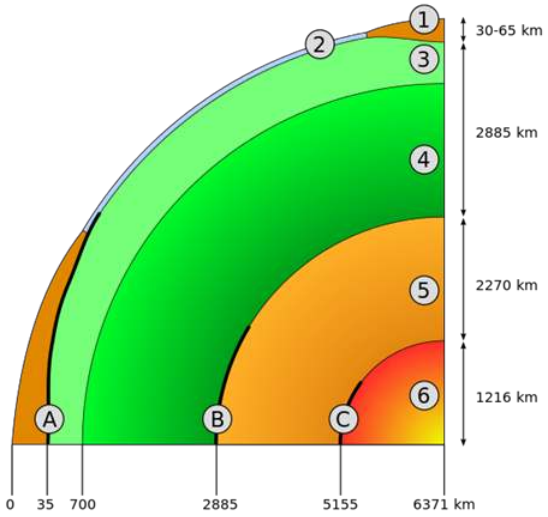


Figure 9.1.1: : Schematic view of the interior of Earth. 1. continental crust 2. oceanic crust 3. upper mantle 4. lower mantle 5. outer core 6. inner core (after Wikipedia).

From the viewpoint of geothermal power, only the properties of the uppermost layer, the *lithosphere*, are important. The lithosphere is the rigid outermost shell of a planet (the crust and upper mantle, with a combined thickness up to about 100 kilometers.). It is not a uniform spherical shell – it is broken into a number of tectonic plates: there are six major of them, and a number of smaller ones. The most important fact is that the pattern of plates is not static, it's *dynamic*. The relative motion of the plates is explained in closer detail in this Web site of the US Geological Service. The regions where there exist the most favorable conditions for harnessing geothermal energy are near the boundaries of the tectonic plates.

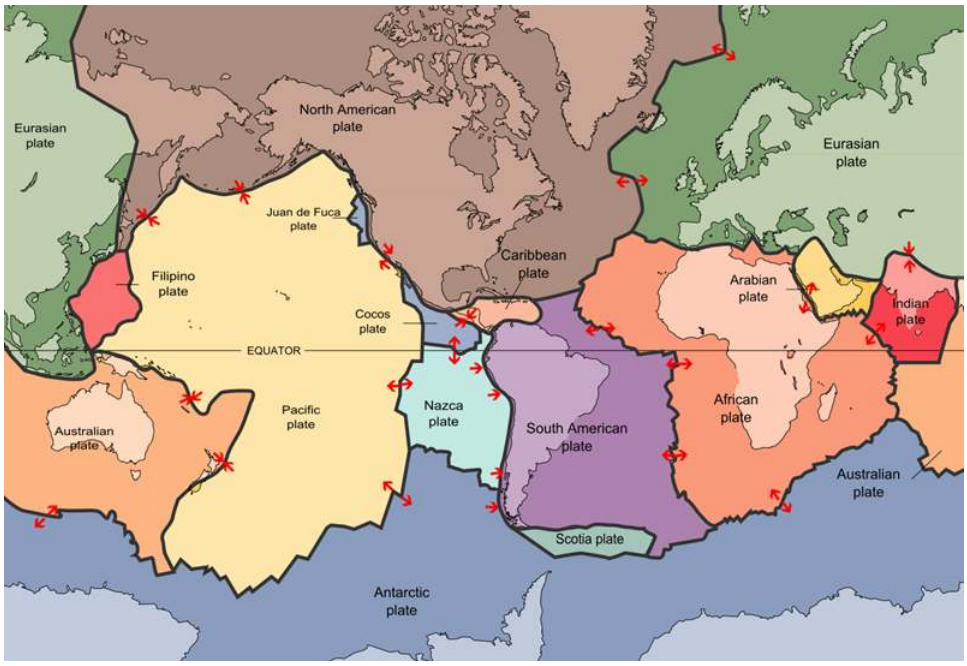


Figure 9.1.2: Map of the tectonic plates using colors for clarity. Small arrows at the boundaries indicate whether the plates are convergent ("colliding") or divergent (moving away from each other). Source: Wikipedia

The boundaries may be *divergent* – meaning that the two plates tend to move away from one another. It would create a deep "cleft" between the two plates, reaching down to the upper mantle – yet, the pressure existing in the upper mantle pushes upwards the liquid and semi-liquid rocks it consists of (such mixture is called *magma*), so that the cleft gets filled

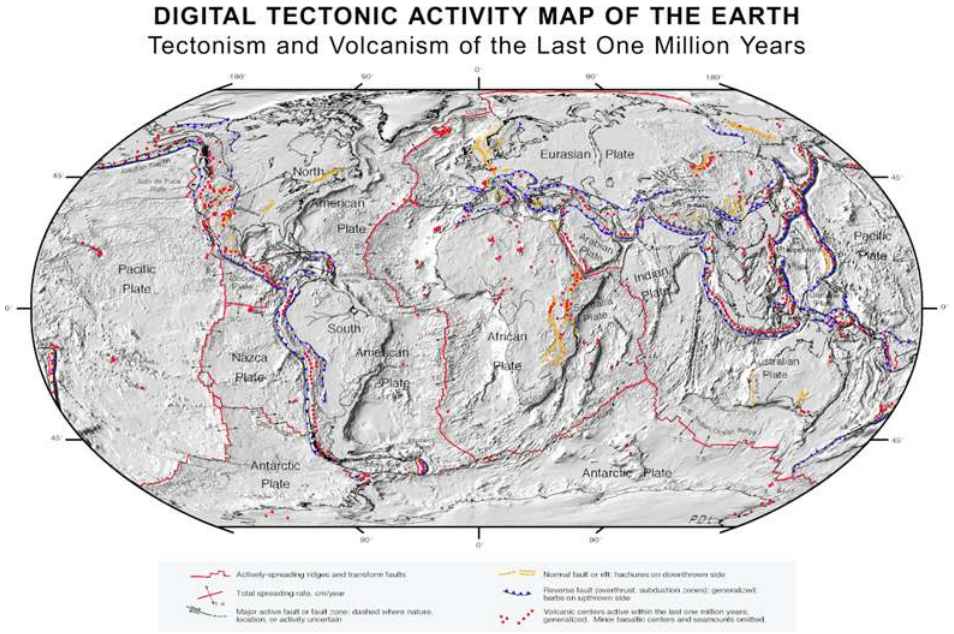


Figure 9.1.3: Another map of the tectonic plate boundaries superimposed on a three-dimensional map of the Earth's solid surface, digitally generated by NASA based on precise satellite measurements. Out of necessity in this text the map is in a low resolution version, but the Reader is encouraged to explore its high resolution version, available at this Wikimedia Commons Web address. In a fully blownout version selected smaller areas of the globe can be explored in a "piecemeal fashion" - e.g., it's certainly of interest to look at the complicated pattern of faults and rifts off the Oregon Coast.

The best known example of such an divergent boundary is the Mid- Atlantic Ridge. The lower end of it is located below the southern tip of Africa, and it runs over the Atlantic Ocean bottom all the way up to the Arctic Ocean. Note that the shape of this boundary closely resembles the contour of the western coastal line of Africa and the contour of the eastern coastal line of South America, indicating that these two continent were once a single larger continent that separated and drifted away, while the magma

emerging from the widening cleft gradually formed the entire bottom of the Atlantic in between. Today, Africa and South America still drift away with a rate of 2.5 cm/year. It's a slow motion, but over a million years it corresponds to 25 kilometers – so, assuming that the drifting rate did not change in time, the creation of the entire part of Atlantic separating the two continents could take about one hundred million years. One more interesting feature of the cleft is that it runs below the island of Iceland. The inhabitants of the island benefit from the presence of a large number of geothermal spots. In Iceland, geothermal energy is used for generating nearly 30% of electric power, and for heating 85% of homes, as well as greenhouses, schools, and a variety of public institutions.

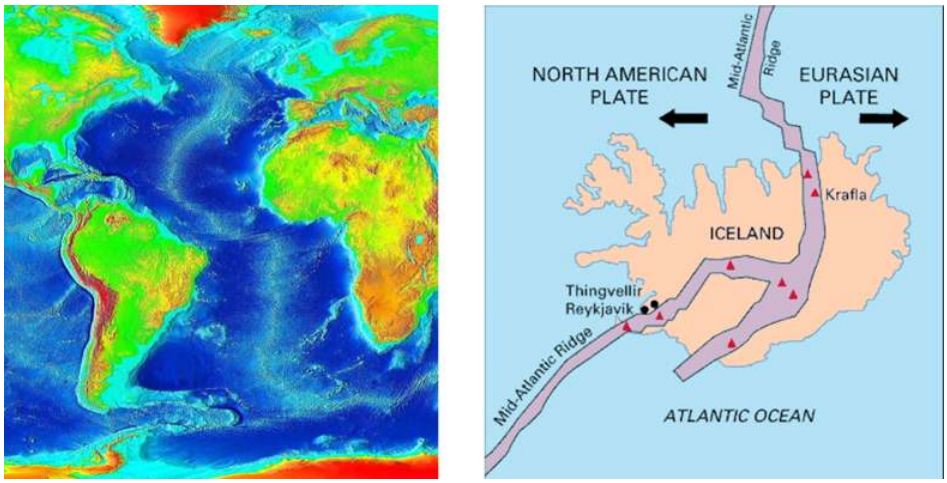


Figure 9.1.4: Left image: the Mid-atlantic Ridge running from below the southern tip of Africa up to the Arctic Ocean, and crossing Iceland (top of the figure). Right image: a small-scale map showing the Ridge crossing Iceland.

If the plates press on each other, the boundary is *convergent*. Sometimes one of the plates is the "winner", the other "sinks" below it, or is *subducted*. The region where it happens is the *subduction zone*. One consequence of subduction is that much magma is pushed upwards and if it reaches the surface, it may create hot spot or even a volcano. In fact, most of the existing volcanoes are located close to tectonic plate boundaries. In particular, the boundary of the Pacific Plate is an especially efficient "volcano maker", about 80% of all active volcanoes on Earth are located close to it, and therefore it is known as the "Ring of Fire".

Divergent tectonic plate boundaries may also give rise to volcanic activity – as is the case, for example, at Iceland. Sometimes, however, a hot spot or volcano occur in the middle of a tectonic plate, far away from its boundary. Known example are the active volcanoes at Hawaii, or a number of volcanoes in Africa– by the way, the linked site makes it possible to go to a "global tour" and to visit most of the existing volcanoes on Earth, these close to the tectonic plate boundaries as well as those located pretty far from them.

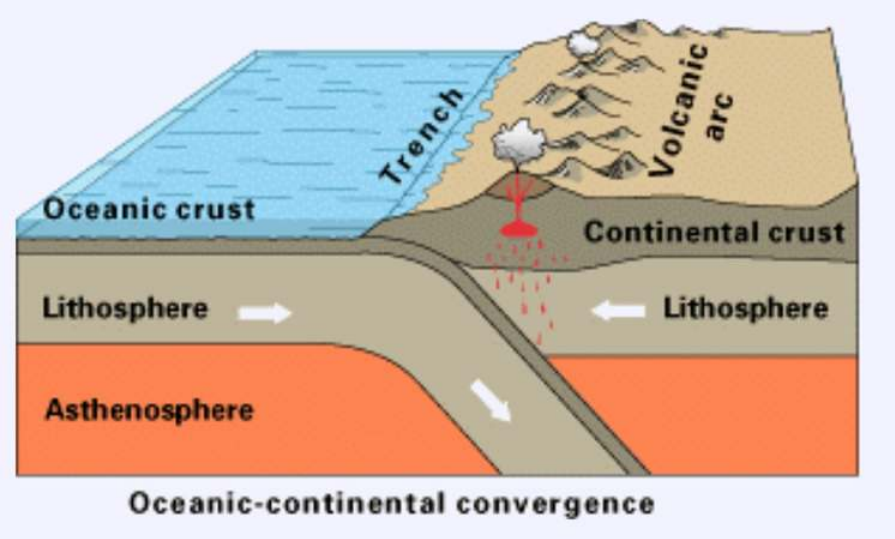


Figure 9.1.1: One tectonic plate subducted under the other at a *convergent* boundary. Since the crust over oceanic plates is thinner is thinner than that on continental plates, it is normal that the oceanic plate is the one "subducted". Magma raising up from the subduction zone gives rise to much volcanic activity in the continent's coastal area (source: Wikimedia Commons).

9.1: The Non-Uniform Distribution of Geothermal Heat Flux is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

9.2: Geothermal Electricity Generation

Geothermal heat sometimes flows “all by itself” from underground caverns, as hot water in hot springs and geysers, or even as dry steam (the very first geothermal electric generator in the world was built in 1904 in Larderello, Italy, on a dry steam source that had been active since the ancient Roman times). More often, to get access to geothermal resources one has to drill wells – in areas with high geothermal activity they may not need to be very deep. But it is possible to access geothermal heat at almost any site of the globe, even if it is located far away from tectonic plate boundaries or other geothermal hot spots. If one digs a well or a shaft, the temperature of the surrounding rocks increases at a rate of 25-30° per kilometer – it’s referred to as the *geothermal gradient*. So, in order to “produce” hot water at boiling temperature, it may be enough to drill a well of 3-4 km depth and to pump water from it – if it happens to be down there. Or, if there is no water, one can drill two parallel wells and pump cold water from the surface down through one of them, and after it is heated, to recover it through the other well.

In order to generate electricity in a geothermal power plant, three “com-ponents” are needed:

- Heat – an underground geothermal “heat reservoir” is needed, with the property of “self-renewability” (meaning that if thermal energy were extracted from it, the loss would be supplemented at the same rate by heat influx from surrounding hot rocks).
- Fluid – sufficient fluid (water) must exist naturally in the reservoir, because only hot water or steam can act as the “carrier” of the thermal energy extracted from the reservoir through a “production well”. But if there in not enough water naturally occurring in the reservoir, extra water can be pumped into it through an additional “injection well”.
- Permeability – the water must be able to come into contact with the hot rocks, so that natural fractures must exist in the reservoir, or they have to be artificially created (e.g., in a similar way as “fracking” is performed in natural gas mining).

Depending on the temperature and the pressure in the reservoir, the thermal energy may be delivered to the surface either as dry steam or as hot water. Steam is convenient in that it may be directly sent to a turbine, as shown schematically in Fig. 9.2.1.

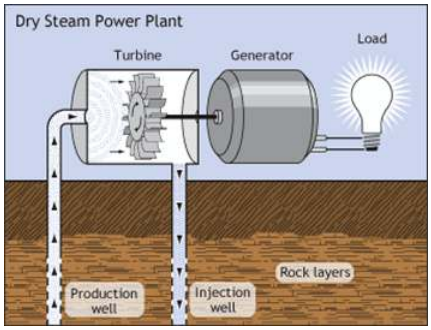


Figure 9.2.1: Dry steam emerging from a geothermal heat reservoir through a “production well” can be sent directly to a steam turbine. The exhaust steam is cooled, so that it condenses back into water, which is then returned to the heat reservoir through the “injection well” – the figure if from the Web page of the U.S. Department of Energy’s Office of Energy Efficiency and Renewable Energy (EERE) via this EIA Web page (public domain).

The dry steam technology was used in the aforementioned world’s first geothermal power plant at Larderello, Italy. The geothermal area in which Larderello is located is pretty large (about 250 square kilometers) and can accommodate many more power plants. Currently, there is a number of plants in the area, running on steam delivered by over 200 production wells. The original Larderello plant commissioned in 1913 delivered 250 kW of electric power from a single production well. Due to many improvements in the original technology (such as reinjection of water and deep drilling – see Page 4 in the Web document linked above) the power yielded by a single well has increased tenfold, so that the entire complex now delivers nearly 600 MW of electric power.

The dry steam technology is also used in the world’s largest geothermal power complex, located in *The Geysers* area, located about 72 miles north of San Francisco, California. There are 22 power plants in the complex, supplied by steam from over 350 production wells. The total nameplate electric generation capacity of the complex is 1517 MW, with an average actual production close to 1000 MW.

If a geothermal reservoir provides dry steam, it means that the temperature of the rocks inside it is greater than 374°C (or 705°F). Why? Because 374°C is the **critical temperature** of water: above this point, water cannot exist in liquid form, only in a gas form, which is commonly referred to as *steam* or as *vapor*. At lower temperature the gas phase, steam, may coexist

with the liquid phase – but then the steam coming out of the production well is no longer dry, it's a mix of steam and suspended miniature droplets of liquid water. But if there is enough water in a reservoir of temperature lower than 374°C, the pressure will rise to a level enabling all water to remain in a liquid form – and it's hot water coming out of the production well, rather than "wet steam".

Coming back to Italy: in the 1950-s and 1960-s another 50 km^2 geothermal field was discovered in the vicinity of Mt. Amiata, about 50 miles south-east from Larderello. Here, in contrast to the situation in the the Larderello-field, the underground heat reservoirs are liquid-dominated. In other words, the "output" from the production well is not dry steam, but hot water. What is the appropriate technology for such a situation? Currently, two different methods are being used, illustrated by the two schemes in Fig. 9.2.2.

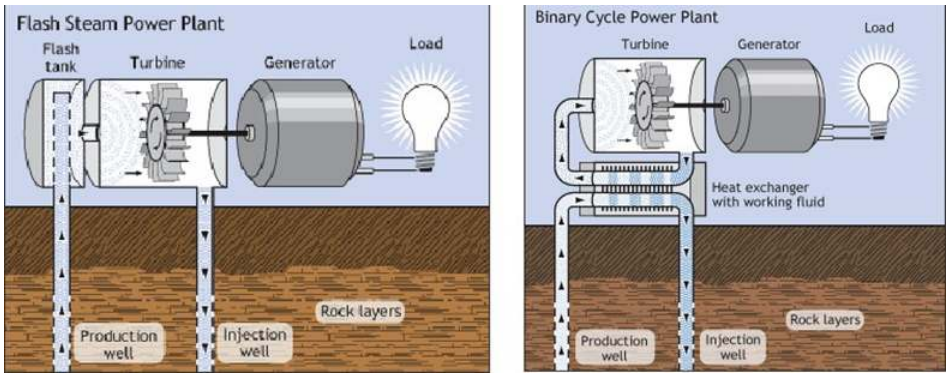


Figure 9.2.2: Left image: a scheme of a *Flash Steam Power Station*. Right image: a scheme of a *Binary Cycle Power Station*. Both figures are from the same source as that in Fig. 9.2.1.

9.2: Geothermal Electricity Generation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

9.2.1: The Flash Steam Method

This method is the most popular one. The lower limit of hot water temperature making it possible to use this technique is given in the literature as 180°C. In order to explain how the method works, it's good to use an example with "real numbers".

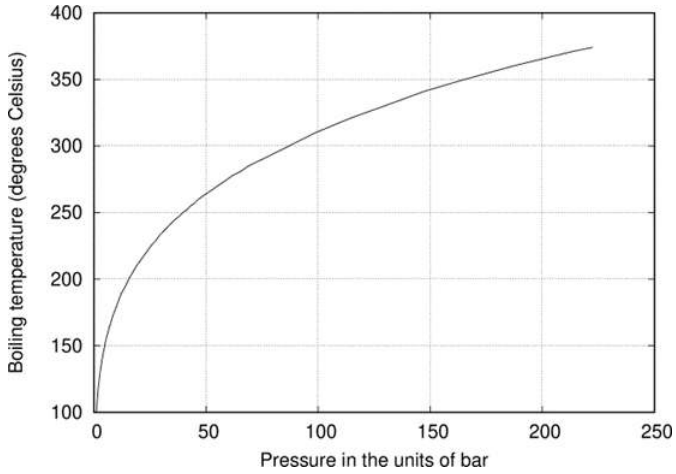


Figure 9.2.1.1: The boiling temperature of water as a function of pressure. It has to be kept in mind that in the process commonly referred to as "boiling", the liquid phase and the gas phase (steam) of water **coexist** – so that a point (*p*, *t*) lying on the curve can be interpreted as: *the pressure p needed to keep water liquid at temperature t* (source: aop).

OK, so let's assume that the water temperature at the outlet of the pro- duction well is 225°C. At this temperature, as one can readily find from the graph in Fig. 9.2.1.1, the pressure needed to keep water in liquid state is 25 bara[1]. Now, the 225°C water is let to flow into a *flash tank*, in which the pressure is much lower – say, 10 bara. What happens then? At 10 bar, as we find from the graph in the engineering source linked above, water can stay liquid only if its temperature is no higher than 185°C. So, after entering the flash tank it starts to boil violently, until its temperature falls down to 185°C. It takes only a split second, hence the term "flash" – and in the process much of the water changes into steam. So, as the result, one obtains water of temperature 185°C as well as a generous portion[2] of steam of 185°C temperature and of 10 bara pressure. This steam is sent to a turbine, whereas the 185°C water is recirculated down to the heat reservoir.

1The *bara* is a unit of absolute pressure – it's not an official unit in the SI system, it's only "tolerated" by it – but engineers like it. One *bara* is equal to 100,000 Pascals, where a Pascal (Pa) **is** an official SI unit of pressure, equal to 1 N/m^2. The reason why the *bara* is so popular among engineers is that it is approximately equal to the atmospheric pressure at standard conditions. will spare you the calculation details, but with the help of another table from an *Engineer's Handbook* one can find that for the temperature and pressure data used in our example 10.1% of the hot water entering the flash tank changes into steam. For higher hot water temperature this percentage may increase significantly. It should be noted that steam turbines "like" high-pressure steam. In thermal coal- fired or gas-fired power plants the water is usually heated *above* the critical point, in order to obtain the so-called *superheated steam* with a pressure even higher than water's *critical pressure* of 220 bara.

9.2.1: The Flash Steam Method is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

9.2.2: The Binary Cycle Method

If the temperature of the hot water retrieved from the geothermal reservoir is lower than 180°C, the temperature and the pressure of the flash steam would be so low that a turbine would not be able to convert its energy to mechanical work with any reasonable efficiency. From the graph in Fig.9.7 it follows that after "flashing" water with temperature lower than about 180°C one would obtain steam with a pressure of, say, 5 bara – which is perhaps a good parameter for driving small model turbines, but definitely too small for industrial-scale turbines[3].

There is a "conceptually simple" solution to the pressure problem

namely, at the head of the production well one should install a device trans- ferring the thermal energy from water to another liquid with a much lower boiling temperature. For instance, ammonia NH_3 is a relatively inexpensive fluid (used in large amounts in fertilizer production) with boiling tempera- ture of -34°C at atmospheric pressure. At 132°C, an equilibrium between liquid and gas state occurs at the pressure of 113 bara. So, by using a 132°C water for heating ammonia liquid one can obtain "vapor" (not steam – the term "steam" is reserved for water vapor) of pressure high enough for run- ning a turbine with high efficiency. The process of heating ammonia – or, more accurately, of transferring the thermal energy from water to ammonia – takes place in device commonly referred to as a "heat exchanger".

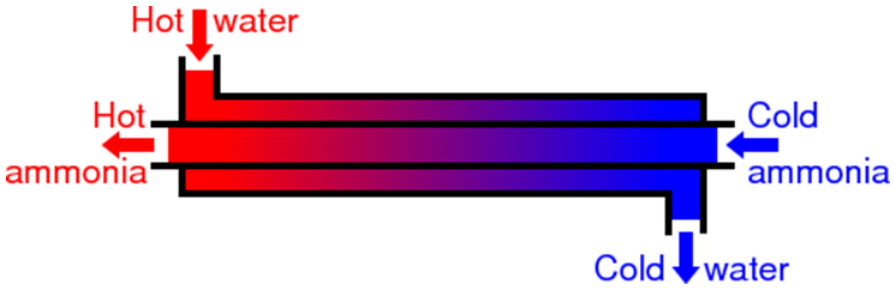


Figure 9.2.2.1: A simple "counterflow" heat exchanger consisting of two concentric tubes. Ammonia flows in one direction in the inner tube, and water flows in the opposite direction in the space between the outer surface of the inner tube and the inner surface of the outer tube (source: aop).

Ammonia is highly toxic, and small installations are prone to leaks. For- tunately, Mother Nature has equipped us with leak detector which are par- ticularly sensitive to ammonia leaks – our noses. The odor of ammonia is incredibly strong and people start detecting ammonia in the air before the concentration attains the toxicity level. There are other safer "working flu- ids" for binary plants, though. In the middle of the 20th Century the ideal candidates seemed to be freons, due to their low toxicity and flammability – however, since then they have been banned because freon leaking from ther- mal engines and refrigerators destroys the atmospheric ozone. But there is much ongoing research see, e.g., this Web document.

The US is the world leader in geothermal electricity generation, with over 3600 MWe installed – almost twice as much as in the Philippines, next on the list. The geothermal reservoirs exploited are all large, capable of delivering enough heat for generating tens of MWatts of electric power using the dry steam or flash steam technology. It is expected that many more such reservoirs will be identified in the western part of the US. Therefore, most of the existing interest in geothermal technologies in the US is focused on the two aforementioned technologies. Similarly, in Europe the resources exploited in the Mediterranean geothermal region have similar characteristics as those in the Western US. However, more to the north in Europe the geothermal resources are of different kind, in most cases they are scattered geothermal aquifers at the depth of 2-4 km and with temperatures close or only slightly higher than 100°C. So, in many places there are are good conditions for building binary cycle installations. They can generate enough electric power for a community of several hundred households and – what is even more important – for heating those houses in wintertime.

In the case of huge, multi-MW dry steam or flash steam installation heating usually does not get much attention because to reach customers, heat should be sent over long distances – which is not always feasible. In contrast, the "waste heat" from a binary cycle plant serving a small community located nearby may become a valuable commodity. A good example illustrating how a small community can profit from geothermal energy is Altheim, a small town (or a large village?) in Austrian alpine region. The binary cycle power plant, installed over there at the end of 1990-s, has been operating until now. It uses geothermal water of temperature 106°C for generating about 500 kWe of electric power, and for heating 800 homes in wintertime. The Altheim installation can be

considered as a “model prototype”, it has been described in several professional publications. For instance, one such report was published in 2002 in the Quarterly Bulletin of Geo-Heat Center (a subdivision of the Oregon TECH). The paper is worth reading, because it’s a highly professional report written in a very “pedagogical” style. Another Altheim paper, from 2015, is a report summarizing 15-year period of the installation’s activity.

9.2.2: The Binary Cycle Method is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

9.3: The Enhanced Geothermal Systems

The basic methods described in the two preceding sections can be applied in areas where there exist **conventional hydrothermal resources (CHR**. A Web page of the US Department of Energy lists three components that a CHR contains naturally: namely, it requires **fluid, heat**, and **permeability**. Further quoting the linked document: A CHR *...can occur in widely diverse geologic settings, sometimes without clear surface manifestations of the underlying resource. In 2008, the U.S. Geological Survey (USGS) esti- mated that 30 GWe of undiscovered geothermal resources exist in the western United States1 – ten times the current installed capacity*.

As of December 2017, the installed geothermal nameplate capacity in the US was 3,600 MW (in the world, over 14,000 MW – see the list of the 10 top countries). So, currently about 0.3% of electric power generated in the US comes from geothermal sources, and, according to USGS, by building new power plants at existing CHRs this contribution can be increased to about 3%.

However, there is an MIT document published in 2006, titled *The Future of Geothermal Energy*, which shows that the contribution of electric power generated by geothermal plants can reach 10% or more of the total usage by the middle of this century. How comes, if, according to the USGS, the capacity of the still undiscovered resources may contribute only about 3%?

The extra capacity needed, as the MIT document claims, can be obtained from **Enhanced Geothermal Systems (EGS)**. What a EGS is is described in great detail in the document, but since it's 372 pages long, we do not encourage the reader to study it page-by-page – we will provide a concise explanation over here. We will do that by quoting three pieces of graphics from another US Dept. of Energy Web page. So, the first piece of graphics, shown in Fig. 9.3.1, graphically explains what are the three components needed in a Natural Geothermal System:

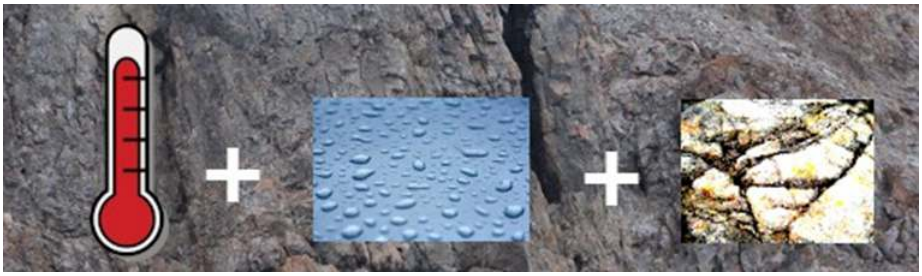


Figure 9.3.1: The three necessary components of a natural geothermal system: (i) abundant heat found in rocks in depth; (ii) fluid to carry heat from the rocks; and (iii) crevices and cracs to allow water to flow through the hot rocks (source: aop)

One of those components is available almost everywhere, not only at the Western part of the US. As was stated before, at locations even far away from geothermal activity zones, there is a *geothermal gradient* of 25-30°C per kilometer. So, if we need, say, a 150°C source for running a binary power plant, we will find it by drilling a well 5-6 km deep. Well, but there is no guarantee that water and permeability will be down there. The situation may appear to be as it is illustrated in Fig. 9.11: there is little or no water and the rock has a monolithic structure. What can be done in such situation?

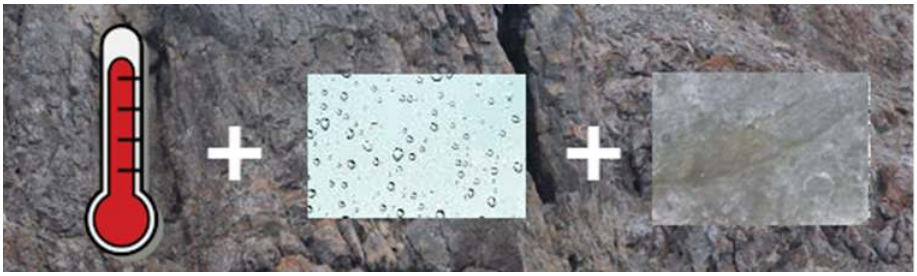


Figure 9.3.2: One can always find rocks hot enough for the purpose of geothermal electric power generation, but what if one of the other two components are missing, or both are missing? (insufficient water, the rock "too solid", or both).

There is a solution – water can be delivered to the hot rocks through a second "injection well", and the permeability can be created artificially in the same way the miners who exploit natural gas or oil trapped in non-permeable rocks do it.



Figure 9.3.3: The situation described in the preceding figure can be repaired, by artificially producing cracks and crevices in the rock by applying the ”fracking” technique used in oil and natural gas mining, and by delivering water from the surface through an extra ”injection well”.

It’s perhaps a bit disappointing that despite the many potential advan- tages of the EGS technology, not much effort has been made towards its practical implementation. Until now, only some early-stage R&D work has been undertaken. One of the leaders in EGS is Australia who plans to launch a 25 MW demonstration project in Cooper Basin in the near future. A list of ongoing demonstration projects in some other countries is given in this Wikipedia article.

Good news is that China has entered the club of countries interested in EGS research. China has started search for Hot Dry Rocks (HDR) re- sources later than many other countries, but then it started pushing forward with a zeal characteristic for many new high-tech ventures in this country. HDR deposits are exactly what is needed for “mining” energy using the EGS technology. Chinese geologists have so far identified deposits of nearly a quadrillion tons of HDRs. China is now the biggest emitter of CO_2 in the world, but it has declared many time its sincerely commitment to signifi- cantly reduce that emission. A quadrillion ton of HDR contains an amount of energy sufficient for satisfying the total energy consumption of China at the current level for many thousand years. Therefore, one can expect that China will soon undertake energetic actions to start extracting energy from those vast resources.

The U.S. has known HDR resources of comparable magnitude to those in China – but not much has been done yet to start “mining” energy from them. One reason for that is perhaps that the U.S. – as has been mentioned above – is expected to have not yet discovered **hydrothermal resources**, capable of increasing **tenfold** the current geothermal generation of 3,600 GWe. And extracting power from a hydrothermal resource is by far easier than from a HDR resource. Anyway, in order to reach the 100 GWe goal set in the aforementioned 2006 MIT report, sooner or later it will be necessary to activate potential existing EGS resources. In fact, first research efforts

focused on HDR energy started as early as in the 1970s, motivated by the oil crisis. The best known of the studies undertaken at that time is the *Fenton Project* originated in the Los Alamos National laboratory. Over the years, the status of the project has been reported in several papers – here we have chosen two: one from 2012 because it’s “pedagogically written”, and one posted in 2015 which has a number of highly instructive pieces of graphics, and even more importantly, a long list of “clickable” references to earlier reports from the Fenton Project and relevant reports from other projects.

However, neither the Fenton Project nor any other American HDR re- search project has ever generated any commercial activity – everything has been R&D activity. So, the good news is that the U.S. Department of Energy has created a partnership of seven National Labs. Under the leadership of Lawrence Berkeley National Laboratory (Berkeley Lab) they will start a new cycle of vigorous research aimed at removing technical barriers to com- mercialization of enhanced geothermal systems. Let’s then hope that their research efforts will progress smoothly and fruitfully.

9.3: The Enhanced Geothermal Systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

9.4: Geothermal Heating

In the preceding sections of this chapter we talked mostly about using geothermal for generating electric power. But another worthwhile application is heating – either the space in people's houses, or household water, or water in public swimming pools, or greenhouses, or the list may be long.

In geothermal heating, one can either use for this purpose all power extracted from the geothermal reservoir, or send this power first to a heat engine for generating electric power and use the waste heat from the engine for heating. As we remember, a heat engine converts only a fraction ε of the input thermal energy into mechanical work. In the best coal-fired or natural gas-fired power plants this fraction may be slightly higher than $\varepsilon = 30\%$, the rest is the "waste heat". But in geothermal systems the temperature of the "input heat" is lower than the steam temperature in coal or gas fired plants. Hence, the efficiency of the heat engines in geothermal installation is also lower – perhaps 20% in the dry steam system, and even less than 10% in the binary cycle systems. So, as much as 80-90% of the input heat ends up as the "waste heat" and it can be used for heating. As follows from the previously discussed example of Altheim, an electricity generation + heating combination may indeed work pretty well.

There is only one problem with heating – namely, in contrast to electric power, heat power cannot be sent over long distances. In other words, all available heat power has to be used locally. In Altheim, the geothermal installation was intentionally built next to the town, so delivering heat to homes easy. But often, in the case of huge American installations like, for instant, the famous *The Geysers* power plant in California, there are no human settlements nearby. So, all the waste heat has to be dispersed in the cooling towers or injected back to the underground reservoir. This is regrettable, because in the energy usage balance of an **average** American household, as can be seen in Fig. \(\PageIndex{1}\ more than 50% energy is used for space heating.

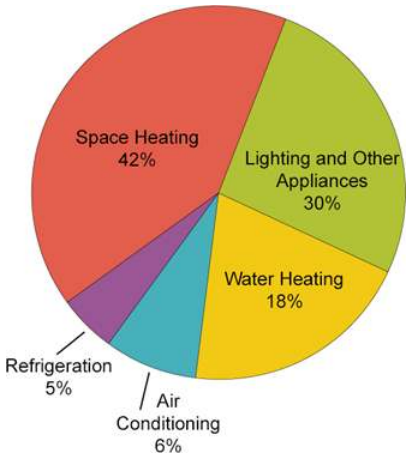


Figure 9.4.1:A "pie-style" chart showing how energy is used in an average American home – 2009 data from the U.S. Energy Information Agency Web page.

If geothermal energy could be used for this purpose instead of electricity, natural gas, or heating oil, the total CO_2 emission from American homes could be significantly reduced.



Figure 9.4.2: The distribution of geothermal resources in the US. Red color indicates very good conditions, orange – favorable conditions for installing binary cycle power plants, or for using geothermal water directly for heating. And the other two colors shows the areas where geothermal resources can be perhaps reached by very deep drilling, which, however, would make the costs of such ventures economically unjustified. Nevertheless, the homeowners in such areas (as well as in the other two areas) may reduce the heating costs considerably by using the so-called *shallow geothermal heating* (source: EIA

Well, but the country's geography is not particularly supportive for such a noble idea! It is evidenced by Fig. 9.4.2, in which the distribution of the "geothermal potential" over the entire U.S. area is shown.

9.4: Geothermal Heating is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

9.4.1: Shallow Geothermal Heating

As was discussed in the Chapter describing heat engines, it is the Second Law of Thermodynamics which does not allow us to take advantage of all energy contained by fuels – one is tempted to say that the 2nd Law "steals" 2/3 or more of the thermal energy one puts into a thermal engine to be converted to mechanical work. Such a theft is outrageous, isn't it?

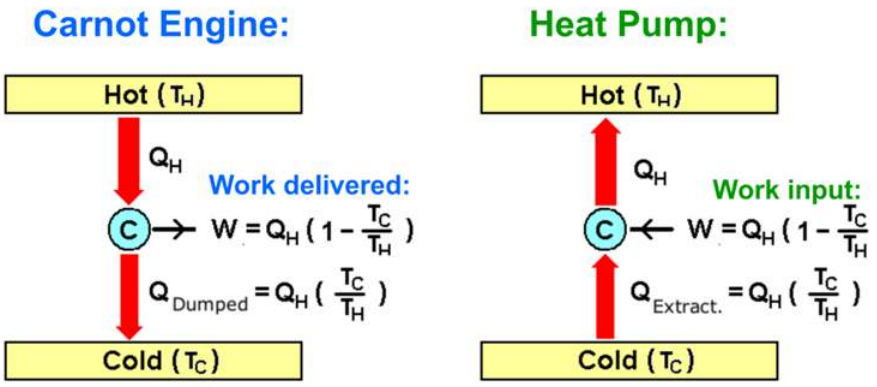


Figure 9.4.1.1: Left image: a Carnot engine – an ideal heat engine converting heat to work with maximum efficiency. Right image: a heat pump – a Carnot engine working in reverse. Work is the input, and now the engine transfers heat from the cold reservoir to the hot reservoir source: aop).

However, it was Lord Kelvin – one of the greatest physicists of the 19th century – who in 1855 pointed out that in certain situations the 2nd Law can work in our favor. Namely, not when we convert heat to mechanical work, but – conversely – when we convert mechanical work to heat. Not always – only when we use a machine called the "heat pump".

Let's us quickly recall how an ideal heat engine – the Carnot engine – works. Most thermal engines we deal with work in *cycles*. Let's analyze what happens in a single cycle. A portion of heat Q_{H} (we will call it the "heat input") is transferred to the engine from a "hot reservoir" of temperature TH. It means that at the same time a portion of entropy $\mathrm{SH} = Q\mathrm{H}/T_{\mathrm{H}}$ "enters" the engine.

Now, the engine converts some of the input heat QH to work W (we will call it the "work output") - and transfers the remaining heat part Qc to a "heat sink", i.e., to "cold reservoir" of temperature T_{C}.

From the First Law of Thermodynamics it follows, of course, that:

$$W = Q_{\mathrm{H}} - Q_{\mathrm{C}} \tag{9.4.1.1}$$

Now the transfer of Q_{C} to the cold reservoir of temperature T_{C} is associated with a transfer of entropy ${}^{1}\mathrm{SC}_{\mathrm{C}} = Q_{\mathrm{C}}/T_{\mathrm{C}}$ to this reservoir.

In order to invoke the Second Law, we have to start thinking of the three elements (hot reservoir) + (the engine) + (cold reservoir) as of a single **SYSTEM**. A portion of entropy SH_{H} is taken out of the hot reservoir, and then a portion S_C is deposited in cold reservoir. Accordingly, the **change** in the entropy of the **SYSTEM** is:

$$\Delta S_{\mathrm{SYSTEM}} = Q_{\mathrm{C}} - Q_{\mathrm{H}} \tag{9.4.1.2}$$

Now, what does the Second Law say? It says that the entropy of the **SYSTEM** may only remain constant, or it may increase. The Second Law strictly forbids it to decrease!

Well, therefore let's think of the most favorable course of events - namely, that the entropy of the **SYSTEM** does not change, i.e., Δ System $= 0$. It means that $S_{\mathrm{C}} = \mathrm{SH}_{\mathrm{H}}$, or:

$$\frac{Q_{\mathrm{C}}}{T_{\mathrm{C}}} = \frac{Q_{\mathrm{H}}}{T_{\mathrm{H}}} \tag{9.4.1.3}$$

The Equations 9.4.1.1 and 9.4.1.3 together constitute a system of two equations we can use to obtain the relation between W and QH. From the latter one, we get that:

$$Q_{\mathrm{C}} = Q_{\mathrm{H}} \frac{T_{\mathrm{C}}}{T_{\mathrm{H}}} \tag{9.4.1.4}$$

and by substituting this result to the Eq. 9.1, we obtain:

$$W = Q_H - Q_H \left(\frac{T_C}{T_H} \right) = Q_H \left(1 - \frac{T_C}{T_H} \right) \tag{9.4.1.5}$$

The ratio of the heat input to the work output we call the efficiency of the thermal engine and we use the symbol of ε for it:

$$\varepsilon = \frac{W}{Q_H} = 1 - \left(\frac{T_C}{T_H} \right) = \left(\frac{T_H - T_C}{T_H} \right) \tag{9.4.1.6}$$

where the subscript " E " indicates that ε is meant to be the efficiency of an engine. The result, often referred to as the Carnot Law 5, tells us that the lower is the difference between the hot source temperature T_{H} and the heat sink temperature T_{C}, the lower is the engine's efficiency. The lower is the difference, the more the Second Law steals from us!

