TLP LIBRARY I



Dissemination of IT for the Promotion of Materials Science (DoITPoMS)
University of Cambridge

LibreTexts™

TLP Library Vol. 1

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: Analysis of Deformation Processes

- 1.1: Introduction to Analysis of Deformation Processes
- 1.2: Lévy-Mises Equations
- 1.3: Slip Line Field Theory
- 1.4: Work Formula Method
- 1.5: Limit Analysis
- 1.6: Finite Element Method
- 1.7: Hencky Relations
- 1.8: Deeper Questions

2: Introduction to Anisotropy

- 2.1: Introduction
- 2.2: Mechanical analogy of anisotropic response
- 2.3: Anisotropic Thermal Conductivity
- 2.4: Derivation of the anisotropy ellipsoid
- 2.5: Anisotropic electrical conductivity
- 2.6: Anisotropic Diffusion
- 2.7: Anisotropic dielectric permittivity
- 2.8: Optical anisotropy and the optical indicatrix
- 2.9: Liquid crystals
- 2.10: Summary
- 2.11: Questions

3: Atomic Force Microscopy

- 3.1: Introduction
- 3.2: Tip Surface Interaction
- 3.3: Modes of Operation
- 3.4: The Scanner
- 3.5: Tip and Cantilever
- 3.6: Feedback
- 3.7: Scanner Related Artefacts
- 3.8: Tip Related Artefacts
- 3.9: Other Artefacts
- 3.10: Summary
- 3.11: Questions

4: Atomic Scale Structure of Materials

- 4.1: Introduction
- 4.2: Single crystals- Shape and anisotropy
- 4.3: Single crystals- Mechanical properties
- 4.4: Single crystals- Optical properties
- 4.5: Polycrystals

- 4.6: Defects
- 4.7: Summary
- 4.8: Questions

5: Avoidance of Crystallization in Biological Systems

- 5.1: Introduction
- 5.2: Nucleation and crystallization
- 5.3: The water-sucrose system
- 5.4: Avoidance of crystallization by glass formation
- 5.5: Avoidance of crystallization by freeze resistance
- 5.6: Summary
- 5.7: Questions

6: Batteries

- 6.1: Introduction
- 6.2: Basic principles
- 6.3: Thermodynamics and kinetics
- 6.4: Primary batteries
- 6.5: Zinc
  - 6.5.01: Zinc
    - 6.5.1: Zinc/carbon batteries
- 6.6: Alkaline
  - 6.6.01: Alkaline
    - 6.6.1: Alkaline/manganese oxide batteries
- 6.7: Questions
- 6.8: Zinc
  - 6.8.01: Zinc
    - 6.8.1: Zinc/silver oxide batteries
- 6.9: Secondary batteries
- 6.10: Lead
  - 6.10.01: Lead
    - 6.10.1: Lead/acid batteries
- 6.11: Lithium batteries
- 6.12: Battery characteristics
- 6.13: The future

7: Bending and Torsion of Beams

- 7.1: Introduction
- 7.2: Pole-vaulting
- 7.3: Bending moments and beam curvatures
- 7.4: Maximising the beam stiffness
- 7.5: Beam deflections from applied bending moments
- 7.6: Twisting moments (torques) and torsional stiffness
- 7.7: Springs
- 7.8: Plastic deformation during beam bending
- 7.9: Summary

- 7.10: Questions

8: Brillouin Zones

- 8.1: Introduction
- 8.2: Reciprocal lattice vectors
- 8.3: Brillouin Zone construction
- 8.4: The general case in three dimensions
- 8.5: Zone Folding
- 8.6: Brillouin Zones in Three Dimensions
- 8.7: Summary
- 8.8: Questions

9: Brittle Fracture

- 9.1: Introduction
- 9.2: When do atomic bonds break?
- 9.3: Why do cracks weaken a material?
- 9.4: Inglis and the crack tip stress idea
- 9.5: Can we calculate the energy changes?
- 9.6: What about tension?
- 9.7: Another way of expressing the energies
- 9.8: Another way of calculating the energies
- 9.8: Why bother if they are the same?
- 9.9: Coping with a scatter in strength
- 9.10: When does the sample fail completely?
- 9.11: Sub-critical crack growth and R-curves
- 9.12: Summary
- 9.13: Questions

10: Casting

- 10.1: Introduction
- 10.2: Heat Transfer
- 10.3: Heat Transfer Simulation
- 10.4: Microstructure and Segregation in Castings
- 10.5: Sand Casting
- 10.6: Die Casting
- 10.7: Continuous Casting
- 10.8: Investment Casting
- 10.9: Other Methods of Casting
- 10.10: Summary
- 10.11: Questions

11: Coating mechanics

- 11.1: Introduction
- 11.2: Misfit Strains
- 11.3: Force and Moment Balances
- 11.4: The Biaxial Modulus
- 11.5: The Stoney Equation - the Thin Coating Limit
- 11.6: Spallation (Interfacial Debonding)

- 11.7: Summary
- 11.8: Questions

12: Creep Deformation of Metals

- 12.1: Introduction
- 12.2: Creep Mechanisms
- 12.3: Constitutive Laws for Creep
- 12.4: Uniaxial Creep Testing - Practical Basics
- 12.5: Multiaxial Creep Testing – The Creeping Coil
- 12.6: Multiaxial Creep Testing – Indentation Creep Plastometry
- 12.7: Designing for Creep Resistance - Nickel Based Superalloys
- 12.8: Summary
- 12.9: Questions

13: Crystallinity in polymers

- 13.1: Introduction
- 13.2: Crystalline and Amorphous Polymers
- 13.3: Spherulites and Optical Properties
- 13.4: Polymer Stress-Strain Curve
- 13.5: Tensile Testing of Polyethylene
- 13.6: Summary
- 13.7: Questions

14: Crystallographic Texture

- 14.1: Introduction
- 14.2: Measurement of Texture
- 14.3: Representing Texture
- 14.4: Origins of Texture
- 14.5: Effect of Texture
- 14.6: Summary
- 14.7: Questions

15: Crystallography

- 15.1: Introduction
- 15.2: Lattices
- 15.3: Unit Cell
- 15.4: Lattice Geometry
- 15.5: Crystal Structure
- 15.6: Close Packing and Packing Efficiency
- 15.7: Symmetry
- 15.8: Combining symmetry
- 15.9: Crystal Systems
- 15.10: Summary
- 15.11: Questions

16: Deformation of honeycombs and foams

- 16.1: Introduction
- 16.2: Compression of a Honeycomb: Experimental

- 16.3: Elastic Behaviour
- 16.4: Yielding and Plateau Behaviour
- 16.5: Densification
- 16.6: Other Porous Structures
- 16.7: Porous Structures in Bending
- 16.8: Summary
- 16.9: Questions

17: Introduction to Deformation Processes

- 17.1: Introduction
- 17.2: Rolling
- 17.3: Forging
- 17.4: Extrusion
- 17.5: Drawing
- 17.6: Other Processes
- 17.7: Summary
- 17.8: Questions

18: Dielectric materials

- 18.1: Introduction
- 18.2: Electric Dipole
- 18.3: Polarisation Mechanisms
- 18.4: Capacitors
- 18.5: The Dielectric Constant
- 18.6: Variation of the Dielectric Constant in Alternating Fields
- 18.7: Effect of Structure on the Dielectric Constant
- 18.8: Effect of Temperature on the Dielectric Constant
- 18.9: Loss in Dielectrics
- 18.10: Dielectric Breakdown
- 18.11: Applications of Dielectrics
- 18.12: Summary
- 18.13: Questions

19: Diffraction and imaging

- 19.1: Introduction
- 19.2: Diffraction Patterns
- 19.3: Two-dimensional Diffraction
- 19.4: Image Formation
- 19.5: Information Obtained from Diffraction Pattern
- 19.6: Applications of the Theory of Optical Diffraction and Imaging
- 19.7: Summary
- 19.8: Questions

20: Diffusion

- 20.1: Fick's First Law of Diffusion
- 20.2: Fick's Second Law of Diffusion
- 20.3: Applications of Diffusion
- 20.4: Interdiffusion

- 20.5: Microstructural Effects
- 20.6: Temperature Effects

21: Dislocation Energetics and Mobility

- 21.1: Introduction
- 21.2: Making a Dislocation
- 21.3: Join the Crystals to Form the Dislocation
- 21.4: Dislocation Width
- 21.5: Form of the Displacement
- 21.6: Change in the Misfit Energy of a Dislocation as it Moves
- 21.7: Peierls Energy
- 21.8: What is the Peierls Stress?
- 21.9: Determining the Peierls Stress
- 21.10: Lattice Resistance
- 21.11: Uses and Limitations of the Atomistic Model
- 21.12: Summary
- 21.13: Questions

22: Introduction to Dislocations

- 22.1: Introduction
- 22.2: Dislocations in 2D
- 22.3: Bubble Raft
- 22.4: Dislocation Glide
- 22.5: Dislocation Motion
- 22.6: Dislocation in 3D
- 22.7: Observing Dislocations
- 22.8: Summary
- 22.9: Questions

23: Elasticity in Biological Materials

- 23.1: Introduction
- 23.2: Hookean Elasticity
- 23.3: S-shaped Curves
- 23.4: Aneurysms
- 23.5: Demonstration of Aneurysms
- 23.6: J-Shaped Curves
- 23.7: Viscoelasticity and Hysteresis
- 23.8: Summary
- 23.9: Questions

24: Electromigration

- 24.1: Introduction
- 24.2: The Theory of Electromigration
- 24.3: Electromigration Damage
- 24.4: Flux Divergence
- 24.5: Avoiding Electromigration Problems
- 24.6: Alternatives to Aluminium Metallization
- 24.7: Summary

- 24.8: Questions

25: Ellingham Diagrams

- 25.1: Introduction
- 25.2: Thermodynamics
- 25.3: Chemical Reactions
- 25.4: The Equilibrium Constant
- 25.5: The Ellingham Diagram
- 25.6: Application
- 25.7: Partial Pressure of Reacting Gas
- 25.8: Reading p_O2 from the Ellingham Diagram
- 25.9: Other Gas Mixtures
- 25.10: Reducing Agents
- 25.11: Other Atmospheres and Non-Standard States
- 25.12: Summary
- 25.13: The Interactive Ellingham Diagram
- 25.14: Questions

26: Expitaxial Growth

- 26.1: Introduction
- 26.2: The Simulation and its Limitations
- 26.3: Summary
- 26.4: Questions

27: Examination of a Manufactured Article

- 27.1: Introduction
- 27.2: Dismantling the Article
- 27.3: Classes of Materials
- 27.4: Metals
- 27.5: Polymers
- 27.6: Ceramics and Composites
- 27.7: Example Article
- 27.8: Summary
- 27.9: Questions

28: Ferroelectric Materials

- 28.1: Introduction
- 28.2: The Dipole Moment
- 28.3: Polarisation
- 28.4: Switching Polarisation
- 28.5: Measurement of Polarisation
- 28.6: Fabrication of a KNO_3 Ferroelectric Capacitor
- 28.7: Temperature Dependence of the Hysteresis Loop
- 28.8: Barium Titanate
- 28.9: Order of Phase Transitions
- 28.10: Ferroelectrics - Why?
- 28.11: Summary
- 28.12: Questions

29: Ferromagnetic Materials

- 29.11: Questions
- 29.1: Introduction
- 29.2: Types of Magnetism
- 29.3: Why are some materials magnetic?
- 29.4: Why are Only Some Materials Ferromagnetic
- 29.5: Curie-Weiss Law
- 29.6: Domains
- 29.7: Domain Walls
- 29.8: Hysteresis
- 29.9: Anisotropy
- 29.10: Summary

30: Finite Element Method

- 30.1: Introduction
- 30.2: Nodes, Elements, Degrees of Freedom and Boundary Conditions
- 30.3: Direct Stiffness Method and the Global Stiffness Matrix
- 30.4: Enforcing Boundary Conditions
- 30.5: Interpolation/Basis/Shape Functions
- 30.6: 1D First Order Shape Functions
- 30.7: 1D Second Order Shapes and Functions
- 30.8: Typical steps during FEM modelling
- 30.9: Summary
- 30.10: Questions

31: Fuel Cells

- 31.1: Introduction
- 31.22: Questions
- 31.2: The Principle
- 31.4: History of The Technology
- 31.4: Types of fuel cells
- 31.5: Solid Oxide Fuel Cells (SOFCs)
- 31.6: Electrolyte
- 31.7: Electrode Materials
- 31.8: Interconnection
- 31.9: Thermodynamics
- 31.10: System and Outlook
- 31.11: Molten carbonate fuel cells (MCFCs)
- 31.12: Electrolyte
- 31.13: Electrochemistry
- 31.14: Electrodes
- 31.15: Proton Exchange Membran Fuel Cells (PEMFCs)
- 31.16: Membrane
- 31.17: Electrodes and Membrane-Electrode Assembly
- 31.18: Efficiency and Reaction Conditions
- 31.19: Fuelling Requirements
- 31.20: Case Study: Fuelling Practicalities
- 31.21: Summary

32: The Glass Transition in Polymers

- 32.1: Introduction
- 32.2: Theory 1
- 32.3: Theory 2
- 32.4: Measurement of Tg
- 32.5: Demostrations
- 32.6: Summary
- 32.7: Questions

33: Granular Materials

- 33.1: Introduction
- 33.2: Terzaghi's Effective Stress Principle
- 33.3: Ground Settlement and Centrifuge Modelling Experiments
- 33.4: Angle of Repose
- 33.5: Principle of Normality
- 33.6: Yielding of Disturbed Saturated Soils in Triaxial Stress Tests
- 33.7: Original Cam-Clay Model
- 33.8: Liquefaction
- 33.9: Wet Sand Drying Underfoot
- 33.10: Summary
- 33.11: Questions

34: Indexing Electron Diffraction Patterns

- 34.1: Introduction
- 34.2: Mathematics
- 34.3: Laue Zones
- 34.4: Kikuchi Lines
- 34.5: Using Polycrystalline Materials in the TEM
- 34.6: Convergent Beam Electron Diffraction (CBED)
- 34.7: Using Other Methods in Conjunction with Electron Diffraction
- 34.8: Summary
- 34.9: Questions

35: The Jominy End Quench Test

- 35.1: Introduction
- 35.2: Hardenability
- 35.3: Jominy Test
- 35.4: Uses of Jominy Data
- 35.5: Example Test Data
- 35.6: Simulations
- 35.7: Summary
- 35.8: Questions

Index

Glossary

Detailed Licensing

Licensing

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

CHAPTER OVERVIEW

1: Analysis of Deformation Processes

https://www.doitpoms.ac.uk/tlplib/metal-forming-3/index.php

Learning Objectives

- Appreciate different approaches, including their inherent assumptions, that can be used to model plastic flow in metal forming operations.
- Gain an insight into these approaches and how they can provide estimates of the necessary deformation loads for different metal forming operations.
- Recognise the relative importance and limitations of the different approaches, ranging from simple work analyses to hodographs, and to become familiar with their use.

Before You Start

You should be familiar with the contents of the Theory of Metal Forming – Stress States and Yielding Criteria TLP and the concepts of stress and strain.

1.1: Introduction to Analysis of Deformation Processes
1.2: Lévy-Mises Equations
1.3: Slip Line Field Theory
1.4: Work Formula Method
1.5: Limit Analysis
1.6: Finite Element Method
1.7: Hencky Relations
1.8: Deeper Questions

This page titled 1: Analysis of Deformation Processes is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

1.1: Introduction to Analysis of Deformation Processes

After extracting metal from their ores and adding different elements to obtain a precise alloy optimised for a given end usage, the alloy often starts as a relatively large billet from which objects have to be fabricated. Hence, the large billets have to be reduced by mechanical deformation processes such as forging, rolling and extrusion, to reduce and change their shapes.

These processes are energy intensive and require expensive machinery. It would be inappropriate to over-design such machinery since that would be unnecessarily expensive, while under-designing would prevent the alloy from being deformed. Therefore, it is important to know the required loads needed to achieve the necessary deformations for different alloys.

This TLP examines various approaches that can be used to provide estimates of the loads (forces or stresses) required when deforming metallic objects. Some of the approaches are two-dimensional, and this introduces the concepts of plane strain and plane stress. In addition, the precise nature of the alloy will affect its mechanical behaviour and so the idea of homogeneous deformation is also included.

The approaches included in this TLP to estimate deformation loads are:

- Levy-Mises equations leading to plane stress and plane strain
- Slip line field theory
- Work formula
- Limit analysis and hodographs
- Finite element analysis

This module addresses how materials deform (change shape) when subjected to applied forces (stresses). Clearly, the nature of such deformations will depend on the class of material (metal, polymer, glass, etc.) as well as the precise microstructure of the specific material. In the following, the deformation is assumed to be homogeneous, i.e. it is independent of microstructural features such as grain size, dislocation density and defects. Thus the properties of the material are assumed to be isotropic.

This page titled 1.1: Introduction to Analysis of Deformation Processes is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

1.2: Lévy-Mises Equations

Once the yield criterion is satisfied, we can no longer expect to use the equations of elasticity. We must develop a theory to predict plastic strains from the imposed stresses.

When a body is subjected to stresses of sufficient magnitude, it will plastically deform (or fracture). The nature of the stresses depend on the particular forces applied to the body and, often, the same resulting deformation may be achieved by applying forces in different ways. For instance, a ductile metallic rod may be extended (elongated) a given amount either by a single force along its axis (i.e. a tensile stress) or by the combined action of several forces acting in different directions (i.e. multi-axial loading). A simple example of the latter multi-axial loading situation to obtain the same extension in the metallic rod as that obtained in pure tension is to apply a reduced tensile stress while simultaneously compressing the rod along its length. Under such multi-axial loading, the behaviour of ductile metallic materials can be described by the *Lévy-Mises equations*, which relate the principal components of strain increments during plastic deformation to the principal applied stresses.

In general, there will be both plastic (non-recoverable) and elastic (recoverable) strains.

However, to a first approximation, we can ignore the elastic strain assuming that the plastic strains will dominate in a deformation processing situation. We can therefore treat the material as a rigid-plastic, i.e. a material which is perfectly rigid prior to yielding and perfectly plastic afterwards.

Since plasticity is a form of flow, we can relate the strain rate, $\frac{\mathrm{d}\varepsilon}{\mathrm{dt}}$ to stress σ.

Plastic flow is similar to fluid flow, except that any rate of flow (strain rate) can occur for the same yield stress.

From symmetry we can show that in an isotropic body, the principal axes of stress and strain rate coincide, i.e. it goes the way you push it.

With respect to principal axes

$$\frac{\dot{\varepsilon}_1}{\sigma'_1} = \frac{\dot{\varepsilon}_2}{\sigma'_2} = \frac{\dot{\varepsilon}_3}{\sigma'_3} \tag{1.2.1}$$

,

where $\dot{\varepsilon}_i = \frac{\mathrm{d}\varepsilon}{\mathrm{dt}} (i = 1, 3)$, the normal strain rate parallel to i^{th} axis.

σ'_i = deviatoric component of normal stress parallel to the ith axis, and

$$\sigma'_1 = \sigma_1 - \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3) \tag{1.2.2}$$

If we consider small intervals of time δt, and call the resultant changes in strain $\delta\varepsilon_1$, $\delta\varepsilon_2$, $\delta\varepsilon_3$, it follows that,

$$\frac{\delta\varepsilon_1}{\sigma_1 - \frac{1}{2}(\sigma_2 + \sigma_3)} = \frac{\delta\varepsilon_2}{\sigma_2 - \frac{1}{2}(\sigma_3 + \sigma_1)} = \frac{\delta\varepsilon_3}{\sigma_3 - \frac{1}{2}(\sigma_1 + \sigma_2)} \tag{1.2.3}$$

the **Lévy-Mises equations**.

As $\frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3)$ is an invariant of the stress tensor, it also turns out that these equations apply even if stresses and strains are not referred to principal axes, so

$$\frac{\delta\varepsilon_{11}}{\sigma_{11} - \frac{1}{2}(\sigma_{22} + \sigma_{33})} = \frac{\delta\varepsilon_{22}}{\sigma_{22} - \frac{1}{2}(\sigma_{33} + \sigma_{11})} = \frac{\delta\varepsilon_{33}}{\sigma_{33} - \frac{1}{2}(\sigma_{11} + \sigma_{22})} \tag{1.2.4}$$

for a general stress tensor and plastic strain increments $\delta\varepsilon_{11}$, $\delta\varepsilon_{22}$ and $\delta\varepsilon_{33}$.

The above Lévy-Mises equations describe precisely the relationships between the normal stresses (arising from any general applied stress situation with respect to a particular set of orthogonal axes) and the resulting normal plastic strains (deformation) of a body referred to the same set of orthogonal axes. In many situations, the precise stresses are not known accurately and so more empirical approaches can be very helpful in describing the deformation of a body when subjected to applied forces. A number of these

approaches are considered in this TLP. However, several require further constraints, in particular the need to work in two dimensions and this introduces the concepts of **plane stress** and **plane strain**.

Plane Stress

In plane stress, one of the principal stresses is zero but there are three finite strains. An example of this is the surface of a thin walled, pressurized cylinder, where the principal stress normal to the surface has a value of zero. More generally, plane stress conditions occur in sheet metal forming when a thin sheet is subjected to uniaxial or biaxial tension.

Plastic Deformation in Plane Stress

Consider the uniaxial tensile behavior of a sheet (Figure 1.2.1).

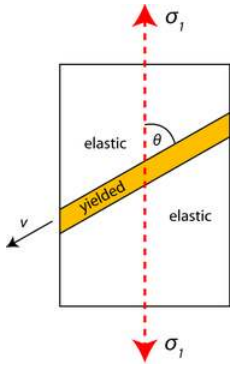


Figure 1.2.1*: Uniaxial Tensile Behavior of a Sheet*

Plastic flow will result once a critical stress is reached. Due to the constraint of neighboring elastic material, the plastically deforming material forms in a band across the sheet at a characteristic angle to the axis of loading.

$$\sigma_1 \neq 0, \sigma_2 = \sigma_3 = 0 \tag{1.2.5}$$

At the boundary between the elastic and the yielded material, longitudinal strains must match for continuity. Therefore, they must be zero since strain is effectively zero in the elastic regions.

Plastic strain along *v* is zero: $\delta\varepsilon_{vv} = 0$

From the Levy-Mises equations

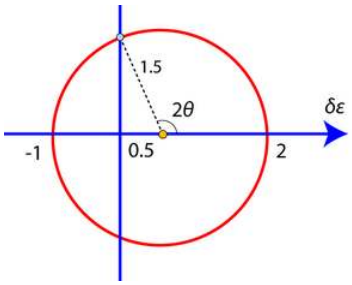
$$\frac{\delta\varepsilon_1}{\sigma_1} = \frac{\delta\varepsilon_2}{-1/2\sigma_1} = \frac{\delta\varepsilon_3}{-1/2\sigma_1} \tag{1.2.6}$$

and so $\delta\varepsilon_1 = -2\delta\varepsilon_2$ in the plane of the sheet.

Hence if we let $\delta\varepsilon_1 = +2$ units of plastic strain,

$\delta\varepsilon_2 = -1$ unit of plastic strain.

Therefore, on a Mohr's circle, we have:



and the longitudinal strain $\delta\varepsilon$ is zero, i.e. $\delta\varepsilon_{vv} = 0$, at an angle θ with respect to the direction parallel to σ_1.

From the diagram,

$$\cos(180-2\theta)=\frac{0.5}{1.5}=\frac{1}{3}$$
$$\Rightarrow \theta = 1/2\cos^{-1}(-1/3)=54.74^{\circ} \tag{1.2.7}$$

and so the longitudinal strain increment is $\delta\varepsilon_{vv}=0$at an angle of with respect to σ_1.

This phenomenon is well known in mild steel (Figure 1.2.2). The bands created are known as **Lüders bands**. These bands require less stress for their propagation than for their formation because of the freeing of dislocations from their solute atmospheres.

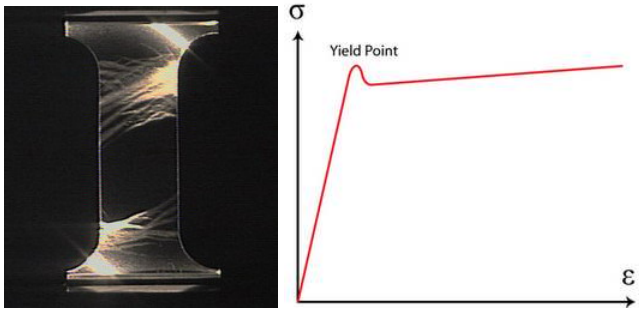


Figure 1.2.2*:* Lüders bands formation in steel, contributed by Mike Meier, University of California, Davis.

It is worth noting that Lüders bands occur in certain types of steel, such as low carbon steel (mild steel), but not in other metallic alloys, such as aluminum alloys or titanium alloys. This is because plastic strain localization is normally suppressed by work hardening, which tends to make plastic flow occur rather uniformly in a metal, particularly in the early stages of plastic flow, i.e., just after yield has taken place.

However, in certain types of low carbon steel at room temperature, Cottrell atmospheres of carbon atoms which have been able to segregate preferentially to dislocation cores pin dislocations until the upper yield point is reached. Once the upper yield point is reached, there is a load drop and then a sudden burst of plastic straining at a constant externally applied load, as cascades of dislocations are able escape their Cottrell atmospheres. This is clearly rather specialized behavior, caused by the ability of carbon atoms to diffuse relatively easily interstitially in these steels, but it is actually necessary behavior for the formation of Lüders bands.

Conventional work hardening in metallic alloys in the early stages of plastic deformation makes any strain localization (as demonstrated by the formation of Lüders bands) unlikely. This is also the case for pure metals, and for metals at high temperature, where large plastic strains can occur without a significant load increase once plastic deformation begins.

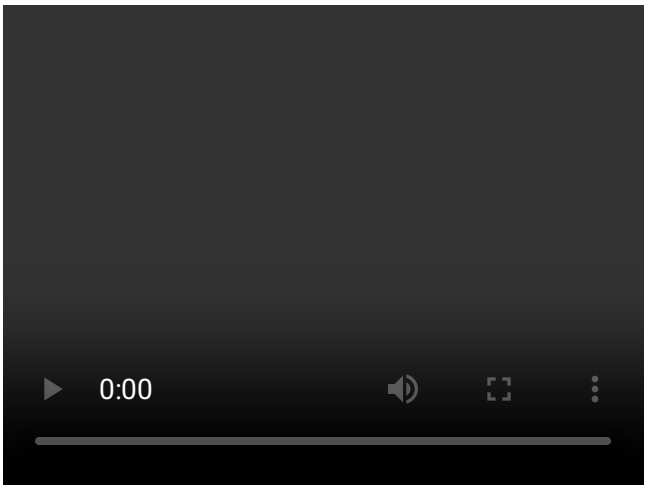
Therefore, Lüders bands only form if a limited burst of plastic straining is able to take place at constant load. Mild steels heated to sufficiently high temperatures (> 400 °C) and then tensile tested do not exhibit Lüders bands.

Plane Strain

Much deformation of practical interest occurs under a condition that is nearly, if not exactly, one of plane strain, i.e. where one principal strain (say ε_3) is zero so that $\delta\varepsilon_3$=0.

Plane strain is applicable to rolling, drawing and forging where flow in a particular direction is constrained by the geometry of the machinery, e.g. a well-lubricated die wall. A specific example of this is in rolling, where the major deformation occurs perpendicular to the roll axis. The material becomes thinner and longer but not wider. Frictional stresses parallel to the rolls (i.e. in the width direction) prevent deformation in this direction and hence a plane strain condition is produced where $\delta\varepsilon_3=0$. This can be seen in the animation below.

HOT ROLLING



Reproduced from Materials Selection and Processing CD, by A.M.Lovatt, H.R.Shercliff and P.J.Withers.

Plastic deformation in plane strain

Here, one principal strain is zero. Let this be ε_3. Then $\delta\varepsilon_3 = 0$.

From the Levy-Mises equation,

$$\frac{\delta\varepsilon_1}{\sigma_1 - \frac{1}{2}(\sigma_2 + \sigma_3)} = \frac{\delta\varepsilon_2}{\sigma_2 - \frac{1}{2}(\sigma_3 + \sigma_1)} = \frac{\delta\varepsilon_3}{\sigma_3 - \frac{1}{2}(\sigma_1 + \sigma_2)} \neq 0 \tag{1.2.8}$$

it follows that $\sigma_3 = \frac{1}{2}(\sigma_1 + \sigma_2)$ in order to avoid $\frac{\delta\varepsilon_1}{\sigma_1 - \frac{1}{2}(\sigma_2 + \sigma_3)} = 0$

Hence σ_3 is the mean of σ_1 and σ_2. By convention we define $\sigma_1 > \sigma_2$ $\sigma_1 > \sigma_3 > \sigma_2$. Therefore the maximum shear stress in the σ_1- σ_2 plane is at 45° to the axes and has magnitude σ1−σ22 .

If we now examine the Tresca and von Mises yield criteria, we find:

- Tresca $\frac{\sigma_1 - \sigma_2}{2} = k = \frac{Y}{2}$ (k = shear yield stress and Y = uniaxial yield stress)
- von Mises $(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 = 6k^2 = 2Y^2$

If

$$\sigma_3 = \frac{1}{2}(\sigma_1 + \sigma_2), \frac{3}{2}(\sigma_1 - \sigma_2)^2 = 6k^2 = 2Y^2 \tag{1.2.9}$$

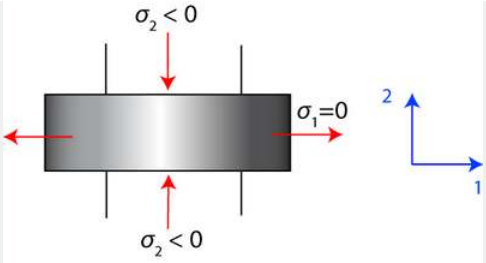
then

$$(\sigma_1 - \sigma_2) = 2k = \frac{2Y}{\sqrt{3}}. \tag{1.2.10}$$

Therefore, if we have plane strain, the Tresca yield criterion and the von Mises yield criterion have the same result expressed in terms of k. It is unnecessary to specify which criterion we are using, provided we use k.

Note

Consider a metal in uniaxial compression where plastic strain only takes place in the 1-2 plane. There is no friction between the work piece and the die faces. (To achieve this experimentally, a sample should be wide in the 3 direction).



$$\Rightarrow \sigma_{ij} = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \frac{\sigma_1+\sigma_2}{2} \end{pmatrix}$$

Hydrostatic stress:

$$\sigma_H = -p = \frac{\sigma_1 + \sigma_2}{2} = \sigma_3 \tag{1.2.11}$$

where p is the hydrostatic pressure.

So at yield, we have $\frac{\sigma_1-\sigma_2}{2} = k$ and since $\frac{\sigma_1+\sigma_2}{2} = -p$, we have

$$\sigma_1 = -p + k = 0 \tag{1.2.12}$$

$$\sigma_2 = -p - k = -2k \tag{1.2.13}$$

since $p = k$ at yield

$$\sigma_3 = -p \tag{1.2.14}$$

So for this example, the stress tensor is

$$\sigma_{ij} = \begin{pmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{pmatrix} + \begin{pmatrix} k & 0 & 0 \\ 0 & -k & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1.2.15}$$

which is the sum of hydrostatic stress (which can vary in magnitude through the object) and deviatoric pure shear stress (which has the same value throughout the material).

The directions of maximum shear therefore lie at 45° to σ_1 and σ_2. These are slip lines along which plastic flow occurs.

We are avoiding additional complexities such as work hardening by assuming the materials are rigid-plastic.

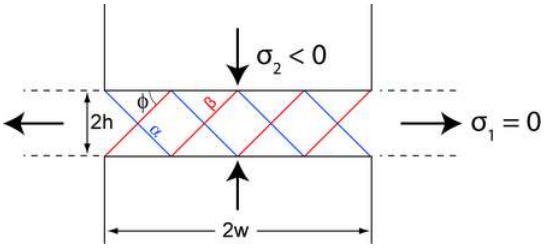
This page titled 1.2: Lévy-Mises Equations is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

1.3: Slip Line Field Theory

This approach is used to model plastic deformation in plane strain only for a solid that can be represented as a rigid-plastic body. Elasticity is not included and the loading has to be quasi-static. In terms of applications, the approach now has been largely superseded by finite element modelling, as this is not constrained in the same way and for which there are now many commercial packages designed for complex loading (including static and dynamic forces plus temperature variations). Nonetheless, slip line field theory can provide analytical solutions to a number of metal forming processes, and utilizes plots showing the directions of maximum shear stress in a rigid-plastic body which is deforming plastically in plane strain. These plots show anticipated patterns of plastic deformation from which the resulting stress and strain fields can be estimated.

The earlier analysis of plane strain plasticity in a simple case of uniaxial compression established the basis of slip line field theory, which enables the directions of plastic flow to be mapped out in plane strain plasticity problems.

There will always be two perpendicular directions of maximum shear stress in a plane. These generate two orthogonal families of slip lines called α-lines and β-lines. (Labelling convention for α and β lines.)



Experimentally, these lines can be seen in a realistic plastic deformation situation, e.g.

- Polyvinyl chloride (PVC) viewed between crossed polars
- Nitrogen-containing steels can be etched using Fry's reagent to reveal regions of plastic flow in samples such as notched bars and thick walled cylinders.
- Under dull red heat in forging, we see a distinct red cross caused by dissipation of mechanical energy on slip planes.

To develop slip line field theory to more general plane strain conditions, we need to recognize that the stress can vary from point to point.

Therefore, p can vary, but k is a material constant. As a result of this, directions of maximum shear stress and the directions of principal stresses can vary along a slip line.

Hencky relations

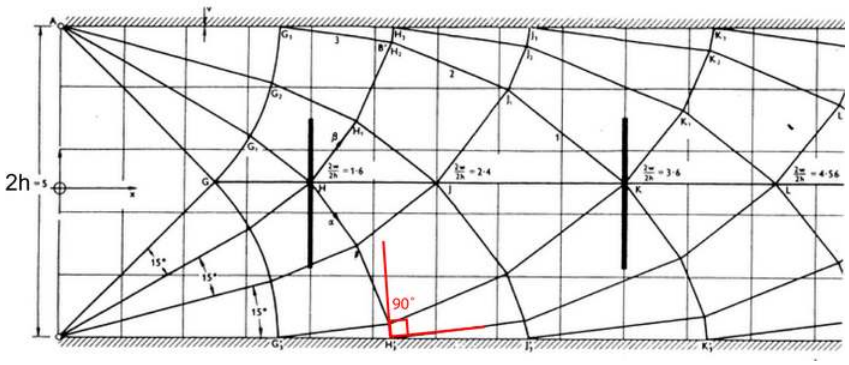
These equations arise from consideration of the equilibrium equations in plane strain.

$$p + 2k\varphi = \text{ constant along an } \alpha \text{ line}$$
$$p - 2k\varphi = \text{ constant along a } \beta \text{ line}$$

Click here for the derivation.

We can apply these relations to the classic problem of indentation of a material by a flat punch. This is important in hardness testing, in foundations of buildings and in forging.

In more general cases, slip lines do not intersect external boundaries at 45° because of friction. In the extreme case, *sticking friction* occurs (a perfectly rough surface) and slip lines are at 90° to the surface.



Slip-line field for compression between a pair of rough parallel platens [1]

The slip line patterns above are very useful for analyzing plane strain deformation in a rigid-plastic isotropic solid. Arrival at this slip-line pattern is rather complex. They are either derived from model experiments in which the slip-line field is apparent or they are postulated from experience of problems with similar geometry.

For a slip line to be a valid solution (but not necessarily a unique solution) the stress distribution throughout the whole body, not just in the plastic region, must not violate stress equilibrium nor must it violate the yield criterion outside the slip-line field.

The resultant velocity field must also be evaluated to ensure that strain compatibility is satisfied and matter is conserved.

These are stringent conditions and mean that obtaining a slip-line field solution is often not simple.

Instead, it is useful to take a more simple approach to analysing deformation processing operations where one or another of the stringent conditions are relaxed to give useful approximate solutions for part of the analysis.

This page titled 1.3: Slip Line Field Theory is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

1.4: Work Formula Method

When deforming a body, work has to be done by the applied forces. In the simplest case, the work done can be estimated from the magnitude of the applied stress(es) and the extent of the deformation. This is analogous to simple mechanics in which the work done is equal to force applied multiplied by the distance moved. Clearly, this simple approach assumes, in the first instance, that all the work done by the applied forces results in deformation; this can be called "useful" work. If it is assumed that all the work done is useful, then the work formula approach leads to an underestimate of the actual forces needed, i.e. a "lower bound". In other words, it would not be possible to deform the body if at least those forces were not applied and hence that work was not done. In practice, frictional forces need to be overcome and heating of the body occurs due to the internal micro-mechanisms of deformation. This non-useful work is called "redundant" work and estimates can be made to allow for this, and so better approximations for the necessary forces can be made.

Consider a uniaxial tensile deformation process:

In this process, $\sigma_1 = Y$, $\sigma_2 = \sigma_3 = 0$ and Y = uniaxial yield stress.



Suppose that at some instant, the bar is of length l and of cross-sectional area A

$$V = Al \tag{1.4.1}$$

If the bar extends by an amount δl, the increment of work done/unit volume, δw, is

$$\frac{F\delta l}{V} = \frac{YA\delta l}{Al} = \delta w$$

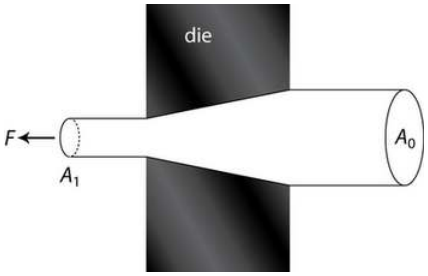
$$\delta w = \frac{Y\delta l}{l} \tag{1.4.2}$$

So if the bar extends from length l_0 to l_1, total work done/unit volume is

$$\int \delta w = Y\int_{l_0}^{l_1} \frac{\mathrm{d}l}{l} = Y\ln\frac{l_1}{l_0} \tag{1.4.3}$$

This can be applied to wire drawing.

In wire drawing, a tensile force is applied to the product rather than a compressive force applied to the billet (such as in extrusion for example).



To extend the wire by distance l_1, we have to feed in a length l_0 to the die where $A_1 l_1 = A_0 l_0$ by conservation of volume.

Work done by $F = Fl_1$

Volume of metal $= A_1 l_1$ drawn by l_1

$$\text{Work done} = YA_1 l_1 \ln\frac{l_1}{l_0} = Fl_1$$

$$\Rightarrow \frac{F}{A_1} = \sigma = Y\ln\frac{l_1}{l_0} = Y\ln\frac{A_0}{A_1} \tag{1.4.4}$$

where σ = stress required, the *drawing stress*.

Therefore we can estimate the maximum reduction possible with perfect lubrication.

If there is no work hardening, $\sigma \leq Y$ then the maximum reduction occurs when

$$\sigma = Y \Rightarrow \ln \frac{A_1}{A_0} = 1$$

$$\Rightarrow \frac{A_0}{A_1} = 2.72 \tag{1.4.5}$$

(e to 3 s.f.)

$$\Rightarrow \frac{A_1}{A_0} = 0.37 \tag{1.4.6}$$

63% is therefore the maximum possible reduction of cross-sectional area with a perfect wire-drawing operation for a rigid-plastic material. Work hardening causes this to increase to a slightly higher value. Friction between the wire and the die reduces this value, causing *redundant work* – work in excess of the minimum necessary to cause the deformation.

In practice, the best reduction obtainable with real dies and good lubrication is approximately 50%.

To allow for redundant work, we can apply empirical corrections, e.g. by using a value of efficiency, η:

$$\eta = \frac{\text{Work formula estimate}}{\text{Actual total work}} \tag{1.4.7}$$

Typical values of η are:

- extrusion $\eta = 45 – 55\%$
- wire drawing $\eta = 50 – 65\%$
- rolling $\eta = 60 – 80\%$

Clearly, the work formula method gives a lower band to the true force required for a given deformation processing operation because we are neglecting ‘redundant work.’

For metalworking, it is often preferable to have an overestimate of the load required in order to be sure that a given operation or process is possible.

This page titled 1.4: Work Formula Method is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

1.5: Limit Analysis

The concept of a lower bound has been introduced with reference to the work formula approach to analyse deformation. This approach generally results in an underestimate of the required load. Clearly, there also will be an “upper bound”, i.e. an overestimate of the load that needs to be applied to effect a given deformation. The two approaches together are called “limit analysis” since the actual loads required will lie between the lower and upper bounds. In practice, limit analysis is much easier to apply to a problem than the slip-line field approach and can be reasonably accurate. The upper bound is particularly useful for the study of metalworking processes in which it is essential to ensure sufficient forces are applied to cause the required deformation. In contrast, the lower bound is valuable in engineering where failure of a component must be avoided and hence an estimate of the minimum collapse load is needed.

The approach taken for estimating the upper bound is based on suggesting a likely deformation pattern, i.e. lines along which slip would be expected to occur for a given loading situation. Then the rate at which energy is dissipated by shear along these lines can be calculated and equated to the work done by an (unknown) external force. By refining the geometry of the deformation pattern, the minimum upper bound can be determined. Frictional forces can be accommodated in this approach. The approach utilises hodographs, which are self-consistent plots of velocity for different regions within a body being deformed; the different regions are assigned by considering how the overall body will deform for a particular deformation process and their relative velocities are estimated by assuming that the applied external force has unit velocity.

For both upper- and lower-bounds, one of the following two conditions has to be satisfied:

- *geometrical compatibility* between internal and external displacements or strains. This is usually concerned with kinetic conditions – velocities must be compatible to ensure no gain or loss of material at any point.
- *stress equilibrium* i.e. the internal stress fields must balance the externally applied stresses (forces).

The basis of limit analysis rests upon two theorems, which can be proved mathematically. In simple terms, these theorems are:

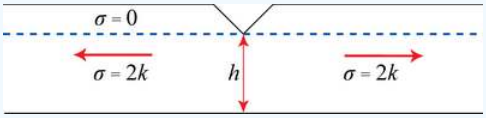
1. *Lower Bound:* any stress system in which the applied forces are just sufficient to cause yielding.
2. *Upper Bound:* Any velocity field that can operate is associated with an upper bound solution.

Example 1.5.1: Notched bar in tension

The plane strain condition is satisfied when breath, \(b » h\), the depth of the bar.

Lower-Bound:

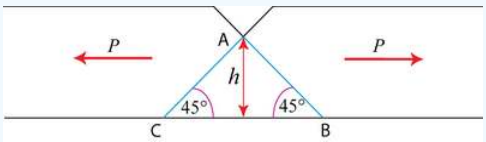
Find a stress system, e.g. $\sigma = 0$ in the length of the bar where there is the notch, \(\sigma = 2k\) elsewhere.



Therefore, for a breadth b, $P = 2khb = \text{load} = \text{stress} \times \text{area}$

Upper-Bound:

Postulate a suitable simple deformation pattern.



Assume yielding by slip on 45º shear planes with shear yield stress k. Let displacement along shear plane $\text{AB} = \delta x$.

Then internal work done = $k.|AB|b\delta x = k\sqrt{2}bh\delta x$, where the force is $k\,|AB|\,b$ acting on the shear plane AB.

Distance moved by the external load $P = \delta x \cos 45^\circ = \frac{\delta x}{\sqrt{2}}$

$$\Rightarrow P\frac{\delta x}{\sqrt{2}} = k\sqrt{2}bh\delta x \tag{1.5.1}$$

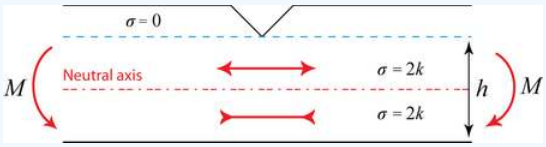
$$\Rightarrow P = 2kbh \tag{1.5.2}$$

So, here we obtain the same result for the upper bound and lower bound $\Rightarrow P = 2kbh$ is the true failure load, the load required to cause plastic flow.

Example 1.5.2: Notched bar in plane bending

Lower-Bound:

The area immediately under the notch, above the neutral axis is in tension $\sigma = 2k$. The area below the neutral axis is in compression $\sigma = 2k$.



where:

h = thickness of slab beneath the notch.

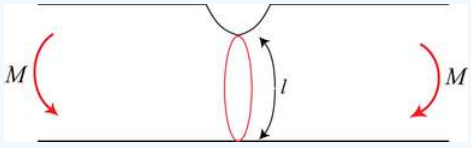
$2k.\frac{h}{2}\cdot b$ = magnitude of forces in tensile and compressive regions.
$\frac{h}{2}$ = distance between the two.

Equating the couples,

$$M = \left(2k.\frac{h}{2}\cdot b\right)\frac{h}{2} = 0.5kh^2b \tag{1.5.3}$$

Upper-Bound:

Assume failure occurs by sliding around a 'plastic hinge' along a circular arc of length *l* and radius *r*.



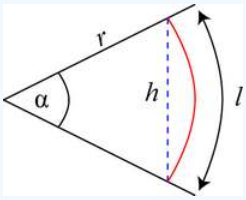
If the rotation is $\delta\theta$, the internal work done $= k.lb.r\delta\theta$ along one arc.

External work $= M\delta\theta$ by one moment.

$$\Rightarrow M = klbr \tag{1.5.4}$$

where no assumptions have been made regarding *l* and *r*.

The upper bound theorem states that whatever values are taken for *l* and *r* will lead to an upper bound. Clearly we wish to find the lowest possible value.



From the above geometry,

$$l = r\alpha \tag{1.5.5}$$

and

$$r = \frac{h}{2\sin\alpha/2} \tag{1.5.6}$$

$$\Rightarrow M = \frac{kh^2b}{4}\frac{\alpha}{\sin^2\alpha/2} \tag{1.5.7}$$

and so to find the lowest possible value of *M*, we minimize the function $\frac{\alpha}{\sin^2\alpha/2}$

Let

$$Y = \frac{\alpha}{\sin^2\alpha/2}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}\alpha} = \frac{1}{\sin^4\alpha/2}\left\{\sin^2\frac{\alpha}{2} - 2\frac{\alpha}{2}\cos\frac{\alpha}{2}\sin\frac{\alpha}{2}\right\}$$

$$= 0 \quad \text{when} \quad \sin\frac{\alpha}{2} = \alpha\cos\frac{\alpha}{2}$$

$$\Rightarrow \tan\frac{\alpha}{2} = \alpha$$

$$\Rightarrow M = \frac{kh^2b}{4}\cdot\frac{1}{\sin\alpha/2\cos\alpha/2} = 1/2\frac{kh^2b}{\sin\alpha} \cong 0.69kh^2b$$

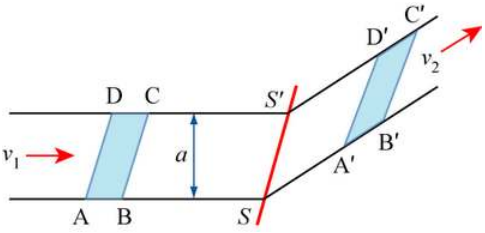
Taking the lower bound and the upper bound as limits, we therefore find

$$\Rightarrow 0.5 \leq \frac{M}{kh^2b} \leq 0.69$$

This forms a good example of constraining the value of the external force between lower bound and upper bound. It is also a good example of how to produce a lower limit on an upper bound calculation.

Hodographs I

A hodograph is a diagram showing the relative velocities of the various parts of a deformation process. To analyze a complicated deformation process with many shear planes, it is worth looking at the basic equation for the rate of energy dissipation in an upper bound situation in more detail.



ABCD is distorted into A'B'C'D' by shear along $\overrightarrow{SS'}$ at a velocity $\underline{v_s}$ in the metal.

Suppose ABCD moves towards the shear plane SS' at a velocity v_1 and suppose that there is a pressure P acting on the area al (where l is the dimension out of the plane of the paper) helping to cause this movement.

Rate of performance of work externally $= pal\left|\underline{v_1}\right|$

Rate of performance of work internally $= k\left|SS'\right| l\left|\underline{v_s}\right|$

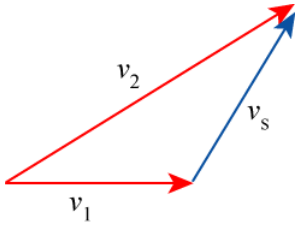
since the only internal work assumed to occur is that required to effect the shear deformation so that ABCD → A'B'C'D'.

Equating these,

$$\Rightarrow pa\left|\underline{v_1}\right| = k\left|SS^*\right|\left|\underline{v_s}\right\| \tag{1.5.8}$$

$$\Rightarrow pa = k\,|SS^*|\,\frac{|v_s|}{|v_1|} \tag{1.5.9}$$

Simple vector algebra relates v_1, v_2 and v_s as on the diagram below:



If in a deformation process there are n such shear planes of the type SS', then $pa = k\sum_n |SS_n^s|\,|v_{sn}|$ setting $\left|\underline{v_1}\right| = 1.0$, i.e. unit velocity.

Rules for constructing a hodograph

The animation below illustrates the seven rules for constructing a hodograph, for the case of a constrained punch.

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

An analysis of the geometry of the hodograph enables an upper bound for the applied force to be calculated.

Let Oq be a velocity vector of unit magnitude in the hodograph, i.e. $\nu_{Oq} = 1$
Due to the dead metal zone, Q and Q' move at the same velocity.
O is a stationary component of the system, anywhere in the surrounding perfectly rigid metal which has not yielded at all.

Oq and Oq' are in essence vectors defining the motion of particles in region Q'.
Or is velocity of a particle in region R.
q'r is a vector defining the shear velocity parallel to Q'R.
Os is velocity of a particle in region S.
rs is a vector defining the shear velocity parallel to RS.

Hence,

$$v_{Or} = \frac{1}{\tan\theta}, v_{q'r} = \frac{1}{\sin\theta} \text{ and } v_{Os} = v_{rs} = \frac{1}{2\sin\theta}$$

$\Rightarrow$ Using the rate of performance of work formula we have:

$$p\left(\frac{b}{2}\right) = k\{Q'Rv_{q'r} + ORv_{Or} + RSv_{rs} + OSv_{Os}\} \tag{1.5.10}$$

where Q'R is the length of the line dividing regions Q' and R, OR is the length of the line dividing regions O and R, RS is the length of the line dividing regions R and S and OS is the length of the line dividing regions O and S

$$p\left(\frac{b}{2}\right) = kb\left\{\frac{1}{2\cos\theta}\cdot\frac{1}{\sin\theta} + 1\cdot\frac{1}{\tan\theta} + \frac{1}{2\sin\theta}\cdot\frac{1}{2\sin\theta} + \frac{1}{2\cos\theta}\cdot\frac{1}{2\sin\theta}\right\}$$

$$= kb\left\{\frac{1}{\sin\theta\cos\theta} + \frac{\cos\theta}{\sin\theta}\right\} = kb\left\{\frac{1+\cos^2\theta}{\cos\theta\sin\theta}\right\}$$

$$\Rightarrow \frac{p}{2k} = \frac{1+\cos^2\theta}{\sin\theta\cos\theta} = f(\theta)$$

$\frac{df}{d\theta} = 0$ and \(f \)is then a minimum when

$$\cos 2\theta = -\frac{1}{3} \Rightarrow \theta = 54.74^\circ \text{ when } \sin\theta = \frac{\sqrt{2}}{\sqrt{3}} \text{ and } \cos\theta = \frac{1}{\sqrt{3}}$$

$\Rightarrow$ minimum $\frac{p}{2k} = 2\sqrt{2} = 2.83$ from this upper bound analysis.

When indenting using a sliding (frictionless) punch, we can postulate a different deformation pattern without the dead metal zone. The system also has a plane of symmetry and a hodograph can be constructed as follows:

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

As before, $O_q = 1.0$

Material in R travels in direction shown with velocity $Or = \frac{1}{\sin\theta}$

$$|Or| = |rs| = \frac{1}{\sin\theta} = |st| = |Ot|$$

$$|Os| = \frac{2\cos\theta}{\sin\theta}$$

$$|qr| = \frac{1}{\tan\theta} = \frac{\cos\theta}{\sin\theta}$$

Lengths in drawing of indent: $QR = SO = \frac{b}{2}$

$$OR = RS = ST = OT = \frac{b}{4\cos\theta} \tag{1.5.11}$$

We therefore have:

$$\frac{pb}{2} = k\{ORv_{Or} + RSv_{rs} + OSv_{Os} + STv_{st} + TOv_{tO}\}$$

$$= kb\left\{\frac{1}{4\cos\theta}\cdot\frac{1}{\sin\theta} + \frac{1}{4\cos\theta}\cdot\frac{1}{\sin\theta} + \frac{1}{2}\cdot\frac{2\cos\theta}{\sin\theta} + \frac{1}{4\cos\theta}\cdot\frac{1}{\sin\theta} + \frac{1}{4\cos\theta}\cdot\frac{1}{\sin\theta}\right\}$$

$$= kb\left\{\frac{1}{\sin\theta\cos\theta} + \frac{\cos\theta}{\sin\theta}\right\}$$

$\Rightarrow \frac{P}{2k} = \left\{\frac{1+\cos^2\theta}{\sin\theta\cos\theta}\right\}$ as before for the case of the constrained punch.

This analysis has assumed that no friction occurs at the punch face to cause particles in R to move parallel to OR. If there is friction, we can take it to be sticking friction, so that there is a shear stress k acting and slippage velocity $= v_{qr}$

$\Rightarrow$ in this case $\frac{P}{2k} = \left\{\frac{2+3\cos^2\theta}{2\sin\theta\cos\theta}\right\}$

$\Rightarrow$ of the three possible upper bound solutions, the 'best' answer is $\frac{P}{2k} = 2\sqrt{2} = 2.83$

This is a 10% overestimate of the true value of $\frac{P}{2k}$ found from slip-line field theory.

Hodographs II

Extrusion is an important working process. A simple form of extrusion used for non-ferrous metals involves a smooth square die.

We define extrusion ratio, *R* = ratio of areas.

$R = \frac{A_0}{A_1} = \frac{H}{h}$ for plane strain (*R* > 1), e.g. *R* = 4 = 75% reduction in area.

For a square die with sliding on the die face in plane strain, the hodograph can be constructed as follows:

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

Click here for a full mathematical analysis of this hodograph.

An alternative approach to an extrusion hodograph assumes there is a 'dead metal' zone.

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

Then

$$p\frac{H}{2} = k\left\{PQv_{PQ} + DQv_{dq} + QRv_{qr}\right\} \tag{1.5.12}$$

After similar algebra to the previous example, we obtain

$$\frac{p}{2k} = \frac{1}{2(\sin\varphi - \cos\varphi)}\left\{\frac{R+1}{\sin\varphi} - 2(R-1)\cos\varphi\right\} \tag{1.5.13}$$

Minimising $\cot\varphi = 1 - \frac{2}{\sqrt{R+1}}$

After more algebra, it is found that

$$\frac{p_{\min}}{2k} = 2(\sqrt{R+1} - 1) \tag{1.5.14}$$

Note that for low R (< 4) this value is less than that for sliding on the die face, even if the die face is frictionless.

⇒ For R < 4 this is a better upper bound solution for extrusion problems.

Click here for a full mathematical analysis of this hodograph.

This page titled 1.5: Limit Analysis is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

1.6: Finite Element Method

Finite element analysis (FEA) is an increasingly powerful computational technique in which the geometry of even a very complex body is represented by a mesh comprising a large number of discrete regions called finite elements. The elements are linked to each other at discrete points called nodes and different geometry elements (with varying numbers of nodes) can be selected for different types of problems. The use of FEA extends well beyond simple deformation since the approach can handle static and dynamic loading (with time-dependent changes in loads), two or three dimensional objects and different types of loading, e.g. thermal, electromagnetic as well as mechanical forces. Many commercial packages are now available.

The animation below depicts a finite-element simulation for the production of gudgeon pins. These are pins which hold a piston rod and a connecting rod together. The process consists of 4 stages [2]:

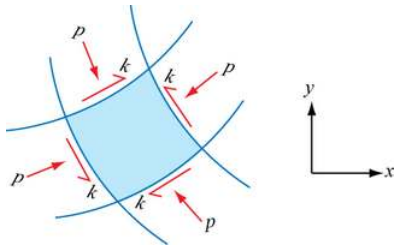
1. Upsetting
2. Indentation
3. Backward Extrusion
4. Punching – The base of the cup is punched out producing the gudgeon pin.

The simulation accounts for steps 1 to 3.

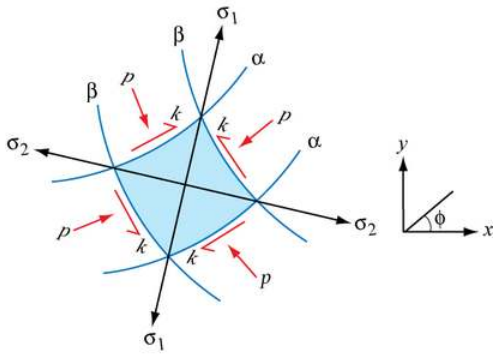
This page titled 1.6: Finite Element Method is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

1.7: Hencky Relations

Unless $\tau_{xy} = \pm k_i$, where k is the shear yield stress, the x- and y-directions (or axes) will not correspond to the directions of the $\alpha-$ and $\beta-$ slip lines, which are themselves at ± 45° to the directions of principal stresses acting on the element. In general, we need to examine the stresses on a small curvilinear element in the x - y plane upon which a shear stress and a hydrostatic stress are acting, and where the principal stresses are $-p–k$ and $-p + k$ for a situation where there is plane strain compression, such as in forging or indentation in which k is a constant but p can vary from point to point:



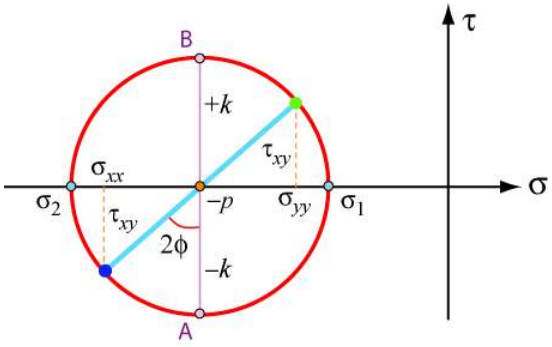
We can then identify on this diagram the directions of principal stress σ_1 and σ_2 (remembering that $\sigma_1 > \sigma_2$, and which of the lines are α-lines and which are β-lines. We can also specify the angle φ of the α-lines with respect to the x-axis:



Suppose that the α-slip line passing through the element makes an angle φ with respect to reference x- and y-axes, as in the above diagram. The β-slip line must then make an angle of 90° $+\varphi$ with respect to the *x*-axis, so that an anticlockwise rotation from the α-slip line to the β-slip line crosses the direction of maximum principal stress, σ_1.

The direction parallel to the principal stress σ_1 makes an angle of 45° $+\varphi$ with respect to the x-axis and the direction parallel to the principal stress σ_2 makes an angle of $135^\circ + \varphi^\circ 45^\circ - \varphi$ with respect to the x-axis.

On a Mohr's circle, this all looks like:



where A and B represent the stress states along the α- and β-lines respectively.

Hence, from the above,

$$\sigma_{xx} = -p - k \sin 2\varphi \tag{1.7.1}$$

$$\sigma_{yy} = -p + k \sin 2\varphi \tag{1.7.2}$$

$$\tau_{xy} = k \cos 2\varphi \tag{1.7.3}$$

while the tensile stress in the z-direction in plain strain plastic yielding is simply:

$$\frac{1}{2}\left(\sigma_{xx} + \sigma_{yy}\right) = -p \tag{1.7.4}$$

Substituting these expressions into the equilibrium equations

$$\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} = 0 \tag{1.7.5}$$

$$\frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} = 0 \tag{1.7.6}$$

and recognizing that k is a constant independent of x and y, we obtain two equations for p and φ as a function of x and y:

$$-\frac{\partial p}{\partial x} - 2k \cos 2\varphi \frac{\partial \varphi}{\partial x} - 2k \sin 2\varphi \frac{\partial \varphi}{\partial y} = 0 \tag{1.7.7}$$

$$-2k \sin 2\varphi \frac{\partial \varphi}{\partial x} - \frac{\partial p}{\partial y} + 2k \cos 2\varphi \frac{\partial \varphi}{\partial y} = 0 \tag{1.7.8}$$

For $\varphi = 0^{\circ}$, in which case the α and β lines coincide with the external x- and y-axes respectively at a particular position, these equations become

$$\frac{\partial}{\partial x}(p + 2k\varphi) = 0 \tag{1.7.9}$$

$$\frac{\partial}{\partial y}(p - 2k\varphi) = 0 \tag{1.7.10}$$

Integrating these equations we find

$$p + 2k\varphi = f_1(y) + C_1 \tag{1.7.11}$$

$$p - 2k\varphi = f_2(x) + C_2 \tag{1.7.12}$$

as the most general form of the solutions of these two partial differential equations. However, we know that when φ is exactly zero, p must have the same value in both Equations 1.7.11 and Eq. 1.7.12. Hence it follows that $f_1(x) = f_2(y) = 0$.

In general for points in a slip-line field we have therefore proved that the Hencky relations have to be satisfied:

Hencky relations

The hydrostatic pressure p varies linearly with the angle φ turned by a slip line.

- $p + 2k\varphi$ is constant along an α line
- $p - 2k\varphi$ is constant along a β line

where the angle φ is in radians.

This page titled 1.7: Hencky Relations is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

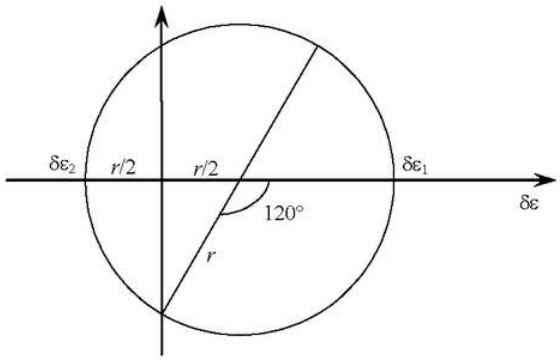
1.8: Deeper Questions

1. A specimen of sheet steel is tested in unequal biaxial tension, and Lüders bands form at 60° to one of the tensile axes. Show that the ratio between the two principal stresses in the plane of the sheet is 1:5.

If the greater of these two principal stresses is 500 MPa and the steel obeys von Mises' yield criterion, show that the yield stress in uniaxial tension is 458 MPa.

Answer

If the specimen is tested in unequal biaxial tension, we have stresses σ_1 and σ_2 which give rise to strain increments $\delta\varepsilon_1$ and $\delta\varepsilon_2$. As before the longitudinal plastic strain along the Lüders band has to be zero. We first use Mohr's circle to calculate the ratio of $\delta\varepsilon_1$ to $\delta\varepsilon_2$:



Now, cos 120° = $-\frac{1}{2}$, and so the centre of the Mohr's circle must be $\frac{r}{2}$ to the right of zero. Therefore, $\delta\varepsilon_2$ must be $\frac{r}{2}$ to the left of zero and $\delta\varepsilon_1$ must be $\frac{3r}{2}$ to the right of zero.

Hence, $\delta\varepsilon_1 = -3\delta\varepsilon_2$.

Using the Lévy-Mises equations, we have:

$$\frac{\delta\varepsilon_1}{\sigma_1 - \frac{1}{2}\sigma_2} = \frac{\delta\varepsilon_2}{\sigma_2 - \frac{1}{2}\sigma_1} \tag{1.8.1}$$

and so if $\delta\varepsilon_1 = -3\delta\varepsilon_2$, we have:

$$-3\left(\sigma_2 - \frac{1}{2}\sigma_1\right) = \left(\sigma_1 - \frac{1}{2}\sigma_2\right) \tag{1.8.2}$$

and so $\sigma_1 = 5\sigma_2$.

If σ_1 = 500 MPa, it follows that σ_2 = 100 MPa.

If Y is the yield stress in uniaxial tension, it follows from the von Mises yield criterion that

$$(\sigma_1 - \sigma_2)^2 + \sigma_2^2 + \sigma_1^2 = 2Y^2 \tag{1.8.3}$$

for this situation since If σ_3 = 0. Hence working in units of σ_2, we have:

$$16\sigma_2^2 + \sigma_2^2 + 25\sigma_2^2 = 2Y^2 \tag{1.8.4}$$

and so

$$Y = \sqrt{21}\sigma_2 = 4.58\sigma_2 = 458\text{MPa(3s.f.)} \tag{1.8.5}$$

2. Use a work formula approach to estimate the minimum pressure required to extrude aluminium curtain rail of I-section, 12 mm high with 6 mm wide flanges, all 1.6 mm thick, from 25 mm diameter bar stock.

[The mean uniaxial yield stress for aluminium for heavy deformation at room temperature is 150 MPa. The minimum pressure, p_{min}, required is given by the formula

$$p_{\min} = Y \ln\left(\frac{A_0}{A_1}\right) \tag{1.8.6}$$

where A_0 is the original cross-sectional area and A_1 is the cross-sectional area of the extruded I-section].

Answer

From the data given,

$$A_0 = \pi\left(\frac{25}{2}\right)^2 \text{mm}^2 \tag{1.8.7}$$

and

$$A_1 = (2 \times 6 \times 1.6) + ((12 - 2 \times 1.6) \times 1.6)\text{mm}^2 = 33.28\text{mm}^2 \tag{1.8.8}$$

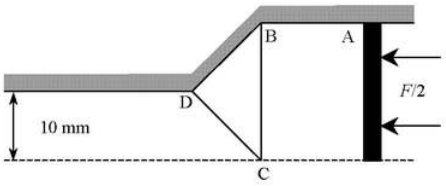
Hence,

$$R = \frac{A_0}{A_1} = 14.75 \tag{1.8.9}$$

Therefore,

$$p_{\min} = 150 \ln(14.75) = 404\text{MPa} \tag{1.8.10}$$

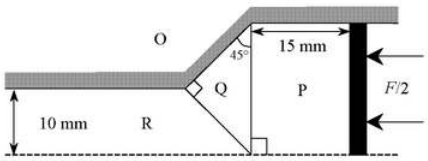
3. The diagram below shows a possible deformation pattern for the direct plane-strain extrusion of a metal slab, initially 40 mm thick, through a symmetrical 45° tapering die, with an extrusion ratio of 2.The diagram shows half the deformation pattern and the angles BCD and CBD are both 45°. The distance AB is 15 mm. The width of the slab is 100 mm and its yield stress in pure shear is 150 MPa.



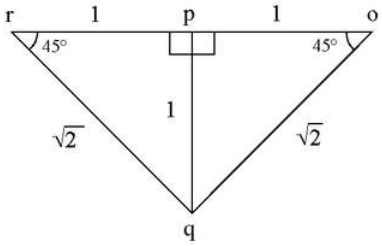
Calculate an upper bound to the extrusion force F acting on the slab if the extrusion process is frictionless.

Hint:

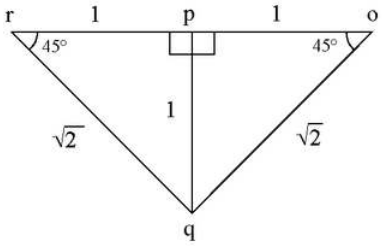
Begin by labelling the various regions of the diagram showing direct extrusion:



O is a stationary component of the system. The hodograph is straightforward for this special geometry:



Answer



If the extrusion process is frictionless, energy is only dissipated on PQ and QR, so applying the usual work formula, we have, per unit depth into the plane of the paper,

$$\frac{F}{2} \cdot \mathrm{op} = k(\mathrm{PQ} \cdot \mathrm{pq} + \mathrm{QR} \cdot \mathrm{qr}) \tag{1.8.11}$$

Now op = pq = 1 and qr = $\sqrt{2}$, PQ = 20 mm and QR = $10\sqrt{2}$ mm, so

$$\frac{F}{2} = k(20.1 + 10\sqrt{2}.\sqrt{2}) \times 10^{-3} = 40 \times 10^{-3}k \tag{1.8.12}$$

per unit depth, with lengths in metres.

Hence, for a 100 mm wide slab whose yield stress, k, in pure shear is 150 MPa,

$$F = \left(80 \times 10^{-3}\right) \times \left(150 \times 10^{6}\right) \times \left(100 \times 10^{-3}\right) \mathrm{N} = 1.2\mathrm{MN} \tag{1.8.13}$$

This page titled 1.8: Deeper Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

2: Introduction to Anisotropy

Learning Objectives

On completion of this TLP you should:

- Understand the concept of anisotropy, and appreciate that the *response* (e.g. displacement) need not be parallel to the *stimulus* (e.g. force)
- Understand the nature of anisotropic behaviour in a range of properties, including electrical and thermal conductivity, diffusion, dielectric permittivity and refractive index, and be aware of a range of everyday examples
- Be familiar with the use of representation surfaces

Before you start

There are no specific prerequisites for this TLP, but you will find it useful to have a basic knowledge of crystal structures, as this will enable a better understanding of the structural origins of anisotropy. Take a look at the Crystallography TLP and the Atomic Scale Structure of Materials TLP.

2.1: Introduction
2.2: Mechanical analogy of anisotropic response
2.3: Anisotropic Thermal Conductivity
2.4: Derivation of the anisotropy ellipsoid
2.5: Anisotropic electrical conductivity
2.6: Anisotropic Diffusion
2.7: Anisotropic dielectric permittivity
2.8: Optical anisotropy and the optical indicatrix
2.9: Liquid crystals
2.10: Summary
2.11: Questions

This page titled 2: Introduction to Anisotropy is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.1: Introduction

Some physical properties, such as the density or heat capacity of a material, have values independent of direction; they are *scalar* properties. However, in contrast, you will see that many properties vary with direction within a material. For example, thermal conductivity relates heat flow to temperature gradient, both of which need to be specified by direction as well as magnitude - they are *vector* quantities. Therefore thermal conductivity must be defined in relation to a direction in a crystal, and the magnitude of the thermal conductivity may be different in different directions.

A perfect crystal has long-range order in the arrangement of its atoms. A solid with no long-range order, such as a glass, is said to be *amorphous*. Macroscopically, every direction in an amorphous structure is equivalent to every other, due to the randomness of the long-range atomic arrangement. If a physical property relating two vectors were measured, it would not vary with orientation within the glass; i.e. an amorphous solid is *isotropic*. In contrast, crystalline materials are generally *anisotropic*, so the magnitude of many physical properties depends on direction in the crystal. For example, in an isotropic material, the heat flow will be in the same direction as the temperature gradient and the thermal conductivity is independent of direction. However, as will be demonstrated in this TLP, in an anisotropic material heat flow is no longer necessarily parallel to the temperature gradient, and as a result the thermal conductivity may be different in different directions.

The occurrence of anisotropy depends on the symmetry of the crystal structure. Cubic crystals are isotropic for many properties, including thermal and electrical conductivity, but crystals with lower symmetry (such as tetragonal or monoclinic) are anisotropic for those properties.

Many (but not all) physical properties can be described by mathematical quantities called *tensors*. A non-directional property, such as density or heat capacity, can be specified by a single number. This is a *scalar*, or *zero rank tensor*. Vector quantities, for which both magnitude and direction are required, such as temperature gradient, are *first rank tensors*. Properties relating two vectors, such as thermal conductivity, are *second rank tensors*. Third and higher rank tensor properties also exist, but will not be considered here, since the mathematical descriptions are more difficult.

This page titled 2.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.2: Mechanical analogy of anisotropic response

A *stimulus* (such as a force or an electrical field) does not necessarily induce a response (such as a displacement or a current) parallel to it.

This can be demonstrated with a simple mechanical model, consisting of a mass supported by two springs.

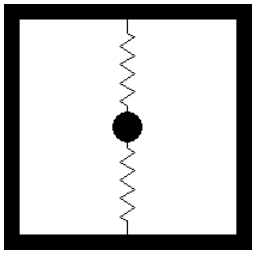


Figure 2.2.1: Mechanical model: a mass supported by two springs

A force, F, applied at an angle θ, to the central mass acts as the stimulus. The response is the displacement, r, of the mass, at an angle φ.

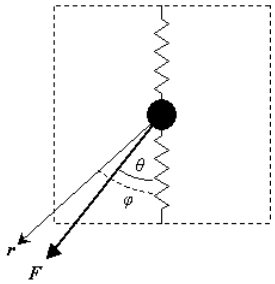
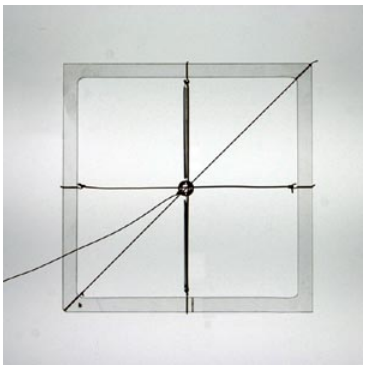


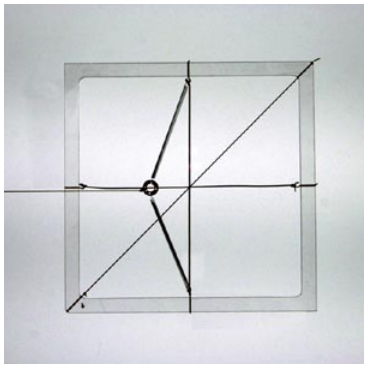
Figure 2.2.2: Diagram showing applied force and resulting displacement

For second rank tensor properties in anisotropic materials, parallel responses occur along orthogonal directions known as the *principal directions*.

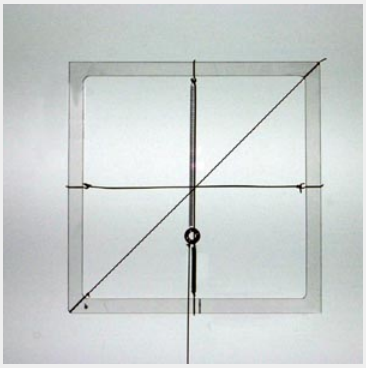
The following photographs show the response of the model under the application of various forces. (Click on an image to view a larger version.)



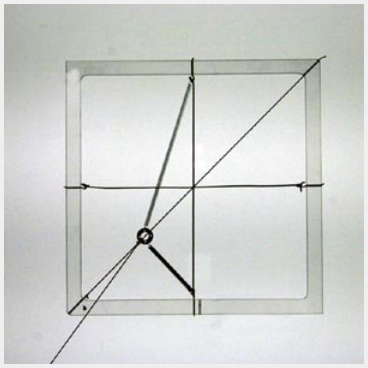
Model with no force applied



Model with horizontal force producing horizontal displacement (parallel response) ($\theta = \varphi = 90^{o}$)



Model with vertical force producing vertical displacement (parallel response) ($\theta = \varphi = 0^{o}$)



Model with 45° displacement from non-45° force (non-parallel anisotropic response) (θ = approx 35°, $\varphi = 45^{o}$)

Note that the displacement of the mass is only parallel to the force when the force acts parallel or perpendicular to the springs. These are the directions of the *principal axes*.

Symmetry

As a rule, the symmetry present in crystalline materials (such as mirror planes and rotational axes) determines or restricts the orientation of the principal axes. In this model, there exist two orthogonal mirror planes perpendicular to the plane of the model, one parallel and the other perpendicular to the springs, and a third mirror plane exists in the plane of the model. The principal axes lie along the intersections of these mirror planes. Real crystals typically show more complicated symmetry, but the orientation of the principal axes is still determined by the main symmetry elements.

Anisotropic properties may be analysed by resolving onto these principal axes.

The symmetry elements of any physical property of a crystal must include the symmetry elements of the point group of the crystal (Neumann's Principle). Thus crystals that, for example, display spontaneous polarization (see later section on anisotropic dielectric permittivity) can belong to only a few symmetry classes. It is worth noting that the absence of a centre of symmetry does not necessarily imply anisotropic second rank tensor properties, nor does the presence of a centre of symmetry rule out anisotropy in such properties.

This page titled 2.2: Mechanical analogy of anisotropic response is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.3: Anisotropic Thermal Conductivity

When a temperature gradient is present in a material, heat will always flow from the hotter to the colder region to achieve thermal equilibrium. As mentioned in the introduction, thermal conductivity is the property that relates heat flow to the temperature gradient. In an isotropic material:

$$J = k\frac{dT}{dr} \tag{2.3.1}$$

where J is the heat flow, k is the thermal conductivity, and $\frac{dT}{dr}$ is the temperature gradient.

Anisotropic thermal conductivity in quartz.

In quartz, perpendicular to the c-axis, the thermal conductivity is 6.5 $Wm^{-1}K^{-1}$. However, the thermal conductivity parallel to c is 11.3 $Wm^{-1}K^{-1}$.

The anisotropic thermal conductivity of quartz can easily be seen using a simple demonstration. Two sections cut from a quartz crystal, one perpendicular to the c-axis, and one parallel to it, are in turn mounted as shown in the diagram below. Pieces of plastic containing a heat sensitive liquid crystal are then glued to the top surfaces and the sections are heated from a point at their center, using a soldering iron. As the quartz heats up, the heat sensitive film changes color, which allows us to see how quickly the heat is conducted away from the center. The colours indicate contours of constant temperature.

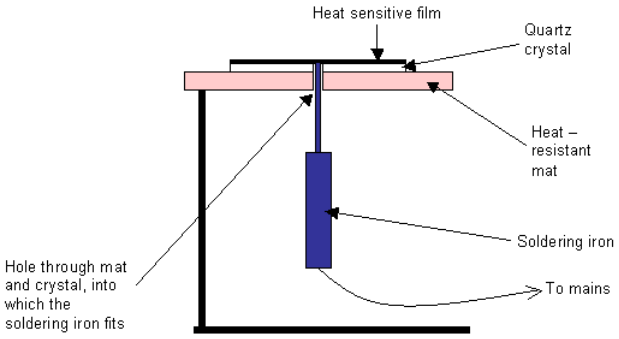


Diagram of experimental apparatus

When heating the section cut perpendicularly to the c-axis, the observed shape is a circle, showing that the thermal conductivity is the same in all directions in this plane. However, when using the section cut parallel to the c-axis, the shape seen is an ellipse, which shows that the thermal conductivity in this plane is direction-dependent.

https://www.doitpoms.ac.uk/tlplib/an...os/Quartz1.mp4

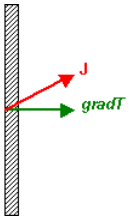
Video of a section of quartz cut perpendicular to the c axis being heated from a point at its center

https://www.doitpoms.ac.uk/tlplib/an...os/Quartz2.mp4

Video of a section of quartz cut parallel to the c axis being heated from a point at its center

The heat flow does not have to be parallel to the thermal gradient. A result of this can be seen by considering one-dimensional conduction in a long rod and a thin plate, both made of the same anisotropic material, arranged so that the normal to the plate and the length of the rod are oriented in an arbitrary general direction.

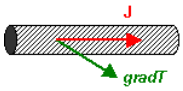
Thin Plate



Here the geometry of the set-up constrains the temperature gradient to be perpendicular to the plate. Due to the anisotropic nature of the material, the heat flux, **J**, will be in the direction shown, say. However, the thermal *conductivity* perpendicular to the plate is defined as the component of the heat flux parallel to the temperature gradient, $j_{\|}$, divided by the magnitude of that gradient. Thus:

$k_{\|} = \frac{j_{\|}}{gradT}$

Rod



Now the heat must flow along the rod, and the temperature gradient will be in a different direction, as shown. Here the thermal *resistivity* is defined as the component of the temperature gradient parallel to the rod, $gradT_{\|}$, divided by the magnitude of the heat flux. Thus:

$\rho_{\|} = \frac{\text{grad}\, T_{\|}}{J}$

where ρ is the resistivity.

It is important to realize that in anisotropic materials

$$\rho_{\|} \neq 1/k_{\|} \tag{2.3.2}$$

except along the principal axes. Only in isotropic materials is the resistivity always the reciprocal of the conductivity, and vice versa.

Note: By using a large, thin plate and a long rod, the effects of the alterations in the directions of heat flow and temperature gradient close to the edges (of the plate) or ends (of the rod) - "edge effects" and "end effects" - affect only a very small proportion of the sample and can be ignored.

This page titled 2.3: Anisotropic Thermal Conductivity is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.4: Derivation of the anisotropy ellipsoid

The variation of anisotropic properties such as conductivity can conveniently be illustrated by a "representation surface". In many cases this is an ellipsoid.

Suppose a three-dimensional temperature gradient, *gradT*, lies along a direction specified by direction cosines l, m and n>, where, for example, l is the cosine of the angle between the x-axis and the temperature gradient vector. Then the components of the temperature gradient parallel to the principal axes will be:

$$\operatorname{grad} T_x = (\operatorname{grad} T)\, \operatorname{lgrad} T_y = (\operatorname{grad} T)\, \operatorname{mgrad} T_z = (\operatorname{grad} T) n \tag{2.4.1}$$

The components of the heat flux are:

$$j_x = k_1 (\operatorname{grad} T) l j_y = k_2 (\operatorname{grad} T) m j_z = k_3 (\operatorname{grad} T) n \tag{2.4.2}$$

where k_1, k_2 and k_3 are the values of thermal conductivity along the principal axes, x, y and z, and are called the principal values.

Hence, resolving back along the direction of the temperature gradient, the heat flux is:

$$j_{\parallel} = j_x l + j_y m + j_z n = \left(k_1 l^2 + k_2 m^2 + k_3 n^2\right) gradT \tag{2.4.3}$$

Thus the value of the thermal conductivity, k_{lmn}, defined by

$$k_{lmn} = \frac{j_{\parallel}}{gradT} \tag{2.4.4}$$

is related to the principal values and the directional cosines by:

$$k = k_1 \cdot l^2 + k_2 \cdot m^2 + k_3 \cdot n^2 \tag{2.4.5}$$

$$l = x/r \quad m = y/r \quad n = z/r \tag{2.4.6}$$

Substituting in our equation for k gives:

$$k = \frac{k_1 x^2}{r^2} + \frac{k_2 y^2}{r^2} + \frac{k_3 z^2}{r^2} = \frac{1}{r^2}\left(k_1 x^2 + k_2 y^2 + k_3 z^2\right) \tag{2.4.7}$$

Setting

$$r = 1/\sqrt{k} \tag{2.4.8}$$

then

$$k >_1 x^2 + k_2 y^2 + k_3 z^2 = 1 \tag{2.4.9}$$

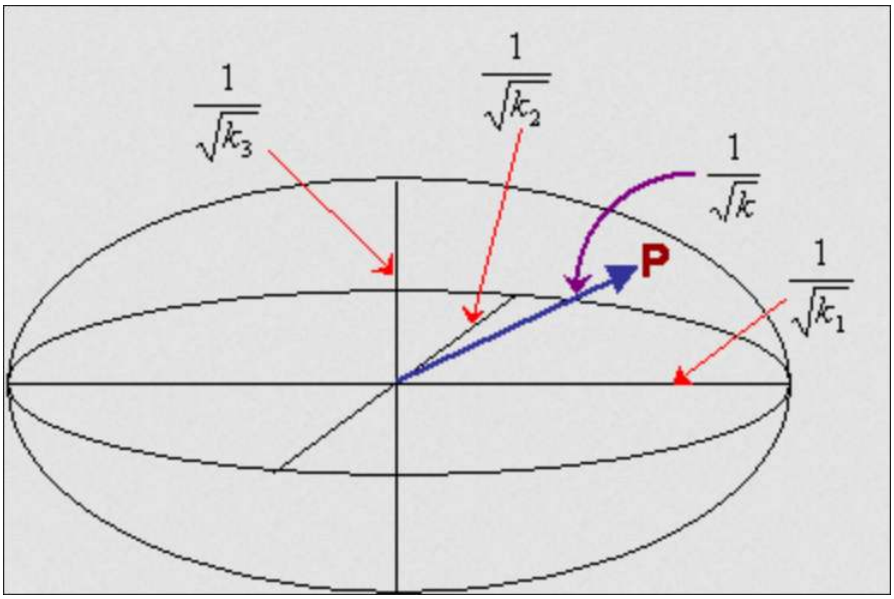
If all the principal values are positive (as they must be for thermal conductivity), then this equation describes the surface of an ellipsoid. The general equation of an ellipsoid (with semi-axes a,b,c) is:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{2.4.10}$$

Thus for this *representation ellipsoid*, the semi-axes are:

$$\frac{1}{\sqrt{k_1}}, \frac{1}{\sqrt{k_2}}, \frac{1}{\sqrt{k_3}} \tag{2.4.11}$$

The radius of this ellipsoid in a general direction is equal to the value of $\frac{1}{\sqrt{k}}$ in that direction. Thus the value of k in a particular direction - the ratio of the component of the heat flow in that direction to the magnitude of the temperature gradient in that direction - can easily be calculated from the radius in that direction.



An equivalent representation surface exists for electrical conductivity, and a similar representation surface exists for refractive index - the optical indicatrix. Both of these are discussed later in this TLP.

Using the representation surface for thermal conductivity

As shown above, the representation surface for thermal conductivity is an ellipsoid with semi-axes $\frac{1}{\sqrt{k_1}}$, $\frac{1}{\sqrt{k_2}}$, and $\frac{1}{\sqrt{k_3}}$. The distance between the centre of the ellipsoid and a point, P, on its surface, is equal to $\frac{1}{\sqrt{k}}$ at this point.

As well as determing the conductivity, the representation surface can be used to relate the directions of heat flow and the temperature gradient. If the temperature gradient is applied radially from the centre of the ellipsoid, then the direction of resulting heat flow is perpendicular to the tangential plane constructed from the point at which the thermal gradient meets the surface of the ellipsoid.

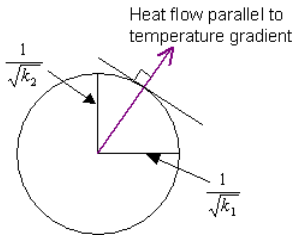
In an isotropic material, the representation surface will be a sphere, and the heat flow will be in the same direction as the temperature gradient. However, in an anisotropic material, heat flow is no longer necessarily parallel to the temperature gradient.

We will now revisit the **anisotropic thermal conductivity of quartz**.

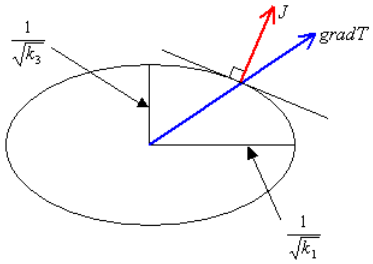
Because of the crystal symmetry of quartz, $k_1 = k_2 \neq k_3$, and so the representation surface for the thermal conductivity of quartz is a uniaxial ellipsoid of revolution. Depending on the relative values of k_1 and k_3, this is a shape either like a rugby ball ($k_3 < k_1$) or a Smartie ($k_3 > k_1$).

Consider sections of the ellipsoid:

a. Since $k_1 = k_2$, this section is a circle, and the direction of heat flow is parallel to the temperature gradient.



b. Since $k_1 \neq k_3$, this section is an ellipse, and the direction of heat flow is no longer parallel to the temperature gradient, except in the directions of the principal axes (which here correspond to the semi-axes of the ellipse).



The direction of the heat flux is always parallel to the normal to the tangential plane drawn at the point at which the direction of the temperature gradient intersects the representation ellipsoid. This is called the radius-normal property.

View proof of the radius-normal property.

This page titled 2.4: Derivation of the anisotropy ellipsoid is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.5: Anisotropic electrical conductivity

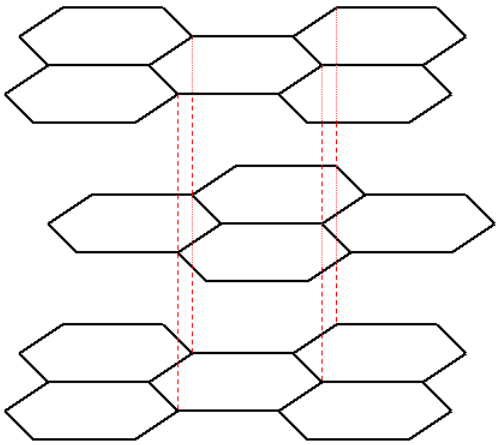
The current density, $\boldsymbol{J}$, is related to the electric field, $\boldsymbol{E}$, by

$$\boldsymbol{J} = \sigma \boldsymbol{E} \tag{2.5.1}$$

In an analogous way to thermal conductivity, the current density does not have to be parallel to the electric field. Three different examples of anisotropic electrical conductivity are described here. These show how the anisotropy is related to the crystal structure.

In metals, conductivity occurs by transport of delocalised electrons through the crystalline lattice, under the influence of an applied electric field. The conductivity is limited by the scattering of the electrons by imperfections in the periodicity of the structure (vibrations, impurities, etc). Because of the high symmetry in cubic metals, the overall drift velocity is parallel to the electric field, i.e. there is an isotropic response. However in hexagonally "close packed" metals, the nature of the symmetry in the crystalline array allows the conductivity to be anisotropic. For example in cadmium, it varies from 1.3 x 10^7 Sm^{-1} along the six-fold axis to 1.5 x 10^7 Sm^{-1} perpendicular to that axis.

Graphite consists of layered planes of carbon atoms with a structure as shown below. The layers are stacked above one another in a staggered fashion, the spacing between layers being about 2.3 times the distance between the adjacent carbon atoms in a layer.



The Hexagonal Structure of Graphite Planes

Here the hexagonal carbon rings provide the delocalised electrons, allowing easy conduction within the planes. Conduction is very much less perpendicular to the planes (around three orders of magnitude smaller) - this is highly anisotropic behaviour. The structure also creates anisotropy in other properties of graphite, such as thermal conductivity and thermal expansion.

This planar anisotropy is also seen in high temperature superconductors like BiSrCaCuO. The copper oxide "ab" planes provide superconducting pathways for electrons, but such pathways are not available perpendicular to the planes.

This page titled 2.5: Anisotropic electrical conductivity is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.6: Anisotropic Diffusion

The rate of diffusion of a specific atomic species is measured in terms of the coefficient of diffusion, *D*, which relates the flux of atoms (number crossing unit area in unit time) to the concentration gradient. In an isotropic material:

$$J = -D(gradc) \tag{2.6.1}$$

where J = flux (number per unit area per unit time) of an atomic species across a plane normal to the concentration gradient, *gradc*. The negative sign indicates that the flux is from high to low concentrations. For an isotropic material, such as an amorphous solid, the diffusion coefficient is independent of temperature.

Flux and concentration gradient are both vectors, so the coefficient of diffusion is a second rank tensor in an anisotropic material. The diffusion is anisotropic and can be described by three principal values of D, in the same way as thermal and electrical conductivity. (*Note*: we are not concerned here with the dependence of diffusion on temperature).

Example

Diffusion of Ni in Olivine, $(Mg,Fe)_2SiO_4$

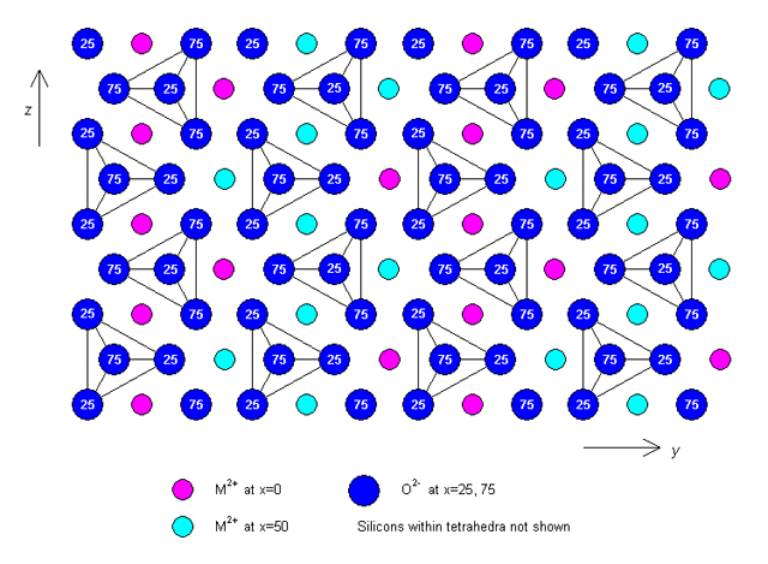
Olivine is the name for a series of minerals between two end members, fayalite (Fe_2SiO_4) and forsterite (Mg_2SiO_4). The two minerals form a solid solution where the iron and magnesium atoms can be substituted for each other without significantly changing the crystal structure. Olivine has an orthorhombic structure. The lattice parameters depend on the precise composition, but a typical set of values is: a = 0.49 nm, b = 1.04 nm, c = 0.61 nm.

The principal values of D for diffusion of nickel atoms in olivine also depend on the precise composition of the olivine. One set of values for an unspecified composition at 1423 K (1150ºC) is:

$$\begin{aligned} D_x &= 4.40 \times 10^{-18} \mathrm{m^2 s^{-1}} \\ D_y &= 3.35 \times 10^{-18} \mathrm{m^2 s^{-1}} \\ D_z &= 124.0 \times 10^{-18} \mathrm{m^2 s^{-1}} \end{aligned} \tag{2.6.2}$$

where the values correspond to the diffusion coefficients along the *x*, *y* and *z* crystallographic axes respectively. The exact values are unimportant for this discussion, but it is important to appreciate that diffusion occurs much faster parallel to the *z*-axis than in directions in the plane perpendicular to it. This happens because of the way in which the atoms are arranged in the crystal structure, a plan view of which is shown below (projected down the *x*-axis).

Note that x = 0, 25, 50 and 75 represent the *x*-coordinates of the atoms (as a percentage of the unit cell dimension *a*).



Crystal structure of olivine (click on image to view a larger version)

https://eng.libretexts.org/@go/page/7791

As you can see, there are chains of M^{2+} sites (where M^{2+} represents a metal ion, in this case either Mg or Fe) parallel to the *z*-axis. Diffusion occurs by Ni^{2+} substituting for M^{2+} along these chains, making diffusion in this direction much faster than in any other.

https://www.doitpoms.ac.uk/tlplib/an...os/Olivine.mp4

(Rotatable model of the olivine structure)

Fast ion conduction

When the structure of a solid material contains a large number of vacant sites (as a consequence of its composition), then it is likely to show high ionic mobility for some species of ion (sometimes an anion, sometimes a cation), even at modest temperatures. A high ionic mobility means that charge can be transferred very easily, and conductivities can approach those of aqueous electrolyte solutions or molten salts.

For a material to be a fast ion conductor, there should be:

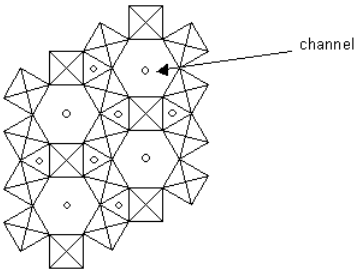
- a high concentration of charge carriers
- a high concentration of vacant sites in the structure
- a low activation energy for ionic migration

There are a number of different types of fast ion conductors. The material may be either a cationic or an anionic conductor, depending on the charge of the mobile ion. The material may also have a fully ionic structure, or the mobile ions may be in a covalent host structure. The dimensionality of the mobility can also vary: the ions may move through channels (1D), within layers of a structure (2D), or throughout the whole structure (3D).

Examples of fast ion conductors

1D fast ion conductors: Tungsten Bronzes, M_xWO_3

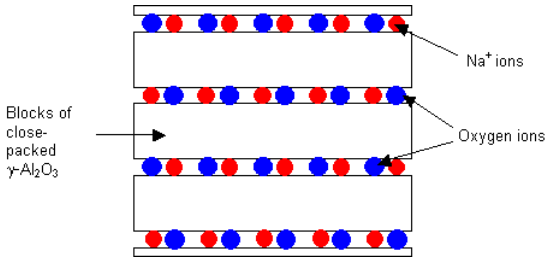
In these materials, WO_3 tetrahedra or octahedra form a covalent network. Usually either M^+ or M^{2+} (for example Na^+ or Cu^{2+}) are the mobile ions, and these move along channels in the structure. As a result, the conductivity is very high in this direction. An example of a tungsten bronze is shown below, projected along the *z*-axis, which is parallel to the channel direction.



Structure of tungsten bronze

2D fast ion conductor: Sodium beta-alumina, Na β-Al_2O_3

Sodium beta-alumina consists of blocks of γ-alumina (which has a spinel structure, the details of which need not be considered here) connected by a layer of bridging oxygen and sodium ions. Not all the Na^+ sites are occupied, and conduction occurs by the movement of the ions within this layer.



This page titled 2.6: Anisotropic Diffusion is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.7: Anisotropic dielectric permittivity

When an electric field, $\boldsymbol{E}$, is applied to a dielectric solid, positive and negative charges are displaced in opposite directions within the solid, creating *polarisation*, $\boldsymbol{P}$. This is defined as the net dipole moment per unit volume. (An electrical dipole is created by a small separation of equal and opposite charges.)

In an isotropic material, these vectors are related by:

$$\boldsymbol{P} = (e-1)e_0\boldsymbol{E} \tag{2.7.1}$$

e_0 is the permittivity of free space, and e is the relative dielectric permittivity (a scalar constant in this case). As with the other examples, in anisotropic materials this scalar has to be replaced by a tensor. Often the occurrence of highly anisotropic dielectric permittivity is associated with *ferroelecticity* (spontaneous polarisation reversible by an electric field) and *pyroelecticity* (temperature dependent generation of polarisation).

Example: barium titanate

The high temperature form of $BaTiO_3$ has the cubic perovskite structure with a primitive cubic lattice. At 150°C, a = 0.401 nm. In the temperature range 0ºC to 120ºC, $BaTiO_3$ is tetragonal. At 100°C it has $a = b = 0.400$ nm and $c = 0.404$ nm.

https://www.doitpoms.ac.uk/tlplib/an...vskite-vid.mp4

(Video of cubic perovskite structure)

https://www.doitpoms.ac.uk/tlplib/an...ovskite-vr.mp4

(Rotatable model of cubic perovskite structure)

https://www.doitpoms.ac.uk/tlplib/an...vskite-vid.mp4

(Video of tetragonal perovskite structure)

https://www.doitpoms.ac.uk/tlplib/an...ovskite-vr.mp4

(Rotatable model of tetragonal perovskite structure)

The tetragonal-cubic phase transition is highlighted in the following video. It shows a thin section of barium titanate viewed between crossed-polars, which is heated through the transition temperature, and then allowed to cool naturally. Initially, the sample is below the transition temperature, and since the domains of the anisotropic tetragonal phase exhibit birefringence, it is brightly coloured when viewed between crossed-polars. When the sample reaches the transition temperature, the isotropic cubic phase forms, which appears black. The heat source is then removed, so the sample cools down and again undergoes a phase transition to return to the anisotropic tetragonal phase.

https://www.doitpoms.ac.uk/tlplib/an...deos/BaTiO.mp4

(Video of barium titanate phase transition)

In the tetragonal form, the Ti ion is displaced by a small distance, from the centre of the surrounding octahedron of nearest neighbour oxygen ions, along the *z*-direction. A spontaneous polarisation along the *z*-axis is generated, but by symmetry there is no polarisation in the *x*-*y* plane. Note that the polarsation can be orientated forwards or backwards along the tetragonal axis. This polarisation is easily changed by applying an electric field parallel to the *z*-axis, but a field applied to the *x*-*y* plane has little effect on the polarisation. Consequently the dielectric permittivity is anisotropic.

The refractive index, *n*, is given by the square root of the relative dielectric permittivity, i.e. $\sqrt{\varepsilon}$. The resulting optical effects are considered in the next section.

This page titled 2.7: Anisotropic dielectric permittivity is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.8: Optical anisotropy and the optical indicatrix

In transparent materials with anisotropic dielectric permittivity, important optical effects can be observed. Recall that a light wave may be considered in terms of oscillating transverse electric and magnetic fields. Here we concentrate on the effects of the electric field. When discussing optical properties it is important to remember that this field is in a direction lying in the wavefront. It is *not* necessarily perpendicular to the direction of propagation. The interaction between the electric field and the material is governed by the dielectric permittivity discussed in the previous section. A large value of the permittivity gives rise to a large refractive index, and consequently the wave travels relatively slowly. (The refractive index n is related to the velocity of light in the medium, v, and the velocity in a vacuum, c, by n = c/v.)

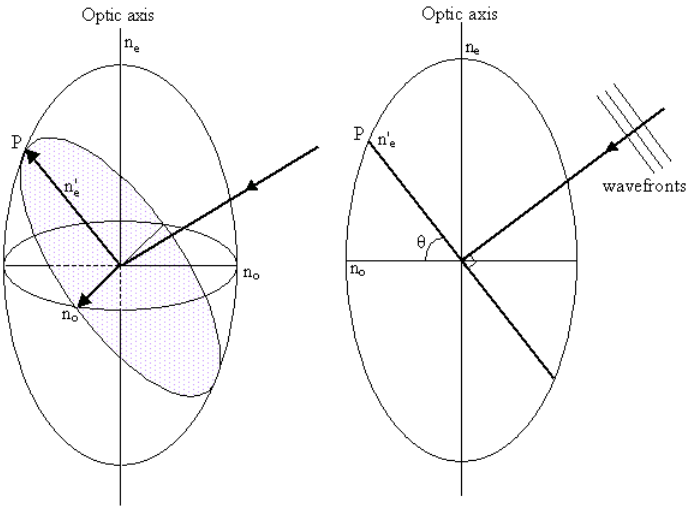
In an anisotropic material the refractive indices can again be illustrated by a representation surface - the *optical indicatrix*. For each (orthogonal) principal direction in the anisotropic material, there is an associated principal refractive index. The variation of the refractive index with the plane of the wavefront can be represented by an ellipsoid. The semi-axes of this optical indicatrix are directly proportional to the principal refractive indices.

Optically isotropic materials (e.g. cubic crystals) have one refractive index, with a spherical indicatrix.

Crystals with one 3, 4 or 6 fold axis of symmetry have a principal axis of the ellipsoid along this symmetry axis. These *uniaxial* crystals have an indicatrix which is an ellipsoid with a circular cross-section perpendicular to the major symmetry axis – an ellipsoid of revolution. They have two principal refractive indices and one *optic axis* (parallel to the symmetry axis and so perpendicular to the circular section).

In general, the electric field of a light wave experiences two *permitted vibration directions*, known as the fast and slow directions, both in the plane of the wavefront, and determined by the shape of the indicatrix. Consider a section passing through the origin of the indicatrix for a uniaxial crystal, and orientated parallel to the wavefronts, as shown by the dotted line in the left-hand figure below. The two permitted vibration directions are given by the major and minor axes of this section. The corresponding refractive indices are the lengths of these axes. The section will be elliptical unless the light is travelling along the optic axis so that the plane of the wavefront coincides with the circular section of the ellipsoid.

The observation of two refractive indices for a general orientation of the wavefront is known as *birefringence*. Related effects such as stress-induced birefringence and photoelasticity are discussed in the Photoelasticity TLP.



The *ordinary* vibration direction lies in the circular section of the indicatrix (i.e. perpendicular to the optic axis) with refractive index n_o. Light travelling along the optic axis experiences just this refractive index - the *ordinary refractive index*.

The *extraordinary* vibration direction lies in the plane of the wavefront and perpendicular to the ordinary vibration direction, and has refractive index n'_e. The value of n'_e is determined from the ordinary refractive index and the principal extraordinary refractive index n_e, as follows.

Consider a cross-section of the indicatrix (as shown in the diagram on the right above), containing the optic axis and the extraordinary vibration direction. The equation for this ellipse will be:

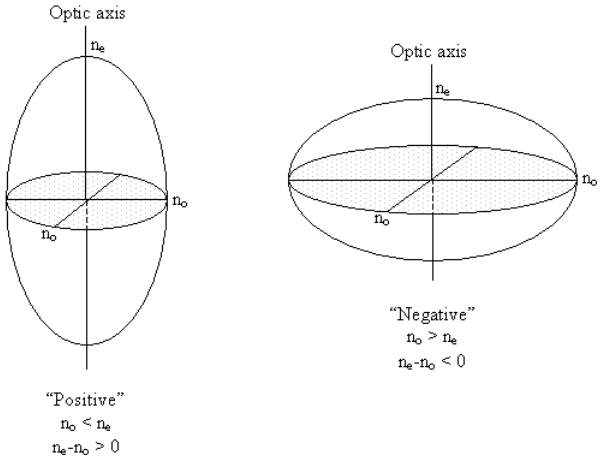
$$\frac{x^2}{n_o^2} + \frac{z^2}{n_e^2} = 1 \tag{2.8.1}$$

For the point P, x = n_e'cosθ, and y = n_e'sinθ. Therefore

$$\frac{(n_e' \cos\theta)^2}{n_o^2} + \frac{(n_e' \sin\theta)^2}{n_e^2} = 1 \tag{2.8.2}$$

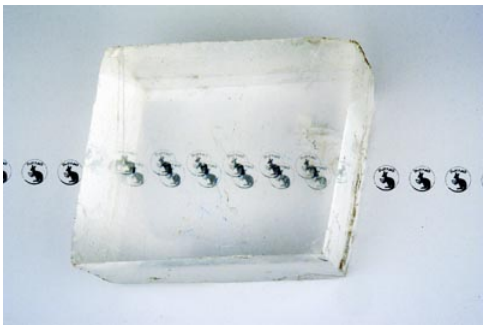
For a general extraordinary wave, the direction in which the light energy travels, the ray direction, is no longer perpendicular to the wavefront. For an explanation see, for example, the link to Optical Birefringence in Going further.

As a side note, the relative magnitudes of n_o and n_e determine whether a material is defined as optically positive or negative, the optical sign.



Example: calcite rhomb

The birefringence (defined as $|n_0 - n_e|$) in calcite is so large that two images can easily be observed when viewing an object through a suitable crystal with the naked eye. One of which is due to the ordinary wave (with electric field vibrating parallel to the ordinary vibration direction) and the other is due to the extraordinary wave (with electric field vibrating parallel to the extraordinary vibration direction).



The DoITPoMS logo viewed through a calcite rhomb (click on image to view larger version)

In calcite, the planar carbonate groups all lie in planes normal to the three-fold axis - the optic axis. The groups are well separated in the direction of the axis. This makes the crystal less polarisable parallel to the axis, so the refractive index for vibrations parallel to the triad axis is smaller than for vibrations perpendicular to it (making the crystal optically negative: $n_e < n_o$).

https://www.doitpoms.ac.uk/tlplib/an...os/Calcite.mp4

(Video of calcite structure)

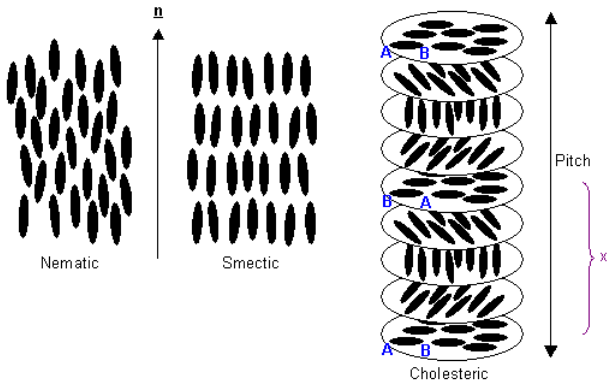
This page titled 2.8: Optical anisotropy and the optical indicatrix is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.9: Liquid crystals

When most solids melt, they form an isotropic fluid, whose optical, electrical and magnetic properties do not depend on direction. However, when some materials melt, over a limited temperature range they form a fluid that exhibits anisotropic properties. These materials generally consist of organic molecules that have an elongated shape, with a rigid central region and flexible ends. The molecules in a *liquid crystal* do not necessarily exhibit any positional order, but they do possess a degree of orientational order.

The anisotropic behaviour of liquid crystals is caused by the elongated shape of the molecules. The physical properties of the molecules are different when measured parallel or perpendicular to their length, and residual alignment of the rods in the fluid leads to anisotropic bulk properties. This residual alignment occurs as a result of preferential packing arrangements, and also electrostatic interactions between molecules that are most favourable (lowest in energy) in aligned configurations.

There are three types of liquid crystal: nematic, smectic and cholesteric. In the liquid crystalline phase, the vector about which the molecules are preferentially oriented, **n**, is known as the "director". The long axes of the molecules will tend to align in this direction.

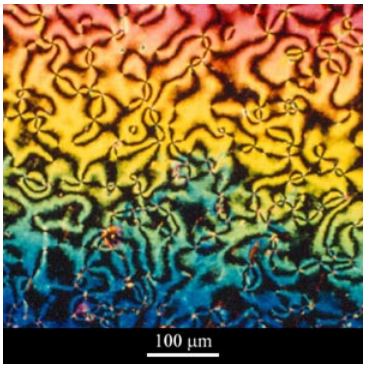


Three types of liquid crystal

In addition to the long range orientational order of nematic liquid crystals, smectic liquid crystals also have one dimensional long range positional order, the molecules being arranged into layers.

A cholesteric (or twisted nematic) liquid crystal is chiral: the molecules have left or right handedness. When the molecules align in layers, this causes the director orientation to rotate slightly between the layers, eventually bringing the molecules back into the original orientation. The distance required to achieve this is known as the *pitch* of the twisted nematic, as seen in the diagram above. The pitch is not equal to the distance marked x, because only 180º of rotation occurs over this length, so the molecules are aligned antiparallel to their starting orientation.

When viewed between crossed polars, thin films (approximately 10µm thick) of liquid crystals exhibit *schlieren textures*, as seen in the micrograph below, which shows a nematic liquid crystalline polymer.



Micrograph of nematic liquid crystalline polymer, courtesy of Professor TW Clyne and the DoITPoMS Micrograph Library (click on image to view larger version, or view full library record for the micrograph)

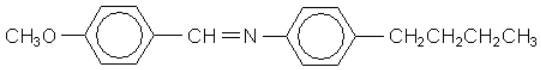
The black brushes are regions where the director is either parallel or perpendicular to the plane of polarisation of the incident radiation, and the points at which the brushes meet are known as disclinations.

If the temperature of a liquid crystal is raised, the constituent molecules have more energy, and are able to move and rotate more, so the liquid crystal becomes less ordered. As a result, the magnitude of the anisotropy of the bulk properties of the liquid crystal decreases, usually eventually resulting in an isotropic fluid.

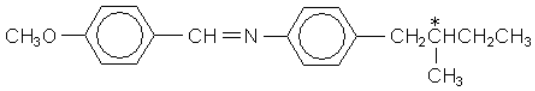
Liquid crystals are used in many different applications, for example the displays on calculators, digital watches and mobile phones.

examples of molecules that form liquid crystals

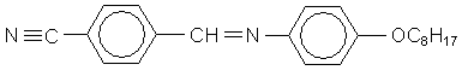
4-methoxylbenzylidene-4'-butylaniline (MBBA) transforms from crystalline to nematic liquid crystal at 20ºC, and from nematic to an isotropic liquid at 74ºC



This very similar molecule forms a chiral nematic (cholesteric) liquid crystal. The chiral centre is marked with a *.



4-cyanobenzylidene-4'-n-octyloxyanaline (CBOOA) forms a smectic liquid crystal below 83ºC and a nematic liquid crystal between 83ºC and 109ºC.



This page titled 2.9: Liquid crystals is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.10: Summary

In some materials a property will be the same, irrespective of the direction in which it is measured, but this is not always the case. On completion of this TLP, you should now understand the concept of anisotropy, and be able to appreciate that a response can be non-parallel to the applied stimulus. Anisotropy in a range of properties has been discussed, including electrical and thermal conductivity, diffusion, dielectric permittivity, and optical properties. You should also now be familiar with the use of representation surfaces for a range of anisotropic properties, including the basis behind their mathematical description.

Anisotropic properties are exploited in many applications. In polarised-light microscopy, a quartz wedge can be used to determine birefringence and optical sign. Liquid crystals have electronic uses such as displays, and the liquid crystalline state has advantages in the processing of polymers (such as Kevlar). The anisotropic thermal conductivity in polymer thin films has use in microelectronic devices, for example, solid-state transducers.

Anisotropic properties described by higher than second rank tensors (not discussed here) can also have useful applications. Examples include:

- Piezoelectricity (relating an applied stress to the induced polarisation)
- The electro-optic effect (when a field causes a change in the dielectric impermeability)
- Elastic compliance and elastic stiffness (relating stress and strain)
- Piezo-optical effect (when a stress induces a change in refractive index)
- Electrostriction (strain arising from an electric field)

Non-tensor properties can also demonstrate anisotropy; for example, yield stress can vary with direction of applied stress.

Going further

Books

- A. Putnis, *Introduction to Mineral Sciences*, CUP, 1992 (specifically Chapter 2, "Anisotropy and physical properties").
- R.E. Newnham, *Structure-Property Relations Relations (Crystal chemistry of non-metallic materials)*, Springer-Verlag, 1975
- D.R. Lovett, *Tensor Properties of Crystals*, IOP, 1999
- P.J. Collings and M. Hird, *Introduction to Liquid Crystals: Chemistry and Physics*, Taylor & Francis, 1997

More advanced and detailed books:

- C. Kittel, *Introduction to Solid State Physics*, 7th Edition 1995
- J.F. Nye, *Physical Properties of Crystals: Their Representation by Tensors and Matrices*, Oxford, 2nd Edition 1985
- E. Hecht, *Optics*, Addison-Wesley, 4th Edition 2001

Websites

- Polymers and Liquid Crystals
  An award-winnning website based at Case Western Reserve University in the USA, with a Virtual Textbook and Virtual Laboratory.
- Introduction to Photoelasticity
  A TLP covering many features of birefringence under polarised light.
- Theoretical methods for Mineral Sciences research
  Summary of phase transitions and formation of domains in perovskites.
- Optical Birefringence
  This comprehensive introduction to optical birefringence is part of the excellent award-winning Molecular Expressions website based at Florida State University in the USA.
- Birefringence in calcite crystals
  Also part of the Molecular Expressions website, this is an interactive Java tutorial.

For some light relief, take a look at:

- The Optical Indicatrix - And Why I Love It So Much

This page titled 2.10: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

2.11: Questions

1. Which of these properties of a crystal may be anisotropic?

a. Density

b. Young's modulus

c. Surface energy

d. Refractive index

e. Electrical conductivity

f. Thermal conductivity

g. Heat Capacity

h. Melting point

i. Coefficient of thermal expansion

Answer

a. No. Density is a *scalar* property so cannot be anisotropic.

b. Yes. Young's modulus relates *vector* quantities so can be anisotropic.

c. Yes. Surface energy relates *vector* quantities so can be anisotropic.

d. Yes. Refractive index relates *vector* quantities so can be anisotropic.

e. Yes. Electrical conductivity relates *vector* quantities so can be anisotropic.

f. Yes. Thermal conductivity relates *vector* quantities so can be anisotropic.

g. No. Thermal conductivity relates *vector* quantities so can be anisotropic.

h. No. Melting point is a *scalar* property so cannot be anisotropic.

i. Yes. Coefficient of thermal expansion relates *vector* quantities so can be anisotropic.

2. Two similar transparent uniaxial crystals show the same (principal) extraordinary refractive index, n_e. However one is optically positive, and the other is optically negative. In which will the light travel faster along the optic axis?

Answer

In the optically positive crystal. In an optically negative material, the extraordinary refractive index is smaller than the ordinary refractive index. The light will therefore experience a larger (ordinary) refractive index than in the optically positive material. A large refractive index implies a slower propagation of light through the medium, so when travelling parallel to the optic axis, the light will travel slower than if the crystal was optically positive.

3. Which of these could not induce anisotropy in an initially isotropic material?

a. Application of a stress

b. Application of an electric field

c. Application of a magnetic field

d. Application of a high temperature

Answer

A is incorrect. For example, in polymers, a stress can align the chains along a common direction, creating anisotropy. This may lead to optical birefringence .

B is incorrect. An electric field can create an overall electric polarisation direction in a material. This may lead to the electro-optic effect.

C is incorrect. A magnetic field can create an overall magnetic dipole direction in a material.

D is correct. Isotropy can be induced in a number of different ways. Most obviously, if the sample is heated through the melting temperature, it will form an isotropic liquid, but isotropy can also be induced in the solid state. For example, in polymers, heating can allow chains to resort to a coiled state, destroying any common alignment in the bulk polymer, and in perovskites such as barium titanate, a phase transition to an isotropic cubic form occurs at high temperature (phase transitions due to an increase in temperature commonly involve an increase in symmetry).

4. Below 0°C a particular material has a crystal structure that gives rise to anisotropic thermal conductivity. At room temperature the thermal conductivity of a sample of this material is found to be isotropic. In what circumstances would the following hypotheses explain this observation?

a. The sample is polycrystalline

b. The sample has undergone a phase transition when brought up to room temperature

c. The principle values of thermal conductivity have changed with temperature

Answer

A. In this case the thermal conductivity is not truly isotropic. If the sample is polycrystalline and the crystallites (grains) are randomly oriented the anisotropy will be averaged out. (Note: in many polycrystalline samples the grains are not randomly oriented but have preferred orientations or "texture".)

B. If the transition is to a cubic crystal structure the material will then display true isotropic thermal conductivity.

C. In this case the thermal conductivity is not truly isotropic. If the changes of the principal values with temperature are such as make them equal within the accuracy of the experimental technique at room temperature the thermal conductivity will appear to be isotropic.

5. Which of these materials will show isotropy in its mechanical properties?

a. Wood

b. Carbon fibre reinforced polymer

c. Window glass

d. Extruded polyethylene

Answer

C

6. An olivine crystal has the following diffusion constants for Ni at a certain temperature:

$$\begin{aligned} D_x &= 6 \times 10^{-8} \mathrm{m^2 s^{-1}} \\ D_y &= 4 \times 10^{-18} \mathrm{m^2 s^{-1}} \\ D_z &= 120 \times 10^{-18} \mathrm{m^2 s^{-1}} \end{aligned} \tag{2.11.1}$$

The unit cell dimensions are as follows:

$$\begin{aligned} a &= 0.5 \mathrm{nm} \\ b &= 1.0 \mathrm{nm} \\ c &= 0.6 \mathrm{nm} \end{aligned} \tag{2.11.2}$$

By considering the representation surface, what will the diffusion constant be when the concentration gradient lies along the [101] direction?

Answer

Using simple geometry, the [101] direction lies 50.19° from the x-axis (to 2 decimal places). Since [101] lies in the *x*-*z* plane, a simplified version of the equation of the ellipsoid representation surface can be used:

$$\frac{x^2}{A^2} + \frac{z^2}{C^2} = 1 \tag{2.11.3}$$

where A and C are equal to $\frac{1}{\sqrt{D}}$ in the *x* and *z* directions respectively.

Hence $x^2D_x + z^2D_z = 1$

Letting the length of the radius in the [101] direction be *r*, the equation becomes:

$$r^2 \cos^2 50.19 D_x + r^2 \sin^2 50.19 D_z = 1 \tag{2.11.4}$$

Since *r* will be equal to $\frac{1}{\sqrt{D}}$ in the [101] direction, substituting in the given values gives a diffusion constant of 73.3 x 10^{-18} m^2s^{-1}.

7. A certain orthorhombic crystal has the following principal values of thermal conductivity:

$$\begin{aligned} k_x &= 6.25 \mathrm{Wm^{-1}K^{-1}} \\ k_\mathrm{y} &= 1.00 \mathrm{Wm^{-1}K^{-1}} \\ k_z &= 1.75 \mathrm{Wm^{-1}K^{-1}} \end{aligned} \tag{2.11.5}$$

Answer

By simple geometry, the lengths of the [110] and the [111] vectors can be calculated. These are 1 nm and $\sqrt{2}$ nm respectively, and using these values, the direction cosines can be obtained. Using equations and notation from the TLP, the direction cosines are found to be:

$$I = \frac{x}{r} = \frac{0.8}{\sqrt{2}}, m = \frac{y}{r} = \frac{0.6}{\sqrt{2}}, n = \frac{z}{r} = \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}} \tag{2.11.6}$$

The thermal conductivity can be calculated using:

$$\begin{aligned} k &= k_\mathrm{x} R + k_\mathrm{y} m^2 + k_z n2 \\ k &= 6.25 \left(\frac{4}{5\sqrt{2}}\right)^2 + 1.00 \left(\frac{3}{5\sqrt{2}}\right)^2 + 1.75 \left(\frac{1}{\sqrt{2}}\right)^2 \\ k &= 2 + 0.18 + 0.875 = 3.055 \\ k &= 3.06 \mathrm{Wm^{-1}K^{-1}} (3\mathrm{s.f}) \end{aligned} \tag{2.11.7}$$

8. Explain how anisotropy is involved in the operation of the following devices:

a. a liquid crystal display
b. a fuel cell that uses a sold state electrolyte
c. a pyroelectric intruder alarm

9. Suggest ways of distinguishing between the answers to Question 4.

This page titled 2.11: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

3: Atomic Force Microscopy

https://www.doitpoms.ac.uk/tlplib/afm/index.php

Learning Objectives

- Understand the basic principles of atomic force microscopy (AFM), including the different modes it can be used in.
- Understand how AFM can be used in materials science.
- Be aware of some of the problems that can be encountered, and how to overcome them.

Before You Start

You should have a basic understanding of the behaviour of piezoelectric materials to understand how the piezo-scanner works in AFM

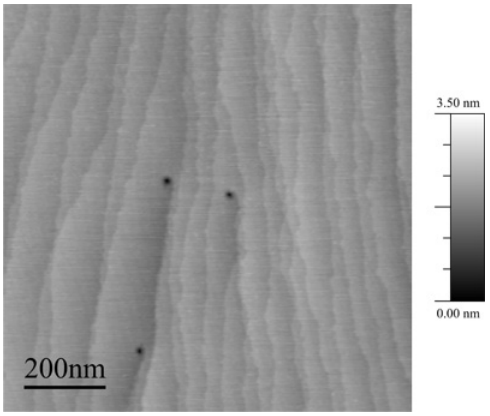
3.1: Introduction
3.2: Tip Surface Interaction
3.3: Modes of Operation
3.4: The Scanner
3.5: Tip and Cantilever
3.6: Feedback
3.7: Scanner Related Artefacts
3.8: Tip Related Artefacts
3.9: Other Artefacts
3.10: Summary
3.11: Questions

This page titled 3: Atomic Force Microscopy is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

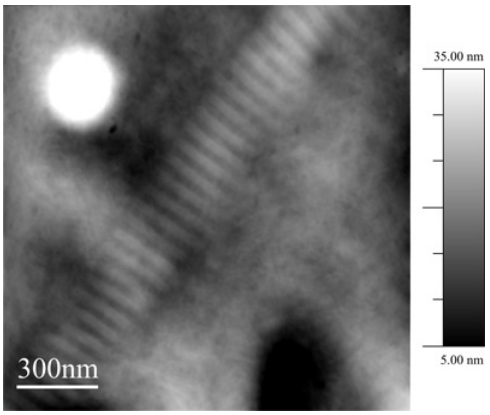
3.1: Introduction

Atomic force microscopy (AFM) is part of the family of techniques known as scanning probe microscopy, and has proved itself extremely valuable and versatile as an investigative tool. The AFM invented by Gert Binnig and others in the mid 1980s differed in many ways from today's instruments, but its basic principles remain the same. Binnig had already received the Nobel Prize in Physics for his creation of the scanning tunnelling microscope (STM), and the first AFMs in fact relied on an integrated STM tip. But the AFM had a major advantage over STM; it could be used for insulating as well as conducting samples.

Over the years, AFM has already had a significant impact in many disciplines, from surface science to biological and medical research. Because of its ability to image samples on an atomic scale, it has been vital to the advance of nanotechnology.



This AFM image shows the surface of a thin film of GaN. The surface morphology is dominated by terraces and steps. The step heights are approximately 0.25 nm, corresponding to one layer of gallium and nitrogen atoms. This illustrates the ability of AFM to measure very small height changes on surfaces.



The figure above is a topographic AFM image of a collagen fibril. The fibril is the striped structure running diagonally across the middle of the image. The periodicity of the narrow stripes or bands seen in the image is 64 nm. AFM can be used to image biological samples such as collagen without requiring a conductive coating to be added. It is even possible to take images of live cells in a fluid environment

In simple terms, the atomic force microscope works by scanning a sharp probe over the surface of a sample in a raster pattern. By monitoring the movement of the probe, a 3-D image of the surface can be constructed.

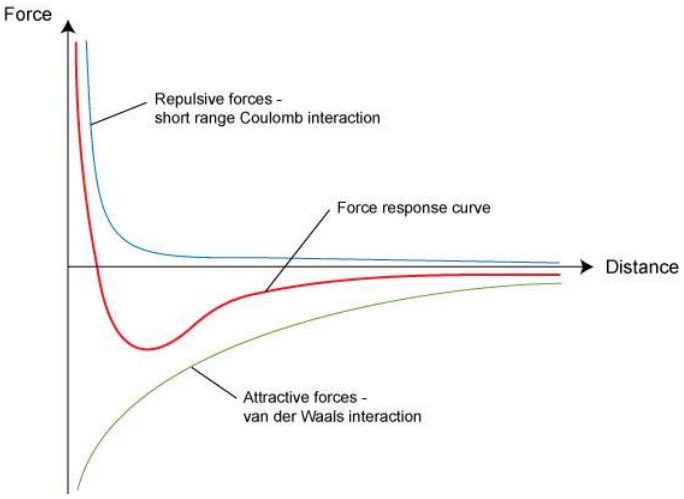
This page titled 3.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.2: Tip Surface Interaction

When the tip is brought close to the sample, a number of forces may operate. Typically the forces contributing most to the movement of an AFM cantilever are the *coulombic* and *van der Waals* interactions.

- **Coulombic interaction:** This strong, short range repulsive force arises from electrostatic repulsion by the electron clouds of the tip and sample. This repulsion increases as the separation decreases.
- **Van der Waals interactions:** These are longer range attractive forces, which may be felt at separations of up to 10 nm or more. They arise due to temporary fluctuating dipoles.

The combination of these interactions results in a force-distance curve similar to that below:



Plot of force against distance

As the tip is brought towards the sample, van der Waals forces cause attraction. As the tip gets closer to the sample this attraction increases. However at small separations the repulsive coulombic forces become dominant. The repulsive force causes the cantilever to bend as the tip is brought closer to the surface.

There are other interactions besides coulombic and van der Waals forces which can have an effect. When AFM is performed in ambient air, the sample and tip may be coated with a thin layer of fluid (mainly water). When the tip comes close to the surface, *capillary forces* can arise between the tip and surface.

It is also possible to detect other forces using the AFM, such as magnetic forces to map the magnetic domains of a sample

This page titled 3.2: Tip Surface Interaction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.3: Modes of Operation

AFM has three differing modes of operation. These are contact mode, tapping mode and non-contact mode.

Contact mode

In contact mode the tip contacts the surface through the adsorbed fluid layer on the sample surface. The detector monitors the changing cantilever deflection and the force is calculated using Hooke's law:

$$\$F = -kx \quad (F = \text{ force}, k = \text{ spring constant}, x = \$cantileverdeflection) \tag{3.3.1}$$

The feedback circuit adjusts the probe height to try and maintain a constant force and deflection on the cantilever. This is known as the *deflection setpoint*.

Tapping mode

In tapping mode the cantilever oscillates at or slightly *below* its resonant frequency. The amplitude of oscillation typically ranges from 20 nm to 100 nm. The tip lightly "taps" on the sample surface during scanning, contacting the surface at the bottom of its swing.

Because the forces on the tip change as the tip-surface separation changes, the resonant frequency of the cantilever is dependent on this separation.

$$\omega = \omega_0 \sqrt{1 - \frac{1}{k}\frac{\mathrm{d}F}{\mathrm{d}z}} \tag{3.3.2}$$

The oscillation is also damped when the tip is closer to the surface. Hence changes in the oscillation amplitude can be used to measure the distance between the tip and the surface. The feedback circuit adjusts the probe height to try and maintain a constant amplitude of oscillation i.e. the *amplitude setpoint*.

Non-contact mode

In non-contact mode the cantilever oscillates near the surface of the sample, but does not contact it. The oscillation is at slightly *above* the resonant frequency. Van der Waals and other long-range forces decrease the resonant frequency just above the surface. This decrease in resonant frequency causes the amplitude of oscillation to decrease.

In ambient conditions the adsorbed fluid layer is often significantly thicker than the region where van der Waals forces are significant. So the probe is either out of range of the van der Waals forces it attempts to measure, or becomes trapped in the fluid layer. Therefore non-contact mode AFM works best under ultra-high vacuum conditions.

Comparison of modes

	Advantage	**Disadvantage**
Contact Mode	• High scan speeds • Rough samples with extreme changes in vertical topography can sometimes be scanned more easily	• Lateral (shear) forces may distort features in the image • In ambient conditions may get strong capillary forces due to adsorbed fluid layer • Combination of lateral and strong normal forces reduce resolution and mean that the tip may damage the sample, or vice versa
Tapping Mode	• Lateral forces almost eliminated • Higher lateral resolution on most samples • Lower forces so less damage to soft samples or tips	• Slower scan speed than in contact mode
Non-contact Mode	• Both normal and lateral forces are minimised, so good for measurement of very soft samples • Can get atomic resolution in a UHV environment	• In ambient conditions the adsorbed fluid layer may be too thick for effective measurements • Slower scan speed than tapping and contact modes to avoid contacting the adsorbed fluid layer

This page titled 3.3: Modes of Operation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.4: The Scanner

The scanner moves the probe over the sample (or the sample under the probe) and must be able to control the position extremely accurately. In most AFMs piezoelectric materials are used to achieve this. These change dimensions with an applied voltage. The diagram below shows a typical scanner arrangement, with a hollow tube of piezoelectric material and the controlling electrodes attached to the surface.

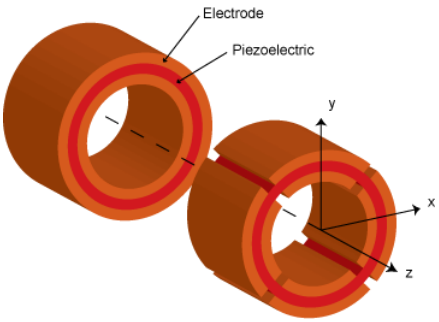


Diagram of a typical piezo scanner (cut into two parts). Separate pairs of electrodes control movement in the x, y and z directions

This page titled 3.4: The Scanner is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.5: Tip and Cantilever

The cantilever is a long beam with a tip located at its apex. In most AFMs the motion of the tip is detected by reflecting a laser off the back surface of the cantilever.

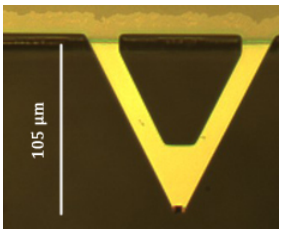
Tip

The tip is generally pyramidal or tetrahedral in shape, and usually made from silicon or silicon nitride. Silicon can be doped and made conductive, allowing a tip-sample bias to be applied for making electrical measurements. Silicon nitride tips are not conducting.

The geometry of the tip greatly affects the lateral resolution of the AFM, since the tip-sample interaction area depends on the tip radius. The radius of the apex of a new tapping mode tip is around 5–15 nm, but this increases quickly with wear. In general the sharper the tip, the higher the resolution of the AFM image.

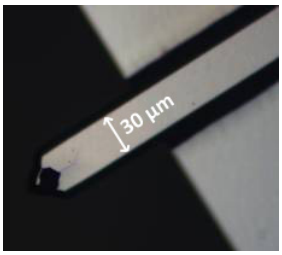
Cantilever

For contact mode AFM the cantilever needs to deflect easily without damaging the sample surface or tip. Therefore it should have a low spring constant, this is achieved by making it *thin* (0.3–2 μm). It also needs a high resonant frequency to avoid vibrational instability, so is typically *short* (100–200 μm). V-shaped cantilevers are often used for contact mode as these can provide low resistance to vertical deflection, whilst resisting lateral torsion.



Optical microscopy image of a triangular cantilever

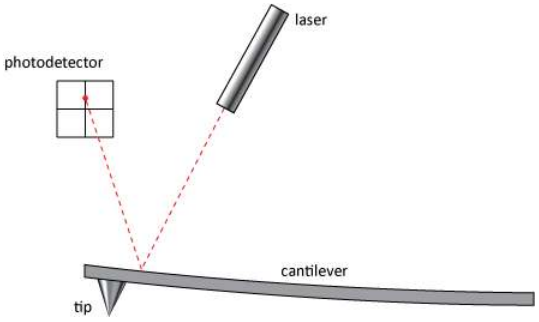
For tapping mode AFM a high spring constant is required to reduce noise and instabilities. Rectangular cantilevers are often used for tapping mode.



Optical microscopy image of a rectangular cantilever

Detection of cantilever deflection

There are a number of ways to detect the deflection of the cantilever in an AFM. The most common method is using a laser beam. A *diode laser* is focused onto the back reflective surface of the cantilever, and reflects onto a photodetector. This is position sensitive, and usually has four sectors. The vertical deflection of the cantilever is determined by the difference in light intensity measured by the upper and lower sectors. It is also possible to measure the lateral deflection of the cantilever by the difference between the left and right sectors of the photodetector; this technique is known as lateral force microscopy (LFM).



How the deflection of the cantilever is measured

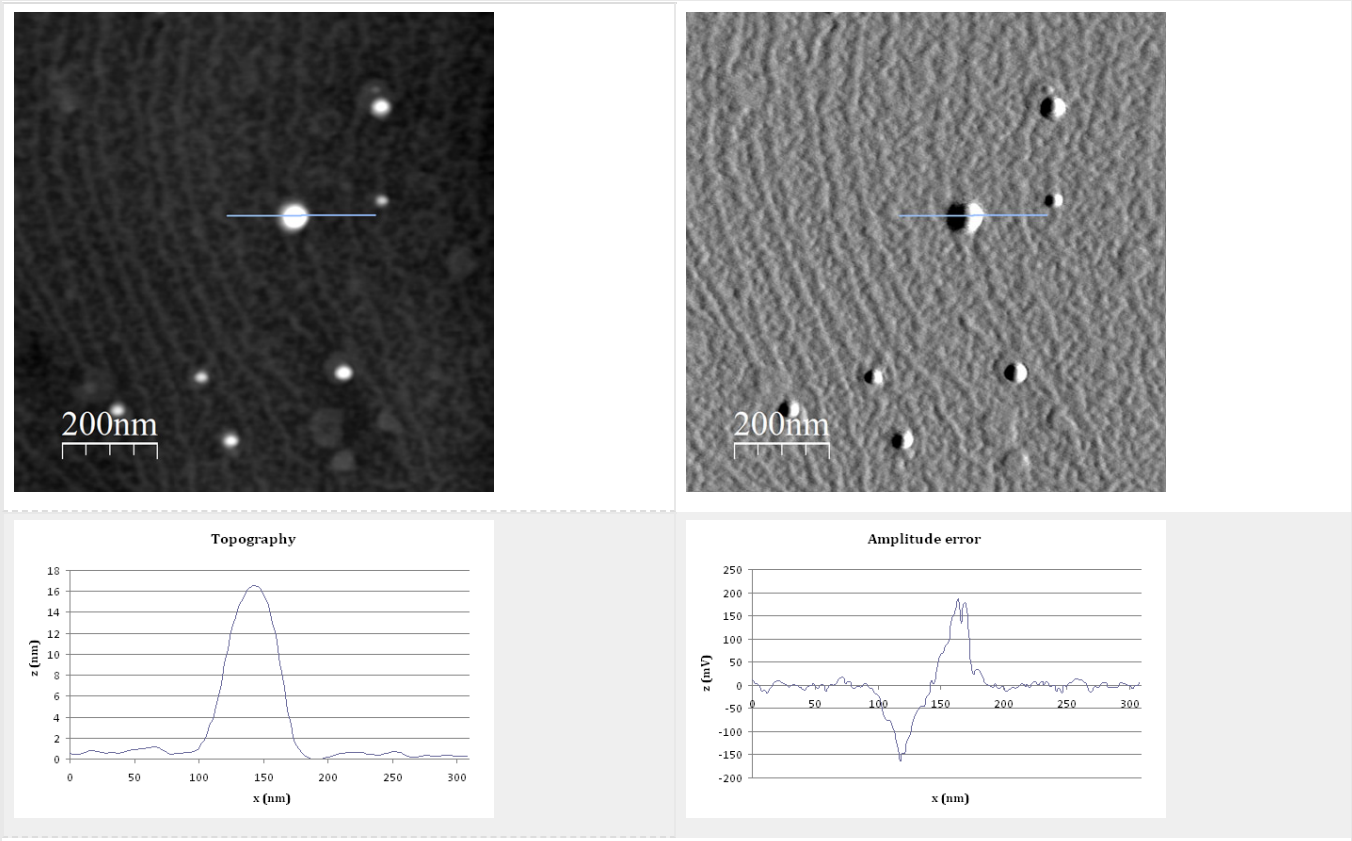
This page titled 3.5: Tip and Cantilever is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.6: Feedback

When the tip contacts the surface directly the tip and/or surface may be damaged. If the tip is blunted or damaged, then the imaging capability of the AFM is reduced. Soft surfaces (e.g. on biological samples) can also be easily damaged.

In almost all operating modes, a feedback circuit is connected to the deflection sensor and attempts to keep the tip–sample interaction constant by controlling the tip–sample distance. This protects both the tip and the sample. Either the cantilever deflection (in static mode) or oscillation amplitude (in dynamic mode) is monitored by the feedback circuit, which attempts to keep this at a setpoint value by adjusting the z height of the probe. The height of the probe is what is recorded to produce a topographic image.

In practice however feedback is never perfect, and there is always some delay between measuring a change from the setpoint and restoring it by adjusting the scanning height. In tapping mode for example this can be measured by the difference between the instantaneous amplitude of oscillation and the amplitude setpoint. This is known as the amplitude error signal, and highlights changes in surface height.



Example images showing the relationship between topography and amplitude error signal. The two line plots demonstrate a slice through the acquired image.

The feedback system is affected by three main parameters:

- *Setpoint* – this is the value of the deflection or amplitude that the feedback circuit attempts to maintain. This is usually set such that the force on the cantilever is small, but the probe remains engaged with the surface.
- *Feedback gains* – the higher these are set, the faster the feedback system will react. However if the gains are too high then the feedback circuit can become unstable and oscillate, causing high frequency noise in the image.
- *Scan rate* – scanning the probe over the surface more slowly gives the feedback circuit more time to react and results in better tracking, but this increases the time needed to acquire an image.

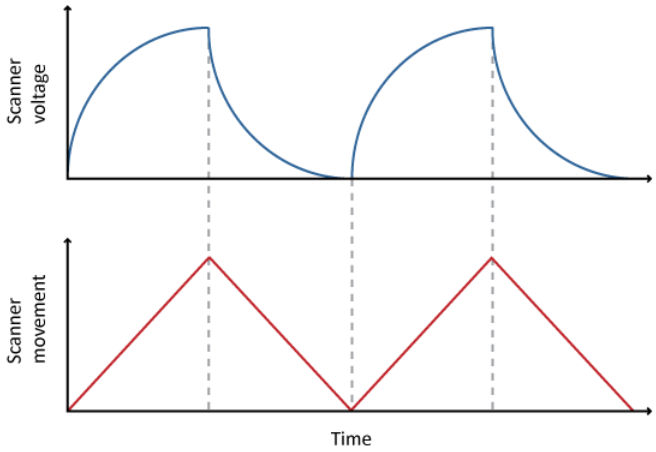
This page titled 3.6: Feedback is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.7: Scanner Related Artefacts

There are a number of problems and artefacts that can arise during atomic force microscopy. This page and the following pages will discuss some of them, and how they can be overcome.

Hysteresis

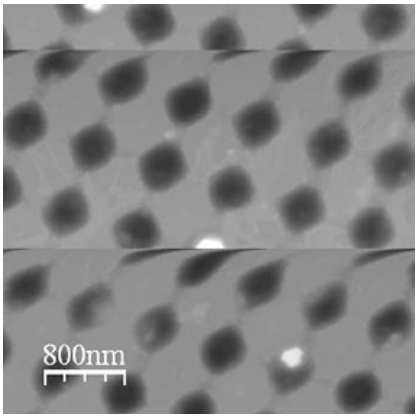
The piezoelectric's response to an applied voltage is not linear. This gives rise to *hysteresis*. Since the scanner makes more movement per volt at the beginning of a scan line than at the end, this can cause artefacts in the images, especially at large scan sizes. This is overcome by using a non-linear voltage waveform calculated during a calibration procedure.



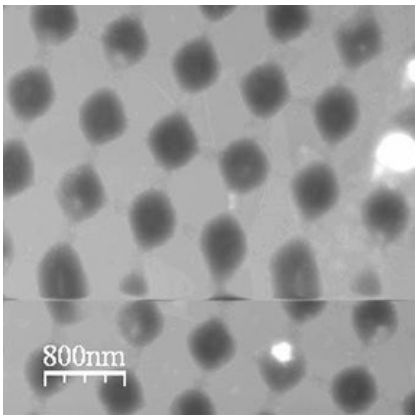
Example of a voltage waveform calibrated to overcome hysteresis

Scanner creep

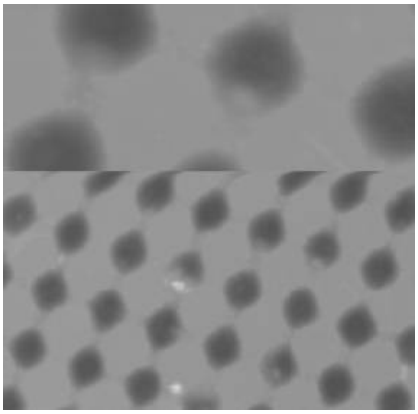
If the applied voltage suddenly changes e.g. to move the scanning position, then the piezo-scanner's response is not all at once. It moves the majority of the distance quickly, then the last part of the movement is slower. If this is done during scanning, then the slow movement will cause distortion. This is known as *creep*.



When a change in x-offset is applied, features are distorted in the x-direction



When a change in y-offset is applied, features are distorted in the y-direction



The scan size is changed abruptly, and features are distorted

Bow and tilt

Because of the construction of the piezo-scanner, the tip does not move in a perfectly flat plane. Instead its movement is in a parabolic arc, as shown in the image below. This causes the artefact known as *scanner bow*. Also the scanner and sample planes may not be perfectly parallel, this is known as *tilt*. Both of these artefacts can be removed by using post-processing software.

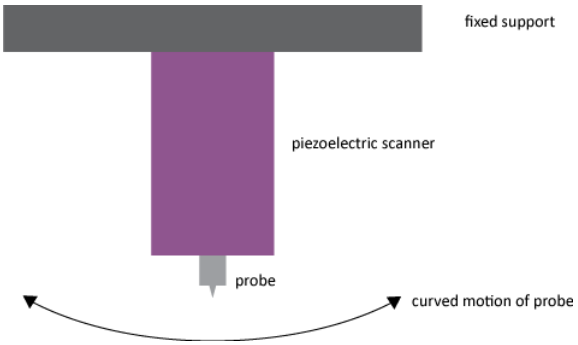
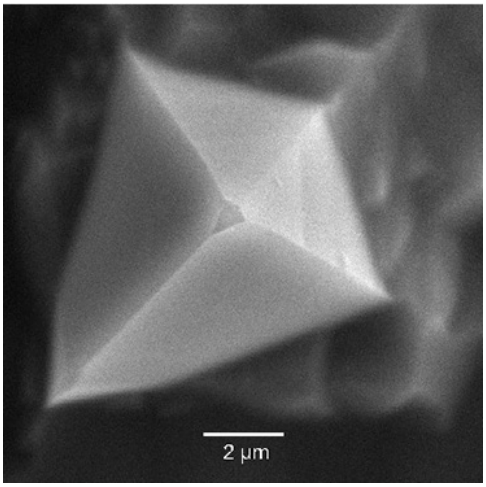


Diagram of scanner bow

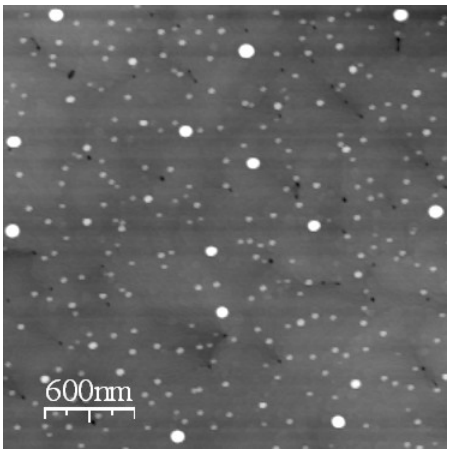
This page titled 3.7: Scanner Related Artefacts is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.8: Tip Related Artefacts

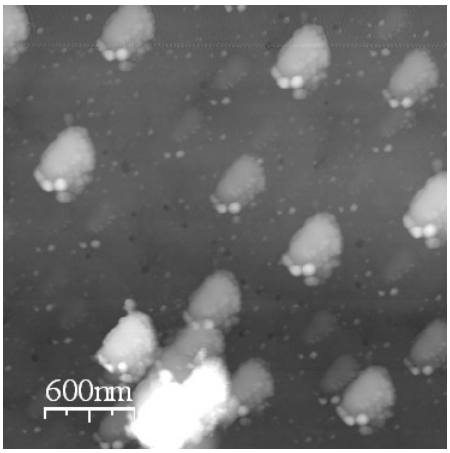
The tip may pick up loose debris from the sample surface. This may be reduced by cleaning the sample with compressed air or N_2 before use. Or the tip can be damaged during scanning, which degrades the images. This may be blunting of the tip, as shown in the SEM image below:



Below is an example of an image taken with a severely damaged tip. The shape due to tip damage appears several times over the image, effectively the sample is imaging the tip rather than the other way round.



Sample imaged with sharp tip



The same sample imaged using a severely damaged tip

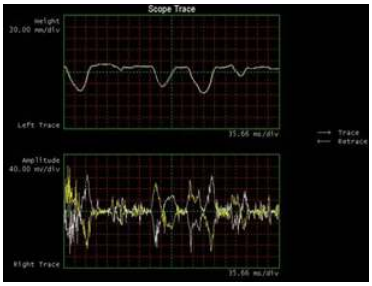
One easy way to check for tip artefacts is to rotate the sample (**not** just the scanning direction) by 90 degrees.

This page titled 3.8: Tip Related Artefacts is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

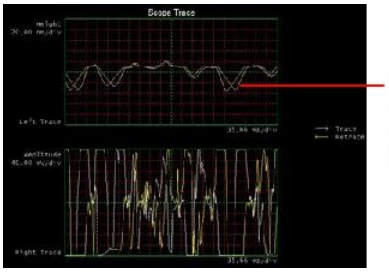
3.9: Other Artefacts

Feedback related

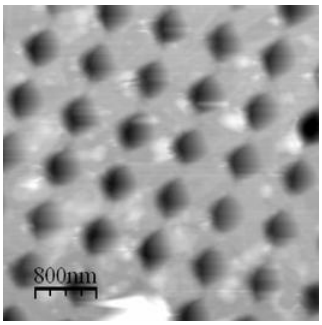
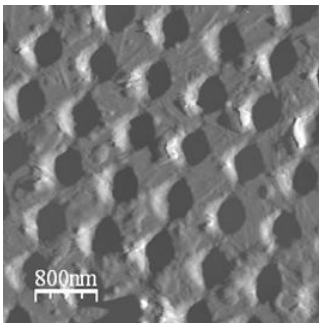
The feedback is supposed to keep the tip-sample interaction at a fixed setpoint by adjusting the z height of the probe, as discussed earlier. However if the scan speed across the sample is fast, then the feedback may not be able to react quickly enough and tracking is poor. This can be seen by comparing the trace and retrace (forward and backward direction) for a single line in the scan. The following image shows the height and amplitude trace (white) and retrace (yellow) when tracking is good. The height trace and retrace are almost identical, and the amplitude retrace is a mirror image of the trace because it is in the opposite direction.



When tracking is poor, the trace and retrace of height no longer overlap. Blurred images result. This can happen because the gains are set too low, or the scan speed is too high.

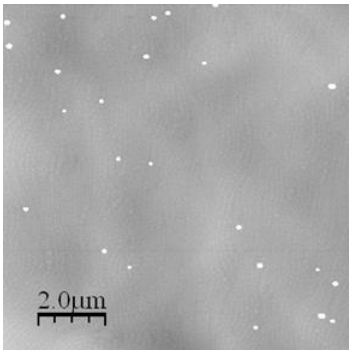


The images below are examples of poor tracking

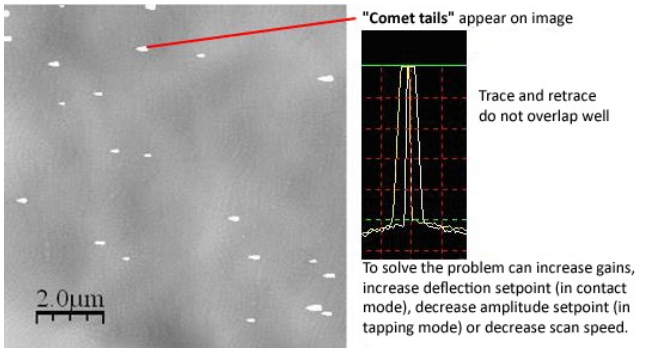
	
Topography	Amplitude error

With sharp slopes, poor tracking may result in overshoot giving rise to “comet tails” in the image. The following images show indium aluminium nitride with small balls of indium on the surface. On the left the gains are set high enough for the scan rate, and

tracking is good. On the right the gains are too low for the scan rate, and the tracking is poor. This results in overshooting off the edges of the indium dots, appearing in the image as comet tails. This can also be seen as the trace and retrace not overlapping.

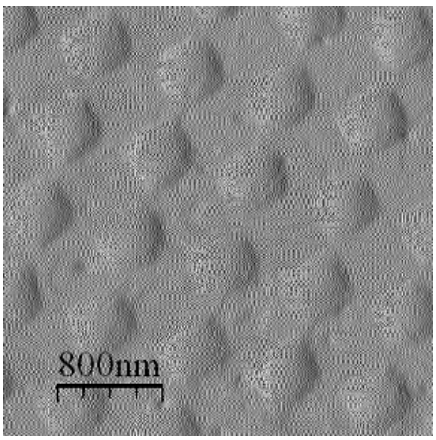


Good tracking



Poor tracking, resulting in “comet tails

However if the gains are set too high, then the feedback circuit can begin to oscillate. This causes high frequency noise.

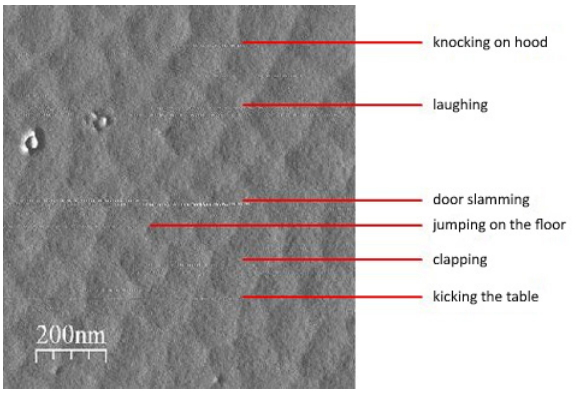


Amplitude error image for a scan with the gains set too high

The precise values used for feedback gains will vary between instruments. A good rule of thumb is to increase the gain until excess noise begins to appear, and then reduce it slightly to get good tracking with low noise.

Vibrations

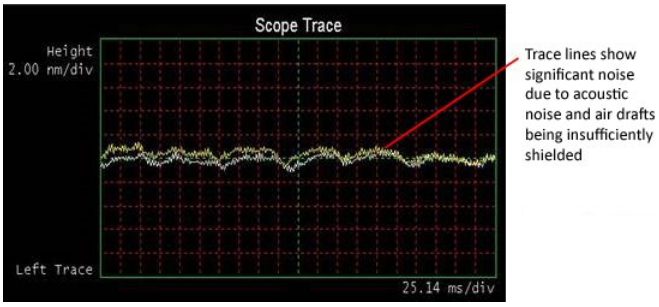
AFMs are very sensitive to external mechanical vibrations, which generally show up as horizontal bands in the image.



Evidence of external vibrations in an amplitude error image

These vibrations may be transmitted through the floor, for example from footsteps or the use of a lift. These can be minimised by the use of a vibrational isolation table, and locating the AFM on a ground floor or below.

Acoustic noise such as people talking can also cause image artefacts, as can drafts of air. An acoustic hood can be used to minimise the effects of both of these.



Acoustic hood open



Acoustic hood closed

This page titled 3.9: Other Artefacts is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.10: Summary

Atomic force microscopy may be used to image the micro- and nano-scale morphology of a wide range of samples, including both conductive and insulating materials, and both soft and hard materials. Successful imaging requires optimisation of the feedback circuit which controls the cantilever height, and an understanding of the artefacts which may arise due to the nature of the instrument and any noise sources in its immediate environment. Despite these issues, atomic force microscopy is a powerful tool in the emerging discipline of nanotechnology.

Going Further

Books

- Meyer, Hug and Bennewitz, *Scanning Probe Microscopy: The Lab on a Tip*, Springer, 2003

Websites

- A guide to the principles and techniques of AFM and other SPM methods, SPM Principles (NT-MDT) including animations
- A Guide to AFM Image Artifacts (Pacific Nanotechnology)]
- Atomic Force Microscopy (nanoHUB.org)] – a presentation/podcast introducing AFM

This page titled 3.10: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

3.11: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

1. Which operating mode allows for the fastest scanning speeds?

a. Contact mode

b. Tapping mode

c. Non-contact mode

Answer

A

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

2. If high frequency noise is seen in an image, what should be done?

a. Increase the feedback gains

b. Decrease the feedback gains

c. Change the tip

d. Recalibrate the AFM

Answer

B. This is due to oscillation of the feedback circuit when the gains are set too high.

This page titled 3.11: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

4: Atomic Scale Structure of Materials

Learning Objectives

- know the differences between single crystal, polycrystalline and amorphous solids
- be able to identify the characteristic features of single crystals and polycrystals
- understand the nature of crystal defects
- appreciate the use of polarised light to examine optical properties

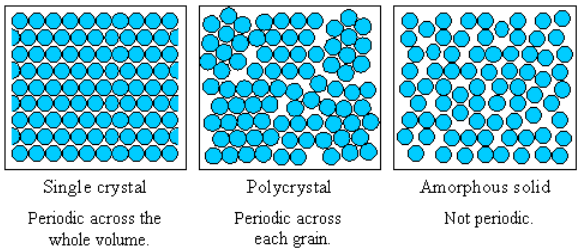
4.1: Introduction
4.2: Single crystals- Shape and anisotropy
4.3: Single crystals- Mechanical properties
4.4: Single crystals- Optical properties
4.5: Polycrystals
4.6: Defects
4.7: Summary
4.8: Questions

This page titled 4: Atomic Scale Structure of Materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

4.1: Introduction

The fundamental difference between single crystal , polycrystalline and amorphous solids is the length scale over which the atoms are related to one another by translational symmetry ('periodicity' or 'long-range order'). Single crystals have infinite periodicity, polycrystals have local periodicity, and amorphous solids (and liquids) have no long-range order.

- An *ideal single crystal* has an atomic structure that repeats periodically across its whole volume. Even at infinite length scales, each atom is related to every other equivalent atom in the structure by translational symmetry.
- A *polycrystalline solid* or *polycrystal* is comprised of many individual *grains* or *crystallites*. Each grain can be thought of as a single crystal, within which the atomic structure has long-range order. In an *isotropic* polycrystalline solid, there is *no relationship* between neighbouring grains. Therefore, on a large enough length scale, there is no periodicity across a polycrystalline sample.
- *Amorphous* materials, like window glass, have no long-range order at all, so they have no translational symmetry. The structure of an amorphous solid (and indeed a liquid) is not truly random - the distances between atoms in the structure are well defined and similar to those in the crystal. This is why liquids and crystals have similar densities - both have *short-range order* that fixes the distances between atoms, but only crystals have long-range order.



The range of crystalline order distinguishes single crystals, polycrystals and amorphous solids. The figure shows how the *periodicity* of the atomic structure of each type of material compares.

Many characteristic properties of materials, such as mechanical, optical, magnetic and electronic behaviour, can be attributed to the difference in structure between these three classes of solid.

This page titled 4.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

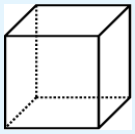
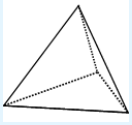
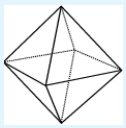
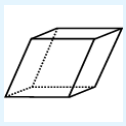
4.2: Single crystals- Shape and anisotropy

A single crystal often has distinctive plane faces and some symmetry. The actual shape of the crystal will be determined by the availability of crystallising material, and by interference with other crystals, but the angles between the faces will be characteristic of the material and will define an ideal shape. Single crystals showing these characteristic shapes can be grown from salt solutions such as alum and copper sulphate.

Gemstones are often single crystals. They tend to be cut artificially to obtain aesthetically pleasing refractive and reflective properties. This generally requires cutting along crystallographic planes. This is known as cleaving the crystal. A familiar example is diamond, from which decorative stones can be cleaved in different ways to produce a wide range of effects.

To see a variety of symmetrical naturally formed minerals, visit the mineral galleries website.

Consider the following three-dimensional shapes:

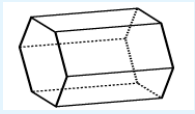
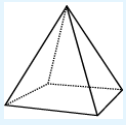
	Cube: 6 identical squares
	Tetrahedron: 4 identical equilateral triangles
	Octahedron: 8 identical equilateral triangles
	Rhombohedron: 6 identical parallelograms with sides of equal length

You can make your own cube, octahedron and tetrahedron by printing the following pages and following the instructions on them.

- Patterns for cube and tetrahedron
- Pattern for octahedron

These three shapes are the most important in materials science, and you should be very familiar with them!

The symmetry exhibited by real single crystals is determined by the crystal structure of the material. Many have shapes composed of less regular polyhedra, such as prisms and pyramids.

	Hexagonal prism: 2 hexagons and 6 rectangles
	Square-based pyramid: 4 triangles and a square

Not all single crystal specimens exhibit distinctive polyhedral shapes. Metals, for example, often have crystals of no particular shape at all.



These quartz specimens show a range of shapes typically exhibited by crystals.

Most single crystals show anisotropy in certain properties, such as optical and mechanical properties. An amorphous substance, such as window glass, tends to be isotropic . This difference may make it possible to distinguish between a glass and a crystal. The characteristic shape of some single crystals is a clue that the properties of the material might be directionally dependent. The properties of polycrystalline samples can be completely isotropic or strongly anisotropic depending on the nature of the material and the way in which it was formed.

This page titled 4.2: Single crystals- Shape and anisotropy is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

4.3: Single crystals- Mechanical properties

Gypsum can be cleaved along particular crystallographic planes using a razor blade. The bonding perpendicular to these cleavage planes is weaker than that in other directions, and hence the crystal breaks preferentially along these planes. Quartz and diamond do not have such distinct cleavage planes, and so cleaving these crystals requires much more effort and care.

There are distinct planes in the gypsum structure, with no bonding between them. These are the cleavage planes. It is much more difficult to cleave gypsum along planes other than these. In contrast, all of the planes in the quartz structure are interconnected and the material is much more difficult to cleave in any direction. This is a demonstration of a way in which the crystal structure of a material can influence its mechanical properties.



Certain crystals, such as gypsum, can be cleaved with a razor blade along particular crystallographically-determined planes. (Click on image to view larger version.)

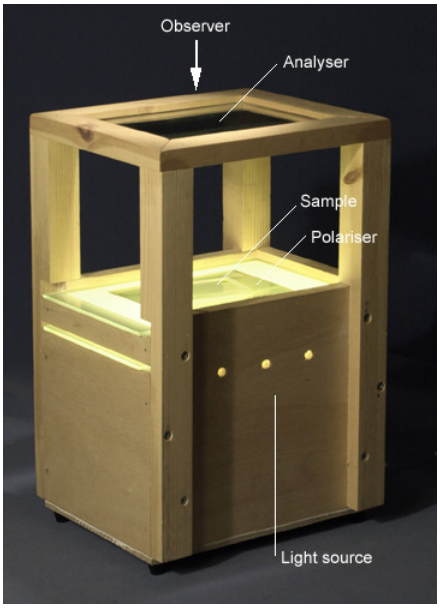
Glass is impossible to cleave. As an amorphous substance, glass has no crystallographic planes and therefore can have no easy-cleavage directions. Glassy materials are often found to be mechanically harder than their crystalline equivalents. This is an example of how mechanical properties of crystals and amorphous substances differ.

This page titled 4.3: Single crystals- Mechanical properties is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

4.4: Single crystals- Optical properties

Quartz crystals are birefringent , so they exhibit optical anisotropy. Consider plane polarised light passing through a birefringent crystal. Inside the crystal, the light is split into two rays travelling along *permitted vibration directions* (p.v.d.s). The two rays are subject to different refractive indices, so the light travelling along each p.v.d. reaches the opposite side of the crystal at a different time. When the two rays recombine, there is a phase difference between the two rays that causes the *polarisation state* of the transmitted light to be different from that of the incident light.

Optical anisotropy in thin samples can be observed by placing the sample between crossed polarising filters in a light box. The bottom filter, between the light source and the sample, is called the polariser . The top filter, between the sample and the observer, is called the analyser . The polariser and analyser have polarising directions perpendicular to one another.



The apparatus used for examining optical anisotropy consists of a white-light source, two polarising filters and a frame to hold them apart so creating a working space.

When no sample is in place the light that reaches the analyser is polarised at 90° to the analyser's polarisation direction, so no light is transmitted to the observer. When a quartz sample (with favourable orientation, see later) is placed between the filters, the crystal changes the polarisation state of the light that is transmitted through it. When this light reaches the analyser, some component of it lies parallel to the polarisation direction of the analyser, and therefore some light is transmitted to the observer.

If a quartz slice shows optical anisotropy, the intensity of light transmitted through the analyser varies as a function of the angle of rotation of the quartz sample in the plane of the filters. At certain orientations, no light is transmitted. These 'extinction positions' are found at 90° intervals.

https://www.doitpoms.ac.uk/tlplib/at..._animation.mp4

Video animation of anisotropic quartz rotated between crossed polarisers

When the same experiment is done using a piece of glass, it is found that light is not transmitted for *any* orientation. This is because the glass is *optically isotropic*, and does not change the polarisation state of the light passing though it.

In quartz, there is one direction of propagation for which no birefringence is observed. If a sample is cut so that the incident light is parallel to this direction, the sample behaves as if it is optically isotropic and no light is transmitted. The crystallographic direction that exhibits this property is known as the *optic axis*.



When the quartz sample is cut so that the incident light is parallel to the optic axis, no light is transmitted in any orientation.

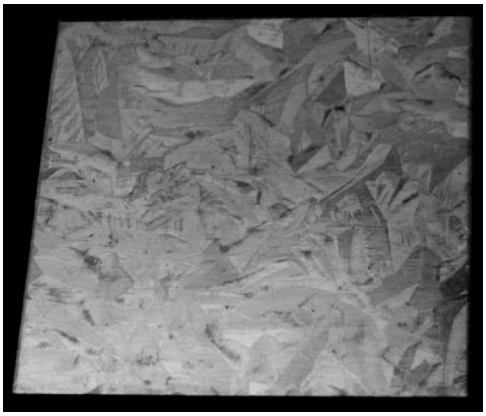
This experiment demonstrates that some single crystals, such as quartz, show anisotropic optical properties. The phenomenon depends on the crystallographic orientation of the crystal with respect to the incident light. Amorphous materials like glass have no 'distinct' crystal directions, so anisotropic properties are generally not observed.

This page titled 4.4: Single crystals- Optical properties is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

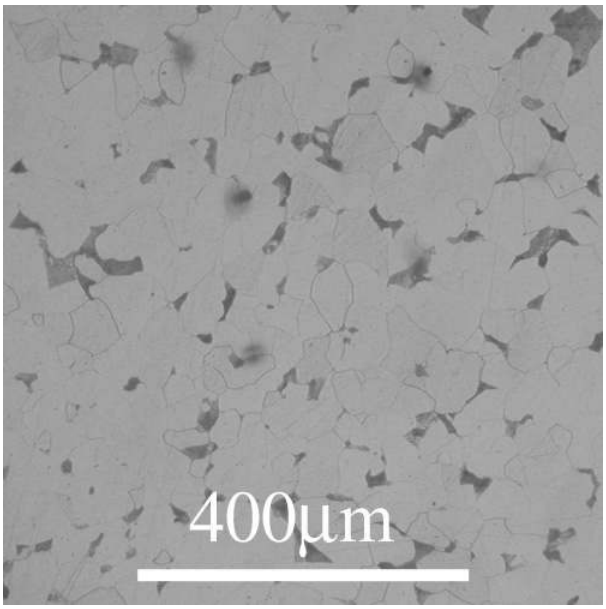
4.5: Polycrystals

Single crystals form only in special conditions. The normal solid form of an element or compound is *polycrystalline*. As the name suggests, a polycrystalline solid or *polycrystal* is made up of many crystals. The properties of a polycrystal are notably different from those of a single crystal. The individual component crystallites are often referred to as grains and the junctions between these grains are known as grain boundaries .

The size of a grain varies according to the conditions under which it formed. Galvanised steel has a zinc coating with visibly large grains. Other materials have much finer grains, and require the use of optical microscopy.



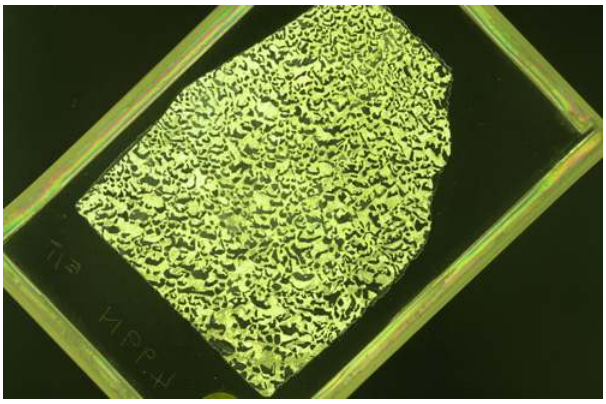
In galvanised steel, the grains are big enough to be seen unaided. The plate measures 5 cm across.