Now, let's consider what the Lord Kelvin's heat pump does. As shown in Fig. 9.4.1.1, the pump is nothing else than the heat engine we discussed above, but working in the opposite direction. In a single cycle, the pump "sucks in" a portion of input heat Q_{C} from the cold reservoir. A portion of input work W is also sent into the engine, where it is converted to heat, and added up to the heat input. The sum becomes the heat output Q_{H} :

$$Q_{\mathrm{H}} = Q_{\mathrm{C}} + W \tag{9.4.1.7}$$

which is deposited into the hot reservoir. And again, how it is with the entropy. When the portion Q C is transferred from the cold reservoir to the pump, it's corresponds to the entropy transfer of $Q\mathrm{C}/T\mathrm{C}$. And the transfer of heat from the pump to the hot reservoir corresponds to an entropy transfer of $Q_{\mathrm{H}}/T_{\mathrm{H}}$. Now, using the same reasoning as before for the heat engine: in the case of an ideal heat pump, the entropy change of the combined system (cold reservoir) $+($ pump $)+($ hot reservoir) is zero, so that:

$$\frac{Q_{\mathrm{H}}}{T_{\mathrm{H}}} = \frac{Q_{\mathrm{C}}}{T_{\mathrm{C}}} \tag{9.4.1.8}$$

As before, the Eqs. 9.5 and 9.6 make a two-equation set, but we solve it now not for W, but for Q H. The reader will easily find that:

$$Q_{\mathrm{H}} = \frac{T_{\mathrm{H}}}{T_{\mathrm{H}} - T_{\mathrm{C}}} W \tag{9.4.1.9}$$

We define the efficiency $\mathcal{E}_{\mathrm{P}}$ of the heat pump as the ratio of the output heat QH to the input work W :

$$\epsilon_{\mathrm{P}} = \frac{Q_{\mathrm{H}}}{W} = \frac{T_{\mathrm{H}}}{T_{\mathrm{H}} - T_{\mathrm{C}}} = \frac{1}{\epsilon_{\mathrm{E}}} \tag{9.4.1.10}$$

So, the efficiency of a heat pump appears to be an inverse of the efficiency of a heat engine. What is the practical meaning of this result? The best answer we get by working an example. Suppose that the temperature of a "cold reservoir" available to a homeowner in Corvallis is $12^{\circ}\mathrm{C}$ (in a moment, it will be shown that this is a reasonable value). The homeowner uses a heat pump and a conventional forced air central heating system that blows air of $60^{\circ}\mathrm{C}$ temperature into her rooms. Let's first convert the temperatures to Kelvins: the temperature of the cold reservoir is $T_{\mathrm{C}} = (273 + 12)\mathrm{K} = 285\ \mathrm{K}$. The temperature of the hot reservoir is $(273 + 60)\mathrm{K} = 333\ \mathrm{K}$. So, the efficiency of the homeowner's heat pump is:

$$\epsilon_{\mathrm{P}} = \frac{333K}{333K - 285K} = 6.94 \tag{9.4.1.11}$$

The homeowner’s pump needs a mechanical work input, right? In all avail- able heat pumps the mechanical work is delivered by electric motors. The efficiency of electric motors is very high. 90% is not unusual. So, by adjusting for the motor efficiency, the heat power the pump delivers to the house is 90% 6.94= 6.25 times the electric power used by the motor.

Before the homeowner installed the heat pump, she had used the old- fashion electric heating – the so-called “resistance electric heaters”. Their efficiency is 100%, meaning that 100% of the input electric power is converted by them to heat power. So, the conclusion is pretty straightforward: now, to get the same heat power as before, she needs 6.25 times less electric power than she needed before! A great deal, isn’t it?

Well, it would be so if her heat pump were *ideal*. But the efficiency of real heat pumps is lower. It may be even as low as 50% of the ideal E_P. So, it would be no 6.25, but only 3.12 times less. It still seems like a pretty good deal.

Professional heat pump installers do not use the symbol ε_P and term "efficiency", instead they use the "Coefficient of Performance" (COF) which is essentially the same as the ε_P defined by us as Q_H/W .

The COF depends both on the temperature difference T_H T_C and on the construction of the heat pump mechanism (i.e., on how close it is to an "ideal heat engine"). In practice, the COP of the real machines varies from 3 to 7.

The homeowner in our example would do even better if she used not the forced air central heating system (the most common in the US), but underfloor heating system in which water of temperature 45° is circulated through a "lattice" of tubes hidden under the floor surface. Then, according to the Eq. 9.7, the ideal COP value would be 9.6, and even if it's lowered by 50% by the non-ideal design of the pump, it still would be an impressive 4.8.

By the way, it's not a common knowledge that one can find a heat pump in almost every American kitchen. Because any refrigerator is nothing else than a heat pump. The cold reservoir is the refrigerator's interior. And the hot reservoir is the ambient air. The heat pumps used for heating homes have a similar design as the machinery that is hidden inside your refrigerator. Only it's much larger. More details of how a heat pump is built and how it

works can be found in several Web pages, for instance, in this one. There are several "clickable" items in this Web page, e.g., if you click on the *Coefficient of Performance* one, you'll find much information about various factors that determine the COP value in practical heat pumps.

Shallow Geothermal Heating Implementation

How can a heat pump be combined with geothermal energy? Well, this is simple. In always all areas inhabited by humans the subterranean tempera- ture at the depth of 10 ft and more is constant all over the year and is equal to the average annual air temperature in a given region. In the Willamette Valley, the annual average temperature of air is about 53°F, which corresponds to about 12°C (it's why we used this temperature in the example above). In the US the subterranean temperature varies from 10 to 16 °C. Therefore, by burying a long enough system of pipes with circulating water at the depth of 10 ft. or more, one can make an excellent "cold reservoir".

An additional advantage of a heat pump is that it can work "in reverse", so that in the summertime it can act as an air conditioner.

Other commonly used "cold reservoirs" are the air outside, or water in a creek or a pond, if they are nearby. Using outside air is the simplest and the most often used option. One can often see heat pumps in the backyards of people's homes – they are equipped with huge fans that one can see and hear if the pump is on. But a disadvantage of such air cooling is that the lower is the temperature outside, the lower is the pump's COP (one can clearly see in the Eq. 9.7 that if the temperature T_C decreases, then also the value of ε_P, or CPO, gets lower. Therefore, in areas where there are severe winters, homeowners have to install backup heaters they can turn on when the outside temperature falls so much that the heat pump practically stops working. And if the system is used as an air conditioner in summertime, a very hot outside air temperature lowers it efficiency.

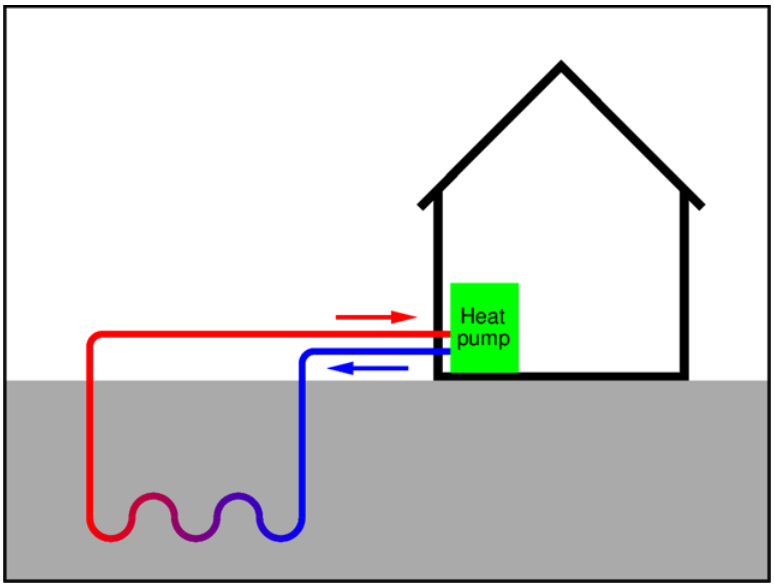


Figure 9.4.1.1: : Shallow geothermal heating scheme. By reversing the action of the heat pump in summertime, the system changes into an air conditioner (source: aop).

Using water in a nonfreezing pond or in a nonfreezing creek as the cold reservoir has the advantage that the water temperature never falls below 0°C (32°F). So, there is no fear that the heat pump may stop working on a very cold winter day. But if the water in a pond or creek becomes warm in summertime, it may lower the efficiency of the system working in the air conditioning mode.

The advantage of the shallow geothermal system is that the temperature of the subterranean cold reservoir stays constant all the year round. So, the outside weather does not affect in any way the performance of the system, no matter if it's in the heat pump mode, or in the air conditioning mode. The costs of installing a system of pipes ten or more feet underground (necessary), and a system of heating tubes under the floor (optional) may be discouraging. Yet, these are only one-time investments that will be paid back due to significantly lower heating bills.

---

1. An important reminder: the temperature in the expression for entropy, $S = Q/T$, must always be taken as the absolute temperature, in Kelvins, and never in centi- grades °C or in Fahrenheits °F.

2. Let's keep in mind that the Carnot Law is the efficiency of an ideal heat engine. In Chapter 3 it has been discussed how a real engine differs from an ideal one - namely, due to the irreversible processes associated with its functioning, it adds a portion of "extra entropy" $S_{\text{EXTRA which}}$ has to be added to the right side of the Eq. 9.3. The calculations yield then an efficiency that is lower than $\boldsymbol{E}_{\text{E}}$ in the Eq. 9.4.

9.4.1: Shallow Geothermal Heating is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

10: Biofuels

The thermal energy we can harness by burning biofuels is yet another form of converted solar energy. How comes? Bioufuels are made of plants, with various degrees of processing. The oldest kind of biofuels known to humanity is firewood – the processing is pretty simple, on has just to cut it, dry it, and then cut to smaller pieces. More complicated processes are needed for making more sophisticated types of biofuels – for instance, for making biodiesel one needs to extract oil from oily seeds (may be the same cooking oil is extracted from), and next the oil, before being ready for pouring it into a fuel tank, has to be subjected to a process called “esterification.” Making ethyl alcohol (ethanol), a substance now added to gasoline, requires even more steps – first, one has to grow corn (maize), then extract starch from the seeds, then convert it to sugar by mixing it with special enzymes, then subject the sugar to alcoholic fermentation using yeast, and finally, to distill out the ethanol. But the first stage is always the same – before any processing may start, a plant has to grow first from a seed sown in soil.

The main problem with fuels is that by burning them we add CO_2 to the atmosphere. No matter whether it is a fossil fuel, or biofuel – burning it always produces CO_2. Yet, there **is** a difference. Before the biofuel can be obtained from a plant, the plant has to grow from a seed – it has to build its trunk, its stalk and its seeds – and where does it get the “bulding material” from? Well, the answer is simple: the two essential components the plant needs are CO_2 and water, H_2O. Plus sunlight. The plants use a biochemical mechanism, catalyzed by **chlorophyll** – a substance that makes all plants green – to combine, with the help of sunlight, CO_2 and H_2O molecules into larger molecules of “carbohydrates”. The simplest carbohydrates is **glucose**, a sugar. By combining several hundred glucose molecules into long chains, the plant make **starch** – and another material, **cellulose**, in which many glucose chains are bundled together, forming fiber-like molecules. Cellulose is a “scaffolding” material from which plants make cell walls, as well rigid structures needed for forming stalks. Another important carbohydrate is lignin. Cellulose and lignin together produce especially rigid structures – one that we all know is wood.

As follows from the above, plants when they grow remove much CO_2 from the atmosphere – and when we burn biofuels, we don’t add new CO_2 to it, we only return to it CO_2 which has been taken away from it some time before. Therefore, if one day we managed to eliminate all fossil fuels and to use biofuels exclusively, CO_2 would be only “cycling” – that released by burning would be used by next plant generations to build their stalks, leaves and seeds, and the concentration of CO_2 in t4he air would stop growing.

10.1: The History of Biofuels

10.2: New Solid Biofuels

10.3: Liquid Biofuels

10.3.1: Bioethanol

10.3.2: Biodiesel Fuel

10.4: Other Ways of Harnessing Plant Energy- “Bioelectricity

10: Biofuels is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

10.1: The History of Biofuels

Humans started using biofuels many tens thousand years ago – twigs, kindlings, and larger pieces of wood for roasting their foods and for heating their caves. An important observation of ancient people was that clay kept for a prolonged time in fire changes into a solid – it led to a real revolution, opened the *era of pottery* about 26,000 BCE. Ceramic vessels added new applications for biofuels. For instance, a pot filled with vegetable or animal fat, with a wick inserted, was an early ancestor of toady's lamps.



Figure 10.1.1: Early pottery oil lamp (source: Wikimedia Commons)

Until the 18th century when people started mining coal, firewood was essentially the only fuel used for heating homes. Whale oil was used in lamps until it was replaced by kerosene in the second half of 19th century. Before the end of that century American steam locomotives used firewood as fuel.

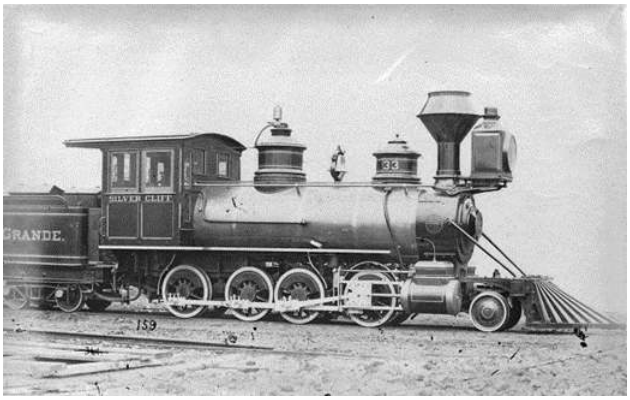


Figure 10.1.1: Copy and Paste Caption here. (Copyright; author via source)

Do you know why the old American locomotives carried a big funnel-shaped smokestack? European locomotives from the same period had much smaller smokestacks.

The answer is: because European locomotives were running on coal, not on firewood. Burning firewood produces large sparks and embers, sometimes they are flaming pieces of wood! When American trains were crossing praeries in the dry season, such sparks could give rise to catastrophic fires. Inside the funny-looking smokestacks there were devices called *spark arrestors* which did not let the sparks to get through. Burning coal produces sparks, too – but not as large as those from firewood. In European locomotives the smokestacks were compact and hidden inside the locomotive's belly.

In the 20th century simple biofuels like firewood were "running out of favor" they were used mostly in leisure activities, e.g., for barbecuing, for having a good "family time" at a fireplace, etc.

10.1: The History of Biofuels is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

10.2: New Solid Biofuels

However, firewood is now “returning with vengeance”, as *wood pellets*. Single- home heating installations using firewood in such form are fully automatic and require little maintenance. The wood-pellet industry in the US is growing fast. Advantages: wood pellet heating systems emit no nitrous oxides (NO_x), no sulfur dioxide SO_2. The only product of combustion is CO_2 – but it does not add “new” CO_2 to the atmosphere, it only – as has been discussed earlier – “recycles” the natural CO_2. Moreover, there is much wood available for making pellets. There is one problem, though – it is called the “pollution by particulate matter”. The term *particulate matter* refers to microscopic solid or liquid particles suspended in the air. Until recently, the pollution by particulate matter was not the principal concern of institutions responsible for the cleanliness of air. But more and more results of studies pointed out to the odd effects of particulates on human lungs.

What first raised the awareness of public opinion to the seriousness of particulate matter pollution were diesel cars. Until recently believed to be the most environmentally friendly of all automobiles, they have been recenly found to emit too much particulate matter (and too much NO_x, either). Some big cities in Europe (where diesel cars are much more popular than in the US) plan in the near future to forbid diesel cars to enter the city centers. Some countries even consider banning sales of new diesel cars altogether in not-so-distant a future.



Figure 10.2.1: Wood pellets in someone’s hands, in order to show what the actual pellets’ size is (source: Wikimedia Commons).

Also, the wood pellet stoves, considered to be “the most green” of all house- hold heating technologies, started attracting attention. Even though they emit no NO_x, there is a growing concern about particulate matter in their exhaust gases. Fortunately, removing the particulates from the gases is tech- nically feasible – so there is no fear that pellet stoves may be banned altogether. But if new laws requires that all new stoves be equipped with filters efficiently removing the particulate matter from the gases, the prices of such heating installations my raise considerably. At the moment this text is writ- ten (2018) it’s not yet clear what may happen – at the moment, we can only “wait and see”.

There are still other possible applications for solid biofuels – we will return to this issue later, after discussing possible applications of liquid biofuels and methods of their production.

10.2: New Solid Biofuels is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

10.3: Liquid Biofuels

Of all sectors of American economy, the two that are the largest emitters of CO_2 are the electric power sector (29%), and transportation (27%). The figures may be different for individual states. The pie chart in Figure 10.4 show how much CO_2 is emitted by each sector of US economy, and Fig. 10.5 how much fuel of different kinds is consumed by the US transportation.

As can be seen in the graph, an overwhelming part (over 90%) of the US transportation is powered by liquid fuels: gasoline, diesel fuel, and jet fuel. In other states the percents may be somewhat different. But the overall picture for the entire country is essentially the same – nearly all American transportation is powered by these three kinds of fuels, and they all are derived from crude petroleum oil.

The good news is that currently there exist technologies of producing biofuels that can replace the currently used petroleum products. For diesel engines the conversion to biofuel is pretty simple – just fill the tank with biodiesel, and that's it. There exist technologies of making bio-aviation fuel equivalent to petroleum jet fuel. And replacing gasoline with its "number one" bio-substitute – namely, bioethanol – may require some modification of the fuel injection system in the engine, but not replacing the entire engine.

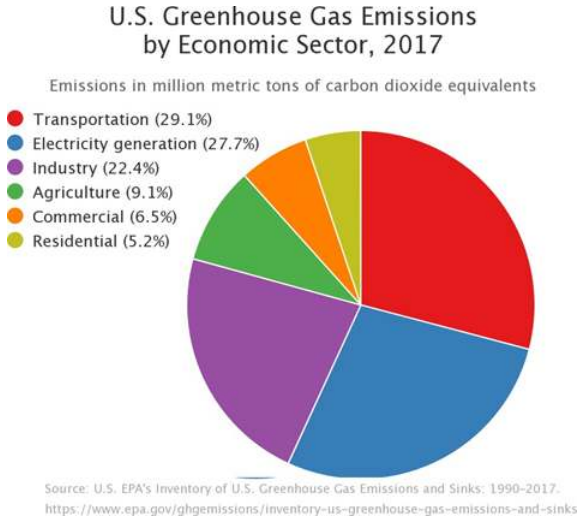


Figure 10.3.1: CO_2 emission of the US economy by sector (source: US EPA).

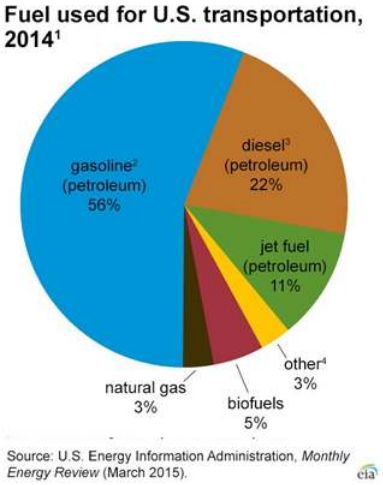


Figure 10.3.2: : The consumption of different kinds of fuels by the US transportation sector (Source: US Energy Information Administration, 2015).

A serious interest in biofuels for cars and trucks has erupted for the first time after the "oil crisis of 1970-s". Since then, the interest keeps alternating – it increases at times when oil and gasoline prices are surging, and dwindles when the prices fall. The prices since 1929 exhibit many maxima – the most recent price surge happened around the year 2007, mostly because the crude oil prices grew to the level well over $100 per barrel, and after a short break they stayed at a level close to $100/barrel until 2014. Over that

period, there was much interest in biofuels in the US. Then the oil and natural gas prices dropped considerably, due to the application of novel “fracking” drilling techniques that have brought the US close to the state of “energy independence”. In the 2007-14 period the prices of biofuels were still higher that those of the petroleum products, but there was much hope that eventually the bio- prices might become competitive with the petroprices. But after 2014 those hopes disappeared, and, understandably, the interest of researchers and industrialists in biofuels markedly dwindled.

Nonetheless, no matter how “uncompetitive” the biofuel prices are, nothing can change the fact that their “emission intensity” will always be lower that of petroleum fuels. And if one day the emission intensity of a fuel be- comes more important than its price, then research on biofuels may be put on the “fast track” again.

10.3: Liquid Biofuels is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

10.3.1: Bioethanol

People have learned how to produce ethanol several thousand years ago. The method was simple – make a pulp out of ripe grapes, put it into a clean vessel, and wait. Grapes contain sugar – most fruits do, it's their method of storing energy. In grapes there are three varieties of sugars, two simple known as *glucose* and *fructose*, and one complex, *sucrose*, whose molecule consists of one glucose and one fructose molecule.

On the grapes' skin, there is always *yeast* – a simple organism single-cell organism, a member of the large kingdom of *fungi*. Yeast feeds on sugars, the energy it acquires from that is used for multiplying the cells. The simplified chemical formula for both glucose and fructose is $C_6H_{12}O_6$. When the grapes are pulped, the yeast comes into direct contact with the sugars. They start to process them – what's happen is called the *ethanol fermentation*

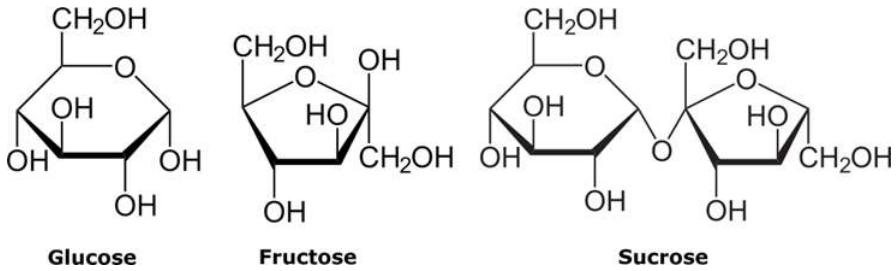


Figure 10.3.1.1: The structure of glucose, fructose, and sucrose molecules

The process involves several steps catalyzed by some yeast enzymes, but the overall chemical reaction can be written as:

$$C_6H_{12}O_6 \longrightarrow 2C_2H_5OH + 2CO_2 \tag{10.3.1.1}$$

where C_2H_5OH is the chemical formula of ethanol. The fermentation of sucrose is a bit more complicated – first, the yeast uses an enzyme to split the molecule into one glucose and one fructose molecule, and then process them in the same way as described above.

What is created as the result of fermenting grape juice is *wine*. The final product usually contains 12%-13% of ethanol. Long time ago, our ancestors discovered that drinking wine brings good mood (sometimes, even too good a mood). So, wine has become an important ingredient in the cultures of many tribes living in areas where grapes can be grown.

But many people lived in regions where the climate was not favorable for grape growing. Apples and berries growing in the north were not good precursors for fermentation. Therefore, German tribes, Slavic tribes, British tribes could not enjoy the benefits of wine. But they invented something else. They learned how to ferment *starch* products.

Mono-, Bi-, and Poly-saccharides

Glucose, fructose, sucrose and starch are all members of a great family of *saccharides*. Glucose and fructose are called *monosaccharides* because their molecules are just single saccharide "unit" (the third monosaccharide "sibling" is *galactose*). As can be seen in Fig. 10.3.1.1, in the monosaccharide molecules there are many -O-H groups, sprouting out of a carbon atom

They are called *hydroxyl groups*. What may happen is that two hydroxyl groups, each from a different monosaccharide molecule, get together, "jet- tison" one H_2O molecule, and leave one oxygen atom which now links to two carbon atoms, each one belonging to a different "parent " molecule (it's referred to as an "oxygen link" or an "O-link"). In Fig. 10.6 the glucose and the fructose molecules are linked by such an O-link, forming the sucrose molecule. Sucrose is the main component of table sugar. But there are many other combinations of two monosaccharides coupled by the same mechanism – for instance, *lactose* made of glucose and galactose – we certainly know it all because it's one component of human milk. But there many other possi- ble combinations of two monosaccharides, and all of them together make a sub-family called *disaccharides*. In the Wikipedia article linked as many as 18 such combinations are listed.

If two can be linked, why not three, four, ...? Yes, they can. Groups made up from 3 to 10 O-linked monosaccharides re called *oligosaccharides* – from *oligo*, meaning "a few" in ancient Greek language. And even larger groups are called *polysaccharides*. And starch is perhaps the best known representative of the latter category.

Many plants create starch as an energy-storing substance by O-linking several hundred glucose molecules into chains – either a single chain, called an *amylose chain*, or a main chain with several sub-chains branching out – forming a molecule of *amylopectin*. In each form, all glucose units are kept together by O-links. Per average, forming an O-link costs each glucose molecule one molecule of H_2O, so that the overall chemical formula for amy- lose or amylopectin is $(C_6H_10O_5)_n$, with n of the order of several hundred for amylose, and up to 200,000 for amylopectin.

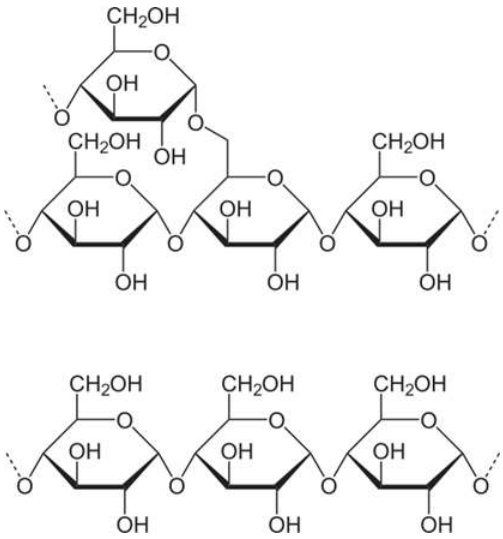


Figure 10.3.1.2: Two types of starch chains. The **lower scheme** is an *amylose* chain, formed by several hundred glucose molecules – each connected in identical fashion with two "neighbor" molecules. The **upper graph** shows a fragment of an *amylopectin* molecule, which consist of a main "backbone" chain similar to that in amylose – but at every 24 to 30 units another glucose chain branches out, forming an O-link with one of the -CH_2OH groups in the main chain. Then, new chain may branch out from each "branch". In some plants due to multiple branching the molecules may grow very large, consisting of up to 200,000 glucose units.

History Continued

As noted, the northern tribes did not have fruits rich in sugar that could be easily fermented. They did have plant products rich in starch, though – but the bad news is that yeast does not agree to ferment starch. Monosaccharides and disaccharides – yes, happily, but starch – no way!

Nonetheless, the norther tribes did find a way to obtain ethanol from starch. The key to success was *barley*, a cereal grain rich in starch – one of the first grains cultivated by humans. The northern people discovered that if barley grains were soaked with water for a short period, then dried and left for some time – they started tasting sweet! The sweet component could be extracted from the grains using hot water, and water solution of sugars obtained in this way could be fermented. And the product of fermentation was nothing else than **beer**.

The first part of this recipe is now called *malting* . What exactly happens in this process? When barley is for the first time soaked with water, the grains start germinating. In the early stage of germination the grains get ready to convert starch into another form of an "energy carrier". Why? Well, because soon the energy will be needed in new parts of the plant growing out of the grain. And starch is a material non-soluble in water. How to get it out of the grain? Mother Nature had found an ingenious trick: convert starch to sugars that are water soluble, and then "pump" the solution to the sprout and to the emerging stalk and root system.

So, in the early stage of germination an enzyme called *amylase* is produced in the grains. Its task is to convert the starch into sugars. The enzymes are essentially catalysts – they allow a chemical reaction to occur, but they are not "consumed" by it. So, if the germinating process is terminated at the right moment – with amylase already there – the enzyme will continue to do its work, as long as there is still any starch remaining. But the sugars will stay in the grain. One should give the malting process enough time to be completed – and the next step is to extract the sugar from the grains. One can do that by immersing the grains into hot water. The sugar gets dissolved in water, and then the sugar solution, after separating it from the grains' leftovers, can be fermented by adding yeast for obtaining beer as the final product.

With time, the process was further refined. Beer typically contains 3%- 5% of ethanol. The next invention was to use distillation of the fermented malt – by such an operation, one can increase the ethanol concentration about tenfold: this is how *whisky* is produced

(the technology was developed around the year 1200, and without much change has been used until today). Other people discovered that the enzymes from germinating barley can be added to starch from other plants – e.g., from potato. The enzyme converts the starch to sugar, yielding a syrup-like product known as *molasses*. It can fermented by adding yeast – and after distillation, one obtains a final product with 40% - 50% of ethanol, known as *vodka*. By making further improvements to the distillation technique – much credit for that should be given to medieval alchemists – people learned how to obtain a distillate with ethanol concentration up to 95%. It is commonly referred to as *spirit*. It's not something one would like to drink, but it has many other applications – as a solvent, or as a disinfectant.

It is also worth mentioning that in some p;arts of the world where sugar cane grows, the process of making vodka may be much simpler. The juice pressed out from the cane contains sugar, so one can skip the process of enzymatic conversion. By fermenting the sugar cane juice and applying distillation, on obtains a product known as *rum*. At some period rum from Caribbean Islands and Middle America became perhaps the most popular alcoholic beverage. Around 1800 it was transported to destinations even as distant as Australia. In order to economize the transport – sailing from Ja- maica to Sydney could take six months or more – the rum was sent out as a concentrate, with about 80% of ethanol – and after reaching Sydney, it was diluted by adding the same volume of water.

Bioethanol as a Fuel

Why did we pay so much attention in the preceding section to the methods of producing ethanol? Well, the answer is simple: when the demand for transportation biofuel has emerged for the first time, no new technology was needed – the existing distilleries simply had to start cranking out more product. But the method remains essentially the same.

Brazil has been the pioneer in biofuel production. As far as the total volume manufactured is concerned, it was the world leader until 2006 – and it still is the world leader in imposing bioethanol fuel economy. One helpful factor that has enabled Brazil to attain this position is the fact that all Brazilian bioethanol is made from sugar cane. Like in the case of rum production, the labor end energy consuming process of enzymatic conversion can be skipped. Therefore, the Brazilian bioethanol has a competitive price and a very favor- able energy balance:

\[\text{energy balance = \dfrac{\text{energy balance}}{\text{energy input} \]

---

the average of which is 8.3 for the entire industry, and 10.2 for the newest installations.

Currently, there are no light cars running on pure gasoline in Brazil. Currently, the minimum ethanol content in the fuel blend is 27% (E27). But the "flex" vehicles that dominate the market of new cars can run on fuel consisting of any proportion of gasoline and alcohol, including pure hydrous[1] ethanol (E100).

The **United States** is presently the larger manufacturer of bioethanol fuel, with the output twice larger than that of Brazil.

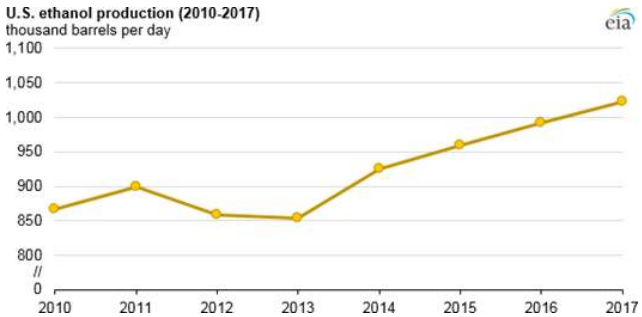


Figure 10.3.1.3: Daily production of bioethanol in the US in recent years (a graph compiled from the data published at the Web page of the U.S. Energy Information Administration)

In the US, over 99% of biofuel is made from corn (maize). Currently, two major techniques are used: wet milling and dry milling. In the former, the grain kernel is separated into its components (germ, fiber, protein, and starch) and then only the starch is further processed – like in the "classical" recipe of vodka making. The latter is a newer process, in which the grain kernels are first ground into flour – then the process is continued the same way as for pure starch, and the non-fermentable parts are removed only at the later stage of the process. Now most of the operating facilities in the US use the dry milling technique.

There is one sad truth about American corn ethanol that should be made clear. Above, it was mentioned that the average energy balance of Brazilian sugar cane ethanol is 8.3, and in the newest production facilities it is as high as over 10. But for American corn ethanol the value is a miserable 1.3. The Author of this text does not want to make any comment – the Reader is kindly asked to draw her/his own conclusion.

What is the corm biofuel made by American distilleries used for? Ten percent of bioethanol is now mandated to be blended into gasoline in most American states. Lowering the CO_2 emission is one goal, but perhaps a more important reason is that ethanol acts as an anti-knocking agent and as a *oxygenate*. Anti-knocking additives enhance the octane number of gasoline, and oxygenates cause that the gasoline burns more completely, thus reducing the tailpipe emission. It is expected that soon E15 will be mandated, i.e., a blend with 15% of ethanol. But a significant reduction of CO_2 the E85 blend is recommended. It is already available in some states, but it can be used only by cars with "flexible fuel" engines – at present, only about 10% of all cars in the US are equipped with such engines.

The data shown in Fig. \(\PageIndex{3}\ indicate that currently there is a clear growing tendency in the production of corn biofuel in the US. However, as follows from a recent analysis, it is hard to expect that eventually there will be enough alcohol to replace most of the gasoline used by Americans. Currently, over 40% of all corn grown in the US is used for making alcohol – with the output approaching 20 billion gallons per year. But the country uses over 130 billion gallons of gasoline, and over 50 gallons of diesel fuel per year. Therefore, even if 100% of corn is used for making fuel, it will be only enough to replace 25-30% of all the petroleum needed by the entire American transportation.

Well, this is not a very optimistic perspective... People and farm animals need to eat something, so even 25-30% is not a realistic figure. Is there any remedy for such a gloomy forecast? Yes, there is. Corn is not the only raw material for making ethanol that is available in the US. Stay tuned, please!

More Polysaccharides – Cellulose and Hemicellulose.

Earlier in this chapter it was mentioned that plants photosynthesize CO_2 and H_2O to make sugars, starch, and what else? Cellulose! The third photo- synthesis product is cellulose, a material used to make cell walls, and stalks needed to grow high. In a ripe corn plant, there is more cellulose than starch and sugar combined.

Is it possible to make ethanol from cellulose? Yes, definitely, because cellulose is made of the same "building blocks" as starch, i.e., of glucose molecules. The structure of cellulose is more complicated, though. The structure of amylose, the linear starch molecule, is shown in Fig. 10.3.1.2. Note that all glucose molecules forming the chain are aligned in the same way.

Cellulose also consists of linear chains of glucose molecules connected by the "O-links" – with one small but essential difference: look at the upper chain in. Fig. 10.9 and imagine that every other glucose molecule is "turned upside down" – in other words, is rotated by 180° around the chain's long axis – as in the chain below.

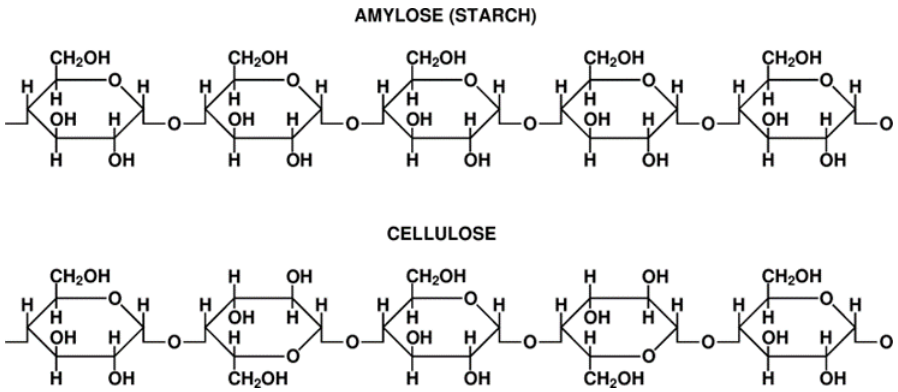


Figure 10.3.1.4: The arrangement of glucose molecules in amylose (starch) and in cellulose strands. Note the hexagonal figures in the plots representing individual glucose molecules: there is an atom in each "corner", but only the oxygen one is shown. By a commonly accepted convention, tne carbon atoms located at the other five corners are not shown, because if they were, the figure would become "too messy (source: plotted by the author).

It is instructive to make a model of a single cellulose strand using small plastic balls, as shown in Fig. \(\PageIndex{5}\. Here the black ball symbolize the carbon atoms, the red balls – the oxygen atoms, and the small gray ball – the hydrogen atoms. The rods

connecting the atoms represent the chemical bonds – the so-called *covalent bonds* which are the main type of coupling that binds atoms in organic molecules together.

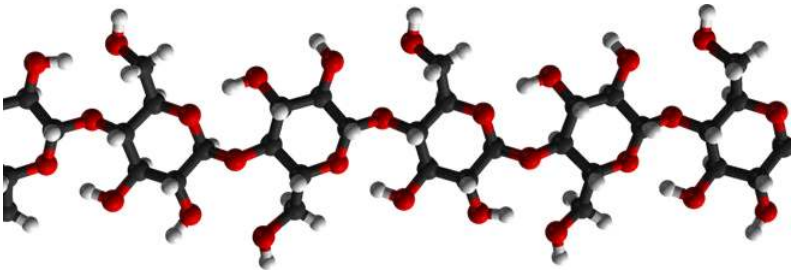


Figure 10.3.1.5: A ball-model showing the arrangement of individual glucose molecules in a single strand of cellulose (source: Wikimedia Commons).

The physical mechanism of covalent bonding is such that each one of the two coupled atoms contributes to the binding one electron from its outer electronic shell, and these electrons are "shared" by the two atoms. The electrons in the outer shell are referred to as the *valence electrons*. Carbon has four of them, therefore it may be coupled with as many as four other atoms. Oxygen has two valence electrons, and hydrogen only a single one.

In Fig. 10.11 there are two parallel elementary cellulose chains. What can be seen in this figure is that some of the hydrogen atoms from the "upper" chain get pretty close to the oxygen atoms from the "lower" chain – and *vice versa*. Between these hydrogen-oxygen pairs there are small orange two- headed arrows. What do they represent?

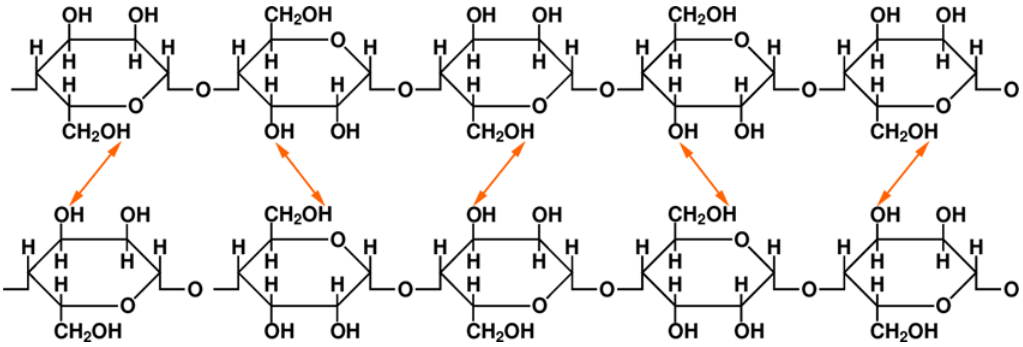


Figure 10.3.1.6: Two parallel strands of glucose molecules in a cellulose macromolecule. The orange two-headed arrows between some H-O pairs show where hydrogen bondings form between the two chains. A plot like this one, in which all constituent molecules are "flat", does nt explain too well **why** the bonds form between the H-O pairs indicated, and not between some other H-O pairs. It's better explained when the strands are plotted in a quasi-3-dimensional perspective –as, e.g., in in this Wikimedia plot – where, in addition to H-O bonds connecting parallel strands, also H-O bonds forming **within** individual strands ("intra-strand" bonds) are shown. Those internal bonds significantly strenghten individual strands. In contrast, such bond cannot form in amylose chains, so it's much easier to break them down into individual glucose molecules.

The H and O atoms forming an -OH group – such combination is commonly referred to as the *hydroxyl group* – are coupled by a covalent bonding, meaning that both H and O contribute one electron and then they "share" the electron pair. But you know how it often is with sharing... Oxygen in this partnership behaves like the proverbial "big brother" – well, let's not go too deep into details, but the truth is that the two electrons are not equally shared by H and O, but they spend more time closer to the O atom than to the H atoms. So what happens? Well, the O atom effectively acquires a bit more negative charge than it normally has, and the H atom has a bit less negative charge than it normally has. Atoms normally are neutral, right? So, effectively the H atoms becomes positively charged, and the O atom becomes negatively charged. And as we know very well, positive and negative charges attract each other, it's called the *Coulomb interaction*. Consequently, due to the attractive forces between individual H-O pairs, there is a net Coulomb coupling between the two chains over their entire length.