In many other metals, such as this hypoeutectoid iron-carbon alloy, the grains may only be seen under a microscope.

These photographs show a polycrystalline sample of quartz mixed with feldspar in which the grains all have optically anisotropic properties. Between the crossed polarisers, each grain allows transmission of light at a slightly different point in the rotation. This gives the strange effect seen here.





This polycrystal contains randomly oriented grains that allow transmission at different angles. Consequently different regions of the polycrystal are seen in these two photographs.

The three-dimensional shape of grains in a polycrystal is similar to the shape of individual soap bubbles made by blowing air into a soap solution contained in a transparent box.

The surface between bubbles is a high-energy feature. If the area of the surface is decreased, the overall energy of the system decreases, so *reduction of surface area is a spontaneous process*. If all the bubbles were the same size, the resulting structure would be a regular *close-packed* array, with 120° angles between the surfaces of neighbouring bubbles.

In practise, *bubble growth* can occur because the surface area of a few large bubbles is lower than that of many small bubbles. Large bubbles tend to grow, and small bubbles tend to shrink. The bubbles are therefore different sizes so there are large deviations from the close-packed structure. On average, however, three bubbles meet at a junction, and the angle between the bubble surfaces is usually within a few tens of degrees of 120°.

The *curvature* of the surfaces is also important. Surfaces with a smaller radius of curvature have a higher energy than those with a larger radius of curvature. As a result, some small bubbles cannot shrink and disappear, even though the surface area would decrease if they did so. This is because the curvature of the boundaries, and the associated energy, would be too high.

In a real polycrystal, the grain boundaries are high-energy features, similar to the surfaces between bubbles. The soap froth is a very good model for the grain structure of a simple polycrystalline material, and many similar features can be observed in the two systems. The soap bubbles are analogous to the grains, and the surfaces of the bubbles are analogous to the grain boundaries.

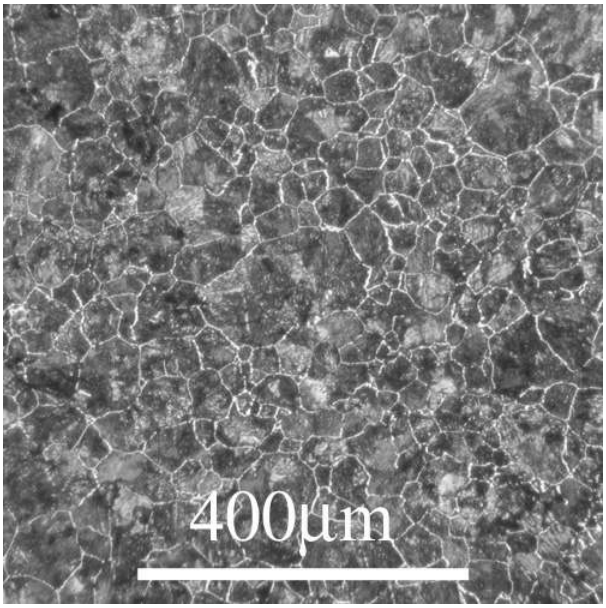
Compare the photographs of the soap bubbles with the micrograph of a polycrystalline material that has been etched to reveal the grain boundaries.







The packing of soap bubbles is somewhat similar to the packing of crystals - both systems seek to minimise their surface area. Note the angles at the junctions of grain boundaries.



The grains of this hypereutectoid iron-carbon alloy are packed in a similar way to the bubbles in the previous photographs.

Grain boundaries in a polycrystalline solid can *relax* (move in such a way to decrease the total energy of the system) when atomic rearrangement by diffusion is possible. In real materials, many other effects can influence the observed grain structure.

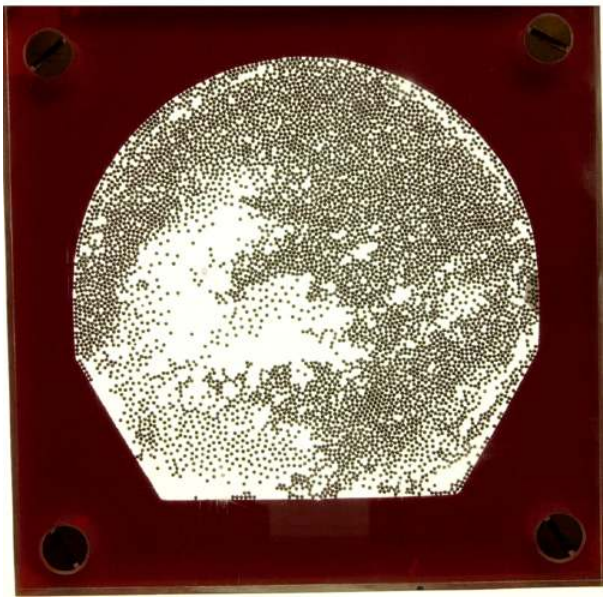
This page titled 4.5: Polycrystals is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

4.6: Defects

Within a single crystal or grain, the crystal structure is not perfect. The structure contains defects such as vacancies , where an atom is missing altogether, and dislocations , where the perfection of the structure is disrupted along a line. Grain boundaries in polycrystals can be considered as two-dimensional defects in the perfect crystal lattice. Crystal defects are important in determining many material properties, such as the rate of atomic diffusion and mechanical strength.

We can use a "shot model" to get a picture of crystal defects. The model consists of many small ball bearings trapped in a single layer between two transparent plates. They tend to behave like the atoms in a crystal, and can show the same kind of defects.

When the shot model is held horizontally, so that the balls flow freely, the resulting structure is similar to a liquid.



Shot model held horizontally. The balls form a liquid-like structure.

As the model is tilted towards the vertical, the balls pack closely together. This represents crystallisation. One or two balls may be suspended above the main body by electrostatic forces: this is comparable to the vapour found above the crystal.

In some places the balls form close-packed regions. Tapping of the model causes minor rearrangements of the balls, especially at the top of the "solid" region. This is similar to diffusion, in which case the tapping is analogous to thermal activation. Occasionally, the "diffusion" process may cause two grains to join together, or for some grains to "grow". The following image sequence shows the behaviour of the shot model as it is rearranged by tapping, starting from a polycrystalline state with many small grains and ending with much larger grains. Note the presence of vacancies in the structure.

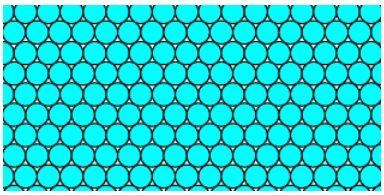
https://www.doitpoms.ac.uk/tlplib/at...rainGrowth.mp4

Grain growth in the shot model

With great care, it may be possible to create a single crystal, as *all* the balls form a single pattern. Note that diffusion occurs mainly near the top of the balls: those towards the middle and bottom do not easily move, as the photographs show.

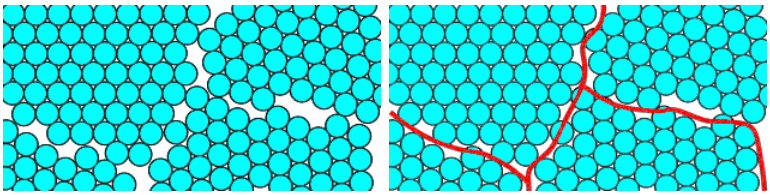
Even in a single crystal, or large-grained sample, there are still vacancies, as the shot model shows. The reason for this involves *entropy*: at all finite temperatures, there will be some disorder in the crystal.

The balls within a grain arrange themselves into close-packed planes. In metals, close-packing of atoms is a very common structure. This pattern is typical of *hexagonal close-packed* and *cubic close-packed* lattices. Note that in this 2-D model, each ball touches six others. In a 3-D crystal, such as a real one, each ball would also touch three on the plane above, and three on the plane below.



A close-packed plane.

In the shot model, the balls are normally arranged in to a polycrystalline form, shown schematically below:



A polycrystal will typically have crystalline regions (grains) bounded by disordered grain boundaries. These boundaries are marked in the picture on the right.

Note that the packing of atoms at the grain boundaries is disordered compared to the grains. At a grain boundary, the normal crystal structure is obviously disturbed, so the boundaries are regions of high energy. The ideal low energy state would be a single crystal.

Polycrystals form from a melt because crystallisation starts from a number of centres or nuclei . These developing crystals grow until they meet. Since they are not usually aligned before meeting, the grains need not necessarily be able to fit together as a single crystal, hence the polycrystalline structure.

After crystallisation the solid tends to reduce the boundary area, and hence the internal energy, by *grain growth*. This can only happen by a process of *atomic diffusion* within the solid. Such diffusion is more rapid at a higher temperature since it is thermally activated .

This page titled 4.6: Defects is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

4.7: Summary

The focus of this package is the difference between single crystals, polycrystals and amorphous solids. This is explained in terms of the atomic scale periodicity: single crystals are periodic across their entire volume; polycrystals are periodic across individual grains; amorphous solids have little to no periodicity at all.

The different atomic structures can have effects on the macroscopic properties. A single crystal may exhibit anisotropy - we have seen mechanical anisotropy of gypsum, and optical anisotrpy of quartz. Polycrystals may also be anisotropic within each grain, as seen when the polycrystalline quartz-feldspar mix was placed between the crossed polarisers. Amorphous solids do not have anisotropic mechanical or optical properties, since they are isotropic on the atomic scale.

Defects may exist in all structures, even single crystals. They include vacancies and grain boundaries, where the regular repeating structure is disrupted.

Going further

Most 'introductory' materials science textbooks will cover the basic material in this package. The following resources cover the subjects in more detail than this teaching and learning package, and may prove useful to the interested student.

Books

Introduction to Mineral Sciences by Putnis (Cambridge University Press, 1992)
Provides a mineral-based treatment of many of the topics introduced in this package. Of particular interest:
Chapter 1 on Periodicity and Symmetry
Chapter 2 on Anisotropy and Optical Properties, including the phenomenon of birefringence
Chapter 5 on Crystal Structures
Chapter 7 on Defects in Minerals

The Structure of Materials by Allen and Thomas (Wiley, 1999)
Gives a thorough mathematical treatment of the noncrystalline and crystalline states (chapters 2 and 3).

Websites

- Steffen Weber's homepage
  Contains java-based applets that allow the structure of common polyhedra and crystals to be explored.
- Crystallography and Minerals Arranged by Crystal Form
  A library of 'crystal forms' - the shapes adopted by natural crystals. Contains Java applets.
- Molecular Expressions: Optical Birefringence
  An excellent tutorial on birefringence. Contains Java applets.

Other resources

The MATTER Project's 'Materials Science on CD-ROM' includes modules on:
Introduction to Crystallography (including Miller Indices etc.)
Introduction to Point Defects
Dislocations
See the MATTER website for details of availability.

This page titled 4.7: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

4.8: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

1. Most single crystals contain:

a	No defects, since they must be perfect crystals.
b	Exactly one defect, hence the term '*single* crystal'.
c	Many defects.
d	More defects than atoms, since every atom must generate at least one defect.

Answer

c

2. If quartz had optical properties such that the refractive indices for all vibration directions were equal, the crossed polariser experiment would show:

a	No light transmitted for any orientation.
b	Varying intensity as the orientation changes.
c	Uniform non-zero intensity of transmitted light regardless of orientation.
d	Circular dark patches to represent the symmetry of the optical properties.

Answer

a

3. Bubbles in a box behave in a similar way to grains in a crystal in several ways, but not in all. Which of the following statements is TRUE?

a	The geometry of the places where bubbles meet one another is different from the geometry of the junctions between grains in a real polycrystal.
b	The shape of the bubbles is different from the typical shape of a grain.
c	The way a bubble deforms when a load is applied is different from the way a grain deforms.
d	The three dimensional structure of bubbles in a box is unlike the three dimensional structure of a polycrystal.

Answer

c

4. Which of the following is false?

a	Quartz crystals have optically anisotropic properties.
b	Glass has no regular repeating crystalline structure.
c	Certain crystals may cleave easily along certain planes, defined by the crystal structure.
d	Crystal defects are not found in single crystals.

Answer

d

5. What does the 'shot model' fail to show?

a	Polycrystallinity
b	Crystalline defects
c	The third dimension of the structure
d	The difference between a vapour and a solid

Answer

c

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

6. Why is window glass transparent?

hint: Think carefully about the materials discussed in this package. Do the transparent ones have a common structure type (single crystal, polycrystal, amorphous)? How does the density of an amorphous material compare to a similar crystalline material?

6.

a	Because it has a single crystal structure and each sheet is cut with the optic axis normal to the plane of the window.
b	Because it has an amorphous structure with large interatomic spacing. Light waves can pass between widely spaced atoms without any interaction with the solid structure.
c	Because sheets of glass are cut thin enough for light to pass through without any significant absorption.
d	Because of the electronic nature of the bonds between the atoms in the glass.

7. A quantity of pure liquid aluminium is cooled slowly through its melting point. The solid is then left at room temperature for 100 years. What is the resulting structure?

a	A polycrystal with grains of identical chemical composition but different crystallographic orientation.
b	A polycrystal consisting of finely spaced lamellae with alternating composition.
c	A single crystal.
d	An amorphous solid with good mechanical strength.

Answer

A is incorrect. Window glass is not a single crystal.

B is incorrect. Think carefully about the materials discussed in this package. Do the transparent ones have a common structure type (single crystal, polycrystal, amorphous)? How does the density of an amorphous material compare to a similar crystalline material?

C is incorrect. Glass remains transparent even when it is very thick (although impurities such as water tend to decrease the transparency). This is why fibre-optic cables, made from glass, can carry signals for kilometres.

D is correct. The optical properties of materials largely depend on their electronic nature. Metals, which have lots of free electrons, interact strongly with light and are opaque and reflective. Window glass and quartz crystals have similar compositions and bonding, with few free electrons, and both are transparent. When glass crystallises, it may lose its transparency because the nature of the bonding at the grain boundaries changes, leading to optically-absorbing electronic states.

8. Self-diffusion is the diffusion of a species within a body of material made from the same species. In general, self-diffusion in a polycrystalline solid can occur through the bulk of the grains (*lattice diffusion*) or along the grain boundaries (*grain boundary diffusion*). Which of the following statements gives the *best* description of the relative contribution of each process to the overall diffusion rate?

hint: Re-read the sections in this tutorial on polycrystals and defects. Are grain boundaries likely to be easy paths for diffusion? Consider what happens when the temperature is increased - what is the overall effect on the rate of diffusion? Look at the micrographs of grains again. Is most of the material 'grain'or 'grain boundary'?

a	The contributions should be about the same in both cases.
b	The contribution from lattice diffusion will always be greater than the contribution from grain boundary diffusion.
c	The contribution from grain boundary diffusion will always be greater than the contribution from lattice diffusion.
d	The relative contributions of the two processes depend upon the temperature of the material.

Answer

A is incorrect. The structure of a grain boundary is very different to that of the bulk lattice, so in general one would expect the rate of bulk diffusion to be different in some way to the rate of lattice diffusion. In addition, the material contains different volume fractions of 'grain' and 'grain boundary'. Only in special cases do the two processes contribute equally to the overall diffusion rate.

B is incorrect. The contribution from lattice diffusion is *sometimes* greater than the contribution from grain boundary diffusion.

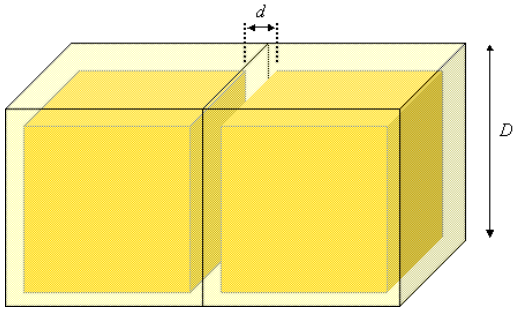
C is incorrect. The contribution from grain boundary diffusion is *sometimes* greater than the contribution from lattice diffusion.

D is correct. At low temperatures, diffusion is much easier along the grain boundaries, which are more disordered and less dense than the bulk lattice. Therefore grain boundary diffusion dominates. As the temperature is increased, the rate of diffusion along both pathways increases - diffusion is *thermally activated*. There is normally much more 'grain' than 'grain boundary' in a material, so at a high enough temperature the contribution to the overall rate of diffusion from lattice diffusion becomes greater than the contribution from grain boundary diffusion. Note that the definition of 'low' and 'high' temperature depends upon the *material* in question (its grain size, crystal structure *etc*.)

9. Imagine a polycrystalline solid with cubic grains of edge length D. When D = 10 μm, what percentage of the volume of solid lies within a grain boundary, if the grain boundary width d is 1 nm? What must the grain size D be if 10% of the volume lies within a grain boundary? Comment on your answers.

Answer

The following diagram shows two cubic grains side-by-side, with the dimensions D and d indicated.



The volume of a single grain is D^3. The volume of the grain boundary on one face of the grain is $1/2D^2d$, where the factor of 1/2 accounts for each grain boundary being shared by two grains. The total volume of grain boundary per grain is therefore 6 x $1/2D^2d$. The percentage of material (by volume) occupying the grain boundaries is then

$$\text{Percentage} = 100 \times 3D^2d / D^3 = 300d / D$$

So for D = 10 μm, the percentage of material on grain boundaries is

$$300 \times 1 \text{ nm} / 10 \text{ μm} = 0.03\%$$

Only a small proportion of the atoms of a polycrystal with a reasonably fine grain size lie within grain boundaries. One might therefore expect the properties of a polycrystal to be virtually identical to those of a single crystal, since the vast majority of the material is well ordered and crystalline within the grains. However, experimentally we observe that this is rarely the case. The grain boundaries in a material are of critical importance in many phenomena - see, for example, question 8 on diffusion.

If the percentage of material on grain boundaries is 10 %, then

$$D = 300d / \text{percentage} = 300 \times 1 \text{ nm} / 10 = 30 \text{ nm}$$

A significant fraction of the material is in a 'grain boundary' environment only for the very finest microstructures. A material with a grain size of 30 nm might be classed as a 'nanomaterial'. Often, the properties of nanomaterials are very different from similar materials with conventional grain sizes. An interesting example is in the superplastic forming of ceramics with ultrafine grain sizes. The grain size of the ceramic in this article is around 200nm.

Which of the following material properties could show anisotropy? *(answer yes or no for each)*

a	Density (see)
b	Young's modulus (see)
c	Electrical conductivity (see)
d	Refractive index (see)

Answer

No. When defined as mass / volume, density is isotropic because its value is independent of the direction in which it is measured (provided that the volume sampled has dimensions greater than the dimensions of a unit cell of the material).

Yes. The Young's modulus of many materials is anisotropic. Graphite, for example, has a Young's modulus of about 1000 GPa in directions parallel to the planes of atoms, compared to 36 GPa in the 'c' direction (perpendicular to the planes). Examine the structure of graphite here. In polycrystals, the individual grains can have anisotropic properties but the bulk material may be isotropic if the grain orientations are truly random.

Yes. Electrical conductivity depends on the electronic nature of the bonding within the structure. If the bonding is different in different directions, as can be seen clearly in the crystal structures shown in this module, then the conductivity is also likely to vary. Polycrystalline materials are less likely to be electrically anisotropic than single crystals. The refractive index depends on the electronic nature of the bonding in the crystal, as seen in section Crystals 3. Materials can be *optically isotropic* (one unique refractive index), or *uniaxially* or *biaxially anisotropic* (two and three unique refractive indices respectively). In polycrystalline samples, each individual grain can exhibit optical anisotropy. Some single crystal materials, such as metallic single crystals, are optically isotropic and have only one unique refractive index.

Yes.

Open-ended questions

The following questions are not provided with answers, but intended to provide food for thought and points for further discussion with other students and teachers.

11. Think about some of the possible applications of materials showing optical anisotropy, like the quartz crystal.
12. How might you control the grain size of a material produced from a melt? How might the cooling rate and the chemical composition affect the results? Can you think of ways to change the grain structure *after* the material has solidified?
13. In this TLP, we have discussed *pure* materials. Real materials almost always contain some impurities. How might these impurities be incorporated into the crystal structure of a material? Consider the relative size of the impurity atoms and the host atoms. Are impurities always undesirable?

14. Graphite is sometimes used as a lubricant, and diamond can be used on the tips of cutting tools. In terms of the crystal structure, why might this be?
15. Why do the individual grains in a polycrystalline material, such as those in the photo of galvanised steel (on the Polycrystals page) appear to be different colours or shades, when the composition of every grain is approximately the same?

This page titled 4.8: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

5: Avoidance of Crystallization in Biological Systems

www.doitpoms.ac.uk/tlplib/bi...stal/index.php

Learning Objectives

- Understand how nucleation and crystallization occur.
- Be aware of the reasons why crystallization in the cells of living systems is usually fatal.
- Appreciate the importance of sugars as solutes in the cellular liquid.
- Understand the ways in which some living systems avoid crystal formation.

5.1: Introduction
5.2: Nucleation and crystallization
5.3: The water-sucrose system
5.4: Avoidance of crystallization by glass formation
5.5: Avoidance of crystallization by freeze resistance
5.6: Summary
5.7: Questions

This page titled 5: Avoidance of Crystallization in Biological Systems is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

5.1: Introduction

Before You Start

Before beginning this TLP, you should:

- Understand basic concepts including osmosis.
- Be familiar with phase diagrams - take a look at the Phase Diagrams and Solidification TLP.
- Be familiar with the formation and structure of glass

Introduction

Many plants and animals live in extreme climate zones, such as the polar regions or subtropical deserts. One of the major problems associated with living in such regions is that most biological systems are largely composed of water. Water is the universal solvent for biological processes, and most living organisms contain a large amount — humans, for example, are 50-75% water.

In the polar regions temperatures can fall to below –20°C. In deserts, the absence of water can lead to the dehydration of plants and animals. The organisms that live in these regions have become adapted to these conditions.

Plants that live in arctic regions must contend with the formation of ice crystals in their cells. This is fatal to living tissue for a variety of reasons, which are discussed in the next section. Plants and animals living in the desert have a similar problem; dehydration can cause salts and sugars to precipitate out of solution inside their cells.

In this TLP we will consider how biological adaptations to cells allow them to survive such conditions by preventing crystallization of ice or salts within cells. We will look at the basic theory behind nucleation and crystallization, and the details of the water-sucrose system that illustrate key features of the cellular liquid (the *cytosol*). We will then examine the specific ways in which some plants and animals avoid crystallization, through glass formation, extracellular crystallization (forming the crystals outside the cell) and the use of antifreeze proteins.

This page titled 5.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

5.2: Nucleation and crystallization

Living cells consist essentially of an aqueous solution contained within a cell membrane. Thus the soft tissues of many living systems can be described as structured water. The human body for instance, is 50-75 % water. This high proportion of water means that crystallization in the body occurs in two main ways:

- by *dehydration*, when minerals crystallize from a saturated solution.
- by *freezing*, when ice crystals are formed.

The formation of crystals in living cells is usually fatal. This is either due to a change in the ionic ratios in the cytosol or due to the bursting of the cell.

During dehydration, water is removed from the cell, leading to a supersaturated solution. Mineral or sugar crystals can then form, changing the ionic ratios in the cytosol.

Ice crystals can form in the cells of both plants and ectothermic (cold-blooded) animals. Since ice is essentially pure H_2O, ice formation can increase the concentration of minerals in the remaining cytosol to a toxic level. The increased mineral concentration in the cytosol will cause water to be drawn in from the surrounding cells by osmosis, which can cause the cell to swell and burst.

In dehydration, the crystals that are formed can puncture the cell membrane causing the cells to burst, leading to death.

Both dehydration and ice formation involve the nucleation and growth of a new, solid phase from an aqueous solution. In the case of ice formation, the situation is effectively that of a solid crystallizing from a melt. In the case of the formation of mineral crystals, the case is that of precipitation from solution. Nucleation is the formation of a small cluster (or nucleus) of the new phase, and these nuclei arise spontaneously. Nuclei that are smaller than a certain size will simply disappear, but if a nucleus is greater than a certain size, it will spontaneously grow and will eventually form a grain. This critical size varies with temperature and the reasons for this are outlined below, using the example of ice forming in water.

Nucleation can occur either homogeneously (nucleation in a uniform phase in which there are no inhomogeneities on which nucleation can preferentially occur) or heterogeneously (in which the new phase nucleates on an inhomogeneity).

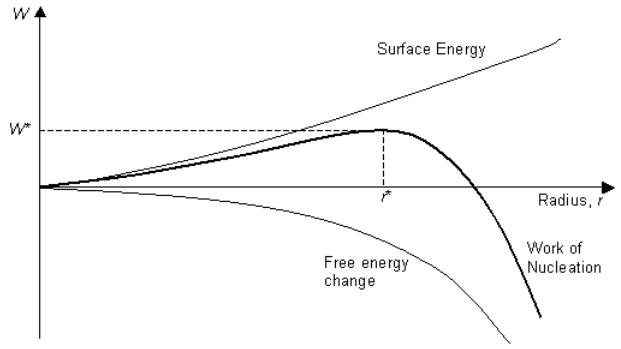
For the nucleation of ice in pure water, the transformation is a structural change only (there is no change in the chemical composition), and the change in free energy per unit volume on transformation is ΔG_v. The interface between the ice and water phase has a free energy γ per unit area. Due to the random motion of the water molecules, nuclei of ice will continually form. Assuming that these nuclei are spherical with radius r, the work done in forming the nucleus is:

Work for nucleation = change in free energy of bulk phases + interface energy

or

$$W = \frac{4}{3}\pi r^3 \Delta G_v + 4\pi r^2 \gamma \tag{5.2.1}$$

Since the interface between the water and liquid can be considered to be a defect, it contributes an excess energy to the system, and γ is positive. γ is approximately constant over the relevant range of temperatures. ΔG_v varies with temperature (as described below), but if the transformation occurs spontaneously, (i.e. if the temperature is below the melting temperature of ice), then ΔG_v is negative, and a graph of <W against r has the form:



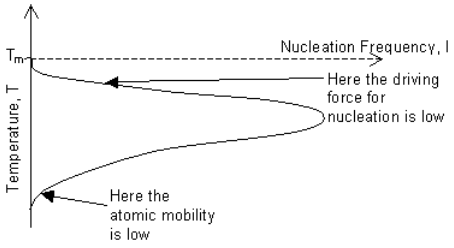
So, if a nucleus is formed, which has r > r*, it will decrease its energy by increasing r, i.e. by growing. Any nuclei with r < r* will decrease in energy by decreasing r and by disappearing. The critical radius, r* occurs when dW/dr = 0, giving:

$$r* = -\frac{2\sigma}{\Delta G_v} \tag{5.2.2}$$

and

$$W* = \frac{16\pi}{3}\frac{\sigma^3}{\Delta G_v^2} \tag{5.2.3}$$

We define ΔG to be the free energy difference between the solid and liquid phases, $\Delta G = \Delta G_{ice} - \Delta G_{water}$. Similarly we define the differences in enthalpy ΔH and entropy ΔS. Since ΔG = ΔH - TΔS and at T_m, the melting point of ice, ΔG = 0, then $\Delta H = T_m \Delta S$. If ΔH and ΔS are independent of temperature, then, at temperature T, $\Delta G = \Delta S(T_m - T) = \Delta S\, \Delta T$, where ΔT is the *supercooling* (also known as *undercooling*). The critical radius and the work for nucleation therefore decrease with decreasing temperature below T_m, and the rate of nucleation would increase with temperature below T_m. This effect is limited by the decrease in atomic mobility at lower temperatures, and the actual variation of nucleation frequency with temperature is shown below:

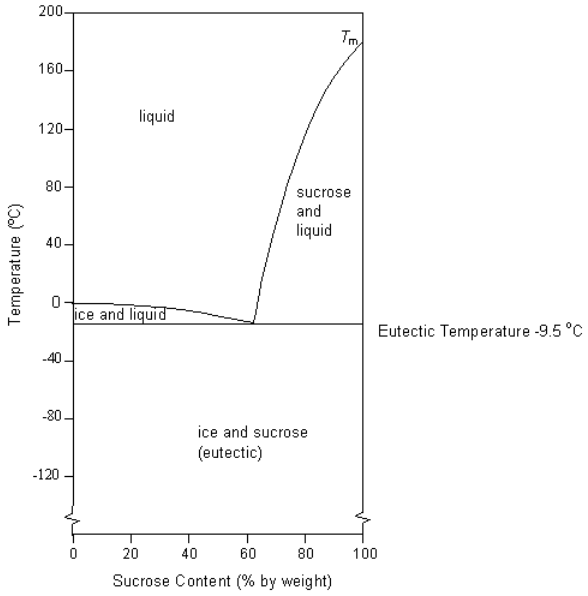


However, this analysis assumes homogeneous nucleation, which occurs only rarely. Usually there are heterogeneities, such as mould walls or cell membranes, in the melt onto which nucleation preferentially occurs. These heterogeneities are points with high excess energy and so the energy required to form the interface between the existing phase and the new phase is not so significant. Removing heterogeneities is one effective way of decreasing the temperature at which ice forms, i.e. increasing the difficulty of freezing.

This page titled 5.2: Nucleation and crystallization is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

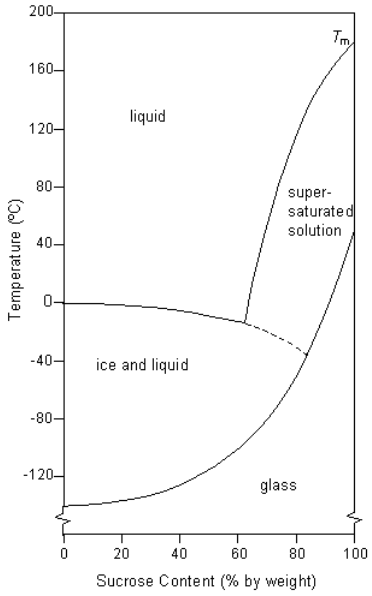
5.3: The water-sucrose system

Key features of the cytosol of cells are represented in the binary water-sucrose system. In practice, of course, cytosol compositions are much more complex. The equilibrium phase diagram for this system is shown below:



Due to kinetic factors, the equilibrium states shown in this phase diagram are rarely reached. Nucleation of sugar crystals is difficult, since the complexity of the sucrose molecule and the viscosity of the liquid make getting a critical mass of molecules with the right orientation uncommon. At high concentrations of sucrose the system becomes sufficiently viscous to be considered (at lower temperatures) as a glass.

The practical phase diagram is:



This page titled 5.3: The water-sucrose system is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

5.4: Avoidance of crystallization by glass formation

As previously discussed, a common biological problem is dehydration.

Crystal formation upon dehydration can be avoided by forming a glass. Glasses are amorphous, and form by a continuous process (no interface or solidification front is involved). Their structure is comparable to that of a liquid (some short-range order is observed, but no long-range order), but their properties are those of a solid.

In response to dehydration, some living systems alter the composition of the cytosol in order to favour glass formation, for example by hydrolysis of starch to sugar. As can be seen on the sucrose-water phase diagram, the higher the sugar content, the higher the temperature at which a glass can be formed. Using this mechanism, complete dehydration can be survived.

Examples of living systems that exploit the formation of glass in order to preserve life

- **Resurrection Plant**
  Various different species fall in the category of "resurrection plant", including the *Rose of Jericho* and *Selaginella lepidophylla*. When dry, they appear brown and lifeless, but after rain they become moist and green. This can be seen in the image below, which shows the *Selaginella lepidophylla* plant in both the moist and dry states.



Image reproduced with permission of Brad Fiero, Pima Community College, Tucson, Arizona

- **Flatworms**
  Under the stress of dehydration, some species of simple flatworm are able to convert starch into sugars, which promotes the formation of glass.
- **Seeds**
  Sucrose is the most abundant sugar within mature seeds, and its ability to form a glass in dry tissues greatly slows the chemical reactions within the seed that could lead to its degradation and hence contributes to the longevity of seeds. When seeds are planted after storage in the dry glassy state, they take up moisture from the ground and become rehydrated, and are then able to germinate.

Glass formation is similarly exploited in dried foods such as pasta, drug storage for drugs such as insulin for inhalation, and organ preservation. In these cases, the glassy state is used to inhibit degradation.

This page titled 5.4: Avoidance of crystallization by glass formation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

5.5: Avoidance of crystallization by freeze resistance

Many organisms exist in habitats where the temperatures fall below the freezing point of water. As previously explained, the formation of ice crystals in cells is lethal and many species have therefore evolved to prevent ice crystals forming in cells.

There are two different types of resistance to freezing temperatures: *freeze avoidance* and *freeze toleration.*

Freeze avoidance

During freezing of water ice crystals nucleate and grow. For pure water, homogenous nucleation occurs at ~ 40 K beneath the thermodynamic freezing point. This is a substantial supercooling, but in most cases, freezing occurs above –40°C due to heterogeneous nucleation.

One way to avoid freezing is to discourage heterogeneous nucleation. Some frost-hardened woods achieve this by dispersing the water in cells, and by the lack of nucleation on the cell walls. As a result, the water in their cells does freeze at –40°C.

Fish, insects and some plants that live in arctic regions have evolved to produce *antifreeze proteins*, which inhibit the growth of ice crystals by adsorption to the ice surface. Adsorption of these antifreeze proteins prevents crystal growth on the primary growth directions, forcing growth to occur parallel to the secondary axes. This inhibits the formation of stable ice crystals and lowers the kinetic freezing temperature.

Freeze toleration

An alternative to freeze avoidance is to promote the freezing of **extracellular** liquid. This protects the cells in two ways:

- By the release of latent heat into the cells, which prevents their temperature from decreasing.
- By drawing water out of the cells and decreasing the temperature at which ice forms. This depression of the liquidus line at higher sucrose concentrations can be seen in the sucrose-water phase diagram.
- Ultimately dehydration of the cells leads to glass formation.

Many biological systems promote heterogeneous nucleation by the presence of a variety of Ice Nucleating Agents (INAs). These INAs may either be adaptive or incidental. Adaptive INAs, which are discussed below, are present in order to promote heterogeneous nucleation, whereas incidental INAs (for example, features such as cell walls) promote heterogeneous nucleation only as a normally unwanted side effect.

Some organisms have evolved to produce adaptive INAs, which nucleate ice crystals between the cells. These INAs can reduce the nucleation supercooling to as little as 1°C. Examples of these are the giant rosette plant (*lobelia telekii*) which grows on Mount Kenya, and the northern wood frog (*rana sylvatica*) which lives in Canadian forests. The northern wood frog's body contains 35-45% ice during the winter months.

Adaptive INAs are generally large proteins with molecular weights of up to 30,000 atomic mass units. The amino acids within the proteins are ordered, forming a template for ice. Thus a thin layer of ice can always form on the surface of an INA. However, this will not lead to spontaneous ice growth unless the INA is of a certain critical size. If the INA is assumed to have a circular surface with radius R, then free ice growth will occur only when R is greater than or equal to r*. The critical radius, r*, is given by the equation:

$$r* = -\frac{2\gamma}{\Delta S_v \Delta T} \tag{5.5.1}$$

where γ is the interfacial energy per unit area, and $\Delta S_v \Delta T$ is the free energy of solidification per unit volume. Nucleation will occur when the supercooling ΔT satisfies the condition:

$$\Delta T \geq -\frac{2\gamma}{\Delta S_v R} \tag{5.5.2}$$

A larger INA (greater R) therefore gives a smaller required supercooling and a higher nucleation temperature.

This page titled 5.5: Avoidance of crystallization by freeze resistance is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and

standards of the LibreTexts platform.

5.6: Summary

In this TLP, you have learnt how the formation of crystals can occur inside cells by cooling or dehydration, and why this is usually fatal to them. The basic theory behind nucleation and crystallization has been introduced, and the water-sucrose system has been described as an approximation to the composition of cell cytosol.

You should appreciate that some plants and animals have adapted in order to avoid crystallization. The three main ways of achieving this are:

- through the formation of a glass.
- through extracellular crystallization (formation of the crystals outside the cell membrane).
- through the use of antifreeze proteins.

Going further

Books

- Fletcher, *Chemical Physics of Ice*, CUP,1970.
- Franks F, *Biophysics and Biochemistry at Low Temperatures*, CUP, 1985.
- Greenwood G W, Greer A L, Herlach D M, & Kelton K F, *Nucleation Control*, Phil. Trans. Royal Society (A 361, no 1804, 15 Mar 2003), pp. 403-633.

Websites

- Cryobiology - A Short Course
  An online course in cryobiology from the University of Calgary. Particularly useful are: chapter 1, which describes biological cells; chapter 6, on freezing as a crystallisation process; and chapter 12, on animal strategies for surviving in cold climates.
- Ice Formation in Higher-Plant Cells
  A page describing freeze avoidance in plants on the Physiology Online website.
- Resurrection Plant
  This page on the Union County College website describes the *Selaginella lepidophylla* resurrection plant.
- Nucleation in Metals and Alloys
  This section from the online manual for the MATTER project's Materials Science on CD-ROM, gives more in-depth information about nucleation. The page refers to the nucleation of metals, but the same applies for nucleation of ice or minerals in cells.
- Rana sylvatica (wood frog)
  This page on the Animal Diversity website of the University of Michigan Museum of Zoology describes the freezing capabilities of the Northern Wood Frog.

This page titled 5.6: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

5.7: Questions

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

1. Crystallization is usually fatal to cells because:

Yes	No	a	Crystals can puncture cell walls
Yes	No	b	The formation of crystals changes the ionic ratios in the cytosol
Yes	No	c	Crystallization speeds up the cell reactions
Yes	No	d	Ice formation causes water to be drawn into the cell by osmosis causing the cell to swell and burst
Yes	No	e	Nucleation of the crystals requires large amounts of energy

Answer

a. Yes

b. Yes

C. No

D.Yes

e. No

2. Ice formation in cells can be limited by:

Yes	No	a	The presence of antifreeze proteins
Yes	No	b	The dispersion of water in the cellular structure, which limits heterogeneous nucleation
Yes	No	c	The formation of starch from sugar
Yes	No	d	The introduction of Ice Nucleating Agents (INAs) into the cells
Yes	No	e	The freezing of extracellular water

Answer

a. Yes

b. Yes

C. No

D.No

e. Yes

3. Crystallization of minerals in the cells can be limited by:

Yes	No	a	The presence of anti-freeze proteins
Yes	No	b	The hydrolysis of starch to sugar
Yes	No	c	The preferential formation of a glass
Yes	No	d	The increased activity of the Golgi Apparatus

Answer

a. No

b. Yes

C. Yes

D.No

In some alpine plants, extracellular ice formation occurs at around –2°C. What is the critical radius for ice nucleation at this temperature?

In these plants, the extracellular ice forms on ice nucleating agents. If an ice nucleating agent is a circular disc and is a perfect template for ice, what size must it be for nucleation to occur at –2°C?

(For the ice-water interface, the interfacial energy γ = 0.028 J m^{-2}; the latent heat of freezing of ice is ΔH_v = -3.34 x 10^8 J m^{-3}.)

hint: The critical radius for nucleation is r^*, where r^* = -2γ /ΔG_v, and γ is the solid-liquid interfacial energy and ΔG_v is the free energy of freezing per unit volume. For further details see the section on nucleation and crystallization.

Answer

r^*, the critical nucleus for nucleation is given by, r^* = -2γ /ΔG_v where γ is the solid-liquid interfacial energy and ΔG_v is the free energy of freezing per unit volume. γ is given in the question and ΔG_v can be determined from ΔH_v. For freezing at a temperature T, the supercooling, ΔT is (T_m - T), where (T_m is the melting temperature. The entropy of fusion per unit volume is $\Delta S_v = \Delta H_v / T_m$. As shown in the section on Nucleation and Crystallization, $\Delta G_v = \Delta S_v \Delta T$. Thus:

$$r^* = -\frac{2\gamma}{\Delta G_v} = -\frac{2\gamma}{\Delta S_v \Delta T} = -\frac{2\gamma T_m}{\Delta H_v \Delta T} \tag{5.7.1}$$

With γ and ΔH_v as given, T_m taken to be 273 K, and T = 271 K, we find ***r** = 22.9 nm**

For ice to grow on the INA, the radius of the INA must be greater than or equal to r^*, so, for ice to grow, the INA should have radius 22.9 nm.

This page titled 5.7: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

6: Batteries

https://www.doitpoms.ac.uk/tlplib/batteries/index.php

Learning Objectives

- In this TLP, it is aimed to consider the basic electrochemical principles involved in the operation, design and use of batteries and the technical criteria relevant for battery selection.

6.1: Introduction
6.2: Basic principles
6.3: Thermodynamics and kinetics
6.4: Primary batteries
6.5: Zinc
6.5.01: Zinc
6.5.1: Zinc/carbon batteries
6.6: Alkaline
6.6.01: Alkaline
6.6.1: Alkaline/manganese oxide batteries
6.7: Questions
6.8: Zinc
6.8.01: Zinc
6.8.1: Zinc/silver oxide batteries
6.9: Secondary batteries
6.10: Lead
6.10.01: Lead
6.10.1: Lead/acid batteries
6.11: Lithium batteries
6.12: Battery characteristics
6.13: The future

This page titled 6: Batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.1: Introduction

Electrical energy/power storage has widespread applications as can be seen in the sheer range and diversity of batteries used as a source of electrical power. Batteries range from tiny button cells storing miliwatthours of energy and delivering microwatt power, to gigantic load levelling batteries situated within a factory size building, storing megawatthours of energy and delivering megawatts of power. It is predicted that in the year 2006 global demand for batteries will reach £33 billion pounds.

This page titled 6.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.2: Basic principles

The electrochemical series

Different metals (and their compounds) have different affinities for electrons. When two dissimilar metals (or their compounds) are put in contact through an electrolyte, there is a tendency for electrons to pass from one material to another. The metal with the smaller affinity for electrons loses electrons to the material with the greater affinity, becoming positively charged. The metal with the greater affinity becomes negatively charged. A *potential difference* between the two electrodes is thus built up until it balances the tendency of the electron transfer between the metals. At this point the potential difference is the *equilibrium potential*: the potential at which the net flow of electrons is 0.

The electrochemical series represents the quantitative expression of the varying affinity of materials relative to each other. In an aqueous electrolyte the standard electrode potential for an electrode reaction is expressed with respect to a reference electrode. Conventionally this is the H_2/H^+ cell, with reaction:

$$H^+ + e \leftrightarrow 1/2H_2 \tag{6.2.1}$$

Reaction	E^o (V)
$Li^+ + e^- \rightarrow Li$	-3.10
$Na^+ + e^- \rightarrow Na$	-2.71
$Mg^{2+} + 2e^- \rightarrow Mg$	-2.36
$½H_2 + e^- \rightarrow H^-$	-2.25
$Mn^{2+} + 2e^- \rightarrow Mn$	-1.18
$MnO_2 + 2H_2O + 4e^- \rightarrow Mn + 4OH^-$	-0.98
$2H_2O + 2e^- \rightarrow H_2 + 2OH^-$	-0.83
$Cd(OH)_2 + 2e^- \rightarrow Cd + 2OH^-$	-0.82
$Zn^{2+} + 2e^- \rightarrow Zn$	-0.76
$Ni(OH)_2 + 2e^- \rightarrow Ni + 2OH^-$	-0.72
$Fe^{2+} + 2e^- \rightarrow Fe$	-0.44
$Cd^{2+} + 2e^- \rightarrow Cd$	-0.40
$PbSO_4 + 2e^- \rightarrow Pb + SO_4^{2-}$	-0.35
$Ni^{2+} + 2e^- \rightarrow Ni$	-0.26
$MnO_2 + 2H_2O + 4e^- \rightarrow Mn(OH)_2 + 2OH^-$	-0.05
$2H^+ + 2e^- \rightarrow H_2$	0.00
$Cu^{2+} + e^- \rightarrow Cu^+$	+0.16
$Ag_2O + H_2O + 2e^- \rightarrow 2Ag + 2OH^-$	+0.34
$Cu^{2+} + 2e^- \rightarrow Cu$	+0.34
$O_2 + 2H_2O + 4e^- \rightarrow 4OH^-$	+0.40
$2NiOOH + 2H_2O + 2e^- \rightarrow 2Ni(OH)_2 + 2OH^-$	+0.48
$NiO_2 + 2H_2O + 2e^- \rightarrow Ni(OH)_2 + 2OH^-$	+0.49
$MnO_4^{2-} + 2H_2O + 2e^- \rightarrow MnO_2 + 4OH^-$	+0.62
$2AgO + H_2O + 2e^- \rightarrow Ag_2O + 2OH^-$	+0.64
$Fe^{3+} + e^- \rightarrow Fe^{2+}$	+0.77
$Hg^{2+} + 2e^- \rightarrow Hg$	+0.80
$Ag^+ + e^- \rightarrow Ag$	+0.80
$2Hg^{2+} + 2e^- \rightarrow Hg^+$	+0.91
$O_2 + 4H^+ + 4e^- \rightarrow 2H_2O$	+1.23
$ZnO + H_2O + 2e^- \rightarrow Zn + 2OH^-$	+1.26
$Cl_2 + 2e^- \rightarrow 2Cl^-$	+1.36
$PbO_2 + 4H^+ + 2e^- \rightarrow Pb^{2+} + 2H_2O$	+1.47
$PbO_2 + SO_4^{2-} + 4H^+ + 2e^- \rightarrow PbSO_4 + 2H_2O$	+1.70

$F_2 + 2e^- \rightarrow 2F^-$	+2.87

Cells using aqueous electrolytes are limited to under 2 V as water is decomposed at higher voltages. Thus Li batteries available in the 2.7 to 4 V use non-aqueous electrolyte. Typical non-aqueous electrolytes have higher impedance unless used as very thin films, eg lithium polymer battery or at higher temperatures.

For more information on this, click here.

What is a battery?

A battery is an electrochemical cell that converts chemical energy into electrical energy. It comprises of two electrodes: an anode (the positive electrode) and a cathode (the negative electrode), with an electrolyte between them. At each electrode a half-cell electrochemical reaction takes place, as illustrated by the animation below.

Discharge

Electrode 1 is an anode: the electrode is oxidised, producing electrons. Electrode 2 is a cathode: the electrode is reduced, consuming electrons. In the fully charged state, there is a surplus of electrons on the anode (thus making it negative) and a deficit on the cathode (thus making it positive). During discharge, electrons therefore flow from the anode to the cathode in the external circuit and a current is produced. Therefore in simple terms batteries work as electron pumps in the external circuit, preferably with only ionic current flowing through the electrolyte. The electrical potential difference between the cathode and the anode, which can drive the electrons in the external circuit, is called *electromotive force* (emf). Once all the active material at the cathode has been reduced, and/or all the active anodic material is oxidised, the electrode has effectively been used up, and the battery cannot provide any more power. It can then be either disposed of or preferably recycled if it is a primary battery, or recharged if it is a rechargeable (secondary) battery.

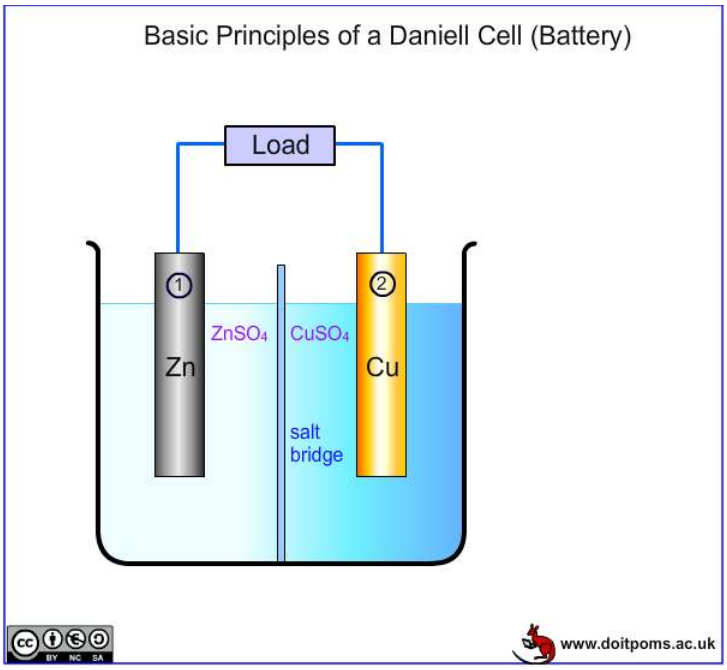
If the anode were zinc and the cathode were copper the half reactions would proceed as follows:

At the anode: $Zn \rightarrow Zn^{2+}(aq) + 2e^- \; E^o = 0.76V$

At the cathode: $Cu^{2+}(aq) + 2e^- \rightarrow Cu \; E^o = 0.34V$

Thus the total potential for this cell is 1.10 V.

During use as a battery, discharge leads to dissolution of Zn at the anode and the deposition of Cu at the cathode. Such a cell is embodied in the Daniell Cell introduced in 1836. As a practical cell this required two electrolytes (typically zinc sulphate and copper sulphate aqueous solutions) to avoid polarisation. The electrolytes are separated from each other by a salt bridge or a porous membrane, which allows the sulphate ions to pass and carry the ionic current, but blocks metallic ions. The Daniell Cell is an effective battery but not practical for portability. More recently, however, the idea of using two separate electrolytes has been resurrected in the form of *redox batteries*.



Discharge

Electrode 1 is an anode: the electrode is oxidised, producing electrons. Electrode 2 is a cathode: the electrode is reduced, consuming electrons. In the fully charged state, there is a surplus of electrons on the anode (thus making it negative) and a deficit on the cathode (thus making it positive). During discharge, electrons therefore flow from the anode to the cathode in the external circuit and a current is produced. Therefore in simple terms batteries work as electron pumps in the external circuit, preferably with only ionic current flowing through the electrolyte. The electrical potential difference between the cathode and the anode, which can drive the electrons in the external circuit, is called *electromotive force* (emf). Once all the active material at the cathode has been reduced, and/or all the active anodic material is oxidised, the electrode has effectively been used up, and the battery cannot provide any more power. It can then be either disposed of or preferably recycled if it is a primary battery, or recharged if it is a rechargeable (secondary) battery.

If the anode were zinc and the cathode were copper the half reactions would proceed as follows:

At the anode: $Zn \rightarrow Zn^{2+}(aq) + 2e^{-}$ $E^{o} = 0.76V$

At the cathode: $Cu^{2+}(aq) + 2e^{-} \rightarrow Cu$ $E^{o} = 0.34V$

Thus the total potential for this cell is 1.10 V.

During use as a battery, discharge leads to dissolution of Zn at the anode and the deposition of Cu at the cathode. Such a cell is embodied in the Daniell Cell introduced in 1836. As a practical cell this required two electrolytes (typically zinc sulphate and copper sulphate aqueous solutions) to avoid polarisation. The electrolytes are separated from each other by a salt bridge or a porous membrane, which allows the sulphate ions to pass and carry the ionic current, but blocks metallic ions. The Daniell Cell is an effective battery but not practical for portability. More recently, however, the idea of using two separate electrolytes has been resurrected in the form of *redox batteries*.

Charge

When the cell potential is depleted the battery can be recharged. When a current is applied to the cell in the opposite direction the anode becomes the cathode, and vice versa. Thus electrode 2 that was oxidised upon discharge is now reduced and the electrode 1 that was reduced is now oxidised so the electrodes are returned to their former state, ready to be discharged again.

This time the anode would be copper and the cathode would be zinc, and the half reactions would proceed as follows:

At the anode: $Zn^{2+}(aq) + 2e^{-} \rightarrow Zn$ $E^{o} = -0.76V$

At the cathode: $Cu \rightarrow Cu^{2+}(aq) + 2e^{-}$ $E^{o} = -0.34V$

The minimum potential required for charging will be 1.10 V, as this is the potential of the cell. In reality much higher potentials will be required to overcome the polarisation.

Example 6.2.1

How to make a potato battery?

Equipment:

- A potato
- A clean copper coin (If necessary clean it by placing it in a fizzy drink for a few minutes, or clean with steel wool)
- Some aluminium foil
- Crocodile clips
- Wires
- A voltmeter or multimeter

Instructions:

Cut the potato in half and place the flat end on the foil. Push the copper coin into the potato. Attach crocodile clips to the coin and the foil, then wires to the crocodile clips. Now attach these wires to the voltmeter. The volt meter should give a reading, showing that the stored energy is available and a current can flow, and that you have produced a battery!

Why this happens?

The potato contains a mild phosphoric acid (H_3PO_4), which acts as the electrolyte.

At the copper coin there are two reactions that could take place. If there is sufficient oxygen present, oxygen ions will be reduced to oxygen gas in the reaction
$O_2 + 4H^{+} + 4e^{-} \rightarrow 2O_2^{-} + 2H_2O$ 1.23 V

If there is not sufficient oxygen, hydrogen ions from the acid are reduced to hydrogen gas in the reaction
$H^{+} + e^{-} \rightarrow \frac{1}{2}H_2$ 0 V

At the foil aluminium metal is reduced to aluminium ions in the reaction
$Al \rightarrow Al^{3+} + 3e^{-}$ 1.66 V

The total theoretical voltage across the potato cell is therefore 2.89 V in sufficient oxygen (Aluminium/air battery), or 1.66 V in insufficient oxygen (Aluminium/Hydrogen battery).

This also works with lemons, tomatoes, apples and other fruits!

This page titled 6.2: Basic principles is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.3: Thermodynamics and kinetics

The overall reaction in the cell can be described by two half-cell reactions: one for the anode and one for the cathode.

The cathodic reaction can be represented as:

$aA + ne^- \longleftrightarrow cC$, Standard electrode potential: E_C^o

where a is the number of moles of A etc.

The cathode is usually a metallic oxide or a sulphide, but oxygen is also used.

The anodic reaction can be represented as:

$bB - ne^- \longleftrightarrow dD$, Standard electrode potential: E_A^o

The anode is generally a metal, which is electrochemically oxidised to form a metal ion which is soluble in the electrolyte.

The anode and the cathode are connected internally through an electrolyte, which is an ionic conductor, thereby providing the medium for transfer of charge as ions between the two electrodes. It is typically a solvent containing dissolved salts, acids or bases. The electronic conduction in the electrolyte should be negligible in order to avoid self-discharge by internal short-circuiting.

The overall reaction is given by:

$$aA + bB \longleftrightarrow cC + dD \tag{6.3.1}$$

The change in standard free energy, ΔG^o, is given by:

$$\Delta G^o = -nFE^o \tag{6.3.2}$$

where E^o is the standard cell potential $= E_{\mathrm{C}}^o - E_A^o$

At conditions other than the standard state, E can be given as:

$$E = E^o - \frac{RT}{ZF}\ln\frac{(aC)^c(aD)^d}{(aA)^a(aB)^b} \tag{6.3.3}$$

where (aC) is the activity of C etc.

This gives:

$$\Delta G^o = -nFE - \frac{nRT}{F}\ln\frac{(aC)^c(aD)^d}{(aA)^a(aB)^b} \tag{6.3.4}$$

The change in ΔG^o of a reaction is the driving force for a battery, which enables it to deliver electrical energy.

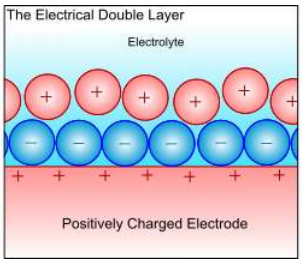
Electrode Kinetics (polarisation and cell impedance)

Thermodynamics can tell us the feasibility of a cell reaction occurring, and the theoretical cell voltage, however it is necessary to consider kinetics to gain a better idea of what the actual cell voltage may be, since rates of charge transfer are usually the limiting factor.

Electrical Double-Layer

When a metal electrode is in an electrolyte, the charge on the metal will attract ions of opposite charge in the electrolyte, and the dipoles in the solvent will align. This forms a layer of charge in both the metal and the electrolyte, called the *electrical double layer*. The electrochemical reactions take place in this layer, and all atoms or ions that are to be reduced or oxidized must pass

through this layer. Thus, the ability to pass through this layer controls the kinetics, and is therefore the limiting factor when controlling the electrochemical reaction.



Rate of reaction

The rates of the chemical reactions are governed by the Arrhenius relationship:
Rate of reaction:

$$\alpha \exp\left(\frac{-\Delta G}{RT}\right) \tag{6.3.5}$$

where:

ΔG^o is the activation energy for the reaction
T is the temperature in Kelvin
R is the universal gas constant

In this case, the rate of the reaction can be measured by the current produced, since current is the amount of charge produced per unit time, and therefore proportional to the number of electrons produced per unit amount of time; i.e. proportional to the rate.

Electrodes away from equilibrium

When an electrode is not at equilibrium an overpotential exists, given by

$$\eta = E - E^o \tag{6.3.6}$$

where η = overpotential, E = actual potential, E^o = equilibrium potential

The Tafel equation

Consider a general reaction for the oxidation of a metal at an anode:

$$M \rightarrow M^{z+} + ze^- \tag{6.3.7}$$

The rate of this reaction, k_a is governed by the Arrhenius relationship:

$$k_a = K \exp\left(\frac{-\Delta G}{RT}\right) \tag{6.3.8}$$

where K is the rate constant.

From Faraday's law:

$$i_o = zFk_a = zFK \exp\left(\frac{-\Delta G}{RT}\right) \tag{6.3.9}$$

If an overpotential is now applied in the anodic direction, the activation energy of the reaction is reduced to $(\Delta G - \alpha zF\eta)$, where α is the "symmetry factor" of the electrical double layer, usually 0.5.

Therefore

$$\begin{aligned} i_a &= zFK \exp\left(\tfrac{-(\Delta G - \alpha zF\eta)}{RT}\right) \\ &= zFK \exp\left(\tfrac{-\Delta G}{RT}\right) \exp\left(\tfrac{\alpha zF\eta}{RT}\right) \\ i_a &= i_o \exp\left(\tfrac{\alpha zF\eta}{RT}\right) \end{aligned} \tag{6.3.10}$$

This is the *Tafel equation*.

By taking natural logs and rearranging, this can be written as:

$$\eta = \left(\frac{RT}{\alpha zF}\right) \ln\left(\frac{i_a}{i_o}\right)$$
$$\eta = b_a \log\left(\frac{i_a}{i_o}\right) \tag{6.3.11}$$

Or, in terms of electrode potential,

$$\ln(i_a) = \ln(i_o) + (E - E^o)\left(\frac{\alpha zF}{RT}\right)$$
$$E = b_a \log\left(\frac{i_a}{i_o}\right) + a_a \tag{6.3.12}$$

, where b_a is the anodic Tafel slope.

Similarly, we can consider the reduction of metal ions at a cathode:

$$\mathrm{M}^{z+} + \mathrm{ze}^- \rightarrow \mathrm{M} \tag{6.3.13}$$

The activation energy will be decreased by $(1-\alpha)zF\eta$ giving:

$$i_c = i_o + \exp\left(\frac{(1-\alpha)zF\eta}{RT}\right) \tag{6.3.14}$$

and

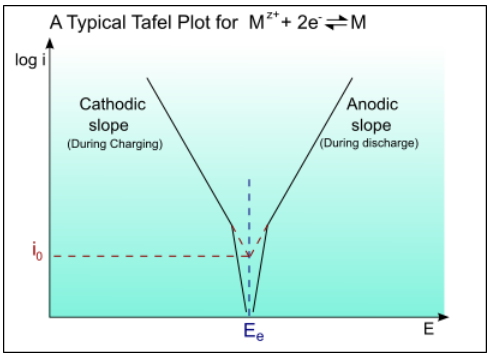
$$\eta = \frac{RT}{(1-\alpha)zF} \ln\left(\frac{i_c}{i_o}\right) \tag{6.3.15}$$

Therefore

$$E = b_c \log\left(\frac{i_c}{i_0}\right) + a_c \tag{6.3.16}$$

, where b_c is the cathodic Tafel slope.

The following is a typical Tafel plot – a plot of log i_o against E:



Thus for an applied potential, the current density can be found from the Tafel plot.

Example 6.3.1

Tafel curve of copper electrode

The half-cell reaction for copper is:

$$\mathrm{Cu}^{2+} + 2\mathrm{e}^- \rightarrow \mathrm{Cu};\ E^\circ = +0.34\mathrm{V} \tag{6.3.17}$$

The exchange current density is $i_o = 1\ \mathrm{A\ m^{-2}}$

For the Tafel equation:

$$\eta = a + b\log(i) \\ = a + b\log\left(\frac{i}{i_o}\right) \tag{6.3.18}$$

The Tafel slope is given by:

$$b = \pm 2.303\left(\frac{RT}{\alpha ZF}\right) \tag{6.3.19}$$

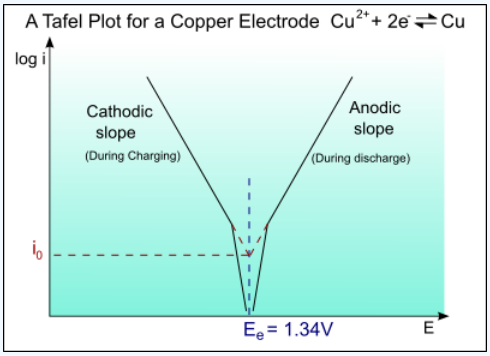
Taking:
T = 300 K
α = 0.5
Z = 2

We get:
b_A = 0.059 V/decade of current
and
b_C = -0.059 V/decade of current

For the anodic curve: $\eta_A = E_A - E^o$
$E_A = E^o + \eta_A$

For the cathodic curve: $\eta_C = E_C - E^o$
$E_C = E^o + \eta_C$

The plot of this is shown on the diagram below:



Other limiting factors:

At very high currents, a limiting current may be reached as a result of concentration overpotential, η_C(conc).

$$\eta_C(\text{conc}) = 2.303\left(\frac{RT}{zF}\right)\ln\left(\frac{i}{i_L}\right) \tag{6.3.20}$$

where i_L is the limiting current (in this case for the cathodic double layer).

The limiting current is diffusion limited, and can be determined by Fick's law of diffusion.

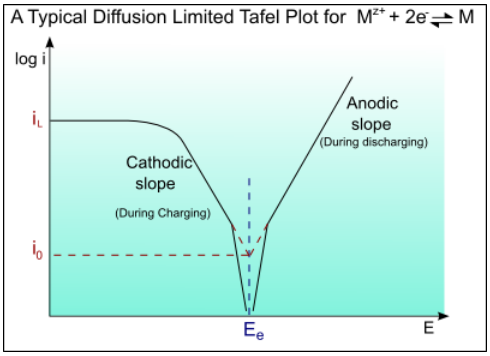
$$i_L = ZFD\left(\frac{C}{\delta}\right) \tag{6.3.21}$$

where:
D = diffusion coefficient of Cu^{2+} in the electrolyte
C = concentration of Cu^{2+} in the bulk electrolyte
δ = the thickness of the double layer

Typical values would be:
$D = 2 \times 10^{-9}\ m^2s^{-1}$
$C = 0.05 \times 10^4\ kg\ m^{-3}$

$\delta = 6 \times 10^{-4}$ m
which gives $i_L = 3.2 \times 10^2$ A m^{-2}

A Tafel curve showing this diffusion limiting of the current is shown below:



Tafel curves for a battery

In a battery there are two sets of Tafel curves present, one for each material. During discharge one material will act as the anode and the other as the cathode. During charging the roles will be reversed. The actual potential difference between the two materials for a given current density can be found from the Tafel curve: The anodic potential, E_A, and cathodic potential, E_C, can be found from the curve. The total cell potential is the difference between the two.

On discharge, the potential is always less than thermodynamics alone predicts. It can be calculated by the equation:

$$V'_{\text{cell}} = E_C - |\eta_C| + E_A - |\eta_A| \tag{6.3.22}$$

Upon discharge the cell potential may be further deceased by the Ohmic drop due to the internal resistance of the cell, r. Thus the actual cell potential is given by:

$$V_{\text{cell}} = V'_{\text{cell}} - iAr \tag{6.3.23}$$

where A = Geometric area relevant to the internal resistance.

Similarly on charging the potential is greater than thermodynamics alone predicts, and can be calculated by the equation:

$$V'_{\text{charge}} = E_C + |\eta_C| + E_A + |\eta_A| \tag{6.3.24}$$

The cell potential may now be increased by the Ohmic drop, and the actual cell potential is given by:

$$V_{\text{charge}} = V'_{\text{charge}} + iAr \tag{6.3.25}$$

This page titled 6.3: Thermodynamics and kinetics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.4: Primary batteries

Primary batteries are not easily rechargeable, and consequently are discharged then disposed of. Many of these are “dry cells” – cells in which the electrolyte is not a liquid but a paste or similar. The cell electrochemical reactions are not easily reversible and cell is operated until the active components in the electrodes are exhausted. Generally primary batteries have a higher capacity and initial voltage than rechargeable batteries.

Applications:

- Portable devices
- Lighting
- Toys
- Memory back-up
- Watches/clocks
- Hearing aids
- Radios
- Medical implants
- Defence related systems such as missiles

Advantages:

- Inexpensive
- Convenient
- Lightweight
- Good shelf life
- High Energy density at low/moderate discharges

Disadvantages:

- Can only be used once
- Leads to large amount of waste batteries to be recycled
- Batteries put into landfill sites have severe environmental impact
- Life cycle energy efficiency < 2 %

The table below demonstrates the properties of various primary batteries:

System	Nominal Cell Voltage (V)	Capacity (Wh/kg)	Advantages	Disadvantages	Applications
Carbon/Zinc	1.50	65	Lowest cost; variety of shapes and sizes	Low energy density; poor low-temperature performance	Torches; radios; electronic toys and games
Mg/MnO_2	1.60	105	Higher capacity than C/Zn; good shelf life	High gassing on discharge; delayed voltage	Military and aircraft receiver-transmitters
$Zn/Alk/MnO_2$	1.50	95	Higher capacity than C/Zn; good low-temperature performance	Moderate cost	Personal stereos; calculators; radio; TV
Zn/HgO	1.35	105	High Energy density; flat discharge; stable voltage	Expensive; energy density only moderate	Hearing aids; pacemakers; photography; military sensors/detectors
Cd/HgO	0.90	45	Good high and low-temperature performance; good shelf life	Expensive; low energy density	
Zn/Ag_2O	1.50	130	High Energy density, good high rate performance	Expensive (but cost effective)	Watches; photography; missiles; Larger space applications
Zn/Air	1.50	290	High Energy density; long shelf life	Dependent on environment; limited power output	Watches; hearing aids; railway signals; electric fences
$Li/SOCl_2$	3.60	300	High Energy density; long shelf life	Only low to moderate rate applications	Memory devices; standby electrical power devices
Li/SO_2	3.00	280	High energy density; best low-temperature performance; long shelf life	High-cost pressurized system	Military and special industrial needs
Li/MnO_2	3.00	200	High energy density; good low-temperature performance; cost effective	Small in size, only low-drain applications	Electrical medical devices; memory circuits; fusing

This page titled 6.4: Primary batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.5: Zinc

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.5: Zinc is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.5.01: Zinc

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.5.01: Zinc is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.5.1: Zinc/carbon batteries

This is commonly known as the Leclanché Cell and despite being the oldest type of battery it is still the most commonly used as it is very low-cost.



Georges Leclanché

The first cell was produced by Georges Leclanché in 1866 and was the first cell to contain only one low-corrosive fluid electrolyte with a solid cathode. This gave it a low self discharge in comparison to previously attempted batteries. The original cell consisted of a solid Zinc anode with an ammonium chloride solution as the electrolyte immobilized in the form of a paste (hence called a "dry cell"), and an 1:1 mixture of powdered carbon and manganese dioxide packed around a carbon rod acting as a cathode. In another version for extra heavy-duty applications, the electrolyte is zinc chloride mixed with a small amount of ammonium chloride. The most common variant is the alkaline cell where the electrolyte is potassium hydroxide.

Characteristics in brief

Voltage: 1.5 – 1.75 V

Discharge characteristics: Generally sensitive to external factors. Generally very sloped. Better when discharged intermittently.

Service Life: 110 min (continuous use)

Shelf life: ~ 1 – 2 years (at room temperature)

Chemistry

The zinc/carbon cell uses a zinc anode and a manganese dioxide cathode; the carbon is added to the cathode to increase conductivity and retain moisture; it is the **manganese dioxide** that takes part in the reaction, **not** the carbon.

The overall reaction in the cell is:

$Zn + 2MnO_2 \rightarrow ZnO + Mn_2O_3$

The exact mechanism for this is complicated, and there is still controversy over the exact mechanism, however the approximate half-cell reactions are:

Anode: $Zn \rightarrow Zn^{2+} + 2e^-$

Cathode: $2NH_4^+ + 2MnO_2 + 2e^- \rightarrow Mn_2O_3 + H_2O + 2NH_3$

However, this is complicated by the fact that the ammonium ion produces 2 gaseous products:

$2NH^{4+} + 2e^- \rightarrow 2NH_3 + H_2$

These products must be absorbed in order to prevent build up of pressure in the vessel. This occurs by 2 mechanisms:

$ZnCl_2 + 2NH_3 \rightarrow Zn(NH_3)_2Cl_2$

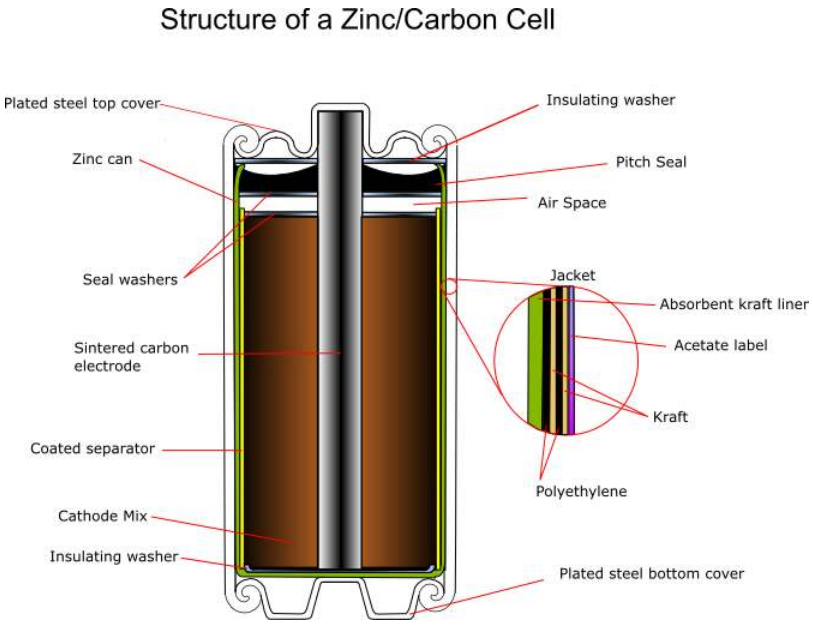
$2MnO_2 + H_2 \rightarrow Mn_2O_3 + H_2O$

Construction

The cell has two basic designs: the cylindrical cell and the flat cell.

Cylindrical Cell

The zinc serves as both the container and the anode. The manganese dioxide/carbon mixture is wetted with electrolyte and shaped into a cylinder with a small hollow in the centre. A carbon rod is inserted into the centre, which serves as a current collector. It is also porous to allow gases to escape, and provides structural support. The separator is either cereal paste or treated absorbent kraft paper (the kind of brown paper used to make large envelopes or grocery bags).



Carbon cathode

This is made of powdered carbon black and electrolyte. It adds conductivity and holds the electrolyte. The MnO_2 to Carbon ratios vary between 10:1 and 3:1, with a 1:1 mixture being used for photoflash batteries, as this gives a better performance for intermittent use with high bursts of current. Historically the carbon black was graphite, however acetylene black is often used in modern batteries as it can hold more electrolyte.

Manganese dioxide

everal grades of MnO_2 are available:

- Natural Manganese Dioxide: ores occur naturally in Gabon, Greece and Mexico with 70 – 85% MnO_2
- Activated Manganese Dioxide
- Chemically synthetic Manganese Dioxide: 90 – 95% MnO_2
- Electrolytic Manganese Dioxide (EMD): higher cell capacity and rate capabilities, and less polarization. Used in industrial applications.

Electrolyte

A standard Leclanché cell uses a mixture of ammonium chloride and zinc chloride in aqueous solution. A zinc-corrosion inhibitor is also added, which forms an oxide layer. This inhibitor is usually mercuric oxide or mercurous chloride.

A typical electrolyte composition is:

NH_4Cl	26.0%
$ZnCl_2$	8.8 %
H_2O	65.2 %
Corrosion inhibitor	0.25 - 1.0 %

Carbon rod

This is inserted into the cathode and acts as a current collector. It also provides structural support and vents hydrogen gas that evolves as the reactions proceed. When raw the rods are very porous, so must be treated with waxes or oils to prevent loss of water, but remain porous enough to allow hydrogen to pass through. Ideally they should also prevent oxygen entering the cell, as this would aid corrosion of the zinc.

Separator

This can be either a gelled paste, or kraft paper coated with cereal. It physically separates the anode and the cathode, but allows ionic conduction to occur in the electrolyte.

Paste: The paste is flowed into the zinc can, and the carbon cathode inserted, forcing the paste up the sides of the can between the zinc and the cathode, where it sets.

Paper: The paper is coated with cereal, or another gelling agent, rolled into a cylinder and along with a circular bottom sheet is added to the can. The carbon cathode is added, and the rod inserted, pushing the paper against the walls of the cans. This compression releases some electrolyte from the cathode mix, soaking the paper.

As the paste is relatively thick, more electrolyte can be held by the paper than the paste, giving increased capacity, thus paper is usually the preferred separator.

Seal

This can be asphalt pitch, wax/resin mix, or plastic (usually polyethylene or polypropylene) An airspace is usually left between the seal and the cathode to allow for expansion. The function of the seal is to prevent evaporation of the electrolyte, and to prevent oxygen entering the cell and corroding the zinc.

Jacket

The jacket provides strength and protection, and will hold the manufacturers label. It contains various components, which can be metal, paper, plastic, mylar, cardboard (sometimes asphalt-lined) and foil.

Electrical contacts

These are the terminals of the battery, and are tin plated steel or brass. They aid conductivity and prevent exposure of the zinc.

This page titled 6.5.1: Zinc/carbon batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.6: Alkaline

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.6: Alkaline is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.6.01: Alkaline

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.6.01: Alkaline is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.6.1: Alkaline/manganese oxide batteries

This system has a higher capacity than the zinc/carbon cell. It has a very good performance at high discharge rates and continuous discharge and at low temperatures. The first modern alkaline cell was developed in the 1960s and by 1970 it was produced all over the world. Currently over 15 billion alkaline cells are used worldwide each year.

Chemistry

The active materials used are the same as in the Leclanché cell – zinc and manganese dioxide. However the electrolyte is potassium hydroxide, which is very conductive, resulting in low internal impedance for the cell. This time the zinc anode does not form the container; it is in the form of a powder instead, giving a large surface area.

The following half-cell reactions take place inside the cell:

At the anode: $Zn + 2OH^- \rightarrow Zn(OH)_2 + 2e^-$
$Zn(OH)_2 + 2OH^- \rightarrow [Zn(OH)_4]^{2-}$

At the cathode: $2MnO_2 + H_2O + 2e^- \rightarrow Mn_2O_3 + 2OH^-$

For full discharge: $MnO_2 + 2H_2O + 2e^- \rightarrow Mn(OH)_2 + 2OH^-$

Overall: $Zn + 2MnO_2 \rightarrow ZnO + Mn_2O_3$

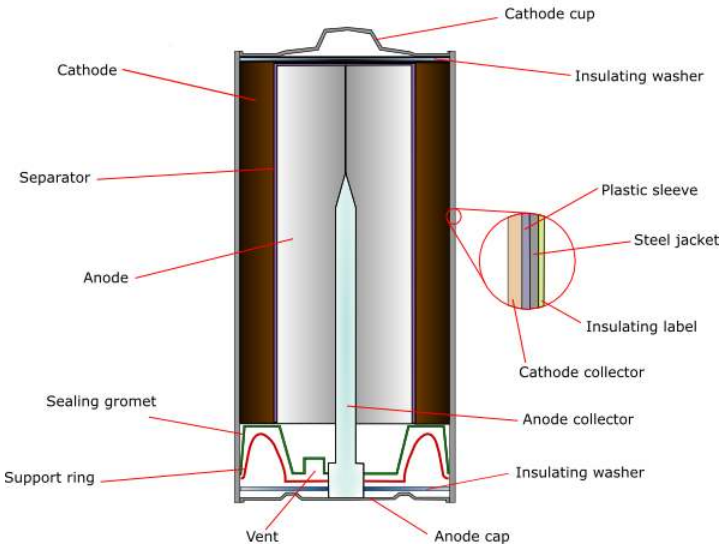
For full discharge: $Zn + MnO_2 + 2H_2O \rightarrow Mn(OH)_2 + Zn(OH)_2$

It is not possible to describe the cathodic reaction on discharge in a simple unambiguous way, despite a lot of research. In fact the discharge curve has two fairly distinct sections corresponding to change in the oxidation state of Mn from +4 to +3 and then to +2 during the reduction of MnO_2. The reality is more complicated than described in the two reactions shown above.

Construction

This cell is “inside out” compared to the Leclanché cell - the manganese dioxide cathode is external to the zinc anode, giving better diffusion properties, and lower internal resistance.

Structure of an Alkaline/Manganese dioxide cell



Cathode

For an alkaline cell electrochemically produced MnO_2 must be used. The ore rhodochrosite ($MnCO_3$) is dissolved in sulfuric acid, and electrolysis is carried out under carefully controlled conditions using titanium, lead alloys or carbon for the electrode onto which the oxide is deposited. This gives the highest possible purity, typically 92 ± 0.3%.

The cathode itself also contains around 10% graphite – more for more powerful batteries. A typical composition would be:

70% MnO_2 (of which 10% is water);
~10% graphite;
1-2% acetylene black;
Balance: binding agents and electrolyte.

Zinc Anode

The zinc must be very pure (99.85 – 99.90%) and is produced by electroplating or distilling. Very small amounts of lead are sometimes added to help prevent corrosion (usually ~0.05%) The zinc is powdered by discharging a small stream of molten zinc into a jet of air “atomising” it. The powder contains particles between 0.0075 and 0.8 mm.

There are two methods of formation of the anodes from the powder:

- **Gelled anodes:** These contain around 76% Zn, 7% mercury, 6% sodium carboxymethyl cellulose and 11% KOH solution. It is extruded into the cell, as the viscosity is high. In very small cells, NaOH is added to reduce creepage around the seal area. However this mixture is not ideal: it does not fully utilize the zinc at high current densities. Two-phase anodes have therefore been developed, consisting of a clear gel phase and a more compact zinc-powder gel phase, which enables 90% zinc usage.
- **Porous anodes:** The zinc powder is wetted with mercury and cold pressed, welding the particles together. The porosity can be controlled by materials such as NH_4Cl or plastic binders if required, which can be removed later. These anodes can carry very high currents.

Separators

These cells usually use “macro porous” separators. These are made from woven or felted materials.

This page titled 6.6.1: Alkaline/manganese oxide batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.7: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

1. Which of the following statements is not true?

a	A battery is an electrochemical cell.
b	If an electrode acted as the anode during discharge it will act as a cathode during charging.
c	During the operation of a battery, one material is oxidised and another is reduced.
d	The voltage needed to charge a battery is less than the voltage produced when it is discharged.

Answer

d

2. What can thermodynamics and kinetics tell us about the reactions that occur in a battery?

a	They both tell us if it is possible for the reaction to occur or not.
b	Thermodynamics tells us if it is possible for a reaction to occur or not, kinetics tells us what the rate of this reaction will be.
c	They both tell us what the rate of the reaction will be.
d	Kinetics tells us if it is possible for a reaction to occur or not, thermodynamics tells us what the rate of this reaction will be.

Answer

b

3. Which of the following pieces of information cannot be gained from a Tafel plot?

a	The maximum voltage that can be expected whilst drawing or supplying a particular current density.
b	The minimum voltage that can be expected whilst drawing or supplying a particular current density.
c	The anodic and cathodic potentials in a battery for a given current density.
d	The diffusion limited current density.

Answer

b

4. Are the following primary or rechargeable batteries?

(a) Zinc/Carbon battery

(b) Lithium ion battery

(c) Lead/acid battery

(d) Alkaline/Managanese dioxide battery

(e) Nickel/Cadmium battery

(f) Zinc/Air battery

(g) Zinc/Silver Oxide battery

(h) Silver/Hydrogen battery

(i) Potato battery

Answer

Primary batteries: a, d, f, g, i

Secondary batteries: b, c, e, h

5. Using the electrochemical series, what would the theoretical potential difference be in the following cells, if only thermodynamics was considered?

(a) Zinc and copper

(b) Nickel and mercury

(c) Zinc and silver

(d) Lithium and silver

Answer

(a) 1.10 V

(b) 1.06 V

(c) 2.36 V

(d) 3.90 V

6. What batteries might you use to power the following applications?

(a) A mobile phone

(b) A hand-held computer console

(b) A wristwatch

(d) To start a car

Answer

(a) Lithium

(b) Zinc/Carbon

(c) Zinc/Silver Oxide

(d) Lead/Acid

This page titled 6.7: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.8: Zinc

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.8: Zinc is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.8.01: Zinc

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.8.01: Zinc is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.8.1: Zinc/silver oxide batteries

The zinc/silver oxide batteries (first practical zinc/silver oxide battery was developed in the 1930's by André; Volta built the original zinc/silver plate voltaic pile in 1800) are important as they have a very high energy density, and can deliver current at a very high rate, with constant voltage. However the materials are high cost, so it is limited to application in button cells, for use in calculators, watches hearing aids and other such applications that require small batteries and long service life.

Characteristics in brief

Voltage: around 1.6 V, linearly dependent on temperature.

Discharge characteristics: Very good – flat discharge curve.

Service life: several thousand hours (continuous use).

Shelf life: several years (at room temperature).

Chemistry

The silver oxide used is usually in the monovalent form (Ag_2O), as it is the most stable.

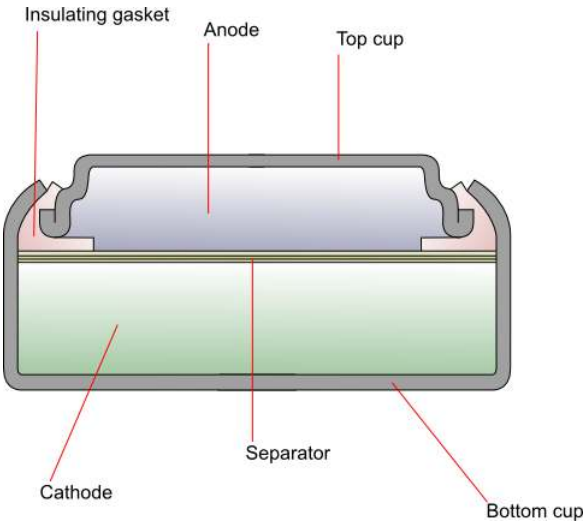
The following reactions take place inside the cell:

At the anode:
$Zn + 2OH^- \rightarrow Zn(OH)_2 + 2e^-$

At the cathode:
$Ag_2O + H_2O + 2e^- \rightarrow 2Ag + 2OH^-$

Overall:
$Ag_2O + H_2O + Zn \rightarrow 2Ag + Zn(OH)_2$

Construction



The cathode is generally composed of monovalent silver oxide with added graphite to improve conductivity. The anode is zinc powder mixed with a gelling agent, which is then dissolved in the alkaline electrolyte. The two are separated by a combination of layers of grafted plastic membrane, treated cellophane and non-woven absorbent fibres. The top cup (negative terminal) is made up of laminated layers of copper, tin, steel and nickel, and the bottom cup (positive terminal) is nickel-plated steel. An insulating gasket prevents contact between the two.

This page titled 6.8.1: Zinc/silver oxide batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.9: Secondary batteries

Secondary (rechargeable) batteries can be recharged by applying a reverse current, as the electrochemical reaction is reversible. The original active materials at the two electrodes can be reconstituted chemically and structurally by the application of an electrical potential between the electrodes to “inject” energy. These batteries can be discharged and recharged many times.

Applications:

These fall into two categories:

(a) The battery is used as an energy storage device. It is constantly connected to an energy source and charged by it. It can then release the stored energy whenever needed, e.g. in

- Car battery used to start engine
- Aircraft systems
- Standby power resources
- Emergency no-fail systems

(b) The battery is used as a primary battery would be but is then recharged instead of being disposed of, e.g. in

- Electric vehicles
- Mobile phones
- Cameras
- Power tools
- Toys
- Portable computers

Advantages:

- High power density
- High discharge rate
- Good low temperature performance

Disadvantages:

- Lower Energy density
- Poorer charge retention
- Safety issues
- Lack of standards
- High initial costs

The table below demonstrates the properties of various rechargeable batteries:

System	Nominal Cell Voltage (V)	Capacity (Wh/kg)	Advantages	Disadvantages	Applications
Pb/Acid	2.00	35	Low cost; good high and low-temperature operation	Low cycle life; low energy density; poor charge retention	Cars; lawn mowers; aircraft
Ni/Cd	1.20	30	Good physical durability; good charge retention; good cycle life	High cost; memory effect	Aircraft; emergency power applications
Ni/Fe	1.20	60	Good physical durability; long cycling and standing life	Low power and energy density; high self discharge; high cost	Stationary applications; fork lift trucks
Ni/Zn	1.60	27	High energy density; low cost; good low-temperature performance	Poor cycle life	Electric scooters/bikes; military vehicles
Zn/AgO	1.50	90	Highest energy density; low self discharge; high discharge rate	High cost; low cycle life; low performance at low temperatures	Military equipment eg torpedo propulsion, submarines
Cd/AgO	1.20	55	High energy density; low self discharge; Good cycle life	High cost; low performance at low temperatures	Portable power tools; satellites
Ni/H_2	1.40	55	High energy density; good cycle life; can tolerate over charge	High initial cost; self discharge proportional to H_2 pressure	Aerospace
Ag/H_2	1.40	80	High energy density; good cycle life	High cost - limited to military and aerospace applications	Aerospace
Li/Poly	up to 4.2	135	High specific energy; good shelf life; moldable; non-volatile	High cost; expensive control methods needed for charge/discharge	Mobile phones

This page titled 6.9: Secondary batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.10: Lead

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.10: Lead is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.10.01: Lead

This page was auto-generated because a user created a sub-page to this page.

This page titled 6.10.01: Lead is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.10.1: Lead/acid batteries

The lead acid battery is the most used battery in the world. The most common is the SLI battery used for motor vehicles for engine **S**tarting, vehicle **L**ighting and engine **I**gnition, however it has many other applications (such as communications devices, emergency lighting systems and power tools) due to its cheapness and good performance.

It was first developed in 1860 by Raymond Gaston Planté. Strips of lead foil with coarse cloth in between were rolled into a spiral and immersed in a 10% solution of sulphuric acid. The cell was further developed by initially coating the lead with oxides, then by forming plates of lead oxide by coating an oxide paste onto grids. The electrodes were also changed to a tubular design.

Characteristics in brief (for an SLI battery)

Voltage: 2 V

Discharge characteristics: Generally quite curved, particularly at higher discharge rate. Best performance with intermittent discharge.

Service Life: Several years

Chemistry

The lead acid battery uses lead as the anode and lead dioxide as the cathode, with an acid electrolyte.

The following half-cell reactions take place inside the cell during discharge:

At the anode: $Pb + HSO_4^- \rightarrow PbSO_4 + H^+ + 2e^-$

At the cathode: $PbO_2 + 3H^+ + HSO_4^- + 2e^- \rightarrow PbSO_4 + 2H_2O$

Overall: $Pb + PbO_2 + 2H_2SO_4 \rightarrow 2PbSO_4 + 2H_2O$

During the charging process, the reactions at each electrode are reversed; the anode becomes the cathode and the cathode becomes the anode.

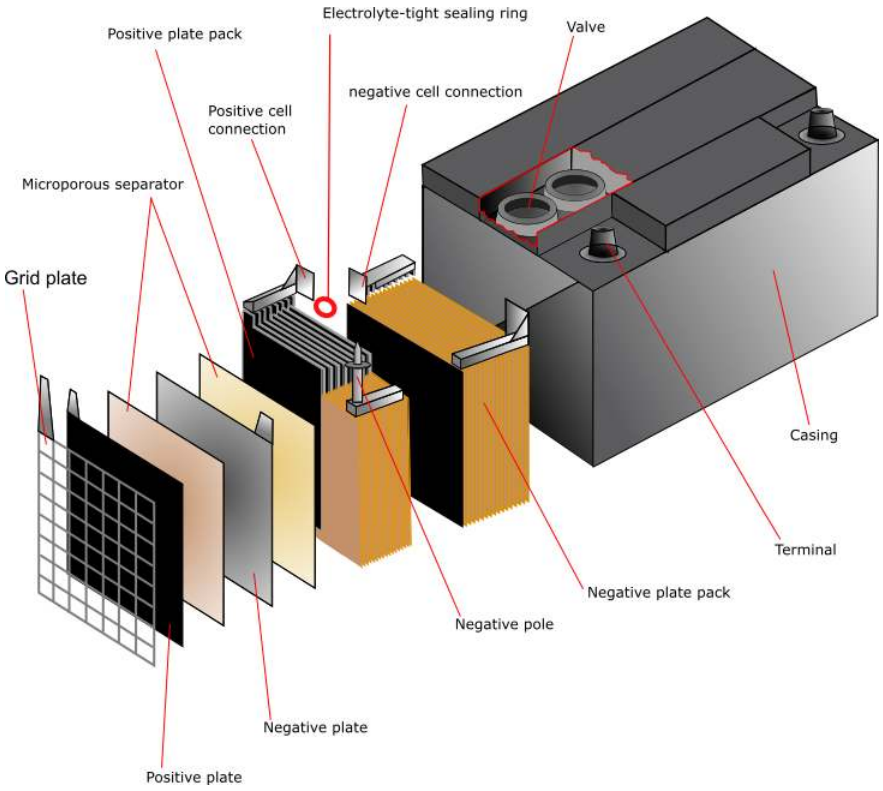
Gassing

During charging, given the high voltage, water is dissociated at the two electrodes, and gaseous hydrogen and oxygen products are readily formed leading to the loss of the electrolyte and a potentially explosive situation. Sealed batteries are made safer by allowing the gases to recombine within the cell.

Sulphation

Under certain circumstances the lead sulphate products at both the electrodes achieve an irreversible state, making the recharging process very difficult.

Construction



Lead

Pure lead is too soft to use as a grid material so in general the lead is hardened by the addition of 4 – 6% antimony. However, during the operation of the battery the antinomy dissolves and migrates to the anode where it alters the cell voltage. This means that the water consumption in the cell increases and frequent maintenance is necessary. There are two possible solutions to this problem:

(1) Using below 4% the battery water consumption is reduced, however it is then necessary to add small amounts of other elements such as sulphur, copper, arsenic and selenium. These act as grain refiners, decreasing the grain size of the lead and thereby increasing its hardness and strength.

(2) Alkaline earth metals such as calcium can be used to stiffen the lead. This is often used for telephone applications, and for no maintenance automotive batteries, since a more stable battery is required. A typical alloy would be 0.03 – 0.10% calcium and 0.5 – 1.0% tin (to enhance mechanical and corrosion properties).

The function of the grid is to hold the active material and to conduct electricity between the active material and the battery terminals. The design is a simple grid framework with a “tab” or “lug” for connection to the terminal post.

“Book mold” casting is the most common method of production for the grid. Permanent steel molds are made from blocks by machining. The molds are closed and filled with sufficient molten lead to fill the mold, leaving some excess to form a sprue, which is then removed by cutting or stamping. Grids can also be formed by mechanical working, either by cutting deep grooves into a sheet of steel, or by rolling up crimped strips and inserting them into holes in a cast plate, see Metal Forming TLP.

Lead Oxide

The lead can be oxidised by two processes: The Barton pot and the ball mill.

- Barton pot: A fine stream of molten lead is inserted into a heated vessel. Each droplet reacts with the air to form an oxide layer, giving 70 – 85% lead oxide.

- Ball milling: Pieces of lead are put into a rotary mechanical mill, forming fine lead flakes, which are then oxidised in air and removed. This also gives 75 – 80% lead oxide.

Red lead (Pb_3O_4) can also be added to the PbO formed by these methods, as it is more conductive. This is produced from PbO by roasting in a flow of air. This process would also increase the percentage of lead oxide in the material.

The oxide is mixed with water, sulphuric acid and a mixer, and then mixed to form a paste. It is then integrated with the grid by extrusion to form a plate. The paste is pressed by a machine into the interstices of the grid. They are partially dried, then stacked for curing. The curing process transforms the paste to a cohesive, porous solid. The most typical form of curing is “hydrosetting”: the grid is left at low temperature and humidity (25 – 40°C and 8 – 20% H_2O) for between 24 and 72 hours.

Assembly

The simplest cell would consist of one cathode plate, one anode plate and a separator between them. In practice, most cells contain up to 30 plates with separators between. The separators are usually cellulose, PVC, rubber, microporous polyethylene or non-woven polypropylene. The plates are stacked and welded together. The tabs that are fixed to the plates are cast, then punched on between the layers and welded together. The plates are suspended inside the case, which is filled with electrolyte in order to activate it.

This page titled 6.10.1: Lead/acid batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.11: Lithium batteries

One of the main attractions of lithium as an anode material is its position as the most electronegative metal in the electrochemical series combined with its low density, thus offering the largest amount of electrical energy per unit weight among all solid elements. In many applications the weight of the battery is a significant percentage of the total weight, and there is great competition to make lighter batteries.

Li cannot be used with the traditional aqueous electrolytes due to the very vigorous corrosive reaction between Li and water with flammable hydrogen as the product. It took many years to develop a suitable electrolyte based on organic solvents with sufficient stability and ionic conductivity. Ionic conductivity is induced by dissolving a suitable Li salt in an organic solvent used in the form of a gel or immobilised within a polymeric separator. In the 1980s progress was made in the use of Li as an anode material with MnO_2, liquid SO_2 or thionyl chlorides as the cathode, and hexaflurophosphate dissolved in propylene carbonate as a typical organic electrolyte. Li cells are generally properly sealed against contact with air and moisture Whilst the primary lithium battery has been well established for nearly two decades, there have been many problems experienced whilst developing rechargeable lithium batteries, mainly due to the extreme reactivity of lithium, and they have only become widely available since mid 1990s.

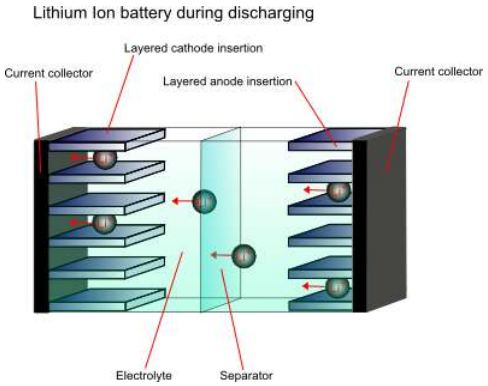
Rechargeable batteries

Li-ion batteries are now used in very high volumes in a number of relatively new applications, such as in mobile phones, laptops, cameras and many other consumer products. The typical Li-ion cells use carbon as the anode and $LiCoO_2$ or $LiMn_2O_4$ as the cathode. The first commercial Li-ion cell introduced by Sony in the 90's used a polymeric gel electrolyte, swollen by a high proportion of organic solvent for the Li salt. The translucent gel achieved a respectable ionic conductivity of 3×10^{-3} S cm^{-1} at 300 K.

Chemistry and construction

In order to overcome the problems associated with the high reactivity of lithium, the anode material is not purely the metal, it is a non-metallic compound, e.g. carbon, which can store and exchange lithium ions. A lithium ion-accepting material, for example CoO_2, is then used as the cathode material, and lithium ions are exchanged back and forth between the two during discharging and charging. These are called intercalation electrodes.

This type of battery is known as a “rocking chair battery” as the ions simply “rock” back and forth between the two electrodes.



Cathode materials

The most common compounds used for cathode materials are $LiCoO_2$, $LiNiO_2$ and $LiMn_2O_4$. Of these, $LiCoO_2$ has the best performance but is very high in cost, is toxic and has a limited lithium content range over which it is stable. $LiNiO_2$ is more stable, however the nickel ions can disorder. $LiMn_2O_4$ is generally the best value for money, and is also better for the environment.

Anode material

The anode material is carbon based, usually with composition $Li_{0.5}C_6$. This lithium content is lower than would be ideal, however higher capacity carbons pose safety issues.

Electrolyte

Since lithium reacts violently with water, and the cell voltage is so high that water would decompose, a non-aqueous electrolyte must be used. A typical electrolyte is $LiPF_6$ dissolved in an ethylene carbonate and dimethyl carbonate mixture.

After initial charging the following reactions take place upon discharge:

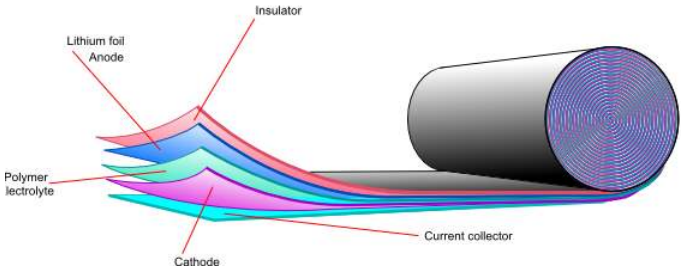
At the cathode: $xLi^+ + Mn_2O_4 + xe^- \rightarrow Li_xMn_2O_4$

At the anode: $Li_xC_6 \rightarrow xLi^+ + 6C + xe^-$

Overall: $Li_xC_6 + Mn_2O_4 \rightarrow Li_xMn_2O_4 + 6C$

Lithium polymer batteries

Another way of overcoming the high reactivity of lithium is to use a solid polymer electrolyte. Using lithium metal gives a higher energy density, higher cell potential and very low self discharge, so if the safety issues can be overcome, it would be the preferred anode material. Another problem to overcome is the high resistivity of the polymer electrolyte. One possible solution is to use the electrolyte as a very thin film to decrease the total resistance.



Cell capacity and specific energy density

It is important to specify the exact steps taken when calculating the theoretical cell capacity and the maximum specific energy density of a given lithium cell. For full lithium utilization, the cell capacity is 3860 mAh/g of lithium, simply calculated by Faraday's laws. Thus, the actual rated capacity of the cell in mAh is determined by the weight of lithium in the cell. The actual specific capacity, on the other hand, is usually calculated as the actual rated capacity divided by the weight of lithium in the cell (and quoted as mAh/g of Lithium) or, less frequently, as the ratio of the rated capacity and the weight of the cell (and quoted as mAh/g of the cell). In a general case, the cell weight can be calculated as follows:

Lithium cell capacity and specific energy density

$$\mathrm{W}_{\mathrm{cell}} = w_{\mathrm{Li}} f_{\mathrm{A}} + w_{\mathrm{Li}} f_{\mathrm{C}} + w_{\mathrm{aux}} \tag{6.11.1}$$

where

w_{Li} is the weight (wt.) of lithium in the cell;
f_A is the multiplier for the anode wt.;
f_C is the multiplier for the cathode wt.;
w_{aux} is the auxiliary wt. of the cell (which includes the wt. of the electrolyte, separator, connectors etc. and typically comprises of 20% of the wt. of the active part of the cell).

Thus, as a first approximation:

\[W_{\text {cell }} \approx 1.2 w_{\mathrm{Li}}\left(f_{\mathrm{A}}+f_{\mathrm{C}}\right)\]

And the specific capacity of the cell (as mAh/g of the cell wt.) is given by:

\[w_{\mathrm{Li}} 3860 / 1.2 w_{\mathrm{Li}}\left(f_{\mathrm{A}}+f_{\mathrm{C}}\right)=3860 / 1.2\left(f_{\mathrm{A}}+f_{\mathrm{C}}\right)\]

The specific energy density (which is energy density, E divided by the weight of the cell) is the product of the specific capacity and the operating voltage in one full discharge cycle. The values are often quoted for an average voltage and less frequently for

the maximum voltage in the discharge sequence. For a constant current drain I, the specific energy density can be written as:

$$\left[I w_{\mathrm{Li}} \int V \mathrm{d}t\right] / w_{\mathrm{cell}} \tag{6.11.2}$$

The integration is carried out over the full specified discharge time. The specified discharge time is related to a maximum and a minimum voltage threshold value to allow for recharging the battery. Outside the threshold values, the battery may attain an irreversible state. The specific energy density (Wh/kg of the cell) is related to the specific capacity as follows:

Specific energy density = specific capacity × average operating voltage

$$= w_{\mathrm{Li}} 3860 / 1.2 w_{\mathrm{Li}} \left(f_{\mathrm{A}} + f_{\mathrm{C}}\right) V_{\mathrm{ave}} \tag{6.11.3}$$

$$= 3860 / 1.2 \left(f_{\mathrm{A}} + f_{\mathrm{C}}\right) V_{\mathrm{ave}} \tag{6.11.4}$$

Li-ion battery

In order to maximize the specific energy density, it is desirable to minimize the weight of the cell, while maximizing the ratio of weight of lithium to the weight of the cell. For the Li-ion cell, for example, the theoretical stoichiometric value of the anodic multiplier (f_A) is 10.3, while for the cathode (f_C) is 25. Thus the maximum theoretical specific energy density for a max 4.2 V is calculated to be between 380 to 460 Wh/kg, depending upon whether the weight of the auxiliary components are taken into account.

The stoichiometric value for the carbon anode arises from the fact that lithium is intercalated into the carbon structural layers at the max possible molar ratio of 1 Li atom to 6C atoms giving rise to the limiting formula of LiC_6. In practice, Li availability in the anode is only 50% of the theoretical maximum corresponding to the formula $Li_{0.5-x}C_6$ where x can vary from 0 (fully charged state) to 0.5 (fully discharged state). For the cathode, Li is intercalated into a perovskite structure of $LiCoO_2$. Although it appears that, theoretically, the maximum ratio of Li to CoO_2 is one, in practice the formula corresponds to $Li_{0.5+x}CoO_2$, where x=0 corresponds to the fully charged state and x=0.5 corresponds to the fully charged state. Thus the available capacities for both the anode and the cathode are more than halved, with the multipliers typically being, f_A>21 and f_C>50. The excess value arises from the weight of the binder and other additives. The operating voltage during discharge decreases from a max value of 4.2 to a cut-off value of 2.8 V, giving an average value of 3.35 V over the discharge cycle. The practical specific energy density is therefore in the region of 160 Wh/kg for a Li-ion cell.

The only practical method for increasing the specific energy density of a Li-ion cell is to decrease the weight of the auxiliary components of the cell. It is widely believed that with a considerable amount of research and development the maximum specific energy density that can be achieved for a Li-ion cell within the next five years will reach 220 Wh/kg of the cell.

The cycle life of Li-ion batteries are between 500 to 1000 cycles.

This page titled 6.11: Lithium batteries is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.12: Battery characteristics

The following battery characteristics must be taken into consideration when selecting a battery:

- Type
- Voltage
- Discharge curve
- Capacity
- Energy density
- Specific energy density
- Power density
- Temperature dependence
- Service life
- Physical requirements
- Charge/discharge cycle
- Cycle life
- Cost
- Ability to deep discharge
- Application requirements

Type

See primary and secondary batteries page.

Voltage

The theoretical standard cell voltage can be determined from the electrochemical series using E^o values:

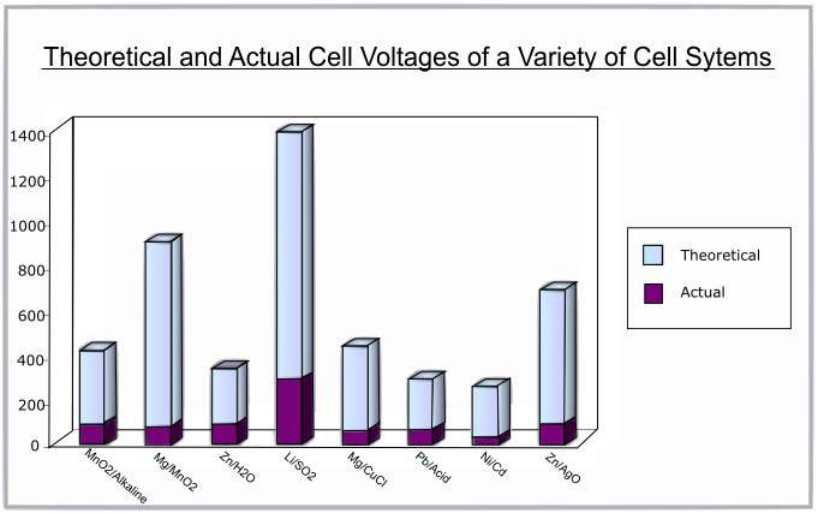
E^o (cathodic) – E^o (anodic) = E^o (cell)

This is the standard theoretical voltage. The theoretical cell voltage is modified by the Nernst equation, which takes into account the non-standard state of the reacting component. The Nernstian potential will change with time either because of use or self-discharge by which the activity (or concentration) of the electro-active component in the cell is modified. Thus the nominal voltage is determined by the cell chemistry at any given point of time.

The actual voltage produce will always be lower than the theoretical voltage due to polarisation and the resistance losses (IR drop) of the battery and is dependent upon the load current and the internal impedance of the cell. These factors are dependent upon electrode kinetics and thus vary with temperature, state of charge, and with the age of the cell. The actual voltage appearing at the terminal needs to be sufficient for the intended application.

Typical values of voltage range from 1.2 V for a Ni/Cd battery to 3.7 V for a Li/ion battery.

The following graph shows the difference between the theoretical and actual voltages for various battery systems:



Discharge Curve

The discharge curve is a plot of voltage against percentage of capacity discharged. A flat discharge curve is desirable as this means that the voltage remains constant as the battery is used up.

Capacity

The theoretical capacity of a battery is the quantity of electricity involved in the electro-chemical reaction. It is denoted Q and is given by:

$$Q = xnF \tag{6.12.1}$$

where x = number of moles of reaction, n = number of electrons transferred per mole of reaction and F = Faraday's constant

The capacity is usually given in terms of mass, not the number of moles:

$$Q = \frac{nF}{M_r} \tag{6.12.2}$$

where M_r = Molecular Mass. This gives the capacity in units of Ampere-hours per gram (Ah/g).

In practice, the full battery capacity could never be realised, as there is a significant weight contribution from non-reactive components such as binders & conducting particles, separators & electrolytes and current collectors & substrates as well as packaging. Typical values range from 0.26 Ah/g for Pb to 26.59 Ah/g for H_2.

Energy density

The energy density is the energy that can be derived peer unit volume of the weight of the cell.

Specific energy density

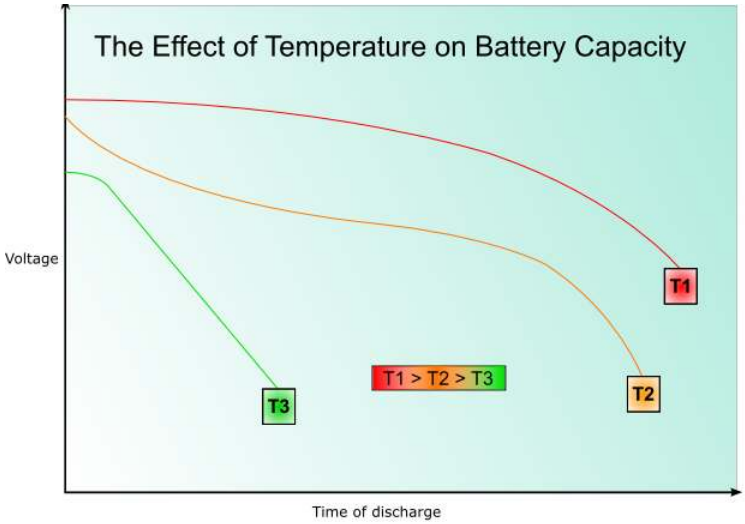
The specific energy density is the energy that can be derived per unit weight of the cell (or sometimes per unit weight of the active electrode material). It is the product of the specific capacity and the operating voltage in one full discharge cycle. Both the current and the voltage may vary within a discharge cycle and thus the specific energy derived is calculated by integrating the product of current and voltage over time. The discharge time is related to the maximum and minimum voltage threshold and is dependent upon the state of availability of the active materials and/or the avoidance of an irreversible state for a rechargeable battery.

Power density

The power density is the power that can be derived per unit weight of the cell (W/kg).

Temperature dependence

The rate of the reaction in the cell will be temperature dependant according to theories of kinetics. The internal resistance also varies with temperature; low temperatures give higher internal resistance. At very low temperatures the electrolyte may freeze giving a lower voltage as ion movement is impeded. At very high temperatures the chemicals may decompose, or there may be enough energy available to activate unwanted, reversible reactions, reducing the capacity.
The rate of decrease of voltage with increasing discharge will also be higher at lower temperatures, as will the capacity- this is illustrated by the following graph:



Service life

The battery cycle life for a rechargeable battery is defined as the number of charge/recharge cycles a secondary battery can perform before its capacity falls to 80% of what it originally was. This is typically between 500 and 1200 cycles.

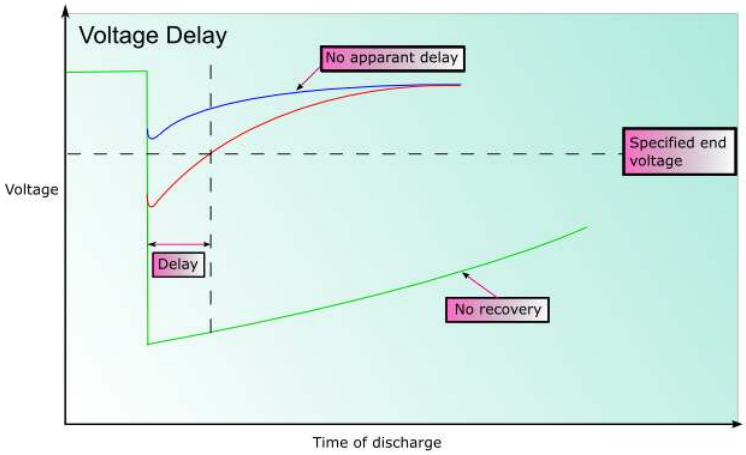
The battery shelf life is the time a battery can be stored inactive before its capacity falls to 80%. The reduction in capacity with time is caused by the depletion of the active materials by undesired reactions within the cell.

Batteries can also be subjected to premature death by:

- Over-charging
- Over-discharging
- Short circuiting
- Drawing more current than it was designed to produce
- Subjecting to extreme temperatures
- Subjecting to physical shock or vibrations

Voltage Delay

In some battery systems *passivation* may occur. Passivation is the process by which the reduced product that forms (often an oxide) does not dissolve into the electrolyte, or fall away from the electrode, but forms a film on the surface of the electrode instead. This can significantly impede the reaction as there is reduced electrical contact within the cell. This can substantially prolong the shelf life of the battery, however when the battery discharges the initial voltage may be lower than expected until the coating is broken. This is known as *voltage delay*.

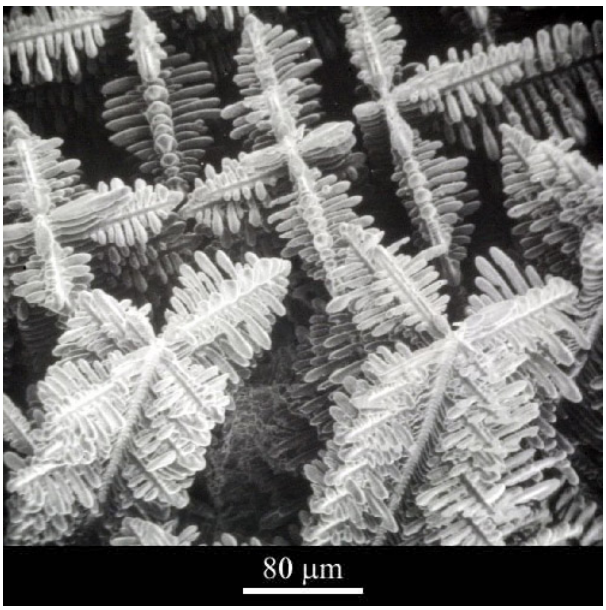


Battery Death due to Aging

Over the life time of the cell the morphology of the components will change, with effects that are detrimental to the functioning of the cell.

- The crystals in the cell grow larger, increasing the impedance,
- Metallic dendrites grow on the cells causing the electrodes to swell, forcing them closer together and increasing the self-discharge,
- The dendrites may penetrate the separator giving very high self discharge, or even a short circuit.

Eventually, the internal impedance and the self discharge will be so high that the battery is no longer usable.



Metallic dendrites (Click micrograph for details)

Physical requirements

This includes the geometry of the cell, its size, weight and shape and the location of the terminals.

Charge/Discharge cycle

There are many aspects of the cycle that need consideration, such as:

- Voltage necessary to charge
- Time necessary to charge

- Availability of charging source
- Potential safety hazards during charge/discharge

Cycle life

The cycle life of a rechargeable battery is the number of discharge/charge cycles it can undergo before its capacity falls to 80%.

Cost

This includes the initial cost of the battery itself as well as the cost of charging and maintaining the battery.

Ability to deep discharge

There is a logarithmic relationship between the depth of discharge and the life of a battery, thus the life of a battery can be significantly increased if it is not fully discharged; for example, a mobile phone battery will last 5-6 times longer if it is only discharged 80% before recharging.

Special deep discharge batteries are available for applications where this might be necessary.

Nickel-cadmium batteries

The exception to this is the nickel-cadmium battery, as subjecting them to partial discharge introduces “memory effects;” the battery appears to “remember” how much charge is being used, and will only recharge that amount rather than the full amount of charge.

In fact the repeated shallow charging causes the crystalline structure in the battery to change: the crystals in the cell grow larger, increasing the impedance and thus reducing its capacity.

Application requirements

The battery must be sufficient for the intended application. This means that it must be able to produce the right current with the right voltage. It must have sufficient capacity, energy and power. It should also not exceed the requirements of the application by too much, since this is likely to result in unnecessary cost; it must give sufficient performance for the lowest possible price.

This page titled 6.12: Battery characteristics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

6.13: The future

The future of battery technology now lies in the concept of fuel cells. A fuel cell, like a battery is an electrochemical cell. It has an anode and a cathode, and an oxidation reaction and a reduction reaction occur, however the electrodes act as catalysts, and are not used up in these reactions. Instead, the reactions take place within a “fuel,” which provides the source for the electrons. This is a very efficient method of producing electricity (up to 85% efficient – that’s over three times as efficient as an internal combustion engine.) They are also very environmentally friendly – the most environmentally friendly cells produce only water and heat. At the moment the only disadvantage is that they are very costly, however it is very likely that in the future fuel cells will be the power source of choice for many applications.

For more information on Fuel cells, see the Fuel Cells TLP.

Going further

Books

- *High Density Energy Lithium Batteries;* KE Aifantis, SA Hackney and RV Kumar (Editors), Wiley-VCH Verlag, 2010

This page titled 6.13: The future is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

7: Bending and Torsion of Beams

www.doitpoms.ac.uk/tlplib/be...ding/index.php

Learning Objectives

On completion of this TLP package, you should:

- Understand the stress distribution within beams subject to bending or torsion.
- Be familiar with the concepts of the radius of curvature of a section of a beam (and its reciprocal, the curvature), second moment of area, polar moment of inertia, beam stiffness and torsional stiffness.
- Be able to calculate the moments acting in a beam subject to bending or torsion.
- Be able to calculate the deflections of a beam on bending and the angle of twist of a bar under torsion.
- Be able to predict the effect of plastic deformation, at least with simple beam geometry.

7.1: Introduction
7.2: Pole-vaulting
7.3: Bending moments and beam curvatures
7.4: Maximising the beam stiffness
7.5: Beam deflections from applied bending moments
7.6: Twisting moments (torques) and torsional stiffness
7.7: Springs
7.8: Plastic deformation during beam bending
7.9: Summary
7.10: Questions

This page titled 7: Bending and Torsion of Beams is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.1: Introduction

Before you start

There are no specific prerequisites for this TLP, but it would be useful to be familiar with stress and strain , elastic strain and Plastic deformation , Young modulus , E and yield stress , σ_Y. While a basic knowledge of mechanical deformation is assumed, this teaching and learning package covers all the fundamentals of beam mechanics.

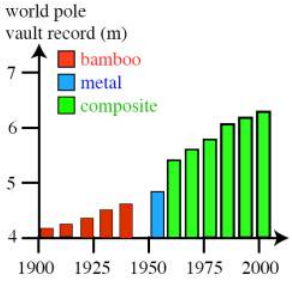
Introduction

Beam stiffness is an important concept for many types of structure, particularly those with slender shapes. Inadequate beam stiffness can lead to large deflections, and may also cause high localized stresses and a danger of failure in that region. In addition to bending moments, such structures may be subjected to twisting, or torsional moments (torques). In fact, virtually all structures, including buildings and many natural structures (trees, bones etc) are commonly subjected to significant applied moments. It is important to recognize the roles of structural shape, applied loads and material properties when predicting the resultant moments, deflections and stress distributions. The aim of this TLP is to provide the necessary information to allow such bending and torsional moments, deflections (both elastic and elastoplastic) and stress distributions to be predicted and understood.

This page titled 7.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.2: Pole-vaulting

Pole vaulting as an athletic activity dates back to the ancient Greeks. Modern competition started around the turn of the 20 th century, when the Olympic Games were restarted. A sharp increase in the achievable height coincided with the advent of composite (fibreglass) poles, about 50 years ago. These are sufficiently strong and flexible to allow substantial amounts of energy (kinetic energy of the athlete) to be transformed into elastic strain energy stored in the deformed pole, and subsequently transformed again into potential energy (height of the athlete) as the pole recovers elastically. The mechanics of beam bending is clearly integral to this operation.





The sharp increase in achievable height that coincided with the switch to composite poles was due to a change in the mechanics of pole vaulting. Bamboo or metal poles with sufficient flexibility to allow significant energy storage would, respectively, be likely to fracture or plastically deform.

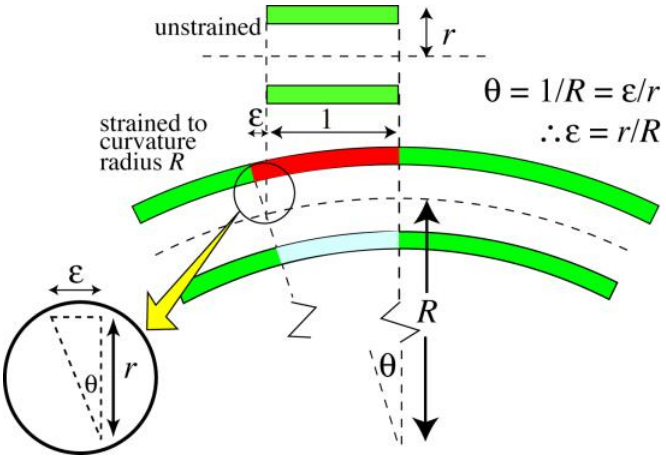
DoITPoMS standard terms of use

Visual inspection of a bent pole (see photo) is all that's needed to estimate the distribution of axial strains (and hence stresses) within its cross-section. The pole has a diameter of about 50 mm and it can be seen in the photo that it is being bent to a (uniform) radius of curvature, R , of the order of 1 m (~ length of the athlete's legs!). Considering a section of unit length (unstrained) in the diagram below, the angle θ (~tan θ) ≈ $1/R$ after bending (where R is the radius of curvature). From the two similar triangles in the diagram, θ is also given by the surface strain ε divided by r , the radius of the pole . The surface strain, ε, is thus given by the ratio r / R , which has a value here of about 2.5 %. This strain is compressive on the "inside" surface of the pole (coloured blue) and tensile on the "outside" surface (coloured red).

The stresses induced by such bending can be high. The axial stress is given by the product of Young's modulus, E, and strain, ε.

$$\sigma = E\varepsilon = Er/R \tag{7.2.1}$$

For example, assuming the composite to have an axial stiffness of ~ 40 GPa, the axial stresses at the inside and outside surfaces of the pole must be about 2.5 % of this, ie ~±1 GPa. Composites are able to sustain such high stresses, although it's not unknown for vaulting poles to fracture.

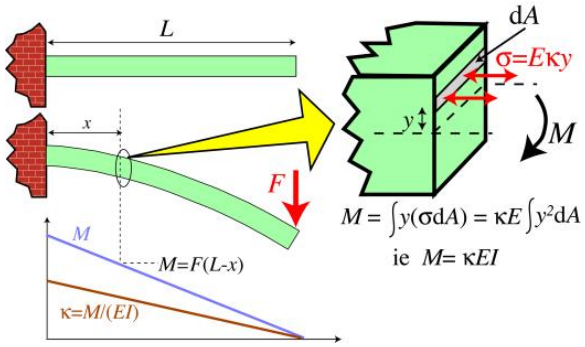


Strains induced during bending of a pole by the application of a bending moment M.

This page titled 7.2: Pole-vaulting is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.3: Bending moments and beam curvatures

Bending moments are produced by transverse loads applied to beams. The simplest case is the cantilever ***beam*** , widely encountered in balconies, aircraft wings, diving boards etc. The bending moment acting on a section of the beam, due to an applied transverse force, is given by the product of the applied force and its distance from that section. It thus has units of N m. It is balanced by the ***internal moment*** arising from the stresses generated. This is given by a summation of all of the internal moments acting on individual elements within the section. These are given by the force acting on the element (stress times area of element) multiplied by its distance from the neutral axis, *y* .



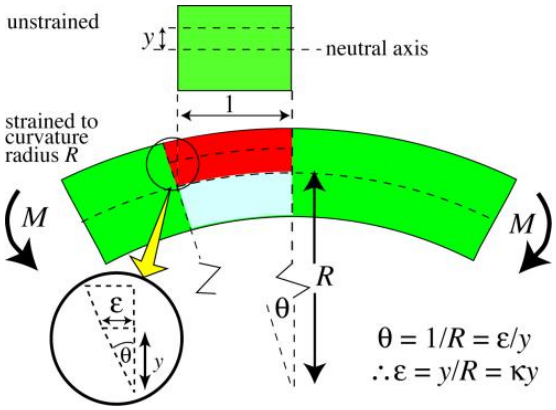
Balancing the external and internal moments during the bending of a cantilever beam

Therefore, the bending moment, *M* , in a loaded beam can be written in the form

$$M = \int y(\sigma dA) \tag{7.3.1}$$

The concept of the ***curvature*** of a beam, κ, is central to the understanding of beam bending. The figure below, which refers now to a solid beam, rather than the hollow pole shown in the previous section, shows that the axial strain, *ε*, is given by the ratio *y* / *R* . Equivalently, *1/R* (the "curvature", κ) is equal to the through-thickness gradient of axial strain. It follows that the axial stress at a distance *y* from the Neutral axis of the beam is given by

$\sigma = E\,\kappa\,y$



Relation between the radius of curvature, R, beam curvature, κ , and the strains within a beam subjected to a bending moment.

The bending moment can thus be expressed as

$$M = \int y(E\kappa y dA) = \kappa E \int y^2 dA \tag{7.3.2}$$

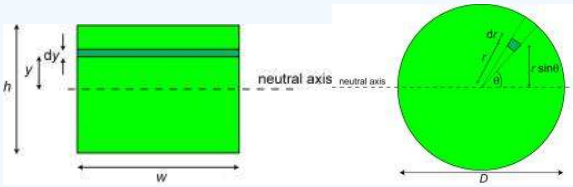
This can be presented more compactly by defining *I* (the ***second moment of area*** , or "***moment of inertia***") as

$$I = \int_0^{y_{max}} y^2 dA \tag{7.3.3}$$

The units of *I* are m^4 . The value of *I* is dependent solely on the beam sectional shape.

Example 7.3.1

Calculate I for a rectangular and a circular beam



Solution

The second moment of area for a rectangular section beam of width *w* and thickness *h* is given by

$$I = \int_0^{h/2} y^2 dA = 2\int_0^{h/2} y^2 w \mathrm{d}y = 2w\left[\frac{y^3}{3}\right]_0^{h/2} = \frac{wh^3}{12} \tag{7.3.4}$$

The corresponding operation for a circular cross-section of diameter *D* gives

$$I = \int_A y^2 \mathrm{d}A = \int_{\theta=0 r-0}^{2\pi D/2} (r\sin\theta)^2[(r\mathrm{d}\theta)\mathrm{d}r] = \int_{\theta=0}^{2\pi}\left\{\int_{r=0}^{D/2} r^3 \mathrm{d}r\right\}\sin^2\theta \mathrm{d}\theta \tag{7.3.5}$$

$$= \frac{D^4}{64}\int_{\theta=0}^{2\pi}\sin^2\theta \mathrm{d}\theta = \frac{D^4}{64}\int_{\theta=0}^{2\pi}\left(\frac{1-\cos 2\theta}{2}\right)\mathrm{d}\theta = \frac{D^4}{64}\left[\frac{\theta}{2} - \frac{\sin 2\theta}{4}\right]_0^{2\pi} = \frac{\pi D^4}{64} \tag{7.3.6}$$

The moment can now be written as

$$M = \kappa EI \tag{7.3.7}$$

These equations allow the curvature distribution along the length of a beam (ie its shape), and the stress distribution within it, to be calculated for any given set of applied forces. The following simulation implements these equations for a user-controlled beam shape and set of forces. The 3-point bending and 4-point bending loading configurations in this simulation are SYMMETRICAL, with the upward forces, denoted by arrows, outside of the downward force(s), denoted by hooks

Drag the loads horizontally with the cursor. The beam length is fixed at 100 cm.

Materials

- Nylon (E=5 GPa)
- Wood (E=15 GPa)
- Aluminium (E=70 GPa)

Loading geometry

- Cantilever
- 3 point bending
- 4 point bending

80

Beam section

w 2.5 cm
h 2.5 cm
I 3.3e-8 m^4
F 80 N

Apply load

0 cm

Reset

Press to add another weight (max 3)

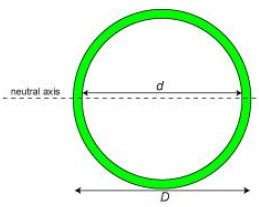
Add weight

50 cm

1.25
0
-1.25

CC BY NC SA

www.doitpoms.ac.uk



A fruitful approach to designing beams which are both light and stiff is to make them hollow. Calculation of the second moment of area for hollow beams is very straightforward, since it is obtained by simply subtracting the *I* of the missing section from that of the overall section. For

example, that for a cylindrical tube is given by

$$I = I_{\text{complete section}} - I_{\text{missing section}} = \frac{\pi D^4}{64} - \frac{\pi d^4}{64} \tag{7.3.8}$$

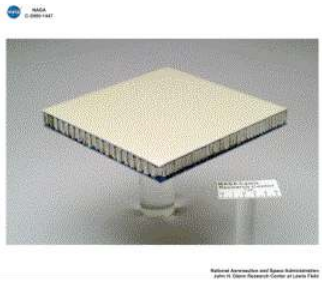
This page titled 7.3: Bending moments and beam curvatures is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.4: Maximising the beam stiffness

The product *EI* is termed the "**beam stiffness**", or sometimes the "flexural rigidity". It is often given the symbol Σ. It is a measure of how strongly the beam resists deflection under bending moments. It is analogous to the Young's modulus in uniaxial loading (with the curvature being analogous to the uniaxial strain and the bending moment being analogous to the uniaxial stress). For a given material, the beam stiffness is maximized by maximizing the value of I . This is done by using sectional shapes for which most of the sectional area is remote from the neutral axis.

For example, a beam of square cross-section is stiffer than a circular beam with the same area, since a circle has a larger proportion of the section near the neutral axis. A hollow square section is even stiffer. Taking this rationale still further leads to I-section beams and sandwich panels.





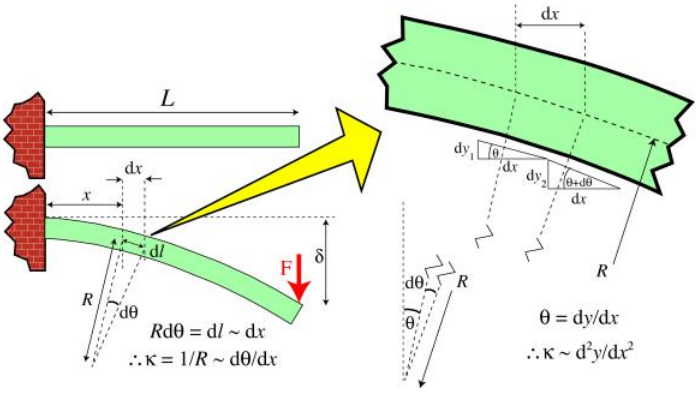
I-beams are commonly used in construction of buildings. Sandwich panels are also in extensive use, for example in surf-boards, aircraft, skis etc

This page titled 7.4: Maximising the beam stiffness is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.5: Beam deflections from applied bending moments

he second derivative (curvature) of the neutral axis line (dotted line in diagram)

$$\kappa = \frac{d^2y}{dx^2} \tag{7.5.1}$$



The approximation involved in equating beam curvature to the curvature of the neutral axis of a beam.

It follows that

$$M = \kappa EI = EI\frac{\mathrm{d}^2y}{\mathrm{d}x^2} \tag{7.5.2}$$

Since the moment at the section concerned can also be written, for a cantilever beam, as M = F (L - x)

it follows that

$$EI\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = F(L-x) \tag{7.5.3}$$

This second order differential equation can be integrated (twice), with appropriate boundary conditions, to find the deflection of the beam at different points along its length. For a cantilever beam, this operation is shown below.

$$EI\frac{\mathrm{d}y}{\mathrm{d}x} = FLx - \frac{Fx^2}{2} + C_1 \tag{7.5.4}$$

$$\text{at } x = 0, \tfrac{\mathrm{d}y}{\mathrm{d}x} = 0, \text{ thus } C_1 = 0$$

$$EIy = \frac{FLx^2}{2} - \frac{Fx^3}{6} + C_2 \tag{7.5.5}$$

$$\text{at } x = 0, y = 0, \text{ thus } C_2 = 0$$

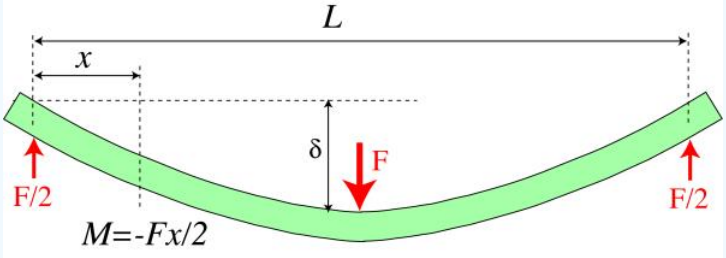
which can be rearranged to give

$$y = \frac{Fx^2}{6EI}(3L - x) \tag{7.5.6}$$

For example, at the loaded end ($x = L$), this gives

$$\delta = \frac{FL^3}{3EI} \tag{7.5.7}$$

Symmetrical 3-point bending



Symmetrical 3-point bend loading

The bending moment is given by

$$M = \frac{-Fx}{2} \tag{7.5.8}$$

It follows that

$$EI\frac{\mathrm{d}^2y}{\mathrm{d}x^2} = \frac{-Fx}{2} \tag{7.5.9}$$

and the integration procedures lead to

$$EI\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{Fx^2}{4} + C_1 \tag{7.5.10}$$

$$\text{at } x = \tfrac{L}{2}, \quad \tfrac{dy}{dx} = 0 \text{ thus } C_1 = \tfrac{FL^2}{16}$$

$$EIy = \frac{-Fx^3}{12} - \frac{FL^2x}{16} + C_2 \tag{7.5.11}$$

$$\text{at } x = 0, y = 0 \text{ thus } C_2 = 0$$

so the equation for the deflection is

$$y = \frac{Fx}{48EI}\left(3L^2 - 4x^2\right) \tag{7.5.12}$$

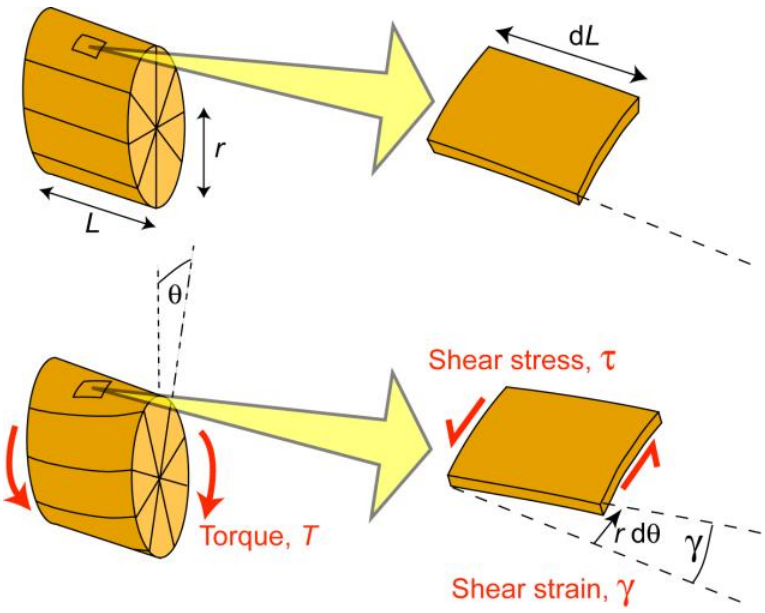
and deflection of the centre of the beam is given by

$$\delta = \frac{FL^3}{48EI} \tag{7.5.13}$$

This page titled 7.5: Beam deflections from applied bending moments is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.6: Twisting moments (torques) and torsional stiffness

Torsion is the twisting of a beam under the action of a torque (twisting moment). It is systematically applied to screws, nuts, axles, drive shafts etc, and is also generated more randomly under service conditions in car bodies, boat hulls, aircraft fuselages, bridges, springs and many other structures and components. A torque, T , has the same units (N m) as a bending moment, M . Both are the product of a force and a distance. In the case of a torque, the force is tangential and the distance is the radial distance between this tangent and the axis of rotation.



Torsion of a Cylindrical Bar

Torsion of a cylindrical bar is illustrated in the figure. It can be seen that the shear strain in an element of the bar is given by

$$\gamma = \frac{r\mathrm{d}\theta}{\mathrm{d}L} \tag{7.6.1}$$

This equation applies both at the surface of the bar, as shown, and also for any other radial location, using the appropriate value of r . Clearly, the shear strain varies linearly with r , from zero at the centre of the bar to a peak value at the free surface.

The shear stress, τ, at any radial location, is related to the shear strain by

$$\tau = G\gamma \tag{7.6.2}$$

The torque, T , can therefore be written as

$$T = \int_A \mathrm{d}T = \int_A \tau r \mathrm{d}A = \int_A Gr^2 \frac{\mathrm{d}\theta}{\mathrm{d}L}\mathrm{d}A \tag{7.6.3}$$

As for the beam bending case, the geometrical integral is represented as a (polar) second moment of area

$$I_\mathrm{P} = \int_A r^2 \mathrm{d}A \tag{7.6.4}$$

For a solid cylinder of diameter w , this can be written as

$$I_\mathrm{P} = \int_A r^2 \mathrm{d}A = \int_0^{d/2} r^2 2\pi r \mathrm{d}r = \pi \left[\frac{r^4}{2}\right]_0^{w/2} = \frac{\pi w^4}{32} \tag{7.6.5}$$

The torque is thus given by

$$T = GI_\mathrm{p}\frac{\mathrm{d}\theta}{\mathrm{d}L} \tag{7.6.6}$$

Comparing this equation with the corresponding one for beam bending

$M = E\,I\,K$

it can be seen that the torsional analogue for the *curvature* of a bent beam is the ***rate of twist*** along the length of the bar. This can be measured experimentally, although not quite so easily as a curvature (because the macroscopic shape of the bar does not actually change - at least when it is straight - see next page for an important example of a case when it is NOT straight).

This page titled 7.6: Twisting moments (torques) and torsional stiffness is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.7: Springs



A collection of assorted springs

An interesting example of torsion is provided by the deformation that takes place during the loading of ***springs*** (torsional coils). Of course, these have a wide range of engineering applications. They are normally made of (high yield stress) metals. (Ceramics are too brittle, while polymers are insufficiently stiff: fibre composites are also unsuitable - see below.) When a spring is loaded (compressed or extended), the deformation experienced by the wire is one of pure torsion. This is illustrated in the diagram below.

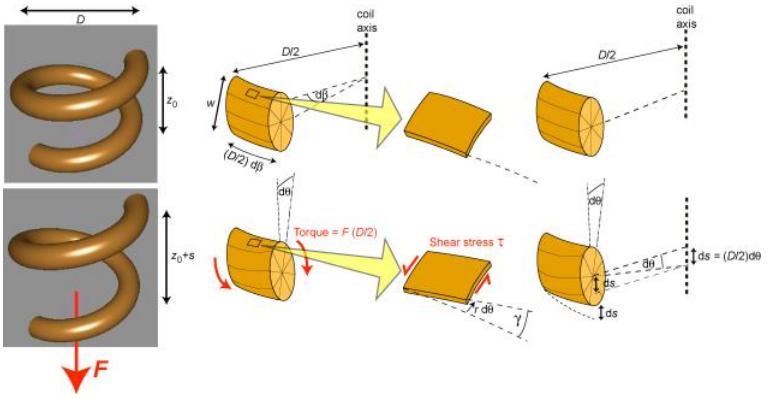


Illustration of how the application of an axial load, F, to a spring generates torsional deformation of the wire and hence axial extension of the spring.

The torque acting on the wire is given by

$$T = F\left(\frac{D}{2}\right) \tag{7.7.1}$$

in which F is the axial force and D is the coil diameter. It can be shown in the example below that the shear stress within the wire (at a distance r from the core) is given by

$$\tau = \frac{Tr}{2I} \tag{7.7.2}$$

in which I is the bending second moment of area (NOT the polar moment), and the shear strain in the wire is related to the change in axial extension of one turn of the coil, s , by the expression

$$\gamma = \frac{2sr}{\pi D^2} \tag{7.7.3}$$

Measurement of the extension (per turn) of a spring, as a function of the applied force (first carried out systematically by ***Robert Hooke*** , in his pioneering work on the nature of elasticity) is a very convenient method of obtaining elastic constants. The ratio of τ to γ , obtained from the above equations, gives the shear modulus, G . The loading geometry is such that a large axial extension (per turn) is generated, while the strains within the material remain low, particularly for springs with a large ratio of D to w . Of

course, this is exactly why springs are of practical use - they accommodate large deflections or displacements without the material being strained beyond its elastic limit (which is small for all materials except rubbers).

It's interesting to note why springs are not normally made of fibre composites. The natural orientation for the fibres would be along the length of the rod (wire) to be formed into a coil. However, these fibres would have very little effect on the shear modulus in a transverse section of the rod, which is the property that controls the elastic extensions of the spring. It might as well have been made solely of the polymeric matrix (although such springs have a very low stiffness). It's only by winding fibres into the hoop direction of the rod that the shear stiffness of transverse sections would be boosted. However, this is impractical, at least for anything but very large scale springs, since it would require the fibres to adopt higher curvatures than would normally be possible.

Example 7.7.1

Consider the small elemental length of the wire shown in the figure, subtending an angle $d\beta$ at the axis of the coil. The torsional shear stress within the wire, τ, can be found by noting that it varies linearly with distance from the centre of the wire ($\tau = \kappa r$, where κ is an unknown constant and r is the distance of the element from the centre of the wire) . The torque can now be expressed in terms of internal forces in the wire. The force in an individual element of the wire is given by the torsional shear stress, τ, multiplied by the area of the element, $2\pi r\,\mathrm{d}r$. Therefore, the elemental torque is simply this force multiplied by the distance of the element from the neutral axis of the wire, r . Summation of these elemental torques gives the total torque:

$$T = \int_0^{w/2} \tau(2\pi r \mathrm{d}r) r = 2\pi K \int_0^{w/2} r^3 \mathrm{d}r = \frac{\pi}{32} K w^4 \tag{7.7.4}$$

$$\tau = Kr = \frac{32Tr}{\pi w^4} = \frac{Tr}{2I} \tag{7.7.5}$$

where I is the second moment of area for a solid cylinder (the wire), given by:

$$I = \frac{\pi w^4}{64} \tag{7.7.6}$$

The local shear strain in the material is given by

$$\gamma = r\frac{\mathrm{d}\theta}{\mathrm{d}L} = \frac{r\mathrm{d}\theta}{(D/2)\mathrm{d}\beta} \tag{7.7.7}$$

in which $\mathrm{d}\theta/\mathrm{d}L$ is the rate of twist along the length of the wire. The incremental angle of twist of the wire is therefore related to the corresponding incremental rotation angle of the coil by

$$\mathrm{d}\theta = \gamma\frac{(D/2)}{r}\mathrm{d}\beta \tag{7.7.8}$$

The axial displacement of the coil, due to the twisting of the elemental section, can therefore be written

$$\mathrm{d}s = \left(\frac{D}{2}\right)\mathrm{d}\theta = \left(\frac{D}{2}\right)\gamma\left(\frac{D}{2r}\right)\mathrm{d}\beta \tag{7.7.9}$$

The axial displacement associated with the twist in one complete turn of the coil is thus

s=\left(\frac{D^{2}}{4 r}\right) \gamma \int_{0}^{2 \pi} \mathrm{d} \beta=\frac{\pi D^{2}}{2 r} \gamma

The local shear strain in the wire is therefore related to the increase in spacing between adjacent turns in the coil by

$$\gamma = \frac{2sr}{\pi D^2} \tag{7.7.10}$$

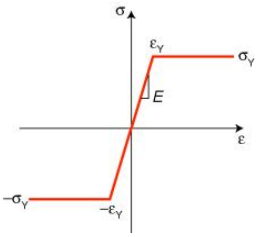
This page titled 7.7: Springs is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.8: Plastic deformation during beam bending



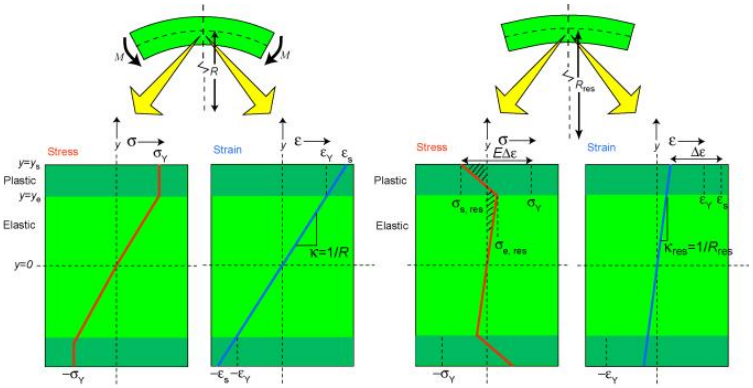
DoITPoMS standard terms of use

If the stresses within a beam exceed the elastic limit, then plastic deformation will occur. This can dramatically change the behaviour. Consider a material exhibiting elastic - perfectly plastic behaviour (ie no work-hardening), as shown below.



Stress-strain curve for an elastic-perfectly plastic material.

Stress and strain distributions before and after applying the moment are shown below. In the outer regions of the beam, the stress will be capped at σ_Y, although the strain will continue to increase linearly with distance from the neutral axis, as in the elastic case. The curvature (strain gradient), κ, induced by a given moment, M, will now be greater, since this increase will be required in order to bring the internal moment back up to the level of the applied moment - i.e. bending will increase.



Distributions of stress and strain within a beam before and after application of a moment sufficiently large to cause plastic deformation

A further difference is observed on removal of the applied moment, since the beam will now retain a ***residual curvature***, κ_{res} , as a result of the plastic deformation. This is due to the presence of ***residual stresses*** . The residual curvature can be calculated, using the fact that the beam is subject to no applied force. It follows that the residual stress distribution must satisfy a ***force balance*** , so that

$$\int_{y=0}^{y_s} \sigma(y)\mathrm{d}y = 0 \tag{7.8.1}$$

which is equivalent to the shaded areas in the diagram being equal. Since the change in stress (at any value of y) on removing the applied moment is given by the change in strain at that depth times the modulus (eg = $E\,\Delta\varepsilon$ at $y = y_s$ - see diagram), these equations allow the residual stress distribution to be established. The following expressions can be obtained (***click here***) for the thickness of the elastic core, the residual curvature, the surface residual stress and the residual stress at the limit of the elastic core.

$$y_e = \frac{\sigma_Y}{E\kappa} \tag{7.8.2}$$

$$\kappa_{\mathrm{res}} = \kappa \left(1 - \frac{y_{\mathrm{e}}}{y_{\mathrm{s}}}\right)^2 \tag{7.8.3}$$

$$\sigma_{\mathrm{s,res}} = \sigma_{\mathrm{Y}} - E y_{\mathrm{s}} \left(\kappa - \kappa_{\mathrm{res}}\right) \tag{7.8.4}$$

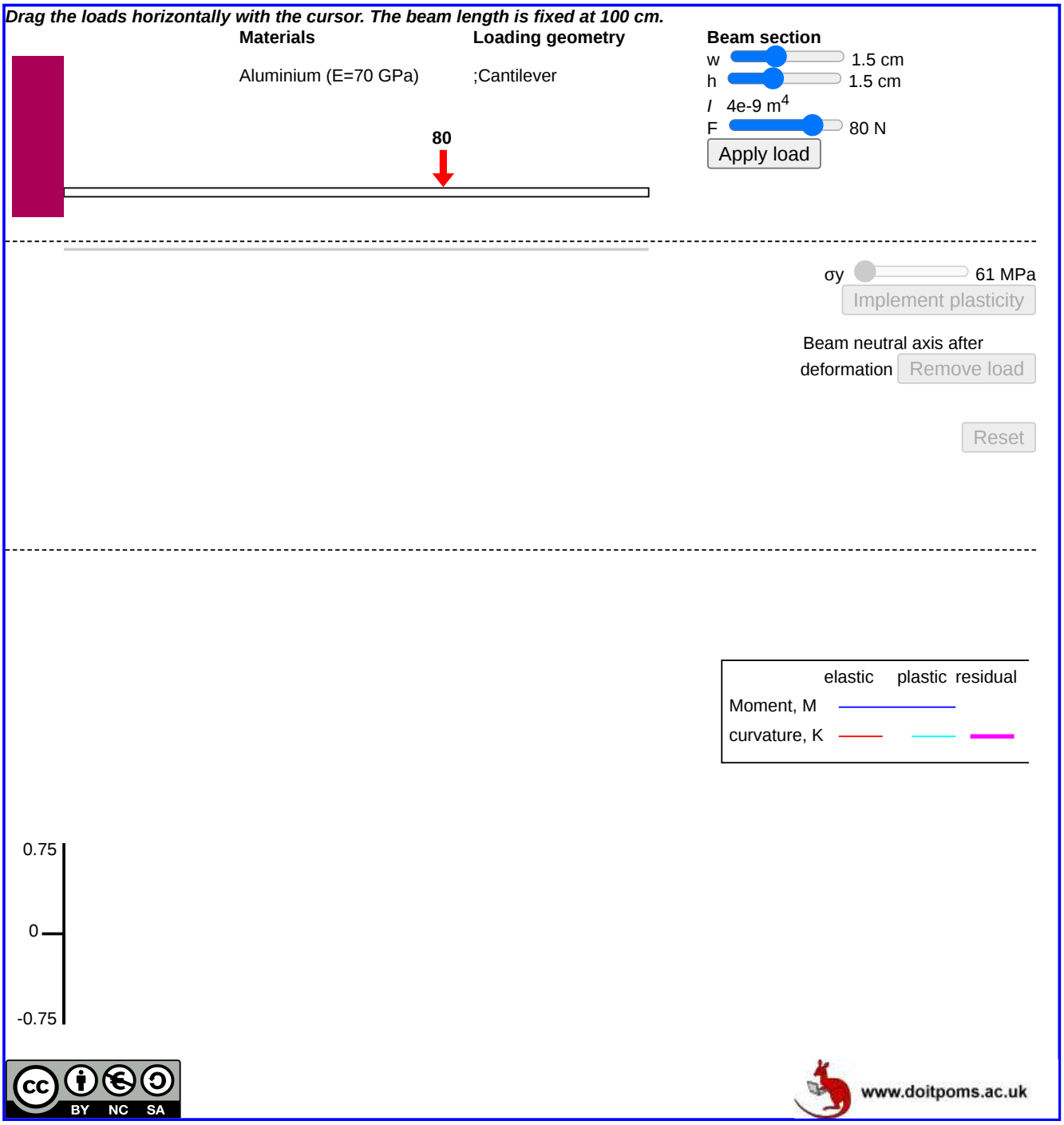
$$\sigma_{\mathrm{e,res}} = \sigma_{\mathrm{Y}} - E y_{\mathrm{e}} \left(\kappa - \kappa_{\mathrm{res}}\right) \tag{7.8.5}$$

Of course, the picture may in practice be complicated by work hardening, more complex sectional geometries, non-prismatic beams etc, but the same principles still apply. Incidentally, it may be noted that, in addition to the force balance, the residual stress distribution in an unloaded beam must also satisfy a ***moment balance*** , so that

$$\int_{y=0}^{y_s} \sigma(y) y \mathrm{d}y = 0 \tag{7.8.6}$$

However, the symmetry of the tensile and compressive sides of the beam ensures that this condition is satisfied, so it is not involved in the solution procedure in this case. In other cases, however, in which the neutral axis is not a plane of symmetry, this condition may also need to be invoked in order to find the solution.

The plastic deformation behaviour of a prismatic beam, with a symmetrical, rectangular section, made of a metal exhibiting no work hardening, can be explored using the plastic version of the beam bending simulation presented in an earlier section.



This page titled 7.8: Plastic deformation during beam bending is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.9: Summary

You should now understand the basic principals of bending and torsion. You should be able to predict how a beam will respond elastically to a bending moment, from a knowledge of the Young's modulus, E , and the sectional geometry of the beam (from which the second moment of area, I , is derived). You should understand the relationship between the (local) bending moment, M , the beam stiffness (flexural rigidity), EI , and the resultant (local) curvature, κ.

The concept of torsion has been introduced, with the analogue of the bending moment being a torque, T , and the analogue of the curvature being the rate of twist of the beam, θ / L . The elastic constant controlling the behaviour is the shear modulus, G , and the sectional geometry analogue of the second moment of area, I , is the polar second moment of area, I_P.

You should also have an appreciation of the nature of the stress distribution within an elastically deformed beam and you should understand that, for a metallic beam, it's possible that these stresses could exceed the yield stress, σ_Y , so that plastic deformation could take place. In this case, there is a change in the relationship between the applied moment and the resultant curvature (so that a given increase in moment gives a larger increase in curvature). Furthermore, on removing the applied moment, the beam retains a residual curvature. Analogous phenomena can occur during torsion. These effects can be quantitiatively predicted.

Some implications of these analyses for the design of components and structures subject to bending moments and torques have been briefly outlined.

Going further

Books

- S.P. Timoshenko, J.N. Goodier, *Theory of Elasticity*, McGraw-Hill International Editions, Third Edition, 1970.
- J.M. Gere, *Mechanics of Materials* , Nelson Thornes, Fifth SI Edition, 2001.

Websites

- Simulation showing the stress distribution in 3-point bending: teaching.ust.hk/~mech101/3%20point%20bending.htm
- Video of the 3-point bending of a brass beam: https://www.doitpoms.ac.uk/vidlib/full_record.php?id=14
- Stress distribution In a polycarbonate bar undergoing 3-point bending: https://www.doitpoms.ac.uk/vidlib/full_record.php?id=36
- 3-point bending simulation for a sandwich beam: mms.ensmp.fr/Mms_scripts/Beam/index.php
- Simulation showing the stress distribution in 4-point bending: teaching.ust.hk/~mech101/4_point_bending.htm
- Stress distribution In a polycarbonate bar undergoing 4-point bending: https://www.doitpoms.ac.uk/vidlib/full_record.php?id=40
- Simulation of shear, moment and deflection on application of different forces to a beam: http://www.engr.umd.edu/HAMLET/BeamAnalysis/
- Sandwich panels: http://oneoceankayaks.com/Sandcore.htm
- The torsional pendulum: http://farside.ph.utexas.edu/teaching/301/lectures/node139.html
- Granta Design case studies by a former Cambridge student

This page titled 7.9: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

7.10: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

1. How do the axial stresses within a vaulting pole vary with distance from the neutral axis?

a	They are zero at the neutral axis, rising to a tensile maximum at the outer surface, and a compressive maximum at the inner surface.
b	They are zero at the inner surface and rise linearly to a maximum at the outer surface.
c	They reach a maximum at the neutral axis, falling to zero at the outer and inner surfaces.
d	They are constant throughout the section.

Answer

A

It is important to maximise the beam stiffness when attempting to minimise the deflection of a beam (of given mass). Which of the following shapes, all with dimensions such that they have the same cross-sectional area, will have the highest beam stiffness?

a	An I-beam
b	A square section beam
c	A hollow cylindrical beam
d	A circular section beam

Answer

A. In general, an I-beam gives a high value of I, assuming that the central section and the flange both have a fairly high aspect ratio.

In which of the following situations is torsion occurring?

a	A tree bending in the wind.
b	A wire hanging under its own weight.
c	A screwdriver being used to tighten a screw.
d	The wings on an aircraft acting as cantilever beams during flight.

Answer

C.

How can the stress distribution in an elastoplastic beam undergoing bending be predicted?

a	The axial strain will vary linearly from the neutral axis to the free surfaces, and so the stress distribution should increase linearly in the same fashion.
b	The axial strain will vary linearly from the neutral axis to the free surfaces, and the stress distribution can be found from this information and the stress-strain curve of the material.
c	The stress distribution cannot be calculated theoretically and must be found by experiment.
d	The beam will be fully plastic and so the stress will be of constant magnitude throughout the section. This stress can be predicted from the Young's modulus and the yield strain of the material.

Answer

B.

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP

Which of the following sectional shapes will give the highest beam stiffness?
a Hollow tube of outer diameter 15 cm and inner diameter 14 cm
b I-beam with a central section 11 cm high by 2 cm wide and flanges 2 cm high by 10 cm wide (giving a total height of 15 cm)

Answer

The I-beam.

The second moment of area of a shape is given by:

$$I = \int_0^{y_{\max}} y^2 \mathrm{d}A \tag{7.10.1}$$

Therefore the second moment of area for the hollow tube is given by:

$$I = \int_0^{D/2} y^2 \mathrm{d}A - \int_0^{d/2} y^2 \mathrm{d}A = \int_0^{d/2} y^2 \pi y \mathrm{d}y - \int_0^{D/2} y^2 \pi y \mathrm{d}y = \int_{d/2}^{D/2} \pi y^3 \mathrm{d}y \tag{7.10.2}$$

$$= \left[\frac{\pi y^4}{4}\right]_{a/2}^{D/2} = \frac{\pi\left((D/2)^4 - (d/2)^4\right)}{4} = \frac{\pi}{64}\left(D^4 - d^4\right) \tag{7.10.3}$$

For the hollow tube of interest here

$$I = \frac{\pi}{64}\left(15^4 - 14^4\right) = 599\mathrm{cm}^4 \tag{7.10.4}$$

The second moment of area of the I-beam is given by:

$$I = \int_0^{H/2} y^2 \mathrm{d}A - \int_0^{(H/2)-t} y^2 \mathrm{d}A = 2\int_0^{H/2} y^2 W \mathrm{d}y - 2\int_0^{(H/2)-t} y^2 (W-t)\mathrm{d}y \tag{7.10.5}$$

$$= 2\left[\frac{y^3 W}{3}\right]_0^{H/2} - 2\left[\frac{y^3 (W-t)}{3}\right]_0^{(H/2)-t} = \frac{1}{12}\left(WH^3 - (W-t)(H-2t)^3\right) \tag{7.10.6}$$

Therefore the second moment of area of the I-beam shown is:

$$I = \frac{1}{12}\left(10(15)^3 - (10-2)(15-4)^3\right) = \frac{1}{12}(33750 - 10648) = 1925\mathrm{cm}^4 \tag{7.10.7}$$

A solid rectangular section beam of length, L = 100 cm, height, h = 5 cm and width, w = 1 cm, is loaded under symmetrical 4-point bending, with 1000 N downward forces applied at 40 cm in from both ends of the bar, which is supported at both ends. A deflection

of 5 mm is measured at the centre of the beam. Using these data, calculate the Young's modulus of the beam. From your answer, suggest a likely material for the beam.

hint: The equation for the deflection of a beam under symmetrical 4-point bending is:

$$\delta = \frac{FL_1^2}{6EI}(2L_1 + 3L_2) \tag{7.10.8}$$

Answer

E = 72 GPa . The beam is probably made of aluminium.

The second moment of area of the beam is:

$$I = \int_0^{h/2} y^2 \mathrm{dA} = 2\int_0^{h/2} y^2 w \mathrm{d}y = 2w\left[\frac{y^3}{3}\right]_0^{h/2} = \frac{wh^3}{12} = \frac{0.01(0.05)^3}{12} = 1.04 \times 10^{-7}\mathrm{m}^4 \tag{7.10.9}$$

For symmetrical 4-point bending the equation for the deflection of the central region of the bar is:

$$\delta = \frac{FL_1^2}{6El}(2L_1 + 3L_2) \tag{7.10.10}$$

where $2L_1 + L_2 = L$

and in this case L_1 =40 cm and L_2 =20 cm
Rearranging to obtain the Young's modulus:

$$E = \frac{FL_1^2}{6I\delta}(2L_1 + 3L_2) = \frac{1000(0.4)^2}{6\left(1.04 \times 10^{-7}\right)\left(5 \times 10^{-3}\right)}(2(0.4) + 3(0.2)) = 72\mathrm{GPa} \tag{7.10.11}$$

The beam has a Young's modulus of 72 GPa, and so therefore is probably made from aluminium. (Glass also has a Young's modulus in the region of 70 GPa, but is unlikely to be used to produce a beam of this type).

Calculate the shear modulus, G , of a material supplied in the form of a hollow tube (length 100 cm, outer diameter 5 cm, wall thickness 0.1 cm), given that, when it is subjected to an applied torque of 1000 N m, an angular twist of 0.10 radians is generated.

hint: Start by calculating the polar moment of inertia for a hollow tube

Answer

G = 108 GPa.

The polar moment of area of the beam is:

$$I_p = \int_A \rho^2 \mathrm{d}A = \int_0^{D/2} \rho^2 2\pi\rho \mathrm{d}\rho - \int_0^{d/2} \rho^2 2\pi\rho \mathrm{d}\rho = \left[\frac{2\pi\rho^4}{4}\right]_{d/2}^{D/2} = \frac{\pi}{32}\left(D^4 - d^4\right) \tag{7.10.12}$$

Therefore for the beam dimensions given, the polar moment of inertia is:

$$I_\mu = \frac{\pi}{32}\left(0.050^4 - 0.048^4\right) = 9.24 \times 10^{-8}\mathrm{m}^4 \tag{7.10.13}$$

The equation for the angle of twist is:

$$\theta = \frac{TL}{GI_p} \tag{7.10.14}$$

Rearranging this to give the shear modulus, G

$$G = \frac{TL}{\theta I_p} \tag{7.10.15}$$

It follows that:

$$G = \frac{1000(1)}{0.10\left(9.24 \times 10^{-8}\right)} = 108\text{GPa} \tag{7.10.16}$$

This page titled 7.10: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

8: Brillouin Zones

www.doitpoms.ac.uk/tlplib/br...ones/index.php

Learning Objectives

- Understand how to generate the reciprocal lattice from the real space lattice
- Know how to construct Brillouin Zones in reciprocal space
- Be familiar with the Brillouin Zones for some simple two- and three-dimensional structures

Before you start

Though this package is largely self-contained, some familiarity with simple crystal lattices is useful. If you want a reminder, look at the teaching and learning package on Crystallography and Miller indices and lattice planes.

8.1: Introduction
8.2: Reciprocal lattice vectors
8.3: Brillouin Zone construction
8.4: The general case in three dimensions
8.5: Zone Folding
8.6: Brillouin Zones in Three Dimensions
8.7: Summary
8.8: Questions

This page titled 8: Brillouin Zones is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

8.1: Introduction

Leon Brillouin (1889-1969) first introduced Brillouin Zones in his work on the general properties of periodic structures. The concept and construction of the zones can appear quite abstract, but this is largely a result of their wide range of useful applications.

Brillouin zones are polyhedra in reciprocal space in crystalline materials and are the geometrical equivalent of Wigner-Seitz cells in real space. Physically, Brillouin zone boundaries represent Bragg planes which reflect (diffract) waves having particular wave vectors so that they cause constructive interference.

This page titled 8.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

8.2: Reciprocal lattice vectors

Every periodic structure has two lattices associated with it. The first is the real space lattice, and this describes the periodic structure. The second is the reciprocal lattice, and this determines how the periodic structure interacts with waves. This section outlines how to find the basis vectors for the reciprocal lattice from the basis vectors of the real space lattice.

Reciprocal lattice vectors, **K**, are defined by the following condition:

$$e^{i\mathbf{K}\cdot\mathbf{R}} = 1 \tag{8.2.1}$$

where **R** is a real space lattice vector. Any real lattice vector may be expressed in terms of the lattice basis vectors, $\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{a_3}$.

$$\mathbf{R} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3 \tag{8.2.2}$$

in which the c_i are integers. The condition on the reciprocal lattice vectors may also be expressed as

$$\mathbf{b}_1 = \frac{2\pi \cdot \mathbf{a}_2 \times \mathbf{a}_3}{|\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3|} \tag{8.2.3}$$

$$\mathbf{b}_2 = \frac{2\pi \cdot \mathbf{a}_3 \times \mathbf{a}_1}{|\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3|} \tag{8.2.4}$$

$$\mathbf{b}_3 = \frac{2\pi \cdot \mathbf{a}_1 \times \mathbf{a}_2}{|\mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3|} \tag{8.2.5}$$

Note that $\mathbf{b_2}$ and $\mathbf{b_3}$ are given by cyclic permutations of the expression for $\mathbf{b_1}$. From this expression it may be seen that the real lattice basis vectors and the reciprocal lattice basis vectors satisfy the following relation:

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij} \tag{8.2.6}$$

where δ_{ij} is the Kronecker delta, which takes the value 1 when *i* is equal to *j*, and 0 otherwise. Any reciprocal lattice vector may then be expressed as a linear sum of these reciprocal basis vectors:

$$\mathbf{K} = h\mathbf{b}_1 + k\mathbf{b}_2 + l\mathbf{b}_3 \tag{8.2.7}$$

in which *h*, *k* and *l* are integers. The set of all **K** vectors defines the reciprocal lattice.

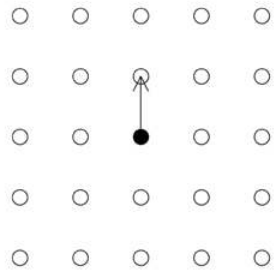
This page titled 8.2: Reciprocal lattice vectors is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

8.3: Brillouin Zone construction

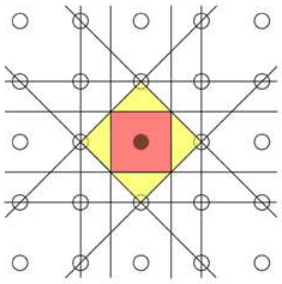
The reciprocal lattice basis vectors span a vector space that is commonly referred to as reciprocal space, or often in the context of quantum mechanics, **k** space. This section covers the construction of Brillouin zones in two dimensions.

The first step is to use the real space lattice vectors to find the reciprocal lattice vectors and construct the reciprocal lattice. One of the points in the reciprocal lattice is then designated to be the origin. When constructing Brillouin zones, they are always centred on a reciprocal lattice point, but it is important to keep in mind that there is nothing special about this point as each reciprocal lattice point is equivalent due to translation symmetry.

Draw a line connecting this origin point to one of its nearest neighbours. This line is a reciprocal lattice vector as it connects two points in the reciprocal lattice.



The second Brillouin Zone is the region of reciprocal space in which a point has one Bragg Plane between it and the origin. This area is shaded yellow in the picture below. Note that the areas of the first and second Brillouin Zones are the same.



The construction can quite rapidly become complicated as you move beyond the first few zones, and it is important to be systematic so as to avoid missing out important Bragg Planes.

Illustration that follows this process to show the first six Brillouin Zones for the 2-D square reciprocal lattice.

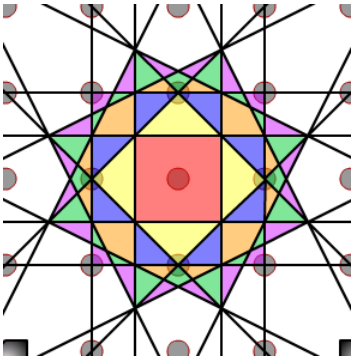
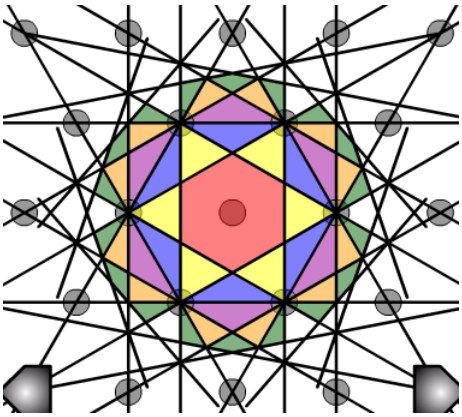


Illustration showing the first six zones for a 2-D hexagonal reciprocal lattice.



This page titled 8.3: Brillouin Zone construction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

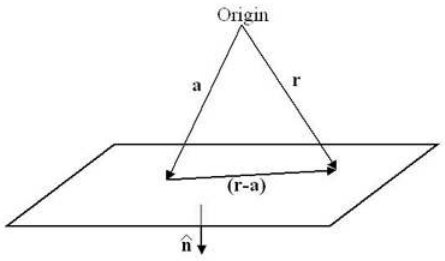
8.4: The general case in three dimensions

After considering these simple examples, it should hopefully be clear how to construct Brillouin Zones. Whilst in two dimensions this geometric method is easy to apply, in three dimensions the lattice cannot be represented on a piece of paper and in general it is much harder to picture the shape of the Brillouin Zones beyond the first. This section considers how the relevant Bragg planes for zone construction may be generated in a systematic fashion.