But can only two chains get coupled? Of course not! Each chain in Fig. 10.3.1.5 may get couple to another chain, these "new" chains to more chains, effectively forming a sheet – well, not necessarily, a sheet may be changed to a fiber by rolling it in such a fashion that the elementary glucose chains are parallel to the fiber axis. Such fibers have a crystal structure – i.e., they represent a highly ordered arrangement of atoms with clear periodicity.

One cellulose fiber contains over 10 000 elementary glucose molecules and is up to 5 micrometers long. From fibers one can make a great variety of good things, as we know. Most of the clothing we wear is made of fibers. Of different kinds, yes, but anything made of fabric is made of fibers.

Cellulose is the main "construction material" that plants use for building objects with a variety of sizes, from cell walls whose dimensions are of the order of a micrometer (10^{-6} m) to tree trunks that in some species (such as, e.g., redwood) reach the height of 100 meters. However, fibers alone can not give plant tissues all necessary rigidity. To attain that, plants combine cellulose fibers with three other structural components: *hemicellulose* and *pectin*, which are polysaccharides, and *lignin*, which is a complicated biopolymer whose structure is not related to sugars.

In this text the interest is focused on biomass that is of potential use for making biofuels. Biomass containing pectins (such as, e.g., hay) has a high value as forage. But there many other "biomasses" that cannot be fed to animals – many kinds of residues e.g., straw, corn or soybean stover, bagasse (what remains after sugarcane is crushed to extract sugar) left after harvesting – as well as sawdust or wood chips. They all contain cellulose and hemicellulose which can be, at least partially, converted to biofuels. For instance, dry corn stover contains about 70% of cellulose and hemicellulose, and 15-20% of lignin. In softwood and in hardwood the proportions are pretty similar. Biomass for making fuels have not necessarily be residues, there are some plants worth growing with the intention of using them whole – here, much attention is paid to *s witchgrass* which does not require good soil and much water, but grows up to 10 feet tall and may yield several tons of biomass per acre.

Cellulosic Ethanol – History

Cellulose and hemicellulose are the two components that can be converted to ethanol. There are several existing technologies, some of them low-tech and some high-tech. The first commercial production of cellulosic ethanol started 120 years ago in Germany. Wood chips were treated with sulphuric acid H_2SO_4 at elevated temperature, which causes a "hydrolysis" of polysaccha- rides – i.e., their decomposition to elementary sugars. The resulting "soup" consists of acid + dissolved sugars + lignin. The latter is not affected by the hydrolysis process and can be later filtered off. The acid can be neutralized by adding crushed limestone, which is nearly pure calcium carbonate $CaCO_3$. It reacts with sulphuric acid, yielding calcium sulfate:

$$CaCO_3 + H_2SO_4 \longrightarrow CaSO_4 + H_2O + CO_2 \qquad (10.3.1.2)$$

which has a very low solubility in water and precipitates from the solution. What remains is a solution of sugars – and now it can be processed the same way as the sugar solution in the techniques of making alcohol from sugar cane or from starch: just by adding yeast and fermenting the sugars. After gaining experience, before the outbreak of World War One the German installations were able to produce up to 50 gallons of ethanol from one ton of sawdust or wood chips. It was called "wood ethanol" and was definitely not good for human consumption because of some nasty contaminants, but it was pure enough to be used in industrial chemical processes or as a solvent.

During World War One the US industry needed a lot of ethanol, so the technology was "imported" from Germany and even though the American installations yielded only about 20 gallons from a ton of wood chips, the total volume of the wood ethanol made in America soon became much higher than that made in Germany. Later, also the American technology was gradually improved and reached the German yield of 50 gallons per ton. During the World War Two years wood alcohol was a crucial ingredient for manufacturing synthetic rubber – it was badly needed regardless of the price. But the price was pretty high, so the production plants were shut down shortly after the war.

In fact, the process of decomposing cellulose into sugars has been known for millions of years before humans started experimenting with it. Wood, as we know (sometimes from sad experience that may cost a homeowner thousands of dollars!) is the food of termites. To be more accurate: the cellulose in wood is what they eat. And larger animals – cattle, sheep – can also feed on a cellulose-rich diet. Well, not exactly: they all employ much smaller "workers" to deal with cellulose – it is, microbe cultures living in their intestines. No large animals can digest cellulose directly. Only some microbes and some fungi can. And because they have been doing that for many hundreds millions of years, the Author believes that it is a right thing to put this piece of information into the "History" subsection.

Cellulosic Ethanol – Today

As noted, the acidic hydrolysis of cellulose was a costly procedure and there fore, once there was no more emergency, it has been abandoned. Well, but if microbes and fungi can decompose cellulose, why cannot we try a similar method – perhaps it will be a cheeper alternative?

The microbes and fungi use enzymes. The generic name for a cellulose- degrading enzyme is *cellulase*. There is a whole variety of them already discovered, and research on them is continuing, so even more of them may be identified in the near future. As all enzymes, cellulases are proteins. Some of them can decompose cellulose and hemicellulose into individual monosaccharides; some other may cut the cellulose chain into longer oligosaccharides consisting of several sugar molecules. From the viewpoint of cellulosic ethanol production, the most interesting are those which convert the cellulose and hemicellulose macromolecules into fermentable sugars.

Efforts of developing practical technologies of producing cellulosic ethanol involving enzymatic hydrolysis have intensified in the opening years of the 21st century. The first step is to prepare the material – in the biomass the cellulose chains are entangled with lignin and hemicellulose. Such structures have to be broken, in order to assure an "easy access" to the cellulose chains and hemicellulose complexes for the enzymes. Only them may be the en- zymes administered – and after they finish their work, the resultant "soup" of sugars can be further processed by yeast, and then distilled to extract the ethanol. A nice graph illustrating the first few steps in the cycle is shown, for instance, in this Web site. As far as the "preparation stage" is concerned – of liberating the polysaccharide chains from the access-blocking wrapping of lignin – several different methods can be used. One possible pre-treatment method uses dilute acids. Another possible approach is to use the so-called "steam explosion". In short, the biomass is packed into a container which is next filled with hot steam of pressure many times higher than the atmo- spheric pressure. The steam penetrates the biomass and fills all the voids in the dead plant tissues. Next, the container is suddenly opened, so that the high-pressure steam, seeking an escape from the voids it has drifted into, rips apart the biomass structure. How such "steam explosion pretreatment" works is illustrated in in this NC State University publication – see, in particular, the Figure 11 – or, (for even more details, see pages 10-11 in this report from the Danish Technological Institute).

But enzymatic decomposition is not the only possible process. Recently, Renmatix, a Pensylvania-based company, has developed a technology in which only "water and temperature" is used for decomposing the cellulose and hemicellulose into elementary constituent sugars. The details of the processes used are explained in the company's Web site. When the page opens, click first on the icon *TECHNOLOGY* – and then, on the *Plantrose process*, to get a verbal description of the technology – and next, on the *WATCH FULL VIDEO* icon. In short, the technology the company uses makes it possible to separate *xylose* sugars from glucose.

A few words about the xylose sugars we have not yet paid much attention to. Cellulose, as has been discussed above, consists exclusively of glucose molecules linked into long chains. But in hemicellulose glucose is only a small fraction of the constituent molecules – the majority are molecules of xylose sugar. In contrast to other monosaccharides found in plants (glucose and fructose), the molecules of which contain six carbon atoms, xylose molecule contains **only five** carbon atoms. There are several other five-carbon sugars in living organisms, and the generic name for them is "pentose sugars".

In the he sugar "soup" obtained by enzymatic decomposition of biomass glucose is the dominant component, but there is also a significant fraction (20-30%) of xylose from the hemicellulose. And in the next step of biofuel production – the fermentation – xylose is only a "neutral bystander", because no known strain of yeast can process five-carbon sugars. It's not surprising: over hundreds of millions of years the yeast fed only on fruits and other plant elements containing monosaccharides and disaccharides, but had little chance to encounter anything containing pentose sugars. So, Mother Evolution or whoever else was taking care of teaching the yeasts what was "good" and what was "not good for eating" did not see any need to teach them to feed on such sugars. But the good news is that this problem may be solved by high-tech biotechnology – it turns out that by modifying their genes some yeast strains can be "educated" to ferment xylose as rapidly as glucose.

Cellulosic Ethanol – Summary

The above subsections on cellulosic ethanol may pay a rosy picture of that technology: there is no need to use corn for making fuel for automobiles – it's immoral! – more than 10% of people in today's world suffer from hunger!.

It turns out that instead of using edible parts of corn for making biofuel, we may produce the same amount of it from corn's stover, i.e., the nod-edible parts of it. And there are plenty of other sources of cellulose – switchgrass which grows well on lands of low

agricultural value, wood debris, even municipal waste.

So, the potential is enormous – and what is being done? – the sad truth is that the situation with cellulosic ethanol is no better than one hundred years ago.

The enthusiasm for cellulosic ethanol was high in the first decade of this century and the mood in the US Government was pretty optimistic. The Energy Policy Act of 2005 created a Renewable Fuel Standard (RFS) for the U.S. The act created mandates ruling how much bioethanol from different sources had to be blended into gasoline in the years to come. However, while the corn ethanol was doing not so bad, the cellulosic ethanol completely failed to meet the expectations. The standards for it, which went into effect in 2010, mandated as many as 5.5 billion gallons in 2017. Yet, the actual production in 2017 reached only a pathetic 10 million gallons – even though two years earlier, in 2015, the nameplate capacity of all cellulosic ethanol facilities was 88 million gallons per year (for more details, please see this document; and the actual production data for the years 2010-2018 are given in this official EPA document). Why did the situation worsen so badly? Well, it seems – because of a conflict between the expectations and the brutal laws of market economy. The nameplate 88 million gallons in 2015 came from adding up the capacities of three facilities – but two of them went bankrupt since then. And what about the future? One can only "wait and see." So far, it's petroleum who is the king, and it does not seem likely that it will voluntarily abdicate in a foreseeable future.

---

1. To blend with gasoline, the ethanol must be *anhydrous*,i.e., contain no water. How- ever, the ethanol obtained by conventional distillation of water-ethanol at normal pressure cannot concentrate alcohol above 95.6%, because at this concentration the mixture becomes an *azeotrope*. So, what distilleries sell as "alcohol" is in most cases a mixture of 95% of ethanol and 5% of water; the proper name for it is rather "hydrous ethanol" – and this is precisely what the Brazilian E100 blend is

10.3.1: Bioethanol is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

10.3.2: Biodiesel Fuel

Of all thermal piston engine types, the diesel engine has the best thermo- dynamic efficiency – significantly higher than that of gasoline engine, and overwhelmingly higher than that of steam engine. The economy of diesel cars, trucks, locomotives, bulldozers and ships is therefore much better than that of their counterparts powered by other engine types. The drivers of diesel cars complained about their “sluggishness”, i.e., their poor acceleration. Therefore, over the recent decades much effort has been devoted to “improving the muscle” of diesel engines for passenger cars. New lightweight turbocharged diesel engines with sophisticated combustion chambers gave the cars a vigor similar to that of gasoline cars. However, it badly worsened the emission of noxious pollutant in the exhaust. Therefore, some European countries are seriously considering closing city centers for diesel cars. Obviously, it would mean a dramatic drop of interest in diesel cars from customers. Not surprisingly, some major manufacturers have already announced that they will discontinue the production of diesel cars.

But in the area of vehicles which are “workhorses” the death of diesel engines doesn’t seem to be so imminent. The engines for buses, trucks, tractors and locomotives don’t need engines giving them a temper of racing cars. In the case of engines that are only supposed to be “workhorses” it is much easier to design such that have lower emission than “muscle” diesels, and it is also easier to develop efficient exhaust filters for them. And there is one more reason why “workhorse” vehicles with diesel engines will not disappear soon from the roads, railway tracks, and construction sites: there is simply no alternative propulsion for them. Small passenger cars with diesel and gasoline engines may be replaced by zero-emission cars (electric or fueled by hydrogen) or low emission cars (such as plug-in hybrids), but there is no viable alternative of such kind for tractor-trailers making nearly one thousand miles per day.

So, if there are no prospects from eliminating diesel vehicles from the road in foreseeable future, what can we do? Well, we can make fuels for such vehicles that will make them as “environmentally friendly” as possible.

Petroleum Diesel Fuel

Natural petroleum crude oil is a mixture of an incredible variety of hydro- carbons, the approximate chemical formula of which can be given as C_nH_{2n}. The overwhelming majority of the molecules have n between 5 and 40. In refineries crude oil is separated into fractions, the lighter of which is gasoline, with the average value of n = 8. Diesel fuels is a mixture of molecules with carbon numbers between n = 8 and n = 24, with the average carbon number

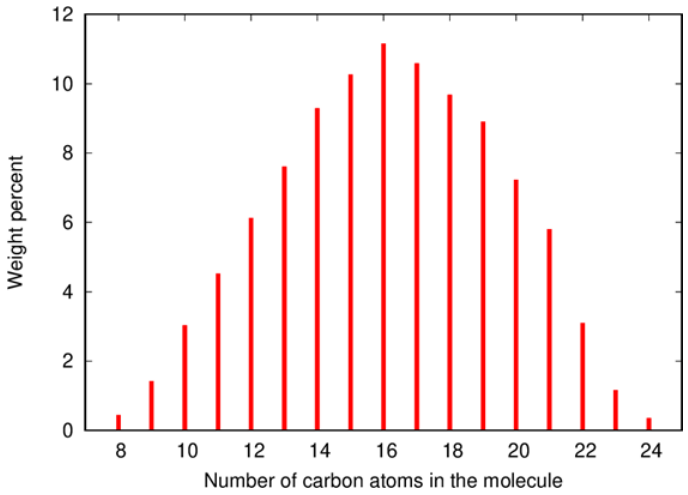


Figure 10.3.2.1: The mass distribution of hydrocarbon molecules with differ- ent numbers of carbon atoms in diesel fuel. Plot based on the results of chromatography and mass spectrometry data presented in the *Kerosene and Diesel Fuel analysis by GCMS* report, July 2017.

Kerosene is pretty similar to diesel fuel, it’s a mixture of molecules with carbon number n between 10 and 16. Kerosene is essentially the same as the aviators call *jet fuel*.

The hydrocarbon molecules in diesel fuel and in kerosene have a large variety of shapes and structures. Some of them are *aromatic* hydrocarbons containing rings made up of 6 carbon atoms. Some are linear chains with or without sub-chains branching out.

Fatty Acids and Fats

Fatty acids are important biomolecules – they do not occur alone plant or animal tissues, but they are essential "building blocs" for a class of substances known as *lipids*. Lipids play important roles in living organisms, both in biochemical processes and as structural components of tissues. For instance, cell membranes in animal (and human) tissues are made of *phospholipids*. *Cholesterol* – everybody has heard of it – is also a lipid. Our brains are made in a large part of lipids. And yet another important sub-class of the lipid family are *fats*, used both by animals and by plants as energy storing substances. In this section, we will limit our attention to fats only.

The most common fatty acid occurring in animal and plant fats is the *palmitic acid*. Its condensed chemical formula is $C_{16}H_{32}O_2$. It consists of a linear hydrocarbon chain with 15 carbon atoms, $CH_3(CH_2)_{14}–$, which constitutes a so-called *alkyl group*[2] to which there is attached a so-called *carboxyl group* –(CO)OH, which occurs in a great many of organic acids, including the well-known acetic acid (vinegar) and citric acid. The structural formula of palmitic acid is shown in Fig. 10.13:

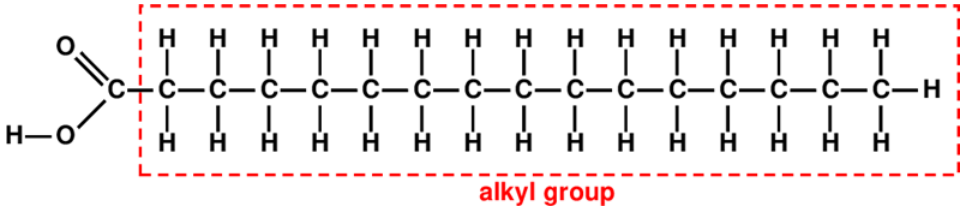


Figure 10.3.2.2: A structural formula of palmitic acid (pbta).

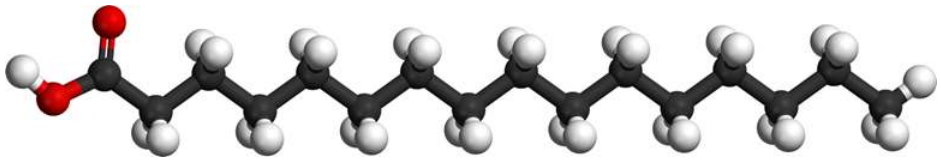


Figure 10.3.2.3: A 3D-ball model of palmitic acid (source: Wikimedia Commons).

In organic chemistry, if we attach a carboxylic group –(CO)OH to any alkyl group, we obtain an acid. But if we attach a *hydroxyl group* –OH in- stead, we obtain an alcohol. For instance, the simplest of all alkyl group (and a very important one!) is the *methyl group* $CH_3–$. By attaching a hy- droxyl group –OH one obtains CH_3–OH, methyl alcohol, or *methanol*. *Ethyl*, $CH_3CH_2–$ plus -OH yields ethyl alcohol, or *ethanol*, our good friend (but not **too good**, I hope!). And *propyl* $CH_3CH_2CH_2–$ plus –OH yields *propanol* [2].

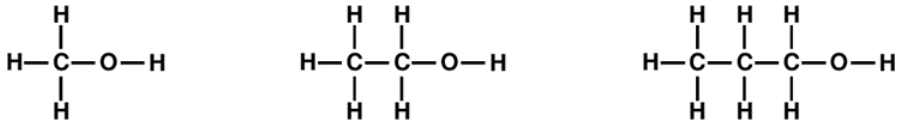


Figure 10.3.2.4: The first three primary alcohols. Primary alcohols are those in which the the carbon atom that carries the -OH group is attached to a single alkyl group (alkyl group – see the Fig. 10.13). (aop).

There may be more -OH groups than a single one – for instance, if one replaces two more hydrogen atoms in propanol by hydroxyl groups, the result is *glycerol*, commonly known as glycerin. It is a viscous, odorless liquid, non- toxic and tasting sweet.

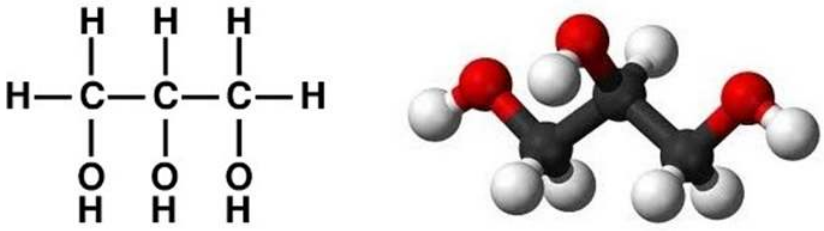


Figure 10.3.2.5: Glycerol, the structural formula and a 3D-ball model (source: Wikimedia Commons).

Now, the only thing that needs to be done before we can explain how fats are built, is to explain what the *esterification reaction* is. It is in certain measure analogous to the acid-base reaction in inorganic chemistry, which produce a salt molecule plus a water molecule, for instance: $NaOH + HNO_3$ $NaNO_3 + H_2O$. The two reagents in esterification are an alcohol and an organic acid. For

example, the esterification reaction between methanol and acetic acid, or – according to the most recent rules in organic chemistry – ethanoic acid, yields an *ester* called *methyl ethanoate* plus a water molecule.

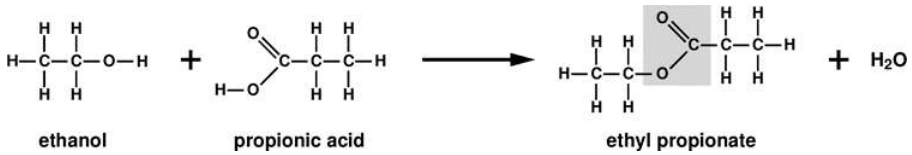


Figure 10.3.2.6: An example of an *esterification* reaction. The group in the shaded box is commonly referred to as the "ester linkage" (aop).

An ester is therefore an analog of a salt in inorganic chemistry.

OK, now we can finally tell what the **fats** are: they are esters called *triglyc- erides* – each molecule is derived from a single glycerol molecule and three fatty acid molecules.

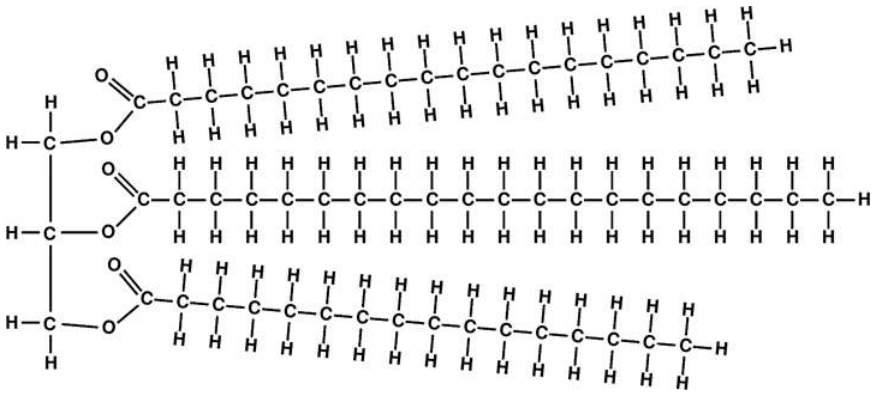


Figure 10.3.2.7: A structural formula of a triglyceride molecule consisting of a triglyceride "backbone" and three fatty acid "tails", attached to it by ester linkages. Natural triglyceride fats are seldom made of a single kind of fatty acid. In this example the three alkyl groups are from *stearic acid* (18 Carbon atoms in the molecule), *arachidic acid* (20 C atoms) and *palmitic acid* (16 C atoms). (aop)

It should be noted that there is a great variety of naturally occurring fatty acids. All those we have discussed so far are *saturated* fatty acid, meaning that there are two hydrogen atoms attached to each carbon atom in the alkyl group, with the exception of the last carbon, which has three hydrogen atoms. In natural fatty acids the number of carbon atoms are always even.

There are ten saturated acids with the number of carbon atoms from 8 to 26.

But the saturated acids are a small minority in the fatty acid family. The thing is that in the hydrocarbon tails there may be *unsaturated bonds* – it means that a pair of carbon atoms is connected not by a single, but by a double bond – and consequently, there is only a single hydrogen atom connected to each carbon atom. Fatty acids with a single bond of such kind are simply called *unsaturated*. But there may be two such bonds in the chain, then the acid is *doubly unsaturated*. And the unsaturated bonds may be at different positions. And last, but not least, the unsaturated bonds may be of two types, *trans* or *cis*. So, the bad news is that there are "zillions" of possible combinations of triglycerides.

The bad news is that all such combinations are not equivalent. At least, depending on the circumstances. If a fat is intended for human consumption, the properties of constituent fatty acids may really matter. American Heart Association recommends limiting consumption of saturated fats, found primarily in animal products. Also, *trans*- fats should be avoided. In short, the terms *trans* and *cis* apply to the properties of *unsaturated* bonds in fatty acids. A *cis*-type double bond produces a characteristic "kink" in the fatty acid hydrocarbon chain – while a chain with a *trans*-type double bond remains straight. Unsaturated plant fats contain almost exclusively *cis*- type fatty acids. *Trans* fats are present in animal fats, most of all in dairy product, and the highest content may be found in the so-called "artificially hydrogenated oils", e.g., in margarines and in shortenings[3].

But the good news is that if fat is to be used as a feedstock for making biodiesel, then it really doesn't matter whether the fatty acids in it are satu- rated or unsaturated and of the *cis*-or the *trans*-type. Even non-edible plant fats can be converted to biodiesel.

Converting Fats to Biodiesel: Transesterification

Essentially, conversion is not absolutely necessary. In one of the first pub- lic demonstrations of his invention, at the 1900 World Fair in Paris, Rudolf Diesel used peanut oil as the engine's fuel. Likewise, some biodiesel ama- teurs fuel their cars with cooking oil (often spent frying oil they buy from fast-food restaurants). It requires some modification of the engine's fuel sys- tem because cooking oil is far more viscous that regular diesel fuel – the car owners solve this problem by installing an extra electric heater in the fuel line that heats up the oil to about 180° F, at which temperature its viscosity is approximately the same as that of petroleum diesel fuel. However, this is considered an unsophisticated and not very elegant method by more ambi- tious biodiesel enthusiast, the more that such converted automobile has to run on cooking oil only, one cannot mix it with regular diesel fuels.

But for those ambitious biodiesel enthusiast, if they have a source of vegetable oil – there is an "elegant" option: namely, one can use a process of *transesterification*. It's relatively simple and can be done even in one's garage using a relatively uncomplicated apparatus.

So, what is the transesterification reaction? Suppose that there is an organic acid $\Phi - COOH$ and an alcohol HOΩ - where Φ and Ω are alkyl groups (let's recall: an alkyl group is a saturated hydrocarbon minus one hydrogen atom). Now, let them form an ester:

$$\Phi - \mathrm{COOH} + \mathrm{HO} - \Omega \longrightarrow \Phi - \mathrm{COO} - \Omega + \mathrm{H_2O} \qquad (10.3.2.1)$$

Like in Fig. 10.3.2.6, right? And now, suppose, another alcohol HO- Δ enters the stage, and something like a "love affair" happens: the acid brakes the relationship (the ester link) with Ω, "dumps" it, and instead form an "ester relationship" with the HO- Δ alcohol:

$$\Phi - \mathrm{COO} - \Omega + \mathrm{HO} - \Delta \longrightarrow \Phi - \mathrm{COO} - \Delta + \mathrm{OH} - \Omega \qquad (10.3.2.2)$$

In other words, the acid de-esterifies itself from the first alcohol, and then it re-esterifies itself with the other alcohol. This is how transesterification works.

The triglycerides are triple esters that forms when three fatty acid molecules combine with a single glycerol molecule which is a triple alcohol. But it turns out that fatty acids have a stronger tendency to "get paired" with some single alcohols than with glycerol. For instance, with methanol, ethanol, or propanol (see Fig. 10.3.2.4).

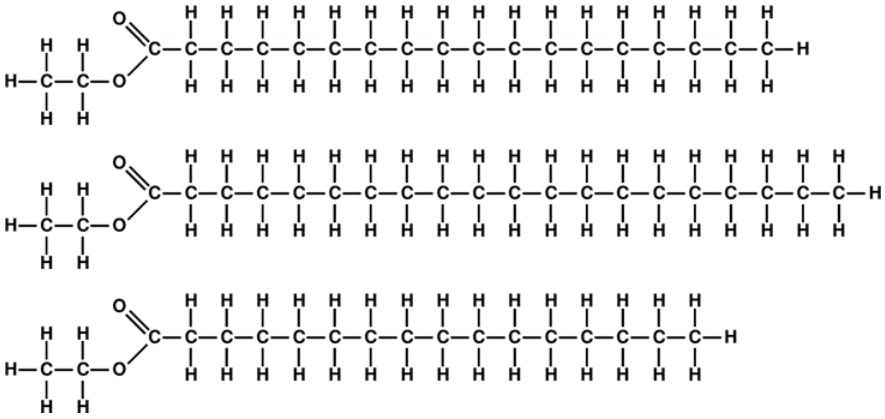


Figure 10.3.2.8: The result of transestrification of the triglyceride molecule from Fig. 10.3.2.7 using ethyl alcohol: the large triglyceride molecule is converted into three smaller ester molecules, plus a glycerol molecule (not shown).

What is the sense of performing transesterification? From Figs. 10.3.2.7 and 10.3.2.8 one can find, after doing some tedious, but straightforward arithmetic, that the chemical formula of the triglyceride molecule is $C_{57}H_{110}O_6$, and its molecular mass is 57 × 12 + 110 × 1 + 6 × 16 = 890. And the average formula for the three ethyl esters is $C_{20}H_{40}O_2$, with the average molecular mass of 312.

The high molecular mass of constituent molecules makes a fluid viscous. By performing transesterification, one obtains a mixture of esters with an average composition quite similar to that of petroleum diesel fuel – with only two extra oxygen atoms per molecule, which in the case of molecules that large makes only a marginal difference.

Practical Methods of Transesterification

As noted, there are transesterification recipies that can be carried out in one's garage or in kitchen. There is one more detail that has not been mentioned in the preceding section: namely, it's not enough to mix oil with alcohol. There has to be a catalyst in the

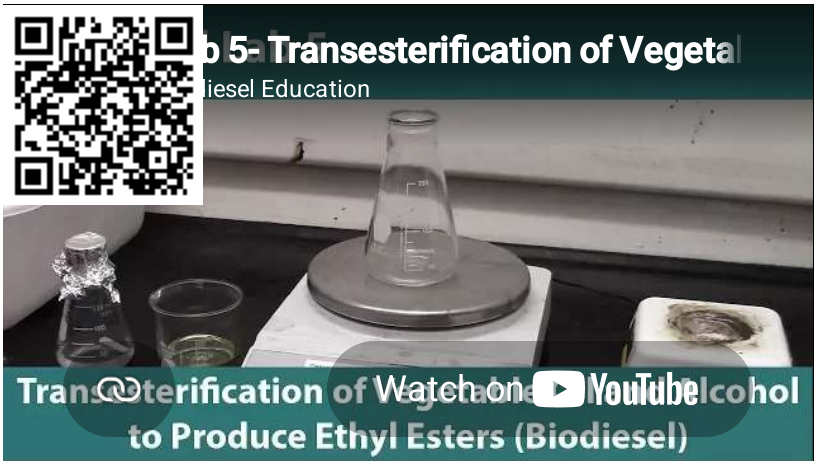
mixture – either inorganic or organic. In home production of biodiesel it is easier to use an inorganic catalyst – its role is to provide a strong enough alkaline environment. NaOH, KOH or even quicklime may be used – the only disadvantage is that a not-so-small amount has to be used. As far as alcohol is concerned, one can use methanol, ethanol or propanol. Ethanol option, if one has no chance of purchasing industrial stuff and goes instead to a liquor store, may be pretty expensive. Industrial methanol is sold by hardware stores – but it is highly toxic. A much better choice may be isopropyl alcohol, widely used as a disinfectant – it's inexpensive and readily available in pharmacies and grocery stores.

It would not make sense to provide a detailed recipe for home-making biodiesel in this text, because a whole bunch of really instructive ones is available at YouTube. The Author has reviewed quite a few of them and can give some recommendations:

A non-nonsense 14m30s movie, the author does not speak a single word, but everything is made clear by how things are shown. A nice thing is that at the end it is shown that the biodiesel really works!

https://www.youtube.com/watch?v=vK8HmWurJow no longer there

University of Idaho tutorial, 5m53s, a nice mixture of theory and practice. KOH and ethanol.



•The University of Idaho Biodiesel Education Program, we use waste vegetable oil from campus dining services and perform the transesterification reaction to turn it into fuel.



Industrial enterprises use more sophisticated technologies – one difference compared with the simple procedures one can carry out in one's garage is us- ing an enzymatic catalyst. Here is a short video published by TransBiodiesel Ltd. It's impressive how little enzyme catalyst is needed – only 1 weight unit per 3000 - 4000 weight units of vegetable oil.



Energy Balance and Global Production of Biodiesel

Industrial companies use oils from many different sources: e.g. canola oil, soybeans oil, or palm oil. The energy balance one can find in literature varies depending on the source. One study finds that for small farms it may be close to 2. The figure given by Wikipedia is 2.6. It's definitely better that 1.3 for corn ethanol, but still far from the value of 8.3 for Brazilian sugar cane ethanol. But it seems that there is still room for optimism, because in one study published in 2011 a record-high value of 5.5 has been reported.

In Table 1 typical yields of biodiesel per acre are shown for several different crops (more data can be found in this Wikipedia article. There is one clear conclusion emerging from the yield data – namely, as with the bioethanol production, the tropical regions are "privileged".

Table 1. Typical yield of biodiesel per acre

Crop	Gallons per acre
Palm oil	508
Coconut	230
Rapeseed (canola)	102
Soybeans	59-99

The global biodiesel production by feedstock data for the years 2007-19 are shown in Fig. 10.20. The current production is approxing 40 biilion liters per year, which is about 1.5% of the US petroleum diesel consumbtion per year and about 0.5% of the global consumption of petroleum diesel fuel. From Fig.

10.20 it follows that it took about 9 years for the global biodiesel production to double. So, assuming that the demand for diesel fuel remains unchanged over the years to come, with the current biodiesel production growth rate it may take about 60 years for the biodiesel production to satisfy 50% of the total demand for diesel fuel. Not a very optimistic scenario, right? But 60 years is a long time horizon. Many things may happen. Global petroleum diesel fuel production has doubled since 1988, so our assumption that it will stay unchanged from now on is a bit risky. On the other hand, global resources of oil will not last forever. New and new petroleum oil deposits are being discovered, but it is hard to make trustworthy predictions reaching more than half a century ahead.

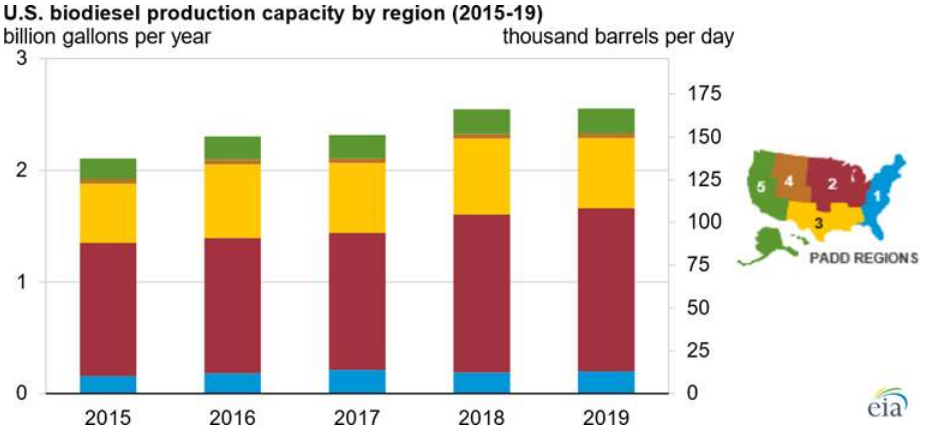


Figure 10.3.2.9: Biodiesel production in the US (2015-2019). Source: US Energy Information Administration.

Algae Diesel Fuel

Currently, about 75% of word biodiesel is derived from vegetable oils – extracted from oil palm fruits, coconuts, rapeseed, soybeans, sunflower seeds, and a few other seeds of oily-rich species. All those plants have been cultivated for centuries for human consumption. When the biodiesel industry emerged, it had an "easy start", because much of the needed infrastructure and "know-how" had already been created: how to cultivate the plants, how to harvest the oily parts and how to extract the oil. What remained to be done was to develop the technology of esterification – and, as follows from the preceding sections, it's a relatively straightforward technological process. All the above-listed plants are terrestrial-based. There are, however, plants living in water environment, which are far more efficient lipid producers than the oil palm, the record-keeping terrestrial-based plant. They are members of the *algae* family, an extra-originally large one with about 100 000 different fresh-water and sea-water species. Some algae, if farmed, may yield a crop per acre as much as 50 or even more times larger than the crop from the best land-based oily plants. The US Department of Energy estimates that algae fuel from farms occupying only 0.42% of the US territory would be able to provide enough fuel for replacing all the petroleum-based fuels currently used in the US. And it's only 1/7 of the land area currently used for corn growing!

If so, why don't we shut down all the farms growing feedstock for corn ethanol and vegetable biodiesel and replace them with algae growing farms? Well, the problem is that the algae biofuel industry is currently at its "earlest infancy". Vegetable oil plants have been harvested for centuries (or even for several thousand years, as is the case with olive trees). But until quite recently, nobody was trying to grow and process algae at an industrial scale. Currently, there is lab research and there are some small-scale pilot research projects going on. However, it is beyond doubt that creating a sustainable algae biodiesel industry will require creating an enormous infrastructure. But almost nothing has been built so far. Even worse, at the moment it's not even clear what should the new industry look like: fields of open ponds, or closed bio-reactors in giant greenhouses? Plus a whole bunch of other ques- tions. There are many visions and concepts, some of them in conflict with other visions and concepts. Among experts, there are optimistic and even enthusiastic voices. Here are some samples: one from Exxon, another one also from Exxon, one from the Department of Energy's Bioenergy Technolo- gies Office. Here is one much more sceptic voice, one clearly pessimistic, and another one expressing similar doubts whether sustainable algal biofuel production is at all possible? Well, even if it is, the "time horizon" given in the "optimistic" Department of Energy Bioenergy Technologies Office's document linked above is not *per se* very optimistic. Please look at Page 27 of the said document, where the goals are stated: *five billion gallons per year of algal biofuel* by the year 2030. Well, today's (2018) biodiesel production according to the US EIA is 1.8 billion gallons, and the total consumption of petroleum diesel fuel is slightly over 60 billion gallons. As follows from these numbers, According to the most ambitious goals, the algal biodiesel production will not even reach 10% of the diesel fuel consumption in 2030. So, we have to be realists! We can start singing "It's a long way to get rid of petroleum diesel", to the tune of *It's a long way to Tipperary*, a famous World War I song. But eventually this goal must be reached!

---

1. An alkyl group is *alkane* minus one hydrogen atom – and alkane, a.k.a. *paraffin*, is any of the series of saturated hydrocarbons C_nH_{n+2}.

2. The physical properties end the smells of all these three alcohols are very similar – and methanol, if consumed, may cause blindness, or even death.

3. Our ancestors living at tree-tops for millions of years fed mostly on plant products, fruits and nuts, which contained almost exclusively *cis* fats. Meat appeared in their diet much later, and diary products with high concentration of *trans* fats only tens of thousands years ago. Perhaps Mother Evolution did not have enough time to teach our organisms to tolerate the *trans* fats, and this is the reason why we should – well, not to avoid *trans* fats completely, but at least show moderation in consuming them.

10.3.2: Biodiesel Fuel is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

10.4: Other Ways of Harnessing Plant Energy- “Bioelectricity

In the opinion of some scientists converting corn to ethanol in order to use it as a fuel for automobiles is not a good idea. In an article published in 2009 in the SCIENCE magazine[1] the authors present the following arguments, backed by graphic illustrations. Because of the tough copyright rules imposed by the journal, we cannot show the graphic part here. Let’s consider the corm plant. The entire sunlight energy recieved by the plant is used by it for building its complete body: the stalk, the leaves, the husk, the cob, the grains and the roots). There is something called the *harvest index HI*, defined as the ratio of the corn grains mass to the total above-the-ground dry biomass:

$$HI = \frac{\text{dry grain weight}}{\text{dry grain weight} + \text{dry weight residue}} \tag{10.4.1}$$

The average HI value of maize in the US is close to 0.5. So, when the grains is separated from the rest of the plant’s body, about 50% of the biomass and its energy is lost. Then, the starch from the grains has to go through the enzymatic conversion to glucose and then to be converted to ethanol by yeast fermentation – at each step some energy is lost. And finally, the alcohol fuel powers the internal combustion (IC) engine in an automobile – the thermal efficiency of an IC using gasoline or ethanol fuel is about 20%, so at the last step, propelling the vehicle, 80% of the energy is lost. It is difficult to tell exactly how much of the original energy contained by the “input biomass” is lost. According to the estimates of Ohlrogge *et al.* only about 10% of the initial energy stored by the plant through the photosynthesis process is used for propelling the vehicle (according to the “gut feelings” of the author of this text, it’s probably a pretty optimistic estimate).

As Ohlrogge *et al.* point out, there is an alternative worth considering. Arguably, the IC engine will not dominate the passenger car sector forever. There is a growing number of zero- and low-emission vehicles such as electric car and plug-in hybrids. At some foreseeable future they may even start outnumbering vehicles powered exclusively by IC engines. So, it makes sense to switch the attention from biofuels to “bioelectricity”.

Bioelectricity ? Is it possible to generate electric power from plant ma- terial? Yes, absolutely! And much of the needed infrastructure allready exists. Instead of growing plants that can provide feedstock for making corn bioethanol, one should cultivate plants with a high biomass yield (such as Napier Grass, or Switchgrass) – and then simply use the biomass as a fuel in a thermal power plant. Either as the principal fuel – as, for instance, in this 32 MW biopower plant fueled by Napier Grass, or by co-firing biomass with coal. The efficiency of modern thermal power plants can exceed 30%. Some power will be lost in transmission, and some in the process of charging car batteries. But the efficiency of motors in electric cars is very high, 90% or even more – so, after taking into consideration all possible losses, we finf that 20-25% of the original biofuel energy will be available for propelling the car.

The paper of J. Ohlrogge *et al.* is not new – it has been published in 2009. But the conclusions are certainly no less relevant today. Actually, taking into account that some biofuel technologies that seemed to have a “bright future” in 2009, but have badly disappointed the planners over the next decade – the conclusions of J. Ohlrogge *et al.* may be even **more relevant** today than they were in 2009.

---

1. John Ohlrogge, Doug Allen,1 Bill Berguson, Dean DellaPenna, Yair Shachar-Hill, and Sten Stymne, *Driving on Biomass*, SCIENCE magazine, vol.324,pp. 1019-21, May 22, 2009.