In vector notation, the equation of a plane may be written as

$$(\mathbf{r}-\mathbf{a})\cdot\hat{\mathbf{n}}=0 \tag{8.4.1}$$

In this expression, **a** is a vector from the origin to a specified (but arbitrary) point in the plane, **r** is a general point in the plane and $\hat{\mathbf{n}}$ is the unit vector normal to the plane.



For Brillouin Zones it is convenient to choose a so that it is the perpendicular vector from the origin to the Bragg plane of interest, i.e., a vector of the form

$$\mathbf{a}=\frac{1}{2}\left(h\mathbf{b}_1+k\mathbf{b}_2+l\mathbf{b}_3\right) \tag{8.4.2}$$

also, the unit normal is then given by

$$\hat{\mathbf{n}}=\frac{\mathbf{n}}{|\mathbf{a}|} \tag{8.4.3}$$

Letting *h*, *k* and *l* be integers (positive or negative) and excluding the point where all three are equal to zero, the relevant Bragg Planes may be generated in a systematic fashion. It is usually sufficient for finding the first three or four zones to have the d_i range between –3 and +3.

Then, given any point in reciprocal space, it may be allocated to a Brillouin zone by determining the number, *N*, of Bragg Planes that lie in between that point and the origin. The point is then in the (*N*+1)th Brillouin Zone.

Whether a Bragg plane lies between a general point, **r**, and the origin may be determined quite simply by considering the projection of the position vector of the point in the direction of the unit normal to the Bragg plane. The scalar product

$$\mathbf{a}\cdot(\mathbf{r}-\mathbf{a}) \tag{8.4.4}$$

is positive if the plane is between the point and the origin, and is negative when the plane does not lie between the point and the origin.

The main complication when extending to Brillouin Zones beyond the first zone is that it is very easy to overlook an important Bragg Plane. This can only really be avoided by being careful and systematic. A useful test to indicate if a mistake has been made is to check that all the zones have the same symmetries as the reciprocal lattice itself. For example the Brillouin Zones for the 2-D Square lattice always have fourfold-rotational symmetry about the origin. A further useful check is to confirm that the area (in 2D) or volume (in 3D) of each Brillouin zone is the same.

Some examples which show the Brillouin Zones for common 2-D lattices may be found by clicking the links below.

- 2-D Square lattice to 10th Brillouin zone (pdf file, 0.53 MB)
- 2-D Hexagonal lattice to 6th Brillouin zone (pdf file, 0.60 MB)

This page titled 8.4: The general case in three dimensions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

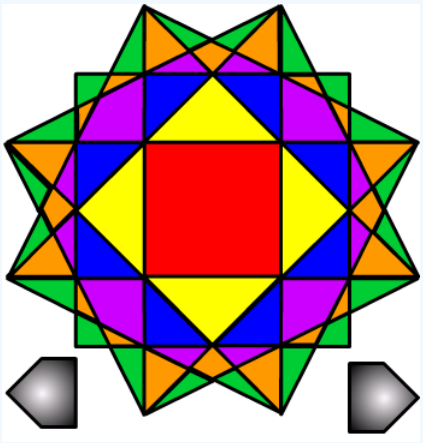
8.5: Zone Folding

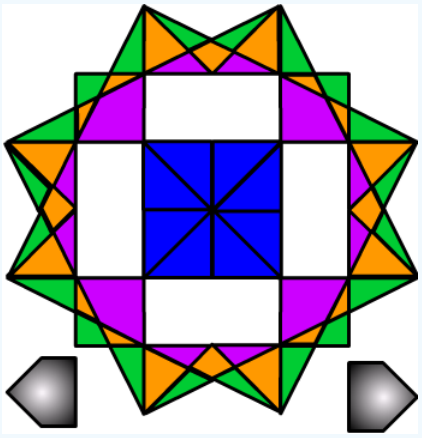
An important property of the Brillouin Zones is that, because the reciprocal lattice is periodic, there exists for any point outside the first zone a unique reciprocal lattice vector that will translate that point back inside the first zone. Each point in reciprocal space is only unique up to a reciprocal lattice vector. Each Zone contains every single physically distinguishable point, and so they all have the same area (in 2-D) or volume (in 3-D).

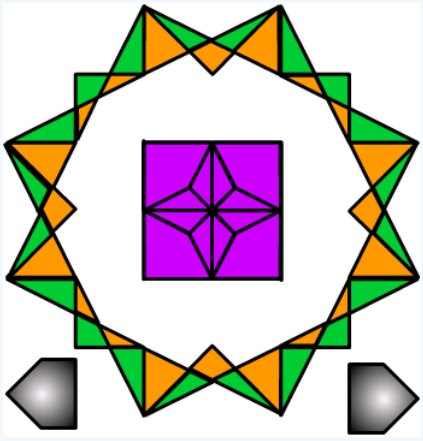
This is easiest to see by example. The illustrations below will show how the first six zones for the 2-D square and hexagonal lattices can be translated or 'folded' back on top of the first zone.

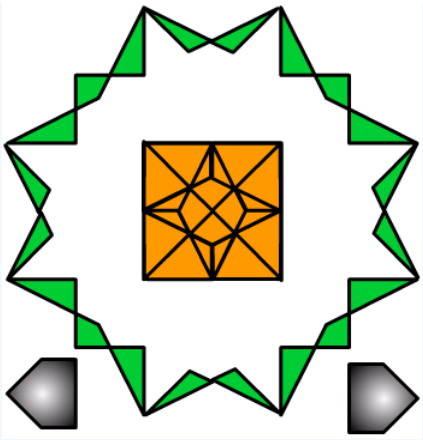
Example 8.5.1

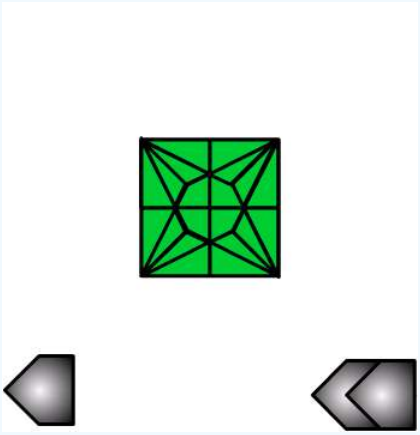
2-D square Zone folding





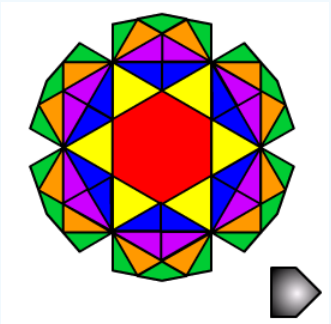


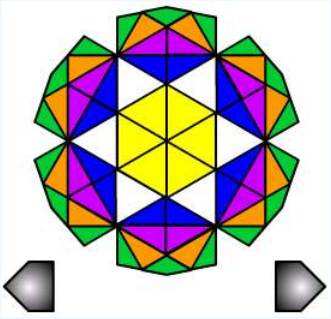


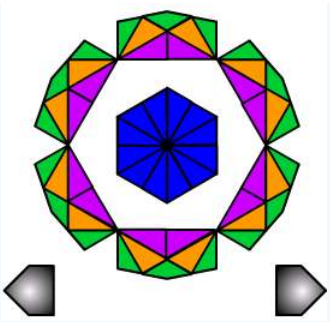


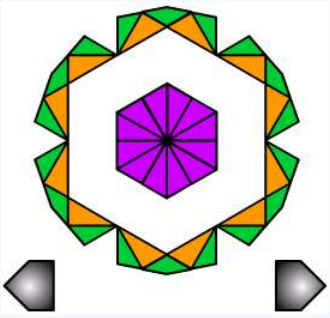
Example 8.5.1

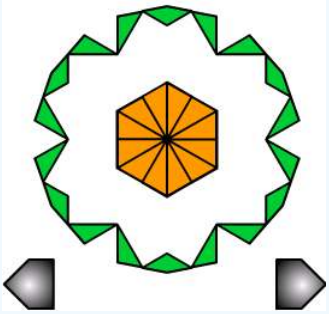
2-D hexagonal Zone folding

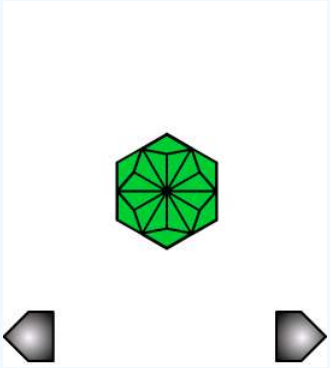












This page titled 8.5: Zone Folding is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

8.6: Brillouin Zones in Three Dimensions

The extension to three dimensions is straightforward by the method outlined in the previous section. However, it is important to remember not to overlook important Bragg Planes. There is also a problem of representation, because the structure of the Zones in 3-D can be quite complicated and hard to visualise.

It is also important to remember that in 3-D the Zones all have the same volume, and that the volume corresponding to the 3rd Zone is the volume between the outer surface of the 2nd Zone and that of the 3rd Zone.

Representations of the Brillouin Zones corresponding to Simple Cubic (SC), Body Centred Cubic (BCC), Face Centred Cubic (FCC) lattices are linked to below. Use the buttons on the left hand side to select which Zone Surface to view, and use the arrow buttons to rotate the view about the [001] axis.

Simple Cubic

Body Centred Cubic

Face Centred Cubic

This page titled 8.6: Brillouin Zones in Three Dimensions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

8.7: Summary

This teaching and learning package has introduced Brillouin zones and their construction in two and three dimensions.

The relationship between the construction of Brillouin zones and diffraction from Bragg Planes has been shown by a consideration of the Bragg equation.

The examples shown in the package itself demonstrate the principles of the construction in two dimensions. In three dimensions the principles of the construction are the same, but the Brillouin Zones are hard to visualise. Rotating the shapes of Brillouin Zones about a cube axis as in the examples we have given in this package helps to improve the visualisation in comparison with simple line diagrams found in textbooks.

Going further

Books

There a number of textbooks which discuss in varying levels of detail Brillouin Zones and their construction and application. The classic text on Brillouin Zones is

- Wave Propagation in Periodic Structures, L. Brillouin (Dover 1953, reprinted 2003)

Other texts are also useful, such as:

- Solid State Physics, J.S. Blakemore (Saunders 1974)
- Electronic Properties of Materials, R.E. Hummel (Springer-Verlag 1992)
- Introduction to Solid State Physics, C. Kittel (Wiley 1996)
- Solid State Physics, N.W. Ashcroft and N. D. Mermin. (Harcourt Brace 1976)

Websites

- Wikipedia - Brillouin zones
  This Wikipedia page has a number of useful external links which should help to reinforce and confirm the concepts introduced in this teaching and learning package.

Further links can be readily found by typing “Brillouin Zones” into Google and searching the web. Examples of such links are:

- http://www.tf.uni-kiel.de/matwis/amat/semi_en/kap_2/illustr/i2_1_3.html
- http://britneyspears.ac/physics/crystals/wcrystals.htm

This page titled 8.7: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

8.8: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

1. If a real space vector has length L, what is the magnitude of the corresponding reciprocal lattice vector as defined in this TLP?

a	$2\pi L$
b	$\frac{\pi}{L}$
c	$\frac{1}{L}$
d	$\frac{2\pi}{L}$

Answer

D

Which of the following statements about the Brillouin zones for a particular reciprocal lattice are correct? True or False.

a	They all have the same area / volume.
b	They all have the same shape.
c	They all have the same symmetry.

Answer

A. True

B. False

C. True

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

By either printing out the lattice in the file linked to here (pdf), or by drawing out a rectangular lattice with the ratio between the short and long sides of 4:5. Use the method outlined earlier to construct the first three Brillouin zones, being careful not to miss any relevant Bragg planes.

Answer

Check your answer against the illustration linked to here (pdf), which shows the Brillouin zones out to the 6th zone.

By either printing out the lattice in the file linked to here (pdf), or by drawing out a parallelogram lattice with the ratio between the short and long sides of 4:5 and an angle between the sides of about 75 degrees. Use the method outlined earlier to construct the first three Brillouin zones, being careful not to miss an relevant Bragg planes.

Answer

Check your answer against the illustration linked to here (pdf), which shows the Brillouin zones out to the 6th zone.

Confirm that when expressed in a Cartesian basis, possible lattice vectors of the primitive unit cell (i.e., the unit cell with one lattice point per unit cell) for a FCC lattice of unit side length are

$a_1 = \begin{bmatrix} 1/2 & 1/2 & 0 \end{bmatrix}$

$a_2 = \begin{bmatrix} 1/2 & 0 & 1/2 \end{bmatrix}$

$a_3 = \begin{bmatrix} 0 & 1/2 & 1/2 \end{bmatrix}$

Confirm that when expressed in a Cartesian basis, possible lattice vectors of the primitive unit cell for a BCC lattice of unit side length are

$\mathbf{d}_1 = \begin{bmatrix} -1/2 & -1/2 & 1/2 \end{bmatrix}$

$\mathbf{d}_2 = \begin{bmatrix} -1/2 & 1/2 & -1/2 \end{bmatrix}$

$\mathrm{d}_3 = \begin{bmatrix} 1/2 & -1/2 & -1/2 \end{bmatrix}$

Using the definition of the reciprocal lattice basis vectors given in the TLP, show that the reciprocal lattice of a FCC lattice is a BCC lattice and vice versa.

For a conventional BCC unit cell with two lattice points per unit cell of side *L*:

a. How many nearest neighbours does each point have?
b. What are their coordinates, and what distance away are they?
c. How many next nearest neighbours are there?
d. What are their coordinates and what distance away are they?

Answer

a. 8

b. [□ □ □], [-□ □ □], [□ -□ □], [□ □ -□], [-□ -□ -□], [-□ -□ □], [-□ □ -□] and [□ -□ -□] all at distance of $\frac{\sqrt{3}}{2}L$.

c. 6

d. [1 0 0], [0 1 0], [0 0 1], [-1 0 0], [0 -1 0] and [0 0 -1] all at distance *L*.

Look at the picture below of the first Brillouin zone for a real space FCC lattice. Using your answers to questions 5 and 6, try to index the planes labelled (a) and (b).

hint: Remember that the Brillouin zones are constructed using the reciprocal lattice not the real space lattice

Answer

(a) (1 0 0) and (b) (1 1 1)

This page titled 8.8: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

9: Brittle Fracture

www.doitpoms.ac.uk/tlplib/br...ture/index.php

Learning Objectives

- that materials break by cracking;
- what determines whether a material will crack or not;
- what determines whether cracking is catastrophic or more gradual;
- the concepts of the *fracture energy*, *strain energy release rate*, *fracture toughness* and *stress intensity factor*.

Before you start

You need not do it now, but you may want to look at the TLP on photoelasticity.

9.1: Introduction
9.2: When do atomic bonds break?
9.3: Why do cracks weaken a material?
9.4: Inglis and the crack tip stress idea
9.5: Can we calculate the energy changes?
9.6: What about tension?
9.7: Another way of expressing the energies
9.8: Another way of calculating the energies
9.8: Why bother if they are the same?
9.9: Coping with a scatter in strength
9.10: When does the sample fail completely?
9.11: Sub-critical crack growth and R-curves
9.12: Summary
9.13: Questions

This page titled 9: Brittle Fracture is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.1: Introduction

The consequences of something breaking can be a pest, or utterly disastrous, as when the pedal drops off one's bike, but without it, biting and crunching, breaking into crisp packets, pulverizing coal, oil drilling and many other processes would be impossible. The most dramatic failures are catastrophic, but sometimes they can be very gradual even in the most brittle materials.

Here we discuss what determines when a material will break, and whether failure will be catastrophic or more gradual. We show cracking is controlled by the energy changes that occur - it is **not** the stress **at** the crack tip that is important..

The emphasis here is on brittle fracture, and although all of this is relevant to metals, we do not discuss the details of ductile fracture.

This page titled 9.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.2: When do atomic bonds break?

Fracture is the separation of atoms. We normally do this by applying a force to the body. How big a force should we need?

We want to know how the energy, U, changes with distance, r, between two atoms. A commonly used expression (known as the Lennard-Jones potential is

$$U(r) = 4\varepsilon\left[\left(\frac{\beta^{12}}{r^{12}}\right) - \left(\frac{\beta^{6}}{r^{6}}\right)\right] \tag{9.2.1}$$

Other potentials exist, but the argument is essentially the same.

This contains a short range repulsive term, and a long range attractive term. The parameter ε is a measure of the depth of the potential well, and β is the non-infinite distance where the interparticle potential equals zero.

Use the animation to see how the energy changes as the distance between two atoms is varied.

The force between a pair of atoms is calculated by taking the derivative of the energy function, giving:

$$\boldsymbol{F} = \frac{\mathrm{d}U(\mathrm{r})}{\mathrm{d}r} = 24\varepsilon\left[-2\left(\frac{\beta^{12}}{r^{13}}\right) + \left(\frac{\beta^{6}}{r^{7}}\right)\right] \tag{9.2.2}$$

Now see how the force between the atoms changes with distance. What is the force needed to separate the atoms in a crystal?

Because we are interested in the material's properties we normally think of failure *stresses*, so in a unit area of crystal we need to know how many bonds in an area of crystal normal to the applied force are carrying the load, say 6.25×10^{-22} m^2.

From the animation it is clear that materials should have breaking stresses of the order of 10 GPa, but a piece of glass normally breaks at a stress of 70 MPa.

Where have we gone wrong?

What have we just estimated? It is the stress required to break the bonds on a given plane simultaneously, what becomes the fracture surface. So do all the bonds on a given plane break at once?

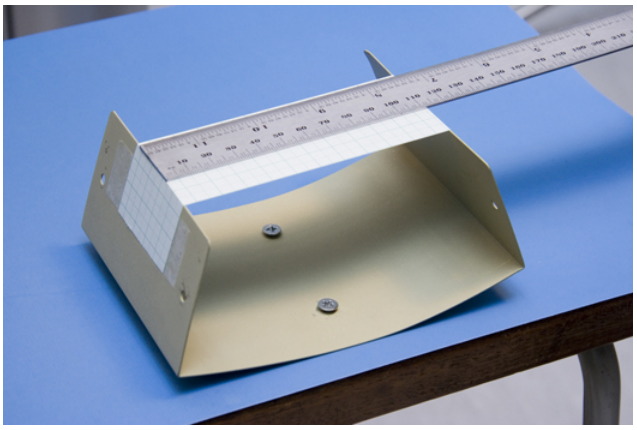
Anybody who has opened a packet of crisps or torn a sheet of paper knows the answer is ‘no’ – the bonds break a few at a time – at the tip of a growing crack.

So the analysis above does not describe what actually happens when a material breaks, which is a relief as the predictions were no good anyway.

This page titled 9.2: When do atomic bonds break? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.3: Why do cracks weaken a material?

Let's watch a crack grow. We've used paper here, because the loads are low and we've applied a force to the paper sheet by stretching it across a metal strip in bending, as shown in the video.



Loading situation of a strip of graph paper.

https://www.doitpoms.ac.uk/tlplib/br...r_fracture.mp4

Video of paper tearing at a preformed crack

You can see that initially the paper is quite capable of withstanding the stress applied to it by the loading jig. And this is fine even when there is quite a big cut in it. But when the cut is lengthened the crack grows quite suddenly.

Is the effect simply one of decreasing the intact section area, causing an increase in stress on the intact section?

Effect of cracks – reducing the intact cross-section

Strips of paper 200 mm long and 70 mm wide were cut. Some had notches cut into one side giving a reduction in the intact cross-section of 5.16×10^{-6} m^2. These and strips without notches were stretched until they broke in a tensile test machine.

The average stress required to break the strip without a notch was measured to be 19.3 MPa.

Where the intact cross-section was reduced by a factor of 1.14 the average breaking stress was measured as 9.4 MPa.

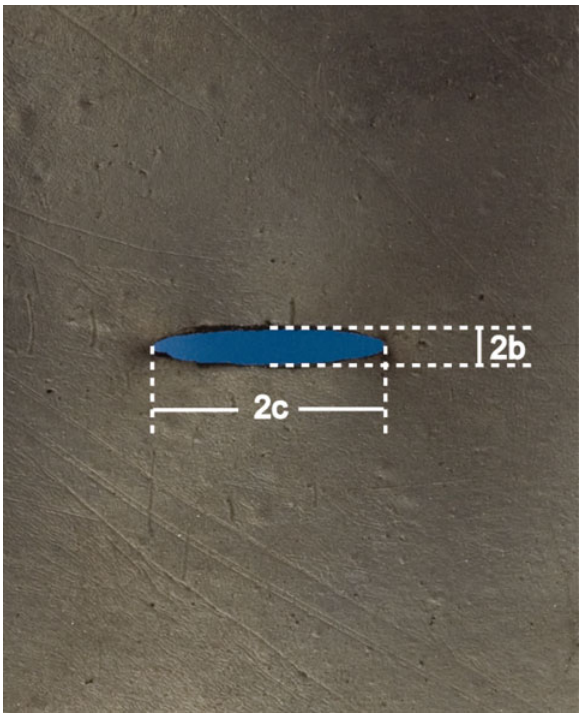
So the breaking strength falls by a greater factor than the reduction in cross-sectional area.

This is why cracks are so dangerous – because they weaken the material more than you would expect from the reduction in intact cross-section. Why?

This page titled 9.3: Why do cracks weaken a material? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.4: Inglis and the crack tip stress idea

In 1913, Inglis calculated what the stresses and strains were in an elastic plate containing an elliptical crack, with semi-axes b and c, and under an applied stress σ - applied vertically in this case.



He found that the stress at the crack tip, σ_t , was given by

$$\sigma_t = \sigma\left(1 + 2\frac{c}{b}\right) \tag{9.4.1}$$

For a sharp crack, i.e. c >> b, the stress would be much greater at the crack tip So failure could occur by cracking because it's only at a crack tip that the stress required to break a bond, or simultaneously break a given number of bonds in a unit area of material, is reached.

So far, so good, but σ_t depends on the SHAPE of the crack, i.e. c / b – and we know that the length is important.

By looking at crack tip stress fields using photoelasticity, we can see that stresses are indeed concentrated at crack tips**.** Inglis was correct about this.

It's the idea that a critical stress to break a bond is needed that is wrong.

This page titled 9.4: Inglis and the crack tip stress idea is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.5: Can we calculate the energy changes?

In the beginning, we calculated the force, and hence the stress, for failure by considering the energy changes as the atoms moved apart.

Why not consider cracks like this?

The first person to do this was A.A. Griffith in 1920. He considered a body in tension, but tension is rather complicated, so let's consider another stress-state – wedging.

Wedging is what you do when you split a piece of wood, or try to peel paper off the wall by getting your fingernail underneath it. It's shown in the animation where a block of thickness h, is being driven in under a layer of thickness d.

So why does the energy, U, of the body change with crack length?

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

From the animation you can see there are 3 contributions:

Pushing the wedge in causes the peeling layer to bend more, increasing the strain energy, U_E. And only the layer between the crack tip and the point where the layer touches the wedge is bending. Using the TLP on beam bending (a more detailed derivation can be found here) gives U_E as

$$U_{\mathrm{E}} = \frac{Ed^3h^2}{8c^3} \tag{9.5.1}$$

What the about work done by the applied force, U_F?

We can see that the action of the wedge is the same as a force applied at the point where the wedge touches the peeling ligament. So as the crack grows the force moves sideways, i.e. perpendicular to its line of action, so no work is done. So

$$U_F = 0 \tag{9.5.2}$$

Finally, as the crack length changes, the energy of the surfaces, U_S, changes, giving

$$U_s = cR \tag{9.5.3}$$

where R is known as the fracture energy, the energy required to create new surfaces.

The total energy, U, of the body is just these 3 terms added up.

So what does this have to do with cracking?

Again explore the animation above and start with the default values.

What will be the crack length, c*, where the body has the lowest energy? We can find c* by differentiating the expression for u with respect to c and setting dU / dc = 0.

What will happen if the crack is shorter than this?

The energy of the body would increase if c < c*, so the crack would grow until c = c*, and then stop.

This is called stable cracking – it happens where there is a stable equilibrium in the energy change with crack length.

But what happens if the initial crack length is longer than the equilibrium value, or if we pull the wedge out?

What is predicted?

What happens if the wedge is pulled out?

If we pull the wedge out, the crack length should decrease; in other words, the crack should heal. To test this, Obreimoff cleaved crystals of muscovite mica by wedging. On pushing the wedge in he found that the crack grew at a constant distance ahead of the crack tip, as predicted, and estimated a value of R for mica close to that measured elsewhere.

Then he pulled the wedge out. His first tests in air showed no crack healing.

Then he tried testing in vacuum – healing really did occur.

Obreimoff thought that this occurred because in air chemical groups attached themselves to the fresh mica surfaces, just as fluff sticks on a toffee in your pocket, so that the surfaces no longer stick.

This showed cracking really was thermodynamic in nature – totally different from the critical stress ideas.

This reversibility, crack healing, seems a strange idea at first. But we know surfaces stick together – that is the cause of friction – and there have been many observations of this on metal sheets, on probes in atomic force microscopes. And it is what we would expect if we think of a surface in terms of unfulfilled chemical bonds, desperate for other electrons.

This page titled 9.5: Can we calculate the energy changes? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.6: What about tension?

It is not as easy to estimate U_E for tension. Griffith used Inglis' analysis for the overall stress and strain fields in a cracked body, under a constant applied load. This gives U_E as

$$U_{\mathrm{E}} = \frac{\pi c^2 \sigma^2}{E} \tag{9.6.1}$$

In other words there is an **increase** in the elastic strain energy as the crack grows.

However when the crack grows, work is done by the applied force, F, and is equal in magnitude to twice the change in elastic strain energy. (Think of the work done by the applied force and the elastic energy changes on a uniform bar loaded in tension.) As it is work done **on** the system, the sign is opposite to that of U_E, so that

$$-U_{\mathrm{F}} + U_{\mathrm{E}} = -\frac{\pi c^2 \sigma^2}{E} \tag{9.6.2}$$

where U_F is the work done by the applied force, U_E is the elastic energy change on cracking and U_S is the work required to create two new surfaces.
The work associated with creating new crack faces, U_S.

$$U_{\mathrm{s}} = 2cR \tag{9.6.3}$$

where R is the fracture energy.

Combining the terms of this energy expression we obtain:

$$U = -\frac{\pi \sigma^2 c^2}{E} + 2cR \tag{9.6.4}$$

The energy function is plotted in the following animation:

How does cracking occur now?

The energy terms above vary with c as shown. Although there is an equilibrium, it is unstable in tension whereas it was stable in wedging.

Adding the energies and differentiating gives the equilibrium crack length c_e as

$$c_e = \frac{ER}{\pi\sigma^2} \tag{9.6.5}$$

This page titled 9.6: What about tension? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.7: Another way of expressing the energies

What we have been doing so far is to calculate how U varies with c, and then find the equilibrium value of c by differentiating U with respect to c. Looking at the terms they fall into two types:

Those that come from the loading system, U_E and U_F: let's add these together and call the sum U_M

- These give the driving force for cracking

Those that are associated with the material, U_S

- This gives the resistance to cracking

Equilibrium will occur when

$$\frac{\mathrm{d}U(c)}{\mathrm{d}c} = 0 \tag{9.7.1}$$

Breaking our energies into the two different types, gives

$$-\frac{\mathrm{d}U_\mathrm{M}}{\mathrm{d}c} = \frac{\mathrm{d}U_\mathrm{s}}{\mathrm{d}c} \tag{9.7.2}$$

From our expression for cracking in tension

$$\frac{\mathrm{d}U_\mathrm{s}}{\mathrm{d}c} = 2R \tag{9.7.3}$$

and

$$\frac{\mathrm{d}U_\mathrm{M}}{\mathrm{d}c} = -\frac{2\pi\sigma^2}{E}c = 2G \tag{9.7.4}$$

where G is the energy per unit area of crack and so is often called the strain energy release rate, or the crack driving force. Now the turning point occurs when

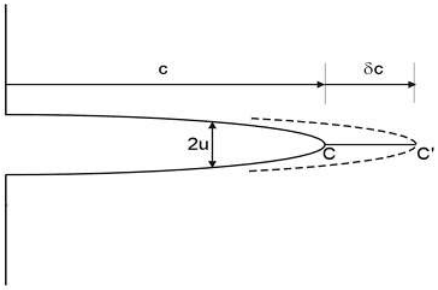
$$G = R \tag{9.7.5}$$

This is our condition for cracking, that the crack driving force, G, equals the fracture resistance of the material, R.

This page titled 9.7: Another way of expressing the energies is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.8: Another way of calculating the energies

Consider a crack growing from C by an amount δc to *C'*. Before cracking occurs there are stresses acting ahead of the crack and across the plane C − C'. After crack growth these stresses have been relaxed to zero as the newly created surfaces move from their initial position at u = 0 to form the crack.



If we knew the stresses ahead of the crack and displacements that relaxed them to zero, we could estimate the total mechanical work, δU_M, when the crack grows by δc.

What are the displacements, u?

To find these we make a sharp cut in a rubber sheet, something compliant enough to get measurable shape changes without breaking, and stretch it, see below



We can plot the opening of the crack, *u*, against the distance from the crack-tip, r. Remember r is taken as positive ahead of the growing crack. The opening is therefore at negative values of r. From the picture

$$u \propto -r^{1/2} \tag{9.8.1}$$

where *r* is the distance from the crack tip, taken as positive in the direction ahead of the growing crack. The crack opening, *u*, at a given point is therefore at a negative value of r. A full elastic analysis gives

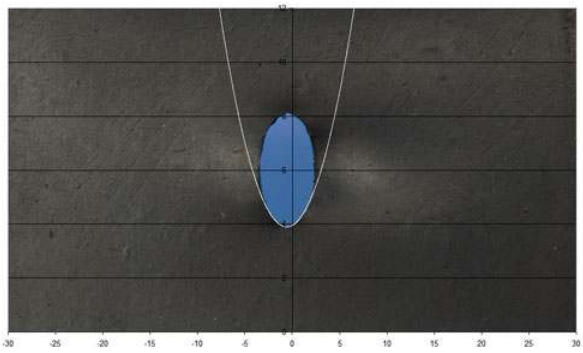
$$u = 4\frac{K}{E}\left(\frac{-r}{2\pi}\right)^{1/2} \tag{9.8.2}$$

where *K* is a constant known as the *stress intensity factor* and E is the Young modulus.

We can see this well by simply stressing a rubber sheet with a cut slit, the shape of the crack being clearly parabolic. A curve with equation

$y = m(x - a)^2 + c$ can be fitted to the tip of the curve.

https://eng.libretexts.org/@go/page/30053



What are the stresses ahead of a crack tip?

If the crack opening is parabolic, it seems not unreasonable to think that the stresses ahead of the crack (not at the crack tip) are too, so

$$\sigma = \frac{K}{(2\pi r)^{1/2}} \tag{9.8.3}$$

We can reassure ourselves by carrying out a Finite Element simulation. Taking the expressions for u and σ and as the elastic energy is proportional to σ × u, we can see that the change in mechanical energy, δU_M, is likely to be proportional to K^2/E. To check we need to integrate δU_M over the increment of crack growth δc. See here. As expected, this gives:

$$G = \frac{K^2}{E} \tag{9.8.4}$$

Now, we said cracking occurred when G = R, an equivalent expression is

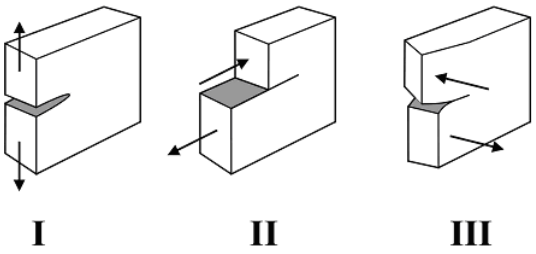
$$K = K_C \tag{9.8.5}$$

where K_C is the critical stress intensity factor at which the crack will grow, known as the fracture toughness. Both criteria are based on the idea that the mechanical energy per increment of crack growth must reach some critical value. Both are entirely equivalent and both calculate the energy changes on cracking.

This page titled 9.8: Another way of calculating the energies is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.8: Why bother if they are the same?

K is useful because stress intensity factors, like stresses, are additive. Irwin showed that any loading state could be broken down into 3 different types of loading that he called modes, I, II and III:



To get the total K, we just sum the contribution from each of the three modes.

This is much more difficult to do with energies, where we would have to think about interaction terms.

This page titled 9.8: Why bother if they are the same? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.9: Coping with a scatter in strength

The Griffith expression shows that the stress required for failure in tension is dependent on the size of the largest flaw, c, according to .

In very brittle materials the flaw sizes cannot be easily measured, so it is sometimes impossible to calculate a minimum strength.

Using a method by Weibull, we can then find the chance of survival of a sample as a function of applied stress. We can then extrapolate back to an acceptable probability of failure and find the corresponding stress.

The linked derivation gives a relationship of

\ln \ln \frac{1}{S_{n}}=m \ln \sigma-\left(m \ln \sigma_{\mathrm{o}}+\ln N\right)

where S_n is the probability of survival, and σ is the fracture stress.

We can treat our results from the tensile of testing of paper with Weibull statistics:

Using Weibull's method

To determine the survival probability associated with each stress we start by testing some samples, in this case the graph paper used in the first video of the TLP. The following table shows all the collected data, and the calculated values needed:

n	UTS	S_n	ln ln $(1/S_n)$	ln σ
	/MPa			
1	21.28	0.07	0.9704	16.8733
2	20.86	0.14	0.6657	16.8531
3	20.20	0.21	0.4321	16.8213
4	20.20	0.29	0.2254	16.8211
5	20.12	0.36	0.0292	16.8173
6	19.30	0.43	-0.1657	16.7755
7	19.03	0.50	-0.3665	16.7615
8	18.87	0.57	-0.5805	16.7531
9	18.53	0.64	-0.8168	16.7347
10	18.49	0.71	-1.0892	16.7330
11	18.11	0.79	-1.4223	16.7122
12	18.10	0.86	-1.8698	16.7112
13	17.74	0.93	-2.6022	16.6912

From Weibull treatment of graph paper test.xls

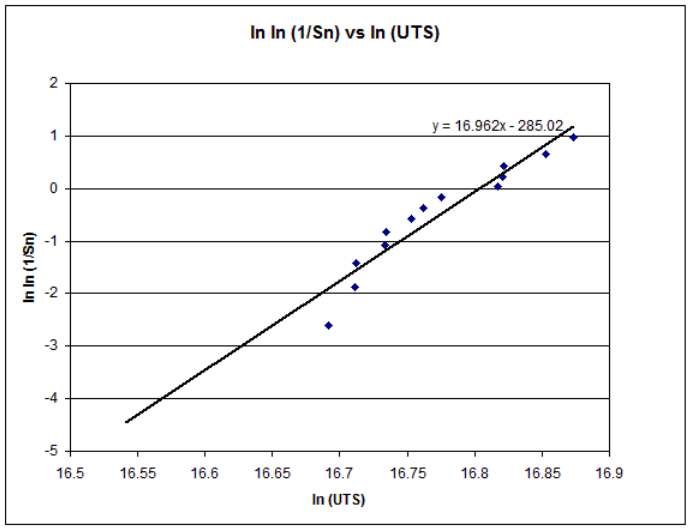
Remember that S_n is the probability of survival at a given stress, given by

S_{n}=\frac{n}{n_{\mathrm{T}}+1}

, where n_T is the total number of samples and σ is the fracture stress of the sample.

The data has been ordered with the highest strength being 1, and the lowest numbered down to n.

When we plot the graph of ln ln $(1/S_n)$ vs ln (UTS), we get the following:



From Weibull treatment of graph paper test.xls

We can then read off the graph for various values of S_n. For instance if we wanted a 99% chance of survival, we read of from –4.60 on the y axis, which corresponds to 15.1MPa.

A fully interactive version of this spreadsheet and graph tool can be downloaded here.

This page titled 9.9: Coping with a scatter in strength is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.10: When does the sample fail completely?

It is incorrect to say that **failure** must occur when

$$G = R \tag{9.10.1}$$

There will be some cracking but complete failure (as in tension) also requires that

$$\frac{\mathrm{d}^2 U(c)}{\mathrm{d}c^2} < 0 \tag{9.10.2}$$

i.e. the energy is at a maximum, or

$$\frac{\mathrm{d}G}{\mathrm{d}c} > \frac{\mathrm{d}R}{\mathrm{d}c} \tag{9.10.3}$$

In other words failure will be catastrophic when the rate of increase of the driving force with crack growth is greater than the change in R with crack growth, which we have taken as a constant.

Alternatively cracking will be stable when

$$\frac{\mathrm{d}^2 U(c)}{\mathrm{d}c^2} > 0 \tag{9.10.4}$$

i.e. the energy is at a minimum, or

$$\frac{\mathrm{d}G}{\mathrm{d}c} < \frac{\mathrm{d}R}{\mathrm{d}c} \tag{9.10.5}$$

That is, as the crack grows, the resistance to cracking, R, increases faster than the driving force, G.

This page titled 9.10: When does the sample fail completely? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.11: Sub-critical crack growth and R-curves

It is commonly observed that cracks can grow stably in a structure over a period of time. Oddly this phenomenon, known as *fatigue*, tends to occur in tougher materials. This suggests that toughening does not simply increase the magnitude of the fracture energy but changes the way in which a crack grows.

Consider a material toughened by crack bridging in which intact ligaments across the crack faces are left behind as the crack grows. Toughening occurs because separating the crack faces then requires extra work in order to either stretch the ligament or pull it out of the matrix in which it is embedded. This type of mechanism occurs in many different materials, including:

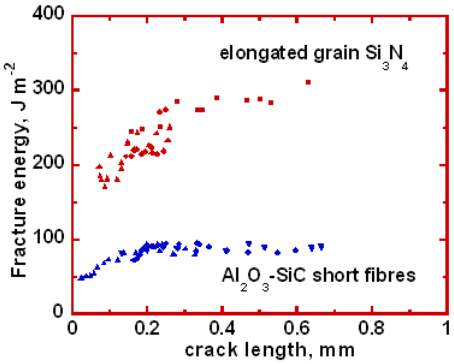
- rubber toughened polymer
- materials containing either fibres or elongated grains

This animation shows how an R-curve is generated when a crack propagates through a material containing ligaments perpendicular to the direction of crack propagation.

After each critical point, the simulation will pause. It can then be continued by clicking the 'proceed' button.

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

This type of behaviour is known as R-curve behaviour. We can see that as we load the material containing a small flaw, it will begin to grow (under an increasing applied stress intensity factor) until the process zone is fully developed. The crack in the process zone has a different shape to that outside it because of the forces that due to the intact ligaments. Once the process zone has developed fully then the whole crack will move forward with the process zone size remaining a constant size. As process zones exist in all toughened materials we might expect that they would all show R-curves and this is the case as shown below.



Showing the R-curve due to grain interface bridging in a silicon nitride containing elongated grains and an alumina containing silicon carbide fibres.

This page titled 9.11: Sub-critical crack growth and R-curves is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.12: Summary

It has been shown that energy arguments describe how cracking occurs, but that there is more than one way of working out the energy changes.

Going further

Books

A best overall book on cracking is that by Brian Lawn, *Fracture of brittle solids*, 1993, Cambridge University Press. You should also look at *Molecular adhesion and its applications* by Kevin Kendall. For the metalllurgist, the short monograph by Withers & Knott is exceptionally useful.

This page titled 9.12: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

9.13: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

The Lennard-Jones potential best describes which of the following:

a	The energy required to break a typical metal
b	How flaws affect the energy of a lattice
c	The energy associated with stretching an atom

Answer

C

Which of the following situations is required for fast fracture to occur?

a	The crack tip is sufficiently sharp
b	dG/dc is greater than dR/dc
c	The applied stress is near the ultimate tensile strength of the material
d	The strain energy release rate G is equal to the fracture energy R.

Answer

B

In the wedging setup, which of the following will *not* influence the critical crack length?

a	The position of the wedge
b	The stiffness of the material
c	The thickness of the peeling layer
d	The surface energy of the material

Answer

A

Which one of Irwin's modes of fracture best describes a car tyre skidding on a dry road?

a	Mode I
b	Mode II
c	Mode III

Answer

B

What is the critical stress intensity factor equal to when no residual stresses are present?

a	Fracture toughness
b	Crack resistance
c	Ultimate tensile strength
d	Fracture energy

Answer

A

What best describes the situation in wedging?

a	As the wedge position is varied, it is always energetically favourable for the crack to return to an equilibrium length
b	The fracture energy dominates at small crack lengths
c	If the crack becomes very long in comparison to the geometry of the setup, unstable fast fracture may occur

Answer

A

This page titled 9.13: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

10: Casting

https://www.doitpoms.ac.uk/tlplib/casting/index.php

Learning Objectives

- Understand the meaning of the Biot number, how it affects the temperature profile in a casting, and the resulting microstructure.
- Be able to explain the formation of the microstructure observed in a cast ingot.
- Be familiar with some common methods of casting, their advantages and disadvantages, and be able to choose a suitable process for manufacturing a variety of metallic components.

Before you start

It will be helpful to have an understanding of solute partitioning and the formation of dendrites. The TLP on Solidification of Alloys covers these topics.

10.1: Introduction
10.2: Heat Transfer
10.3: Heat Transfer Simulation
10.4: Microstructure and Segregation in Castings
10.5: Sand Casting
10.6: Die Casting
10.7: Continuous Casting
10.8: Investment Casting
10.9: Other Methods of Casting
10.10: Summary
10.11: Questions

This page titled 10: Casting is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.1: Introduction

Casting processes are essential in forming solid ingots or billets of metals for further processing, but moreover, a wide variety of casting processes are used in the production of finished or semi-finished components. Casting allows the production of complex shapes, without the need to introduce weaknesses when joining separate pieces, and is conservative of material compared with machining or cutting a shape from a large piece of metal. The methods available can achieve various standards of surface finish, microstructure, start-up cost, unit cost, and production volume, in order to suit many applications.

This page titled 10.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.2: Heat Transfer

The rate of solidification of a metal during casting is dictated by;

- The excess heat in the liquid metal on pouring,
- The amount of heat produced by the solidification of the metal (the latent heat of fusion),
- The rate at which this heat can be dissipated from the metal.

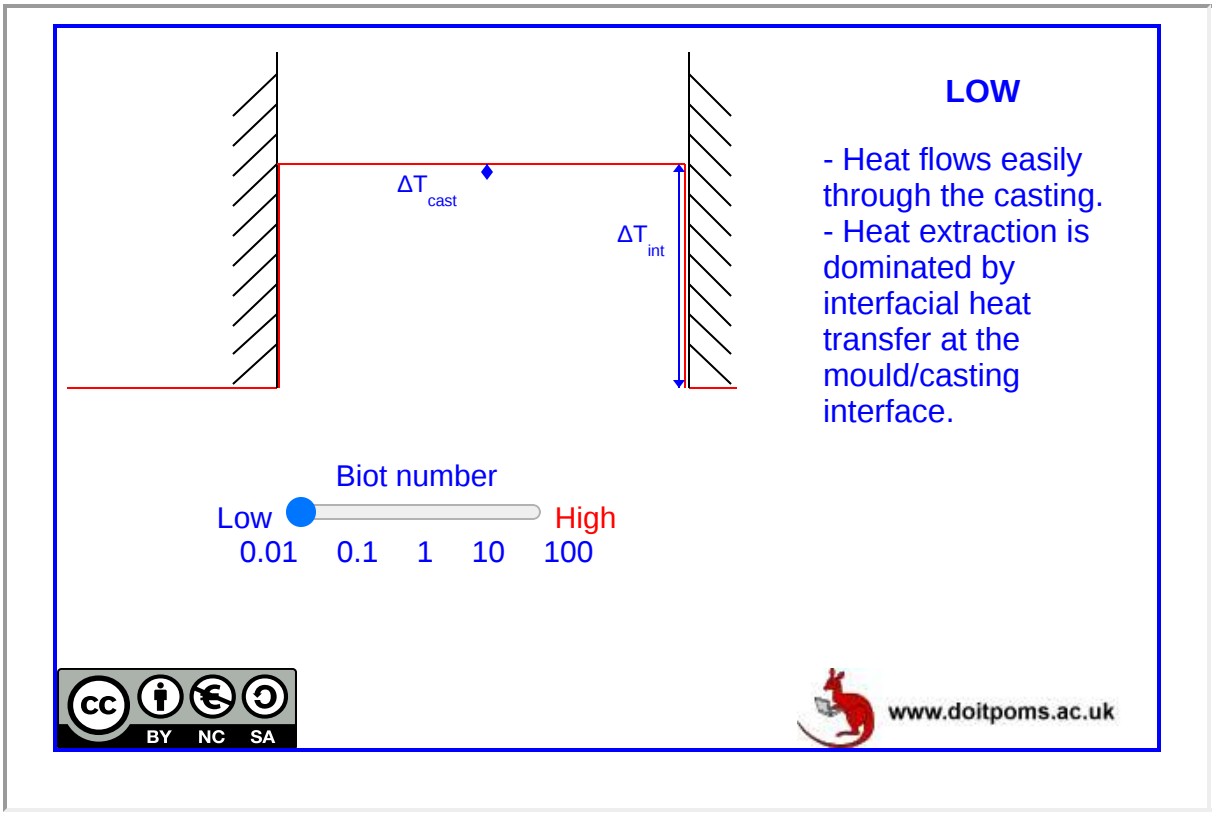
A simple way to predict the way in which a casting will solidify, is using the Biot number, Bi, given by:

$$Bi = \frac{h}{K/L} \tag{10.2.1}$$

where h is the heat transfer coefficient between the metal and the mould wall, K/L is the thermal conductance of the casting, calculated from K, the thermal conductivity of the liquid metal, and L, the length of the casting in the direction of the heat flow.

When the Biot number is large, heat is transferred quickly from the metal to the mould, but takes longer to reach the mould wall from the centre of the casting, resulting in a significant temperature gradient in the casting, and only a small temperature difference across the interface.

When the Biot number is small, the transfer of heat to the edge of the casting is faster than the transfer of heat from the metal to the mould, resulting in a large temperature difference across the interface, and only a small temperature gradient within the casting. This is illustrated in the movie below:



For the situation where Bi << 1, we can consider the cooling and solidification separately.

Firstly, the liquid metal cools uniformly from its pouring temperature, *T*p, to its melting temperature, *T*m We know that:

$$q = h\Delta T \tag{10.2.2}$$

and

$$q = \frac{dT}{dt} C_V L \tag{10.2.3}$$

so that

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{h\Delta T}{LC_V} \tag{10.2.4}$$

where dT / dt is the cooling rate, and CV, is the heat capacity of the liquid.

The time, tC, taken for the casting to reach Tm is given by:

$$t_C = \frac{LC_V}{h} \ln \left| \frac{T_W - T_M}{T_W - T_P} \right| \tag{10.2.5}$$

where TW is the temperature of the mould wall. To see how this is derived click here.

The solidification stage can be understood by equating the heat transferred across the interface, with the heat generated by the solidification:

$$q = h\Delta T \tag{10.2.6}$$

$$q = v\Delta H_{\mathrm{F}} \tag{10.2.7}$$

so that

$$v = \frac{h\Delta T}{\Delta H_F} \tag{10.2.8}$$

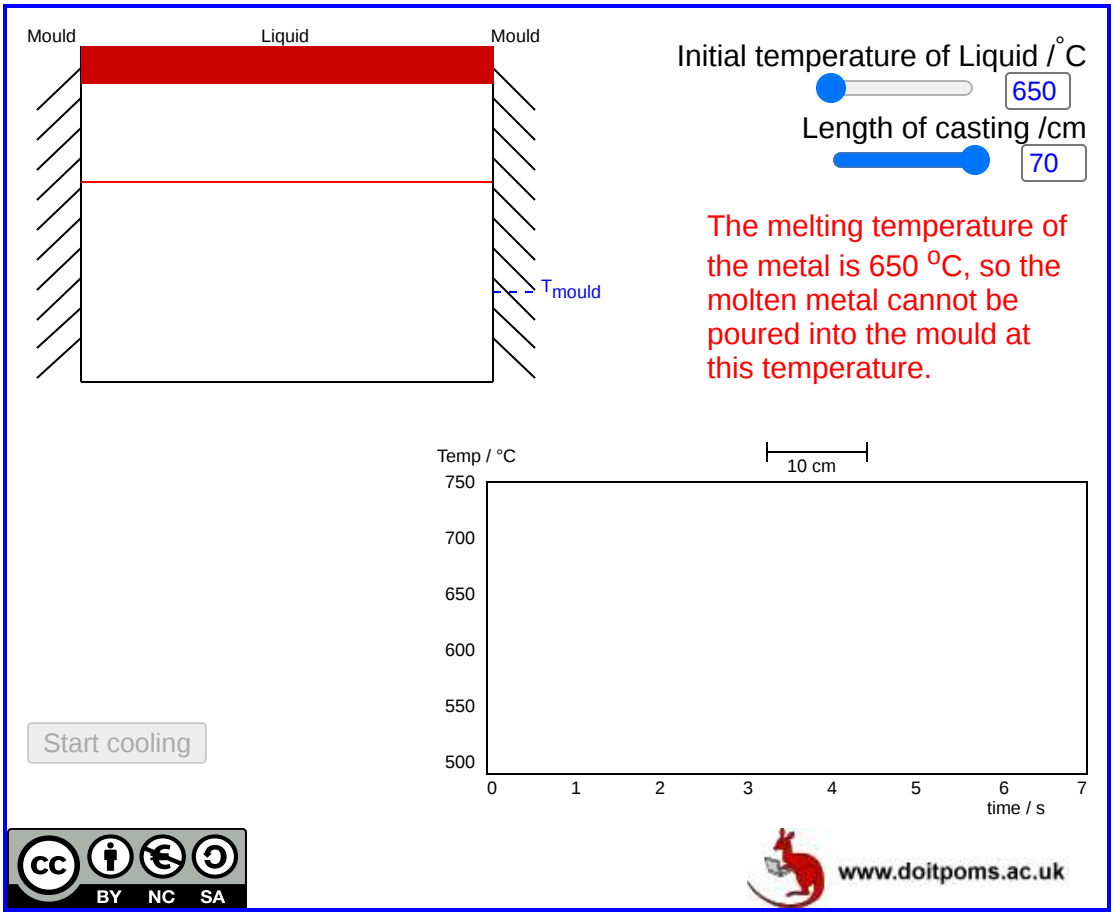
where v is the speed of the solidification front, and ΔHF is the latent heat of fusion of the metal.

The time, tS, taken for the metal to solidify once it has reached Tm is given by:

$$t_S = \frac{\Delta H_F L}{h\left(T_W - T_M\right)} \tag{10.2.9}$$

The total time taken for the casting to solidify once the metal has been poured in is:

$$t = \frac{L}{h}\left(C_V \ln \left| \frac{T_W - T_M}{T_W - T_P} \right| - \frac{\Delta H_F}{(T_W - T_M)}\right) \tag{10.2.10}$$



This page titled 10.2: Heat Transfer is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.3: Heat Transfer Simulation

Here we demonstrate how heat transfer through the mould wall determines the temperature change in a casting during solidification. The simulation assumes ***Newtonian cooling*** where heat transfer is limited by the interface between the metal and the mould.

The simulation shows the effect of varying parameters such as the interfacial heat transfer coefficient, h, the casting length, L and the amount of superheat (determined by the pouring temperature, Tp). It can be carried out for a wide range of metals to study the effect of properties such melting temperature Tm, latent heat of fusion per unit volume, $\Delta H_{f,V}$, thermal conductivity, K and heat capacity per unit volume, $C_{p,V}$. Click here to see how the heat capacity per unit volume is related to specific heats measured relative to other amounts.

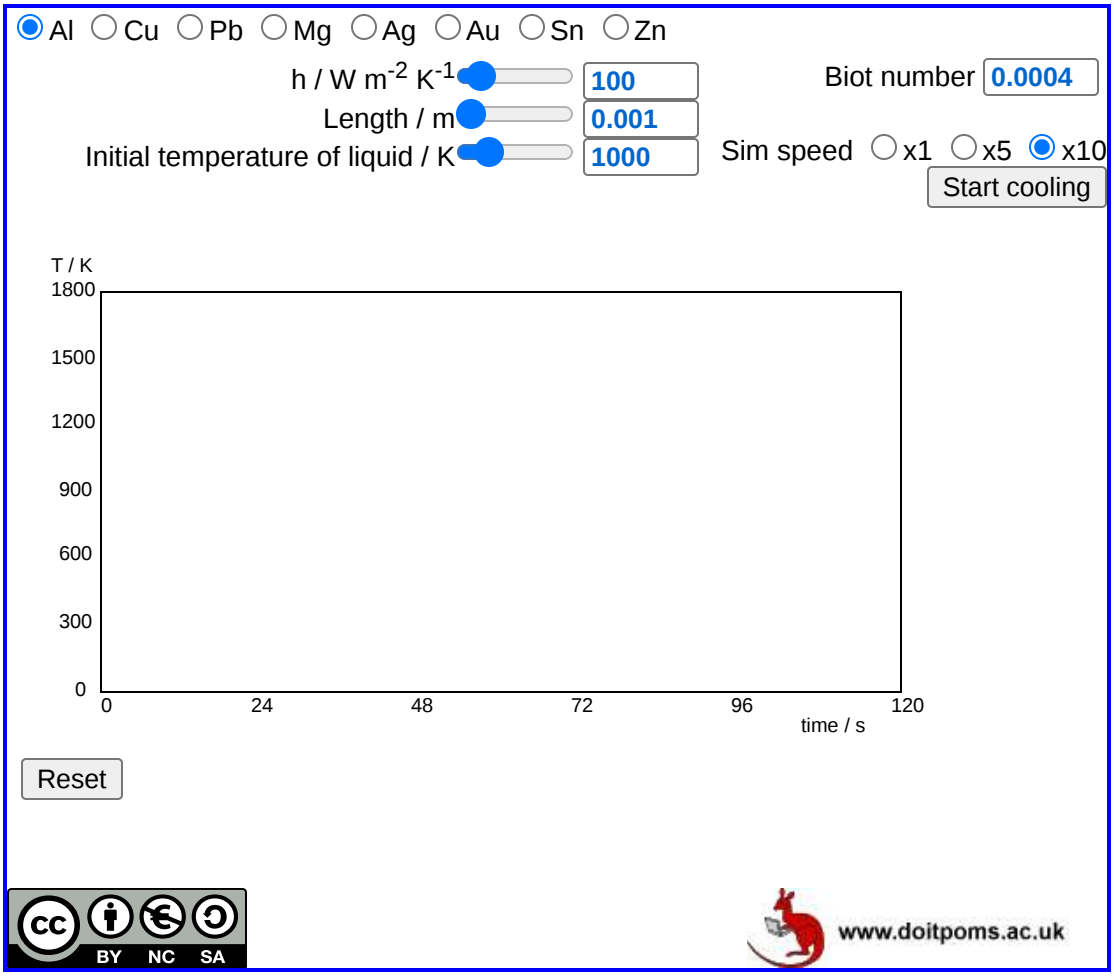
The relevant thermal properties of several pure metals are shown below

	Melting temperature	**Latent Heat of fusion**	**Volumetric Heat capacity**	**Thermal conductivity**
	K	MJ m^{-3}	kJ m^{-3} K^{-1}	W m^{-1} K^{-1}
Ag	1235	1100	2920	422
Au	1337	1200	2800	272
Al	933	1070	3100	240
Cu	1358	1842	4080	395
Mg	922	640	2240	154
Pb	601	260	1590	34
Sn	505	440	1780	63
Zn	693	812	3070	112

In the simulation you can select any of these materials and its properties will be displayed. You may then vary parameters such as the casting length, the interfacial heat transfer coefficient between the solid and the mould wall and the pouring temperature in order to see how long it takes for the casting to solidify and cool.

There are a couple of provisos: Firstly because this is a Newtonian cooling simulation you must ensure that the Biot number is low enough for this assumption to be valid (hence for the metals which have poor thermal conductivities, you must keep the casting length relatively low). Secondly, of course, you must ensure that the metal is poured above its melting temperature!

There are a number of assumptions and simplifications we are making in this simulation which may not be the case for a real casting.

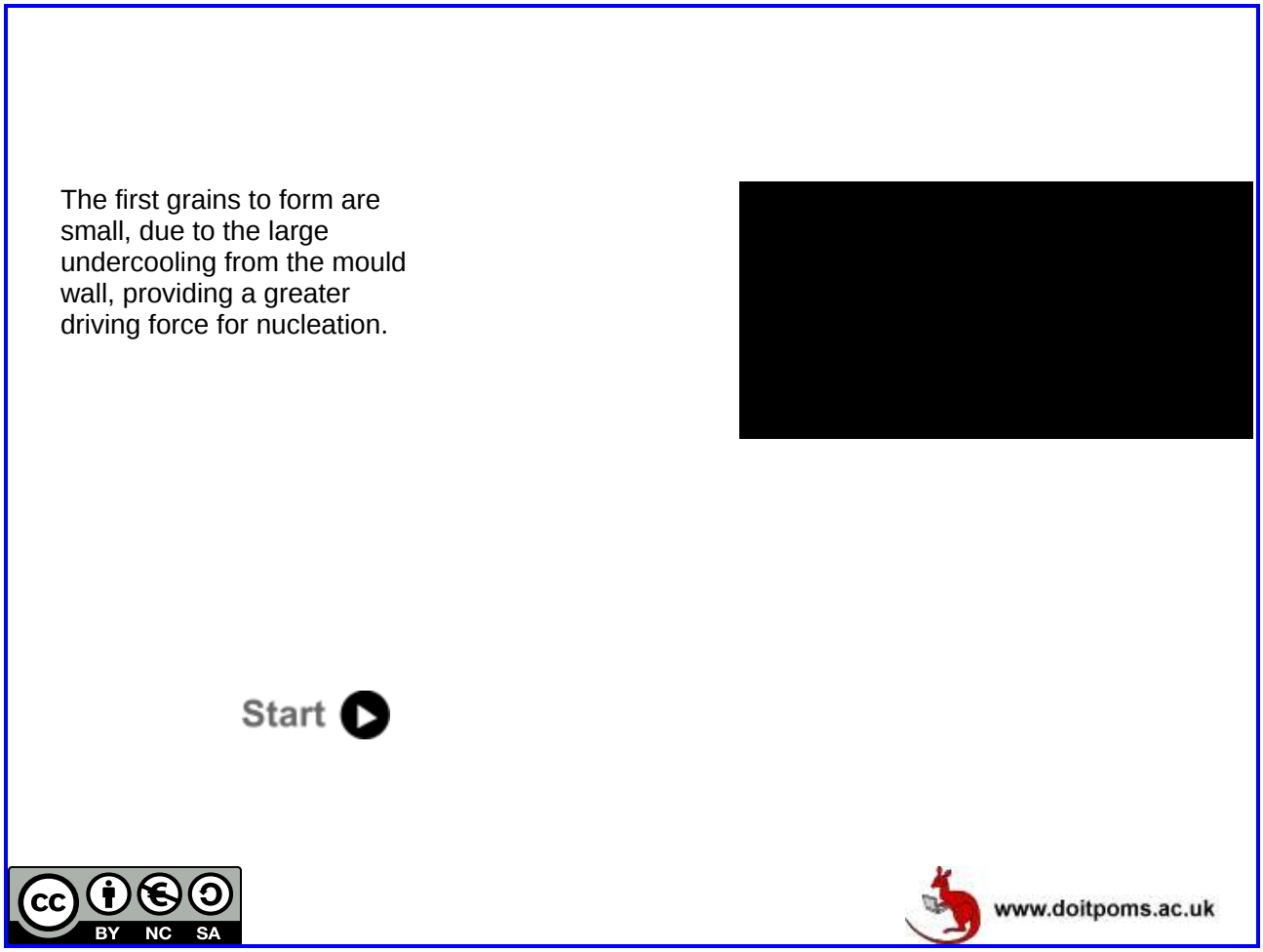


This page titled 10.3: Heat Transfer Simulation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.4: Microstructure and Segregation in Castings

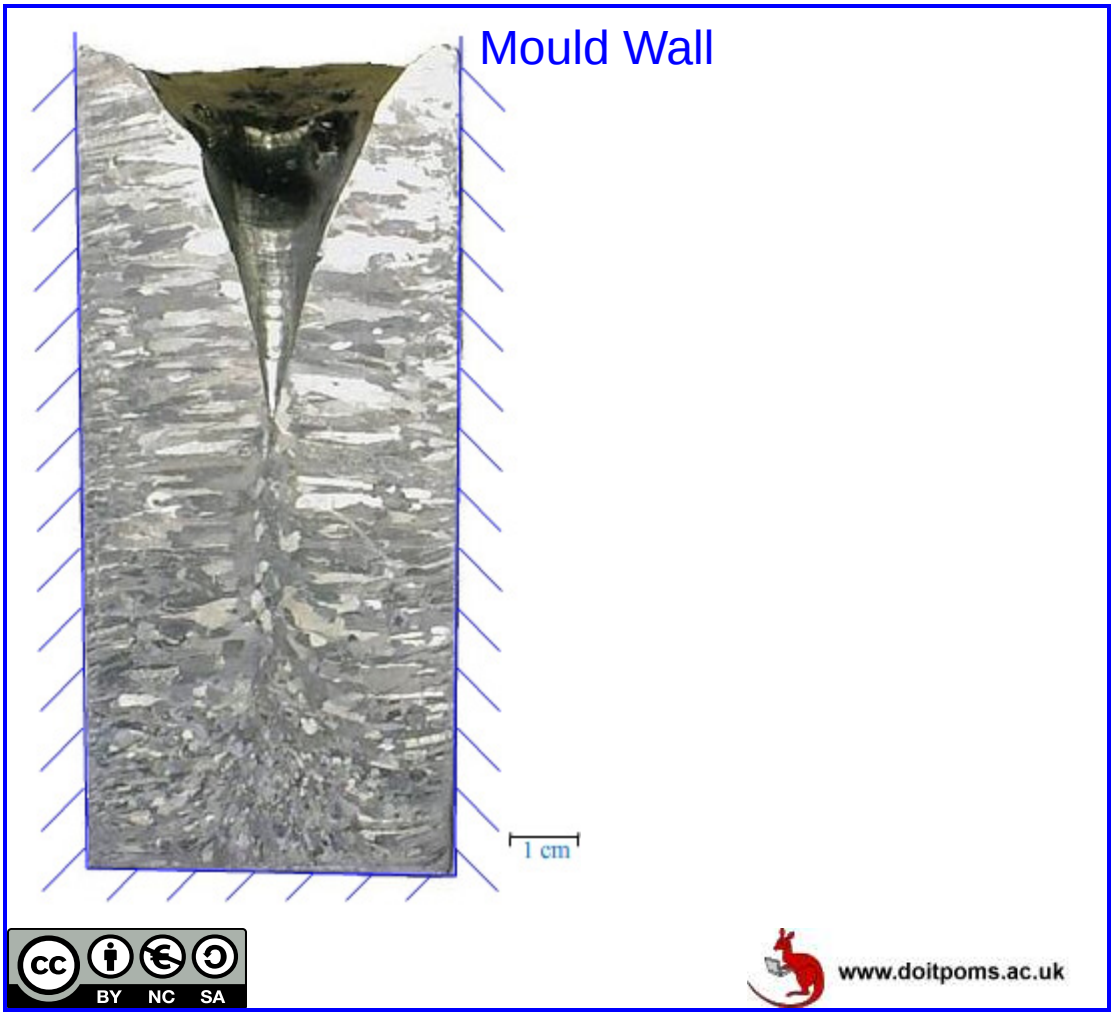
Microstructure of castings

The animation below shows how the microstructure of an ingot develops during solidification.



For a closer look at dendrite formation look at this animation in the Solidification of alloys TLP.

Below is a picture of a real cast Al ingot. Move your mouse over the various parts to see how they are formed.



Segregation in castings

When casting an alloy, segregation occurs, whereby the concentration of solute is not constant throughout the casting. This can be caused by a variety of processes, which can be classified into two types:

Microsegregation; which occurs over distances comparable to the size of the dendrite arm spacing. This occurs as a result of the first solid formed being of a lower concentration than the final equilibrium concentration, resulting in partitioning of the excess solute into the liquid, so that solid formed later has a higher concentration. More about microsegregation can be found in the TLP on Solidification of alloys.

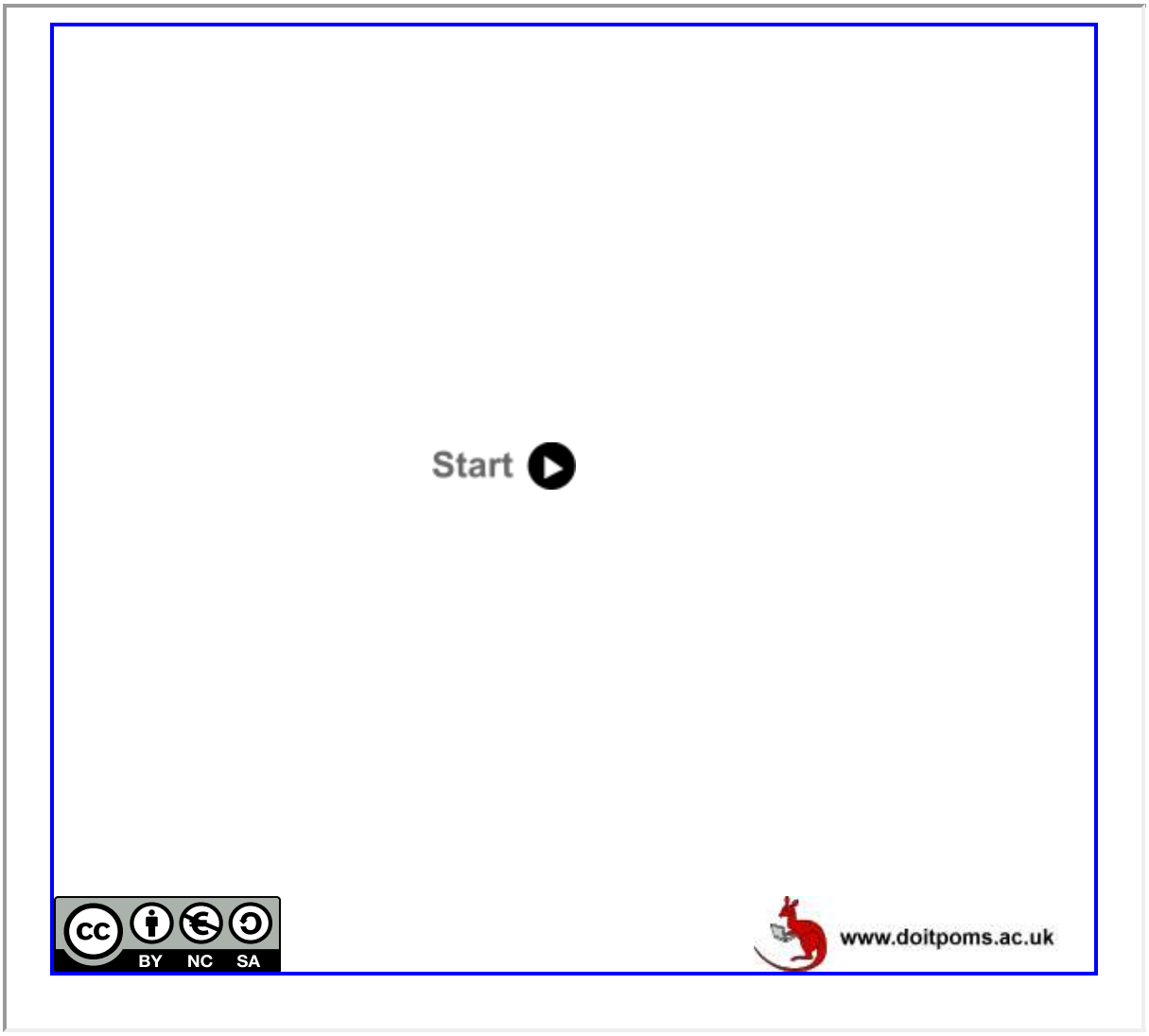
Macrosegregation occurs over similar distances to the size of the casting. This can be caused by a number of complex processes involving shrinkage effects as the casting solidifies, and a variation in the density of the liquid as solute is partitioned. We will not discuss these processes further.

It is desirable to prevent segregation during casting, to give a solid billet that has uniform properties throughout. Microsegregation effects can be removed after casting, by homogenisation, carried out at by annealing at high temperatures where the diffusivity is higher. Macrosegregation effects occur over larger distances so cannot be removed in this way, but can be reduced by control of the casting process and mixing during solidification, often by electromagnetic stirrers. Ultrasound is sometimes used to break up dendrites as they grow, reducing the scale of the dendritic structure and the extent of microsegregation.

This page titled 10.4: Microstructure and Segregation in Castings is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.5: Sand Casting

Sand castings are formed using compacted sand as the mould, more details of the process can be found in the movie below:



Typical components produced by sand casting are automobile engine blocks and ship propellers.

Advantages:

- Low capital investment means that short production runs are viable.
- Use of sand cores allows fairly complex shapes to be cast.
- Large components can be produced.

Disadvantages:

- The process has a high unit cost, as it is labour intensive and time consuming.
- The sand mould leaves a poor surface finish, which often requires further processing.
- Cannot make thin sections.



The drag (left) and cope (right) of a casting flask



A sand casting with risers and the runner system still attached



A crankshaft for an auto engine which has been produced by sand casting

Different types of sand are used in sand casting;

- Petro-bond – This is a mixture of quality sand and oil or synthetic resin.
- Green sand – A mixture of sand, clay, water and sometimes other additives. It is called green sand because it is re-usable. The right amount of water has to be added to prevent porosity.
- Sand mixed with water glass ($NaO.nSiO_2.mH_2O$) can be hardened with CO_2 gas through the chemical reaction;

$$\mathrm{Na_2O \cdot nSiO_2 \cdot (mn + x)H_2O + CO_2 = Na_2CO_3 \cdot xH_2O + n\,(SiO_2 \cdot mH_2O)} \tag{10.5.1}$$

  This transforms the sand into a solid mould, which can be used after the cope & drag are removed.
- A parting powder similar to talcum powder is put on the pattern to make it easier to remove.

More complicated castings can be made by producing a pattern out of polystyrene foam, which is left in place when the molten metal is added. When the metal is poured into the mould the heat vaporizes the foam a short distance away from the metal surface, allowing the metal to fill the mould cavity.

Permanent Mould casting

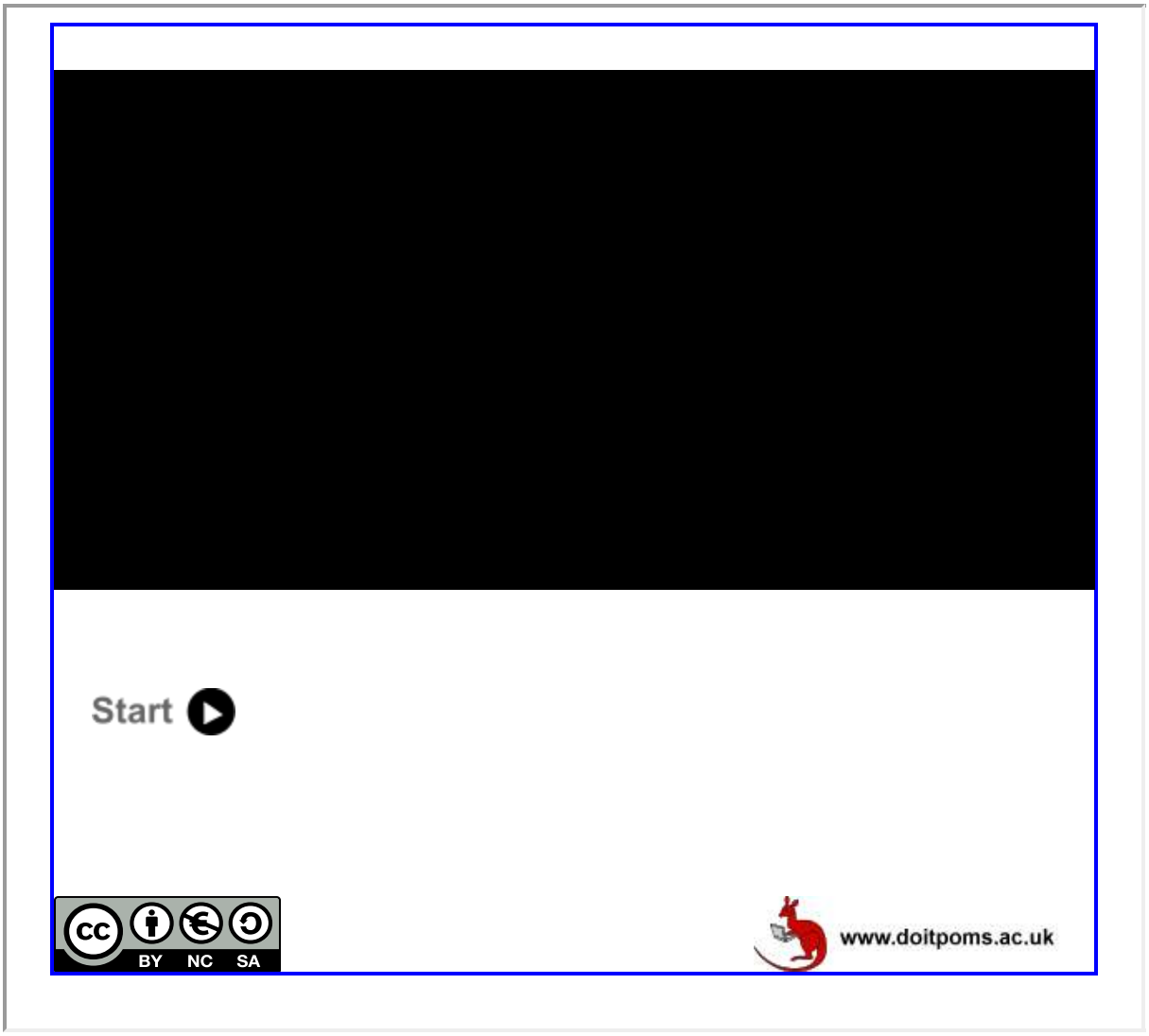
The basic process of permanent mould casting is similar to that for sand casting, except that a permanent metallic mould is used. This mould is made of two parts, containing the necessary gates and risers to ensure the correct flow of metal. The metal moulds

give a better surface finish than sand casting, and selective cooling rates can give control over the scale and morphology of microstructure formed.

This page titled 10.5: Sand Casting is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.6: Die Casting

Die casting is an automated process that forms castings under high pressure in a metal mould. More details of the process can be found in the movie below:



Typical items produced by die casting are toys and small precision parts such as sprockets and gears.

Aluminium, copper and zinc alloys are commonly used in die casting, ferrous alloys are harder to cast in this way as the moulds are made from hardened steel.

There are various methods of die casting;

- Gravity fed – Similar to permanent mould casting, the material is poured into the top of the mould & forced downwards by gravity.
- Pressure – The metal is forced into the mould under pressure.
- Cold chamber – Molten metal is poured into the system and then forced into the mould.
- Hot chamber – The metal is heated within the system & then forced from the crucible into the die.
- Squeeze – Similar to injection moulding, a given amount of material is forced into the mould.

Advantages:

- Very low unit cost.

- High definition & surface finish.
- Excellent dimensional accuracy.
- Cool metal mould gives fast solidification, leading to a fine grain structure.
- Can produce thin sections.

Disadvantages:

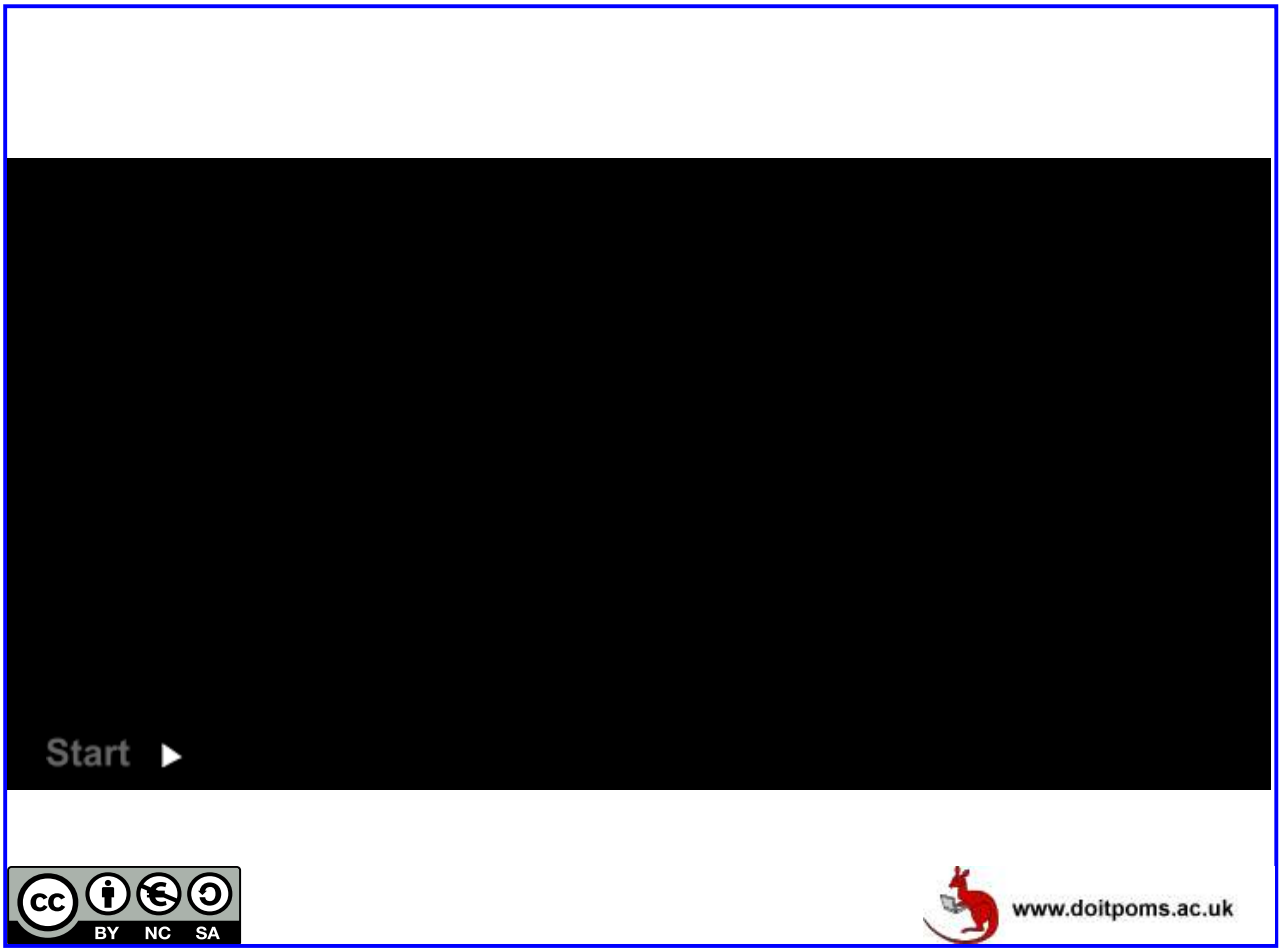
- A large capital investment is required to set up a die casting process.
- It is difficult to control the microstructure of the solid.
- The alloys used must have a low melting point, often at the expense of other properties, such as strength and stiffness.
- Cannot be used for complex shapes, as the casting couldn’t be ejected from the mould.
- Cannot be used for large castings.

This page titled 10.6: Die Casting is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.7: Continuous Casting

Continuous casting is used to produce very large quantities of metals in simple shapes, by casting & rolling the metal continuously in one process. The slabs are then generally further processed by methods such as rolling, extrusion or another casting method to make more useful items.

More about process methods can be found in the TLP on Metal Forming – Introduction to Deformation Processes.



Advantages:

- High yield of casting for a given volume of liquid, mainly thanks to the lack of a contraction pipe.
- Good surface finish.
- Extremely low unit cost due to the very high volume of metal that can be cast.

Disadvantages:

- A large capital investment is required to set up the process.

This page titled 10.7: Continuous Casting is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.8: Investment Casting

Investment casting is used to make precision parts with a good surface finish. It is used to make turbine blades from a single crystal Ni-based superalloys. As the moulds are made from ceramic, metals with high melting points can be cast in this way.

- A pattern of the required component is formed out of wax usually through injection moulding into a metal mould. A runner and gate are included within the wax pattern. Pre formed ceramic cores can also be added, so that hollow castings can be made.
- Many wax patterns are then connected together to form a tree, and this is then dipped into a ceramic slurry, which sets through drying. Different layers of ceramic are created around the wax patterns, increasing in coarseness.
- The wax patterns are then melted out, ready for the molten metal to be poured in to create the castings.
- The ceramic mould is broken to reveal the castings, and any ceramic cores added at the start can be chemically etched out to leave a hollow structure.



Investment wax



Investment mould



Investment blades

https://www.doitpoms.ac.uk/tlplib/ca...entcasting.mp4

Advantages

- Metals with a high melting temperature can be cast due to ceramic mould
- Complex shapes can be formed by using ceramic liners in the original wax patterns
- Good surface finish can be obtained using fine ceramic material

Disadvantages

- Expensive as mould cannot be reused
- Time consuming (drying times for ceramic range roughly 24 hrs)

This page titled 10.8: Investment Casting is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.9: Other Methods of Casting

Melt Spinning

Melt spinning is the rapid cooling of a metal to form a metallic glass ribbon. A cooling rate of 107 K s^{-1} can be achieved by casting onto a water-cooled, rotating Cu wheel. Heat extraction is very rapid, and thin sections are formed (~50 mm), so Newtonian cooling is observed due to the small length scale.

A slowed down video of the process is shown below;

https://www.doitpoms.ac.uk/tlplib/ca...t_multiple.mp4

The fast cooling rate means that amorphous metal (metallic glass) is formed. This has very different properties compared to the alloy with a crystal structure. The metallic glass ribbon is often used in electronic applications, as when a magnetic metal (iron, cobalt, nickel) is used within the casting alloy, the amorphous material forms a magnet with low coercivity , due to the lack of regular structure. This means that the magnetic field can be easily removed.

Centrifugal Casting

Centrifugal casting is used to create hollow, cylindrical parts. The mould rotates, and is only partially filled with molten metal depending on the required thickness. The centrifugal force generated by the mould forces the metal against the wall, giving a good surface finish.

Centrifugal forces can also be used in casting when creating solid objects, the force being used instead of pressure or gravity to push the metal into the mould.

Advantages

- Good external detail
- High control of the microstructure
- No core required to make hollow component

Disadvantages

- Poor internal surface quality
- Inner cross section must be circular

This page titled 10.9: Other Methods of Casting is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.10: Summary

There are 4 main ways of casting metals; sand casting, die casting, continuous casting and investment casting. Each has advantages and disadvantages, and is suited for producing different components or using different materials.
The Biot number predicts how a casting will solidify, and is equal to

$$\frac{h}{K/L} \tag{10.10.1}$$

where: h = heat transfer coefficient
L = length of the casting
K = thermal conductivity

Commonly 3 zones form as the casting solidifies;

- Chill zone – small crystals form on mould wall.
- Columnar zone – larger grains growing in optimum crystallographic orientations.
- Equiaxed zone – small grains that solidified early but detached and moved back into the liquid.

Going further

Books

Manufacturing with Materials by Lyndon Edwards, Mark Endean Butterworth-Heinemann Limited, ISBN 0750627549 (0-7506-2754-9)
Chapter 2 is an excellent coverage of casting

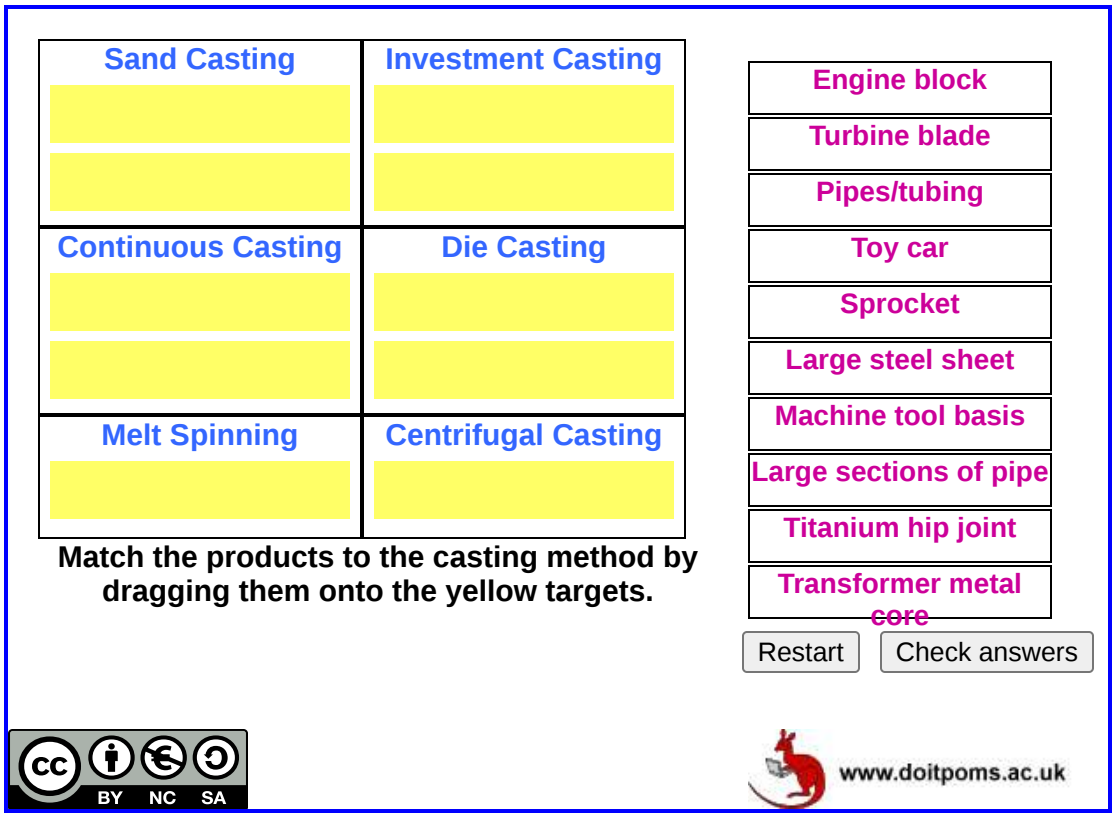
Websites

Continuous Casting
This site contains more information about continuous casting, and a very advanced simulation of the process.

Prof. H.K.D.H. Bhadeshia's teaching pages

This page titled 10.10: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

10.11: Questions



Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

What is the Biot number equal to?

a	hL/K
b	Kh/L
c	K/Lh
d	h/KL

Answer

A

What is h?

a	heat transfer coefficient
b	heat flux
c	heat capacity
d	latent heat
e	thermal conductivity

Answer

A

Which is the first solid to form in a casting?

a	Equiaxed zone
b	Columnar zone
c	Chill zone
d	Pipe

Answer

C

Which casting method uses a wax pattern?

a	Sand casting
b	Die casting
c	Investment casting
d	Continuous casting

Answer

C

Which gas is often used to harden sand casting moulds ?

a	Hydrogen
b	Carbon Monoxide
c	Oxygen
d	Carbon Dioxide

Answer

D

Cold chamber is a type of which casting method?

a	Sand casting
b	Die casting
c	Investment casting
d	Continuous casting

Answer

B

This page titled 10.11: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

11: Coating mechanics

www.doitpoms.ac.uk/tlplib/co...nics/index.php

Learning Objectives

- Appreciate the meaning of a misfit strain between the stress-free (in-plane) dimensions of two layers
- Understand how misfit strains can arise and how they give rise to stresses and strains in the layers, depending on their elastic properties and thicknesses
- See how the general case is simplified when one of the layers (a coating) is much thinner than the other (a substrate) and understand the Stoney equation
- Be able to predict the curvature that tends to arise in such systems
- Be able to predict whether debonding is likely to occur, given the value of the interfacial fracture energy

Before you start

It may be helpful to study the Beam Bending TLP before you start, particularly the page covering Bending Moments and Beam Curvatures.

11.1: Introduction
11.2: Misfit Strains
11.3: Force and Moment Balances
11.4: The Biaxial Modulus
11.5: The Stoney Equation - the Thin Coating Limit
11.6: Spallation (Interfacial Debonding)
11.7: Summary
11.8: Questions

This page titled 11: Coating mechanics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.1: Introduction

This TLP covers some basic mechanics of multi-layered systems. It relates primarily to a 2-layer system, such as a coating on a substrate, although it should be fairly clear how the concepts involved can be extended to multi-layer systems. The treatments presented are for the general case, although the behaviour expected for the special case of one layer being much thinner than the other (eg a thin coating on a massive substrate) is also described.

The TLP is focused on the distributions of (in-plane) stress and strain that can arise within the two constituents under various imposed conditions. These could include applying well-defined external forces, such as an in-plane load or a bending moment. In those cases, the response of the system (in-plane extensions or out-of-plane curvatures) can be obtained from well-known expressions related to the mechanics of composites or of beam bending. However, in layered systems, such as coatings on substrates, the imposed conditions are often more complex than this. For example, heating or cooling of the system will lead to differential thermal expansion or contraction, generating a *misfit strain* - that's to say, a difference between the stress-free (in-plane) dimensions of the two constituents. Misfit strains, which can also arise in other ways (such as plastic deformation of one of the layers), are important in the mechanics of layered systems. They can give rise to relatively complex through-thickness distributions of stress and strain, and also to both in-plane length changes and out-of-plane curvature.

It may also be noted that misfit strains can arise simultaneously in more than one in-plane direction. In fact, it is common for a given misfit strain to be generated in all in-plane directions - this would normally be the case for differential thermal contraction, assuming in-plane isotropy of the thermal expansion coefficients. As a consequence of Poisson effects, this effectively raises the in-plane stiffness (ratio of stress to strain in any given direction), which can be handled by simply using a *biaxial stiffness* in place of the conventional one.

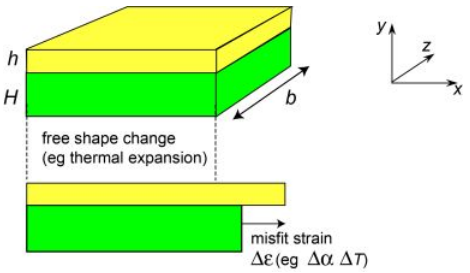
This page titled 11.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.2: Misfit Strains

The Concept of a Misfit Strain

An important concept in layered (and other) systems is that of a misfit strain - ie a difference between the stress-free dimensions of two or more constituents that are bonded together. It is relevant to composite materials and also to macroscopic systems such as two or more components that are bolted or welded together in some way. In general, this strain is a tensor, with principal axes and three principal values. For a layered system, however, the focus is on a single (in-plane) direction, so that the strain can be treated as a scalar.

A simple type of misfit strain is that arising from differential thermal expansion (in a 2-layer system). In general, one layer will expand more during heating than the other (in the direction concerned). If the two layers were not bonded together, then they would behave as shown in the figure below. The misfit strain (in the x-direction) is given by the product of the difference in expansivity between the two constituents and the temperature change. It is often written as Δε. The fact that the two layers are actually bonded together leads to creation of internal stresses and strains, and to changes in the shape of the system - see the next page.



Sources of Misfit Strains

What are effectively misfit strains can arise in a number of ways. One of the simplest is differential thermal contraction, but anything that creates a difference between the (in-plane) stress-free dimensions of a substrate and a coating (or a surface layer of a substrate) has a similar effect. These include phase changes, plastic deformation, creep etc, as well as phenomena, such as atomic bombardment, that can create stresses during formation of a coating.

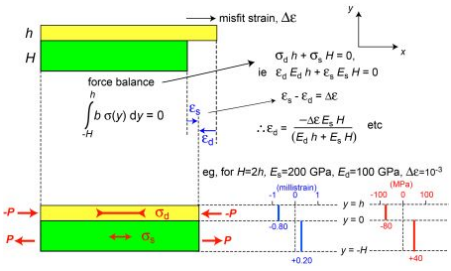
- $\Delta\alpha\Delta T$
- Phase transformations (eg solidification, resin curing, martensitic transformations)
- Plastic deformation (eg shot peening)
- Creep
- (Such deformation can also modify existing Δε values.)

This page titled 11.2: Misfit Strains is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.3: Force and Moment Balances

Force Balance

The fact that the two layers are in fact bonded together must now be taken into account. Assuming first that the layers remain flat, there is clearly a requirement that both must end up with the same length (in the x-direction). This will require a tensile stress to be created in one layer (the one that is initially shorter) and a compressive stress in the other. The forces acting in each layer (in the x-direction) must add up to the externally applied force (ie a *force balance* must apply). Since there is no applied force in the case we are currently considering, the forces in the two layers must sum to zero. In addition, the misfit strain must be partitioned in some way between the two layers. There are thus two simultaneous equations, allowing the solution to be found. This is illustrated in the figure below. (The subscripts d and s represent “deposit” and “substrate” - ie the upper and lower layers.)



Moment Balance

The force balance is relatively straightforward, but the situation depicted in the figure above does not represent complete static equilibrium. This is because the forces being exerted by the stresses acting in the two layers produce a bending moment that tends to create curvature in the x-y plane - for the case shown, the top surface would become convex. The distributions of stress and strain must therefore change, creating an internal balancing moment (and also creating curvature), while maintaining a force balance.

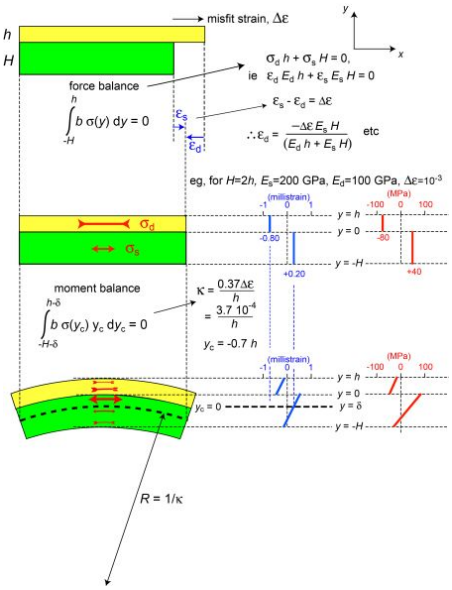
Using the moment balance to find the curvature (and the associated stress and strain distributions) is slightly more complex than applying the force balance. Derivation of the expression for the curvature is shown here. The outcome is:

$$\kappa = \frac{6E_dE_s(h+H)hH\Delta\varepsilon}{E_d^2h^4 + 4E_dE_sh^3H + 6E_dE_sh^2H^2 + 4E_dE_shH^3 + E_s^2H^4} \tag{11.3.1}$$

It’s important to recognize that the curvature of a beam is equal to the *through-thickness gradient* of the associated (in-plane) *strain distribution*. The other key concept here is that of the *neutral axis* of the beam. This is the location - strictly, it’s a plane (in 3-D), rather than an axis - where no (in-plane) strains arise from the bending (adoption of curvature). Its location, for a 2-layer system, is derived here. The result is:

$$\delta = \frac{(h^2E_d - H^2E_s)}{2\,(hE_d + HE_s)} \tag{11.3.2}$$

Using these equations, the final outcome (of imposing a misfit strain) can be obtained. This is illustrated in the figure below. It should be noted that adoption of curvature does NOT lead to zero strain at the neutral axis, but rather to NO CHANGE there - ie it remains at the value resulting from the force balance. (If we simply applied an external bending moment, rather than an internal misfit strain, then there would be no strain at the neutral axis.)



Force and Moment Balances



1. Select materials

Materials:	Thermal Expansivity		Young's Moduli		Poisson Ratio	(In plane) σ/ε ratio	
aluminium	22	$x10^{-6}\ K^{-1}$	69	GPa	0.334	104	GPa
tungsten	4	$x10^{-6}\ K^{-1}$	405	GPa	0.28	563	GPa

2. Select mode

- ○ Uniaxial
- ◉ Biaxial

3. Zoom in

Zoom In

Please look here for further information about the bi-axial stress

CC BY NC SA

www.doitpoms.ac.uk

This page titled 11.3: Force and Moment Balances is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.4: The Biaxial Modulus

Biaxial Stress States

So far, attention has been concentrated on a single (in-plane) direction. For conventional (in-plane) uniaxial loading of a bilayer sample, or for applying a bending moment to it, this is appropriate. It's also possible to generate a misfit strain in a single (in-plane) direction. However, while this is possible, it's actually rather unusual. More commonly, the same misfit strain is generated simultaneously in all in-plane directions, a state that can be represented by creating the strain in two arbitrary (in-plane) directions that are normal to each other. This will lead to an *Equal Biaxial* stress state (since all in-plane directions are clearly equivalent and the through-thickness stress, σ_y, is often taken to be zero - there is no normal stress at a free surface). Differential thermal contraction would normally have this effect. It's also possible to create an *Unequal Biaxial* stress state. This would arise, for example, during differential thermal contraction with one or both of the layers exhibiting in-plane anisotropy in thermal expansivity, so that the misfit strain would be different in different in-plane directions. (Also, anisotropy in stiffness would lead to different stresses in different in-plane directions, even if the misfit strains were equal.) It is, however, common to at least assume that all in-plane directions are equivalent, in terms of both properties and misfit strains.

Poisson Effects

The main reason why the case of an equal biaxial misfit strain differs from that of a uniaxial one is related to Poisson effects. The strains arising in the selected in-plane direction (the x-direction) will be accompanied by Poisson strains in the other two (principal) directions. That in the through-thickness (y) direction is often of little consequence, but in the other in-plane (z) direction, it will need to be added to the outcome of the effects arising in that direction (from the misfit strain in that direction). The upshot of this is actually rather simple. By symmetry, the two in-plane stresses (and strains) must be equal - ie $\sigma_x = \sigma_z$. For isotropic elastic properties and no through-thickness stress ($\sigma_y = 0$), the strain in the x-direction can be written in terms of the three principal stresses:

$$\varepsilon_x E = \sigma_x - \nu\left(\sigma_y + \sigma_z\right) = \sigma_x(1 - v) \tag{11.4.1}$$

where v is the Poisson ratio. The ratio of stress to strain in the x-direction (and all in-plane directions) can therefore be expressed

$$\frac{\sigma_x}{\epsilon_x} = \frac{E}{(1-\nu)} = E' \tag{11.4.2}$$

This modified form of the Young's modulus, E' (often termed the *Biaxial Modulus*), is applicable in expressions referring to substrate/coating systems having an equal biaxial stress state. The effective stiffness (stress/strain ratio) has been raised by this Poisson effect. This higher value should be used in place of E throughout the formulations in the preceding pages (when the misfit strain is generated in all in-plane directions).

This page titled 11.4: The Biaxial Modulus is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.5: The Stoney Equation - the Thin Coating Limit

Origin of the Stoney Equation

The Stoney equation is still in widespread use. It relates the curvature of a substrate with a thin coating to the stress level within the coating (for an equal biaxial case). It was proposed in 1909 - a long time before the relationship described in earlier pages (for the general case in which the coating thickness is not negligible compared to that of the substrate) was established. However, it is easy to show that the Stoney equation can be derived from that relationship, by imposing the h << H condition, which allows all of the denominator terms except the last one to be discarded and (h + H) = H to be assumed, so that

$$\kappa = \frac{6E'_{\mathrm{d}}h\Delta\epsilon}{E'_{\mathrm{s}}H^2} \tag{11.5.1}$$

with the biaxial moduli now being used. Furthermore, the $h \ll H$ condition allows the assumption to be made that all of the misfit strain is accommodated in the coating (deposit), so only the coating is under stress. In addition, the Stoney equation is based on an equal biaxial stress state, so that the biaxial versions of these Young's moduli should be used and the misfit strain can be expressed as

$$\Delta\epsilon = \frac{-\sigma_d}{E'_{\mathrm{d}}} \tag{11.5.2}$$

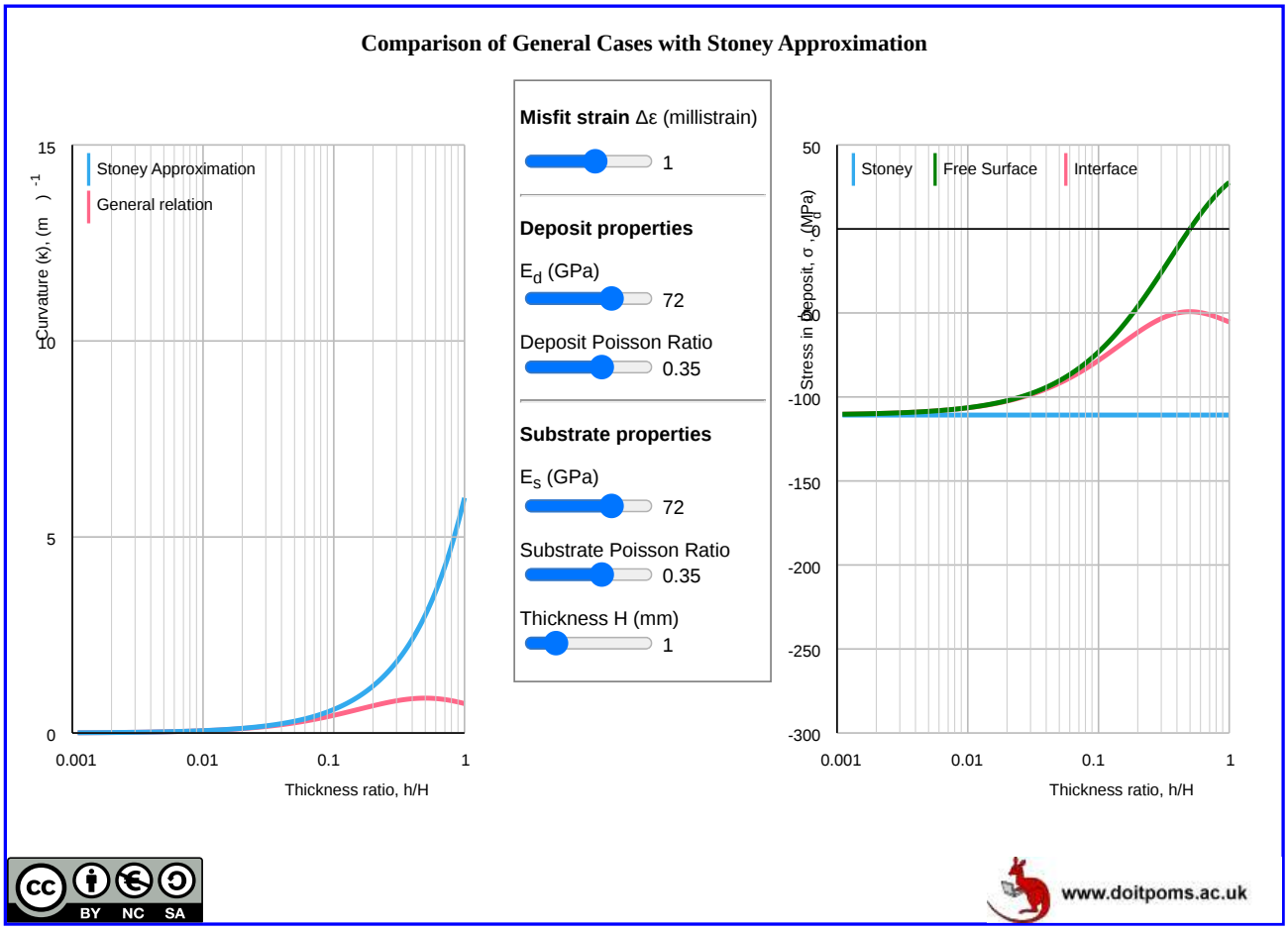
recognizing that, with the convention we are using for $\Delta\epsilon$, a positive value will generate a negative value for σ_d (ie a compressive stress). Substitution of this then leads to the Stoney equation:

$$\kappa = \frac{-6E'_{\mathrm{d}}h\sigma_{\mathrm{d}}}{E'_{\mathrm{s}}H^2E'_{\mathrm{d}}} = \frac{-6h\sigma_{\mathrm{d}}}{E'_{\mathrm{s}}H^2} \quad \therefore \sigma_{\mathrm{d}} = \frac{-E'_{\mathrm{s}}H^2}{6h\left(1-\nu_{\mathrm{s}}\right)}\kappa \tag{11.5.3}$$

(The minus sign is not always included, but the curvature should have a sign and, using the convention that a convex upper surface corresponds to positive curvature, this implies a negative deposit stress.) This equation allows a coating stress to be obtained from a (measured) curvature. Only and values for the substrate are needed - this is convenient, since they are often known (whereas those of the coating may not be). A single stress value is obtained - if the coating is thin, then any through-thickness variation in its value is likely to be small. It is really the misfit strain that is the more fundamental measure of the characteristics of the system, but a stress value is often regarded as more easily interpreted.

Approach to Stoney Conditions

The Stoney equation is easy to use and, indeed, is widely used. However, it does have the limitation of being accurate only in a regime in which the curvatures tend to be relatively small. In some applications – such as with semiconductor wafers – surfaces are very smooth, so that highly accurate optical methods of curvature measurement are feasible and this is not such a problem. However, when curvatures are high (or need to be high for reliable measurement), the Stoney equation should not be used. The simulation below allows exploration of the conditions under which the Stoney equation gives a good or poor approximation to the actual behaviour. It should be noted that, while the stresses are scale-independent, the curvatures are not - changing the substrate thickness thus affects the plots on the left, but not those on the right.

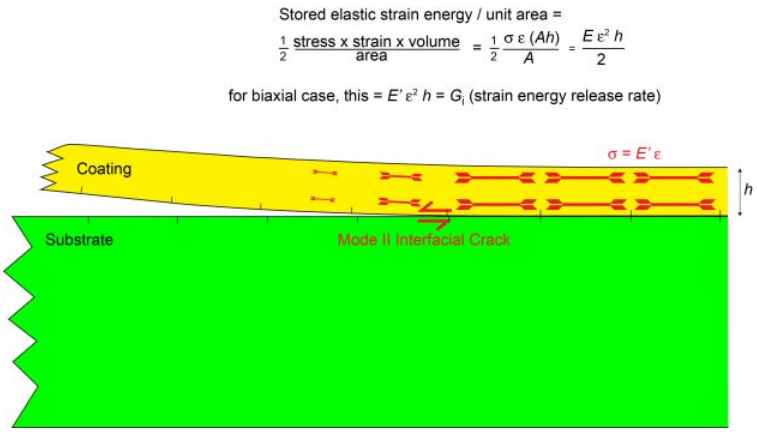


This page titled 11.5: The Stoney Equation - the Thin Coating Limit is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.6: Spallation (Interfacial Debonding)

Driving Force for Spallation

The presence of residual stresses in a substrate/coating system constitutes a driving force for debonding (spallation), since such stresses will almost certainly be at least partially relaxed when this occurs, releasing stored elastic strain energy. The key process is that of propagation of a crack along the interface, driven by the associated release of this stored energy. This propagation is illustrated below, for a (Stoney) case in which there is just a single (uniform) stress in the coating.



It can be seen that propagation of this interfacial crack will be energetically favoured if the driving force (strain energy release rate) is equal to or greater than the (mode II - ie shearing mode) fracture energy of the interface, G_{ic}:

$$E'_d \epsilon_d^2 h \left(= \frac{\sigma_d^2 h}{E'_d} \right) \geq G_{\text{ic}} \tag{11.6.1}$$

This takes no account of any barrier to initiation of the crack. In many cases, however, there are likely to be relatively large defects present in the interface, so the above condition may well lead to spallation. It certainly means that the coating is (thermodynamically) unstable. Immediate implications are (unsurprisingly) that high stresses and brittle interfaces (low G_{ic}) make debonding more likely. Also clear (and widely observed) is that thicker coatings are more likely to debond than thinner ones.

Debonding for non-Stoney Cases

The same concept can be applied to more general cases - ie when the h<<H condition does not apply. Debonding will still tend to allow a reduction in the stored elastic strain energy, constituting a driving force for spallation. However, since there is now a distribution of stress (and strain) in the through-thickness (y) direction, an integration is needed to evaluate the driving force

$$G_i = \int_{-H}^{h} \frac{\sigma(y)^2}{E'(y)} \mathrm{d}y \tag{11.6.2}$$

This is based on the assumption that these stresses become totally relaxed during debonding. This might be the case - for example, the sets of stresses and strains that have been predicted to arise from imposition of a uniform misfit strain would completely disappear if the coating could debond. In practice, however, the stress distribution within the system may be more complex than this. For example, it's common for stresses to be created in a coating during its formation. As such a coating gets thicker, balancing of forces and moments takes place progressively, so that conditions change and the final stress distribution is not one corresponding to imposition of a uniform misfit strain. In such cases, debonding may leave some residual stresses (and residual curvature of the coating and possibly of the substrate). The net driving force may then be written

$$G_i = \int_{-H}^{h} \frac{\sigma(y)^2}{E'(y)} \mathrm{d}y - \int_{-H}^{h} \frac{\sigma_r(y)^2}{E'(y)} \mathrm{d}y \tag{11.6.3}$$

where $\sigma_r(y)$ is the residual stress distribution after debonding.

Debonding of coatings is, of course, commonly observed, since virtually all coatings contain at least some stresses and the associated strain energy can rise above the critical level - for example, as a result of temperature changes, thickening (eg oxide growth), stiffening (eg due to sintering), applied forces or bending moments etc. Also, the toughness (fracture energy) of the interface (G_{ic}) may fall with time - eg due to chemical attack etc. The video clips below give some examples of how coating spallation can be observed and analysed.

https://www.doitpoms.ac.uk/tlplib/co...os/debond1.mp4

Spontaneous Debonding. This video shows a set of coated samples (zirconia on alumina substrates) being withdrawn from a furnace and cooled by gas jets. The cooling creates differential thermal contraction stresses and, once these have become sufficiently large, coatings can spontaneously debond. Such an event can be seen at the end of this short video.

https://www.doitpoms.ac.uk/tlplib/co...eos/kappa1.mp4

In Situ Curvature Measurement. This video, which has a commentary, explains how, provided the substrate is relatively thin, the deposition of a coating (in this case by thermal spraying of ceramics onto metal substrates) generates curvature, which can be monitored as the coating thickness increases. In conjunction with a numerical model, this can be used to infer the residual stress levels in coatings produced under a range of conditions.

https://www.doitpoms.ac.uk/tlplib/co...eos/kappa2.mp4

Debonding during Cooling of Thin Substrate with Coatings. This video, which has a commentary, explains how stored residual stresses provide a driving force for debonding (spallation). It is also explained how estimates can be made of the interfacial toughness (fracture energy) from observations of debonding during cooling.

https://www.doitpoms.ac.uk/tlplib/co...eos/kappa3.mp4

Debonding under Applied Load of Coatings with Residual Stresses. This video, which has a commentary, goes briefly through the 4-point bend delamination test for coatings that already contain stored residual stresses. It is shown how this can be used to measure the interfacial fracture energy.

This page titled 11.6: Spallation (Interfacial Debonding) is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.7: Summary

A coating on a planar substrate is an important special case of a composite system. Attention is concentrated here on the mechanical (stress-related) effects that commonly arise, with the possibility of stresses (in-plane only) and strains within both constituents being generated in a variety of ways. In the treatment presented here, edge effects are ignored, which simplifies matters, but considerable attention is paid to the way in which the system may become curved, and to the relationship between the curvature and the internal stresses and strains.

Two main cases are considered, depending on whether it is possible for the assumption to be made that the coating is very much thinner than the substrate. If this is the case, then certain assumptions can be made and some relatively simple equations can be used to describe the behaviour. On the other hand, it is emphasized that there are many practical cases for which this condition cannot be assumed, although (slightly more complex) analytical treatments can then be employed and these are described here in some detail.

It is also worth noting that a further distinction can be drawn, depending on whether the stress state within the coating (and substrate) comprises a uniaxial (in-plane) stress or an equal biaxial set of (in-plane) stresses, with the latter being much more common. In this case, it is explained that the ratio of stress to strain in any given (in-plane) direction is given by the biaxial modulus, rather than the Young's modulus (for a uniaxial stress state).

Finally, a brief outline is presented of how the stored elastic strain energy associated with the presence of stresses in a coating (and substrate) constitutes a driving force for spallation (interfacial crack propagation). A simple criterion is presented for advance of such a crack and some examples are given of how this can be utilized in some practical cases.

Going further

Books

There are not really any books that specifically cover all of the material in this TLP, although there are, of course, books that treat various aspects of coatings, and of surface engineering more generally. The three below respectively present a review of developments since the original Stoney equation, modelling of progressive deposition of a coating onto a thin substrate and the anisotropy that arises with thin films on (cubic) single crystal substrates.

MR Begley & JW Hutchinson, *The Mechanics and Reliability of Films, Multi-layers and Coatings*, CUP (2017) ISBN: 9781107131866.

J Mencik, *Mechanics of Components with Treated or Coated Surfaces*, Springer (2010) ISBN-13: 978-9048146116.

L. B. Freund & S. Suresh, *Thin Film Materials: Stress, Defect Formation and Surface Evolution*, CUP (2004) ISBN-1139449826

Other resources

There are many journal papers that cover the issue of stresses and strains in surface coatings and also the broader topics of how they can provide various kinds of protection, including thermal, environmental and tribological. The two below respectively present a review of developments since the original Stoney equation and an outline of how progressive deposition of a coating onto a thin substrate can be treated.

GCAM Janssen, MM Abdalla, F van Keulen, BR Pujada & B van Venrooy, *Celebrating the 100th Anniversary of the Stoney Equation for Film Stress: Developments from Polycrystalline Steel Strips to Single Crystal Silicon Wafers*, Thin Solid Films, **517** (2009) p.1858-1867, https://doi.org/10.1016/j.tsf.2008.07.014

YC Tsui & TW Clyne, *An Analytical Model for Predicting Residual Stresses in Progressively Deposited Coatings .1. Planar Geometry*, Thin Solid Films, **306** (1997) p.23-33, https://doi.org/10.1016/S0040-6090(97)00199-5

KM Knowles, *The Biaxial Moduli of Cubic Materials Subjected to an Equi-biaxial Elastic Strain*, J. of Elasticity, 124 (2016) p.1-25, https://doi.org/10.1007/s10659-015-9558-x

This page titled 11.7: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

11.8: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

What is meant by a "misfit strain"?

a	A strain corresponding to the difference in stress-free dimensions of two constituents that are bonded together in some way.
b	A strain that needs to be imposed on a coating in order to make it stress-free.
c	A strain resulting from imposition of the condition that coating and substrate remain bonded together.
d	A strain that arises between coating and substrate as a result of temperature change

Answer

A

Why does curvature tend to arise in a coating-substrate system, as a result of the stresses in each constituent caused by imposing a force balance?

a	The curvature arises because it allows the stress in the coating to be completely relaxed.
b	The force balance cannot be satisfied unless it is accompanied by curvature.
c	There is a lateral separation between the axes along which the corresponding forces act, so a bending moment is generated, creating a curvature.
d	The forces that arise act along axes that are not parallel, so a bending moment is generated, creating a curvature.

Answer

C

Which of these conditions is sufficient to ensure that the Stoney equation is a good approximation for the relationship between coating stress and curvature?

a	The stress state is an equal biaxial one.
b	The coating is thinner than the substrate.
c	The stiffness of the coating is lower than that of the coating
d	The magnitude of the average stress in the coating is much greater than that in the substrate.

Answer

D. The other three conditions will all tend to favour conditions in which the Stoney equation is likely to be reliable, but they are not sufficient individually for this to be the case.

What is meant by the "Biaxial Modulus" of a coating (or a substrate) and why is it greater than the conventional (Young's modulus)?

a	It corresponds to the stress-strain ratio in any in-plane direction, for an equal biaxial stress state, and it is larger than the Young's modulus because of Poisson effects.
b	It corresponds to the stress-strain ratio (in an elastic system) in any in-plane direction, for an equal biaxial stress state, and it is larger than the Young's modulus because of Poisson effects.
c	It corresponds to the stress-strain ratio (in an elastic system) in any in-plane direction and it is larger than the Young's modulus because of Poisson effects.
d	It corresponds to the stress-strain ratio (in an elastic system) in any in-plane direction, for an equal biaxial stress state, and it is larger than the Young's modulus because of the Poisson contraction arising from a stress in another in-plane direction.

Answer

B

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

How is the neutral axis (strictly, the neutral plane) of a beam defined?

a	It is the line along which, when an initially unstressed beam is subjected to a bending moment, no strains arise while the beam becomes curved.
b	It is the line along which there are no shear strains in the beam.
c	It is the line along which there is no strain when the beam becomes curved.
d	It is the line along which there is no stress when the beam becomes curved.

Answer

A. It is important to note that, while the adoption of curvature will not change the (in-plane) strain at the neutral axis, the strain will only be zero at that location if that was the case before the curvature was created.

Which of these definitions of the curvature of a beam is correct?

a	Beam stiffness times applied moment.
b	Through-thickness gradient of the in-plane stress (in the plane of curvature) induced by the adoption of curvature.
c	Beam thickness divided by radius of curvature.
d	Through-thickness gradient of the in-plane strain (in the plane of curvature) induced by the adoption of curvature.

Answer

D. It is often instructive to think of the curvature as a through-thickness strain gradient (rather than in terms of the location of the centre of curvature).

Which of these statements regarding debonding (spallation) of coatings is incorrect?

a	If the residual stress in a coating is compressive, rather than tensile, then it is less likely to debond from the substrate.
b	Coatings are more likely to debond if the coating / substrate interface is brittle (has a low toughness).
c	The main driving force for debonding of many coatings is the stored elastic strain energy associated with the residual stresses in them.
d	Thick coatings tend to debond more readily than thin ones.

Answer

A. The magnitude of the driving force for debonding is independent of the sign of the stress - it appears in a squared term. A compressive stress may be preferable in terms of conferring resistance to certain types of damage, such as through-thickness cracking, but only the magnitude of the stress is relevant to the driving force for spallation.

This page titled 11.8: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

12: Creep Deformation of Metals

https://www.doitpoms.ac.uk/tlplib/creep/index.php

Learning Objectives

- Have an appreciation of the mechanisms involved in creep deformation, and their general dependence on temperature and stress
- Understand what is meant by Primary, Secondary and Tertiary Creep
- Know how creep curves can be represented by (empirical) constitutive laws, and how the values of parameters in them, such as stress exponents and activation energies, can be obtained from experimental data
- Know how a conventional uniaxial creep test is carried out
- Be familiar with a particular set-up for experimental study of the creep characteristics of a material, available in wire form < li=""> /li>
- Understand the basics of Indentation Creep Plastometry

Before you start

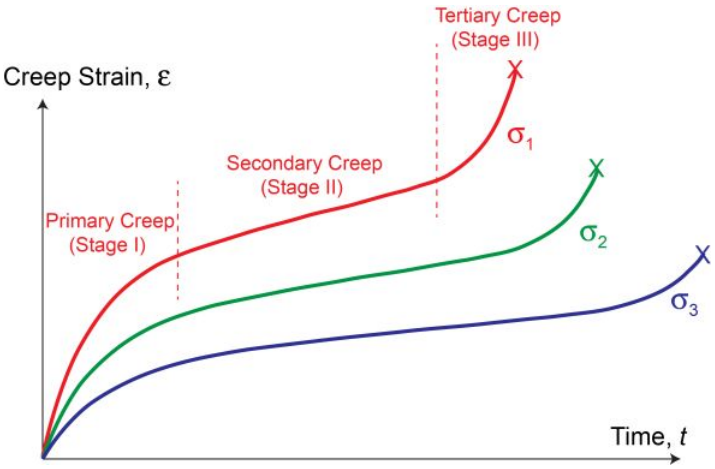
The following TLPs are relevant and could be consulted before you start:
Mechanical Testing of Metals
Mechanisms of Plasticity
Stress Analysis and Mohr's Circle
Finite Element Method

12.1: Introduction
12.2: Creep Mechanisms
12.3: Constitutive Laws for Creep
12.4: Uniaxial Creep Testing - Practical Basics
12.5: Multiaxial Creep Testing – The Creeping Coil
12.6: Multiaxial Creep Testing – Indentation Creep Plastometry
12.7: Designing for Creep Resistance - Nickel Based Superalloys
12.8: Summary
12.9: Questions

This page titled 12: Creep Deformation of Metals is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.1: Introduction

When a material is subjected to a stress that reaches the yield stress, it deforms ***plastically*** (permanently). Provided the stress is kept below this level, then in principle it should only deform elastically. However, if the **homologous temperature** is relatively high (above ~0.4) then in practice plastic deformation can occur, even if the applied stress is lower than the yield stress. This deformation is usually progressive with time and is commonly known as ***creep***. During loading under a constant stress, the strain tends to vary with time approximately as shown below, where the effect of changing the applied stress is also indicated. The graph below is a plot of *creep strain*. The elastic strain has been omitted. In practice, it is quite common to do this (for plasticity, as well as for creep). It's worth noting that elastic strains rarely exceed a small fraction of a %, at least for metals, whereas both plastic strains and creep strains commonly reach the range of several tens of %.



Schematic creep curves, for 3 different levels of applied stress ($\sigma_1 > \sigma_2 > \sigma_3$)

The terms "***Primary***", "***Secondary***" and "***Tertiary***" creep are widely used. At a simple level, they are often associated respectively with the concepts of: (i) setting up some kind of mechanistic balance, (ii) steady state (constant strain rate) deformation occurring once this balance has been set up and (iii) the breakdown of this balance, often with defects starting to appear and failure rapidly following.

In reality, even without going into details of the mechanisms involved, this picture is simplistic and potentially misleading. For example, the transition between primary and secondary regimes is often poorly-defined and indeed a true steady state may never be set up. Furthermore, creep tests are commonly carried out with a constant applied load (rather than a **constant true stress**). Thus, for tensile tests, the tertiary regime may actually be a consequence of the fact that the true stress is rising throughout the test, with the rate of increase becoming greater towards the end of the test. At least in some cases, the tertiary regime may be associated with the true stress starting to approach or exceed the yield stress of the material. The situation can be further complicated by the possibility that significant microstructural changes (such as recrystallization), which could strongly affect the mechanical response, may occur during the test.

These issues are examined in this TLP, together with some details concerning creep mechanisms and the ways in which creep testing can be carried out, and the resultant experimental data interpreted.

This page titled 12.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.2: Creep Mechanisms

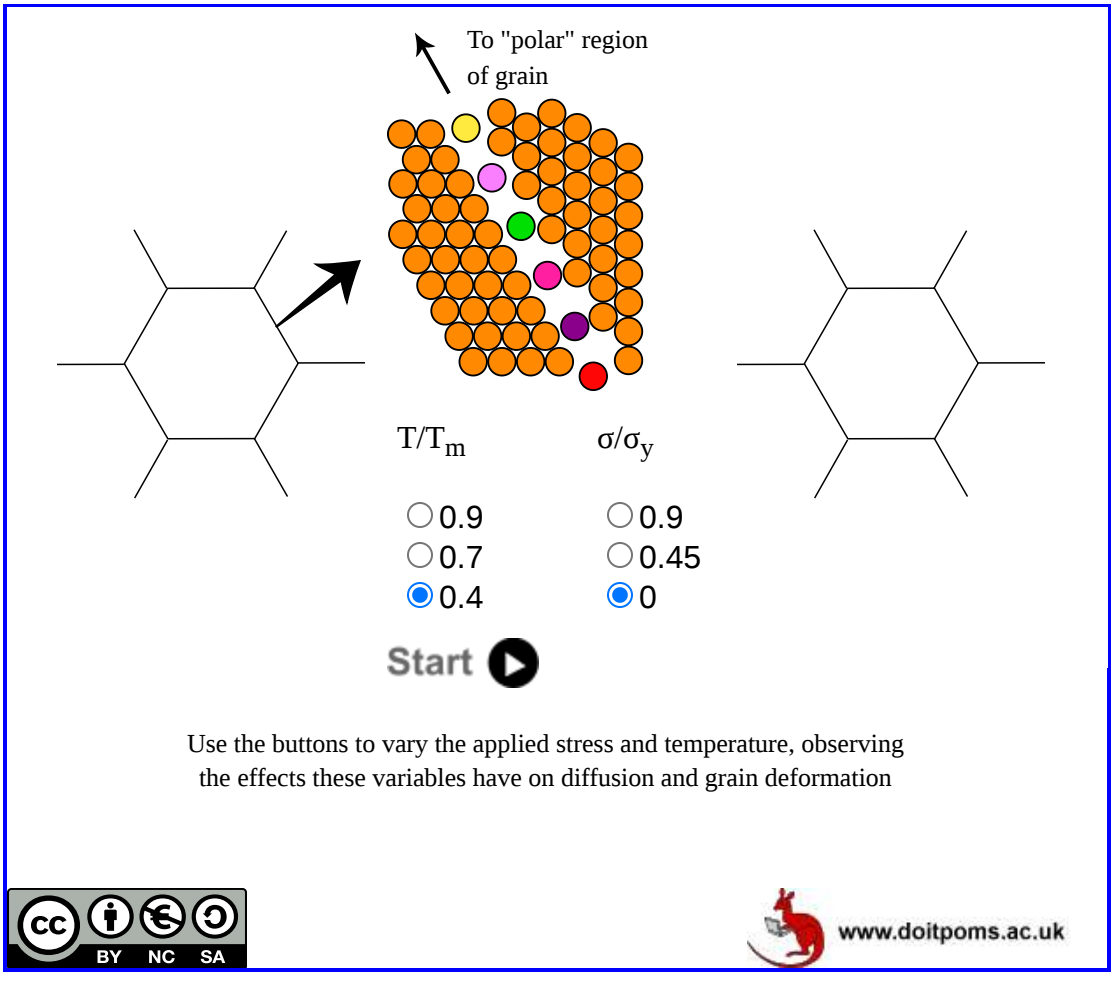
The detailed mechanisms responsible for creep tend to be complex. However, they almost always involve diffusion of some sort. This is how the time-dependence arises, since diffusional processes are progressive with time. Further information about the fundamentals of diffusion is available in the **Diffusion TLP**.

Creep deformation is caused by the deviatoric (shape changing) component of the stress state applied to a sample – the ***von Mises stress***. The hydrostatic component has no effect, meaning that during creep deformation occurs at ***constant volume***. However, whilst the overall deformation is in response to the deviatoric stress, local variations in in hydrostatic stress do affect creep behaviour, as will be outlined below.

Coble Creep

Creep deformation often involves various defects, particularly ***dislocation cores*** or ***grain boundaries***. These may simply act as ***fast diffusion paths***, or play a larger role in creep mechanisms (some of which are beyond the scope of this TLP), depending on factors such as dislocation density, grain size, grain shape and temperature. The shape change experienced by the sample may arise simply from atoms becoming redistributed by diffusion. When this occurs on the scale of a grain, with the diffusion occurring mainly via grain boundaries, then this is commonly referred to as ***Coble Creep***. The simulation below shows how this tends to cause the sample to extend under an applied load.

Note: Coloured atoms are no different from those in the bulk, but merely allow each atom to be identified easily.



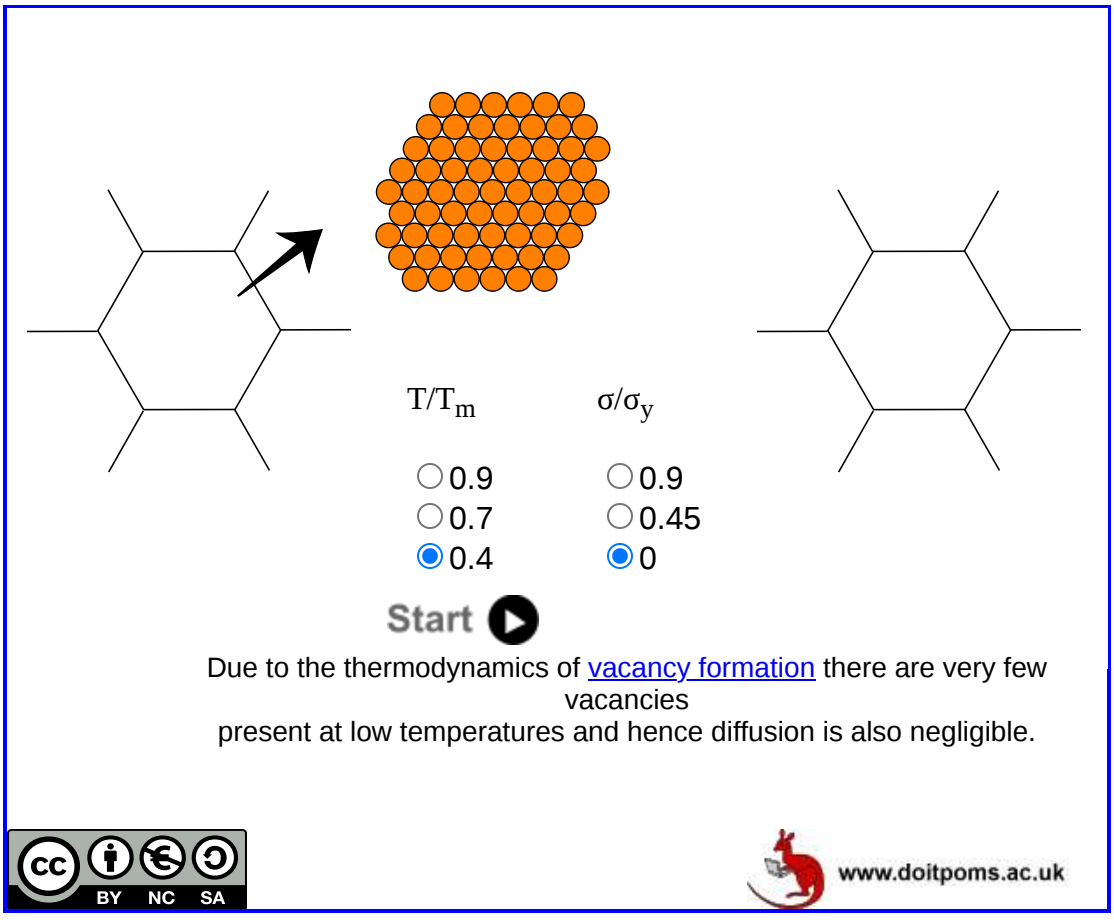
Simulation demonstrating Coble creep on the atomic scale

It can be seen that raising the applied stress accelerates the rate of deformation. The driving force for this net migration of material (from the "***equatorial***" regions of grains to the "***polar***" regions) is that an applied tensile stress like this creates hydrostatic compression in the equatorial regions and hydrostatic tension in the polar regions. The hydrostatic tension can be thought of as arising from the applied tensile stress, where the compression arises from the lateral contraction of the sample (due to volume conservation). The atoms then tend to move from the more "crowded" to the more "open" regions. The diffusive flux can be considered as a migration of vacancies in the opposite direction to the motion of atoms, although the concept of vacant sites is less well-defined in a grain boundary than in the lattice. We could equally imagine how hydrostatic compression could arise in the ***"polar"*** regions, and hydrostatic tension could arise in the ***"equatorial"*** regions, via the application of a compressive stress to the sample. Hence diffusive flux would be in the opposite direction, and the sample would deform in the opposite fashion, with the grains becoming "squashed" by the compressive stress.

It's also clear that raising the temperature increases the creep rate. This is simply due to the rates of diffusion becoming higher as a consequence of the Arrhenius dependence (click **here** for details). The activation energy for grain boundary diffusion is low, and the cross-sectional area available for diffusion along grain boundaries is much less than for diffusion through the bulk. Therefore this type of creep is often the dominant one at ***relatively low temperatures*** and for samples with a ***fine grain size***.

Nabarro-Herring Creep

A similar type of creep deformation to that described above can occur with the diffusion being predominantly within the interior (crystal lattice) of the grains, rather than in the grain boundaries. This is often termed ***Nabarro-Herring creep*** (N-H creep). It is depicted in the simulation below.

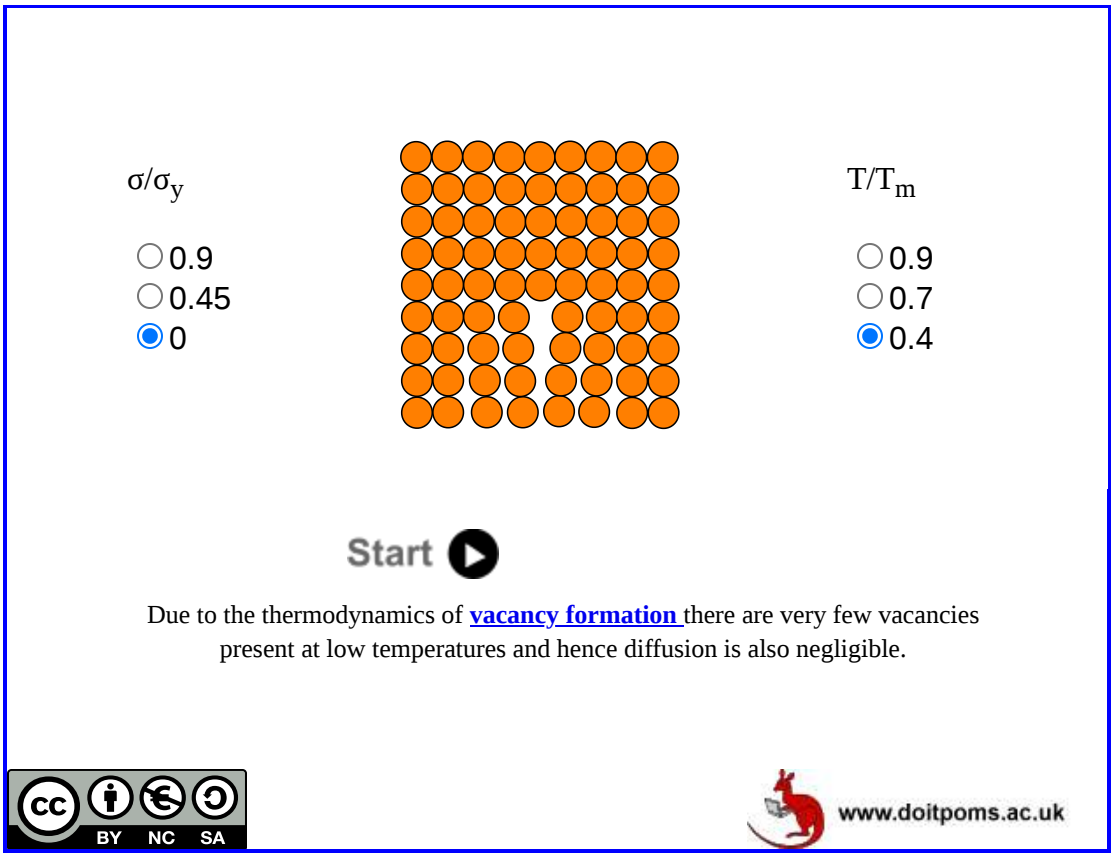


Simulation demonstrating Nabarro-Herring creep on the atomic scale

A similar dependence on temperature and stress is observed to that for Coble creep. The diffusion of atoms in one direction can be more easily pictured as the diffusion of vacancies in the other direction during N-H creep. There is a considerably greater sectional area available via crystal lattice paths, particularly if the grains are relatively large. On the other hand, the activation energy is higher, so diffusion rates tend to be low, particularly at low temperature. This type of creep tends to dominate over Coble creep at ***relatively high temperature,*** and with ***large grains or single crystals***.

Dislocation Creep

Purely diffusional creep (Coble and Nabarro-Herring) is fairly simple, and does occur under certain conditions - usually with relatively low applied stresses. With higher stresses, it is common for a type of creep to occur that involves motion of dislocations, particularly in metals, where dislocation densities tend to be high. Provided the stress is below the yield stress, conventional macroscopic plasticity, occurring predominantly via dislocation glide, should not occur. However, with stresses that are starting to approach the yield stress, and are maintained for extended periods, progressive dislocation motion, and hence macroscopic plastic deformation, can occur, often facilitated by extensive ***climb*** (absorption or emission of vacancies at the core) of individual dislocations. It should be noted that climb does not refer only to vertical motion of the dislocation, and can refer to horizontal motion too. One of the ways dislocation creep can occur via climb is shown in the simulation below:



Simulation demonstrating Dislocation Creep on the atomic scale

In this example, the shear stress provides the driving force for diffusion into the dislocation core, rather than hydrostatic compression or tension as in the case of Coble or N-H creep. In detail, there are several different ways in which combinations of dislocation glide and diffusion in the vicinity of dislocations can promote creep. Some of these have been given specific names, but these often relate to observed dependences on the main variables (e.g. ***temperature*** or ***grain size***), rather than being clear about the precise mechanisms involved. In general, they all involve some combination of dislocation climb and glide, although, in particular cases, factors such as the presence of ***obstacles*** (e.g. fine precipitates), the ease of ***cross-slip*** etc. may affect the observed behaviour. ***Dislocation density*** may also affect the behaviour. However, at the high homologous temperatures at which creep typically occurs,

the dislocation density may decrease somewhat with time, or could indeed drop sharply if recrystallization were to occur. It might be imagined that this would reduce the rate of deformation. However, in practice the associated decrease in the yield stress might well promote the onset of conventional plasticity - ie extensive dislocation glide - such that the rate of deformation increased.

This page titled 12.2: Creep Mechanisms is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.3: Constitutive Laws for Creep

Effects of Microstructure on Creep

As with plastic deformation, creep is a complex process that is strongly affected by the microstructure of the material. (Some of the microstructural effects that influence plasticity are summarised in the **TLP on Mechanical Testing**.) As with plasticity, however, ***guidelines*** can be identified concerning features likely to affect (inhibit) creep, and some of these are similar for the two. For example, a ***fine array of precipitates***, which will inhibit dislocation glide and hence raise the yield stress, is also likely to inhibit (dislocation) creep. However, there are limits to such linkage. For example, precipitates might ***dissolve*** at the high temperatures involved in creep. More fundamentally, some features can affect creep and plasticity quite differently. For example, while a fine grain size tends to raise the yield stress, as a result of ***grain boundaries*** acting as ***obstacles*** to dislocation glide, it can cause accelerated creep in a diffusion-dominated regime, since such boundaries also constitute ***fast diffusion paths***.

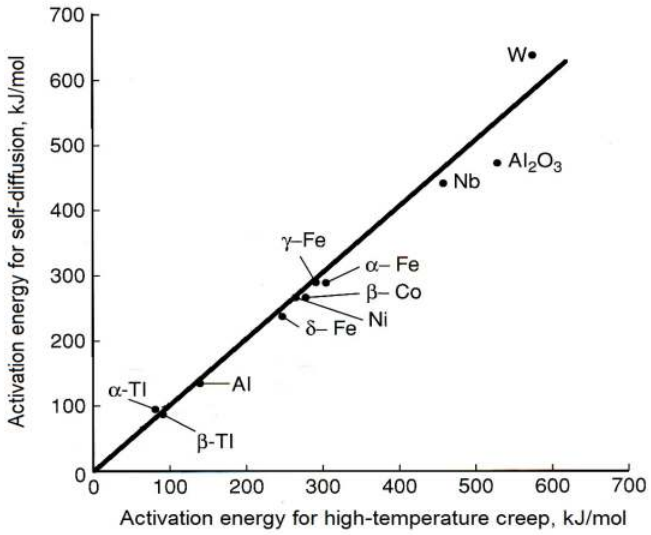
Nevertheless, as with plasticity, empirical ***constitutive laws*** can be used to model and predict creep behaviour. There is sometimes scope for interpreting the ***values of parameters*** in these laws in terms of the dominant ***mechanisms*** involved. In particular, if rates of creep are measured over a range of temperature, then it may be possible to evaluate the ***activation energy***, Q (in an Arrhenius expression), which could in turn provide information about the type of diffusional process that is rate-determining. It is also often claimed that the value of a ***stress exponent***, *n*, obtained via creep rate measurements over a range of applied stress, is indicative of the dominant mechanism, with a low value (~1-2) indicating ***pure diffusional creep*** and a higher value (~3-6) suggestive of ***dislocation creep***. The theoretical basis for such conclusions may sometimes be questioned, and in any event these laws must be recognised as essentially empirical, but it is certainly important to be able to characterize the creep response of a material, and to be clear about the regime of temperature and stress for which a particular law is valid.

General (Steady State) Creep Law

The creep strain rate (rate of change of the von Mises plastic strain) in the steady state (Stage II) regime is often written

$$\dot{\varepsilon} = A\sigma^n \exp(-Q/RT) \tag{12.3.1}$$

where A is a constant, σ is the applied (von Mises - click **here** for definition) stress, Q is the ***activation energy*** and *n* is the ***stress exponent***. This is a relatively simple equation, but several caveats should be added. The most important of these are apparent in Fig.1 of the Introduction page - ie it relates only to the steady state (secondary) regime. It is sometimes stated that the overall creep life is often ***dominated*** by this regime. In practice, this may or may not be true. In particular, it's worth noting that, not only can the primary regime extend over a significant fraction of the creep lifetime, but also, since the creep rate is often much higher during primary creep, its contribution to the overall creep strain can be substantial, and even dominant. Depending on a number of factors, simply ignoring primary creep may be highly inappropriate.



Plot of self-diffusion activation energy vs. high temperature creep activation energy, with a line drawn along which the two are equal. [1]

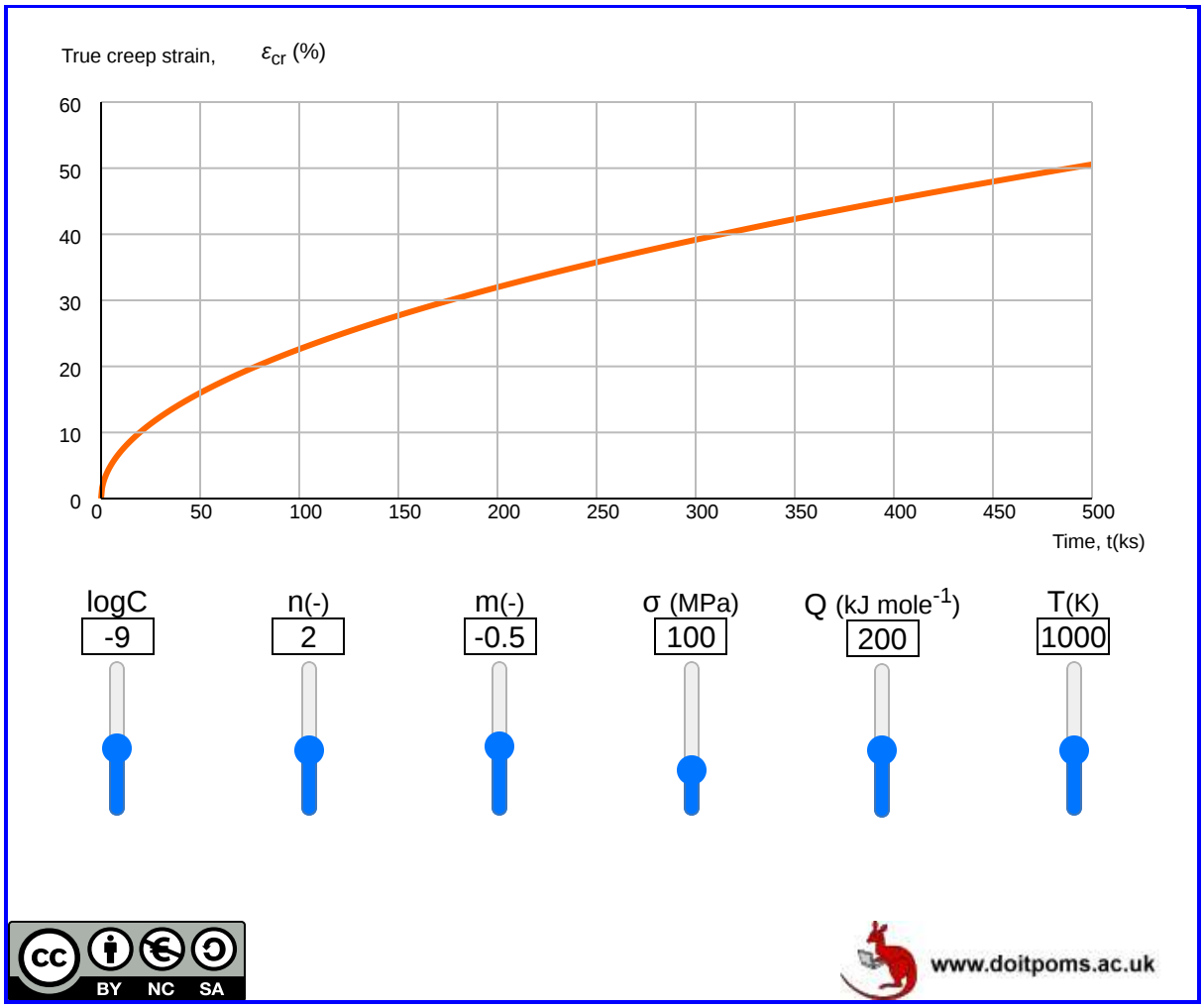
It may be noted that the activation energy for creep at high temperatures is often found to agree closely with that for bulk diffusion - see, for example, the data in the figure. This is consistent with the concept of N-H creep (diffusion through the lattice) dominating Coble Creep (grain boundary diffusion) at high temperatures.

Laws Capturing both Primary and Secondary Regimes

In view of the potential importance of the primary regime, there is strong interest in using modelling approaches that incorporate it. Several expressions have been proposed for capture of both primary and secondary regimes, and of the transition between them. One that can be taken as representative is the following equation, which is sometimes termed the ***Miller-Norton law***

$$\varepsilon_{cr} = \frac{C\sigma^n t^{m+1}}{m+1}\exp\left(\frac{-Q}{RT}\right) \qquad (12.3.2)$$

In this expression, *C* is a constant (units of $Pa^{-n}\ s^{-(m+1)}$), t is the time (s), n is the stress exponent and m is a dimensionless constant. The simulation below, in which this equation is plotted, can be used to explore Miller-Norton creep strain plots as the 6 parameters involved are varied.



Simulation showing behaviour predicted by M-N law with constant true stress

A number of features should quickly become apparent, such as the high sensitivity to temperature and that the sensitivity to the applied stress increases as the value of n is raised.

One issue here is whether the applied stress is a nominal or a true value. It certainly should be a true value, as this is implicit in the M-N law. However, it is common during testing to fix the load (often in the form of a ***dead weight***), rather than the true stress. Also, most uniaxial creep tests tend to be carried out in ***tension***. Neglecting any inhomogeneity that might arise, such as a necking effect - which is not common during creep testing - the true stress will rise as straining occurs and the cross-sectional area reduces. Depending on the value of *n*, this could have the effect of causing the strain rate to rise (whereas it would otherwise be falling and approaching a constant value).

This effect is modeled in the simulation below. The M-N law can be differentiated with respect to time, to give

$$\dot{\varepsilon}_{\mathrm{cr}} = C\sigma^n t^m \exp\left(\frac{-Q}{RT}\right) \tag{12.3.3}$$

Therefore, by stepping in time and repeatedly re-evaluating the true stress, and hence the strain rate, the full creep strain curve can be built up (although it can no longer be expressed as a single analytical equation). In order to implement this, the relationship between true and nominal stresses is needed:

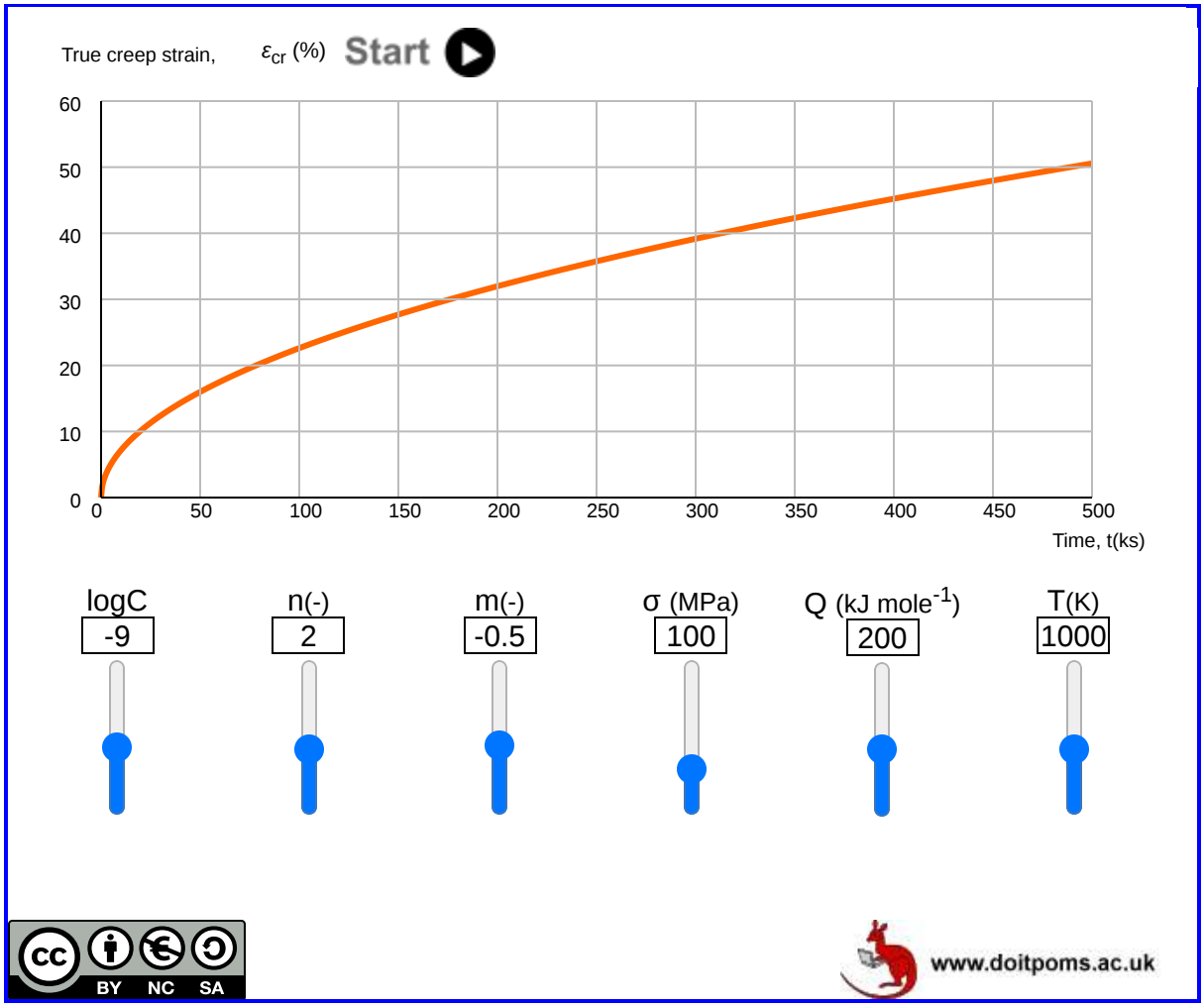
$$\sigma_T = \frac{F}{A} = \frac{FL}{A_0 L_0} = \frac{F\left(L_0 + \varepsilon_N L_0\right)}{A_0 L_0} = \frac{F}{A_0}\left(1+\varepsilon_N\right) = \sigma_N\left(1+\varepsilon_N\right) \tag{12.3.4}$$

and also that between true and nominal strains

$$\varepsilon_T = \int_{L_0}^{L} \frac{\mathrm{d}L}{L} = \ln\left(\frac{L}{L_0}\right) = \ln\left(1+\varepsilon_N\right) \tag{12.3.5}$$

The above plot can therefore be modified using Eqns. (3), (4) and (5), with the nominal stress taken as constant. This is done by stepping forward in small increments of time (ie this is a numerical procedure, rather than just the plotting of an analytical equation). The sequence is as follows. After an initial small time increment, true stress and true strain are taken to be equal to the nominal values. For the next increment, the strain rate is obtained using Eqn.(3), and hence the increment of (true) strain obtained on multiplying this by the time increment. In order to use Eqn.(3), the true stress is needed. This is obtained using Eqn.(4), with the nominal strain obtained from the true strain using Eqn.(5). This operation is repeated after every time step, with the true stress progressively rising.

In the simulation below, the stress selected is a nominal value. Depending on several factors (particularly the value of n), a significant rise in the creep strain rate can be seen. This is sometimes (mis-)interpreted as a “Tertiary” regime, although a similar increase could also arise as a result of microstructural changes (e.g. ***cavitation***). In the case of uniaxial compressive testing, the strain rate will tend to fall, due to a drop in true stress, and hence no such regime will be seen. Of course, these effects will only be noticeable at relatively high strains (say, >~10%), although in practice these are common during uniaxial creep testing.



This figure compares the (tensile) creep strain plot of the standard M-N equation, using a constant nominal stress, with that obtained by stepping through time, repeatedly evaluating the true stress and then using that in the differential form of the equation

It should also be noted that this analysis is based on the Miller-Norton equation being valid, over the range of stresses being considered. In fact, if the stress rises significantly, then this may not be the case. In particular, if the true stress starts to reach the yield stress, then conventional plasticity may be stimulated, which would probably be apparent in a creep strain plot as a sharply increasing strain rate. In fact, this may be responsible for much observed “Tertiary Creep”, rather than it being an effect that can be fully captured by applying the Miller-Norton equation while taking account of the changing true stress.

[1] https://www.intechopen.com/books/sup...ed-superalloys

This page titled 12.3: Constitutive Laws for Creep is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.4: Uniaxial Creep Testing - Practical Basics

Many of the practical issues outlined in the **TLP on Mechanical Testing** (for Plasticity) apply equally to Creep Testing. For tensile testing, the gauge length must have a smaller sectional area than the region that is gripped, such that the latter undergoes only elastic deformation. This is not the case for compressive testing, where the sample usually has a uniform section along its length, and is located between hard platens. For Creep Testing, however, there are additional challenges. For example, it is commonly carried out at high temperature, so a furnace (with good thermal stability) is needed, and all of the sample must be held at the selected temperature. Also, the load must be sustained for long periods - perhaps just a few hours, but in some cases periods of many weeks or even many months might be needed. Such conditions bring slightly different challenges from those of conventional (stress-strain) testing.

Nominal or True Levels of Stress and Strain?

As for conventional testing, there is the issue of true or nominal versions of both stresses and strains. However, there is a difference with Creep Testing. For Stress-Strain Testing, the applied load is being ramped up continuously, so the issue reduces to that of converting any particular load level to a stress in the sample. Provided the stress is uniform throughout the sample, a simple equation can be used to convert an applied load to a true stress. A similar operation allows an extension (nominal strain) to be converted to a true strain. There is a rather more fundamental problem with Creep Testing. As noted in the Introduction page of this TLP, creep is characterised by a series of strain-time curves for different levels of (true) stress. Under fixed load conditions, the true stress changes continuously during the test, so these curves will be unreliable (for tests in which relatively large strains are generated).

Fixed Load Machines

Many creep facilities are based on a fixed load, applied via a dead weight (with a lever arrangement such that the actual applied force is considerably larger than the weight itself). A typical facility of this type is shown below. The actual weight used is unlikely to be much more than a few tens of kg. However, with the lever arm arrangement giving a mechanical advantage of at least 10, and perhaps considerably more, such weights can produce an applied force on the sample of up to several tens of kN. This is usually sufficient for most situations (sample dimensions and required stress levels), although the limitations associated with having a fixed nominal stress (varying true stress) should be noted.



A typical (dead weight) creep frame (produced by Medilab Enterprises)

Variable Load Machines

It is, of course, possible to have a loading frame in which the force (usually produced either by screw-driven or hydraulic systems) can be varied during the test. A typical facility of this type is shown below. Such machines can readily generate forces of up to

hundreds of kN. With the strain (extension) being continuously monitored, software control can be used to change the applied load such that a true stress is maintained. However, it is worth noting that such machines tend to be expensive. They're not well-suited to very long term tests, both because of potential wear on the loading system and because tying up such expensive facilities for long periods is not economically attractive. If a number of samples are to be tested over periods of weeks or months, which is not unusual for creep testing, then it is much more likely that a set of dead weight machines will be used.



A typical (variable load) creep loading frame (produced by IBERTEST)

This page titled 12.4: Uniaxial Creep Testing - Practical Basics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.5: Multiaxial Creep Testing – The Creeping Coil

Background

*Note: this experiment deals with torsion which can be explored further in the **Bending and Torsion of Beams TLP here***

This experiment is a simple and convenient one to carry out. The sample is in the form of a wire that is wound into a coil, which creeps under its own weight. Using a low melting point material such as lead or solder, this occurs at relatively low temperatures (without heating, or using a simple hot air system). Several simplifying assumptions are made in carrying out the analysis. Each turn of the coil is taken to experience a constant stress - the average value acting throughout the section of the wire. Furthermore, it is assumed that the steady state regime of behaviour is immediately established everywhere - ie primary creep is neglected. These are crude assumptions and the experiment cannot be used to obtain quantitative creep characteristics. However, it is possible to obtain fairly reliable indications of the dependence of creep rates on temperature (ie values of Q) and, to some extent, on stress (ie values of n)

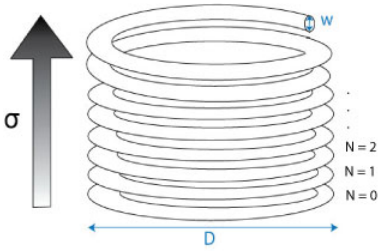


Diagram showing the coil with each turn assigned a number starting from N=0 when the stress is zero

The stress in a particular turn is proportional to its number, N, where the turns are numbered beginning from zero at the bottom turn and ending at the top. The shear stress in each turn varies from zero at the centre of the turning to a maximum value at the edge of the coil and it's average, τ, is given by:

$$\tau = BN \tag{12.5.1}$$

(**Derivation**)

where B is a constant of proportionality (for this experiment). The coil is then allowed to creep over a fixed amount of time (e.g. one minute) and at the end of this time the spacings, s, between the turns are measured.The average local shear strain γ in each turn is a function of s and is given by

$$\gamma = Cs \tag{12.5.2}$$

(**Derivation**)

where C is a constant. Assuming that steady state creep dominates, the strain rate is simply the strain divided by the time. This equation can now be used to explore the relationship between strain rate and shear stress

$$\dot{\gamma} = C\frac{s}{t} \tag{12.5.3}$$

Using the steady state creep equation:

$$\dot{\gamma} = A\tau^n \exp\left(-\frac{Q}{RT}\right) \tag{12.5.4}$$

Taking natural logarithms:

$$\ln\dot{\gamma} = \ln(A) + n\ln(\tau) - \frac{Q}{RT} \tag{12.5.5}$$

using the expressions for average shear stress and average strain rate:

$$\ln\left(\frac{s}{t}\right) = K + n\ln(N) - \frac{Q}{RT} \tag{12.5.6}$$

where

$$K = \ln(A) + n\ln(B) - \ln(C) \tag{12.5.7}$$

This equation can be used to evaluate the creep parameters from (plots of) experimental data.

Experimental Set-Up

https://www.doitpoms.ac.uk/tlplib/creep/videos/28.mp4

https://www.doitpoms.ac.uk/tlplib/creep/videos/85.mp4

In order to create a temperature-controlled environment, the coil is placed inside a Perspex tube. Once the temperature in the tube has stabilised, the coil is allowed to creep under its own weight for one minute. The videos above show qualitatively how creep rate varies with temperature.

Evaluating Parameters

The spacings between the turns of the coil can be obtained from photos or more simply by just rotating the support cylinder until it is horizontal (so that creep will stop). Spacings are then readily measured directly with a ruler. The spacing of the Nth turn is taken as the distance between the Nth and (N-1)th turn, i.e. the spacing of the 1st turn is the distance between the 0th and 1st turn. Using the equation above marked with an asterisk, a plot can be constructed of ln(s/t against ln N), at a given temperature, with the stress exponent, n, obtainable from the gradient. A similar operation can be carried out using data for a number of different temperatures, with ln(s/t) plotted against 1/T, for a selected value of N, to obtain the activation energy, Q

This page titled 12.5: Multiaxial Creep Testing – The Creeping Coil is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.6: Multiaxial Creep Testing – Indentation Creep Plastometry

Indentation creep plastometry is an fairly novel procedure involving use of an ***indenter***, as opposed to conventional uniaxial testing machines. The process has similarities to indentation plastometry, described in the **Mechanical Testing TLP**. The advantages noted there, arising from the procedure being non-destructive, applicable to small samples of simple shape and offering scope for mapping of properties across a surface, also apply here. Furthermore, while uniaxial creep testing requires a series of tests at different stress levels (and experimental difficulties in maintaining a constant true stress), a single indentation creep experiment allows full characterisation of the creep behaviour (as captured in a constitutive law).

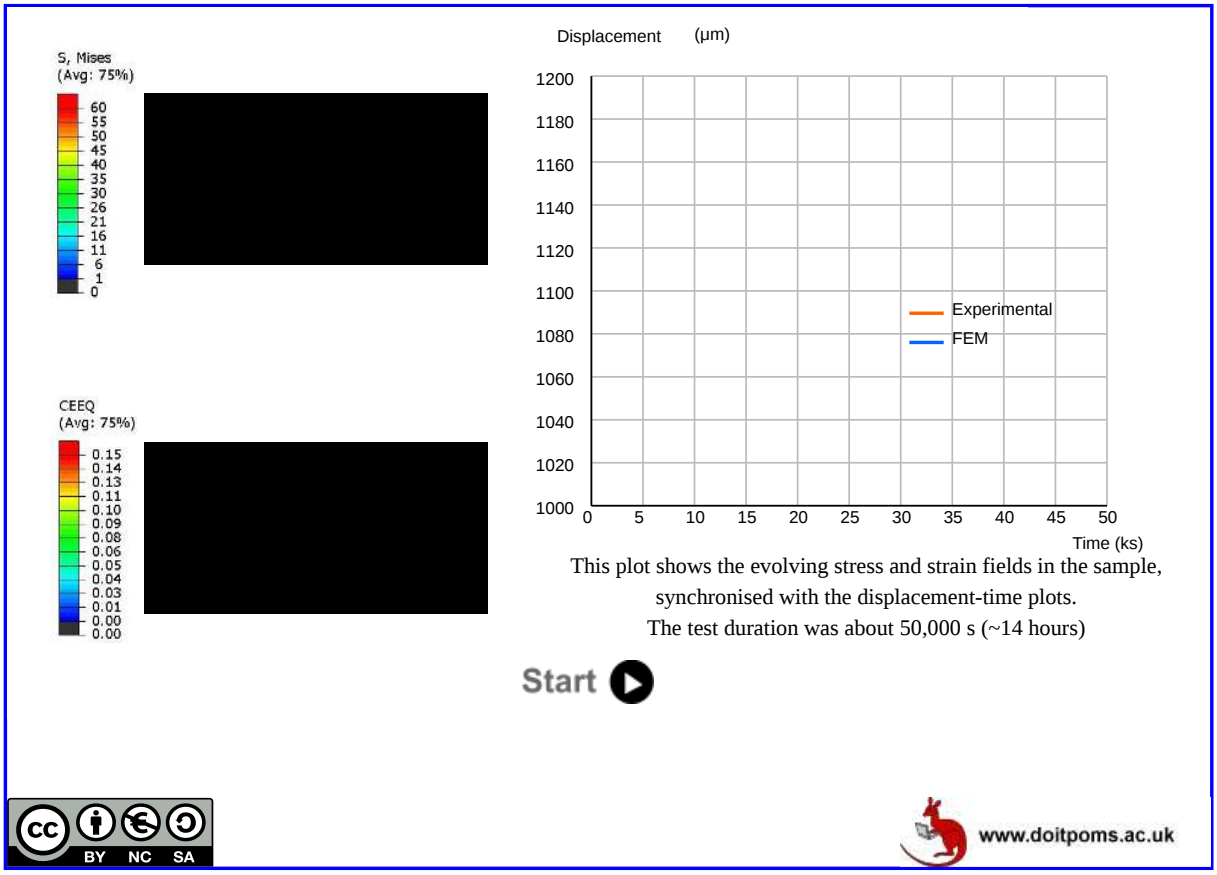
Method

Recess Creation

A spherical indenter is normally used. As with conventional creep testing, it is important to avoid ***conventional (time independent) plastic deformation***. This is particularly an issue in the initial stages, when the contact area would be very low as the indenter penetrated a flat surface, and the stresses correspondingly high. In order to avoid this problem, a (spherical) recess is first created in the sample, matching the indenter. This procedure can be used to ensure that the stresses in the sample never reach the yield stress (at the temperature concerned).

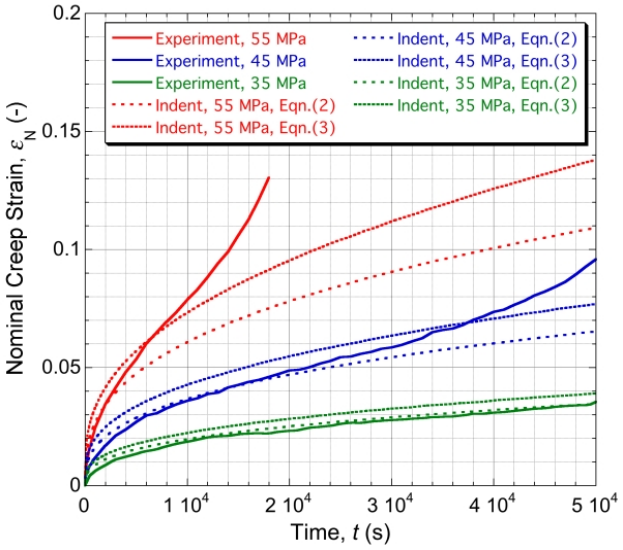
Test Procedure

The selected load is quickly applied and the penetration recorded as a function of time. As with Indentation Plastometry, the core of the procedure is iterative numerical simulation of the test, using the **FEM** method. A constitutive creep relationship, such as the M-N law, is used in the model, with trial, and then improved, values of the (three) parameters in it, with the goodness of fit parameter (S) between predicted and measured displacement-time curves being used to guide the optimisation. The simulation below shows the outcome of this operation for a particular case (a Ni sample at 750°C, with a 1 kN load, an indenter radius of 2 mm and an initial recess depth of 1 mm, with a 2 mm radius of curvature). The evolving von Mises stress and strain fields are shown, together with measured and modelled displacement-time plots, with the optimised M-N parameter set



Predicted and mesured curves agree quite well - the value of S is $10^{-3.2}$, which is acceptably close to zero. The parameter set that gave this level of agreement is shown below:

Cexp(-Q/RT) / Pa^{-n} $s^{-(m+1)}$	n	m
4.3 10^{-8}	2.46	-0.65



Graph showing how our predictions compare with the experimental data for uniaxial tensile creep testing

Using this parameter set, a prediction can be made (using the M-N law directly, with no need for FEM simulation), of the outcome of uniaxial testing, with any particular applied stress level. The outcome of such operations is compared with experimental data, for this material and test temperature, in the plot shown here. Two sets of predicted plots are shown for each stress level. The first set corresponds to use of Eqn.(2) on the page "Constitutive Laws for Creep" - ie the standard M-N law. In fact, the experimental data were obtained during testing with a fixed load (constant nominal stress), so the appropriate plots are really those marked in the figure as referring to Eqn.(3), which were obtained using the differential form of the M-N law - ie Eqn.(3) on the page "Constitutive Laws for Creep". It can be seen that the agreement with these plots is good, apart from periods towards the end of the test with the higher applied stress levels, when tertiary" behaviour is seen, probably because the true stress was approaching the yield stress (~65 MPa), so that "plasticity" was stimulated.

This page titled 12.6: Multiaxial Creep Testing – Indentation Creep Plastometry is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.7: Designing for Creep Resistance - Nickel Based Superalloys

With over 100,000 flights a day, jet engines are commonplace in today's world. Jet engines work more efficiently at higher temperatures. Turbine entry temperatures can be of the order of 1,500°C and, while cooling channels and Thermal Barrier Coatings ensure that the blades do not reach these temperatures, they may be at around 1,000°C - a temperature at which most metals are likely to undergo rapid creep under the stresses concerned (~100 MPa). The gradual deformation of these blades is not only financially costly, as they will have to be replaced, but could lead to catastrophic failure and hence, in some cases, loss of life. This is why it's important to understand the creep behaviour of the materials we use and moreover how we can produce materials with improved creep resistance. Diffusion and dislocation motion are the main mechanisms by which the creep of materials occurs and as such limiting them is key to designing creep resistant materials. Jet engine blades are made from Nickel-based superalloys. A brief outline is presented here of how they are designed, given that resistance to creep at very high temperatures is a key requirement.

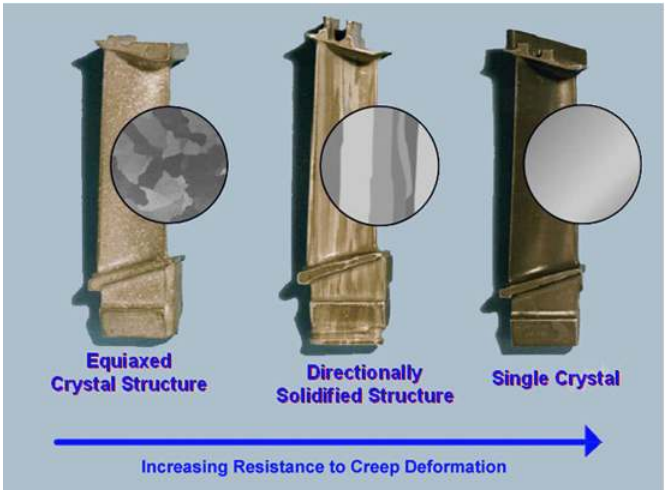
Limiting Diffusion

1.Homologous temperature

The (substitutional) ***diffusivity*** of a material depends on how easily atoms move between vacant sites in the lattice (**see more here**), and on the vacancy concentration. Both jump rates and vacancy concentrations are higher at higher temperatures, approaching limiting values close to the melting temperature. It is thus the ***homologous temperature*** (T/T_m) that is important. For example, diffusion (and hence creep) rates are relatively high at room temperature for Pb (T/T_m ~0.5), but negligible for Ni (T/T_m <0.2). There are clearly advantages in using materials with high melting temperatures."

2.Grain structure

There are two important points to be made regarding grains. Firstly, grain boundaries themselves are detrimental to creep resistance because they provide ***fast diffusion paths*** through the material and thus accelerate (Coble) creep. Secondly, the length of a grain in the direction of applied stress dictates the distance an atom must diffuse in order to reach the "polar" regions from the "equatorial" regions. Elongating the grains parallel to the applied stress direction will thus tend to reduce the creep rate. The latter effect can be obtained with a columnar grain structure, which is relatively easy to create using controlled directional solidification. Even better is a ***single crystal*** (with a selected crystallographic orientation). This requires somewhat greater control during the casting process, although it can be done and this is now routine for many types of blade. The figure below illustrates this.



Progression in manufacture of creep-resistant microstructure over time

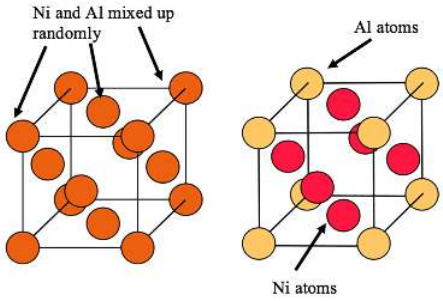
3.Crystal Structure

In addition to the homologous temperature, there are other factors that can affect the diffusivity. These include the crystal structure, which dictates both the packing density and the nature of the paths taken during atomic jumps between vacant sites. In general, diffusive jumps are more difficult in close-packed, high-symmetry structures, such as ***fcc***. There are also additional factors related

to crystal structure. For example, diffusion is difficult in ***ordered structures***, since diffusion tends to create disorder (that is thermodynamically opposed). The microstructure of Ni-based superalloys is typically a 2-phase one, composed of an intimate mixture of γ (fcc Ni) and γ' (an ordered structure based on Ni with Al and/or Ti), with coherent or semi-coherent interfaces between them. These structures are illustrated below.

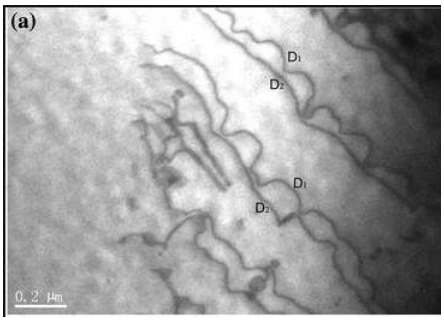
Limiting Dislocation Motion

While the two are often inter-related during creep, it is important to inhibit dislocation motion, as well as diffusion. This happens in several ways in Ni-superalloys. One of these concerns the ordered γ' phase. An ordered structure is degraded, not only by diffusion, but also by dislocation glide. Such motion is therefore opposed, an effect often described as ***order hardening***. (In fact, if dislocations move in pairs, sometimes termed superdislocation pairs, then the second of the pair may restore the order, but this does impose a constraint on the motion - see below.) There is also an effect of the coherency of the γ /γ' interface, which creates lattice strain in the vicinity. This also tends to inhibit dislocation glide.



Unit cell of γ on the left and that of γ' on the right [2]

Precipitates of γ' exist within a γ matrix, and so dislocations will have to pass from γ to γ' in order to move through the alloy. The shortest ***Burgers vector*** in γ is (a/2)<110>, but in γ' this is not a ***lattice vector***. As a dislocation passes from γ to γ' its Burgers vector must be maintained and it will create a region with unfavourable bonding (an ***anti-phase boundary***) and hence an associated energy cost. Its motion is therefore slowed considerably in γ' compared to that in γ, and it can only travel at any appreciable rate through γ' once another dislocation passes into γ which restores the unfavourable bonding back to its ordered state – as they move through together the anti-phase boundary maintains a constant width and there is thus no added cost to their motion through γ'. This can be seen in the micrograph below:



Micrograph showing dislocations passing through a Ni superalloy [3]

It can be seen that dislocations D_1 are impeded by γ' precipitates with D_2 following unimpeded. This mechanism results in much slower dislocation glide through the alloy as a whole and so is effective in reducing the rate of dislocation creep. The finer the dispersion of γ' precipitates the more effective the strengthening will be. The two dislocations that pass through γ' together are called ***super-partial*** dislocations and the sum of their burgers vectors is a lattice vector - a<110> - in the γ' phase. It is important to note that the γ - γ' interface is coherent and therefore coarsening does not occur over time.

[2] Y. M. Wang-Koh (2017) *Understanding the yield behaviour of L12-ordered alloys*, Materials Science and Technology, 33:8, 934-943, DOI: 10.1080/02670836.2016.1215961
[3] Lv, X., Sun, F., Tong, J. et al. J. of Materi Eng and Perform (2015) 24: 143. https://doi.org/10.1007/s11665-014-1307-y

This page titled 12.7: Designing for Creep Resistance - Nickel Based Superalloys is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.8: Summary

This TLP covers the mechanisms involved in creep deformation, how it can be modelled and how it can be investigated experimentally. There is also coverage of the need for creep-resistant materials and how they can be designed.

Firstly, the fundamental mechanisms of creep are covered, including the dependence of each on levels of stress and (homologous) temperature. The key role of diffusion is emphasized, leading to a dependence on time that is not exhibited by conventional plastic deformation.

Secondly, the concept of (empirical) constitutive laws to characterize creep was introduced. The emphasis is often on Stage II (steady state) creep, but in practice the primary regime often constitutes an important part of the overall behaviour and a law is presented that covers both regimes. Mention is also made of the so-called tertiary regime (immediately before final rupture), and the possible ways in which it can arise.

Thirdly, there are descriptions of the experimental procedures that can be used to obtain creep characteristics. The most common of these are the conventional uniaxial (tensile or compressive) tests, which need to be carried out with a series of different applied stresses in order to obtain the values of parameters in constitutive laws. These procedures are rather cumbersome and time consuming. It is also shown that test procedures exist in which the stress and strain fields are more complex, but can be analysed – either using a set of equations or via Finite Element Modelling. The creeping coil experiment and the Indentation Creep Plastometry procedures are described as examples of these, the latter having the potential to at least partly replace conventional creep testing.

Finally, the importance of creep resistance in technological applications was illustrated using the well-known example of Ni-based superalloys in turbine blades for aero-engines and power generation plants. The microstructure of these components, and the ways in which their design and production have been approached, are related to the optimisation of creep resistance, based on various principles outlined earlier in the TLP.

Going further

There are many publications covering creep, over a wide range of depths. Some go into much greater detail than this TLP. The following books
provide a good overview:

Books

Fundamentals of Creep in Metals and Alloys, Michael Kassner, Butterworth-Heinemann, 2015, ISBN: 9780080994277

Creep of Metals and Alloys, RW Evans & B Wilshire, CRC Press, 1985, ISBN-10: 0904357597"

Regarding Indentation Creep Plastometry, which is a very recent development, there are as yet no published books and indeed the software necessary to implement the technology is not yet widely available in user-friendly, commercially mature form. However, there are websites that describe the methodology, where such access is likely to become available in due course. Notable among these is https://www.plastometrex.com/.

This page titled 12.8: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

12.9: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which of the following could be reasons tertiary creep is observed?

a	Rising true stress can cause the strain rate to rise
b	Microstructural damage such as cavitation can occur
c	Rising true stress can exceed the yield stress, causing conventional plastic deformation
d	All of these"

Answer

D

In the γ/γ' structure of Ni-based superalloys, why is the existence of a coherent interface important?

a	It leads to lattice strains that can inhibit dislocation motion
b	The ordered structure of the γ' phase inhibits diffusion
c	The ordered structure of the γ' phase inhibits dislocation motion
d	All of these

Answer

A

What is the function of the recess in Indentation Creep Plastometry?

a	It prevents tertiary creep
b	It allows the stresses created during the early stages of the process to be kept relatively low (below the yield stress)
c	It ensures that the indenter is located properly before indentation starts
d	It removes the near-surface region, which could have different properties from the bulk
e	It eliminates primary creep, so that the steady state regime is immediately established

Answer

B

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

Why is creep deformation different from conventional deformation?

a	Strain develops at a much slower rate in the material
b	Strain produced depends on the time for which the stress is applied
c	Creep cannot occur at room temperature
d	The strain produced doesn’t depend on the applied stress

Answer

A. Incorrect. For conventional deformation any strain is produced instantaneously as a result of a change in applied stress.

B. Correct. Creep differs from conventional plastic deformation because it is time-dependent deformation.

C. Incorrect. Creep can occur, regardless of absolute temperature, provided the homologous temperature is above (roughly) 0.4

D. Incorrect. The creep strain rate is proportional to applied stress raised to the stress exponent.

Which of the following statements is true?

a	Only Coble creep is controlled by diffusion
b	Only Nabarro-Herring creep is controlled by diffusion
c	Only dislocation creep is controlled by diffusion
d	All of the above mechanisms are controlled by diffusion

Answer

D. All creep mechanisms tend to be controlled by diffusional processes, which explains why creep deformation is time-dependent – it takes time for atoms to diffuse a given distance."

Which of the following is not an assumption that we make during the creeping coil experiment?

a	Assume constant true stress when we have constant load
b	Assume that the stress is uniform throughout each turn
c	Assume that only steady state creep occurs throughout
d	Assume that the creep regime is the same throughout the coil

Answer

A. Because the deformation occurs via torsion there is no reduction in cross sectional area and hence, although we are under constant load, we are also under constant true stress conditions

Which of the following features of Ni superalloy turbine blades helps to make them resistant to creep? (There may be more than one)

Yes	No	a	Fine grain structure
Yes	No	b	Single crystal
Yes	No	c	High dislocation density
Yes	No	d	High purity
Yes	No	e	Order hardening due to γ' precipitates in γ matrix
Yes	No	f	Low inherent diffusivity of FCC structure

Answer

A. No

B. Yes

C. No

D. No

E. Yes

F. Yes

This page titled 12.9: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

13: Crystallinity in polymers

https://www.doitpoms.ac.uk/tlplib/polymers/index.php

Learning Objectives

On completion of this TLP you should:

- Understand the concept of crystallinity in the context of polymeric materials
- Be able to recognise and describe spherulites in semicrystalline polymers
- Understand the mechanical behaviour of a polymer subjected to a uniaxial tensile stress, and the effects this has on the polymer's optical properties.

Before you start

You should be familiar with basic optical microscopy techniques and the concept of birefringence and photoelasticity in polymers. Awareness of X-ray diffraction and mechanical testing techniques would also be beneficial. The TLPs listed below contain useful, relevant material.

- Optical Microscopy and Specimen Preparation
- Introduction to Photoelasticity
- X-Ray Diffraction
- Introduction to Mechanical Testing
- The Stiffness of Rubber
- The Glass Transition in Polymers

13.1: Introduction

13.2: Crystalline and Amorphous Polymers

13.3: Spherulites and Optical Properties

13.4: Polymer Stress-Strain Curve

13.5: Tensile Testing of Polyethylene

13.6: Summary

13.7: Questions

This page titled 13: Crystallinity in polymers is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

13.1: Introduction

Polymers are widely used materials. Their versatility makes them suitable for a whole range of applications, and comes from the capability of manufacturers to tailor microstructures and therefore properties through control of the processing conditions.

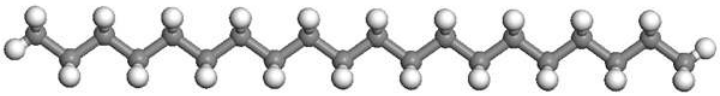
An understanding of polymer crystallinity is important because the mechanical properties of crystalline polymers are different from those of amorphous polymers. Polymer crystals are much stiffer and stronger than amorphous regions of polymer. For example, high strength fibres can be produced from polyethylene whereas it is more commonly associated with applications such as carrier bags and plastic cups, where low cost and ease of manufacture are the key considerations in the choice of material. This TLP covers the formation of crystals in polymers, how they can be observed under the optical microscope, and the response of a semicrystalline polymer sample to uniaxial stress.

This page titled 13.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

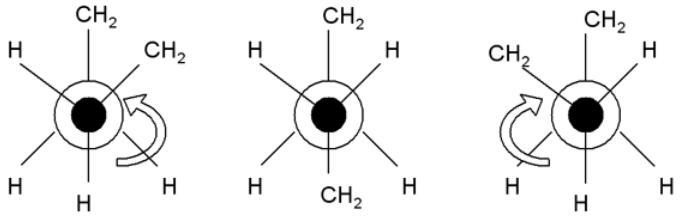
13.2: Crystalline and Amorphous Polymers

In ceramics or metals, a crystalline solid comprises repeating unit cells that contain each of the component atoms in the material. Each unit cell is composed of one or more molecular units. In a polymer this is not possible; the molecules are chains containing potentially millions of formula units. There is, however a repeating unit in a polymer - the monomer from which it was made. This must be the basis of both long and short-range order in a polymeric material.

For example, a short section of linear poly(ethylene) looks like this:



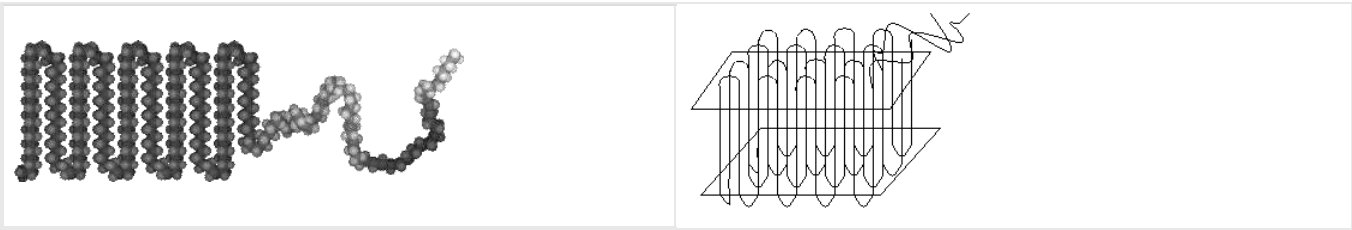
However, the conformation of the bonds around each carbon atom can be represented schematically as follows:



These diagrams are called Newman projections. The circle is a single C-H bond; and this diagram represents a projection along it. These two structures thus represent one half of the backbone continuing on either side of a C-C bond (trans), or both halves on the same side (gauche). Note that there are two possible gauche states, labelled gauche (-) and gauche (+).

Whilst the trans conformation has a lower energy (since it's easier to position the hydrogen atoms on the carbon backbone further apart), an all-trans conformation would be a considerably more ordered structure than a random one - that is, it has a much lower entropy.

Amorphous polymers are generally found in a random coil conformation and have a disordered chain structure. This is the most common structure of many polymers. Crystalline polymers are predominantly in the all-trans conformation, and the chains are arranged in lamellae, as below:



The polymer crystal is made up from one-dimensional chain-folded sequences, shown on the above left, where the repeat distance is given by the chain spacing. To the above right is shown a schematic arrangement of folded chains into a two-dimensional lamella.

An amorphous polymer has the maximum entropy conformation (given by the Boltzmann distribution), and the chains are arrayed randomly throughout the material, making atomic positions quasi-random as in any other glassy material.

As a result of the difference between the amorphous and crystalline arrangements of polymer chains, the X-ray diffraction patterns of the two phases are very different. The amorphous phase contains no long-range order, meaning that there are no regular crystalline planes to diffract X-rays. Thus the incident X-rays are scattered randomly and there are no sharp peaks in the diffraction

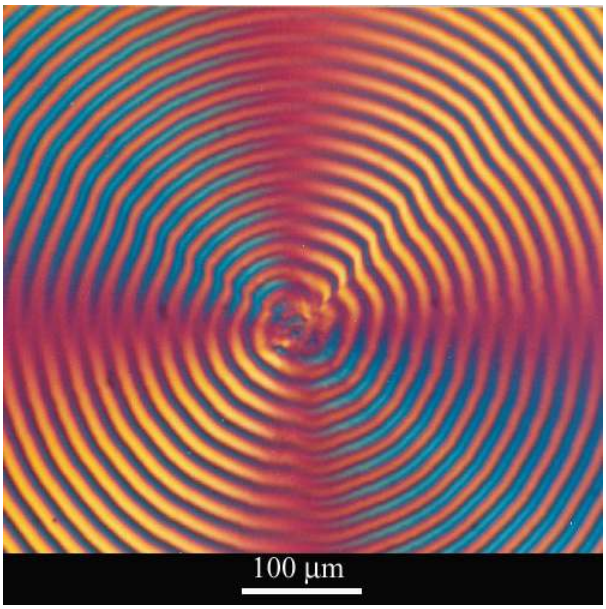
pattern. In the crystalline phase, the repeating lamellar chains provide a regular structure, thus the diffraction pattern will contain sharp, prominent signature peaks, the position of which depends on the exact spacing between chains.

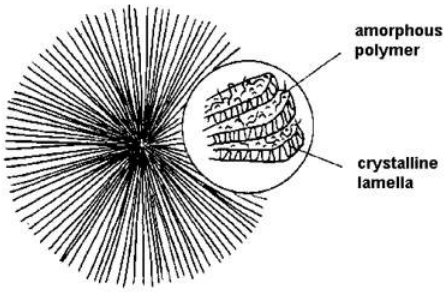
As the degree of crystallinity of a polymer affects its properties, accurately determining it is important. X-ray diffraction can be used to determine the degree of crystallinity of a sample. Thermal analysis techniques such as differential scanning calorimetry (DSC) can also be used. The two determinations may not necessarily be in agreement, and the reasons for this are complex.

This page titled 13.2: Crystalline and Amorphous Polymers is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

13.3: Spherulites and Optical Properties

Since crystallisation in polymers follows a different process to that in metals - the laying down of successive lamellar layers of polymer chain - it produces a different structure. After nucleation, growth in most polymers is faster in one preferred direction. By convention this is called the b-axis. The other two axes (the c- and a-axes) grow at the same speed, and have no set direction provided they are orthogonal to the b-axis. Thus they are free to rotate. This means that polymer crystals grow in helical strands radiating from a nucleation point. Such growth leads to the formation of structures called spherulites.

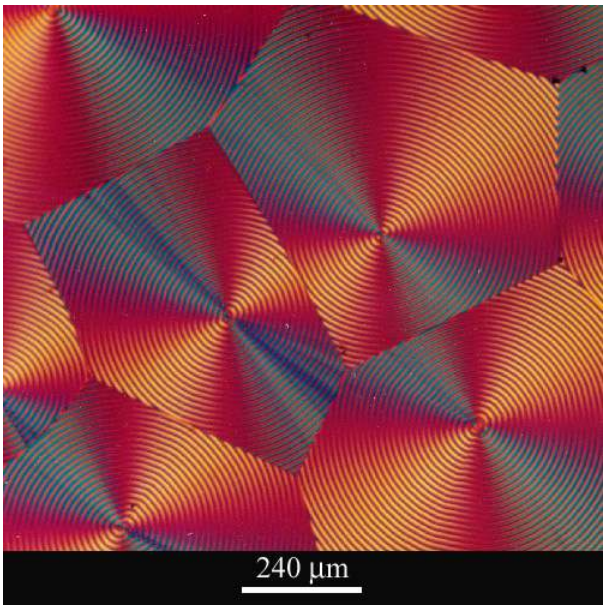




On the left is a transmitted cross-polarised light micrograph of a spherulite in polyhydroxybutyrate (PHB), micrograph no 600 in the DoITPoMS Micrograph Library, where further details of the sample's history can be found. The photograph displays banding and a Maltese cross pattern. These features are characteristic of polymer spherulites viewed with cross-polarised light. The orientations of the polymer chains within a spherulite are shown schematically on the above right. Note that the lamella are growing radially, interspersed with amorphous material.

The Maltese cross is seen because polymers are birefringent. Polarised light cannot travel through a crystalline polymer if the direction of the polarisation of the light is perpendicular to the direction of carbon chain in the polymer. As a result, when a sample is studied under crossed polars, only those polymer chains perpendicular to neither polariser nor the analyser are visible - these are at approximately 45° to each polaroid. The Introduction to Photoelasticity TLP goes into more depth on the subject of birefringence.

The banded appearance of the image is also a consequence of birefringence. Due to a regular helical twist of the lamellae growing radially there are regions in which the polymer chain will be orthogonal to the polarisation of the light in the x-z plane, even if it is running at 45° to it in the x-y plane. These positions will occur at periodic intervals, once every full 360° rotation of the polymer chain. This translates to a given length outwards along each strand. This is then observed under crossed polars as alternating dark and light areas.



The image above is micrograph 601 in the DoITPoMS Micrograph Library. It is an image of PHB spherulites viewed with a transmitted cross-polarised light microscope. A Maltese cross can be seen in each spherulite and each one has a banded appearance. The photograph shows a good example of impingement, which occurs when spherulites growing outwards from single nucleating points meet each other. Due to impingement they are unable to continue growing out radially in all directions. A polygonal microstructure is formed as seen in this photograph.

This page titled 13.3: Spherulites and Optical Properties is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

13.4: Polymer Stress-Strain Curve

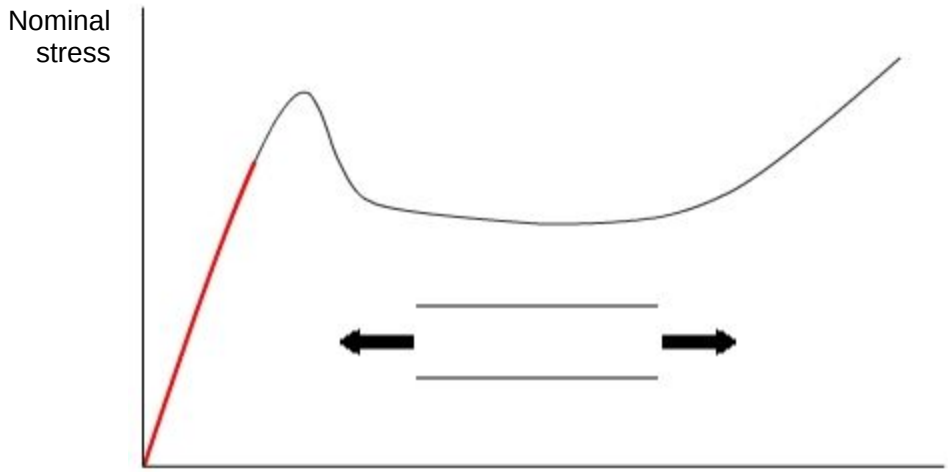
Stress-strain curves show the response of a material to an applied (usually tensile) stress. They allow important information such as a material's elastic modulus and yield stress to be determined. Accurate knowledge of these parameters is paramount in engineering design.

Polymer stress-strain curves are produced by stretching a sample at a constant rate through the application of a tensile force. By using a constant rate of testing the strain-rate dependency of polymer behaviour is not allowed to dominate. However it should be appreciated that polymers have a marked inherent time-dependence in their response to deformation, which sets their behaviour apart from other classes of material.

The following interactive object shows an idealised polymer stress-strain curve and allows you to explore the different stages in the curve as the strain is increased.

Polymer stress-strain curve

Next



Elastic Region. In this region stress is proportional to strain and the deformation is completely reversible; Hooke's law is obeyed. The gradient of the straight-line section of the curve gives the elastic modulus of the material. The entropic mechanism for rubber elasticity is explained in the Rubber Elasticity TLP. In practice the elastic region of the curve is unlikely to be perfectly straight; the elastic modulus is taken as the tangent to the first part of the curve.

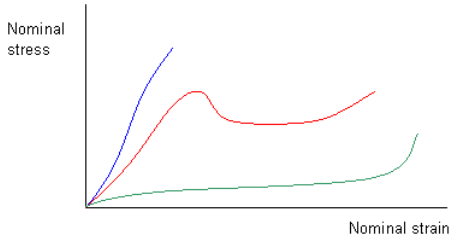
BY NC SA

www.doitpoms.ac.uk

Considère's construction is a useful way of considering the phenomenon of necking and cold drawing.

Temperature dependence

As well as having significant time dependence, polymer behaviour is also temperature dependent. For example the plot below shows schematically different types of polymer stress-strain behaviour, which occur for different types of polymers and at different relative temperatures. The blue curve is for a glassy or semi-crystalline polymer below its glass transition temperature, the red curve is for a semi-crystalline polymer above its glass transition temperature and the green curve is for a rubber.



Stress-strain curve showing behaviour of different polymers and the effect of temperature

The exact form of the stress-strain curve depends on the polymer under investigation. Indeed not all polymers exhibit necking and cold drawing. Polycarbonate is an example of a brittle polymer.

This page titled 13.4: Polymer Stress-Strain Curve is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

13.5: Tensile Testing of Polyethylene

Method

In this experiment samples of polyethylene are subjected to a tensile stress. They are stretched at a constant rate until they fail. The sample is dumbbell-shaped so that it can be firmly held by each set of grips.

The test is carried out using a tensometer. The photograph below is of the Hounsfield H20K-W apparatus, which was used to perform the experiment.



First the sample is secured in place. The strain rate for the test is programmed into the machine, after which the test may begin. After the sample has failed and the test is complete the data is transferred to a computer, where it can be treated and presented as a stress-strain curve.

The test is performed at different strain rates in order to demonstrate the time-dependence of the sample's behaviour.

Video clips

You can view videoclips of a typical test run displaying necking, drawing and fracture. The video of the whole test has been speeded up by a factor of 48 - in reality these thick polyethylene samples took around 30 minutes to break at a strain rate of 10 mm/min. By dragging the video position marker to and fro it become easier to see certain features such as the decrease in the amount of the sample of original width as necking proceeds. Note the position where fracture occurs.

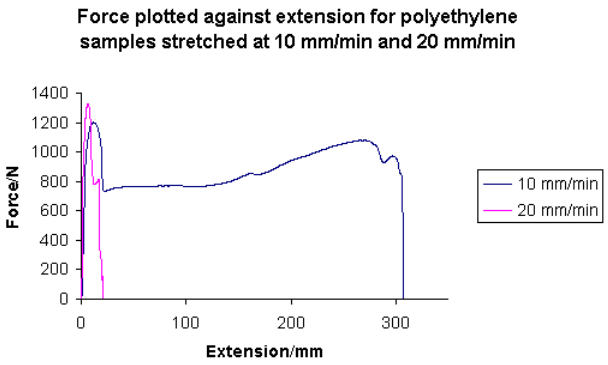
https://www.doitpoms.ac.uk/tlplib/po...eos/vid001.mp4

https://www.doitpoms.ac.uk/tlplib/po...eos/vid002.mp4

Fracture of polyethylene in tensile test

Results

The figure below shows applied force plotted against extension for two polyethylene samples that underwent a tensile test until failure. The blue line is for a sample stretched at 10 mm/min, the pink is for 20 mm/min.



The blue line shows necking and an extensive period of cold drawing, which are characteristic features of the stress-strain behaviour of a semi-crystalline polymer above its glass transition temperature (T_g). For polyethylene, T_g = 148 K and the tests were carried out at room temperature (approx. 293 K). The pink line has a very different form, without the long flat section due to cold drawing seen in the blue line. At the higher strain rate (pink line) almost no cold drawing was able to take place before the sample snapped. This is a good demonstration of the time-dependent nature of the properties of polymers i.e. the observed behaviour is dependent on the strain rate.

Discussion

The results presented here show the difference in behaviour of polyethylene as a result of a different applied strain rate. All other variables were the same for both tests. No cold drawing is observed at the higher strain rate as it is not slow enough for the polymer chains within the sample to disentangle and reorganise themselves parallel to the direction of applied stress; at the lower strain rate there is enough time for cold drawing to occur.

These experiments have highlighted the effect of changing the strain rate on the response of polyethylene to an applied tensile stress. A series of similar experiments could also be performed in order to compare the behaviour of a number of different polymers.

This page titled 13.5: Tensile Testing of Polyethylene is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

13.6: Summary

Polymers contain ordered crystalline regions; chains fold to produce lamellae. When lamellae are nucleated from a small area, round agglomerates known as spherulites are formed. In a fully crystalline material they impinge on each other and appear polygonal under the microscope. In this TLP, polymer crystallisation and spherulite formation have been discussed. Micrographs of spherulites showing the distinctive Maltese cross were presented and their appearance was explained in terms of birefringence.

From the section of the TLP dealing with mechanical testing, you should have gained an understanding of the mechanism of necking and cold drawing by which a semicrystalline polymer deforms under a uniaxial tensile stress above its glass transition temperature. Furthermore, Considère's construction, which is a simple way of determining whether a polymer sample will neck and cold draw, was derived. The stress-strain curves produced by such experiments should now be familiar, for varying strain rates and relative temperatures.

Going further

Books

- *Plastics Microstructure and Engineering Applications*, N J Mills, Edward Arnold, 1993 (2nd edition).
- *Fundamentals of Polymer Science - An Introductory Text*, Paul C Painter and Michael M Coleman, Technomic, 1994.
- *Introduction to Polymers*, R J Young and P A Lovell, Chapman & Hall, 1991 (2nd edition).

Websites

- The Macrogalleria
  Describing itself as "a cyberwonderland of polymer fun", this multi-award-winning site contains a wealth of information about polymers, including a section on Differential scanning calorimetry.

This page titled 13.6: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

13.7: Questions

What is the main reason for the predominance of *trans* states in crystalline polyethylene?

a	Allows better chain packing to produce a denser structure
b	Each *trans* state is lower in energy due to steric hindrance effects
c	The all-*trans* conformation is entropically favoured
d	Formation of chain-folded lamellae requires *trans* states

Answer

B. Whilst statements (a) and (d) are certainly true (and (c) is definitely false) it is the lower energy of the *trans* states due to steric interactions between constituents that *drives* crystallisation in most polymers. Chain folding is a kinetic necessity, as the nucleation of chain-extended crystals would be too slow for polymers of high molecular weight. In fact, polymers containing bulky side-groups are generally inhibited from crystallising and instead form glassy structures.

Rank these structural features of a quenched high molecular weight semi-crystalline polymer in order of their size, smallest first (based on average values - there will be some overlap in ranges):

A. Chain length
B. Lamellar thickness
C. Spherulite radius
D. Root-mean-squared end-to-end distance of chains

a	B, D, C, A
b	A, C, B, D
c	D, C, A, B
d	C, B, D, A

Answer

a. Provided the sample is quenched, the lamellar thickness should be smaller than the r.m.s. end-to-end distance. This is because the lamellae are much wider and broader than they are thick, and there is some memory of the random coil configuration in the melt, with each chain participating in several different lamellae. The spherulite radius is typically 1-10 microns, depending on the polymer and crystallisation conditions, whereas the chain length could be much larger than this for a high molecular weight polymer.

PHB (polyhydroxybuturyate) is a naturally occurring, biodegradable polymer produced by bacteria, which melts at around 200°C and forms large spherulites readily on cooling. What sort of heat treatment would result in a solid with mechanical properties most suitable for a making a fizzy drinks bottle?

a	Annealing at 100°C for one hour
b	Heating to just above 200°C until melted, followed by rapid quench to room temperature
c	Heating to just above 200°C until melted, followed by rapid quench to room temperature, and annealing at 100°C for one hour
d	Heating to just 200°C until melted, followed by a slow cooling to room temperature

Answer

C. Annealing at 100°C for one hour would allow some transformation of the polymer to its crystalline form, but since the thermal history of the sample is unknown it is best to melt the solid first. A rapid quench would result in the nucleation of a large number of small spherulites, giving a high amorphous content solid with poor mechanical properties. Slow cooling would improve matters, but the initial spherulite density would be low so the final solid would contain large spherulites and tend to be rather weak and brittle. An initial fast quench, followed by annealing, should produce a highly crystalline solid with small spherulites, which is the best possible microstructure to achieve good strength and stiffness to resist the pressure inside a fizzy drinks bottle.

During a tensile test of a polyethylene dumbbell sample at room temperature using a strain rate of 10 mm/min, at what position relative to the sample is failure most likely to occur? Justify your answer in terms of the changes in polymer chain orientation in the sample during tensile draw.

a	Exactly at the mid-point between the jaws of the tensile tester
b	Exactly in the middle of the neck
c	Equally likely at either end of the dumbbell, where the neck meets the amorphous region
d	At one end of the dumbbell, furthest from the point at which the necking originated

Answer

D. Watch the movie of the tensile test again, slowly. The polymer chains in the neck are highly aligned, and material is drawn out from the amorphous regions at either end of the dumbbell (notice that the polymer in the neck itself does not stretch, but the neck simply extends as more amorphous material is aligned). The end furthest from the origin point of the neck contains the most amorphous material, and therefore will therefore extend until no more material can be draw out, at which point brittle failure occurs at the junction between the neck and the remaining material around the jaws.

How are polymer chains able to form crystallites?

Answer

Polymers are long chain molecules that are folded and twisted along their length. A bulk sample contains an enormous number of chains of varied lengths and with a range of molecular masses. Crystallites occur in regions where there is an ordered arrangement of sections of polymer chains. A single chain can pass through a single crystallite on many occasions and will pass through both amorphous and crystalline regions of a bulk polymer. When a polymer crystallises, lamellae are generally the first structure to be formed, which grow into more complex three-dimensional structures such as spherulites. In a crystallite, chains can be arranged parallel to each other or in other arrangements such as helices in which sections of chains are twisted around each other with a regular repeated sequence. For crystallites to form on cooling, a sample must be cooled slowly enough to allow ordered regions to arrange themselves. In general a quenched polymer will cool too quickly to allow any crystallites to form.

Which distinctive shape is seen when observing a spherulite between crossed polars? Explain why this shape is seen.

Answer

When observing a spherulite between crossed polars, a distinctive Maltese cross pattern is seen. The arms of the cross are parallel to the crossed polarising filters below and above the specimen. This pattern occurs as a consequence of the orientation of crystal lattice axes and the fact that in an anisotropic material, e.g. a polymer crystal, each lattice axis has a different refractive index. That the cross remains stationary relative to the filters if the polymer is rotated on the microscope stage, demonstrates that the optic axes are tangential and radial in the spherulite.

A second feature of spherulites is visible in polarised light. The radiating fibrous appearance occurs due to lamellae within the spherulite having a variety of orientations.

State two techniques that could be used to determine the degree of crystallinity of a polymer.

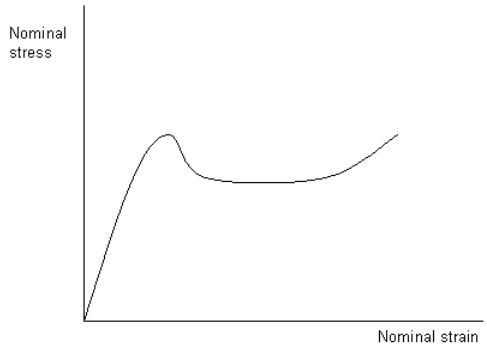
Answer

Two methods that could be used to determine the degree of crystallinity of a polymer are x-ray diffraction (XRD) and differential scanning calorimetry (DSC).

Describe and explain the form of a stress-strain curve for a semi-crystalline polymer subjected to a uniaxial stress above its glass transition temperature.

Answer

The general form of the stress-strain curve is shown below. Hooke's law is obeyed in the initial elastic region; deformation is uniform and reversible. The dip in the curve occurs after yield and is a consequence of the reduction in local area that occurs when the sample begins to neck. There is then a flat region of the curve, during which the sample is cold drawing as the sample necks along its entire length and the chains align themselves parallel to the direction of the applied force. The final rise in the curve before failure is due to strain hardening.



A stress-strain curve for a semi-crystalline polymer subjected to a uniaxial stress above its glass transition temperature

Describe how Considère's construction can explain the phenomenon of necking and stable cold drawing in polymers. What other types of behaviour may be observed for a polymer sample subjected to a uniaxial stress?. What other types of behaviour may be observed for a polymer sample subjected to a uniaxial stress?

Answer

For a sample under uniaxial tension, there is an observed drop in nominal stress at the onset of necking. At this point:

$$\frac{d\sigma}{d\lambda} = 0 \tag{13.7.1}$$

Where σ is the nominal stress and λ is the extension ratio, which is defined as the current length divided by the initial length and is therefore related to the nominal strain; an extension ratio of 1 corresponds to 0% strain. By conservation of volume it can be shown that:

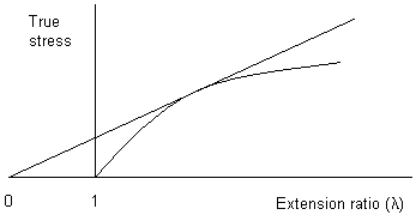
$$\sigma = \frac{\sigma_t}{\lambda} \tag{13.7.2}$$

Therefore

$$\frac{d\sigma}{d\lambda} = \frac{1}{\lambda}\frac{d\sigma_t}{d\lambda} - \frac{\sigma_t}{\lambda^2} \tag{13.7.3}$$

In these expressions σ_t is the true stress, which is equal to the applied force divided by the instantaneous area. Since $\frac{d\sigma}{d\lambda} = 0$ at the onset of necking, it follows that at the onset of necking the condition below is satisfied.

$$\frac{d\sigma_t}{d\lambda} = \frac{\sigma_t}{\lambda} \tag{13.7.4}$$



True stress plotted against extension ratio, showing Considère's construction for the onset of necking

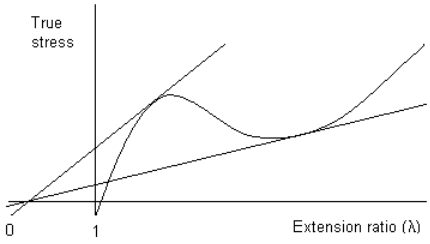
From the plot above, of true stress versus extension ratio, it is possible to determine whether necking will occur. A tangent to the curve constructed from $\lambda = 0$, touches the curve at the point where necking starts. This is Considère's construction.

As in the plot below, it is sometimes possible to draw a second tangent to the curve from $\lambda = 0$, which touches the curve at a higher extension ratio. In the region between the points where the tangents touch the curve, the polymer exhibits stable cold drawing; the point where the second tangent touches the curve corresponds to the limiting strain in the neck. In conclusion three types of behaviour may be recognised:

If $\frac{d\sigma_t}{d\lambda} > \frac{\sigma_t}{\lambda}$ at all times, no necking occurs; the material extends uniformly until failure.

If $\frac{d\sigma_t}{d\lambda} = \frac{\sigma_t}{\lambda}$ at one point on the curve, then the material will neck and fail.

If $\frac{d\sigma_t}{d\lambda} = \frac{\sigma_t}{\lambda}$ at two points, then necking will be followed by cold drawing.



True stress plotted against extension ratio, showing the double tangent construction for stable cold drawing

Compare and contrast features of the crystallisation in polymers with that in other types of materials, such as metals and ceramics.

This page titled 13.7: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

14: Crystallographic Texture

www.doitpoms.ac.uk/tlplib/cr...ture/index.php

Learning Objectives

On completion of this TLP package, you should:

- understand the concept of crystallographic texture;
- be aware of different methods of measuring and representing texture;
- be aware of the effects of texture;
- be able to identify specific textures from pole figures and crystal orientation distribution functions.

Before you start

You should be familiar with the use of stereograms to plot poles of crystals and have a good knowledge of atomic structure and be able to index planes with Miller indices. You should also be familiar with mechanisms of slip in single crystals.

It would be helpful to read the following TLPs: Atomic scale structure, Crystallography, Miller indices, The Stereographic Projection, and Slip in Single crystals.

Material in this TLP relates to a package found on the AluMATTER website explaining how stereograms can be used to represent texture: AluMATTER - representing texture

14.1: Introduction
14.2: Measurement of Texture
14.3: Representing Texture
14.4: Origins of Texture
14.5: Effect of Texture
14.6: Summary
14.7: Questions

This page titled 14: Crystallographic Texture is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

14.1: Introduction

What is crystallographic texture?

In a polycrystalline material, the crystallographic axes of the grains can be oriented randomly with respect to each other, or they can be oriented so that there is a non-random distribution. If there is a preferred orientation, then we say that the material has *crystallographic texture*.

This page titled 14.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

14.2: Measurement of Texture

In the past, optical methods and etching have been used to determine grain orientation, but recently texture is almost exclusively measured by diffraction techniques. Diffraction of x-rays, electrons, and neutrons will all be discussed in this TLP.

X-ray diffraction

The most common method of measuring texture uses x-ray diffraction and is known as the “Schultz reflection method”. The apparatus used is known as a four-angle diffractometer or a *Eulerian cradle.*

The source of x-rays and the detector are oriented so that a particular value of 2θ is specified. This allows for a single Bragg reflection to be measured. The stage of the cradle is tilted and rotated systematically, so that all angular orientations of the sample are investigated.

The animation below shows how the Eulerian cradle is used for measurement of texture using the reflection method.

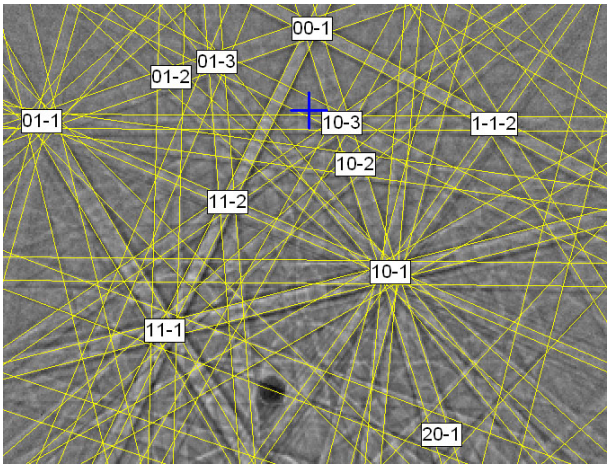
When the specified lattice plane of a crystallite fulfils the Bragg condition, the detector will record the reflection. For a polycrystalline material, the intensity of detected x-rays will increase when there are more crystallites in a specific orientation. The intensity for a given orientation is proportional to the volume fraction of crystallites with that orientation. Areas of high and low intensity suggest a preferred orientation, while constant intensity at all angles would occur in a random polycrystalline aggregate.

X-ray diffraction may be carried out so that the x-rays are reflected from the surface of the sample, or they may penetrate via transmission. Transmission is only suitable for thin films or wires because of the high absorption of the x-rays by many materials.

In some materials the bulk and surface textures may be different, e.g. in some rolled textures. Therefore, it is important to identify which texture is of interest. Different sources of radiation can lead to different degrees of penetration, and hence allow the measurement of either bulk or surface textures.

Electron Backscatter Diffraction (EBSD)

An alternative method of texture determination is that of electron backscattered microscopy using a scanning electron microscope. Within a single grain, the electron beam is fixed at a point on the surface. At particular angles the beam is diffracted, so that there is a change in the intensity of the reflection measured. This leads to the formation of a *backscatter Kikuchi pattern* made up of Kikuchi lines.



Boron doped (001) Si wafer. The blue cross is at 90° to the incident electron beam. The wafer is oriented so that [001] is normal to the stage. The stage is held at an angle of 70°, so that the blue cross is 20° from the stage normal. Poles are determined and plotted by the commercial EBSD software using symmetry and vector addition considerations. Image courtesy of Dr Jeff Wheeler.

The location and symmetry of the backscatter Kikuchi pattern allows specific bands of Kikuchi lines to be indexed unambiguously. The indexed patterns are used to describe the orientation of the grain within typically an experimental error of ±2°.

The information stored by the commercial software using this method includes specimen coordinates as well as the crystal orientation. Therefore, it is possible to build up a two dimensional map of the orientation of grains on the surface of a polycrystalline material.

In this image of recrystallised stainless steel, grains, annealing twins and grain boundaries are apparent.

The orientation at each pixel in the image is represented by a different colour. Each coloured pixel is defined by a single EBSD measurement. The colours can be separated into their red, green, and blue constituents by fractal analysis. The orientation of the crystal can be determined from the key. Each grain orientation is described with reference to an external frame of reference. Here, the orthogonal axes of the external frame of reference with respect to which the grain orientations are defined are (i) the direction

normal to the plane of the sheet, (ii) the direction parallel to the tensile axis, and (iii) the transverse direction parallel to the plane of the sheet.

Thus, for example, there is a large grain coloured mostly red in the lower left-hand corner of the picture. Within this grain there are subtle changes in rotation giving rise to regions which are more orange than red in colour: this indicates the occurrence of low-angle sub-grain boundaries within this large grain. More significantly, there are grey, green and blue-grey regions abutting this red grain. Each of these three regions is twinned with respect to the red grain.

Neutron diffraction

Neutron diffraction can be used in a similar way to x-ray diffraction. There is a large reduction in absorption but a much higher angular resolution in neutron diffraction in comparison with x-ray diffraction. In-situ texture changes due to environmental factors (e.g. temperature changes and stress) can be measured using neutron diffraction.

This page titled 14.2: Measurement of Texture is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

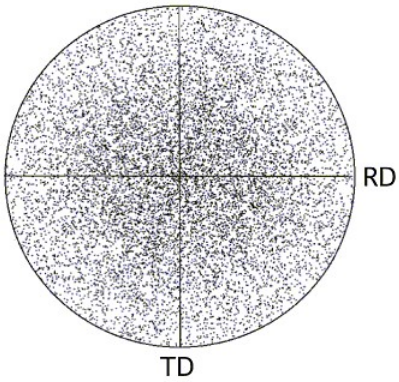
14.3: Representing Texture

Pole figures

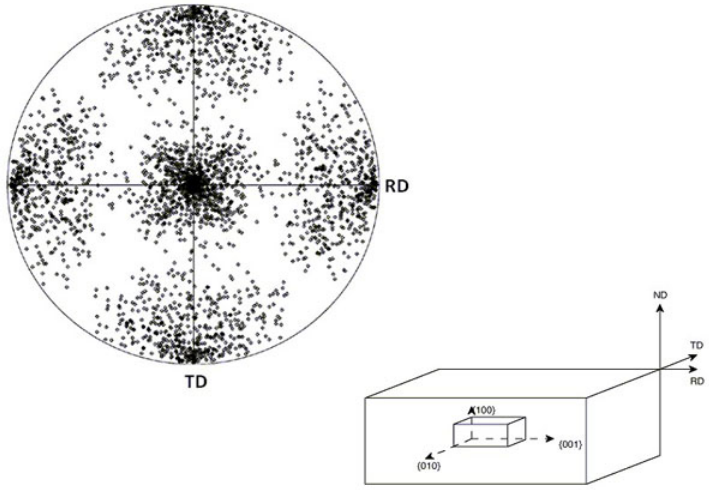
A pole figure is simply a stereogram with its axes defined by an external frame of reference with particular *hkl* poles plotted onto it from all of the crystallites in the polycrystal. Typically, the external frame is defined by the normal direction, the rolling direction, and the transverse direction in a sheet (ND, RD and TD respectively. Occasionally, CD meaning cross direction is used instead of TD.)

The animation below shows the relationship between the orientation of the crystal and the stereographic projection obtained for the <100> poles. Drag an atom in the green sphere to reorientate the unit cell of the grain under consideration. This will alter the projections of the [100], [010] and [001] directions on the stereogram inside the rectangle. Press 'Add grain' to add the [100], [010] and [001] directions of another grain, up to a maximum of four additional grains. Try altering their orientations so that all five are similar and then different, and notice how the positions of the poles change.

A pole figure for a polycrystalline aggregate, which shows completely random orientation, does not necessarily appear as might naively be expected. Angular distortions inherent in the stereographic projection result in the accumulation of points close to the centre of the pole figure as shown in the image below.



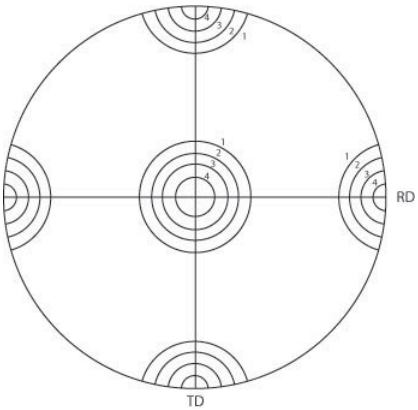
If the material shows a degree of texture, the resultant pole figure will show the accumulation of poles about specific directions.



A 100 pole figure showing “cube” texture – the {100} poles of the crystallites are oriented so that they are aligned with the axes defined by the rolling, transverse, and normal directions.

A single crystal can be plotted on the pole figure and there is no ambiguity regarding its orientation. However, as more crystallite poles are plotted onto the pole figure, the specific orientation of a particular crystallite can no longer be defined.

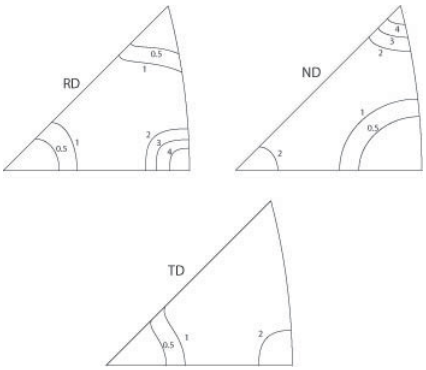
For a large number of grains in a polycrystal, poles may overlap on the pole figure, so that the true orientation density is not clearly represented. In this case, contours tend to be used instead. Regions of high pole density have a high number of contours, while regions with low pole density have a few, greatly spaced contours.



100 pole figure showing “cube” texture and pole density represented using contours rather than discrete points.

Inverse pole figures

Instead of plotting crystal orientations with respect to an external frame of reference, inverse pole figures can be produced which show the rolling, transverse, and normal directions (RD, TD and ND respectively) with respect to the crystallographic axes. Typically, these are plotted on a *standard stereographic triangle* as shown below



Crystal Orientation Distribution Function (CODF)

As previously mentioned, pole figures do not give information about the orientation of a particular crystal relative to another. More information can be gathered from a CODF. CODFs are constructed by combining the data from several pole figures. This requires intensive use of mathematics. More details can be found in the references listed in the Going further section.

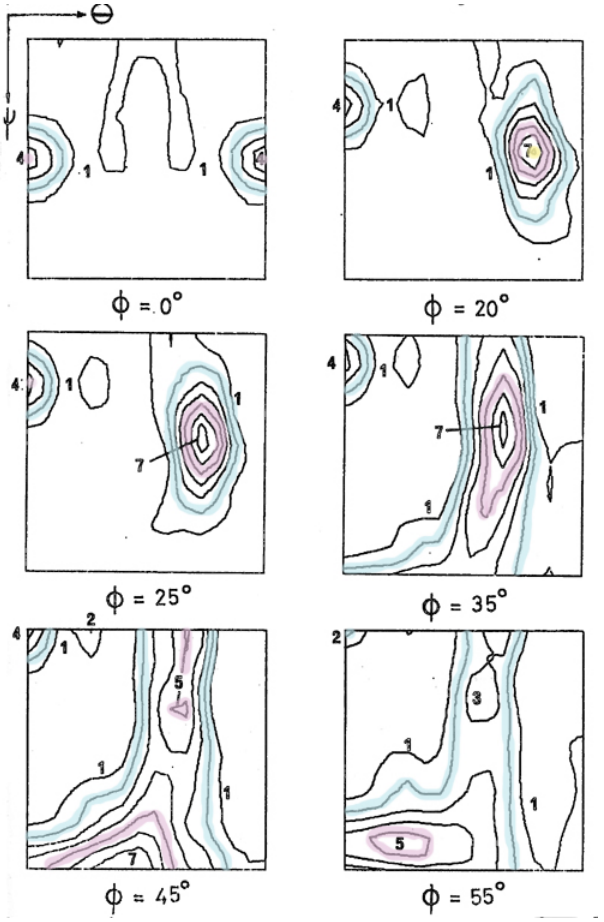
CODFs describe the orientation of each crystal relative to three *Euler angles* (φ,ψ, and θ). The Euler angles define the difference in orientation between the crystal axes and the deformation axes (i.e. the RD, the ND and the TD).

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

One convention for Euler angles (and the convention described here) is known as the Roe convention. An alternative convention can be used where the θ-rotation occurs about the x1 direction; this is known as the Bunge convention. These two conventions are related by:

$\psi_{Roe} = \psi_{1,Bunge} - \pi/2$
$\theta_{Roe} = \varphi_{Bunge}$
$\varphi_{Roe} = \psi_{2,\ Bunge} + \pi/2$

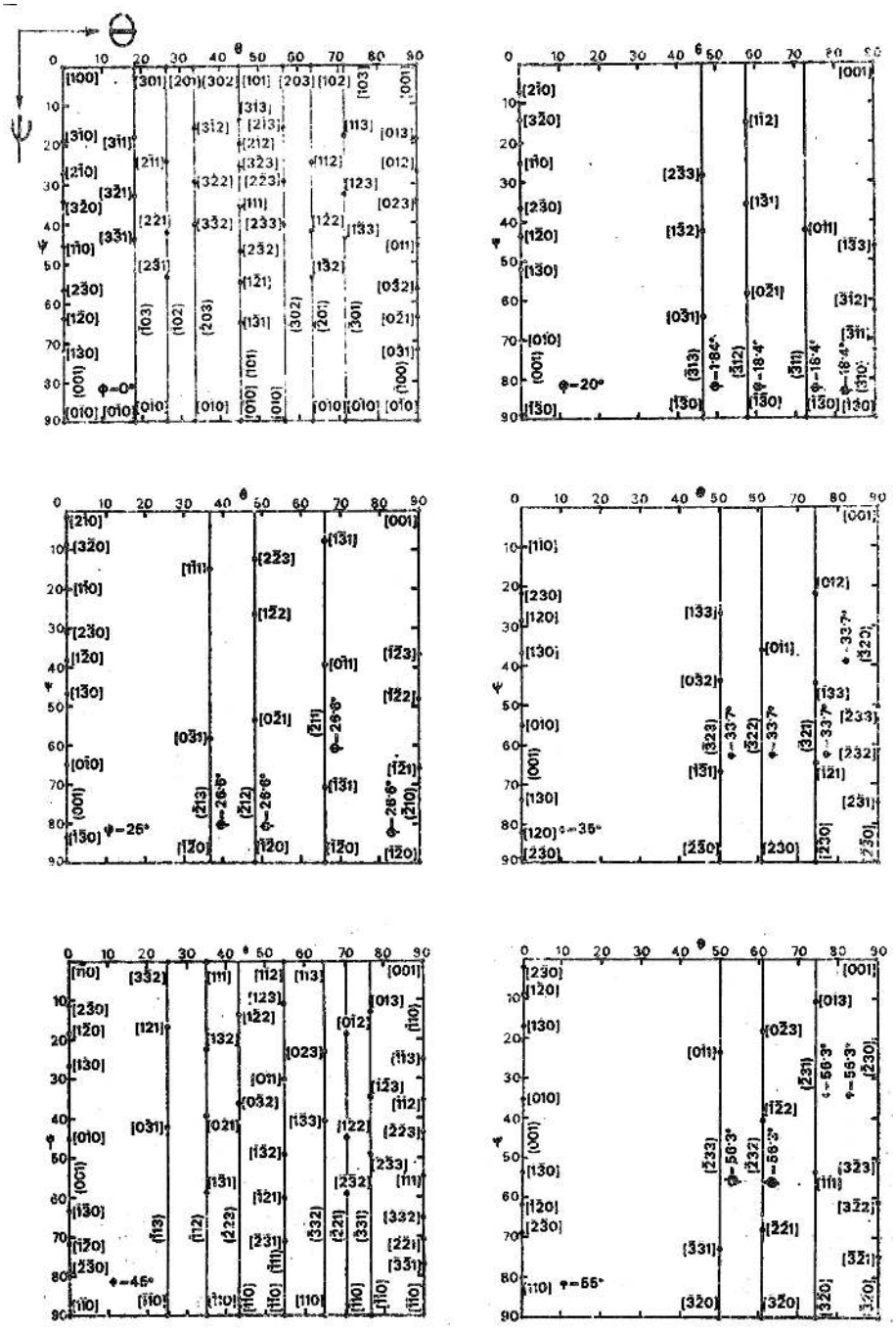
A single crystal is completely described by a point in a cube with axes of φ, ψ and θ. This cube is referred to as *Euler space* and is often shown as a series of cross-sections.



These sections of a CODF are for a steel sheet cold-rolled to an 80% reduction in thickness. Any values of *φ* between 0° and 90° can be used to produce the sections. In the image above, the values of φ that have been chosen are: 0°, 20°, 25°, 35°, 45°, and 55°. The contours of the sections make up part of a three-dimensional surface in Euler space as seen in the following animation. The highlighted contours on the sections correspond to the similarly coloured surface in the 3D plot below. The area of highest density and hence strongest texture is bound by the yellow surface centred around φ = 26.6°, ψ = 39.2°, and θ = 65.9°. *The original figure can be found in D.J. Goodwill Ph.D. thesis, University of Cambridge (1972) – The relationship between texture and properties of steel sheets.*

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

Texture diagrams such as those produced by Davies, Goodwill, and Kallend (1971) (See Going further) can be used to identify the texture present. In this case it is (211)[011]. (211) describes a plane which is orientated parallel to the plane of the sheet, while [011] is a direction parallel to the rolling direction.



This page titled 14.3: Representing Texture is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

14.4: Origins of Texture

Texture can arise whenever there is a preferred crystallographic orientation within a polycrystalline material.

Solidification

Directional solidification leads to texture when columnar grains grow in a preferred crystallographic direction in the heat flow direction. Crystals with a fast growth direction parallel to that of the heat flow will dominate the final structure. For more details see: Microstructure and segregation in castings in the Casting TLP

The preferred growth direction of dendrites in cubic metals is <100>. In the animation above it can be seen that close to the chilled mould texture is random, while at great distances from the mould, <100> fibre texture is prominent. This variation in texture occurs because nucleation at the mould wall is a random process with respect to orientation, whereas growth of the grains is strongly dependent on the orientation.

Mechanical deformation

The effect of slip during deformation is described in The stereographic projection TLP and in Slip in single crystals TLP. Slip planes of individual crystals will rotate, so that the direction of slip will rotate towards the tensile axis. If the axis along which force is applied is constrained, so that it does not deviate from its original direction, then the crystal must rotate with respect to the axis. In a polycrystalline sample, an originally random distribution of crystal orientation will become non-random and tend towards an *ideal orientation*. The ideal orientation depends upon a number of factors including the crystal structure of the material, its temperature, the alloying additions, and the processing route.

The effect of wire drawing, rolling, and annealing on the texture of c.c.p. and b.c.c. metals can be investigated in the flowing animation.

Thin films

Pronounced texture is very common in thin films. Even when there is no crystallographic matching between the crystal structure of the growing film and the substrate, the grains in the film show a strong preference for a particular plane to be parallel to the substrate surface. For example, films of c.c.p. metals prefer to have a close packed plane parallel to the surface of the substrate.

This page titled 14.4: Origins of Texture is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

14.5: Effect of Texture

Mechanical properties

The effects of anisotropy can either cause problems for the mechanical properties of metals, or they can be exploited to the benefit of the manufacturer.

The phenomenon of “earing” in deep drawing, where a wavy edge forms on the top of a drawn cup, allows the effect of texture to be seen easily. Depending on the degree of preferred orientation, two, four or six ears will form and extensive trimming is required to produce a uniform top. However, in beverage can and automobile body manufacture, plastic instability is avoided by controlling the texture in the thickness direction. This allows for very thin sheets to be produced without fracture.

Magnetic properties

In transformers, the texture of grain-oriented silicon steel (GOSS steel) is controlled to minimise core loss.Two textures can be developed in silicon steels: the cube on face or {001}<100> texture in which a {001} plane lies in the sheet plane, or the cube on edge {011}<100> texture (also known as the Goss texture, after its discoverer, Norman P. Goss) in which a {011} plane lies in the sheet plane. The Goss texture arises initially as a small component during rolling, but forms large grains during recrystallisation. By promoting the development of the Goss texture, the magnetic flux density can be increased by up to 30% relative to steel without this texture.

This page titled 14.5: Effect of Texture is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

14.6: Summary

Texture describes the orientation distribution of crystals within a polycrystalline aggregate. This can be measured using x-ray diffraction, EBSD, and neutron diffraction and represented in the form of pole diagrams, inverse pole diagrams, and crystal orientation distribution functions.

Texture can arise from processes such as solidification, mechanical deformation, annealing, and in thin films. The presence of texture may be problematic, but if the textures present and their effects are understood, then it can be exploited to great benefit.

Going further

Books

A readable book covering measurement and analysis of textures as well as modelling the development of texture during plastic deformation

- *Texture and Anisotropy – Preferred orientations in polycrystals and their effect on materials properties* by U.F Kocks, C.N Tomé and H.-R Wenk, Cambridge University Press, Cambridge 1998

This book is a good introduction to crystallography and the basics of texture.

- *Crystallography and Crystal Defects*, 2nd edition, A. Kelly and K.M. Knowles, John Wiley and Sons, Chichester, 2012

This book covers the mathematical theory behind texture analysis including the construction of CODFs

- *Texture Analysis in Materials Science – Mathematical methods* by H. J. Bunge, Butterworths & Co. 1982

Original papers containing charts to determine ideal orientations from crystallite orientation distribution functions

- *Charts for analysing crystallite distribution function plots for cubic materials* by G.J Davies, D.J Goodwill, and J.S Kallend (*J. Appl. Cryst.* **4** (1971) 67-70)
- *Charts for analysing crystallite orientation distribution function plots for hexagonal materials* by G.J Davies, D.J Goodwill, and J.S Kallend (*J. Appl. Cryst.* **4** (1971) 193-196)

Websites

- An introduction to the representation of texture in pole figures and CODF aluminium.matter.org.uk/content/html/eng/default.asp?catid=100&pageid=1039432491

This page titled 14.6: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

14.7: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

What is crystallographic texture?

a	The distribution of orientations of crystallites within a polycrystalline sample.
b	The orientation of a polycrystalline sample.
c	The distribution of orientations of crystallites within a single crystal.
d	The orientation of a surface of a sample.

Answer

A. The distribution of orientations of crystallites within a polycrystalline sample. A random texture describes a sample in which the crystallographic axes of the crystallites are oriented randomly, whereas a non-random distribution means that the material has crystallographic texture.

Which of these are limitations of pole figures? (Select all that apply)

a	Specific *hkl* planes cannot be plotted
b	Poles in the final plot for a polycrystalline material are not identified with particular crystals
c	Information about the crystallite location in a sample is absent
d	The orientation of a crystal must be described relative to another

Answer

B and C

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

A sample of rolled a-brass (ccp) has a very strong (110) [$1\bar{1}2$] texture. Using a Wulff net, sketch stereograms showing the distribution of 111 and 200 poles with respect to the normal to the sheet and the rolling direction.

Hint:

When plotting the stereogram, begin by plotting the rolling direction (RD) [$1\bar{1}2$] on the primitive of the stereogram and the normal direction (ND) [110] at the centre.
Remember that rolled sheets always have *mmm* symmetry

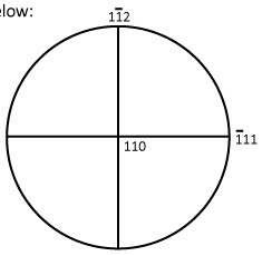
Answer

Determination of 111 pole figure for α-brass with strong (110)[1$\bar{1}$2] texture

Rolling direction is: [1$\bar{1}$2]

Plane parallel to the plane of the sheet is the plane (110). As α-brass is cubic, the normal direction is: [110]

These directions will be located on the stereogram as shown below:



The cross product of [1$\bar{1}$2] and [110] would give the direction perpendicular to these two. This direction is [$\bar{1}$11].

The angle between the 111 pole and the [1$\bar{1}$2] direction can be calculated using the dot product as the α-brass is ccp:

$$\frac{(111)\cdot(1\bar{1}2)}{|111||1\bar{1}2|} = \cos\theta$$

$$\theta = 61.8°$$

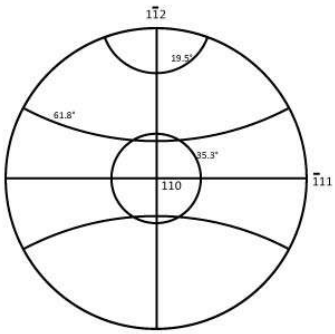
The angles between the 111 type poles and the rolling direction are similarly calculated:

($\bar{1}$11)	θ = 90°
(1$\bar{1}$1)	θ = 19.5°
(11$\bar{1}$)	θ = 118.1°

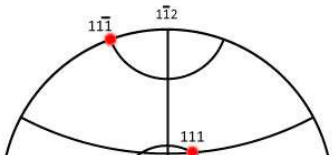
The angles between the same 111 poles and the normal direction [110] are calculated:

(111)	θ = 35.3°
($\bar{1}$11)	θ = 90°
(1$\bar{1}$1)	θ = 90°
(11$\bar{1}$)	θ = 35.3°

Using a Wulff net, the small circles representing these angles are plotted onto the stereogram:

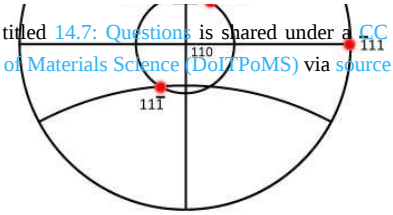


The intersection of the small circles give the position of the poles and hence shows where the regions of high pole density will appear on the pole figure:



https://eng.libretexts.org/@go/page/31576

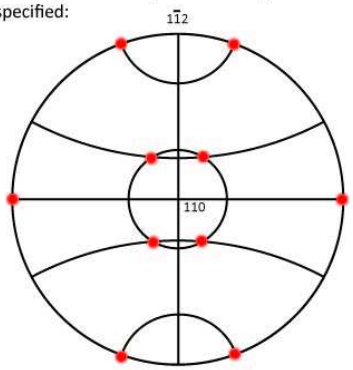
This page titled 14.7: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.



111 and $11\bar{1}$ lie on opposite ends of the diameter of the circle with angular radius 35.3° with 110 between them

Since there is an angle of 109.4° between $1\bar{1}1$ and $\bar{1}11$, the two poles must be located on opposite sides of $1\bar{1}2$ on the primitive

As rolling textures must have *mmm* symmetry with respect to RD, TD and ND the positions of maxima due to symetrical equivalent <11> poles can be specified:



This predicted pole figure is consistent with experimental results

CHAPTER OVERVIEW

15: Crystallography

www.doitpoms.ac.uk/tlplib/cr...phy3/index.php

Learning Objectives

On completion of this TLP you should:

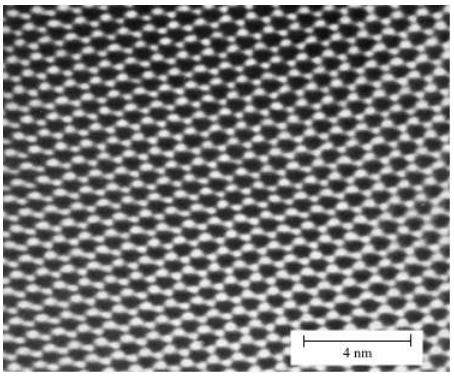
- Have an understanding of the basic concepts of crystallography, i.e. lattices, motifs, symmetry elements etc.
- Be able to identify lattices and symmetry elements within those lattices.
- Know about the different types of unit cell.
- Understand the idea of close-packing and packing efficiency.
- Be familiar with the different crystal systems and Bravais lattices.

15.1: Introduction
15.2: Lattices
15.3: Unit Cell
15.4: Lattice Geometry
15.5: Crystal Structure
15.6: Close Packing and Packing Efficiency
15.7: Symmetry
15.8: Combining symmetry
15.9: Crystal Systems
15.10: Summary
15.11: Questions

This page titled 15: Crystallography is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.1: Introduction

Crystalline materials are characterised by a regular atomic structure that repeats itself in all three dimensions. In other words the structure displays **translational symmetry**. Translational symmetry is illustrated in this image of the crystal structure of the mineral cordierite ($Mg_2Al_4Si_5O_{18}$) taken with a high resolution transmission electron microscope. The image is a projection through a very thin slice (~200 Å thick) of the atomic distribution. Black spots correspond to hollow channels through the structure and white spots correspond to regions of high electronic density, arranged around the channels in 6-fold rings (Scale: the distance between the black spots is ~ 9.7 Å). The structure consists of a simple group of atoms that repeats itself periodically in space. This periodicity can be revealed using the concept of a **lattice**.



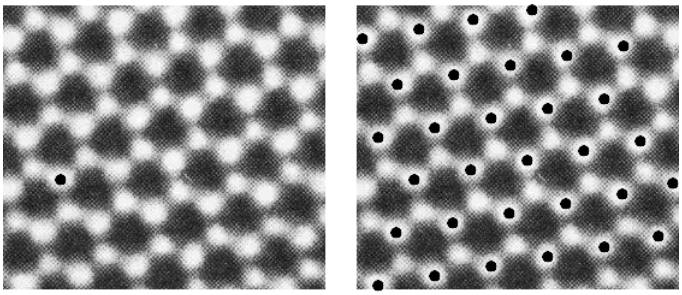
A. Putnis, *Introduction to Mineral Sciences*, Cambridge University Press, 1992 - frontispiece

This page titled 15.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.2: Lattices

Crystalline structures are characterised by a repeating pattern in three dimensions. The periodic nature of the structure can be represented using a lattice.

To generate the lattice from any repeating pattern, we choose an arbitrary reference point and examine its environment. We then simply mark in all the points in the pattern that are identical to the chosen reference point. The set of identical points is the lattice, and each point within it is a **lattice point**.



A. Putnis, *Introduction to Mineral Sciences*, Cambridge University Press, 1992

Note that not all white discs within this pattern are exactly equivalent, and therefore they are not all lattice points. The discs marked with a black spot have different arrangements around them than those that are unmarked (each is surrounded by 3 others in a triangle, but the orientation of the triangles is different).

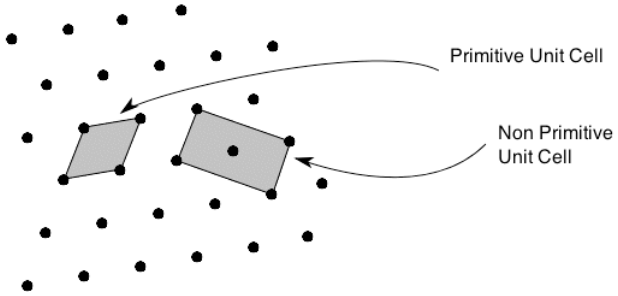
To practice identifying the lattice points within a more complex repeating pattern, try the following game! The brick pattern corresponds to an unusual style of Danish bricklaying where after each normal brick, the next is laid breadthwise,and so on. The original image was taken from the wall of the Centre for Electron Nanoscopy in Copenhagen (thanks to Dr Rafal Dunin-Borkowski).

This page titled 15.2: Lattices is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.3: Unit Cell

The structure of a crystal can be seen to be composed of a repeated element in three dimensions. This repeated element is known as the **unit cell**. It is the building block of the crystal structure. We define the unit cell in terms of the lattice (set of identical points). In three dimensions the unit cell is any parallelepiped whose vertices are lattice points, in two dimensions it is any parallelogram whose vertices are lattice points.

Of course this definition means that there are an infinite number of possible unit cells. So, in general, the unit cell is chosen such that it is the smallest unit cell that reflects the symmetry of the structure. There are two distinct types of unit cell: **primitive** and **non-primitive**. Primitive unit cells contain only one lattice point, which is made up from the lattice points at each of the corners. Non-primitive unit cells contain additional lattice points, either on a face of the unit cell or within the unit cell, and so have more than one lattice point per unit cell.

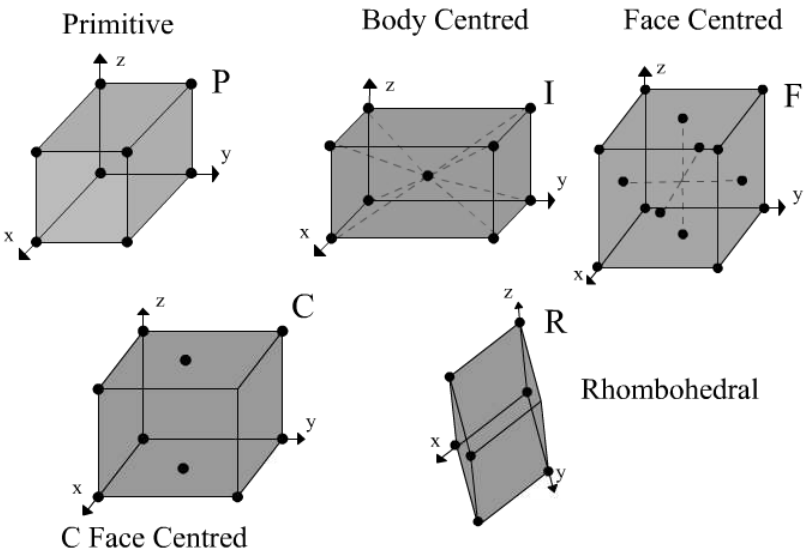


It is often the case that a primitive unit cell will not reflect the symmetry of the crystal structure. A suitable non-primitive unit cell will be picked in such cases.

It was mentioned above that the (eight) lattice points at the corners of the unit cell contribute only one lattice point to the cell. This is because the lattice points at the corners are shared between eight unit cells. Each corner lattice point therefore is equivalent to 1/8 of a lattice point per unit cell. Similarly lattice points on the edge of a unit cell are shared among four unit cells and are worth 1/4 of

a lattice point per unit cell. Lattice points on the face of a unit cell are shared between two unit cells and are worth 1/2 of a lattice point per unit cell. Lattice points contained entirely within the unit cell are worth one lattice point per unit cell.

The most common types of unit cell are the **primitive**(P) unit cell with one lattice point per unit cell; the **face centred**(F) unit cell with additional lattice points **at the centre of each face** and four lattice points per unit cell; and the **body centred**(I) unit cell with a lattice point in the middle of the unit cell and two lattice points per unit cell. Other cell types are the C face centred unit cell and the rhombohedral unit cell.



This page titled 15.3: Unit Cell is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.4: Lattice Geometry

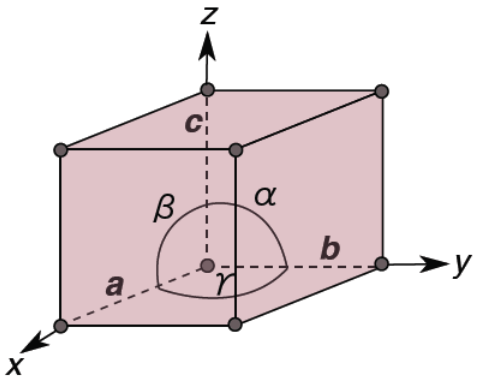
To define the geometry of the unit cell in 3 dimensions we choose a right-handed set of crystallographic axes, x, y, and z, which point along the edges of the unit cell. The origin of our coordinate system is at one of the lattice points.

Lattice parameters

The length of the unit cell along the x, y, and z direction are defined as a, b, and c. Alternatively, we can think of the sides of the unit cell in terms of vectors **a**, **b**, and **c**. The angles between the crystallographic axes are defined by:

α = the angle between **b** and **c**
β = the angle between **a** and **c**
γ = the angle between **a** and **b**

a, b, c, α, β, γ are collectively known as the **lattice parameters** (often also called 'unit cell parameters', or just 'cell parameters').



Lattice vectors

A lattice vector is a vector joining any two lattice points. Any lattice vector can be written as a linear combination of the unit cell vectors **a**, **b**, and **c**:

t = U **a** + V **b** + W **c**

In shorthand, lattice vectors are written in the form:

t = [UVW]

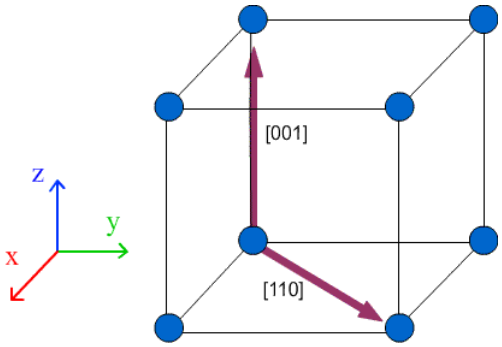
Negative values are not prefixed with a minus sign. Instead a bar is placed above the number to denote that the value is negative:

t = −U **a** + V **b** − W **c**

This lattice vector would be written in the form:

t = [UVW]

Lattice directions are written the same way as lattice vectors, in the form [UVW]. The direction in which the lattice vector is pointing is the lattice direction. The difference between lattice directions and lattice vectors is that a lattice vector has a magnitude which can be shown by prefixing the lattice vector with a constant. By convention U, V and W are integers.



Many crystal systems have elements of symmetry. In these systems, certain sets of directions are symmetrically equivalent to each other. The set of directions that are symmetrically related to the direction [UVW] are written <UVW>.

This page titled 15.4: Lattice Geometry is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.5: Crystal Structure

The structure of a crystal can be described by combining the following elements: the lattice type, the lattice parameters, and the motif.

The **lattice type** defines the location of the lattice points within the unit cell.

The **lattice parameters** define the size and shape of the unit cell.

The **motif** is a list of the atoms associated with each lattice point, along with their fractional coordinates relative to the lattice point. Since each lattice point is, by definition, identical, if we add the motif to each lattice point, we will generate the entire structure:

Plan view

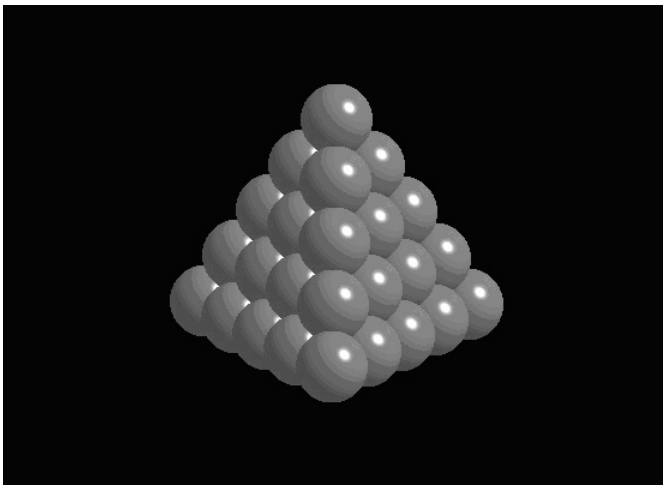
Knowing the motif and lattice it is possible to construct a **Plan view** of the crystal structure. The **Plan view** is the standard representation of a crystal structure and is very easy to produce. It is generally the 2D projection looking down the [001]/z-axis of the unit cell. Note this is equivalent to constructing a projection on the (001) plane. Refer to Lattice Planes and Miller Indices TLP for information on lattice planes.

The Plan view generally displays a 2×2 array of unit cells. The heights of the atoms within the unit cell are represented by fractions next to them, the fraction indicating that atom's fractional height in terms of the unit cell height (c) (atoms at the top and bottom of the unit cell have no numbers next to them). On constructing the plan view it is essential to not only indicate heights of atoms within the unit cell but also define the crystallographic axes you are using along with tracing out the unit cell.

This page titled 15.5: Crystal Structure is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

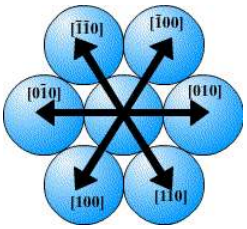
15.6: Close Packing and Packing Efficiency

In many cases the atoms of a crystal pack together as tightly as possible. Approximating atoms as hard spheres they will achieve this by forming a **close-packed** structure. This is the case for most metallic structures.

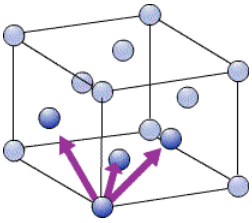


The main ideas of close packing are demonstrated in the animation below

In a close-packed structure the **close packed directions** are the directions in which atoms are touching. For a **hcp** structure the close packed directions are [100], [010] and [110] and their negatives. Directions that are related by symmetry are represented using the notation <UVW>. The close packed directions for **hcp** are then <100>.



The close packed directions for **ccp**, which has a **fcc** unit cell, are along the diagonals of each face, [110], [101], [011]… etc. The set of directions that are related to these by symmetry are the <110> set.



Packing Efficiency

The packing efficiency of a crystal structure tells us how much of the available space is being occupied by atoms. It is usually represented by a percentage or volume fraction.

The packing efficiency is given by the following equation:

The steps usually taken are:

- Calculate the volume of the unit cell
- Count how many atoms there are per unit cell
- Calculate the volume of a single atom and multiply by the number of atoms in the unit cell
- Divide this result by the volume of the unit cell

The steps are straightforward. The main source of difficulty is expressing the volume of the unit cell in terms of the radii of the atoms (or vice versa). Knowing the close-packed directions makes this step easier for us. The animation below demonstrates how to calculate the packing efficiency of hcp, ccp and bcc structures.

Note: If you know the motif, an easy way to find the number of atoms per unit cell is to multiply the number of atoms in the motif by the number of lattice points in the unit cell.

This page titled 15.6: Close Packing and Packing Efficiency is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

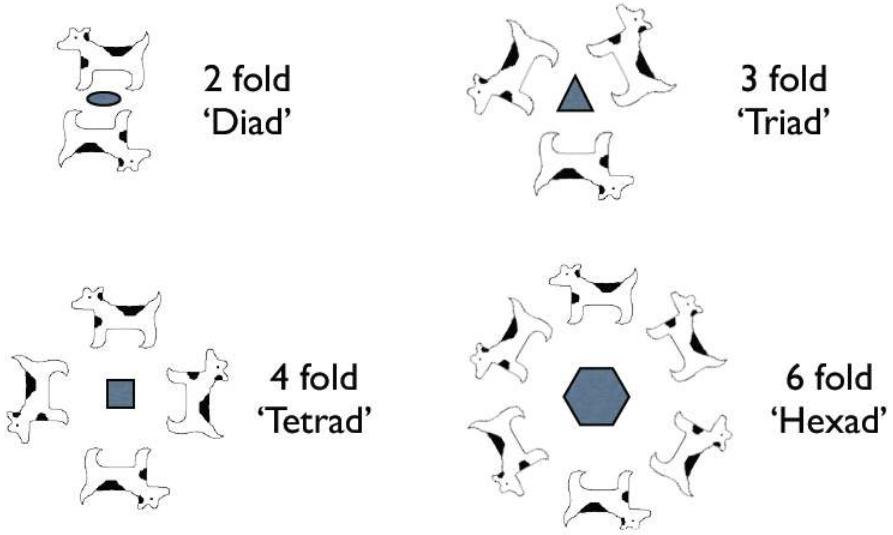
15.7: Symmetry

We have already met the concept symmetry in relation to crystal structures: the lattice generates the translational symmetry—the motif is repeated on every lattice point.

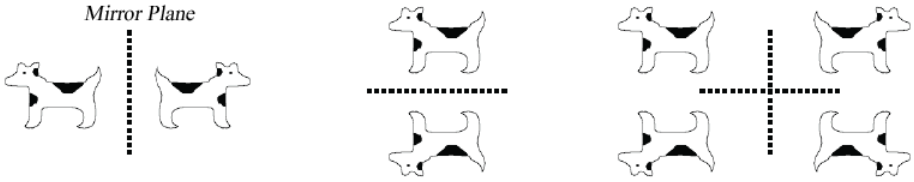
Other types of symmetry exist, including:

- rotation axes
- mirror planes
- centre of symmetry
- inversion axes (combination of rotation and centre of symmetry operations)

An n fold rotational symmetry operation rotates an object by 360°/n. Only n = 1, 2, 3, 4, and 6 are permitted in a periodic lattice



An object has mirror symmetry if reflection of the object in a plane brings it into coincidence with itself:



Some objects have special symmetry about an origin such that, for any point at position *x*, *y*, *z*, there is an exactly similar point at –*x*, –*y*, –*z*. The origin is called a centre of symmetry ("inversion centre"). Such an object is said to be centrosymmetric:

An n-fold inversion axis is a combination of a rotation by 360/n followed by a centre of symmetry operation. An example of a 4-fold inversion axis is show in the following animation:

https://eng.libretexts.org/@go/page/31577

This page titled 15.7: Symmetry is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.8: Combining symmetry

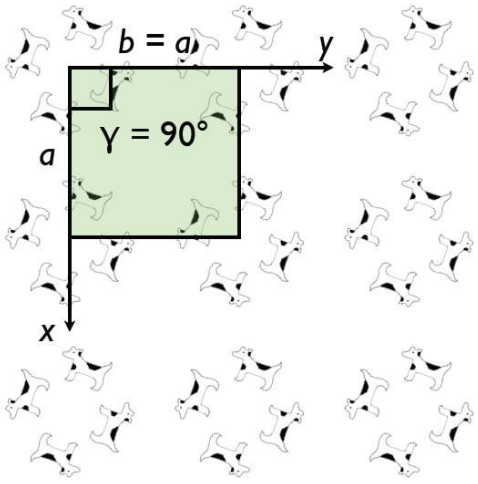
Only certain combinations of symmetry operation can exist in a crystal structure. This is because one symmetry element operating on another will generate a third symmetry element in the structure and this can end up generating an infinite number of symmetry elements, as shown in the animation below:

In fact, there are only 32 permitted combinations of mirror planes, rotation axes, centres of symmetry and inversion axes. These are known as the 32 **point groups**. Each point group is a finite set of mutually compatible symmetry elements. When the symmetry elements of a point group are operated on each other, they simply generate one of the other elements within the group.

This page titled 15.8: Combining symmetry is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

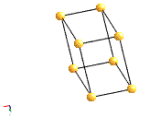
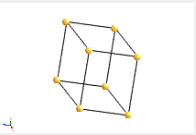
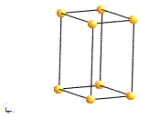
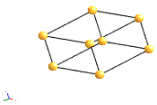
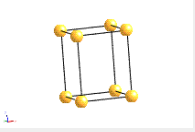
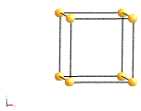
15.9: Crystal Systems

The rotational symmetry of a crystal places constraints on the shape of the conventional unit cell we choose to describe the structure. On this basis we divide all structures into one of 7 crystal systems. For example, for crystals with 4 fold symmetry it will always be possible to choose a unit cell that has a square base with $a = b$ and $\gamma = 90°$:

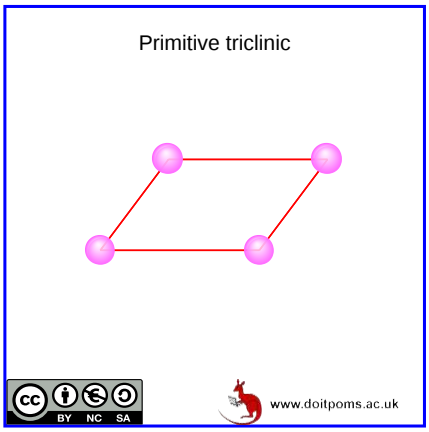


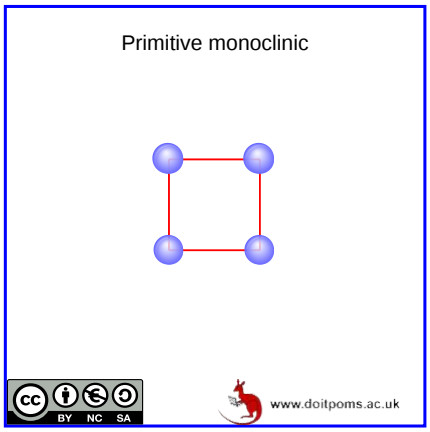
There are 14 unique combinations of the 7 crystal systems with the possible types of primitive and non-primitive lattices. These are referred to as the 14 Bravais lattices.

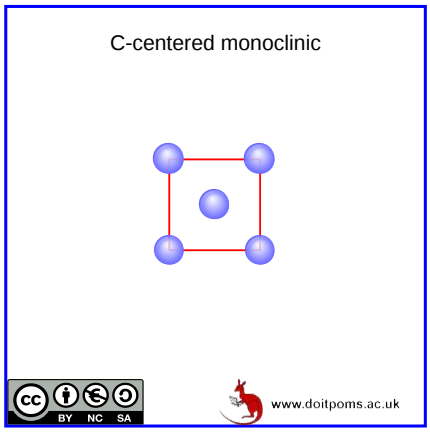
Crystal systems, lattices and symmetry elements

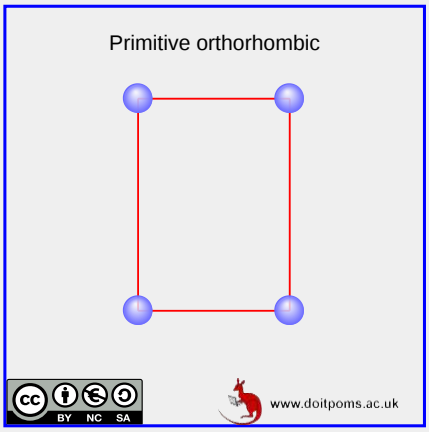
Crystal System	Defining Symmetry	Unit Cell Geometry	
Triclinic	Translational Only	a≠b≠c; α≠β≠γ	
Monoclinic	A diad axis (parallel to [010])	a≠b≠c; α=γ=90°; β>90°	
Orthorhombic	3 diads (each should be parallel to each axis)	a≠b≠c; α=β=γ=90°	
Trigonal For more information click here	1 triad (parallel to [001])	a=b≠c; α=β=90°; γ=120° **(or** a=b=c; 120° > α=β=γ ≠ 90°)	
Hexagonal	1 hexad (parallel to [001])	a=b≠c; α=β=90°; γ=120°	
Tetragonal	One tetrad (parallel to the [001] vector)	a=b≠c; α=β=γ=90°	
Cubic	4 triads (all parallel to <111> directions)	a=b=c; α=β=γ=90°	

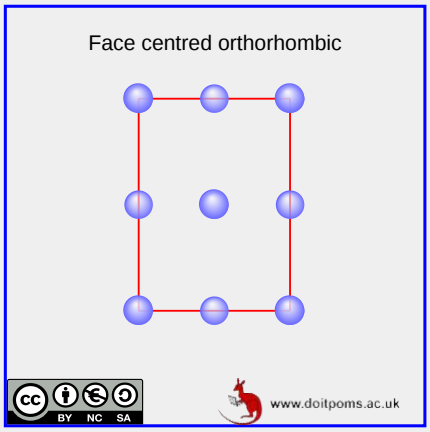
https://eng.libretexts.org/@go/page/31579

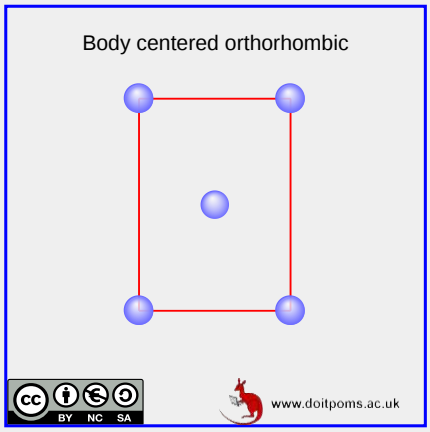

Primitive triclinic
www.doitpoms.ac.uk

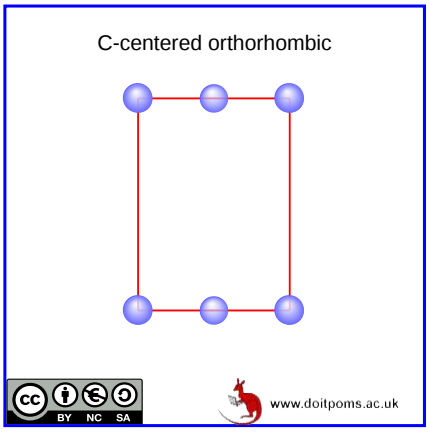

Primitive monoclinic
www.doitpoms.ac.uk

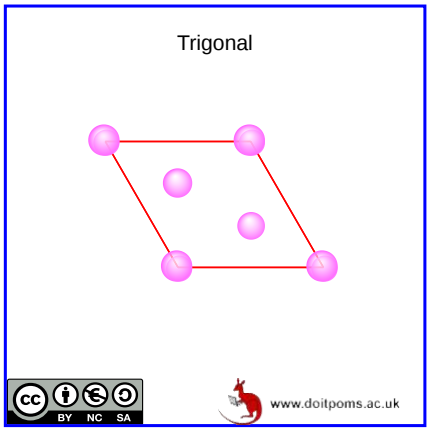

C-centered monoclinic
www.doitpoms.ac.uk

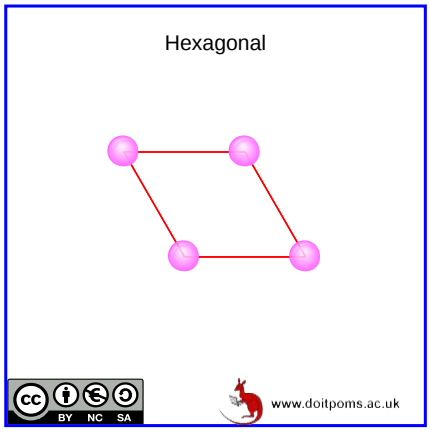

Primitive orthorhombic
www.doitpoms.ac.uk

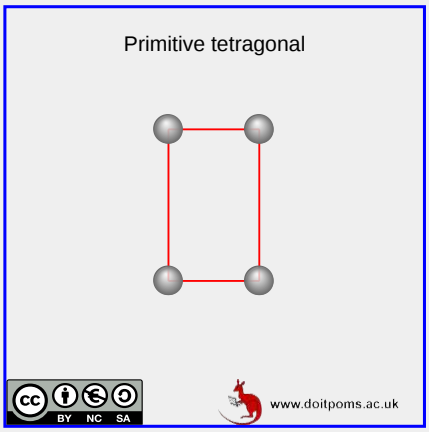

Face centred orthorhombic
www.doitpoms.ac.uk

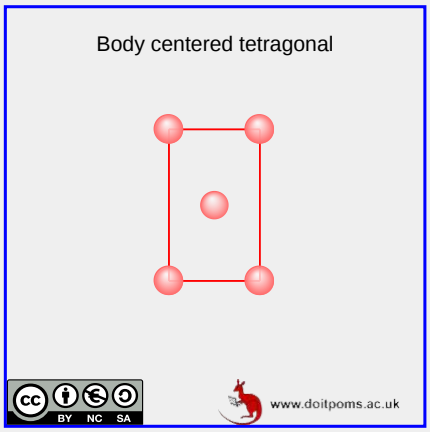

Body centered orthorhombic
www.doitpoms.ac.uk

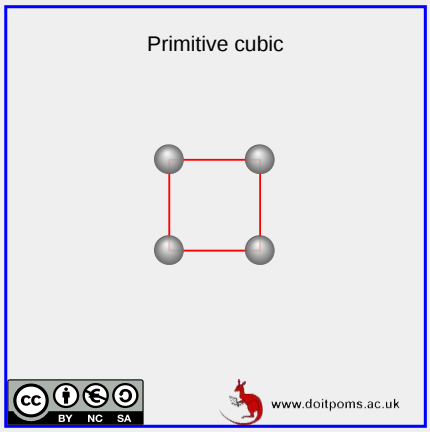

C-centered orthorhombic
www.doitpoms.ac.uk

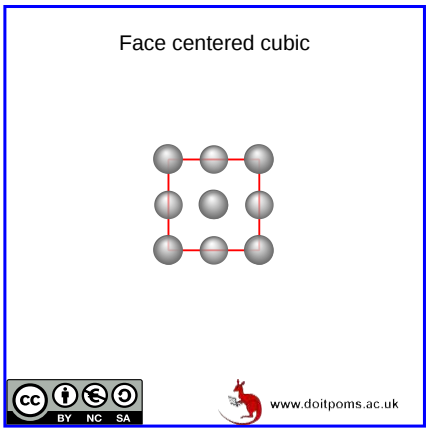

Trigonal
www.doitpoms.ac.uk

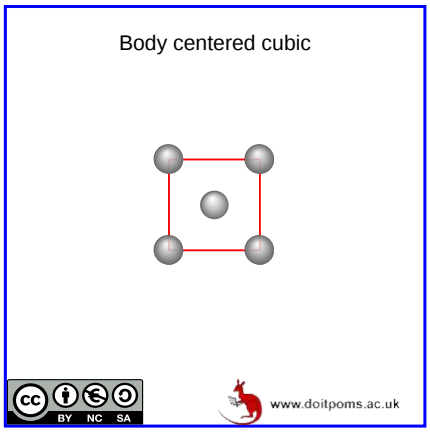

Hexagonal
www.doitpoms.ac.uk





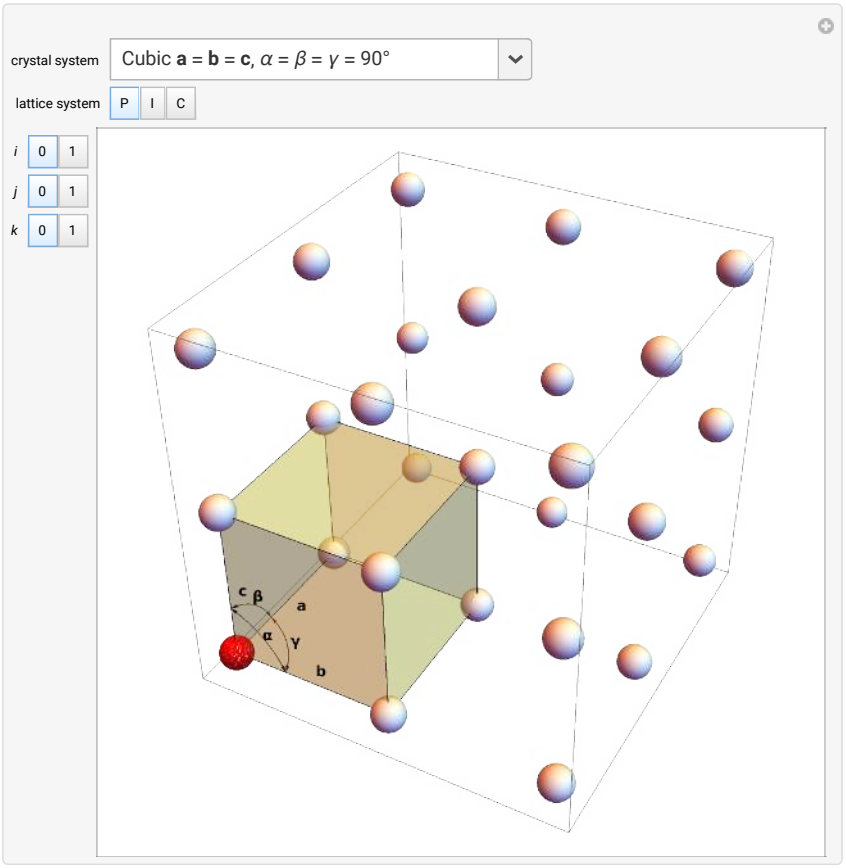






And you can use the Wolfram Demonstration Project Viewer to look at the Bravais lattices

The 14 3D Bravais Lattices



This Demonstration shows the characteristics of 3D Bravais lattices arranged according to seven crystal systems: cubic, tetragonal, orthorhombic, monoclinic, triclinic, rhombohedral and hexagonal. Each crystal system can be further associated with between one and four lattices by adding to the primitive cell (click "P"): a point in the center of the cell volume (click "I"), a point at the center of each face (click "F") or a point just at the center of the base faces (click "C"). The points located at the center/faces are highlighted in blue; each point is also a vertex or center of the cell/face, therefore each point is equivalent to every other point.

Crystal systems are determined by the relative lengths of the basis vectors ***a***, ***b***, ***c*** and the angles (α, β, γ) between them [1].

It is possible to shift the cell by one unit along a basis vector by selecting the *i*, *j*, *k* values. When repeated, this can generate the entire lattice.

Get this Notebook
Powered by Wolfram

This page titled 15.9: Crystal Systems is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.10: Summary

This TLP has covered the following topics:

- The mathematical description of the atomic structure of a crystal.
- The different lattice types.
- What a lattice vector, unit cell & motif are and how they are relevant to describing crystals.
- Close packing and the packing efficiency of a crystal structure.
- What symmetry operators are present within a crystallographic lattice and how only specific combinations of such symmetry elements may exist.
- Bravais lattice and Point Groups.
- How to construct the plan view of a 2×2 unit cell.

Going further

Books

- A. Putnis, *Introduction to Mineral Sciences*, Cambridge University Press, 1992
- Feynman, Leighton, Sands, *Lectures on Physics Vol. II, Chapter 30*, Addison-Wesley Publishing Company, 1964 [Note: This book follows a different convention when referring to body-centred and face-centred structures]
- C. Hammond, *The Basics of Crystallography and Crystallography*, III Ed, Oxford University Press, 2009
- B.D Cullity, S.R Stock, *Elements of X-Ray Diffraction*, Prentice Hall, 2001
- McKie, McKie, *Essentials of Crystallography*, Blackwell Scientific Publications, 1986

Website of **International Tables for Crystallography** http://it.iucr.org/A/

This page titled 15.10: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

15.11: Questions

https://eng.libretexts.org/@go/page/31582

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which of these is **not** a lattice type?

a	H
b	F
c	I
d	C

Answer

A.

What is a lattice ?

a	A transforming matrix
b	A set of atoms
c	A group of symmetry elements
d	An infinite array of identical points repeated throughout space

Answer

D.

How many Bravais lattices (3D) are there ?

a	36
b	28
c	14
d	16

Answer

C.

How many point groups (3D) are there ?

a	24
b	32
c	12
d	19

Answer

B.

What is a unit cell ?

a	A unit of volume
b	Any parallelepiped with lattice points at its corners
c	A parallelepiped containing only one lattice point
d	The angle between lattice vectors

Answer

B.

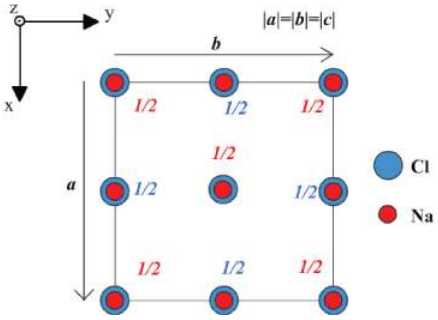
Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

Construct a plan view of NaCl (sodium chloride).
NaCl has a **face-centred cubic** lattice. The motif is:
Cl @ (0,0,0);
Na @ (0,0,1/2);
Note 1: The motif coordinates are positions relative to each lattice point
Note 2: In a face centred cubic structure the lattice points are located at:
(0,0,0), (1/2,1/2,0), (1/2,0,1/2), (0,1/2,1/2)

Answer

Your construction should look something like this

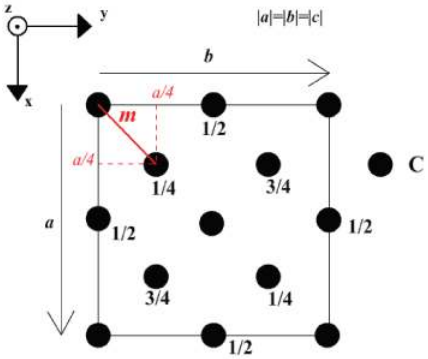


It should include:
A set of axes
The ***a*** and ***b*** lattice vectors
Two separate symbols for Cl and Na plus a key explaining the symbols
The fractional heights of the atoms.

Diamond has a **face centred cubic** lattice. Its motif is
C @ (0,0,0), (1/4,1/4,1/4)
Construct a plan view of the diamond unit cell.
Treating the carbon atoms as hard spheres calculate the packing efficiency of diamond.
Note: In a face centred cubic structure the lattice points are located at:
(0,0,0), (1/2,1/2,0), (1/2,0,1/2), (0,1/2,1/2)

Answer

Your construction should look something like this

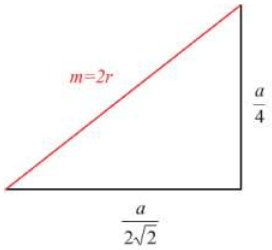


Packing Efficiency:
The volume of the unit cell is a^3.
The volume of a single atom is

$$\frac{4}{3}\pi r^3 \tag{15.11.1}$$

There are 8 atoms per unit cell.
Note that the solid red line extends to a height of a/4. The total length of m is equal to 2r (the centre-centre distance of two touching atoms is 2r). The distance covered by m in the xy plane is .

$$\frac{a}{2\sqrt{2}} \tag{15.11.2}$$



We can see from this diagram that

$$2r = \frac{\sqrt{3}}{4}a \tag{15.11.3}$$

therefore

$$r = \frac{\sqrt{3}}{8}a \tag{15.11.4}$$

The volume of a single atom is

$$\frac{4}{3}\pi\left(\frac{\sqrt{3}}{8}a\right)^3 = \frac{\sqrt{3}}{128}\pi a^3 \tag{15.11.5}$$

The volume occupied by the 8 atoms is

$$\frac{\sqrt{3}}{16}\pi a^3 \tag{15.11.6}$$

Therefore the packing efficiency is

$$\frac{\sqrt{3}}{16}\pi a^3 \div a^3 \approx 0.34 \tag{15.11.7}$$

$$\Rightarrow 34\% \tag{15.11.8}$$

This page titled 15.11: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

16: Deformation of honeycombs and foams

www.doitpoms.ac.uk/tlplib/de...tion/index.php

Learning Objectives

On completion of this TLP you should:

- Understand what processes determine the elastic behaviour of a honeycomb structure;
- Know what processes cause the onset of yielding;
- Understand how these ideas for simple structures might be extended to more complex ones, such as foams.

16.1: Introduction
16.2: Compression of a Honeycomb: Experimental
16.3: Elastic Behaviour
16.4: Yielding and Plateau Behaviour
16.5: Densification
16.6: Other Porous Structures
16.7: Porous Structures in Bending
16.8: Summary
16.9: Questions

This page titled 16: Deformation of honeycombs and foams is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.1: Introduction

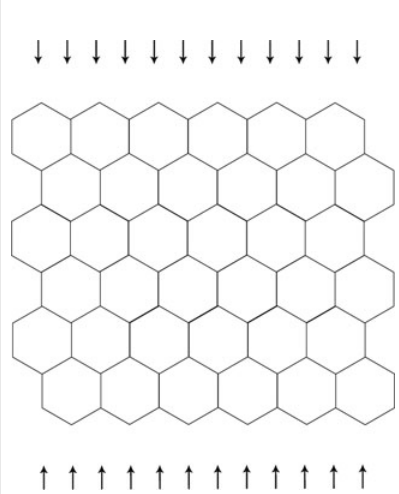
There is an important class of materials, many biological, that are highly porous and made by bonding rods, ribbons or fibres in both regular and irregular structures. These include paper, bone, wood, packaging foam and insulating fibre mats and they can be made of polymers, metals, ceramics and natural materials. Despite the different structures and materials, there are many similarities in how they behave. An important class of these materials is where the rods or ribbons form cells, co-called cellular structures. Here we explain how such structures deform in compression. To understand the deformation processes more easily the behaviour of a regular honeycomb structure is described before extending the ideas to structures such as foams.

This page titled 16.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

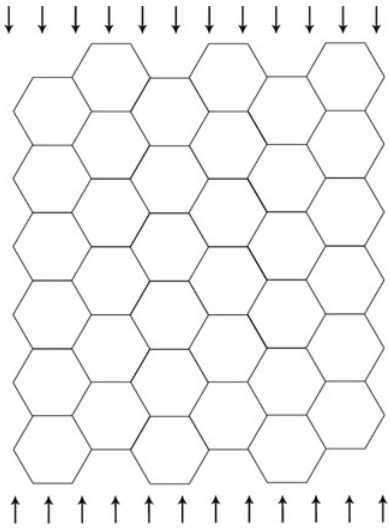
16.2: Compression of a Honeycomb: Experimental

The honeycomb studied here is an array of regular hexagonal cells, with the cell walls made of thin strips of aluminium. The structure is not quite as simple as it first appears because of the way in which it is made (see details). This causes one in every three cell walls to consist of two layers of aluminium bonded with adhesive, making some walls stiffer than others. In the honeycomb used here the thickness of an individual sheet, *t*, was 0.09 mm, and the length of each of the cell faces, *l*, was 6.30 mm. The relative density, the measured density as a fraction of the density of the solid material, is 0.008.

There are two different directions in which a hexagonal honeycomb can be compressed in the plane of the honeycomb.



Cell walls lie parallel to the loading axis



Cell walls lie diagonal to the loading axis

Although quantitatively different, the basic deformation processes are similar in both cases so only the situation where some cell walls lie parallel to the loading axis is described here.

Squares of honeycomb with 6 cells along each side were cut from a sheet of material. The samples were then compressed between flat, parallel platens at a constant displacement rate of 1 mm min^{-1}, giving the stress-strain curve below.

The stress-strain curve has 3 distinct regions:

- an initial *elastic region*;
- followed by the onset of irreversible leading to region where the stress does not change with increasing strain, known as the *plateau region*;
- and lastly a region where the stress again begins to rise rapidly with increasing strain, known as *densification*.

Elastic region

The elastic region is characterised by the effective Young modulus of the material, that is the Young modulus of a uniform material that for the same imposed stresses gives rise to the same strains. This is found by taking the slope of the unloading curve, which helps to reduce the effects of any local plastic flow.

The measured Young modulus was 435 kPa.

Plateau region

With continued loading the stresses in the faces increase. Eventually these reach the flow stress of the material and irreversible yielding begins. The stress at which this occurred was approximately 15 kPa. The structure then began to collapse at an approximately constant stress, σ_p.

Densification

This continued up to a strain of ~ 0.7, at which point the stress started to rise much more rapidly as the faces from opposite sides of the cells were pressed up against one another.

This three-stage deformation behaviour is typical of virtually all highly porous materials, even ones made of very brittle materials. The next step is to try and quantitatively describe the observed behaviour.

This page titled 16.2: Compression of a Honeycomb: Experimental is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.3: Elastic Behaviour

To start let us assume that the predominant contribution to the elastic strain comes from the axial compression of the vertical struts, as shown below.

We can estimate the magnitude of this strain as the cross-sectional area of solid material, A_V, in a cut across just the vertical faces is less than that if the material were completely solid by the ratio of the cell wall thickness, t, to the horizontal distance across each cell, 2 l cos θ.

As t << l cos θ, this is given by

$$A_V = t/2l\cos\theta \tag{16.3.1}$$

Using the measured values of t (= 0.09 mm) and l (= 6.30 mm), and taking θ = 30° as the cells are hexagonal gives A_V as 0.008. Taking the Young modulus of aluminium as 70 GPa, this predicts the Young modulus of the honeycomb to be 560 MPa. This is greater than the observed value of 435 kPa by more than 3 orders of magnitude and shows that axial compression of the vertical faces makes a negligible contribution to the elastic strain.

If the axially loaded faces do not deform, then clearly it must be those at an angle to the loading axis that do, the diagonal faces. And because they are at an angle to the loading axis they will bend.

The bending of each face beams must be symmetrical about the mid-point of the face and can be estimated using beam bending theory. To do this each face is described as two beams, cantilevered at the vertices of the hexagonal cell and loaded at the centre point. Note that one beam (i.e. a half cell wall) is pushed upward, the other downward.

This [derivation] gives the Young modulus of the honeycomb as

$$E = \frac{4}{\sqrt{3}} E_S \left(\frac{t}{l}\right)^3 \tag{16.3.2}$$

Using the measured values of t(= 0.09 mm) and l (= 6.30 mm) and taking E_S as 70 GPa, the elastic modulus is predicted to be 471 kPa. This is within 10% of the measured value for the sample.

The predominant contribution to the Young modulus is therefore from the bending of the diagonal faces.

This page titled 16.3: Elastic Behaviour is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.4: Yielding and Plateau Behaviour

The aluminium honeycomb will start to plastically deform if the stress in the faces anywhere exceeds the flow stress, σ_Y, of the aluminium cell wall. We have already shown that the predominant contribution to the elastic strain is the bending of the diagonal faces. (see here). Furthermore we could estimate this if each face were considered to be made up of two beams, each of length $l/2$, cantilevered at the end connected to the vertical cell wall and acted upon by a force of magnitude $F \cos \theta$, where F is the force applied at the ends of the sample and θ is the angle between the diagonal face and the horizontal. It is clear then that the stress will be a maximum where the moment is greatest, that is at the vertices of the hexagonal cells.

It can be shown that the applied stress, σ, when the maximum stress in each face reaches the flow stress, σ_Y, of the material making up the cell walls is given by (derivation)

$$\sigma = \frac{4}{9} \cdot \left(\frac{t}{l}\right)^2 \sigma_Y \tag{16.4.1}$$

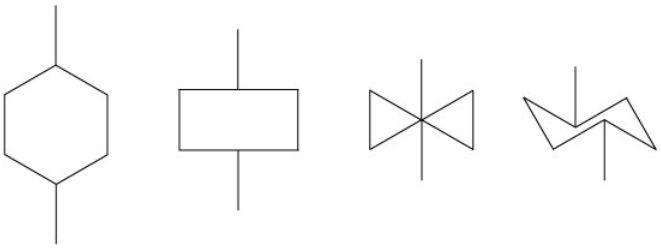
Using the measured values of t (= 0.09 mm), l (= 6.30 mm) and σ_Y (= 100 MPa), predicts the yield strength of the honeycomb, σ, to be 9 kPa, somewhat lower than the measured value of 15 kPa. The stress we have estimated is the stress at which plastic flow will start in the outer surfaces at the cantilevered point. To enable plastic flow to spread through the thickness of the cell face requires that the stress is increased further by a factor of 1.5, giving a macroscopic flow stress of 13.5 kPa, much closer to the measured value.

Once the material has started to yield the cell walls begin to collapse. This occurs at an approximately constant stress until the cell walls impinge on one another when the stress begins to rise more rapidly with increasing strain.

This page titled 16.4: Yielding and Plateau Behaviour is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.5: Densification

As the honeycomb yields in the plateau region, the regular hexagonal cell with a height (l + 2l sin θ) changes shape with the protruding apices being pressed toward one another to give cells with the shape of a bow-tie and a height l.



If the cells deform uniformly then the strain at which this occurs, ε_D, is given by

$$\varepsilon_D = \ln\left(\frac{l}{l+2l\sin\theta}\right) = \ln\left(\frac{1}{1+2\sin\theta}\right) \tag{16.5.1}$$

Note true strain is used because the strains are large and compressive. As θ = 30°, ε_D, is predicted to have a magnitude of 0.7. Further increases in strain cause opposing cell walls to be pressed against one another and the stress required for further deformation increases rapidly. As can be seen in the stress strain curve below this prediction gives good agreement with the observed stress-strain curve.

It can be seen that this collapse does not occur uniformly throughout the whole structure, but layer by layer of cells. This behaviour is rather dependent on the size of the cell compared to that of the sample. Increasing the number of cells in a cross-section causes the behaviour to become more uniform as might be expected.

It is now possible to quantitatively understand the entire stress-strain behaviour of a simple honeycomb. The next step is to extend these ideas to less regular structures, such as foams and fibrous structures.

This page titled 16.5: Densification is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.6: Other Porous Structures

Many other porous structures show the same type of stress-strain behaviour. The basic reasons are similar. The initial behaviour is elastic, until the stresses in the struts reach their flow or fracture stress. There is then a plateau region as the cells collapse, until the struts from opposite sides of the cells impinge on one another and the applied stress increases more rapidly. However the details can be very different. For instance the struts in ceramic foams tend to break, but a plateau region is still seen.

Many combinations of material and cell structure are possible. Seeing how the cells deform in a foam is more difficult than in the simple honeycomb. However this has been done using X-ray tomography as shown in the short video.

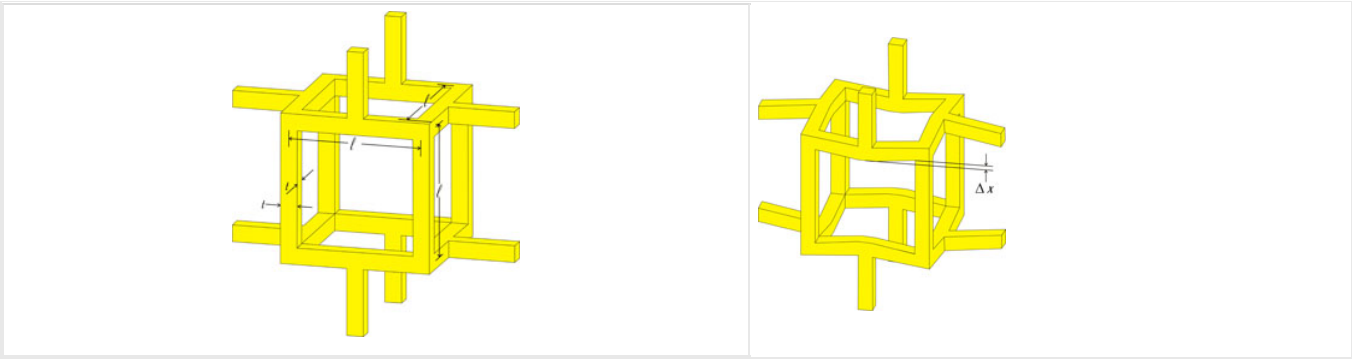
https://www.doitpoms.ac.uk/tlplib/de...s/compress.mp4

Deformation of cells in foam

(For further details see J.A. Elliott et al, "*In-situ* deformation of an open-cell flexible polyurethane foam characterised by 3D computed microtomography", *J. Mater. Sci.* **37** (2002) 1547-1555.)

Looking at the large cell on the right-hand side, it is clear that the deformation of the foam is similar to the honeycomb and the strain comes predominantly from the bending of the struts transverse to the loading axis.

For simplicity, consider the open-cell foam as having a cubic unit cell as shown below.

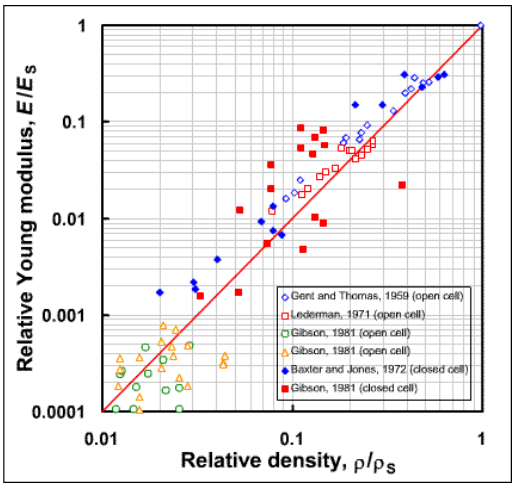


Note that each transverse strut has a vertical strut half-way along it, so that axial loading causes the struts transverse to the loading axis to bend, as shown in the diagram above. The Young modulus can now be estimated in a similar way to that for the honeycomb, except that the struts are assumed to have a square cross-section, rather than being rectangular as before and θ, the angle between the transverse strut and the horizontal is 0.

This gives an expression for the relative Young modulus, E/E_S, as

$$\frac{E}{E_S} = k\left(\frac{\rho}{\rho_S}\right)^2 \tag{16.6.1}$$

where E is the Young modulus of the porous material and E_S that of the solid material and k is a numerical constant, experimentally found to approximately equal to 1 (derivation).

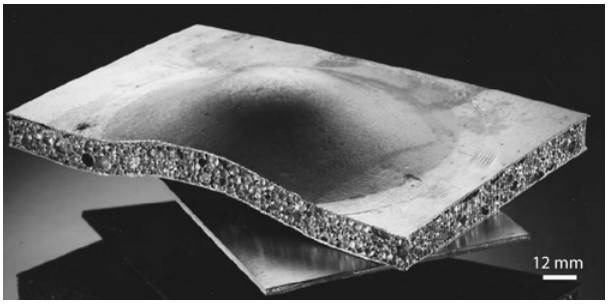


As can be seen in the graph above, experiments show this is correct for isotropic, open-cell foams and even appears to be obeyed where the struts are not slender beams and also, at least approximately, where the cells are closed rather than open. This is thought to arise because in most closed-cell foams most of the material is still along the edges of the cells, rather than being uniformly distributed across the faces. (The data is taken from various sources cited in L.J. Gibson and M.F. Ashby, "On the mechanics of three-dimensional cellular materials", *Proc. Roy. Soc. A*, **382**[1782] (1982) 43-59.)

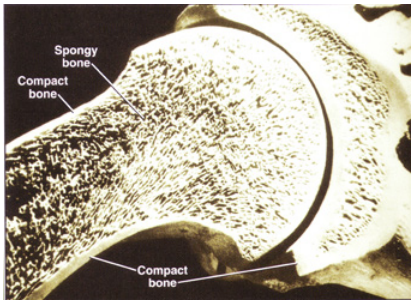
This page titled 16.6: Other Porous Structures is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.7: Porous Structures in Bending

Porous structures are often used as a lightweight core separating two strong, stiff outer layers to form a sandwich panel. Like the I-beam, such structures have a greater resistance to bending per unit weight of material than a solid beam and so are useful where weight-saving and stiffness are important. Typical applications include flooring panels in aircraft or rotors in helicopter blades. Sandwich structures are also common in biological structures, such as leaves or spongy bone.



From J. Banhart, Manufacture, Characterisation and Application of Cellular Metals and Metal Foams, *Progress Mater. Sci.*, 2001, 46, pp.559-632.



From *Cell Biology* by Thomas D. Pollard and William C. Earnshaw, Saunders 2004, pp.540 (Figure 34-4), courtesy of D.W. Fawcett, Harvard Medical School.

However some porous solids, such as wood are used without the stiff, strong outer layers. We might ask whether it is generally true, or under what conditions a porous rod will be stiffer in bending (i.e. give a smaller deflection for a given applied force) than a solid rod of the same length and overall mass.

Consider two rods one is porous and the other is solid. As each rod has the same length and mass and a circular cross-section, the porous one must have a larger radius.

The deflection, δ, of a cantilevered beam of length L under an imposed force W is given by

$$\delta = \frac{1}{3}\frac{WL^3}{EI} \tag{16.7.1}$$

For given values of W and L an increase in beam stiffness requires a higher value of the product EI. For a beam of circular cross-section <I = $\pi r^4/4$. The porous beam has a larger radius and therefore a larger moment of area than the solid beam. However the porous beam also has a lower Young modulus. For the porous beam to be stiffer in bending, the rate at which the moment of area increases with radius must therefore be greater than the rate at which the Young modulus decreases.

If the density of the porous beam is ρ and solid beam is ρ_S and both have the same length and mass, then the ratio of the radius of the porous beam, r, to that of the solid beam, r_S, is

$$\left(\frac{\rho_S}{\rho}\right) = \left(\frac{r}{r_S}\right)^2 \tag{16.7.2}$$

As $I \propto r^4$ the ratio of the second moments of area of the porous and solid beams, I and I_S respectively, is

$$\frac{I}{I_S} = \left(\frac{\rho_S}{\rho}\right)^2 \tag{16.7.3}$$

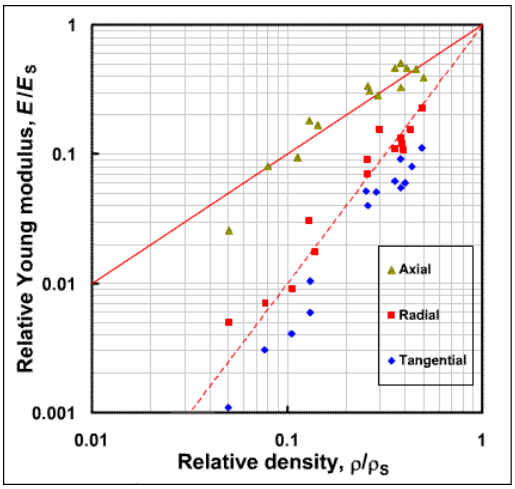
In other words I/I_S increases as the inverse square of the relative density, ρ/ρ_S.

Now the expression derived above for the elastic modulus of an open-cell porous body was

$$\frac{E}{E_S} = k\left(\frac{\rho}{\rho_s}\right)^2 \tag{16.7.4}$$

That is E/E_S decreases as the square of the relative density. In other words although I is increasing with decreasing density, E is decreasing at the same rate. In this case there would be no advantage in using such a material in bending compared with the solid material.

As nothing can be done about the change in radius, and hence I, with relative density, a higher bending stiffness can only be obtained by ensuring that E/E_S varies with ρ/ρ_S by a power less than 2. This is the case for the **axial** Young modulus of wood where the exponent lies closer to 1 rather than 2, as shown below.



(The data is taken from K.E. Easterling et al, "On the mechanics of balsa and other woods", *Proc. Roy. Soc. A*, **383**[1784] (1982) 31-41.) Such changes can be brought about by varying the cell structure, for instance by elongating the cells, as occurs in wood. However in the **transverse** (radial and tangential) directions E/E_S for wood decreases much more rapidly with decreasing ρ/ρ_S. Here the exponent lies between 2 and 3.

This page titled 16.7: Porous Structures in Bending is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.8: Summary

In this TLP, the elastic, yielding and densification behaviour of a simple honeycomb structure has been studied experimentally. It is shown that the deformation of a honeycomb structure is made up of 3 main regions: an elastic region, which ends when the maximum stress in the cell faces becomes equal to the flow stress of the material, allowing the cells compact at a constant stress, followed by a region in which the load rises rapidly with increasing strain, as the honeycomb is compacted.

Quantitative descriptions of the behaviour have been derived and compared with the experimental measurements. These show that the deformation behaviour of a honeycomb is determined not by the axial compression of those faces parallel to the loading axis, but by the bending of faces lying at some angle to the loading axis.

It has been shown that these ideas can be extended to describe the deformation behaviour of more irregular structures, such as foams. For foams that are isotropic and have open cells, it is predicted that the relative elastic modulus varies with the square of the relative density, consistent with observations in the literature.

The uses of such structures are described and it is shown that in isotropic open-cell foams, sandwich structures are required to obtain improved specific stiffness in bending. The enhanced stiffness of cellular structures such as wood arises from modifications to the structure that gives a different dependence of the relative stiffness on the elastic modulus.

Going further

Books

- L.J. Gibson and M.F. Ashby, *Cellular solids: structure and properties*, Cambridge University Press, 2nd edition (1997).
  Covers honeycombs and foams, both open and close celled, as well as the effects of gases and liquids in the cells. It also discusses the properties of bone, wood and the iris leaf as highly porous solids.
- K.K. Chawla, *Fibrous materials*, Cambridge University Press, 2nd edition (1998).
  Covers fibrous and some woven structures.
- D. Boal, *Mechanics of the Cell*, Cambridge University Press, 2002.
  See chapter 3 on two-dimensional networks.

Websites

- Properties of a chiral honeycomb with a Poisson's ratio -1
  Website on amusing elastic properties.

This page titled 16.8: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

16.9: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

How does elastic deformation of a honeycomb with hexagonal cells with some faces aligned parallel to the loading axis take place?

a	By elastic compression of the vertical faces
b	By elastic buckling of the vertical faces
c	By elastic bending of the diagonal faces

Answer

C.

What is the criterion for the onset of yielding in a honeycomb?

a	That somewhere the stress should exceed the flow stress through the thickness of the film
b	That the maximum stress in the face exceeds the material flow stress
c	That the stress at the centre of the face should exceed the material flow stress

Answer

A.

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

How might elongating the cells of a hexagonal honeycomb in the direction of loading change E/E_S for a given relative density?

a	Decrease it
b	Have no effect
c	Increase it

Answer

C. Increase it. The length of the beam that must bend is decreased whilst the original length of the cell is increased.

For a highly porous structure E/E_S is proportional to $(\rho/\rho_S)n$. In which case will the structure show an increased bending stiffness? Hence explain why many porous materials are often used as the core in sandwich structures.

a	$n > 2$
b	$n = 2$
c	$n < 2$

Answer

In many open-cell structures n = 2, so that there is no benefit in using the porous structure alone. However advantage can be obtained if the surfaces of the material are covered with a stiff, strong material.

The structure will be stiffer in bending if $n < 2$. In this case the relative second moment of area will increase as the density is reduced faster than the Young modulus will decrease.

Consider a honeycomb loaded with some faces parallel to the compressive loading direction, in which the shape of the hexagonal cell is such that $\theta < 0$, how would the material deform elastically in the transverse direction?

a	contract inwards
b	no tranverse movement
c	expands outwards

Answer

A. contract inwards, in other words the material has a negative Poisson ratio.

This page titled 16.9: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

17: Introduction to Deformation Processes

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

- Be aware of major deformation processes and the ways in which they are used to form metallic objects:
- - Rolling
  - Extrusion
  - Forging
  - Drawing
- Appreciate the advantages and disadvantages of each.
- Understand the reasons for selecting a process for a particular shape of end product.

Before you start

A basic understanding of plastic deformation (including work hardening) is assumed and can be found in the Introduction to Mechanical Testing TLP.

17.1: Introduction
17.2: Rolling
17.3: Forging
17.4: Extrusion
17.5: Drawing
17.6: Other Processes
17.7: Summary
17.8: Questions

This page titled 17: Introduction to Deformation Processes is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.1: Introduction

A forming operation is one in which the shape of a metal sample is altered by plastic deformation. Forming processes include stamping, rolling, extrusion and forging, where deformation is induced by external compressive forces or stresses exceeding the yield stress of the material. Drawing is a fundamentally different process in that the external forces are tensile in nature and hence the ultimate tensile strength of the material cannot be exceeded.

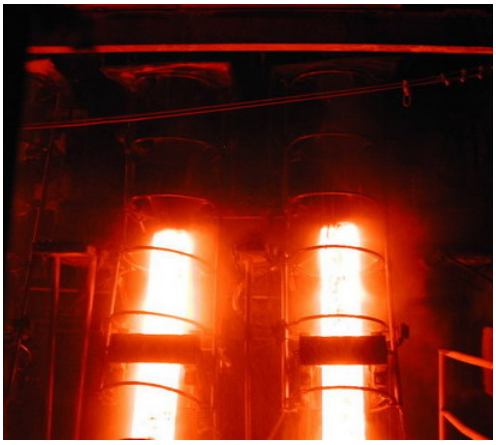
Metals or alloys used in forming processes require a moderate level of ductility to enable plastic deformation with no fracture.



Some formed articles

Forming can be divided into two categories:

Hot Working



Hot ingots from Phase Transition Group web page

Deformation is carried out at a temperature high enough for fast recrystallisation to occur. A crude estimate for a hot working temperature T for a particular metal or alloy is that it must be greater than $0.6T_m$ where T_m is the melting point in degrees Kelvin. This lower bound for the hot working temperature varies for different metals, depending on factors such as purity and solute content. Thus, a highly pure metal will undergo recovery and recrystallisation at a particular hot working temperature more readily than an alloyed metal.

Deformation energy requirements for hot working are less than that of cold working. At hot working temperatures, a metal remains ductile through dynamic reforming of its grain structure, so repeated, large deformations are possible. The strain rates of many metal-working processes are so high that there is insufficient time for the metal to recrystallise as it deforms. However, recovery and recrystallisation do occur in the time period between repeated hot working operations.

Hot working achieves both the mechanical purpose of obtaining the desired shape and also the purpose of improving the physical properties of the material by destroying its original cast structure. The porous cast structure, often with a low mechanical strength,

is converted to a wrought structure with finer grains, enhanced ductility and reduced porosity. Depending on the final hot working temperature, an annealed microstructure can be obtained.

At elevated temperatures, most metals experience some surface oxidation, which results in a poor surface finish as well as a loss of material. Processing in an inert atmosphere is possible, but it is very expensive and is usually avoided unless the metal is very reactive.

Cold Working

This is the term for processes that are performed at room temperature (or up to about 200°C for some metals). Cold working leads to anisotropy and increased stiffness and strength in a metal. There is a corresponding decrease in ductility and malleability as the metal strain hardens. Advantages over hot working include a better quality surface finish, closer dimensional control of the final article and improved mechanical properties.

Cold working processes can be divided into two broad classes:

1. Those in which cold working is carried out for the purpose of shaping the article only.

- Here, any strain hardening effects are not desired and may have to be removed by annealing both between the various stages of plastic shaping as well as after the final cold working shaping operation.

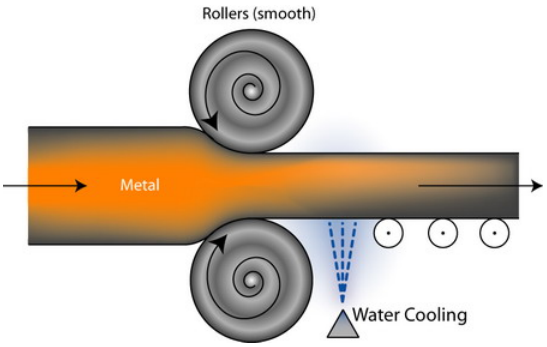
2. Those in which the object of cold rolling is not only to obtain the required shape but also to strain harden and strengthen the metal.

- Plastic deformation must not be carried beyond a certain point or brittle fracture is likely to result. In order to avoid this, total deformation can be accomplished in a series of steps in which the article is successively cold worked by a small amount and then process annealed in order to reduce hardness and increase ductility, thereby permitting further cold working as required.

This page titled 17.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.2: Rolling

Rolling is the most widely used deformation process. It consists of passing metal between two rollers, which exert compressive stresses, reducing the metal thickness. Where simple shapes are to be made in large quantity, rolling is the most economical process. Rolled products include sheets, structural shapes and rails as well as intermediate shapes for wire drawing or forging. Circular shapes, 'I' beams and railway tracks are manufactured using grooved rolls.



Rolling

Hot Rolling

Initial breakdown of an ingot or a continuously cast slab is achieved by hot rolling. Mechanical strength is improved and porosity is reduced. The worked metal tends to oxidise leading to scaling which results in a poor surface finish and loss of precise dimensions. A hot rolled product is often pickled to remove scale, and further rolled cold to ensure a good surface finish and optimise the mechanical properties for a given application.

https://www.doitpoms.ac.uk/tlplib/me...os/rolling.mp4

Hot Rolling

Reproduced from Materials Selection and Processing CD, by A.M.Lovatt, H.R.Shercliff and P.J.Withers.

Cold Rolling

Cold rolling is often used in the final stages of production. Sheets, strips and foils are cold rolled to attain dimensional accuracy and high quality surface finishes.

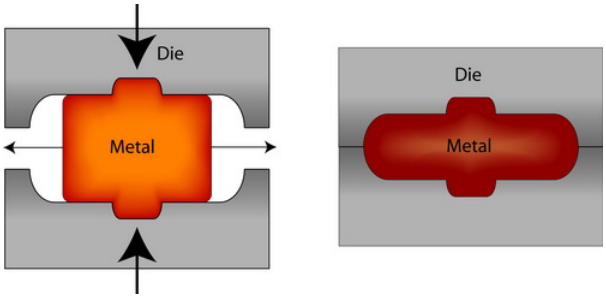
With softer metals such as lead and copper, a succession of cold-rolling passes can impose very large deformations. For many materials, however, the rolling sequence has to be interrupted for intermediate annealing in order to prevent fracture.

This page titled 17.2: Rolling is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.3: Forging

In this operation, a single piece of metal, normally hot, is deformed mechanically by the application of successive blows or by continuous squeezing. Forged articles range in size from nuts and bolts, hip replacement prostheses and crankshafts to (traditionally) gun barrels. Most engineering metals and alloys can be forged readily and include most steels, aluminium and copper alloys and certain titanium alloys including 6-4 (Ti-6 wt.%Al-4 wt%V) and 6-2-4-2 (Ti-6 wt.%Al-2 wt.%Sn-4 wt.%Zn-2 wt.%Mo). Strain-rate and temperature-sensitive materials, such as magnesium and nickel based superalloys, may require more sophisticated forging processes such as radial forging. Forged articles have excellent mechanical properties, combining fine grain structure with strengthening through strain hardening.

Closed Die



Closed-die forging

A force is brought to bear on a metal slug or preform placed between two (or more) die halves. The metal flows plastically into the cavity formed by the die and hence changes in shape to its finished shape. Examples of the machinery used include hydraulic presses, mechanical presses and hammers.

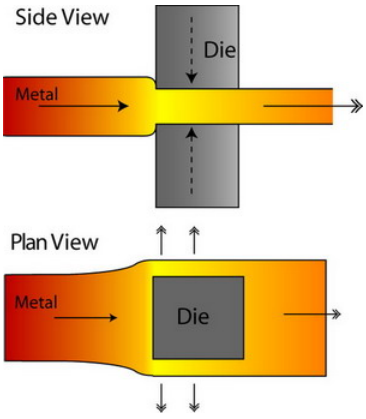
https://www.doitpoms.ac.uk/tlplib/me...os/forging.mp4

Close Die Forging

Reproduced from Materials Selection and Processing CD, by A.M.Lovatt, H.R.Shercliff and P.J.Withers.

Possible geometries range from simple spherical blocks and discs to intricate components incorporating thin webs, holes, cavities, pockets and ribs. As metal flow is restricted by the die contours, closed-die forging can produce complex shapes and higher tolerances than the shapes and tolerances achieved using open-die forging processes.

Open Die



Open-die forging

Open-die forging is performed between flat dies with no pre-cut profiles. The dies do not confine the metal laterally during forging. Deformation is achieved through movement of the workpiece relative to the dies. Parts up to thirty metres in length can by

hammered or pressed into shape in this way. Open-die forging comprises many process variations, enabling an extremely broad range of shapes and sizes to be produced. Unlike rolling operations, which generally produce continuous plates, sheets, strip, or various structural cross-sections, forging operations produce discrete parts.

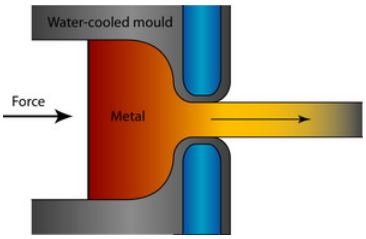
In addition to round, square, rectangular, hexagonal bar and other basic shapes, open-die processes can produce:

- Spindles or rotors, whose diameter increases or decreases at multiple locations along the longitudinal axis.
- Hollows cylindrical in shape, usually with length much greater than the diameter of the part.
- Ring-like parts resembling washers or approaching hollow cylinders in shape, depending on the height:wall thickness ratio.

This page titled 17.3: Forging is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.4: Extrusion

In extrusion, a bar or metal is forced from an enclosed cavity via a die orifice by a compressive force applied by a ram. Since there are no tensile forces, high deformations are possible without the risk of fracture of the extruded material. The extruded article has the desired, reduced cross-sectional area, and also has a good surface finish so that further machining is not needed. Extrusion products include rods and tubes with varying degrees of complexity in cross-section.



Extrusion

Examples of metals that can be extruded include lead, tin, aluminium alloys, copper, brass and steel. The minimum cross-sectional dimensions for extruded articles are approximately 3 mm in diameter for steel and 1 mm in diameter for aluminium. Some metals such as lead alloys and brass lend themselves to extrusion rather than drawing or rolling.

https://www.doitpoms.ac.uk/tlplib/me...s/extrusio.mp4

Extrusion

Reproduced from Materials Selection and Processing CD, by A.M.Lovatt, H.R.Shercliff and P.J.Withers.

Hot extrusion is carried out at a temperature T of approximately $0.6T_m$ and the pressures required range from 35 to 700 MPa. Under these demanding conditions, a lubricant is required to protect the die. Oil and graphite lubricants function well at temperatures up to 150°C, but borate glass or hexagonal boron nitride powders are favoured at higher temperatures where carbon-based lubricants oxidise.



Extruded products

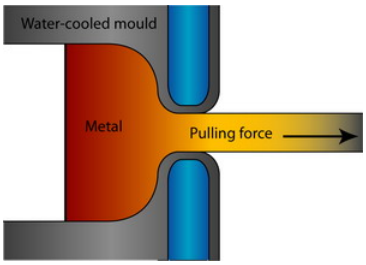
Cold extrusion is performed at temperatures significantly below the melting temperature of the alloy being deformed, and generally at room temperature. The process can be used for most materials, provided that sufficiently robust machinery can be designed. Products of cold extrusion include aluminium cans, collapsible tubes and gear blanks.

This page titled 17.4: Extrusion is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.5: Drawing

Drawing is the pulling of a metal piece through a die by means of a tensile force applied to the exit side. A reduction in cross-sectional area results, with a corresponding increase in length. A complete drawing apparatus may include up to twelve dies in a series sequence, each with a hole a little smaller than the preceding one. In multiple-die machines, each stage results in an increase in length and therefore a corresponding increase in speed is required between each stage. This is achieved using “capstans” which are used both to apply the tensile force and also to accommodate the increase in the speed of the drawn wire. These speeds may reach 60 ms^{-1}.

Dies must be very hard so they tend to be made from steel or chilled cast iron. However, tungsten carbide and even diamond are increasingly used because of their greater ability to retain shape. A typical lubricant used for drawing is tallow, a soap/fat paste-type material that has a formulation of 5 wt% soap, 25 wt% oil, 25 wt% water, and 45 wt% solids.



Drawing

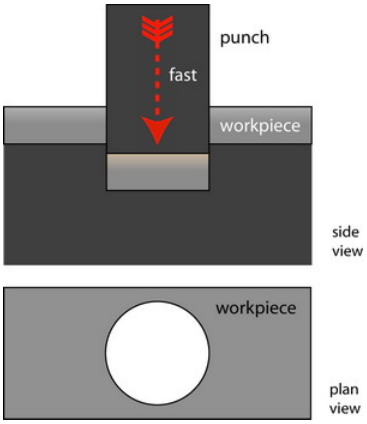
Metals can be formed to much closer dimensions by drawing than by rolling. Shapes ranging in size from the finest wire to those with cross-sectional areas of many square centimetres are commonly drawn. Larger artefacts may be drawn to square, round and even irregular cross sections. Drawn products include wires, rods and tubing products. Large quantities of steel and brass are cold drawn. Seamless tubing can be produced by cold drawing when thin walls and very accurate finishes are required.

This page titled 17.5: Drawing is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.6: Other Processes

Stamping

Stamping is used to make high volume parts such as aviation or car panels or electronic components. Mechanical or hydraulic powered presses stamp out parts from continuous sheets of metal or individual blanks. The upper die is attached to the ram and the lower die is fixed. Whereas mechanical machinery transfers all energy as a rapid punch, hydraulic machinery delivers a constant, controlled force.

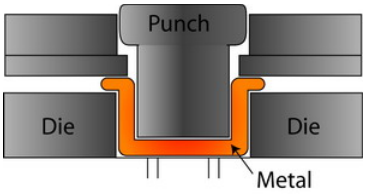


Stamping

Deep Drawing

For deep drawing, the starting sheet of metal is larger than the area of the punch. A pressure plate, fixed to the machine, prevents wrinkling of the edges as the plug is drawn into a top die cavity. The outer parts of the sheet are drawn in towards the die as the operation proceeds. The process is limited by the possibility of fracture occurring during drawing; the maximum sheet width is rarely more that twice the die diameter.

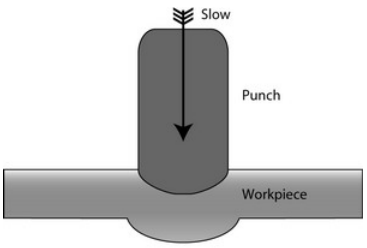
Many shapes are possible including cups, pans, cylinders and irregular shaped products.



Deep drawing

Pressing

A sheet of metal is deformed between two suitably shaped dies usually to produce a cup or dish shaped component. A thick pad of rubber may replace one of the dies, giving reduced tooling costs and allowing larger deformations to be imposed.



Pressing

This page titled 17.6: Other Processes is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.7: Summary

The production of the vast majority of metallic objects at some stage involves one of these four deformation processes:

- Rolling
- Extrusion
- Forging
- Drawing

The intention of this TLP is to introduce these four processes and to provide a brief description of their strengths, limitations and suitability for various applications.

Going further

CD-ROM

- Lovatt A.M., Shercliff H.R. and Withers P.J. (2000), "Material Selection and Processing", CD-ROM and supporting booklets, Technology Enhancement Programme (part of Gatsby Technical Education Project), London.

Websites

- Forging
  Different forging processes, including still graphics, animated sequences and videos.
- Metal Forming Processes
  Various metal forming processes videos.

This page titled 17.7: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

17.8: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

What would happen to a brittle metal such as white cast iron, if it were formed by closed die forging?

Answer

It would shatter when the load was applied at high loading rate, and the final object would be composed of shards pushed together it would not be strong enough.

Discuss how the metal pieces of this article were made.



Answer

The walls of the filing cabinet were made by rolling. The components of the lock were forged. The rails inside (that hold the drawers) were extruded.

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

The melting temperature of a low-carbon steel is 1534°C. Above what temperature can we use hot working to form it, and why?

Answer

Hot working requires the temperature to be >0.6T_m. In the case of this steel, T_m is 1807 K so 0.6 T_m is 1084.2 K hot working can be performed above this temperature. Above this temperature, recovery, recrystallisation and grain growth occur as the deformation process is happening. It is also interesting to note that the range of temperature in which hot working can occur is mostly in the austenite field. Once a sample of this steel has been hot worked for long enough for the transformation to take place, it is austenitic and will again undergo the ferritic transformation on cooling to room temperature. The properties of the steel will change between the end of hot working and the piece s use.

What processes could be used to make this shape? What else could affect the choice of process?



Answer

Deep drawing, closed die forging and open die forging could be used. The cylindrical part could also be extruded and the bottom welded on afterwards.

The choice of process could be influenced by:

- The size of the shape; a small cup could be made by deep drawing, a 10m diameter tank must be open die forged.
- The metal used to make the shape; some metals are better suited to processes The application the shape will be used for; the work hardening exerted during cold working processes may make the shape more suitable to its final purpose (e.g. if it were a pressurised tank).

At 30 ms^{-1}, what weight of 1 mm diameter copper wire could be drawn in an hour? The density of copper is 8920 kg m^{-3}.

Answer

One metre of the wire has volume of $\pi \times r^2 \times 1$ m = 7.85×10^{-7} m^3. To find the mass per metre, we multiply by the density to get 7.01×10^{-3} kg per metre of wire. If 30 metres are drawn in a second, then $30 \times 60 \times 60 = 108000$ metres are drawn in an hour. Thus 757 kg of wire can be drawn per hour.

Do you think steel reinforcing bars for concrete are drawn or extruded?

Answer

They are extruded. If they were drawn they would end up coiled on a large drum; they would then need to be unrolled and cut into lengths at the site. This would be difficult because they are by necessity very stiff. If they are extruded they can be cut as they emerge and this makes them easier to make into buildings.

This page titled 17.8: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

18: Dielectric materials

www.doitpoms.ac.uk/tlplib/di...rics/index.php

Learning Objectives

- Understand the meaning of the terms dielectric constant, dielectric loss and dielectric breakdown.
- Recognise that the properties of dielectrics are due to polarisation and understand how this polarisation arises on the microscopic scale.
- Understand how material structure, temperature and frequency affect the properties of dielectrics.
- Be aware of some practical applications of dielectric materials.

18.1: Introduction
18.2: Electric Dipole
18.3: Polarisation Mechanisms
18.4: Capacitors
18.5: The Dielectric Constant
18.6: Variation of the Dielectric Constant in Alternating Fields
18.7: Effect of Structure on the Dielectric Constant
18.8: Effect of Temperature on the Dielectric Constant
18.9: Loss in Dielectrics
18.10: Dielectric Breakdown
18.11: Applications of Dielectrics
18.12: Summary
18.13: Questions

This page titled 18: Dielectric materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.1: Introduction

A dielectric material is any material that supports charge without conducting it to a significant degree. In principle all insulators are dielectric, although the capacity to support charge varies greatly between different insulators for reasons that will be examined in this TLP.

Dielectric materials are used in many applications, from simple electrical insulation to sensors and circuit components.

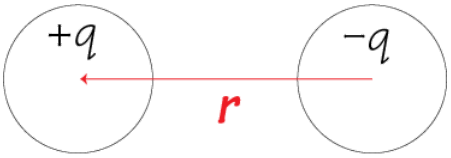
This page titled 18.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.2: Electric Dipole

A dielectric supports charge by acquiring a polarisation in an electric field, whereby one surface develops a net positive charge while the opposite surface develops a net negative charge. This is made possible by the presence of electric dipoles – two opposite charges separated by a certain distance – on a microscopic scale.

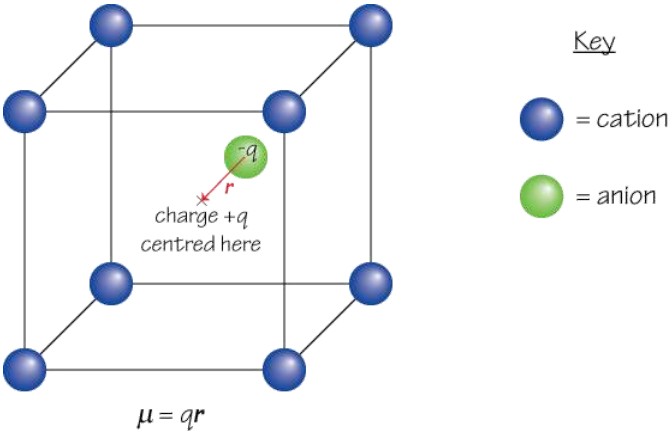
A mathematical treatment of dipole moment can be found in the TLP on ferroelectrics. For the purposes of this TLP it is worth noting that a dipole can be considered in two ways:

1. If two discrete charged particles of opposite charges are separated by a certain distance, a dipole moment **μ** arises.



$$\mu = qr$$

2. If the centre of positive charge within a given region and the centre of negative charge within the same region are not in the same position, a dipole moment **μ** arises. For example, in the diagram below the centre of positive charge from the 8 cations shown is at X, while the centre of negative charge is located some distance away on the anion.



$\mu = qr$

The second view of dipole moment is more useful, since it can be applied over a large area containing many charges in order to find the net dipole moment of the material, and can also be used in situations where it is inappropriate to consider the charges as belonging to discrete particles – e.g. in the case of the electron cloud that surrounds the nucleus in an atom, which must be described by a wavefunction.

Note that in the equation for dipole moment, **r** is a vector (the sign convention is that **r** points from negative to positive charge) therefore the dipole moment **μ** is also a vector. The polarisation of a material is simply the total dipole moment for a unit volume.

$$P = \frac{\sum \mu}{V} \tag{18.2.1}$$

where V is the overall volume of the sample.

Since Σ**μ** is a vector sum, a material may contain dipoles without having any net polarisation, since dipole moments can cancel out.

This page titled 18.2: Electric Dipole is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.3: Polarisation Mechanisms

There are three main polarisation mechanisms that can occur within a dielectric material: electronic polarisation, ionic polarisation (sometimes referred to as atomic polarisation) and orientational polarisation. The animation below illustrates how each of these mechanisms functions on the microscopic scale.

All non-conducting materials are capable of electronic polarisation, which is why all insulators are dielectric to some degree. In contrast, the ionic and orientational modes are only available to materials possessing ions and permanent dipoles respectively.

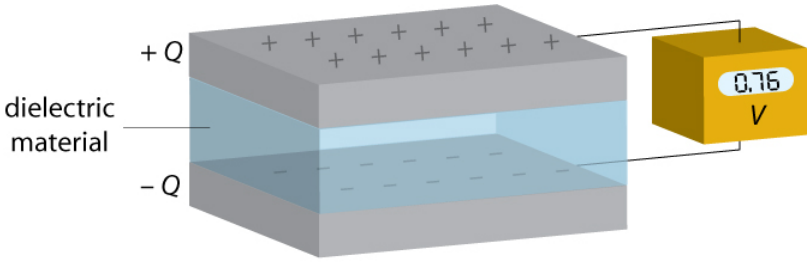
Another contribution to polarization is space charge, or the accumulation of mobile charges at structural surfaces and interfaces. Rather than being a direct property of a material this is only a feature of heterostructures, and hence is not discussed further here.

This page titled 18.3: Polarisation Mechanisms is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.4: Capacitors

A capacitor is a device used for storing charge. It normally consists of two conducting plates with a dielectric material between them, although an “empty capacitor” – one with a vacuum between the plates – may also be used in some applications.

Each capacitor has a capacitance C, the standard units of which are Farads (F). The capacitance is defined by the relationship $Q = C V$ where Q is the charge on each capacitor plate and V is the voltage between capacitor plates. Therefore $1 F = 1 CV^{-1}$.



$$Q = CV$$

The capacitance is affected by various factors, such as the capacitor geometry, however here we shall only deal with the effect of the dielectric material chosen to occupy the space between the plates.

Increasing the capacitance in this way is desirable, since it allows a greater electric charge to be stored for a given field strength.

This page titled 18.4: Capacitors is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

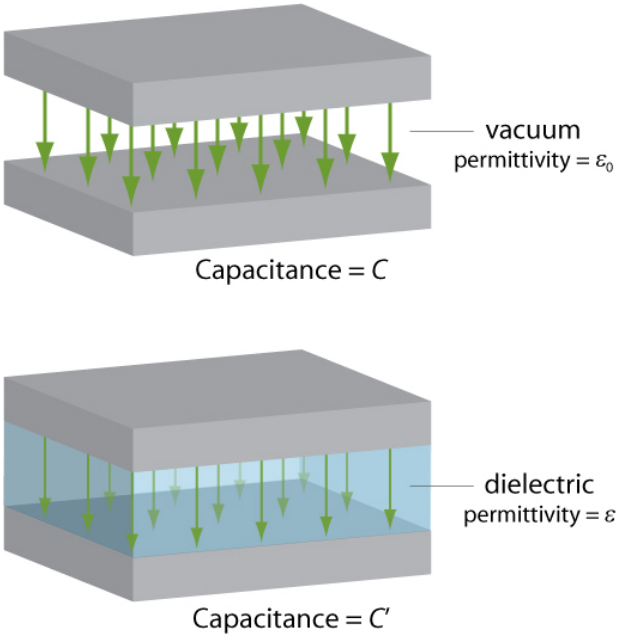
18.5: The Dielectric Constant

The dielectric constant of a material provides a measure of its effect on a capacitor. It is the ratio of the capacitance of a capacitor containing the dielectric to that of an identical but empty capacitor.

An alternative definition of the dielectric constant relates to the permittivity of the material. Permittivity is a quantity that describes the effect of a material on an electric field: the higher the permittivity, the more the material tends to reduce any field set up in it. Since the dielectric material reduces the field by becoming polarised, an entirely equivalent definition is that the permittivity expresses the ability of a material to polarise in response to an applied field. The dielectric constant (sometimes called the 'relative permittivity') is the ratio of the permittivity of the dielectric to the permittivity of a vacuum, so the greater the polarisation developed by a material in an applied field of given strength, the greater the dielectric constant will be.

There is no standard symbol for the dielectric constant – you may see it referred to as κ, ε, ε' or ε_r. In this TLP κ shall be used to avoid confusion with the absolute permittivity, which may also be given the symbol ε.

The two definitions of the dielectric constant are illustrated by the diagram below (the green arrows represent the electric field).



$$\kappa = C'/C \qquad \kappa = \varepsilon/\varepsilon_0$$

In general, the more available polarisation mechanisms a material possesses, the larger its net polarisation in a given field will be and hence the larger its dielectric constant will be.

The dielectric constant of a material and its refractive index are closely linked by the equation $\kappa = n^2$ (click here for derivation). However, care must be taken in applying this equation. It is only strictly accurate when the dielectric constant and the refractive index are measured under the same conditions. Specifically, since the dielectric constant can vary significantly with frequency (for reasons discussed in the next section of this TLP), we must measure the dielectric constant under alternating current at the same frequency that we measure the refractive index at – the frequency of visible light, ~10^{15} Hz. However, quoted values of the dielectric constant normally refer to the static dielectric constant – that is, the dielectric constant under direct current. This is often very different from the value of the dielectric constant at 10^{15} Hz.

The exception to this is for materials that possess only the electronic mode of polarisation. For these materials, the dielectric constant does not vary significantly with frequency below visible frequencies, and $\kappa_S \approx n^2$ where κ_S is the static dielectric constant.

To summarise: the equation $\kappa = n^2$ can be applied to the static dielectric constants of non-polar materials only, or to the high-frequency dielectric constants of any dielectric.

This page titled 18.5: The Dielectric Constant is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.6: Variation of the Dielectric Constant in Alternating Fields

We know that a dielectric becomes polarised in an electric field. Now imagine switching the direction of the field. The direction of the polarisation will also switch in order to align with the new field. This cannot occur instantaneously: some time is needed for the movement of charges or rotation of dipoles.

If the field is switched, there is a characteristic time that the orientational polarisation (or average dipole orientation) takes to adjust, called the relaxation time. Typical relaxation times are $\sim 10^{-11}$ s. Therefore, if the electric field switches direction at a frequency higher than $\sim 10^{11}$ Hz, the dipole orientation cannot 'keep up' with the alternating field, the polarisation direction is unable to remain aligned with the field, and this polarisation mechanism ceases to contribute to the polarisation of the dielectric.

In an alternating electric field both the ionic and the electronic polarisation mechanisms can be thought of as driven damped harmonic oscillators (like a mass on a spring), and the frequency dependence is governed by resonance phenomena. This leads to peaks in a plot of dielectric constant versus frequency, at the resonance frequencies of the ionic and electronic polarisation modes. A dip appears at frequencies just above each resonance peak, which is a general phenomenon of all damped resonance responses, corresponding to the response of the system being out of phase with the driving force (we shall not go into the mathematical proof of this here). In this case, in the areas of the dips, the polarisation lags behind the field. At higher frequencies the movement of charge cannot keep up with the alternating field, and the polarisation mechanism ceases to contribute to the polarisation of the dielectric.

As frequency increases, the material's net polarisation drops as each polarisation mechanism ceases to contribute, and hence its dielectric constant drops. The animation below illustrates these effects.

At sufficiently high frequencies (above $\sim 10^{15}$ Hz), none of the polarisation mechanisms are able to switch rapidly enough to remain in step with the field. The material no longer possesses the ability to polarise, and the dielectric constant drops to 1 – the same as that of a vacuum.

The resonances of the ionic and electronic polarization mechanisms are illustrated below.

This page titled 18.6: Variation of the Dielectric Constant in Alternating Fields is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.7: Effect of Structure on the Dielectric Constant

We have already seen that the more available polarisation mechanisms a material possesses, the larger its dielectric constant will be. For example, materials with permanent dipoles have larger dielectric constants than similar, non-polar materials.

In addition, the more easily the various polarisation mechanisms can act, the larger the dielectric constant will be. For example, among polymers, the more mobile the chains are (i.e. the lower the degree of crystallinity) the higher the dielectric constant will be.

For polar structures, the magnitude of the dipole also affects the magnitude of polarisation achievable, and hence the dielectric constant. Crystals with non-centrosymmetric structures such as barium titanate have especially large spontaneous polarisations and so correspondingly large dielectric constants. Conversely, a polar gas tends to have smaller dipoles, and its low density also means there is less to polarise, therefore polar gases have lower dielectric constants than polar solids or liquids. The density argument also applies for non-polar gases when compared with non-polar solids or liquids.

This page titled 18.7: Effect of Structure on the Dielectric Constant is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

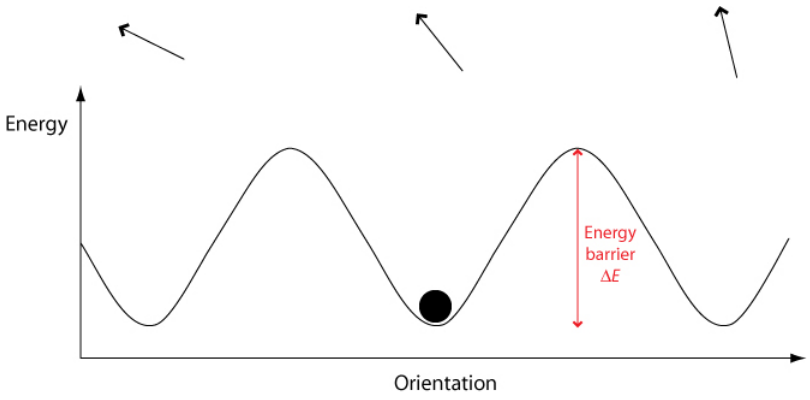
18.8: Effect of Temperature on the Dielectric Constant

For materials that possess permanent dipoles, there is a significant variation of the dielectric constant with temperature. This is due to the effect of heat on orientational polarisation.

However, this does not mean that the dielectric constant will increase continually as temperature is lowered. There are several discontinuities in the dielectric constant as temperature changes. First of all, the dielectric constant will change suddenly at phase boundaries. This is because the structure changes in a phase change and, as we have seen above, the dielectric constant is strongly dependent on the structure. Whether κ will increase or decrease at a given phase change depends on the exact two phases involved.

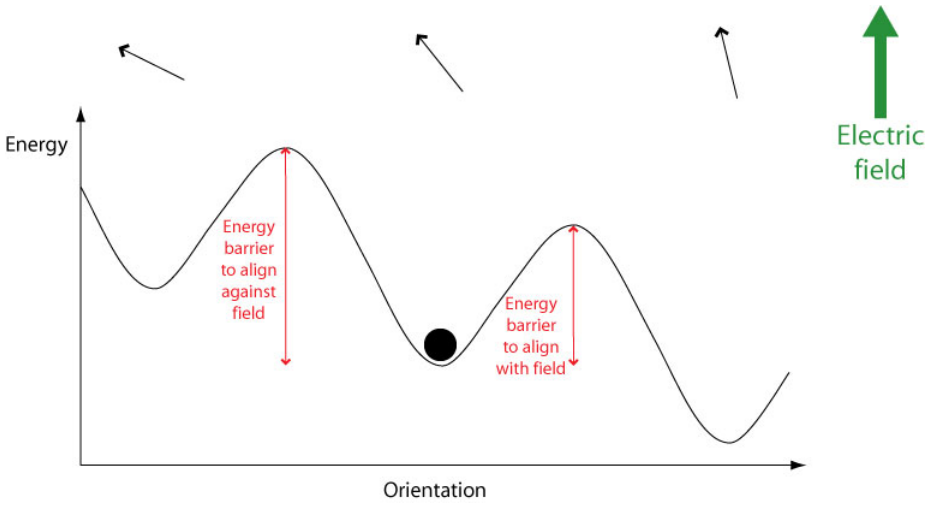
There is also a sharp decrease in κ at a temperature some distance below the freezing point. Let us now examine the reason for this.