10.4: Other Ways of Harnessing Plant Energy- “Bioelectricity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

CHAPTER OVERVIEW

11: Energy Storage

11.1: Introduction
11.2: Overcoming the Challenges of Wind Power
11.3: Problems with Solar Power- the \Duck Curve"
11.4: Energy Storage Technologies
11.4.1: Battery Storage
11.4.1.1: The Voltaic Cell
11.4.1.2: Lead-Acid Batteries
11.4.1.3: Lithium-Ion Batteries
11.4.1.4: The “Families” of Li-ion Batteries Manufactured
11.4.1.5: Batteries for Industry- and Grid-scale Storage
11.4.2: Electricity → potential energy → back to electricity (gravitational energy storage and compressed air energy storage).
11.4.3: Electricity → Kinetic Energy → Electricity Storage Scheme
11.4.4: Electricity → Thermal Energy → Electricity Storage Scheme
11.4.4.1: Molten Silicon
11.4.4.2: Heated Volcanic Rocks
11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage
11.4.4.4: Small and Medium Scale Thermal Energy Storage
11.4.5: Electricity → Chemical Energy
11.4.5.1: The evolution of the hydrogen economy policy in the 21st century
11.4.5.2: Methods of Green Hydrogen Production
11.4.5.3: The recovery of energy stored in the form of hydrogen
11.4.5.4: Hydrogen storage and transportation issues
11.4.5.5: Hydrogen as a “green” fuel suitable for sea- sonal storage

---

11: Energy Storage is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.1: Introduction

Some experts on acquiring energy from new resources have declared that in 2018 there was an “Energy Storage Revolution”.

What does it mean? That energy storage was invented? No, definitely not, large scale energy storage devices – such as pumped storage hydropower plants (see Chapter 5, Section 5.4) – have been built since the late 1800s. Perhaps there was a sudden extraordinary increase in the construction of new energy storage facilities? No, more were indeed built than a year before, but it can hardly be viewed as a “revolutionary increase”. So what?

In the opinion of this Author, there has been a ***revolutionary increase in awareness*** that a full “decarboniation” of energy production may never be achieved without the development of a grand-scale storage technology.

Why may it never be achieved? Well, let’s take a look at the zero-emission methods of generating electric power. According to the Energy Information Agency, their current (2018) contribution to the overall power generation in the US is 36%; the numbers in parentheses show the contribution of each of the listed sources:

1. Wind turbines (6.5%);
2. Solar PV panels and CSP plants (1.5%);
3. Hydropower plants (7.0%);
4. Nuclear power plants (19.4%);
5. Methods utilizing biomass (1.4%);
6. Methods using geothermal sources (0.4%).
7. Methods utilizing ocean waves, tides and tidal currents (< 0.1%);

A number of USA States (e.g., California, Oregon and Washington, Massachusetts, and many others) have already made commitments to attain a 100% or 80% economy-wide greenhouse gases emission reduction by 2050. European Union has just (December 13, 2019) announced an ambitious goal to “make Europe the very first climate neutral continent by 2050”.

So, if all fossil-fuel burning power plants are to be eliminated, they must be replaced by zero-emission power generation methods listed above. Con- sider the scale of the problem. There will not only be need to create new zero-emission with total power of 64% of the current US generation. Much more will be needed, because gasoline and diesel cars and trucks have to be replaced by electric ones. So, even more generation will be needed for charg- ing all those vehicles. How much? There are only some estimates, e.g., in in a CITYLAB Web page. It seems that not only 64% of the current generation will be needed, it may be 100% or even more.

Which of the seven methods listed above offers the greatest chance that it will be able to satisfy these enormous new requirements? Let’s evaluate it for each of them, from the bottom up.

Ocean waves and tides, as presented in Chapter 7, may provide huge amounts of energy. However, the methods of harnessing them are not yet sufficiently developed. They may play a meaningful role in global economy in a more distant future than the 30 or 50 years from now, which is a widely accepted necessary “target date” for attaining global decarbonization.

Geothermal energy, as discussed in Chapter 8, has the potential of satisfying all the needs of humanity – yet, reaching these resources would require drilling a huge number of very deep wells. Again, heat and electricity generated from geothermal resources may become major contributors in the future global “energy budget”, but a realistic time period needed to develop this sector is definitely much longer than the 30 or 50 years in question.

Biofuels are a great sources of “carbon neutral” energy. They will play an important role in the 30-year plans of decarbonization: most likely, there will be no other way to make air transportation fully eco-friendly than to replace current jet fuel derived from petroleum by fuel of nearly identical properties, but derived from biological products. It can be expected that the production of biofuels will increase several times over the next 30 years – but to replace all energy based on fossil fuels it would have to increase **several dozen times**.

Hydropower is patently zero-emission. Currently, the electric energy it provides is more than one tenth of that obtained from fossil fuel burning. It produces about 16% of the electricity generated globally (about a quarter of electric energy obtained from fossil fuels). But existing resources with easy access have already been used to a large extent. Also, there are other concerns – even though hydropower does not emit any CO_2, it’s not com- pletely “benign” for the environment. Currently, some hydropower

facilities in Oregon are being destroyed because of their negative impact on biosphere. As far as new hydropower ventures are concerned, the attention is focused primarily on pumped-storage plants (see Chapter 5) which do not generate new power, but only store surplus power for later usage.

Nuclear power. It's zero-emission and generates more than 10% of electric energy globally. In principle, it's capable of generating much more – it could even replace all fossil-fuel generation if enough new reactors were built. The only limitations here are the high up-front costs – which is probably not an insurmountable obstacle – and the negative public perception of nuclear energy. The latter seems to be an obstacle much harder to overcome than the financial barrier. Some nations, e.g. China, or , Russia, have far-reaching plans to develop nuclear power. So, nuclear energy will certainly play a role in global decarbonization, but most likely it won't be the **principal player** in this process.

Wind and solar power. By eliminating all other candidates, we come to the conclusion that the main burden of global decarbonization must rest on these two technologies. Possible? Yes, definitely, the existing resources have only been used to a minimal extent. There are huge "resources" of off-shore wind power. There are large areas of lands (e.g., deserts) where stationary PV panels may be deployed and many bodies of water available for "floating" PV panels mounted on big "rafts" – without any depletion of areas that can be used for agriculture, for recreation, or for new human settlements.

But there is one major problem: namely, generation by means fossil fuel burning can deliver power in a stable manner, according to consumer require- ments. In the case of generating electricity by geothermal, hydropower and nuclear methods the situation is very similar. But not with wind and solar power. **And this is the heart of the problem!**

Why? In contrast to other methods of harnessing renewable energy, wind power and solar power are intermittent. In the case of solar power, there is regular intermittency (it's on at daytime, and off at night) and random intermittency, due to cloudy skies. The latter is troublesome especially in medium latitudes (e.g., in most European countries). The intermittency of wind power may be even more chaotic. There are regions on the globe where the wind is highly regular and predictable – namely, in areas where tradewinds blow (see Chapter 6). Unfortunately, they blow mainly over open oceans – which is good perhaps for sailing ships, but definitely not good for wind turbines.

11.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.2: Overcoming the Challenges of Wind Power

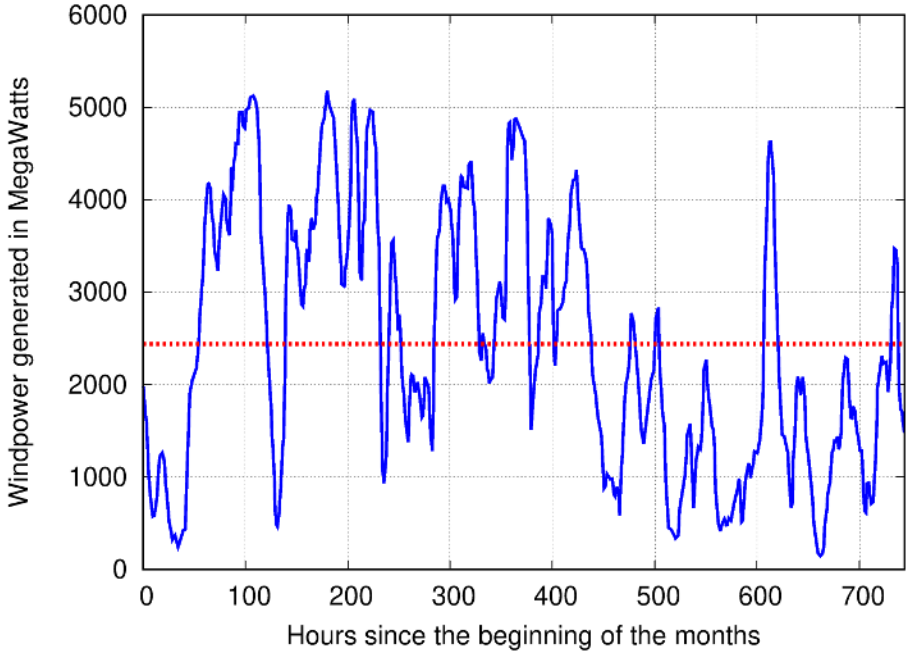


Figure 11.2.1: generated over the month can be obtaining by calculating the area below the curve (i.e., by integrating), which yields 1,8126 GWh,corresponding to the average power of 2.436 GW (shown as the red line). Source: hourly data published by the operator of the national grid in Poland "Polskie Sieci Elektroenergetyczne (PSE)" and systematically monitored on Mrs. Jozefa Sokolowska's blog – chart based on the data collected in her blog, with her permission.

Above lands, winds can be quite capricious - especially in medium latitudes where different atmospheric fronts often clash with one another. It leads to periods of high winds alternating irregularly with light winds, or even with completely windless periods. An example of electric power generated by wind in March 2019 in Poland, a country located centrally in Europe, is shown in Fig. 11.2.1. It illustrates irregular fluctuations in the power, between the maximum of 5 GW and the minimum of less than 0.5 GW. Such fluctuating patterns are typical for other countries from the same region – in particular, for Germany and Denmark, two nations in which the energy systems rely heavily on the wind power.

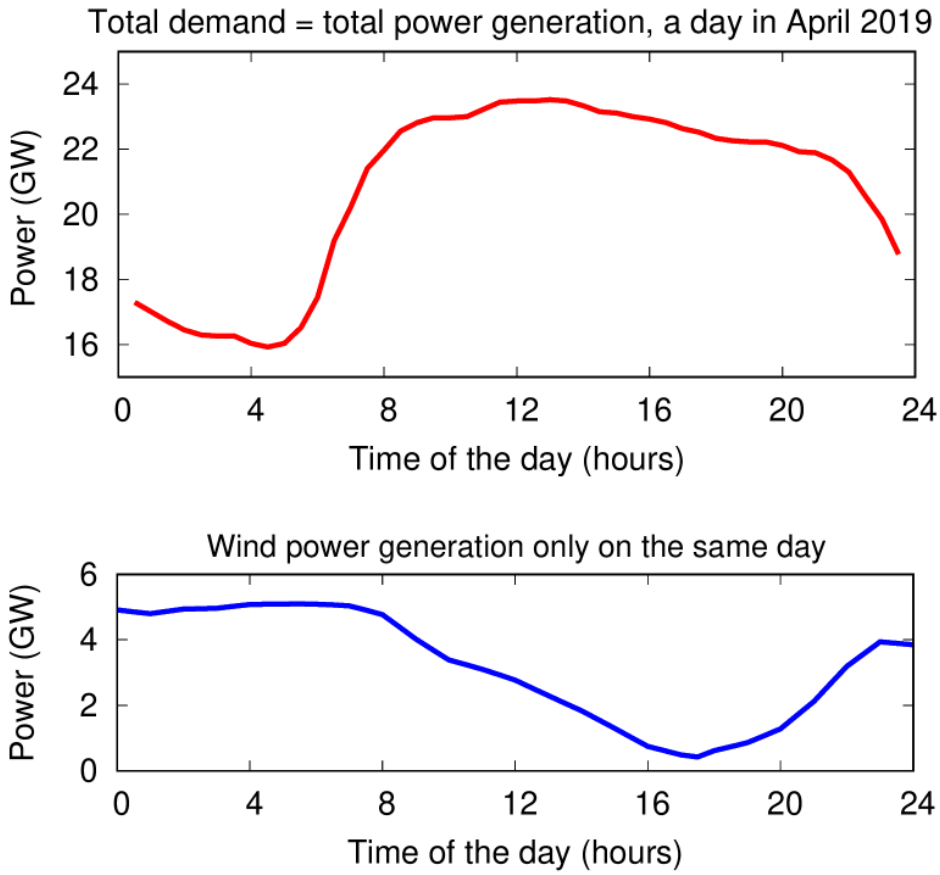


Figure 11.2.2: Upper panel: the total demand for power in Poland on a day in April 2019; lower panel: the power provided by wind farms only in the same day. Over the period in which the wind power ramps down by almost 5 GW, the other available sources – all of them burning fossil fuels – must ramp up by the same amount (Source: as in the Fig. 11.2.1)

Is energy storage necessary to deal with such wind fluctuations in Poland? Well, no. Not yet! To explain what we have in mind, let's take a look at another graph (Fig. 11.2.2), illustrating the situation in Poland on one day in April 2019. The red curve in the upper plot shows the total generation over the 24-hour period – generation equal to customers' needs. The lower graph shows that in the morning hours the wind sector delivered about 5 GW, nearly 1/3 of the power demand on 5 a.m. – but then, while the demand was strongly increasing, about 8:00 a.m. the wind power generation started suddenly "ramping down" and over the next 9 hours it fell to less than 0.5 GW.

This unexpected loss of energy from wind farms forced other sectors of the national utility system (mainly coal-fired power plants) to increase output power to almost it's maximum capacity. Fortunately, necessary reserves in the coal-fired plants still existed and a power outage did not happen.

But let's try to imagine a similar situation in the future. The European Union has set itself the goal of achieving total "decarbonization" by 2050. Poland, being a member of the Union, must therefore gradually reduce the share of coal-fired plants in the entire power generating "mix". Then, say, in 2040 the coal-fired sector needs to be reduced to about 1/3 of the current power, from about 27 GW to 9 GW. The 27 - 9 = 18 GW power deficit which arises as the result of such transformation must be covered by increased wind farm generation.

As noted before, currently the wind farm system in Poland is capable of delivering about 5 GW under the most favorable circumstances. But such "most favorable" circumstances may not occur at the moment when all 18 GW are needed. More realistically, one should take into accuont the *average* of 2.4 GW. So, the existing wind-farm capacity should be increased at least **eight-fold** or even more to offer sure guarantees that 18 GW will be delivered when needed. Well, but as can be seen in Fig. 11.1, more than once in the month of March 2019 the existing wind farms were able to deliver even less than 0.5 GW. So, it may happen that the enlarged system, even with the capacity increased by a factor as big as 10, will be able to provide only some 5 GW out of the necessary 18 GW. And then what?

Well, the currently existing utility systems – those still dominated by fossil fuel burning sources – are prepared to deal with sudden dramatically increased power demand. They have special generators in reserve, called “peakers”, capable of switching from zero to full power in minutes. Presently, most of them are natural gas turbines – or, sometimes, huge Diesel engines. Such “peakers” may offer a remedy for sudden “incapacities” of wind farms. But let’s keep in mind that the objective is to reach a full decarbonization, so any “remedies” using fossil fuels are certainly not welcome! But are there zero-emission “peakers”? The answer is ***yes, there are*** - such a task can be fulfilled by ***energy storage devices***.

Note that in the example scenario discussed the number of wind farms is increased so that the *average* generation meets the needs. Accordingly, there will be periods in which the energy generation will significantly **exceed** the needs. This is not a small problem! What to do with that surplus energy? In nations which already have a substantial windpower potential (e.g., in Germany), such situations occur from time to time. But, as was emphasized earlier, in an efficient system the generated power must match the needs of consumers: not only there should not be less power, but also there must not be no more than needed! Over-generation would destabilize the grid, potentially leading shutdowns and even to damage of power lines and distribution centers. Therefore, an often used “emergency remedy” is to curtail the wind generation: either by changing the angles of the wind turbine blades to lower their efficiency, or – in a more brutal fashion – by stopping some turbines altogether. Such curtailing, of course, makes little sense, because it is simply a waste of energy. The surplus energy should be stored This excess electricity should be stored in appropriate “rechargeable” reservoirs capable of returning the stored energy on demand. Then such “reservoirs” can act as “peakers” in times when the wind is so weak that wind farms are no longer able to generate the needed power.

The idea is pretty clear. But it has only recently been realized that with- out the existence of extensive energy storage systems achieving a significant reduction in CO_2 emissions will not be realistic. Thus, efforts to create such systems at the scale needed are only now gaining momentum. Fortunately, several viable technologies that can be employed for this purpose already exist and additional opportunities are still emerging thanks to intensive research. An overview of such technologies is presented in the following sections of this chapter - but only after discussing the problems arising from the intensive use of solar energy - which is also characterized by intermittence, but of a somewhat different nature than that in the case of wind.

11.2: Overcoming the Challenges of Wind Power is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.3: Problems with Solar Power- the \Duck Curve"

The intermittence of solar power has two components: one – let's call it "regular" – is due to the daylight-nighttime cycle and is completely predictable.

The other component is due to clouds. In medium latitude non-solar sources regions, such as, e.g., Central Europe, where the weather is often cloudy, the curves illustrating the daytime power generated by photovoltaic farms may be similar to the patterns of windpower fluctuation (shown previously). But in regions with prevailing sunny weather and high insolation, as, e.g., Southern California, in Arkansas, or in Hawaii.

The abundance of solar energy in these regions combined with the rapidly falling prices of solar panels have resulted in a very rapid growth of the solar energy sector. In California there has been an "explosive growth" of solar generation. Between 2014 and 2019 it has increased five-fold, reaching about 13 GW in the record months of 2019. This is more than 50% of the demand (a graph showing the current power demand in California can be watched at this Web page of California ISO. Is this good news? Yes, but only in the middle of the day. In the morning and late afternoon hours there is less power generated. Then the Generation ceases completely shortly before sunset and resumes only some time after sunrise. However, the demand exhibit a completely different trend with a maximum only **after** the sunset. Te maximum, as can be seen in the California ISO site, can be as high as 30 GW in the winter months, and even higher in summer months. The situation can be best explained using a graph such as the one in the Fig. \(\PageIndex{1}\).

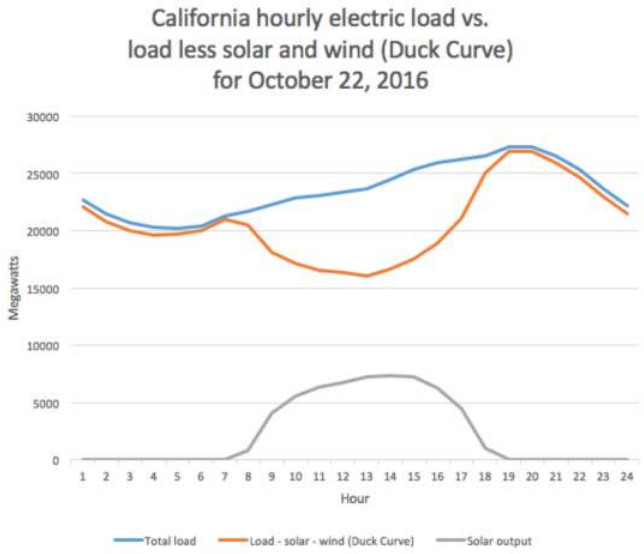


Figure \(\PageIndex{1}\): The demand for power for a 24-hour period in California in October 2016. The upper blue curve is the total demand and the gray curve is the solar power generation. The difference between the two – i.e., all power provided by – is shown by the orange curve (from Wikipedia's Duck Curve page).

As noted before, a utility system cannot generate more energy than the consumers need. Therefore, the non-solar generation must "take a dip" – which makes the orange curve look somewhat like a duck. This duck-like shape becomes even better visible if one plots similar "total minus solar generation" curves for several consecutive years, as is done in the Fig. \(\PageIndex{2}\).

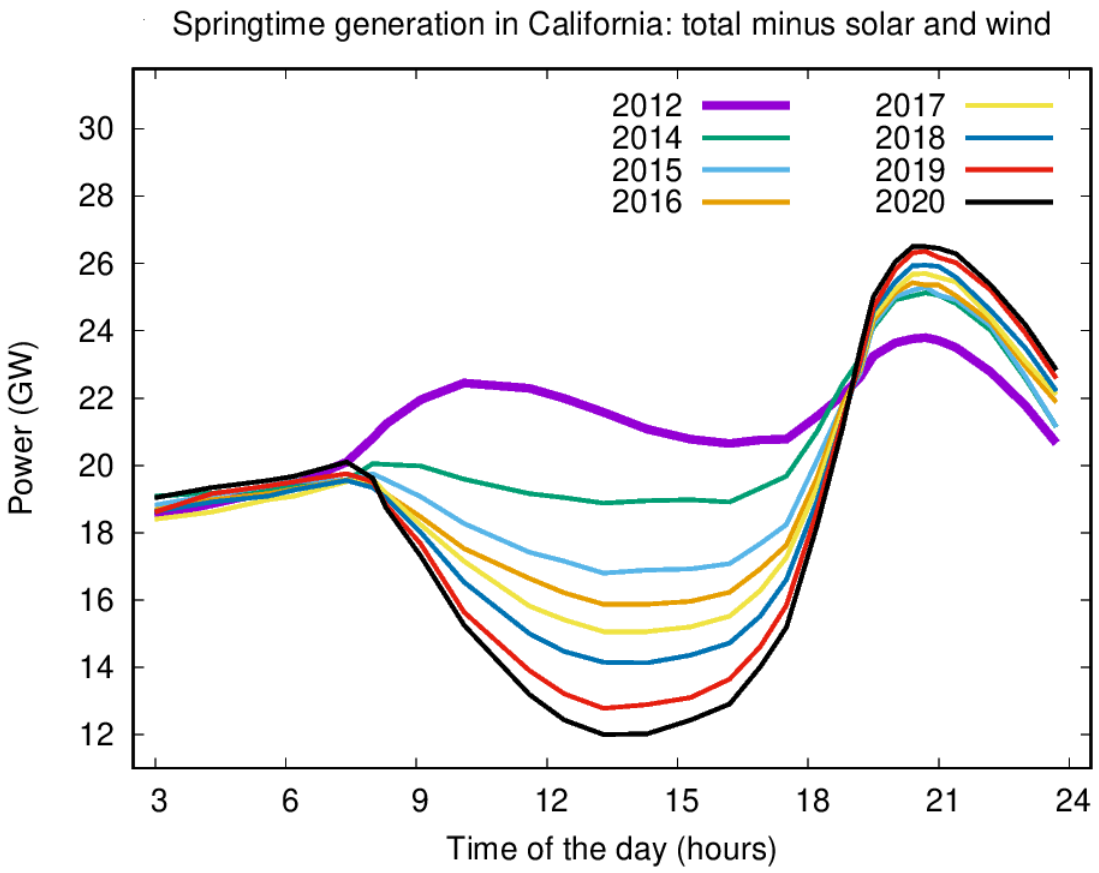


Figure \(\PageIndex{2}\): Springtime "duck curves" for California for several years, starting from the year 2012. By 2020, the installed solar capacity has increased more than tenfold (plot inspired by videos posted at YouTube by the Vox Group and the US Department of Energy).

What is wrong with the "duck curve"? Well, let's take a closer look at the curve for the year 2020 in Fig. 11.4. What happens over the seven hours between 2 p.m. and 9 p.m.? The sunlight generation dies out completely, and the demand reaches its evening maximum – the generation has to be ramped up by about 15 GW. And who is going to provide this huge amount of electricity? Well, at the present time (2020) this burden lies overwhelmingly on "peakers". They are aided by power drawn from energy storage devices, for example, pumped storage power plants – but currently, they are able to provide only a fraction of the power needed. In California there are three types of power plants whose output must remain constant or nearly constant because the intensity of the physical processes employed for generating energy cannot be easily varied "up" and "down". Those three sectors are: nuclear power, hydropower and geothermal generation, with 2.2 GW, 3.2 GW and 1.5 GW of output power, respectively, and a total of 6.9 GW (let's call it "the baseline generation"). Accordingly, as it follows from the cited data, at about 2 p.m. the overall California utility system needs about 5 GW generated by burning fossil fuels (mainly natural gas), while by 9 p.m. it already requires about 20 GW from fossil fuels.

Taking into consideration how fast the installation of solar power in California has grown, it can be anticipated that in a few years there will be so much energy from the sun and wind that in the middle of afternoon generation of electricity from natural gas won't be needed any more. And further installation of new solar and wind farms will already create excess energy in the middle of the day. However, regardless of how much solar energy is deployed, in the evening the state will need 20 GW produced by other means. And this power may be gradually reduced later, but only some time after the next sunrise it won't be needed any more.

What are those "other means"? – mainly, natural gas peakers. So, even with duplication or even a tripling of installed solar power will not free California's energy systems from emitting a significant amount of CO_2 making it difficult to achieve its ambitious plans to completely decarbonize the electricity generation. The solution may be to use another zero-emission method of generation, but the only such replacement for natural gas peakers that can be taken into account is wind energy. Yet, the average power of current California's wind farms is about 1.6 GW (with a total installed power of about 4.5 GW). Hence, to eliminate the gas

peakers, the potential of wind farms would have to be increased at least 10 times, which does not seem realistic in a foreseeable future.

California, however, has a very strong resolve to get 100% carbon-neutral by the year 2045. So, the CO_2 emitting peakers must be replaced. By what? There is one more option: by energy-storing devices. The deployment of new solar farms should continue with the same or even faster rate than before. The mid-day "overproduction" should not be curtailed or imported, but instead the energy has to be stored in "rechargeable" devices, from which the power could be later re-drawn. Are there such devices?

What should be the capacity of such a storage facility? Well, the answer is simple: it should be at least the same as the energy provided by the CO_2 emitting sources in California in one 24-hour cycle. One can get a rough estimate of that using a simple model. Based on what was said before, California needs to ramp up fossil-fuel generation from 5 GW at 2 p.m. to 20 GW at 9 p.m., and then the power may be gradually decreased back to 5 GW by the following noon – as illustrated by the simple graph in Fig. 11.5. One can readily carry out a simple integration, from which one gets that the total fossil-fuel energy generated over a 24 hour period is 285 GWh. Please keep in mind that this is only an estimate. When using very simple models, one has to be prepared that the deviations may be significant, 10%, perhaps 20%? But for the considerations that we carry out here, such accuracy is

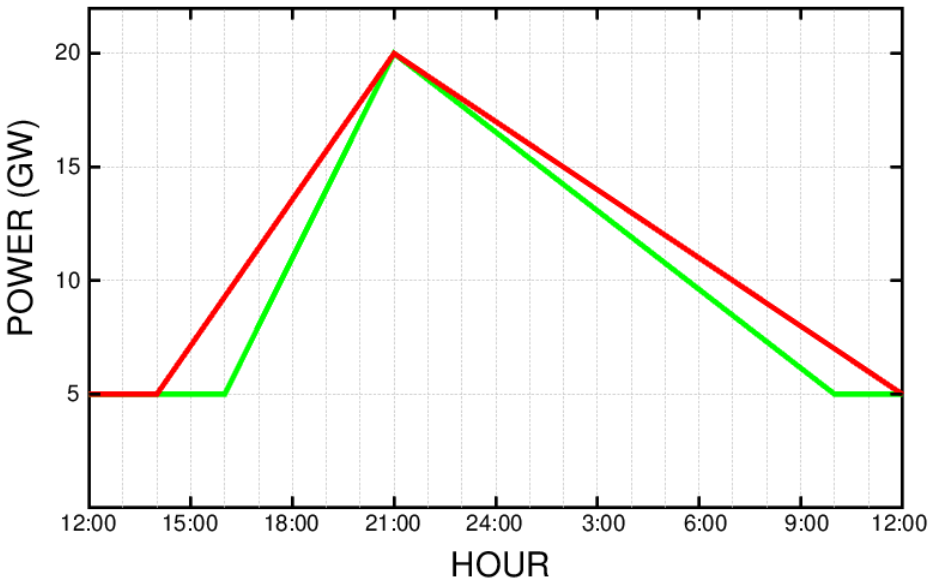


Figure \(\PageIndex{3}\): A simple graph illustrating the power from fossil fuel generation needed by the California utility system over a 24 hour period. The "triangle" is only a rough approximation of the actual characteristics, the shape of which is rounded and not completely smooth – but such a "triangle" makes the integration straightforward. For the meaning of the red and the green lines, see the text. quite sufficient. We can round up the number we have gotten to 300 GW, it will be easier to remember.

It may be interesting to think of how this picture may change if much more solar generation is added – say, so much that the gas "peakers" won't be needed in the middle of the day. Then, the "triangle" in Fig. \(\PageIndex{3}\) will start from zero level and perhaps become somewhat narrower – as shown using green lines in Fig. \(\PageIndex{3}\). Then the area encompassed by the blue line corresponds to 200 GWh. Adding more solar will reduce the need for using gas "peakers", but only by 1/3. Now not 300 GWh of storage capacity will be needed to eliminate the CO_2 emission, but only 200 GWh. Still, a huge capacity!

Is it possible to build a energy storage device of such capacity? Lets look at the largest manufacturer of lithium-ion storage batteries (made primarily for Tesla cars), the Gigafactory 1 plant located in Nevada, not far from the California's borderline. In 2019 it manufactured batteries of an overall capacity of 35 GWh, and its target capacity planned to be reached by the middle of the current decade is 150 GWh/year. Thus, the creation of storage batteries needed for reaching a total decarbonization of California's utility lies completely in the capabilities of todays industry. The cost will be astronomical – but there is an old truth: it's better to pay than to regret later.

Lithium-ion batteries represent only one of several existing technologies that can be employed for storing energy at a level needed by a carbonneutral utility system. On overview of such technologies – existing, as well of emerging – is presented in the following Sections.

11.3: Problems with Solar Power- the \Duck Curve" is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

11.4: Energy Storage Technologies

There are several types of devices that can be used to store energy. In practice, the input may be either electrical energy (EE), or heat (Q) = flow of thermal energy (TE). The same applies to the output. Then, there are different forms of energy contained in the "storage reservoir". Electrical energy cannot be stored "as it is". It must be converted to a different energy form. The "storable" forms are thermal energy, potential energy due to gravity (PEG) or due to compression of the storing medium (PEC), kinetic energy of spinning bodies (KESB), or – last, but definitely not least – chemical energy (ChE). There is a large variety of possible combinations of input energy, stored energy and output energy. Here, we list those that are of practical importance, using a scheme: input → stored energy → output:

- EE → ChE → EE: it is how electric batteries work, and currently it is the fastest growing storage technology (further on, battery storage is discussed as a separate topic). Another method is to produce and to store a "green" fuel, i.e., hydrogen, and later on convert it to electricity or heat.
- EE → PEG → EE: this is the scheme according to which pumpedstorage power plants work; currently, considering the total storage capacity of all such plants worldwide, this is the "Number One" storage technology (however, storage in batteries may outweigh it some time from now).
- H → ThE → EE: this method is currently used primarily in Concentrated Solar Power technology (see Chapter 7).EE → PEC → EE: here compressed air is used as the "carrier" of the stored energy.
- EE → ThE → EE: there are several variants of this scheme, now only in the early stage of development.

All the methods listed will be discussed in closer detail in the following subSections and sub-sub-Sections.

11.4.1: Battery Storage

11.4.1.1: The Voltaic Cell

11.4.1.2: Lead-Acid Batteries

11.4.1.3: Lithium-Ion Batteries

11.4.1.4: The "Families" of Li-ion Batteries Manufactured

11.4.1.5: Batteries for Industry- and Grid-scale Storage

11.4.2: Electricity → potential energy → back to electricity (gravitational energy storage and compressed air energy storage).

11.4.3: Electricity → Kinetic Energy → Electricity Storage Scheme

11.4.4: Electricity → Thermal Energy → Electricity Storage Scheme

11.4.4.1: Molten Silicon

11.4.4.2: Heated Volcanic Rocks

11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage

11.4.4.4: Small and Medium Scale Thermal Energy Storage

11.4.5: Electricity → Chemical Energy

11.4.5.1: The evolution of the hydrogen economy policy in the 21st century

11.4.5.2: Methods of Green Hydrogen Production

11.4.5.3: The recovery of energy stored in the form of hydrogen

11.4.5.4: Hydrogen storage and transportation issues

11.4.5.5: Hydrogen as a “green” fuel suitable for sea- sonal storage

11.4: Energy Storage Technologies is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

11.4.1: Battery Storage

According to the orthodox definition, the the current i basic “building block” of an electric battery is an *electrochemical cell*. The battery consists of several such cells connected together to increase their performance. Accordingly, the popular AA or C “batteries” are not betteries, but single cells. Yet, today little attention is paid to this orthodox division and we use the term “battery” for single cells as well. In this Chapter we will stick to that habit and a single cell will often be referred to as a “battery”.

Progress in Battery Storage Technology

There is one area in which "revolutionary progress" is currently taking place -- namely, energy storage, and specifically, electric storage battery technology. Sodium-ion batteries, almost nonexistent four years ago, are already starting to become commercially available (in Chinese electric cars). The liquid electrolyte and graphite anodes in lithium-ion batteries are being replaced with increasing success by solid electrolytes and anodes made of metal itself, allowing for drastic increases in battery capacity and shorter charging times.

Such developments have taken place well after this book was written. We hope that they will be covered in future editions, but until then the interested reader can read a review on the Recent Advances in Battery Development and Energy Storage Technologies by George G. Njema, Russel Ben O. Ouma and Joshua K. Ki which appeared in the Journal of Renewable Energy (2024)

A conventional electrochemical cell consists of essential parts: two *electrodes* and *electrolyte*. The electrodes are made of metals or current-conducting solids. The electrolyte is an ion-conducting substance placed between the two electrodes. Most often it’s a liquid or it has consistency of a gel. A potential difference (the same as *voltage*) arises between the electrodes immersed in the electrolyte. The one with a higher potential is called the *positive electrode* or the “plus” (+) electrode. The other is the negative electrode or the “minus” one. Now, if the two electrodes are connected by a metal wire, an electric current starts flowing[1]. Electrons in the wire travel from the “minus” electrode to the “plus” one. But the charge cannot accumulate, therefore it must be carried back – now, through the electrolyte, by ions.

In texts describing galvanic cells as well as other electric devices two important terms are often found: namely, the ***anode*** and the ***cathode***. How are they defined? Suppose that the electrodes of the device in question are connected by a wire. Then, current flows: as mentioned above, it’s the electrons flowing from the “minus” electrode to the “plus” electrode. But the “conventional current” – the concept introduced by Franklin – “flows” in the opposite direction. The concepts of the anode and the cathode refer specifically to this “conventional current”. Namely, the *anode* is the electrode **into which** the “conventional current” enters from outside. And the *cathode* is the electrode from which the “conventional current” flows out. Accordingly, if we think of an AA battery or any other disposable one, the “+” terminal is the cathode, and the “-” terminal is the anode. But with rechargeable batteries, as will be discussed below, things become more tricky.

---

1. Benjamin Franklin, one of the Founding Fathers of the United States, was also a great scientist. He conducted research of fundamental importance on electricity. In the 18th century when he lived it was impossible to determine what was the exact nature of the charges carrying electric current. Franklin assumed that the current flowed from the positive electrode to the negative one – and it became an established rule. Only 150 years later it have been proved beyond any doubts that in metals the current is carried by negative particles, *the electrons*, who travel from the negative electrode to the positive one. But that doesn’t mean Franklin was completely wrong because there are non-metallic conductors in which electric currents are carried by positive particles – e.g., by positive ions, or by *holes* in *p-type* semiconductors

11.4.1.1: The Voltaic Cell

11.4.1.2: Lead-Acid Batteries

11.4.1.3: Lithium-Ion Batteries

11.4.1.4: The “Families” of Li-ion Batteries Manufactured

11.4.1.5: Batteries for Industry- and Grid-scale Storage

11.4.1: Battery Storage is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.1.1: The Voltaic Cell

The invention of electrochemical cell is credited to Alessandro Volta, an italian scientist, who published it in 1799. Volta used copper and zinc electrodes. In his early cells they were immersed in brine (a water solution of ordinary salt) which served as electrolyte, and in later versions in acids.

An experiment of little practical, but of really high educational value is to make the so-called “lemon battery”. Lemon juice, containing citric acid, is a good electrolyte. In order to make a cell, it’s enough to make two incisions in the lemon’s skin and insert a copper plate into one, and a zinc plate into the other (instead of plates, one can use nails, pieces of wire, etc.). By using several lemons and connecting the “cells” in series, one gets a battery with enough voltage to lit up a small LED. There are several videos in the YouTube worth watching, for instance, lemon battery, or another lemon battery.

Furthermore, Volta invented the “Voltaic pile”. He made the electrodes in the form of discs, and put a put a cardboard or felt disc soaked with electrolyte in between each copper-zinc pair – and then, stacked them up into a pile, as shown in Web page. A single pair produced about one Volt, so by building high piles one could obtain a source of a really high voltage – the record was about 2000 elementary disc pairs. The Voltaiuc piles were essential for all research on electric currents until about 1870, when more powerful sources, electric dynamos, took over.

We have not told what is the physical mechanism that forces the electrons to run from the anode to the cathode, and the ions in the electrolyte to take the charge back from the cathode to the anode. But here we will limit the explanation to a simplified “phenomenological” description – i.e., to one based primarily on empirical facts. Namely, there is a property characterizing electrode materials known as the Standard Electrode Potential (a Hyperphysics page) – in short, it’s the voltage supplied by a cell made of the given electrode material, and of a “standard electrode” – a widely acceptedone is the Standard Hydrogen Electrode. Tables listing the electrode potentials (denoted by the symbol E°) of different electrode materials can be found in many Web pages, e.g., in a very extensive table presented by Wikipedia, or, in a more concise table in the Hyperphysics page linked above. In addition, in the Hyperphysics portal there is a useful tutorial on how to calculate the voltage supplied by a cell made of two different electrode materials.

However, our natural curiosity clearly raises the question: “but what causes that a given material has just such a value of the E° parameter? Yet, there is no way to answer this question in a concise way, covering no more than a few paragraphs in this chapter. A Reader who would like to learn more about the in-depth theory of electrode potential should be referred to an electrochemistry textbook at the academic level – it should be noted, however, that the material in question may constitute a significant part of such textbook. Articles devoted to this material can also be found in the Web, and they are definitely not something for “short reading” – here is one example. Another good tutorial worth recommending is a sequence of Chapter 1.1 1.8 in Unit 1: Electrochemistry in Chemistry LibreTexts, a Web publication from University of California Davis.

The voltaic cells and batteries are all disposable, no rechargeable, so they cannot be used for energy storage. However, it is worth to start with them the discussion about batteries in general, because of their simplicity. This makes it easy to introduce a number of basic concepts that are important for all classes of electric batteries.

11.4.1.1: The Voltaic Cell is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.1.2: Lead-Acid Batteries

Lead-acid battery, the very first type of a rechargeable cell, was invented in France in 1859 by Gaston Plan´e. The positive electrode in such cell is lead dioxide PbO_2[2], and the negative one is metallic lead. The electrolyte is water solution of sulfuric acid H_2SO_4. When the cell is discharging, the PbO_2 in the "+" electrode is gradually converted to lead sulfide, $PbSO_4$, and the same compound forms on the lead "-" plate. In the process, most H_2SO_4 is removed from the electrolyte, so in a discharged battery the water/acid solution is much more diluted. But a discharged cell can be recharged by sending electric current through it in the "opposite direction". Now, the $PbSO_4$ in the "+" plate changes back to PbO_2, and the $PbSO_4$ in the "-" plate back to metallic Pb. In both reactions H_2SO_4 is "returned" to the electrolyte, so that at the end of charging it has the same concentration as at the beginning of the discharging process. More details concerning the chemical reactions occurring in the electrodes and the electrolyte can be found in the article linked above.

Even though the lead-acid technology is over 160 years old, such batteries are widely used today and they continue to be "respected members of the family of rechargeable current sources". This technology is mature and one knows exactly what can be expected from a lead-acid batteries. They tolerate harsh conditions well, such as, e.g., high temperature under the hood of a car or salty moisture on small ocean-going ships. Their electrolyte is nonflammable, in contrast to that in lithium-ion batteries. Also unlike the latter they tolerate well overcharging and do not require sophisticated electronic chargers. Their disadvantages include heavy weight and and relatively short lifespan (measured in the number of charge-recharge cycles).

As energy-storage devices, they may be a good least-expensive option for a single-home "off-grid" solar power installation. An "off-grid" system is designed to meet the needs of one home with no connection to the utility grid (the opposite is a "grid-tied" solar installation, which, in principle, does not require batteries).

There are two main types of lead-acid batteries: "automotive" and "deep cycle" ones. The former (also known as "cranking batteries") are designed to provide very strong currents, hundreds or even over one thousand of Amps, for a relatively short time – such currents are typically needed for starting a car engine.