In a crystalline solid, there are only certain orientations permitted by the lattice. To switch between these different orientations, a molecule must overcome a certain energy barrier ΔE.



When an electric field is applied, the potential energy of orientations aligned with the field is lowered while the energy of

orientations aligned against the field is raised. This means that less energy is required to switch to orientations aligned with the field, and more energy required to switch to orientations aligned against the field.



Therefore over time molecules will become aligned with the field. However, they must still overcome an energy barrier in order to do this. If a molecule possesses an energy less than the height of any energy barrier, it cannot cross the energy barrier therefore cannot change its orientation. Hence the orientational mode becomes “frozen out” and can no longer contribute to overall polarisation, leading to a drop in the dielectric constant.

These effects are summarised in the graph below.

This page titled 18.8: Effect of Temperature on the Dielectric Constant is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and

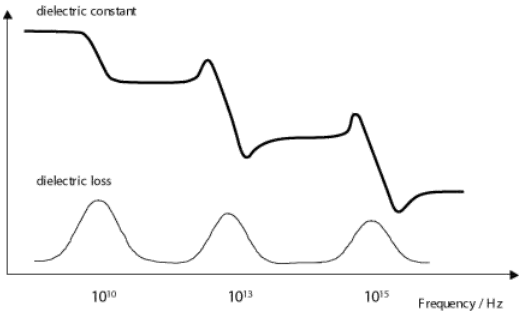
standards of the LibreTexts platform.

https://eng.libretexts.org/@go/page/31628

18.9: Loss in Dielectrics

An efficient dielectric supports a varying charge with minimal dissipation of energy in the form of heat. There are two main forms of loss that may dissipate energy within a dielectric. In conduction loss, a flow of charge through the material causes energy dissipation. Dielectric loss is the dissipation of energy through the movement of charges in an alternating electromagnetic field as polarisation switches direction.

Dielectric loss is especially high around the relaxation or resonance frequencies of the polarisation mechanisms as the polarisation lags behind the applied field, causing an interaction between the field and the dielectric's polarisation that results in heating. This is illustrated by the diagram below (recall that the dielectric constant drops as each polarisation mechanism becomes unable to keep up with the switching electric field.)



Dielectric loss tends to be higher in materials with higher dielectric constants. This is the downside of using these materials in practical applications.

Dielectric loss is utilised to heat food in a microwave oven: the frequency of the microwaves used is close to the relaxation frequency of the orientational polarisation mechanism in water, meaning that any water present absorbs a lot of energy that is then dissipated as heat. The exact frequency used is slightly away from the frequency at which maximum dielectric loss occurs in water to ensure that the microwaves are not all absorbed by the first layer of water they encounter, therefore allowing more even heating of the food.

This page titled 18.9: Loss in Dielectrics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.10: Dielectric Breakdown

At high electric fields, a material that is normally an electrical insulator may begin to conduct electricity – i.e. it ceases to act as a dielectric. This phenomenon is known as dielectric breakdown.

The mechanism behind dielectric breakdown can best be understood using band theory. A detailed explanation of this can be found in the TLP on semiconductors although not all of this is relevant to the content of this TLP, therefore the aspects of band theory needed to understand dielectric breakdown are presented here.

For each material, there is a characteristic field strength needed to cause dielectric breakdown. This is referred to as the breakdown field or dielectric strength. Typically values of the dielectric strength lie in the range 10^6 – 10^9 Vm^{-1}. The exact value of the dielectric strength depends on many factors – most obviously the size of the energy gap, but also the geometry and microstructure of the sample and the conditions it is subjected to.

The phenomenon of dielectric breakdown is utilised in cigarette lighters and similar devices where a spark must be produced in order to ignite the fuel. The “spark gap” is a small air gap between two electrodes. Charge is built up on the electrodes on either side of the spark gap until the strength of the field across the spark gap exceeds the dielectric strength of air (the mechanism used to create this field is not directly relevant to this TLP, but interested readers may find an explanation of it here). At this point the air within the spark gap becomes capable of conduction, resulting in a spark.

This page titled 18.10: Dielectric Breakdown is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

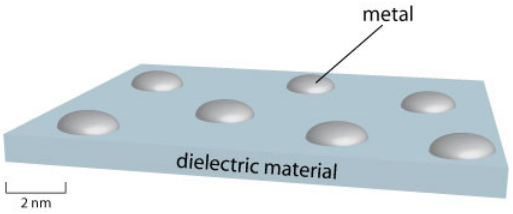
18.11: Applications of Dielectrics

A major use of dielectrics is in fabricating capacitors. These have many uses including storage of energy in the electric field between the plates, filtering out noise from signals as part of a resonant circuit, and supplying a burst of power to another component. The TLP on ferroelectrics shows how the last of these functions is utilised in a camera flash system.

The larger the dielectric constant, the more charge the capacitor can store in a given field, therefore ceramics with non-centrosymmetric structures, such as the titanates of group 2 metals, are commonly used. In practice, the material in a capacitor is in fact often a mixture of several such ceramics. This is due to the variation of the dielectric constant with temperature discussed earlier. It is generally desirable for the capacitance to be relatively independent of temperature; therefore modern capacitors combine several materials with different temperature dependences, resulting in a capacitance that shows only small, approximately linear temperature-related variations.

Of course in some cases a low dielectric loss is more important than a high capacitance, and therefore materials with lower values of κ – and correspondingly lower dielectric losses – may be used for these situations.

Some applications of dielectrics rely on their electrically insulating properties rather than ability to store charge, so high electrical resistivity and low dielectric loss are the most desirable properties here. The most obvious of these uses is insulation for wires, cables etc., but there are also applications in sensor devices. For example, it is possible to make a type of strain gauge by evaporating a small amount of metal onto the surface of a thin sheet of dielectric material.



Electrons may travel across the metal by normal conduction, and through the intervening dielectric material by a phenomenon known as quantum tunnelling. A mathematical treatment of this phenomenon is outside the scope of this TLP; simply note that it allows particles to travel between two “permitted” regions that are separated by a “forbidden” region and that the extent to which tunnelling occurs decreases sharply as distance between the permitted regions increases. In this case the permitted regions are the solidified metal droplets, and the forbidden region is the high-resistance dielectric material.

If the dielectric material is strained, it will bow causing the distances between the metal islands to change. This has a large impact on the extent to which electrons can tunnel between the islands, and thus a large change in current is observed. Therefore the above device makes an effective strain gauge.

This page titled 18.11: Applications of Dielectrics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.12: Summary

- Dielectrics are electrical insulators that support charge.
- The properties of dielectrics are due to polarisation.
- There are three main mechanisms by which polarisation arises on the microscopic scale: electronic (distortion of the electron cloud in an atom), ionic (movement of ions) and orientational (rotation of permanent dipoles).
- A capacitor is a device that stores charge, usually with the aid of a dielectric material. Its capacitance is defined by $Q = C V$
- The dielectric constant κ indicates the ability of the dielectric to polarise. It can be defined as the ratio of the dielectric's permittivity to the permittivity of a vacuum.
- Each of the polarisation mechanisms has a characteristic relaxation or resonance frequency. In an alternating field, at each of these (materials dependent) frequencies, the dielectric constant will sharply drop.
- The dielectric constant is also affected by structure, as this affects the ability of the material to polarise.
- Polar dielectrics show a decrease in the dielectric constant as temperature increases.
- Dielectric loss is the absorption of energy by movement of charges in an alternating field, and is particularly high around the relaxation and resonance frequencies of the polarisation mechanisms.
- Sufficiently high electric fields can cause a material to undergo dielectric breakdown and become conducting.

Going further

Website

- Source for the first part of the derivation for the relationship between the dielectric constant and the refractive index.

Books

- *Dielectrics*, P. J. Harrop, 1972 (Butterworths)
  Contains a more mathematical treatment of dielectrics, as well as information on many other potential applications.
- *The Solid State*, Second Edition, H. M. Rosenberg, 1978 (OUP)
  Chapter 13, "Dielectric properties", provides a good overview of many of the subjects discussed here and contains the latter part of the derivation for the relationship between the dielectric constant and the refractive index.
- *Electronic and Magnetic Behaviour of Materials*, A. Nussbaum, 1967 (Prentice-Hall) pp.70-77
  Provides a more detailed look at how the properties of dielectrics arise from their microscopic polarisation.

This page titled 18.12: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

18.13: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

A Ca^{2+} cation and an O^{2-} anion are separated by a distance of 2.4 Å. Calculate the resultant dipole moment. (Charge on an electron = 1.6×10^{-19} C)

Answer

7.7×10^{-29} Cm

Consider a capacitor in a computer power supply, possessing a capacitance of 2200 μF. If a voltage of 10 V is applied to this capacitor, what will the charge on the positive plate be? (2 sig figs)

Answer

0.022 C

In which of the cases below does A have a higher static dielectric constant than B, assuming that both A and B are dielectrics? (note that more than one answer may be correct)

a	A has a higher permittivity than B.
b	A is a non-polar gas; B is a polar gas.
c	Two capacitors, X and Y, have identical geometry. Capacitor X contains a sample of A and has a capacitance of 200 nF. Capacitor Y contains a sample of B and has a capacitance of 0.6 μF.
d	A is a sample of water at a temperature of 20 °C; B is a sample of water at a temperature of 80 °C.
e	An electric field of strength 1000 V m^{-1} passes through both A and B. The field strength is reduced less on passing through A than it is on passing through B.

Answer

A and D

Under what conditions is the refractive index related to the dielectric constant by $\kappa \approx n^2$?

a	Any dielectric at low frequency, non-polar dielectrics at all frequencies
b	Any dielectric at optical frequency, non-polar dielectrics at all frequencies
c	Any dielectric at low frequency, polar dielectrics at all frequencies
d	Any dielectric at optical frequency, polar dielectrics at all frequencies

Answer

B

A polar liquid is subjected to an alternating current at 50 Hz. The current frequency is then increased to just above the relaxation frequency of the orientational mode of polarisation. Which of these best describes the behaviour of the dielectric constant as the frequency is increased?

a	It remains roughly constant.
b	It gradually decreases.
c	It remains roughly constant until the relaxation frequency of the orientational mechanism is reached, at which point it drops sharply.
d	It gradually decreases until the relaxation frequency of the orientational mechanism is reached, at which point it drops sharply.

Answer

C

And which best describes the behaviour of the dielectric loss as the frequency is increased?

a	It gradually decreases.
b	It gradually increases.
c	It shows a dip around the relaxation frequency of the orientational mechanism.
d	It shows a peak around the relaxation frequency of the orientational mechanism.

Answer

D

You need to make a capacitor that will operate at low electric field strengths and store a large quantity of charge. Energy efficiency does not need to be high (i.e. loss can be tolerated). Which of the following are you most likely to place between the capacitor plates?

a	Nothing - I would use an empty capacitor.
b	A ceramic with a non-centrosymmetric structure.
c	A polymer with polar chains and an amorphous structure.
d	A polymer with polar chains and a crystalline structure.

Answer

B

This page titled 18.13: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

19: Diffraction and imaging

www.doitpoms.ac.uk/tlplib/di...tion/index.php

Learning Objectives

On completion of this TLP you should:

- Be familiar with one-dimensional and two-dimensional diffraction using a laser.
- Be able to relate features of the diffraction pattern to the diffracting object.
- Understand the concepts of the back focal plane; and bright and dark field image formation.

Before you start

Whilst a very basic knowledge of the physics of waves and optics is assumed, this teaching and learning package covers the fundamentals of diffraction and imaging.

19.1: Introduction
19.2: Diffraction Patterns
19.3: Two-dimensional Diffraction
19.4: Image Formation
19.5: Information Obtained from Diffraction Pattern
19.6: Applications of the Theory of Optical Diffraction and Imaging
19.7: Summary
19.8: Questions

This page titled 19: Diffraction and imaging is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

19.1: Introduction

The phenomenon of diffraction was first documented in 1665 by the Italian Francesco Maria Grimaldi. The use of lasers has only become common in the last few decades. The laser's ability to produce a narrow beam of coherent monochromatic radiation in the visible light range makes it ideal for use in diffraction experiments: the diffracted light forms a clear pattern that is easily measured.

As light, or any wave, passes a barrier, the waveform is distorted at the boundary edge. If the wave passes through a gap, more obvious distortion can be seen. As the gap width approaches the wavelength of the wave, the distortion becomes even more obvious. This process is known as diffraction. If the diffracted light is projected onto a screen some distance away, then interference between the light waves create a distinctive pattern (the diffraction pattern) on the screen. The nature of the diffraction pattern depends on the nature of the gap (or mask) which diffracts the original light wave.

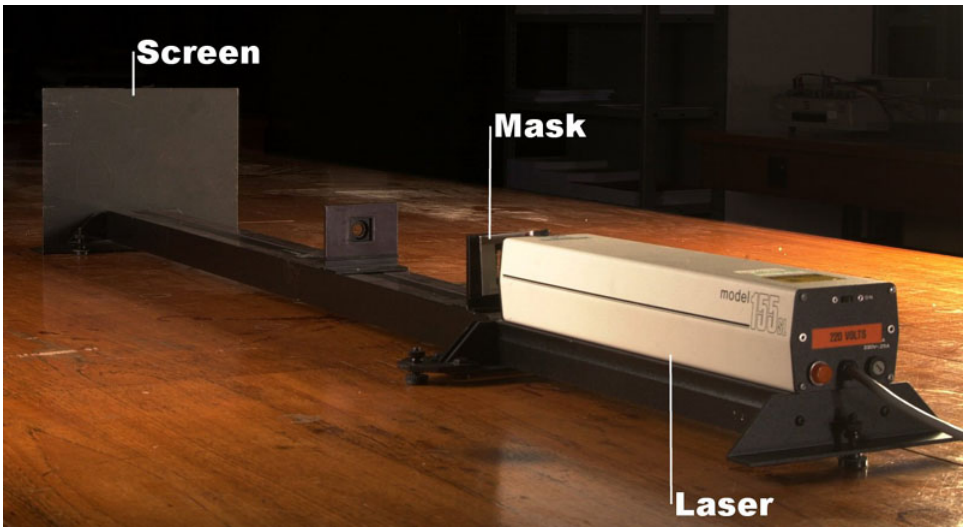
Diffraction patterns can be calculated by from a function representing the mask. The symmetry of the pattern can reveal useful information on the symmetry of the mask. For a periodic object, the pattern is equivalent to the *reciprocal lattice* of the object.

In conventional image formation, a lens focuses the diffracted waves into an image. Since the individual sections (spots) of the diffraction pattern each contain information, by forming an image from only particular parts of the diffraction pattern, the resulting image can be used to enhance particular features. This is used in *bright and dark field imaging*.

This page titled 19.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

19.2: Diffraction Patterns

Laser diffraction experiments can be conducted using an optical bench, as shown below. Light from the laser (of wavelength λ) is diffracted by a mask (usually a small aperture or grating) and projected onto the screen, located at a large distance away, such that Fraunhofer geometry applies. The light on the screen is known as the diffraction pattern.

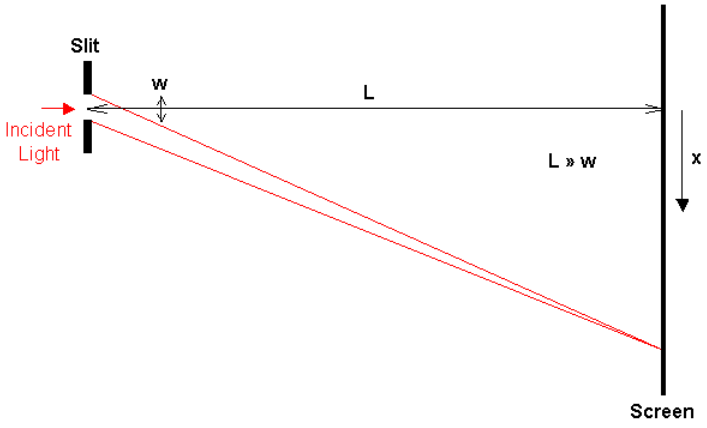


The form of the diffraction pattern from a single slit mask, of width w, involves the mathematical "sinc function", where

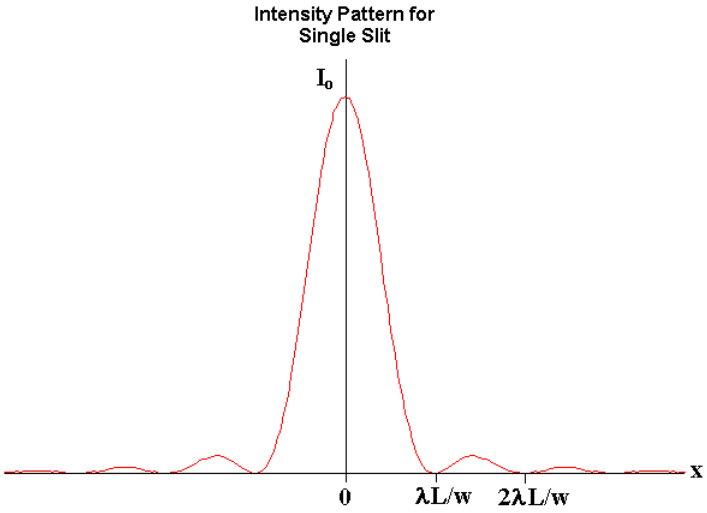
$$\operatorname{sinc}(z) = \frac{\sin(z)}{z} \tag{19.2.1}$$

The observable pattern projected onto the screen (a distance L away) has an intensity pattern as follows, where x is the distance from the straight-through position:

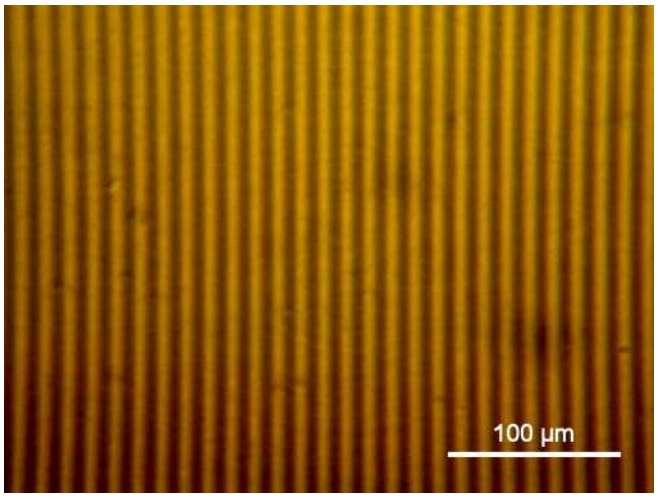
$$I(x) = I_o \operatorname{sinc}^2\left(\frac{\pi x w}{\lambda L}\right) \tag{19.2.2}$$



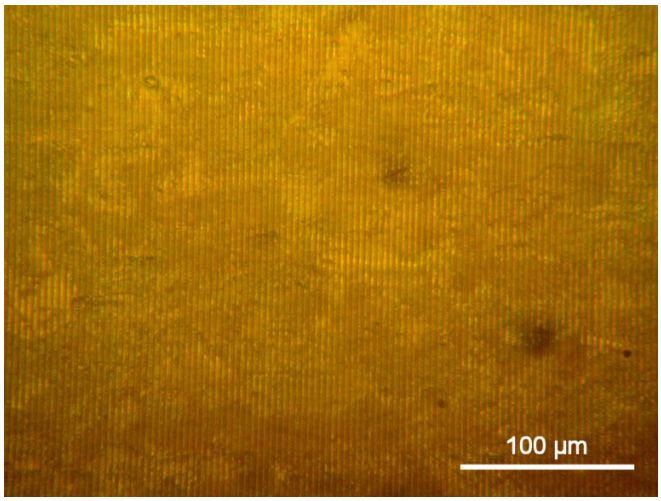
Note that sinc(0) = 1.



Diffraction patterns can be calculated mathematically. The operation that directly predicts the amplitude of the diffraction pattern from the mask is known as a Fourier Transform (provided the conditions for Fraunhofer Diffraction are satisfied). The derivation of some simple patterns can be found here.

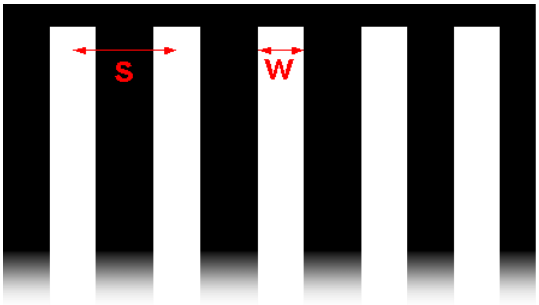


(a) s = 12 μm

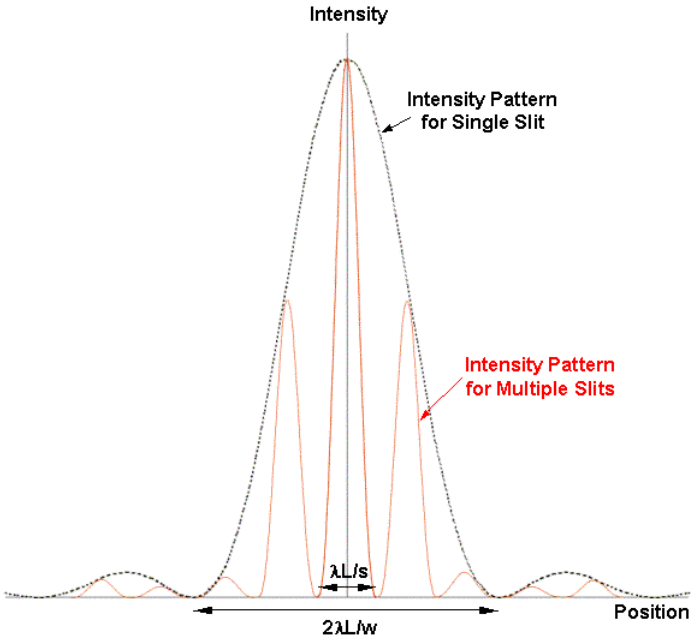


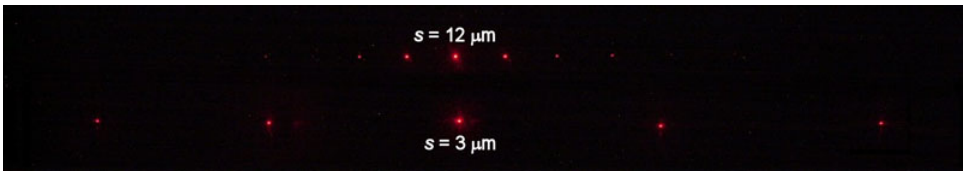
(b) s = 3 μm

A diffraction grating is effectively a multitude of equally-spaced slits. The diffraction pattern from a complex mask such as a grating can be constructed from simplier patterns via the convolution theorem. The observed diffraction pattern is composed of repeated "sinc-squared" functions. Their positions from the central spot are determined by *s* (the spacing between slits) and their relative intensity is dependent on *w* (the width of individual slits).



Slit spacing s and slit width w

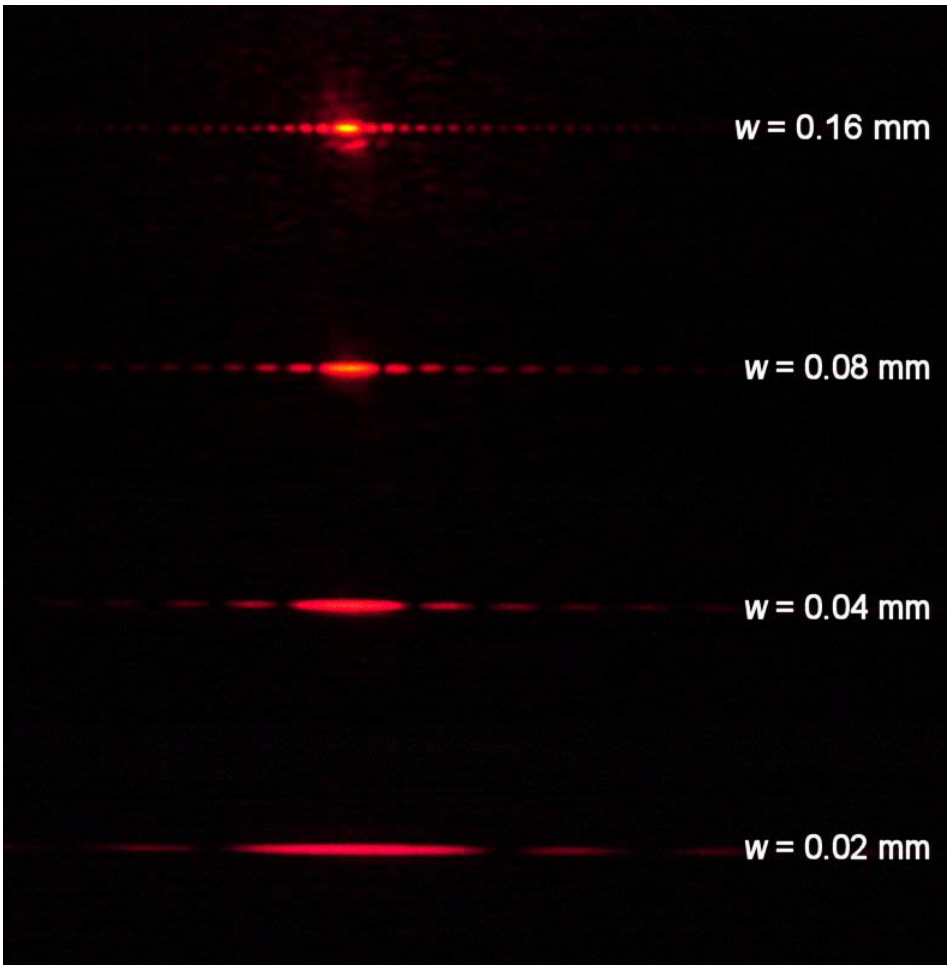




Diffraction patterns from gratings (a) and (b).

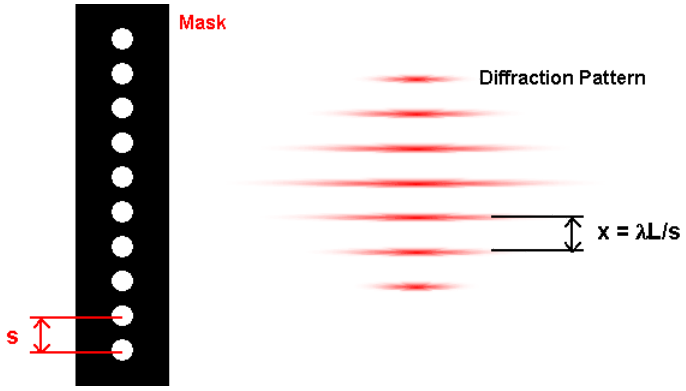
By considering diffraction from a grating, the reciprocal nature of the pattern can be derived. This relationship can be seen in the diffraction patterns of the slits: small features of the diffracting object give wide spacings in the diffraction pattern

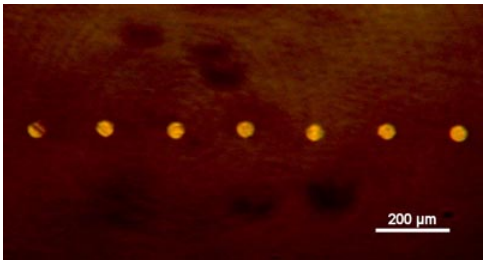
$$s = \frac{\lambda L}{x} \tag{19.2.3}$$



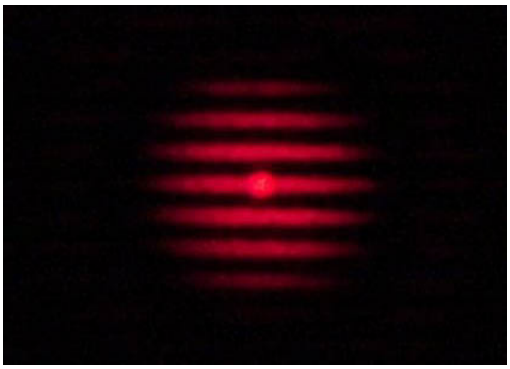
Diffraction patterns from slits of different widths.

More complicated masks, for example a periodic row of apertures, will show more intricate diffraction patterns, but still follow the same basic inverse relationship.





Mask consisting of periodic row of apertures



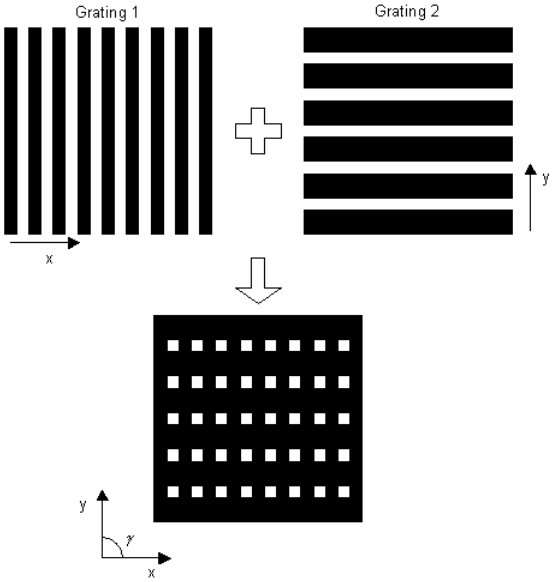
Diffraction pattern for periodic row of apertures

This page titled 19.2: Diffraction Patterns is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

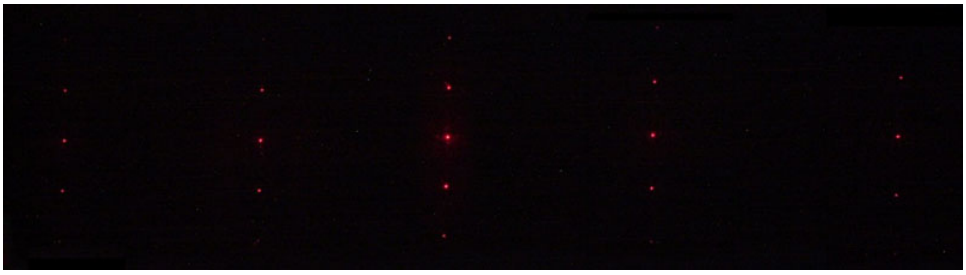
19.3: Two-dimensional Diffraction

The resulting diffraction pattern of a complex mask can be predicted by considering the individual diffraction patterns associated with the components that make up the shape of the mask. This can be seen in the diffraction pattern of a row of apertures.

If two diffraction gratings are superimposed perpendicularly, they form a two-dimensional periodic array of apertures.

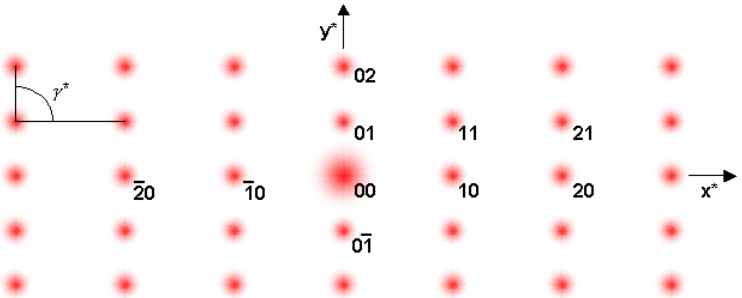


The observed diffraction pattern is neither the sum nor the product of the original patterns of the individual gratings, but the separate patterns are repeated to form a two-dimensional array.

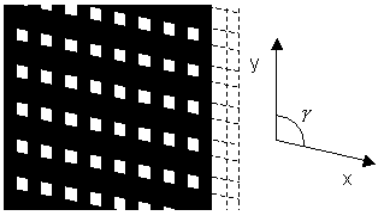


Two-dimensional diffraction pattern

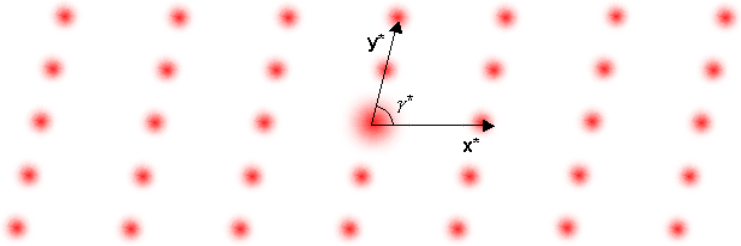
This diffraction pattern of a two-dimensional array of apertures is analogous to the reciprocal lattice of the array, and can be labelled (indexed) as such. Inverse axes are therefore created (where x^* is perpendicular to y, and y^* is perpendicular to x). In a reciprocal lattice, the magnitude of the reciprocal lattice vector is inversely proportional to the magnitude of the original vector. This inverse relationship is evident between the pattern and the mask (the x-axis repeat is smaller than in the y-axis, whereas the x^*-axis repeat is larger than in the y^*-axis). See the X-ray Diffraction and Reciprocal Space TLPs for an explanation of the reciprocal lattice in terms of diffraction.



The angle γ^* in the diffraction pattern is complementary to the angle between the grating axes, γ. i.e. $\gamma + \gamma^* = 180^o$ This can be seen by rotating the gratings with respect to each other.



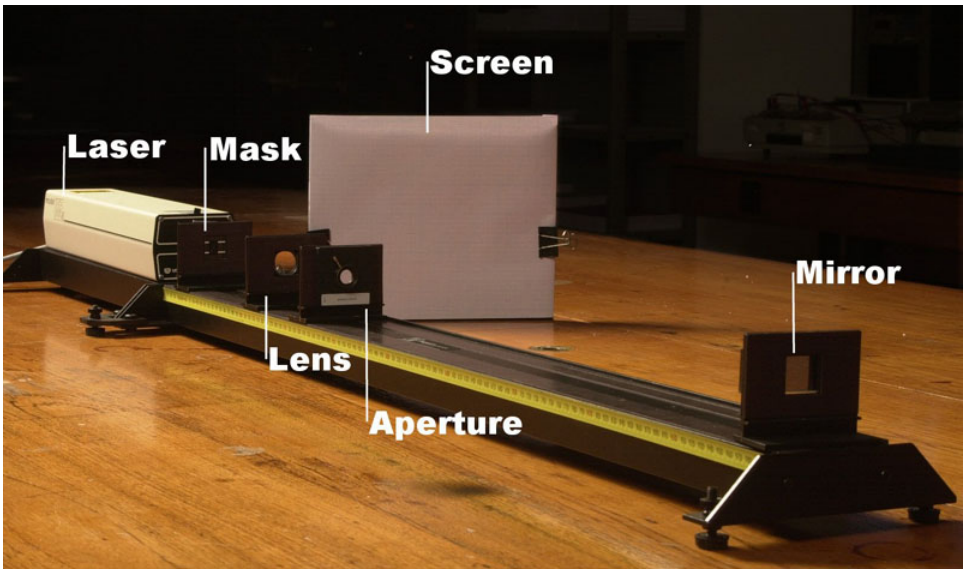
New mask created by rotating grating 2



This page titled 19.3: Two-dimensional Diffraction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

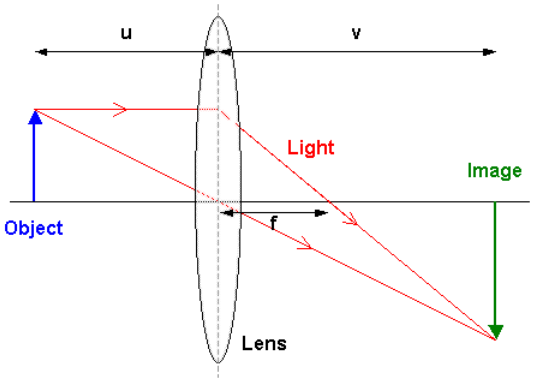
19.4: Image Formation

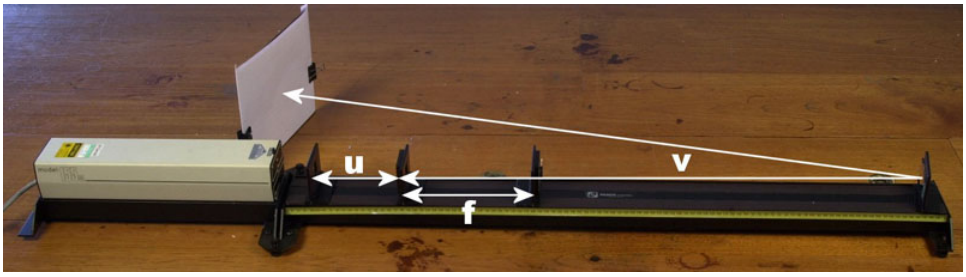
When a convex lens is placed between the mask and the screen, the optical bench can form magnified images of the mask onto the screen. Use of a mirror can simply extend the effective screen distance. Note: caution should be taken when using a mirror to reflect laser light.



Optical bench set up for image formation

The distance between the object and lens (u), the distance between the image and lens (v) and the focal length of the lens (f) are related by the equation

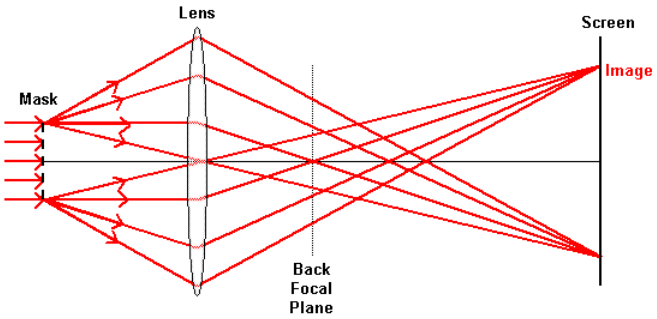




Distances involved

A lens will focus light from infinity to the 'focal point', at a distance from the lens known as the focal length, f. Located at the focal point, is the *back focal plane of the lens* where the diffraction pattern is visible (by using a screen). The diffraction pattern acts as a

source of light that propagates to the screen where the image is formed. This theory was first described by Ernst Abbe in 1872.



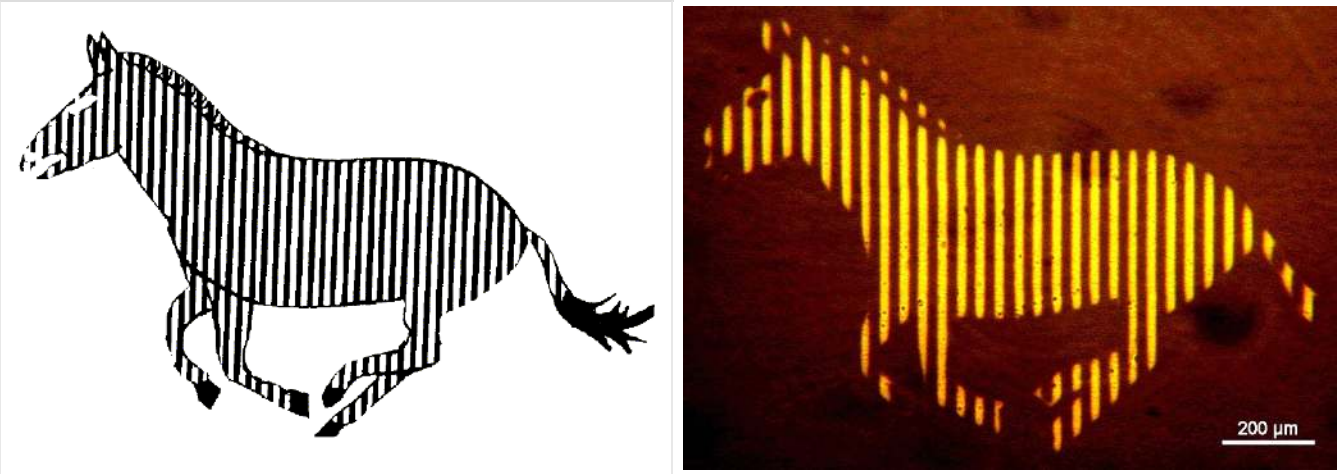
The diffraction pattern of a mask without a centre of symmetry will still be symmetrical. This can be seen in the mathematics of calculating the pattern. The non-centrosymmetric nature of the mask will however cause non-centrosymmetric variations in the phase.

This page titled 19.4: Image Formation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

19.5: Information Obtained from Diffraction Pattern

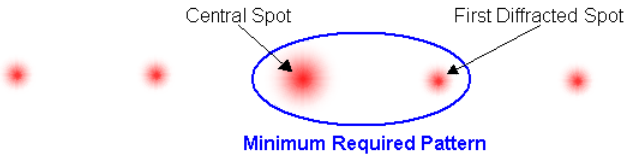
The nature of the diffraction pattern (shape, symmetry, dimensions, etc.) is determined by the nature of the mask that diffracts the light. A lens can recombine the (accessible) diffracted light to generate a magnified image of the mask. However, by forming the image from a limited proportion of the pattern, then elements of the mask can be enhanced.

A mask containing many different geometrical elements is shown here:



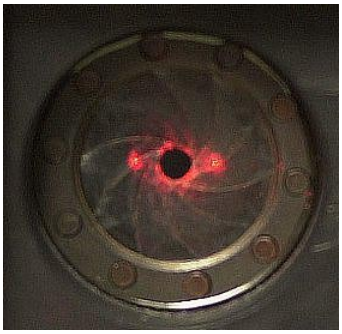
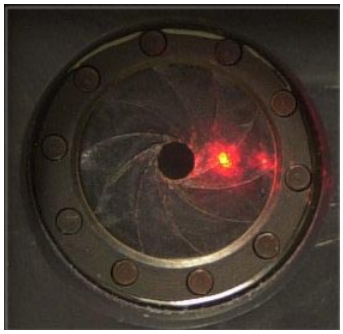
The zebra

A variable aperture can be placed at the back focal plane. Thus the aperture can be adjusted to limit the region of the diffraction pattern that goes on to form the image. The minimum area of the pattern necessary to form a “full” image of the zebra (with overall shape and stripes visible) contains the undiffracted beam and one of the first diffraction spots. In order to properly resolve the features of the mask, both first order diffracted spots should be included.



If only one diffraction spot is allowed through the back focal plane then no information about the spacing of the slits is passed on to the image and individual slits will not be resolved. Note, however, that each diffraction spot is made up of beams scattered from all parts of the object. Therefore, information about the size and shape of the object as a whole is passed on to the image through a single diffraction spot.

If the central (zero order) spot (undiffracted straight-through beam) is solely used, the resulting image is known as the bright field image . If a non-zero order diffraction spot is solely used then a dark field image .

	
Spot selection for bright field imaging	Spot selection for dark field imaging

This page titled 19.5: Information Obtained from Diffraction Pattern is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

19.6: Applications of the Theory of Optical Diffraction and Imaging

The principles of optical diffraction and image formation are equally valid for other waves: for example, neutrons, electron beams and X-rays. The similarity in the diffraction behaviour means that the theory presented here is applicable to them as well.

The symmetry of a diffraction pattern can reveal useful information on the symmetry of the mask. This is exploited in the electron diffraction of crystals, where the pattern can reveal the nature of the crystallographic symmetry, e.g. the periodicity of the structure; the distribution of atoms in the unit cell; and the shape of the crystal. X-ray diffraction patterns are used to measure spacing between layers or rows of atoms, and to determine crystal orientations and structures.

Electron diffraction patterns are two-dimensional sections of the reciprocal lattice of the diffracting crystal. X-ray diffraction patterns are simply 3-dimensional extensions of Fraunhofer diffraction. With X-rays, the crystal only diffracts in a few directions.

The nature of diffraction from a single slit allows macro-scale measurements to be used to calculate micro-scale dimensions. This has important implications - for example, allowing microscopes resolve to very fine scale (nanometre scale).

In optics, the basic shape of the mask is preserved in the bright field image, and some fine detail is lost. In electron diffraction, the contrast of the bright field image is due entirely to thickness and density variations in the sample. A convex glass lens is typically used to focus laser light, but magnetic fields are required to focus electron beams. By selecting individual diffraction spots, dark field images can be used in electron microscopy to distinguish phases (such as characterising two phase intergrowths in crystals).

Multi-beam images (composed of various spots, and known as 'high resolution images') are commonly used to study dislocations. Dark field imaging can be used to highlight the dislocation lines, and by tilting the electron beam, the Burgers vector can be determined. These techniques are common in Transmission Electron Microscopy.

This page titled 19.6: Applications of the Theory of Optical Diffraction and Imaging is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

19.7: Summary

The basic features of diffraction and imaging have been presented in this package. When a wave, such as light, passes through a small aperture, it will be distorted. It will form a distinctive pattern on a screen, known as the diffraction pattern. This pattern contains information on the diffracting aperture (such as a mask or grating), with an inverse relationship in dimensions. The form of the intensity pattern can be predicted mathematically.

A lens can be used to form an image of the mask onto the screen. The diffraction pattern of the mask can be seen in the back focal plane of the lens. By forming the image from selected portions of the diffraction pattern in the back focal plane, particular information present in the image can be enhanced.

The theories involved can be applied to electrons and X-rays, as well as optics.

Going further

Books

- C. Hammond, *The Basics of Crystallography and Diffraction*, Oxford University Press 1997.
- R. Steadman, *Crystallography*, Van Nostrand Reinhold, student edition, 1982.
- J.S. Blakemore, *Solid State Physics*, Cambridge University Press, 1985.

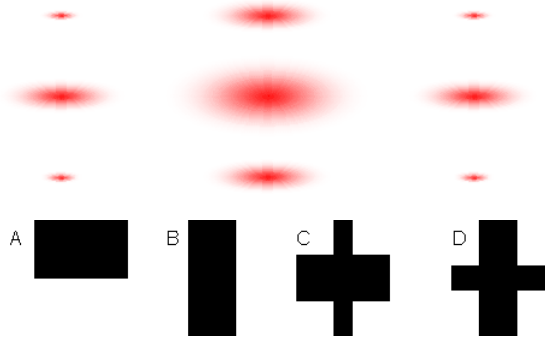
Websites

- X-ray Diffraction
  A DoITPoMS teaching and learning package on the application of diffraction principles to X-ray diffraction.
- Optical Microscopy and Specimen Preparation
  A DoITPoMS teaching and learning package that demonstrates some microscopy techniques involving diffraction.
- Diffraction
  A MATTER module providing an in-depth look at diffraction, including optical, X-ray and electron diffraction.
- Molecular Expressions: Light and Color
  A lot of material on diffraction and interference, among many other topics, including interactive Java applets to help visualise the fundamentals of diffraction patterns, such as:
  - Diffraction of Light
  - Double Slit Interference
- How Lasers work Background information on LASERs on the Marshall Brain's HowStuffWorks website.

This page titled 19.7: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

19.8: Questions

Which of the masks below could have produced the following diffraction pattern?



Answer

B will produce the diffraction pattern, since its dimensions correspond to the inverse relationship.

A's dimensions do not correspond to the inverse relationship.

C and D are cross-shaped apertures which will result in a more complex diffraction pattern.

Increasing the spread of the diffraction pattern allows for more accurate measurement of the spot spacing. Which *one* of these will achieve this?

a	Analysing the pattern in the back focal plane of a lens
b	Moving the laser closer to the aperture
c	Using a blue light laser instead of a red one
d	Moving the screen away from the mask

Hint: Take another look at Diffraction pattern formation.

Answer

A is incorrect. The diffraction pattern visible in the back focal plane is smaller in size than the pattern produced on a screen without a lens.

B is incorrect. This will have no effect on the diffraction pattern (as long as the screen distance is unaffected).

C is incorrect. Red light has a longer wavelength, and hence the spot spacing will be smaller with blue light.

D is correct. This increases the screen distance, and hence the separation between the spots.

When the diffraction pattern from a particular grating is projected onto a screen, the even diffraction spots are missing, i.e. the second, fourth, etc. What is the relationship between the width of the slits (w) of the grating and the distance between each slit (s)?

Answer

Diffraction pattern from a grating: the convolution theorem shows how the intensity function depends on the slit width and the slit spacing. The second maxima of the pattern (i.e. the second diffraction spot) lie $2\lambda L/s$ on either side of the central position. The first minima of the binding sinc function lie at $\lambda L/w$ on either side of the central position. If these overlap, then all the even diffraction spots will be missing, due to the periodic nature of the pattern. Hence:

$$2\lambda L/s = \lambda L/w \tag{19.8.1}$$

I.e.

$$s = 2w \tag{19.8.2}$$

The diffraction pattern from a single slit is projected onto a screen 0.90 m from the slit. Seven spots are visible, with a bright central spot. The maxima of the outer spots are a distance of 15 cm apart. A He-Ne laser is used, which produces light of a wavelength of 0.6328 μm. What is the width of the slit?

Hint: $w = 7\frac{\lambda L}{x}$

Answer

From the diagram of the intensity pattern shown in the Diffraction 1 section it can be seen that the intensity is zero for

$$x = \frac{\lambda L}{w}, \frac{2\lambda L}{w} \tag{19.8.3}$$

and so on. Outside the central bright spot, the other bright spots occur almost exactly half way between the zeros, i.e. close to

$$\frac{3}{2}\frac{\lambda L}{w}, \frac{5}{2}\frac{\lambda L}{w}, \frac{7}{2}\frac{\lambda L}{w} \tag{19.8.4}$$

Thus, if seven spots are visible, the distance x between the outermost pair on opposite sides of the central spot will be

$$2 \times \frac{7}{2}\frac{\lambda L}{w} \tag{19.8.5}$$

Hence the width of the slit w will be given by

$$7\frac{\lambda L}{x} \tag{19.8.6}$$

Separation of outermost spots, x = 15 cm
Screen distance, L = 90 cm
Wavelength of diffracted light, λ = 0.6328 x 10-6 m
Therefore Slit width is

$$w \approx 7\left(\frac{\lambda L}{x}\right) = \frac{7 \times 0.6328 \times 90}{15} = 26.6\mu\text{m} \tag{19.8.7}$$

A convex lens has a focal length of 150 mm. If it is placed 180 mm from a mask on an optical bench, where must the screen be placed in order to focus the diffracted light into a sharp image?

Answer

Object distance, *u* = 180 mm
Focal distance, *f* = 150 mm

The lens equation is

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \tag{19.8.8}$$

Therefore, image distance,

$$v = \left(\frac{1}{f} - \frac{1}{u}\right)^{-1} = \left(\frac{1}{150} - \frac{1}{180}\right)^{-1} = 900\text{mm} \tag{19.8.9}$$

i.e. 1080 mm from the object.

As in the previous question, a convex lens has a focal length of 150 mm and is placed 180 mm from a mask on an optical bench. Where must the screen be placed in order to observe the diffraction pattern?

Answer

The diffraction pattern can be observed at the back focal plane, located at the focal distance away from the lens, on the opposite side from the aperture, i.e. 150 mm from the lens, on the opposite side from the aperture.

As in the previous questions, a convex lens has a focal length of 150 mm and is placed 180 mm from a mask on an optical bench, giving an image distance of 900 mm. What is the magnification of the object in this set-up?

Hint: Magnification, $m = v / u$

Answer

Image distance, v = 900 mm
Object distance, u = 180 mm

Therefore, magnification

$m = v / u = 900 / 180 = 5$

As in the previous questions, a convex lens has a focal length of 150 mm and is placed 180 mm from a mask on an optical bench, giving an image distance of 900 mm. If the above lens is 56 mm in diameter, what is the finest grating size that could be resolved theoretically using light of wavelength 0.6328 μm?

Hint: In order to resolve the grating, the two first-order diffraction spots must be included in forming the image.

Answer

Screen distance, L = 0.90 m
Pattern spacing, x = 0.5 x lens diameter = 28 mm (to include 2 diffracted spots)
Wavelength, λ = 0.6328 μm

Therefore

$$s_{\min} H'' \lambda L / x = 0.6328 \times 10^{-6} \times 0.90 / 0.028 = 20.03 \text{mm} \tag{19.8.10}$$

The theoretical resolution of a microscope is given by

$$d_{\min} = \frac{\lambda}{n \sin \alpha} \tag{19.8.11}$$

where n is the refractive index of the medium (n = 1 for air) and sinα is known as the numerical aperture, N.A. (commonly printed on the side of a lens). If a microscope can just resolve a "400 lines per mm" grating, what would the N.A. of the lens be?

Answer

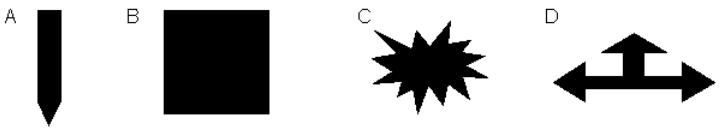
Resolution,

$$d_{\min} = \left(400 \text{mm}^{-1}\right)^{-1} = 2.5 \text{mm} \tag{19.8.12}$$

Therefore, N.A.

$$\sin \alpha = \lambda \left(n d_{\min}\right) = 0.6328 \times 10^{-6} / \left(1 \times 2.5 \times 10^{-6}\right) = 0.253 \tag{19.8.13}$$

A diffraction pattern shows just two-fold symmetry. Which *one* of these apertures could not have produced such a pattern?

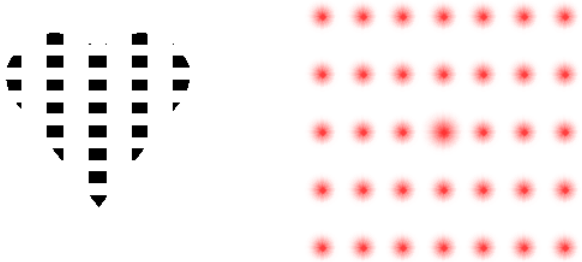


Answer

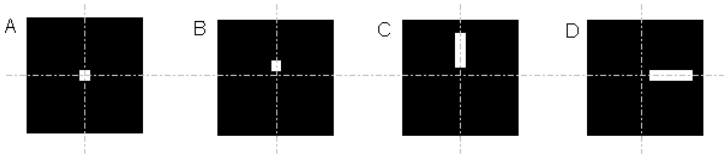
B. This aperture will produce a four-fold symmetry diffraction pattern, whereas the others will produce two-fold patterns. Diffraction patterns always have at least two-fold symmetry, even if the aperture has no symmetry. This can be seen

mathematically in the nature of the Fourier Transform.

The following heart-shaped aperture produces the adjacent diffraction pattern.



Which of the following masks should be placed in the back focal plane in order to best study the horizontal stripes of the aperture in the image? Dashed lines are shown to identify location of central diffraction spot with respect to mask.



Answer

C will be best. This will give a dark field image with respect to the horizontal stripes of the aperture, but with higher detail than B since more diffracted spots are going into making up the image.

A will give the bright field image in which the contrast of the strips will not be optimal.

B will give the dark field image with respect to the horizontal stripes of the aperture, with high contrast.

D will give a dark field image with respect to the vertical stripes of the aperture.

This page titled 19.8: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

20: Diffusion

https://www.doitpoms.ac.uk/tlplib/diffusion/index.php

20.1: Fick's First Law of Diffusion

20.2: Fick's Second Law of Diffusion

20.3: Applications of Diffusion

20.4: Interdiffusion

20.5: Microstructural Effects

20.6: Temperature Effects

This page titled 20: Diffusion is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

20.1: Fick's First Law of Diffusion

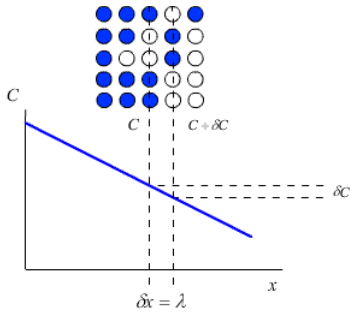
We have looked at the mechanisms of on an atomic scale. We now want to examine the emergent properties of these mechanisms when there are a lot of atoms. The first thing to note before we start is that no real material has a perfect structure. There will some amount of vacancies or other imperfections present. Therefore if we add impurity atoms to the material, they will be able to move around the material at some rate. If they are interstitial they will move around at a faster rate since they do not require any vacancies to move.

The second thing to note is that if the impurity atoms are distributed evenly throughout the material such that there is no concentration gradient, their random motion will not change the concentration of the material. Nor will there be a net movement of atoms through the material. We are interested in the cases where there is some kind of energy difference in the material, which causes a net movement of atoms. This can be caused by concentration differences, electric fields, chemical potential differences etc.

We will first look at the case of concentration differences. The fact that a concentration difference causes diffusion should be familiar to everyone, particularly in the case of liquids and gases. Consider adding a drop of ink to a bowl of water. The ink will diffuse through the water until the concentration is the same everywhere. There is no force causing the ink particles to diffuse through the water. It is in fact a statistical result of the random motion of the particles.

We can use Fick's laws to quantitatively examine how the concentrations in a material change.

Consider a crystal lattice with a lattice parameter λ, containing a number of impurity atoms. The concentration of impurity atoms, C (atoms m^{-3}) may not be constant over the whole crystal. In this case there will be a concentration gradient across the crystal, which will act as a driving force for the diffusion of the impurity atoms down the concentration gradient (i.e. from the area of high concentration to the area of low concentration).



Fick's first law relates this concentration gradient to the flux, J, of atoms within the crystal (that is, the number of atoms passing through unit area in unit time)

Fick's first law (derivation here) is

$$J \equiv -D\left(\frac{\partial C}{\partial x}\right) \tag{20.1.1}$$

D is the diffusivity of the diffusing species.

Our equation relating the mean diffusion distance to time can now be modified to be in terms of this parameter:

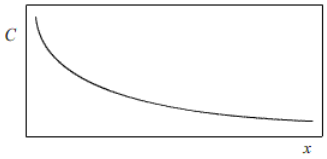
$$\begin{aligned} \overline{x} &= \lambda\sqrt{\nu t} \\ D &= \tfrac{1}{6}\nu\lambda^2 \\ \overline{x} &= \sqrt{6Dt} \\ \overline{x} &\approx \sqrt{Dt} \end{aligned} \tag{20.1.2}$$

The animation below demonstrates Fick's 1st Law with respect to a fluid.

This page titled 20.1: Fick's First Law of Diffusion is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

20.2: Fick's Second Law of Diffusion

Fick’s second law is concerned with concentration gradient changes with time.



By considering Fick’s 1st law and the flux through two arbitrary points in the material it is possible to derive Fick’s 2nd law.

$$\frac{\partial C}{\partial t} = D\left(\frac{\partial^2 C}{\partial x^2}\right) \tag{20.2.1}$$

This equation can be solved for certain boundary conditions:

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

1. “Thin source”

Consider a semi-infinite bar with a small, fixed amount of solute material diffusing in from one end.



The amount of solute in the system must remain constant, therefore

$$\int_0^\infty C\,(x,t)\;dx = B \tag{20.2.2}$$

where B is a constant.

The initial concentration of solute in the bar is zero, therefore

$$C\left(x, t=0\right)=C_0 \tag{20.2.3}$$

These boundary conditions give the following solution:

$$C\left(x, t\right)=\frac{B}{\sqrt{\pi D t}} \exp\left(\frac{-x^2}{4Dt}\right) \tag{20.2.4}$$

2. “Infinite source”

A semi-infinite bar with a constant source (i.e. constant concentration) of solute material diffusing in from one end.



In this case, the solution is obtained by stacking a series of “thin sources” at one end of the bar, and summing the effects of all of the sources over the whole bar.

The initial concentration of solute in the bar is C_0, therefore

$$C\left(x, t=0\right)=C_0 \tag{20.2.5}$$

The concentration of solute at the end of the bar is a constant, C_s, therefore

$$C\left(x=0, t\right)=C_s \tag{20.2.6}$$

These boundary conditions give the following solution:

$$C\left(x, t\right)=C_s-\left(C_s-C_0\right) erf\left(\frac{x}{2\sqrt{Dt}}\right) \tag{20.2.7}$$

erf{x} is known as the *error function* and results from the summation of the thin sources at the end of the bar. It is defined as

$$erf\left(x\right)=\frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2) du \tag{20.2.8}$$

The integral can only be solved numerically with a computer, so erf tables are used to solve the diffusion equation where necessary.

This animation shows the applications of Fick’s 2nd law and its solutions.

3. Non-Analytical Solutions

For more complicated situations we cannot obtain an analytical solution for Fick’s 2nd law. In these cases numerical analysis is used. Solutions obtained in this way are approximations, however, they can be made as precise as needed.

The following demonstration shows how numerical analysis can be used to approximate solutions for various conditions.

Note: This animation requires Adobe Flash Player 8 and later, which can be downloaded here.

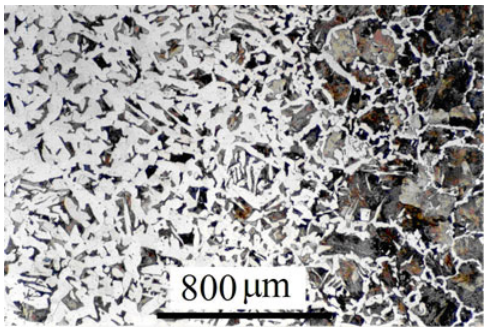
This page titled 20.2: Fick's Second Law of Diffusion is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

20.3: Applications of Diffusion

Diffusion is a key process in much of materials science. We will examine some applications more closely here:

Carburisation

Carburisation is the process by which carbon is diffused into the surface of steel in order to increase its hardness. The carbon forms carbide precipitates (particularly if the steel contains carbide forming elements such as manganese or molybdenum) which pin dislocations and prevent slip, thus making the material harder. However, the increased carbon content reduces the toughness of the material. In most applications it is important that the surface of the steel is hard, but the bulk material can remain softer without detriment to the properties of the component. Thus, carbon is often diffused in from the outer surfaces to obtain a material that is hard on the surface but tough in the bulk.



A carburised steel, showing increased carbon content on the outer surface See Micrograph No 289

This is done by heating the steel in a carbon atmosphere, so that there is a concentration gradient of carbon across the interface. Carbon diffuses into the steel, and the elevated temperature speeds up the process. The concentration profile of carbon is governed by Fick's second law, as there is effectively an infinite source of carbon.

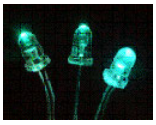
Nuclear Waste

In this case diffusion causes a problem that needs to be solved. Radioactive waste from nuclear energy production must be stored in such a way that the radioactive atoms do not diffuse out of the container until the radioactivity levels have sufficiently dropped. This can be a very long time indeed: often around 1000 years. Thus, the container must be made of a material in which the diffusivity of the atoms is very low (and the container must be very thick, to increase the diffusion distance). This will ensure that the time taken for atoms to diffuse out of the container is as long as possible.

Generally, the radioactive atoms are suspended in a glass matrix, such as borosilicate glass. The diffusivity of the atoms in this glass is low, thus the atoms are less likely to diffuse out of the glass before they have ceased to be radioactive. The glass is then sealed inside steel containers and buried deep in the ground under rocks, in remote areas away from populated regions.

Semiconductors

Semiconductors can be made by doping one material (often silicon) with a small number of atoms of another material of a different valency. This is known as doping, and means that there is an excess of charge carriers in the material (electrons if the valency of the dopant is greater than that of the silicon, or holes if it is less). For more details on this see the TLP on semiconductors.



Gallium-Nitride semiconductor LEDs from the Cambridge centre for gallium nitride web site

The doping is often carried out by diffusion methods: the silicon is placed in a gas of the dopant atoms and heated to high temperatures. The dopant atoms diffuse down the chemical potential gradient into the silicon. As with carburisation, this process follows Fick's second law.

In practice, the diffusion process occurs in two steps. After the initial diffusion described above, the atoms will be concentrated mainly on the surface of the silicon. The sample must therefore be annealed in order to “drive in” the atoms, so that they penetrate beyond the surface.

This page titled 20.3: Applications of Diffusion is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

Welcome to the Engineering Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Welcome to the Engineering Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

Welcome to the Engineering Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

CHAPTER OVERVIEW

21: Dislocation Energetics and Mobility

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

On completion of this TLP you should:

- Understand why a dislocation forms and how the dislocation width can be determined using an atomistic model, by minimisation of misfit energy due to in-plane stress and misalignment.
- Appreciate how the energy changes as a dislocation moves and how the parameters of atomic spacing, Shear modulus and Poisson ratio affect the total misfit energy.
- Be familiar with the concepts of Peierls stress and lattice resistance and compare theoretical and experimental values. Also understand the difference between the linear elastic analytic solution used by Peierls and the atomistic model.

Before you start

- Look through the TLP on Introduction to Dislocations
  - https://www.doitpoms.ac.uk/tlplib/dislocations/index.php
- These pages on dislocation width and Peierls Nabarro stress will be particularly useful.
  - https://www.doitpoms.ac.uk/tlplib/dislocations/dislocationwidth.php
  - https://www.doitpoms.ac.uk/tlplib/dislocations/slip_via_dislocation.php
- This LD shows what happens when dislocations interact to cause work hardening. It also covers the formation of dislocations (Frank Read sources).
  - https://www.doitpoms.ac.uk/ldplib/work-hardening-polycrystals/index.php

21.1: Introduction
21.2: Making a Dislocation
21.3: Join the Crystals to Form the Dislocation
21.4: Dislocation Width
21.5: Form of the Displacement
21.6: Change in the Misfit Energy of a Dislocation as it Moves
21.7: Peierls Energy
21.8: What is the Peierls Stress?
21.9: Determining the Peierls Stress
21.10: Lattice Resistance
21.11: Uses and Limitations of the Atomistic Model
21.12: Summary
21.13: Questions

This page titled 21: Dislocation Energetics and Mobility is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.1: Introduction

Plastic deformation of crystals generally proceeds by the propagation of dislocations along the slip planes. In most brittle crystals and some metals, it is the changes in misfit energy as the dislocation moves that are the predominant obstacle to dislocation motion. These arise from atoms in the crystal lattice being displaced from their equilibrium positions, so this effect is known as lattice resistance.

Why do dislocations ever form? Deformation can be thought of a kinetic process since dislocation motion (the most common cause of plastic flow) requires the breaking and reforming of bonds, like in a chemical reaction.

In this TLP we will be setting up an atomistic model, rather than using continuum elasticity as in the Peierls Nabarro model, where it is assumed that in-plane strains do not change as the dislocation moves.

It can be shown that despite the general form of the expression it is the same as that derived by Peierls, but with a slightly lower magnitude.

The misfit energy is made up of in-plane strains and misalignment strains (across the slip plane). This energy changes as the dislocation moves. The misalignment energy increases as the dislocation width increases, so act to localise the misfit strains. Whereas the in-plane strain energy decreases as dislocation width increases, so acts to spread the misfit strain over a larger region. The final arrangement of atoms results in a minimisation of the overall misfit energy for a given dislocation width.

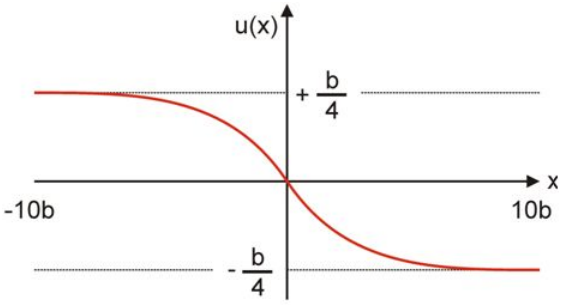
Considering the energy changes as the dislocation moves allows us to calculate the Peierls stress, required to move the dislocation, and see that it is exponentially dependent on the dislocation width. The Peierls stress depends on the atomic spacing both normal to the slip plane and parallel to it – in fact it is extremely sensitive to the ratio of lattice parameters b/d.

This page titled 21.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.2: Making a Dislocation

We will consider the changes in atomic positions in a given plane of atoms when an edge dislocation is formed in a simple cubic lattice. If the extension of the dislocation is large compared with the atomic spacing b, the horizontal displacement u varies slowly from atom to atom. The relative displacement of neighbouring atoms within each half crystal is much smaller than b.

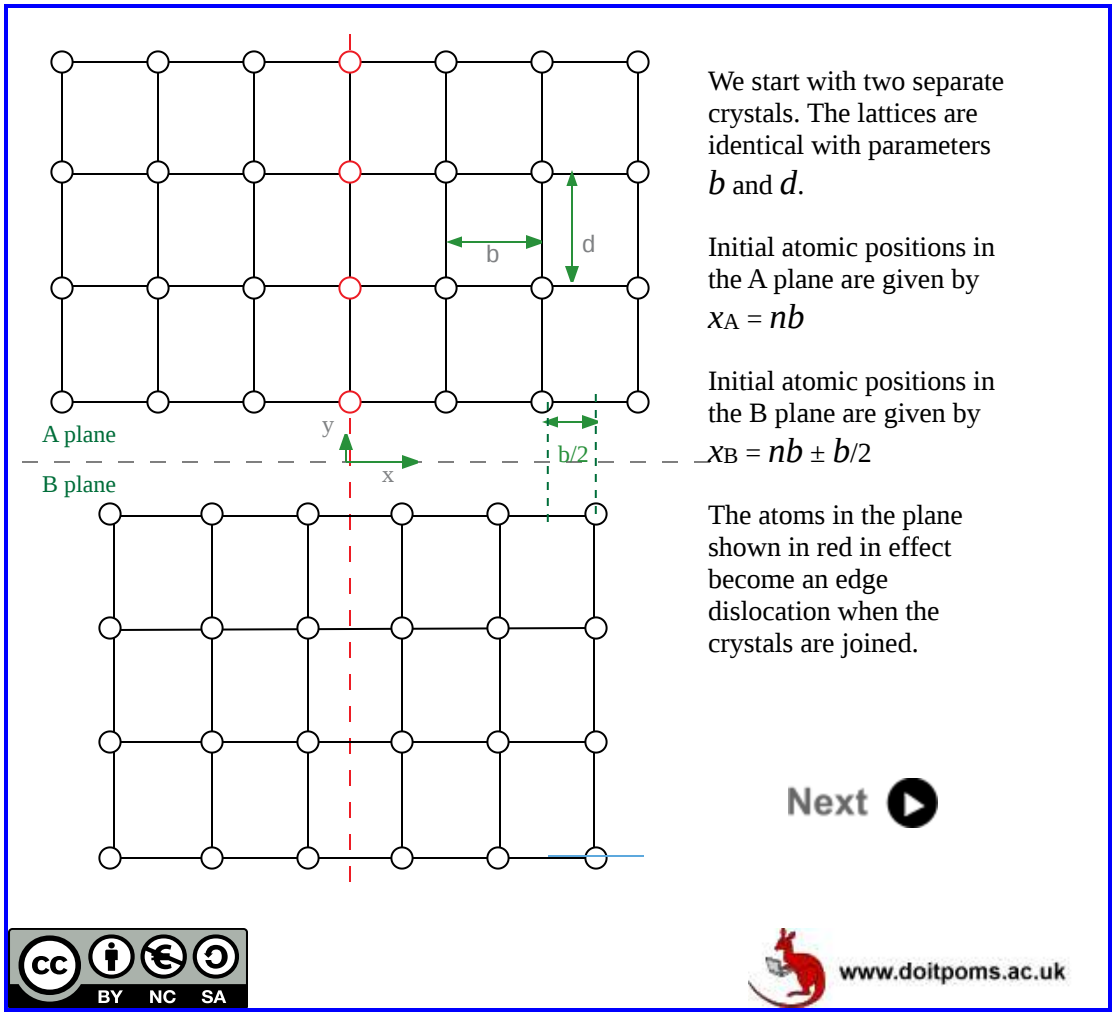
We start out by considering the initial atomic positions in just two planes **A** and **B**.



The variation of the displacement u with distance from the dislocation line

It is clear from the above figure that a tan function would be consistent with the displacement. Peierls showed this was the case.

- The following animation shows how the model is set up using the initial atomic positions in just two planes; here we are making the approximation that the displacements of atoms in the planes on either side (in the y direction) are small enough to be neglected from the overall energy. We are able to show that the final displacement (in the x direction) varies as an arctan function, as demonstrated by Peierls.



This page titled 21.2: Making a Dislocation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.3: Join the Crystals to Form the Dislocation

We have determined that the displacements of the atoms are:

$$u_A(x) = -\frac{b}{2\pi}\tan^{-1}\left(\frac{x_A}{w}\right) \tag{21.3.1}$$

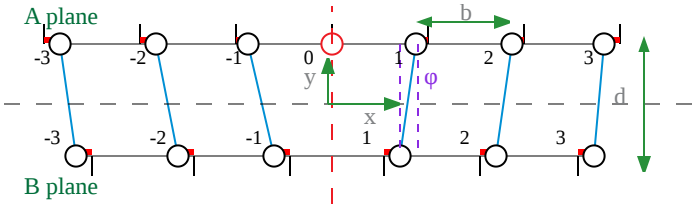
$$u_B(x) = -\frac{b}{2\pi}\tan^{-1}\left(\frac{x_B}{w}\right) \tag{21.3.2}$$

From below, the displacements of the atoms on the **A**-plane are symmetrical on either side of the dislocation line, and are zero at the centre. It is clear that the misfit around the dislocation is of two types: a strain in the planes above and below the dislocation and a misalignment of the atoms across the slip plane.

The process of determining the energies is shown in the following animation:

We know that the misfit energy of a dislocation is equal to $\alpha G b^2$ where α is about 1/2 and b is the Burgers vector, derived using continuum linear elasticity.

In continuum, the properties are the same everywhere, so there would be no energy change from place to place. Therefore we need to think atomistically.



There are two components of the misfit energy. These are:

In-plane strains between atoms in the same plane (parallel to slip plane)

Misalignment of atoms across the slip plane

Next

www.doitpoms.ac.uk

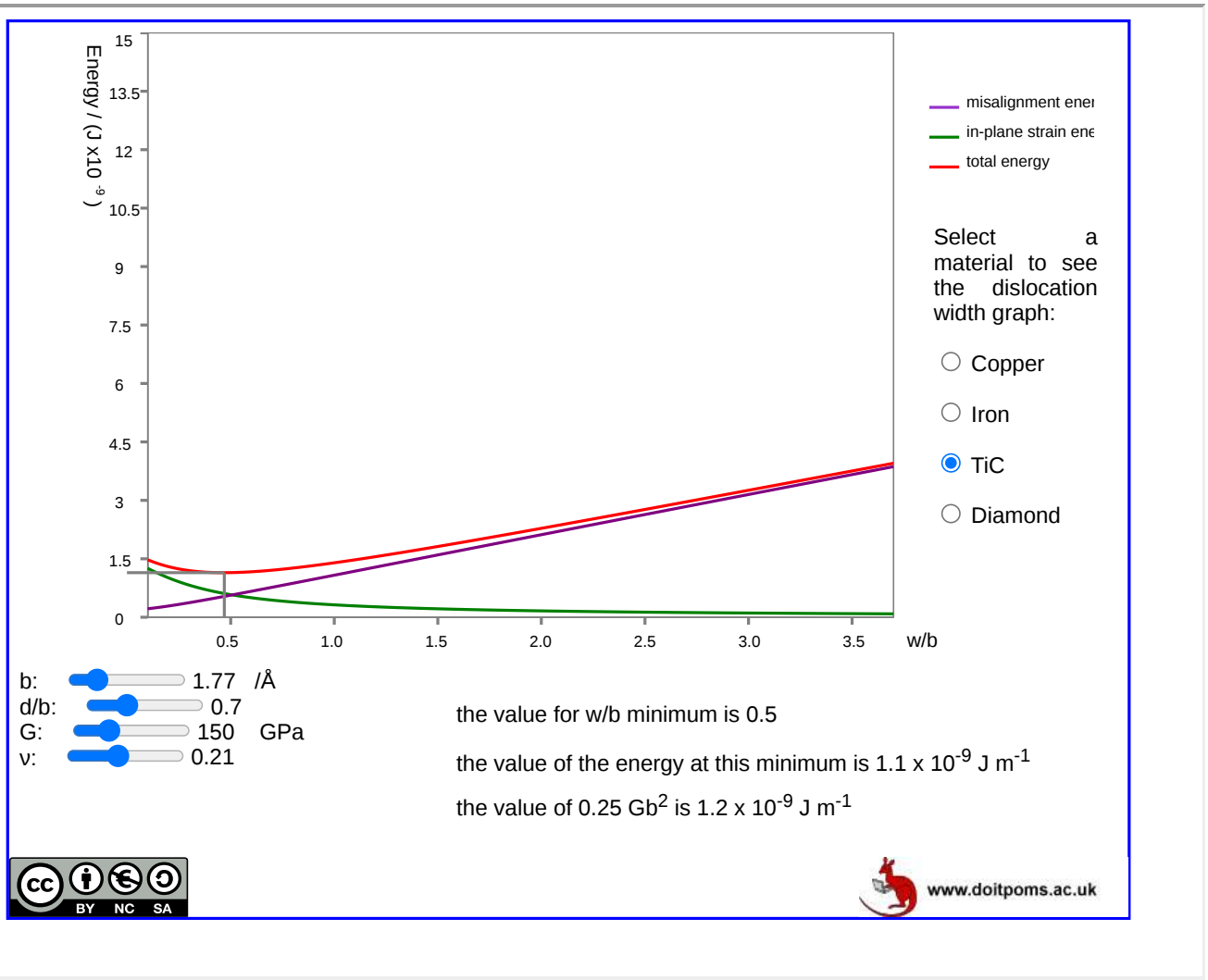
This page titled 21.3: Join the Crystals to Form the Dislocation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.4: Dislocation Width

For a given atomic configuration, we can work out the total energy and then try different values of w until we find the minimum. Practically (see graph below), this can be done by summing the energy over planes from n=-1000 to n=+1000 either side of the “extra plane of atoms”, since the effects of increasing the number of planes beyond this is negligible.

As w increases, the decrease in energy associated with localising the in-plane strains (which is why we said a dislocation should never form) is offset by the increase in misalignment energy, which increases with w. There is therefore a minimum in the misfit energy, which gives the dislocation width.

The following graph shows the sums of the misfit energies (in-plane, misalignment and total) resulting from the displacements of planes either side of the initial “extra half plane of atoms”. We can determine the dislocation width (w/b) for given parameters by considering the minimum of the total misfit energy curve.



If we look at the value of $0.25G\mathbf{b}^2$ for the four materials given, it is similar to the energy value calculated by the model. This shows an approximate agreement with the linear elastic solution.

What determines w/b?

- What we are interested in is the ratio of d/b. As d/b increases, w/b increases. d/b is determined by crystal structure.
  - **b** is the Burger's vector
  - d is the spacing of the close packed planes
  - We also need to take into consideration partial dislocations. For example, copper is ccp so the slip system is {111} <110>. But the dislocations dissociate into two partial dislocations $\frac{a}{6}$<112>; hence the magnitude of b will be $\frac{a}{\sqrt{6}}$ where a is the lattice parameter.
  - Therefore, the ratio of d/b in copper will be $\frac{a}{\sqrt{3}} \div \frac{a}{\sqrt{6}} = \sqrt{2}$
- Poisson ratio (ν) affects the in-plane component of the misfit energy. A higher Poisson ratio means the width of the dislocation is greater.

What determines the total energy?

- The total energy is affected by the magnitude of $G\mathbf{b}^2$, where G is the shear modulus and **b** is the Burger's vector.

The following clip shows the free surface of a Cd single crystal subject to tensile testing. Slip is occurring on a particular set of planes, and the set of ridges that form on the free surface are created by the arrival of sets of dislocations.

https://www.doitpoms.ac.uk/tlplib/di...-1_Cd_slip.mp4

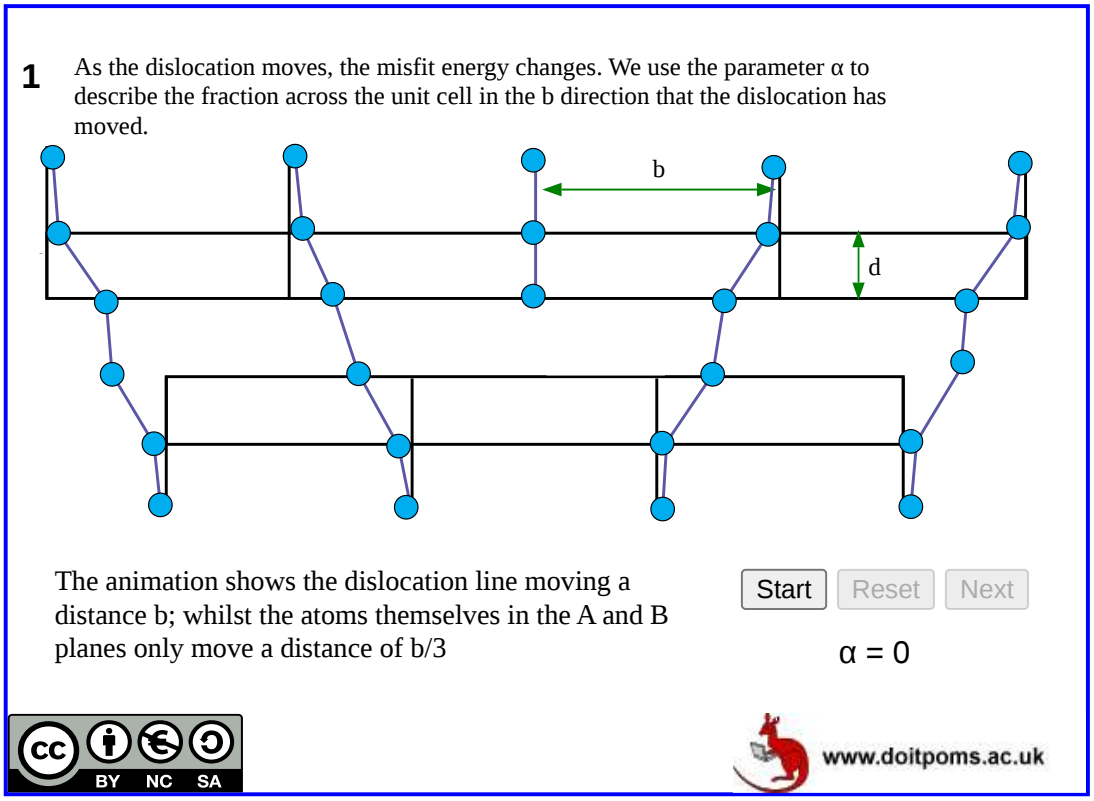
This page titled 21.4: Dislocation Width is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.5: Form of the Displacement

We now need to modify the stationary dislocation model to estimate the change in misfit energy of a dislocation as it moves.

The half-plane of atoms moves by much less than the dislocation. The half-plane of atoms was used as the origin, so we think of the dislocations as a moving origin from which we can estimate the displacements. We use the parameter alpha to describe the fraction across the unit cell across which the dislocation has moved.

As shown in the following animation, when the dislocation has moved through b (where α = 1) the extra half plane of atoms is now moved to the plane adjacent to the new half plane.



The dislocation has moved about twice as far as the extra half plane.

This page titled 21.5: Form of the Displacement is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

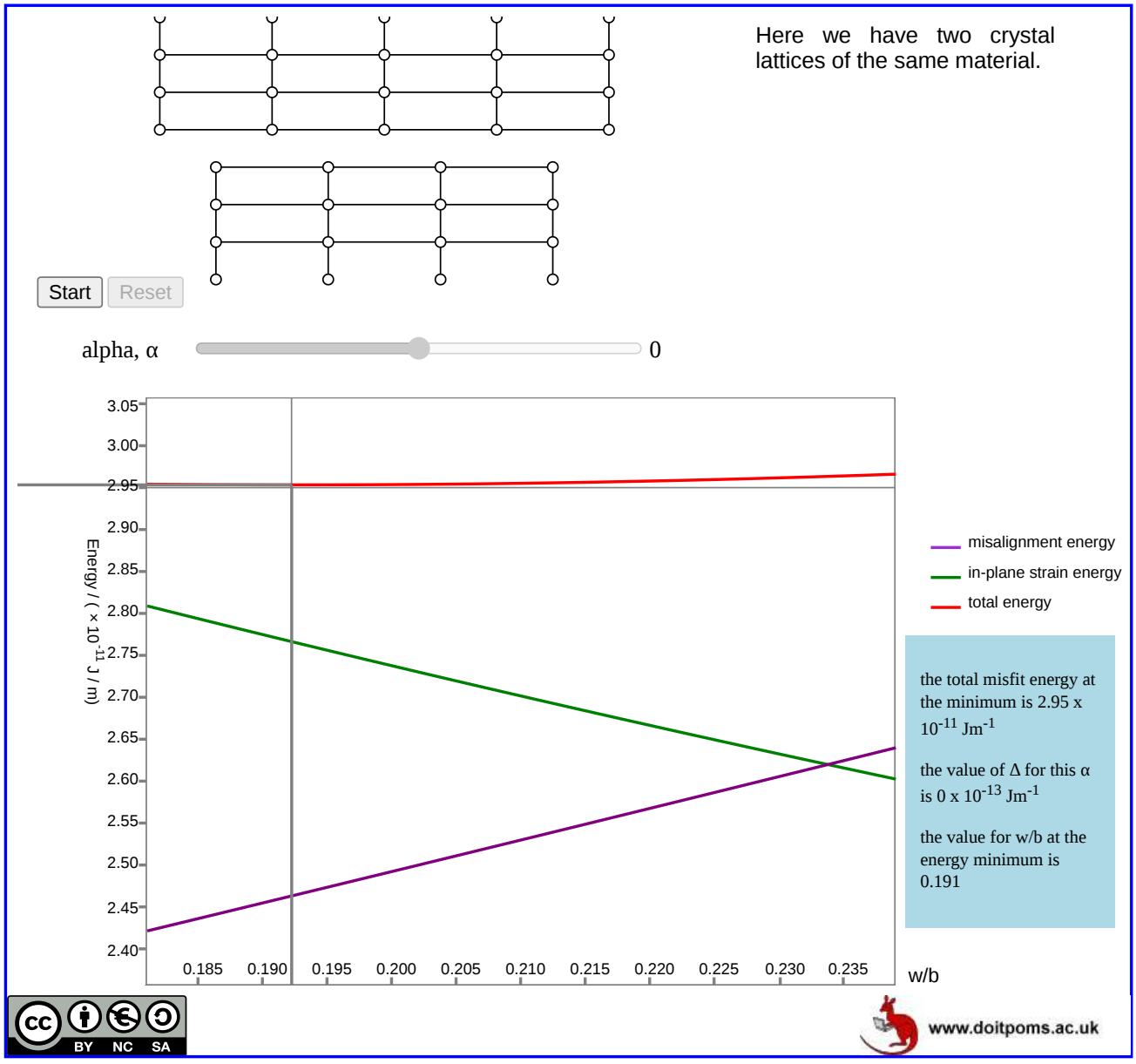
21.6: Change in the Misfit Energy of a Dislocation as it Moves

The dislocation remains fixed (at the origin) and the lattice moves around it, therefore the initial positions of atoms xA and xB are given by

$$x_A = nb - \alpha b \tag{21.6.1}$$

$$x_B = nb - b/2 - \alpha b \tag{21.6.2}$$

We use the same method as for the α = 0 case to determine the in-plane displacement and misalignment. As a result, the potentials for different α values can be determined. The following animation (which is a magnified version of the strain energy graph on the previous page) demonstrates how changing the parameter alpha affects the position of the energy minimum and the dislocation width.



By calculating the misfit energy for each value of alpha, we can determine the change in misfit energy as the dislocation moves.

- $\Delta U_T(x)$ is the difference in total energy from α = x and α = 0.
- At first, as the dislocation moves across the unit cell $0<|\alpha|<0.25$, ΔU_T increases to a maximum.
- At the maximum (α = 0.25), the value of ΔU_T is the Peierls energy, ΔU_P
- Then for $0.25<|\alpha|<0.5$, ΔU_T decreases. At α = 0.5, ΔU_T = 0.

- w also changes as the dislocation moves across the cell as seen. This is because the configuration of the dislocation changes as the dislocation moves across the unit cell. For the energy maxima, the dislocation width is also a maximum.

This page titled 21.6: Change in the Misfit Energy of a Dislocation as it Moves is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.7: Peierls Energy

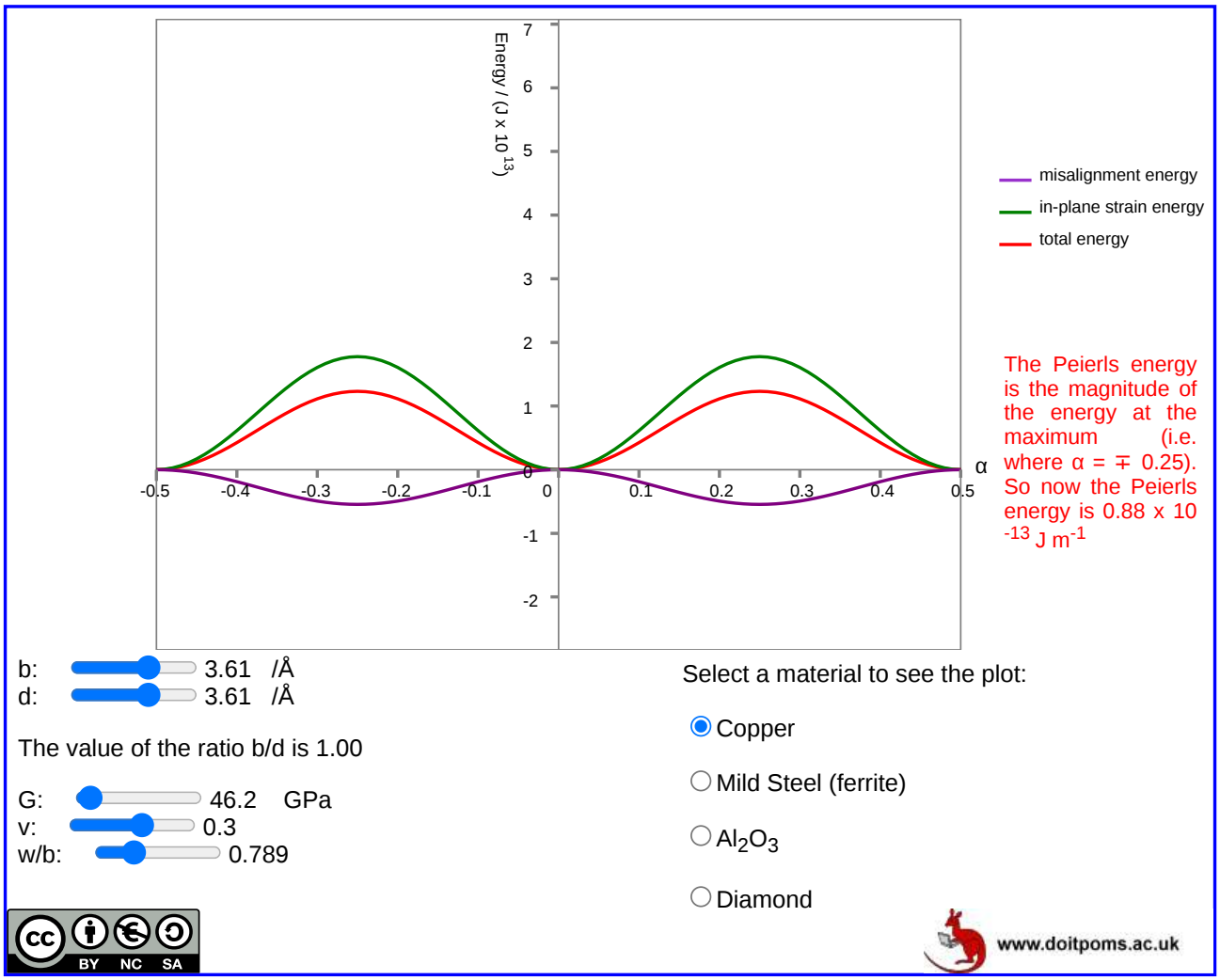
The maximum change in the total misfit energy is called the Peierls energy . It is the energy required to move unit length of dislocation over the resistance of the lattice. The following graph shows how the fraction of the unit cell across which the dislocation has moved affects the misfit energy. Adjusting parameters demonstrates how these affect the Peierls energy.

The misfit energy associated with in plane strains varies sinusoidally, as does the total energy. Whereas the misfit energy associated with the misalignments is greatest at the position of lowest overall energy. The period is b/2. Hence:

$$\Delta U_T(\alpha) = \frac{1}{2}\Delta U_P(1 - \cos 4\pi\alpha) \tag{21.7.1}$$

$\Delta U_T(\alpha) = \frac{1}{2}\Delta U_P(1 - cos4\pi\alpha)$

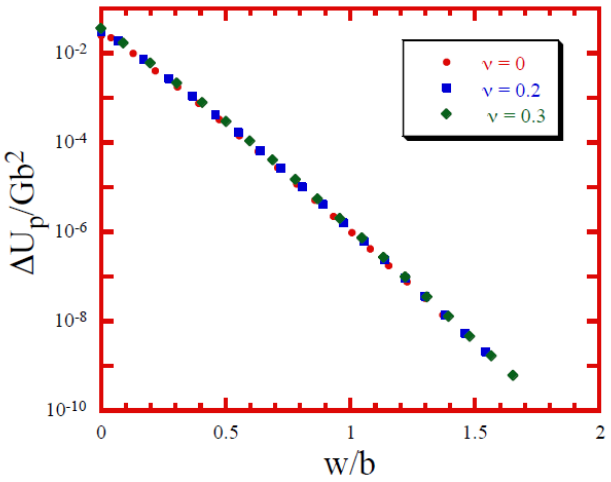
This is shown in the following interactive graph.



As can be seen in the magnified strain energy graph, the dislocation width also changes as the dislocation moves, although for simplicity simulations often assume it is constant.

The magnitude of the Peierl's energy (and hence the misfit energy) scales with G**b**^{2}. The Peierl's energy varies exponentially with w/b.

$$\frac{\Delta U_P}{Gb^2} \propto \exp\left(-\frac{w}{b}\right) \tag{21.7.2}$$



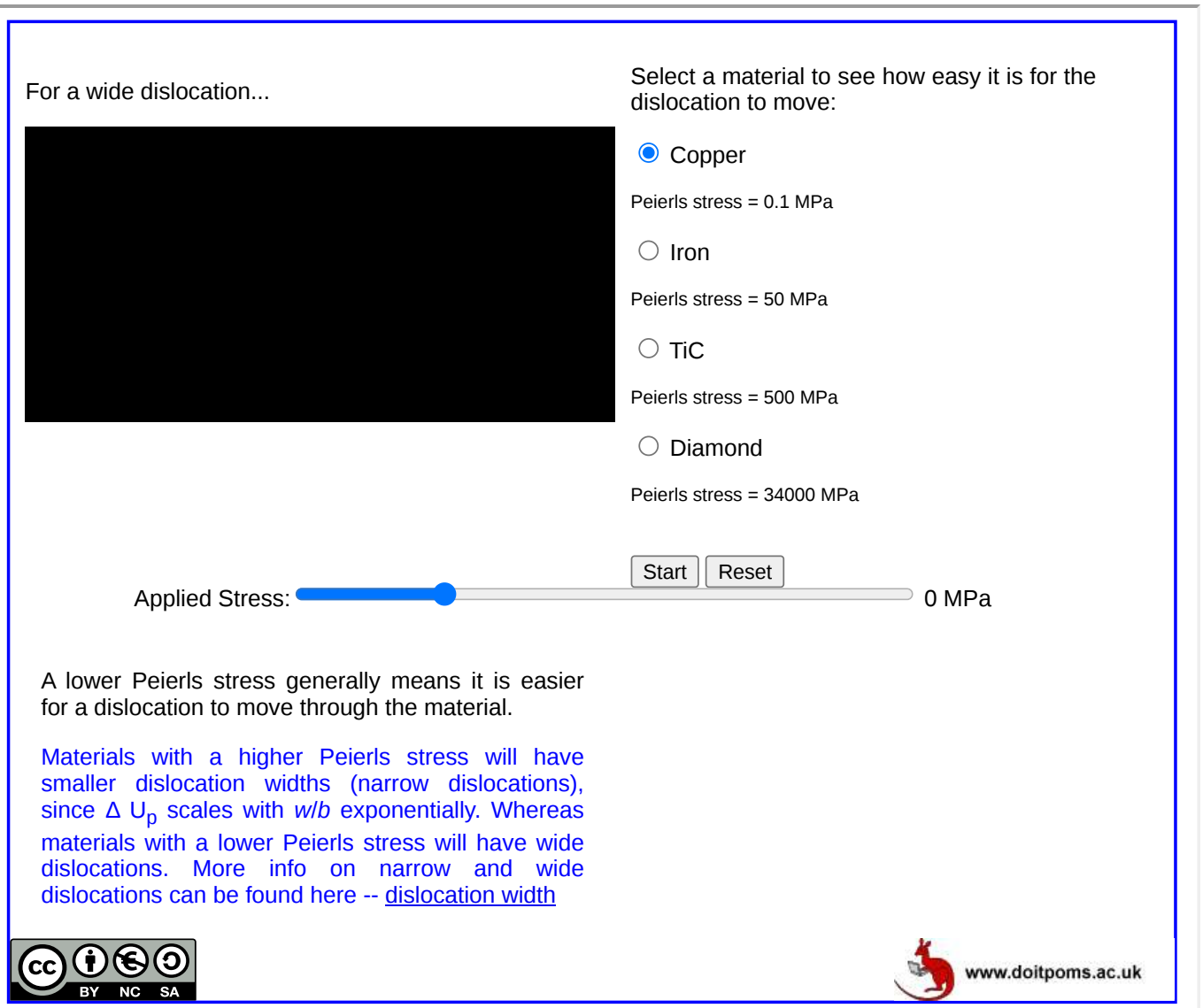
As can be seen in the misfit energy graph animation, the dislocation width also changes as the dislocation moves, although for simplicity, simulations often assume it is constant. For example, in the energy calculation, we assumed that w/b was a constant for any value of α.

This page titled 21.7: Peierls Energy is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.8: What is the Peierls Stress?

The Peierls stress is the minimum shear stress required to move a single dislocation of unit length in a perfect crystal. The magnitude of the Peierls stress determines the ability of the lattice to resist dislocation motion – the lattice resistance in the absence of thermal activation.

Another way is to think about the resistive force to dislocation motion as the magnitude of the gradient of strain energy curve. This means that the Peierls stress is proportional to the maximum gradient of the misfit energy curve. If we use the continuum model, the energy of the dislocation is independent of position. Any stress, however small, would set the dislocation into motion, because the dislocation is always in neutral equilibrium and so will move under any force. We know this is not right. We need to take into account the crystal structure on an atomic level such that the energy of the dislocation depends on its exact position. The dislocation will have several stable equilibrium positions which persist up until a stress is applied that exceeds a certain magnitude (the Peierls stress), at which point the dislocation can move.



This page titled 21.8: What is the Peierls Stress? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.9: Determining the Peierls Stress

The force required to move the dislocation is

$$F = \frac{\delta \Delta U_T(\alpha)}{\delta(\alpha b)} \tag{21.9.1}$$

Plotting the derivative of the energy-alpha graph gives a plot of the force required to move a dislocation. The corresponding graph gives the stress required to move the dislocation per unit length of dislocation.

$$\tau = \frac{1}{b^2}\frac{\delta \Delta U_T(\alpha)}{\delta(\alpha b)} \tag{21.9.2}$$

As we saw in expression (1) on the "Peierls energy" page.

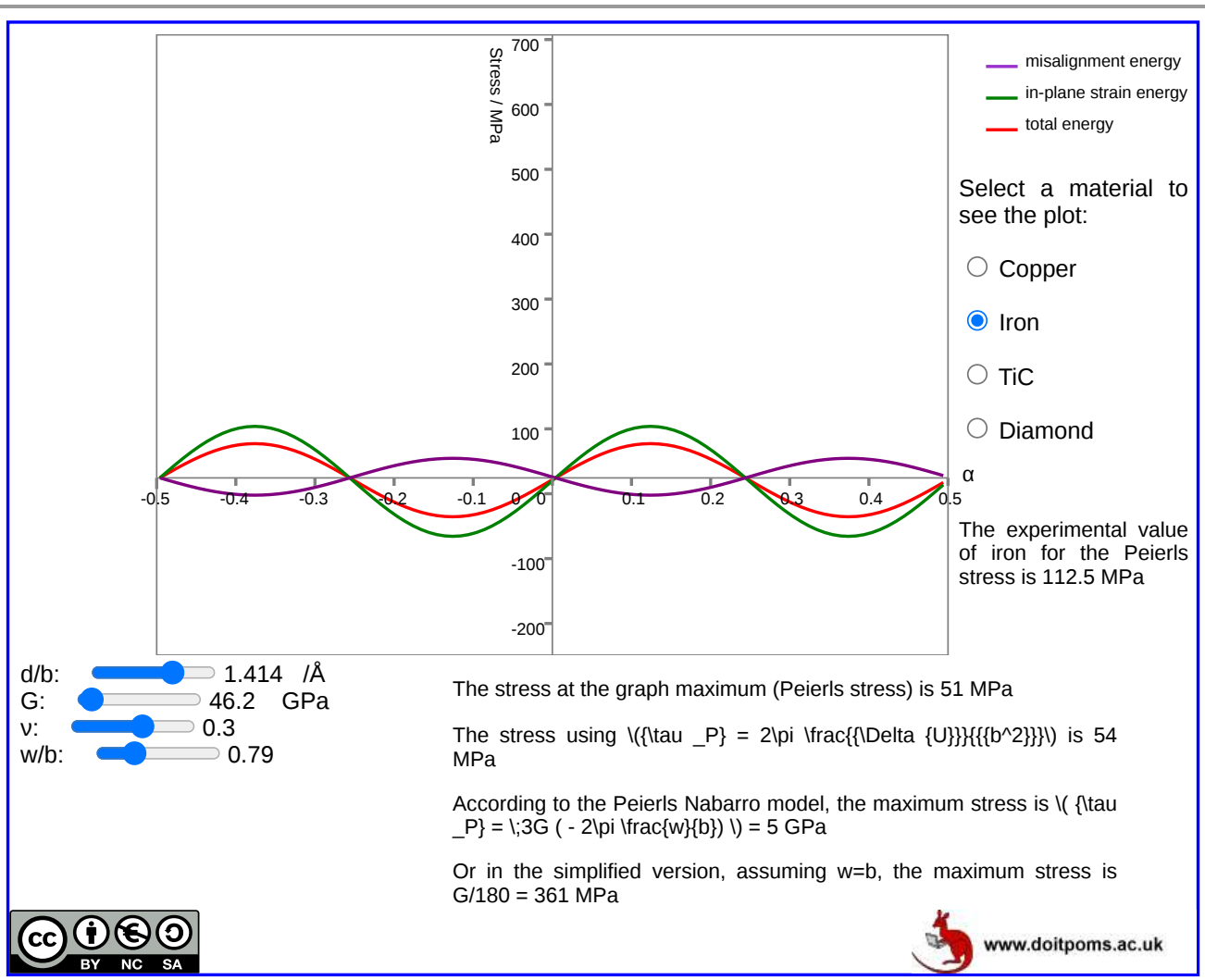
$$\Delta U_T(\alpha) = \frac{1}{2}\Delta U_P(1 - \cos 4\pi\alpha) \tag{21.9.3}$$

Hence, differentiating, we get

$$\tau_P = \frac{2\pi}{b^2}\Delta U_P \sin(4\pi\alpha) \tag{21.9.4}$$

The maximum stress is where $sin(4\pi\alpha) = 1$ The expression is often written as

$$\frac{\tau_P}{G} = \frac{2\pi}{Gb^2}\Delta U_P \tag{21.9.5}$$



*A note on this graph: It's important to remember that even though we are multiplying by G**b**^{2}, d/b also affects the dislocation width - increasing d/b increases the dislocation width as seen on the "determining w" graph. So for this graph if we impose a w/b, then we can see the effects of changing the magnitude of G**b**^{2} by adjusting the parameters G and b/d. Likewise if we impose a b/d we can see the effects of altering w/b through changing something else like the Poisson ratio.*

The graph shows the stress calculated using the maximum of the graph (expression 4) and also the differentiated expression for ΔU (equation 7). They are very similar, which is expected.

Therefore, using expression (2) from the "Peierls energy" page.

$$\frac{\tau_P}{G} \propto 2\pi \exp\left(-\frac{w}{b}\right) \tag{21.9.6}$$

The analytical solution provided by the "continuum elasticity" model (www.doitpoms.ac.uk/tlplib/dislocations/slip_via_dislocation.php) gives the constants as:

$$\tau_P = 3G \exp\left(-2\pi\frac{w}{b}\right) \tag{21.9.7}$$

which is sometimes simplified to $\frac{G}{180}$ if we assume that w = b (which is not in fact true as we have seen!). This solution gives us a theoretical shear stress – i.e. the stress required for uniform slip – but as can be seen from the above graph, it is several orders of magnitude higher than the actual Peierls stress.

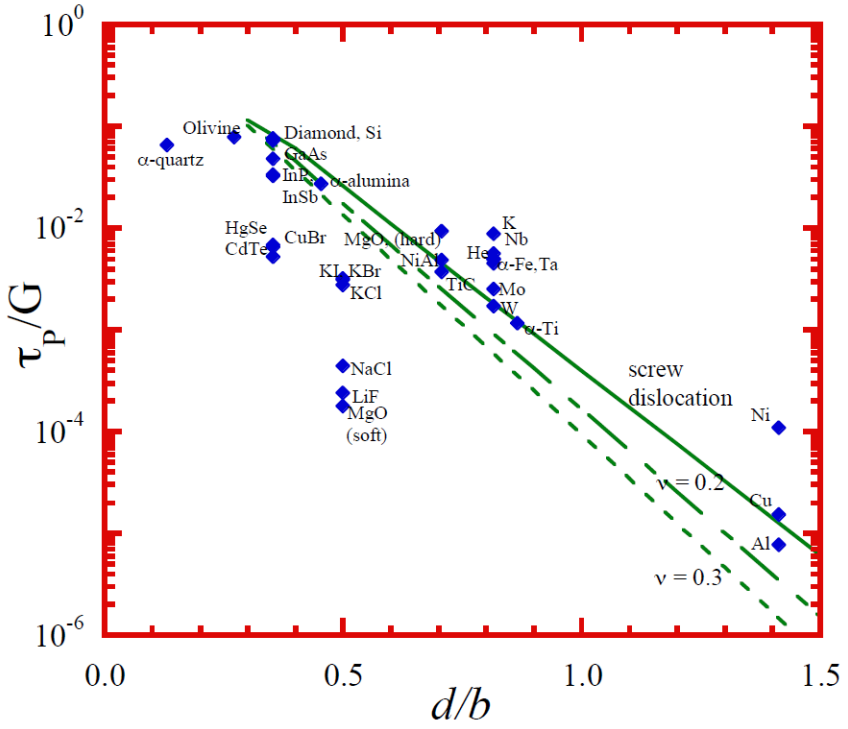
The analytical solution is from Peierls 1940 and Nabarro 1947. In essence, they assume that the in-plane strain is the surface of a semi-infinite elastic continuum and use that and the sum of misalignment energies to find the width at the equilibrium point. The elastic continuum (i.e. in-plane energy) is assumed to be fixed as is the width; therefore, the changes in energy are due only to the misalignment summation. This gives a wrong answer and indeed the maximum and minimum positions swap (U-tot Vs U-misalignment are out of phase). Since both the in-plane strain energy and the width change as the dislocation is displaced these should be included.

Compared with real values, the atomistic model is actually a small underestimate of the Peierls stress, but the analytical solution is a vast overestimate, which renders it is not very useful.

This page titled 21.9: Determining the Peierls Stress is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.10: Lattice Resistance

The higher the Peierls stress the higher the lattice resistance. The model showed an exponential dependence of Peierls stress on w/b (and hence d/b since we found that was the key factor affecting dislocation width).This exponential dependence is very important and is seen in materials experimentally.



That τ_P can be predicted for both metallic and covalent materials using the atomistic model; there is no obvious effect of bonding. d/b is the important factor. We often think of covalent materials as the "strongest", but this graph shows otherwise. Experimental determination of the Peierls stress is done by considering the yield stress at zero kelvin. It is seen that the materials tend to follow the trend for giant covalent crystals (diamond, TiC); but it is more complicated when there are coulombic forces coming in to play, for example in ionic crystals.

It is important to consider what slip system we are using. For example, halides (NaCl, LiF) and MgO have two slip systems – {100} <110> (for which there is good agreement) and {110} <110> (for which the model predicts too high a Peierls stress by $\approx 10^2$).

This page titled 21.10: Lattice Resistance is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.12: Summary

There are two types of energy associated with a dislocation

- In-plane energy – decreases as dislocation width increases, so acts to spread the misfit strain over a larger region
- Misalignment energy – increases as the dislocation width increases, so acts to localise the misfit strains
  - The dislocation width will be the value for which the sum of the two types of energy is a minimum
  - w/b is strongly dependent on d/b, where b is the atom spacing parallel to the slip plane and d normal to it.
  - Changes in misfit energy are the primary obstacle to dislocation motion.
  - Using the atomistic model with a moving origin allows us to estimate the energy as the dislocation moves, hence we can determine the Peierls energy and the Peierls stress.
  - Peierls stress increases exponentially as the dislocation width w/b decreases.

Going further

Books

Derek Hull and, D.J.Bacon, *Introduction to Dislocatioins* (*Volume 3 of Materials Science and technology, 5th Edition*), Elsevier, 2011, ISBN: 008096673X, 9780080966731

H.J.Frost and M.F.Ashby, *Deformation-mechanism maps: the plasticity and creep of metals and ceramics*, First Edidtion, Pergamon Press, 1982, ISBN: 0080293379, 9780080293379, http://engineering.dartmouth.edu/defmech/

Ulrich Messerschmidt, *Dislocation Dynamics During Plastic Deformation*, Springer-Berlin, Heidelburg, 2010, ISBN: 978-3-642-03176-2, 978-3-642-03177-9, https://doi.org/10.1007/978-3-642-03177-9

Paper

Howie,P.R., Thompson, R.P., Korte-Kerzel,.S., & Clegg,W.J. (2017), Softening non-metallic crystals by inhomogenenous elasticity. *Scientific Reports*, 7(1), 11602. https://doi.org/10.1038/s41598-017-09453-1

This page titled 21.12: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

21.13: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

What is the main cause of lattice resistance?

a	Change in the dislocation width as the dislocation moves
b	Changes in misfit energy as the dislocation moves
c	Change in the Peierls stress as the dislocation moves
d	Dislocations interacting and causing work hardening

Answer

B

What is the main problem with the Peierls continuum elasticity model?

a	It only includes the nearest-neighbour misalignment energy and does not include the next-nearest-neighbour term.
b	It assumes the in-plane energy is negligible and can be ignored
c	It assumes the total energy remains constant as the dislocation moves
d	It assumes that in-plane energy is constant and does not vary as the dislocation moves

Answer

D

How do the misalignment and in-plane energies change with w/b?

a	As w/b increases, both misalignment and in-plane energies increase
b	As w/b increase, both misalignment and in-plane energies decreases
c	As w/b increases, misalignment energy increases; in-plane energy decreases
d	As w/b increases, misalignment energy decreases; in-plane energy increases

Answer

C

What is the relationship between dislocation width and Peierls stress?

a	Peierls stress increases linearly as the dislocation width increases.
b	Peierls stress decreases exponentially as the dislocation width increases.
c	Peierls stress increases exponentially as the dislocation width increases.
d	Peierls stress decreases linearly as the dislocation width increases.

Answer

B

What is the Peierls stress?

a	The minimum stress for which fracture occurs
b	The tensile stress at which slip starts to occur
c	The minimum shear stress required to move a dislocation across the slip plane
d	The minimum stress at which plastic deformation begins

Answer

C

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

6. For what value(s) of alpha (fraction the dislocation has moved across the unit cell) is the misfit energy maximum?

◯ Yes	◯ No	a	Alpha = 0
◯ Yes	◯ No	b	Alpha = 0.25
◯ Yes	◯ No	c	Alpha = 0.5
◯ Yes	◯ No	d	Alpha = 0.75

Answer

A. no

B. yes

C. no

D. yes

This page titled 21.13: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

22: Introduction to Dislocations

https://www.doitpoms.ac.uk/tlplib/

Learning Objectives

On completion of this TLP you should:

- understand the nature of dislocations both conceptually and in a real crystal structure
- appreciate the motion of a dislocation under an applied stress
- be aware of the methods that can be used to reveal dislocations in a crystal structure

Before you start

Most of this package assumes no prior knowledge. However, some of the more detailed explanations use the following:

- Miller index notation for specification of crystal planes and directions
- Bravais lattice types

It should be possible to make use of the package even without knowledge of these concepts.

22.1: Introduction
22.2: Dislocations in 2D
22.3: Bubble Raft
22.4: Dislocation Glide
22.5: Dislocation Motion
22.6: Dislocation in 3D
22.7: Observing Dislocations
22.8: Summary
22.9: Questions

This page titled 22: Introduction to Dislocations is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

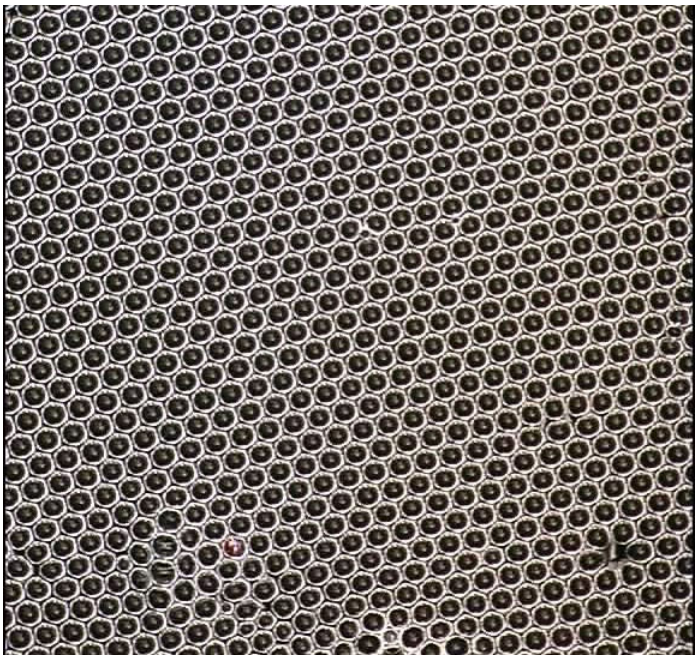
22.1: Introduction

The concept of the dislocation was *invented* independently by Orowan, Taylor and Polanyi in 1934 as a way of explaining two key observations about the plastic deformation of crystalline material:

- The stress required to plastically deform a crystal is much less than the stress one calculates from considering a defect-free crystal structure
- Materials *work-harden*: when a material has been plastically deformed it subsequently requires a greater stress to deform further.

Not until 1947 was the existence of dislocations experimentally verified. It took another ten years before electron microscopy techniques were advanced enough to show dislocations moving through a material.

The first section of this package deals with dislocations in the simplest way - in two dimensions. Much of the early work on dislocations was done using simple models such as bubble rafts. These can be tremendously instructive, and we use video clips and still pictures to demonstrate how plastic deformation occurs via dislocation motion.



A bubble raft. (Click on image to view a larger version.)

In real crystals, dislocations are three-dimensional. The structure of a dislocation in 3D can be more difficult to visualise. We look briefly at these structures.

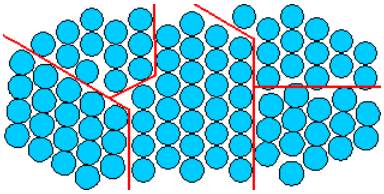
In the final part of this package, we consider some of the observations one can make that were first used to verify the presence of dislocations in real crystals.

This page titled 22.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.2: Dislocations in 2D

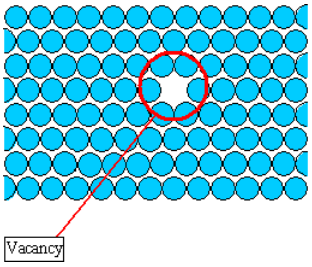
A 'raft' of equally sized bubbles floating on the surface of a liquid is a good large-scale model of a single plane of atoms in a crystal structure. The forces between the bubbles mimic the forces between atoms in a crystal. The bubbles pack to form a close-packed plane. If the raft is made carefully, it is possible to see a variety of structural features in the raft that also occur in real crystal structures, such as *grain boundaries*, *vacancies*, *dislocations* and *solute 'atoms'*.

A grain boundary in a 2D lattice is the interface between two regions of crystalline order. Each region or 'grain' has a different orientation with respect to some arbitrary axis perpendicular to the plane of the lattice.



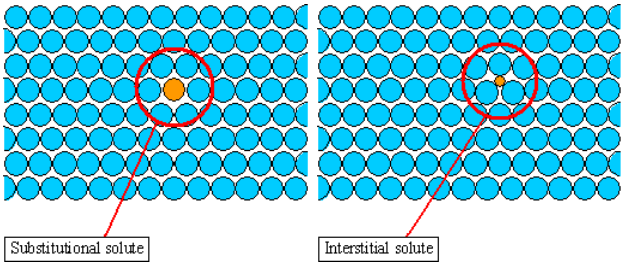
Grain boundaries

A *vacancy* is a point defect that arises when an atom is 'missing' from the ideal crystal structure.



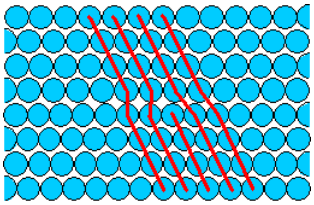
A vacancy

A *solute atom* in a crystal structure is an atomic species that is different from the majority of atoms that form the structure. Solute atoms of similar size to those in the host lattice may substitute for host atoms - these are known as *substitutional solutes*. Solute atoms that are much smaller than the host atoms may exist within normally empty regions (*interstices*) in the host lattice, where they are called *interstitial solutes*.



Substitutional and interstitial solutes. Note that some distortion of the host lattice occurs around the solutes.

A dislocation in a 2D close-packed plane can be described as an extra 'half-row' of atoms in the structure. Dislocations can be characterised by the Burgers vector which gives information about the *orientation* and *magnitude* of the dislocation.



A dislocation

This page titled 22.2: Dislocations in 2D is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.3: Bubble Raft

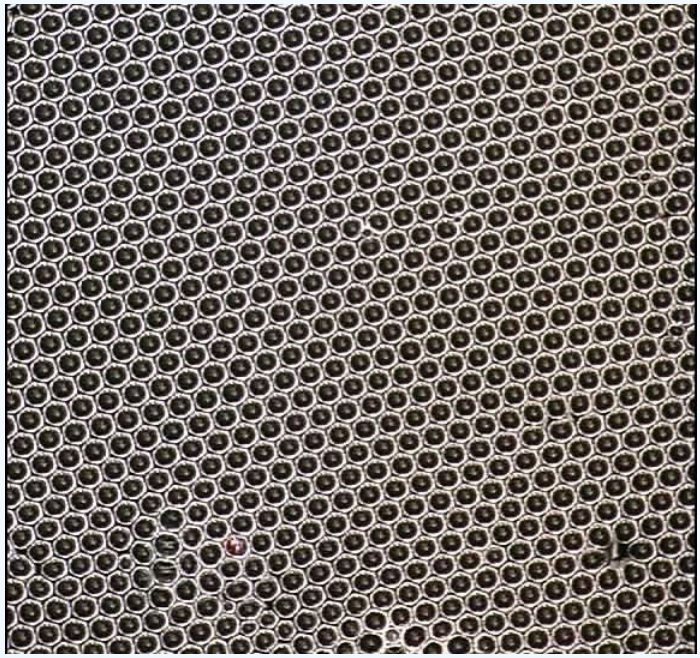
A bubble raft can be made by bubbling air through a soap solution, using a small air pump connected to a hollow needle. The size of the bubbles can be controlled by varying the flow of air through the needle, and by varying the depth below the surface of the liquid that the needle is submerged. Two bars at each end allow forces to be applied to the bubble raft.



Creating the bubble raft.

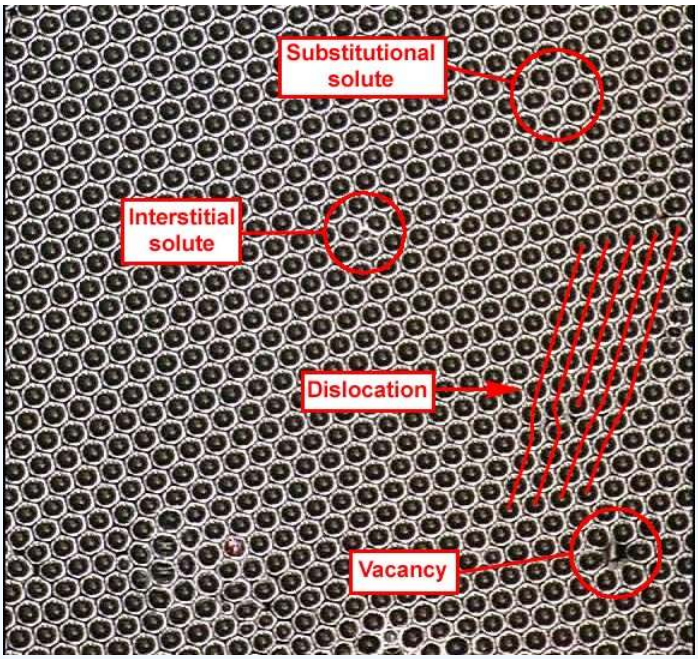
Examine the following still photograph of the bubble raft. The bubbles have been arranged approximately into a single crystal using gentle movement of the bars. The raft shows several defects that are analogous to crystalline defects. Try and identify *vacancies*, *dislocations*, *substitutional solutes* and *interstitial solutes*.

Example 22.3.1



There is at least one example of a vacancy, a dislocation, a substitutional solute and an interstitial solute in this bubble raft.

Solution



A vacancy, dislocation, substitutional solute and interstitial solute identified.

This page titled 22.3: Bubble Raft is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.4: Dislocation Glide

Dislocation motion along a crystallographic direction is called *glide* or *slip*. In the bubble raft experiment, dislocations glide when the raft is deformed. There must be a *local shear stress* in an appropriate direction on the dislocation for glide to occur. Dislocation glide allows plastic deformation to occur at a much lower stress than would be required to move a whole plane of atoms past another. These animations compare how plastic shear deformation occurs in a 2D primitive square lattice with and without dislocation glide.

https://www.doitpoms.ac.uk/tlplib/di...true-glide.mp4

Animation of slip by dislocation glide

https://www.doitpoms.ac.uk/tlplib/di...alse-glide.mp4

Animation of slip by movement of whole lattice planes

The stress required to cause slip by moving entire planes past one another, and the stress required to cause slip by dislocation motion can be estimated. The calculation shows that the stress required for slip is much lower when the mechanism of slip is dislocation motion, and from this we can conclude that slip *does occur by dislocation motion*.

This page titled 22.4: Dislocation Glide is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.5: Dislocation Motion

Watch the video clips of the bubble raft undergoing compressive, tensile and shear deformation. It may help you to watch each clip several times.

https://www.doitpoms.ac.uk/tlplib/di...t-stresses.mp4

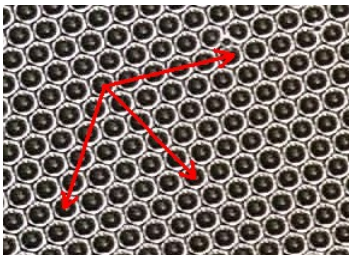
Video of bubble raft undergoing compressive and tensile deformation

https://www.doitpoms.ac.uk/tlplib/di...r-stresses.mp4

Video of bubble raft undergoing shear deformation

At small strains, the arrangement of the bubbles does not change. This is elastic deformation of the raft. The bubbles change shape and move slightly apart in an effort to maintain the lowest energy close-packed configuration.

At larger stains, plastic deformation occurs. The bubble raft rearranges by *dislocation motion*. Notice how dislocation motion occurs along three directions in the raft. These are the close-packed directions, along which the distance between bubble centres is smallest.



Close-packed directions in a bubble raft

In the different loading conditions, dislocations tend to move mainly along different sets of directions. In each case, the direction along which dislocations generally move is that with the highest resolved shear stress . Dislocations may nucleate near a different type of crystalline defect, such as a grain boundary, solute atom or vacancy.

This page titled 22.5: Dislocation Motion is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.6: Dislocation in 3D

In three dimensions, the nature of a dislocation as a line defect becomes apparent. The *dislocation line* runs along the *core* of the dislocation, where the distortion with respect to the perfect lattice is greatest.

There are two types of three-dimensional dislocation. An *edge* dislocation has its Burgers vector perpendicular to the dislocation line. Edge dislocations are easiest to visualize as an extra half-plane of atoms. A *screw* dislocation is more complex - the Burgers vector is parallel to the dislocation line. *Mixed* dislocations also exist, where the Burgers vector is at some acute angle to the dislocation line. In a 2D model such as the bubble raft, *only edge dislocations can exist*.

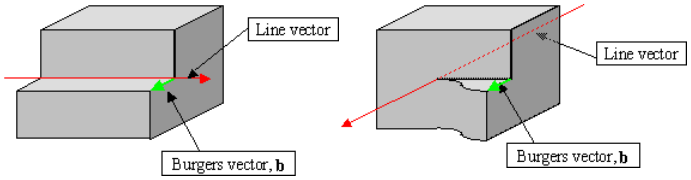


Figure 22.6.1: Edge and screw dislocations with line and Burgers vectors shown.

https://www.doitpoms.ac.uk/tlplib/di...islocation.mp4

Virtual reality model of an edge dislocation

https://www.doitpoms.ac.uk/tlplib/di...islocation.mp4

Virtual reality model of a screw dislocation

When a dislocation moves under an applied shear stress:

- individual *atoms* move in directions *parallel to the Burgers vector*;
- the *dislocation* moves in a direction *perpendicular to the dislocation line*.

An edge dislocation therefore moves in the direction of the Burgers vector, whereas a screw dislocation moves in a direction perpendicular to the Burgers vector. The screw dislocation 'unzips' the lattice as it moves through it, creating a 'screw' or helical arrangement of atoms around the core.

The ease of dislocation glide is partly determined by the degree of distortion (with respect to the perfect lattice) around the dislocation core. When the distortion is spread over a large area, the dislocation is easy to move. Such dislocations are known as *wide* dislocations, and exist in ductile metals.

Width of a dislocation

The width of a dislocation gives a measure of the degree of disruption a dislocation creates with respect to the perfect lattice. If the dislocation is regarded as the transition between slipped and unslipped areas of a slip plane, then the width of the dislocation is a measure of the sharpness of the transition. Formally, the edge dislocation width w is defined as the distance over which the disregsitry is greater than one quarter of the magnitude of the Burgers vector, b. The disregistry is the magnitude of the displacement of the atoms from their perfect crystal positions.

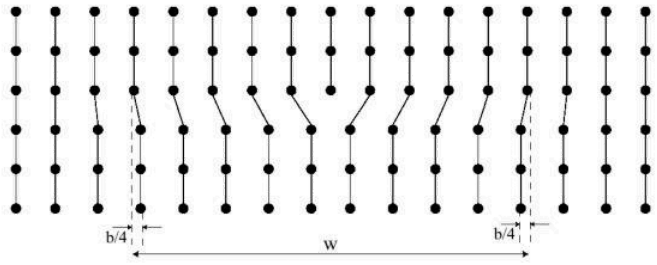


Diagram showing width of a wide dislocation

At the centre of a perfect edge dislocation, the disregistry is always b/2.

When w is several atomic spacings in dimension, the dislocation is wide; if w is of the order of one or two atomic spacings, it is narrow. Dislocation glide occurs most easily in wide dislocations - these are found in simple metals with simple close-packed crystal structures, hence these materials are ductile . Ceramics, for example, tend to have narrow dislocations, and are hard and brittle as a result. A mathematical treatment of this relationship can be seen here.

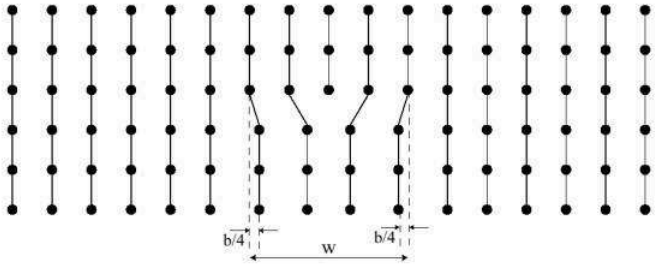


Diagram showing width of a narrow dislocation

This page titled 22.6: Dislocation in 3D is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.7: Observing Dislocations

A number of ways of 'seeing' dislocations in real materials have developed since the 1950s. Only in the last few years have electron microscopy techniques advanced sufficiently to allow the atomic structure around a dislocation to be resolved.

Optical microscopy - Etch pits in Sodium Chloride

Sodium chloride can be chemically etched to reveal some crystallographic features. Where a dislocation line intersects with a crystal surface, the core of the dislocation etches more rapidly than the surrounding dislocation-free crystal. This results in a small *etch pit*, large enough to be visible under low magnification in the optical microscope. The dislocations themselves are on the atomic scale - orders of magnitude too small to be visible with optical microscopy.

Sodium chloride single crystals can be cleaved using a razor blade tapped with a hammer. This creates a 'fresh' surface, with no environmental damage - the surface of the crystal is attacked rapidly by moisture in the air. Sodium chloride cleaves along {100} planes [link to Miller Indices]. Under the optical microscope, *cleavage steps* can be seen.

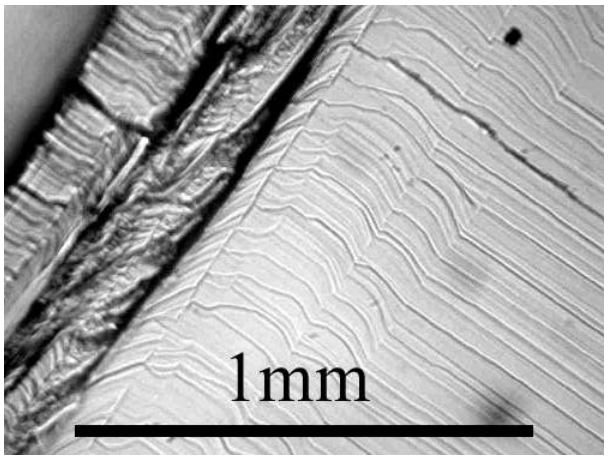
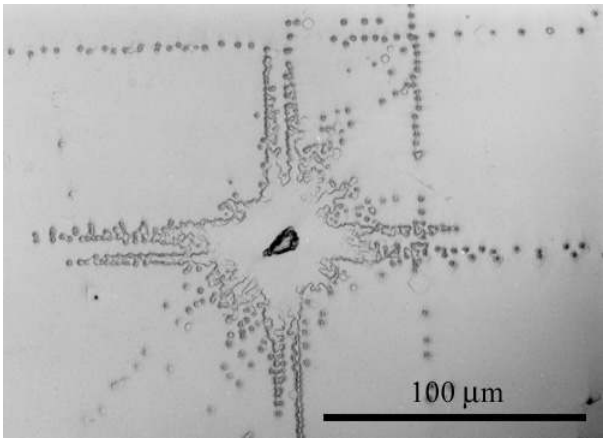


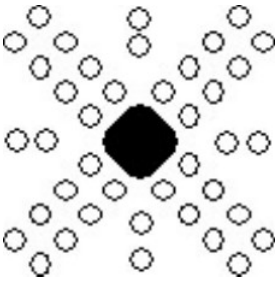
Figure 22.7.1: Micrograph of sodium chloride showing cleavage steps.

Dropping a few particles of silicon carbide grit onto the surface from a height of about 10 cm causes localised plastic deformation of the surface. This causes dislocations in the crystal to move. The surface can then be etched with iron (III) chloride in glacial acetic acid, and the etchant washed away with acetone. Under the optical microscope at a magnification of around 100x, 'rosettes' of etch pits may be observed, with their centres at the sites of impact of the silicon carbide. Elsewhere, more randomly distributed etch-pits may be present due to pre-existing dislocations and damage during cleaving.



Micrograph of sodium chloride showing a rosette.

The etch pits around the deformed region are aligned along particular directions, shown schematically below:

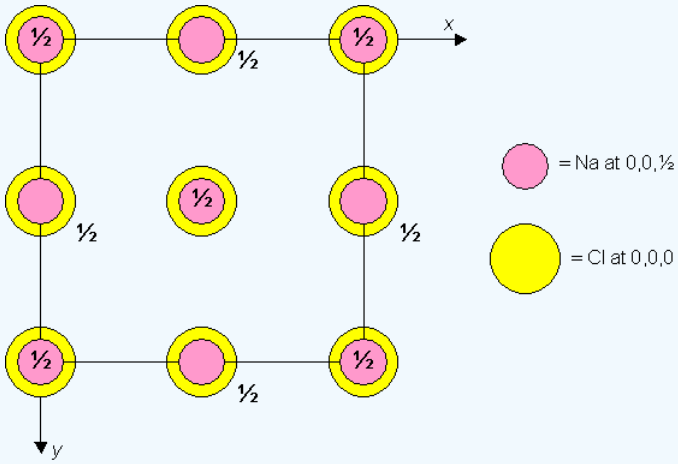


Schematic showing alignment of etch pits around deformed region.

Examining the orientation of the rosettes with respect to the surfaces of the crystal (which will be {100}) can be interpreted in terms of the *slip systems* in sodium chloride - that is, the *direction* of the Burgers vector and the *slip plane* on which the dislocation moves. The key point is that the *dislocations only move on specific crystallographic planes in specific crystallographic directions*.

Slip Systems in Sodium Chloride

Sodium chloride is cubic F with Cl atoms at 0,0,0 and Na atoms at 0,0,½. A unit cell of the material is shown below.



Unit cell plan for sodium chloride.

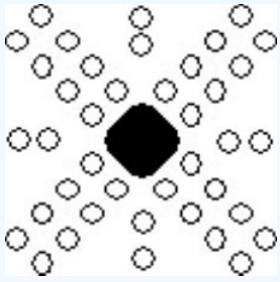
Dislocation glide (or slip) generally occurs in the direction of the shortest lattice vector. In the sodium chloride structure, therefore, the shortest lattice vectors connect lattice points such as 0,0,0 and 0,0,½, across half of the face diagonal. Try and identify these vectors on the 3D model of the unit cell.

https://www.doitpoms.ac.uk/tlplib/di...nacl-model.mp4

Virtual reality model of a sodium chloride unit cell

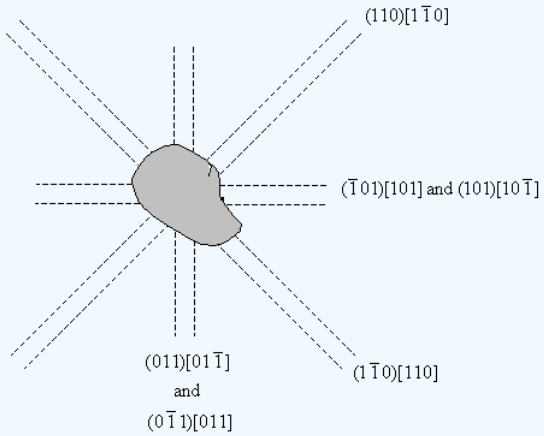
Using the Miller index notation, these vectors are in <1 bar1 0> directions. The planes on which slip occurs are slip planes - in sodium chloride, these are {110} planes. Therefore we say that sodium chloride slips on {110}<1 bar1 0> slip systems.

The etch pits in the deformed surface appear schematically as follows:



Schematic showing alignment of etch pits around deformed region.

The diagrams have been drawn with the sides of the crystal horizontal and vertical, and it is assumed that the surface is a (001) plane. Etch pits arise when a dislocation line intersects with the surface, and since the dislocation line must lie in the slip plane, the orientation of the etch pits can be matched to the slip planes and the slip directions, as shown in the diagram below. A row of etch pits in the rosette corresponds to the intersection of slip planes with a specific orientation intersecting with the surface.



Miller indices for etch pit orientations.

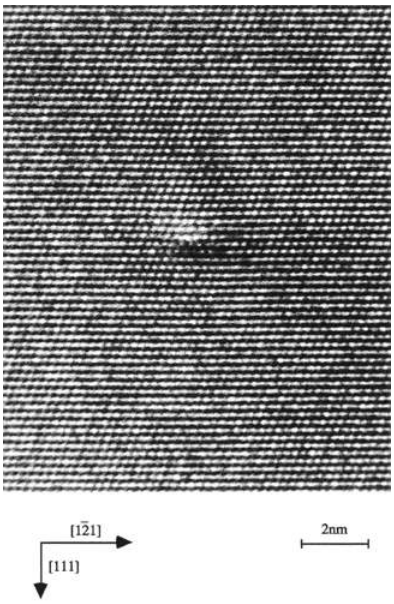
More on slip systems in sodium chloride ...

Transmission electron microscopy

Dislocations can be observed in the transmission electron microscope (TEM). Due to the lattice distortion of around the core of the dislocation, some Bragg diffraction of the electron beam occurs in a localised region around the core. Intensity is therefore directed away from the 'straight through' beam, so dislocations appear as dark lines in bright field TEM images. Crytstallographic information about the dislocation such as the direction of the Burgers vector can be determined from these TEM images. For some examples, see the video clips on Professor Harry Bhadeshia's website.

Being able to see dislocations as they move through a structure gives materials scientists a fascinating insight into the mechanisms of plastic deformation.

Recently, high resolution TEM has allowed microscopists to actually image the crystal planes and atomic positions within materials. The method can be exceedingly complex, making use of the phase difference between several diffracted beams caused by the atomic structure. In the example micrograph, the atomic positions in an edge dislocation in TiAl can be seen.



HRTEM image of a dislocation in TiAl. It is a b = ½ [110] dislocation in taken with the beam down the [101] direction. (Source: Beverly Inkson, PhD Thesis, University of Cambridge, 1994) (Click on image to view larger version.)

Scanning tunnelling microscopy (STM) is a high-resolution surface imaging technique. It allows the atomic surface structure to be deduced, revealing the disruption of the lattice at the surface where dislocation lines intersect with it.

This page titled 22.7: Observing Dislocations is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.8: Summary

In this package we have seen that a dislocation is a defect found in crystals. Dislocations are *line defects*, extending through a crystal for some distance along a *dislocation line*.

The *Burgers vector* specifies the magnitude and direction of the atomic movements that occur as the dislocation moves through the lattice. The angle between the line vector and Burgers vector characterises the nature of a dislocation - when the dislocation line and Burgers vector are perpendicular, the dislocation is known as an *edge* dislocation. When they are parallel, the dislocation is a screw dislocation. Between these two ideal angles, the dislocation is *mixed*.

A slice through an edge dislocation perpendicular to the dislocation line reveals that the dislocation is like an extra half-plane of atoms inserted between full planes, which are distorted to accommodate the dislocation.

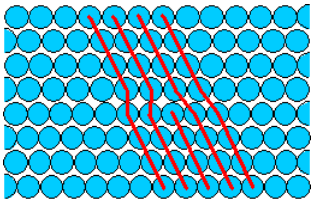


Figure 22.8.1: A 2D schematic representation of an edge dislocation in a close-packed plane.

The bubble raft experiment shows how dislocations and other defects occur in a close-packed plane. Application of stress to the raft shows how dislocations move under an applied stress, and it can be shown that the stress required to move a dislocation is less than that required to create a similar motion *via* movement of whole planes of atoms simultaneously. Dislocations explain the observation of plastic deformation at lower stress than would be required in a perfect lattice, and the phenomenon of work hardening.

Dislocations can be observed by a number of methods. Only high-resolution TEM lattice images or STM surface images can show dislocations directly, but etching methods and optical microscopy can be used to elucidate the presence of dislocations, for example on sodium chloride crystal surfaces.

Going further

Books

Many general materials science texts cover introductory material on dislocations. It is worth studying a few books because different authors present the material in different ways, and you may be able to visualise the concepts more easily when presented in a particular way.

Also consider

- Hull and Bacon, Introduction to Dislocations (4th edition), Pergamon, 2001.
  A comprehensive text covering material at introductory through to advanced level.
- Cahn, The Coming of Materials Science, Pergamon, 2001
  A 'history and portrait' of the subject, including the story of the emergence of the dislocation concept.

For the interested student, a library search for books by Cottrell, Read, Nabarro or Friedel will produce some classic texts on dislocations.

This page titled 22.8: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

22.9: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which of the following statements is *false*?

a	A bubble raft can demonstrate vacancies and solute atom defects.
b	Plastic deformation in a bubble raft occurs by dislocation motion.
c	Screw dislocations can be seen in the bubble raft.
d	Bubbles in a well-formed bubble raft are close-packed.

Answer

C

Which of the following statements best describes the nature of dislocations in an *amorphous material*?

a	Amorphous materials cannot contain dislocations.
b	The dislocation density in an amorphous material is normally less than the dislocation density in a crystalline material with the same composition.
c	The dislocation density in an amorphous material is normally greater than the dislocation density in a crystalline material with the same composition.
d	A dislocation in an amorphous material must be of the *edge* type.

Answer

A

What are the conventional units of dislocation density?

a	m^{-2}
b	m^{-3}
c	kg m^{-3}
d	kg m^{-2}

Hint: Dislocation density is defined as the length of dislocation line per unit volume of material.

Answer

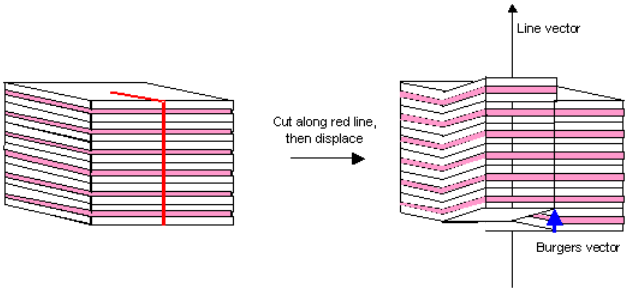
A

How would you make a stack of ham sandwiches look like a screw dislocation?

Hint: You need to make one cut in the stack, then displace the sandwiches on one side of the cut by a Burgers vector. What must the Burgers vector be?

Answer

The Burgers vector is one sandwich thickness. The cut is made halfway through the stack, as shown below. After displacement, one can follow the ham continuously along the 'screw'.



Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

Do the explanations of these experimental observations involve the concept of dislocations? *(Answer yes or no for each)*

Yes	No	a	The yield stress of a material is lower than that calculated from assuming a perfect lattice
Yes	No	b	In some cases, the stress required to continue plastic deformation increases as deformation proceeds
Yes	No	c	The yield stress of most metals decreases as the temperature increases
Yes	No	d	Ceramics tend to be brittle, whereas metals tend to be ductile

Answer

A. Yes. Dislocations can move through a lattice under a significantly lower applied stress than would be required to cause slip in a perfect lattice. Plastic deformation therefore occurs by dislocation motion in the majority of circumstances. (Note that other mechanisms of plastic deformation may occur, such as *diffusion creep* and *twinning*).

B. Yes. This is the phenomenon of *work hardening*. The mechanisms of work hardening involve mobile dislocations interacting to form *sessile* dislocations that cannot move under an applied stress, and the creation of *forest* dislocations - these intersect other slip planes and act as obstacles to dislocation motion on those planes, raising the stress required for dislocation glide.

C. Yes. The yield stress is the stress at which the metal begins to plastically deform, and since plastic deformation occurs (in most cases) by dislocation motion, it can be deduced that the stress required to drive dislocations through the metal is temperature dependent. As the temperature is increased, dislocation *climb* becomes easier. Climb allows dislocations to move onto parallel slip planes in order to overcome obstacles, and hence the yield stress is temperature dependent. (The ability of dislocations to climb over obstacles at high temperatures determines a material's susceptibility to *creep*, *i.e.* plastic deformation at stresses below the yield stress.)

D. Yes. The *width* of a dislocation determines the stress required for glide. Narrow dislocations, commonly found in ceramics, require greater applied stress to glide than wide dislocations, which are often found in metals. The stress required for plastic flow in ceramics normally exceeds the stress needed to drive a crack through the material hence brittle fracture occurs. In ductile metals, plastic flow occurs before brittle fracture.

The energy per unit length, U, associated with an edge dislocation is given by

$U \sim 0.5(Gb^2)$

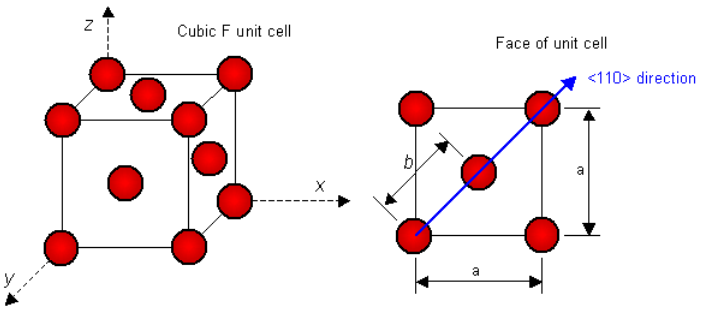
where b is the magnitude of the Burgers vector **b** and G is the shear modulus. Estimate the energy per unit length of a dislocation in silver.

Data for silver: Crystal system is cubic F, a = 0.409 nm. **b** lies parallel to <110> directions. Shear modulus G = 28.8 GPa.

Answer

This question requires the magnitude of the Burgers vector to be calculated. The magnitude is the distance between adjacent lattice points in the direction of **b**. In the cubic crystal systems, the <110> directions lie along the face diagonals. In cubic F, there are lattice points at 0,0,0 and ½,½,0 (*i.e.* at the corners and face centres of the unit cell). The possible Burgers vectors can

therefore be written as **b** = ½<110>. The magnitude of the Burgers vector can be deduced by considering the atoms on one face of the unit cell:



By geometry, the magnitude of the Burgers vector is $b = a/2^{1/2}$ = 0.289 nm. The energy of the dislocation per unit length of dislocation line is then

$U = ½Gb^2 = ½ \times 28.8 \times 10^9 \times (0.289 \times 10^{-9}) = 1.2 \times 10^{-9}$ J m^{-1}

Determine whether the following dislocations in sodium chloride are *edge*, *screw* or *mixed*. Identify the slip plane in which the dislocation lies. (Sodium chloride is cubic F and slips on {110} slip systems.)

	Burgers vector ***b*** parallel to:	Line vector ***l*** parallel to:
a	[110]	[110]
b	[001]	[110]
c	[100]	[111]

Hint: Edge dislocations have ***b*** perpendicular to ***l***, screw dislocations have ***b*** parallel to ***l***. Mixed dislocations have ***b*** at some acute angle to ***l***. Determine the angle ***b*^*l*** using the scalar (dot) product or otherwise. Both ***b*** and ***l*** must lie in the slip plane of the dislocation. The is one method of determining whether a vector lies in a plane when both vector and plane are specified by Miller indices and the crystal system is cubic.

Answer

a) Screw (since ***b*** is parallel to ***l***). Slip plane is (bar1 1 0), since this is the only allowed slip plane in which both ***b*** and ***l*** lie.

b) Edge (since ***b*** is perpendicular to ***l***). Slip plane is (110), since this is the plane specified by the two vectors.

c) Mixed - the angle between ***b*** and ***l*** is $\cos^{-1}(1/3^{½})$ = 54.7°. Slip plane is (0 bar1 1), since this is the plane specified by the two vectors.

Which of the following statements is true? *(answer yes or no for each)*

Yes	No	a	Using high-resolution microscopy techniques, it is possible to probe the structure of a dislocation.
Yes	No	b	The number of dislocations in a sample of material affects the ease of diffusion of impurities through the sample.
Yes	No	c	The diffusion of impurity atoms in a sample of a material affects the ease of motion of dislocations under an applied stress.

Answer

A. Yes. Modern transmission electron microscopy techniques have made it possible to create lattice images showing atoms in dislocations.

B. Yes. The atomic lattice is disrupted in the region of a dislocation - strain fields are present that make the lattice more 'open' or more 'closed' than the perfect structure. This can affect the diffusivity of species in the region of the dislocation. A sample of a

material with a high dislocation density may exhibit higher diffusion coefficients for solutes than a sample of the same material with a lower dislocation density.

C. Yes.

Several effects are important:

- Dislocation motion is hindered by the presence of obstacles, such as precipitates of a second phase. The diffusivity of atoms making up this second phase affects the distribution and stability of precipitates, and hence affects the extent of *precipitate strengthening*.
- Dislocations can overcome obstacles like precipitates by *climb*, in which the dislocation moves through the structure onto other parallel slip planes until it can glide past the obstacle. Climb involves the diffusion of vacancies.
- *Solution strengthened* materials rely on the size difference between the host lattice atoms and substitutional solutes. The strain fields in the lattice around the solutes interact with the strain fields around dislocations, hindering dislocation motion. Diffusion is important in determining the distribution of these substitutional solutes and the ability of the dislocations to climb past them.
- Impurities can diffuse to the core of a dislocation, where the lattice is under dilatational strain (in the case of an edge dislocation). The dislocation is a sink for impurities. This build-up of solutes is called a *Cottrell atmosphere*. The presence of these impurities at the core makes it more difficult for the dislocation to move. This phenomenon is the origin of *dynamic strain ageing* (or the *Portevin-Le Chatelier effect*). Here, carbon diffusion to dislocation cores in mild steel occurs rapidly enough at elevated temperatures to cause 'jerky' plastic flow. Dislocations break free from their Cottrell atmospheres at some applied local stress, but can then glide at a lower applied local stress. If the dislocation motion slows or pauses, the Cottrell atmosphere around the dislocation core builds up again as carbon diffuses to the sink, and a higher stress must again be applied before glide can proceed.

Open-ended questions

The following questions are not provided with answers, but intended to provide food for thought and points for further discussion with other students and teachers.

9. Is dislocation glide always the mechanism of plastic flow? How might other mechanisms operate?
10. Why might the presence of dislocations in materials used for electronic components (such as integrated circuits) be a problem? How might the problem be reduced or solved?
11. Plastically deforming a material requires energy input. In what ways is this energy dissipated? Is any of it stored in the material?

This page titled 22.9: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

23: Elasticity in Biological Materials

https://www.doitpoms.ac.uk/tlplib/

Learning Objectives

On completion of this TLP you should:

- Appreciate that many materials in living systems do not show Hookean elasticity.
- Understand the concept of J-curves and be familiar with examples of biomaterials that show this behaviour, in particular arterial wall.
- Understand how aneurysms arise in hardened arteries.
- Appreciate that some biomaterials, for example hair and spiders' silk, exhibit hysteresis, and understand how this can be essential to their performance and applications.

Before you start

Before beginning this TLP, you should understand the elasticity of rubber, including the concept of an entropy spring. Take a look at The Stiffness of Rubber TLP.

You should also be aware of how a tensile test is carried out. Take a look at the Introduction to Mechanical Testing TLP.

23.1: Introduction
23.2: Hookean Elasticity
23.3: S-shaped Curves
23.4: Aneurysms
23.5: Demonstration of Aneurysms
23.6: J-Shaped Curves
23.7: Viscoelasticity and Hysteresis
23.8: Summary
23.9: Questions

This page titled 23: Elasticity in Biological Materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.1: Introduction

As you will already know, many engineering materials, such as metals, show Hookean elasticity in which the tensile stress applied to a sample is directly proportional to the resultant strain. Within the range of Hookean elasticity, the stress-strain curves on loading and unloading are identical. Such linear elasticity is the usual assumption in engineering design. However, the elasticity of most materials in living systems is much more complicated.

In this TLP, you will learn that some biomaterials exhibit non-linear stress-strain curves. Mammalian skin, for instance, exhibits a *J-shaped stress-strain curve*, as do healthy arterial walls. Materials with J-shaped curves are usually tough, and have other advantages. For example, materials with different elastic behaviour such as S-shaped stress-strain curves are prone to elastic instabilities such as aneurysms when used for tubes under pressure.

This TLP also discusses viscoelasticity. Many biomaterials show time-dependent stress-strain curves. Associated with this, the loading and unloading curves do not superimpose on each other. Although deformation is elastic (i.e. recoverable), energy is absorbed during the deformation. This is particularly important in spider silk: when a fly hits the web its energy should be absorbed by the deformation of the web. If spider silk showed elasticity without energy absorption, then the web instead would act as a trampoline!

This page titled 23.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.2: Hookean Elasticity

For many materials loaded in uniaxial tension, the tensile stress on the material, σ, is directly proportional to the tensile strain , ε.



A sample loaded in uniaxial tension

The linear relationship between stress and strain is known as Hooke's Law,

$\sigma \propto \varepsilon$

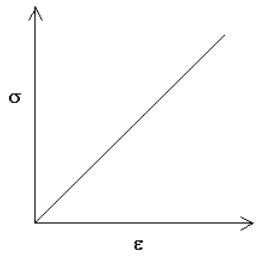
The constant of proportionality in this equation for simple tension is the Young Modulus of the material, E:

$$E = \frac{\sigma}{\varepsilon} \tag{23.2.1}$$

The Young Modulus of a material has values ranging from approx. 0.01 GPa for rubbers to approx. 1000 GPa for diamond.

Hooke's Law further states that the stress response of a material is independent of time and that the strain of a material disappears completely on removal of the applied stress (i.e. a Hookean material shows elastic deformation).

This leads to a linear stress-strain curve with a gradient of E. Loading and unloading occur along the same curve.



A stress-strain curve for a Hookean material

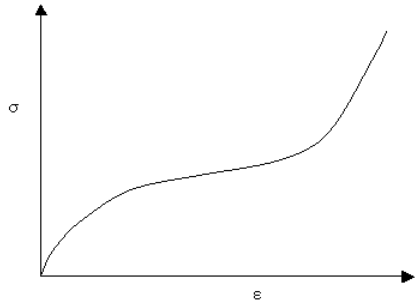
Most materials are Hookean only at small strains (typically less than 1%). Metals, for which fully elastic behaviour is only for very small strains (typically <0.2%), show Hookean behaviour. In this region, the extension is usually both linear and recoverable. At larger strains, extension is non-Hookean (i.e. either non-recoverable, or non-linear, or both).

Although many materials used in engineering applications show Hookean behaviour, only a few biomaterials approximate to it (wood and bone being the two most common). Many biomaterials exhibit a J-shaped stress-strain curve, but firstly, we shall consider the S-shaped stress-strain curve seen in rubbery materials.

This page titled 23.2: Hookean Elasticity is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.3: S-shaped Curves

S-shaped stress-strain curves occur in rubbery materials (lightly cross-linked polymers). Materials with S-shaped stress-strain curves are particularly susceptible to elastic instabilities, which are of interest in analyzing phenomena such as aneurysms in arteries. The curves have the form:



An S-shaped stress-strain curve

The initial part of this curve, where the stiffness decreases with increasing load can be predicted theoretically, by considering the rubber as an entropy spring, (shown in detail in The Stiffness of Rubber TLP). This treatment assumes that all extension occurs via conformational changes (i.e. that there is no bond stretching) and also assumes that the chain is composed of a series of joined links which are equally likely to lie in any direction (the random walk assumption).

Using the random walk assumption, the probability distribution for the end-to-end length of a polymer chain with a certain number of links can be obtained, and the Boltzmann expression ($S = k \ln W$) can be used to determine the change in entropy on extending the chain and moving it to a less probable conformation. The total entropy change can be found by multiplying the result for one chain by the total number of chain segments N_t.

If we assume that the extension occurs without significant change in the enthalpy, pressure, volume or temperature, then the change in Gibbs free energy (ΔG) can be related to the entropy change (ΔS) via the equation $\Delta G \approx -T\Delta S$. If deformation occurs only in the x direction then the force required for deformation is

$$F = \frac{\partial \Delta G}{\partial x} \tag{23.3.1}$$

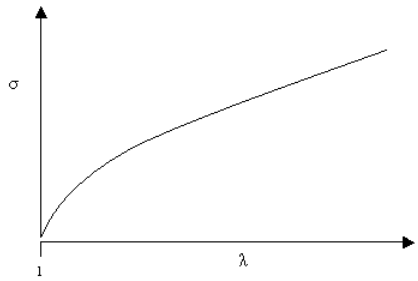
and an expression for the stress required for a given extension (with the rubber in uniaxial tension) can be obtained.

If N is the number of chain segments per unit volume and λ is the extension ratio of the rubber, then the nominal stress for a rubber loaded in uniaxial tension is:

$$\sigma = kTN\left[\lambda - \frac{1}{\lambda^2}\right] \tag{23.3.2}$$

where k is the Boltzmann constant and T is the temperature.

This gives a stress strain curve of the form:



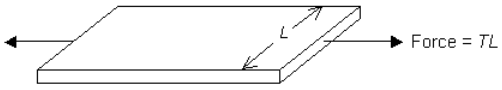
The theoretically predicted stress-extension curve for rubbery materials

Thus the theory predicts that the stiffness of rubber varies a little (particularly at low extensions), but that it tends to a limiting value at higher extensions. At lower extensions, the theoretical stress-extension curve is fairly similar to the S-shaped curves obtained experimentally. At extensions of $\lambda = 4$ the experimental and theoretical curves diverge, with much larger stiffness seen experimentally than predicted theoretically. This occurs because at larger extensions the assumptions of the model are no longer valid: the polymer chains are mostly aligned with the applied stress and so applying higher stress stretches strong intra-molecular bonds.

This page titled 23.3: S-shaped Curves is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.4: Aneurysms

Hardened and weakened arteries can show elastic instabilities such as aneurysms. A hardened artery can be modelled as a long cylindrical balloon, with radius r and internal pressure P. Consider the tension, T, in the polymer sheet as a function of the extension ratio, λ. The advantage of using T is that it allows us to ignore changes in the thickness of the polymer sheet, which may be considerable at large λ. The units of T are force/length.

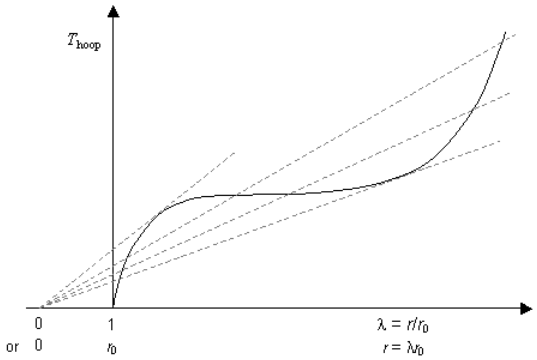


For a cylinder of radius r, the pressure is related to the tension in the hoop direction, Thoop, by:

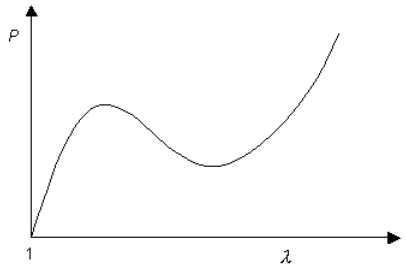
$$P = \frac{T_{\text{hoop}}}{r} \tag{23.4.1}$$

An increased hoop tension for a given radius, or a decreased radius for a given hoop tension gives an increased pressure.

A plot of Thoop against λ for the balloon rubber would show a characteristic S-shaped curve, similar (but not identical) in shape to the stress - strain (σ — ε) curve:



The gradients of the dashed lines are proportional to the pressure inside the balloon at the values of extension ratio (or radius) at which they cross the curve. This allows us to obtain the curve showing the variation of pressure with extension ratio:



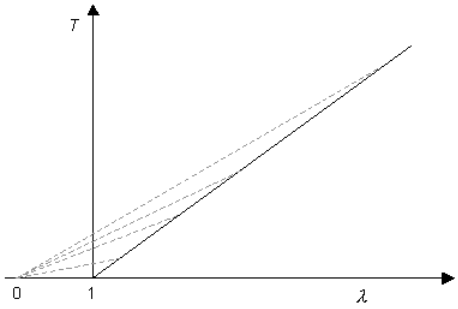
Although the tension in the balloon only ever rises with λ, it can be seen that this is not the case with pressure. The fact that there are regions of the graph for which

$$\frac{\mathrm{d}P}{\mathrm{d}\lambda} < 0 \tag{23.4.2}$$

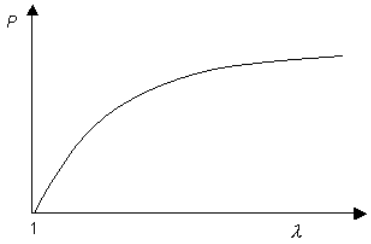
allows the occurrence of different radii at equal pressures, which leads to the formation of aneurysms.

This behaviour can be observed in modelling balloons, as shown in the next section.

Now consider Hookean behaviour:



This then gives:



In this case, there is no region of

$$\frac{\mathrm{d}P}{\mathrm{d}\lambda} < 0 \tag{23.4.3}$$

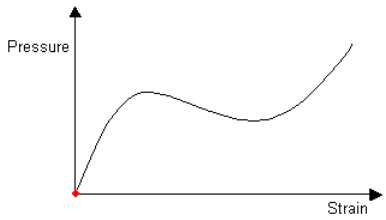
(although the gradient does tend to zero at large extension ratios). Thus if an artery exhibited Hookean elasticity, it should theoretically be stable against aneurysms, but only just, as the pressure inside the tubes is almost independent of its radius. However, since a normal artery exhibits natural variations along its length, Hookean behaviour in the arterial wall would not provide a strong safeguard against aneurysms. In fact, arteries exhibit J-shaped stress-strain curves as described later. For J-shaped curves dP/d λ is again always greater than zero but now does not tend to zero at large extension ratios. This provides greater stability against aneurysms.

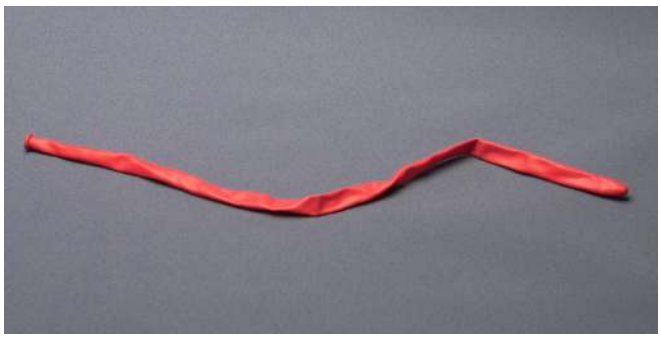
This page titled 23.4: Aneurysms is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.5: Demonstration of Aneurysms

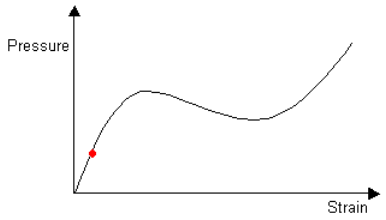
As mentioned on the previous page, the behaviour of a modelling balloon can be used to provide a simple demonstration of an aneurysm. This is shown in the series of photographs below.

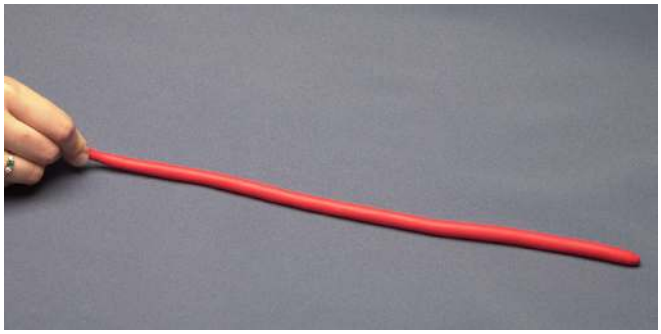
Initially, before any air is pumped into the balloon, it is completely deflated:





As the balloon is slowly inflated, it initially has one stable radius:

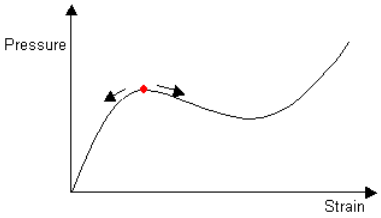


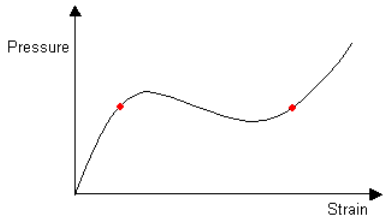


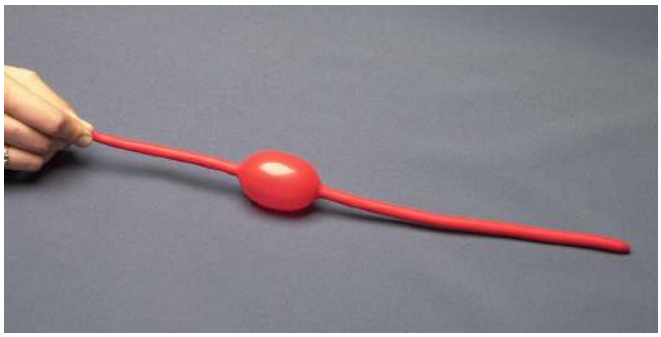
As air is blown into the balloon, the pressure continues to increase. When the balloon has an internal pressure equal to that at the maximum in the pressure-strain curve, an instability is introduced. The balloon develops an aneurysm, which greatly increases the internal volume and so the pressure goes down. The aneurysm grows to give a shape which is stable.



For the cylinder of small and large radius, and for the transition between the two, the local conditions of T and the curvature of the rubber are such as to correspond to a uniform internal pressure. Although the pressure is constant along the length of the balloon, two cylindrical radii are stable and so aneurysms are formed.

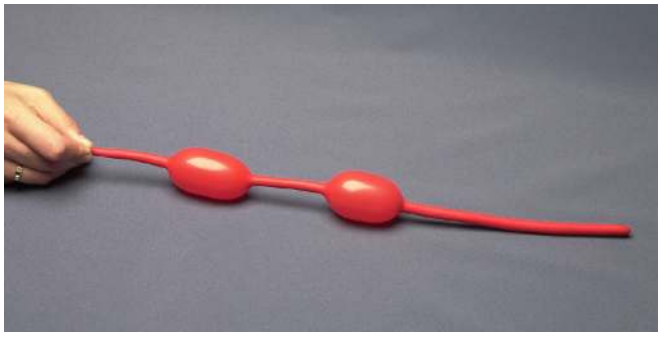






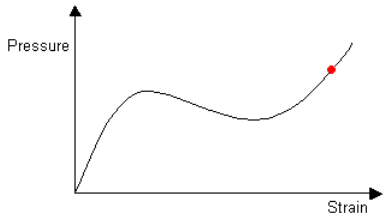
As more air is forced into the balloon the pressure within the balloon remains constant and the two radii stay the same size. The extra air is accommodated by the lengthening of the part of the balloon with greater radius. The lengthened aneurysm can also be split into two.





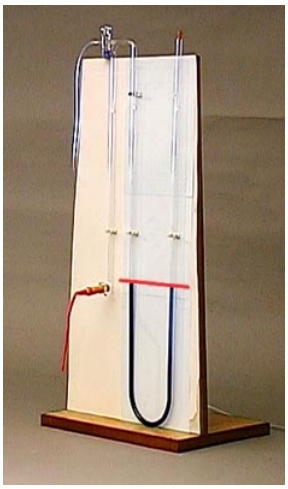
Aneurysms that form in arteries do not usually increase in length but instead increase spherically in radius until they burst. This is because the pressure-strain curve for hardened arteries does not have a local minimum after the maximum, but instead decreases continuously from the maximum.

Once the balloon is completely inflated, it again has one stable radius. Further increases in pressure lead to uniform expansion.





The pressure inside the balloon as it inflates can be measured using a manometer. The balloon was fixed to the apparatus shown, and a pump was used to inflate the balloon at a constant rate.



As shown in The Stiffness of Rubber TLP, the difference in height of the two menisci in the manometer is proportional to the pressure inside the balloon. The videos below give a qualitative idea of how the pressure varies as the balloon is blown up. The pressure drop as the aneurysm forms and the constant pressure as the aneurysm lengthens are particularly obvious. The red lines in the video and on the picture show the initial position of the menisci.

https://www.doitpoms.ac.uk/tlplib/bi...eos/vid001.mp4

Video showing the inflation of balloon on the manometer

https://www.doitpoms.ac.uk/tlplib/bi...eos/vid002.mp4

Video giving a close-up of the manometer as the balloon expands

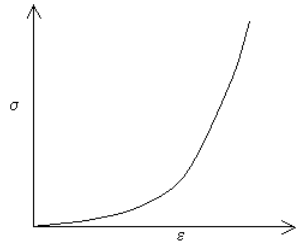
---

This page titled 23.5: Demonstration of Aneurysms is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.6: J-Shaped Curves

Descriptions of J-Shaped Curves

Many biomaterials exhibit the following type of stress-strain curve, which is known as a J-shaped curve:



A J-shaped stress-strain curve

The curve shows that initially, small increases in stress give large extensions, however, at larger extensions the material becomes stiffer, and more difficult to extend. Mammalian skin and flesh are two of the many biomaterials that exhibit a J-shaped curve. If you pinch your earlobe and try to pull it downwards, you will see that initially it is quite easy to stretch the earlobe, but that at larger extensions it becomes more difficult to stretch.

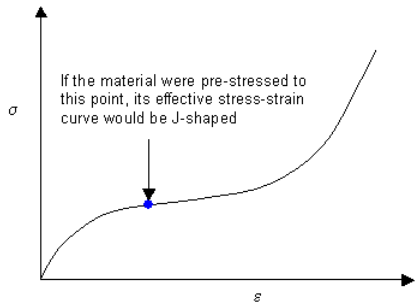
Loading and unloading occurs along the same curve, i.e. the loading is completely reversible and elastic. This ensures that all the energy used in extending the system is returned once the load is removed. It is clearly important that there should not be too much energy absorption in arterial walls. The elastic properties of arterial wall are important not only to protect against aneurysms, but to smooth out variations in blood pressure and blood flow rate.

Advantages of J-shaped curves for biomaterials

J-shaped stress-strain curves cause biological membranes to be extremely tough, even though the fracture energy for many materials is not particularly high (around 10 kJ m^{-2}). This toughness arises for the following reasons:

- The lower part of the J-shaped curve gives very large extension for low applied stress, so the shear modulus in this region is very low and so there is no mechanism whereby the released strain energy on fracture can be transmitted to the fracture zone.
- The material gets stiffer as the failure point approaches ensuring that very large extensions require large stresses, so that extensions which are likely to cause harm occur infrequently.
- Since the J-shaped curve is concave, the area under the curve up to a given extension is far lower than that for the equivalent Hookean curve meaning that the energy released in the fracture of a material with a J-shaped stress-strain curve is far lower than the energy released when an equivalent Hookean material fails. Since the release of energy drives crack propagation, a material that releases less energy on fracture is tougher.
- The J-shaped stress-strain curve does not lead to the elastic instabilities such as aneurysms which arise with S-shaped curves. Hence J-shaped curves are favoured for arteries.

A J-shaped curve can be compared to an S-shaped curve in which the material has been pre-stressed, causing the effective origin of the graph to be further along the curve. This is shown below. Arteries are naturally pre-stressed as even the minimum blood pressure must be sufficiently positive to raise blood to the highest point of the body.



How Does the J-Shaped Curve Arise?

The J-shaped curve can arise at a number of different structural levels. For a J-shaped curve to arise, there must simply be the progressive recruitment of strain-resistant components. This can occur by the progressive alignment of polymer chains with the stress, and indeed, the stress-strain curve of a rubber is (at higher extensions) J-shaped. Many biomaterials are pre-stressed rubbery materials (i.e. in their neutral position they are still under tension). These rubbery materials, by virtue of being pre-stressed, show only J-shaped behaviour. A similar effect occurs in materials that contain fibres in a soft matrix. Initially the stress acts only against the soft matrix, but with time the fibres align in the direction of the stress and so further pulling works against the stiffer fibres. This effect is used in arterial walls, where the collagen fibres act as the stiffer fibres. (More information on the structure of arterial walls.)

Examples of Materials With J-Shaped Stress-Strain Curves

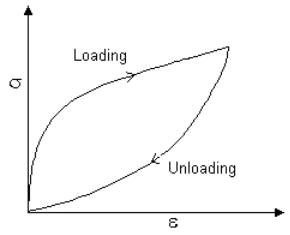
Textile materials provide some non-biological examples of materials with J-shaped stress-strain curves. Knitted materials and woven fabrics pulled at 45° to the warp and weft have J-shaped curves and are thus quite tough. Knitted fabrics usually fail by unravelling not by tearing, and when woven fabrics tear it is usually along the warp and weft directions, even though these directions are those in which the force is aligned with the strong threads.

Many biological soft tissues, show J-shaped stress-strain curves. Skin and arterial walls have already been cited as examples. Viscid spider silk (the capture threads) is a further example. At low strains, both collagen and tendon show J-shaped curves, in fact, the toughness of raw meat is due to the presence of collagen fibres within the meat. On cooking, the collagen disintegrates, and the meat becomes tender.

This page titled 23.6: J-Shaped Curves is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.7: Viscoelasticity and Hysteresis

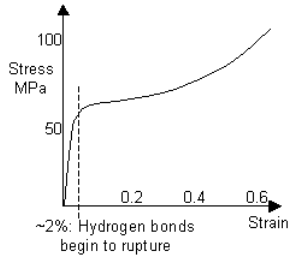
Some biomaterials show a time-dependent elastic behaviour. Although, in the elastic regime, the strain is recoverable, the stress-strain curve is not the same for loading and unloading. Such materials instead exhibit *viscoelasticity*, involving both elastic and viscous components, which at normal loading and unloading rates leads to *hysteresis*. A typical hysteresis curve is shown below, and the energy absorbed during one loading-unloading cycle is given by the area within the loop. The shape of the loop depends on the rates of loading and unloading (unlike normal time-independent elasticity).



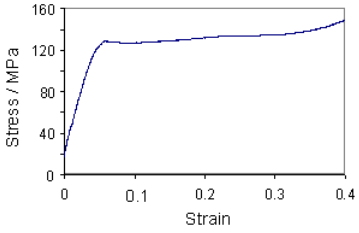
Hair

The effect of hysteresis can clearly be seen in hair. Hair consists of keratin, which is a type of protein. There are two main forms of keratin: α-helices, which are found in hair, and β-sheets. As the names suggest, α-helical keratin (also known as α-keratin) contains keratin molecules arranged in helices and held in place by hydrogen bonds, whereas β-sheet keratin (also known as τ-keratin) contains keratin molecules arranged in flat sheets, in which adjacent molecules are antiparallel due to more favourable interactions between side-groups.

As hair is stretched, the hydrogen bonds between the α-helices rupture, causing the helices to unravel to form β-sheets. The curve for human hair, and a curve obtained for a horse hair from a violin bow are shown below.



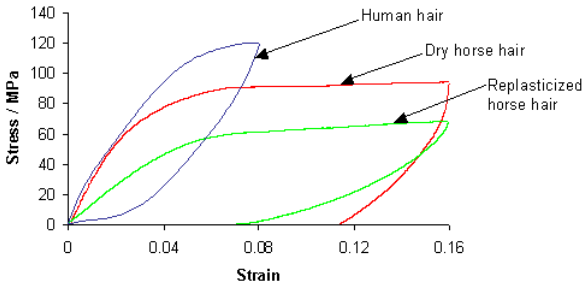
Stress-strain curve of a human hair (*Structural biomaterials*, Vincent)



Measured stress-strain curve of a horse hair from a violin bow

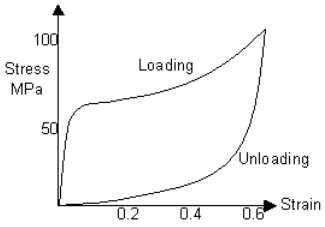
When the stress on the hair is removed, the helices re-form over time, and since the unravelling of α-helices to form β-sheets is a high-energy-absorbing process, a large area is contained within the hysteresis loop. As a result, this energy is unavailable for fracture, giving a high toughness. This is particularly important for hooves and horns, which are also made up of keratin.

The following graph shows experimentally measured stress-strain curves for human hair, dry horse hair (as used in violin bows) and replasticized horse hair (the same type of horse hair but soaked in water).



As you can see, the replasticized horse hair requires less force to stretch than the dry horse hair (for the same extension). This is because the presence of the water allows the unravelling and reravelling of the helices to occur more easily.

If the human hair had been stretched to a higher strain, its hysteresis curve would look more like this:



Spiders' silk

There are two main types of silk: spiders' silk and the silk produced by silk worms. Both types have very good mechanical properties, are durable and readily available, and both form fine threads (typically about 1 μm in diameter) that are biodegradable and biocompatible, making them ideal for use as suture materials.

The two silks are proteins and are very similar in composition: they are made up of similar proportions of the same amino acids. However, the sequence of these amino acids in spiders' silk is much less regular than that from silk worms, impeding the formation of crystallites (*micelles*) and making it much more extensible.

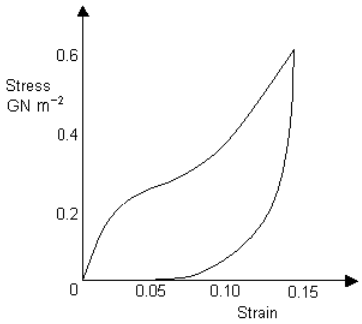
Spiders produce silk in several different glands, each of which produces silk for a different purpose (wrapping prey, producing drag lines, forming frame threads and capture threads etc). Slightly above the concentration of protein found in the gland, the silk proteins form a liquid crystalline phase. This liquid crystalline phase is formed as the silk passes through the duct leading to the spinnaret, and the silk crystallizes as it passes through the spinnaret itself to form an insoluble β-sheet. The higher the draw ratio through the spinnaret, the higher the orientation of the fibres.

Frame threads contain well-aligned molecules, and are highly crystalline, dry and relatively thick. In contrast, *capture threads* contain less-aligned molecules, and are less crystalline, plasticised, relatively thin, and exist as coiled fibres within an 'aqueous glue' layer.

The *coefficient of restitution*, or *resilience*, describes the fraction of energy returned elastically and can vary by a very large amount, as seen in the table below.

Resilin	Found in the wing hinges of insects	97%
Collagen	Found in tendons, ligaments, skin etc.	93%
Elastin	The main elastic protein of vertebrates	76%
Viscid silk	Capture threads	35%

Capture threads are made from a material known as *viscid silk* and are used to capture prey. Viscid silk has a very low coefficient of restitution, making it ideal for absorbing the energy of an insect flying into the web (instead of catapulting it away!). Its high strength also helps to ensure that the web is not destroyed upon impact. The hysteresis curve for viscid silk is shown below.



Silk from silkworms



silkworm starting spinning a cocoon © Michaal Cook www.wormspit.com

Silk is used in nature by silk worms to form their cocoons, and so must be strong and not easily breakable, as this would kill the silkworm, preventing it maturing into a moth and reproducing. Silkworms originate from China, India and Japan, and have been used by humans to make silk since at least 3,000 BC. Although silkworms only live for two months, they manage in this time to eat roughly 30,000 times their initial weight. It is estimated that 2,500 to 3,000 cocoons are needed to make just one yard of silk fabric, so despite silk being an excellent material for making fibres it is also expensive to produce.

This page titled 23.7: Viscoelasticity and Hysteresis is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.8: Summary

In this TLP the elastic properties of biological materials were studied. You will have learnt that few biological materials exhibit Hookean behaviour either because they do not have linear stress-strain curves, or because they do not have reversible stress-strain curves.

Many biomaterials show J-shaped curves and the reasons why these arise were discussed, in terms of material structure. The S-shaped curve has been covered, and concepts such as the entropy spring introduced in The Stiffness of Rubber TLP have been revised. You will have seen that S-shaped stress-strain curves lead to elastic instabilities such as aneurysms, and that this can be demonstrated experimentally, using cylindrical balloons.

Horse hair, human hair and spider silk are examples of materials which show viscoelasticity and tensile tests were carried out on human hair and horse hair. You should now be aware of the structure of hair and spider silk and the reasons for their low coefficient of restitution.

Going further

Books

- S Vogel, *Comparative Biomechanics: Life's Physical World*, Princeton University Press, 2003.
- JE Gordon, *The Science of Structures and Materials*, Scientific American Books, 1988.
- JFV Vincent, *Structural Biomaterials*, Princeton University Press, 1990.

Websites

- Mechanical properties of keratin
  A site maintained by Richard Bonser about the mechanical properties of keratin, with particular reference to the mechanical properties of feathers.
- Silk biomechanics
  An article on the American Museum of Natural History website.

This page titled 23.8: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

23.9: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

What type of stress-strain curve does a normal arterial wall show?

a	Hookean
b	S-shaped
c	J-shaped

Answer

C

What type of stress-strain curve does natural rubber show?

a	Hookean
b	S-shaped
c	J-shaped

Answer

B

What type of stress-strain curve do metals at small extensions show?

a	Hookean
b	S-shaped
c	J-shaped

Answer

A

What type of stress-strain curve does bone show?

a	Hookean
b	S-shaped
c	J-shaped

Answer

A

What type of stress-strain curve does skin show?

a	Hookean
b	S-shaped
c	J-shaped

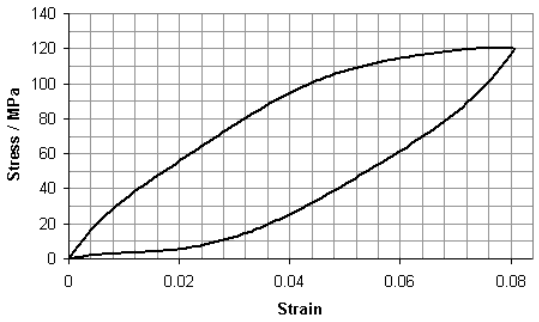
Answer

C

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

6. Estimate the coefficient of restitution (sometimes known as the resilience) of a human hair using the hysteresis curve below:



Answer

The area under the loading curve is the work done per unit volume of the stressed material, which is equal to about 166 squares.

The area under the unloading curve is the recovered work, which is equal to about 72 squares.

The coefficient of restitution is therefore found by dividing the area under the unloading curve by the area under the loading curve, giving an answer of about 43%.

Why are hairs used in violin bows?

Answer

Hairs are used in violin bows because this low resilience enables the damping of resonances. The desired resonances are in the violin strings!

A fly of mass 30 mg hits a spider's web at 0.8 m s^{-1} and is stopped by two capture threads each of diameter 0.7 micrometres and length 5 cm. Calculate the maximum tensile strain in the threads, assuming that the tensile loading curve of the capture threads can be approximated as a straight line with gradient 0.1 GPa.

Hint: Consider the kinetic energy of the fly and the work done in extending the capture threads.

Answer

If the tensile loading curve is linear, then the work done in extending a thread, which is equal to the area under the curve, can be calculated using 0.5 x load x extension, or 0.5 x stress x strain, or 0.5 x Young's modulus x $strain^2$

At 0.8 m s^{-1}, the kinetic energy of the fly is equal to $\frac{1}{2}mv^2$

= 9.6 x 10^{-6} J.

The total volume of the two threads

= 2 x 0.05 x π(0.35 x 10^{-6})^2
= 3.85 x 10^{-14} m^3

Work done per unit volume of thread

= 9.6 x 10^{-6}/3.85 x 10^{-14}
= 2.49 x 10^8 J m^{-3}
= ½$E\varepsilon^2$

Using the fact that E = 0.1 GPa, ε = 2.23

(.. continuation from previous question) Given that the coefficient of restitution of a capture thread is 35%, calculate the total energy dissipated in the two threads.

Answer

The energy dissipated is 65% of the kinetic energy of the fly = 6.24 x 10^{-6} J

(.. continuation from previous question) Since the threads are very thin, this energy would easily be dissipated as heat to the surrounding air. However, supposing that the process is adiabatic (no heat transfer occurs), estimate the temperature rise that would occur. The density of viscid silk is 1500 kg m^{-3} and the specific heat capacity is 1400 J K^{-1} kg^{-1}.

Answer

The total heat capacity of both threads

= total volume x density x heat capacity
= 3.85 x 10^{-14} x 1500 x 1400 = 8.09 x 10^{-8} J K^{-1}

The adiabatic temperature rise would be

= energy/heat capacity
= 6.24 x 10^{-6}/8.0 x 10^{-8}
= 65 K

(.. continuation from previous question) Comment on the effects of using cables with a large diameter made of viscid silk, for example to decelerate aircraft landing on carriers.

Answer

Thick cables made of viscid silk would be unable to dissipate the heat quickly to the surroundings, and very large temperature rises would be seen.

This page titled 23.9: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

24: Electromigration

https://www.doitpoms.ac.uk/tlplib/

Learning Objectives

On completion of this TLP you should:

- understand the process of electromigration
- be aware of how damage is caused by electromigration and the ways in which such damage is minimized
- know the reasons for changing from Al-based to Cu-based metallization.

Before you start

It would help to look through the Diffusion section within the Matter software, though the Diffusion TLP, but this TLP is mostly self-explanatory.

24.1: Introduction
24.2: The Theory of Electromigration
24.3: Electromigration Damage
24.4: Flux Divergence
24.5: Avoiding Electromigration Problems
24.6: Alternatives to Aluminium Metallization
24.7: Summary
24.8: Questions

This page titled 24: Electromigration is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

24.1: Introduction

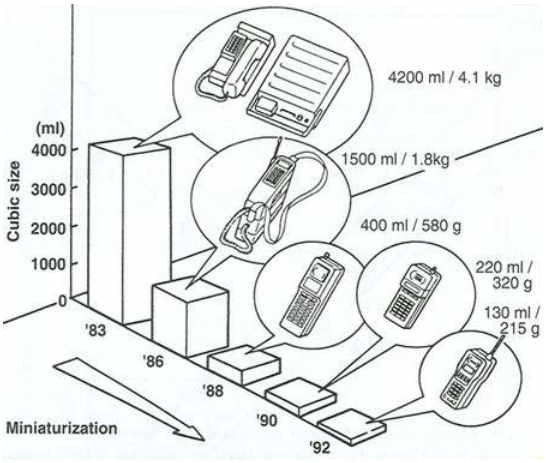
The use of microelectronic devices within our daily lives has increased vastly resulting in the transformation of our lives by recent scientific and technological advances. Electromigration is a principal wear-out mechanism of integrated circuits (IC), thus limiting their reliability. Reliability is important for the success in the microelectronics industry as a product is expected to work for an extended period of time without failure. This makes electromigration an area of intense research as more and more is demanded from microprocessors – to be faster, with smaller components and cheaper.

History

An IC contains various interconnected semiconductor components, such as transistors, resistors, capacitors and diodes. The first integrated chip contained tens of transistors within a chip area of 1 $cm^{2,}$ and can now be called "Small-Scale Integration" (SSI). To find out more about the different parts of an integrated circuit, click here.

With time, there has been a reduction of the dimensions of personal computing systems, from the size of a desktop, to a notebook, to a palmtop, to a credit card, to a watch, eventually to the size of a finger ring!

Concurrently, there has been a change in volume and complexity of wireless communication system.



Reprinted from 'Reliability and Failure of Electronic Materials and Devices' by Milton Ohring, Copyright 1998, with permission from Elsevier.

Currently, a chip area of 1 cm^2 or smaller contains hundreds of thousands of transistors to several million – "Very Large-Scale Integration" (VLSI) – making it possible to fabricate a Central Processing Unit (CPU) on a single integrated circuit. The industry now uses the terminology of "Ultra-Large-Scale Integration" (ULSI), for chips containing more than a million transistors and to emphasize chip complexity.

This continuous stream of achieving higher levels of integration brings about an exponential growth in the power of computing and communications technology to consumers and businesses worldwide. This is described by Moore's Law.

Moore's Law states that:

The number of transistors on an integrated circuit doubles about every two years.

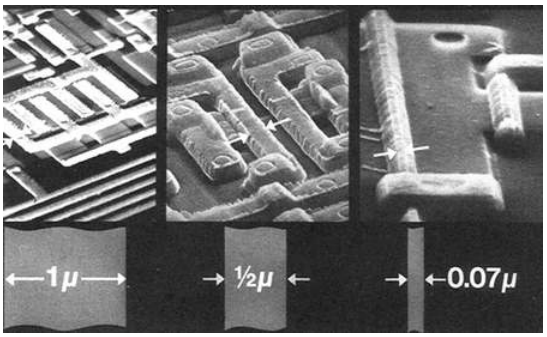
Today, adding transistors in pace with Moore's Law continues to drive increased functionality, performance and decreased cost of computing and communications technology. The price of an individual transistor has decreased drastically from $45 (in the 1950s) to less than a hundred-thousandth of a cent (currently - 2005)!

Click this link to view a graph depicting the rapid growth in the number of transistors integrated onto a single chip.

The dramatic miniaturization has involved not only the semiconductor components of an IC but also the conducting lines linking them (the *interconnects* – part of the *metallization* on a device). Changes relevant for electromigration-induced failure of interconnects include:

- line widths shrinking from 10 mm to 0.19 um, which will become even finer in the future.

- line width shrinking from 10 μm to 0.25 μm, which will become even finer in the future;
- total length of interconnect lines on a single IC increasing from several cm to several km (reaching lengths of about 5 km).
- current densities increasing from 108 A m^{-2} to 1010 A m^{-2}.
- number of metal levels increasing from 1 to 7 – this is because the number of components has increased so vastly that a single level on the top of the silicon substrate is insufficient to interconnect all the semiconductor components in the substrate, thus requiring multiple levels of metallization lines.



Line widths decreasing from 1 μm to 0.5 μm (for research purposes line widths as small as 0.07 μm have been used).
Reprinted from 'Reliability and Failure of Electronic Materials and Devices' by Milton Ohring, Copyright 1998, with permission from Elsevier.

These changes all increase the likelihood of electromigration-induced failure of the metallization and they drive further research to reduce the effect of electromigration. The latest development is the move away from aluminium-based to copper-based metallization lines.

This page titled 24.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

24.2: The Theory of Electromigration

Electromigration is the transport of material in a conductor under the influence of an applied electric field. All conductors are susceptible to electromigration, therefore it is important to consider the effects the electrical current resulting from the applied field may have on the conductor.

The net force exerted on a single metal ion in a conductor has two opposing contributions: a *direct force* and *wind force*.

Direct Force

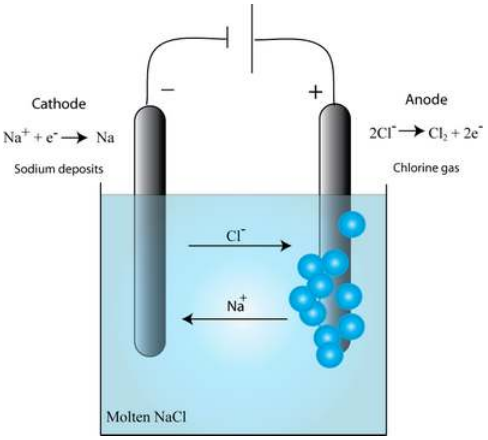
Application of an electric field results in an electrostatic pull being exerted on the metal ion core. The direction and magnitude of the electrostatic pull depend on the charge on the ion core, modified by screening effects. Positively charged ion cores (cations) are pulled towards the cathode, whilst negatively charged ion cores (anions) are pulled towards the anode. The direct force, F_d, is given by:

$$F_d = aZeE = aZejp \tag{24.2.1}$$

where a = factor accounting for screening (a<<1); Z = actual valence of the atom; e = electron charge (1.6 x 10^{-19} C); j = current density (A m^{-2}); ρ = resistivity (Ω m).

When electromigration in conductors was discovered, it was thought that the migration effects would be analogous to those in liquid electrolysis.

Consider for example, the electrolysis of the molten salt, NaCl. The Na^+ ions move towards the cathode, whilst the Cl^- ions move towards the anode (as shown in the diagram below), as the electric field is applied.



It was thought that the metal ions within a wire would always move towards the cathode as in electrolysis. While this is found in some cases, in the metallization of interest for ICs (which are based on good metal conductors) the migration was observed to be in the opposite direction – towards the anode. This startling effect results from the wind force.

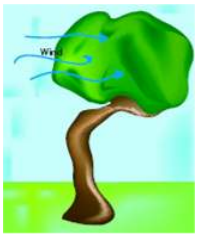
Wind Force

Electrons move along the metallization line, carrying the current. These electrons tend to scatter. Electron scattering takes place at imperfections within the lattice: vacancies, impurities, grain boundaries, dislocations and even phonon vibrations of the metal ions from their ideal positions! The scattering of electrons gives us electrical resistance but it also results in a force exerted on the metal ion core.

An electron changes direction as a result of a scattering event. This change in direction is accompanied by an acceleration, which results in a force. The electrons are also accelerated within the electric field. When the overall electron drift velocity is established, the force on the ions due to electron scattering is in the direction of the electron flow – this is known as the wind force, F_w, which can be described as:

$$F_{\mathrm{w}} = -\mathrm{e}n_{\mathrm{e}}\lambda\sigma_t E \tag{24.2.2}$$

where n_e = density of electrons;σ_i = cross-section for collision; λ = mean free path.



The term *electron wind* comes from an analogy with, for example, a tree being blown in the wind. The wind is analogous to the electron current, the leaves on the tree to the metal ions in a conductor.

Net Force

In good conductors, such as the ones used in IC metallization, the electron wind force is the dominant force felt by the ion cores, resulting in atomic migration towards the anode. The net force on the ions can be represented as:

$$F_{net} = F_{\mathrm{wind}} + F_{\mathrm{direct}} = (Z_w + Z_d)\, ejp = Z^* ejp \tag{24.2.3}$$

where Z_{wind}, Z_{direct} *and* Z^* respectively refer to the effective valences for the wind force, direct force and the net force.

Diffusion

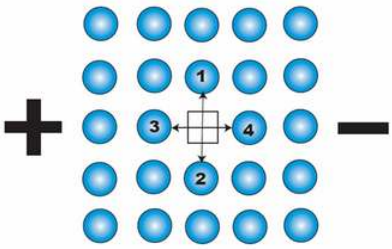
The migration of the metal ion cores occurs by diffusive jumps. In the cases of interest in metallization, this is self-diffusion or substitutional diffusion by the vacancy mechanism.

To find out more about diffusion, refer to Diffusion in the MATTER Software.

The rate of diffusion is described by the diffusion coefficient (or diffusivity) with units of m^2 s^{-1}. The diffusivity, D, has an Arrhenius dependence on temperature according to:

$$D = D_0 \exp(-Q/RT) \tag{24.2.4}$$

where D_0 = constant (m^2 s^{-1}); Q = activation energy for diffusion (J mol^{-1}); R = gas constant (8.314 J mol^{-1} K^{-1}); T = absolute temperature (K).
The diffusion coefficient and the activation energy are dependent on the nature of the material.



We depict above a 2-D section through a simple cubic metallic lattice showing a single vacant site. The vacancy can jump into either position 1, 2, 3 or 4 (i.e. it can exchange position with the atom in position 1, 2, 3 or 4). Without any external influence, the probability of the vacancy jumping into each site is equal, as its energy on every site is the same and the activation energy for a jump from site to site is always the same.

When an electric field is applied, the activation energies for a jump into sites 1 and 2 remain identical, whereas the activation energy for a jump into site 4 is smaller than that into site 3. This results in a biasing of vacancy diffusion jumps towards the cathode. This net flux of vacancies towards the cathode corresponds to the net flux of atoms (or ions) towards the anode. This flux of metal ions can be considered to be due to the effective charge, Z^*, on the ion and the associated net force as described above.

$$J = CDF_{net}/RT = CDZ^* e\rho j/RT \tag{24.2.5}$$

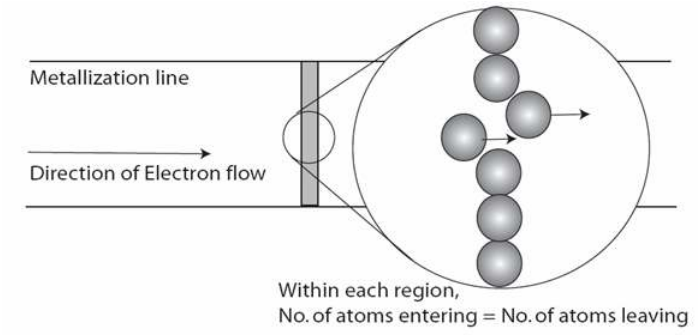
where J = atomic flux (atoms m^2 s^{-1}) ; C = atomic concentration (atoms m^{-3}); D = diffusivity (m^2 s^{-1}).

Note that this overall net flux is exceedingly minute. There is only a very slight biasing of atomic movement e.g. in 1000 atoms 499 would diffuse towards the anode whilst 501 diffuse towards the cathode, resulting in a net of 2 atoms migrating towards the cathode. Electromigration is a very slow process, taking time.

This page titled 24.2: The Theory of Electromigration is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

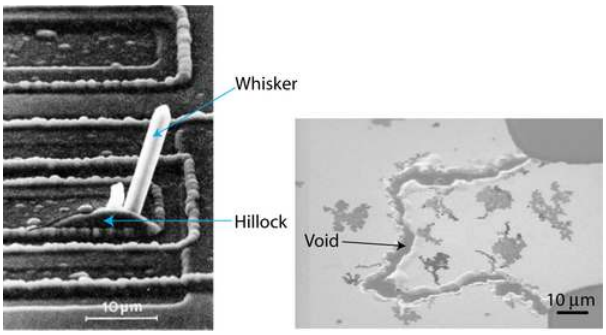
24.3: Electromigration Damage

Electromigration is the mass transport in a metallic conductor due to the momentum transfer between conducting electrons and diffusing metal atoms. Uniform electromigration within the metallization lines, if it could be maintained, would not be damaging: in steady-state, no damage should be observed other than at the beginning and end of the metallization line. This is because along the metallization line the number of atoms arriving in a given local volume is equal to the number of atoms leaving the volume, as shown in the diagram below.



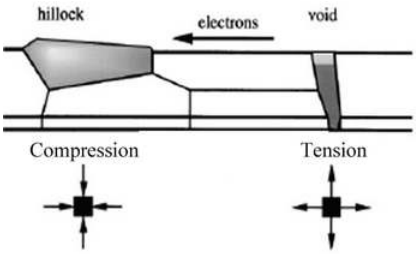
Damage to the metallization lines is caused by *divergences* in atomic flux. When the amounts of matter leaving and entering a given volume are unequal, the associated accumulation or loss of material results in damage.

When atomic flux into a region is greater than the flux leaving it, the matter accumulates in the form of a hillock or a whisker. If the flux leaving the region is greater than the flux entering, the depletion of matter ultimately leads to a void. These features are shown in the SEM micrographs below.



Source of images: (left) Microelectronic Materials by CRM Grovenor, IOP Publishing Ltd, Bristol (UK); (right) Dr Lo Veng Cheong

Regions of void formation are usually associated with neighbouring regions of material accumulation, as atoms are transported from one region to the other. Stresses develop within the metallization line as a result of the mass transport.



Source of image: O. Kraft, J.E. Sanchez Jr., M. Bauer, E. Arzt: Quantitative analysis of electromigration damage in Al-based conductor lines. J. Mat. Res. 12 (1997) p.2027-2037.

A stress gradient builds up within the metallic line and opposes the electromigration force. The formation of voids and hillocks partially relieves these stresses. A void forms to relieve tensile stresses, whilst hillock growth relieves compressive stresses.

The growth of voids and hillocks can be viewed in the videos below:

https://www.doitpoms.ac.uk/tlplib/el...12K_Stream.mp4

Video showing hillock growth (wide metallization lines)

https://www.doitpoms.ac.uk/tlplib/el...12k_stream.mp4

Video showing void growth and migration (wide metallization lines)

https://www.doitpoms.ac.uk/tlplib/el...12k_stream.mp4

Video showing void growth (modern, narrow metallization lines)

Voids and hillocks are detrimental to the metallization lines because a growing hillock could come into contact with other metallization lines resulting in a short circuit, an unintended pathway for the electricity to flow. As a void grows, the effective cross-sectional area of the metallization line decreases. This results in both *an increase in resistance* and current density within the system. The void ultimately leads to an open circuit when no material bridges across it.

These effects disrupt the intended functioning of the integrated circuit and result in failure. The nature of a microprocessor chip makes repair impossible and the failed chip has to be replaced.

This page titled 24.3: Electromigration Damage is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

24.4: Flux Divergence

Divergences in atomic flux, J, within the metallization line result in damage. The origins of non-steady-state mass transport (i.e. transport such that the concentration of atoms, *C* changes locally) can be understood in terms of the equation of continuity; for example, in one dimension:

$$\frac{\partial C}{\partial t} = -\frac{\partial}{\partial x}\left\{\frac{CDZ * e\rho j}{RT}\right\}\Big|_{T=\text{ constant}} - \frac{\partial}{\partial T}\left\{\frac{CDZ * e\rho j}{RT}\right\}\Big|_{x=\text{ constant}} \frac{\partial T}{\partial x} \tag{24.4.1}$$

The first term is from local structural property gradients at constant T and the second term is from T gradients and thermally dependent properties

Therefore when:

$\partial C/\partial t < 0$: mass depletion occurs and voids form (positive flux divergence)

$\partial C/\partial t > 0$: mass accumulates in growths – hillocks and whiskers (negative flux divergence)

$\partial C/\partial t = 0$: there is no change in atomic concentration and no damage occurs.

All the metal parts through which current flows in integrated circuits are potentially susceptible to electromigration effects. Different aspects are important in interconnects, contacts and vias.

As an added complication, the system observed is never static. Interactions within the metallization line result in the:

- movement of grain boundaries (grain growth)
- induction of heating
- recrystallization of the grain structure
- subtle evolution of microstructure and chemical composition.

The main contributions to flux divergence within metallization lines are: variations in *microstructure*, *material* or *temperature*.

Variation in Microstructure

Grain boundaries within Al-based metallization lines act as fast diffusion paths compared to the bulk system. The atomic environment at a boundary is less confining and contains fewer obstacles to diffusion. Accordingly, the activation energy for grain boundary diffusion is lower than that for bulk diffusion. As diffusion at grain boundaries far exceeds transport through the bulk of grains, the overall rate of atomic transport is greatly affected by the grain size, which determines the area of grain boundary in a given volume of sample.

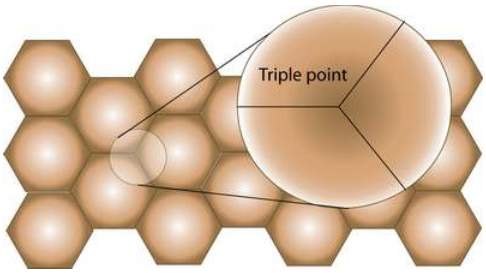
The grain structure within metallization lines can vary from place to place: for example, the deposition technique would affect the degree of variation of microstructure. The associated property variations also lead to electromigration damage. Differences in microstructure or properties can be very large or barely perceptible, involving variations in grain orientation, grain size, chemical composition, atomic diffusivity, effective valency and vacancy generation within the grain structure.

Electromigration-induced damage is profoundly affected by grain structure: for example single-crystal aluminium stripes exhibit “infinite” life. This is because atomic diffusion and drift, for a metallization based on Al, are dominated by transport along grain boundaries rather than transport through the bulk of grains.

Some examples of microstructural variations are:-

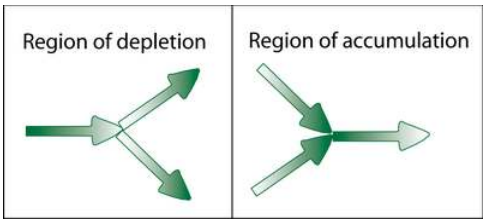
Triple points

A film with a uniform grain size has a network of grain boundaries meeting at triple points (strictly, triple *lines* in 3-D). In annealed thin films it is usual for the grain structure to be in a 2-D pattern, with each grain occupying the full thickness of the film.



For a common metallization, such as Al-Cu, where the migrating atoms are solely confined to grain boundaries, it is possible to get divergences in mass transport at the triple points found in interconnect and contact lines. The atomic diffusivity in a given boundary can vary widely, depending on the structure of the boundary, which is related to the crystallographic misorientation of the grains. But even if we assume that all the grain boundaries are high-angle boundaries with similar diffusivities, it can easily be understood that flux divergences must arise at triple points.

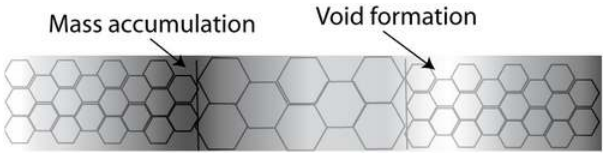
When the direction of migration is such that one grain boundary leads into and two boundaries lead away from a triple point, the negative flux divergence can lead to the formation of a void. The converse is true when two boundaries lead into a triple point and one leaves, giving mass accumulation and possible hillock formation.



It would be assumed that a metallization line consisting of numerous triple points would be full of defects. This is not the case, as described by the concept known as the *Blech length*. Often, no damage is observed at triple points because they are closer than this critical length for damage e.g. for a current density of 1 x 10^9 A m^2, a line length of several hundred micrometers will not be susceptible to electromigration. A longer metallization line would have a larger number of triple points and hence an increased likelihood of this kind of failure. However, other factors come into effect as metallization line length increases, thus resulting in yet further damage.

Differences in grain size

A fine-grained region contains more grain boundaries for atomic migration, than a coarse-grained region. Accumulation of atoms therefore occurs when atomic migration is from a fine-grained region moving into a coarser grained region. Conversely, voiding occurs when the migration is from a coarse to a finer grain size.



To avoid damage-inducing mass divergences, a bamboo-like grain structure (as shown in the SEM micrograph below) is desirable. The lack of continuous grain boundary paths for diffusion results in negligible mass transport along grain boundaries. The driving force for electromigration is predominantly perpendicular to the grain boundaries.



SEM micrograph of a metallization line containing a bamboo structure - grain boundaries are perpendicular to electron flow. (*Reprinted with permission from: A.G. Domenicucci et.al: Effect of copper on the microstructure and electromigration lifetime Ti-AlCu-Ti fine lines in the presence of tungsten diffusion barriers. J. Appl. Phys. 80, p4952. Copyright 1998. American Institute of Physics.*)

The animation below shows bamboo structure and its resistance to EM-damage.

Variations in Material

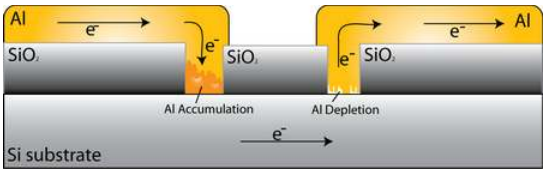
Different materials are used within an integrated circuit: for example, the Si substrate, W vias and Al interconnects. Differences in diffusion rates between two materials result in atomic flux divergence at the interface between them.

If the current flows from a material with a higher diffusivity to one with a lower diffusivity, the interface between the materials is a region of mass accumulation. Conversely when current flows from material with a lower diffusivity to another material with a higher diffusivity, void formation takes place at the interface. Electromigration-induced damage is most evident where the change in diffusivity is very large, e.g. at a Cu/barrier interface or an Al line/ W via interface.

Two main examples of this can be seen in: *contacts* and *vias*.

Contacts

When current flows through the contacts, electromigration causes metal atoms to move away from and towards the semiconductor interface. The inability to replenish metal that has been removed from and the inability to remove metal brought to the semiconductor interface results in a region of mass divergence. At the contact window either accumulation or depletion of Al occurs, depending on the direction of current flow.



Vias

The IC architecture requires use of successive interconnect levels and therefore of vias that enable current to flow between the layers. For Al-based metallization, W plug vias are commonly employed due to their high reliability, though structures consisting of

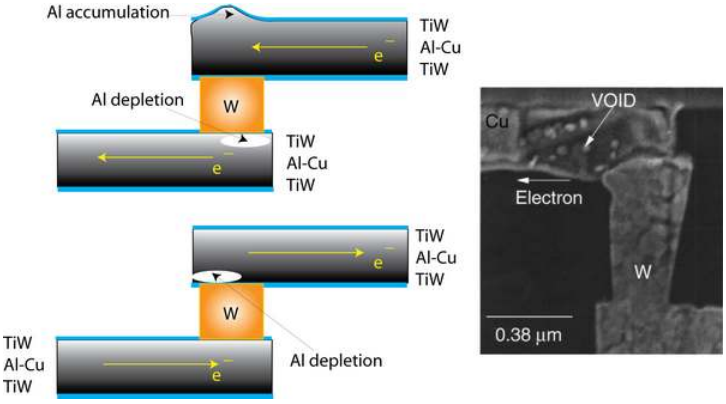
Al-Cu vias can be used but are more difficult to make. As Al metallization tends to migrate whilst W exhibits negligible atomic transport (as shown in values of self-diffusion below), via-interconnect interfaces become sites of potentially large mass divergences.

The bulk lattice diffusivity values of Al and W at 300 K (assumed service temperature of the IC) are*:

$D_{Al,\ bulk} = 3.2\times10^{-29}\ m^2\ s^{-1}$

$0 < D_{W,\ bulk} < 5\times10^{-88}\ m^2\ s^{-1}$ (negligible)

**Values calculated from data in Smithells Metals Reference Book, 6th Ed., Mechanisms of Diffusion p.13-11; E.A. Brandes (editor), Butterworths, London (UK), 1983.*



Source of SEM image of void formation at a via; R. Rosenberg et. al.: Copper metallization for high performance silicon technology, Annu. Rev. Mater. Sci. 20 p.229, 2000.

Degradation at vias is dependent not only on composition and grain structure, but also on the direction of current flow. The two modes of electromigration damage are shown above (in a schematic form and from an SEM micrograph). Voids form where the electrons flow away from the via, while a hillock forms when electrons flow towards the via.

In addition, thinned metal conductors and corners induce excessive *Joule-heating effects* and current crowding that lead to accelerated degradation.

Variations in Temperature

Temperature differences along the metallization line cause flux divergence because the diffusion coefficient is dependent on temperature. At higher temperatures, diffusion rates are increased. If there is a variation of temperature along the metal line, regions of accumulation and void growth would develop over time. Although the silicon substrate acts as a very good heat sink, the temperature along a line can vary because of heat generation in the underlying semiconductor components and because of local heating due to *current crowding* in the metallization itself.

This page titled 24.4: Flux Divergence is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

Welcome to the Engineering Library. This Living Library is a principal hub of the LibreTexts project, which is a multi-institutional collaborative venture to develop the next generation of open-access texts to improve postsecondary education at all levels of higher learning. The LibreTexts approach is highly collaborative where an Open Access textbook environment is under constant revision by students, faculty, and outside experts to supplant conventional paper-based books.

24.6: Alternatives to Aluminium Metallization

Technological pressures on the speed and reliability of integrated circuits has caused a need for changes to be made in the choices of materials used for metallization lines today. In selecting a material for metallization, it is necessary to consider what its desirable properties would be. These include:

- low resistivity
- mechanical stability, good adherence, and low stress
- easy to form
- easy to etch for pattern generation
- stable throughout processing, including: high temperature sinter, dry or wet oxidation, gettering, phosphorus glass (or any other materials) passivation, and metallization
- should not contaminate devices, wafers or working apparatus
- should be stable in oxidation ambients and is able to form a stable oxide
- no reaction with other components
- good surface smoothness
- low contact resistance, minimal junction penetration and low electromigration for use in window contacts
- very low cost

During the growth of the microelectronics industry, metallization lines have moved from the use of *pure Al* to *Al-Cu alloy* and currently to *pure Cu*.

Using Pure Aluminium

Commercially, it would be assumed that a material having the lowest resistivity would be used for IC interconnections. This is because, the lower the resistivity of the material, the smaller the resultant RC-delay, and the faster the signal transmission between devices – in order to be useful, resistivities must be below 50 μΩ cm.

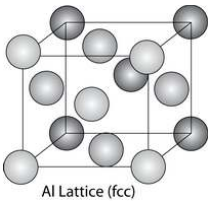
Metal	Electrical resistivity (at 20 ºC) / μΩ cm
Al	2.65
Cu	1.67
Au	2.35
Ag	1.59
W	5.65

Data obtained from Metals Handbook, 8th Edition, ASM, Metals Park, Ohio.

Yet, the standard interconnection material previously used was Al, despite having other materials such as Ag, Cu and Au with lower resistivity values (as shown in the table). This is because it was found that:

- Ag is too prone to attack by S and O, and could not maintain its low resistivity over the device lifetime.
- Cu is prone to oxidation resulting in a vast increase in resistivity of the material.
- Ag and Au are very difficult to deposit as very low resistance films.

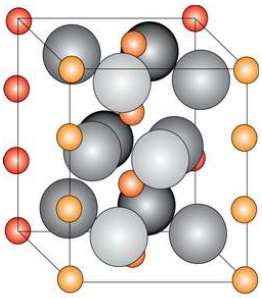
Al has low resistivity and can be easily deposited. It can be dry etched, does not contaminate Si as it forms a protective Al_2O_3 oxide layer which prevents further oxidation, has excellent adhesion to dielectrics, and is able to form ohmic contacts.



These factors resulted in the universal use of Al-based metallization in ICs. Unfortunately, Al has a problem forming contacts with shallow junctions, has difficulty achieving a good mechanical and electrical connection as a vacuum deposited thin film and is very prone to electromigration failure.

Using Aluminium-Copper Alloys

Due to problems associated with electromigration in pure Al metallization, there was a real need to improve the metallization. In the mid 1960s, a mis-focused electron beam evaporated some Cu, instead of the charge metal (Al). The resultant Al-Cu alloy films were found to be significantly resistant to electromigration.



This is shown in the table below, as the amount of Cu content increases the critical product $(jl)_c$ is increased and the effective charge, Z^*, is decreased.

	Cu content (wt%)	Critical product $(jl)_c$/ A cm^{-1}	Effective charge number, Z^*
Al standard	0	244	18
AlSiCu cold	0.5	211	13
AlSiCu hot	0.5	298	13
AlCu	2	833	5

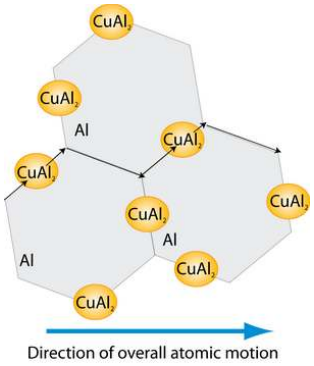
Data from: 'Quantitative analysis of electromigration damage in Al-based conductor lines' by O. Kraft et.al. J. Mater. Res., 12 (1997) 2027.

Reducing the effect of electromigration

A commonly used Al-Cu alloy is Al-1.5at%Si-4at%Cu. Despite its increase in resistivity from 2.86 μΩ cm (pure Al) to 3 μΩ (Al-Cu alloy), the improvement of electromigration resistance made Al-Cu alloys the system of choice in metallization.

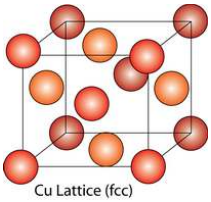
There are several reasons for the increase in electromigration resistance:

- As most mass transport in Al occurs along grain boundaries, the addition of Si and Cu reduced the grain boundary diffusion rate, by increasing the activation energy of grain-boundary diffusion.
- The addition of Cu to the Al alloy, results in the formation of $CuAl_2$ precipitates. These precipitates form primarily on grain boundaries, impeding grain boundary diffusion. The precipitates also act as reservoirs of Cu, delaying any damage, as the precipitate must dissolve before the level of Cu in grain boundaries can fall significantly. Dissolution is rather slow as a steady-state flow of Cu can be established from precipitate to precipitate before migration takes place.



Using Pure Copper

The trend towards narrower interconnections, and faster and more reliable devices resulted in the examination of the possibility of using Cu-based metallization. It was found that the change to a Cu-based material greatly improved the reliability of devices.



A comparison (see table below) shows the reasons for the currently increasing use of Cu-based metallization in the microelectronics industry:

Advantages of Cu	Disadvantages of Cu
• Cu is more conductive than Al, thus allowing finer metallization with lower resistive losses (ρ_{Cu} = 1.67 μΩ cm, ρ_{Al} = 1.65 μΩ cm) • Atomic migration in Al occurs along grain boundaries and surfaces. There is little or no bulk transport in Cu. (Bulk self diffusivity of Al: 1.9×10^{-12} m^2 s^{-1}, Cu: 1.8×10^{-16} m^2 s^{-1} at 933 K) • Al is very susceptible to electromigration, (get rapid formation of hillocks and voids), whilst Cu is less vulnerable as it has higher mass and a higher melting point. • Cu is also less likely to fracture under stress.	• Cu diffuses rapidly into Si and SiO_2, causing deep-level defects as it contaminates the Si. • The main transport path in Cu is the top surface of metallization lines. This results in some electromigration damage. Al-based metallization does not exhibit this, as Al forms a protective oxide layer preventing surface transport. • As Cu cannot be dry etched, it was necessary to develop an electroplating process for making copper networks, the dual-damascene chemical-mechanical polishing (CMP) process, and an effective linear material for use as a copper diffusion barrier and to promote adhesion.

Due to the favourable properties of Cu, it is possible for the chip size to be reduced, whilst increasing the speed and complexity of the device.

Cu has proved to be an excellent metallization material as it has an improved current carrying capability and high electromigration resistance. The disadvantages have been overcome using new thin-film technology and careful materials selection. It is therefore possible for component size to be further reduced, increasing the speed and complexity of the device.

This is sufficient for now, though there is a continuing need to achieve high conductivity and minuscule dielectric constants for future devices. Therefore, in order for the microelectronics industry to keep up with Moore's Law and the ever increasing consumer needs, there needs to be introduction of new materials and processes.

This page titled 24.6: Alternatives to Aluminium Metallization is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

24.7: Summary

This TLP has covered several points:-

- The reliability of microelectronic devices is important, as our lifestyles depend on the current technology and we expect continuing improvements in performance. Electromigration reduces device reliability, which is a problem for the electronic industry.
- Electromigration is the transport of material in a conductor under the influence of an applied electric field.
- Electromigration phenomena occur in all conductors on application of an electric field. Damage is commonly observed within narrow metallization lines in integrated circuits, as they are exposed to particularly high current densities.
- The electromigration force, which causes atomic migration, is a combination of two forces – the direct force and the wind force, related by:

$$F_{\text{net}} = F_{\text{wind}} + F_{\text{direct}} = Z^* ejp \tag{24.7.1}$$

  where Z^* = *effective charge; j = current density* (A m^{-2})*;* and ρ = resistivity (Ω m).

  Mass atomic transport towards the cathode occurs as the net force biases the net diffusion direction.
- Damage induced by electromigration results in hillock and void formation. Electromigration-induced damage is caused by divergences in atomic flux, themselves caused by place-to-place variation in:
  - Microstructure
  - Material
  - Temperature
- Choosing the best materials for use in an integrated circuit, the most suitable processing methods and good device design can assist in reducing the resultant electromigration damage.
- The device lifetime can be roughly estimated using Black's Law to extrapolate median time to failure at service temperature from accelerated test conditions:

$$t_{50} = cj^{-n} e^{\frac{E_a}{kT}} \tag{24.7.2}$$

- where t_{50} = median time to failure of metal lines subjected to electromigration (hrs); c = constant based on the metal line properties (units depend on exponent n); j = current density (A m^{-2}); n = value between 1 and 7 (though commonly 2); E_a = activation energy (J) [within the range 0.5–0.7 for Al]; k = Boltzmann constant (1.38×10^{-23} J K^{-1}); and T = temperature (K).
  - Due to the increased miniaturization of microelectronic devices, there is a need for new materials and processes to be used for metallization lines. One such development is the change from Al-based to Cu-based metallization in high performance devices. Further speed, reliability and miniaturization requirements are spurring on the search for the use of new materials within integrated circuits.

Going further

Books and Papers:

- *Reliability and Failure of Electronic Materials and Devices* by Milton Ohring, Academic Press, San Diego, 1998.
  Provides a good overall explanation on the topic of electromigration
- Review Article: Electromigration in integrated circuit conductors by J R Lloyd (*J. Phys. D: Appl. Phys*. **32** (1999) R109-R118).
  A good summary on electromigration in integrated circuit conductors
- VLSI Technology by S.M. Sze (editor), Mc-Graw-Hill Book Company, New Jersey, 1988.
  Contains a good section on metallization and its choice of use for different materials

Website:

- Article on electromigration in Wikipedia

This page titled 24.7: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

24.8: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which force/forces act on metal ions within a conductor, during electromigration – under the application of an electric field? (You may pick more than one answer)

(a) Wind force, $F = an_e\lambda d\sigma_i j\rho$

(b) Direct force, $F = aZej\rho$

(c) Mechanical back force, $F = \Omega\Delta\sigma/\Delta x$

(d) Moment force, F = Moment/distance from pivot

(e) None of the above

Answer

(a) wind force – scattering electron events within the conductor result in exerting a force on the metal ion.

(b) direct force – due to the electrostatic pull from the electric field.

(c) mechanical back force – results from the build up of stresses as mass atomic migration takes place within the conductor.

The median life of pure Al stripes of width, w (in cm) and thickness, d (in cm) at 50 °C is given by:

$t_{50} = 4.4\times10^{12} wdj^{-n} \exp(E_e/kT)$

where $n = 2$, $E_e = 0.49$ eV, $j = 1$ x 105 A cm^{-2}, $w = 0.4$ µm and $d = 0.5$ µm.

What is the median life of the interconnect?

(a) 0.38 hrs

(b) 3.83 hrs

(c) 38.3 hrs

(d) 380 hrs

Hint: Ensure that you are using the correct units:

w is in cm (4×10^{-5} cm) and d is in cm (5×10^{-5} cm).
T is in K (323 K)
E_e needs to be converted from eV to J using the relation: 1 eV = 1.6 x 10^{-19} J.
k = 1.38 x 10^{-23} J K-1

Answer

(c) 38.3 hrs

How does the median life of pure Al stripes compare with that of Al-Si interconnects of the same dimensions as above, for which:

$$t_{50} = 2.2 \times 10^{15} \mathrm{wdj}^{-\mathrm{n}} \exp(0.54\mathrm{eV}/\mathrm{kT}) \tag{24.8.1}$$

a	the same
b	less by 10%
c	more by 10%
d	more by 100,000%

Answer

D

Which material properties are desirable for use as interconnects?

a	low availability
b	low resistivity
c	low melting point
d	low electromigration resistance

Answer

B

Which material is the mostly used in vias?

a	Aluminium-based alloys
b	Tungsten
c	Titanium Nitride
d	Silver

Answer

A

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

In a solid conductor, the mechano diffusion force, F_m, acting on an atom, in a gradient of hydrostatic stress, $d\sigma_H$ / dx, is given by:

$$F_m = \frac{1}{C_0}\frac{d\sigma_H}{dx} \tag{24.8.2}$$

where C_0 = no. of atoms per unit volume.

A conductor has along it, a segment of length, L, which has a raised atomic diffusivity. The material of the conductor develops damage when the difference in hydrostatic stress between the ends of such a segment reaches the value $(\Delta\sigma_H)_{crit}$. Derive the critical condition for the damage to develop in terms of L and current density, j, in the conductor.

Answer

The mechano diffusion force, F_m, opposes the electromigration force $F_{wind} + F_{direct} = Z^* ej\rho$.

At L and $(\Delta\sigma_H)_{crit}$ damage is not yet observed. Above this critical value, damage would occur. Therefore at this point, both forces have to be equal and opposite:

$$\frac{1}{C_0}\frac{d\sigma_H}{dx} = Z^* e\rho j \tag{24.8.3}$$

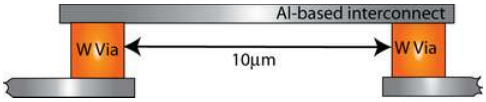
Assuming $(d\sigma_H / dx)$ can be approximated to $(\Delta\sigma_H)_{crit} / L$,

$$\frac{1}{C_0}\frac{\Delta\sigma_{h,rrit}}{L} = Z^* e\rho j \tag{24.8.4}$$

$$(jL) = \frac{1}{C_0}\frac{\Delta\sigma_{H,crit}}{Z^* e\rho} \tag{24.8.5}$$

This is the critical condition for damage to develop.

An IC has several levels of Al-based metallization linked by W vias. In a test, electromigration damage, develops in an Al-based conductor linking two vias when the current density flowing through it from one via to the other 10 μm away exceeds 9×10^9 A m^{-2}. It is planned to use the same materials with a current density of 1.2×10^{10} A m^{-2}. What condition must be met by the length of links between vias for damage to be avoided? How will the temperature of the operation affect the development of electromigration damage?



Answer

The critical condition for damage to develop within the system is:

$(jL)_{crit} = (9\times10^9)(10\times10^{-6}) = 90{,}000$

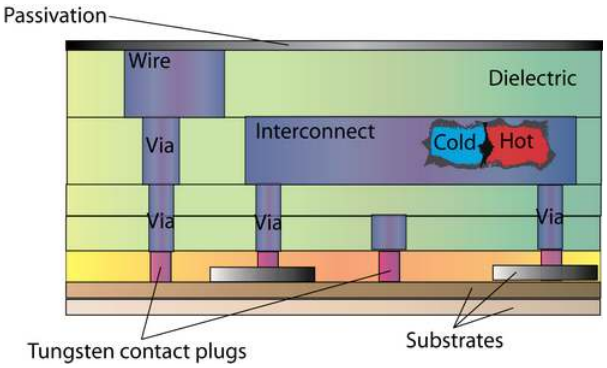
Therefore to avoid any damage, the jL product has to be below the critical value. For current density, $j = 1.2\times10^{10}$:

$jL = 90{,}000$
$1.2\times10^{10}L = 90{,}000$
$L = 7.5\times10^{-6}$ m

Therefore, in order for damage to be avoided, the length must be below 7.5 μm. This ensures that the mechano diffusion forces counters the electromigration force, thus preventing damage.

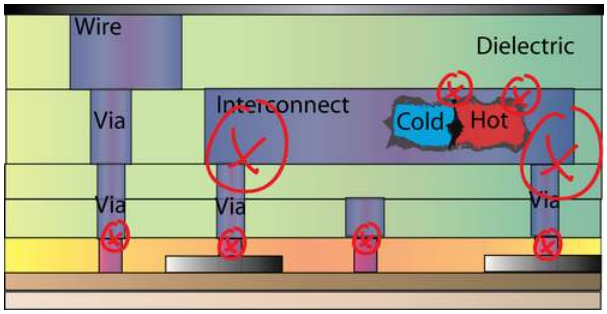
The flux of atoms is dependent on temperature, within the terms of $D = D_0\exp(E/kT)$ and T. Therefore a higher temperature results in a faster rate of flux and faster occurrence of electromigration damage.

In the schematic diagram of a microelectronic device (below), where would electromigration-induced damage be observed? Give your reasons. What type of material would you use for the vias and interconnects, and why?



Answer

The answers are marked in the diagram below:



1. Interface between the tungsten and metallic via – differences in diffusivity rates result in negligible electromigration within tungsten, whilst atomic migration occurs in the metallic via.

2. Hot and cold regions within the metallization line – different rates of atomic migration result in an overall atomic flux, where hillocks and voids form in volumes of mass accumulation and depletion.

3. Bends within the metallization line – a decrease in metallization cross-sectional area results in an increase in Joule-heating and temperature. Current crowding takes place causing a flux divergence.

Cu is the most suitable for use as interconnects and vias, as it has a low resistivity (ρ = 1.67 $\mu\Omega$ cm), is resistant to electromigration, is easily processed into metallization lines (using newly developed technology) and is a relatively cheap material.

Alternatively, Al-Cu can be used for interconnects, coupled with W vias. This combination is less electromigration resistive than Cu.

This page titled 24.8: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

25: Ellingham Diagrams

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

On completion of this TLP you should:

- Be aware of the structure, origin and use of the Ellingham Diagram;
- Be able to use the interactive diagram included to find thermodynamic data quickly and effectively including:
  - ΔG, k, p_{O_2} for a standard and non-standard reactions across a range of temperatures;
  - Relative stabilities of elements with respect to oxidation etc or conversely that of oxides, sulphides etc;

25.1: Introduction
25.2: Thermodynamics
25.3: Chemical Reactions
25.4: The Equilibrium Constant
25.5: The Ellingham Diagram
25.6: Application
25.7: Partial Pressure of Reacting Gas
25.8: Reading p_O2 from the Ellingham Diagram
25.9: Other Gas Mixtures
25.10: Reducing Agents
25.11: Other Atmospheres and Non-Standard States
25.12: Summary
25.13: The Interactive Ellingham Diagram
25.14: Questions

This page titled 25: Ellingham Diagrams is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.1: Introduction

The Ellingham Diagram, originally constructed for oxides, is a tool to find a variety of thermodynamic data quickly, without the need for repetitive calculation.

The diagram is essentially a graph representing the thermodynamic driving force for a particular reaction to occur, across a range of temperatures. With the data for several reactions plotted on the same graph, the relative stabilities of different elements with respect to, for example, their oxides, can be seen. It is also possible to compare the relative driving force for an element for oxidation or sulphidation in an environment containing both oxygen and sulphur as reactants.

This page titled 25.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.2: Thermodynamics

Here we shall go through the basic thermodynamics that lies behind use of the Ellingham diagram. First, we will establish the link between the thermodynamics of a reaction and its chemistry.

The Gibbs free energy, *G*, of a system can be described as the energy in the system available to do work. It is one of the most useful state functions in thermodynamics as it considers only variables contained within the system, at constant temperature and pressure.

It is defined as:

$$G = H - TS \tag{25.2.1}$$

Here, *T* is the temperature of the system and *S* is the entropy, or disorder, of the system. *H* is the enthalpy of the system, defined as;

$$H = U + \mathrm{p}v \tag{25.2.2}$$

Where *U* is the internal energy, p is the pressure and *v* is the volume.

To see how the free energy changes when the system is changed by a small amount, we can differentiate the above functions:

$$dG = dH - TdS - SdT \tag{25.2.3}$$

and;

$$dH = dU + pdv + vdp \tag{25.2.4}$$

From the first law,

$$dU = dq' - dw \tag{25.2.5}$$

and from the second law

$$dS = \frac{dq'}{T} \tag{25.2.6}$$

we see that,

$$dG = dq' - dw + pdv + vdp - TdS + SdT \tag{25.2.7}$$

$$= -dw + pdv + vdp - SdT \tag{25.2.8}$$

$$= vdp - SdT \tag{25.2.9}$$

Since work, dw = pdv

The above equation shows that if the temperature and pressure are kept constant, we see that the free energy does not change. This means that the Gibbs free energy of a system is unique at each temperature and pressure.

At a constant temperature d*T* = 0 and so

$$dG = vdp \tag{25.2.10}$$

We can find *G* for the system by integration. To do this we need the system's equation of state, to give a relationship between *v* and p.

We will consider an ideal gas. For one mole of an ideal gas the equation of state is;

$$v = \frac{RT}{p} \tag{25.2.11}$$

so

$$dG = RT\frac{dp}{p} \tag{25.2.12}$$

Integrating:

$$GRTln(p) + const. \tag{25.2.13}$$

This we can express as

$$G = G^o + RTln(\frac{p}{p^o}) \tag{25.2.14}$$

This page titled 25.2: Thermodynamics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.3: Chemical Reactions

We will now see why studying the free energy of a system is useful in determining its behaviour.

The free energy change, ΔG, of a chemical reaction is the difference in free energy between the products of the reaction and the reactants. If the free energy of the products is less than the free energy of the reactants there will be a driving force for the reaction to occur.

For the reaction

$$A + B \rightarrow C \tag{25.3.1}$$

the free energy change,

$$\Delta G = G_C - G_A - G_B \tag{25.3.2}$$

$$= \Delta G^o + RTln(\frac{\frac{p_c}{p_c^0}}{\frac{p_A p_B}{p_A^0 p_B^0}}) \tag{25.3.3}$$

$$\Delta G^o + RTln(\frac{p_c}{p_A p_B}) \tag{25.3.4}$$

if the standard states

$$p_A^0 = p_B^0 = p_C^0 = 1bar \tag{25.3.5}$$

We see that the free energy change of a reaction is determined by the relative quantities of reactants and products.

This page titled 25.3: Chemical Reactions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.4: The Equilibrium Constant

A chemical reaction will occur if the total free energy of the products is less than the total free energy of the reactants. (ie. The free energy change for the reaction is negative.) If the system containing the reactants and products is closed (if there is no input of reactants, for example), the concentration of reactants will decrease and the concentration of products will increase as the reaction proceeds. This will alter the state of the system and therefore alter the free energy change for the reaction (see equation 14, above).

The reaction will continue if the free energy change remains negative. Hence, the system proceeds down a free energy *gradient* with respect to composition and this gradient provides the *driving force* for the reaction to proceed. The system alters the quantities of reactants and products in response to the driving force until a minimum in free energy is reached and the gradient is zero. This is a point of equilibrium.

At equilibrium the free energy change for the reaction is equal to zero:

$$\Delta G = \Delta^o + RTln(\frac{p_C}{p_A p_B}) \tag{25.4.1}$$

$$= 0 \tag{25.4.2}$$

Therefore

$$\Delta G = -RTln(\frac{p_C}{p_A p_B})_{equilibrium} \tag{25.4.3}$$

$$RTln(K_p) \tag{25.4.4}$$

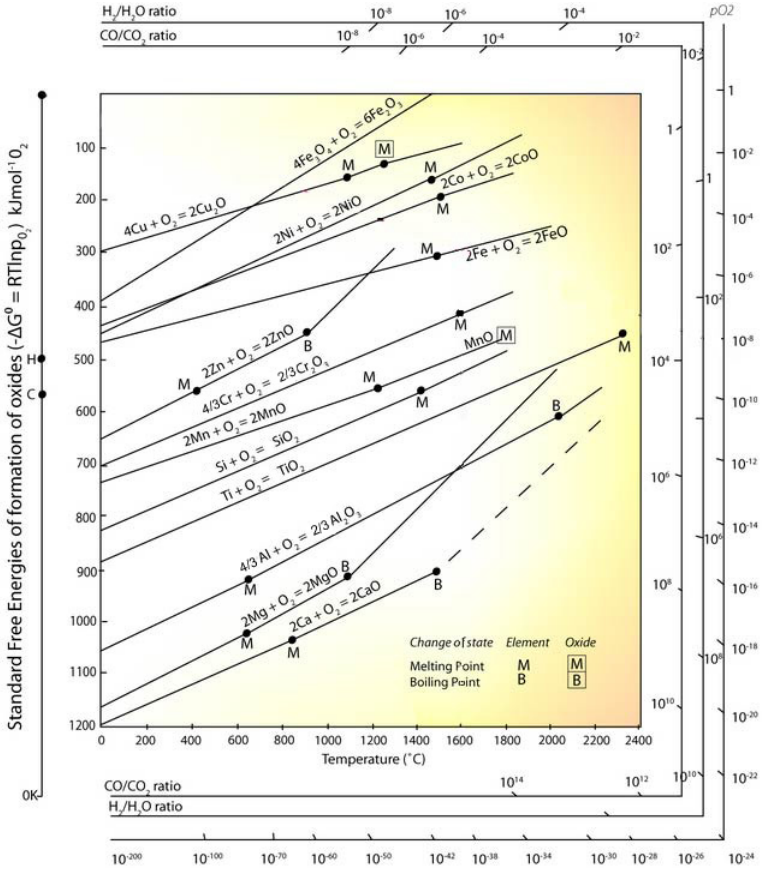
For the composition at equilibrium, the quotient is equal to G_P - the equilibrium constant for the reaction at constant pressure. We see that the equilibrium composition of the system is defined by the standard free energy change, ΔG°. Equation 15 provides a link between the thermodynamics of a reaction and its chemistry. ΔG° for a reaction is hence a very useful value to know.

This page titled 25.4: The Equilibrium Constant is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.5: The Ellingham Diagram

The Ellingham diagram plots[1] the standard free energy of a reaction as a function of temperature. Originally the values were plotted for oxidation and sulphidation reactions for a series of metals, relevant to the extraction of metals from their ores (extraction metallurgy). These reactions generally involve the reaction of a gaseous phase (the oxidising gas) with (almost) pure condensed phases (the metal and oxidised compounds).

By using the diagram the standard free energy change for any included reaction can be found at any temperature. Along with allowing you to calculate the equilibrium composition of the system, the data on the diagram is useful in other ways, as we shall see.



Note: Due to shortage of reliable experimental data, some of these lines are not plotted over the complete range of temperature.

This page titled 25.5: The Ellingham Diagram is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

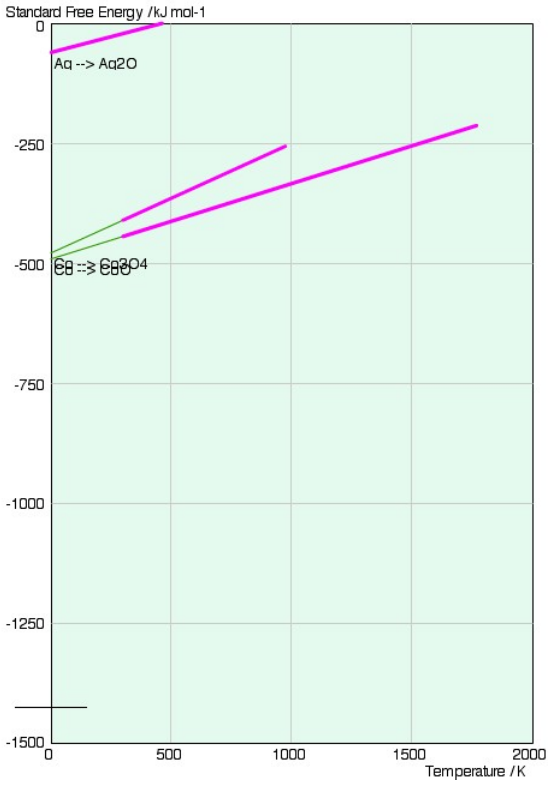
25.6: Application

We shall consider the following reactions on the Ellingham diagram below: the oxidation of silver to form $Ag_2O_{(s)}$; and the oxidation of cobalt to form $CoO_{(s)}$ and Co_3O_4:

$4Ag_{(s)} + O_2 = 2Ag_2O_{(s)}$;

$2Co + O_{2\,(g)} = 2CoO_{(s)}$

$3Co + 2O_{2\,(g)} = Co_3O_{4\,(s)}$.



The graph consists of lines of

$$\Delta G^o = \Delta H^o - T\Delta S^o \tag{25.6.1}$$

Although ΔS and ΔH vary with temperature (see here to see why), the changes are so small as to be negligible, and so the lines are approximately straight. This means that they are of the form

$$y = mx + c \tag{25.6.2}$$

We can immediately see that the standard free energy change is greater (more negative) for the cobalt reaction relative to that of silver at all temperatures. This means that at all temperatures the equilibrium constant is larger for the cobalt reaction-the composition is further weighted towards the products of the reaction.

This is the reason that metals that appear higher up on the diagram are more stable than those metals that appear lower down, and are more likely to be found in their pure solid form. The gradient of the two lines is approximately the same.

We can see that the gradient of the lines is simply the standard entropy change for the reactions;

$$\frac{\partial \Delta G^o}{\partial T} = -\Delta S^o \tag{25.6.3}$$

This is evident from the reactions, which both involve the elimination of one mole of gas - a large decrease in entropy. This is the reason for the positive slope of the lines.

The reason for the change in slope is the change in phase of a component of the system, which alters the entropy change.

As the standard free energy change for both reactions is still negative (below 460K) the large decrease in entropy must be counteracted by a large enthalpy of reaction. This is indeed the case. The intercept of the lines with 0K gives the enthalpy of the reaction:

$$\Delta G^{\circ}|_{0K} = \Delta H^{\circ} \tag{25.6.4}$$

We can hence see that the relative stability of the *oxide* of cobalt compared with the oxide of silver is due to the much larger standard enthalpy of reaction.

This page titled 25.6: Application is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.7: Partial Pressure of Reacting Gas

Using equations (15) we can see that the equilibrium constant is related to the partial pressures of reacting gases:

$$K_p = \frac{p_C}{p_A p_B} \tag{25.7.1}$$

for the reaction

$$A + B \rightarrow C \tag{25.7.2}$$

(Remember that these pressures must be related to a standard state.)

For a metal oxidation reaction

$2M_{(s)} + O_{2\,(g)} = 2MO_{(s)}$,

the equilibrium constant has the form

$$K_p = \frac{1}{p_{O_2}} \tag{25.7.3}$$

Click here for a more detailed discussion of the equilibrium constant.

We can therefore find the equilibrium partial pressure of oxygen at a particular temperature from the value of $\Delta G°$:

$$p_{O_2}|_{eq.,T} = exp(\frac{\Delta G^\circ}{RT}) \tag{25.7.4}$$

The equilibrium partial pressure of oxygen is the pressure at which the driving force for the reaction is zero. From equation 14 we see that if the partial pressure of oxygen is greater than this value, the free energy change for the reaction is negative and there is a driving force for the reaction to take place. Metal will be oxidised, and the partial pressure of oxygen will drop until it reaches equilibrium. This is effect described by Le Chatelier's principle.

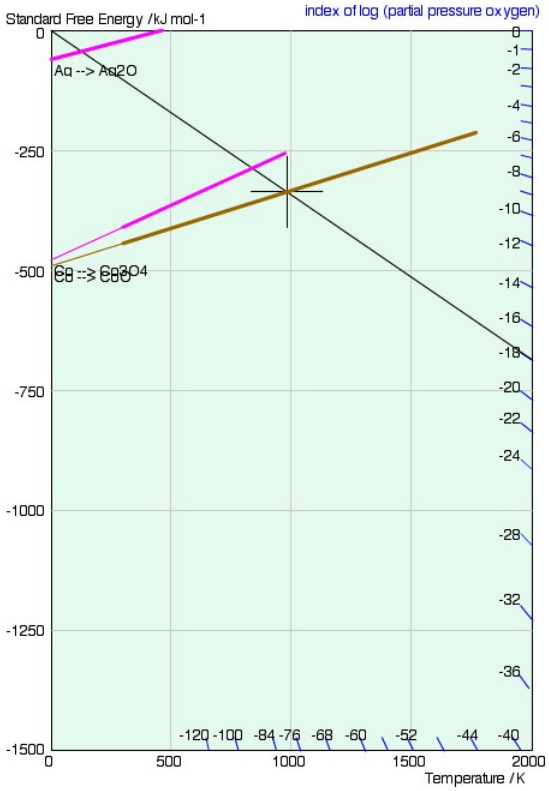
If the partial pressure of oxygen is below the equilibrium value, oxidation is avoided. (In fact, the metal oxide will disassociate to form metal plus oxygen gas-this is because there is a driving force for the reaction to proceed *backward*. For this reason the equilibrium partial pressure is often known as the dissociation pressure.)

This page titled 25.7: Partial Pressure of Reacting Gas is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.8: Reading p_O2 from the Ellingham Diagram

In to avoid calculating the equilibrium partial pressure for each value of $\Delta G°$, Richardson[2] added a nomographic scale to the Ellingham diagram. The equilibrium partial pressure is found as follows:

A line is drawn from the origin of the graph (T = 0, ΔG = 0) through the point on the Ellingham line of interest, at the required temperature. The equilibrium partial pressure is read off at the point where the drawn line crosses the nomographic scale.



The scale is simply derived by considering the change in free energy of one mole of ideal gas, from p = 1 atm to p = P. We know that

$$\Delta G^{\circ} = RTln(pO_2) \tag{25.8.1}$$

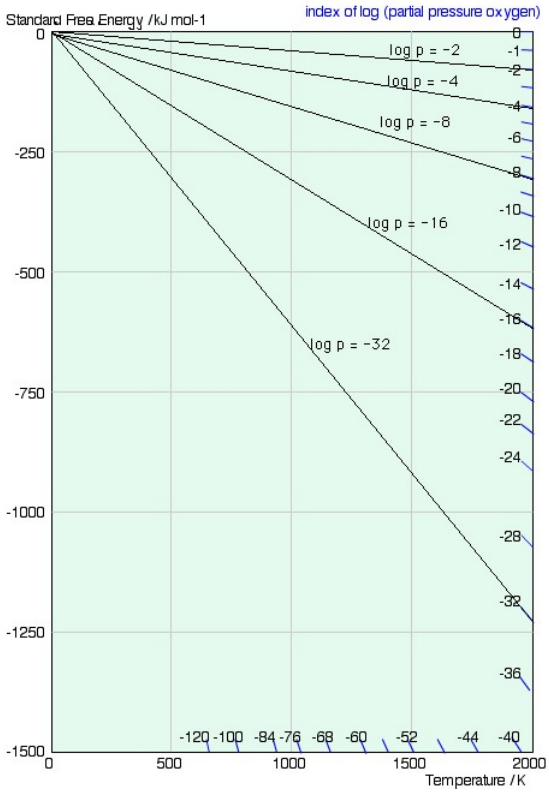
and using equation 12

$$G = G^{\circ} + RTln(\frac{p}{p^{\circ}}) \tag{25.8.2}$$

we find that this is equal to the difference in free energy of a mole of gas at 1 atm and a mole of gas at P atm:

$$\Delta = RTlnP \tag{25.8.3}$$

ie. a line from the origin with gradient *R*lnP. A series of lines for different values of the partial pressure, P, is shown below.



[2] F.D. Richardson and J.H.E. Jeffes, "The Thermodynamics of Substances of Interest in Iron and Steel Making from 0°C to 2400°C: I-Oxides," J. Iron and Steel Inst. (1948), **160** 261.

This page titled 25.8: Reading p_O2 from the Ellingham Diagram is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.9: Other Gas Mixtures

The oxygen required to cause oxidation in the gas phase need not to come from oxygen gas. Consider the following reaction:

$2CO_{(g)} + O_{2(g)} = 2CO_{2(g)}$

For this reaction,

$$K_{\frac{CO}{CO_2}} = \frac{p^2_{CO_2}}{p^2_{CO} p_{O_2}} \tag{25.9.1}$$

or

$$p_{O_2} = \frac{p^2_{CO_2}}{p^2_{CO}} K_{\frac{CO}{CO_2}} \tag{25.9.2}$$

and hence

$$ln\frac{1}{p_{O_2}} = lnK_{\frac{CO}{CO_2}} + 2ln\frac{p_{CO}}{p_{CO_2}} \tag{25.9.3}$$

$$= \frac{-\Delta G}{RT} \tag{25.9.4}$$

We see that p_O2 is equivalent to a ratio:

$$\frac{p_{CO_2}}{p_{CO}} \tag{25.9.5}$$

Another nomographic scale may be added to the diagram, with a new origin, C where the CO/CO_2 line crosses the y-axis.

Similarly for the reaction $2H_2 + O_2 = 2H_2O$; p_{O_2} is equivalent to

$$\frac{p_{H_2O}}{p_{H_2}} \tag{25.9.6}$$

Adding a further nomographic scale to the diagram, we see that the equilibrium pressure ratios of CO and CO_2 or H_2 and H_2O for a given oxidation of metal, or reduction of an oxide, can be deduced at a given temperature from the diagram.

This page titled 25.9: Other Gas Mixtures is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.10: Reducing Agents

A major application of the Ellingham diagram is the determination of the conditions required to reduce metal compounds, such as oxides or sulphides, to obtain pure metal. This is often the basis of extraction metallurgy, the extraction of metals from their ores. It is also very important in the recycling of metals. See the Ellingham Diagram section of Recycling of Metals TLP for more information.

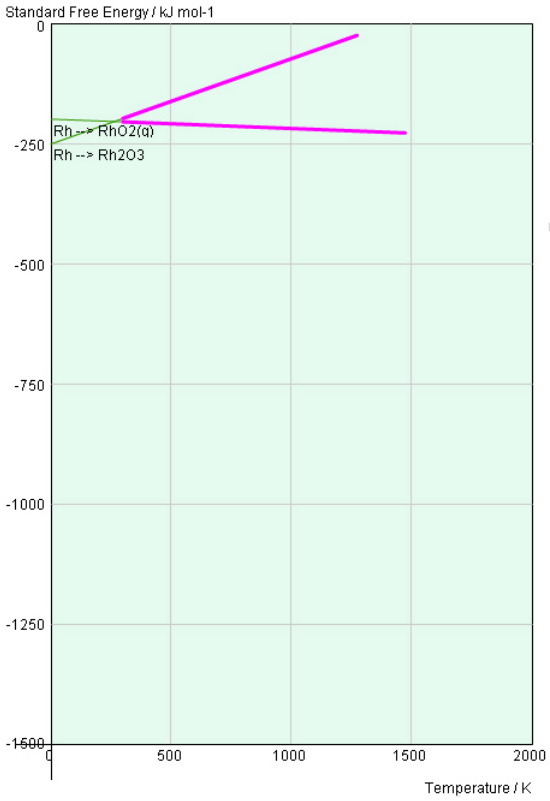
A chemical is a reducing agent with respect to a particular metal when the free energy change for its oxidation is more negative than the free energy change of oxidation of the pure metal. This means that when the reducing agent is placed in a closed system containing the metal ore there is a driving force for the dissociation of metal and oxidiser (ie. oxygen or sulphur gas, for example).

Consider the two oxidation reactions below, whose lines on the Ellingham diagram cross each other:

$2A + O_2 = 2AO$ (a)

and

$B + O_2 = BO_2$. (b)



As we can see, the y-intercept, $\Delta H°_{(a)}$, of the first reaction is greater than the y-intercept of the second reaction, $\Delta H°_{(b)}$. Since the lines cross, the gradient of the second line, $\Delta S°_{(b)}$, is greater than that of the first.

At the point that the lines cross the standard free energy changes for two reactions are equal. This means that a closed system containing the metals A and B will be at equilibrium. This can be shown by considering the reaction below, obtained by subtracting reaction (a) from reaction (b):

$B + 2AO = 2A + BO_2$

At $T = T_E$, ΔG for this reaction is zero, and no reaction occurs. However above this temperature A is reduced by B, and below it B is reduced by A.

Drawing a line from the origin through the point at which the lines cross and extending it to the nomographic scale for oxygen pressure gives the partial pressure of oxygen at equilibrium.

Consider the reaction

$3C + 2Fe_2O_3 = 3CO_2 + 4Fe$

ΔG is only negative for this reaction above T=1020K. Hence, steel furnaces operate above 1020K.

This page titled 25.10: Reducing Agents is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.11: Other Atmospheres and Non-Standard States

An Ellingham diagram may be drawn for other systems where the gaseous phase is not oxygen but another gas such as chlorine. This is useful for industrial processes where the environment is a mixture of gases.

In calculating the various thermodynamic quantities above we have assumed that the condensed metal phase is pure.

In many cases, however, the metal is in the form of an alloy – perhaps in solid solution with other metals. In this case the metal is not in its standard state, and its ability to participate in a reaction – its activity – is reduced. This affects the free energy change for the reaction (which is no longer the *standard* free energy change) and means that the pressure of reacting gas required for equilibrium between the metal in solution and the pure product is changed.

These values can be found by substituting the modified activity of the metal into the expression for the equilibrium constant and ΔG:

$$K = \frac{1}{p_{B(Minstandardstate)}} = \frac{1}{a_M p_{B(Minalloy)}} \tag{25.11.1}$$

$$\Rightarrow p_{B(Minalloy)} = \frac{p_{B(Minstandardstate)}}{a_M} \tag{25.11.2}$$

and

$$\Delta G = \Delta G^\circ - RTlna_M \tag{25.11.3}$$

where a_M is the activity of the metal in the alloy.

Graphically, a decrease in activity has the effect of rotating the Ellingham line for the reaction anti-clockwise around its intersection with T = 0.

This page titled 25.11: Other Atmospheres and Non-Standard States is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.12: Summary

In the preceding pages we have seen that the standard free energy change for a reaction is a very useful quantity to know, in that its value affects the equilibrium constant and gives composition of the system.

We have also seen that the Ellingham diagram is a convenient way of displaying the standard free energy change of many reactions at different temperatures. We have also seen how to read the value of p_{O_2} from the Ellingham diagram.

Finally, we have seen how the Ellingham diagram is used in extraction metallurgy to find the conditions needed for reduction of metal ores.

Going further

Books

- D.R. Gaskell, *"Introduction to the Thermodynamics of Materials"* (Taylor and Francis, 1995) Third ed.
  Provides a good treatment of the Ellingham diagrams pp.347-395.
- E.T. Turkdogan, *"Physical Chemistry of High Temperature Technology"* by (Academic Press, 1980)
  Main data source for the interactive diagram and deals with the physical chemistry of iron and steelmaking in ch.9.

This page titled 25.12: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.13: The Interactive Ellingham Diagram

The interactive Ellingham diagram included within this TLP is a teaching and learning tool, which you can use to obtain a variety of useful thermodynamic information pertaining to a wide range of reactions.

Click on this link to launch the interactive Ellingham diagram in a new window.

Click to on this link to open the diagram's User Guide in a new window.

Please note that the diagram is for educational purposes only, and the accuracy of the data contained within it is not guaranteed.

Using the interactive Ellingham diagram tool

Displaying reactions

The diagram consists of two screens. On the first screen, you can select the reactions you wish to include on the Ellingham diagram. First of all, select the system or systems you want to investigate. This will be dictated by the gaseous phases present in the system. Position the cursor over the name of each type of compound to highlight the elements for which data is available.

Then you can select the metals involved in the reactions. In order to keep the diagram simple you may select up to two elements at one time, along with carbon and hydrogen.

Press the “see Ellingham diagram” button to proceed to the second screen, where the selected reactions are displayed on the Ellingham diagram.

On the second screen, the standard free energy of each reaction, in kJ mol^{-1} of reacting gas, is plotted as a function of temperature in Kelvin. Each selected metal may have several reactions, or none, associated with each gaseous phase.

A change in gradient of any line may be associated with a phase change, either melting or boiling of the metal which is indicated by an m or b at the point, or melting or boiling of the resulting compound, indicated by (m) or (b).

The temperature range over which data is available for each line is shown in purple. Each line is extrapolated to the absolute zero of temperature (thin green line) for simple comparison of the standard enthalpy changes for different reactions.

Obtaining precise reaction data

In order to obtain specific data for a particular reaction, its line on the diagram may be selected by clicking anywhere along its length. The selected line is shown in brown and a crosshair appears at one end. Moving the slider bar along the x-axis of the diagram can move this crosshair to the temperature of interest, and an accurate readout of the standard free energy of the selected reaction is given at the chosen temperature.

Data including the equilibrium constant for the reaction, the equilibrium partial pressure of the reacting gas and the equation of the line in the form A + BT is also given. These values are obtained by simple operations on the standard free energy value - see section on Partial pressure of reacting gas.

A different line may be selected by simply clicking anywhere along its length.

Non-standard reactions

If the metal is not in its standard state, the free energy change for the reaction changes, along with the partial pressure of reacting gas at equilibrium. Entering the activity of the metal in the blue-coloured box and pressing the “Compute” button can find non-standard reaction data at the selected temperature.

Gas mixtures

If the oxidation reactions of carbon or hydrogen are selected, the Ellingham lines for each shown in black or red respectively, the ratio of the pressures of the gases CO and CO_2, or the gases H_2 and H_2O is given as well as the pressure of O_2 gas.

The diagram can be restarted, so that different reactions can be seen, by pressing the “Restart” button.

This page titled 25.13: The Interactive Ellingham Diagram is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

25.14: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

The Ellingham diagram shows values of which thermodynamic quantity as a function of temperature?

a	Standard electrode voltage;
b	Standard free energy change of reaction;
c	Partial pressure of gas;
d	Enthalpy change of reaction.

Answer

B

For a closed system at equilibrium at a temperature T, which of the following statements are true?

a. $\Delta G = 0$;
b. $\Delta G° = 0$;
c. $\Delta H = T\Delta S$;
d. $\Delta S = 0$.

Answer

a. is true; and hence c. is true. At equilibrium, there is no driving force for a reaction to proceed in any direction - no decrease in free energy occurs. Therefore, $\Delta G = \Delta H - T\Delta S = 0$. The enthalpy change of the reaction is balanced by the entropy change of the reaction. $\Delta G°$ is the standard free energy change for the reaction, which is the free energy change when all the reactants and products are in their standard states.

Why are the slopes of many of the lines on the Ellingham diagram almost identical?

a	Most reactions involve the elimination of one mole of gas, so there is a similar standard enthalpy change of reaction.
b	Most reactions involve the elimination of one mole of gas, so there is a similar standard entropy change of reaction.
c	The activity of most of the metals is the same.
d	The partial pressure of the reacting gas is the same for all reactions.

Answer

B. Most of the reactions on the Ellingham diagram involve the elimination of one mole of gas to form a solid product. Since the entropy of a gaseous state is very much greater than a solid state, but the entropy of different gases is similar, most reactions have a similar entropy change.

What thermodynamic quantity does the intercept at T=0 K for any standard free energy vs T line signify?

a	The approximate value of the standard entropy change;
b	The approximate value of the standard enthalpy change;
c	The equilibrium constant for the oxidation reaction;
d	Heat capacity of the oxide.

Answer

B. Since $\Delta G = \Delta H - T\Delta S$, at $T = 0$, $\Delta G = \Delta H$

What is the decomposition temperature (to the nearest 50 K) for Ag_2O? (This question should be completed with the help of the interactive Ellingham diagram included with this TLP).

a	460 K
b	500 K
c	540 K
d	620 K

Hint: On the Ellingham diagram, select the 'oxides' reactions, and then select the elements Ag from the periodic table. The decomposition temperature is the temperature at which the standard free energy of formation of the oxide becomes positive. ie. $\Delta G° = 0$.

Answer

A

What is the decomposition temperature (to the nearest 50 K) for PdO? (This question should be completed with the help of the interactive Ellingham diagram included with this TLP).

a	1050 K
b	1100 K
c	1150 K
d	1200 K

Hint: On the Ellingham diagram, select the 'oxides' reactions, and then select the elements Pd from the periodic table. The decomposition temperature is the temperature at which the standard free energy of formation of the oxide becomes positive. ie. $\Delta G° = 0$.

Answer

C

Which of the following elements can be used to produce Cr from Cr_2O_3 at 1200K? (This question should be completed with the help of the interactive Ellingham diagram included with this TLP).

a. Mg
b. Fe
c. Co
d. Al

Hint: Display the oxide reactions for chromium along with each of the other elements on the Ellingham diagram. What does the relative stability of the oxides of each element and Cr_2O_3 (shown by their relative position on the diagram) tell you?

Answer

a and d. The standard free energy change for the formation of MgO and Al_2O_3 is more negative than for Cr_2O_3. This means that either of these compounds will reduce Cr_2O_3 to Cr.

Explain, using the Ellingham diagram, how a mixture of Cl_2/O_2 gas may be used to separate Zn from Fe in a galvanised scrap. (This question should be completed with the help of the interactive Ellingham diagram included with this TLP).

Answer

On the Ellingham diagram, the free energy of formation for iron oxide is always lower than the free energy of formation of iron chloride. Hence, iron will always react to form its oxide in a Cl_2/O_2 gas mixture. Conversely, the free energy of formation of zinc oxide is always greater than the free energy of formation of zinc chloride, so zinc will always form its chloride in a Cl_2/O_2 gas mixture.

In addition, any iron chloride will be reduced by zinc metal, and any zinc oxide will be reduced by iron.

This page titled 25.14: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

26: Expitaxial Growth

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

This TLP is designed to help you learn about epitaxial growth; the growth of a (usually thin) single crystalline layer in the same crystallographic orientation as its single crystal substrate. Epitaxial growth is widely used in the electronics industry to enable the deposition of precisely controlled thin layers of semiconductors or oxides for use in devices such as thin film transistors, diodes and lasers.

One of the major film deposition techniques, which is implied in this TLP, is molecular beam epitaxy (MBE). In this technique low energy gas phase beams of atoms or molecules are directed at the crystalline substrate, usually in a vacuum. The resultant thin films are required to be single crystalline and may, or may not, be atomically flat. The exact form of the thin film (flatness, composition, strain, band gap ...) depends sensitively on the growth parameters, of which the most important is temperature.

The aim of this TLP is to allow you to explore the significance of key parameters in the process, such as the substrate temperature and the various bond energies between the atoms involved. At the heart of the package is a two-dimensional simulation of the deposition of the atoms. Using this you can explore the effect of these parameters on the growth of a generic crystal which models many features of "real" systems involving semiconductors such as Si, GaAs and AlGaAs.

The perfection of the thin film depends crucially on the mode of growth of the deposit. Three such modes are usually identified:

- Frank-van der Merwe (one perfect monolayer at a time);
- Volmer-Weber (island growth), and a combination of the two called
- Stranski-Krastanov (one or more perfect layers followed by island growth).

You are encouraged to explore the simulation until you can identify the physical reasons why each of these occurs, and thus the conditions most likely to lead to the preferred growth mode and thin film. Your knowledge will be applicable to other epitaxial growth techniques such as MOCVD and VPE, and not simply MBE.

The questions associated with this TLP should help you to decide whether you have understood the key messages.

Before you start

You can view the epitaxy simulation simply as a rather pretty animation. However in order to fully appreciate what is going on you need to be familiar with the concepts of interatomic bonding, the migration of a single atom across the surface of a crystal (surface diffusion) and the thermal activation of processes (Arrhenius behaviour).

You do not need to know anything specifically about semiconductors or opto-electronics.

26.1: Introduction

26.2: The Simulation and its Limitations

26.3: Summary

26.4: Questions

This page titled 26: Expitaxial Growth is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

26.1: Introduction

Epitaxy by MBE involves expensive vacuum deposition equipment such as that shown in the figure.



Figure 1 : Epitaxial growth equipment at the University of Liverpool

It is used to deposit thin layers (usually less than a micrometre thick) intended to form the active layers in optoelectronic devices. Such layers must be flat, of precise composition with the appropriate concentration of dopant atoms, and may involve abrupt changes of composition, for instance in order to form quantum wells. The details of the arrangement of the deposited atoms is therefore of great importance: Not only must they adopt the same crystal structure and orientation as their substrate, but the surface of the growing crystal must either be flat or , if not , then rough in a controlled and predictable way.

When the deposited material is of identical composition to the substrate we refer to homoepitaxy, while if the deposit is different from the substrate we refer to heteroepitaxy.

The images in figures 2 and 3 show high resolution TEM images of typical layers, in which the columns of atoms can be seen.

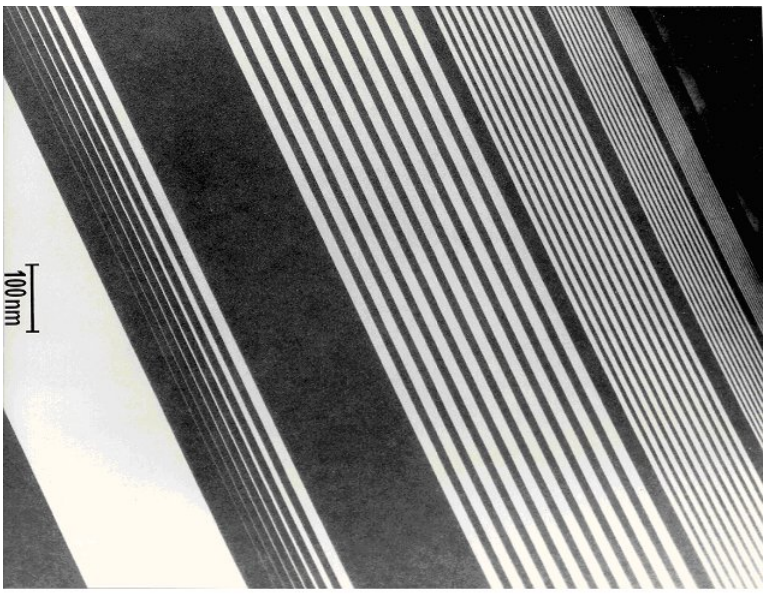


Figure 2 . TEM micrograph of quantum wells in AlGaAs/GaAs imaged so that there is strong contrast between the layers of different composition.

The following image (Figure 3) shows the columns of atoms in alternating epitaxial layers of GaAs and AlAs. The interfaces can be seen to be flat to about one monolayer. The appearance of the AlAs is different from that of the GaAs despite their identical crystal structure because different planes are imaged in the two phases.

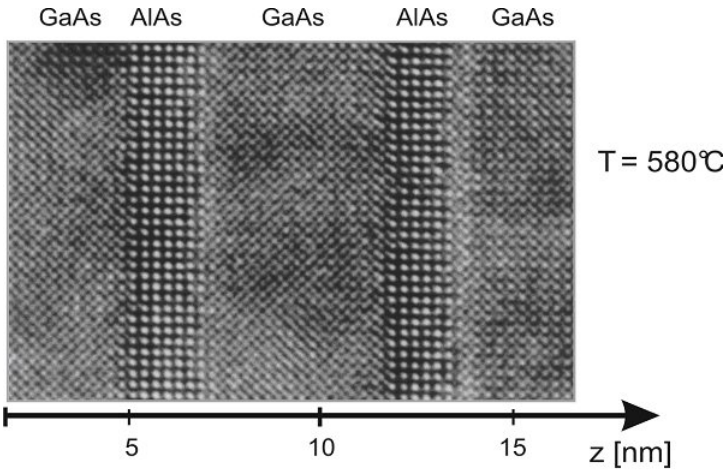


Figure 3 : A high resolution TEM image of alternating GaAs and AlGaAs layers (Simone Montanari PhD thesis (2005))

Figure 4 illustrates the epitaxial growth of a complex compound. The substrate plane is (001) and the heavy atoms show as white dots in a square array. The interface is clearly very flat.

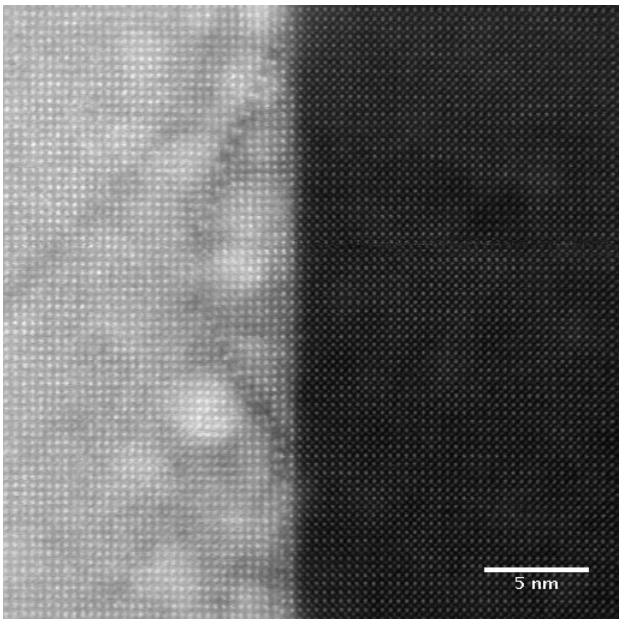


Figure 4: $Bi_{0.5}Mn_{0.5}FeO_3$ film (left hand side, lighter contrast), epitaxially grown on a strontium titanate ($SrTiO_3$) single crystal substrate.

This page titled 26.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

26.2: The Simulation and its Limitations

The most obvious limitation of the simulation is its restriction to two dimensions. Surface diffusion on the screen thus only takes place along a line, and atoms are rather more likely to meet each other than in three-dimensional reality. A second less-obvious consequence is that the crystallography of real semiconductors is not reproduced on the screen (silicon and GaAs adopt the "diamond cubic" structure whereas our simulation implies a close-packed fcc structure). Thirdly we have not allowed you to change the rate of production of the deposited atoms. In an MBE reactor this would be controlled via the temperature of the (solid or liquid) sources of gas phase atoms. The simulation must be regarded as a "model" which demonstrates many (but certainly not all) of the behaviours of the real crystals. Another simplification is that the possible existence of impurity or dopant atoms is ignored. This is obviously of huge importance in the device industry, but makes very little difference to the behaviour simulated in this TLP.

However despite all these limitations the model is based soundly on the physics of atomic interactions, and incorporates key aspects of the movement of atoms across surfaces ("surface diffusion") including thermal activation. Atoms will therefore show a greater probability of moving from their original position, across the surface, at higher temperatures. It also takes into account the possibility that the relaxed lattice parameter of the deposited crystal might not be the same as that of the substrate crystal. This often occurs in heteroepitaxy, leading to the build-up of strain in the growing layer. The simulation correctly models many of the behaviours observed during the epitaxial growth of real crystals.

The three common growth modes have the following characteristics:

- Frank-van der Merwe: each layer of deposited atoms is completed before the next layer starts to form. The surface at any instant will be flat or will contain a few monolayer steps.
- Volmer-Weber (island growth): each deposited atom attaches to an island (or incipient particle); islands grow appreciably before joining up to cover the substrate completely. The surface at any instant will not be flat, and the substrate may not be entirely covered.
- Stranski-Krastanov: initially the deposited atoms form one or more perfect layers but this is followed by island growth. The surface is therefore initially flat but develops to become less flat.

Click on this link to **launch the epitaxial growth simulation in a new window.**

Please note that the simulation is for educational purposes only, and the accuracy of the data contained within it is not guaranteed.

Note: This animation requires Adobe Flash Player 9 and later, which can be downloaded here.

An effective way of using this TLP is to experiment for a few minutes with the simulation, in order to familiarise yourself with its operation, before attempting to answer the following questions, many of which will require that you experiment with the effect of specific variables on the simulated growth, in order to explore different growth regimes. The simulation offers you by default 100 atoms to deposit (three to four atomic layers), but if you need more to see how the film develops, simply increase the number using the slider control.

Notes about the simulation parameters:

- Temperature is shown as Tm, the temperature of the substrate expressed as a fraction of its melting temperature.
- Bond strengths are in arbitrary units on a scale from 0 to 100, where zero implies no bonding and 100 is a strong bond which is unlikely to be broken except at high temperatures.
- Lattice parameter difference is the percentage difference between the natural lattice parameters of the substrate material and the deposit material. If this is non-zero then for epitaxy to occur the deposit must be strained to fit the substrate and therefore strain energy will accumulate as growth occurs.

This page titled 26.2: The Simulation and its Limitations is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

26.3: Summary

After working through this TLP you should understand:

- the importance of growing thin films which are flat at the atomic scale;
- the reasons why three different types of growth can occur during epitaxial deposition;
- the factors which determine the growth mode and hence the flatness of the resultant film;
- the importance of strain in the growth of heteroepitaxial films.

If you read some of the recommended further material (below) you will also appreciate the range of applications for epitaxial films and the existence of alternative growth techniques.

Going further

Most of the literature relating to epitaxy has arisen in the context of semiconducting or oxide systems. There are currently good articles on Wikipedia which can be found if you search for the following key words:

More on the material covered in this TLP:

- Stranski-Krastanov growth
- Epitaxy
- Molecular beam epitaxy (*don't search for* MBE)

For some ideas on why we grow epitaxial thin films:

- Quantum well
- Laser diode

And for some background on other approaches to growth, search for:

- Thin film solar cell

This page titled 26.3: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

26.4: Questions

Click on this link to **launch the epitaxial growth simulation in a new window.** You may need it to answer some of the questions,

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

At a medium temperature, say 0.5 Tm, which of these conditions favours layer (Frank-van der Merwe) growth?

		Adatom-adatom bond strength	Adatom-substrate bond strength	Lattice parameter difference
a		50	50	0
b		20	10	0
c		30	40	0
d		20	50	0.6%

Answer

B

At a low temperature, say 0.3 Tm, which of these conditions favours island (Volmer-Weber) growth?

		Adatom-adatom bond strength	Adatom-substrate bond strength	Lattice parameter difference
a		50	10	1%
b		50	50	0.3%
c		20	60	0
d		30	30	0.1%

Answer

A

At a high temperature, say 0.8 Tm, which of these conditions favours Stranski-Krastanov growth?

	Adatom-adatom bond strength	Adatom-substrate bond strength	Lattice parameter difference
a	0	40	0
b	70	30	0
c	80	20	0.1%
d	50	50	0.6%

Answer

D

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

At 0.7 Tm, what is the effect of lattice parameter difference in the following cases? Increase the number of atoms from 100 to 150 to see this effect.

	Adatom-adatom bond strength	Adatom-substrate bond strength	Lattice parameter difference

Answer

50	50	0

What happens at 0.5 Tm in the following conditions? Is this physically reasonable?

	Adatom-adatom bond strength	Adatom-substrate bond strength	Lattice parameter difference
	100	0	0

Answer

Particles, some detached from substrate. Yes, they are nanoparticles which might indeed float free for a while (although eventually subject to gravity)

What happens as more and more atoms are added in the conditions of Q7? What drives this process?

Answer

The particles coarsen, driven by surface energy which is lower per unit of material (i.e. number of atoms) for larger particles.

Run the simulation with parameters which give you either pronounced SK growth or clear island formation. What shape are the resultant particles or islands?

Answer

Faceted but tending to be equi-axed

What would you expect to happen to the growing layer if the lattice parameter difference was 5%?

Answer

It would start as island growth, but the islands would not be in perfect alignment as they eventually met and defects (dislocations or twins) would have to form.

If the substrate had been polished so that its surface was not exactly a low-index plane (e.g. perhaps 0.5 degrees away from exact {111}, would there be any occasions during growth when there were no steps on the growing surface?

Answer

No, steps would move constantly across the surface, maintaining approximately the same spacing the whole time.

Open-ended questions

The following questions are not provided with answers, but intended to provide food for thought and points for further discussion with other students and teachers.

Try to explain the results you saw in the sixteen runs of Questions 1 to 4. Predict a set of parameters which should give Stranski-Krastanov growth at 0.2Tm. e.g. [50, 50, 1]

Predict and confirm a set of conditions which might give a gas rather than a deposit [0,0,0] or [0,0,1] at any temperature, because no bonding of adatoms either to substrate or each other]

This page titled 26.4: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

27: Examination of a Manufactured Article

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

On completion of this TLP you should:

- be able to identify the materials used in common objects using suitable techniques;
- be able to identify the processes used to produce a particular shape or microstructure within components of an article;
- be able to understand why materials and processes might be chosen to produce a given article.

Before you start

This TLP brings together considerations from a range of other topics, and so many other TLPs might be helpful in understanding the properties and techniques referred to. There will be links to the relevant TLPs at appropriate points.

27.1: Introduction
27.2: Dismantling the Article
27.3: Classes of Materials
27.4: Metals
27.5: Polymers
27.6: Ceramics and Composites
27.7: Example Article
27.8: Summary
27.9: Questions

This page titled 27: Examination of a Manufactured Article is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.1: Introduction

Most manufactured items contain a large number of individual components, often using a surprisingly wide variety of materials. Examining a manufactured item (article) can help to understand what materials are commonly used in manufacturing and why.

It is important to keep in mind economic factors as well as the properties of materials and the specific purpose of each component. This means it can be quite a complex procedure to identify, with confidence, the materials within an article. It is usually done by considering a range of all appropriate factors and making use of some of a broad range of identification techniques.

This TLP suggests some factors that may be considered in attempting to identify the materials used in various articles. It will also give some examples of where these factors may be important.

The other DoITPoMS TLPs provide a wide variety of information about the functional and mechanical behaviour of materials, as well as a range of information on techniques for examining materials. This TLP will not aim to describe all of these, but will link to them.

This page titled 27.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.2: Dismantling the Article

The first step to determining the materials and reasons for their use in an article is to take it apart. This is an important stage because it is a chance to see where components fit, and what their purposes might be.

An exploded diagram of components and where they fit in relation to the rest of the article is often extremely helpful. Once the article has been dismantled it may not be possible to reassemble it, and knowing the position of a component in the article is vital to understanding what its function might be. The exploded diagram helps to sort out what the purpose of components are, and therefore what might be required of their properties.

The diagram should be clear and well labelled, including scale bars.

Important properties to consider for each component could include:

- **Mechanical properties:** such as strength or toughness
- **Electrical properties**: conducting or insulating
- **Aesthetic properties**: is the component on display, and would the appearance be important?
- **Corrosion resistance**: the component will probably need to last for at least the lifetime of the product
- **Density of component**: it may be important that the article is lightweight
- **Specific functions**: some materials are chosen because they fulfil a specific function within the article (for example piezoelectric or ferromagnetic materials)

It is also important to consider economic factors, such as the costs of the raw materials and processing. The cost of a material can influence choice as much as its physical properties; for example diamond is an extremely hard material, but it is not widely used due to cost.

Each component will probably be reasonably easily identifiable as belonging to one of four classes of materials:

- Metals
- Polymers
- Ceramics
- Composites

This page titled 27.2: Dismantling the Article is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.3: Classes of Materials

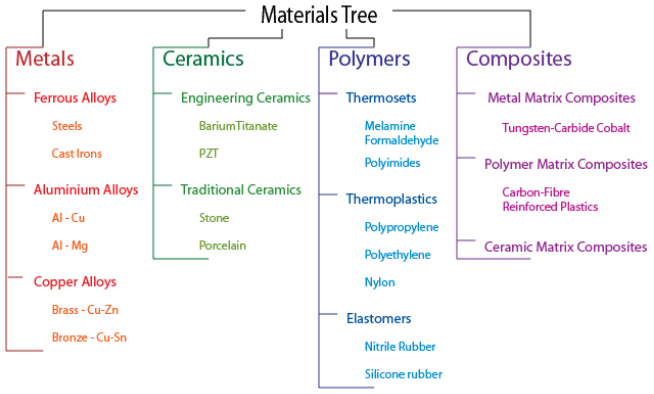
It can often be quite straightforward to tell materials classes apart by look or feel. Metals are usually more reflective or 'metallic' looking, ceramics are commonly matte and polymers may be shiny or matte, but are typically less dense than either metals or ceramics. Composites may be harder to immediately identify, but the surface may appear non-uniform and/or sectioning the sample may reveal fibres or particles.

It is useful to note that taking a cross-section can often be helpful in identifying materials used in components as the internal material and/or microstructure may differ from that at the edge.

Some materials may not be quite so easily identified simply from appearance and texture, but considering the factors mentioned can help to narrow down the possibilities. The table below gives a rough guideline to the kinds of properties you would expect from each class, and a few examples:

Class	Common Properties	Examples
Metal	Hard, ductile and conduct heat and electricity	Copper (wires), stainless steel (cutlery)
Polymer	Widely variable, often soft and flexible	Polystyrene (cups), polycarbonate (CDs), polyethylene (plastic bags)
Ceramic	Hard, brittle, resistant to corrosion, electrically non-conductive	Concrete (buildings), PZT (piezoelectric used in lighters and ultrasonic transducers), porcelain (vases, teacups)

The tree-diagram below shows an overview of a variety of materials that might be encountered:



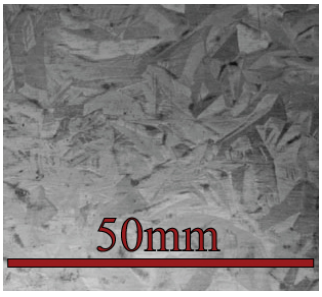
Coatings

Many components may have some kind of coating; a covering of another material designed to improve the surface qualities of the item. The improvement could be for many reasons including: corrosion resistance, appearance, adhesion, wear resistance and scratch resistance.

Different kinds of coating will have different processing methods. It is often possible to deduce the method from the composition (of the coating and the bulk of the component) and the shape of the component.

Common coating methods include:

Hot dip coating – a method used for coating metals (commonly ferrous alloys) with a low melting point alloy. The component is dipped in a bath of the molten coating alloy. For example zinc is often hot dipped onto steel (called ‘galvanising’). This also offers sacrificial corrosion protection and gives a distinctive ‘spangled’ appearance (which can be prevented by including particles to encourage nucleation in the electrolyte).



Steel coated with zinc

Electroplating – reduction of cations in an electrolytic solution onto conducting components. For example silver plated cutlery.

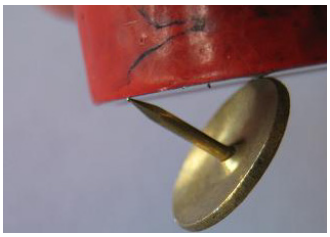
Anodising – commonly used for aluminium components, an electrochemical cell is set up which drives the oxidation of the metal, increasing the thickness of the protective oxide layer.

Vacuum deposition – also known as PVD – physical vapour deposition. For example, 'evaporation' involves the heating of the coating metal in a vacuum, so that it evaporates and is deposited onto the surface of the component that is positioned above. This process is used to make mirrors, depositing a thin layer of metal, usually aluminium.

Thermal Spraying - powder particles are fed into a high temperature torch (combustion or plasma), where they melt and are accelerated against the substrate. It is mainly used to produce relatively thick ceramic and metallic layers.

Enamelling - a powder is distributed on a surface, which is then heated so that the powder melts and bonds to the substrate. The resultant layer is usually glassy. Originally developed in ancient Egypt, and extensively used for jewelry, it is also widely employed for cooking utensils and various domestic items, especially those subjected to high temperature.

One example of a coated metal is shown below; a drawing pin, which appears to be brass, was found to be magnetic and so the surface was abraded, revealing a grey metal within – steel. Steel, which has very good mechanical properties, is covered with brass for aesthetic reasons and also protects the surface from corrosion.



Magnetic drawing pin

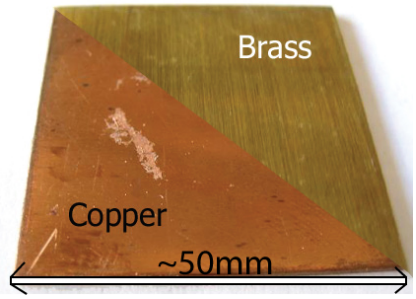


Abraded pin revealing steel

This page titled 27.3: Classes of Materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.4: Metals

Metals are extremely widely used in manufacturing, often for mechanical or electrical properties. They are often easily shaped and have good mechanical properties so they may be used in ‘structural’ elements of an article, and they also have good conduction (electrical and thermal) and so may also be used, for example, in electrical wiring. Familiar metals can often be reasonably well identified by eye (see examples below), but there are more complex methods of metal identification available too.



Techniques for the Identification of Metals

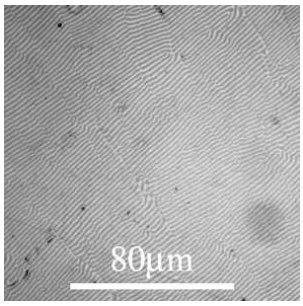
One very easy test for a metal is to see if the component is magnetic, this narrows down the possible materials to those that are ferromagnetic (most commonly iron or nickel).

Simple corrosion tests involving immersing a scratched sample (to remove any coating) in water (or some other electrolyte) can be helpful. Leaving a sample in water overnight might reveal rusting. For example a scratched zinc coated steel sample would not rust due to the zinc offering sacrificial protection. However, a scratched tin coated steel sample would rust, because the tin is supposed to act as a barrier between the steel and air. This is especially useful for ferrous alloys as corrosion resistance is very often a concern for these (and they are very common).

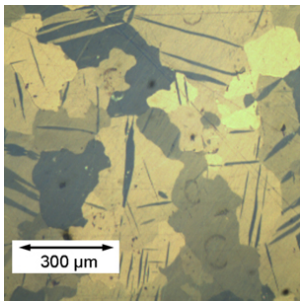
Optical Microscopy

This involves looking at mounted, polished and etched samples under a light microscope. It reveals the microstructure of the sample; this can give information on both the composition and processing of the component.

See Optical Microscopy and Specimen Preparation TLP for more information on how to go about this.



Al-Cu Eutectic composition
This is an Al-Cu alloy showing a very clear eutectic lamellar microstructure.
(See micrograph library entry for more information)



Cold rolled zinc showing deformation twins
This is zinc, it has been cold rolled as can be seen from lenticular deformation twins

See the DoITPoMS Micrograph Library for further examples.

The benefits of this method are that optical micrographs can reveal a large amount of information about a metallographic sample, and it is possible to find known examples (see above links) to compare your work to. After an initial examination by eye and consideration of properties, optical microscopy is an important step in the characterisation of metals. It can reveal many things that the initial examination does not.

Scanning Electron Microscopy (SEM)

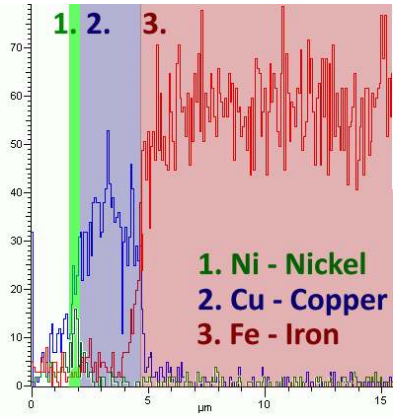
Scanning Electron Microscopy uses a focussed beam of high-energy electrons to form images of samples. Electron Microscopy is not limited by the wavelength of light, so very closely spaced features can be resolved, so this method gives very clear high magnification images when set up correctly. It also gives a large depth of field , so rough surfaces can still be in focus.

The SEM can give very good high magnification images, again revealing more than optical microscopy could. One limitation is that the sample must be electrically conducting; the mounting polymer and the sample must both conduct electricity.

Energy Dispersive X-ray Spectroscopy (EDS)

This is a technique often used in conjunction with the SEM, with an electron beam of ~20 keV. The beam strikes the sample resulting in X-rays being emitted; the X-rays are collected and the intensities and energies examined. The results can determine the atomic composition of the sample at the point of beam-sample interaction.

Examples: It can reveal the composition of a very thin coating; for example by taking a linescan. The diagram below shows results for a linescan taken across a coating, showing a layer of copper and a finer layer of nickel on an iron alloy (all other elements were ignored). This may not be easy to tell from optical microscopy alone.

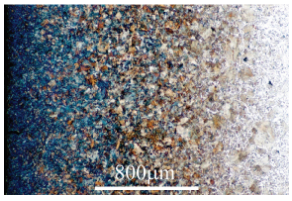


EDS can be a very helpful method of characterisation, but it is not always absolutely reliable. Elements of low atomic number (less than about 11, i.e. below sodium) are difficult to detect by EDS. There is often contamination of elements like carbon from the environment. It is important to use common sense when interpreting the EDS results; most of the time is it unlikely that very heavy elements like uranium are actually present. Results may be very good qualitatively, but care must be taken when trying to obtain and interpret quantitative results. The quantitative analysis results depend upon things like the set-up of the SEM and the geometry of the sample.

Examples of Fabrication Processes

The ease with which metals are shaped leads to a wide range of processing techniques for different end products; any coatings will also be applied by one from a range of processes. It may be possible to deduce the method of processing from the shape and the properties of the metal. The microstructure may also give further clues:

Process	Description	Features
Deformation	Includes a variety of techniques including forging, extruding and drawing, see the TLP Introduction to Deformation Processes for more information	May expect to see directionality or squashed grains in places that have been stressed (see the animation)
Machining	Includes, for example, laser cutting and water-jet cutting as well as more conventional methods like sawing or grinding.	These methods would not result in larger scale microstructural directionality, but may show localised deformation. Can give a very good finish
Casting	The molten metal is set in a cast of some kind, see the Casting TLP for further details on the different kinds of casting.	There is a lot of variety of microstructure from cast products, which ranges from single crystal components to those with clear chill, columnar and equiaxed zones.
Carburisation – a surface heat treatment	A surface treatment in which carbon is diffused into the surface of a steel object above the ferrite-austenite transition temperature (~ 740 °C). This is done by heating the steel in a C-rich atmosphere, (e.g. in CO gas). The result is a hard, high carbon surface several hundreds of microns thick, surrounding a tough, low carbon interior. To improve the hardness, the surface may be quenched, which helps the production of martensite .	See micrograph below



Micrograph #271
An example of carburisation; see the description in the micrograph library for further information.

This page titled 27.4: Metals is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.5: Polymers

Polymers are very widely used in many areas today. They have a range of properties that can often be controlled by additives, blending or copolymerisation. Many structures and chemical compositions are seen in polymers, but they can be separated into three main groups:

Thermoplastics: These are the most widely used polymers due to the ease of processing (especially for injection molding). Thermoplastics can, once they have been set (solidified) for the first time, be re-melted and remoulded (unlike thermosets). Some examples of thermoplastics are: polyethylene, polystyrene and PET.

Thermosets: These differ from thermoplastics in that they do not re-melt after they have been set (or cured). This is due to the long polymer chains forming cross links on curing. One example is melamine formaldehyde, which is used in domestic electrical plugs.

Elastomers: These polymers have a glass transition temperature below room temperature (see The Glass Transition in Polymers TLP). Rubbers are examples of commonly used elastomers (for more information see the Stiffness of Rubber TLP)

Techniques for identifying polymers

Polymer tests

The polymer tests are a simple way to identify polymers, or at least to narrow down the possibilities. Some of the steps rely on recognising smells, which can be difficult, and it is important to remember that some tests (for example transparency) can be unhelpful due to additives like dyes.

The polymer identification chart goes through a series of simple tests, which should be carried out on a small sample of the polymer. Below is an interactive version of the identification chart. It is important to connect the results of the test to the function and cost of the item.

Infra Red (IR)

In an IR spectrometer IR radiation excites covalent bonds, causing them to vibrate at their resonant frequency. This frequency depends on the exact nature of the bonds (e.g. single/double and the atomic masses of the elements involved). The output is a graph of intensities at different wavelengths (and therefore energies) of infrared radiation. This plot shows the transmitted intensities, so at resonant frequencies, where the energy is absorbed, there is a peak. This allows the bonds to be identified and therefore the sample identified.

Here is a collection of IR spectra for some common polymers:

IR spectrometry is often a very quick method of polymer identification (depending on the equipment available). Preparing a sample for IR spectroscopy may be very simple. A small sample of the polymer with any coatings removed should be placed in the IR machine, and analysed. If it is likely that the plastic contains plasticizers and colours, placing the polymer in ether for an hour and then fully drying it may remove them prior to carrying out IR spectroscopy on the sample. (Test this with a small piece of your sample polymer first though, as some polymers are soluble in ether).

Differential Scanning Calorimetry (DSC)

DSC measures specific heat capacity and how it varies with temperature. As a polymer is heated through its glass transition point, it experiences a sudden change in heat capacity, as chain rotation allows it to take up more energy. This means that DSC allows us to identify the glass transition temperature of a polymer. It can also aid the interpretation of the type of a copolymer (e.g. block copolymer, random copolymer, graft copolymer).

See Macrogalleria's page on DSC for an explanation of this technique.

Examples of processes

Polymers are usually processed by moulding methods:

Process	Description	Features
Injection moulding	Polymer granules are melted and forced into a mould. This is extremely widely used to mass produce small, precise polymer components.	It gives a good finish and the injection points where excess material has been cut off are often visible. In transparent polymers a residual stress field may be visible under crossed polars (see the Introduction to Photoelasticity TLP).
Blow moulding	Cylinders of polymer are inserted into a die and hot air is forced in, pushing the polymer out to the walls of the die.	Gives hollow components, such as bottles or containers. It is only used for thermoplastics.

For further examples see plastics processes on Plastipedia.

Additives and Blends:

Polymers very often have some form of additive, even if it is simply to add colour. These may or may not impair the ability to identify the polymer. When identifying any material it is important to think about the cost and properties, but blending or additives can change the properties of a polymer.

One very common example of a polymer commonly found both with and without additives is polyvinylchloride (PVC). This polymer is used in its rigid, un-plasticized form in plastic guttering and water and gas piping, but is also often found with added plasticizers in a variety of applications from clothing to coating electrical wires.

This page titled 27.5: Polymers is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.6: Ceramics and Composites

Ceramics cover a very wide range of materials from structural materials like concrete to technical ceramics like PZT – a piezoelectric. Usually they are defined as solids with a mixture of metallic or semi-metallic and non-metallic elements (often, although not always, oxygen), that are quite hard, non-conducting and corrosion-resistant.

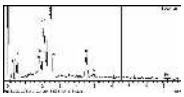
Techniques for identifying ceramics

It is effectively impossible to identify ceramics by eye. Optical microscopy will allows the examination of the microstructure to identify the method of processing, however, it does not allow the identification of different phases.

The most useful technique for finding the composition of a ceramic is energy dispersive x-ray spectroscopy (EDS). Note that for non-conducting ceramics the surface of the sample must be covered with a metallic coating (often gold) to prevent charge build-up.

Here is an example EDS for PZT – a piezoelectric ceramic: $Pb[Zr_xTi_{1-x}]O_3$, this data gives the formula to be: $Pb_{0.7}[Zr_{0.49}Ti_{0.44}]O_3$. For the piezoelectric ceramic we would expect to have $x \sim 0.52$.

EDS data for PZT		
Element	**Weight%**	**Atomic%**
O K	17.26	63.49
Ti K	7.58	9.32
Zr L	16.15	10.42
Pb M	59.01	16.77
Totals	100.00	100.00



Another appropriate method is X-ray diffraction. This allows you to detect the phase or phases present as well as measuring lattice parameter(s) in order to specify precise compositions.

Processing techniques for ceramics

Ceramics are mostly made by powder processing techniques, for example sintering . It may be possible to identify the kind of processing from directionality or porosity in the sample.

Composites

Composites are often used in applications that require specific ‘conflicting’ properties such as a high strength and high toughness. The properties may be conflicting because having a high yield stress sometimes relies on trapping and tangling dislocations, but these reduce the ductility and toughness of the material. Composites often consist of a matrix and fibres or particles that affect the properties (see the TLP on the Mechanics of Fibre-Reinforced Composites).

Usually for composites, once they have been identified as such, it is better to treat each part of the composite as a separate material, and then subsequently look at costs of manufacture and processing.

One important distinction to make is the structure of the two parts that make up the composite – i.e. is it a matrix with long, aligned fibres? Or a matrix with particles? etc

This page titled 27.6: Ceramics and Composites is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.7: Example Article

The best way to understand the concepts in this TLP is to try analysing something. Here is a 'virtual' article, which can be clicked through:

This page titled 27.7: Example Article is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.8: Summary

Many articles can be analysed using the relatively simple techniques described here. This can help with determining the types of material used for different components, their composition, and also processing history. The examination of an article can start to put into use the methods and theory of materials science. Looking at the mechanical, thermal and aesthetic properties of materials can help materials scientists design similar items.

The range of techniques available today is very large, but often a reasonable amount of understanding can be gained from fairly simple techniques and using some common sense. It is essential not to forget the importance of stepping back from results and considering whether or not they are logical; do they fulfil the requirements in terms of mechanical, thermal, aesthetic and economic properties?

Going further

Websites

Material Selection and Processing
Information on selection, design and processing
Identifying Metals
Some simple ways to identify metals
How to Identify Plastics
Plastic Materials Identification Chart
Some simple ways to identify polymers
Introduction to Deformation Processes
This TLP covers the fundamentals of metal forming processes
Differential Scanning Calorimetry
Macrogalleria's explanation of using DSC for sudying thermal transitions of polymers
Plastipedia
"Probably" the Web's largest plastics encyclopedia, including plastics processes
MetPrep Support
Metallography advice, including sample preparation methods and choice of standard etchants

This page titled 27.8: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

27.9: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which of these methods of characterisation would be helpful in identifying a ceramic?

a	Infra red spectroscopy
b	Differential Scanning Calorimetry
c	Energy Dispersive X-ray Spectroscopy
d	Scanning Electron Microscopy

Answer

Correct. Ceramics are very hard to identify by eye, using EDS can give information about what elements are present and the proportions in which they are present. SEM will give an idea of microstructure and may help to identify processing methods, but does not give information about the actual composition. IR spectroscopy and DSC are commonly used in characterising polymers: IR identifies materials by their bonds, DSC by the relationship between heat capacity and temperature.

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

Match the material to the properties, try to think of reasons why the material may have been chosen.

Material 1 has a shiny 'spangled' appearance due to large dendritic grains on the surface, it is a structural component within the article and must withstand relatively high stresses, and may experience wear in service, it must not corrode in warm air and is in a low cost article.

a	Tinned Steel
b	Stainless Steel
c	Aluminium
d	Galvanised Steel

Answer

D. Galvanised Steel: The spangled appearance is due to large zinc grains on the surface, the metal is galvanised to prevent corrosion. Steel is a better structural material than pure aluminium, and it is much cheaper. Stainless steel might also fulfil the requirements of the material, but it is more expensive. Coating the steel with tin provides corrosion protection, but only if the coating is not damaged.

Tinned steel usually has a fairly uniform appearance; it may be shiny, but does not have visible grains on the surface. It is not the best choice due to the stresses and wear the component must undergo. The corrosion protection comes from a layer of tin, if this is damaged the steel may corrode.

Stainless steel offers very good corrosion resistance due to a layer of chromium oxide forming on the surface. Steel is a very good structural material, but stainless steel is reasonably expensive (some of the other choices are more economical).

Unalloyed aluminium is not a good structural material, and it is more expensive than steels.

Match the material to the properties, try to think of reasons why the material may have been chosen.
Material 2 is a brightly coloured, low-density component, it must be tough, rigid and non-toxic, the recycling mark is number 7.

Answer

B. ABS: the recycling mark immediately rules out PE (identified by recycling marks 2-HDPE and 4-LDPE) and PET (identified by recycling mark 1). PTFE is more commonly known as Teflon, and it is usually used as a very low-friction coating. ABS is the only material that satisfies the evidence and requirements: it is a very tough, easily coloured polymer.

The recycling mark for PE is 2 for High Density Polyethylene and 4 for Low Density Polyethylene.

The recycling mark for PET is 1.

PTFE is more commonly known as Teflon, and it is usually used as a very low-friction coating. It is unsuitable for this application.

Match the material to the properties, try to think of reasons why the material may have been chosen.
Material 3 is an electrical component, it forms a contact that is subjected to a reasonable amount of wear, it must not corrode in air.

a	Low carbon steel
b	Copper
c	Brass: Copper/Zinc alloy
d	Stainless steel

Answer

C. Low carbon steel and stainless steel create non-conducting oxides, so without some other coating they are not good choices for electrical contacts. Copper has a very high conductivity, but the addition of zinc greatly increases the wear resistance.

Low carbon steel may corrode in air and forms a non-conducting oxide so without some other coating this is not a good choice for an electrical contact.

Copper has a very high conductivity, but not very good mechanical properties.

Stainless steel forms a non-conducting oxide so without some other coating this is not a good choice for an electrical contact.

====

This page titled 27.9: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

28: Ferroelectric Materials

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

On completion of this TLP you should:

- Be aware of the atomic requirements for ferroelectricity.
- Understand how this atomic structure leads to spontaneous polarisation, and how this polarisation can be switched.
- Understand how this polarisation leads to the interesting electrical properties displayed by ferroelectrics.
- Understand how these properties are made use of in today's technologies.

Before you start

This TLP should be fairly self-contained, but some knowledge of crystal structures is assumed. It may be helpful to read the teaching and learning package on Crystallography and Atomic Scale Structure Of Materials first.

28.1: Introduction
28.2: The Dipole Moment
28.3: Polarisation
28.4: Switching Polarisation
28.5: Measurement of Polarisation
28.6: Fabrication of a KNO_3 Ferroelectric Capacitor
28.7: Temperature Dependence of the Hysteresis Loop
28.8: Barium Titanate
28.9: Order of Phase Transitions
28.10: Ferroelectrics - Why?
28.11: Summary
28.12: Questions

This page titled 28: Ferroelectric Materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

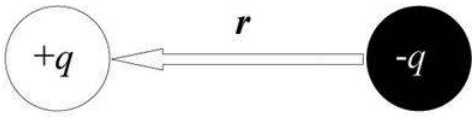
28.1: Introduction

The ferroelectric effect was first observed by Valasek in 1921, in the Rochelle salt. This has molecular formula $KNaC_4H_4O_6{\cdot}4H_2O$. The effect was then not considered for some time, and it wasn't until a few decades ago that they came into great use. Nowadays, ferroelectric materials are used widely, mainly in memory applications. This TLP will show how the ferroelectric effect arises, and how it is usefully used.

This page titled 28.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.2: The Dipole Moment

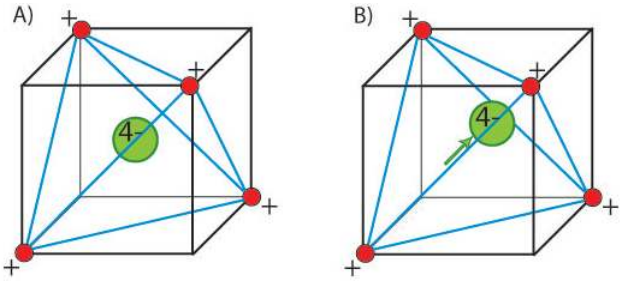
To be ferroelectric, a material must possess a spontaneous dipole moment that can be switched in an applied electric field, i.e. spontaneous switchable polarisation. This is found when two particles of charge q are separated by some distance $\boldsymbol{r}$, i.e.:



The dipole moment, $\boldsymbol{\mu}$ is:

$$\mu = qr \tag{28.2.1}$$

In a ferroelectric material, there is a net permanent dipole moment, which comes from the vector sum of dipole moments in each unit cell, $\Sigma\boldsymbol{\mu}$. This means that it cannot exist in a structure that has a centre of symmetry, as any dipole moment generated in one direction would be forced by symmetry to be zero. Therefore, ferroelectrics must be non-centrosymmetric. This is not the only requirement however. There must also be a spontaneous local dipole moment (which typically leads to a macroscopic polarisation, but not necessarily if there are domains that cancel completely). This means that the central atom must be in a non-equilibrium position. For example, consider an atom in a tetrahedral interstice.



In (A) the structure is said to be non-polar. There is no displacement of the central atom, and no net dipole moment. In (B) however, the central atom is displaced and the structure is polar. There is now an inherent dipole moment in the structure. This results in a polarisation.

This page titled 28.2: The Dipole Moment is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

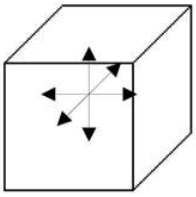
28.3: Polarisation

Polarisation may be defined as the total dipole moment per unit volume, i.e.

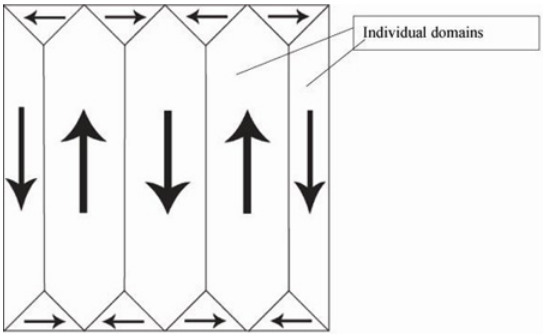
$$P = \frac{\sum \mu}{V} \tag{28.3.1}$$

Materials are polarised along a unique crystallographic direction, in that certain atoms are displaced along this axis, leading to a dipole moment along it. Depending on the crystal system, there may be few or many possible axes. As it is the most common and easy to see, let us examine a tetragonal system that forms when cooled from the high temperature cubic phase, through the Curie temperature, of e.g. T_c=120°C in $BaTiO_3$.

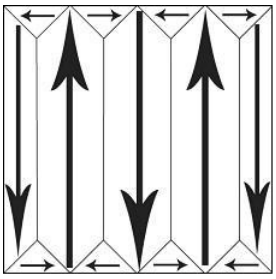
In this system, the dipole moment can lie in 6 possible directions corresponding to the original cubic axes:



In a crystal, it is likely that dipole moments of the unit cells in one region lie along a different one of the six directions to the dipole moments in another region. Each of these regions is called a domain, and a cross section through a crystal can look like this:



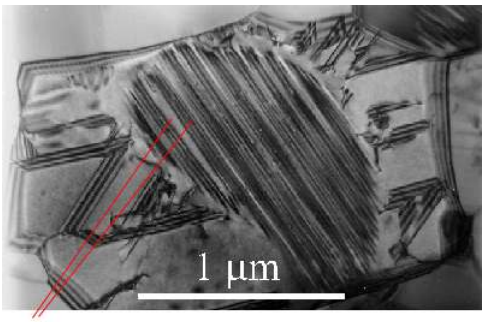
A domain is a homogenous region of a ferroelectric, in which all of the dipole moments in adjacent unit cells have the same orientation. In a newly-grown single crystal, there will be many domains, with individual polarisations such that there is no overall polarisation. They often appear:



The polarisation of individual domains is organised such that +ve heads are held near -ve tails. This leads to a reduction in stray field energy, because there are fewer isolated heads and tails of domains. This is analogous to the strain energy reduction found in dislocation stacking.

Domain boundaries are arranged so that the dipole moments of individual domains meet at either 90°or 180°. In a polycrystal (one with more than one crystallographic grain), the arrangement of domains depends on grain size. If the grains are fine (<< 1 micron), then there is usually found to be one domain per grain. In larger grains there can be more than one domain in each grain.

This is a micrograph showing the domains in a single grain.



Individual domains

This micrograph is reproduced from the DoITPoMS Micrograph Library (entry number: 199).

In this grain, the domains are twinned in such a way as to reduce the overall stray electric field energy. As each domain possesses its own dipole moment, we may switch dipole moments in order to encode information.

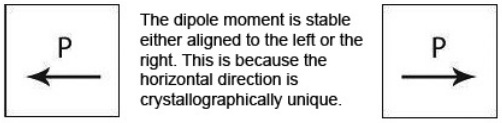
This page titled 28.3: Polarisation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.4: Switching Polarisation

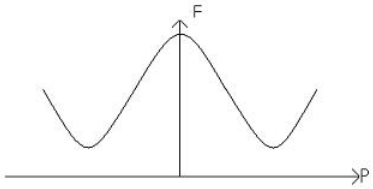
In an electric field, E, a polarised material lowers its (volume-normalized) free energy by $-P.E$, (where P is the polarisation). Any dipole moments which lie parallel to the electric field are lowered in free energy, while moments that lie perpendicular to the field are higher in free energy and moments that lie anti-parallel are even higher in free energy, ($+P.E$).

This introduces a driving force to minimise the free energy, such that all dipole moments align with the electric field.

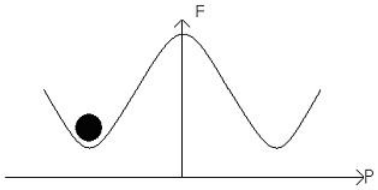
Let us start by considering how dipole moments may align in zero applied field:



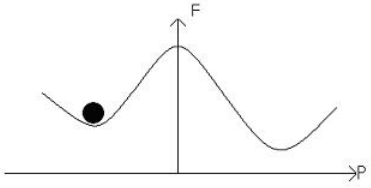
These two moments are stable, because they sit in potential energy wells. The potential barrier between them can be represented on a free energy diagram:



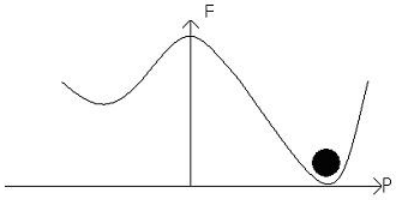
This material is considered to be homogenous. If the polarisation points left then we have:



The electric field alters the free energy profile, resulting in a ‘tilting’ of the potential well:



An increase in the electric field will result in a greater tilt, and lead to the dipole moments switching, leading to:

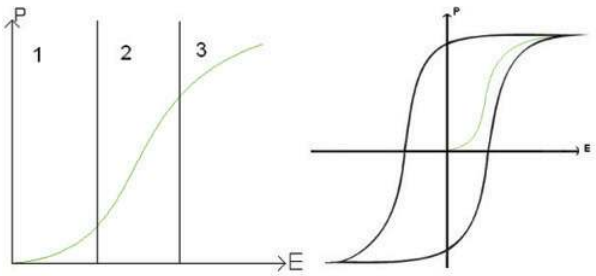


Next we must look at the more realistic scenario in which domains form.

Consider a material which is fully polarised, so that all of the dipole moments are aligned in the same direction. Then apply a reversed electric field over it. New domains with a reversed polarisation nucleate inhomogenously. This requires a certain amount of time, in the same manner as any nucleation process. When the fluctuating nuclei reach a certain critical radius, they grow outwards, forming needle-like structures. When they reach the other side of the ferroelectric, they begin to grow outwards.

This shows the origin of the hysteresis loop. The removal of the field will leave some polarisation behind, and only when the field is reversed does the polarisation start to lessen as new, oppositely poled domains form. They grow quickly however, giving a large change of polarisation for very little electric field. But to form an entirely reversed material, a large switching field is required. This is because of both defects in the crystal structure, in a manner similar to zener drag, and also to do with stray field energy. The polarisation of the material goes from a coupled pattern, with 180° boundaries, to a state in which many heads and tails are separated. This leads to the increase in stray field energy. Therefore, to attain this state, lots of energy has to be put in by a larger field.

Here we show a how a minor hysteresis loop fits into the major loop above.



https://eng.libretexts.org/@go/page/32818

The part of curve shown fits into the major hysteresis curve.

There are three sections to this curve.

1) Reversible domain wall motion.

2) Linear growth of new domains.

3) New domains reaching the limit of their growth.

This page titled 28.4: Switching Polarisation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.5: Measurement of Polarisation

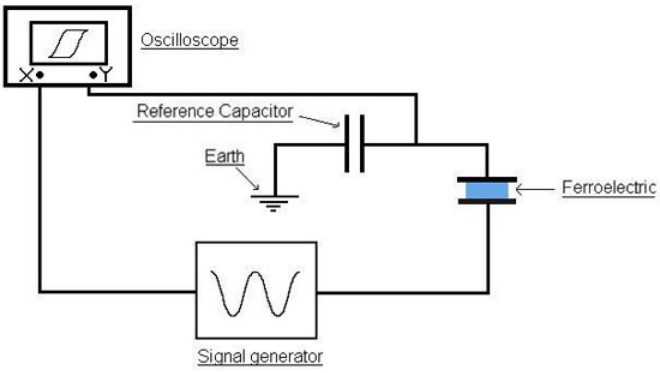
Polarisation may be defined as:

$$P = \frac{Q}{A} \tag{28.5.1}$$

where Q is the charge developed on the plates (Coulombs) and A is the area of the plates (m^2).

A good ferroelectric has 10 μC cm^{-2} < P < 100 μC cm^{-2}.

We can measure the polarisation by using the classic Sawyer-Tower circuit:



In this experiment, the voltage is cycled by the signal generator. Its direction is reversed at high frequency, and the voltage across the reference capacitor is measured. The charge on the capacitor must be the same as the charge over the ferroelectric capacitor, as they are in series. This means the charge on the ferroelectric can be found by:

$$Q = C * V \tag{28.5.2}$$

where C is the capacitance of the reference capacitor, and V is the voltage measured over this capacitor. We can therefore represent the polarisation of a material in an oscillating electric field, by plotting the voltage applied to the material on the x-axis of the oscilloscope, and the surface charge on the y-axis. This can be done because the capacitance of the reference capacitor is much higher than the capacitance of the ferroelectric, so most of the voltage lies over the ferroelectric. It is only possible to measure P by cycling the polarisation through cycling the voltage across the ferroelectric. We cannot measure absolute values instantaneously , but can deduce absolute values from the changes measured when cycling the polarisation.

This page titled 28.5: Measurement of Polarisation is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.6: Fabrication of a KNO_3 Ferroelectric Capacitor

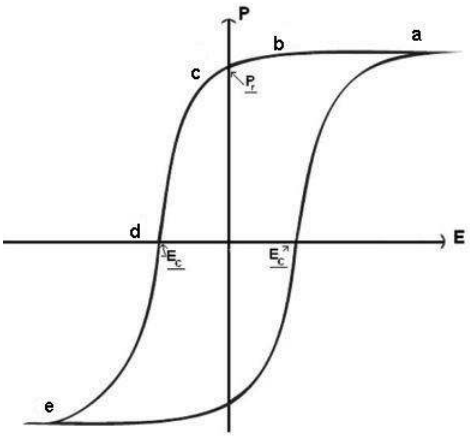
A capacitor can be made from potassium nitrate (KNO_3), which is ferroelectric below 120°C. (The temperature dependence of ferroelectrics will be explained later.) The following video clip shows the construction of a KNO_3 capacitor, and the hysteresis loop it displays. The circuit used is the standard Sawyer-Tower circuit.

https://www.doitpoms.ac.uk/tlplib/fe...kno3-large.mp4

Demonstration to construct a KNO_3 capacitor

The result is a hysteresis loop. This arises from the fact that a system does not respond immediately to a given set of external conditions. Rather, there is a history dependence and this is the basis for memory (two states are possible in E=0).

The final hysteresis loop appears like this:



When the field is removed, the polarisation does not disappear like a dielectric. (a→ c). The polarisation which remains after a material has been fully polarised and then had the field removed is called the remanent polarisation (P_r).
Only after a field is applied in the opposite direction to the original polarising field does the polarisation diminish significantly. There is a specific field which results in zero net polarisation (d). This is called the coercive field (E_C).
Finally, if a sufficiently strong electric field is applied in the reverse direction, the polarisation will reach its maximum value in the opposite direction (e).

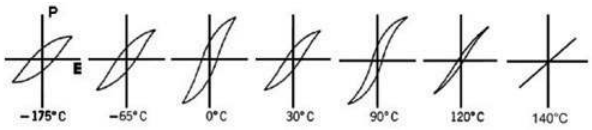
To understand how the polarisation switches we must consider domains more fully.

This page titled 28.6: Fabrication of a KNO_3 Ferroelectric Capacitor is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.7: Temperature Dependence of the Hysteresis Loop

We have now seen the way in which the hysteresis loop arises. However, there are more aspects to the hysteresis loop.

We have only observed it at one particular temperature, one at which the material is ferroelectric. What happens if the temperature is raised? The hysteresis loop changes with temperature, becoming sharper and thinner, and eventually disappearing, i.e.:



As you can see, the polarisation increases at 90°C, as a result of a phase transition. Between this temperature and room temperature, the polarisation increases steadily, as a direct relation with temperature, such that:

$\Delta P = p\,\Delta T$

where p = pyroelectric coefficient ($C\,m^{-2}\,T^{-1}$).

Click here to find out more about pyroelectrics.

Why should this be?

This is a general behaviour (that does not apply to KNO_3) that can arise for two reasons depending on the material.

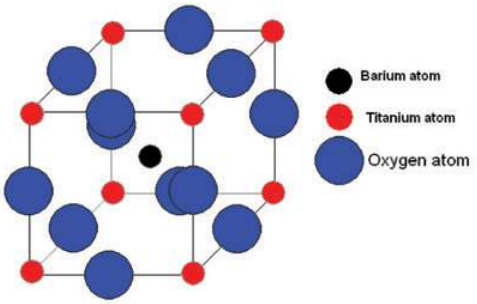
i. Disorder. Each unit cell has its own dipole moment, which, when there is a net polarisation, are described as ordered. At high T, the direction of the dipole moments randomises, giving a disordered material with no net polarisation.
ii. Phase transitions that can open up new possibilities for dipole moments to form. In this case, there is a jump at 0°C, and at 90°C, where the loop becomes taller.

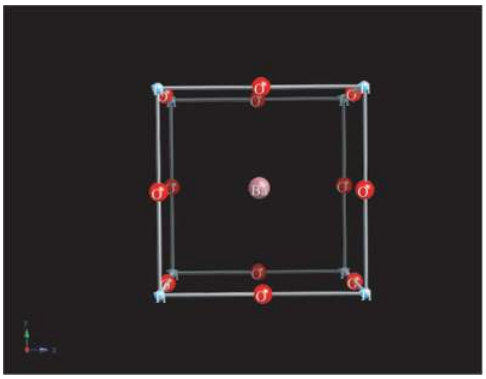
This page titled 28.7: Temperature Dependence of the Hysteresis Loop is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.8: Barium Titanate

Let us consider one of the most well-known ferroelectrics, barium titanate, ($BaTiO_3$).

It has this perovskite structure:

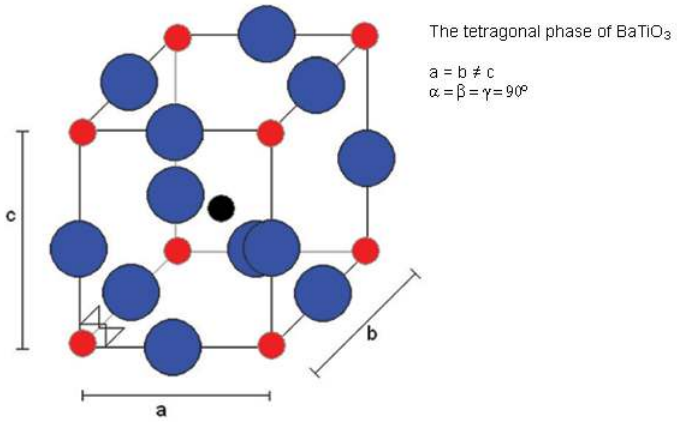




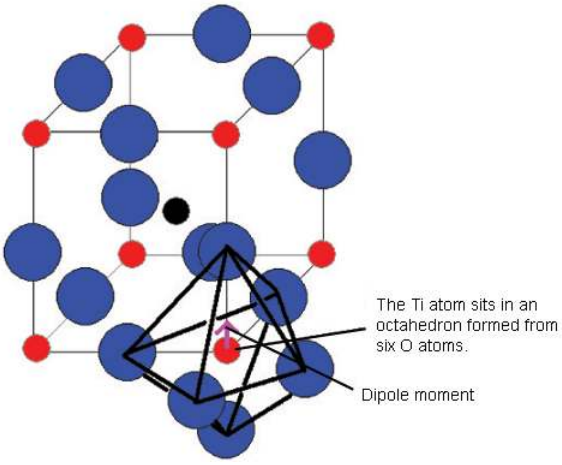
The temperature at which the spontaneous polarisation disappears is called the Curie temperature, T_C.

Above 120°C, barium titanate has a cubic structure. This means it is centro-symmetric and possesses no spontaneous dipole. With no dipole the material behaves like a simple dielectric, giving a linear polarisation. T_C for barium titanate is 120°C.

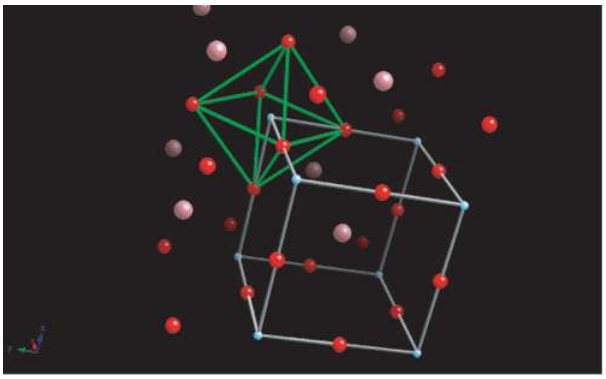
Below 120°C, it changes to a tetragonal phase, with an accompanying movement of the atoms. The movement of Ti atoms inside the O_6 octahedra may be considered to be significantly responsible for the dipole moment:



Cooling through 120°C causes the cubic phase of barium titanate to transform to a tetragonal phase with the lengthening of the c lattice parameter (and a corresponding reduction in a and b). The dipole moment may be considered to arise primarily due to the movement of Ti atoms with respect to the O atoms in the same plane, but the movement of the other O atoms (i.e. those O atoms above and below Ti atoms) and the Ba atoms is also relevant.

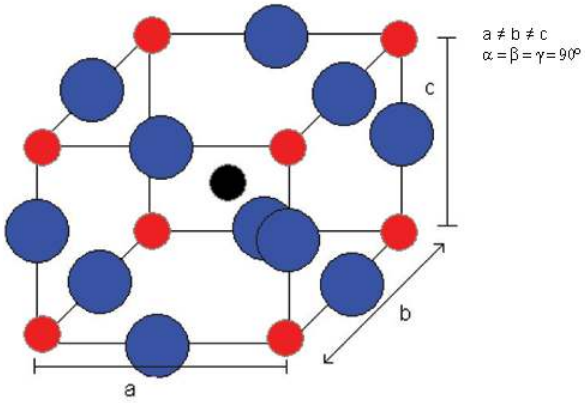


This shows the $BaTiO_3$ structure with an O_6 octahedron surrounding the important Ti atom.

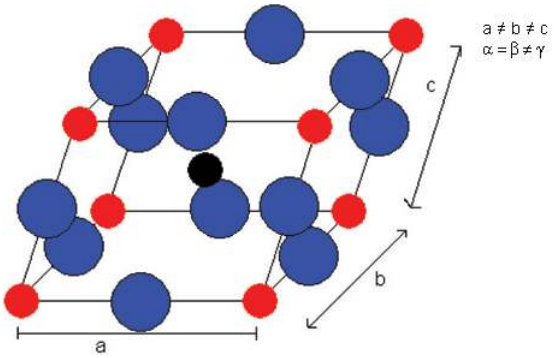


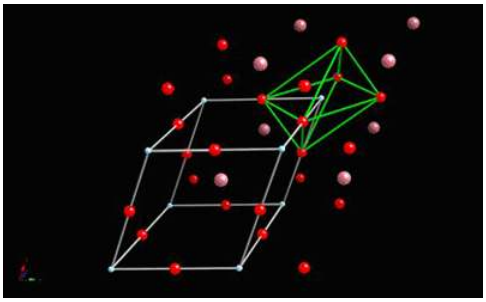
The switching to a cubic structure is the reason for the polarisation spontaneously disappearing above 120°C. Barium titanate has two other phase transitions on cooling further, each of which enhances the dipole moment:

The phase which is reached after cooling to ~ 0°C from tetragonal is orthorhombic.



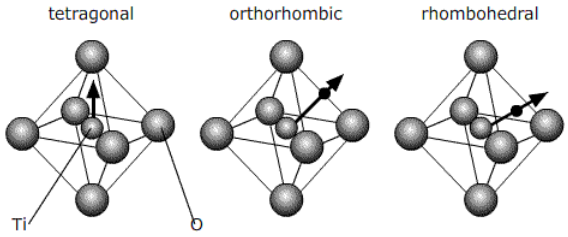
And then rhombohedral below -90°C:





https://eng.libretexts.org/@go/page/32822

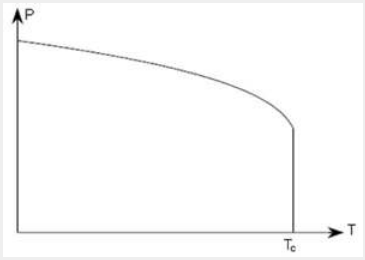
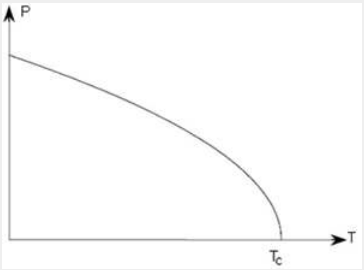
All of these ferroelectric phases have a spontaneous polarisation based to a significant extent on movement of the Ti atom in the O_6 octahedra in the following way (using pseudo-cubic notation):



This page titled 28.8: Barium Titanate is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.9: Order of Phase Transitions

Two common types of phase transition may be identified. These are named depending on how the order changes during the transition. In ferroelectrics this is the polarisation, which is called the order parameter. A first order transition is one which has a discontinuity in the order parameter itself, while a second order transition is one which has a discontinuity in the first derivative of the order parameter. Plots of the spontaneous polarisation vs. temperature appear as follows.

First order	Second order
	

In a first order transition the polarisation varies continuously, until the Curie temperature at which there is a discontinuity. In a second order transition, the order parameter itself is a continuous function of temperature, but there is a discontinuity in its first derivative at T_C.

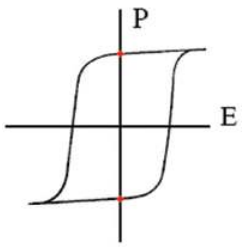
This page titled 28.9: Order of Phase Transitions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.10: Ferroelectrics - Why?

We have now seen how ferroelectrics behave and why they behave in this way. The question to ask now is ‘So what?’. The answer is simple. Ferroelectrics are very useful for devices and are used in many different ways today. If a ferroelectric is used in its linear region, above T_C it makes a very good capacitor, as its dielectric constant can be very high indeed. These are often used in cameras as a way of powering a flash. A battery slowly stores a charge on a capacitor, which when connected to a bulb, releases a burst of high current, creating the flash.

First, the battery charges the ferroelectric capacitor and then, once fully charged, the ferroelectric is connected to the bulb and causes it to flash.

Ferroelectrics are also being considered as a form of non-volatile computer memory, FeRAM (Ferroelectric Random Access Memory). Here, a ferroelectric can be in one polarised state or another, which can act as a 0 or 1, allowing use as binary.



The two states may be distinguished by cycling the charge. They store non-volatile data, i.e. data even when the power is switched off. However, this is not as fast as current DRAM structures, so it is still under development, but it is actually used in e.g. the Sony Playstation 2, because a range of considerations permit it to fulfil this role.

This page titled 28.10: Ferroelectrics - Why? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.11: Summary

Ferroelectrics have now been in use for decades, but they are still an expanding field. Their use in computing will only increase as they are miniaturised. However, to do this, the way in which their properties vary on the microscale has to be understood, so this will be a target for future research. Ferroelectrics will be used for a long time to come, as their properties are unique.

Going further

Books

- *Ferroelectrics: An Introduction to the Physical Principles* by J.C. Burfoot, D. Van Nostrand Company Ltd., 1967

Websites

- You may wish to look at the related TLP on Piezoelectric Materials.

This page titled 28.11: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

28.12: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

What symmetry element must be absent for a material to be ferroelectric?

a	An axis of rotation.
b	A mirror plane.
c	A centre of symmetry.
d	An improper axis of rotation.

Answer

C

Which of these is not a correct definition of polarisation?

a	The net dipole moment per unit volume.
b	The surface charge per unit area.
c	The movement of atoms giving rise to a dipole moment.
d	The net charge per dipole moment.

Answer

D

Why are domains in crystals found in a manner such that their polarisation is coupled?

a	To reduce stray field energy.
b	To reduce dislocation strain energy.
c	To grow the crystal in a regular manner.
d	To give small grains.

Answer

A

What is needed to make ferroelectric domains useful as a binary memory store?

a	Large domains.
b	A square hysteresis loop.
c	A large coercive field.
d	Small domains.

Answer

B

If a ferroelectric, 50mm by 10 mm, has a measured surface charge of 2.5 x 10^{-4} Coulombs, and a lattice parameter of 5 x 10^{-10}, what is the dipole moment in a single cubic unit cell?

a	6.25 x 10^{-29}
b	6.25 x 10^{-35}
c	1.5625 x 10^{-35}
d	1.5625 x 10^{-29}

Answer

A

What field has to be applied to give a ferroelectric, with a square hysteresis loop, zero net polarisation?

a	The switching field.
b	The polarising field.
c	The coercive field.
d	The stray field.

Answer

C

This page titled 28.12: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

29: Ferromagnetic Materials

Learning Objectives

On completion of this TLP you should:

- Understand ferromagnetism as a type of magnetism and some of the reasons an element is ferromagnetic
- Be aware that magnetism is affected by temperature
- Understand the factors contributing to the formation of magnetic domains
- Know why hysteresis occurs, and the factors which affect it

Before you start

Most of this package does not assume any previous understanding of ferromagnetism. However a brief understanding of atomic and electronic structure in elements may aid some explanations. However it is not necessary!

29.1: Introduction
29.11: Questions
29.2: Types of Magnetism
29.3: Why are some materials magnetic?
29.4: Why are Only Some Materials Ferromagnetic
29.5: Curie-Weiss Law
29.6: Domains
29.7: Domain Walls
29.8: Hysteresis
29.9: Anisotropy
29.10: Summary

This page titled 29: Ferromagnetic Materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.1: Introduction

Magnets are used by millions of people worldwide everyday, but what they may not realise every time they open their fridge, listen to their mp3 player or turn on their computer is that they are using a quantum mechanical phenomenon! The understanding of magnetism is extremely important in the continuing development of modern technology. For example electronic device memory is an area of intensive research as there is continuous demand for more memory in smaller devices.

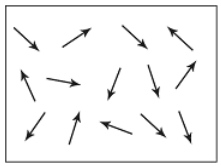
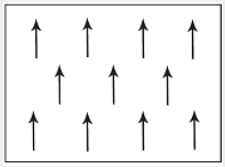
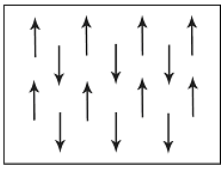
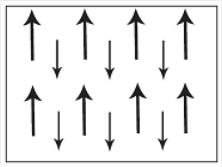
This page titled 29.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.11: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which picture best describes ferromagnetism?

a	
b	
c	
d	

Answer

B. In a ferromagnetic material below the Curie temperature all the dipole moments are aligned parallel to each other.

Which 3 elements are ferromagnetic at room temperature?

a	Mn, Cr, Fe
b	Mg, Ti, Ni
c	Fe, Ni, Co
d	Co, Ce, Mn

Answer

C. There are only 4 ferromagnetic elements at room temperature. Of these, iron (Fe), nickel (Ni), and cobalt (Co) are shown above. The fourth is gadolinium (Gd).

What happens to ferromagnetic materials above the Curie Temperature?

a	They become diamagnetic
b	They become gases
c	They become antiferromagnetic
d	They become paramagnetic

Answer

D. Above the Curie temperature there is a large enough contribution from the thermal energy to overcome the cooperative ordering of the magnetic dipole moments and so the material becomes paramagnetic.

What is the main driving force for the formation of domains?

a	Magnetostatic energy
b	Magnetocrystalline energy
c	Magnetostrictive energy
d	All of the above

Answer

A. While domains do form to minimize the total energy, the principal driving force is the Magnetostatic energy. The magnetocrystalline energy and magnetostrictive energy influence the shape and size of the domains.

This page titled 29.11: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.2: Types of Magnetism

All magnetic materials contain *magnetic moments*, which behave in a way similar to microscopic bar magnetis. In order to define a ferromagnetism as a class of magnetism, it is easiest to compare the various properties of different possible types of magnetic material. These are principally: paramagnets, ferromagnets, antiferromagnets and ferrimagnets.

Paramagnetism

In a paramagnet, the magnetic moments tend to be randomly orientated due to thermal fluctuations when there is no magnetic field. In an applied magnetic field these moments start to align parallel to the field such that the magnetisation of the material is proportional to the applied field.

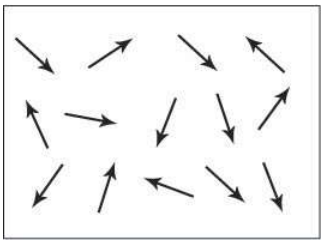


Figure A. Schematic showing the magnetic dipole moments randomly aligned in a paramagnetic sample.

Ferromagnetism

The magnetic moments in a ferromagnet have the tendency to become aligned parallel to each other under the influence of a magnetic field. However, unlike the moments in a paramagnet, these moments will then remain parallel when a magnetic field is not applied (this will be discussed later).

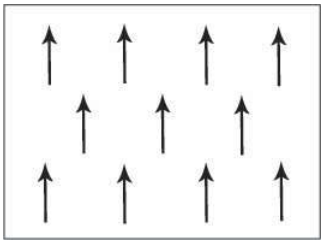


Figure B. Schematic showing the magnetic dipole moments aligned parallel in a ferromagnetic material.

Antiferromagnetism

Adjacent magnetic moments from the magnetic ions tend to align anti-parallel to each other without an applied field. In the simplest case, adjacent magnetic moments are equal in magnitude and opposite therefore there is no overall magnetisation.

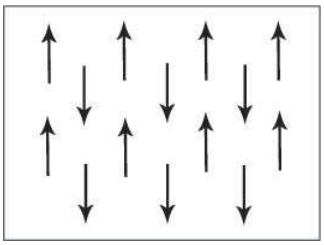


Figure C. Schematic showing adjacent magnetic dipole moments with equal magnitude aligned anti-parallel in an antiferromagnetic material. This is only one of many possible antiferromagnetic arrangements of magnetic moments.

Ferrimagnetism

The aligned magnetic moments are not of the same size; that is to say there is more than one type of magnetic ion. An overall magnetisation is produced but not all the magnetic moments may give a positive contribution to the overall magnetisation.

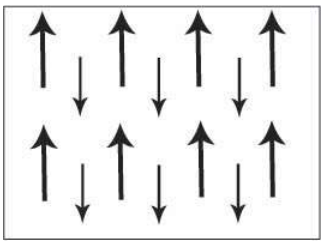


Figure D. Schematic showing adjacent magnetic moments of different magnitudes aligned anti-parallel.

Below is a periodic table showing the elements and the types of magnetism at room temperature:

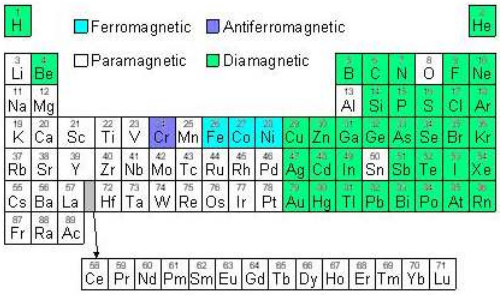


Figure E. Diagram of a periodic table showing elements coloured according to the type of magnetism they show at room temperature.