The capacity of batteries (lead-acid, as well many other types) is usually given in in the **Ah** units, i.e., Amp/hours. For instance, if the capacity is given as 50 Ah, it means that the battery can supply the current of 1 A for 50 hours – or, say of 5 A for 10 hours. But since the advent of solar power and of electric cars another unit that expresses more directly the performance of a given battery began to be used: namely, the kilowatt-hour (kWh), a unit of energy. If we remember that 1 Volt multiplied by 1 Ampere gives 1 Watt, and we know the battery's voltage, then the conversion from Ah to kWh is very simple – e.g., a 12 V and 50 Ah cranking battery has the capacity of:

$$\text{12 V × 1 A × 50 hours = 12 W × 50 h = 600 Wh = 0.6 kWh}$$

Can we therefore use it as a battery for storing 0.6 kWh of energy? Now, it's time to **to make important warning**: if this is an automotive battery, then the answer is **no**. Charging the battery repeatedly with 0.6 kWh and then and then extracting from it the same energy portion would ruin it after a small number of such charge/discharge cycles! Automotive batteries are not designed for "deep cycling". If intended to be used for storage, it should be a **deep cycle battery**.

Because lead-acid batteries have been used for years, almost everyone is familiar with them – therefore, they play one more important role – as a convenient "reference standard" helpful for assessing other types of rechargeable batteries. The two most important parameters of such batteries from the point of view of practical applications are:

- Gravimetric energy density – i.e., the amount of energy that can be stored per mass unit.
- The number of charge/discharge deep cycles the battery guarantees. The energy density of practical lead-acid batteries is 25-40 Wh/kg, and the manufacturers usually guarantee a "lifetime" of 300-500 charge/discharge cycles.

For comparison: the popular nickel-metal hydride rechargeable batteries (widely available in AA and AAA sizes) have the energy density more than twice that of the lead-acid. The lithium-ion batteries of the 18650 type made by Panasonic and used in the Tesla cars have a record-high energy density of 265 Wh/kg. A table listing essential parameters of different rechargeable battery types can be found, e.g., in this Wiki page, or in this Electropaedia chart. Note that in most such tables the lead-acid battery is listed at the top as a familiar reference standard.

Due to the important role that lead-acid batteries have played in the “technosphere” for over 100 years, and the fact that there is a huge industry around the world involved in their production, the US Department of Energy and Argonne National Laboratory have recently launched an initiative to open a new chapter in research on these batteries , with the goal of significantly increasing their capacity and lifetime.

2It should be noted that most oxides of metals are insulators – PbO_2 is one of the few exceptions from that rule. It does conduct electricity, yes, although its conductivity, a parameter used to describe the “goodness” of electric conductors – is over 1000 times worse than that of copper, but is still good enough for an electrode material.

11.4.1.2: Lead-Acid Batteries is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.1.3: Lithium-Ion Batteries

For most of the 20th century, research into new types of batteries continued to be stagnated. In fact, the Ni-Cd rechargeable battery (invented in 1899) was the only mass-produced type with energy density higher(by about 50%) than the lead-acid.

In 1980 an American scientist, John Goodenough, invented a battery with entirely new lithium-based chemistry. Right from the beginning it was clear that the energy density of this new type was far superior to that of all previously known kinds of rechargeable batteries. The new "lithium ion" technology was first commercialized by SONY in 1991. The lightweightness of the new power sources made possible the emergence of new generations of portable electronics (e.g., smartphones, tablets), cordless power tools, electric cars, and last but definitely not least, of industrial-scale energy-storage batteries capable of smoothing out the fluctuations in power provided by intermittent sources such as solar panels and wind turbines.

In electrochemical cells normally more than one species of ions participates in transmitting current between the electrodes across the electrolyte. For instance, in lead-acid batteries these are the negative HSO_4ions and positive H^+ ions. In contrast, in the family of lithium-ion batteries only a single kind is involved – namely, positive Li^+ ions. Another difference is that in the lithium-ion cells the material of the electrodes is not a subject to chemical reactions. The electrodes are made of materials that can act as "sponges" for the Li ions. Just like a bathroom sponge that absorbs water or releases it when squeezed, the electrodes can absorb the ions from the electrolyte, or they can release them back.

The physio-chemical process underlying such "sponge-like" action is known as *intercalation*. There is a number of materials in which the constituent atoms or molecules are organized into layers, stacked on the top of each other – in such a fashion that that relatively wide "gaps" remain between the layers. Wide enough that they can accommodate small atoms of other elements. The best known "host" of such kind is *graphite*, one of the bulk crystalline forms of carbon (the other is diamond).

In a single layer (monolayer) in graphite the carbon atoms are arranged into a two-dimensional (2D) hexagonal lattice (often referred to as a "honeycomb lattice"). A single layer of this kind, if individually prepared, is known as *graphene*[3]. Graphite can be then thought of as a solid consisting of a large number of graphene monolayers. The three-dimensional stacking arrangement in graphite is such that at **above the center** of each elementary hexagon in one layer (call it A) there is an atom belonging to the next layer (call it B) – and, inversely, **under the center** of each hexagon in Layer B there is an atom belonging to Layer A. The arrangement of atoms in graphene and in graphite is very well illustrated by figures and animations in the two Web sites linked above, it is strongly recommended to watch them.

The distance between two carbon atom nearest neighbors in each layer is 0.2461 nanometers (nm), while the layer-layer distance is 0.3353 nm. Such "gap" is wide enough to accommodate lithium atoms – a bulk specimen of graphite can absorb one lithium atom per each 6 carbon atoms, so that the LiC_6 symbol is often used for graphite fully "saturated" with intercalated Li atoms.

Graphite is used as the negative electrode (anode) in almost all sub-types of lithium-ion batteries. There is a variety of materials used for positive electrodes. One particularly interesting such compounds is cobalt dioxide CoO_2 which was used in the very first lithium-ion batteries by their inventor, John Goodenough. The material also has a layered atomic structure. The "basic block" in it is an *octahedron*. The octahedron can be thought of as two pyramids sharing a common square base.

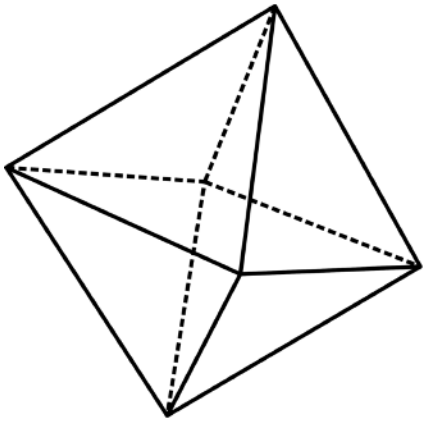


Figure 11.4.1.3.1: A regular octahedron. Each of the eight identical faces is an equilateral triangle.

In CoO_2 the basic octahedron is formed by six oxygen atoms located at its six vertices, with the cobalt atom residing in the center of the common square base. The octahedra are arranged into layers with relatively wide "gaps" between them (see the Fig. 11.4.1.3.1). Therefore CoO_2 can become an "intercalation host" for lithium. The gaps can accommodate one lithium atom for each cobalt atom – so, when fully intercalated, the formula of the system becomes $LiCoO_2$ (lithium cobalt dioxide). But the intercalation process is fully reversible, CoO_2 may be thought of as a "fully de-intercalated" lithium cobalt dioxide. The system may exist in any intermediate form between CoO_2 and $LiCoO_2$, and then the formula is Li_xCoO_2, where x is a number $0 < x < 1$.

The anatomy of a lithium-ion battery. The two intercalation electrodes are separated by a layer of electrolyte, which is a solution of a lithium salt (most often, it's lithium hexafluorophosphate $LiPF_6$) in an organic solvent. The electrons flow in or out of the electrodes through *contact plates*. These plates are directly connected to the cell terminals. Inserted into the electrolyte, between the two electrodes, is a *separator* – a thin porous flexible plate or a mesh through which ions can pass freely, but it prevents short circuits between the two electrodes (i.e., touching one another).

The arrangement of the electrodes, the contact plates, the electrolyte and the separator is shown schematically in the Fig. 11.7. The parts (a) and (b) of the figure show the flow of Li^+ ions in the electrolyte, and of electrons in the external circuit when, respectively, current is drawn from the battery during normal operation, and during battery recharging. In the next Figure 11.4.1.3.2 the flow of Li^+ ions and electrons is shown only for a normal currentdrawing operation, but here more attention has been paid to showing the atomic structure of the graphite anode and the Li_xCoO_2 cathode.

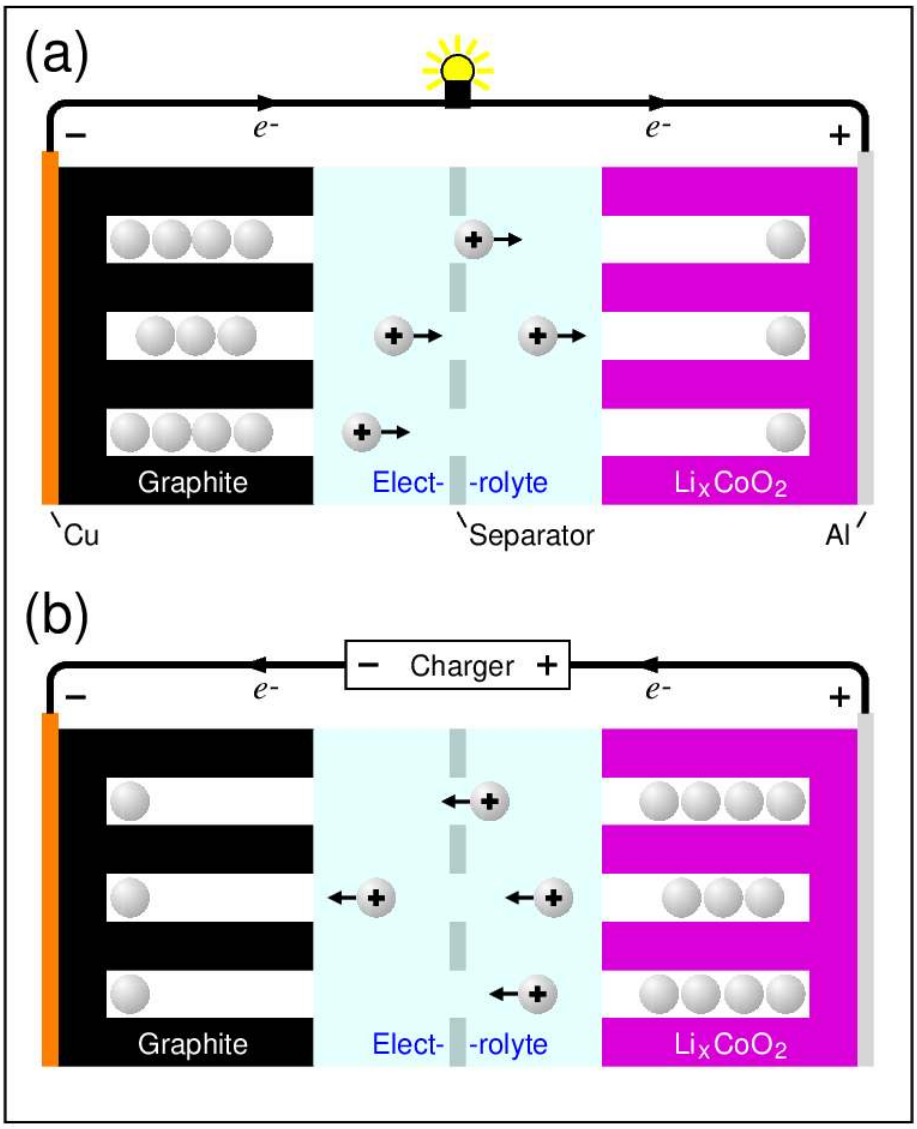


Figure 11.4.1.3.2: Schemes showing the arrangement of intercalation electrodes in Li-ion batteries and the motion of Li^+ ions between the electrodes in (a) the discharging, or "current-supplying" mode, and (b) in the recharging phase (for more details, see the text).

When current is drawn from the battery the electrons flow out of the graphite anode through the external circuit (in the Fig. 11.4.1.3.2 (a) it contains a light bulb) and reach the cobalt oxide cathode. But charge cannot get accumulated in the electrodes! Therefore, the influx of the negative charge of electrons must be "neutralized" by the influx of the same amount of *positive charge* – and this charge is supplied by positive Li^+ ions entering the electrode from the electrolyte.

A reverse process occurs at the same time in the anode: the outflow of electrons would charge it positively therefore, the electrode prevents the accumulation of this charge by ejecting the positive Li^+ ions. As follow from the above, the current-supplying phase may last as long as there are still ions in the anode, and there is still space in the cathode for accommodating additional ones.

In the battery charging phase (Fig. 11.4.1.3.2 (b) all these processes are reversed. A device known as a "charger" now forces electrons to leave the cathode and enter the anode – so that the Li^+ ions are now ejected from the cathode and re-enter the anode. Note that the statement *electrons leave the cathode* means, at the same time, that *classical current flows* ***into it*** – in other words, during the re-charging process the cathode and anode change their roles, the cobalt oxide electrode becomes the anode, and the graphite electrode becomes the cathode.

Note that in contrast to the lead-acid battery the electrolyte does not participate in any chemical reactions – it acts only as a "passive" medium allowing the passage of Li^+ ions between the electrodes.

In Fig. 11.4.1.3.3 there is another graph showing the "anatomy" of a Li-ion cell, with more attention paid to the atomic structures of the electrodes that allows them to serve as "intercalation hosts" for the Li^+ ions: namely, the hexagonal graphene-like atomic layers in the graphite anode, and the layers built of elementary CoO_6 tetrahedra in the cobalt oxide cathode.

In fact, the anatomy of real commercially available Li-ion batteries is not very similar to what is shown in the Fig. 11.4.1.3.3. The most common battery versions are cylindrical in shape. A special five-digit designation code has been created for such Li-ion cells: the first two digits are the cylinder diameter in millimeters, and the other three digits indicate the cylinder height in tenths of a millimeter. Accordingly, the most popular cell (i.e., being produced in the largest number and used, among others, in Tesla cars) is the "18650", slightly larger than the AA battery (whose "code name", if the same rule was used, would be 14500).

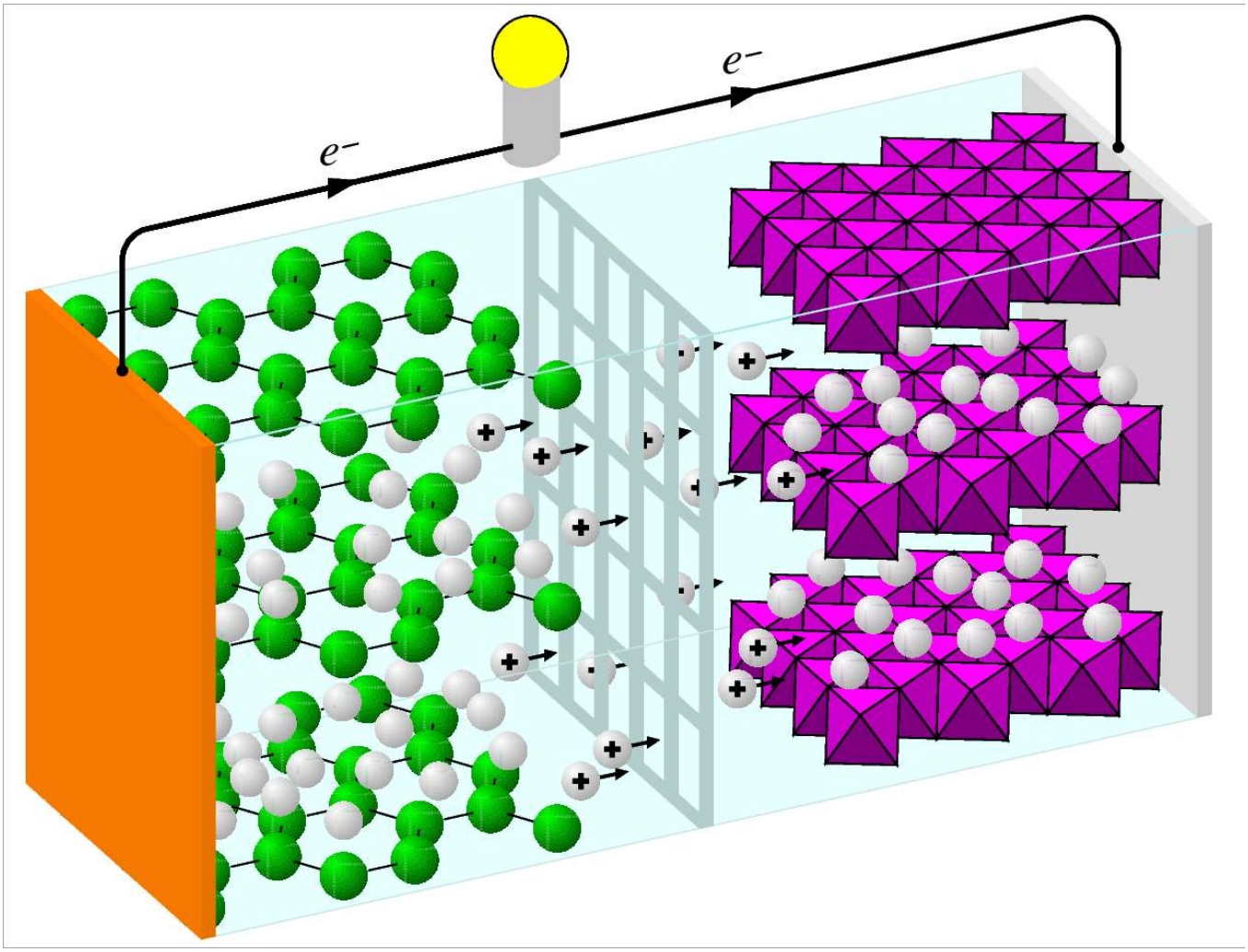


Figure 11.4.1.3.3: A graph showing the anatomy of a Li-ion cell in closer detail. The green spheres are the carbon atoms in a graphite electrode. The magenta colored layers consist of CoO_6 octahedra. White spheres are lithium atoms (if between the carbon or CoO_6 layers) or ions (if traveling through the electrolyte which is colored light-blue). On the left there is a copper plate collecting electrons from the graphite anode, and on the right side there is an aluminum plate from which electrons flowing in from the external circuit enter the cobalt oxide electrode. The grid-like object in the middle is the separator.

The electrodes in such batteries are made of graphite and CoO_2 ground to powders consisting of microcrystals and mixed with a plastic-like compound which makes it possible to prepare them in the form of long thin "ribbons", or stripes. The contact plates are also replaced by strips made of aluminum (Al) an copper (Cu) foil. The remaining two elements, the electrolyte and the separator, can be prepared jointly in the form of a stripe of fabric (e.g., made of glass fibers) soaked with the electrolyte. Then, the stripes are stacked on top of each other in the following order: Al foil stripe, the CoO2 cathode stripe, separator with electrolyte stripe, graphite anode stripe, and the Cu foil stripe. The idea is to roll the stacked stripes into a tight "scroll". However, in such a scroll the Al and Cu foils would touch one another; therefore, one extra stripe made of insulating foil is put at the very bottom of the stack.

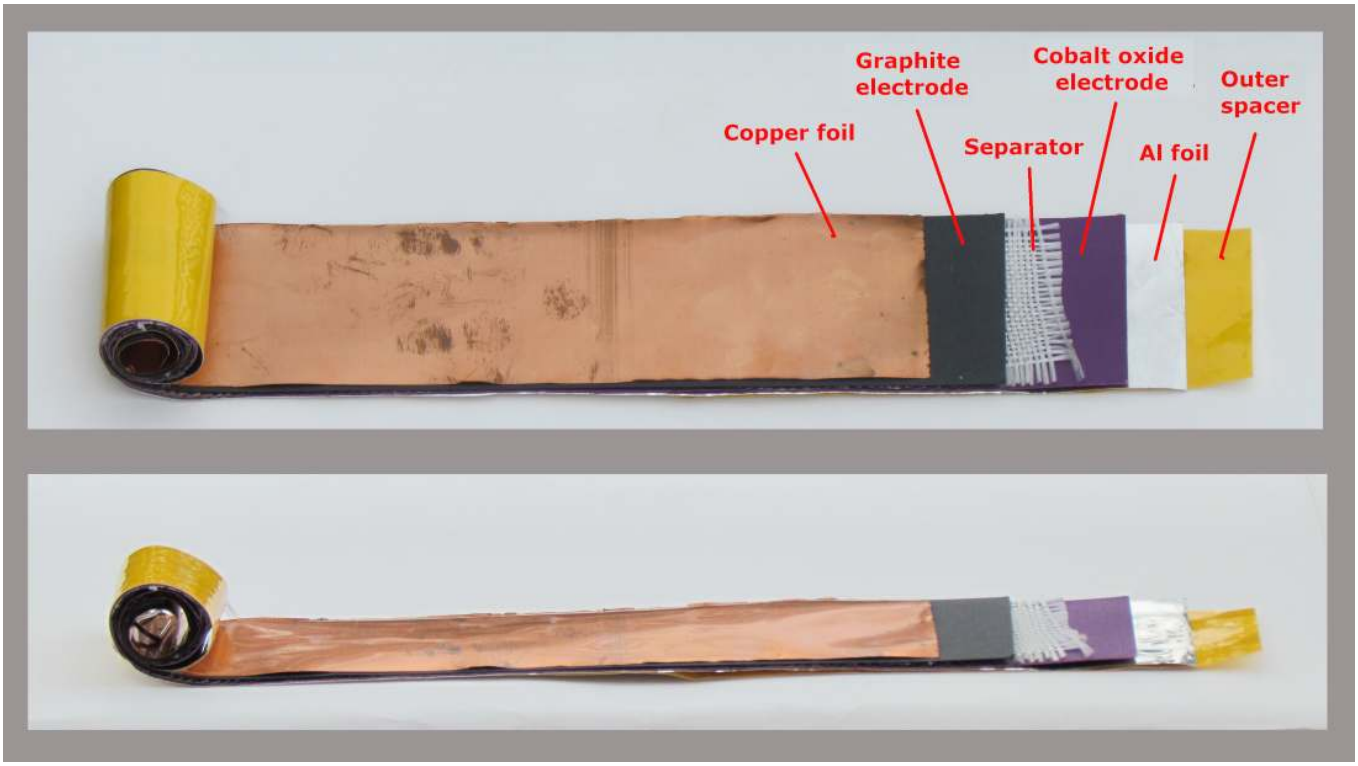


Figure 11.4.1.3.4: A model explaining how cylindrical Li-ion batteries are manufactured. Several stripes are stacked in succession: aluminum foil; the cathode – a cobalt oxide electrode (microcrystals in a conducting plastic matrix); the separator (e.g., fiberglass fabric soaked with electrolyte; the anode – a graphite electrode (fine powder in a conducting matrix); and copper foil. The whole stack is rolled into a tight "scroll" – for that reason, an "outer separator" made of insulateing foil must be added, otherwise the aluminum and copper foils would touch each other. The cathode is not necessarily made of cobalt oxide, oxides of other metals can be used, or a mixture of different oxides. The scroll, when ready, is put into a cylindrical steel capsule. The two photos of the model are the view from above and from the side.

After being rolled into a tight "scroll", the whole assembly is put into a cylindrical capsule, most often made of steel – with the terminals at the two cylinder bases. But "rolling" is not the only possible way of processing the stripe stack – another method used by some manufacturers is "accordion folding" at which one gets a "sandwich" that can be put into a flat capsule. Such li-ion batteries are commonly referred to as "prismatic". Some electric vehicle (EV) manufacturers (e.g., Tesla) use cylindrical cells because when thousands of them are assembled into large batteries there is space left for easy coolant circulation (during a rapid charge operation the cells must be intensively cooled). Other (e.g., Nissan Leaf or BMW) prefer the prismatic type because they offer a better stored energy per unit volume ratio. Prismatic cells are also the obvious choice for portable electronics devices, such as, e.g., cell phones or tablets.

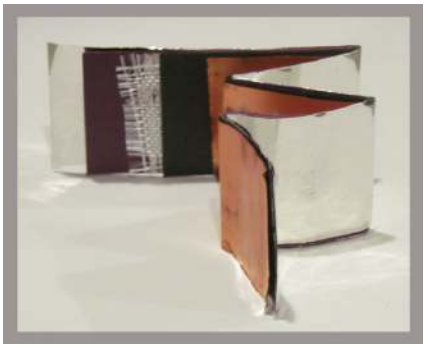


Figure 11.4.1.3.5: "Accordion-type" folding of the stack, used in prismatic and pouch Li-ion batteries. Note that the "outer separator" is missing – it's not necessary because in this type of folding the aluminum and copper foils can not touch each other.

3. It is not trivial to prepare large specimens such two-dimensional layers; Andre Geim and Konstantin Novoselov who first mastered the technique in 2004, were awarded a 2010 Nobel Prize in physics for that

11.4.1.3: Lithium-Ion Batteries is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.1.4: The “Families” of Li-ion Batteries Manufactured

Most Li-ion batteries currently are several sub-types currently manufactured Li-ion batteries in which different cathode materials, anode materials, or electrolyte type are used. Depending of what the battery is going to be used for, the users must take into account several important parameters, or “figures of merit” – namely:

- Gravimetric energy density – i.e., how much energy can be stored in a battery of unit mass, usually expressed in the units of Wh/kg);
- Volumetric energy density – energy per unit volume, usually given in Wh/liter);
- Power density (maximum power that can be drawn from a battery of unit mass or unit volume – respectively, in W/kg or W/liter);
- The “lifetime” – i.e., the guaranteed number of charge/discharge cycles at which the energy density will stay above a certain threshold usually 80% of the original density;
- Charging time – the shortest recharging time the battery can tolerate (too fast a charging leads to overheating and damage to the battery);
- The charge/discharge rate[4] in the so-called “C units”;
- Cost;
- Safety.

An even more comprehensive list of important parameters can be found in this MIT Web page.

For EV makers, for instance, the gravimetric energy density and the charging time are highly important. For aircraft makers who want to install Li-ion batteries in a new model, the safety is of utmost importance. For a not-too-wealthy retired hobbyist working on a small project the cost may be a meaningful factor.

Below the main existing types of Li-ion batteries are discussed.

Graphite anodes + 3d transition metal oxide cathodes. The “3d transition metals” are those with numbers from 21 to 30 in the Periodic Table (the “3d” in the name does not mean “three dimensions”, but refers to the 3d electronic orbitals in their atoms). The first Li-ion cells created by John Goodenough and coworkers in the 1980-s had intercalation cathodes made of Li_xCoO_2 (commonly referred to as “Lithium Cobalt Oxide”, or LCO) – as described in the text above. Cobalt, however, is expensive and raises some moral objections – namely, most of it comes from the Democratic Republic of Congo, in which “democracy” can be found primarily in its name, but not much in the functioning of the state. In particular, human rights are not respected in the cobalt mines, where the use of child labor is commonplace. In search for less controversial substitutes of cobalt, researchers started experimenting with some other transition metal oxides. One intercalation host that was found to be a good cathode material was Lithium Manganese Oxide (LMO). Many other efforts were focused on mixtures, or alloys of several 3d metal oxides which did not eliminate cobalt completely, but needed much less of it. A number of such cathodes from oxides have been developed, such as, e.g., Lithium Nickel Manganese Cobalt Oxide (NMCO), or a composition with the addition of aluminum (Al), Lithium Nickel Cobalt Aluminum Oxide (or NCAO). Aluminum does not belong to the family od 3d transition metals, but a certain fraction of Al in the oxide mixture has been found to result of a record-high performance of such cathode. The batteries used in the Tesla cars have NCAO cathodes and a record-high energy density of 265 Wh/kg. A comprehensive review of Li-ion battery sub-types, with the discussion of the “figures of merit” listed above for each sub-type, is presented in this Web article.

Graphite anode + Lithium Iron Phosphate($LiFePO_4$) cathode. In 1996 scientists from the University of Texas discovered a viable non-oxide based cathode material, $LiFePO_4$ (LFP). The energy density of batteries with such cathodes is only about one-half of that of the best batteries with oxide-based ones, but the offer several meaningful advantages – e.g., higher tolerance to heat and therefore they are much less likely to catch fire than the other types. Also, they experience much slower degradation, their “lifetime” can reach thousands of charge/recharge cycles. Yet another advantage are their high charge/discharge rates, higher than those typically seen in cells with oxide cathodes.

The nominal voltage of a LFP cell is 3.2 V (slightly lower than the 3.7-3.8 V of Li-ion cells with oxide cathodes). Four of them make a 12.8 V battery that – due to a high discharge C rate – may offer a “lightweight replacement” for a 12 V lead-acid one typically used in automobiles. With one caveat, though – lead batteries in cars are constantly recharged with 14.4 V, which they

tolerate very well while in phosphate iron, such a regime leads to excessive charging which quickly damages them. Therefore, such batteries should be installed in cars only with additional electronics controlling the charging process.

Also, due to their reliability and tolerance for "harsh environment", Iron phosphate batteries are especially recommended as replacement of lead-acid batteries in marine applications, in yachts and small vessels.

Batteries with Lithium Titanate anodes ($Li_4Ti_5O_{12}$). They are the first commercially available Li-ion batteries with a non-graphite anode. Even though their capacity is only about the same as that of nickel metal hydride batteries – in other words, not impressively high – but they are well suited for certain special applications due to their exceptional tolerance to low temperatures, and their very high charge/discharge rates.

Lithium polymer (LiPo) batteries. The name is a bit misleading because the LiPo-s are essentially Li-ion batteries using the same anode and cathode materials as the "regular" Li-ion ones. The main difference is in the type of electrolyte used. The "regular" cells use a porous separator material soaked with liquid electrolyte. In the Lithium-polymer cells the electrolyte has the form of a thin film made of a gelled ion-conducting substance. Such an electrolyte effectively protects against "short-circuiting", i.e., a direct contact between the cathode and anode. Thus, no separator is needed as an additional element. Some commercially available LiPo batteries are packed not in metal capsules, but in flexible plastic envelopes that one can bend within certain limits Without fear that the internal structure will be damaged.

Trends in research on new types of L-ion batteries.

Lithium metal anodes. In the early efforts of creating lithium-ion batteries anodes made of bulk lithium metal, not of intercalated lithium atoms were used. Those attempts were unsuccessful, mainly because during the recharging process lithium ions were not deposited uniformly on the flat anode surface, but instead formed "three-like structures" called *dendrites*. After several charge/discharge cycles the dendrites created shortcuts between the anode and the cathode, which not only rendered the battery useless, but might even set in on fire. The dendrite-growing process is well illustrated in this short video made in the Oak Ridge National Laboratory.

Recently, attempts of using pure lithium metal anodes have been renewed. The stake is high! Why? Let's keep in mind that in the currently used graphite anodes in a fully charged state there is one Li atom per each **six** carbon atoms. The atomic weight of Lithium is 7, and of carbon it is 12. Hence, for a fully charged anode, LiC_6, the total molecular weight is 6 x 12 + 7 = 79. The fraction of lithium in it is thus 7/79 = 8.9% of the total weight. Accordingly, replacing the intercalated material by pure lithium would reduce the weight of the anode more than tenfold! Even if all other component of the battery are not changed, the anode of pure lithium may nearly double the gravimetric energy density of the battery. The total mass of the battery pack in the famous Tesla car is 540 kg and if fully charged, it can propel the car over a distance of 400 miles. A 540 kg package of batteries with lithium-metal anodes would nearly double that range! So, as the old proverb says, *the game is worth the candle!* Scientist, in this recent article, are investigating several possible ways of suppressing the dendrite growth – one of them may be, for instance, using not a liquid, but a "solid state electrolyte" sturdy enough that the dendrites would not be able to penetrate it. John Goodenough, the inventor of the lithium-ion battery, recently announced that a team led by him had developed such an electrolyte.

Silicon anodes. Silicon can become "lithiated" – not through intercalation, but through another physicochemical mechanism it can bind as many as 3.75 lithium atoms per one silicon atom, while graphite can host one Li atom per six carbon atoms. The atomic weight of silicon is 28. Hence, if a silicon crystal is in a "fully lithiated" state the, the fraction of lithium in it is (3.75 7)/(3.75 7+ 28) = 48.4% by weight – more than five times a better figure than the 8.9% for graphite. So, it raises hopes that silicon may become the anode material in next generations of "lithium silicon batteries". However, Mother Nature has a perverse sense of humor – the volume of a fully lithiated silicon is 300% of that of the pure element (for comparison, the volume of graphite increases by only 10% when it is saturated with lithium). Yet, the lithium-silicon battery concept is so attractive that researchers are trying to design it in such a way that the increase in anode volume would not interfere with its operation – see, e.g., this report.

The Lithium-Sulfur battery. In 2008, the British BBC agency reported a new sensational world record: namely, a pilotless solar powered aircraft Zephyr-6 stayed aloft for three days at an altitude of 60 000 feet. What? An overnight flight of a solar-powered aircraft? Yes, it was possible, thanks to a new ultra-light rechargeable battery pack. During daylight hours the solar panels on wings yielded enough power to operate the engines and to recharge the battery – with enough energy to keep the plane aloft overnight.

The new sensational battery had the energy density twice as high as the best li-ion ones available in 2008. A metallic lithium anode and a sulfur cathode were used. In later years the record was significantly improved with the plane remaining in the air for 11 days. The energy density of the battery pack used was 500 Wh/kg, almost two times better than the 265 Wh/kg of the best Li-ion cells used in the Tesla cars.

So why aren't these lithium-sulfur batteries widely used in electric cars?

They could increase the range of the Tesla auto to almost 800 miles which would probably satisfy even the most demanding customers. Unfortunately, they are still haunted by problems. The most serious is that they offer too few charge/discharge cycles. For current electric vehicles the required battery lifespan is at least 500 charge/discharge cycles. The Li-S cells made in laboratories sometimes attain that many, but a technology that would make it possible to manufacture them on industrial scale with required parameters has still not been mastered. Some optimistic news have been recen.tly heard from Australia, and a British company Oxis Energy promises to manufacture Li-S rechargeable batteries with 600 Wh/kg within five years. Let's keep our fingers crossed and watch the news!

Further reading: for a deeper understanding of the physicochemical processes occurring in lithium-ion cells, it is worth recommending the papers linked below. They present, in advanced manner, both an approach based on macroscopic thermodynamic theory in terms of Gibbs free energy, as well as an approach based on a detailed analysis of microscopic electron states in electrode materials in the light of quantum theory of conducting solids.

• ChaofengLiu *et al.* (2016).

• Gennady Cherkashinin *et al*. (2019).

• Arumugam Manthiram (2020).

---

4. The battery capacity can be expressed in terms of energy units – but another, perhaps even more common measure are ampere-hours (A h). A capacity of *N* A h means that a fully charged battery will be completely discharged by delivering a current of *N* Amps for one hour. The charging/discharging rate unit "C is simply related to the A h capacity measure. For instance, if a battery of, say, 10 A h capacity has a charging rate of 0.5C, it means that the maximum charging current for it should be 0.5*10 A = 5A. A discharge rate of 5C would mean that the maximum current safely drawn from this battery cannot exceed 5*10A = 50 A.

11.4.1.4: The "Families" of Li-ion Batteries Manufactured is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.1.5: Batteries for Industry- and Grid-scale Storage

The “size spectrum” of Lithium-ion batteries is very broad, it covers all possible range of sizes – starting from those in miniature portable electronic devices and ending with huge “grid-tied” storage ones with capacities of hundreds of MWh. However, in the latter group there are neither the only possible technology nor the largest existing devices. The lithium-ion technology is the best known because we constantly deal with such batteries. The other types are less known because they exist only as large stationary devices, but not in the portable form. However, they can successfully compete with the li-ion technology offering a number of advantages: lower material costs, much longer “lifespan” calculated in the number of charge/discharge cycles. Below, three such types are presented: Sodium-Sulphur (NAS) batteries, the so-called “redox flow batteries”, and molten metal batteries.

The Sodium-Sulfur (NAS) batteries. They were invented in Ford Motor Company in 1966. Due to their high gravimetric energy density, several times better than that of lead-acid batteries, it was hoped that would allow developing electric vehicles with a range above 100 miles. However, there were several reasons that made the use of such batteries in automobiles highly risky and, last but not least, pretty inconvenient for the car user. But they may be attractive when it comes to stationary energy storage – the best proof for that is that the largest currently (March 2020) existing storage battery is just of the NAS type.

Most of the existing battery types use solid electrodes and liquid electrolytes. In an elementary NAS cell this scheme is reversed – namely, the electrodes are molten sodium metal and sulfur and the electrolyte is a solid ceramic material known as sodium β-alumina. The design of a NAS battery is pretty simple, as illustrated by graphic schemes in this Web page of Zhang’s Research Group or in ScienceDirect Web article. The liquid sodium electrode fills a cylindrical vessel which is made of β-alumina, and this vessel is immersed in molten sulfur contained in a cylindrical vessel of larger diameter.

The β-alumina is a material obtained by sintering aluminum oxide and sodium oxide. Its unique property is that at elevated temperatures it is not permeable for to sodium and sulfur atoms while it permeates *sodium ions* (a sodium ion Na^+ is created when the single outer electrons is “pealed off” a sodium atom – therefore, the ion’s radius is much smaller than that of the atom). In the discharge phase the Na^+ ions penetrate the β-alumina partition, “leaving behind” their electrons. Those electrons flow out of molten sodium and through an external circuit they enter the sulfur electrode, where sulfur atoms, sodium ions and electrons react and form molecules of sodium polysulfate Na_2S_4:

$$2Na^+ + 2e^- + 4S \rightarrow Na_2S_4$$

During a recharging process these reactions are reversed.

Note that the NAS batteries are made of widely available and inexpensive components. Their other significant advantages are longevity (if properly exploited, up to 15 years, or 5000 charge/recharge cycles) and high gravimetric energy density, comparable with the best Lithium-ion batteries currently on the market. So, where is the problem?

The problem is the temperature, 300˚ 350˚C that has to be maintained to keep the electrodes molten and the β-alumina partition between them in an ion-conducting state. Well, you may know that since the medieval ages molten sulfur was believed to a terrible substance, used by devils torture the sinners’ souls in the hell. In fact, it is a devilish fluid indeed, highly flammable and emitting toxic and caustic SO_2 fumes when burning. And what medieval people were ignorant of – because they had no idea that such a thing as sodium metal does exist – molten 300˚C sodium is even far worse than molten sulfur! One cannot even use water to extinguish it, because water, instead of suppressing the fire, adds fuel to it!

The high temperature would be a major inconvenience for owners of electric vehicles with NAS batteries because heating up the battery pack before taking off in the car would take many minutes. But what might have happened if the the battery pack were damaged in a traffic collision would be an authentic drama!

But dealing with huge scale stationary industrial high-temperature installation is nor a problem and engineers have much necessary experience in this field. In the smelting of iron, ammonia synthesis, cement production, and many other industrial processes furnaces and installations with masses of hundreds or even thousands of tons are heated to temperatures even much, much higher than 300˚C.

Suppose that a storage pack with a capacity of 100 MWh is needed. NAS with energy density as high as 200 Wh/kg are available. So, it would be 5 kg per 1 kWh, 5 000 kg = 5 metric tons for 1 MWh, and 500 metric tons for 100 MWh. It does not seem likely that an engineer would panic when given a task of heating up a 500 ton piece of hardware to 300˚C!

In fact, NAS batteries have been used for building **the largest in the world** (as of March 2020) electric energy storing facility in Abu Dhabi. Its total capacity is 648 MWh, five times larger than the Tesla-supplied world largest Lithium-ion storage installation in Hornsdale, Australia.

Catastrophic fires may happen in installations using NAS batteries. There was one such serious accident in Tsukuba, Japan on Oct. 21, 2011. It caused a serious disruption in the Japanese energy storage development program. The creators of the Abu Dhabi system demonstrated wisdom and caution by building it not as a single structure, but in the form of 15 separate "blocks" located at a safe distance from one another in a desert terrain. Therefore even in the event of a catastrophic fire in one of them, neither other blocks nor any human settlements will be threatened.

Vanadium redox flow batteries. The word "redox" has been coined from two others, "reduction" and "oxidation". In chemistry, if an atom loses one or more electrons and becomes a *positive* ion, such a process is called "oxidation". The *oxidation state* of an ion is simply the number of electrons it has lost. Reduction is an opposite process – if an atom **gains** one or more electrons and becomes a *negative* ion, the process is called "reduction". Also, if a positive ion at a high oxidation state gains an electron and still remains a positive ion, but with a lower oxidation state, it's also called "reduction".

The oxidation state of an ion is denoted by a number lost electrons written as a right-hand superscript to the chemical symbol of the element in question, followed by the plus sign. The same code may be used for negative ions, i.e., such that gained one or more extra electrons – but then after the number of gained electrons goes a minus sign. So, negative ions are in a negative oxidation state.

Vanadium can occur in various water-soluble compounds in four different oxidation states: V^{2+}, V^{3+}, V^{4+} and V^{5+}. One more piece of information important for the further discussion is the *ionization energy* of each of these states – i.e., how much energy is needed for removing the given number of electrons from the atom. These energies are, respectively, 6.7, 14.6, 29.3, 46.7 and 65.3, all in electron-Volts (from the www.webelements.com Web portal where molar enthalpy values of the vanadium oxidation states are given in kJ/mole units – they can be readily converted into individual ion energies in eV using, e.g., this on-line tool).

The exact theory of the vanadium-redox flow battery is pretty sophisticated, so here we will not even try to make a detailed presentation of it. In fact, it's not needed. The thing that really matters is making it clear how energy is stored in the battery using a possibly simple "pedagogical tools".

OK, so suppose that there is a liquid containing two types of vanadium ions, V^{2+} and V^{5+}. Now, what needs to be done to convert a V^{2+} ion to a V^{3+} one? Well, the energy of the former is 14.6 eV, and of the latter is 29.3 eV. So, if we want to remove yet another electron from the V^{2+}, an energy input of 29.3 eV 14.6 eV = 14.7 eV is needed. Such a transition process can be described by an equation:

$$V^{2+} + 14.7\ \text{eV} \rightarrow V^{3+} + e^{-}$$

Conversely, an electron can be added to a V^{5+} ion with the energy of 65.3 eV to lower its oxidation state to V^{4+} with 46.7 eV of energy. Thus, in the process 65.3 eV 46.7 eV = 18.6 eV of energy will be released. Again, one can write an equation:

$$V^{5+} + e^{-} \rightarrow V^{4+} + 18.6\ \text{eV}$$

Then, suppose that there is a "close encounter" between V^{2+} and V^{5+}. The two ions may then exchange one electron. An equation describing such a process can be obtained by adding the two above equations, viz.:

$$V^{2+} + 14.7\ \text{eV} + V^{5+} + e^{-} \rightarrow V^{3+} + e^{-} + V^{4+} + 18.6\ \text{eV}$$

After reducing the e^{-} appearing on both sides and grouping the two energies on the right side, we obtain:

$$V^{2+} + V^{5+} \rightarrow V^{3+} + V^{4+} + 3.9\ \text{eV}$$

So, in such an encounter between the two ions and "electron swapping" 3.9 eV of energy is released. It means that such a reaction is – as chemists call it – "exothermic", i.e., energy is released. Exothermic reactions occur spontaneously – therefore, by mixing solutions containing equal numbers of V^{2+} and V^{5+} ions, one obtains a mixture containing V^{3+} and V^{4+} ions, somewhat warmer than the two initial solutions.

But where is the battery?! – one moment! If we mixed the solutions, the ions would directly swap the electrons. Let's then keep the solutions in two separate vessels, and give the ions another option of how to swap the electrons. How? By wire! Let's dip the ends

of a wire in the two vessels. Electrons will start flowing down the wire from the vessel containing V^{2+} ions to the one with the V^{5+} ions! Hurray! We do have our battery!

Just a moment, just a moment! There is still one problem. If the electrons starts flowing, soon in the "electron donor" vessel a positive charge will start accumulation, and in the "electron acceptor" one negative charge will. Positive charge will tend to pull the electrons back, negative charge will repel them – so, after a brief moment, the current will stop flowing in the wire. Things will work only if we find a method of "neutralizing" the charges that build up.

The credit for solving the problem – end, effectively, for the invention of redox flow battery – goes to an Australian scientist, Maria Skyllas Kazakos. She obtained a patent for the invention in 1986. The ingenious solution of the charge build-up problem was to keep both solution in the same vessel, but separated by a partition permeable to hydrogen ions H^+ – they carry the excess positive charge away from the V^{2+} chamber into the V^{5+} chamber, where they neutralize the negative charge building up.

But why it's called a "flow" battery? It's because fresh solution containing V^{2+} ions (called the "Anolyte")and solution containing V^{5+} ions (called the "Catholyte") are continually pumped through the reaction chamber, in order to replace the V^{3+} and V^{4+} that "have already done their work". There is one important conclusion that can be reached right away: what is the energy capacity of such a battery? The answer: there is no upper limit! If one needs to increase the energy storage capacity of the battery, it can be done simply by adding more anolyte and catholyte tanks!

We have only discussed the processes in the discharge phase. But the battery is fully rechargeable, instead of drawing current from it, current from an external source should be run through it – while the "spent" anolyte and catholyte are pumped back through the chamber. Over there, the V^{3+} ions will be converted back to V^{2+} ones, and the V^{4+} ions back to V^{5+} ones.

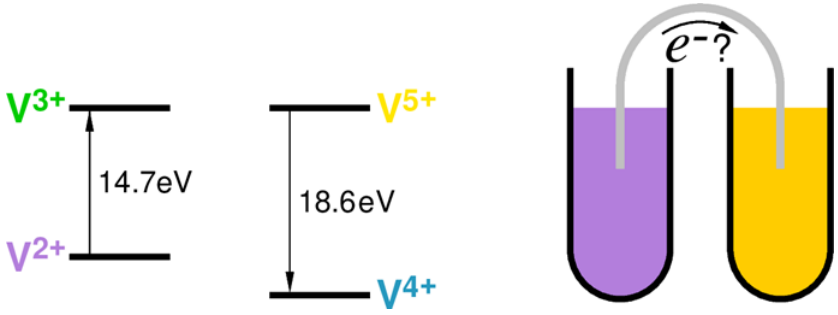


Figure 11.4.1.5.1: From the left there is a graphic representation of the oxidation energy levels of Vanadium ions in the anolyte (V^{2+} and V^{3+} and V^{3+}) and in the catholyte (V^{4+} and V^{5+}). The colors of the symbols labeling the levels are about the same as the colors of solutions containing such ions. Any physical system always tends to reach its lowest energy – therefore, the electrons tend to "jump out" of the V^{2+} level and "land" at the V^{4+} level, which lowers the energy of the pair of ions involved by 3.9 eV. Yet, connecting a solution containing V^{2+} ions with one containing V^{5+} with a wire to "help" the electrons to make such a jump is not a way of creating a battery because due to the flow of electrons the lavender-colored solution acquires a positive charge, and the yellowish-colored solution accquires a negative charge. After a small number of electrons flows across the wire, voltage builds up that stops any further flow.

A nice property of vanadium is that each oxidation state gives the solution a distinct color: 2+ is lavender, 3+ is green, 4+ is blue and 5+ is yellow (see, e.g., the picture in this chemicool.com Web page. It makes plotting graphs explaining how VRFBs work, because the four essential different phases of anolyte and catholyte can be shown by distinct colors – see Figs. 11.4.1.5.1: and 11.4.1.5.2:

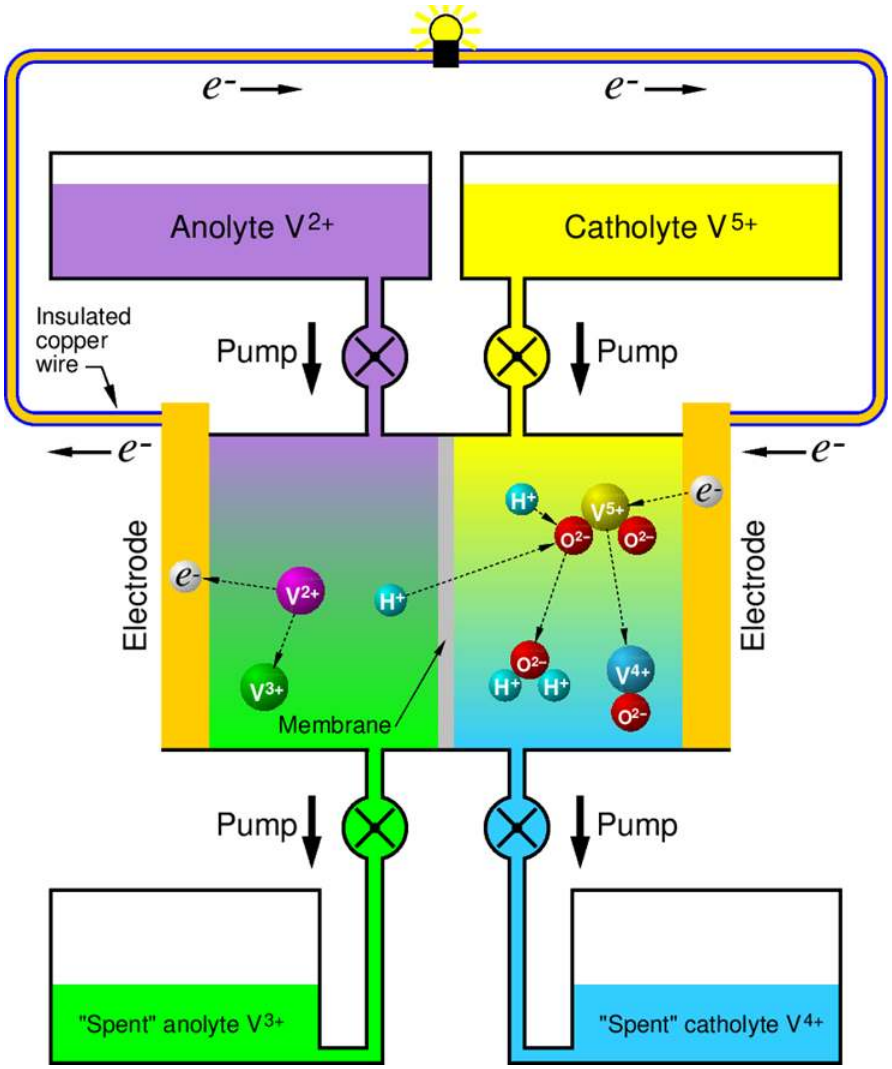


Figure 11.4.1.5.2: A scheme explaining the functioning of a Vanadium-redox rechargeable cell. See the text for details.

A simple concept of a battery shown in Fig. 11.4.1.5.1 would not work –unless there is an additional mechanism preventing the build-up of charges in the anoand catho-lyte.

As noted, the problem was solved by Maria Skyllas Kazakos, and a graph explaining in closer detail how her “recipe” works is presented in the Fig. 11.4.1.5.2. In both the anolyte and the catholyte the respective V^{2+} and the V^{5+} ions are kept in strongly acidic solutions – the reason is to create an environment with an abundance of hydrogen H^+ ions. The two solutions are pumped into a “reaction chamber” in which there are two compartments separated by a membrane permeable to H^+ ions. In the left compartment, the V^{2+} V^{3+} transition occurs and the released electron is picked up by the electrode. Then it continues its trek across the wire to the electrode of the other compartment, from which it enters the catholyte. At the same time a positive hydrogen ion passes through the membrane, thus neutralizing the positive charge created in the chamber by “escape” of the electron.

In the right compartment the V^{5+} V^{4+} transition involves a somewhat more complicated process. The thing is that the V^{5+} ion in the catholyte forms a complex with two negative oxygen ions, $V^{5+}O(^{2-})_2$. When the V^{5+} absorbs the electron arriving from the other compartment, it becomes “less positive”, which weakens bonds with the oxygen ions. One of these gets released from the complex, it attaches two hydrogen ions and forms a neutral molecule of water H_2O (in the Fig. 11.12 one of these H^+ ions is the one arriving from the other compartment; of course, it does not have to be this way, in the catholyte there is an abundance of of other H^+ ions). The vanadium ion, now V^{4+}, remains combined with the other oxygen and it becomes a part of the “spent” catholyte.

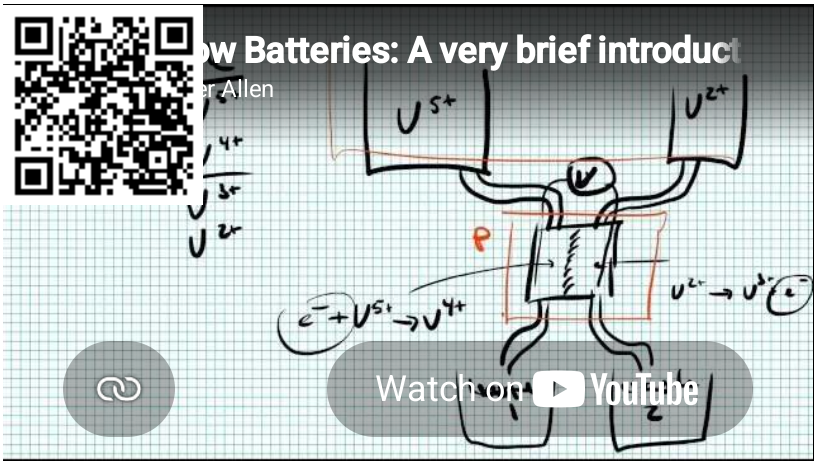
The “spent” anolyte and catholyte contain, respectively, V^{2+} and V^{4+} ions. After leaving the reaction chamber they are stored in separate tanks. In the re-charging phase, all liquids are pumped back through the reaction chamber and current from a recharging

source is sent through it – now electrons enter the anolyte compartment, and exit the catholyte compartment. The “refreshed” fluids are returned to their original tanks.

In the scheme in Fig. 11.12 there are four tanks. In literature and in videos on the subject most often a two-tank version is shown, where the spent fluids are returned to the same tanks from which the fresh fluids have been taken. The Author of this book, though, believes that the four-tank version is more “pedagogical”. At least then the question does not arise: *and how is mixing of fresh and spent anolyte and catholyte prevented?*

There are several short videos on YouTube explaining how redox flow batteries work. The Readers are particularly advised to watch three of them, preferably in the order as listed below:

• Video 1 by Peter Allen, a scientist and blogger;



• Video 2 from a YouTube Channel of Seeker;



• Video 3 by XRG Technologies.



We have paid much attention to the vanadium redox flow battery because it may play an important role in future energy storage systems. A few VRF batteries of single-digit MWh capacity have already been deployed in energystorage facilities (see the Wikipedia list of energy storage projects). One such battery in China is expected to become the record-keeper in 2020, with capacity of 800 GWh, beating the aforementioned 648 GWh of the sulfursodium complex in Abu Dhabi.

China is known to posses the world-richest resources of vanadium ore, which may the reasons for the technology chosen. But vanadium is not inexpensive, so that other countries may show less enthusiasm for the VRF method. The good news, however, is that the VRF battery is only one of the several existing options of redox-flow technology (for review, see, e.g., this article by Dapeng Zhang *et al.*. In some of those materials considerably cheaper than vanadium compounds can be used – e.g., in all-iron redox flow batteries described below. The reasons why we have chosen to discuss the VRF technology in closer detail are the following: one, because it was the earliest technology implemented in practical energy storage installations, and two, because the "chemistry" underlying the processes involved is arguably the simplest.

Typically, the energy density of redox flow batteries is comparable to that of lead-acid ones. This may not sound particularly impressive today – but it should be kept in mind that redox flow batteries are primarily intended for use not in cars, but in large-scale stationary installations, where maximization of gravimetric/volumetric energy density is not the main goal so it does not necessarily have to be seen as a disadvantage. However, their unquestionable advantages should be emphasized namely, a typical lifetime of 10-20 thousand charge/discharge cycles, the non-flammability of liquids used as anolytes and catholytes, as well as the ease of their recycling if the anolyte gets contaminated with ions from the catholyte, or vice versa.

All-iron redox flow batteries – another candidate for the "Pride of Oregon" title. As noted, the vanadium-based redox flow technology is highly reliable – but the fact that vanadium is neither cheap nor widely available may be a major obstacle to the widespread use of this energy storage method. That is why the news that there exists a similar technology based on iron and a few other inexpensive elements – and it's already in the marketable phase – sounds very optimistic. Namely, it has been pioneered by ESS^{INC}, a start-up company located at Wilsonville, Oregon (a few m iles south of Portland). The product already offered by ESS is a unit of 400 kWh capacity, housed in a standard 40-ft shipping container. Its guaranteed service life is 20 years, or over 20,000 charge / discharge cycles. The reliability of this guarantee is confirmed by Munich RE, a world-renowned provider of reinsurance, primary insurance and insurance-related risk solutions. ESS has won a vote of confidence from prestigious and well-heeled backers, the Japanese SoftBank's SB Energy and Bill Gates-funded Breakthrough Energy Ventures, who awarded it with a $30 million grant for further development of the innovative technology. No doubt, the Oregon-based ESS has a potential of becoming a major player in the rapidly developing industrial sector of energy storage!

In the Web-published material ESS has not revealed too many details of the chemistry involved in the operation of its iron-based redox flow battery. Which is understandable, because at an early stage of new technology development companies do not want to reveal the exact "recipes" used in it, for the sake of protecting their fresh patents. However, it is likely that, in general, the technology used by ESS closely resembles the process described in this review article and illustrated in Figure 2 in that report – with the guess that ESS has additionally introduced certain proprietary improvements to this process.

Molten metal batteries. A highly interesting innovative battery type was invented at MIT by professor Donald R. Sadoway. There is some similarity between the Sadoway's battery and sodium-sulphur battery, which we discussed earlier. Namely, in both these

types the electrodes are molten substances. However, in the sodium-sulfur cell the electrolyte is a solid material, which allows the electrodes to be placed in two concentric cylinders separated by an electrolyte also in the shape of a cylinder. But in the molten-metal battery the electrolyte – molten salt – is also in a liquid form! So how can it effectively separate the electrodes? The answer is simple it is possible due to different geometry and the aid of gravity. In all prototypes the Sadoway's team tested the anode was always a light metal: calcium (Ca), magnesium (Mg), or lithium (Li), and the cathode was a heavy metal: antimony (Sb), lead-antimony alloy (Pb+Sb), or lead-bismuth alloy (Pb+Bi). As the electrolyte, the team used mixtures of lithium halides (LiF, LiCl, LiI) mixed in various proportions. The point is that such a molten salt combination is immiscible (= it does not mix) either with the molten anode metal or with the molten cathode metal. In addition, the density of the molten electrolyte is higher than that of anode metal and lower than that of cathode metal. So, if the three components are poured into a vessel, gravity automatically separates them – the cathode remains at the bottom, the anode flows up to the top, and a layer of molten electrolyte stays in between. In the discharge phase an anode atom gives away an electron (if it's a Li anode), or two electron (for Ca or Mg anodes). The electrons flow to the cathode across an external circuit, while the Li^+, Ca^{2+} or Mg^{2+} ions pass through the electrolyte layer to the cathode, where they "recover" their electrons and, again becoming neutral atoms, they get alloyed with the cathode metal. In the recharging phase electron is run through the cell in the opposite direction: electrons are forced to travel from the cathode to the anode and the cathode gets "di-alloyed", the Li, Ca, or Mg atoms travel back as ions through the electrolyte to their "parent" anode. To learn more about the liquid metal batteries it's worth to watch Prof. Sadoway's lecture on YouTube. Also, It should be noted that a startup company Ambri has been created for marketing the batteries. The product offered and its technical parametersa are presented in the Ambri's Web page. The basic unit is a self-contained "module" of a weight of 56 lbs. 60 such modules are arranged into a "tray", and 10 "trays" are packed in a standard 10-foot container, forming a unit of a 400 kWh capacity. As can be readily calculated, its gravimetric capacity is about 26 Wh/kg. This may not seem particularly impressive, considering that "old" lead-acid batteries exhibit a similar capacity. On the other hand, the important advantages of these batteries should be emphasized. First of all, they are intended for stationary energy storage, in which weight is not a parameter of primary importance. However, the undoubted advantages of Sadoway's batteries are: essentially maintenance-free operation, life expectancy of at least 25 years, or 20,000 charge / discharge cycles and last but not least, definitely lower price for the storage capacity (a parameter usually expressed in the units of Dollars per kWh) than that of lithium-ion batteries. Therefore, this moltenmetal technology certainly has the potential of playing a meaningful role infuture energy storage systems.

Zinc-air batteries. Non-rechargeable zinc-air batteries have been known for a long time. They achieve very high energy density, reaching up to 500 Wh / kg i.e. almost twice the record high density of 265 Wh/kg in lithiumion cells used in Tesla cars. Therefore, such disposable batteries are readily used, for example, in miniature electronic devices such as hearing aids (as the author knows from his own experience). The high value of energy density is due to the fact that only one of the reagents namely, zinc is housed in the cell itself. The second, oxygen, is taken straight from the air when the cell supplies electric current. The batteries are made of readily available and inexpensive components and have a fairly simple structure: they consist of two electrodes, between which there is a paste of powdered zinc mixed with electrolyte (most often, an aqueous solution of potassium hydroxide KOH). One of the electrodes, the cathode, is made of porous metal, thanks to which oxygen from the air can reach the inside of the cell. When discussing other cells, we always started with the reactions taking place at the anode. But in the present case it is perhaps a better idea to start with what happens at the cathode. When the cell is functioning, electrons from the external circuit flow into the cathode, and oxygen flows in through the pores in the metal. In the " paste " in the vicinity of the cathode, there occurs a reaction between water and oxygen molecules with the participation of electrons:

$$H_2O + 2O_2 + 2e- \rightarrow 2OH^- \qquad (11.4.1.5.1)$$

that is, with the participation of one oxygen molecule, four OH^- ions commonly referred to as hydroxyl ions are formed. In fact, in the "paste" there are already abundant hydroxyl ions because in water solutions the molecules of KOH dissociate:

$$KOH \rightarrow K + +OH^- \backslash notag \qquad (11.4.1.5.2)$$

In turn, at the other side of the cell, the hydroxyl ions react with zinc metal, forming a complex ion called "zincate", and releasing two electrons to the anode:

$$Zn + 4OH^- \rightarrow Zn(OH)_2^- + 2e^- \qquad (11.4.1.5.3)$$

Away from the anode, the zincate ions decompose, forming a molecule of zinc oxide, a molecule of water, and two hydroxyl ions:

$$Zn(OH)_4^{2-} \tag{11.4.1.5.4}$$

\to ZnO + H_{2})+2OH^{-}\]

The same operation can be used for chemical equations as for algebraic equations. So, let's now add up all components at the left sides of Eqs. 11.1, 11.2 and 11.3, and than do the same all components at their right sides, which will give us the following "superequation":

$$H_2O + \frac{1}{2}O_2 + 2e^- + Zn + 4OH^- + Zn\,(OH)_4^{2-}$$
$$\rightarrow 2OH^- + Zn\,(OH)^{2-} + 2e^- + ZnO + H_2O + 2OH^-$$

Now, as we do with algebraic equations, we can cancel out all terms that appear at both sides. It leaves us with the following compact overall reaction equation:

$$Zn + \frac{1}{2}O_2 \rightarrow ZnO \tag{11.4.1.5.5}$$

so we can say using a somewhat theatrical and anthropomorphic mannerism: the hydroxyl and zincate ions only play the role of intermediaries who enable the react ion to occur – but having rendered their service, they disappear from the stage.

So, from the practical point of view, as the result of discharge in the cell remains "spent paste" no longer with zinc, but with zinc oxide instead. But in such a battery as the one described, processes cannot be reversed by simple re-charging (i.e. by passing current through the cell in the opposite direction). Until recently, the only way was "mechanical recharging" namely, replacing the spent "paste" with fresh one. Due to the simple construction of the cell, it was relatively easy. But even those "mechanically rechargeable" batteries are not suitable to be used as energy storage units 'tied" to the power grid: such devices must be able to switch from the recharging mode to power-supplying mode in milliseconds.

However, it seems that quite recently the situation has changed in favor. Due to the considerable advantages of zinc-air batteries – their high storage capacity and the fact that they are made of inexpensive, widely available and "environmentally friendly" materials, efforts continued to create a version that would be fully rechargeable in the conventional meaning. The Canadian company *Zinc8 Energy Storage* has just announced that it has succeeded in this matter. The case is completely new and so far (March 2020) the details of their technological solution have not been made public. However, they were certainly revealed to the New York authorities. This state has an ambitious plan of achieving full decarbonization of electricity generation by 2040 and is therefore looking for energy storage technology that would warranty this goal. This new Canadian technology was apparently found to be highly promising. The state has began cooperating with the Zinc8 Company and will finance the construction of its first prototype installation in New York City. It will be worth following this project because if these new-generation zinc-air batteries indeed appear to be cheap and reliable, they might greatly facilitate the creation of storage systems at a global scale

11.4.1.5: Batteries for Industry- and Grid-scale Storage is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.2: Electricity → potential energy → back to electricity (gravitational energy storage and compressed air energy storage).

Harnessing the potential energy due to gravity. Known forms of mechanical potential energy are those associated with gravity, or those associated with elastic deformation of bodies. Both are used for energy storage, but the former on a much larger scale in pumped storage power plants. We've discussed these power plants extensively in Chapter 5. So let's just briefly remind you that these power plants are the largest energy storage facilities in existence today. The largest in the USA (and in the world) in Bath County, Virginia can generate 2.4 GW of power for 10 hours, i.e. it can store 24 GWh of energy. The Raccoon Mountain power plant can generate 1.65 GW, but has an even larger capacity, 36 GWh. In 2018, as reported by the U.S. Energy Information Administration, the total capacity of pumped storage power plants in the USA was 22.9 GW. The total capacity is not given, but knowing that typically such power plants have a water supply for at least 10 hours it can be estimated that this capacity is at least 230 GWh. It is planned to build several new pumped storage power plants in the USA, including a 393 MW project in Oregon in the Klamath Falls area and the 1.2 GW Goldendale Energy Storage Project in Washington, not far from the John Day Dam at Columbia River.

Potential energy can be stored not only by raising water, one can do that as well by raising solid objects. Such a method has been used for centuries, e.g., for storing energy needed for the operation of clocks at church towers. Now in the Web one can find ideas of using the same method at a much larger scale. Suppose, for instance, that there is a shaft in a now defunct mine – say, 500 meters deep. It is conceivable to build a dedicated elevator and load it with 1000 tons of rocks, and use its potential energy to drive an electric power generator. The total energy released by lowering such weight would be:

$$E = 10^6\ kg \times 500\ m \times 9.81\ m/s2 = 4905\ MJ = 1.363\ MWh \tag{11.4.2.1}$$

Such a device, operating, say, for 4 hours, would deliver 341 kW of power – good for satisfying the evening-hours needs of a community of some 100 homes. Another idea – if there is a village not far from a hill with its summit 300 m above the village center (similar to the Dimple Hill or Vineyard Mountain in the Corvallis area), then it is possible to build a ramp with railway tracks and use two freight cars, each loades with 1000 tons of rocks and pulled uphill by cables, to generate a similar power as in the preceding example. One can find quite a few such ideas in the Web – but, so far, no reports of those ideas being implemented. But, who knows? – Maybe someday somewhere in the world the local community will wish to implement such a project?

Harnessing the potential energy of compressed bodies – of air, in particular. As mentioned above, potential mechanical energy is also contained in elastically deformed bodies, for example, helical or spiral springs. And indeed, until recently, they were very widely used for energy storage but only in small devices such as watches, turntables, portable film cameras or toy cars they all required "winding up" (which either stretches the spring, or – more often – compresses it). However, due to technological progress today we can find these things mainly in attics or in the museum.

It should be noted that springs are very efficient energy storing devices – those made of high quality steel can return over 99% of the input energy. This can be confirmed by a very simple demonstration – namely, the observation of oscillations of a weight suspended on a spring, as, e.g., in this YouTube clip. Such a weight can keep oscillating for long minutes, the energy dissipation is indeed very slow. Now, please take a look at the list of the largest energy storage installations in the world is given in this Wikipedia site. Note that if the installation capacity in GWh is taken into account, the devices using the CAES technology (Compressed Air Energy Storage) come only second after pumped storage power plants.

Compressed air behaves much like a compressed spring, which can be easily checked by squeezing a portion of air in a pump for inflating bicycle tires with the outlet blocked. Yet, the properties of a gas are not exactly the same as those of an elastic body. The point is that when gas is rapidly compressed, some of the input energy changes into thermal energy – in other words, it is used for heating up the gas. In theory, if the compressed gas were then kept in an adiabatic tank, i.e., a perfectly thermally insulated from the surroundings, it would be possible to recover 100% of the energy input. In theory, let us stress. But if the tank is not insulated and heat escapes from it, then even in theory much less energy is recovered than that inserted. A more detailed thermodynamic analysis of CAES can be found in this article extensively discussing various aspects of this technology.

In practice, devices for storing hundreds or even thousands of MWh are needed. In a tank withstanding the pressure of 100 bars, the energy stored per one cubic meter is 4 5 kWh. Then, for storing 1 GWh, reservoir volume as large as 200 000 m^3 would be needed. Making such a sturdy structure of steel is conceivable – but, probably, the cost would be very high. A much better idea is to use

existing underground reservoirs such as salt caverns left out in existing salt bed deposits after mining operations. The salt is mined by drilling a well down into the formation, and pumping water through the completed well to dissolve the salt which returns to the surface as brine. Many such currently existing caverns are used as reservoirs for storing natural gas. They can be used as well as tanks for large CAES installation. In fact, there are two such major objects currently operating. One, located in Huntorf, Germany, capable of storing 1.16 GWh, was commissioned in 1978 and it has been operating since then. It uses two caverns of a total volume of 310,000 m^3. The other is the CAES installation in McIntosh, Alabama, USA, operating since 1991, with 560,000 m^3 caverns and the storage capacity of 2.64 GWh.

In order to improve the recovery efficiency the Huntorf and the McIntosh facilities both use the so-called “diabatic” operating regime, in which the turbine operation is aided by the combustion of natural gas. The energy recovery efficiency is still not spectacular: 42% in Huntorf and 54% in McIntosh.

However, the operation of both of these installations is apparently costeffective since both have been kept in operation for years. One argument is that in such a “diabatic cycle” the cost of natural gas needed to obtain an energy unit is four times lower than for a gas-only power plant. It should also be taken into account that if the utility system (no matter whether it uses renewable energy or conventionally produced energy) is able to generate surplus energy but this capacity is not used because there are no surplus storage devices then the efficiency is ... zero percent! 42 % or 54 % are not actually spectacular figures, but they are always better than 0 %.

An interesting extensive discussion about the sense of using CAES and its future prospects can be found in this series of articles published by POWER magazine. Also very much worth recommended is this quite recent Web document devoted to these matters.

Geothermally Coupled Compressed Air Energy Storage. A very interesting concept seems to be the combination of CAES technology with the use of geothermal heat. If the underground caverns for storage will be located in geothermal bed rocks, then the problem of lowered storage efficiency due to heat loss at compression can be significantly eased, and if the rocks are hot enough, the storage efficiency may even exceed 100%. Particularly favorable conditions for the use of such hybrid technology exist in California and the Pacific Northwest (for example, by using hot basalts in the Columbia River Gorge area). To date, no such project has been implemented, but there are serious studies that can motivate for creating such hybrid storage reservoirs – interested readers may be advised to consult reports prepared by Pacific Northwest National Laboratory, by the U.S. Department of Energy, or Section 7 in this report compiled by Idaho National Laboratory.

11.4.2: Electricity → potential energy → back to electricity (gravitational energy storage and compressed air energy storage). is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.3: Electricity → Kinetic Energy → Electricity Storage Scheme

Another form of mechanical energy that can be stored is that associated with the body motion – i.e., the kinetic energy. For obvious reasons, rectilinear motion cannot be taken into account – but rotary motion is well suited for this purpose. Heavy spinning wheels began to be used as a kind of "energy buffers" over 200 years ago and the term "flywheel" was already created. Well, what flywheels were used for was not yet energy storage in the exact sense in which we understand it today. Their role was initially to "smooth" the motion of steam engines, which without it exhibited unpleasant jerks.



Figure 11.4.3.1: A huge flywheel in a British steam engine from the beginning of the 19th Century (from Wikimedia Commons).

For the first time, the flywheel was used as a "reusable battery" i.e. a system that could be charged with energy, keep it stored for some time and finally return it in Switzerland around 1950 in the so-called "gyrobuses". A 1500 kg flywheel was installed in the bus. An electric motor accelerated the wheel to 3000 revolutions per minute. Then the same motor acted as a generator and the current from it propelled the vehicle. From a single "charge" the gyrobus could make about 4 miles. But at each stop along the bus' route , the wheel was recharged for a minute or two, so that the vehicle could operate a line even as long as 20 miles. The gyrobuses offered a nice low-noise and zero-emission means of urban public transportation, but the world in the 1950s was flooded with cheap oil, so they eventually succumbed to the competition with diesel buses.

The kinetic energy stored in a spinning flywheel is:

$$K = \frac{4\pi^2 f^2 I}{2}$$

where I is the wheel's moment of inertia, and f is the number of revolutions per second. Most often, however, the speed of rotation is given not in revolutions per second, but in revolutions per minute (RPM). If a wheel spins too fast, it may be torn apart by centrifugal forces. The maximum speed at which a given wheel can spin depends on its radius, shape and the strength of material from which it is made. The energy density i.e., the maximum energy that can be stored per unit weight for steel flywheels is higher than for lead batteries. And when using materials with extremely high tensile strength (such as, e.g., carbon fiber composites), the energy density of such flywheels may reach some 50% of that of the best lithium-ion cells manufactured today (written in April 2020). However, there are problems that hinder the widespread use of such energy storage. Until recently, only mechanical bearings were available, which always lose energy due to friction. Even when using the highest quality bearings, the "self-discharging" rate of such flywheels could amount to as much as 20-50% energy loss per hour.

The technology of making "flywheel batteries" capable of storing energy for many hours or even days have only recently emerged, thanks to the invention of zero-friction bearings based on the phenomenon of magnetic levitation. As is well known, magnets facing each other with identical poles experience a repulsive force. The repulsion can be strong enough to overcome the force of gravity and one magnet can float above the other without touching it. So, there is essentially no friction, with the exception of that caused by surrounding air (which can be eliminated by putting the spinning object in evacuated space).

A very good demonstration of the phenomenon of magnetic levitation is the educational toy known as the “levitating spinning top”. There are a number of short videos on YouTube, here we provide links to such (one from the US, the other from China) in which hobbyists make such spinning tops. Magnetic levitation can lift not only spinning tops, but also bricks and even flywheels with a mass of hundreds of kilograms. The latter generally are enclosed in an evacuated container in order to eliminate air friction.

Magnetic levitation bearings poorly absorb shocks and changes in the orientation of the spin axis. Therefore, flywheels with such bearings are not suitable for use in vehicles, but they can perform very well in stationary systems. Another limitation for bearings is weight certain limits cannot be exceeded. Therefore, flywheel manufacturers concentrate on the production of modular systems that can be scaled up that is, a number of individual modules can be combined and form one large assembly.

For example, the Californian company Amber Kinetics produces M-32 modules, each of which can store 32 kWh of energy and provide an output of 8 kW. According to the M-32 data sheet, the “self-discharge” power loss of such a module is less than 100 W, i.e. after 24 hours the energy loss will be less than 2.4 kWh, or below 7.5 % of the initially stored energy. 125 M-32 modules will form a group with an output of 1 MW.

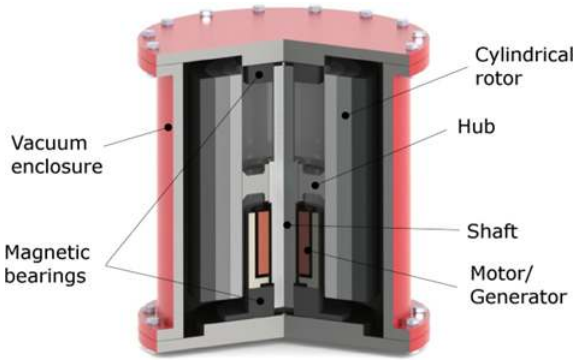


Figure 11.4.3.1: A graph showing the anatomy of a modern flywheel, similar to those produced by Amber Kinetics (from Wikimedia Commons).

The German company STORNETIC produces modules called Enwheel 22, capable of storing 3.6 kWh and providing an output of 22 kW. These modules can be grouped in standard shipping containers. 28 modules packed into a standard 40-foot container deliver 616 kW of output power.

As follows from the above, flywheel based systems are unlikely to be suitable for long-term energy storage. But this is not their main task an important role they play is stabilization of the power grid specifically, frequency stabilization. If the grid in certain area is operating at the limit of its performance and the load suddenly increases even more, then the generators in power plants stop keeping up, their rotational speed drops and so does the frequency of the alternating current they generate. This is a very dangerous phenomenon that can lead to a collapse of the entire power grid in a smaller or larger area. The remedy is a sudden injection of additional power flywheel based systems are great for this, they can react in a split second. The boost of extra power they deliver can prevent a dangerous situation.

However, there is a special category of vehicles in which flywheels are well suited for energy storage namely, space ships. NASA is absolutely serious about this and it has created a prototype that can be viewed in this YouTube video.



11.4.3: Electricity → Kinetic Energy → Electricity Storage Scheme is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

11.4.4: Electricity → Thermal Energy → Electricity Storage Scheme

If we look at the list of the largest energy storage devices in Wikipedia, it turns out that if we take into account the amount of energy stored in a single installation – then, after pumped storage hydroelectric plants and after installations using compressed air, the third place is taken by devices storing energy in molten salts heated to high temperatures. We already talked about this method in Chapter 7 (see Section 7.3.1) when discussing concentrated solar power (CSP) plants. Solar power is not converted in them directly into electricity, as in photovoltaic (PV) panels, but into thermal energy, which in turn is converted into electricity by a heat engine. The advantage of such a scenario is that the conversion to electricity doesn't need to happen right away – as in PV facilities – but some part of the thermal energy collected may be stored and used later, which allows the CSP plants to provide electric power for several hours after the sunset.

The molten salt technology has been developed in the late 1990-s in Sandia National Laboratory. Soon afterwards, it began to be used successfully in CSP power plants. However, it has only recently been realized that this technology can also be used to store energy from other sources – for example, power surpluses generated on a windy day by wind turbines. And secondly, that the medium in which thermal energy is deposited does not necessarily have to be molten salt. In this context, two exciting recent news come from Australia, Germany and the United Kingdom.

11.4.4.1: Molten Silicon

11.4.4.2: Heated Volcanic Rocks

11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage

11.4.4.4: Small and Medium Scale Thermal Energy Storage

11.4.4: Electricity → Thermal Energy → Electricity Storage Scheme is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.4.1: Molten Silicon

In Australia, a startup company CCT Energy Storage has created an unusual “thermal battery” using not molten salt, but ... molten silicon. This is a remarkable achievement, given that the melting point of silicon is as high as 1410 degrees Celsius (or 1683 K), almost twice as high as the highest temperature used in the molten salt technology. In addition, silicon’s unique feature is its unusually high latent heat of fusion (denoted as L_f , it’s the energy needed to melt a material after heating it up to its melting point; its value is usually given in Joules per mole, or Joules per kg). For silicon, L_f = 1787 kJ/kg, or 0.496 kWh/kg. The density of silicon at its melting temperature is about 2300 kg/m^3 – taken together, it means that for melting one cubic meter of silicon the energy of about 1.2 MWh is needed – and, of course, the same amount of energy can be recovered on the transition from the molten phase back to the solid phase. And it should be stressed that during the entire process of “extracting” heat from the “battery” the temperature will remain at the constant level of 1683 K.

If this silicon “thermal battery” is to be used to store electricity, then “charging” it will be a simple operation: resistance heaters convert electricity into thermal energy with 100% efficiency. But recovering electricity will be more complicated because a heat engine driving an electric generator must be used for this. A heat engine, as discussed in Chapter 3 (see Section 3.4.6 and the Equation. 3.15), will never convert all heat into mechanical energy. But the high temperature of silicon is a favorable factor here. It can be expected that at T_h as high as 1787 K, the efficiency of the heat engine can be at least 50%. Thus, one can realistically expect that from such “thermal battery” about 0.6 MWh could be recovered from every cubic meter of silicon.

Let’s try to compare this with the performance of a molten salt installation. In contrast to the silicon thermal battery, which uses the phase transformation process, i.e. the transition from solid to liquid state taking place at a constant temperature, in the molten salt technology only the liquid phase is involved and the temperature changes it increases when heat is added and decreases when it is extraced. On the Stanford University website the volumetric heat capacity of the salt mixture used in this method is given as 3770 kJ/m^3 K. Assuming that the temperature decreases by 200 K, this gives the total amount of stored heat energy as 754 MJ/m^3, or 0.209 MWh/m^3. At the same time, the efficiency of converting thermal energy into electricity in this temperature range, which is used in the molten salt method, can be at most 30%, which effectively gives about 0.06 MWh/m^3 so using molten silicon you can get the same storage capacity with a **ten times smaller volume** of the storage medium! Let us add that the needed silicon does not have to be highly purified, i.e. it is easily available and inexpensive. So one can expect that these silicon “thermal batteries” will find a number of interesting practical applications.