This page titled 29.2: Types of Magnetism is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.3: Why are some materials magnetic?

It was discovered by Oersted in 1820 that a magnetic compass needle was deflected when placed near to an electric current; this was a breakthrough in the understanding of magnetism that later led to Ampère observing that the magnetic field of a solenoid being identical to that of a magnet.

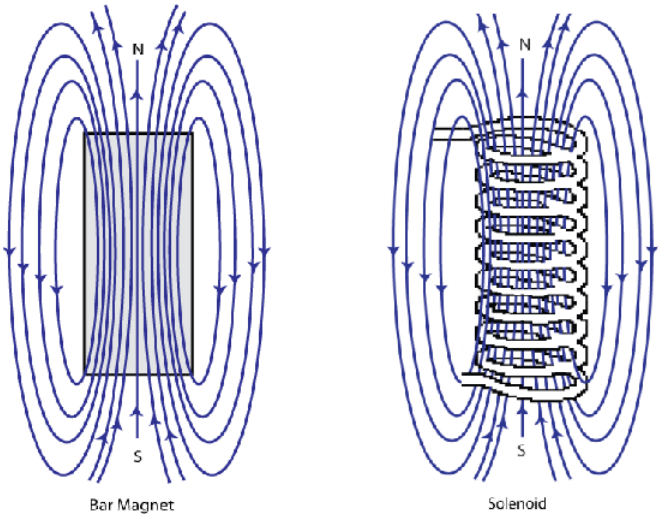


Figure F. Schematic showing the shape of the magnetic field around a bar magnet and a solenoid are identical.

Ampère then hypothesised that all magnetic effects were due to current loops and that the magnetic effects in materials must be due to “molecular currents”, attributed to the movement of electrons. However, the currents predicted by this model were unfeasibly large and a second origin of magnetism was required. This second magnetic mechanism was spin, postulated by Dirac in 1928, after he solved the fully relativistic quantum mechanical equations governing the electron.

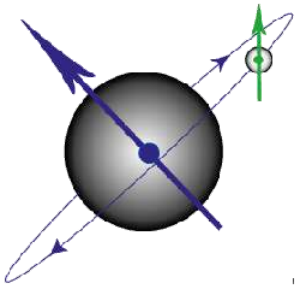


Figure G. Diagram to show the magnetic moment produced by an electron orbiting the nucleus and that produced by the spin of the electron.

The spin of an electron is hard to visualise, but has the properties of a small magnetic moment pointing either “up” or “down”. Within an atom, electrons are arranged in *orbitals*, with a maximum of two electrons with opposite spin occupying each orbital (due to the Pauli Exclusion Principle). The orbitals are further grouped into *shells*. In all atoms except for hydrogen there is more than one electron and these electrons can interact with each other as well as with the nucleus, leading to “coupling”.

Coupling

As outlined above, the total magnetic moment of a free atom has two contributions from each electron:

1. The angular momentum as the electron orbits the nucleus (strictly, the momentum of the nucleus relative to the orbiting electron). This is effectively Ampère’s molecular current and is known as the orbital contribution.
2. The ‘spin’ of the electron itself

In an atom with a single electron, there are just two magnetic fields produced which can interact (see above figure). The magnetic field from the electron's spin interacts with the magnetic field from its movement around the nucleus, leading to so-called *spin-orbit* coupling.

In an atom with more than one electron, the total magnetic moment of the atom will depend on the spin-orbit (intra-electron), spin-spin (inter-electron) and orbit-orbit (also inter-electron) coupling. The spin-orbit coupling is weak for light atoms and generally can be ignored in calculating the total angular momentum.

The total magnetic moment can be determined by simple vector addition of the fields.

Important results of this coupling are as follows. When working out the moment on an atom, only *incomplete electron shells (groups of orbitals) contribute*. Furthermore, electrons arrange themselves in shells in such a way as to maximise total spin. Putting these two results together, we see that once a shell is more than half full, spins begin to pair up within orbitals and the available magnetic moment decreases. This underlies the existence of magnetic order in elements in the middle of the 3d series and the middle of the 4f series.

This page titled 29.3: Why are some materials magnetic? is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.4: Why are Only Some Materials Ferromagnetic

For ferromagnetism to occur there must be an internal driving force that causes parallel alignment of the spins of the electrons to be the more favourable state. Weiss explained this observed behaviour by proposing the presence of an internal interaction between the localised moments called a *molecular field*. Whilst Weiss' phenomenological model represented a significant step forward, the microscopic explanation of his molecular field required the laws of quantum mechanics. The important term in the interaction of the localised moments is called the *exchange interaction*.

The exchange interaction occurs due to the *Pauli Exclusion Principle*; if two electrons have different spins then they can occupy the same orbital (angular momentum state), hence be closer to each other and they will therefore have a stronger Coulomb repulsion.

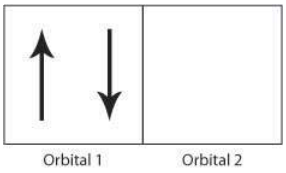


Figure H. Schematic showing two electrons occupying orbital 1 producing a strong Coulomb repulsion compared to Figure I.

If the electrons have the same spin then they will occupy different orbitals and therefore have less Coulomb repulsion as they will be further apart. In this way, the exchange energy (the energy due to the repulsion between the two electrons) is minimsed.

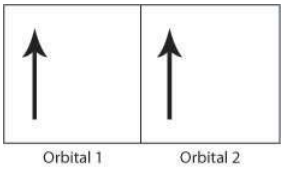


Figure I. Schematic showing two electrons occupying different orbitals.

Therefore the Coulomb repulsion force favours the parallel alignment of all the electron spins as the exchange energy is minimised. The above rules help to explain the strong ferromagnetic order seen in iron, cobalt and nickel. The Flash demo below allows the electronic structure of the transition elements in the 3rd row of the periodic table and the numbers of unpaired electrons to be seen.

Taking the elements shown in the demo above as an example, it is, however, not straightforward to predict the presence or form of magnetic order from the electronic configuration alone as not all elements with unpaired electrons are ferromagnetic. There are other factors, principally the atomic structure, that go towards determining the type of magnetism exhibited by the elements. See the discussion on band theory in reference 1.

This theory can be summarised into rules:

- In the elements, electrons maximise their total spin and so will occupy orbitals with one electron per orbital with all the spins parallel until all the available orbitals contain an electron. Once this has occurred the electrons pair up in these orbitals in pairs of opposite spin.
- For atoms with filled electron shells the total atomic orbital angular momentum and total spin are zero; therefore there is no magnetic dipole moment.
- For atoms with incomplete outer shells, only the incomplete shells need to be considered in calculating the total angular momentum and hence magnetic dipole moment.
- If an atom has no incomplete shells then there is no permanent dipole. These atoms are called diamagnetic.
- Materials that do contain unpaired electrons are ferromagnetic only if the atomic structure allows the cooperative parallel alignment of the magnetic dipole moments.

This page titled 29.4: Why are Only Some Materials Ferromagnetic is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.5: Curie-Weiss Law

Above a critical temperature T_c, the Curie temperature, all ferromagnetic materials become paramagnetic. This is because thermal energy is large enough to overcome the cooperative ordering of the magnetic moments.

The susceptibility of a material, χ, indicates how dramatically a material responds to an applied magnetic field, and is defined as the ratio of the magnetisation of the material, M, and the applied magnetic field, H.

$$\chi = \frac{M}{H} \tag{29.5.1}$$

The magnetisation of a material, M, is defined as the magnetic moment per unit volume or per unit mass of a material and is dependent on the individual magnetic dipole moments of the atoms in the material and on the interactions of these dipoles with each other.

Above the Curie Temperature there will be a change in the susceptibility as the material becomes paramagnetic, therefore giving the equation:

$$\chi = \frac{C}{T - T_c} = \frac{M}{H} \tag{29.5.2}$$

where C is a constant.

The graph below shows the saturation magnetisation (ie that obtained in a high magnetic field) of a ferromagnetic element, nickel, as a function of temperature. We see that the saturation magnetisation decreases with increasing temperature until it falls to zero at the Curie temperature where the material becomes paramagnetic:

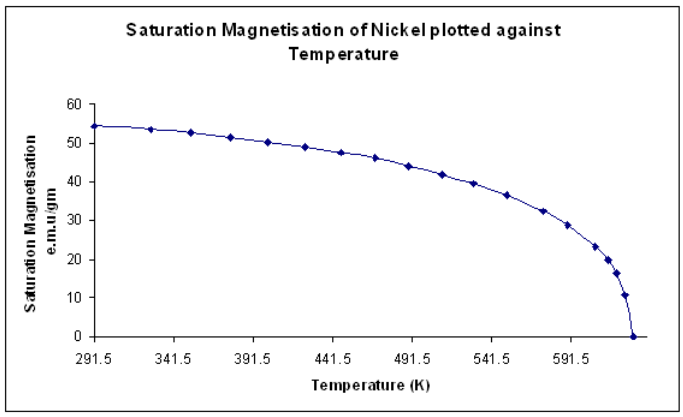


Figure J. Variation of saturation magnetisation with temperature for Nickel. (Data from Weiss and Forrer, 1926)

Differentiation of Equation 2 with respect to temperature shows that the susceptibility is a maximum at the Curie temperature. This is no surprise; it is easiest to increase the magnetic moment of the material by applying a magnetic field when the material is undergoing a transition between magnetic order and disorder. The graph below for nickel shows the susceptibility tending to infinity as the temperature moves closer to the Curie temperature:

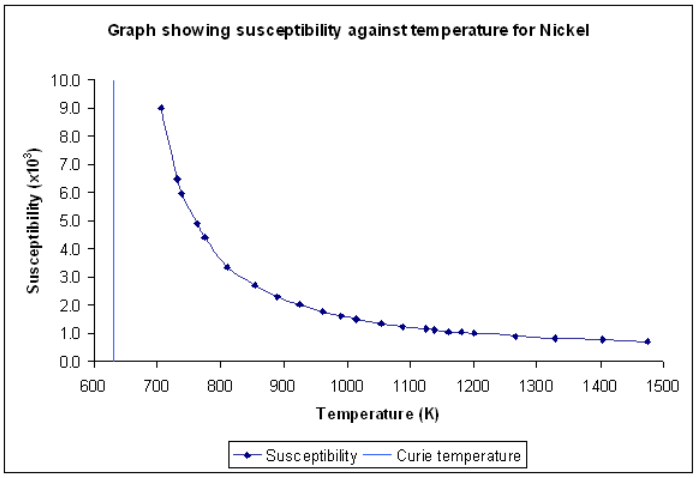


Figure K. Variation of susceptibility with temperature for Nickel (Sucksmith and Pearce, 1938)

This page titled 29.5: Curie-Weiss Law is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.6: Domains

Domains are regions of a ferromagnetic material in which the magnetic dipole moments are aligned parallel. When the material is demagnetised the vector summation of all the dipole moments from all the domains equals zero. When the material is magnetised the vector summation of the dipoles gives an overall magnetic dipole (we will discuss this in more detail later).

Why do domains occur?

From our previous discussion of the exchange interaction it would appear that the most stable state would be that of a single domain where all the electron spins would be aligned parallel. However, while this minimises the energy contribution arising from the exchange interaction, there are other contributions to the total magnetic energy that must be considered:

- Magnetostatic energy
- Magnetostrictive energy
- Magnetocrystalline energy

We will now consider each in turn:

Magnetostatic energy

If a material consists of a single domain then it behaves as a block magnet (Figure L (i) below) and so a "demagnetising field" (the blue arrows) must be present around the block. This external, demagnetising field has a magnetostatic energy that depends on the shape of the sample and is the field that allows work to be done by the magnetised sample (e.g. lifting another ferromagnetic material against the force of gravity). In order to minimise the total magnetic energy the magnetostatic energy must be minimised. This can be achieved by decreasing the external demagnetising field by dividing the material into domains (Figure L (ii)). Adding extra domains increases the exchange energy, as the domains cannot align parallel, however the total energy has been decreased as the magnetostatic energy is the dominant effect. The magnetostatic energy can be reduced to zero by a domain structure that leaves no external demagnetising field (Figure L (iii)).

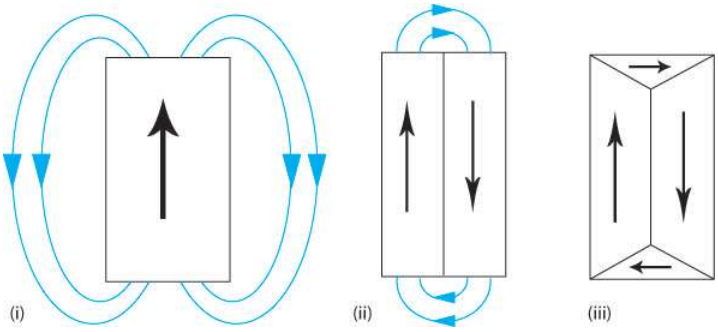


Figure L. Schematic showing how the addition of domains can reduce the external demagnetising field therefore reducing the magnetostatic energy.

This is the main driving force for the formation of domains.

Magnetocrystalline energy

Up until now, we have ignored the influence of the atomic lattice structure; however this also has an affect on the total energy of a magnetised sample. A ferromagnetic material has '*easy*' crystallographic directions along which it is preferred that the magnetisation vector points and '*hard*' directions along which a higher field is required to achieve the same magnetisation. Therefore, it is easiest to magnetise a ferromagnetic material along these easy axes, as shown by the schematic below.

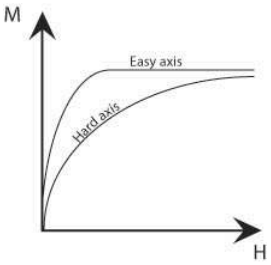


Figure M. Schematic showing the difference in size of field required to achieve the same magnetisation along easy and hard axes.

There is an energy difference associated with magnetisation along the hard and easy axes which is given by the difference in the areas under (*M*,*H*) curves. This is called the magnetocrystalline energy. This energy can be minimised by forming domains such that their magnetisations point along the easy crystallographic directions. The ideal material might have easy crystallographic directions perpendicular to one another as then both the magnetostatic and magnetocrystalline energies could both be minimised (Figure L (iii) above). In the regions bounding the domains, the *domain walls*, there must be a change in the direction of the magnetisation. Therefore, it cannot be aligned along easy axes and so large domains with few domain walls minimise the magnetocrystalline energy.

Below a table shows a summary of the easy, hard and intermediate directions in iron, nickel and cobalt:

	Fe	Ni	Co
	bcc	fcc	hexagonal
Easy	<100>	<111>	<1000>
Intermediate	<110>	<110>	
Hard	<111>	<100>	<1010>

Magnetostrictive energy

When a ferromagnetic material is magnetised it changes length. This is known as magnetostriction; an increase in length along the direction of magnetisation is positive magnetostriction (e.g. in Fe), and a decrease in length is negative magnetostriction (e.g. in Ni). These length changes are usually extremely small; in the range of tens of parts per million. However, they do affect the domain structure of the material.

In iron the change in length causes the domains of closure to attempt to elongate horizontally (shown in blue in Figure N (i)) and the vertical domains to attempt to elongate vertically (shown in green in Figure N (i)). It is impossible for both to be accommodated and so this causes elastic strain in the material. The elastic strain energy is proportional to the volume of the domains of closure hence magnetostrictive energy can be minimised by decreasing the size of these domains. Reducing the volume of the domains of closure also requires the primary domains to decrease in size and increase in number (Figure N (ii)). The addition of extra domain walls increases the magnetocrystalline and exchange energy contributions to the total energy.

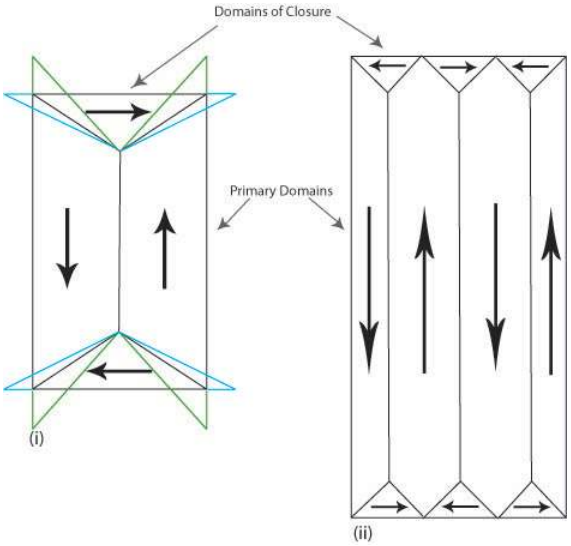


Figure N. Schematic showing the attempted change in shape of the domains (i) and the actual change (ii) due to magnetostriction.

The actual domain structure is a compromise of all the aforementioned energy contributions.

This page titled 29.6: Domains is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.7: Domain Walls

Domain walls are the regions between domains where the direction of magnetisation must change, usually by either 180° or 90°.

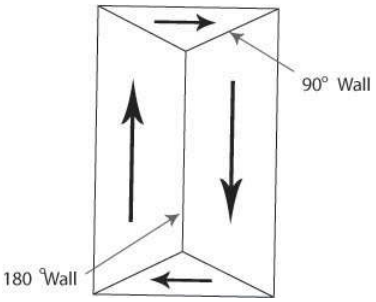


Figure O. Diagram showing domain walls.

The width of domain walls is controlled by the balance of two energy contributions:

- Exchange energy
- Magnetocrystalline energy

The exchange energy favours wide walls where adjacent magnetic dipole moments can be as close to parallel as possible, whereas the magnetocrystalline energy favour sharp changes in the dipole moments between the favoured directions in the crystal so that as few dipole moments as possible point along “non-easy” directions. The actual width is determined by the minimum of the total energy.

The most favourable domain walls are those which do not require an external demagnetising field. Three walls, or “boundaries” of this type are discussed below.

1. Twist Boundary
   The magnetisation perpendicular to the boundary does not vary across the domain wall hence no demagnetising fields are generated.

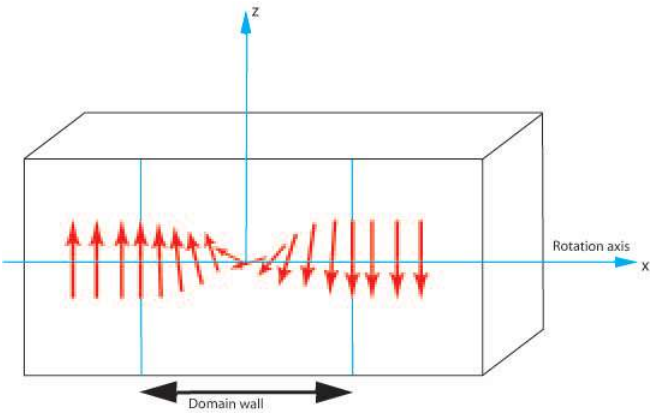


Figure P. Diagram showing the rotation of magnetic moments through a 180° domain wall.

2. Tilt Boundary
   The magnetic moments rotate in such a manner that a constant angle is maintained between them and both the wall normal and the surface.

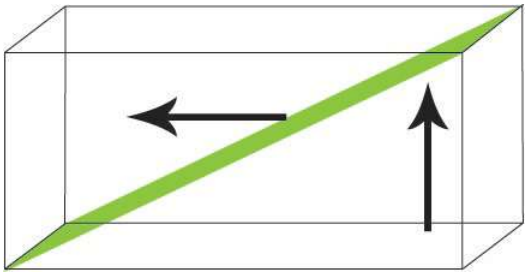


Figure Q. Diagram showing the rotation of magnetic moments in a tilt boundary

3. Néel Wall

   In thin films a Néel wall occurs, and the magnetic dipole moments rotate around an axis perpendicular to the surface of the film. These are favourable in thin films because the free poles are formed on the domain wall surface rather than the film surface, causing a reduction in magnetostatic energy.

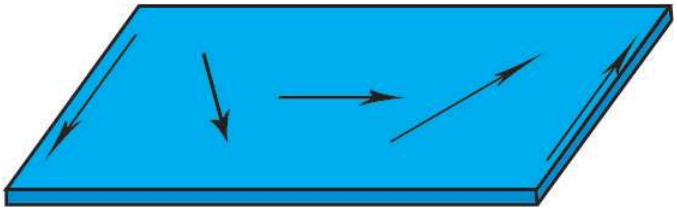


Figure R. Diagram showing the rotation of magnetic moments in a thin film in a plan view.

This page titled 29.7: Domain Walls is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.8: Hysteresis

Magnetic hysteresis is an important phenomenon and refers to the irreversibility of the magnetisation and demagnetisation process. When a material shows a degree of irreversibility it is known as *hysteretic*. We will now explore the physics behind ferromagnetic hysteresis.

When a demagnetised ferromagnetic material is placed in an applied magnetic field the domain that has a direction closest to that of the applied field grows at the expense of the other domains. Such growth occurs by motion of the domain walls. Initially domain wall motion is reversible, and if the applied field is removed the magnetisation will return to the initial demagnetised state. In this region the magnetisation curve is reversible and therefore does not show *hysteresis*.

The crystal will contain imperfections, which the domain boundaries encounter during their movement. These imperfections have an associated magnetostatic energy. When a domain wall intersects the crystal imperfection this magnetostatic energy can be eliminated as closure domains form. This pins the domain wall to the imperfection, as it is a local energy minima. The applied magnetic field provides the energy to allow the domain wall to move past the crystal imperfection, but the domains of closure cling to the imperfection forming spike-like domains that stretch as the domain wall moves further away. Eventually these spike domains snap off and the domain wall can move freely. As the spike domains snap off there is a discontinuous jump in the boundary leading to a sharp change in the magnetic flux, which can be detected by winding a coil around the specimen connected to a speaker. In doing so, crackling noises are heard corresponding to the spike domains breaking away from the domain walls. This phenomenon is known as the Barkhausen effect.

Eventually all the domain walls will have been eliminated leaving a single domain with its magnetic dipole moment pointing along the easy axis closest to the direction of the applied magnetic field. Further increase in magnetisation can occur by this domain rotating away from the easy direction to an orientation parallel to that of the externally applied field. The magnetisation of the material at this stage is called the saturation magnetisation (see Figure J) . The ease of this final rotation is dependent on the magnetocrystalline energy of the material; some materials require a large field to reach this saturation magnetisation.

If the external applied field is removed the single domain will rotate back to the easy direction in the crystal. A demagnetising field will be set up due to the single domain, and this field initiates the formation of reverse magnetic domains as these will lower the

magnetostatic energy of the sample by reducing the demagnetising field. However the demagnetising field is not strong enough for the domain walls to be able to grow past crystal defects so they can never fully reverse back to their original positions when there is no external applied field. This results in the *hysteresis* curve as some magnetisation will remain when there is no external applied field. This magnetisation is called the remanent magnetisation, B_r. The field required to reduce the magnetisation of the sample to zero is called the coercive field H_c. And the saturation magnetisation B_s is the magnetisation when all the domains are aligned parallel to the external field. These are shown on the schematic below:

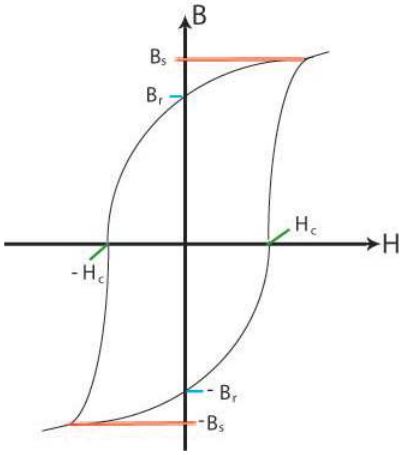


Figure S. Schematic showing the general shape of the hysteresis curve with some relevant points marked; B_s, the saturation magnetisation; B_r, the remanent magnetisation; H_c, the coercive field.

The movie clip below shows the domain structure as the material is subjected to a hysteresis cycle.

Hard and soft magnets

Ferromagnetic materials are further classified as soft or hard. Hard magnets are also called permanent magnets. These require a large field to demagnetise (and magnetise). Generally hard magnets have a large remenant magnetisation. Soft magnets, however,

are easily magnetised and demagnetised, they have a small coercive field, and generally a small remanent magnetisation. Soft and hard magnets have different applications depending on the ease of magnetisation and demagnetisation required. See the animation below for examples of hard and soft magnets, their hysteresis curves and applications.

This page titled 29.8: Hysteresis is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.9: Anisotropy

Magnetocrystalline anisotropy

We have already met magnetocrystalline anisotropy in the section about domains where it was stated that there were easy and hard axes in a crystalline material such that a magnetic moment would preferentially orient along the former. The magnetocrystalline anisotropy energy is the energy difference between samples magnetised along the easy and hard directions. It is important to note that the spontaneous magnetisation is the same for both directions but the external applied field required to reach this value is different.

To produce the spontaneous magnetisation the applied field must re-orient the atomic orbital in order to re-orient the direction of the electron spin; this is due to the spin-orbit coupling. The atomic orbitals are strongly coupled to the lattice and so resist the change in orientation. In most magnetic materials the spin-orbit coupling is weak; however in some heavy rare-earth metals the coupling is strong. In the latter case, the resistance to re-orientating the domains away from the easy crystallographic directions is high and requires a large coercive field. Such materials have uses as permanent magnets.

Magnetocrystalline anisotropy decreases with increasing temperature, until at temperatures approaching the Curie temperature there is no preferred orientation of the magnetisation.

Shape Anisotropy in Polycrystals

If a sample is non-spherical it is easier to magnetise with the dipole moments pointing along a long axis. The field required to magnetise the sample must overcome the demagnetising field, which is minimised if it the magnetisation lies along the long axis. The magnetisation produces surface poles, and the magnitude of the demagnetising field decreases as the distance between the poles increases. This therefore has consequences for the shape anisotropy of the sample: the more prolate a sample is, the closer the distance between poles becomes, and the larger the demagnetising field and shape anisotropy is (Figure T).

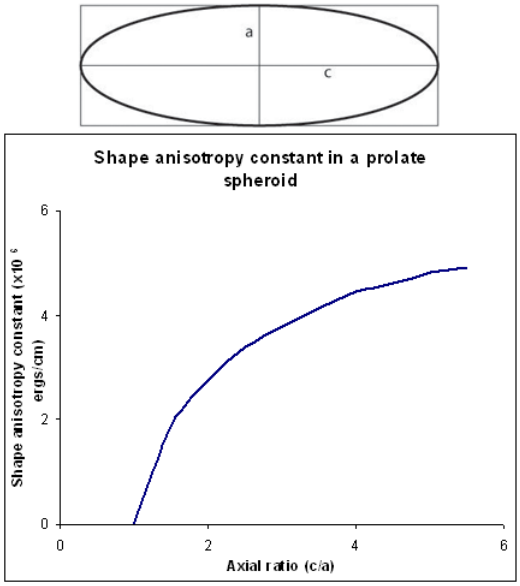


Figure T. Schematic showing the shape anisotropy constant for a prolate spheroid

Induced Magnetic Anisotropy

It is possible to produce anisotropy in polycrystals by using a treatment with directional characteristics giving texture to the sample. Possible processes include:

- Casting
- Wire-drawing
- Annealing in presence of a magnetic field giving easy axis parallel to applied field

- Rolling

This page titled 29.9: Anisotropy is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

29.10: Summary

This TLP has covered the basic points of ferromagnetism:

1. In a magnetic atom there are two contributions to the magnetic dipole moment; firstly the spin of the electrons themselves and secondly that of electrons orbiting the nucleus.
2. Ferromagnetism occurs in materials where all the magnetic dipole moments align parallel below the Curie temperature.
3. Ferromagnetic ordering was explained by Weiss via a hypothetical average field which acts to cause the parallel alignment. However, the microscopic explanation for this can be found by looking at the Pauli Exclusion Principle; it is energetically more favourable for electrons be placed in different orbitals as this reduces the Coulomb repulsion energy and allows for the alignment of the electron spins.
4. The magnitude of the magnetisation is dependent on temperature and modelled by the Curie-Weiss law:

$$\chi = \frac{C}{T - T_c} = \frac{M}{H} \tag{29.10.1}$$

1. The formation of domains is driven by the minimisation of energy, with the main driving force often being that of the magnetostatic energy.
2. Magnetic hysteresis is seen, due to the defects found in crystals, as these hinder the movement of domain walls.
3. Hard magnets have a large coercive field, whereas soft magnets are easily demagnetised and so has a small coercive field.
4. Both magnetocrystalline anisotropy and shape anisotropy give directions in a material along which it is easier to magnetise a sample.

Going further

Books and Papers

1. Magnetic Materials : Fundamentals and Device Applications by N.A. Spaldin, Cambridge University Press (2003)
   A good overall explanation of ferromagnetism which also covers other types of magnetism and some applications.
2. The Feynman Lectures on Physics by R.P. Feynman, R.B. Leighton and M. Sands, Addison-Wesley Publishing Company (1970)
   Provides a complete explanation of magnetism including the mathematics and qualitative explanations.
3. P.A.M. Dirac, The quantum theory of the electron, Proc. R. Soc. London A 117 610-612 (1928)
4. P.A.M. Dirac, The quantum theory of the electron Part II, Proc. R. Soc. London A 118 351-361 (1928)

These are the original papers by Dirac in which the spin of the electron was derived.

Websites

- Get Perpendicular: Hitachi Global Storage Technologies
  This website gives a fun summary of one of the latest developments in magnetic memory.

This page titled 29.10: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

30: Finite Element Method

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

On completion of this TLP you should:

- Understand the basics of the finite element method
- Be familiar with the concepts of nodes, elements and discretisation
- Understand the direct stiffness method
- Be able to construct an element stiffness matrix and a global stiffness matrix for 1-dimensional elements
- Appreciate the importance of boundary conditions
- Understand shape (interpolation) functions for 1-dimensional elements
- Understand the difference between linear and non-linear static finite element problems
- Be familiar with some common reasons for non-convergence

Before you start

There are no special prerequisites for this TLP, but it would be useful to be familiar with stress and strain, beam bending mechanics and matrices (see Tensors in Materials Science).

30.1: Introduction
30.2: Nodes, Elements, Degrees of Freedom and Boundary Conditions
30.3: Direct Stiffness Method and the Global Stiffness Matrix
30.4: Enforcing Boundary Conditions
30.5: Interpolation/Basis/Shape Functions
30.6: 1D First Order Shape Functions
30.7: 1D Second Order Shapes and Functions
30.8: Typical steps during FEM modelling
30.9: Summary
30.10: Questions

This page titled 30: Finite Element Method is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

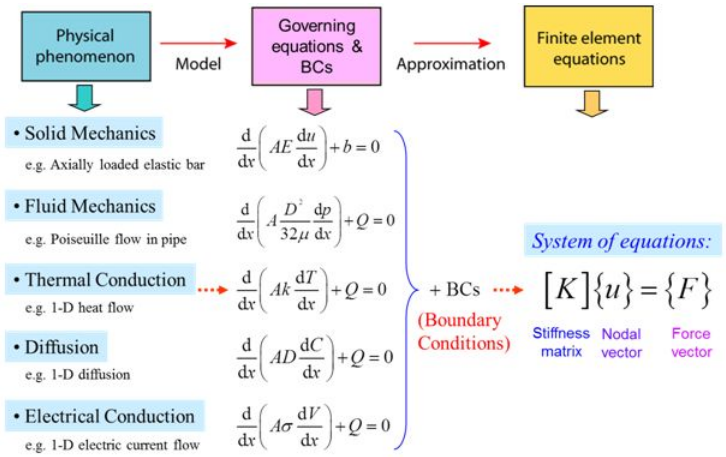
30.1: Introduction

Overview

The Finite Element Method (FEM) is a mathematical (numerical) technique for finding approximate solutions to partial differential equations. It is a technique which is very well suited to the study and analysis of complex physical phenomena, particularly those exhibiting non-linearity in geometry and/or material behaviour (which is often the case for most “real-world” situations). It is used frequently to tackle problems that are not readily amenable to analytical treatments. Such problems can be structural, thermal, electrical, magnetic, acoustic etc., either in isolation or when coupled. Coupled examples include (but are not limited to) thermomechanical – where constrained differential thermal expansion generates thermal-stresses, thermoelectric – where heat is generated in a material due to its resistance to current flow, and conjugate heat transfer – where moving fluids can remove heat from hot objects over which they pass. When the Finite Element Method is used to solve problems of this type it is often referred to as Finite Element Analysis (FEA).

Premise

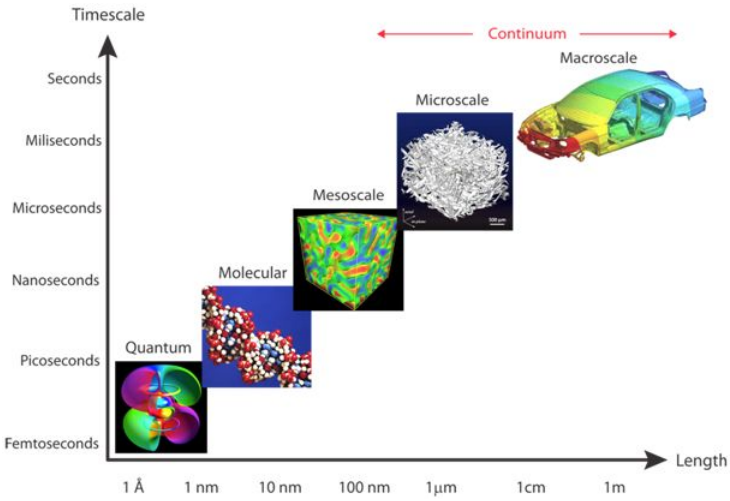
The premise is very simple; continuous domains (geometries) are decomposed into discrete, connected regions (or ***Finite Elements***). An assembly of ***element-level*** equations is subsequently solved, while conditions of kinematic compatibility are enforced, in order to establish the response of the complete domain to a particular set of boundary conditions. The premise is described to some degree in the figure below, which also lists the governing equations (in differential form) for a range of common physical phenomena (in 1-D).



Time and Lengthscales

Typical time and lengthscales for application of the finite element method are shown in the following figure.

https://eng.libretexts.org/@go/page/32860



Example Videos:

The videos below have been produced from finite element analyses and include:

https://www.doitpoms.ac.uk/tlplib/fe...erforation.mp4

FEM of plate perforation

https://www.doitpoms.ac.uk/tlplib/fe...2_40_Mises.mp4

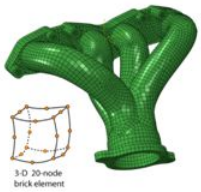
Punch

This page titled 30.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.2: Nodes, Elements, Degrees of Freedom and Boundary Conditions

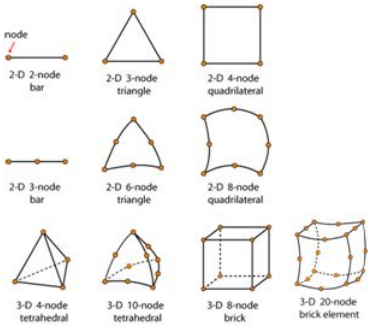
"Nodes", "Elements", "Degrees of Freedom" and "Boundary Conditions" are important concepts in Finite Element Analysis.

When a domain (a geometric region) is meshed, it is decomposed into a series of discrete (hence finite) **ELEMENTS**. The meshed geometry of an exhaust manifold is shown in the figure below. It can be seen that the elements in the mesh conform very well to the geometry, and represent therefore a good approximation of the geometry. The manifold in this instance is meshed with a 3-dimensional brick element which contains 20 **NODES**. Adjacent elements are connected to each other **AT** the nodes. A node is simply a point in space, defined by its coordinates, at which **DEGREES OF FREEDOM** are defined. In finite element analysis a degree of freedom can take many forms, but depends on the type of analysis being performed. For instance, in a structural analysis the degrees of freedom are displacements (U_x, U_y and U_z), while in a thermal analysis the degree of freedom is temperature (T). In the exhaust manifold example, there are 4 degrees of freedom at each node – U_x, U_y, U_z and T, since the analysis is a coupled temperature-displacement analysis (due to thermal expansion effects). These **FIELD VARIABLES** are calculated at every node from the governing equation. Field variable values between the nodes and within the elements are calculated using interpolation functions, which are sometimes called shape or base functions.



Manifold and node map

There are of course many types of element, covering the complete range of space dimension. The most common are shown in the figure below, along with the position of the nodes, where It can be seen that some of the elements have "midside" nodes – i.e. nodes positioned midway between the corner nodes. The edges of these "***higher order***" elements can therefore curve – making them suitable for capturing complex geometrical shapes (as in the manifold above). This is possible since these elements permit the solution between the nodes to vary in non-linear ways (see section on shape functions), which is an important feature when field variables change rapidly.



Examples of FEM element types

Boundary Conditions

Boundary conditions are specified values of the field variables (or their derivatives) on the boundaries of the field (the geometry). They fall in to three categories: Dirichlet conditions – where you prescribe the variable for which you are solving; Neumann conditions – where you prescribe a flux, which is the gradient of the dependent variable, and Robin conditions – which is a mixture of the two, where a relation between the variable and its gradient is prescribed. The following table features some examples from various physical fields that show the corresponding physical interpretation.

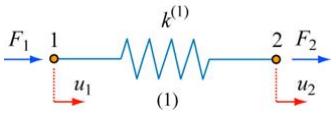
Physics	Dirichlet	Neumann	Robin
Solid Mechanics	Displacement	Traction (Stress)	Spring
Heat Transfer	Temperature	Heat Flux	Convection
Pressure Acoustics	Acoustic Pressure	Normal Acceleration	Impedance
Electric Currents	Fixed Potential	Fixed Current	Impedance

This page titled 30.2: Nodes, Elements, Degrees of Freedom and Boundary Conditions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.3: Direct Stiffness Method and the Global Stiffness Matrix

Although there are several finite element methods, we analyse the Direct Stiffness Method here, since it is a good starting point for understanding the finite element formulation. We consider first the simplest possible element – a 1-dimensional elastic spring which can accommodate only tensile and compressive forces. For the spring system shown, we accept the following conditions:

- *Condition of Compatibility – connected ends (nodes) of adjacent springs have the same displacements*
- *Condition of Static Equilibrium – the resultant force at each node is zero*
- *Constitutive Relation – that describes how the material (spring) responds to the applied loads*



Model spring system

The constitutive relation can be obtained from the governing equation for an elastic bar loaded axially along its length:

$$\frac{d}{du}(AE\frac{\Delta l}{l_0}) + k = 0 \tag{30.3.1}$$

$$\frac{\Delta l}{l_0} = \varepsilon \tag{30.3.2}$$

$$\frac{d}{du}(AE\varepsilon) + k = 0 \tag{30.3.3}$$

$$\frac{d}{du}(A\sigma) + k = 0 \tag{30.3.4}$$

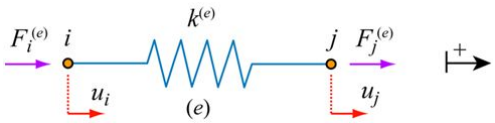
$$\frac{dF}{du} + k = 0 \tag{30.3.5}$$

$$\frac{dF}{du} = -k \tag{30.3.6}$$

$$dF = -kdu \tag{30.3.7}$$

The spring stiffness equation relates the nodal displacements to the applied forces via the spring (element) stiffness. The minus sign denotes that the force is a restoring one, but from here on in we use the scalar version of Eqn.7.

Derivation of the Stiffness Matrix for a Single Spring Element



From inspection, we can see that there are two degrees of freedom in this model, u_i and u_j. We can write the force equilibrium equations:

$$k^{(e)}u_i - k^{(e)}u_j = F_i^{(e)} \tag{30.3.8}$$

$$-k^{(e)}u_i + k^{(e)}u_j = F_j^{(e)} \tag{30.3.9}$$

In matrix form

$$\begin{bmatrix} k^e & -k^e \\ -k^e & k^e \end{bmatrix} \begin{Bmatrix} u_i \\ u_j \end{Bmatrix} = \begin{Bmatrix} F_i^{(e)} \\ F_j^{(e)} \end{Bmatrix} \tag{30.3.10}$$

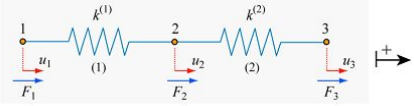
The order of the matrix is [2×2] because there are 2 degrees of freedom. Note also that the matrix is symmetrical. The '**element**' stiffness relation is:

$$[K^{(e)}]\,[u^{(e)}] = [F^{(e)}] \tag{30.3.11}$$

Where $K^{(e)}$ is the element stiffness matrix, $u^{(e)}$ the nodal displacement vector and $F^{(e)}$ the nodal force vector. (The element stiffness relation is important because it can be used as a building block for more complex systems. An example of this is provided later.)

Derivation of a Global Stiffness Matrix

For a more complex spring system, a 'global' stiffness matrix is required – i.e. one that describes the behaviour of the complete system, and not just the individual springs.



From inspection, we can see that there are two springs (elements) and three degrees of freedom in this model, u_1, u_2 and u_3. As with the single spring model above, we can write the force equilibrium equations:

$$k^1 u_1 - k^1 u_2 = F_1 \tag{30.3.12}$$

$$-k^1 u_1 + (k^1 + k^2) u_2 - k^2 u_3 = F_2 \tag{30.3.13}$$

$$k^2 u_3 - k^2 u_2 = F_3 \tag{30.3.14}$$

In matrix form

$$\begin{bmatrix} k^1 & -k^1 & 0 \\ -k^1 & k^1 + k^2 & -k^2 \\ & -k^2 & k^2 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} \tag{30.3.15}$$

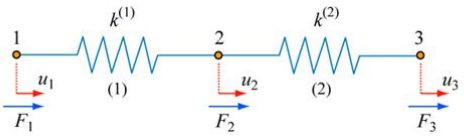
The '**global**' stiffness relation is written in Eqn.16, which we distinguish from the 'element' stiffness relation in Eqn.11.

$$[K]\{u\} = \{F\} \tag{30.3.16}$$

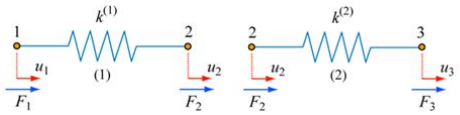
Note the shared k^1 and k^2 at k^{22} because of the compatibility condition at u_2. We return to this important feature later on.

Assembling the Global Stiffness Matrix from the Element Stiffness Matrices

Although it isn't apparent for the simple two-spring model above, generating the global stiffness matrix (directly) for a complex system of springs is impractical. A more efficient method involves the assembly of the individual element stiffness matrices. For instance, if you take the 2-element spring system shown,



split it into its component parts in the following way



and derive the force equilibrium equations

$$k^1 u_1 - k^1 u_2 = F_1 \tag{30.3.17}$$

$$k^1 u_2 - k^1 u_1 = k^2 u_2 - k^2 u_3 = F_2 \tag{30.3.18}$$

$$k^2 u_3 - k^2 u_2 = F_3 \tag{30.3.19}$$

then the individual element stiffness matrices are:

$$\begin{bmatrix} k^1 & -k^1 \\ k^1 & k^1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \end{Bmatrix} \tag{30.3.20}$$

and

\[

$$\begin{bmatrix} k^2 & -k^2 \\ k^2 & k^2 \end{bmatrix} \tag{30.3.21}$$

$$\begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} \tag{30.3.22}$$

=

$$\begin{Bmatrix} F_2 \\ F_3 \end{Bmatrix} \tag{30.3.23}$$

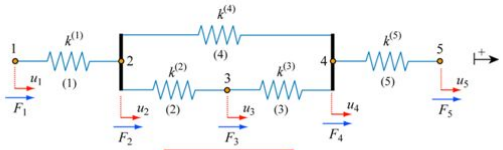
\]

such that the global stiffness matrix is the same as that derived directly in Eqn.15:

$$\begin{bmatrix} k^1 & -k^1 & 0 \\ -k^1 & k^1+k^2 & -k^2 \\ 0 & -k^2 & k^2 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix}$$

(Note that, to create the global stiffness matrix by assembling the element stiffness matrices, k^{22} is given by the sum of the direct stiffnesses acting on node 2 – which is the compatibility criterion. Note also that the indirect cells k_{ij} are either zero (no load transfer between nodes i and j), or negative to indicate a reaction force.)

For this simple case the benefits of assembling the element stiffness matrices (as opposed to deriving the global stiffness matrix directly) aren't immediately obvious. We consider therefore the following (more complex) system which contains 5 springs (elements) and 5 degrees of freedom (problems of practical interest can have tens or hundreds of thousands of degrees of freedom (and more!)). Since there are 5 degrees of freedom we know the matrix order is 5×5. We also know that it's symmetrical, so it takes the form shown below:



$$[K]\{u\} = \{F\}$$

$$\begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \end{Bmatrix}$$

We want to populate the cells to generate the global stiffness matrix. From our observation of simpler systems, e.g. the two spring system above, the following rules emerge:

- The term in location ii consists of the sum of the *direct stiffnesses* of all the elements meeting at node i
- The term in location ij consists of the sum of the indirect stiffnesses relating to nodes i and j of all the elements joining node i to j
- Add a negative for reaction terms ($-k_{ij}$)
- Add a zero for node combinations that don't interact

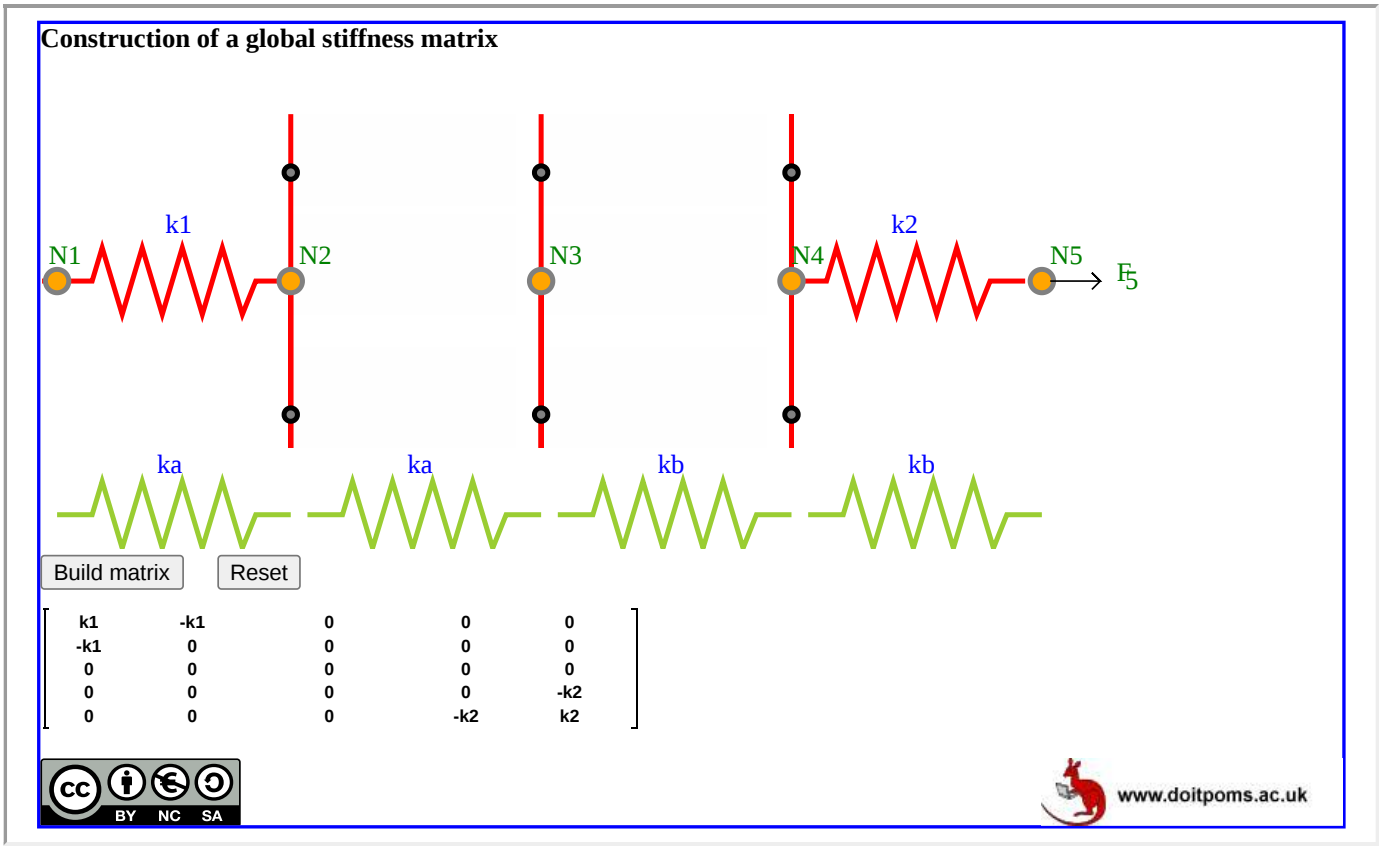
By following these rules, we can generate the global stiffness matrix:

$$\begin{bmatrix} k^{(1)} & -k^{(1)} & 0 & 0 & 0 \\ -k^{(1)} & k^{(1)}+k^{(2)}+k^{(4)} & -k^{(2)} & -k^{(4)} & 0 \\ 0 & -k^{(2)} & k^{(2)}+k^{(3)} & -k^{(3)} & 0 \\ 0 & -k^{(4)} & -k^{(3)} & k^{(3)}+k^{(4)}+k^{(5)} & -k^{(5)} \\ 0 & 0 & 0 & -k^{(5)} & k^{(5)} \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \end{Bmatrix}$$

This type of assembly process is handled automatically by commercial FEM codes

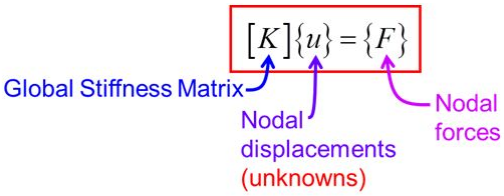
Drag the springs into position and click 'Build matrix', then apply a force to node 5. You will then see the force equilibrium equations, the equivalent spring stiffness and the displacement at node 5.

https://eng.libretexts.org/@go/page/32862



Solving for (u)

The unknowns (degrees of freedom) in the spring systems presented are the displacements u_{ij}. Our global system of equations takes the following form:



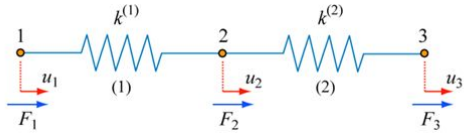
To find {u} solve

$$u = F[K]^{-1} \tag{30.3.24}$$

Recall that

$$[k][k]^{-1} = I = IdentityMatrix = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{30.3.25}$$

Recall also that, in order for a matrix to have an inverse, its determinant must be non-zero. If the determinant is zero, the matrix is said to be singular and no unique solution for Eqn.22 exists. For instance, consider once more the following spring system:



We know that the global stiffness matrix takes the following form

$$\begin{bmatrix} k^1 & -k^1 & 0 \\ -k^1 & k^1+k^2 & -k^2 \\ 0 & -k^2 & k^2 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} \tag{30.3.26}$$

The determinant of [K] can be found from:

$$det\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = (aei + bfg + cdh) - (ceg + bdi + afh) \tag{30.3.27}$$

Such that:

$$(k^1(k^1 + k^2)k^2 + 0 + 0) - (0 + (-k^1 - k^1k^2) + (k^1 - k^2 - k^3)) \tag{30.3.28}$$

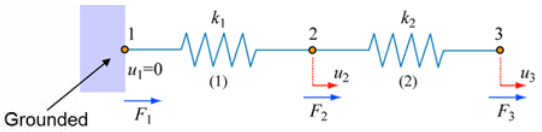
$$det[K] = (k^{1^2}k^2 + k^1k^{2^2}) - (k^{1^2}k^2 + k^1k^{2^2}) = 0 \tag{30.3.29}$$

Since the determinant of [K] is zero it is not invertible, but singular. There are no unique solutions and {u} cannot be found. If this is the case in your own model, then you are likely to receive an error message!

This page titled 30.3: Direct Stiffness Method and the Global Stiffness Matrix is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.4: Enforcing Boundary Conditions

By enforcing boundary conditions, such as those depicted in the system below, [K] becomes invertible (non-singular) and we can solve for the reaction force F_1 and the unknown displacements u_2 and u_3, for known (applied) F_2 and F_3.



No BC:

$$\begin{bmatrix} k^{(1)} & -k^{(1)} & 0 \\ -k^{(1)} & k^{(1)}+k^{(2)} & -k^{(2)} \\ 0 & -k^{(2)} & k^{(2)} \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} \rightarrow \det[K] = 0$$

Enforce BC: $u_1 = 0$

$$\begin{bmatrix} k^{(1)} & -k^{(1)} & 0 \\ -k^{(1)} & k^{(1)}+k^{(2)} & -k^{(2)} \\ 0 & -k^{(2)} & k^{(2)} \end{bmatrix} \begin{Bmatrix} 0 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ F_2 \\ F_3 \end{Bmatrix} \rightarrow \det[K] \neq 0$$

$$[K] = \begin{bmatrix} k^1 + k^2 & -k^2 \\ -k^2 & k^2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{30.4.1}$$

$$det[K] = ad - cb \tag{30.4.2}$$

$$det[K] = (k^1 + k^2)k^2 - k^{2^2} = k^1 k^2 \neq 0 \tag{30.4.3}$$

Unique solutions for F_1, $\{u_2\}$ and $\{u_3\}$ can now be found

$$k^1 u_2 = F_1 \tag{30.4.4}$$

$$(k^1 + k^2)u_2 - k^2 u_3 = F_2 = k^1 u_2 + k^2 u_2 - k^2 u_3 \tag{30.4.5}$$

$$-k^2 u_2 + k^2 u_3 = F_3 \tag{30.4.6}$$

Unique solutions:

$$-k^{(1)}u_2 = F_1$$

$$u_2 = \frac{1}{k^{(1)}}(F_2 + F_3)$$

$$u_3 = \frac{F_2}{k^{(1)}} + \left(\frac{1}{k^{(1)}} + \frac{1}{k^{(2)}}\right)F_3$$

In this instance we solved three equations for three unknowns. In problems of practical interest the order of $[K]$ is often very large and we can have thousands of unknowns. It then becomes impractical to solve for $\{u\}$ by direct inversion of the global stiffness matrix. We can instead use Gauss elimination which is much more suitable for solving systems of linear equations with thousands of unknowns.

Gauss Elimination

We have a system of equations

$$x - 3y + z = 4 \tag{30.4.7}$$

$$2x - 8y + 8z = -2 \tag{30.4.8}$$

$$-6x + 3y - 15z = 9 \tag{30.4.9}$$

We wish to create a matrix of the following form

$$\begin{bmatrix} 1 & -3 & 1 & 4 \\ 2 & -8 & 8 & -2 \\ -6 & 3 & -15 & 9 \end{bmatrix} \tag{30.4.10}$$

$$\begin{bmatrix} 11 & 12 & 13 & 1 \\ 0 & 22 & 23 & 2 \\ 0 & 0 & 33 & 3 \end{bmatrix} \tag{30.4.11}$$

Where the terms below the direct terms are zero. We need to eliminate some of the unknowns by solving the system of simultaneous equations. To eliminate x from row 2 (where R denotes the row)

-2(R1) + R2

$$-2(x-3y+z)+(2x-8y+8z)=-10 \tag{30.4.12}$$

$$-2y+6z=-10 \tag{30.4.13}$$

so that

$$\begin{bmatrix} 1 & -3 & 1 & 4 \\ 0 & -2 & 6 & -10 \\ -6 & 3 & -15 & 9 \end{bmatrix} \tag{30.4.14}$$

To eliminate x from row 3

6(R1) + R3

$$6(x-3y+z)+(-6x-3y-15z)=33 \tag{30.4.15}$$

$$-15y-9z=33 \tag{30.4.16}$$

$$\begin{bmatrix} 1 & -3 & 1 & 4 \\ 0 & -2 & 6 & -10 \\ 0 & -15 & -9 & 33 \end{bmatrix} \tag{30.4.17}$$

To eliminate y from row 2

R2/2

$$-y+3z=-5 \tag{30.4.18}$$

$$\begin{bmatrix} 1 & -3 & 1 & 4 \\ 0 & -1 & 3 & -5 \\ 0 & -15 & -9 & 33 \end{bmatrix} \tag{30.4.19}$$

To eliminate y from row 3

R3/3

$$-5y-3z=-11 \tag{30.4.20}$$

$$\begin{bmatrix} 1 & -3 & 1 & 4 \\ 0 & -1 & 3 & -5 \\ 0 & -5 & -3 & 11 \end{bmatrix} \tag{30.4.21}$$

And then

-5(R2) + R3

$$-5(-y+3z)+(-5y-3z)=36 \tag{30.4.22}$$

$$-18z=36 \tag{30.4.23}$$

$$\begin{bmatrix} 1 & -3 & 1 & 4 \\ 0 & -1 & 3 & -5 \\ 0 & 0 & -18 & 36 \end{bmatrix} \tag{30.4.24}$$

$$-2=z \tag{30.4.25}$$

Substituting z = -2 back in to R2 gives y = -1

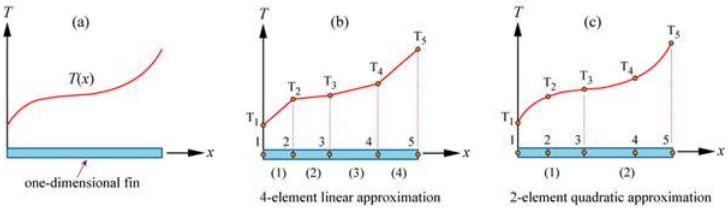
Substituting y = -1 and z = -2 back in to R1 gives $x = 3$

This process of progressively solving for the unknowns is called ***back substitution***.

This page titled 30.4: Enforcing Boundary Conditions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.5: Interpolation/Basis/Shape Functions

Consider the temperature distribution along the one-dimensional fin in Fig.1.



A one-dimensional continuous temperature distribution with an infinite number of unknowns is shown in (a). The fin is discretised in (b) – i.e. divided into 4 subdomains (or elements). The nodes are numbered consecutively from left to right, as are the elements. The elements are *first order* elements; the interpolation scheme between the nodes is therefore linear. Note that there are only 5 nodes for this system, since the internal nodes are shared between the elements. Since we are only solving for temperature, there are only 5 degrees of freedom in this model of the continuous system. It should be clear that a better approximation for *T(x)* would be obtained if the number of elements was increased (i.e. if the element lengths were reduced). It is also apparent that the nodes should be placed closer together in regions where the temperature (or any other unknown solution) changes rapidly. It is useful also to place a node wherever a step change in temperature is expected and where a numerical value of the temperature is needed. It is good practice to continue to increase the number of nodes until a converged solution is reached.

In (c), the fin has been divided into two subdomains – elements 1 and 2. However, in this instance we have chosen to use a *second order* (quadratic) element. These elements contain 'midside' nodes as shown, and the interpolation between the nodes is quadratic which permits a much closer approximation to the real system. For this model system there are still just 5 degrees of freedom. However, the analysis takes longer for (c) than it does for (b) because the quadratic interpolation (which calculates the temperature at locations between the nodes) is more demanding than the corresponding linear case.

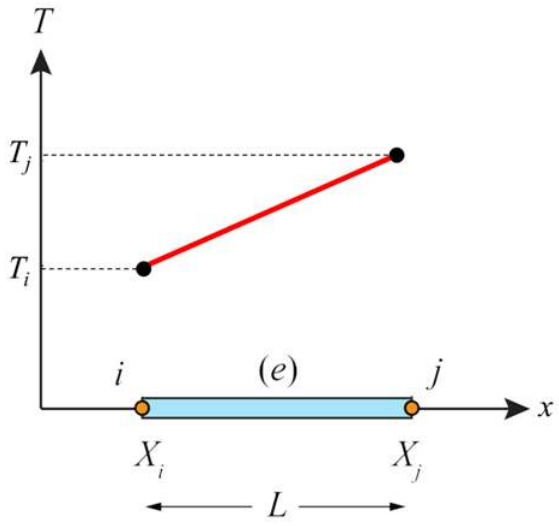
(There is often a trade-off between a high number of first order elements requiring little computation and a smaller number of second order elements requiring more heavy computation to be made, which affects both the analysis time and the solution accuracy. The choice depends to a large extent on the problem being solved.)

This page titled 30.5: Interpolation/Basis/Shape Functions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.6: 1D First Order Shape Functions

We can use (for instance) the direct stiffness method to compute degrees of freedom at the element nodes. However, we are also interested in the value of the solution at positions inside the element. To calculate values at positions other than the nodes we interpolate between the nodes using shape functions.

A one-dimensional element with length L is shown below. It has two nodes, one at each end, denoted i and j, and known nodal temperatures T_i and T_j. We can deduce automatically that the element is first order (linear) since it contains no 'midside' nodes.



One dimensional linear element with temperature degrees of freedom

We need to derive a function to compute values of the temperature at locations between the nodes. This interpolation function is called the shape function. We demonstrate its derivation for a 1-dimensional linear element here. Note that, for linear elements, the polynomial inerpolation function is first order. If the element was second order, the polynomial function would be second order (quadratic), and so on.

Since the element is first order, the temperature varies linearly between the nodes and the equation for T is:

$$T(x) = a + bx \tag{30.6.1}$$

We can therefore write:

$$T_i = a + bX_i \tag{30.6.2}$$

$$T_j = a + bX_j \tag{30.6.3}$$

which are simultaneous. To determine the coefficients a and b:

$$\frac{T_i - a}{X_i} = b \tag{30.6.4}$$

$$\frac{T_j - a}{X_j} = b \tag{30.6.5}$$

$$\frac{T_i - a}{X_i} = \frac{T_j - a}{X_j} \tag{30.6.6}$$

$$(T_i - a)X_j = (T_j - a)X_i \tag{30.6.7}$$

$$T_iX_j - aX_j = T_jX_i - aX_i \tag{30.6.8}$$

$$T_iX_j - T_jX_i = a(X_j - X_i) \tag{30.6.9}$$

$$\frac{T_i X_j - T_j X_i}{(X_j - X_i)} = a \tag{30.6.10}$$

$$\frac{T_i X_j - T_j X_i}{L} = a \tag{30.6.11}$$

and

$$T_i - bX_i = a \tag{30.6.12}$$

$$T_j - bX_j = a \tag{30.6.13}$$

$$T_i - bX_i = T_j - bX_j \tag{30.6.14}$$

$$bX_j - bX_i = T_j - T_i \tag{30.6.15}$$

$$b(X_j - X_i) = T_j - T_i \tag{30.6.16}$$

$$b = \frac{T_j - T_i}{(X_j - X_i)} \tag{30.6.17}$$

$$b = \frac{T_j - T_i}{L} \tag{30.6.18}$$

Substitution of Eqns.11 and 18 into Eqn.1 yields:

$$T(x) = \frac{T_i X_j - T_j X_i}{L} + \frac{T_j - T_i}{L} x \tag{30.6.19}$$

$$T(x) = \frac{T_i X_j - T_j X_i}{L} + \frac{T_j x - T_i x}{L} \tag{30.6.20}$$

$$T(x) = \frac{T_i X_j}{L} - \frac{T_j X_i}{L} + \frac{T_j x}{L} - \frac{T_i x}{L} \tag{30.6.21}$$

$$T(x) = \frac{T_i X_j}{L} - \frac{T_i x}{L} + \frac{T_j x}{L} - \frac{T_j X_i}{L} \tag{30.6.22}$$

$$T(x) = T_i(\frac{X_j - x}{L}) + T_j(\frac{x - X_i}{L}) \tag{30.6.23}$$

It should be clear from Eqn.23 that the nodal temperature values are multiplied by linear functions of x– the shape functions. The functions are denoted by S with a subscript to indicate the node with which a specific shape function is associated. In the case presented:

$$S_i = (\frac{X_j - x}{L}) \tag{30.6.24}$$

$$S_j = (\frac{x - X_i}{L}) \tag{30.6.25}$$

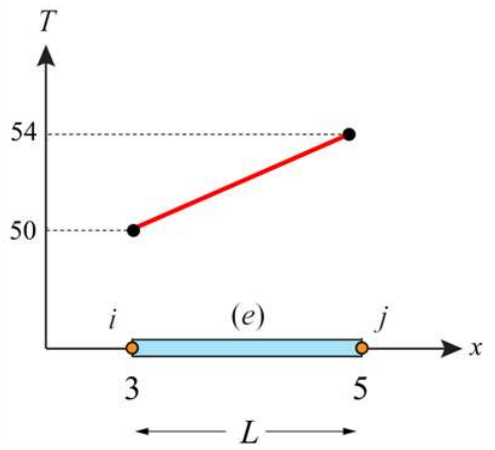
And Eqn.23 becomes

$$T(x) = S_i T_i + S_j T_j \tag{30.6.26}$$

In matrix form

$$T_x^e = [S_i \quad S_j] \begin{Bmatrix} T_i \\ T_j \end{Bmatrix} \tag{30.6.27}$$

For the case shown below, calculate T at x = 3.3.



1-dimensional linear element with known nodal temperatures and positions

$$T(x=3.3)=T_i(\frac{X_j-x}{L})+T_j(\frac{x-X_i}{L}) \tag{30.6.28}$$

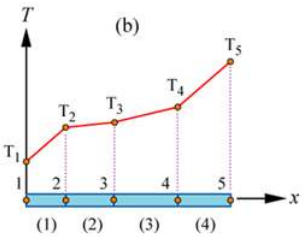
$$T(x=3.3)=50(\frac{5-3.3}{2})+54(\frac{3.3-3}{2}) \tag{30.6.29}$$

From inspection of Eqn.26 we can deduce that each shape function has a value of 1 at its own node and a value of zero at the other nodes. The sum of the shape functions sums to one. The shape functions are also first order, just as the original polynomial was. The shape functions would have been quadratic if the original polynomial had been quadratic.

A continuous, piecewise smooth equation for the one dimensional fin first shown above can be constructed by connecting the linear element equations. We know that the temperature at any point in any element can be found from the nodal temperatures T_i and the shape functions S_i. For the following system:

$$T_x^e=S_iT_i+S_jT_j \quad X_i \quad \leq x \leq X_j \tag{30.6.30}$$

	Node	
Element #	**i**	**j**
1	1	2
2	2	3
3	3	4
4	4	5



4-element linear approximation

$$T_x^1=S_1^1X_1+S_2^1X_2 \quad S_1^1=\tfrac{X_2-x}{X_2-X_1} \quad S_2^1=\tfrac{x-X_1}{X_2-X_1} \quad X_1 \quad \leq x \leq X_2 \tag{30.6.31}$$

$$T_x^2=S_2^2X_2+S_3^2X_3 \quad S_2^2=\tfrac{X_3-x}{X_3-X_2} \quad S_3^2=\tfrac{x-X_2}{X_3-X_2} \quad X_2 \quad \leq x \leq X_3 \tag{30.6.32}$$

$$T_x^3=S_3^3X_3+S_4^3X_4 \quad S_3^3=\tfrac{X_4-x}{X_4-X_3} \quad S_4^3=\tfrac{x-X_3}{X_4-X_3} \quad X_3 \quad \leq x \leq X_4 \tag{30.6.33}$$

$$T_x^4 = S_4^4 X_4 + S_5^4 X_5 \quad S_4^4 = \frac{X_5 - x}{X_5 - X_4} \quad S_4^3 = \frac{x - X_4}{X_5 - X_4} \quad X_4 \quad \leq x \leq X_5 \tag{30.6.34}$$

$$S_2^1 \neq S_2^2, S_3^2 \neq S_3^3, S_4^3 \neq S_4^4 \tag{30.6.35}$$

The temperature gradient through an individual element $\frac{dT^e}{dx}$ can be found from the derivative of Eqn.33

$$T_x^e = T_i(\frac{X_j - x}{L}) + T_j(\frac{x - X_i}{L}) \tag{30.6.36}$$

$$T_x^e = \frac{X_j T_i}{L} - \frac{x T_i}{L} + \frac{x T_j}{L} - \frac{X_i T_j}{L} \tag{30.6.37}$$

$$T_x^e = \frac{x T_j}{L} - \frac{x T_i}{L} + \frac{X_j T_i}{L} - \frac{X_i T_j}{L} \tag{30.6.38}$$

$$T_x^e = \frac{x}{L}(T_j - T_i) + \frac{1}{L}(X_j T_i - X_i T_j) \tag{30.6.39}$$

such that

$$\frac{dT_x^e}{dx} = \frac{T_j - T_i}{L} \tag{30.6.40}$$

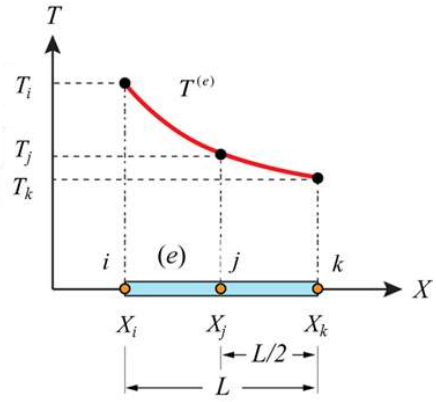
We can check that our shape functions are correct by knowing that the shape function derivatives sum to zero.

This page titled 30.6: 1D First Order Shape Functions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.7: 1D Second Order Shapes and Functions

A one-dimensional quadratic element is shown in Fig.4. We can deduce immediately that the element order is greater than one because the interpolation between the nodes in non-linear. We can determine from inspection that the element is quadratic (second order) because there's a 'midside' node. We know therefore that the function approximating the solution is a second order polynomial:

$$T_x^e = a + bx + cx^2 \tag{30.7.1}$$



The shape functions S_i cam be determined bu solving Eqn. 1 using known T_i at known X_i to give:

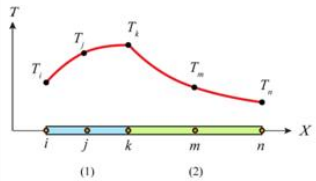
$$T_x^e = S_i T_i + S_j T_j + S_k T kj \tag{30.7.2}$$

$$S_i = \frac{2}{L^2}(x - X_k)(x - X_j) \tag{30.7.3}$$

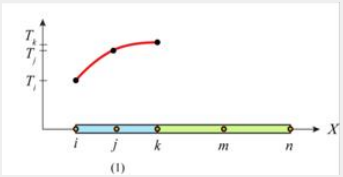
$$S_j = \frac{-4}{L^2}(x - X_i)(x - X_k) \tag{30.7.4}$$

$$S_k = \frac{2}{L^2}(x - X_i)(x - X_j) \tag{30.7.5}$$

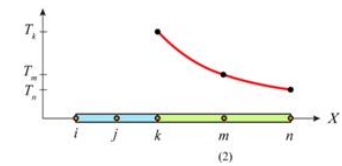
Using the quadratic shape functions for a single element (Eqns.3-4), we can assemble a corresponding set of equations for a larger system:



	Nodes		
Element #	***i***	***j***	***k***
1	*i*	*j*	*k*
	Nodes		
Element #	***i***	***j***	***k***
2	*k*	*m*	*n*



$$T_x^1 = \begin{bmatrix} S_i^1 & S_k^1 & S_j^1 \end{bmatrix} \begin{Bmatrix} T_i \\ T_k \\ T_j \end{Bmatrix} \tag{30.7.6}$$



$$T_x^2 = \begin{bmatrix} S_k^2 & S_m^2 & S_n^2 \end{bmatrix} \begin{Bmatrix} T_k \\ T_m \\ T_n \end{Bmatrix} \tag{30.7.7}$$

$$T_x^1 = S_i^1 T_i + S_j^1 T_j + S_k^1 T_k \tag{30.7.8}$$

$$S_i^1 = \frac{2}{L^2}(x - X_k)(x - X_j) \tag{30.7.9}$$

$$S_j^1 = \frac{-4}{L^2}(x - X_i)(x - X_k) \tag{30.7.10}$$

$$S_k^1 = \frac{2}{L^2}(x - X_i)(x - X_j) \tag{30.7.11}$$

$$T_x^2 = S_k^2 T_k + S_m^2 T_m + S_n^2 T_n \tag{30.7.12}$$

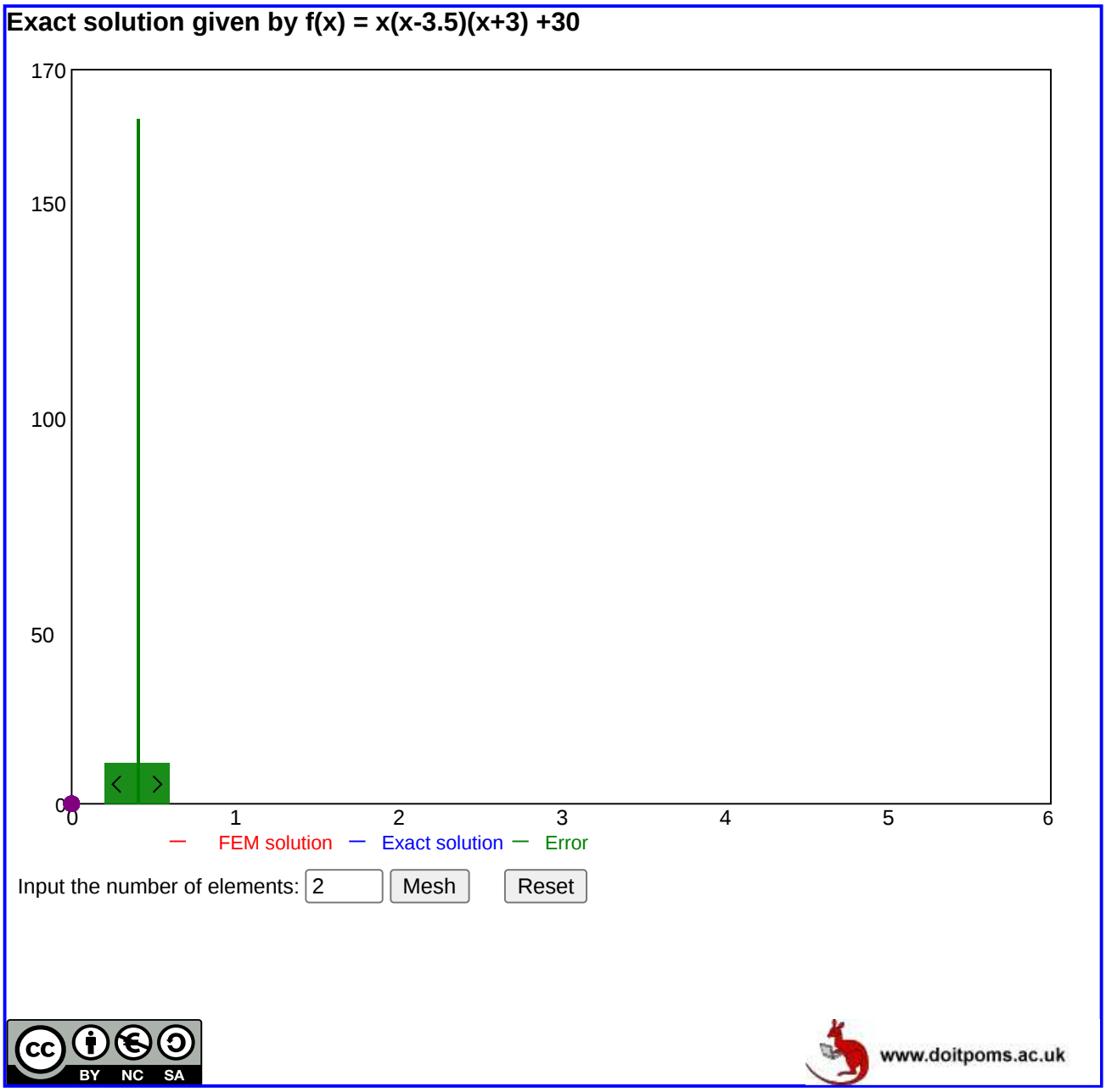
$$S_k^2 = \frac{2}{L^2}(x - X_n)(x - X_m) \tag{30.7.13}$$

$$S_m^2 = \frac{-4}{L^2}(x - X_k)(x - X_n) \tag{30.7.14}$$

$$S_n^2 = \frac{2}{L^2}(x - X_k)(x - X_m) \tag{30.7.15}$$

The element shape functions are stored within the element in commercial FE codes. The positions X_i are generated (and stored) when the mesh is created. Once the nodal degrees of freedom are known, the solution at any point between the nodes can be calculated using the (stored) element shape functions and the (known) nodal positions.

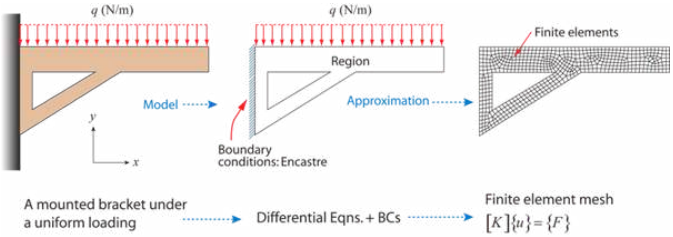
The order of the element and the number of elements in your geometric domain can have a strong effect on the accuracy of the solution. This is demonstrated in the following application which demonstrates how the number of elements (mesh density) can affect the accuracy of finite element model predictions. It compares the exact solution to the equation shown (which has an analytical solution) to predictions using the finite element method. In the application you can discretise (mesh) the "domain" using any number of elements between 1 and 20. The shape functions used are first order ones. The error between the exact solution and the FEM-predicted solution can be found by dragging the green cursor left and right through the domain.



This page titled 30.7: 1D Second Order Shapes and Functions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.8: Typical steps during FEM modelling

Consider a wall mounted bracket loaded uniformly along its length as in the figure below:

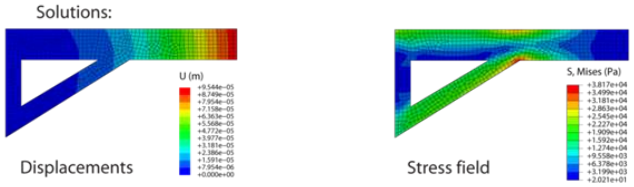


Wall mounted bracket

The geometry (field) is defined for us and is (relatively) complex. The boundary conditions are also defined and are:

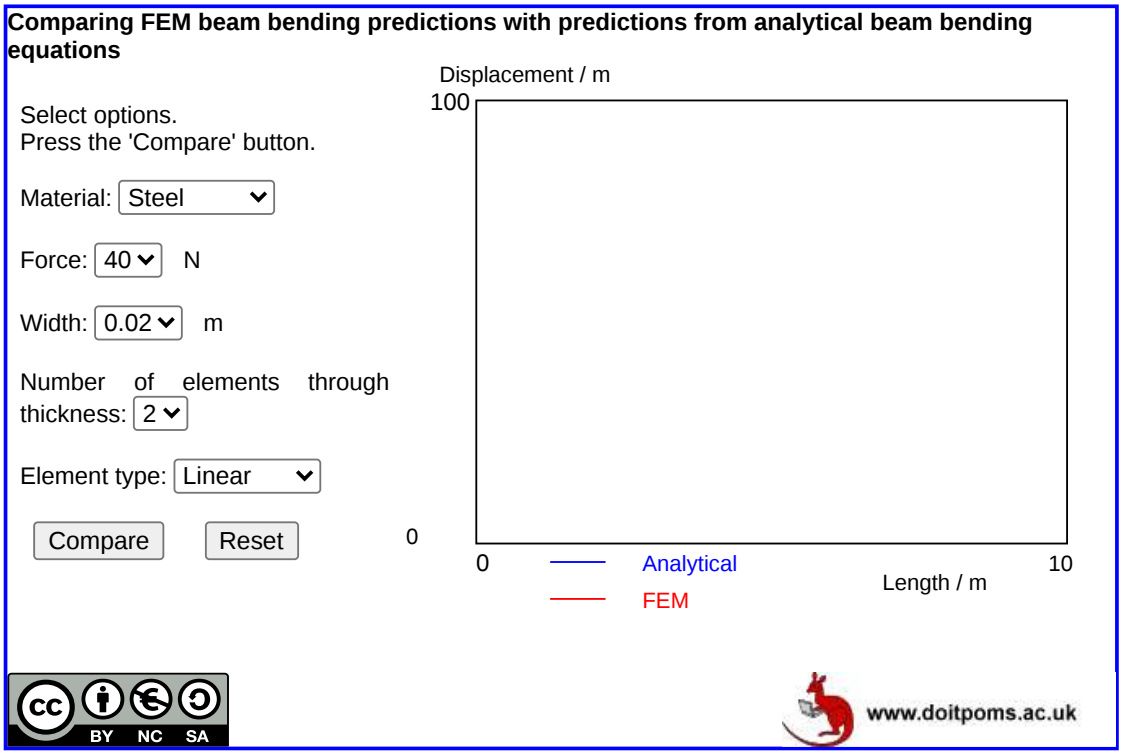
- A uniform force per unit length (Neumann condition) along the upper edge
- Fixed x and y displacements along the clamped edge (Dirichlet condition)

It is apparent that the bracket will respond mechanically under the action of the applied load and a system of internal stresses will develop (to balance the applied load). To calculate the stresses that develop we must first mesh (discretise) the domain, assemble the global stiffness matrix [K], and then determine the nodal displacements {u} and resultant forces {F} using some iterative numerical technique (Gauss elimination, for instance). It is then a relatively trivial exercise to compute the nodal stresses from the nodal displacements and to find the solution between the nodes and within the elements using shape functions.



However, it is important to be aware that certain combinations of the specified number of elements (mesh density) and the specified element order can give rise to solutions that are highly inaccurate. It is highly advisable (and good practice) to perform a mesh sensitivity study, whereby the effect on the solution of successively finer meshes is analysed in order to eliminate any mesh sensitivity. The following application demonstrates this point:

The application below concerns the deflection of a CANTILEVER beam loaded at its end with an applied force. The width of the beam can be altered, although the beam length and beam depth are fixed constant. The beam can be made of either steel or aluminium and it can be loaded with a force of either 40 N or 80 N. The beam can then be meshed with either 2 elements through its thickness or 4 elements through its thickness. The elements can be either first or second order. For any combination chosen, a prediction of the beam deflection based on finite element calculations will be compared to predictions from (analytical) ordinary beam bending equations. It should become apparent that the error between the two decreases as the mesh density and element order are increased.



This page titled 30.8: Typical steps during FEM modelling is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.9: Summary

- You should now understand the basics of the finite element method
- You should be able to build a global stiffness matrix for a combination of 1-dimensional springs
- You should understand the importance of boundary conditions (including types of boundary conditions)
- You should be familiar with first and second order shape functions and you should be able to use these functions to calculate the value of a solution between nodes.
- You should understand and appreciate that the finite element method provides approximate solutions only, and that the accuracy of your solution will depend on element type and mesh density.

This page titled 30.9: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

30.10: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which is true? For second order elements:

a	The interpolation between the element nodes is linear
b	The shape function derivatives sum to 1
c	The solution between the nodes can be approximated using a second order polynomial
d	The solution between the nodes has no dependence on the solution at the nodes

Answer

C

Which is true? For first order elements:

a	The shape functions sum to 0
b	The shape function derivatives sum to 1
c	Field variables vary linearly between the element nodes
d	The elements contain midside nodes

Answer

C

Which is true? The error between an exact and finite element solution will always be reduced by:

a	Decreasing the element density
b	Increasing the element density
c	Increasing the element order
d	Decreasing the element order

Answer

B

If your finite element mesh contains 16 nodes, how many degrees of freedom are there in a coupled thermomechanical simulation:

a	64
b	48
c	24
d	16

Answer

A

In a structural mechanics analysis, what type of boundary condition is an applied pressure:

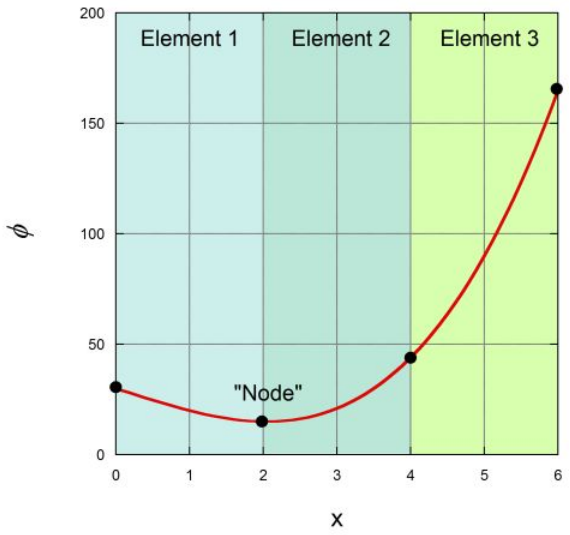
a	Dirichlet
b	Neumann
c	Robin
d	Dirichlet + Neumann

Answer

B

Discretise the following function using three equal length elements between$0 \leq x \leq 6$. Assume the elements are linear (first order), and calculate $\phi(x = 3.2)$ using the finite element method. Compare your answer to the exact solution.

$$\phi = x(x - 3.5)(x + 3) + 30 \tag{30.10.1}$$



hint:

Answer

FEM Solution (3 first order elements) = 32.4, Exact Solution = 24.05, Error = 34%

Discretise the same function using six equal length elements and find $\phi(x = 3.2)$ using the finite element method. Compare your answer to the exact solution and to the answer obtained using a three element discretisation.

Answer

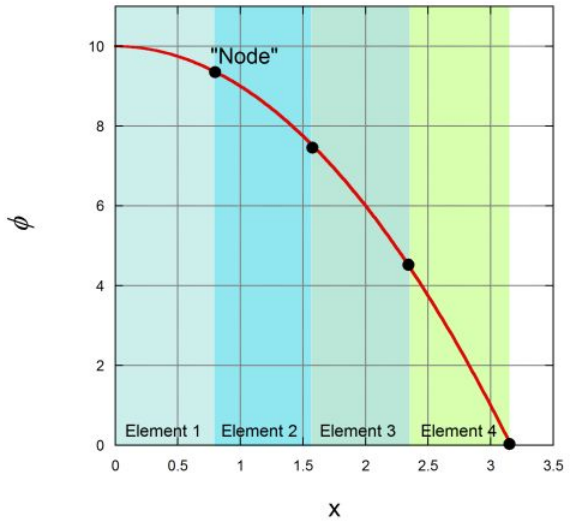
FEM Solution (6 first order elements) = ,25.6, Exact Solution = 24.05, Error = 6.4%

Discretise the same function using three equal length but QUADRATIC elements. Calculate $\phi(x = 3.2)$ and compare your answer to the ones obtained previously.

Answer

FEM Solution (3 second order elements) = 24.24, Exact Solution = 24.05, Error = 0.8%

Using an equal length, 4-element discretisation of $f(x) = 10 - x^2$, calculate $f(x = 0.6)$ and the error between the finite element and exact solutions.

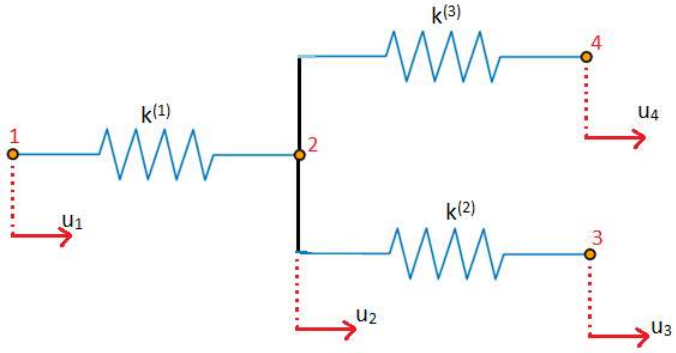


hint:

Answer

FEM Solution (4 first order elements) = 9.525, Exact Solution = 9.64, Error = 1.2%

For the system of springs below, determine the global stiffness matrix.



Answer

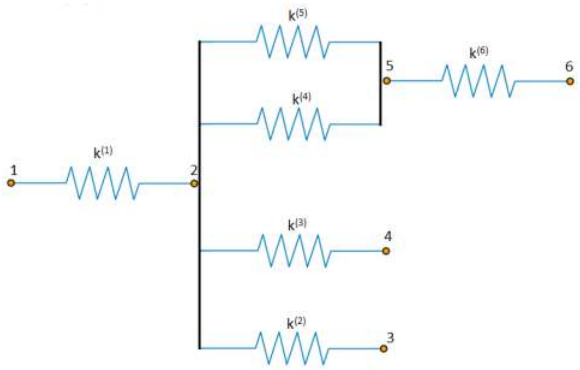
$$K = \begin{pmatrix} k1 & -k1 & 0 & 0 \\ -k1 & k1+k2+k3 & -k2 & -k3 \\ 0 & -k2 & k2 & 0 \\ 0 & -k3 & 0 & k3 \end{pmatrix} \tag{30.10.2}$$

A system of 1-dimensional springs has the following global stiffness matrix. Draw the system of springs.

$$\begin{pmatrix} k1 & -k1 & 0 & 0 & 0 & 0 \\ -k1 & k1+k2+k3+k4+k5 & -k2 & -k3 & -(k4+k5) & 0 \\ 0 & -k2 & k2 & 0 & 0 & 0 \\ 0 & -k3 & 0 & k3 & 0 & 0 \\ 0 & -(k4+k5) & 0 & 0 & k4+k5+k6 & -k6 \\ 0 & 0 & 0 & 0 & -k6 & k6 \end{pmatrix} \tag{30.10.3}$$

Answer

https://eng.libretexts.org/@go/page/32956



This page titled 30.10: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

31: Fuel Cells

https://www.doitpoms.ac.uk/tlplib/

Learning Objectives

On completion of this TLP you should:

- The principle behind fuel cells and what constitutes a fuel cell.
- A brief history of the technology.
- Function and specifics of four main types of fuel cell:
  - Solid oxide fuel cell (SOFC)
  - Molten carbonate fuel cell (MCFC)
  - Proton exchange membrane/ polymer electrolyte membrane fuel cell (PEMFC)
  - Direct methanol fuel cell (DMFC)
- The advantages, problems and applications associated with each of the different electrolytes commonly used in fuel cells, including discussion of the factors influencing their suitability.
- The electrochemical explanation of the fuel cell processes.
- The requirements of fuelling a fuel cell system including “balance of plant” considerations, reforming and ways of storing the fuel.
- Building a simple fuel cell.

Before you start

- You should be familiar with the principles of electrochemical reactions
- You should understand the principles of thermodynamics

31.1: Introduction
31.2: The Principle
31.22: Questions
31.4: History of The Technology
31.4: Types of fuel cells
31.5: Solid Oxide Fuel Cells (SOFCs)
31.6: Electrolyte
31.7: Electrode Materials
31.8: Interconnection
31.9: Thermodynamics
31.10: System and Outlook
31.11: Molten carbonate fuel cells (MCFCs)
31.12: Electrolyte
31.13: Electrochemistry
31.14: Electrodes
31.15: Proton Exchange Membran Fuel Cells (PEMFCs)
31.16: Membrane
31.17: Electrodes and Membrane-Electrode Assembly
31.18: Efficiency and Reaction Conditions
31.19: Fuelling Requirements

31.20: Case Study: Fuelling Practicalities
31.21: Summary

This page titled 31: Fuel Cells is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.1: Introduction

Conventional power plants convert chemical energy into electrical energy in three steps:

- Production of heat by burning fuel
- Conversion of heat into mechanical energy
- Conversion of mechanical energy into electrical energy

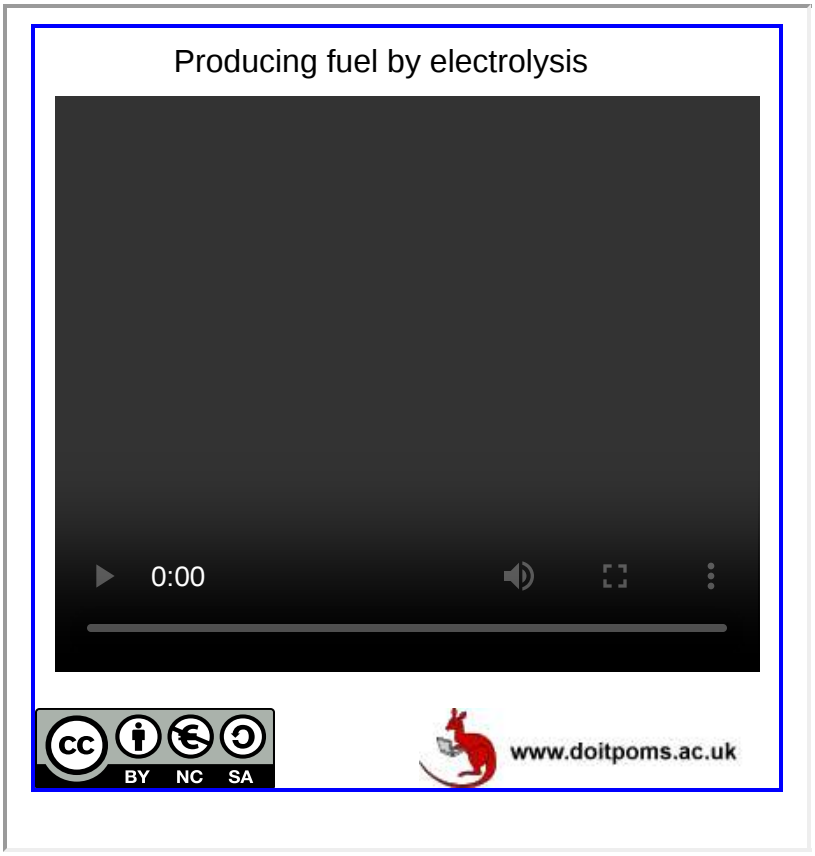
The efficiency of the second step is limited (by the Second Law of Thermodynamics) to the Carnot efficiency, since the conversion of heat into mechanical energy occurs in a closed-cycle heat engine. An efficiency of about 41% can be reached by modern systems. A fuel cell is an electrochemical device that converts the chemical energy in fuels (e.g. hydrogen, methane, butane or even gasoline and diesel) into electrical energy. It exploits the natural tendency of oxygen and hydrogen to react to form water. The direct reaction is prevented by the electrolyte, which separates the two reactants. Therefore two half-reactions occur at the electrodes:

- Anode: Fuel (e.g. H_2, CO, CH_4) is oxidised
- Cathode: Oxygen is reduced

The ions are transported to the other electrode through the electrolyte. The fuel cell contains no moving parts and only four active elements: cathode, anode, electrolyte and interconnect; it is a simple and robust system. Fuel cells have a number of advantages compared to conventional electricity generation:

- Negligible air pollution (if fossil fuels are used, otherwise none)
- Reduced weight, especially in mobile applications
- 100% theoretical efficiency, 80% efficiency in high temperature turbine hybrid systems, that can use the generated heat
  High efficiency in low power systems
- Constant efficiency at low load
- Flexible output with fast adjustment
- Low maintenance cost and very few moving parts (or none)
- Quiet or completely silent

Fuel cells have many interesting applications. This short video shows a demonstration fuel cell car. Note that hydrogen and oxygen being used up by the reactions.



This page titled 31.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.2: The Principle

Oxygen and hydrogen, when mixed together in the presence of enough activation energy have a natural tendency to react and form water, because the Gibbs free energy of H_2O is smaller than that of the sum of H_2 and $½O_2$ added together (Hence, we don't smoke our pipes on Zeppelins!). If hydrogen and oxygen were combined directly, we would see combustion:

$H_2 + ½O_2 \rightarrow H_2O$

Combustion involves the direct reaction of H_2 gas with O_2. The hydrogen donates electrons to the oxygen . We say that the oxygen has been reduced and the fuel oxidised. This combustion reaction releases heat energy.

The fuel cell separates hydrogen and oxygen with a gas-impermeable electrolyte through which only ions (e.g. H^+, O^{2-}, CO_3^{2-}) can migrate. Hence two half reactions occur at the two electrodes. The type of reactions at the electrodes is determined by the type of electrolyte.

Grove's fuel cell is one of the simplest examples.

The half-reaction at the anode:
$H_2 \rightarrow 2H^+ + 2e^-$
The half-reaction at the cathode:
$O_2 + 4e^- + 4H^+ \rightarrow 2H_2O$
The net reaction is the combustion reaction:
$H_2 + ½O_2 \rightarrow H_2O$

Activation polarization is caused by the energy intensive activity of the making and breaking of chemical bonds: At the anode, the hydrogen molecules enter the reaction sites so that they are broken into ions and electrons. The resulting ions form bonds with the catalyst atoms and the electrons remain in the vicinity until new hydrogen molecules start bonding with the catalyst, breaking the bond between the earlier ion. The electrons migrate through the bipolar plate if the bonding energy of the ion is low enough and the ions diffuse through the electrolyte. A similar process occurs at the cathode: Oxygen molecules are broken up and react with the electrons from the anode and the protons that diffused through the electrolyte to form water. Water is then ejected as a waste product and the fuel cell runs (can supply a current), as long as fuel and oxygen is provided.

The exact reactions at the electrodes depend upon which species can be transported across the electrolyte. Fuel cells are classified according to the type of electrolyte (see Types of Fuel Cells). The most common electrolytes are permeable for protons and the reactions are as discussed above. The second most common electrolytes, found in solid oxide fuel cells (SOFCs), are permeable for oxide ions and the following half-reactions occur:

The half-reaction at the anode:
$H_2 + O^{2-} \rightarrow H_2O + 2e^-$
The half-reaction at the cathode:
$O_2 + 4e^- \rightarrow 2O^{2-}$

The net reaction is the same as before:
$H_2 + ½O_2 \rightarrow H_2O$

A third type of electrolyte, used for molten carbonate fuel cells at high temperatures conducts carbide ions (CO_3^{2-}):

The half-reaction at the anode:
$H_2 + CO_3^{2-} \rightarrow H_2O + CO_2 + 2e^-$
The half-reaction at the cathode:
$½O_2 + CO_2 + 2e^- \rightarrow CO_3^{2-}$
The net reaction is the combustion reaction:
$H_2 + ½O_2 \rightarrow H_2O$

We also commonly see alkaline electrolytes, across which OH^- is the transported species. In this case the half-reactions would be:

The half-reaction at the anode:
$H_2 + 2OH^- \rightarrow 2H_2O + 2e^-$
The half-reaction at the cathode:
$O_2 + 4e^- + 2H_2O \rightarrow 4OH^-$
The net reaction is the combustion reaction:
$H_2 + ½O_2 \rightarrow H_2O$

Building a Simple Fuel Cell - Jut Like Billy Grove Did

Collect the necessary components:

This is the most challenging part of the project, due to the difficulty many students will have in procuring the two lengths of platinum (Pt) wire required. Persist.

- Two short lengths of platinum wire – 2 cm a piece is sufficient
- A beaker
- Weak NaCl solution (table salt and tap water will suffice)
- Crocodile clips ×2
- Wires ×2
- Voltmeter
- DC power source, e.g. 6-volt battery

Step by step:

1. Fill beaker with salt solution
2. Clamp the ends of the Pt wires in the crocodile clips
3. Place the Pt wires in the solution in such a way that they will not move and so that only the Pt is in contact with the water. Plasticine (or Blu-Tac) would be helpful at this point. If the crocodile clips are in contact with the solution they will corrode quickly and taint the water.
4. Verify that the voltage across the two electrodes is zero (two wires made of the same metal have no galvanic difference between them and hence no voltage will be seen).
5. Connect each Pt wire to one of the battery's terminals and watch the gas bubbles from the electrodes.
6. Verify that the voltage across the electrodes is now equal to that of the battery, i.e. 6 V.
7. Disconnect the battery and be careful not to knock the beaker.
8. Verify that a voltage remains across the two electrodes.

This remaining voltage is due to bubbles of chlorine gas adhering to the cathode and to bubbles of hydrogen gas adhering to the anode. HCl is produced as the two atoms combine, via H^+ ions travelling through the solution from anode to cathode, to produce HCl.

Things to consider:

An ammeter will confirm that we only obtain a tiny current in the above experiment (although the voltage remains set to the value we can calculate using the Nernst equation). Factors limiting the current in this case are:

- the small surface area of the electrode - small amount of gas
- wide spacing of the electrodes and electrolyte's ionic resistance.

In production cells, the electrodes are made larger and flat so as to maximise contact area, and the thickness of the electrolyte is kept to a minimum so as to reduce resistance.

This page titled 31.2: The Principle is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.22: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

For each type of fuel cell below, what are the charge carrying species (electrolyte type)?

(a) Polymeric exchange membrane

(b) Immobilised alkaline solution

(c) Phosphoric acid

(d) Molten carbonate

(e) Ceramic solid oxide

Answer

(a) H^+

(b) OH^-

(c) H^+

(d) CO_3^{2-}

(e) O^{2-}

Which cells are able to internally reform?

Answer

High temperature, solid oxide or molten carbonate cells.

The efficiency limit of H_2 cells reduces as operating temperature is increased. Why are high efficiency systems often run at very high temperatures?

Answer

The heat produced can be more useful than the efficiency gained at lower temperatures. Running a system at a higher temperature allows the hot exhaust gasses to be utilised in a turbine-hybrid system, raising the overall efficiency of the system up as high as 80% or more.

Which of the following statements are TRUE about the development of fuel cells?

(a) Fuel cells have been around since the 19th century.

(b) A fuel cell tractor was built in 1959.

(c) PEM fuel cells are so expensive because of the platinum needed as a catalyst.

(d) The first fuel cells were developed to run on hydrogen.

(e) Fuel cell cars are at least a decade away.

Answer

(a) True. The phenomenon was discovered in 1800 and the first attempts to engineer a fuel cell were in 1889.

(b) True.

(c) False. The amount of platinum used in a fuel cell is actually small compared to the cost of the bipolar plates; typically about £10 for a 1 kW stack.

(d) False. The first cells were developed to run on coal gas. It was never successful however. The first successful cell was run on pure hydrogen and oxygen by Bacon in 1932.

(e) False. Honda are already marketing a fuel cell powered automobile.

Give two reasons why we don't all use fuel cells in our cars already?

Answer

(i) High temperature fuel cells are not easily portable and low temperature ones require hydrogen as the fuel in order to produce any useful power. Hydrogen is not yet a readily available fuel. It is also difficult to store.

(ii) Fuel cells also suffer from low reaction rates and therefore low power output.

Why is nickel used as the anode in SOFCs? How is it treated in order to aid bonding to YSZ?

Answer

It is cheap and performs well but flakes off unless mixed with zirconia.

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

Give as many ways as you can think of to provide a mobile fuel cell with pure hydrogen.

Answer

(i) Pure H_2 may be stored as used either under pressure, cryogenically, in a metal hydride, in a high surface area carbon nano-tube arrangement or even in glass nano-spheres.

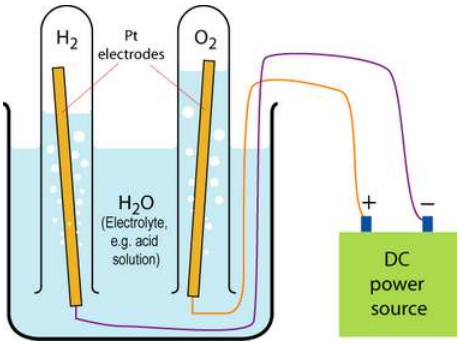
(ii) Many fuels may be carried and reformed before use e.g. alcohols, alkali metal hydrides, ammonia, sodium borohydride, methane and even hydrazine.

(iii) Alternately the cell could be run directly on methanol.

This page titled 31.22: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.4: History of The Technology

The fuel cell concept was first demonstrated by William R. Grove, a British physicist, in 1839. The cell he demonstrated was very simple, probably resembling this:



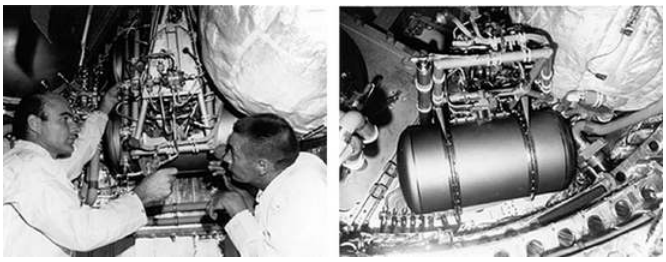
Electrolysis setup

By application of a voltage across the two electrodes, hydrogen and oxygen could be extracted (the process is called **electrolysis**) and captured as shown (William Nicholson first discovered this in 1800). The fuel cell, or "gas battery" as it was first known, is the reverse of this process. In the presence of platinum electrodes, which are necessary as catalysts, the electrolysis will essentially run in reverse and current can be made to flow through a circuit between the two electrodes.

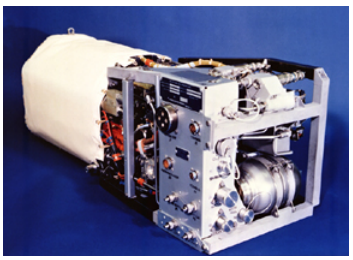
Nobody tried to make use of the concept demonstrated by William R Grove until 1889 when Langer and Mond tried to engineer a practical cell fuelled by coal gas. Further early attempts carried on into the early 1900's but the development of the internal combustion engine made further research into the technology sadly unnecessary.

Francis Bacon developed the first successful fuel cell in 1932, running on pure O_2 and H_2 and using an alkaline catalyst and nickel electrodes. It was not until 1959 that Bacon and his colleagues first demonstrated a 5 kW device; the 27 year delay is perhaps an indication of just how difficult it is to make progress in this field of development. Harry Karl Ihrig demonstrated a 20 bhp fuel cell tractor in the same year.

Around about this time, NASA started researching the technology with a view to produce a compact electricity generator for use on spacecraft. Due to their astronomical budget, it was not long before they got the job done. The Gemini program used early PEM fuel cells (PEMFCs) in its later missions, and the Apollo program employed alkaline fuel cells. On a spacecraft the water produced by the reaction was available for the spacemen to drink. NASA continued to use alkaline cells in the space shuttle until the 90's when PEMFC development meant a switch back to PEMs was considered a possibility, however, the high cost of design, development, test and evaluation prevented the switch, in spite of several technical advantages.



PEM fuel cells being installed in a Gemini 7 spacecraft (Source: Smithsonian Institution, from the Science Service Historical Images Collection, courtesy of General Electric)



The alkaline fuel cell system as used on the space shuttles. Three such modules were installed in each shuttle

Recent developments are thick and fast as the technology begins to come to fruition. Automotive applications are high on the agenda due to the huge consumer market and the need for an environmentally friendly, renewable alternative to the internal combustion engine and fossil fuels.



Honda fuel cell car

This page titled 31.4: History of The Technology is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.4: Types of fuel cells

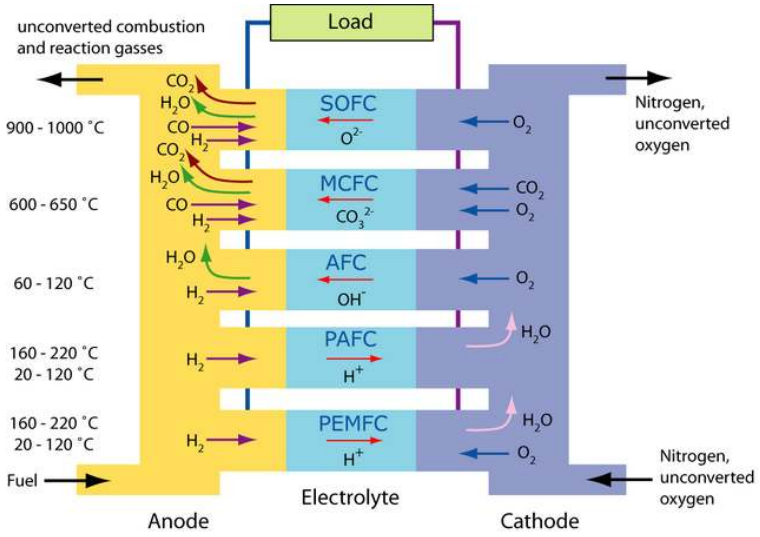
Fuel cells are categorised according to their type of electrolyte, since it is the property-determining component. The six main types of fuel cells are outlined below.

Fuel cell type	**DMFC**	**PEMFC**	**AFC**	**PAFC**	**MCFC**	**SOFC**
Electrolyte type	Polymeric ion exchange membrane	Polymeric ion exchange membrane	Immobilised alkaline salt solution	Immobilised liquid phosphoric acid	Immobilised liquid molten carbonate	Ceramic
Operating temperature (°C)	20 – 90	30 – 100	50 – 200	~220	~650	500 – 1000
Charge carrier	H^+	H^+	OH^-	H^+	H_3^{2-}	O^{2-}
Power range (W)	1 – 100	1 – 100k	500 – 10k	10k – 1M	100k – 10M+	1k – 10M+
Applications and main advantages:						
Portable electronics	Higher energy density than batteries and faster recharge.					
Cars boats and spaceships		Zero emissions and higher efficiency.				
Domestic CHP			Efficiency and reliability			
Distributed power generation, CHP, busses					Efficiency, emissions and less noise	
Able to internally reform CH_4 (see Fuelling section)	×	×	×	×	√	√

The phosphoric acid fuel cell (PAFC) is not covered in detail in this package. It was however the first type of fuel cell to be commercially produced and enjoys widespread terrestrial use. Many 200 kW systems are in place in USA and Europe.

This package also doesn't cover the alkaline fuel cell (AFC) in any detail. These particular cells use potassium hydroxide solution as the electrolyte. This means that any CO_2 at the cathode, even the levels present in the air, will react with the OH^- in the solution to produce carbonates and prevent the cell functioning. This isn't a huge problem in spacecraft, where pure oxygen can be supplied to the cathode reliably, but this characteristic flaw makes the AFC unsuitable for practical terrestrial use.

The diagram below shows the mechanisms by which the different fuel cell types operate:



Other types of fuel cells?

What's the catch?

As fossil fuel resources become more and more pressed upon to deliver the worlds energy needs, as CO_2 and global warming loom ever nearer and as cities become ever increasingly crowded with polluting automobiles the fuel cell seems to offer a golden solution to the world's energy problems. It's efficient, it's clean, hydrogen can be produced by renewable energy and the technology wouldn't require any huge change in our way of life.

So why don't we all drive fuel cell cars already? The technology has two fundamental flaws:

- Slow reaction rate, leading to low currents and power.
- Hydrogen is not a readily available, or easily stored fuel.

We'll discuss ways of getting around these problems in the package. Each type of fuel cell has a different solution, but also brings its own set of difficulties.

Temperature differences

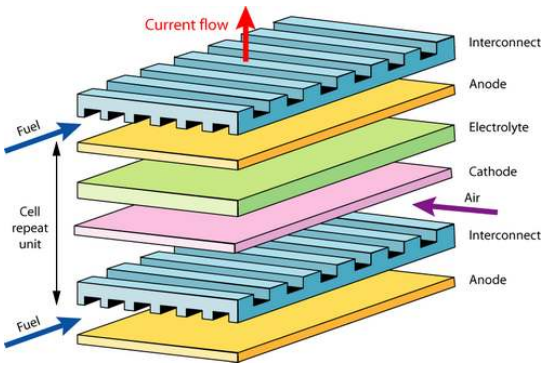
High temperature cells (solid oxide and molten carbonate electrolytes) operate by very different mechanisms to low temperature cells, and have different applications accordingly. The requirements of the "balance of plant" (i.e. the additional fuel processing equipment necessary to fuel a fuel cell) are also different. We therefore split the TLP in two and consider high temperature cells separately from low temperature cells.

This page titled 31.4: Types of fuel cells is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.5: Solid Oxide Fuel Cells (SOFCs)

High temperature cells

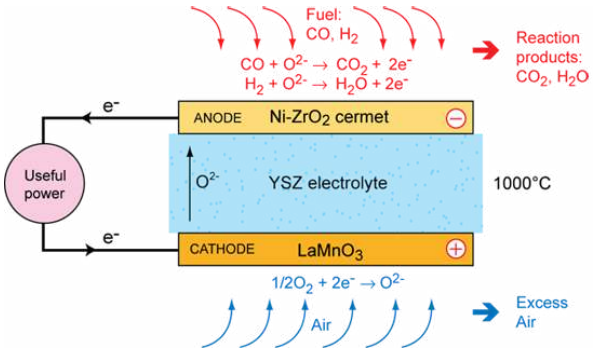
In the late nineteenth century, conduction was not yet understood. Later, Nernst observed at the University of Göttingen, that stabilized zirconia (ZrO_2 doped with Ca, Mg, Y) was an insulator at room temperature, an ionic conductor from 600–1000 °C and a mixed conductor (both electronic and ionic) at around 1500 °C. The main part of the solid oxide fuel cell was therefore discovered. The fuel cell concept was demonstrated by Baur and Preis in the 1930's using zirconium oxide, but many improvements were necessary to make a competitive device. In the 1950's, simple, straightforward design made cheaper manufacturing processes possible: the flat plate fuel cell.



Flat plate solid oxide fuel cell

There are a few problems with the flat plate design when used for larger devices: sealing, around the edges, thermal expansion mismatch and cracking (intrinsically brittle ceramics are used). Tubular designs have been developed to solve these problems (see animation below).

SOFCs are the most efficient devices yet invented, that can convert chemical energy into electrical energy. Both electrodes (cathode and anode) and the electrolyte are made of ceramic materials, since the high operating temperature prevent the use of cheaper metals. The big advantage of the SOFC over the MCFC is that the electrolyte is solid and there are no pumps required to circulate the hot electrolyte. The anode contains nickel, for better electron conduction and catalysis. The operating temperature is between 600 and 1000 °C, depending on the generation of the fuel cell (first, second and third, with decreasing operating temperature). However, thermal cycling can cause cracking of the brittle ceramic components. Both hydrogen and carbon monoxide serve as fuels. Common hydrocarbon fuels can be used in SOFC (diesel, natural gas, gasoline, alcohol etc).



Operation of a SOFC

The operation of the solid oxide fuel cell is straightforward: oxygen atoms are reduced on the porous cathode surface by electrons. The oxide ions diffuse through the electrolyte to the fuel rich and porous anode, where they react with the fuel (hydrogen) and give off electrons to an external circuit. A large amount of heat is produced by the electrochemical reaction, which can be used by an integrated heat management system. Since it takes a long time to reach its operating temperature, the best applications for SOFCs are ones that use both the heat and electricity generated:

stationary power plants, auxiliary power supplies. Start-up time problems could be solved by using supercapacitor batteries for the first few minutes of operation in mobile applications.

This page titled 31.5: Solid Oxide Fuel Cells (SOFCs) is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.6: Electrolyte

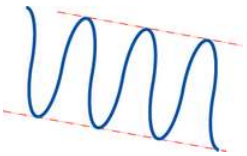
This section of the TLP describes the fundamental properties of the most common materials used as electrolytes in SOFCs.

There are several criteria that the electrolyte has to meet. It must be:

- Dense and leak tight
- Stable in reducing and oxidising environments
- A good ionic conductor at operating temperatures
- Non-electron conductor
- Thin to reduce ionic resistance
- Extended in area for maximum current capacity
- Thermal shock resistant
- Economically processable

The materials used, are solid, ion-conducting ceramics. There are two main groups of such ion conductors: fluorite structured and perovskite structured, besides new materials such as hexagonal structured oxides. The three most common electrolyte materials are: doped ceria (CeO_2), doped lanthanum gallate ($LaGaO_3$) (both are oxygen ion conductors) and doped barium zirconate ($BaZrO_3$) (a proton conductor). The concentration and type (ionic radius) of the dopants influence the material properties strongly. Dopants, that cause the least strain and hence the least influence on the potential energy landscape of the parent lattice, have the biggest effect on the conductivity.

In an oxide ion conductor, current flows by the movement of oxide ions through the crystal lattice. This is a thermally activated process, where the ions hop from one lattice site to the other (from one potential valley to the other) in a random way. When an electric field is applied, there is a drift in one direction superimposed on the random thermal motion.



Potential energy in an electric field in a periodic crystal

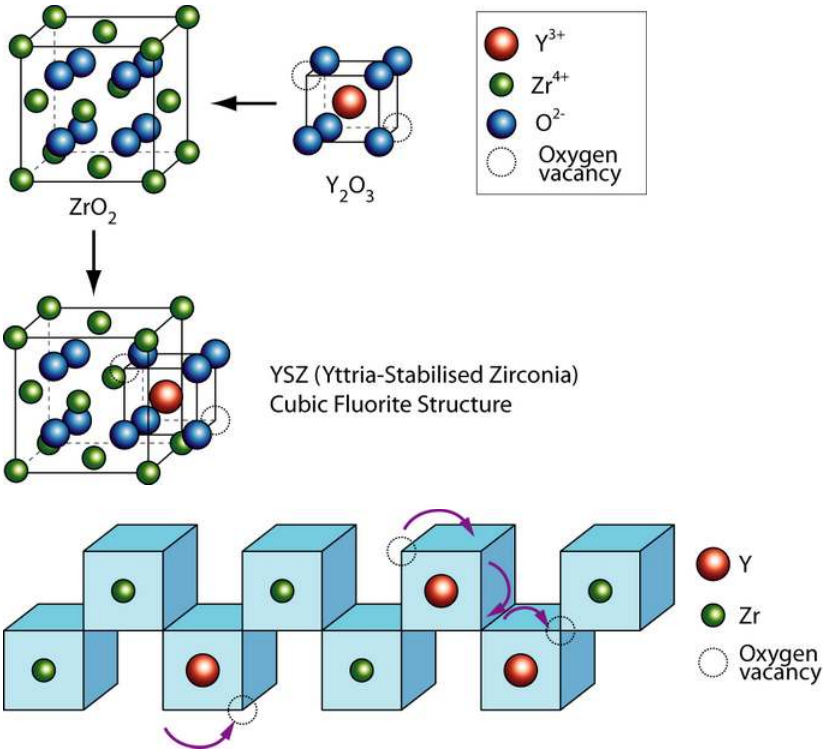
Ionic conduction depends on the mobility of the ions and therefore on temperature. At high temperatures, the conductivity can reach 1 S cm^{-1}, which is of the same order of magnitude as for liquid electrolytes. The crystal has to contain unoccupied sites that are equivalent to the occupied sites by lattice oxygen ions. The energy barrier for migration from an occupied site to an unoccupied site must be small (≤1 eV). This might seem unusual since the relative size of the oxygen ions is big and it seems more likely that the smaller metal ions migrate in an electric field. That is why there are only a few special structures that make oxygen ion migration possible: fluorite structured oxides, perovskites, LAMOX family and BIMEVOXes.

Fluorite oxides are the most common and classical oxygen ion conducting materials. The crystal structure consists of a cubic oxygen lattice with alternate body centres occupied by eight coordinated cations. The cations are arranged into a face centred cubic structure with the anions occupying the tetrahedral sites. This leaves a rather open structure with large octahedral interstitial void.

https://www.doitpoms.ac.uk/tlplib/fu...s/Zirconia.mp4

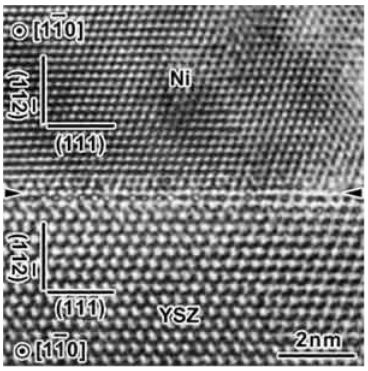
Rotating zirconia lattice

The general formula has the form AO_2, where A is usually a big tetravalent cation, e.g. U, Th, Ce. Since Zr^{4+} is too small to sustain the fluorite structure at low temperatures, it has to be partly substituted with a larger cation, called dopant. Doping involves usually substituting lower valence cations into the lattice. In order to maintain charge neutrality oxygen vacancies have to be introduced, which allow oxygen ion migration.



Vacancy transport in YSZ

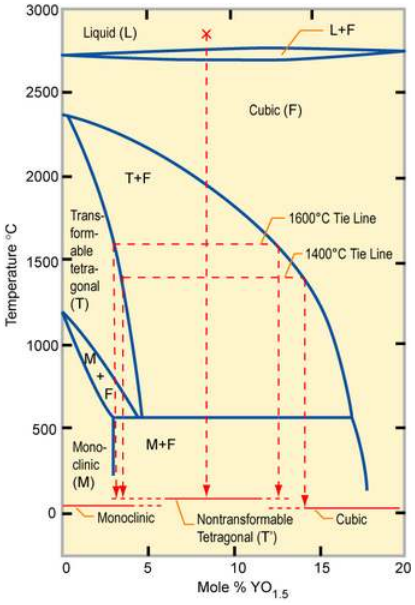
An interesting feature of the fluorite structure is that it can sustain a high degree of substitution. A very disordered and open structured material results from this, which promotes ionic conduction.



High resolution transmission electron micrograph depicting the nickel and YSZ interface (Image courtesy of Technical University of Tokyo)

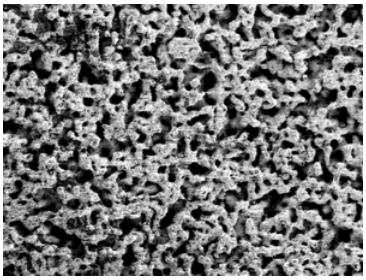
By substituting the host cation sites with either rare earth or an alkaline earth element, just as with yttria-stabilised zirconia (YSZ), an increase of ionic conduction can be achieved. Zirconia (zirconium dioxide, ZrO_2) in its pure form has a high melting temperature and a low thermal conductivity. The applications of pure zirconia are restricted because it shows polymorphism. It is *monoclinic* at room temperature and changes to the denser *tetragonal* phase from circa 1000 °C. This involves a large change in the volume and causes extensive cracking. Hence zirconia has a low thermal shock resistivity. The addition of some oxides results in stabilising the cubic phase and the creation of one oxygen vacancy, i.e. $2[Y'_{Zr}] = [V_{O^{-}}]$

$$Y_2O_3(ZrO_2) \leftarrow 2Y'_{Zr} + 3O^x_o + V_o \tag{31.6.1}$$



Phase diagram of partially stabilized zirconia (PSZ)

Partially stabilized zirconia (PSZ) is a mixture of zirconia polymorphs: a cubic and a metastable tetragonal ZrO_2 phase is obtained, since an insufficient amount of stabilizer has been added. PSZ is also called tetragonal zirconia polycrystal: TZP. PSZ is a transformation-toughened material since the induced microcracks and stress fields absorb energy. PSZ is used for crucibles because it has a low thermal conductivity and a high melting temperature. The addition of 16 mol% CaO or 16 mol% MgO or 8 mol% Y_2O_3 (8YSZ) is enough to form fully stabilized zirconia. The structure becomes cubic solid solution, which has no phase transformation when heating from room temperature up to 2500 °C. Because of its high oxide ion conductivity, YSZ is often used for oxygen sensoring and solid oxide fuel cells.

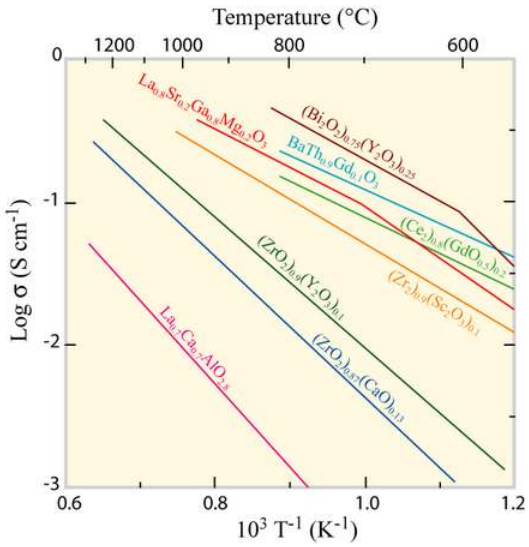


Scanning electron micrograph showing 8 mol% Y_2O_3 (8YSZ) surface (Image courtesy of Altair Nanotechnologies, Inc)

It might be expected that an increase of the dopant concentration would lead to an increase of conductivity. This correlation only applies to low dopant concentrations because at higher levels, the first and second electron coordination shells dopants start interacting with the oxygen vacancies and the conductivity decreases. The conductivity can be calculated as follows:

$$\sigma = \frac{A}{T}[V_{O^-}][\bar{V_{O^-}}]exp(\frac{-E}{RT}) \tag{31.6.2}$$

where E is the activation energy for conduction, T is the temperature, R and A are constants $[\bar{V_{O^-}}]$ is an unoccupied oxygen vacancy.



Conductivity as a function of temperature, data taken from Ref.[3].

In present fuel cells, the electrolyte of choice is zirconia, stabilised by either 3 mol% Y_2O_3 (3YSZ) or 8 mol% Y_2O_3 (8YSZ). YSZ is not the best ion conductor, but it is the cheapest to process and has low enough electronic conductivity. There are many other materials that conduct oxides but the advantages of YSZ: abundance, chemical stability, non-toxicity and economics make it the most suitable material at present. Drawbacks are high thermal expansion coefficient and hence problems with sealing the fuel cell. The world demand for YSZ is rising, but luckily, Zr is one of the most common elements of the Earth's crust usually in the form of silicate zircon ($ZrSiO_4$). This material has to be purified since SiO_2 tends to block the ionic and electron paths.

Yttria is the main stabilizer used and about 13-16 wt% have to be added to give a fully stabilized cubic material. More rare dopants' supply such as scandia could present a problem in the future.

Another interesting fluorite structured material is CeO_2 doped with 10 mol% GdO (GCO). It is especially useful for lower temperature applications. But GCO is an electron conductor in the reducing environment at the anode and hence short-circuiting is a problem.

Fabrication of zirconia electrolyte films is usually done by *tape casting* or *vapour deposition*.

The second highly interesting group of solid state ion conductors is perovskites. The general perovskite stoichiometry is ABO_3.

https://www.doitpoms.ac.uk/tlplib/fu...Perovskite.mp4

Rotating perovskite lattice

Due to the number of combinations (2+4, 5+1, 3+3) to have the total charge of +6 on AB, the high stability of the structure and the wide variety of cations that can be accommodated within, perovskites have a wide range of properties, which are suitable not only for SOFCs but also as ferroelectrics, oxidation catalysts or superconductors. High ionic conductivity in perovskites is achieved by doping the material with trivalent elements, such as Y on the Zr site of $BaZrO_3$ for example, so that oxygen vacancies are introduced. The conductivity of ABO_3 perovskites strongly depends on the size of the A and less so on the size of the B cation, since the oxides have to migrate through a triangular space, consisting of two large A cations and one smaller B cation. The enlargement of this triangular space facilitates the migration of oxide ions through the lattice. Hence we can expect higher ion conductivity with larger lattice dimensions.

In order to incorporate hydroxyl groups into onto the vacant oxide sites, the material is exposed to humid atmospheres. The second proton of the water molecule attaches to some other oxygen atom in the structure. Due to the loose bonding between the hydrogen ion and the oxygen atom, conduction occurs easily, by hydrogen ions jumping from one oxygen to the other.

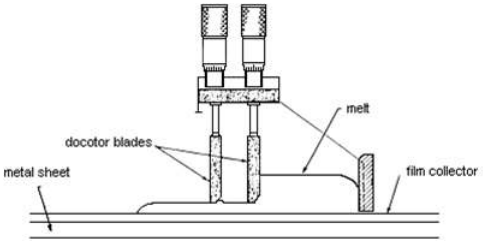
The perovskites $LaGaO_3$ and $BaZrO_3$ proved to be highly interesting since their structure is very tolerant and can accommodate large concentrations of dopants but less than in YSZ (hence the ion conductivity of perovskites is always less than that of YSZ).

A particular composition: $La_{0.9}Sr_{0.1}Ga_{0.8}Mg_{0.2}O_{3-d}$ (LSGM) has been found to have similar, purely ionic conductivities to CeO_2, also at low temperatures (<600 °C). The evaporation of Ga causes problems with stability though. Other problems, such as LSGM reactivity with Nickel electrodes have been solved by adding a CeO_2 buffer layer between the two materials.

In case of an electrolyte-supported design, the electrolyte has to be 120–150 mm thick. Operation at low temperatures requires a thin electrolyte since the conductivity is proportional to temperature. For operation temperatures between 600–800 °C, the electrolyte layer thickness cannot exceed 20 mm to ensure high enough conductivity.

Electrolyte fabrication process

The latest production method involves electrochemical vapour deposition used to make tubular cells (at Westinghouse). Doped lanthanum manganite (cathode material) is placed in a low-pressure chamber and zirconium chloride plus yttrium chloride vapour is passed along the outside of the tube, as water vapour is passed on the inside of the tube. A more conventional production method is tape casting:



Tape casting

This page titled 31.6: Electrolyte is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.7: Electrode Materials

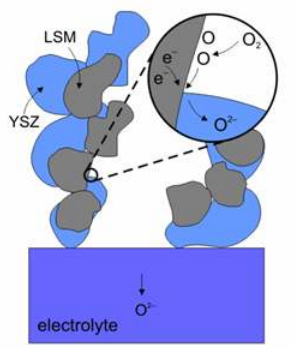
This section of the TLP introduces some of the most common materials used as electrodes in SOFCs.

The electrodes have to possess the following properties to ensure a smooth operation of the fuel cell:

- High electrical conductivity
- High catalytic activity
- High surface area
- Compatibility with the electrolyte (and interconnect)

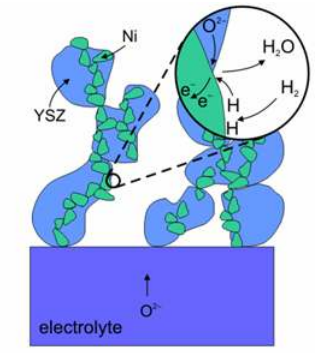
For many years, platinum was the most common and only material that was used as the electrode. The biggest disadvantage of platinum is its cost: it is economically not viable to build fuel cells with platinum electrodes. Hence the pressure to find cheaper electrode materials such as perovskites (e.g. $LaCoO_3$ and $LaMnO_3$) was high. Further development of the perovskites was abandoned because of fast degradation due to their reactivity with YSZ.

The **cathode** material is very important because the oxidation reaction determines the efficiency of the fuel cell. Since cathodes operate in a highly oxidising environment, it is impossible to use cheap base metals. The best compromise is semiconducting oxides, such as **doped** lanthanum cobaltites and lanthanum manganites. A new material: $La_{0.8}Sr_{0.2}MnO_3$ (LSM) has the suitable properties: good electronic conductivity and matching heat expansion coefficient. Since the operating temperature of the fuel cells was reduced to below 1000 °C, it is possible to mix LSM with YSZ at 50/50 proportion to form the first surface on the electrolyte. Other materials have been used, such as $La_{0.6}Sr_{0.4}Co_{0.2}Fe_{0.8}O_3$ (LSCF). LSCF has the advantage of lower power losses at lower temperatures and less susceptibility to poisoning by chromium and it is used with ceria based electrolytes.



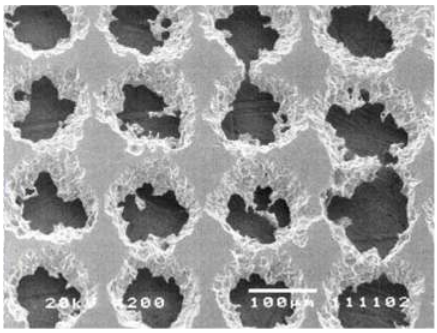
Reduction reaction on the surface of a cathode made of LSM-YSZ.(Image courtesy of Risø National Laboratory)

The loss of cathode performance is mainly due to changes in the microstructure and the phase composition of the material at load conditions. If stainless steel interconnects are used, degradation caused by Cr poisoning occurs, since it evaporates from the steel and condensates preferably on the cathode. In general, the rate at which Cr poisoning occurs decreases with decreasing operating temperature.



Oxidation reaction on the surface of an anode made of Ni-YSZ (Image courtesy of Risø National Laboratory)

Nickel is used as the **anode** because it is economical and exhibits high performance, although due to reasons of adherence and different expansion coefficients, it flakes off easily from the electrolyte unless it is mixed with zirconia, creating a cermet. Ni-YSZ anode allows a rapid and clean connection with the fuel and is a good electronic conductor although Ni is susceptible to become coated with a carbon layer when it is reacting with carbon based fuels, which could prevent further reaction. Certain additives to the Ni+YSZ cermet, such as 5% ceria or 1% molybdena inhibit this process. Besides catalysing the oxidation of hydrogen, Ni is also active in reforming of carbon containing fuels. Anodes, completely made of ceramics show good oxidation-reduction cyclability.



Scanning electron micrograph showing a nickel anode

The contact sides of the electrolyte with the fuel and the oxidant are coated with the electrode material. To form *porous* contact layers, partially sintered materials are used and to allow a gradient of properties, such as heat expansion coefficient several layers with different compositions are laid down. Since the gas atoms discharge (or absorb) electrons at the anode (or cathode), a three-phase boundary zone is required:

- Gas phase (high porosity of electrolyte required for better access)
- Electrolyte phase for ion transport
- Metal phase for electron conduction

Hence a volumetric three dimensional region has to be provided for the reaction. The most common methods of applying the electrode layer on the electrolyte are *plasma spraying*, *vapour deposition*, *solution coating* and *colloidal ink* methods.

This page titled 31.7: Electrode Materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.8: Interconnection

For the interconnection, an inert and impervious material is needed. It should withstand both oxidising and reducing environments. Lanthanum chromite seems to have the necessary properties in systems operating at 1000 °C. Depending on doping, this material matches the thermal expansion coefficient of LSM. For lower temperatures, metallic based alloys can be used. Again, plasma spraying is the most economic method of applying the interconnect layer on the electrode. Although lanthanum provides cell life times of up to 70,000 h, it is not perfectly inert: it expands in the presence of hydrogen, causing cracking, especially at large planar stacks.



Section of SOFC stack with interconnect. From top to bottom: steel, Ni-mesh, cell, contact paste, interconnect steel (Image courtesy of Forschungszentrum Jülich)

Large lanthanum chromite interconnects are made from fine powder, which is prepared as a mixture of the desired components: lanthanum, strontium and chromium nitrate. This mixture is reacted with glycene at high temperatures. This can be compacted to form plates or extruded to make tubes. It is difficult to sinter the powder to full density.

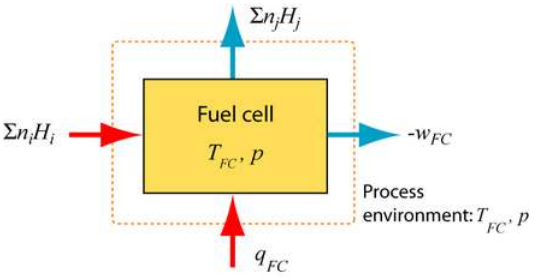
This page titled 31.8: Interconnection is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.9: Thermodynamics

A SOFC is an electrochemical device that converts chemical energy of the fuel and oxidant directly and reversibly into electrical energy. They are not a better or improved heat engine, they are fundamentally different. Hydrogen and oxygen are going to be used to illustrate the thermodynamics of the fuel cell.

The ideal reversible SOFC – Basic derivation of potential and efficiency

The first and second law of thermodynamics describe the reversible SOFC. The reactants (fuel and air) deliver total enthalpy $\sum n_i H_i$ and the total enthalpy $\sum n_j H_j$ leaves the fuel cell, so the change in enthalpy, $\Delta H = \sum n_j H_j - \sum n_i H_i$. The heat q_{FC} has to be extracted from the fuel cell and the reversible work w_{FC} delivered.



1st Law of thermodynamics:

$$q_{FC} + w_{FC} = \Delta H \tag{31.9.1}$$

2nd Law of thermodynamics:

$$\oint dS = 0 \tag{31.9.2}$$

The reaction entropy has to be compensated by the transport of heat to the environment:

$$\Delta S - \frac{q_{FC}}{T_{FC}} \tag{31.9.3}$$

The reversible work is from the above equations:

$$\omega_{FC} = \Delta H - T_{FC} \cdot \Delta S \tag{31.9.4}$$

The Gibbs enthalpy is equal to the reversible work of the reaction. The reversible efficiency is equal to the ratio of the Gibbs and the reaction enthalpy:

$$\nu_{FC} = \frac{\Delta G}{\Delta H} = \frac{\Delta H - T_{FC} \cdot \Delta S}{\Delta H} \tag{31.9.5}$$

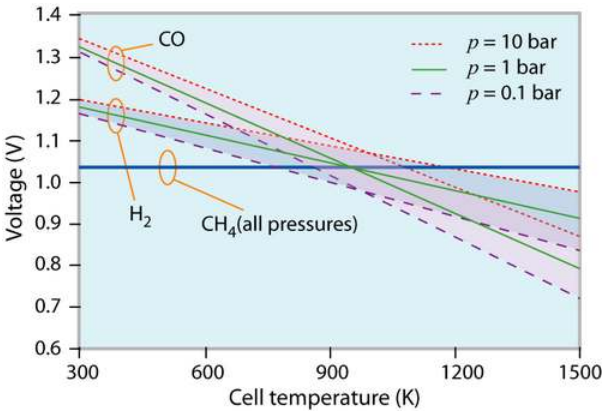
The fuel cell is an electrical device, for which the processes can be fully described by thermodynamic principles.

Hydrogen is absorbed at the anode, it is ionized and the electrons are conducted away to do useful work. Oxygen atoms that arrive at the cathode are ionized by the electrons coming from the anode. The protons and oxide ions react and form water.

Anode: $H_2 \leftrightarrow 2H^+ + 2e^-$
Cathode: $½O_2 + 2e^- \leftrightarrow O^{2-}$
The net reaction is: $2H^+ + O^{2-} \leftrightarrow H_2O$

This shows that the molar flow of hydrogen is twice the molar flow of oxygen. The electric current is therefore:

$$I = \dot{n}_{el} \cdot (-e) \cdot N_A = -2\dot{n}_{H_2} \cdot F \tag{31.9.6}$$



The reversible cell voltage of different fuels at different states (p, T) of the environment

The electric current is a measure of the rate at which fuel is spent. The electric and thermodynamic quantities are matched by considering reversible power:

$$P = V \cdot I = \dot{n}_{H_2} \cdot \omega = \dot{n}_{H_2} \cdot \Delta G \tag{31.9.7}$$

Hence the reversible voltage:

$$V = \frac{-\dot{n}_{H_2} \Delta G}{\dot{n}_{el} \cdot F} = \frac{-\Delta G}{\dot{n}_{el} \cdot F} \tag{31.9.8}$$

Using the assumption that we have (near) ideal gases, a more accurate equation for the Gibbs energy is:

$$\Delta G(T,p) = \Delta H(T) - T \cdot \Delta S(T,p) \tag{31.9.9}$$

where

$$S(T,p) = S^0 + \int {T_0}^T \frac{C_p(T)}{t} dt - R \cdot ln(\frac{p}{p_0}) \tag{31.9.10}$$

We get for the Gibbs energy:

$$\Delta G(T,p) = \Delta G(T) + T \cdot R \cdot ln(K) \tag{31.9.11}$$

where K is the equilibrium constant.

We get from the above analysis the Nernst potential:

$$V_N = \frac{-\Delta G(T)}{n_{el} \cdot F} - \frac{R \cdot T \cdot ln(K)}{n_{el} \cdot F} \tag{31.9.12}$$

Going further: The chemical thermodynamics of SOFCs.

This page titled 31.9: Thermodynamics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.10: System and Outlook

The operating temperature of an SOFC is relatively high.

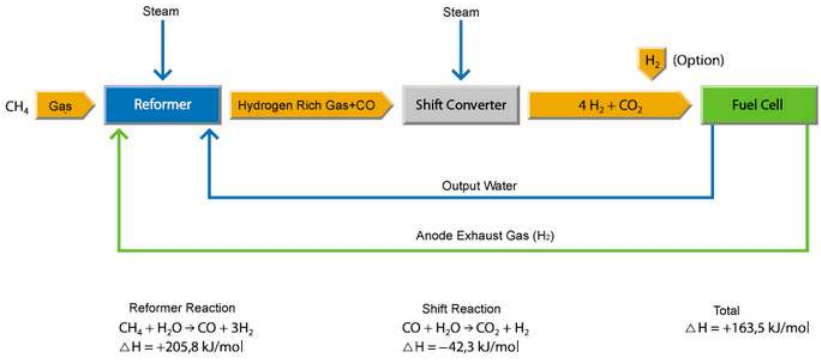
A typical SOFC power plant is fuelled with natural gas because of the lack of a hydrogen infrastructure. A plant must have three main components:

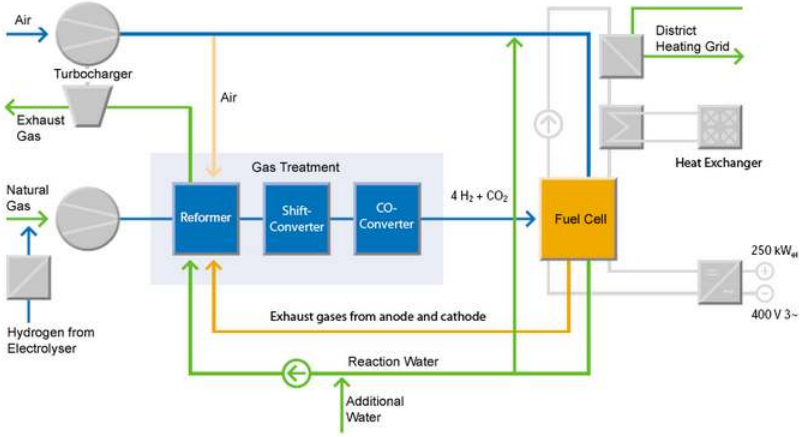
1. The preheater raises the temperature of the fuel and air to near the operating temperature. At the same time, the preaheater reforms the gas by steam reforming to hydrogen. Steam reforming constitutes of two steps:

Methane Reforming: $CH_4 + H_2O \rightarrow CO + 3H_2$
Water Gas Shift: $CO + H_2O \rightarrow CO_2 + H_2$
Overall Reaction: $CH_4 + H_2O \rightarrow CO_2 + 4H_2$

2. The cell stack electrochemically oxidises the hydrogen stream, drawing oxide ions through the electrolyte from the air stream.

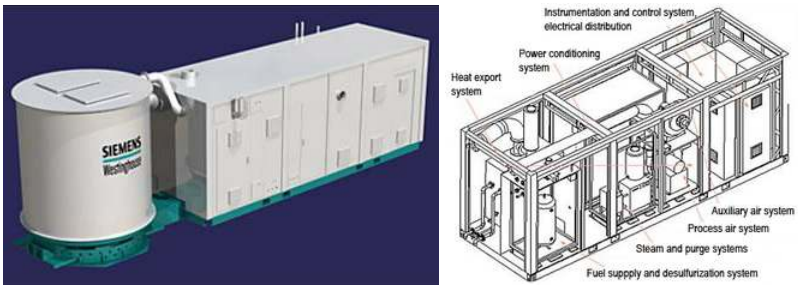
Electrochemical reaction: $H_2 + ½O_2 \rightarrow H_2O$





The schematic diagram above depicts a complete, 250 kW fuel cell system (Source: Innovationspark-Brennstoffzelle).

3. The lower cycle utilises the exhaust energy. The exhaust gases are so hot that gas turbines can be driven to generate additional electrical energy and thus increasing the efficiency of the fuel cell system up to 80%.



SFC-200, a 125 kW SOFC cogeneration system (Source: Siemens Westinghouse)

The durability of the SOFC is mainly determined by the processes occurring during thermal cycles, oxidation-reduction cycles and the sulphur contamination (even at high temperatures, sulphur is absorbed by the anode).

Fuel

One of the great advantages of the SOFC is that it can use a big range of fuels, depending on the cathode composition. Due to the high operating temperature, internal reforming can take place at the anode, when steam is added to the fuel. The reaction of methane is as follows:

$$CH_4 + H_2O \rightarrow CO + 3H_2$$

Both hydrogen and carbon monoxide can react with the oxide ions. A shift reaction also occurs at the anode since the reaction of CO is slow, producing more hydrogen.

$$CO + H_2O \rightarrow CO_2 + H_2$$

The disadvantage of using hydrocarbon fuels is the possible formation of coke on the anode:

$$2CO \rightarrow CO_2 + C$$
$$CH_4 \rightarrow 2H_2 + C$$

As mentioned above, impurities, such as sulphur are also damaging to the SOFC. Only desulphurised natural gas can be used as fuel. Other additives (more than 100 different molecules are present in commercial gasoline) can have damaging effects on the nickel anode.

The activity of the nickel anode decreases due to sintering and coke formation when carbon containing fuels are used. The ceramic parts can easily break if vibrational forces are present. This is one reason, why SOFCs are best suited for stationary applications rather than mobile applications.

The ultimate goal is to build a decentralised network of medium sized power generating SOFCs that can supply a small community with electricity with a much higher reliability and minor consequences in case of failure compared to the current system of few but very large power plants.

This page titled 31.10: System and Outlook is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.11: Molten carbonate fuel cells (MCFCs)

High temperature cells

Molten Carbonate Fuel Cells (MCFCs) are another type of high temperature fuel cell. A molten mixture of salts: lithium, sodium, potassium carbonate is used as the electrolyte. These salts melt and conduct carbonate ions (CO_3^{2-}) from the anode to the cathode when heated to about 600°C. Hydrocarbons have to be used as part of the fuel since the charge carriers in the electrolyte are carbonate ions. Hydrogen is also needed at the anode. It is gained by internal reforming of hydrocarbon based fuels. The electrodes should be resistant to poisoning by carbon. The high exhaust temperature makes cogeneration of electricity with turbines possible; hence the efficiency (60% without and 80% with hybrid technology) is relatively high compared to other fuel cell systems. MCFCs are mainly used for stationary power generation in the 50 kW to 5 MW range. Since it uses a liquid and high temperature electrolyte, it is rather unsuitable for mobile applications. The main problem with MCFC is the slow dissolution of the cathode in the electrolyte. Most of the research is therefore in the area of more durable materials and cathodes.



Molten Carbonate Fuel Cell

Historical summary

Both the solid oxide and the molten carbonate fuel cells are high temperature devices. Their development followed similar lines until the late 1950's. First, E. Baur and H. Preis experimented with solid oxide electrolytes in Switzerland. The technical problems they encountered were again tackled by the Russian scientist O.K. Davtyan without success though. In the late 1950's, Dutch scientists G.H.J. Broers and J.A.A. Ketelaar focused on molten carbonate salts as electrolyte. By 1960, they reported the first MCFC prototype. In the mid-1960's, the US Army's Mobility Equipment Research and Development Center (MERDC) tested several MCFCs made by Texas Instruments ranging from 100 to 1000 Watts. Ishikawjima Heavy Industries showed in Japan in the early 1990s that a 1000 Watt MCFC power generator can operate for 10000 hours continuously. Other large power plants with outputs of up to 3 megawatts are already planned.



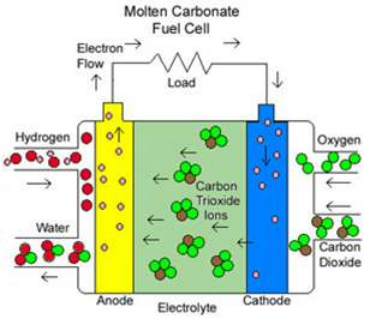
M-C Power's molten carbonate fuel cell power plant in San Diego, California, 1997. Smithsonian Institution, from the Science Service Historical Images Collection, courtesy of National Energy Technology Laboratory.

The MCFC has been under development for 15 years as a stationary electric power plant. Although when most problems with the Solid Oxide Fuel Cell are solved, work on the MCFC might be stopped.

This page titled 31.11: Molten carbonate fuel cells (MCFCs) is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.12: Electrolyte

In most cases, the electrolyte of the MCFC is made of a lithium-potassium carbonate salt heated to 600-1000°C. At this temperature, the salt is in liquid phase and can conduct ions between the two electrodes. The typical mixture ratio of the electrolyte is 62 m/o Li_2CO_3 and 38 m/o K_2CO_3 (62/38 Li/K). This particular mixture of carbonate salts melts at 550°C and when it is mixed with lithium aluminate ($LiAlO_2$ is a ceramic matrix retaining the molten salts, it can be used both as an ion-conducting electrolyte and gasketing for the fuel cell stack. Negative carbonate ions (CO_3^{2-}) are responsible for conduction. As discussed above, the long term performance is an issue for MCFCs.



The following properties of the electrolyte have to be taken into account:

- Volatility of different alkali metal hydroxides generated in the moist cathode atmosphere
- Solubility of the cathode (NiO)
- Segregation of the electrolytes
- Both oxygen and carbon dioxide solubility in the electrolyte
- Oxygen reduction kinetics

Loss of electrolyte mainly occurs at the cathode via hydrolysis:

$$MeCO_3 + H_2O \rightleftharpoons 2MeOH + CO_2$$

Especially the electrolyte that has been used unaltered for a long time (since Ketelaar and Broers), Li_2CO_3/K_2CO_3 (38/62) eutectic, has a relatively high volatility. This causes the fuel cell to dry out.

The partial pressure due to MeOH varies with the square root of water vapour to carbon dioxide vapour pressures:

$$p(Me_iOH) = K_i(T)\sqrt{\frac{p_{H_2O}}{p_{CO_2}}} \tag{31.12.1}$$

$K_i(T)$ is the equilibrium coefficient of the carbonate ion in the melt according to the equilibrium equation:

$$CO_3^{2-} + H_2O \rightleftharpoons CO_2 + 2OH^-$$

The anode off-gas on the other hand can be mixed with air after combustion and reused in the cathode chamber with a sufficiently high air excess. Here oxygen and carbon dioxide are consumed in a ratio of 1:2 (molar) by the cathode process.

The change in composition of the electrolyte due to segregation and the volatility of certain species may result in a change in melting temperature. The electrolyte can solidify, causing the fuel cell to malfunction and sometimes allowing gases to break through. Carbonate ions from the electrolyte are used up in the reactions at the anode, so that we have to compensate for the loss by injecting carbon dioxide at the cathode.

Segregation occurs as the potassium concentration increases near the cathode. This leads to an increased cathode solubility and hence a decline of cell performance.

Recent studies show, that using Na instead of K can decrease the amount of segregation. In a molten binary salt with a common anion, segregation occurs due to the difference in mobilities of the cations. The heavier and bigger potassium and sodium cation is

faster lithium in mixtures that have a higher potassium concentration that K_2CO_3=0.32 (Chemla effect), whereas below this isotachic concentration, lithium is faster.

The partial pressure of carbon dioxide has a much smaller effect on the current voltage curves than the partial pressure of oxygen. Mass transfer limitations are not observed for CO_2 but for oxygen at low partial pressures. CO_2 transport also hinders the cathode operation much less than oxygen transport since this gas has a much better solubility in carbonate melts than O.

Only the Li_2CO_3/Na_2CO_3 eutectic with much lower sodium vapour pressures could assure long term performance of fuel cells. We can conclude that the Li/Na electrolyte is more reliable and safer than Li/K. Since the ionic conductivity of Li/Na carbonate melts is higher than that of Li/K carbonate melts, Li/Na is preferred material as the electrolyte. But since the electrolyte is in the liquid phase, the fuel cell needs a more complex design, compared to other technologies using a solid electrolyte.

This page titled 31.12: Electrolyte is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.13: Electrochemistry

External fuel processors are not needed for MCFCs since the fuels can be reformed internally at the high operating temperatures. Internal reforming includes converting methane and steam into a hydrogen rich gas.

$$CH_4 + H_2O \rightleftharpoons CO + 3H_2$$

At the anode, hydrogen reacts with the carbonate ions to produce water, electrons and carbon dioxide.

$$H_2 + CO_3^{2-} \rightleftharpoons H_2O + CO_2 + 2e^-$$

The electrons are conducted away by an external circuit to do useful work to the cathode. Oxygen from the air and carbon dioxide from the anode react at the cathode with electrons to form water and carbonate ions.

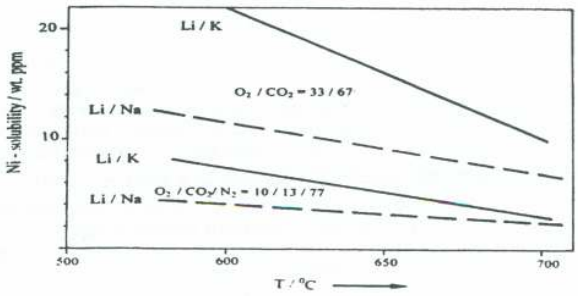
$$O_2 + CO_2 + 2e^- \rightleftharpoons CO_3^{2-}$$

The carbonate ions migrate through the electrolyte to the anode, and complete the electrical circuit. CO_3^{2-} is used up at the anode; CO_2 is needed at the cathode.

This page titled 31.13: Electrochemistry is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.14: Electrodes

A significant advantage of the MCFC is that non-noble metals can be used as electrodes. At the high operating temperature, a Nickel anode and the Nickel oxide cathode is able to promote the electrochemical reaction. This means lower production costs compared to low temperature fuel cell, where the catalyst electrode is usually made of platinum. The Ni electrodes are less prone to CO poisoning, hence coal based fuel can be used, especially since internal reforming can take place.



Solubility of electrode in electrolyte

The main problem with the electrodes is their solubility in the electrolyte by Ostwald ripening, which is a dissolution/reprecipitation process. It decreases the internal surface of the porous nickel oxide cathode, causing it to deteriorate. The solubility of nickel oxide (cathode material) is dependant on the cathode potential and temperature. The solubility of Ni and NiO in Li/Na was found to be lower than in Li/K melts.

Although Li/Na melts have been found to have superior performance compared to Li/K melts, the lower oxygen solubility reduces the cathode performance on lean gas with a low oxygen partial pressure (below 0.1 bars).

This page titled 31.14: Electrodes is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.15: Proton Exchange Membran Fuel Cells (PEMFCs)

Low temperature cells

The proton exchange membrane (a.k.a. polymer electrolyte membrane) fuel cell uses a polymeric electrolyte. This proton-conducting polymer forms the heart of each cell and electrodes (usually made of porous carbon with catalytic platinum incorporated into them) are bonded to either side of it to form a one-piece membrane-electrode assembly (MEA). A quick overview of some key advantages that make PEMs such a promising technology for the automotive markets:

- Low temperature operation, and hence
- Quick start up
- No corrosive liquids involved
- Will work in any orientation (or zero g for that matter)
- Thin Membrane-electrode assemblies allow compact cells

Brief history

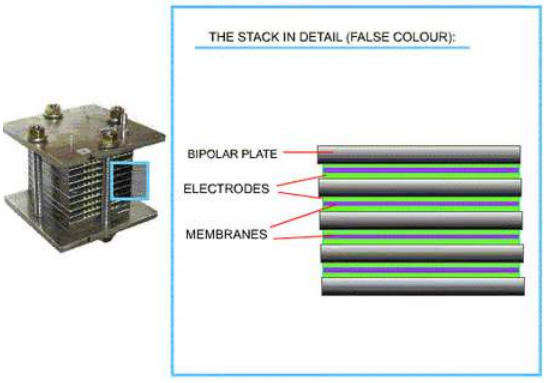
The PEM fuel cell was developed in the 1960's in General Electric's labs. As with so many technologies, the space program and military funded research fast-forwarded it's development. PEM membranes were first applied to a US Navy project and projects for the US Signal Corps. PEM cells were used in NASA's Gemini program, which was to serve as a means of testing technology for the Apollo missions. Batteries were not suitable for a journey to the moon because of the extended flight duration. Early PEM systems were, however, unreliable and plagued with leakages and contamination. The systems installed in Gemini spaceships had an operational lifetime of just 500 hrs, although this was considered suitable. Another issue was the water management systems, which are required to keep the membrane hydrated to the correct extent. Apollo designers opted for the more mature technology of AFCs, as did the Space Shuttle designers in the 70's. Recently however, as part of NASA's program of continuous upgrade on the Shuttles, PEM systems have replaced the aging AFC technology as the primary power source for the Shuttles' systems. GE decided to abandon their research on PEMFCs in the 70's, probably due to the cost. At that time, the catalysis required 28 mg of Platinum per cm^2 of electrode, compared to the current figure of 0.2 mg cm^{-2}, or less.

Automobiles are arguably one of the most important consumer products on the planet. The finite fuel reserves, which they are chewing through, are not currently a limiting factor, but they will be soon. Much investment has been aimed at developing fuel cell technology for the automotive industry and the electrolyte of choice is the PEM. We'll look at the problems which automotive companies need to overcome before fuel cell cars hit the street.

However, recent developments in PEMFCs have brought their current densities up to around 1 A cm^{-2} and cut the platinum requirement to 1% of what used to be needed. The scope of PEMFCs is, arguably, wider than that of any other power supply technology; with the potential to power a range of devices from mobile phones and laptops to busses, boats and houses.

Construction of the PEM cell

The PEMFC is constructed in layers of bipolar plates, electrodes and membranes:



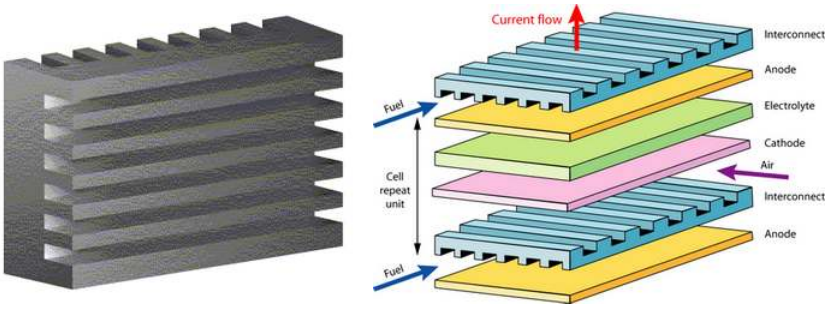
PEMFC components

Each individual cell produces about 0.7 V EMF when operating in air, as calculated by the expressions outlined in the efficiency section. In order to produce a useful voltage, the electrodes of many cells must be linked in series. In addition to connecting the cells, we must ensure that reactant gases can still reach the electrodes and that the resistance of the electrodes has a minimal effect. If, for example, we were to simply wire up the edge of the anode of one cell to the cathode of another, electrons would have to flow across the face of the electrodes. Each cell only produces ~0.7 V, even a small reduction in this isn't permissible, so cells aren't normally wired up this way.

A *Bipolar Plate* is used to interconnect the anode of one cell to the cathode of the next. It must evenly distribute reactant gases over the surface of the anode, and oxygen/air over the cathode. Bipolar plates may also need to carry a cooling fluid, and in addition, need to keep all these gases and cooling fluids separate. Design considerations:

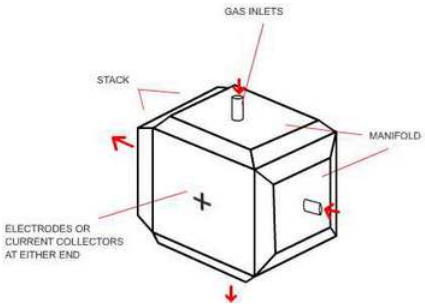
- The electrical contacts should be as large as possible
- The plate should be thin to minimise resistance
- Gas needs to flow easily across the plate

Often these factors are antagonistic to each other, for instance, large contact area would reduce the width of the gas channels. A very simple bipolar plate might look like this:



A typical bipolar plate (Left) found in a plate-type PEM assembly (Right)

Reactant gases flow at right angles to each other. In a simple plate design as above, the channels extend right to the edge. The reactant gases would probably be supplied to the system via *external manifolding* in this case.



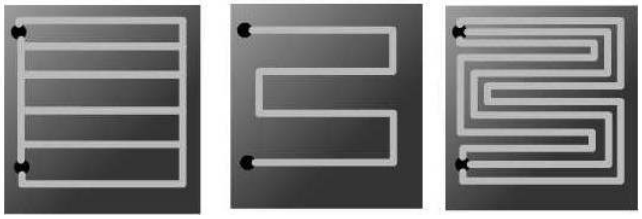
External manifolding

External manifolding is a very simple solution, and therefore carries out the job cheaply, but the technique has two major disadvantages. 1) The gaskets needed to seal the plates don't form a tight seal where the channels come to the edge of the plate, leading to localised leaks of the reactant gases. 2) Additional channels for cooling fluids are very difficult to incorporate into an externally manifolded system, so all the cooling must be done by the air flowing across the cathode. This means more air than is necessary for the reaction must be pumped through the channels, which in turn means the channels must be wider, that the chance of leaks is increased and that some of the energy produced must be used to power blowers. Whilst simplicity is always a bonus, external manifolding is rarely used in modern systems.



In this image of a Ballard Nexus™ fuel cell system, the fan used to blow air through the stack for cooling is visible on the left of the stack.

Most modern bipolar plates make use of *internal manifolding*. The three examples below show how this might be achieved. In each case, the channels do not run to the edge of the plates so a gasket could be fitted here and a gas-tight seal would be more easily achieved.



Internal manifolding

- The design on the left is a fairly simple parallel channels design; reactant gases would be blown into one end of the channel through one hole, and removed at the other hole. There are many different designs possible, and designers of bipolar plates are yet to reach an agreement on which type is best. In parallel designs, water or gas may build up along one of the channels causing a temporary blockage. In this case the reactants will happily continue to pass through the other channels and not clear the blockage.
- The second design, a serpentine design, guarantees that if reactants are flowing at all, they're flowing all along the channel and blockages are easily cleared. The problem in this case is that it takes more effort to push reactants through the long, winding path.
- The third design is more of a compromise between the two and is the type of thing often seen in bipolar plate design. The channels are typically about 1 mm in width and depth. The pressure difference between the start and end of a channel must be engineered to overcome the surface tension of water droplets forming on the channel walls in order to clear blockages. Ballard, for example, achieve this pressure difference with rectangular plates in which the gases run across the long axis in a long parallel design.

The material properties of a bipolar plate, as summed up by Ruge and Büchi (2001), must take into account several important factors:

- Electrical conductivity >10 S cm^{-1}
- Heat conductivity of 20 W m^{-1} K^{-1} if cooling fluid is integrated, 100 W m^{-1} K^{-1} if heat is removed from the edges.
- Gas permeability < 10^{-7} mbar L s^{-1} cm^{-2}
- Resistant to corrosion in an environment of acidic electrolyte, hydrogen, oxygen, heat and humidity.
- Reasonably high stiffness E > 25 MPa
- As ever, it should cost as little as possible.

The plates must also be manufactured so that they are:

- Thin for maximum stack volume
- Light for minimum stack mass
- Able to be produced quickly with a short cycle time

These various and difficult specifications which must be met, along with the fact that modern electrodes require very little catalytic platinum, mean that the bipolar plate is the most expensive part of a modern fuel cell.

Materials for constructing bipolar plates

PEMFCs without bipolar plates

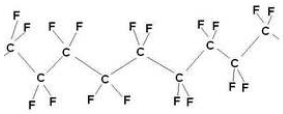
As discussed, bipolar plates may provide excellent contact between cells, but they are expensive and complex. Some manufacturers, often on the smaller industrial scale, choose different techniques to link their cells. Cells could be connected simply edge to edge, reducing the possibility of leakage. One manufacturer (Intelligent Energy) produces cells with stainless steel bases through which hydrogen channels pass. The cathode current collector is a porous metal and these individual cell units are simply stacked with a piece of corrugated stainless steel between them. It's a simple solution which may gain popularity.

In conclusion, we should note that although a broad range of bipolar plates techniques exist, none of them fully meet the criteria set above. There is lots of development still to be done in this area before we meet a new industry standard.

This page titled 31.15: Proton Exchange Membran Fuel Cells (PEMFCs) is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

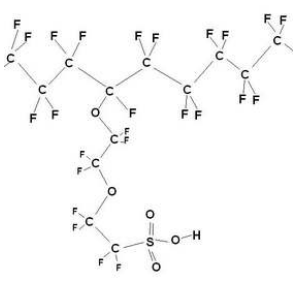
31.16: Membrane

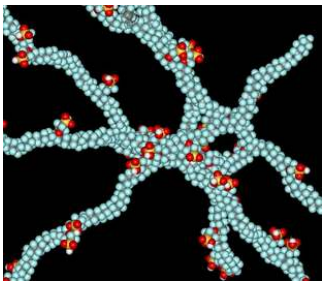
Dupont's Nafion™ ion exchange membrane forms the basis of the proton exchange membrane fuel cell. Each company involved in the development of PEMFCs may have their own variation on Nafion, however, they're all based on the same sulphonated fluoropolymers and Nafion remains something of an industry standard in membranes, to which all others are compared (although it is not always most suitable). Nafion is a polymer based on PTFE (polytetrafluoroethylene).



PTFE

Nafion is essentially PTFE containing a fraction of pendant sulphonic acid groups. (Nomenclature: "sulphonic acid group" usually refers to the un-dissociated SO_3H group, where as "sulphonate" refers to the ionised SO_3^- group after the proton has dissociated). The ion containing fraction is normally given in terms of equivalent weight (i.e. number of grams of dry polymer per mole of acidic groups). The useful equivalent weight for Nafion ranges from 800-1500 g mol^{-1}.





Nafion structure (Left) and a fluoropolymer (Right), made by DOW chemical company, also used in PEMFCs.

The length of and the precise nature of the side chains vary between different brands of polymer. Common to all is the PTFE based fluorocarbon "backbone" of the polymer that has several desirable properties:

- PTFE is hydrophobic - this means the hydrophilic sulphonate groups are effectively repelled by the chains and cluster together.
- PTFE is extremely resilient to chemical attack – the environment within the membrane is hostile and very acidic. Hydrocarbon-based polymers would tend to degrade rapidly.
- PTFE is a thermoplastic with high mechanical strength – meaning very thin membranes can be produced, reducing the thickness of each cell and increasing the power density of the stack.

Transport through the membrane

The animation below demonstrates schematically the mechanism of proton transport in the proton exchange membrane.

In reality, the protons would be strongly associated with water molecules and transported in the form of H_3O^+ hydronium ions, or even higher order cations. The Zundel ($H_5O_2^+$ - basically a protonated water dimer) and the Eigen ($H_7O_3^+$) cations are thought to be particularly important in transfer of protons from one hydronium to another.