11.4.4.1: Molten Silicon is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.4.2: Heated Volcanic Rocks

In Germany, the Siemens-Gamesa (S-G) Company has built and launched a prototype energy storage installation using an innovative technology. The method is essentially the same as in molten salt technology but volcanic rock blocks are used instead of salt. The rocks are heated up to a temperature as high as 800 degrees Celsius by electric heaters using surplus power generated by wind turbines. The energy is later recovered by passing the heat contained in the rocks to superheated steam which is used next to propel steam turbines and electric generators in the same way as in a conventional coalfired power plant. Undoubtedly, the advantage of this new method is its simplicity compared to the technology of molten salt. The average specific heat and density of the molten "solar salt" in the relevant temperature range can be taken, respectively, as 1.54 kJ/kg K and 1750 kg/m^3. The S-G report does not specifically identify what kind of volcanic rocks are being used, but typical values for them are, respectively, 1.1 kJ/kg K (see Page 72 of the report), and 2400 kg/m^3, so that the volumetric heat capacities of the salt and the rocks seem to be quite similar. Certainly, however, the construction of a thermally insulated "reservoir" for these rock pieces will be a much easier undertaking than making tanks for molten salts: there is no risk of leaks, there is no risk of corrosion and the price of that solid "storage medium" is certainly much lower than that of "solar salt". Let's estimate how much thermal energy U can be stored stored in this S-G prototypical installation. Let's assume that the temperature range used will be 300 K, from 500 to 800 degrees Celsius:

$$U = 1.1\ kJ/kg \bullet K \times 1000\ kg/tonne \times 1000\ tonne \times 300\ K =$$

$$= 3.3 \times \times 10^8\ kJ = 330\ GJ = 91.7\ MWh$$

It should be kept in mind, however, that the efficiency of converting thermal energy to electreic power using steam turbines is only 30%, so the "recoverable" energy will be only 30 MWh – or 10 MW of power for three hours. Not a spectacular result, perhaps, but let's remember that the S-G installation we are talking about is only a prototype. The 1000 tons of rocks occupy less than 500 cubic meters which is about the volume of an American single family house – – but not of a very large family. From a really serious grid-tied installation, we would expect, say, 1 GW of power rather than 10 MW – this means multiplying everything by 100, i.e., 50,000 cubic meters. Which should not frighten us, there are buildings on the OSU campus whose volume is much larger! So building such a thermal reservoir for 1 GW installation will certainly not be a major engineering problem today. In addition, such a reservoir can be hidden partially or even completely underground. The relatively low efficiency of 30% of this type of energy storage remains a problem (this also applies to "molten salt" technology). But if you use such installations to store excess energy from wind turbines and solar farms, the recovery efficiency of 30% is still better than 0%, which we would have had if this energy was wasted. Remember that a general usage system must generate as much power as consumers need. It couldn't be less! if too little power is generated, some of the users simply have to be disconnected from the grid. But if windmills and solar farms generate too much power then some of them have to be disconnect from the grid! which is just a waste, isn't it?

There is another important aspect of this new S-G technology. Namely, it can provide a "second youth" to coal-fired power plants that are currently to be decommissioned in accordance with the decarbonisation program. But all you have to do is eliminate their coal boilers and replace them with an energy store. All other machines needed by the new technology turbines, generators are the same as in a traditional power plant. This may be a particularly attractive option, especially for Germany, which are planning to shut down coal power plants in the near future, although they are still in excellent condition. But for many other countries such an option is perhaps no less attractive.

11.4.4.2: Heated Volcanic Rocks is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage

In the United Kingdom, researcher developed two entirely new technologies of thermal energy storage.

One of these ideas is a radical remedy for the low efficiency of energy recovery in molten salt technology and the technology developed by SiemensGamesa (SG) described in the previous section. The conversion of thermal energy into electricity in both methods is the same as in conventional coalfired power plants – which is shown schematically in Figure X. As discussed in Chapter 3, from a portion of heat Q using an ideal heat engine using the famous Carnot Equation one can get the maximum mechanical energy W[1] amounting to:

$$W = \left(1 - \frac{T_c}{T_h}\right) \times Q = \epsilon_i \times Q$$

where T_h is the absolute temperature of the "hot source", and T_c is the absolute temperature of the "heat sink". The temperature of the latter in the case of using steam turbines is roughly 100°C = 373 K. The highest T_h in the S-G method is, as mentioned, 800°C = 1073 K, and then, as heat is drawn from the hot reservoir, it drops to 400°C = 673 K – which corresponds, respectively, to E equal 0.65, and 0.45. Yet, as discussed in Chapter 3, for real thermal engines a more realistic is the modified Carnot formula:

$$W = \left(1 - \sqrt{\frac{T_c}{T_h}}\right) \times Q = \epsilon_r \times Q$$

from which, for the two temperatures –the initial and the final – we obtain, respectively, the E_r values of 0.41 and 0.26, with an average of 0.34

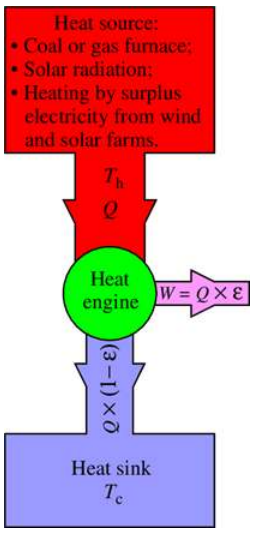


Figure 11.4.4.3.1: A general scheme of a heat engine, which can also be an energy storage device if the "hot source" is a reservoir, "charged" with thermal energy which is produced from surplus energy from wind or solar farms.

In the molten salt storage method the figures are even lower, because the the highest value of T_h is 550°C = 823 K, and if it drops to 250°C = 523 K in the course of the heat extraction process, the average value of E_r is only about 0.24. In other words, in these methods from 2/ to 3/ of the energy storedis deposited to the heat sink – which simply means "lost".

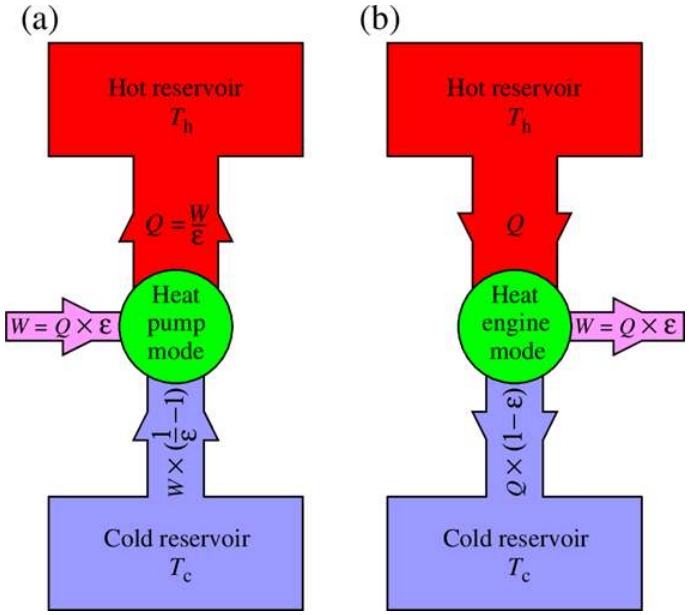


Figure 11.4.4.3.1: The idea of the pumped heat energy storage system. Plot (a) is essentially a heatpump system in which mechanical work (from an electric motor) drives a heat pump, i.e., a heat engine acting "in revers". Plot (b): the thermal energy that has been transferred by the heat pump to a reservoir playing the role of a "hot source", now is used for operating a heat engine. In an ideal case such a storage system makes it possible to recover 100% of the energy used to drive the heat pump.

The idea of the new British PHES method is shown in the Fig. 11.16. The point is that the electrical energy to be stored is not deposited directly to the "hot source" – instead, a heat pump is used. A heat pump, as was discussed in Chapter 8, is a thermal engine working "in reverse". In normal operating mode, a thermal engine withdraws a portion of heat Q from the "hot source", converts E Q to mechanical work W , and deposits (1 E) Q in the heat sink. In the pump mode (Fig. 11.16 (a)) , things get reversed – the input of a mechanical work portion W_{in} into the pump causes that it "withdraws" from the sink (now called the "cold reservoir") a portion of heat equal:

$$Q_{\text{sink}\,\rightarrow\,\text{pump}} = W_{\text{in}} \times \left(\frac{1}{\epsilon} - 1\right)$$

In the pump, this portion of thermal energy adds up to the mechanical energy input and this sum:

$$W_{\text{in}} + Q_{\text{sink}\,\rightarrow\,\text{pump}} = W_{\text{in}} + W_{\text{in}} \times \left(\frac{1}{\epsilon} - 1\right) = \frac{W_{\text{in}}}{\epsilon}$$

is now transferred "up" to the hot source (now called the "hot reservoir").

Now, in order to withdraw energy, the pump is switched over to the "heat engine" mode (Figf. 11.16 (b)). If the same amount of heat $Q = W_{\text{in}}/E$ that was stored in the hot reservoir is now withdrawn from it and passed to the heat engine, the work output yielded by the engine will be:

$$W_{\text{out}} = Q \times \epsilon = \frac{W_{\text{in}}}{\epsilon} \times \epsilon = W_{\text{in}}$$

i.e., 100% of the stored energy is recovered! But only if ideal heat pumps and heat engines are employed – which, of course, do not exist. However, it should be expected that a PHES system built even from existing materials and components will provide much higher energy recovery efficiency than the S-G technology and molten salt technology. The whole idea outlined above is so simple that it must have been known for a long time the only problem was that it was not really known how to implement it and create a working installation. And finally, the British researchers managed to do that! Here one can find a more detailed description of the design of their prototype and of the physical processes used in it. This is another article worth reading. It turns out that the sophisticated PHES technology offers an efficiency of 60-65%, which is indeed a substantial gain in comparison with the other methods. Also interesting are the economic aspects of PHES technology, its cost-effectiveness compared to other methods – such an analysis is presented in this Sciencedirect report.

The other **highly interesting new British idea** is the thermal storage technology using – paradoxically! – not a very hot medium to store energy, but a **very cold one**! Yes, very cold, because we are talking of liquid air, the temperature of which is -194°, or 79 K. How can energy be stored in a cold medium? It may come as a surprise, but the answer is pretty simple. Consider the basic theory of heat engines. What does a heat engine need to work? Well, a "hot source" and a "heat sink", yes? And in order to get a decent efficiency *E*, what is needed? A temperature difference between the two, of course. So, why don't we treat the liquid air as the "heat sink", and the ambient environment, with its temperature of 20 °C, or 293 K, as the "hot source"? Then, we get the realistic efficiency of a heat engine:

$$\epsilon = 1 - \sqrt{\frac{T_c}{T_h}} = 1 - \sqrt{\frac{79\ \mathrm{K}}{293\ \mathrm{K}}} = 0.48 \qquad (11.4.4.3.1)$$

The efficiency of 48% is definitely not a bad result!

More technical details about the liquid air energy storage (LAES), a.k.a. "cryogenic" energy storage (CES), are given in the Web page of Highview Power the company who has pioneered this unusual technology – and their YouTube video.

The technology of liquefying air is well established and is widely used by many industries. The process of storing energy is thus nothing else than running a liquefying machinery and saving the output in tanks – there are so-called "Dewar vessels" in which liquid air can be stored for prolonged periods with minimal losses. However, the overall efficiency will be lower than 48% because our estimate did not include the energy initially needed for liquefying the air. On the other hand, there are possible ways of increasing the efficiency: above, we used the "ambient temperature" of 20°C = 293 K as T_h, but there are many industrial facilities that "dump" large amounts of "waste heat". So, if a liquid air storage installation could use such heat to augment its "hot source" by, say, 100 K, the efficiency would increase from 48% to 55%. The aim of the Highview Power engineers, as Javier Cadava, the COE of Highview Power, declares in this article, is to reach 70-75% of overall efficiency (or "roundtrip efficiency", as Cadava calls it).

There are certainly many advantages of LAES – to list a few, the medium needed for implementing this method is available everywhere in the world in unlimited quantities; it is patently non-toxic; and it is 100% recyclable without leaving any nasty byproducts. So, let's keep our finger crossed and let's wish the Highview Power all possible successes in further development of LAES and in making it one of the leading energy storage technologies on a global scale.

1The output mechanical energy from a thermal engine still needs to be converted to electrical energy, but the efficiency of modern generators used for this conversion is very high and can be taken with a good approximation as 100%.

11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.4.4: Small and Medium Scale Thermal Energy Storage

In previous sections, we focused on large-scale heat storage devices. However, there is a whole variety of small scale heat storage devices. For example, residential roof-mounted solar water heaters are popular in many countries with high sunlight.



Figure 11.4.4.4.1: A rooftop solar water heater at a building on the Mo'orea Island in the French Polynesia (photograph taken by the author).

We will not discuss in detail the existing methods of small and intermediate scale thermal energy storage, but a good review of such technologies can be found in an article by Ioan Sarbu and C. Sebarchievici available at a research- gate.net Web page. After clicking on the link, one will see only the abstract but in the upper right corner of the screen there is an icon "Download full-text PDF", after clicking on which the whole text will appear on the screen.

11.4.4.4: Small and Medium Scale Thermal Energy Storage is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

SECTION OVERVIEW

11.4.5: Electricity → Chemical Energy

In this section, we want to focus on technologies in which energy "input" is converted into fuel that can be stored and someday converted back to energy in a zero-emission process – or used in industry to replace the CO_2-spewing operations with carbon-neutral ones. The only such fuel is hydrogen, the most abundant element in the universe. The combustion reaction produces only water vapor, H_2O, which in no way contaminates the air.

One may argue that water vapor is a greenhouse gas, just like CO_2. It's true, but man-made carbon dioxide accumulates in the atmosphere and oceans while Mother Nature always makes it sure that there is not too much water vapor in the air: if there is, the excess is simply removed in the form of precipitation.

However, there is one major problem: namely, hydrogen does not occur in nature in a "ready to use" form. It has yet to be produced at the expense of energy. Then it can be stored (which is not so simple, however) and finally the energy input can be recovered. That is why we are talking about the possibility of storing energy in the form of hydrogen. However, at all stages: energy conversion into hydrogen, hydrogen storage and finally energy recovery, there are serious challenges that have yet to be faced.

The idea that hydrogen will be the "fuel of the future" has been talked about for a long time. The author of this textbook remembers that he had read about it as long ago as in high school (from which he graduated in 1963). However, it turned out that these were mainly discussions between scholars. To some disappointment of the author, the industry was clearly not interested in these ideas. But in 2003, there seemed to be a clear shift in the situation: namely, President George W. Bush, announced in his "State of the Union Address" a program to open a policy which he called "Hydrogen Economy".

He said, among other things: *A simple chemical reaction between hydrogen and oxygen generates energy, which can be used to power a car producing only water, not exhaust fumes. With a new national commitment, our scientists and engineers will overcome obstacles to taking these cars from laboratory to showroom, so that the first car driven by a child born today could be powered by hydrogen, and pollution-free.*

Unfortunately, after this declaration nothing has really changed. Indeed, the industry after Bush's speech consumed significant amounts of hydrogen, but did the same before the speech. In 1975, hydrogen production in the US amounted to around 6 million tons, followed by a gradual increase to current production of around 10 million tons. But this hydrogen was not produced for energy storage or as a clean fuel for cars. It was entirely consumed by industry, mainly for the production of ammonia (NH_3), which is an important fertilizer, and in oil refineries, where it is used to remove sulfur from gasoline and diesel fuel. Moreover, almost all this hydrogen is called "gray", which means that it is produced from fossil fuel (natural gas) in a process called steam reforming of the gas. Natural gas, which is almost a pure methane CH_44, is mixed with water vapor, heated to a high temperature, and then the chemical reactions occurs:

$$CH_4 + H_2O \rightarrow CO + 3H_2$$

The carbon monoxide molecule created then reacts with another water molecule, yielding one more hydrogen molecule:

$$CO + H_2O \rightarrow H_2 + CO_2$$

The two reaction equations can be combined into a single general equation:

$$CH_4 + 2H_2O \rightarrow CO_2 + 4H_2$$

As follows from the last equation, four H_2 molecules with a total molecular weight of 8 and one CO_2 molecule with a molecular weight of 44 come out from the reaction. Translating to ordinary mass units, it means that as much as 44/8 = 5.5 tons of CO_2 is produced per a single ton of the hydrogen acquired.

But this is not the end, because these reactions are *endothermic*. That is, to split the interatomic bonds in the reactants, more energy is needed than that which is released when new bonds are formed in the reaction products. In other words, an extra boost of energy – in the form of heat – is needed for the reaction to occur. The easiest way to provide this additional heat is by burning an extra

portion of gas – which even further increases the CO_2 emissions. In practice, as much as seven tonnes of CO_2 are produced for acquiring a ton of hydrogen.

As can be readily calculated (see Chapter 3, Section 3.2.1), the methane input needed to produce 7 tons of CO_2 is 2.54 tons. The energy contained in this amount of methane (i.e., the heat of combustion, see Section 3.2) is 2.54 tons 55.5 GJ/ton = 141.2 GJ. And the heat of combustion of one ton of hydrogen is 141.8 GJ – essentially, the same. In view of this, it is immediately apparent that producing hydrogen in such a way in order to *store energy* would make little sense: burning methane directly would release the same amount of energy and emit the same amount of CO_2, without the need for going through the entire steam reforming operation.

The only solution acceptable from the viewpoint of future programs treating hydrogen as a "storable means" must be based on the production of the so-called "green hydrogen" in a completely zero-emission process. As will be discussed further on, water electrolysis, i.e. the decomposition of H_2O into H_2 and O_2 by electric current, is a process that does not produce CO_2 at all. This is a method known for over 200 years and it has been used in industry for well over 100 years. Now, it becomes a major player in the efforts of creating a "hydrogen economy" of global reach. Sometimes heat is easier available than electricity – therefore, a number of thermochemical H_2O decomposition methods have been proposed and tested, but none has yet been implemented at industrial scale. A novel method worth consideration is the pyrolysis of methane. It uses fossil fuel, yes, but in contrast to the CO_2-yielding steam reforming, in pyrolysis the CH_4 molecule is broken down into hydrogen and carbon **in solid form** (a substance similar to sooth, a.k.a. "carbon black") which can be disposed off using methods not harmful to the environment.

11.4.5.1: The evolution of the hydrogen economy policy in the 21st century

11.4.5.2: Methods of Green Hydrogen Production

11.4.5.3: The recovery of energy stored in the form of hydrogen

11.4.5.4: Hydrogen storage and transportation issues

11.4.5.5: Hydrogen as a "green" fuel suitable for sea- sonal storage

---

11.4.5: Electricity → Chemical Energy is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.5.1: The evolution of the hydrogen economy policy in the 21st century

After the speech of President G.W. Bush in 2004, virtually nothing happened over the next 10 years. Finally, things started moving forward in 2014. But not in the US – in Japan. Three years earlier, Japan was hit by a terrible disaster, the earthquake and tsunami on March 11, 2011, which caused, among other things, the catastrophe at the Fukushima-Daiichi nuclear power plant. As a result, Japan concluded that in this country, often haunted by earthquakes, the expansion of nuclear power as the backbone of a zero-emission economy is too risky. From debates it emerged that a viable alternative would be to focus efforts on a rapid transition of the nation to an economy in which hydrogen will play the role as a major energy carrier. A document titled “Basic hydrogen strategy” was created in 2014 – it has been updated several times from then. It is certainly worth getting acquainted with a version of this document in which this Japan’s economy conversion plan is outlined in a compact transparent form aided by graphical schemes. The Australians have realized that they have vast areas where the average population density is extremely low (well below 1 inhabitant per square mile), but where there is plenty of sunshine and wind. It is enough to supply the entire Australian population and industry with electricity – and even much, much more. Unfortunately, there is no way to export this surplus electricity to other nations, because all Australia is surrounded by ocean... And then the idea arose to convert this surplus power to fuel – what kind? – hydrogen! And to transport this hydrogen is across the seas to countries that will need it, to Japan first of all. How to transport it? We will talk about this a little later. Anyway, it’s all included in the official Australian state project entitled “National hydrogen strategy”. The latest (Novcember 2019) version of the governmental publication describing all aspects of the general plan can be found in this PDF document. It’s a big file, 99 pages long – but it reads very well, because of the “user-friendly text structuring” and plenty of graphic illustrations.

There are other countries that have recently established official plans of a strategy for developing hydrogen economy at the national level. There is a website with a list – in which,in addition to Japan and Australia – there are eight additional countries where such plans have begun to be implemented. But there should be more nations on this list than just 10. Definitely, it should include the Netherlands, which also intend to implement an ambitious strategy described in this official comprehensive government document. The country wants to use excess power from large off-shore windmill farms to produce green hydrogen. The Dutch neighbor, Belgium, has similarly ambitious plans. It also intends to use the power from off-shore windmill farms to produce green hydrogen and even wants to become a global leader in this field. Belgium has not yet published such an extensive official document on the planned strategy as the Netherland, but its plans are described in many more compact Web pages, for example in this Web article, or in another one. Two Scandinavian countries known for their contribution to the development of zero-emission technologies, Denmark and Norway, also have far-reaching intentions regarding hydrogen economy.

In the Web list of the 10 nations linked above, one country is conspicuously absent: the United States. Which can be understood, given that the US is a “superpower” in the production of natural gas, and therefore hydrogen as a fuel seems to only attract its marginal attention. But this is not true for all States – in California, hydrogen does attract much interest, because this State – as discussed earlier in this chapter – clearly aims to eliminate natural gas as a source of energy, striving to achieve full decarbonization of its economy by the year 2045.

11.4.5.1: The evolution of the hydrogen economy policy in the 21st century is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.5.2: Methods of Green Hydrogen Production

Electrolytic hydrogen. The only truly zero emission method of obtain ing hydrogen is the electrolysis of water – i.e., the decomposition of H_2O molecules into molecules of the constituent elements, H_2 and O_2. This occurs when electric current flows through water . Pure water, however, is a very bad conductor of electricity. To make the current flow, dissolve some acid, alkali or salt in the water to form an electrolyte. The molecules of such substances undergo dissociation in water that is, they split into positive and negative ions ("cations" and "anions", respectively):

$$H_2SO_4 \leftrightarrow 2H^+ + SO_4^{2-}$$

$$NaOH \leftrightarrow 2Na^+ + OH^-$$

$$Na_2SO_4 \leftrightarrow 2Na^+ + SO_4^{2-}$$

When current runs through the electrolyte between two immersed electrodes, the positive cations (H^+, Na^+) flow towards the "cathode", i.e., the negative electrode, and the negative anions (SO^{2-}, OH^-) flow towards the "anode", i.e., the positive electrode. At the electrodes additional chemical reaction may occur (more details are given in the YouTube video linked above), resulting in liberating hydrogen gas H_2 at the cathode, and liberating oxygen gas O_2 at the anode.

The water electrolysis device is called an electrolyzer. Large plants with units of a total capacity of the order of 10 MW are already being built, and facilities with a capacity as large as 100 MW are expected to emerge in the near future in countries actively implementing the "hydrogen economy".

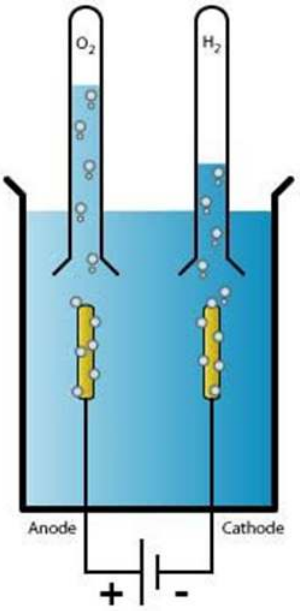


Figure 11.4.5.2.1: Electrolysis of water – i.e., decomposing H_2O molecules to H_2 and O_2 by sending electric current through a water solution of an acid (e.g., H_2SO_4) or a base (e.g., NaOH) – from Wikimedia Commons.

At an industrial scale, three types of electrolyzers are currently used. We will give only their names and their performance, but technical details concerning the design can be found on the web pages linked. So these three types are: alkaline electrolyzers, Proton Exchange Membrane (PEM) electrolyzers, and cells for high temperature electrolysis (HTE) usually referred to as solid oxide electrolytic cells (SOEC).

It is instructive to examine the energy r elations in water electrolysis. They are presented, e.g., in this Hyperphysics website. As can be seen in this page, an electric energy input of 237.13 KJ is needed brake down one mole of H_2O into hydrogen and oxygen. In the same illustration one can also see a value of ΔH = 285.83 kJ. What's the meaning of this symbol? It's simply the heat of

combustion of hydrogen, i.e. the thermal energy released when one mole of hydrogen is burned in the process of combining it back with oxygen. So, does it mean that from the latter reaction one gets 285.83 kJ 237.13 kJ MJ = 48.7 kJ of thermal energy more than the electric energy that was put in? At first glance, this may seem like a violation of energy conservation! But no, everything is fine: as can be seen again on the Hyperphysics Website, it's the "environment" which provides this missing energy. As is indicated at the Web page, the process considered by Hyperphysics is carried out at a temperature of 298 K – meaning that there is a "thermal energy component" present in the water subjected to electrolysis. In the course of the process, part of this thermal energy contained in water is taken away from it to augments the electrolysis – which would result in lowering the water temperature, if not for the new heat still flowing in from outside.

In view of the above, we may conclude that the "practical efficiency" (also called the "voltage cell efficiency") of water electrolysis – it is, the ratio of the energy contained in the released hydrogen to the electric energy put in is (285.83 kJ/ 237.13 kJ) 100% = 120.5%. Well, it would indeed be so at ideal conditions. In real situations, there are many additional factors that lower the efficiency. It depends on the type of the electrolyzer, and it is also correlated with the price one has to pay for it:

- Alkaline electolyzers are the least expensive ones. They use water solution of NaOH or KOH as the electrolyte, work at temperature 60-80°C, at pressure up to 30 bar; and their efficiency may may be up to 70%. The design of the electrolyzer most closely resembles the diagram in Figure 11.16, the only major difference is a membrane between the anode and cathode, which is transparent for ions but prevents gas mixing.
- PEM electrolyzers work at similar temperatures and pressures as the alkaline ones, but the arrangement of the electrodes is quite different than in a "classical" electrolysis experiment. It uses pure water that is suplied only to the anode, where it is split by a catalyzer to O_2 molecules and H^+ ions. The latter, which are "bare" protons, migrate through the PEM membrane to the cathode, where they recombine with incoming electrons and form H_2 molecules (the functioning of a PEM electrolyzer is clearly explained by an animated diagram in this Wikipedia article). The voltage cell efficiency of PEM electrolyzers reportedly reaches 80% and this is not expected to be the upper limit because intensive R&D efforts are still conducted on this technology. However, the electrolyzers are more expensive than the alkaline ones.
- The HTE electrolyzers, also known as "Solid Oxide Electrolytic Cells", are even more expensive than the PEM onesa. They work at temper-atures as high as 500-850°. They use not liquid water but steam which is fed fed to cathode and there it is split to H_2 molecules and O^- ions. The latter migrate across the solid oxide layer to the anode, where give away their electrons and form O_2 molecules. Due to the high temperature, the amount ΔH of thermal energy augmenting the reaction is much higher than in the other two methods, so that the "cell voltage efficiency" in an HTE electolyzer may exceed 100%. However, the "environment" is not able to supply the thermal energy needed, it must be provided by an additional heat source. Therefore, good conditions for the deployment of HTE technology exist where therer is no need to pay extra for heat for example, close to installations that produce a lot of waste heat, or in places where there is access to geothermal heat.

In industrial practice, the amount of hydrogen is not measured in moles, but in cubic meters which the gas occupies under standard conditions. One such cubic meter contains 44.6 moles of hydrogen. Therefore, to produce it under ideal conditions, 237.13 kJ / mol 44.6 mol = 10.59 MJ = 2.94 kWh of electric energy is needed. With the efficiency offered by the technologies considered, electricity consumption will be then from 3 to 5 kWh/m^3. Another measure is 1 kg of hydrogen, which occupies 11.2 cubic meters – so that the electricity consumption is 34 56 kWh/kg depending on the method used. The average price per 1 kWh paid by industrial consumers in the US is about 7 cents/kWh. Hence, the cost of 1 kg of electrolytic hydrogen produced at industrial scale is between $2.40 and $3.90. And this is only the price for electricity; the cost of operating the equipment and amortization of the costs invested in its purchase will certainly add something to this price. This does not compare favorably with the price of hydrogen obtained by steam reforming of natural gas, which – as can be found in various sources – may be as low as $1.25/kg.

To sum up as long as the price is a relevant factor, "green" electrolytic hydrogen loses in competition with that produced by steam reforming of natural gas. But it wouldn't lose if electricity were for free ... And this is not just a hypothetical situation, but quite real. In countries that are actively implementing a zero-carbon economy and are increasingly replacing fossil fuel power plants with wind and solar farms, this is already happening. Thanks to the whims of intermittent weather! Every once in a while there is too much wind and sun. The renewable sources suddenly start generating too much power, the consumers are not able to use the surplus. What is being done then? Energy production gets curtailed, some windmills and PV panels are simply shut down. Which is an obvious waste, this surplus power can be sent, after all, to electrolysis plants. At electricity cost reduced to zero or nearly zero, the hydrogen obtained in such situations will certainly be cheaper than that obtained from methane. There will be more and more of

this cheap electrolytic hydrogen as the potential of the renewable energy sector increases. And not only will it be obtained from excess power generated by strong winds or too much sunshine but, with time, dedicated wind turbine and solar farms will be built with the main purpose of supplying electricity to large water electrolysis plants. These are plans for the future in Australia, Germany, UK, Belgium, the Netherlands... – and in California. Interesting analyzes regarding the perspectives of the developing "hydrogen economy" in the context of prices and other economic aspects can be found, for example, in this British publication, or in this report compiled by Akin, Gump, Strauss, Hauer & Feld, a renowned Washington, D.C. "think tank".

Carbon-neutral Hydrogen derived from methane. The world today is literally "flooded" with cheap methane. When the production of cheap electrolytic hydrogen develops to such an extent that it begins to threaten the dominance of methane, its producers certainly will not want to "give up the field" without a fight. If stringent restrictions against CO_2 emissions are implemented, they may fail. But there is still a way to stay on the market in this situation namely, the conversion of methane into hydrogen without releasing CO_2 into the atmosphere. Is it possible? Yes! A reaction known as methane pyrolysis can be used for this purpose. At temperatures as high as 800-1000°C in the presence of metallic catalysts, the following reaction takes place:

$$\mathrm{CH_4} \cdot -2\mathrm{H_2} \cdot + \cdot \mathrm{C_{solid}}$$

The catalyst can be copper alloy with bismuth, it can also be lead. At high temperatures, these metals are molten. They are placed in a high vertical column, Methane under high pressure is blown into the column from below. Bubbles of gas travel through the molten metal, decomposing into hydrogen and carbon in solid form along the way. Hydrogen is discharged from the upper end, and carbon in a sooth-like form accumulates above the metal from where it can be removed. For every four kilograms of methane processed, one kilogram of hydrogen and three kilograms of solid carbon are obtained. Blowing methane through molten metal is not the only possible method of carrying out the pyrolysis. A German chemical giant BASF is experimenting with a whole spectrum of other methods, as described in this presentation.

But let's do the numbers. The heat of combustion of hydrogen is 141.6 MJ/kg and of methane it's 55.6 MJ/kg. So from pyrolysis conversion on gets a fuel that will release only 141.6/(4x55.6) = 63.7% of the energy that would be released by burning the methane used as the input. Yet, say, if California imposed a total ban on methane fuel, natural gas suppliers might prefer to convert their stocks to zero-emission fuel rather than to lose the market entirely.

A question arises, what to do with the solid carbon byproduct? It's not an acute problem, though, because solid carbon can be utilized in many ways. For instance, if mixed with top soil, it may act as sort of fertilizer, facilitating circulation of water and air.

Nearly carbon-neutral hydrogen from biomass. Biomass is a term that covers a whole variety of things mainly of plant origin: agricultural residues such as stems and leaves; residues from logging and the wood industry such as branches and sawdust; municipal waste, such as leaves removed in autumn; garbage left after metal, glass and plastic have been removed; and many other residues.

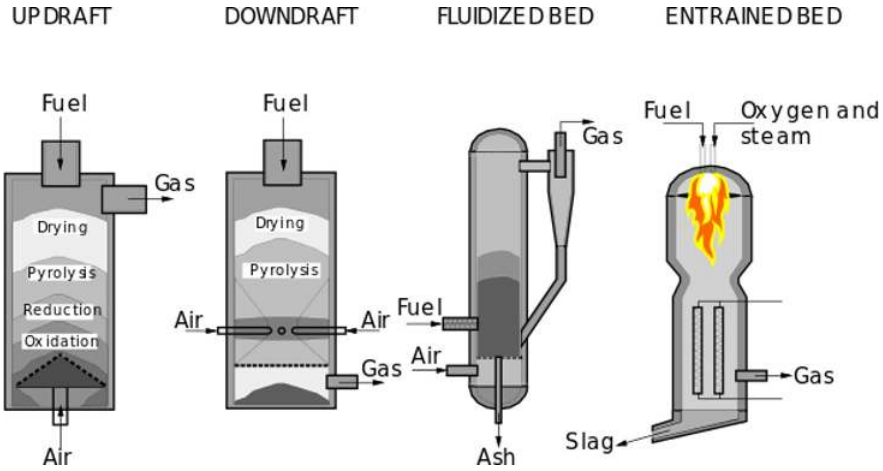


Figure 11.4.5.2.2: Different types of gasifiers used for processing biomass (from commons.wikimedia).

The main components of biomass are cellulose, hemicellulose and lignin.

In Chapter 10, we discussed one way to use biomass, namely, obtaining bioethanol, for which only one ingredient can be used – the cellulose. But we have not mentioned yet another technology, called biomass gasification, in which all its components are used.

The biomass in it is subjected to high temperature (> 700°C) in the presence of steam and oxygen in a device called “gasifier”.

Cellulose is a polymer consisting of chained “links” of glucose molecules, $C_6H_{12}O_6$, and in a gasifier it yields hydrogen, carbon dioxide oxide and carbon monoxide – and the latter is then oxidized:

$$C_6H_{12}O_6 \cdot + \cdot 2H_2O \cdot + O_2 \cdot \rightarrow 8H_2 \cdot + \cdot 4CO_2 \cdot + \cdot 2CO$$

$$2CO \cdot + \cdot 2H_2O \cdot \rightarrow 2H_2 \cdot + \cdot 2CO_2 \uparrow$$

The two equations can be combined into a single overall one:

$$C_6H_{12}O_6 \cdot + \cdot 4H_2O \cdot + \cdot O_2 \cdot \rightarrow 1OH_2 \cdot + \cdot 6CO_2$$

The building blocks for hemicellulose are five-carbon (pentose) sugars, their common chemical formula is $C_5H_{10}O_5$. The overall equation for the pentose sugars is:

$$C_5H_{10}O_5 \cdot + \cdot H_2O \cdot + \cdot \cdot 2O_2 \cdot 6H_2 \cdot + \cdot 5CO_2$$

As can be readily calculated, the mass of hydrogen obtained from the gasification of cellulose and hemicellulose is about 11%, and 8%, respectively, of the mass of the “feedstock” put into the gasifier.

Lignin has a very complex atomic structure. The proportion of carbon, oxygen and hydrogen atoms in lignin from different plants may vary, so that it is not possible to formulate a single gasification equation. The process has to be studied experimentally, as described, e.g., in this research report. The relative mass of hydrogen obtained from lignin gasification is lower than than that obtained from the other two biomass components. But the relative percent of lignin in different kinds of biomass is usually lower than 20%, so that it is reasonable to assume that the mass of hydrogen obtained from gasification is about 10% of the mass of the “feedstock” used.

The 10% or so of hydrogen in the output gas fron the gasifier has to be separated from the other components – there are several methods of doing this, e.g., by using porous ceramics or a technique known as Pressure Swing Adsorption (PSA). The latter is widely used to separate hydrogen from CO_2 in gas obtained from steam reforming og natural gas. Note that the gases produced by biomass gasification and natural gas reforming contain similar fractions of hydrogen (˜10% and ˜15% by weigh, respectively), while the remainder is mainly CO_2.

Just a moment, just a moment! At the beginning of the story, it was clearly stated that it would be about “nearly carbon neutral” method – and now it turns out that about 90% of the exhaust from the gasifier is CO_2. It makes no sense!

Well, it does make. Note that the CO_2 produced by biomass gasification is not “a new CO_2” added to the atmosphere. In order to create biomass, plants absorbed CO_2 from the air – and now this CO_2 is only given back to it. It’s the same story as with all biofuels. The release of CO_2 into the atmosphere when burning biofuels should be regarded as a ”loan return”, and not as “littering”. CO_2 released to the atmosphere is not “evil in its own right” – it’s evil only if it’s produced by burning fossil fuels. If we do not collect biomass to produce hydrogen, but instead we leave it where it has grown then **we** will not release CO_2 to the atmosphere, but, eventually, it will get there anyway after a somewhat longer period, due to the decaying processes caused by fungi, molds, bacteria and a range of other microorganisms.

11.4.5.2: Methods of Green Hydrogen Production is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.5.3: The recovery of energy stored in the form of hydrogen

Hydrogen → Heat: the **easiest method** is simply to burn hydrogen for heat. Just like we do it with natural gas we use it to heat homes, in kitchen stoves, in hot water heaters. Hydrogen can replace natural gas – and it does not have to happen right away – one can start with blending, say, 20

But further increasing the hydrogen content of the gas will not be quite easy, for several reasons. The first is the low heat of hydrogen combustion yes, low. How comes? After all, methane has a heat of combustion of 55.1 MJ/kg, and hydrogen up to 141.6.1 MJ / kg, 2.57 times higher! Right! But let's not forget that the gas bill is not for kilograms, but for cubic feet.

With gases it is so that at a certain temperature and pressure, one mole of gas always occupies the same volume. A mole of hydrogen is 2 grams, and a mole of methane 16 grams eight times more. Thus, one cubic foot of methane has a heat of combustion of 8 / 2.57 = 3.1 times greater than that of hydrogen. A more professional way is to say: *hydrogen's* ***volumetric heat of combustion*** *is 3.1 times lower than that of methane.*

What does it mean? To provide the users with the same amount of energy as with methane, you will need to supply them with three times the volume of hydrogen. Existing pipelines for something like this are not prepared therefore, complete transition from natural gas to hydrogen will require changing pipes in gas pipelines with new ones of a larger diameter. Similarly, home heating stoves, kitchen stoves, hot water heaters and even gas grills will require far-reaching alterations. It will be a huge and very expensive work that will take years. But the alternative is to do nothing and accept the increasing concentration of CO_2 in the atmosphere. Moneywise, the cost will be less, but the overall price the society will have to pay not necessarily in money may turn out to be higher.

Hydrogen → Electric power. The energy stored in hydrogen can also be recovered in the form of electricity. We will present two types of devices that enable such conversion.

The first of these are **fuel cells**. They can be thought of as devices in which water electrolysis "takes place in reverse". An electrolyzer takes electric energy as an input, and delivers oxygen and hydrogen as the output products. In a fuel cell, hydrogen and oxygen are the input, and electric energy is the output product (with water vapor as a byproduct).

The oldest "ancestor" of modern fuel cells was invented as early as in 1838, but the first usable practical one was the 5 kW "alkaline" cell developed by a British engineer F.T. Bacon in 1932. More intensive research on fuel cell technology began around 1950. As a result, a whole spectrum of new types was created (some of them may use other fuels than hydrogen, e.g., methanol, but here we limit our attention only to those using hydrogen).

All hydrogen fuel cells have a similar structure: they feature two electrodes, usually made of porous metals (so that gases can penetrate them), separated by a layer of electrolyte. Hydrogen is fed to one electrode, oxygen to the other. Most often, instead of pure oxygen, air is used because nitrogen and other air components do not interfere with the cell action.

In a **PEM-type fuel cell** (PEM = Proton Exchange Membrane a.k.a. Polymer Electrolyte Membrane), the electrolyte is a made of a substance that enables a flow of H^+ ions, i.e., of protons. Hydrogen gas is fed to the anode, in which, due to the presence of platinum catalyst, H_2 molecules are ionized by giving away two electrons. These electrons travel through the external circuit to the cathode, where they – again aided by a catalyst – join oxygen atoms to form O^{2-} ions. At the same time, the H^+ ions flow across the PEM membrane to the cathode, where they recombine with the O^{2-} ions, forming H_2O molecules. In short, in the cell oxygen and hydrogen combine, creating water, and causing a flow of electric current through an external circuit. The PEM cells typically operate at temperatures of 80 100˚C.

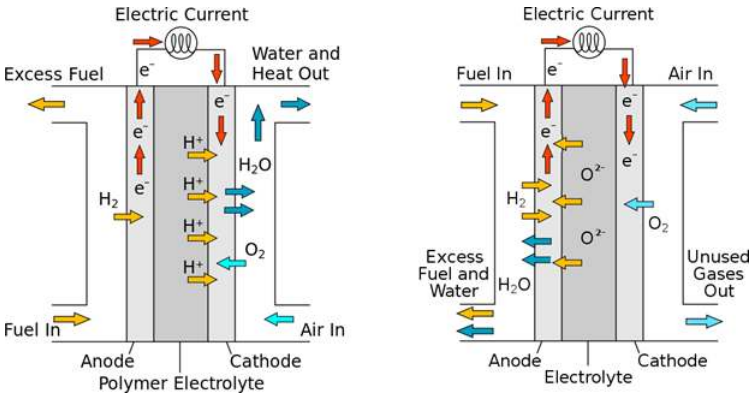


Figure 11.4.5.3.1:The schemes of a PEM fuel cell (left plot) and the SOFC cell (right plot), showing how the gases are fed in, and how ions travel across the electrolyte (from wikimedia.commons).

In **solid oxide fuel cells** (SOFC), in contrast, it's the O_2 ions that travel from the cathode to the anode, through the electrolyte made of metal oxide ceramics. Such ceramics acquires ionic conductivity only at high temperatures, so that the SOFC cells operate in the temperature range between 700°C and 1000°C.

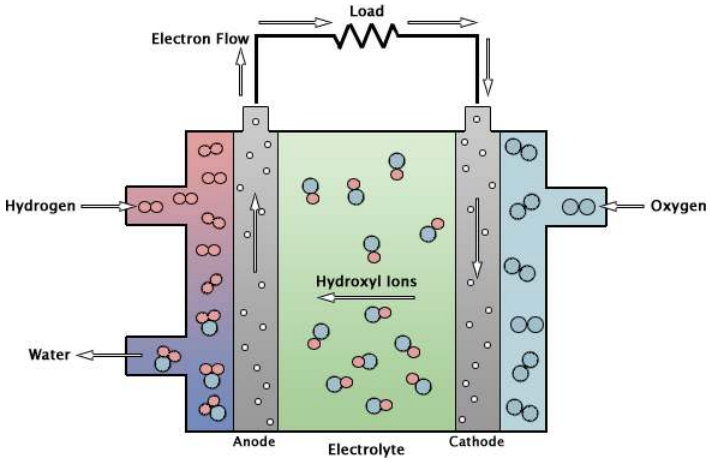


Figure 11.4.5.3.2: An alkaline fuel cell.

Yet another type of ionic conductivity is used in **alkaline fuel cells**. Invented in 1932, they have the simplest design but despite the lack of sophistication, they, as the only fuel cells to date, traveled to the Moon aboard the "Apollo" spacecrafts. The electrolyte in these cells is a porous substance saturated with an aqueous solution of potassium hydroxide KOH. In this solution, KOH dissociates into K^+ and OH^- ions – so that here ionic conductivity is a natural phenomenon. In the cathode, into which electrons arrive through an external circuit, water, oxygen and electrons react to form hydroxyl ions OH^-. Next, these ions flow through the electrolyte to the anode, where there are H^+ ions awaiting them (they are created from hydrogen which, aided by a catalyst, donates electrons to the external circuit). As the result of H^+ and OH^- recombination water, is formed.

Phosphoric acid fuel cells (PAFC) are highly reliable sources of electric current. Their scheme is essentially the same as that of the PEM cell, but they use as the electrolyte a porous matrix soaked with phosphoric acid H_3PO_4. They operate at 150 200°C.

Comparison of Fuel Cell Technologies

Fuel Cell Type	Common Electrolyte	Operating Temperature	Typical Stack Size	Efficiency	Applications	Advantages	Disadvantages
Polymer Electrolyte Membrane (PEM)	Perfluoro sulfonic acid	50-100°C 122-212° typically 80°C	< 1kW–100kW	60% transportation 35% stationary	• Backup power • Portable power • Distributed generation • Transporation • Specialty vehicles	• Solid electrolyte reduces corrosion & electrolyte management problems • Low temperature • Quick start-up	• Expensive catalysts • Sensitive to fuel impurities • Low temperature waste heat
Alkaline (AFC)	Aqueous solution of potassium hydroxide soaked in a matrix	90-100°C 194-212°F	10–100 kW	60%	• Military • Space	• Cathode reaction faster in alkaline electrolyte, leads to high performance • Low cost components	• Sensitive to CO_2 in fuel and air • Electrolyte management
Phosphoric Acid (PAFC)	Phosphoric acid soaked in a matrix	150-200°C 302-392°F	400 kW 100 kW module	40%	• Distributed generation	• Higher temperature enables CHP • Increased tolerance to fuel impurities	• Pt catalyst • Long start up time • Low current and power
Molten Carbonate (MCFC)	Solution of lithium, sodium, and/or potassium carbonates, soaked in a matrix	600-700°C 1112-1292°F	300 kW-3 MW 300 kW module	45-50%	• Electric utility • Distributed generation	• High efficiency • Fuel flexibility • Can use a variety of catalysts • Suitable for CHP	• High temperature corrosion and breakdown of cell components • Long start up time • Low power density
Solid Oxide (SOFC)	Yttria stabilized zirconia	700-1000°C 1202-1832°F	1 kW–2 MW	60%	• Auxiliary power • Electric utility • Distributed generation	• High efficiency • Fuel flexibility • Can use a variety of catalysts • Solid electrolyte • Suitable for CHP & CHHP • Hybrid/GT cycle	• High temperature corrosion and breakdown of cell components • High temperature operation requires long start up time and limits

Figure 11.4.5.3.3: Comparison of the most important properties of the five fuel cell technologies described in the text (from Wikimedia Commons).

An interesting high-temperature fuel cell type is the **Molten Carbonate Fuel Cell**. It uses a mixture of carbonate salts as the electrolyte and an unusual ionic conductivity mechanism. Namely, the gas delivered to its cathode is a mixture of oxygen O_2 with carbon dioxide CO_2. There the gases react with incoming electrons to form CO^{2-} ions that travel across the molten electrolyte. At the anode, they react with H^+ ions, forming molecules of H_2O and CO_2. The latter is recycled to the cathode.

Essential parameters of the above four types, such as the electric power they deliver and the efficiency of they offer are given in the Table in Fig. 11.22. Note that fuel cells can be expected to convert into electricity up to 60% of the energy contained in the stored hydrogen.

Why only that much? Because in this case it's physics that comes to the fore! As we remember, in the case of water electrolysis, the energy contained in the hydrogen produced[1] may be greater than that brought in by the electricity used, because, as has been mentioned before, a certain slice of energy is added "by the environment". In other words, it's Mother Nature who adds some "free extra bonus" to the energy of electrolytically produced hydrogen. But if you want to use this hydrogen to energize a fuel cell, then Mother Nature will order you to give her that "bonus" back! More seriously it's physics that rules that some part of the chemical energy the hydrogen and oxygen bring together into the system **must** be converted to heat, not to electricity. Note that here there is a certain analogy with heat engines, in which as we have already discussed in Chapter 2 only part of the input energy can be converted to mechanical work, and the rest **must** be discharged as "waste heat" (but this is only a superficial analogy, because the detailed theory of heat engines and the famous Carnot Equation does not apply to fuel cells!). A closer analysis of the thermodynamic relationships in the fuel cell (which also apply to water electrolysis) is presented in this Hyperphysics website.

The efficiency of fuel cells may, however, be regarded as higher if we do not consider the heat emitted in them as " waste heat ", but as a valuable commodity. Everybody who lives in cooler climate appreciates the value of heat, especially in the wintertime!

In view of this, it is conceivable to employ a larger fuel cell installation as a source of electricity **and heat** for a number of households. Such an idea is not anything new – it is known as cogeneration and it has been implemented in many locations all over the world, serving communities as large as entire city districts, as well as villages with only a hundred households or so. The existing cogeneration systems are usually build around conventional fossilfuel thermal power plants (and sometimes geothermal plants).

Cogeneration using fuel cells has already found some applications – for example, it provides heating in electric cars and buses powered by such cells. However, this is a very modest "niche". Stationary installations could certainly serve household

communities. An average household in the US uses an average of 1400 W of heating power – as in summertime it's not needed, so in wintertime it is probably of the order of 3000 W. As follows from the table in Fig. 11.22, the electric power provided by an installation using the SOFC technology may be a high as 2 MW. Suppose that the co-generated heat power is only 1 MW – it still will be enough for over 300 households in the wintertime.

It may sound as an attractive option for some rural communities and townships – but it seems that there is not much interest for such an idea. Why? Probably because, as for now, all hydrogen produced worldwide is used by a few industrial branches and is not so cheap. Well, perhaps when

the "hydrogen economy", now in its infancy, will grow in strength and begin to flex it muscles, such installations will begin to emerge...

Gas turbines fueled by hydrogen. Fuel cells offer a direct method of converting the energy contained in hydrogen into electricity – an alternative way, less direct but of similar efficiency is to use a gas turbines. Such turbines, widely used by utility companies, now use natural gas as fuel. The latest models have the highest efficiency of all heat engines, even above 60%.

Manufacturers of these turbines already are aware that governments, as part of decarbonization policy, will impose increasingly stringent restrictions on the use of methane fuel. So they have the choice of either stopping the production of turbines or reworking them so that they can use pure hydrogen instead of methane. They clearly opt for the latter. Intensive research is underway on how to manage this type of conversion. For now, they are testing the use of fuel with a relatively small amount of hydrogen, 10-20%. Turbines tolerate this well, but further increasing the concentration of hydrogen up to 100% will already require major alterations of the fuel system. We will not describe the current status of this research here, but we provide links to four recent reports discussing these issues: Report number one, Report number two, Report number three, and Report number four.

In summary when the "hydrogen economy" reaches its maturity and gathers strength, it may become a stage of competitive struggle between hydrogen fuel cells and hydrogen gas turbines. We have to wait to see...

1If we say that "hydrogen contains energy", it should be kept in mind that this is only a "figure of speach". The actual carrier of the energy stored is **a pair of gases**, hydrogen and oxygen. But we don't store the latter because for recovering the energy we can simply use the oxygen which is abundant in the air.

11.4.5.3: The recovery of energy stored in the form of hydrogen is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.5.4: Hydrogen storage and transportation issues

Production of green hydrogen and recovering energy from it are not the only challenges. If these operations are to be carried out as part of an ”energy storage project”, then obviously there must be a third challenge – namely, the storage of hydrogen itself. For well over a century, it has been known how to store hydrogen and other gases in steel cylinders . Everyone has certainly seen such cylinders, as they are used for many purposes, in hospitals, in welders' shops, and in various other places. They are usually 10 inches in diameter and are 50 inches tall and have an internal volume of about 50 liters. Depending on the strength of the steel from which they are made, they can withstand a pressure of 200 to 400 bar, which corresponds to a gas volume of about 10 m^3 to 20 m^3 at normal pressure.

Pressure cylinders are, so far, the only possible fuel tanks in fuel cell electric vehicles (FCEVs). It is obvious that the reservoirs themselves should be as light as possible, but withstand the highest possible pressure. Conventional steel cylinders are far too heavy. The current state-of-the-arts solution to the problem are cylinders with walls made of carbon fiber composites, withstanding pressure up to 70 MPa (MegaPascals; 1 MPa 10 bar). At this pressure, 1 kg of hydrogen occupies a volume of approx. 26 liters. In Tyota Mirai, a state-of-the-arts FCEV, the two fuel reservoirs of total weight of 87.5 kg can store 5 kg of hydrogen. It means that these reservoirs can store hydrogen the weight of which is only about 5.7% of their own weight. In the so-called tube trailers, specially designed for transporting compressed hydrogen (with a capacity of typically 500 1000 kg of H_2), the (H_2 weight)/(tube weight) ratio may be better, even exceeding 10 %.

Another possible approach to the transport and storage issues is to use liquefied hydrogen. For liquefaction, hydrogen needs to be cooled to 20 K and this process requires a minimum of 15 MJ of energy per kilogram in an ideal process, and about three times as much using the best technology available today. Considering that the energy contained in hydrogen alone is 142.6 Mj / kg, around 30% is lost on the liquefaction alone. Liquid hydrogen has to be stored in special cryogenic tanks – currently, if the best available thermal insulation is used, the evaporation losses in such tanks can be as low as 0.5-1% per day. But one more problem is the density of liquid hydrogen, as low as 71kg/m^3 – compared with 422 kg/m^3 for liquid methane. Hydrogen's heat of combustion is higher than that of methane by a factor of 2.57, but if we want to transport liquid methane and liquid hydrogen of the same energy content, the volume of the hydrogen tanks has to be 2.3 times larger than that of the methane tanks.

Nevertheless, hydrogen transport in liquid form appears to be one of the few options for long-distance shipping by sea. Japan, who already has an agreement with Australia and is counting on serious supplies of Australian hydrogen in the future, is very interested in this possibility. With this in mind, the first liquid hydrogen tanker was just launched at the Kawasaki shipyard. It will be used for testing the feasibility of this transportation mode.

Transportation of chemically converted hydrogen – this is a method that allows transporting hydrogen over long distances, or storing it for long periods of time **without (in principle) any losses**. For this, one needs a substance that acts as a ”carrier”, and allows hydrogen atoms to get attached to its molecules in a reversible way – i.e., such that the H atoms can be after some time relatively easily “disatached” from these molecules, leaving them unchanged.

One candidate for such a carrier that has recently attracted much interest, in particular in Australia, is nitrogen. To a single nitrogen atom one can attach three hydrogen atoms, making a molecule of ammonia, NH3. Australia has already much experience with synthesizing ammonia for other purposes – esprecially, for using it as a fertilizer.

But now, when the prospect of a very significant increase in green hydrogen production appears, the Australians also spot another possible application the use of ammonia as the “carrier” , which will greatly facilitate the export of “green hydrogen” to remote recipients (see Australia's National Hydrogen Strategy, Page 40 and other pages alltogether, ammonia is mentioned in this document 36 times).

The heat of combustion of ammonia is 22.48 MJ/kg. The heat of combustion of hydrogen is 142.6 MJ/kg, over six times higher. But note that in 1 kg of ammonia there is only 3/17 kg of hydrogen. Burning such an amount of hydrogen would release 142.6 MJ/kg (3/17) kg = 25.16 MJ, which means that only 11% of energy originally contained in hydrogen is lost. But instead of burning ammonia (which itself can be used as a fuel), one can also “dis-attach” hydrogen atoms from nitrogen and continue to use pure hydrogen only.

One of the significant advantages of ammonia from the viewpoint of maritime transport is the ease of its liquefaction. To keep it liquefied, no low temperature is needed, only increased pressure: at 25 C, a pressure of about 10 bar is sufficient. Transport tanks

must be therefore sturdy, but they do not need thermal insulation.

Another candidate for a hydrogen carrier – more specifically, for a “Liquid Organic Hydrogen Carrier” (LOHC) is toluene, C_7H_8. One of the not-so-nice applications of toluene is that it is used for the production of TNT, a powerful explosive that killed countless victims in the First and Second World Wars, and in many other wars. It is pleasant, therefore, that now it can be used in an absolutely peaceful role: three molecules of H_2, six hydrogen atoms can be attached to one molecule of C_7H_8, yielding C_7H_{14} which is known as “methylcyclohexane”. This product can be loaded into tankers and sent to the recipient where the 6 hydrogen atoms can be detached relatively easily. And toluene can be returned to the sender for reuse.

In this method, the molecular weight of C_7H_{14} is 98, and the total weight of three H_2 molecules is 6, so that the mass of hydrogen in the shipment is only 6/98 = 6.1% of the total mass of the substance shipped, which is perhaps not a spectacular proportion, given that in the case of ammonia this ratio is 3/17 = 17.6%. But the advantage is that methylcyclohexane can be transported at ambient temperature and ambient pressure – therefore, one can use any containers, even ordinary plastic barrels. Such transport has just arrived in recent days from Brunei to Kawasaki, Japan, covering a distance of 4,000 miles.

11.4.5.4: Hydrogen storage and transportation issues is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

11.4.5.5: Hydrogen as a “green” fuel suitable for sea- sonal storage

There is an article in a recent issue of the pv magazine titled “Hydrogen is the first viable option for seasonal storage”. In fact, note that all methods which have been discussed in this Chapter so far make it possible to store an amount of sufficient for a day or a few days at most in the case of normal demand. A lithium battery in which one could accumulate a supply “ for the entire winter” would have to be of truly gargantuan size!

Hydrogen actually offers the first feasible method of storing “ green energy ”, which can be maintained for a long time without loss, and when it needs to be used, it will meet the demand for months.

There is no need to create any fancy new technology, because almost all necessary ”puzzle pieces” already exist. What are we talking about? Well, about this old-fashioned method of storing compressed hydrogen. But not in tanks made of steel or carbon fiber composite large enough could not be built in any way.

Let us recall, however, what the method of energy storage using compressed air, presented earlier in the Section in this chapter, taught us. Namely, that compressed gases can be stored in underground caverns excavated in natural salt deposit. It’s not a great problem to make such a cavern because no other tools ade needed than water – skilfully used, it will flush out salt, leaving a cavern of the desired size ans shape.

In Utah, in the city of Delta region, a huge salt dome exists deep underground. Its top is at a depth of about 750 meters below ground level, its thickness is about 1500 m, and it is about 5 km across. Currentrly, several companies are interested in the possibility of creating caverns in this salt deposit for various purposes.

As far as hydrogen storage is concerned, plans have been laid out by Mitsubichi Hitachi Power Systems America to excavate kn the Delta salt dome a storage cavern that could fit the entire Empire State Building. So let’s make a quick estimate. This whole skyscraper has a volume of 37,000,000 cubic feet (approx. 1,050,000 m^3). Suppose we fill this cavern with hydrogen at 100 bar (10 MPa). At this pressure, one cubic meter holds 7.3 kg of hydrogen, i.e. the mass of all hydrogen in the cavern will amount to about 7.5 million kg. The heat of combustion of hydrogen is 141.6 MJ / kg, multiplying these two numbers we get 1.062,000 GJ, or 295 GWh.

As we have estimated earlier at the beginning of this chapter, 300 GWh is the amount of energy that California needs in the evening and nighttime hours to completely free itself from the dependence on power generation from fossil fuels. Thus, if just one such cavern were used for storing hydrogen from electrolyzers energized by PV cells during the sunshine hours, and then this hydrogen were used for generating power after the sunset, the goal California has set for itself could be achieved.

It is estimated that in the Delta city region alone up to one hundred such caverns could be built. And similar storage reservoirs can be created in many other places where there are salt deposits. What’s more, not only salt caverns are suitable for storing hydrogen. Natural gas reservoirs from which gas has already been extracted are also suitable. Since such formation were able to hold methane for thousands or millions of years, one can now inject hydrogen into them without much fear that it would escape. And the storage capacity of such depleted gas fields can be enormous! In a recent British Web document (see Page 9) it is estimated that in one depleted gas field off the coasts of the British Islands the energy content of the hydrogen stored would be not several hundred GWh (as in the example discussed earlier), but over **ten thousand GWh!** In view of the above, we can be confident that there are storage methods capable of providing necessary energy not only for a day or a week, but even over long winter months!

********** *Finis Opus Coronat* **********

11.4.5.5: Hydrogen as a “green” fuel suitable for sea- sonal storage is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Tom Giebultowicz.

Index

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: Energy Alternatives

Webpages: 154

All licenses found:

- CC BY 4.0: 97.4% (150 pages)
- Undeclared: 2.6% (4 pages)

By Page

- Energy Alternatives - *CC BY 4.0*
  - Front Matter - *CC BY 4.0*
    - TitlePage - *CC BY 4.0*
    - InfoPage - *CC BY 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Energy Alternatives { what's the meaning?} - *CC BY 4.0*
    - 1.1: Brief semantic considerations - *CC BY 4.0*
    - 1.2: Things to be chosen" What things, what choice? - *CC BY 4.0*
    - 1.3: Back to the definitions- which one? 2a or 2b? - *CC BY 4.0*
  - 2: General Remarks - *CC BY 4.0*
    - 2.1: Energy and power - *CC BY 4.0*
      - 2.1.1: Energy Conservation - *CC BY 4.0*
      - 2.1.2: Energy and Work - *CC BY 4.0*
      - 2.1.3: Energy Units - *CC BY 4.0*
      - 2.1.4: Power - *CC BY 4.0*
      - 2.1.5: Energy expressed in terms of power - *CC BY 4.0*
    - 2.2: Forms of Energy- Physicist's View . - *CC BY 4.0*
      - 2.2.1: Potential Energy - *CC BY 4.0*
      - 2.2.2: Kinetic Energy - *CC BY 4.0*
      - 2.2.3: Energy of an Oscillator - *CC BY 4.0*
      - 2.2.4: Free Fall, Satellite Motion and the Mechanism of Tides - *CC BY 4.0*
        - 2.2.4.1: Free Fall - *CC BY 4.0*
        - 2.2.4.2: Circular Motion - *CC BY 4.0*
        - 2.2.4.3: Projectile Motion- Very Long Range - *CC BY 4.0*
        - 2.2.4.4: Satellite Motion - *CC BY 4.0*
        - 2.2.4.5: Tidal Forces - *CC BY 4.0*
        - 2.2.4.6: Extreme Tidal Effects - *CC BY 4.0*
    - 2.3: Thermal Energy and Heat - *CC BY 4.0*
      - 2.3.1: Forms of Energy and Power- an Engineer's Approach - *CC BY 4.0*
  - 3: Fossil Fuels and Thermal Engines - *CC BY 4.0*
    - 3.1: Fossil Fuel Resources - *CC BY 4.0*
    - 3.2: How much energy there is in fossil fuels? - *CC BY 4.0*
      - 3.2.1: How much CO2 is released by burning various fuels - *CC BY 4.0*
    - 3.3: Heat Engines - *CC BY 4.0*
    - 3.4: The Efficiency of Thermal Engines - *CC BY 4.0*
      - 3.4.1: The First Law of Thermodynamics. - *CC BY 4.0*
      - 3.4.2: Entropy - *CC BY 4.0*
      - 3.4.3: The Second Law of Thermodynamics - *CC BY 4.0*
      - 3.4.4: Sad Consequences of the Second Law - *CC BY 4.0*
      - 3.4.5: The Efficiency of an Ideal Thermal Engine - *CC BY 4.0*
      - 3.4.6: Even sadder news - *CC BY 4.0*
  - 4: Nuclear Power - *CC BY 4.0*
    - 4.1: A Note from the Author - *CC BY 4.0*
    - 4.2: Nuclear Physics- Brief Review of Fundamental Facts - *CC BY 4.0*
    - 4.3: Radioactive Decay - *CC BY 4.0*
      - 4.3.1: Energy Units in Nuclear Physics - *CC BY 4.0*
      - 4.3.2: The Decay Constants and Half-Lives - *CC BY 4.0*
    - 4.4: Nuclear Cross Sections - *CC BY 4.0*
    - 4.5: Nuclear Fission - *CC BY 4.0*
      - 4.5.1: Fission and Fusion Energy - *CC BY 4.0*
      - 4.5.2: How to calculate the mass defect? - *CC BY 4.0*
      - 4.5.3: Energy Released by the Fission of a Single Nucleus - *CC BY 4.0*

- 4.5.4: Nuclear Fusion - *CC BY 4.0*
- 4.6: Controlling the Fission Chain Reaction- Nuclear Reactors - *CC BY 4.0*
  - 4.6.1: Essential Cross Sections - *CC BY 4.0*
  - 4.6.2: No magic tricks are necessary! - *CC BY 4.0*
  - 4.6.3: Controlling the Chain Reaction and Cooling the Assembly - *CC BY 4.0*
  - 4.6.4: Power Reactors - *CC BY 4.0*
  - 4.6.5: Generation IV Reactors and Thorium Fuel - *CC BY 4.0*
- 4.7: NuScale's Small Modular Reactors – an Oregon-born Project - *CC BY 4.0*
- 4.8: The Pros and Cons of Nuclear Power - *CC BY 4.0*

5: Hydropower - *CC BY 4.0*
- 5.1: Introduction - *CC BY 4.0*
- 5.2: The Origin of Hydropower - *CC BY 4.0*
- 5.3: Methods of Harnessing Hydropower - *CC BY 4.0*
  - 5.3.1: The Kaplan and Francis Turbines - *CC BY 4.0*
  - 5.3.2: The Pelton Turbine - *CC BY 4.0*
- 5.4: About One Famous Hydropower Plant Using Pelton Turbines - *CC BY 4.0*
- 5.5: Pumped Storage Hydroelectric Plants (PSHP) - *CC BY 4.0*
- 5.6: New Ideas for Hydraulic Energy Storage - *CC BY 4.0*

6: Windpower - *CC BY 4.0*
- 6.1: Applications- Winds Over the Seas - *CC BY 4.0*
- 6.2: Winds Blowing over the Lands - *CC BY 4.0*
- 6.3: Applications- Land and Off-shore Use - *CC BY 4.0*
- 6.4: The Physics of a Wind Turbine - *CC BY 4.0*
  - 6.4.1: Problems with Wind Power - *CC BY 4.0*

7: Direct Solar Power - *CC BY 4.0*
- 7.1: Introduction - *CC BY 4.0*
- 7.2: Harnessing Solar Power - *CC BY 4.0*
- 7.3: The Concentrated Solar Power (CSP) Technology - *CC BY 4.0*
  - 7.3.1: Solar Towers Molten Salt Heat Storing Technology - *CC BY 4.0*
  - 7.3.2: Parabolic Troughs and Fresnel Mirrors - *CC BY 4.0*
  - 7.3.3: The Linear Fresnel Mirror CSP Technology - *CC BY 4.0*
  - 7.3.4: CSP – Pros and Contras - *CC BY 4.0*
- 7.4: Photovoltaic Conversion of Solar Power - *CC BY 4.0*
  - 7.4.1: Theoretical Background the Nature of Light and the Photoelectric Effect - *CC BY 4.0*
  - 7.4.2: Semiconductors and Doped Semiconductors - *CC BY 4.0*
  - 7.4.3: The p-n Junction - *CC BY 4.0*
  - 7.4.4: Internal Photoelectric Effect in Semiconductors - *CC BY 4.0*
  - 7.4.5: PV Cells - *CC BY 4.0*
  - 7.4.6: Solar Cell Efficiency - *CC BY 4.0*
  - 7.4.7: Inverters - *CC BY 4.0*
  - 7.4.8: Pros and Cons of PV Solar Power Harnessing - *CC BY 4.0*

8: Harnessing Power from Oceans and Seas - *CC BY 4.0*
- 8.1: Introduction - *CC BY 4.0*
- 8.2: Wavepower - *CC BY 4.0*
  - 8.2.1: Methods of Harnessing Wavepower - *CC BY 4.0*
  - 8.2.2: Attenuators - *CC BY 4.0*
  - 8.2.3: Point Absorbers - *CC BY 4.0*
  - 8.2.4: Other Wave Power Capturing Devices - *CC BY 4.0*
- 8.3: Tidal Power - *CC BY 4.0*
  - 8.3.1: A Bit More of Physics - *CC BY 4.0*
  - 8.3.2: Tides in Coastal Regions, Tidal Currents - *CC BY 4.0*
  - 8.3.3: Methods of Harnessing Tidal Power - *CC BY 4.0*
  - 8.3.4: Tidal Power – Conclusions - *CC BY 4.0*
- 8.4: Ocean Thermal Energy Conversion (OTEC) - *CC BY 4.0*
- 8.5: Other Possible Ways of Extracting Power from Oceans - *CC BY 4.0*
  - 8.5.1: Ocean Currents - *CC BY 4.0*
  - 8.5.2: Osmotic (Salinity Gradient) Power Generation - *CC BY 4.0*

9: Geothermal Power - *CC BY 4.0*
- 9.1: The Non-Uniform Distribution of Geothermal Heat Flux - *CC BY 4.0*
- 9.2: Geothermal Electricity Generation - *CC BY 4.0*
  - 9.2.1: The Flash Steam Method - *CC BY 4.0*
  - 9.2.2: The Binary Cycle Method - *CC BY 4.0*
- 9.3: The Enhanced Geothermal Systems - *CC BY 4.0*
- 9.4: Geothermal Heating - *CC BY 4.0*
  - 9.4.1: Shallow Geothermal Heating - *CC BY 4.0*

10: Biofuels - *CC BY 4.0*
- 10.1: The History of Biofuels - *CC BY 4.0*
- 10.2: New Solid Biofuels - *CC BY 4.0*

- 10.3: Liquid Biofuels - *CC BY 4.0*
  - 10.3.1: Bioethanol - *CC BY 4.0*
  - 10.3.2: Biodiesel Fuel - *CC BY 4.0*
- 10.4: Other Ways of Harnessing Plant Energy- “Bioelectricity - *CC BY 4.0*

11: Energy Storage - *CC BY 4.0*

- 11.1: Introduction - *CC BY 4.0*
- 11.2: Overcoming the Challenges of Wind Power - *CC BY 4.0*
- 11.3: Problems with Solar Power- the \Duck Curve" - *CC BY 4.0*
- 11.4: Energy Storage Technologies - *CC BY 4.0*
  - 11.4.1: Battery Storage - *CC BY 4.0*
    - 11.4.1.1: The Voltaic Cell - *CC BY 4.0*
    - 11.4.1.2: Lead-Acid Batteries - *CC BY 4.0*
    - 11.4.1.3: Lithium-Ion Batteries - *CC BY 4.0*
    - 11.4.1.4: The “Families” of Li-ion Batteries Manufactured - *CC BY 4.0*
    - 11.4.1.5: Batteries for Industry- and Grid-scale Storage - *CC BY 4.0*
  - 11.4.2: Electricity → potential energy → back to electricity (gravitational energy storage and compressed air energy storage). - *CC BY 4.0*
  - 11.4.3: Electricity → Kinetic Energy → Electricity Storage Scheme - *CC BY 4.0*
  - 11.4.4: Electricity → Thermal Energy → Electricity Storage Scheme - *CC BY 4.0*
    - 11.4.4.1: Molten Silicon - *CC BY 4.0*
    - 11.4.4.2: Heated Volcanic Rocks - *CC BY 4.0*
    - 11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage - *CC BY 4.0*
    - 11.4.4.4: Small and Medium Scale Thermal Energy Storage - *CC BY 4.0*
  - 11.4.5: Electricity → Chemical Energy - *CC BY 4.0*
    - 11.4.5.1: The evolution of the hydrogen economy policy in the 21st century - *CC BY 4.0*
    - 11.4.5.2: Methods of Green Hydrogen Production - *CC BY 4.0*
    - 11.4.5.3: The recovery of energy stored in the form of hydrogen - *CC BY 4.0*
    - 11.4.5.4: Hydrogen storage and transportation issues - *CC BY 4.0*
    - 11.4.5.5: Hydrogen as a “green” fuel suitable for sea- sonal storage - *CC BY 4.0*

Back Matter - *CC BY 4.0*

- Index - *CC BY 4.0*
- Glossary - *CC BY 4.0*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*

Detailed Licensing

Overview

Title: Energy Alternatives

Webpages: 154

All licenses found:

- CC BY 4.0: 97.4% (150 pages)
- Undeclared: 2.6% (4 pages)

By Page

- Energy Alternatives - *CC BY 4.0*
  - Front Matter - *CC BY 4.0*
    - TitlePage - *CC BY 4.0*
    - InfoPage - *CC BY 4.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Energy Alternatives { what's the meaning?} - *CC BY 4.0*
    - 1.1: Brief semantic considerations - *CC BY 4.0*
    - 1.2: Things to be chosen" What things, what choice? - *CC BY 4.0*
    - 1.3: Back to the definitions- which one? 2a or 2b? - *CC BY 4.0*
  - 2: General Remarks - *CC BY 4.0*
    - 2.1: Energy and power - *CC BY 4.0*
      - 2.1.1: Energy Conservation - *CC BY 4.0*
      - 2.1.2: Energy and Work - *CC BY 4.0*
      - 2.1.3: Energy Units - *CC BY 4.0*
      - 2.1.4: Power - *CC BY 4.0*
      - 2.1.5: Energy expressed in terms of power - *CC BY 4.0*
    - 2.2: Forms of Energy- Physicist's View . - *CC BY 4.0*
      - 2.2.1: Potential Energy - *CC BY 4.0*
      - 2.2.2: Kinetic Energy - *CC BY 4.0*
      - 2.2.3: Energy of an Oscillator - *CC BY 4.0*
      - 2.2.4: Free Fall, Satellite Motion and the Mechanism of Tides - *CC BY 4.0*
        - 2.2.4.1: Free Fall - *CC BY 4.0*
        - 2.2.4.2: Circular Motion - *CC BY 4.0*
        - 2.2.4.3: Projectile Motion- Very Long Range - *CC BY 4.0*
        - 2.2.4.4: Satellite Motion - *CC BY 4.0*
        - 2.2.4.5: Tidal Forces - *CC BY 4.0*
        - 2.2.4.6: Extreme Tidal Effects - *CC BY 4.0*
    - 2.3: Thermal Energy and Heat - *CC BY 4.0*
      - 2.3.1: Forms of Energy and Power- an Engineer's Approach - *CC BY 4.0*
  - 3: Fossil Fuels and Thermal Engines - *CC BY 4.0*
    - 3.1: Fossil Fuel Resources - *CC BY 4.0*
    - 3.2: How much energy there is in fossil fuels? - *CC BY 4.0*
      - 3.2.1: How much CO2 is released by burning various fuels - *CC BY 4.0*
    - 3.3: Heat Engines - *CC BY 4.0*
    - 3.4: The Efficiency of Thermal Engines - *CC BY 4.0*
      - 3.4.1: The First Law of Thermodynamics. - *CC BY 4.0*
      - 3.4.2: Entropy - *CC BY 4.0*
      - 3.4.3: The Second Law of Thermodynamics - *CC BY 4.0*
      - 3.4.4: Sad Consequences of the Second Law - *CC BY 4.0*
      - 3.4.5: The Efficiency of an Ideal Thermal Engine - *CC BY 4.0*
      - 3.4.6: Even sadder news - *CC BY 4.0*
  - 4: Nuclear Power - *CC BY 4.0*
    - 4.1: A Note from the Author - *CC BY 4.0*
    - 4.2: Nuclear Physics- Brief Review of Fundamental Facts - *CC BY 4.0*
    - 4.3: Radioactive Decay - *CC BY 4.0*
      - 4.3.1: Energy Units in Nuclear Physics - *CC BY 4.0*
      - 4.3.2: The Decay Constants and Half-Lives - *CC BY 4.0*
    - 4.4: Nuclear Cross Sections - *CC BY 4.0*
    - 4.5: Nuclear Fission - *CC BY 4.0*
      - 4.5.1: Fission and Fusion Energy - *CC BY 4.0*
      - 4.5.2: How to calculate the mass defect? - *CC BY 4.0*
      - 4.5.3: Energy Released by the Fission of a Single Nucleus - *CC BY 4.0*

- 4.5.4: Nuclear Fusion - *CC BY 4.0*
- 4.6: Controlling the Fission Chain Reaction- Nuclear Reactors - *CC BY 4.0*
  - 4.6.1: Essential Cross Sections - *CC BY 4.0*
  - 4.6.2: No magic tricks are necessary! - *CC BY 4.0*
  - 4.6.3: Controlling the Chain Reaction and Cooling the Assembly - *CC BY 4.0*
  - 4.6.4: Power Reactors - *CC BY 4.0*
  - 4.6.5: Generation IV Reactors and Thorium Fuel - *CC BY 4.0*
- 4.7: NuScale's Small Modular Reactors – an Oregon-born Project - *CC BY 4.0*
- 4.8: The Pros and Cons of Nuclear Power - *CC BY 4.0*

5: Hydropower - *CC BY 4.0*
- 5.1: Introduction - *CC BY 4.0*
- 5.2: The Origin of Hydropower - *CC BY 4.0*
- 5.3: Methods of Harnessing Hydropower - *CC BY 4.0*
  - 5.3.1: The Kaplan and Francis Turbines - *CC BY 4.0*
  - 5.3.2: The Pelton Turbine - *CC BY 4.0*
- 5.4: About One Famous Hydropower Plant Using Pelton Turbines - *CC BY 4.0*
- 5.5: Pumped Storage Hydroelectric Plants (PSHP) - *CC BY 4.0*
- 5.6: New Ideas for Hydraulic Energy Storage - *CC BY 4.0*

6: Windpower - *CC BY 4.0*
- 6.1: Applications- Winds Over the Seas - *CC BY 4.0*
- 6.2: Winds Blowing over the Lands - *CC BY 4.0*
- 6.3: Applications- Land and Off-shore Use - *CC BY 4.0*
- 6.4: The Physics of a Wind Turbine - *CC BY 4.0*
  - 6.4.1: Problems with Wind Power - *CC BY 4.0*

7: Direct Solar Power - *CC BY 4.0*
- 7.1: Introduction - *CC BY 4.0*
- 7.2: Harnessing Solar Power - *CC BY 4.0*
- 7.3: The Concentrated Solar Power (CSP) Technology - *CC BY 4.0*
  - 7.3.1: Solar Towers Molten Salt Heat Storing Technology - *CC BY 4.0*
  - 7.3.2: Parabolic Troughs and Fresnel Mirrors - *CC BY 4.0*
  - 7.3.3: The Linear Fresnel Mirror CSP Technology - *CC BY 4.0*
  - 7.3.4: CSP – Pros and Contras - *CC BY 4.0*
- 7.4: Photovoltaic Conversion of Solar Power - *CC BY 4.0*
  - 7.4.1: Theoretical Background the Nature of Light and the Photoelectric Effect - *CC BY 4.0*
  - 7.4.2: Semiconductors and Doped Semiconductors - *CC BY 4.0*
  - 7.4.3: The p-n Junction - *CC BY 4.0*
  - 7.4.4: Internal Photoelectric Effect in Semiconductors - *CC BY 4.0*
  - 7.4.5: PV Cells - *CC BY 4.0*
  - 7.4.6: Solar Cell Efficiency - *CC BY 4.0*
  - 7.4.7: Inverters - *CC BY 4.0*
  - 7.4.8: Pros and Cons of PV Solar Power Harnessing - *CC BY 4.0*

8: Harnessing Power from Oceans and Seas - *CC BY 4.0*
- 8.1: Introduction - *CC BY 4.0*
- 8.2: Wavepower - *CC BY 4.0*
  - 8.2.1: Methods of Harnessing Wavepower - *CC BY 4.0*
  - 8.2.2: Attenuators - *CC BY 4.0*
  - 8.2.3: Point Absorbers - *CC BY 4.0*
  - 8.2.4: Other Wave Power Capturing Devices - *CC BY 4.0*
- 8.3: Tidal Power - *CC BY 4.0*
  - 8.3.1: A Bit More of Physics - *CC BY 4.0*
  - 8.3.2: Tides in Coastal Regions, Tidal Currents - *CC BY 4.0*
  - 8.3.3: Methods of Harnessing Tidal Power - *CC BY 4.0*
  - 8.3.4: Tidal Power – Conclusions - *CC BY 4.0*
- 8.4: Ocean Thermal Energy Conversion (OTEC) - *CC BY 4.0*
- 8.5: Other Possible Ways of Extracting Power from Oceans - *CC BY 4.0*
  - 8.5.1: Ocean Currents - *CC BY 4.0*
  - 8.5.2: Osmotic (Salinity Gradient) Power Generation - *CC BY 4.0*

9: Geothermal Power - *CC BY 4.0*
- 9.1: The Non-Uniform Distribution of Geothermal Heat Flux - *CC BY 4.0*
- 9.2: Geothermal Electricity Generation - *CC BY 4.0*
  - 9.2.1: The Flash Steam Method - *CC BY 4.0*
  - 9.2.2: The Binary Cycle Method - *CC BY 4.0*
- 9.3: The Enhanced Geothermal Systems - *CC BY 4.0*
- 9.4: Geothermal Heating - *CC BY 4.0*
  - 9.4.1: Shallow Geothermal Heating - *CC BY 4.0*

10: Biofuels - *CC BY 4.0*
- 10.1: The History of Biofuels - *CC BY 4.0*
- 10.2: New Solid Biofuels - *CC BY 4.0*

- 10.3: Liquid Biofuels - *CC BY 4.0*
  - 10.3.1: Bioethanol - *CC BY 4.0*
  - 10.3.2: Biodiesel Fuel - *CC BY 4.0*
- 10.4: Other Ways of Harnessing Plant Energy- “Bioelectricity - *CC BY 4.0*

11: Energy Storage - *CC BY 4.0*

- 11.1: Introduction - *CC BY 4.0*
- 11.2: Overcoming the Challenges of Wind Power - *CC BY 4.0*
- 11.3: Problems with Solar Power- the \Duck Curve" - *CC BY 4.0*
- 11.4: Energy Storage Technologies - *CC BY 4.0*
  - 11.4.1: Battery Storage - *CC BY 4.0*
    - 11.4.1.1: The Voltaic Cell - *CC BY 4.0*
    - 11.4.1.2: Lead-Acid Batteries - *CC BY 4.0*
    - 11.4.1.3: Lithium-Ion Batteries - *CC BY 4.0*
    - 11.4.1.4: The “Families” of Li-ion Batteries Manufactured - *CC BY 4.0*
    - 11.4.1.5: Batteries for Industry- and Grid-scale Storage - *CC BY 4.0*
  - 11.4.2: Electricity → potential energy → back to electricity (gravitational energy storage and compressed air energy storage). - *CC BY 4.0*
  - 11.4.3: Electricity → Kinetic Energy → Electricity Storage Scheme - *CC BY 4.0*
  - 11.4.4: Electricity → Thermal Energy → Electricity Storage Scheme - *CC BY 4.0*
    - 11.4.4.1: Molten Silicon - *CC BY 4.0*
    - 11.4.4.2: Heated Volcanic Rocks - *CC BY 4.0*
    - 11.4.4.3: Pumped Heat Energy Storage and Liquid Air Energy Storage - *CC BY 4.0*
    - 11.4.4.4: Small and Medium Scale Thermal Energy Storage - *CC BY 4.0*
  - 11.4.5: Electricity → Chemical Energy - *CC BY 4.0*
    - 11.4.5.1: The evolution of the hydrogen economy policy in the 21st century - *CC BY 4.0*
    - 11.4.5.2: Methods of Green Hydrogen Production - *CC BY 4.0*
    - 11.4.5.3: The recovery of energy stored in the form of hydrogen - *CC BY 4.0*
    - 11.4.5.4: Hydrogen storage and transportation issues - *CC BY 4.0*
    - 11.4.5.5: Hydrogen as a “green” fuel suitable for sea- sonal storage - *CC BY 4.0*

Back Matter - *CC BY 4.0*

- Index - *CC BY 4.0*
- Glossary - *CC BY 4.0*
- Detailed Licensing - *Undeclared*
- Detailed Licensing - *Undeclared*