Points to note:

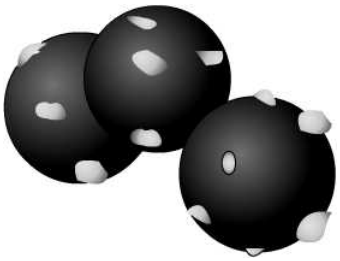
- Sulphonic acids are highly acidic (pKa ~ –6 in Nafion) meaning they have a high tendency to dissociate into anions and protons (the effect of the aliens' blood in the "Alien" films was produced with chlorosulphonic acid). It is of course, these protons that act as the charge carriers through the membrane.
- In order for the polymer to conduct H^+ it must be hydrated to the correct degree, in order to promote dissociation of ionic groups and provide a mechanism for proton transport. Proton conductivity is strongly dependant on the water content of the membrane. The water in the membrane is localised to the hydrophilic groups, where the protons dissociate and are transported in a vehicular manner (by diffusion of hydrated protons) and also structurally (via proton transfer between hydrated clusters).
- Typical PEMs have conductivity in the order of 0.01–0.1 S cm^{-1} at 80–90 °C, which is a far lower temperature than other solid-state (usually ceramic) electrolytes.

This page titled 31.16: Membrane is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.17: Electrodes and Membrane-Electrode Assembly

Catalyst

In the first fuel cells, platinum was used in relatively large quantities. This perhaps led to the false belief that most of the cost of a fuel cell is down to the platinum in it. Generally this is not the case. Platinum particles are deposited very finely onto carbon powders so that the platinum is very finely divided with a maximal surface area. With catalysts produced in this way, the raw material platinum cost is just $10 for a 1 kW cell stack.

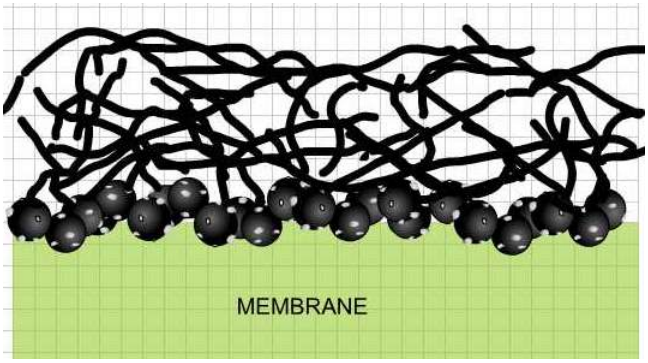


Catalyst made of carbon powders deposited with platinum particles

Bonding

Before the catalyst layer is applied to the electrolyte, a coating of soluble electrolyte is brushed onto it. This ensures that there is good contact between the platinum and the electrolyte to achieve the important three-phase interaction between gas, catalyst and electrolyte necessary for the reaction to proceed.

The catalyst can be applied to the membrane in one of two ways: Either the catalyst powder can be applied directly to the membrane, by rolling, spraying or printing, and then have the supporting electrode structure (often called the gas-diffusion layer) added afterwards, or the electrodes can be assembled separately and bonded to the membrane in complete form by hot pressing. The catalyst powder is sometimes mixed with PTFE to drive out product water and prevent the electrode becoming water logged. The “gas diffusion layer” is added between the catalyst and the bipolar plate to provide some rigidity to the MEA and to ensure ease of diffusion. This layer is usually composed of carbon cloth or carbon paper 0.2–0.5mm thickness, with more PTFE added to expel water.



The membrane electrode assembly, once completed.

This page titled 31.17: Electrodes and Membrane-Electrode Assembly is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.18: Efficiency and Reaction Conditions

The proton exchange membrane is a solid-state electrolyte that functions at around 80 °C. Compared to the 1000 °C at which the solid oxide electrolytes become conductive, this is a low temperature.

What temperature is preferable?

One of the key banners under which fuel cells are marketed is their efficiency. We must consider how this efficiency comes about and what factors influence it. To do this we must look at the thermodynamics governing the fuel cell.

- **Considering the energy of the system - the Gibbs free energy change.**

Let's consider the energy of a hydrogen fuel cell system as follows:

INPUTS:	PROCESS:	OUTPUT:
Hydrogen	FUEL CELL	Electrical Energy = VIt
		Heat
Oxygen		Water

It's easy to calculate the electrical power and energy output of the system:

Power = VI ; Energy = VIt

The "chemical energies" of inputs and outputs are a little more difficult to define. It is the change in Gibbs free energy that we must consider in this case (or more precisely G_f – the energy of formation - because we use the convention of comparison to pure elements in their standard states), which is the energy available to do external work. In the case of a fuel cell system this external work is pushing electrons through the external circuit, and past whatever impedances we put in their way. Work done by changes in volume and temperature between inputs and outputs is not harnessed by the fuel cell (although this is possible in turbine hybrid systems).

The G_f of both O_2 and H_2 is zero, a useful result when dealing with a hydrogen oxygen fuel cell. ΔG_f refers to the difference in Gibbs free energy of formation between the inputs and the outputs, and is therefore a specific measure of the energy released by the reaction.

ΔG_f = G_f of products – G_f of reactants

We usually consider this quantity to be per a mole of chemical. Let us find the chemical energy released during a nominal fuel cell reaction:

$H_2 + ½O_2 \rightarrow H_2O$

The Gibbs free energy of a system is defined as:
G = H – TS

Which leads to the change in free energy being expressed as:

$$\Delta \bar{g}_f = \Delta \bar{h}_f - T\Delta \bar{s} \tag{31.18.1}$$

Note that we've gone lower-case and added little lines above the letters. This is to signify that we are dealing with molar quantities, so that the units will be Joules per mole or something similar.

The value Δh_f of is the difference between h_f of the products and h_f of the reactants. So for the reaction $H_2 + ½O_2 \rightarrow H_2O$, we have:

$$\Delta \bar{h}_f = (\bar{h}_f)_{H_2O} - (\bar{h}_f)_{H_2} - \frac{1}{2}(\bar{h}_f)_{O_2} \Delta \bar{h}_f = (\bar{h}_f)_{H_2O} - (\bar{h}_f)_{H_2} - \frac{1}{2}(\bar{h}_f)_{O_2} \tag{31.18.2}$$

And Δs is the difference between entropy of the products and reactants so that:

$$\Delta\bar{s} = (\bar{s}_{H_2O} - (\bar{s})_{H_2} - \frac{1}{2}(\bar{s})_{O_2} \Delta\bar{s} = (\bar{s})_{H_2O} - (\bar{s})_{H_2} - \frac{1}{2}(\bar{s})_{O_2} \tag{31.18.3}$$

These values of *s* and $h_f$vary with temperature and pressure according to the equations given below. A full derivation of these equations is beyond the scope of this TLP but can be found in thermodynamics textbooks. It should be noted that we use 298 K as the standard temperature, which is necessary as an integration limit. The “T” subscript to the enthalpy, *h*, means the enthalpy at temperature T.

This page titled 31.18: Efficiency and Reaction Conditions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.19: Fuelling Requirements

https://eng.libretexts.org/@go/page/33001

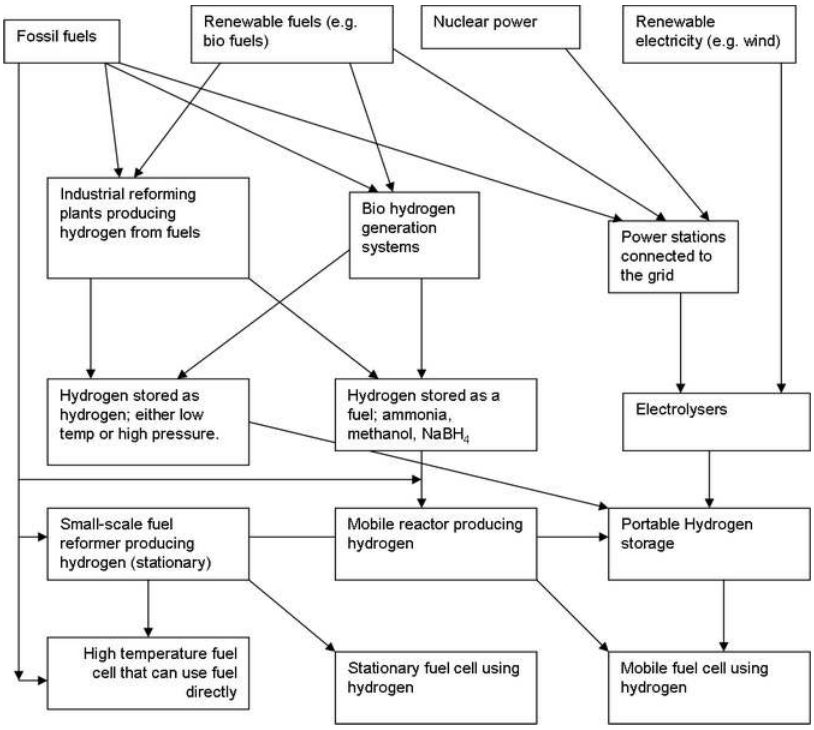
<table>
<tr><th>Cell Type</th><th colspan="5">Fuelling requirements</th><th>Solution</th></tr>
<tr><td>AFC</td><td colspan="5">Alkaline fuel cells, using KOH as the electrolyte, require pure H_2 and O_2 as their fuel. This is because even the small amounts of CO_2 in the air (about 300 ppm) are enough to prevent the cell from functioning properly. This is because CO_2 reacts with the KOH as follows:
$2KOH + CO_2 \rightarrow K_2CO_3 + H_2O$
The presence of potassium carbonate in the electrolyte may reduce cell activity by:
• Reducing the concentration of OH^-
• Increasing the viscosity of the solution, hindering the diffusion of ions through it.
• Reducing the solubility of oxygen in the solution
• Precipitation of the K_2CO_2, which would reduce the surface area of the electrodes.</td><td rowspan="3">Use of Pure Oxygen at the cathodes is necessary. This can be achieved by:
• Regenerative methods i.e. using some external power source (e.g. photovoltaic cells) to hydrolyse H_2O in situ and store the resulting gases.
• Remove the CO_2 from the air. One proposed method takes advantage of the heat exchanging stage necessary in using cryogenically stored hydrogen (the H_2 must be warmed and the cell must be cooled) to freeze the CO_2 out of the air.</td></tr>
<tr><td>Species:</td><td>H_2</td><td>CO</td><td>CH_4</td><td>CO_2 and H_2O</td><td>S (e.g. H_2S and COS)</td></tr>
<tr><td>Effect:</td><td>Fuel</td><td>Poison</td><td>Diluent</td><td>Poison</td><td>Unknown</td></tr>
<tr><td>PAFC</td><td colspan="5">The phosphoric acid (H_3PO_4) systems are almost always fuelled by a hydrocarbon, which must undergo some sort of fuel processing (add link) to release the H_2 gas that reacts at the anodes. The PAFC systems can tolerate CO_2 and un-reacted hydrocarbons (e.g. methane). These act as diluents to the fuel and their concentrations should be minimised. CO gas however will poison the Pt catalysts at concentrations from 0.5%</td><td rowspan="3">• In Situ fuel possessing required (costly)
• CO removed by further processing and shift reaction.</td></tr>
<tr><td>Species:</td><td>H_2</td><td>CO</td><td>CH_4</td><td>CO_2/H_2O</td><td>S etc</td></tr>
<tr><td>Effect:</td><td>Fuel</td><td>Poison (>0.5%)</td><td>Diluent</td><td>Diluent</td><td>Poison (>50 ppm)</td></tr>
<tr><td>MCFC</td><td colspan="5">CH4 can either be a diluent or it may be internally reformed (see SOFCs). Very low sulphur tolerance in the MCFC due to poisoning, especially in the catalysts that aid fuel reforming. The reactions of the MCFC requires CO_2 to be present in the fuel:
$H_2 + ½O_2 + CO_2$ (cathode) $\rightarrow H_2O + CO_2$ (anode)
Recall that CO_3^{2-} ions are transported through the electrolyte from cathode to anode (see section on high temperature fuel cells). This requirement for CO_2 contrasts with the AFC where it must be excluded. The CO_2 is usually recycled externally by passing exhaust gases through a combustor to convert unused fuel into water and CO_2, which can be fed back to the cathode inlet. This also serves to preheat the reactant air.</td><td rowspan="3">• CO_2 recycled externally
• Exhaust gases combusted to yield more CO_2
• Sulphur removed from fuel stream by fuel processors.</td></tr>
<tr><td>Species:</td><td>H_2</td><td>CO</td><td>CH_4</td><td>CO_2/H_2O</td><td>S etc</td></tr>
<tr><td>Effect:</td><td>Fuel</td><td>Fuel (via shift reaction)</td><td>Can be internally reformed</td><td>Diluent</td><td>Poison (>0.5 ppm)</td></tr>
<tr><td>SOFC</td><td colspan="5">SOFCs (and MCFCs) run at high enough temperatures to internally reform CO and hydrocarbons (e.g. petrol or methane) via reaction with H_2O, producing CO_2 and H_2 via “shift” (or oxygenolysis) reactions:</td><td>• Sulphur removed from fuel stream by fuel processors.</td></tr>
</table>

Cell Type	Fuelling requirements					Solution
	$C_nH_m + nH_2O \rightarrow nCO + (m/2 + n)\ H_2$ $CO + H_2O \rightarrow CO_2 + H_2$					
Species:	H_2	CO	CH_4	CO_2/H_2O	S etc	
Effect:	Fuel	Fuel (via shift reaction)	Can be internally reformed	Diluent	Poison (>1.0 ppm)	
PEMFC	PEM Cells generally use pure Hydrogen as the fuel, especially in portable applications where complicated reforming apparatus would be impractical. CO poisons PEM systems very easily because they rely on platinum catalysts. CO has a high affinity for Pt and occupies catalytic sites, preventing hydrogen fuel from preaching them. The processing equipment needed to reduce CO partial pressures to less than 10 ppm adds considerably to the cost of the system. Pure oxygen is used in air independent applications such as submarines and space shuttles. Whilst difficult to implement, use of pure O_2 improves the performance of PEM cells significantly: • The open circuit voltage increases due to increased O_2 partial pressure, as described by the Nernst equation. • Activation over-potential reduces because of better use of catalyst sites. • Limiting current increases, reducing concentration over-potential losses, due to the absence of nitrogen gas.					• CO must be removed to ensure long life of cell. • Pure Hydrogen can be used in either a Compressed form or a cryogenic or metal-hydride stored form.
Species:	H_2	CO	CH_4	CO_2/H_2O	S etc	
Effect:	Fuel	Poison (>10 ppm)	Diluent	Diluent	Unknown	

This page titled 31.19: Fuelling Requirements is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.20: Case Study: Fuelling Practicalities

Supplying the cell's anodes with a constant and consistent supply of hydrogen is no small task. The following diagram shows some ways this can be achieved:



This diagram should be considered it in terms of three levels; where we obtain the energy, how it is used to make hydrogen, and how that hydrogen is stored and used at the cell.

Crude oil derived fuels still account for more than half of the world's total energy supply (petrol, diesel, aviation fuel, kerosene). These consist of simple and aromatic hydrocarbons of varying length. Add to this the consumption of coal and natural gases and the fossil fuel contribution to global energy supply goes well above 80%. Fossil fuels can either be utilised directly in some high temperature fuel cells, or reformed to hydrogen (see notes on reformers), or used in conventional power plants to produce electricity, which could then power electrolysers.

Renewable fuels are useful bio-matters, such as wood or plants, and renewable gas sources such as the methane produced as waste in landfill sites decomposes. These can again be used to make hydrogen by reforming, bio-generation or by conventional power plants and electrolysers. Nuclear power can power electrolysers to make hydrogen indirectly. The animation below summarises the above:

Fuel storage

Once we have the hydrogen, there are several ways in which it can be stored:

- Compression in gas cylinders
- Cryogenic liquid
- A metal absorber as a metal hydride
- Potentially in carbon nano-fibres (still under development)
- Glass micro-spheres

The specific energy of pure hydrogen (the energy per kilo) is higher than any other fuel at about 120 MJ kg^{-1}. However, its energy density (energy per m^3) is very low.

Form of storage	Energy density by weight kWh kg^{-1}	Energy density by volume kWh L^{-1}
Gas (20 MPa)	33.3	0.53
Gas (24.8 MPa)	33.3	0.64
Gas (30 MPa)	33.3	0.75
Cryogenic liquid (-253 °C)	33.3	2.36
Metal hydride	0.58	3.18

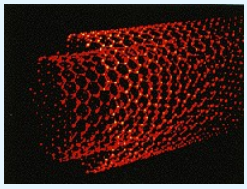
It’s difficult to get a large mass of hydrogen into a small volume. Whilst fossil fuels can be processed to provide an on-demand supply of hydrogen, or even utilised directly by some cells, there are many applications for which a pure hydrogen supply is the only viable solution. These applications might include anywhere where there is limited space for reforming equipment or where it would be too costly, where emissions must be tightly controlled perhaps or where adding the additional fuel processing equipment

would unacceptably reduce the efficiency of the system. Electrical energy output from wind turbines and hydroelectric generators might also be stored as hydrogen.

The safe storage of pure hydrogen presents a tempting challenge to engineers and material scientists. Either very high pressure must be used to squeeze the required mass into a required volume, or very low temperatures used to liquefy it. Hydrogen however, does not liquefy above 22K and even in its liquid phase has rather a low density of 71 kg m^{-3}.

Ways of storing pure H_2

Method:	Pros:	Cons:
High-pressure storage of gas. Hydrogen gas can be stored at pressures of up to 30 MPa (300 atm). Composite tanks can be used with aluminium liners and a Kevlar-epoxy shell.	• Simple • No problems with long-term storage. • No need for high purity hydrogen • Can pressurise the system from the tank	• Tanks must be very carefully manufactured • Failure can be very dangerous • Diffusion of hydrogen into the surface of the aluminium leading to embrittlement • Compression requires energy • Low energy density
Liquid hydrogen (LH_2) stored at 22 K in thermally insulated tanks, usually like large vacuum flasks. State of the art tanks cool the 22 K LH_2 with liquid air at 78 K.	• Widest used method of storing large quantities of hydrogen. • Containment failure less dangerous than with high-pressure storage.	• Gas will slowly boil off so use would require large-scale distribution of filling stations. • Liquefaction requires a lot of energy • Safety concerns of cryogenic liquids • Still low energy density
Metal hydrides More about metal hydrides	• Very safe storage as if the tank ruptures, the fuel will not discharge rapidly. • Constant pressure • Very pure hydrogen is released • Indefinitely reusable.	• decreasing effectiveness over time • Quite a heavy way of storing H_2; only about 1-2% of tank's overall mass is hydrogen. • Still requires a high pressure tank
Glass micro-spheres can be encouraged to absorb H_2 gas at higher temperature, then retain it at lower temperatures before releasing it again when the temperature is raised. A technology yet to be proven.	• Safe storage • Very reusable and un-contaminable. • Lots of H_2 per unit volume	• Complicated • Expensive • Delicate

Method:	Pros:	Cons:
Carbon nano-fibres There was a lot of hype about these after a paper in 1998 claimed amazing results. The idea is that they provide a huge surface for H_2 to bind to and increase the amount of H_2 that can be stored for a given pressure in a tank. Maybe these will work one day, but currently the results are cause for scepticism.  Structure of carbon nano-fibres	• Allegedly capable of absorbing 67% times their own mass of hydrogen. • Safe, inert storage once again	• An unproven technology with results that so far have not been reproducible enough. • Some long-term safety concerns with nano-fibres. They're thought to be carcinogenic. • Carbon nano-fibres are currently rather expensive.

Chemical storage of fuels

This page titled 31.20: Case Study: Fuelling Practicalities is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

31.21: Summary

Problems such as CO_2 emissions and the sustainability of our current energy supplies are going to result in the western world adopting lasting changes to the way we harness energy. It is likely that fuel cell technology has an important part to play in this future.

This TLP should have taught the following:

- The principles, construction and thermodynamics of a fuel cell
- Differences between types of fuel cell – high temperature, low temperature, the key components of each.
- How the technology can be applied and possible future applications
- The complications involved in the fuel cell system as a whole
- The basis of a future hydrogen economy
- Problems associated with storing hydrogen

It is important to remember that the sums spent on developing fuel cells thus far have been very large indeed but, as of yet, returns are sparse. Few people will encounter fuel cells as part of their every day lives, but perhaps this is set to change in near future.

Going further

Books:

1. J. Larminie and A. Dicks, Fuel Cell Systems Explained, 2nd edition, John Wiley and Sons Ltd (2003).

2. M.H. Westbrook, The Electric Car: Development and Future of Battery, Hybrid and Fuel-Cell Cars, London: Institution of Electrical Engineers; Society of Automotive Engineers (2001).

3. High-temperature Solid Oxide Fuel Cells: Fundamentals, Design and Applications by S.C. Singhal and K. Kendall (Eds.), Publisher: Elsevier Science (2004).

This page titled 31.21: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

32: The Glass Transition in Polymers

Learning Objectives

On completion of this TLP you should:

- understand what is meant by the *glass transition* in amorphous polymers, and its causes
- understand kinetic and thermodynamic analyses of the formation of glasses
- appreciate that the value of the *glass transition temperature* depends on the method of measurement, and also on the strain rate or cooling or heating rate
- understand how the glass transition affects the properties of rubber

32.1: Introduction

32.2: Theory 1

32.3: Theory 2

32.4: Measurement of Tg

32.5: Demostrations

32.6: Summary

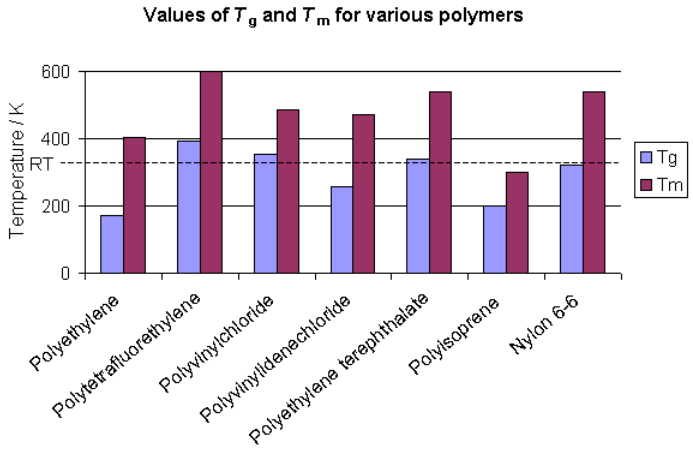
32.7: Questions

This page titled 32: The Glass Transition in Polymers is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

32.1: Introduction

Collections of molecules can exist in three possible physical states: solid, liquid and gas. In polymeric materials, things are not so straightforward. For example most polymers will decompose before they boil, and cross-linked polymers decompose before they melt. For many polymers the transition between the solid and liquid states is rather diffuse and difficult to pinpoint. Amorphous polymers are viscous liquids when they are held at temperatures above their *glass transition temperature*, T_g. Below T_g, the material is solid, yet has no long-range molecular order and so is non-crystalline. In other words, the material is an amorphous solid, or a glass.

The glass transition temperature is different for each polymer, but many polymers are above T_g at room temperature. In many cases the polymers are at least partially crystalline at room temperature and the temperature at which the crystals melt (T_m) is above room temperature. The graph below shows how some polymers are above T_g but below T_m at room temperature. Such polymers are rubbers (so long as they are largely amorphous) at room temperature. However, the polymer may flow like a liquid over long time periods as its amorphous component relaxes under the polymer's weight.



The glass transition of a polymer is related to the thermal energy required to allow changes in the conformation of the molecules at a microscopic level, and above T_g there is sufficient thermal energy for these changes to occur. However, the transition is not a sharp one, nor is it thermodynamically well defined. It is therefore different from melting a crystal, as will be explained later.

A distinct change from rubbery (above T_g) to glassy (below T_g) behaviour is readily observable in a wide range of polymers over a relatively narrow temperature range. In the following sections the behaviour of polymers around the glass transition temperature will be explored. The effects of strain rate, cooling or heating rate and other factors affecting the glass transition temperature will also be explained.

This page titled 32.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

32.2: Theory 1

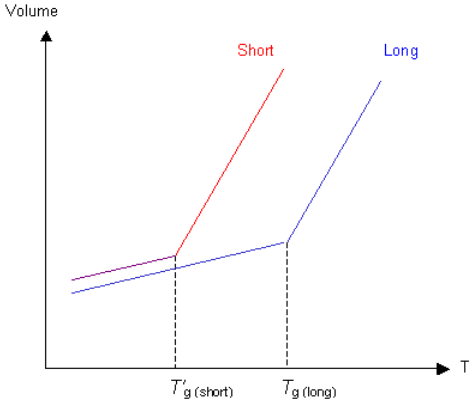
When we extend or compress a polymer elastically (i.e. there is no permanent deformation), we try to move the chain-ends apart or together. For a simple polymer chain to change its conformation, individual C-C bonds must twist from the *trans* to *gauche* position or vice versa - i.e. the torsion angles must change. This is a thermally activated process. At low temperatures, there is not enough thermal energy available to allow torsion angle changes, so the conformation becomes frozen in. The temperature above which the torsion angles can change is called the glass transition temperature. The changes in conformation also depend on time-scale, so the apparent value of T_g depends on the time-scale over which the behaviour is being monitored.

The strain in a polymer is accommodated by the change in shape of the individual molecules, but it should be noted that the response of the bulk polymer is influenced by the interactions between the molecules. This affects the ability of the bonds to rotate, and also the viscosity of the bulk polymer.

Therefore T_g depends on the polymer's architecture, and there are several factors influencing the transition:

Chain Length

Each chain end has some free volume associated with it. A polymer with shorter chains will have more chain ends per unit volume, so there will be more free volume. Hence T_g' for shorter chains will be lower than T_g for long chains. Note that the shorter-chained polymer also has more free volume frozen in below T_g than the long-chained polymer.



Chain Flexibility

A polymer with a backbone that exhibits higher flexibility will have a lower T_g. This is because the activation energy for conformational changes is lower. Therefore, conformational changes can take place at lower temperatures.

Side Groups

Larger side groups can hinder bond rotation more than smaller ones, and therefore cause an increase in T_g. Polar groups such as Cl, CN or OH have the strongest effect.

Branching

Polymers with more branching have more chain ends, so have more free volume, which reduces T_g, but the branches also hinder rotation, like large side groups, which increases T_g. Which of these effects is greater depends on the polymer in question, but T_g may rise or fall.

Cross-linking

Cross-linking reduces chain mobility, so T_g will be increased. It also affects the macroscopic viscosity of the polymer, since if there are cross-links between the chains, then they are fixed relative to each other, so will not be able to slide past each other.

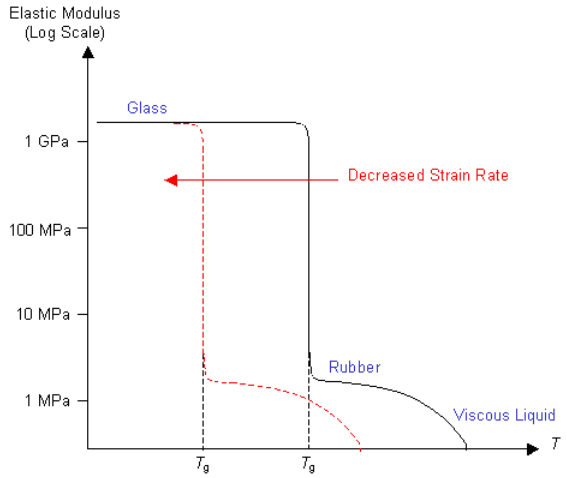
Plasticisers

Small molecules, typically esters, added to the polymer increase the chain mobility by spacing out the chains, and so reduce T_g.

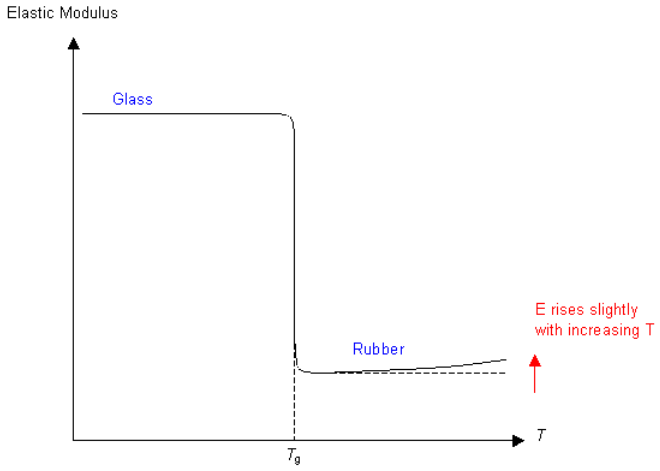
Time Effects

The properties of an amorphous polymer above T_g can change with time

- At very short loading times the polymer can still be glassy because there is not time for the chains to move.
- At intermediate times the polymer may be rubbery - i.e. chains can uncoil and recoil between entanglements, which remain stable.
- At very long times, the chains can move past each other permanently, and so the polymer behaves as a viscous liquid.



For a useful rubbery material it is necessary to suppress chain sliding. One way of doing this is to increase the amount of cross-linking in the polymer. In order for the chains to slide the cross-linking bonds must first be broken. So increasing the number of cross-links decreases the chains' ability to slide over each other. This extends the rubbery region at higher temperatures, so the graph looks like this:



More about the slight rise in elastic modulus with increasing temperature in rubber is available in the Stiffness of Rubber TLP.

This page titled 32.2: Theory 1 is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

32.3: Theory 2

The formation of glasses can be understood from both a kinetic approach and a thermodynamic approach. In this section we will review both.

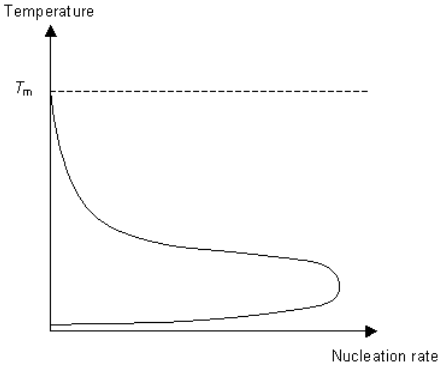
Kinetic Approach to Glasses

Consider a polymeric liquid being cooled towards its melting temperature. Once the temperature of the liquid reaches T_m the solid crystalline phase is thermodynamically favourable. In order for the liquid to undergo a phase transition to the solid state a two-step process must take place:

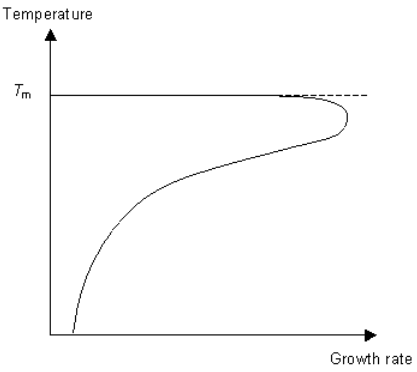
1. Nucleation of solid seeds
2. Growth of the seeds

Nucleation is the formation of small crystalline solid particles in the liquid. As a result a new interface is formed in between the solid particle and the liquid. This interface has an associated energy, the interfacial energy. For a successful nucleation event to occur the polymer must find extra energy from somewhere to "pay" for the energy cost of creating a new interface.

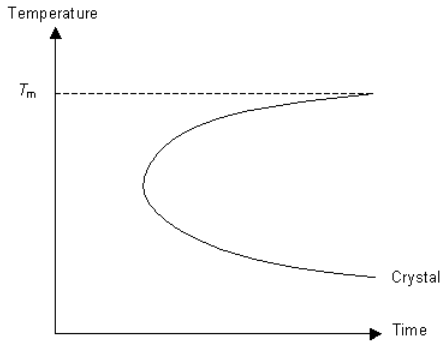
This energy is called the "driving force" for nucleation. As we cool the liquid just below T_m we see no nucleation at first. Instead nucleation occurs at a temperature $T < T_m$, below the expected temperature. When the polymer is still liquid below T_m it is said to be undercooled. The amount of "undercooling" is limited by the nucleation rate.



The second stage of the phase transition is the growth of the nucleated seeds. This is a thermally activated process, which means that its rate is dependent on the temperature. So for fast crystal growth the ideal method is to cool the polymer to just above T_g to allow for nucleation to occur and then raise the temperature to just below T_m to allow growth to occur.

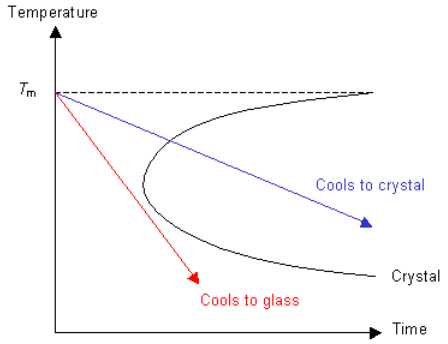


Understanding these concepts we can now think about how best to cool a liquid in order to form a glass. To ensure we form a glass we want to reduce the number of nucleation events as much as possible (otherwise we will have crystals not an amorphous solid) and not allow growth to occur. By cooling the liquid very quickly (i.e. *quenching* it) it is possible to reduce the mobility of the molecules to the point where they can not move around to order themselves periodically as they do not have the energy to diffuse far enough. This concept can be expressed on a TTT (time temperature transformation) graph.



The graph represents the degree of crystallinity of the polymer as it is cooled. The line marked "crystal" indicates a specific degree of crystallinity (for example it may indicate 90% crystallinity in the sample). The TTT graph is used for isothermal transformations, but here it is loosely applied to continuous cooling.

If the cooling rate is fast enough then the polymer can be cooled so that it does not enter the crystal region of the graph. In this case, although the polymer is below T_m, it does not crystallise because the molecules can not move to order themselves periodically (as in a crystal). Instead the polymer has formed an amorphous solid, or a glass.

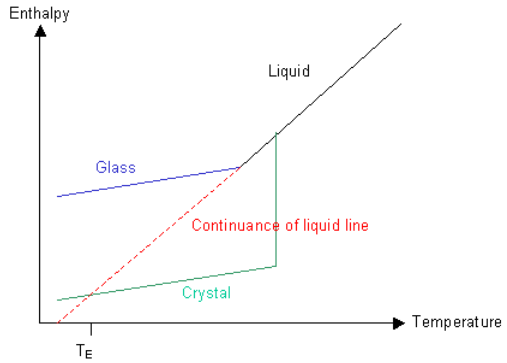


Thermodynamic Approach to Glasses

Consider the red line from the graph above. Above T_m the polymer is a liquid. At temperatures a long way below T_m the polymer behaves like a solid, although it is not crystalline. In this state the glass has the properties of a solid, but may exhibit aspects of liquid behaviour over long time scales.

(There is a common misconception that inorganic glasses behave in this same way. This liquid behaviour is incorrectly thought to be responsible for the windows of old buildings often being thicker at the bottom than at the top. This is discussed briefly, with citations, for those who are interested on a separate page in this TLP.)

Have a look at the graph of the polymer's enthalpy at different temperatures:



If nucleation can occur and the rate of cooling is not too high, then the green line will be followed and the polymer will crystallise below T_m to form a crystalline solid.

If the polymer is quenched so that ordering of the molecules cannot take place then the blue line will be followed and the polymer will form a glass below T_m.

Let us now imagine what would happen if we could follow the red line and keep cooling the liquid without forming a crystalline solid. There is a point, T_E, at which the enthalpy of the supercooled liquid falls below that of the crystalline solid. Similarly, there is a temperature, T_K, the Kauzmann temperature, below which the entropy of the liquid would be less than that of the corresponding solid. In other words, below the Kauzmann temperature the liquid should be more ordered than the corresponding solid. As we know, solids are highly ordered and liquids are not, so the liquid cannot reach this condition.

The paradox is avoided because by the time the Kauzmann temperature has been reached on cooling, the liquid has gone through transition into a glassy state.

Important points relating to phase transitions within polymers

- Glass formation can be explained both kinetically and thermodynamically
- Glasses can be formed by quenching (rapidly cooling) the polymer so that the molecules do not have time to order themselves into a periodic crystal
- Glasses exhibit the properties of a solid, but over long time scales can flow like a liquid
- Below the Kauzmann temperature the polymer must either be a crystalline solid or a glass as the liquid phase is thermodynamically unstable

---

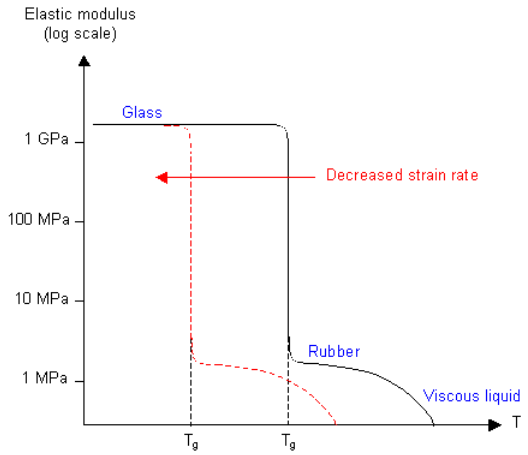
This page titled 32.3: Theory 2 is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

32.4: Measurement of Tg

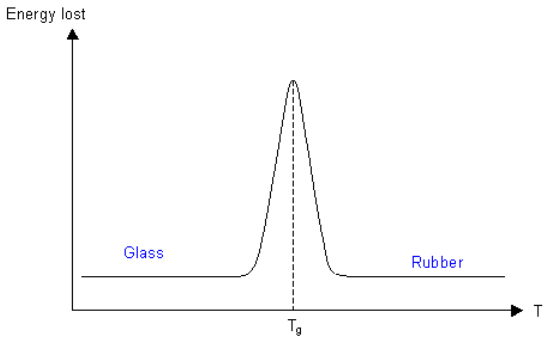
There are several methods available to measure the glass transition temperature, some of which are given below. Since the value of the glass transition temperature depends on the strain rate and cooling or heating rate, there cannot be an exact value for T_g.

Mechanical Methods

It is possible to calculate a value for the glass transition temperature by measuring the elastic modulus of the polymer as a function of the temperature, for example by using a torsion pendulum. Around T_g there is a large fall in the value of the modulus. The frequency of the oscillation is important, since T_g depends on the time allowed for chain segment rotation.



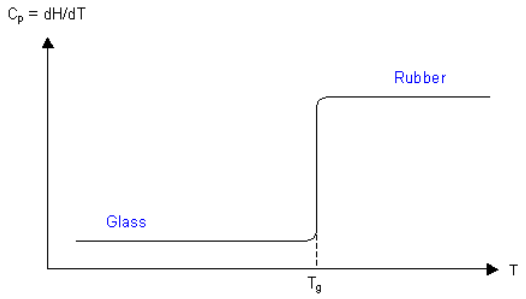
A more common method is *dynamic mechanical thermal analysis* (DMTA), which measures the energy absorbed when a specimen is deformed cyclically as a function of the temperature, and a plot of energy loss per cycle as a function of temperature shows a maximum at T_g.



Thermal Methods

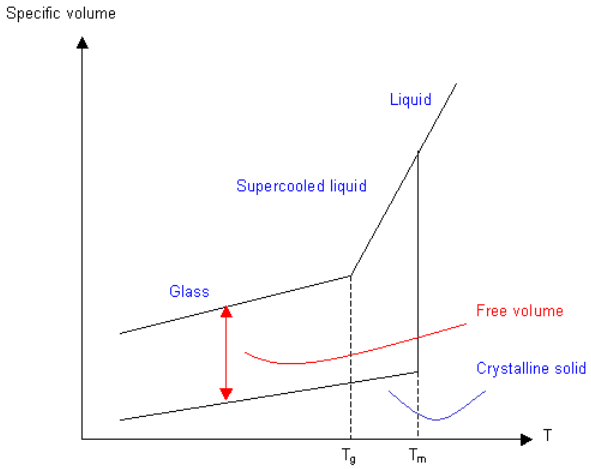
As was shown in the thermodynamic approach to glasses, the enthalpy of a polymer decreases as the temperature decreases, but with a change in slope in the graph at T_g. Taking the derivative of this graph with respect to temperature, the specific heat capacity can be plotted, as below. The specific heat capacity, C_p, can be measured using calorimetry, e.g. differential scanning calorimetry (DSC). The value of T_g depends on the heating or cooling rate.

https://eng.libretexts.org/@go/page/33015



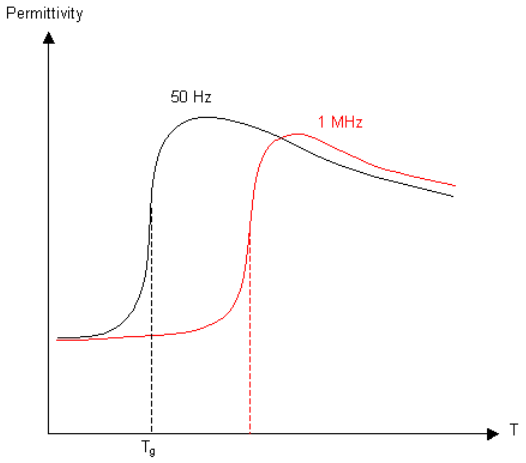
Volume Methods

The changes in conformation that occur above T_g require more volume, so plotting a graph of specific volume or thermal expansion coefficient against temperature will give a value for T_g. The actual volume of the molecules stays the same through T_g, but the *free volume* (the volume through which they can move) increases.



Dielectric Constant

If a varying electric field is applied to a polymeric material, any polar groups will align with the field. Below T_g rotation of the bonds is not possible, so the permittivity will be low, with a big increase around T_g. At higher temperatures the increased thermal vibrations cause the permittivity to drop again. If the frequency of the field is increased, the polar groups have less time to align, so the glass transition occurs at a higher temperature.

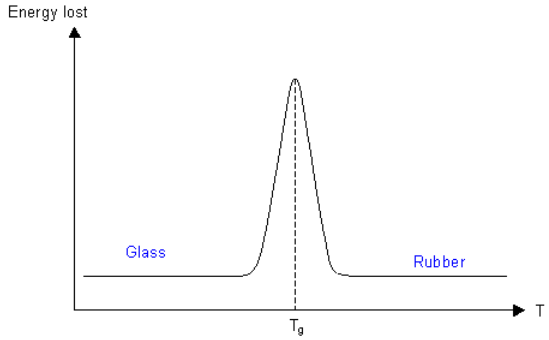


This page titled 32.4: Measurement of Tg is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

32.5: Demostrations

Bouncy Balls

A bouncy ball is made of a polymer that is above its T_g at room temperature, with cross-linking keeping the ball in its spherical shape. The height that the ball rebounds to when dropped is dependent on how much energy is lost during the bounce. At room temperature the ball loses little energy when it deforms, so can rebound to a large fraction of its original height. As the temperature is reduced, the viscosity of the polymer increases, so more of the elastic strain energy is dissipated, and the ball does not bounce as high.



At T_g nearly all the energy is dissipated, and the ball barely bounces at all. As the temperature is reduced more to below T_g, there is not enough energy for conformational changes to occur, and the ball becomes glassy. The energy losses above T_g are due to the viscosity as the conformation changes and the polymer chains move past each other, and since these movements do not occur, the energy is not dissipated, so the ball bounces again.

The videos show how high a bouncy ball bounces at different temperatures. The videos were filmed at 500 frames a second, and played back at 50 frames a second, so are at one tenth of actual speed.

https://www.doitpoms.ac.uk/tlplib/gl...videos/25c.mp4 Ball bouncing at 25°C, well above T_g	https://www.doitpoms.ac.uk/tlplib/gl...ideos/-50c.mp4 Ball bouncing at -50°C, just above T_g
https://www.doitpoms.ac.uk/tlplib/gl...ideos/-70c.mp4 Ball bouncing at -70°C, close to T_g	https://www.doitpoms.ac.uk/tlplib/gl...deos/-190c.mp4 Ball bouncing at -190°C, well below T_g

The second set of videos are close ups of the bounce itself, showing the deformation of the ball. The ball bounces both above and below T_g, although the process is very different. Above T_g the conformation of the polymer chains changes, and the deformation of the ball can clearly be seen. Below T_g, the conformation is frozen, so it is only the interatomic bonds that are strained, and the macroscopic deformation is much less, and is not observable in the videos. bouncing

https://www.doitpoms.ac.uk/tlplib/gl...ber-bounce.mp4 Rubbery bounce, well above T_g	https://www.doitpoms.ac.uk/tlplib/gl.../tg-bounce.mp4 Bounce near T_g
https://www.doitpoms.ac.uk/tlplib/gl...ass-bounce.mp4 Glassy bounce, well below T_g	

Silly Putty

The time dependence on the deformation of a material that exhibits a glass transition can be demonstrated with "Silly Putty". If a ball of putty is left at rest at room temperature, then it will slowly deform plastically under its own weight. However, if the ball of putty is hit with a hammer, again at room temperature, then it is possible to smash the putty into pieces.

https://www.doitpoms.ac.uk/tlplib/gl...utty-small.mp4 Silly putty deforming under its own weight	https://www.doitpoms.ac.uk/tlplib/gl...tty-bounce.mp4 Silly putty striking the ground
https://www.doitpoms.ac.uk/tlplib/gl...utty-smash.mp4 Silly putty being hit by a hammer	

The first video shows a small ball of putty (initially approximately 40 mm diameter) deform under its own weight at room temperature. The deformation took just over 40 minutes, so the film has been speeded-up 80 times. The height of the ball can be judged against the background. The initial deformation is much more rapid than towards the end, as the following numbers show:

Loss of height / mm	Time / mins:secs
5	1:10
10	5:30
15	25:00
16	42:00

The other two videos were both filmed at 1000 frames per second, and played back at 4 frames per second, so are at 1/250th of real speed.

The second video is of a ball of putty hitting the ground after being dropped from a height of 1 m. The ball deforms, but returns to its original shape as it rebounds. When a ball of putty is dropped it rebounds to a large fraction of its original height.

The third video is of a ball of putty being hit with a hammer, also at room temperature. The theory states that at fast strain rates the material will be glassy, and this is shown in this video, although it may not seem like it at first. The ball deforms quite a lot initially, and then breaks apart. This can be explained by the putty being glassy, with an elastic modulus low enough for there to be a visible strain before a critical strain is reached and brittle fracture occurs.

This page titled 32.5: Demostrations is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

32.6: Summary

You should now be familiar with the following concepts:

- Above T_g polymers are rubbery, whereas below T_g they are glassy
- Rubbery behaviour arises from the polymer's ability to change its conformation at high temperatures
- Glassy behaviour arises from the polymer's lack of ability to change its conformation at low temperatures
- Many factors affect the T_g of a polymer. Some factors are to do with the chemistry of the polymer:
  - chain length
  - chain flexibility
  - side groups
  - branching
  - cross-linking
  - the presence of plasticisers

  Other factors are to do with the method of measuring T_g:
  - strain rate
  - cooling or heating rate
- You should also understand the theory of kinetics, which explains how glasses form and noted that there are many different experimental methods for measuring the glass transition temperature in polymers.
- Finally you should have observed how the behaviour of a rubber ball varies when it is bounced above, below and near its T_g and seen how the behaviour of silly putty varies at different strain rates.

Going further

Websites

- The Glass Transition
  Part of the Macrogalleria based at the Department of Polymer Science at the University of Southern Mississippi.
- Crazy Aaron's Puttyworld
  Includes video of a huge ball of silly putty being dropped from a tall building!

This page titled 32.6: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

32.7: Questions

Explain why cooling a liquid very fast (quenching) can lead to the formation of a glass. Your answer should use what you have learnt about the kinetics of glass formation.

Answer

At lower temperatures the mobility of the polymer molecules is reduced. By quenching a liquid the mobility of the molecules is reduced in less time than it takes for the molecules to order themselves in a crystalline manner. Because of this the polymer becomes an amorphous solid, or glass.

What is meant by the 'free volume' of a polymer?

Answer

Unlike the ordered molecular arrangement of a crystalline material, amorphous polymers are made up of disordered polymer chains. At the end of each chain there is a small amount of unfilled space, which is called free volume. The sum of all the small free volumes makes a total free volume for the material.

What is the significance of the theoretical 'Kauzmann temperature'?

Answer

The Kauzmann temperature is the temperature below which the entropy of a liquid would be less than that of the corresponding solid. This is paradoxical and cannot be the case. The glass transition must occur above the Kauzmann temperature.

Why is the measured value of T_g dependent on the method of measurement?

Answer

The glass transition temperature is dependent on the strain rate during measurement. Measurements which take place over a large timescale will result in a lower glass transition temperature, and hence rubbery behaviour at lower temperatures. Measurements which take place on a short timescale will result in a higher value of T_g and hence glassy behaviour at higher temperatures. This can be seen by watching the videos of silly putty flowing like a liquid over a long timescale and smashing like a glass over a short timescale. Both experiments were performed at room temperature.

Using what you understand about the way polymer chains behave above and below T_g, why do you think glasses tend to be brittle whereas rubbers are not?

Answer

The polymer chains in a glass are immobile. They cannot change their conformation so the glass cannot deform plastically. When the glass suffers an impact mechanical energy is transferred into the glass. This energy cannot be dissipated as the molecules are immobile and so the glass will crack, forming new interfaces to expend the input mechanical energy.

In contrast the polymer molecules in a rubber are able to change their conformation and slide past each other. When mechanical energy is transferred into the rubber the energy is expended by the conformation changes taking place in the polymer chains. Macroscopically this leads to plastic deformation and rubbery, rather than brittle behaviour.

What is meant by the conformation of a polymer? Using a Newman projection show what is meant by the 'trans' and 'gauche' states.

Answer

The conformation of a polymer molecule is the arrangement of the bond orientations, both within the backbone, and any side-groups and branches. There are three specific orientations, called trans, +gauche and -gauche, as represented in the Newman projections below. The conformation can be changed by rotation about single bonds.

Explain how the dielectric constant of a polymer is affected by the transition to the glassy state.

Explain why the glass transition temperature is strain-rate dependent.

Below and above the T_g of a rubber ball it will rebound after impact. At T_g the ball will not rebound very much at all. Explain how the ball bounces in these three cases from an energetic point of view.

This page titled 32.7: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

33: Granular Materials

www.doitpoms.ac.uk/tlplib/me...ng-2/index.php

Learning Objectives

The aim of this TLP is to provide an introduction to the static behaviour and flow behaviour of granular materials. Granular materials are seen in many forms in everyday life, such as in piles of powders, sands, soils and mineral ores and pastes such as toothpaste. Other materials such as cement and concrete are in paste form prior to their setting and hardening cement reactions.

On completion of this TLP you should

- Understand the concept of the angle of repose of piles of granular material and the factors which determine this angle.
- Understand the ways in which soils fail and the importance of this knowledge in geotechnical design.
- Understand the phenomenon of liquefaction.
- Understand how the concept of dilatation explains the phenomenon of the surface drying under your feet when you walk on wet sand.

Before you start

There are no special prerequisites for this TLP, although there is a link to the TLP on Stress Analysis and Mohr's Circle in which the Coulomb yield criterion (also known as the Mohr-Coulomb criterion) is discussed.

33.1: Introduction
33.2: Terzaghi's Effective Stress Principle
33.3: Ground Settlement and Centrifuge Modelling Experiments
33.4: Angle of Repose
33.5: Principle of Normality
33.6: Yielding of Disturbed Saturated Soils in Triaxial Stress Tests
33.7: Original Cam-Clay Model
33.8: Liquefaction
33.9: Wet Sand Drying Underfoot
33.10: Summary
33.11: Questions

This page titled 33: Granular Materials is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.1: Introduction

For a wide range of engineering materials, their processing requires that they are in powder form or granular form at some stage. For example, recycled plastics are routinely pelletised and nickel ore in granular form is routinely transported by ship containers across the Pacific Ocean from the Philippines, the largest exporting country of nickel ore, to China, where it is used in the manufacture of stainless steel. Soils are routinely piled into heaps during the construction of new roads and bridges and to conceal industrial plants.



Pelletised recycled plastic (image from: http://www.plasgranltd.co.uk/how-is-plastic-recycled/)

It is vitally important to understand the conditions under which soil and heaps of material in granular form remain stable because the conditions under which they are no longer stable can have devastating consequences.

For example, under suitably adverse conditions in the transport of bulk cargo such as nickel ore and iron ore in ship containers, the cargo can transform abruptly from a solid state to an almost fluid state, i.e., it can liquefy. If this occurs, the stability of the vessel transporting the cargo will be affected. The consequences of this can be dramatic – the ship's structure can be damaged and, under severe conditions, the ship can capsize with the loss of life. Fortunately, there are now operational guidelines for the transport of mineral ores which will make the incidence of such catastrophic events much less in the future.

Many countries around the world are susceptible to landslides and mudslides where, as a consequence of severe rainfall, the pore pressure in the soil rises so that the soil is unable to bear both an equal-all-round compressive spherical stress and any shear stress – the soil becomes a slippery mess. The consequences can be catastrophic to local communities with loss of life, unless areas at risk of mudslides and landslides can be evacuated beforehand.

A third example is that of the stability of piles of granular material. It is not unusual to find such stockpiles failing suddenly, such as in the example below.



Failure of a pellet feed stockpile at the port of Vitoria in south-east Brazil (courtesy Evandro Moraes da Gama, Matheus Henrique de Castro, Carlos Gomes, Felipe Abbas da Gama, *Geomaterials*, **4**, 18-26 (2014))

Coal waste is also routinely piled into heaps. Again, under adverse circumstances, water can build up in the heap so that it is unable to bear both an equal-all-round compressive spherical stress and any shear stress. The Aberfan disaster in Wales (https://en.Wikipedia.org/wiki/Aberfan_disaster) occurred after a prolonged period of heavy rain in October 1966, so that liquefaction occurred in the material in a coal heap adjacent to the village of Aberfan. 144 people died as a consequence of the collapse of this coal heap.

Another way in which the flow behaviour of granular material is relevant in everyday life is when considering the nature of quicksand (https://en.Wikipedia.org/wiki/Quicksand). The two essential components of quicksand are a fluid, such as water, and fine particles, such as clay or fine sand. Central to its behaviour is its ability to liquefy, so that materials on top of it can sink into it without being fully submerged.

Clearly, therefore, understanding the mechanical behaviour of granular material is of intense practical interest, as well as being of interest in its own right.

This page titled 33.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

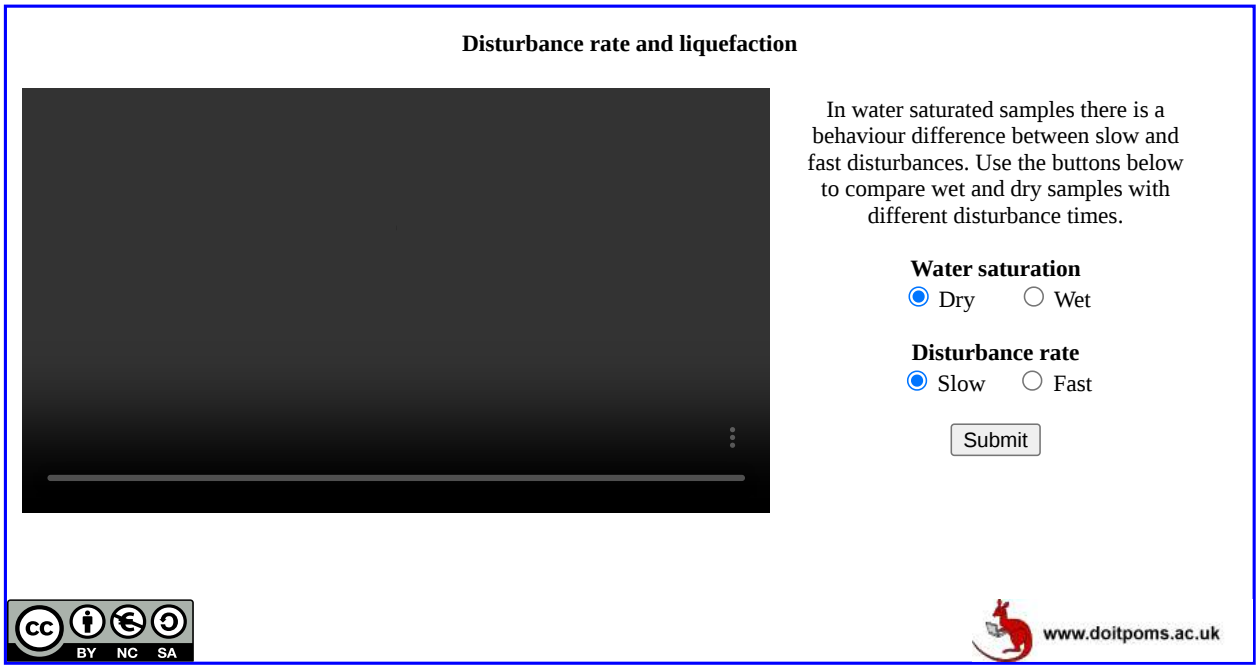
33.2: Terzaghi's Effective Stress Principle

Rocks disintegrate with weathering. In the processes of sedimentary geology when rivers discharge sediments into a pool, various grades of soil grain are deposited at different distances from the point of discharge. The ancient 'soft rock' termed the Cambridge gault clay in the city of Cambridge in England is a 20 m thick layer of silt or clay size soil grains, deposited on the bed of an ancient ocean. It now outcrops around Girton, just west of Cambridge. Above it, when the ancient warm sea bed was far away from points where rivers were depositing sediment, coral could grow. This coral formed the chalk rock that extends far across Europe. Locally near Cambridge, it outcrops in the Gog Magog Hills and along a line through Newmarket, Royston and Luton. There are ripples in the chalk that rise up to form the North Downs and dip in the South Downs. This chalk rock forms the White Cliffs of Dover. Vertical cracks form when lateral pressure falls in this chalk. These cracks can leave a heavy cliff face resting on a foundation of gault clay that, when it fails, lets the cliff face fall into the English Channel.

Anywhere that a sediment of strong durable grains accumulates as a uniform aggregate of sand or gravel that can be excavated for use in construction, such aggregate will be mixed with water and cement powder as a component of mortar in brickwork, or in structural concrete reinforced with mild steel bars. Where natural ground is soft enough, ungraded soil can be excavated, hauled to site and spread out in layers to be compacted with rollers to build up large road or dam embankments. In natural or compacted soft ground the strong soil grains form an effectively stressed aggregate structure in which forces are transmitted from grain to grain through the ground. In a volume of ground part of the volume is occupied by solid soil grains and the pore space between grains in saturated ground contains incompressible pore water. Whenever soft ground is loaded there will be pore pressure gradients. Terzaghi's primary consolidation theory analysed transient flow of ground water and surface settlements.

Terzaghi’s effective stress principle applied to sand recognises that sand needs to have strength if a slope of sand is not to slump. A slope can only be stable if there are intergranular compressive stresses. The total compressive stress that is applied normal to a particular plane, σ, is equal to $\sigma' + u$, where σ' is an effective compressive stress normal to a plane and u is pore water pressure. This is Terzaghi’s effective stress principle.

This principle is illustrated by the behaviour of sand in the two bottles shown on the section titled Angle of Repose and also below. The air can easily move from pore space to pore space in the sand aggregate, but the water will take time to do so.



These two bottles show that if there is time for the pore water pressure to diffuse (i.e., there is drainage), the same slope at repose is achieved with and without the presence of water. If there is not enough time, liquefaction will occur, as could happen in the rocking of a bulk carrier, i.e., a merchant ship designed to transport unpackaged bulk cargo such as metal ores.

This page titled 33.2: Terzaghi's Effective Stress Principle is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.3: Ground Settlement and Centrifuge Modelling Experiments

In a mud slide, pore pressure does not have time to drain because the grains of earth are small and the pathways for drainage between the grains are too narrow. Therefore, under suitably adverse conditions, gravity can cause a mud slide where there is soft clay and where structures for building such as a road embankment are built.

In general with disturbed soil, as the pore water drains and the grains of earth are able to grow into close contact, settlement of the ground surface occurs over time. Where tunnels are constructed in soft clay beneath urban areas, there is the possible risk that above these tunnels, settlement of the ground surface can cause problems such as the partial or total collapse of buildings at street level.

The engineering problem of ground settlement can be modelled at reduced scale and increased acceleration using a large centrifuge with a hopper which places a sand embankment on a clay model foundation. Since mathematically, settlement is a diffusion problem, the shearing in compression of the clay over time will take place in centrifuge experiments at a reduced time scale of $1/n^2$ where n in the multiple in terms of gravitational force g to which the soil is subjected in the experiments.

Therefore, in a centrifuge experiment in which testing is undertaken at $100g$, the reduction in time is a factor of 10^4. Hence, a suitably designed seven hour centrifuge experiment at $100g$ can enable reliable predictions to be made about ground settlement over a period equivalent to 8 years at ground level on the Earth under g.

This page titled 33.3: Ground Settlement and Centrifuge Modelling Experiments is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.4: Angle of Repose

When a container full of granular matter is poured onto a flat horizontal surface from a point source such as a funnel, it will form a conical pile of material. This conical pile will have a characteristic angle of repose, or equivalently, slope of repose. The angle of repose is the angle between the horizontal surface and the sloping surface of the pile. The tangent of this angle is the slope of repose.

https://www.doitpoms.ac.uk/tlplib/gr...deos/large.mp4

https://www.doitpoms.ac.uk/tlplib/gr...deos/small.mp4

It is found experimentally that the angle of repose is determined by a number of factors. One obvious factor is friction, and a second is cohesion caused by the presence of liquid bridges between granules which enable the granules to stick together.

Simple experiments with screw-top glass bottles half-full of fine grain loose sand are able to show various aspects of the flow behaviour of granular materials. In the videos below there are two bottles of sand, one half-full of sand and half-full of air, and the second half-full of sand and half-full of water.

If, to begin, both bottles are rolled slowly on their sides on a horizontal table, the sand in both bottles will move to form packings of the granules with flat horizontal surfaces.

If the bottle half-full of sand and half-full of air is tilted to stand on its end, it can be seen that the sand will form a slope with a particular angle of repose. If the bottle is further tilted in the same direction, the assembly of granules on the slope will be seen to flow in such a way that the individual granules *rotate* as they flow down the slope, rather than *slide down* the slope in order to maintain the angle of repose.

If the bottle half-full of sand and half-full of water is tilted to stand on its end, the outcome depends on the timescale over which the tilt occurs.

If the timescale is short, e.g., less than a second, the surface of the sand is disrupted in such a way that the granules pack down eventually to a horizontal surface so that the surface behaves as though it is a liquid.

If the tilt occurs gradually, e.g., over a period of a few seconds, it can be seen that the slope attained by the sand when the bottled is tilted to stand on its end is the same within experimental error as the bottle half-full of sand and half-full of water. If the bottle is further tilted slowly, the assembly of granules on the slope behaves as they do in air.

Disturbance rate and liquefaction



In water saturated samples there is a behaviour difference between slow and fast disturbances. Use the buttons below to compare wet and dry samples with different disturbance times.

Water saturation

(•) Dry ( ) Wet

Disturbance rate

(•) Slow ( ) Fast

Submit

CC BY NC SA

www.doitpoms.ac.uk

Angles of repose



Use the buttons below to compare the repose angle for samples of different grain sizes

Grain size

(•) Small ( ) Large

Material

(•) Recycled glass ( ) Recycled vitreous china

Submit

CC BY NC SA

www.doitpoms.ac.uk

In clays such as London clay consolidated over some 50 million years (en.Wikipedia.org/wiki/London_Clay), large lumps of sheared blocks of clay known as 'greasy-backs' can fall from faces of tunnels during excavation. The bottom surfaces of these sheared blocks consists of regions of soil where local expansion has occurred and where the failure process on shearing is a consequence of the rotation of granules in these regions, as if they were tumbling down an angle of repose, rather than the translation of the granules. Examination of the surfaces of these greasy-backs after failure shows that they have the characteristics of plasticised clay paste resembling heavy engine grease (C.N.P. Mackenzie, Traditional Timbering in Soft Ground Tunnelling: A Historical Review, British Tunnelling Society, 2014).

Predictions of the angle of repose

Predicting the angle of repose for a particular family of granules is not straightforward: most angles of repose are acquired from experimental measurements, such as those quoted in the Wikipedia entry on 'Angle of repose' (en.Wikipedia.org/wiki/Angle_of_repose). For many granular materials a typical angle of repose will be between

25° and 40°.

For smooth spherical particles all of the same size, a very simple model can be used to appreciate how the packing of spheres in three dimensions has consequences for the stability of slopes. Suppose three identical spherical particles on a slope of angle θ relative to the horizontal pack together so that they form the base of a regular tetrahedron if a fourth identical sphere is added on top of these three particles. Depending on the orientation φ of the base of the tetrahedron with respect to the slope, the tetrahedron will be stable if the angle θ is increased from zero (when the particles are on a horizontal base) to a critical angle θ_c at which the fourth sphere rolls away down the slope from the three identical spheres (all presumed to be (just) held by friction on the slope at this angle).

At a particular angle of φ and θ, simple geometry shows that the top sphere is only stable if the vector defining the gravitational force points through the projection of a suitable base triangle on the horizontal plane. The corners of this base triangle are defined by the positions of the centres of mass of the three identical spheres on the base of the tetrahedron.

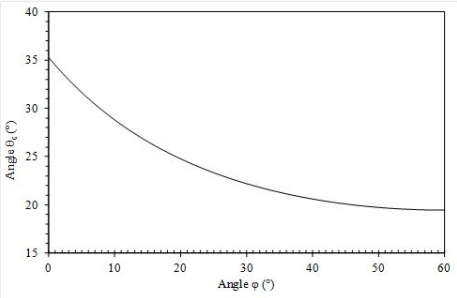
For a particular φ, the maximum angle of stability for this geometry is

$$\theta_c = arctan\frac{1}{2sqrt2cos(60^\circ - \varphi)}) \tag{33.4.1}$$

A proof of this formula is given here.

Hence, when φ = 0°, when one of the corners of the base triangle is at the lowest point on the slope θ_c = 35.26°, and when φ = 60°, when one of the sides of the base triangle is at the lowest point in the slope, θ_c = 19.47°.

A graph of θ_c as a function of φ is shown below.



If it is assumed that the orientation of bases of such tetrahedra are random on a pile of granular matter, then the observed angle of repose for smooth non-cohesive particles can be argued to be the average value of θ_c over the interval 0° < φ < 60°, i.e., 23.8°, determined numerically using Simpson's rule or the Trapezium rule. This equates to an effective coefficient of friction of tan θ_c of 0.44.

This page titled 33.4: Angle of Repose is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.5: Principle of Normality

The principle of normality follows from a detailed consideration of the yield surfaces of work-hardening materials and ideal plastic materials by Daniel Charles Drucker in 1951 in a paper entitled 'A more fundamental approach to plastic stress strain relations', Proc. 1st U.S. Nat. Congr. Of Appl. Mech., pp. 487-491. In this paper Drucker established that these yield surfaces must be convex and that the vector sum of the plastic strain increments (or flow increments) at any point on the yield surface is normal to the yield surface. This is the principle of normality.

To illustrate what this means in practice for an ideal plastic metal, in which hydrostatic stress does not cause plastic deformation (see here), we can consider the von Mises yield criterion in plane stress.

In plane stress in principal stress space with principal stresses σ_1 and σ_2 and where $\sigma_3 = 0$, the von Mises yield surface is defined by the equation

$$\sigma_1^2 + \sigma_2^2 - \sigma_1\sigma_2 = 1 \tag{33.5.1}$$

if the uniaxial yield stress, *Y*, is take to be unity (see here).

At and beyond yield, the flow behaviour of ideal plastic metals is governed by the Lévy-Mises equations (see here):

$$\frac{\delta\varepsilon_1}{\sigma_1 - \frac{1}{2}(\sigma_2 + \sigma_3)} = \frac{\delta\varepsilon_2}{\sigma_2 - \frac{1}{2}(\sigma_3 + \sigma_1)} = \frac{\delta\varepsilon_3}{\sigma_3 - \frac{1}{2}(\sigma_1 + \sigma_2)} \tag{33.5.2}$$

for principal plastic strain increments $\delta\varepsilon_1$, $\delta\varepsilon_2$ and $\delta\varepsilon_3$ parallel to the principal stresses σ_1 and σ_2 and σ_3 respectively.

Examining the equation for the von Mises yield surface, the tangent at a point (σ_1, σ_2) on the yield locus can be found be differentiating equation (1) implicitly:

$$2\sigma_1 d\sigma_1 + 2\sigma_2 d\sigma_2 - \sigma_2 d\sigma_1 - \sigma_1 d\sigma_2 = 0 \tag{33.5.3}$$

Rearranging this,

$$\frac{d\sigma_2}{d\sigma_1} = \frac{2\sigma_1 - \sigma_2}{2\sigma_2 - \sigma_1} \tag{33.5.4}$$

Now, examining equation (2) for the situation where $\sigma_3 = 0$, it follows that

$$\frac{\delta\varepsilon_2}{\delta\varepsilon_1} = \frac{2\sigma_2 - \sigma_1}{2\sigma_1 - \sigma_2} \tag{33.5.5}$$

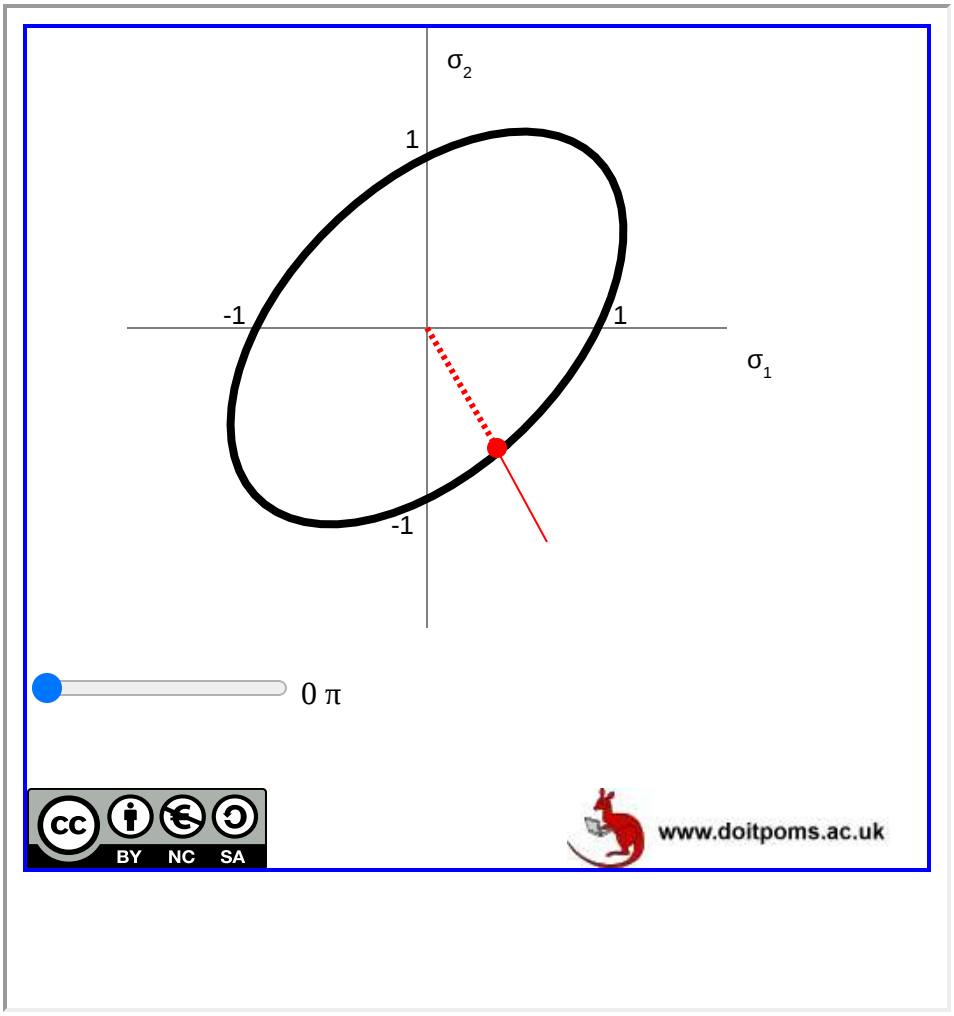
so that

$$\frac{\delta\varepsilon_2}{\delta\varepsilon_1} \cdot \frac{d\sigma_2}{d\sigma_1} = -1 \tag{33.5.6}$$

Hence, in words, the product of the gradient of the tangent at a point (σ_1, σ_2) on the yield surface and the gradient of the vector [$\delta\varepsilon_1$, $\delta\varepsilon_2$] defining the plastic flow increments in the (σ_1, σ_2) plane is minus one, i.e., these two gradients are perpendicular.

Since the third plastic strain increment $\delta\varepsilon_1$ is parallel to the principal stress σ_3, it follows that the vector sum of plastic flow increments [$\delta\varepsilon_1$, $\delta\varepsilon_2$, $\delta\varepsilon_3$] is normal to the yield surface at (σ_1, σ_2) , i.e., the principle of normality is proved for the von Mises yield criterion.

The animation below shows how in two dimensions in (σ_1, σ_2) space, the vector defining the plastic flow increments (in red) is normal to the von Mises yield surface (in black). The angle can be varied by moving the cursor in the bottom left of the animation.



In the language of soil mechanics, the principle of normality is known as the associated flow rule (see, for example, A. Schofield, *Disturbed Soil Properties and Geotechnical Design*, Thomas Telford Ltd., London, 2005, p. 91). A further result which follows from the convexity of yield surfaces and this principle is that

$$d\sigma_{ij}d\varepsilon_{ij}^{P} \geq 0 \tag{33.5.7}$$

where $d\sigma_{ij}$ are stress increments created by an external agency which produce very small plastic strain increments in strain, $d\varepsilon_{ij}^{P}$ (as well as elastic strain increments). In words, this is a statement that plastic work done by an external agency is always positive.

Applying equation (7) to a triaxial stress test on a soil where q is the axial compressive stress, σ_a, minus the radial compressive stress, σ_r, and where p' is the mean effective compressive stress generates the condition

$$dp'dv + dqd\varepsilon \geq 0 \tag{33.5.8}$$

where dv is the incremental change in the specific volume of the soil aggregate and where dε is the incremental change in pure distortion. The specific volume is the ratio of the volume occupied by soil to that which it would occupy if the voids were eliminated. In soil mechanics the convention is that dp', dq , dv and dε are all positive.

Equation (8) is relevant to the Original Cam-Clay model of how soils at the critical state and on the 'wet' side of the critical state fail.

This page titled 33.5: Principle of Normality is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.6: Yielding of Disturbed Saturated Soils in Triaxial Stress Tests

A schematic yield surface ABCDEFG of a soil plotted as a function of two parameters: η and v_λ is shown in the diagram below.

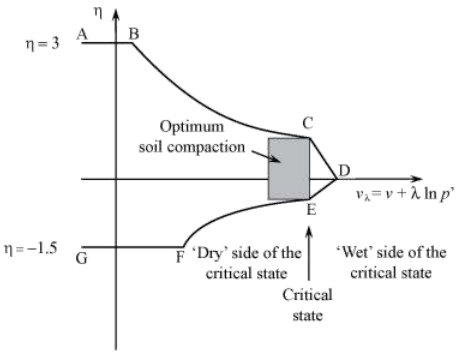
The vertical axis is η, the stress obliquity, defined as q/p', where in a triaxial test on a cylinder of soil, q is the axial compressive stress, σ_a, minus the radial compressive stress, σ_r, and p' is the mean total compressive stress, p, minus the pore water pressure, u. p' is also known as the mean effective compressive stress. It is useful to note that the effective axial and compressive stresses are $\sigma_a' = \sigma_a - u$ and $\sigma_r' = \sigma_r - u$ respectively, and that

$$q = \sigma_a - \sigma_r \tag{33.6.1}$$

and

$$p' = \frac{\sigma_a + 2\sigma_r}{3} - u \tag{33.6.2}$$

The convention in soil mechanics is to take compressive stresses as positive quantities. This is an important difference from the convention used in the TLP Stress Analysis and Mohr's Circle.



The horizontal parameter v_λ is a linear function of the specific volume, v, and the logarithm of the mean effective compressive stress, p'. The specific volume is the ratio of the volume occupied by soil to that which it would occupy if the voids were eliminated. Therefore, values of v of 2 are typical for soils.

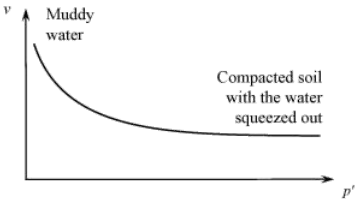
We now need to define what we mean by the 'Dry' side of the critical state and the 'Wet' side of the critical state for soils. 'Wet' means that when you shear the soil, grains expel water so that the soil feels sticky to the touch. 'Dry' does not literally mean dry: it means the pores are full of water, but when you shear the soil, the grains ride over one another - this is called dilation. The void space increases and fluid is taken into the soil. Remoulding the soil in your hands will dry your hands. At the critical state, the soil neither dries your hands nor feels sticky to the touch.

For soils at the critical states C and E in the above diagram and soils on the 'wet' side of C and E, it is found experimentally that there is a linear relationship between the specific volume of the soil aggregate, v, occupied by the soil and the logarithm of the mean effective compressive stress. Hence, a graph of v against ln p' for such soils is experimentally found to fit an equation of the form

$$v = v_\lambda - \lambda ln p' \tag{33.6.3}$$

where by convention p' is usually normalised with respect to a standard stress of 1 kPa. Micromechanically, such an equation in v − ln p' space can be rationalised in terms of particle breakage as p' increases in crushable aggregates, and the fractal nature of the process of breakage in such aggregates, in which the tensile strength of the smallest particles within an assembly of particles constituting an aggregate of soil grains determines the yield stress of the aggregate (G.R. McDowell and M.D. Bolton, 'On the micromechanics of crushable aggregates', *Géotechnique* **48**, 667-679 (1998)).

Physically, it can be appreciated that mud saturated with water has a high v relative to the same mud with water squeezed out of it as a consequence of an increase in p', and that there will be a lower asymptotic limit of v achieved by mud at large mean effective compressive stresses with the water squeezed out. Schematically, we might imagine the following:



Each part of this yield surface in (v_λ, η) space will now be explained in more detail, going clockwise around the yield surface, starting with A and finishing with G:

Line AB

When $\sigma_a >> \sigma_r$, $\eta \rightarrow 3$, as $\sigma_r/\sigma_a \rightarrow 0$. Under these circumstances, axial compression will cause a test cylinder of soil to split on axial planes, just as logs split for firewood. Open cracks are therefore produced in this part of (v_λ, η) space.

Curve BC

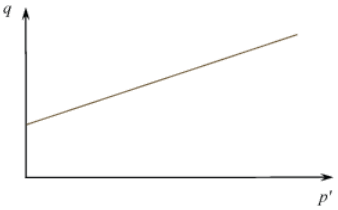
In this region of (v_λ, η) space the Mohr-Coulomb failure criterion is assumed to hold:

$$\tau = \tau * + \sigma_n' tan\phi' \tag{33.6.4}$$

where in this equation the compressive stress is taken here to be positive using the conventions of soil mechanics, and where the dash denotes that pore water pressure has been taken into account, so that is an effective normal compressive stress.

This criterion is discussed further on the TLP page https://www.doitpoms.ac.uk/tlplib/metal-forming-1/yield_non_metals.php .

For the interpretation of this criterion for triaxial stress tests on soils using the convention that in soil mechanics, pressures are taken to be positive, it is reasonable to map Mohr-Coulomb behaviour in (τ,) space onto (q, p') space so that a graph of q against p' is taken to be of the form



$$q = q_0 + mp' \tag{33.6.5}$$

A model for why this is reasonable is shown here.

There are a number of caveats in making the assumption that the Mohr-Coulomb behaviour holds. The most important caveat is that the Mohr-Coulomb criterion does not have a physical base – it is an empirical equation based on the interpretation of experimental observations. In addition, in the Mohr-Coulomb yield criterion the intermediate principal stress is not required, whereas it clearly is required for the interpretation of triaxial stress tests.

Notwithstanding these caveats, rearranging (1), we have:

$$\eta = \frac{q}{p'} = \frac{q_0}{p'} + m \tag{33.6.6}$$

If we now define ln p' to be x, then $p' = e^x$, and so (2) can be rearranged in the form

$$\eta = q_0 e^{-x} + m \tag{33.6.7}$$

Hence, a graph of η against x has a gradient of , so that the gradient is negative, decreasing in magnitude exponentially as x increases. Hence plotting η against v + λ the Mohr-Coulomb yield criterion has the same form, since v is taken to be a constant and λ is a dimensionless constant.

Eventually, as ln p' increases while η decreases, a critical state C is reached where physically the soil is sufficiently loosely packed that there is no specific volume change during a triaxial stress test, i.e., the soil neither contracts nor dilates. Hence, there is a critical value of v_λ at C. This critical value of v_λ is given the symbol Γ.

During both clockwise and anticlockwise shearing, soils dilate for $v_\lambda < \Gamma$, i.e., on the dry side of critical and contract for $v_\lambda > \Gamma$, i.e., on the wet side of critical.

Line CD

Line CD represents the behaviour of soft soil during plastic yielding and flow at C, and on the wet side of the critical state C, through the Original Cam-Clay model. The proof that CD is a straight line can be found here. As shown here, C is defined by a point (Γ, M) in (v_λ, η) space, while, as will also be shown below, D is defined by a point (Γ + λ - κ, 0).

In (v_λ, η) space, CD is the straight line

$$(\lambda - \kappa)\eta = M v_\lambda - M(\Gamma + \lambda - \kappa) \tag{33.6.8}$$

It is useful to have an idea of the values of some of the parameters in this model. For a material like London Clay, M = 0.89, λ = 0.161, κ = 0.062 and Γ is 2.759 when p'_C is 1 kPa (A. Schofield, *Disturbed Soil Properties and Geotechnical Design*, Thomas Telford Ltd., London, 2005, p. 100).

Line DE

Line DE is the equivalent of line CD, i.e., Original Cam Clay yield behaviour but for when $\eta < 0$, i.e., for when $\sigma_a << \sigma_r$. The critical state E occurs at a position (Γ, M_E) in (v_λ, η) space where there is no physical reason for M_E and M to be the same.

Curve EF

Line EF is the equivalent of curve BC, i.e., the yield surface is defined by Mohr-Coulomb failure, but under circumstances where $\eta < 0$. At F a stress state is reached where $\sigma_a/\sigma_r = 0$.

Line FG

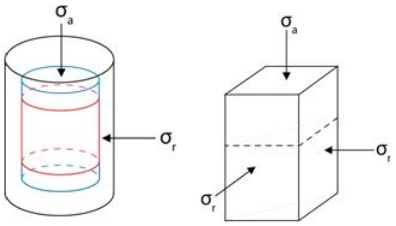
When $\sigma_a << \sigma_r$, $\eta \rightarrow -1.5$ is reached as $\sigma_a/\sigma_r \rightarrow 0$. Under these circumstances, radial compression causes cracking in planes perpendicular to the axis of a test cylinder: it can crack into many discs. This is termed spalling – it has the same nomenclature as that used to describe the failure in compression of thin films on substrates because of failure due to in-plane biaxial compressive stresses.

Finally, note that F and B occur at different values of v_λ: spalling occurs over a larger range of mean normal effective pressures than its equivalent process of cracking when $\sigma_a >> \sigma_r$.

This page titled 33.6: Yielding of Disturbed Saturated Soils in Triaxial Stress Tests is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.7: Original Cam-Clay Model

The original Cam-clay model (OCC) was developed by Andrew Schofield in the 1960s as a description of the behaviour of saturated soil and sands. It shows how, depending on water content, soils can fail by spalling or by plasticity and liquefaction.



Consider a cylinder of water-saturated sand in a triaxial testing regime as in the above figure.

The cylinder is subjected to the total axial stress, σ_a, and total radial stress, σ_r. It is more useful to work in terms of the effective stress which takes pore water pressure, u, into account

$$\sigma' = \sigma - u \tag{33.7.1}$$

This allows us to define the general mean effective compressive stress as

$$p' = \frac{1}{3}(\sigma'_1 + \sigma'_2 + +\sigma'_3) \tag{33.7.2}$$

this case $\sigma_a = \sigma_1$ and $\sigma_r = \sigma_2 = \sigma_3$ and so

$$p' = \frac{1}{3}(\sigma'_a + 2\sigma'_r) \tag{33.7.3}$$

A deviator stress, *q*, can also be defined, given by the equation

$$q = \sigma'_a - \sigma'_r \tag{33.7.4}$$

In triaxial yield testing it is found that sands and soils on the 'wet' side of the critical state yield on a ductile-plastic continuum. The plastic deformations arise as a change in the specific volume of the sample and a strain along its length. We can therefore define the following two strains:

Axial strain:

$$\delta\varepsilon_a = \frac{\delta l}{l} \tag{33.7.5}$$

Volumetric strain:

$$\frac{\delta v}{v} = \varepsilon_v = \delta\varepsilon_a + 2\delta\varepsilon_r \tag{33.7.6}$$

To find the triaxial shear strain, ε_s, we separate the axial and volumetric strains:

$$\begin{aligned} \delta\varepsilon_s &= \delta\varepsilon_a - \frac{1}{3}\delta\varepsilon_v \\ &= \frac{2}{3}\delta\varepsilon_a - \delta\varepsilon_r \end{aligned} \tag{33.7.7}$$

The work done per unit volume by elastic straining is

$$\begin{aligned} \delta W &= p'\delta\varepsilon_v + p\delta\varepsilon_s \\ &= \frac{1}{3}\sigma'_a + 2\sigma'_r\delta\varepsilon_a + 2\delta\varepsilon_r + \frac{2}{3}\sigma'_a - \sigma'_r\delta\varepsilon_a - \delta\varepsilon_r \\ &= \sigma'_a\delta\varepsilon_a + \sigma'_r 2\delta\varepsilon_r \end{aligned} \tag{33.7.8}$$

For work done by plastic straining at failure for soils at the critical state or wetter than the critical state, the relevant dissipation function is defined by the equation

$$p'\delta\varepsilon_v + q\delta\varepsilon_s = \delta W = Mp'\delta\varepsilon_s \tag{33.7.9}$$

where M is the general coefficient of friction. The work done against friction per unit volume is defined by this equation. The work done in producing a volume change does not explicitly appear in this model because it is a consequence of the interlocking of particles.

In OCC the associated plastic flow vector is locally orthogonal to the tangent of the yield locus so that

$$dp'\delta\varepsilon_v + dq\delta\varepsilon_s \geq 0 \tag{33.7.10}$$

In words, this equation is a recognition that the scalar product of the plastic flow normal to the yield locus at (*p*', *q*) and the incremental loads (*dp*', *dq*) causing failure at (*p*', *q*) must be positive.

To derive the OCC we combine Equations (8) and (9). First we divide equation (8) by p'$\delta\varepsilon_s$:

$$\frac{\delta\varepsilon_v}{\delta\varepsilon_s} + \frac{q}{p'} = M \tag{33.7.11}$$

Rearranging Equation (9) after setting the inequality to zero, we have:

$$\frac{\delta\varepsilon_v}{\delta\varepsilon_s} = -\frac{dq}{dp'} \tag{33.7.12}$$

so that eliminating $\delta\varepsilon_v/\delta\varepsilon_s$ in these two equations, we produce the equation

$$\frac{q}{p'} - \frac{dq}{dp'} = M \tag{33.7.13}$$

This is a differential equation on which we impose limits and introduce the stress ratio, $\eta = q/p'$. Differentiating η:

$$\begin{aligned} &\tfrac{d\eta}{dp} = \tfrac{1}{p'}\left(\tfrac{dq}{dp'} - \tfrac{q}{p'}\right) = -\tfrac{M}{p'} \\ &\Rightarrow \tfrac{d\eta}{M} + \tfrac{dp'}{p'} = 0 \end{aligned} \tag{33.7.14}$$

As we are finding the locus for 'wetter than critical' states the integral is as follows:

$$\begin{aligned} &\int_M^\eta \tfrac{1}{M}d\eta = -\int_{p'_c}^{p'} \tfrac{1}{p'}dp' \\ &\tfrac{\eta}{M} - 1 = -\ln\left(\tfrac{p'}{p'_c}\right) \\ &\tfrac{q}{Mp'} = 1 - \ln\left(\tfrac{p'}{p'_c}\right) \end{aligned} \tag{33.7.15}$$

Therefore, when *p*' is equal to p'_c, $\eta = M$. Also, when $q = 0$, so that the soil cannot withstand any shear stress at failure, $p' = e\,p'_c$,where e is the base of the natural logarithm, 2.71828 … .

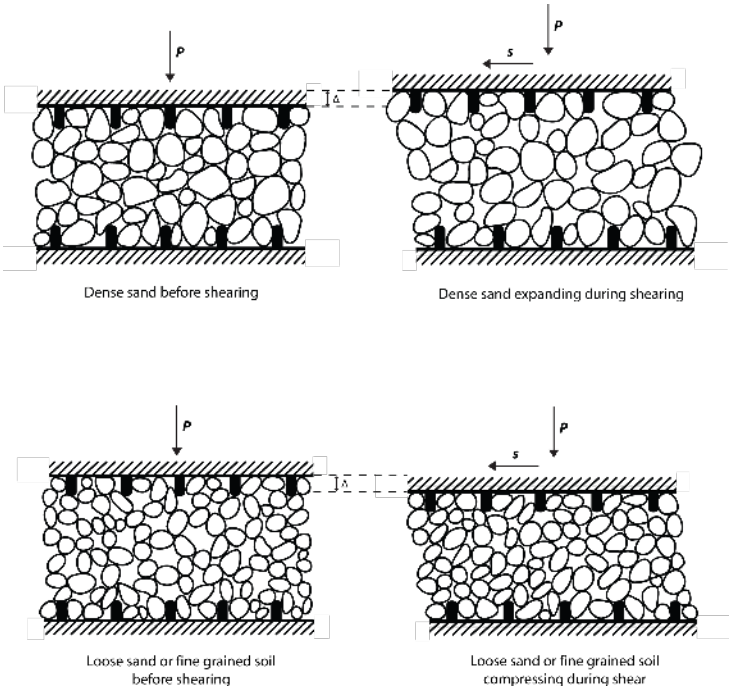
In practice $q = 0$ is actually a situation difficult to attain – see A. Schofield, *Disturbed Soil Properties and Geotechnical Design*, Thomas Telford Ltd., London, 2005, p. 106.

This page titled 33.7: Original Cam-Clay Model is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

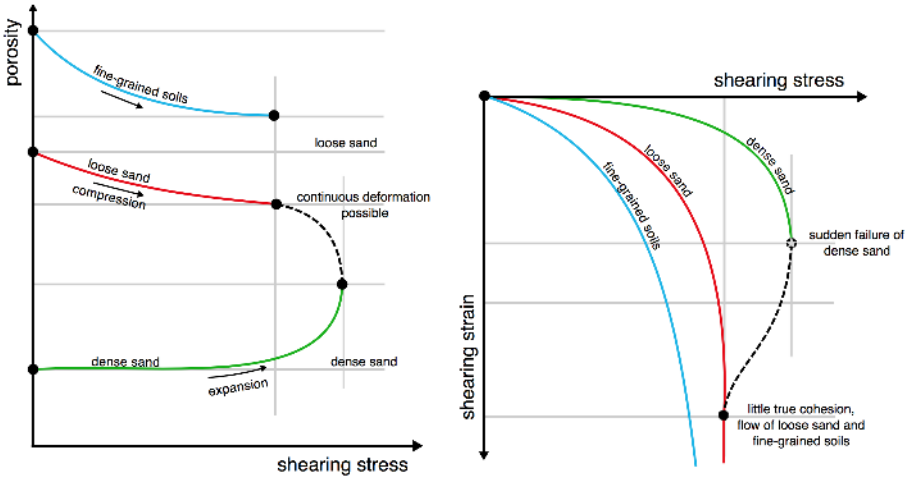
33.8: Liquefaction

The phenomenon of soil liquefaction describes failure of saturated or partially saturated soil that, upon application of an applied stress, behaves as a liquid when it fails. In the real world these failures usually result from earthquakes leading to landslides.

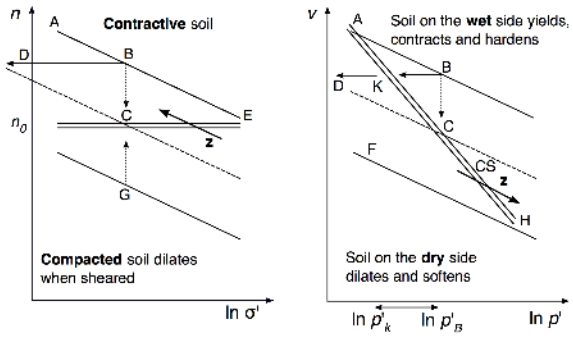
It was proposed by Casagrande that grain rotation can result in a change from an interlocking grain structure to a flowing grain structure. He tested this hypothesis by applying an impact force to a tank filled with loose sand. The aggregate remained solid until the instant the force was applied. The application of the force caused a high transient pore pressure to be generated within the aggregate, reducing the effective stress to near zero, so that liquefaction flow occurred.



A large body of soil is not safe from liquefaction unless it is compacted below the critical void value. This is due to the pore pressure changes during shear. Casagrande studied this using triaxial compression tests. He showed how with shear at a constant normal stress the peak strength (starred point) of dense sand increased as it expands. Loose sand showed a fall in porosity and dense sand showed an increase. After the displacement both loose and dense sand have reached a peak porosity n_0 where the sand flows under constant shear stress.



This critical porosity n_0 is independent of the effective pressure. If silty sand is compacted below n_0, it will not be at risk of liquefaction. Compression with drained shear is the most effective way of reducing the risk of liquefaction of an existing sample as the water content is also reduced.



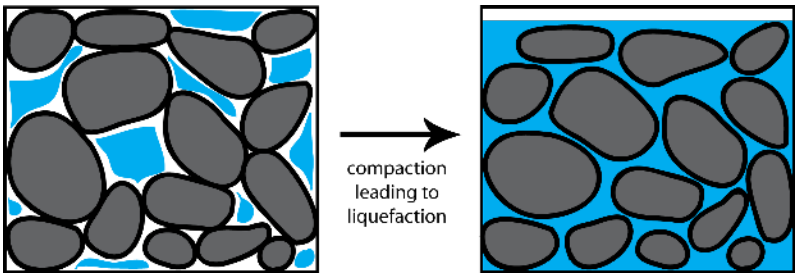
In the left-hand graph above, the dashed line BC represents a drained test and the solid line BD an undrained test. The double horizontal line in the left-hand graph is the critical void ratio n_0. In the graph on the right, the double line at an angle to both the horizontal and the vertical is a λ-line representing soil at critical state, CS, as a function of the specific volume of aggregate, v, and the logarithm of the effective compressive stress p' applied to the aggregate relative to a standard compressive stress state of 1 kPa.

For the critical void ratio theory, compression with large deformation depends on the initial state, while in CS theory an aggregate of grains does not retain any record of its initial structure. The states predicted under Casagrande's model differ from those predicted by the line AH in CS theory. Under CS theory the dashed line still follows the same path BC but the solid line instead stops short at BK and does not continue from K to D. In 1975 Casagrande revised his theory of constant critical porosity to adopt the CS line which combines specific volume v and effective pressure p'. The slope of the double λ-line in CS is greater than the slope of the elastic κ compression lines, shown in the right-hand graph as lines parallel to FH.

The bold arrows marked **z** show the main differences between the two initial theories. In the Casagrande model, the vector **z** shows that the reduction of pressure is unsafe as it results in a swelling of the aggregate. This moves from a safe dilative state to an unsafe contractive state. There is a likelihood of liquefaction in the contractive state. In the CS theory, the vector **z** shows that increasing the effective pressure moves the aggregate from a safe dilative state to an unsafe contractive state. This means that even very dense sand can liquefy if the effective pressure is high enough; this is especially relevant for aggregates with a high water content.

Concerns of liquefaction in the real world

As well as natural disasters there are other situations where liquefaction needs to be taken into account. Liquefaction in bulk cargo carriers has a catastrophic impact on ship stability that can result in large losses of life. Nickel ore is one of the most dangerous granular materials to transport due to the fine grain size and moisture content that sits in the pores. There are many precautions and safety checks before a cargo is approved to leave port. Before the cargo experiences compaction at sea it appears in a solid dry state due to frictional forces between grains.



During transport, the motions of a ship on the waves cause compaction of the cargo so that the pore size between the grains is reduced. Since the grains themselves can be very fine, there can be limited permeability, and so very little water drainage can occur. Once the pore water has been compressed, the pressure will start to push the grains apart, causing the grains no longer to be in contact, so the surface friction is lost. The shear strength of the cargo is reduced to zero and so it flows out to a level surface when

liquefaction has occurred. Fortunately, this is often a transient phenomenon because the cargo can settle back into a more dense state.

Some nickel ores have a clay-like consistency. In these materials liquefaction is instead a fatigue process where after a certain number of stress cycles the material will fail as the cohesion of the material collapses. This can result in the entire cargo becoming liquid simultaneously. As the cargo does not often fill the entire hold, there is sufficient space for the liquefied material to flow and destabilise the ship. Once this has occurred, it is very difficult to regain stability and it can result in the capsize of a ship in a matter of minutes.

This page titled 33.8: Liquefaction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.9: Wet Sand Drying Underfoot

When walking on wet sand, you may have noticed that the surface goes dry under your feet. This is not due to water being squeezed out into the surrounding sand. Instead, the water is absorbed into the sand directly under your feet. This is because you are applying a pressure to the sand under your feet as you walk. When pressure is applied to this sand under your feet, water flows into the pores that have expanded due to the relative shearing of sand particles, as shown in the diagram at the top of this page. Once your foot has been removed, the disturbed grains will settle back into a more dense arrangement and the water will be pushed out of the now smaller pores, and the sand appears wet again.

https://www.doitpoms.ac.uk/tlplib/gr...experiment.mp4

Video provided with kind permission of Ruben Meerman, 'The Surfing Scientist'. rubenmeerman.com/

The video is also available directly from YouTube: https://www.youtube.com/watch?v=1ssTdOt9DKI&feature=youtu.be

This page titled 33.9: Wet Sand Drying Underfoot is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.10: Summary

After working through this teaching and learning package, you should:

- Understand the concept of dilatation
- Have a feeling for 'typical' angles of repose for granular materials
- Recognise how friction, cohesive forces and the shapes of grains all affect the angle of repose of a pile of granular material
- Be able to appreciate the various ways in which soil can fail in compression and how the choice of failure mode is a function of the water content of the soil
- Be able to recognise the factors which cause soil to fail catastrophically and flow like a liquid – the phenomenon of liquefaction
- Recognise that quicksand is dangerous, but that if you ever happen to fall into quicksand, you can reassure yourself that you will not be sucked beneath the surface.

Going Further

Books

- A. Mehta, *Granular Physics*, CUP, 2007.
- A. Schofield, *Disturbed Soil Properties and Geotechnical Design*, Thomas Telford Ltd., London, 2005.
- A. Schofield and C.P. Wroth, *Critical State Soil Mechanics*, McGraw-Hill, Maidenhead, 1968.
- D.M. Wood, *Soil Behaviour and Critical State Soil Mechanics*, CUP, 1990.

Websites

- and
- en.Wikipedia.org/wiki/Soil_liquefaction

  are both useful web sites to help to go further into the subject of granular flow and to cover aspects of granular materials not covered in detail in this TLP, such as how forces are transmitted between particles in granular media.

Other resources

For those interested in going much deeper into the subject of granular flow, these research references should be useful:

R. Albert, I. Albert, D. Hornbaker, P. Schiffer and A. Barabási, 'Maximum angle of stability in wet and dry spherical granular media', *Phys. Rev. E* **56**, R6271-R6274 (1997).

A. Barabási, R. Albert, and P. Schiffer, 'The physics of sand castles: maximum angle of stability in wet and dry granular media', *Physica A* **266**, 366-371 (1999).

M.D. Bolton, 'The strength and dilatancy of sands', *Géotechnique* **36**, 65-78 (1986).

H.W. Chandler and D.E. Macphee, 'A model for the flow of cement pastes', *Cem. Conr. Res*. **33**, 265-270 (2003).

B.-P. Dai, J. Yang and C.-Y. Zhou, 'Micromechanical origin of angle of repose in grnular materials', *Granular Matter* **19**: art. 24 (2017).

P.A. Arias García, R.O. Uñac, A.M. Vidales and A. Lizcano, 'Critical parameters for measuring angles of stability in natural granular materials', *Physica A* **390**, 4095-4104 (2011).

T.C. Halsey and A.J. Levine, 'How sandcastles fall', *Phys. Rev. Lett*. **80**, 3141-3144 (1998).

B.C. Johnson, C.S. Campbell and H.J. Melosh, 'The reduction of friction in long runout landslides as an emergent phenomenon', *J. Geophys. Res.: Earth Surf*. **121**, 881-889 (2016).

A. Khaldoun, E. Eiser, G.H. Wegdam and D. Bonn, 'Liquefaction of quicksand under stress', *Nature* **437**, **635** (2005).

A. Mehta and G.C. Barker, 'The dynamics of sand', *Rep. Prog. Phys*. **57**, 384-416 (1994).

C.N.P. Mackenzie, Traditional Timbering in Soft Ground Tunnelling: A Historical Review, British Tunnelling Society, 2014.

G.R. McDowell and M.D. Bolton, 'On the micromechanics of crushable aggregates', *Géotechnique* **48**, 667-679 (1998).

G.R. McDowell, M.D. Bolton and D. Robertson, ‘The fractal crushing of granular materials’, *J. Mech. Phys. Solids* **44**, 2079-2102 (1996).

S. Nowak, A. Samadan and A. Kudrolli, ‘Maximum angle of stability of a wet granular pile’, *Nature Physics* **1**, 50-52 (2005).

D.A. Robinson and S.P. Friedman, ‘Observations of the effects of particle shape and particles size distribution on avalanching of granular media’, *Physica A* **311**, 97-110 (2002).

C.M. Sands, A.R. Brown and H.W. Chandler, ‘The application of principles of soil mechanics to the modelling of pastes’, *Granular Matter* **13**, 573-584 (2011).

Z.Y. Zhou, R.P. Zou, D. Pinson and A.B. Yu, ‘Angle of repose and stress distrbution of dandpiles formed with ellipsoidal particles’, *Granular Matter* **16**, 695-709 (2014).

This page titled 33.10: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

33.11: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which of these factors do not influence the angle of repose taken up by a pile of granular particles?

a	Shape
b	Friction
c	Density
d	Cohesion

Answer

C. Density. Shape is important - the more 'angular' the particles, the more likely the angle of repose will be higher. Conversely, the more rounded the particles, the lower the angle of repose, all other factors remaining unchanged. Friction is vitally important - without friction, even a simple arrangement such as a regular tetrahedron on a flat surface is unstable. Cohesion is also important, and is particularly important if there is a steep angle to be maintained by a pile of granular particles, e.g., when building a sandcastle.

Which of the following systems of granular materials exhibit dilatancy?

a	Soil which when picked up from the ground is dry to the touch
b	Tomato ketchup
c	Sparkly nail polish
d	Soil which when picked up from the ground is wet to the touch

Answer

D. Soil which when picked up is dry to the touch has too low a water content to be able to dilate. When packed together, this type of soil will take in water to fill the pores between the grains of soil. Shearing this soil will enable grains to move relative to one another so that the packing efficiency is improved and so there is a decrease in the overall porosity of the soil aggregate. Tomato ketchup is a suspension of edible solids within water at a level typically of 25% by volume. It is a shear-thinning fluid, meaning that its viscosity decreases as it is sheared, in contrast to shear-thickening fluids such as dense wet sand, the viscosity of which increases as it is sheared. Sparkly nail polish has a suspension of particles such as mica, BiOCl (bismuth oxychloride) and aluminium powder, but the amount of solids per unit volume in the fluid is too low to cause the phenomenon of dilatancy - nail polish is in essence a liquid containing small granular particles, rather than being a wet assembly of granular material.

Inspired by this TLP, you purchase some corn starch and mix it with water in the ratio by volume of 1 part water to 2 parts corn starch to make a thick paste of corn starch which you pour into a large plastic container. The surface of the corn starch is then hit

quickly with a rubber mallet so that the time of contact with the surface is 100 milliseconds or so. What happens to the surface of the corn starch?

a	The rubber mallet bounces back from the surface. Everything happens too quickly to work out what is going on!
b	Water retreats from where the impact is made by the rubber mallet because of the phenomenon of dilatancy, so that in this region the surface goes from shiny to matt after the hammer has struck. The rubber mallet causes a small impression in the corn starch which over time gradually disappears as water returns. Eventually, the surface returns to being shiny and it is difficult to determine where the rubber mallet had struck.
c	The rubber mallet causes a mess everywhere because it has gone deep into the paste initially, causing paste to be ejected from the surface into the surrounding environment, and has then been forcibly removed.
d	The surface of the cornstarch is lowered where the rubber mallet has struck, but it returns to its initial state quickly. There is no change to the shininess of the surface anywhere.

Answer

B. However, this answer will be heavily dependent on making a sufficiently dense paste of corn starch. If the paste is not dense enough, so that the density of corn starch particles per unit volume is not high enough, the rubber mallet will be able to go deep into the paste and it will splash (answer (c)). Answers (a) and (d) would suggest that you have not paid enough attention when undertaking the experiment!

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

Ginkaku-ji ('The Silver Pavilion') is a Zen temple in Kyoto, Japan renowned for its meticulously maintained raked dry sand garden called Ginshadan known as the 'sea of silver sand' next to which there is a massive 1.8 m high sand cone platform called Kōgetsudai said to symbolise Mount Fuji.



Kōgetsudai



Ginshadan

The 'angle of repose' of Kōgetsudai is impressively high, as are the angles of repose in Ginshadan. Which of the following are the most likely explanations for how these angles of repose are maintained?

- a The sand is highly angular, very cohesive and has a high coefficient of friction because it is unusually rough to enable the sand particles to form such steep structures.
- b Moisture, e.g., water, is added to the dry sand daily to enable the structures to be maintained, just like the shapes of sandcastles can be maintained by adhesive forces generated by capillarity between grains.
- c The dry sand contains some cementitious binding agent such as cement, gypsum mortar or lime mortar cunningly disguised in the white sand.
- d The custodians of such dry sand gardens have secrets handed down to them by their predecessors over the years which science has yet to establish.

Answer

All of the above, apart from (d). A variety of special rakes are used by custodians of dry sand gardens and is it evident that water is used in the process of designing and maintaining these gardens - see www.japanesegardening.org/reference/sand/ and http://www.japanesegardening.org/site/pushing-the-line-a-theoretical-approach-to-raking-a-karesansui-garden/, so that areas which require raking need only be raked on a weekly basis. Therefore, the term 'dry sand garden' is somewhat of a misnomer. Where there are steep sides to dry sand garden features such as at Ginkaku-ji and at the sand mounds at Honen-in, it would make sense not to have to alter the underlying shape of the features, so that the underlying shape is a rigid assembly of sand and cement, i.e., concrete, on the surface of which there is a relatively thin dusting of moistened sand raked into shape. Custodians of these dry sand gardens clearly have expertise in maintaining them which can take time to learn, but which is explicable in terms of the science behind the art.

Soil in an embankment is made progressively more wet by unusually heavy rain. What measures should be in place during the construction of the embankment and its subsequent maintenance to reduce the risk of possible failure of this embankment?

Answer

This is a question which is of direct relevance to the construction of the levees which broke when Hurricane Katrina struck New Orleans in August 2005. In constructing and maintaining a soil embankment a number of engineering problems require attention: the foundations, the nature of the material used to construct the embankment and the degree of compaction, the need for adequate drainage, the monitoring of erosion and the monitoring of pore water pressure. A relevant research paper on the levees which failed in New Orleans is G. L. Sills, N. D. Vroman, R. E. Wahl and N. T. Schwanz, Overview of New Orleans levee failures: lessons learned and their impact on national levee design and assessment, *J. Geotech. Geoenv. Engng.* **134**, 556-565 (2008).

This page titled 33.11: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

34: Indexing Electron Diffraction Patterns

Learning Objectives

On completion of this TLP you should:

- Understand why the spots on an electron diffraction pattern appear where they do.
- Know how to index a diffraction pattern from a sample with a known lattice.

Before you start

It is strongly recommended that you read through the TLPs on Diffraction and Imaging and X-ray Diffraction before reading this TLP.

34.1: Introduction
34.2: Mathematics
34.3: Laue Zones
34.4: Kikuchi Lines
34.5: Using Polycrystalline Materials in the TEM
34.6: Convergent Beam Electron Diffraction (CBED)
34.7: Using Other Methods in Conjunction with Electron Diffraction
34.8: Summary
34.9: Questions

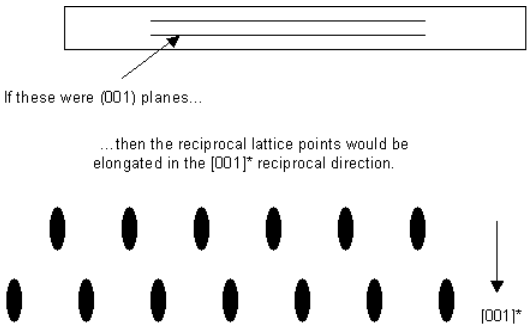
This page titled 34: Indexing Electron Diffraction Patterns is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

34.1: Introduction

Electrons can act as waves as well as particles; this is a consequence of quantum mechanics. A series of electrons hitting an object is exactly equivalent to a beam of electron waves hitting the object and it produces a diffraction pattern in the same way as a beam of X-rays does.

The two important differences between electron and X-ray diffraction are that (1) *electrons have a much smaller wavelength than X-rays*, and (2) *the sample is very thin in the direction of the electron beam* (of the order of 100 nm or less) - it has to be thin so that enough electrons can get through to form a diffraction pattern without being absorbed. These factors conspire to have a fortunate effect on the Ewald sphere construction (see The Ewald sphere in the Reciprocal Space TLP) and diffraction pattern:

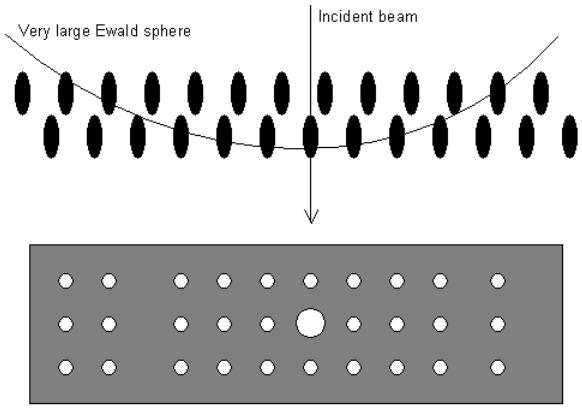
1. The thin sample makes the reciprocal lattice points longer in the reciprocal direction corresponding to the real-space dimension in which the sample is thin:



   It should be noted that there is not necessarily always a particular plane oriented like this. However, it is usual for identification of crystalline phases in a sample to orient the sample so that the electron beam is parallel to a low index lattice direction, as this makes the electron diffraction pattern easier to interpret.

2. The small electron wavelength makes the radius of the Ewald sphere very large (recall its radius is $1/\lambda$). The small electron wavelength also makes the diffraction angles θ small (1-2°); this can be seen by substituting a wavelength of 2.51 x 10^{-12} m into the Bragg equation (see Bragg's law in the X-ray Diffraction TLP).

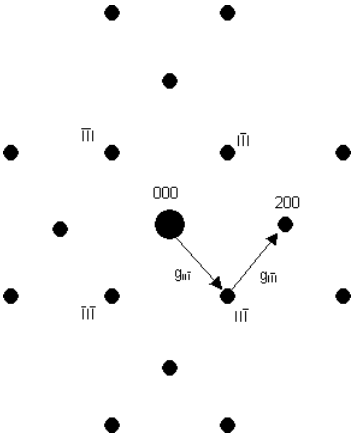
These make the Ewald sphere diagram look like this so that whole layers of the reciprocal lattice end up projected onto the film or screen:



Note that the large (strong) spot in the middle is the straight-through beam (the beam which has passed through the sample without diffracting). This always has the index 000.

Caution 1: systematic (kinetic) absences appear in electron diffraction patterns just as in X-ray diffraction patterns, for the same reason: the various features of the lattice or motif diffract electrons in the same direction but the phase factors from the various features cancel, leaving an absence.

Caution 2: sometimes where there should be a systematic absence, the spot appears to be still there. This is because of the strong interaction between electrons and atoms: there is a small but significant probability that an electron will be diffracted twice, from two planes one after another - i.e. in two different reciprocal lattice directions one after another. These two directions can add up so that the twice-diffracted electron may arrive at a position in reciprocal space where there is a systematic absence. As an example, the diagram below is a schematic of the [011] electron diffraction pattern of silicon: the 200 type reflections are systematically absent. The intensity at the 200 reflections is caused by double diffraction (arising from the addition of the two reciprocal lattice vectors shown). Thus, in words, intensity can occur in the 200 reflection from, firstly, diffraction from the 111 planes, followed by, secondly, diffraction by the 111 planes as the electron wave passes through the specimen.



This page titled 34.1: Introduction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

34.2: Mathematics

Relation 1

The distance, r_{hkl}, on the pattern between the spot *hkl* and the spot 000 is related to the interplanar spacing between the *hkl* planes of atoms, d_{hkl}, by the following equation: (Derivation)

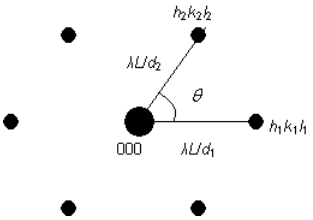
$$r_{hkl} = \frac{\lambda L}{d_{hkl}} \tag{34.2.1}$$

where L is the distance between the sample and the film/screen.

We can therefore say that *the diffraction pattern is a projection of the reciprocal lattice with projection factor* λL, because reciprocal lattice vectors have length $1/d_{hkl}$.

Relation 2

Since the diffraction pattern is a projection of the reciprocal lattice, *the angle between the lines joining spots* $h_1k_1l_1$ *and* $h_2k_2l_2$ *to spot 000 is the same as the angle between the reciprocal lattice vectors* $[h_1k_1l_1]$* *and* $[h_2k_2l_2]$*. This is also equal to the angle between the $(h_1k_1l_1)$ and $(h_2k_2l_2)$ planes, or equivalently the angle between the normals to the $(h_1k_1l_1)$ and $(h_2k_2l_2)$ planes. This angle is θ in the diagram below.



Using these two relations between the diffraction pattern and the reciprocal lattice, we are now able to index the electron diffraction pattern from a specimen of a known crystal structure.

The two pages linked to here refer only to indexing the central region of the diffraction pattern - the rest will be dealt with later.

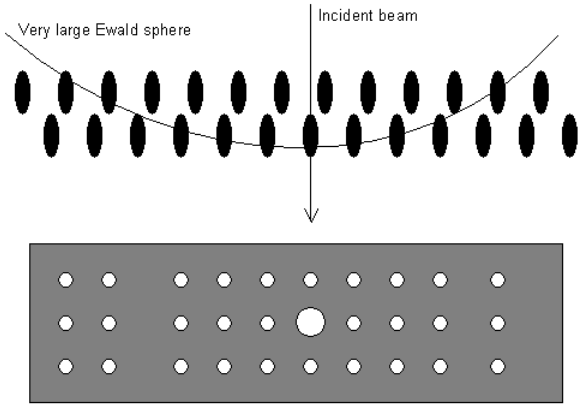
Indexing with the orientation of the electron beam *known*

Indexing with the orientation of the electron beam *unknown*

This page titled 34.2: Mathematics is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

34.3: Laue Zones

So far we have been looking at the central region of the diffraction pattern. This is only a part of the total diffraction pattern. If we look again at the Ewald sphere construction, we have:



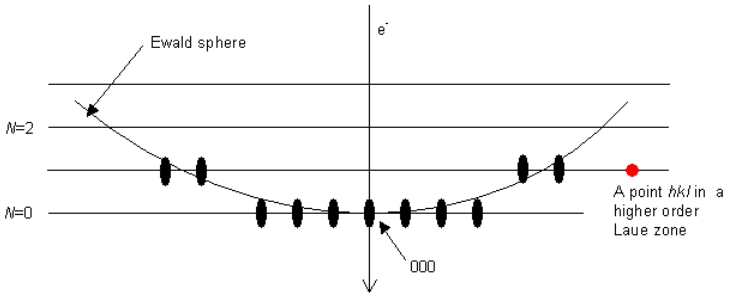
We have been indexing the portion in the middle with the 000 spot in it. However, there are also areas of diffraction spots at the edges of the film, caused by the Ewald sphere intersecting points in an adjacent parallel plane containing reciprocal lattice points. (If the film was small or the camera length large it is possible that it did not catch these spots at the side, so that we sometimes only have the middle part.)

These outlying parts of the diffraction pattern are called Higher Order Laue Zones (HOLZs). Each of the HOLZs can be described by an equation of the general form

$$hu + kv + lw = N \tag{34.3.1}$$

where:

- N is always an integer, and is called the *order* of the Laue zone.
- [*uvw*] is the direction of the incident electron beam.
- *hkl* are the co-ordinates of an allowed reflection in the *N*th order Laue zone.



The middle part of the diffraction pattern, with 000 in it, is the zero order Laue zone (ZOLZ), because it comes from the plane for which $N = 0$: an allowed reflection *hkl* in the ZOLZ is joined to the origin 000 by a reciprocal lattice vector that lies in the ZOLZ. For the ZOLZ the electron beam [uvw] and the allowed reflection *hkl* satisfy the Weiss zone law $hu + kv + lw = 0$. The next layer up has a value $N = 1$, then $N = 2$, and so on, as shown.

From the geometry of the way in which the Ewald sphere intersects the HOLZs, the radius of the Nth HOLZ ring, R_n, in reciprocal space, is given to a very good approximation by the formula

$$R_n = \sqrt{\left(\frac{2N}{\lambda |uvw|}\right)} \tag{34.3.2}$$

assuming that the wavelength of the electrons is much less than the modulus |*uvw*| of the direction [*uvw*] in the crystal parallel to the electron beam direction.

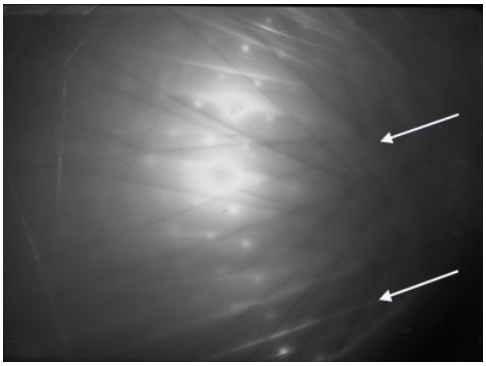
Thus, HOLZs are seen more easily at lower voltages (e.g. 100 kV rather than 300 kV) and when the electron beam is parallel to a relatively high index direction in a crystal.

It is possible to index the reflections in the HOLZs on a diffraction pattern. Examples of such indexing are given in the book *Transmission Electron Microscopy of Materials* by D B Williams and C B Carter.

This page titled 34.3: Laue Zones is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

34.4: Kikuchi Lines

Kikuchi lines often appear on electron diffraction patterns:



An example of a "two-beam" electron diffraction pattern with a number of Kikuchi lines. A pair of Kikuchi lines is arrowed.
[The term "two-beam" denotes the fact that the straight-through beam, 000, and one diffraction spot are both diffracting very strongly. The intensity of all spots in this electron diffraction pattern are significantly weaker by comparison with these two beams.]
(Click on image to view larger version)

We will not learn to index the Kikuchi lines in this TLP. Instead, we will explain their origin and behaviour with the help of the following animation.

Kikuchi lines are interesting because of what they do when the crystal is moved in the beam. Diffraction spots fade or become brighter when the crystal is rotated or tilted, but stay in the same places; the Kikuchi lines move across the screen.

The difference in behaviour can be explained by the position of the effective source of the electrons that are Bragg-scattered to produce the two phenomena. The diffraction spots are produced directly from the electron beam, which either hits or misses the Bragg angle for each plane; so the spot is either present or absent depending on the orientation of the crystal. The source of the electrons that are Bragg-scattered to give Kikuchi lines is the set of inelastic scattering sites within the crystal. When the crystal is tilted the effective source of these inelastically scattered electrons is moved, but there are always still some electrons hitting a plane at the Bragg angle - they merely emerge at an angle different to the one that they did before the crystal was tilted.

This page titled 34.4: Kikuchi Lines is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

34.5: Using Polycrystalline Materials in the TEM

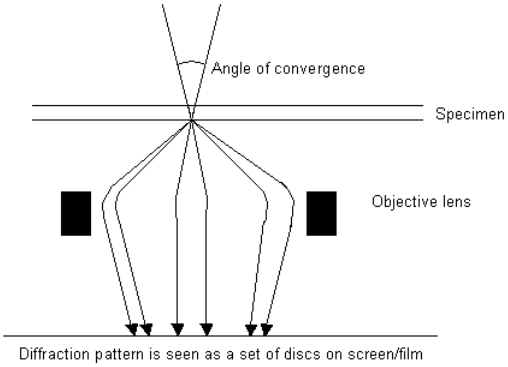
Just as with X-rays, a completely isotropic fine-grained polycrystalline sample will give a diffraction pattern of concentric rings in the zero order Laue zone (ZOLZ), as the many small crystals at random orientations produce a continuous angular distribution of hkl spots at distance $1/d_{hkl}$ from the 000 spot - a ring of radius $1/d_{hkl}$ around the 000 spot for each allowed reflection. The rings are then indexed according to the order of allowed reflections within the ZOLZ.

As the grain size increases, the rings within the diffraction pattern break up into discontinuous rings containing discrete reflections. If there is any texture (preferred orientation) within the specimen, arcs may be seen instead of complete rings.

This page titled 34.5: Using Polycrystalline Materials in the TEM is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

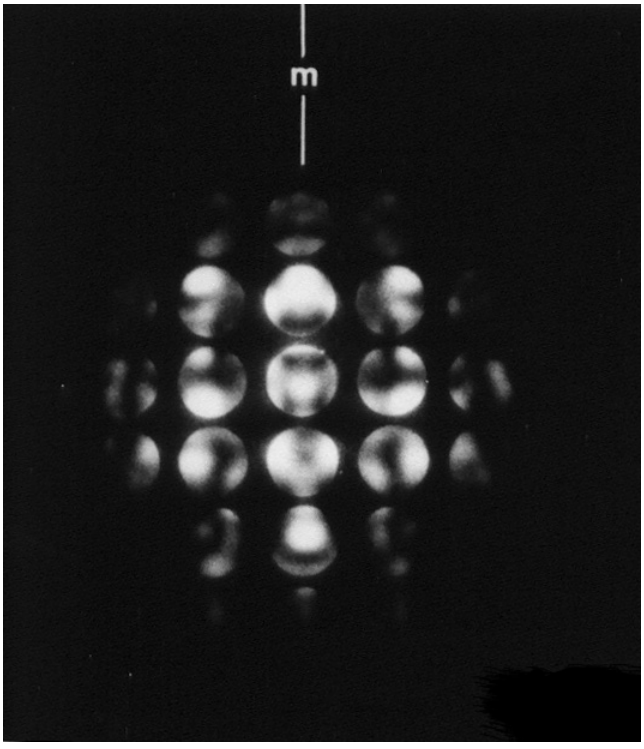
34.6: Convergent Beam Electron Diffraction (CBED)

When a convergent beam is used instead of a parallel beam of electrons, the rays converge to a point within the specimen and come out the other side inverted like a camera. However, we do not look at the inverted image; we look at the diffraction pattern, with the spots magnified:

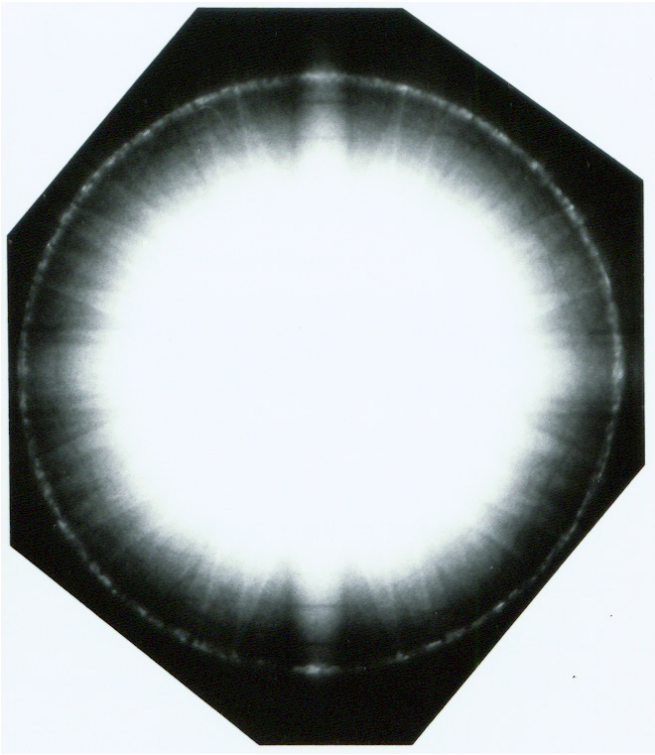


Depending on the camera length chosen, either the zero order Laue zone can be examined or the zero order Laue zone and higher order Laue zones. Two examples of CBED images are shown below. The symmetry seen is such patterns can be related to the space group symmetry of the specimen.

Examples of CBED images:



Diffraction pattern showing a zero order Laue zone
with a mirror in the pattern as shown
(from H.H. Hng, Ph.D. thesis, University of Cambridge, 1999.)



Diffraction pattern showing a first order Laue zone
from H.H. Hng, Ph.D. thesis, University of Cambridge, 1999.)

This page titled 34.6: Convergent Beam Electron Diffraction (CBED) is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and

standards of the LibreTexts platform.

https://eng.libretexts.org/@go/page/34941

34.7: Using Other Methods in Conjunction with Electron Diffraction

Electron diffraction is a powerful technique - but other techniques must be used with it to put the results in context. This is a brief synopsis of how other methods can be used to help.

X-ray Diffraction

The majority of novel / unknown crystalline materials are indexed by single crystal X-ray diffraction methods. Originally, photographs were taken along crystallographic axes and indexing performed manually. Nowadays, data are collected with a single point or two-dimensional electronic detector and the data are indexed with automatic computer programs.

When single crystals cannot be grown, data from a polycrystalline sample may be used to determine unit cells of crystal structures. Simple structures, such as cubic crystal structures, can be indexed manually by looking for integer relationships between interplanar spacings. For more complex structures, such as orthorhombic, monoclinic and triclinic, there are several different types of computer program. However, as there is a substantial loss of information in going from single crystal diffraction data to powder diffraction data, indexing powder diffraction data is demanding. Single crystal X-ray diffraction data are the first choice.

Information from both electron and X-ray diffraction are sometimes combined to tackle difficult crystal structures. Hints from other methods maybe also be useful. Indexing X-ray reflections from both single crystals and powders also gives information about the space group of the material under investigation from considerations of symmetry and systematic absences. Measurement of accurate unit cell lattice parameters can also be undertaken - this requires high quality, high angle X-ray diffraction data.

References

[1] B.D. Cullity and S.R. Stock, 'Elements of X-ray diffraction', 3rd editon, Prentice Hall (2001)
[2] L.S. Dent Glasser, 'Crystallography and its applications', Van Nostrand Reinhold (1977).
[3] International Union of Crystallography website, www.iucr.org
[4] CCP 14: Available Software for Powder Diffraction Indexing including a Literature Search List, www.ccp14.ac.uk/solution/indexing

Optical imaging

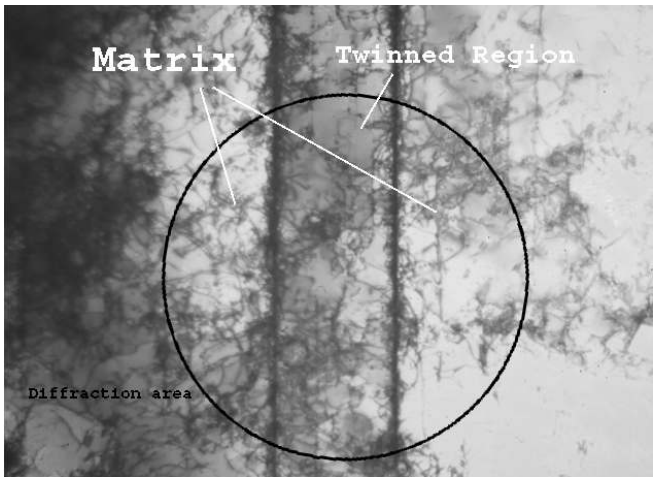
This is a very important way of analysing a specimen. Using the naked eye and optical microscopes we can determine down to a point-to-point resolution limited by the wavelength of light how many phases there are and how they relate to one another. We can also infer what type of material they are likely to be and how they may have been processed.

Chemical analysis

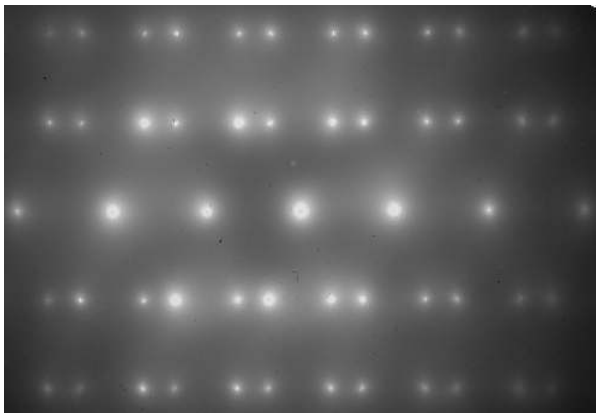
A wide range of chemical techniques can be used to find out what components are present in the different phases and in what proportions. This will narrow the field of possible elements that we need to consider when analysing our diffraction results. These techniques range from simple chemical tests, through infrared spectroscopy of organic samples, to a wide variety of chemical characterisation techniques that can often be performed within the transmission electron microscope.

TEM imaging

Using the TEM to image the same area of sample that is being used to produce the diffraction pattern is an invaluable technique:



Nitrided surface layer of austenitic stainless steel



Diffraction pattern from nitrided surface layer of austenitic stainless steel

Using the image to verify that the double dots in the diffraction pattern are being caused by the two crystal structures either side of the twin boundary, we can index the pattern and determine the twin plane and the crystal structures either side of it.

This page titled 34.7: Using Other Methods in Conjunction with Electron Diffraction is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

34.8: Summary

In this teaching and learning package we have considered how electron diffraction patterns are formed in the transmission electron microscope. The principles of how to index spot electron diffraction patterns have been discussed in some detail. Although we have considered how to index electron diffraction patterns from relatively simple crystal structures to illustrate the basic principles, these principles are generic and can therefore be applied to any crystal structure. We have also considered other features of electron diffraction patterns such as the formation of Kikuchi lines, the formation of convergent beam electron diffraction patterns and the formation of higher index Laue zones.

Going further

Books

- *Electron Microscopy of Thin Crystals* by P B Hirsch, A Howie, R B Nicholson, D W Pashley & M J Whelan, published by Krieger
  Now out of print but explains everything very clearly.
- *Transmission electron microscopy of materials* by D B Williams and C B Carter
  This is a series of four books that contains very detailed descriptions of what happens and how to operate the microscope. Explanations are often in quantum mechanical terms and can be hard going if you want a quick reminder of how something works.
- *Electron microscopy and analysis* by P J Goodhew, J Humphreys and R Beanland
  A clear guide to the principles and phenomena involved in electron microscopy.

This page titled 34.8: Summary is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

34.9: Questions

Quick questions

You should be able to answer these questions without too much difficulty after studying this TLP. If not, then you should go through it again!

Which of the following is *not* an effect of the small wavelength of the electron?

a	The angle of diffraction is very small
b	The reciprocal lattice points are elongated
c	The Ewald sphere is very large
d	The diffraction pattern is produced from a plane of reciprocal lattice points

Answer

B.

Why might you get a diffraction spot where you thought there would be an absence?

a	The atoms moved across a little
b	Inelastic scattering
c	Some electrons were diffracted two or more times
d	There is a fault in the lens

Answer

C.

Where do higher order Laue zones come from?

a	The Ewald sphere cutting parts of reciprocal lattice planes other than the one tangent to the sphere
b	A different scattering mechanism
c	Internal reflection in the microscope
d	Small crystals on the edge of the "single" crystal

Answer

A.

Deeper questions

The following questions require some thought and reaching the answer may require you to think beyond the contents of this TLP.

4. If you increase the camera length, what happens to the diffraction pattern? What about the Kikuchi lines?

Answer

The diffraction spots move further apart, because the spot spacing is related to the camera length by

$$r = \frac{\lambda L}{d} \tag{34.9.1}$$

The Kikuchi lines also move further apart, and stay in the same places with respect to the diffraction spots, as they are a result of diffraction from the same crystal.

Why would it not be possible to index a zero order Laue zone in a diffraction pattern from a cubic crystal, knowing the two points 111 and 111 ?

Answer

These two points are collinear and cannot be added to produce any points not in the line that passes through both of them - so we could not index any of the pattern not on that line!

Open-ended questions

The following questions are not provided with answers, but intended to provide food for thought and points for further discussion with other students and teachers.

We have seen how electron diffraction relates to X-ray diffraction; how do you think neutron diffraction compares to electron diffraction? (Assume that we have a sufficient vacuum to enable the neutrons to reach the sample and be diffracted, and that we have some means of detecting the diffracted neutrons.)

This page titled 34.9: Questions is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

CHAPTER OVERVIEW

35: The Jominy End Quench Test

Learning Objectives

On completion of this TLP you should:

- be able to describe how the Jominy test is conducted and how the information that it provides is obtained and presented
- be able to describe the general effects of alloying and prior heat treatment on the jominy test results

35.1: Introduction
35.2: Hardenability
35.3: Jominy Test
35.4: Uses of Jominy Data
35.5: Example Test Data
35.6: Simulations
35.7: Summary
35.8: Questions

This page titled 35: The Jominy End Quench Test is shared under a CC BY-NC-SA 2.0 license and was authored, remixed, and/or curated by Dissemination of IT for the Promotion of Materials Science (DoITPoMS) via source content that was edited to the style and standards of the LibreTexts platform.

35.1: Introduction

The Jominy end quench test is used to measure the hardenability of a steel, which is a measure of the capacity of the steel to harden in depth under a given set of conditions. This TLP considers the basic concepts of hardenability and the Jominy test.

Knowledge about the hardenability of steels is necessary to be able to select the appropriate combination of alloy steel and heat treatment to manufacture components of different size to minimize thermal stresses and distortion. The Jominy end quench test is the standard method for measuring the hardenability of steels. This describes the ability of the steel to be hardened in depth by quenching.

Hardenability depends on the chemical composition of the steel and also be can affected by prior processing conditions, such as the austenitizing temperature. It is not only necessary to understand the basic information provided from the Jominy test, but also to appreciate how the information obtained can be used to understand the effects of alloying in steels and the steel microstructure.

35.1: Introduction is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

35.2: Hardenability

Hardenability is the ability of a steel to partially or completely transform from austenite to some fraction of martensite at a given depth below the surface, when cooled under a given condition. For example, a steel of a high hardenability can transform to a high fraction of martensite to depths of several millimetres under relatively slow cooling, such as an oil quench, whereas a steel of low hardenability may only form a high fraction of martensite to a depth of less than a millimetre, even under rapid cooling such as a water quench. Hardenability therefore describes the capacity of the steel to harden *in depth* under a given set of conditions.

Steels with high hardenability are needed for large high strength components, such as large extruder screws for injection moulding of polymers, pistons for rock breakers, mine shaft supports, aircraft undercarriages, and also for small high precision components such as die-casting moulds, drills and presses for stamping coins. High hardenability allows slower quenches to be used (e.g. oil quench), which reduces the distortion and residual stress from thermal gradients.

Steels with low hardenability may be used for smaller components, such as chisels and shears, or for surface hardened components such as gears.

Hardenability can be measured using the Jominy end quench test.

35.2: Hardenability is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

35.3: Jominy Test

The test sample is a cylinder with a length of 102 mm (4 inches) and a diameter of 25.4 mm (1 inch).



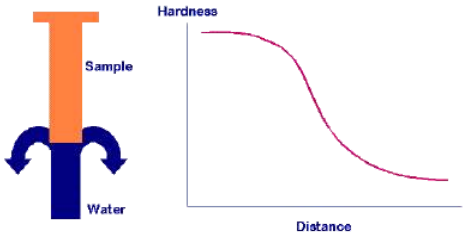
Jominy test specimen

The steel sample is *normalised* to eliminate differences in microstructure due to previous forging, and then *austenitised*. This is usually at a temperature of 800 to 900°C. The test sample is quickly transferred to the test machine, where it is held vertically and sprayed with a controlled flow of water onto one end of the sample. This cools the specimen from one end, simulating the effect of quenching a larger steel component in water.



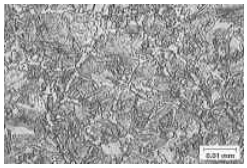
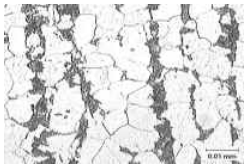
Jominy test machine

The cooling rate varies along the length of the sample from very rapid at the quenched end, to rates equivalent to air cooling at the other end.



The round specimen is then ground flat along its length to a depth of 0.38 mm (15 thousandths of an inch) to remove decarburised material. The hardness is measured at intervals from the quenched end. The interval is typically 1.5 mm for alloy steels and 0.75 mm for carbon steels.

High hardness occurs where high volume fractions of martensite develop. Lower hardness indicates transformation to bainite or ferrite/pearlite microstructures.

	
Martensite	Ferrite/pearlite

Jominy end quench hardness data for two steels of different hardenability can be seen in a later section of this TLP, with images of the microstructure variation along the length of the sample.

Similar tests have been developed in other countries, such as the SAC test, which uses a sample quenched from all sides by immersion in water. This is commonly used in the USA.

Video clips of the Jominy test procedure

Help on viewing video clips

Video clip 1: Transferring the sample from furnace to quenching machine



The specimen is suspended from a wire and held in a furnace to austenitise the microstructure at around 900°C. It is then carefully and quickly moved to the quenching machine and positioned above a water jet.The water jet is started and sprayed onto the bottom of the specimen until the specimen is cool.

https://www.doitpoms.ac.uk/tlplib/jo...os/jominy1.mp4

Transferring the sample from furnace to quenching machine

Video clip 2: Quenching the sample



As the water jet sprays onto the end of the hot, glowing specimen, a cold dark region spreads up the specimen. The cold region has transformed from austenite to a mixture of martensite, ferrite and pearlite. The proportions of the phases at any position depends on the cooling rate, with more martensite formed where the cooling rate is fastest. Ferrite and pearlite are formed where the cooling rate is slower.

https://www.doitpoms.ac.uk/tlplib/jo...os/jominy2.mp4

Quenching the sample

Video clip 3: Jominy end quench test



This alternative longer video clip (contributed by Oxford Brookes University) shows both the transfer of the sample from furnace to Jominy machine, and the jet spraying one end of the sample.

https://www.doitpoms.ac.uk/tlplib/jo...videos/jet.mp4

Jominy end quench test

35.3: Jominy Test is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

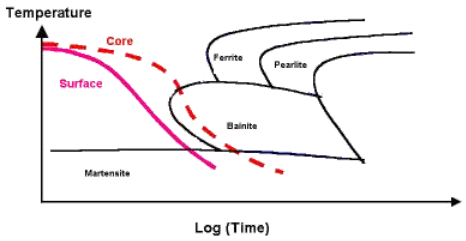
35.4: Uses of Jominy Data

Measurement of Hardenability

Data from the Jominy end quench test can be used to determine whether a particular steel can be sufficiently hardened in different quenching media, for different section diameters.

For example, the cooling rate at a distance of 9.8 mm from the quenched end is equivalent to the cooling rate at the centre of an oil-quenched bar with a diameter of 28 mm. Full transformation to martensite in the Jominy specimen at this position indicates that a 28 mm diameter bar can be *through hardened*, i.e. hardened through its full thickness.

A high hardenability is required for through hardening of large components. This data can be presented using CCT (**C**ontinuous **C**ooling **T**ransformation) diagrams which are used to select steels to suit the component size and quenching media. Slow quenching speeds are often chosen to reduce distortion and residual stress in components.



CCT diagram for an alloy steel

Slower cooling rates occur at the core of larger components, compared to the faster cooling rate at the surface. In the example here, the surface will be transformed to martensite, but the core will have a bainitic structure with some martensite.

Effects of Alloying and Microstructure

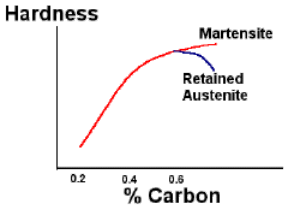
Jominy end quench test can also be used to demonstrate the effects of microstructure and alloying variables on the hardenability of steels. These include alloying elements and grain size.

Alloying elements.

The main alloying elements which affect hardenability are carbon, boron and a group of elements including Cr, Mn, Mo, Si and Ni.

Carbon

Carbon controls the hardness of the martensite. Increasing the carbon content increases the hardness of steels up to about 0.6wt%. At higher carbon levels, the formation of martensite is depressed to lower temperatures and the transformation from austenite to martensite may be incomplete, leading to retained austenite. This composite microstructure of martensite and austenite gives a lower hardness to the steel, although the microhardness of the martensite phase itself is still high.



Effect of carbon content (wt%) on hardness

Carbon also increases the hardenability of steels by retarding the formation of pearlite and ferrite. However, the effect is too small be be commonly used for control of hardenability. High carbon steels are prone to distortion and cracking during heat treatment, and can be difficult to machine in the annealed condition before heat treatment. It is more common to control hardenability with other elements, and to use carbon levels of less than 0.4wt%.

Boron

Boron is a very potent alloying element, typically requiring 0.002 to 0.003wt% to have an equivalent effect as 0.5wt% Mo. The effect of boron is also independent of the amount of boron, provided sufficient is added, and the effect of boron is greatest at lower carbon contents. It is typically used with lower carbon steels.

Boron has a very strong affinity for oxygen and nitrogen, with which it forms compounds. Boron can therefore only affect the hardenability of steels if it is in solution. This requires the addition of "gettering" elements such as aluminium and titanium to react preferentially with the oxygen and nitrogen in the steel.

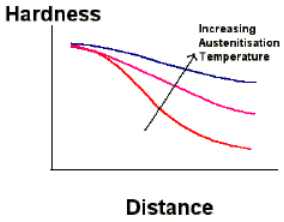
Chromium, molybdenum, manganese, silicon, nickel, vanadium

The elements Cr, Mo, Mn, Si, Ni and V all retard the phase transformation from austenite to ferrite and pearlite. The most commonly used elements are Cr, Mo and Mn.

The retardation is due to the need for redistribution of the alloying elements during the diffusional phase transformation from austenite to ferrite and pearlite. The solubility of the elements varies between the different phases, and the interface between the growing phase cannot move without diffusion of the slowly moving elements. There are quite complex interactions between the different elements, which also affect the temperatures of the phase transformation and the resultant microstructure. Steel compositions are sometimes described in terms of a *carbon equivalent* which describes the magnitude of the effect of all of the elements on hardenability.

Grain size

Increasing the austenite grain size increases the hardenability of steels. The nucleation of ferrite and pearlite occurs at heterogeneous nucleation sites such as the austenite grain boundaries. Increasing the austenite grain size therefore decreases the available nucleation sites, which retards the rate of the phase transformation. This method of increasing the hardenability is rarely used since substantial increases in hardenability require large austenite grain size, obtained through high austenitisation temperatures. The resultant microstructure is quite coarse, with reduced toughness and ductility.



Effect of austenite grain size on hardenability

The austenite grain size can be affected by other stages in the processing of steel, and therefore the hardenability of a steel also depends on the previous stages employed in its production.

35.4: Uses of Jominy Data is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

35.5: Example Test Data

A plain carbon steel and an alloy steel were assessed using the Jominy end quench test.

The hardness of the samples was measured as a function of the distance from the quenched end to demonstrate the different hardenability of the two steels. The data is shown as Vickers and Rockwell hardness.

The alloy compositions are given in the table below.

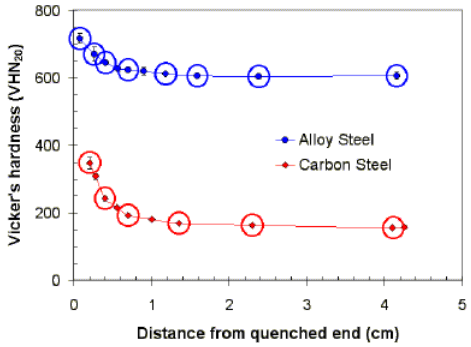
(wt%)	C	Mn	Cr	Ni	Si	Mo	P	S
Plain carbon Steel	0.3	0.7	0.1	0.14	0.26	0.03	0.003	0.02
Alloy steel	0.3	0.6	0.7	3.5	0.26	0.35	0.01	

Vickers Hardness

The *Vickers* hardness test uses a square pyramidal diamond indentor. The recorded hardness depends on the indentation load and the width of the square indentation made by the diamond. The indentation load is typically between 10 and 30 kg. The hardness number is usually denoted by HV20 for **H**ardness **V**ickers **20** kg, for example. The Vickers test is most commonly used in the UK. The Rockwell hardness of a metal can also be determined using a similar technique.

The variation of hardness was measured with distance from the quenched end. The results are plotted in the graph below. *Click on the circled data points to see how the microstructure varies with distance from the quenched end.*

The alloy steel clearly has the highest hardenability, forming martensite to a greater depth than the plain carbon steel. Look at both the microstructures at high magnification, and try to observe the relationship between the volume fraction of martensite and the hardness of the steel.



Rockwell Hardness

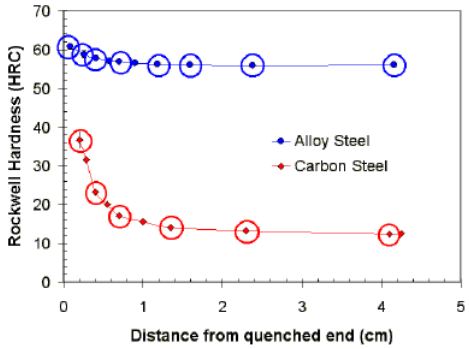
The Vickers hardness scale is not the only scale used to measure hardness in metals.

The *Rockwell* hardness test measures a number which depends on the difference in the depth of an indentation made by two loads, a minor load followed by a major load. There are different scales for the Rockwell hardness test. For example, the commonly used Rockwell C test uses a minor load of 10 kg, followed by a major load of 150 kg. The number is denoted as HRC for **H**ardness **R**ockwell **C** scale. The indentor is either a conical diamond pyramid, or a hardened steel ball. The Rockwell test is commonly used in the USA.

Other tests include the Brinell and Knoop hardness tests.

There are conversion charts between the hardness scales. These can be found in standards, such as the British Standards, and reference works such as the ASM Metals handbook. It's important to use the correct conversion chart for different materials, since the hardness test causes plastic strain, and therefore varies with the strain hardening properties of the material.

The graph below gives the Jominy end quench data in terms of the Rockwell hardness number. Clicking on the circled data points will take you to images of the microstructure at that location in the sample.



35.5: Example Test Data is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

35.6: Simulations

Heat flow simulation

In this heat flow simulation you can adjust various parameters and observe the effect on the heat flow and cooling of the specimen. The simulation ignores the effect of heat loss from the sides of the specimen, i.e. it employs a one-dimensional model of heat flow through the specimen.

The bar is divided into 25 equal length elements, and, at each time step of the simulation, for each element, a new temperature, resulting from heat transfer at either end, is calculated. The size of the time step is set to the maximum allowed while ensuring numerical stability of the simulation.

Quench test simulation

In this simulation (courtesy of SteelMATTER, where it is also available), you can follow through the stages in carrying out a Jominy end quench test.

The Jominy Test involves heating a test piece from the steel (25mm diameter and 100mm long) to an austenitising temperature and quenching from one end with a controlled and standardised jet of water. Take a sample from the furnace and place it on the Jominy test fixtures and observe the cooling pattern.

After quenching the hardness profile is measured at intervals from the quenched end after the surface has been ground back to remove any effects of decarburisation (0.38mm is removed from the surface).

The hardness variation along the test surface is a result of microstructural variation which arises since the cooling rate decreases with distance from the quenched end. The cooling rate along the Jominy test specimen varies from about 225 °C s^{-1} to 2 °C s^{-1}.

35.6: Simulations is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

35.7: Summary

The Jominy end quench test is the standard method for measuring the hardenability of steels. This describes the ability of the steel to be hardened in depth by quenching. The hardenability depends on the alloy composition of the steel, and can also be affected by prior processing, such as the austenitisation temperature.

Knowledge of the hardenability of steels is necessary in order to select the appropriate combination of alloy and heat treatment for components of different size, to minimise thermal stresses and distortion.

Going further

Reading

- Honeycombe R W K and Bhadeshia H K D H, *Steels: Microstructure and Properties*, Edward Arnold, 1995.
- Thelning, Karl-Erik, *Steel and its Heat Treatment*, Butterworths, 1975.
- Llewellyn D T and Hudd R C, *Steels: Metallurgy and Applications*, 3rd Edition, Reed Educational and Professional Publishing, 1998.
- ASM Handbook, Volume 4: Heat Treating, ASM International, 1991.

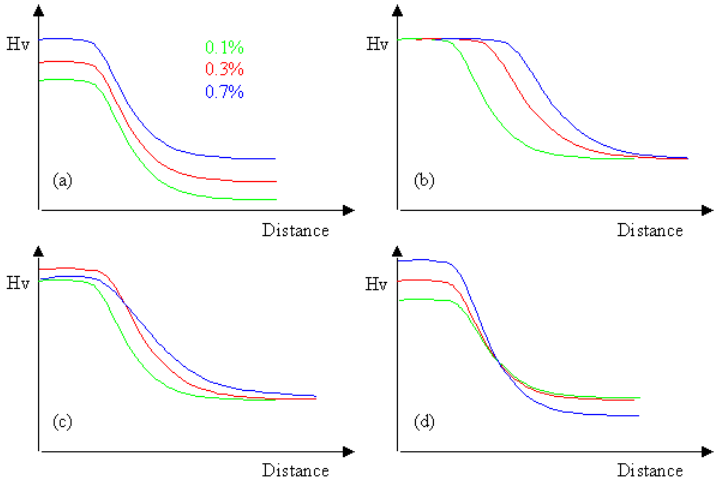
Websites

- Understanding The Jominy End Quench Test
  An article by James Marrow published on the Industrial Heating website, 9/7/2001. Much of the material is in common with the above website and this TLP.
- Jominy hardenability (quench) test
  Part of the SteelMATTER website, this page has a Java applet simulation which allows you to perform a series of Jominy end quench tests on different grades of steel to see how composition affects hardenability.
- Jominy Testing: The Practical Side
  An article by Daniel Herring also published on the Industrial Heating website, 10/10/2001.

35.7: Summary is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

35.8: Questions

Three low alloy steels, which differ only in their carbon content (0.1, 0.3 and 0.7 wt% carbon) are characterised using the Jominy end quench test. Select the plot of hardness variation along the test specimen that best describes their behaviour.

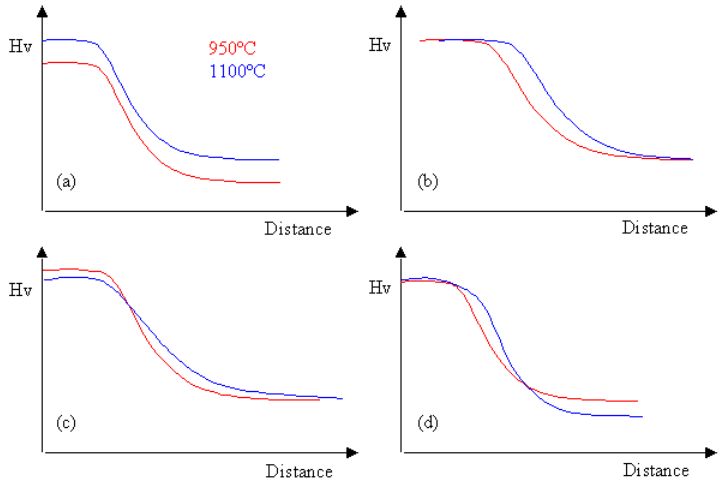


a	Plot (a)
b	Plot (b)
c	Plot (c)
d	Plot (d)

Answer

C. Increasing the carbon content increases the hardenability (depth of hardening) and also the hardness of the untempered martensite. However, at high carbon levels (0.7%), the martensite can auto-temper, resulting in a softer microstructure.

Two specimens of a low alloy steel with 0.3wt% carbon are characterised using the Jominy end quench test. One was austenitised at 950°C, and the other was austenitised at 1100°C. Select the plot of hardness variation along the test specimen that best describes their behaviour.

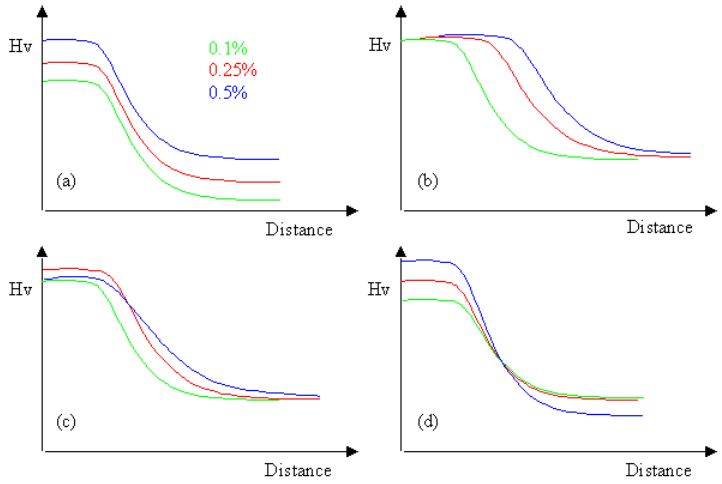


a	Plot (a)
b	Plot (b)
c	Plot (c)
d	Plot (d)

Answer

D. Increasing the grain size with the high temperature anneal has two effects. It increases the hardenability (hardness in depth) by reducing the nucleation rate of ferrite at grain boundaries, which encourages martensite formation, but decreases the hardness of the non-martensitic microstructure (ferrite and pearlite) due the Hall-Petch effect.

Three medium carbon steels (0.3wt%) that differ only in their Chromium content (0.25, 0.5 and 1 wt%) are characterised using the Jominy end quench test. Select the plot of hardness variation along the test specimen that best describes their behaviour.



a	Plot (a)
b	Plot (b)
c	Plot (c)
d	Plot (d)

Answer

B. Increasing the Cr content increases hardenability (hardness in depth) by reducing the diffusion controlled transformation rate of ferrite and pearlite and thus encouraging martensite formation. There is also a small increase in strength due to solid solution strengthening by the Cr.

You have three steels. Select the most appropriate steel to achieve the necessary levels of mechanical properties, residual stress and distortion in a 1mm diameter wood-working drill.

a	1% C, 0.4% Si, 1% Mn, 5% Cr, 1% Mo
b	0.4% C, 0.4% Mn, 0.3% Si
c	0.5% C, 4% Cr, 6% Mo

Answer

B. A low hardenability steel (cheap) suitable for water quenching small component with transformation to martensite across the full section.

Again, you have three steels. Select the most appropriate steel to achieve the necessary levels of mechanical properties, residual stress and distortion in an injection moulding die for a mobile phone plastic case.

a	1% C, 0.4% Si, 1% Mn, 5% Cr, 1% Mo
b	0.4% C, 0.4% Mn, 0.3% Si
c	0.5% C, 4% Cr, 6% Mo

Answer

A. An oil-hardening steel (moderately high hardenability), suitable for transforming to martensite through thickness with a slow quench, which reduces the level of distortion in the mould. This is essential for high precision injection mouldings.

Again, you have three steels. Select the most appropriate steel to achieve the necessary levels of mechanical properties, residual stress and distortion in a tool for high speed milling of steel components.

a	1% C, 0.4% Si, 1% Mn, 5% Cr, 1% Mo
b	0.4% C, 0.4% Mn, 0.3% Si
c	0.5% C, 4% Cr, 6% Mo

Answer

C. A high hardenability steel, which maintains its hardness at high operating temperatures when tempered (stable, secondary hardening Cr-Mo carbides).

35.8: Questions is shared under a not declared license and was authored, remixed, and/or curated by LibreTexts.

Index

A

Anisotropic thermal conductivity
2.3: Anisotropic Thermal Conductivity
Anisotropy
2: Introduction to Anisotropy

C

Carburization
20.3: Applications of Diffusion
cleavage steps
22.7: Observing Dislocations
contact mode (AFM)
3.3: Modes of Operation

D

diffusion coefficient
2.6: Anisotropic Diffusion
direct stiffness method
30.3: Direct Stiffness Method and the Global Stiffness Matrix
dislocation line
22.6: Dislocation in 3D

E

edge dislocation
22.6: Dislocation in 3D

F

Fick's First Law of Diffusion
20.1: Fick's First Law of Diffusion

G

global stiffness matrix
30.3: Direct Stiffness Method and the Global Stiffness Matrix

H

Hencky Relations
1.7: Hencky Relations

I

Interdiffusion
20.4: Interdiffusion
interfacial debonding
11.6: Spallation (Interfacial Debonding)

L

Leclanché cell
6.5.1: Zinc/carbon batteries

N

narrow dislocation
22.6: Dislocation in 3D

P

Pole vaulting
7.2: Pole-vaulting
polycrystal
4.5: Polycrystals
pyroelecticity
2.7: Anisotropic dielectric permittivity

R

residual curvature
7.8: Plastic deformation during beam bending

S

screw dislocation
22.6: Dislocation in 3D
spallation
11.6: Spallation (Interfacial Debonding)
Stoney equation
11.5: The Stoney Equation - the Thin Coating Limit

T

tapping mode (AFM)
3.3: Modes of Operation

W

wide dislocation
22.6: Dislocation in 3D

Z

zone folding
8.5: Zone Folding

Glossary

Sample Word 1 | Sample Definition 1

Detailed Licensing

Overview

Title: TLP Library I

Webpages: 396

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY-NC-SA 2.0: 97% (384 pages)
- Undeclared: 3% (12 pages)

By Page

- TLP Library I - *CC BY-NC-SA 2.0*
  - Front Matter - *CC BY-NC-SA 2.0*
    - TitlePage - *CC BY-NC-SA 2.0*
    - InfoPage - *CC BY-NC-SA 2.0*
    - Table of Contents - *Undeclared*
    - Licensing - *Undeclared*
  - 1: Analysis of Deformation Processes - *CC BY-NC-SA 2.0*
    - 1.1: Introduction to Analysis of Deformation Processes - *CC BY-NC-SA 2.0*
    - 1.2: Lévy-Mises Equations - *CC BY-NC-SA 2.0*
    - 1.3: Slip Line Field Theory - *CC BY-NC-SA 2.0*
    - 1.4: Work Formula Method - *CC BY-NC-SA 2.0*
    - 1.5: Limit Analysis - *CC BY-NC-SA 2.0*
    - 1.6: Finite Element Method - *CC BY-NC-SA 2.0*
    - 1.7: Hencky Relations - *CC BY-NC-SA 2.0*
    - 1.8: Deeper Questions - *CC BY-NC-SA 2.0*
  - 2: Introduction to Anisotropy - *CC BY-NC-SA 2.0*
    - 2.1: Introduction - *CC BY-NC-SA 2.0*
    - 2.2: Mechanical analogy of anisotropic response - *CC BY-NC-SA 2.0*
    - 2.3: Anisotropic Thermal Conductivity - *CC BY-NC-SA 2.0*
    - 2.4: Derivation of the anisotropy ellipsoid - *CC BY-NC-SA 2.0*
    - 2.5: Anisotropic electrical conductivity - *CC BY-NC-SA 2.0*
    - 2.6: Anisotropic Diffusion - *CC BY-NC-SA 2.0*
    - 2.7: Anisotropic dielectric permittivity - *CC BY-NC-SA 2.0*
    - 2.8: Optical anisotropy and the optical indicatrix - *CC BY-NC-SA 2.0*
    - 2.9: Liquid crystals - *CC BY-NC-SA 2.0*
    - 2.10: Summary - *CC BY-NC-SA 2.0*
    - 2.11: Questions - *CC BY-NC-SA 2.0*
  - 3: Atomic Force Microscopy - *CC BY-NC-SA 2.0*
    - 3.1: Introduction - *CC BY-NC-SA 2.0*
    - 3.2: Tip Surface Interaction - *CC BY-NC-SA 2.0*
    - 3.3: Modes of Operation - *CC BY-NC-SA 2.0*
    - 3.4: The Scanner - *CC BY-NC-SA 2.0*
    - 3.5: Tip and Cantilever - *CC BY-NC-SA 2.0*
    - 3.6: Feedback - *CC BY-NC-SA 2.0*
    - 3.7: Scanner Related Artefacts - *CC BY-NC-SA 2.0*
    - 3.8: Tip Related Artefacts - *CC BY-NC-SA 2.0*
    - 3.9: Other Artefacts - *CC BY-NC-SA 2.0*
    - 3.10: Summary - *CC BY-NC-SA 2.0*
    - 3.11: Questions - *CC BY-NC-SA 2.0*
  - 4: Atomic Scale Structure of Materials - *CC BY-NC-SA 2.0*
    - 4.1: Introduction - *CC BY-NC-SA 2.0*
    - 4.2: Single crystals- Shape and anisotropy - *CC BY-NC-SA 2.0*
    - 4.3: Single crystals- Mechanical properties - *CC BY-NC-SA 2.0*
    - 4.4: Single crystals- Optical properties - *CC BY-NC-SA 2.0*
    - 4.5: Polycrystals - *CC BY-NC-SA 2.0*
    - 4.6: Defects - *CC BY-NC-SA 2.0*
    - 4.7: Summary - *CC BY-NC-SA 2.0*
    - 4.8: Questions - *CC BY-NC-SA 2.0*
  - 5: Avoidance of Crystallization in Biological Systems - *CC BY-NC-SA 2.0*
    - 5.1: Introduction - *CC BY-NC-SA 2.0*
    - 5.2: Nucleation and crystallization - *CC BY-NC-SA 2.0*
    - 5.3: The water-sucrose system - *CC BY-NC-SA 2.0*
    - 5.4: Avoidance of crystallization by glass formation - *CC BY-NC-SA 2.0*
    - 5.5: Avoidance of crystallization by freeze resistance - *CC BY-NC-SA 2.0*

- 5.6: Summary - *CC BY-NC-SA 2.0*
- 5.7: Questions - *CC BY-NC-SA 2.0*

- 6: Batteries - *CC BY-NC-SA 2.0*
  - 6.1: Introduction - *CC BY-NC-SA 2.0*
  - 6.2: Basic principles - *CC BY-NC-SA 2.0*
  - 6.3: Thermodynamics and kinetics - *CC BY-NC-SA 2.0*
  - 6.4: Primary batteries - *CC BY-NC-SA 2.0*
  - 6.5: Zinc - *CC BY-NC-SA 2.0*
    - 6.5.01: Zinc - *CC BY-NC-SA 2.0*
      - 6.5.1: Zinc/carbon batteries - *CC BY-NC-SA 2.0*
  - 6.6: Alkaline - *CC BY-NC-SA 2.0*
    - 6.6.01: Alkaline - *CC BY-NC-SA 2.0*
      - 6.6.1: Alkaline/manganese oxide batteries - *CC BY-NC-SA 2.0*
  - 6.7: Questions - *CC BY-NC-SA 2.0*
  - 6.8: Zinc - *CC BY-NC-SA 2.0*
    - 6.8.01: Zinc - *CC BY-NC-SA 2.0*
      - 6.8.1: Zinc/silver oxide batteries - *CC BY-NC-SA 2.0*
  - 6.9: Secondary batteries - *CC BY-NC-SA 2.0*
  - 6.10: Lead - *CC BY-NC-SA 2.0*
    - 6.10.01: Lead - *CC BY-NC-SA 2.0*
      - 6.10.1: Lead/acid batteries - *CC BY-NC-SA 2.0*
  - 6.11: Lithium batteries - *CC BY-NC-SA 2.0*
  - 6.12: Battery characteristics - *CC BY-NC-SA 2.0*
  - 6.13: The future - *CC BY-NC-SA 2.0*
- 7: Bending and Torsion of Beams - *CC BY-NC-SA 2.0*
  - 7.1: Introduction - *CC BY-NC-SA 2.0*
  - 7.2: Pole-vaulting - *CC BY-NC-SA 2.0*
  - 7.3: Bending moments and beam curvatures - *CC BY-NC-SA 2.0*
  - 7.4: Maximising the beam stiffness - *CC BY-NC-SA 2.0*
  - 7.5: Beam deflections from applied bending moments - *CC BY-NC-SA 2.0*
  - 7.6: Twisting moments (torques) and torsional stiffness - *CC BY-NC-SA 2.0*
  - 7.7: Springs - *CC BY-NC-SA 2.0*
  - 7.8: Plastic deformation during beam bending - *CC BY-NC-SA 2.0*
  - 7.9: Summary - *CC BY-NC-SA 2.0*
  - 7.10: Questions - *CC BY-NC-SA 2.0*
- 8: Brillouin Zones - *CC BY-NC-SA 2.0*
  - 8.1: Introduction - *CC BY-NC-SA 2.0*
  - 8.2: Reciprocal lattice vectors - *CC BY-NC-SA 2.0*
  - 8.3: Brillouin Zone construction - *CC BY-NC-SA 2.0*
  - 8.4: The general case in three dimensions - *CC BY-NC-SA 2.0*
  - 8.5: Zone Folding - *CC BY-NC-SA 2.0*
  - 8.6: Brillouin Zones in Three Dimensions - *CC BY-NC-SA 2.0*
  - 8.7: Summary - *CC BY-NC-SA 2.0*
  - 8.8: Questions - *CC BY-NC-SA 2.0*
- 9: Brittle Fracture - *CC BY-NC-SA 2.0*
  - 9.1: Introduction - *CC BY-NC-SA 2.0*
  - 9.2: When do atomic bonds break? - *CC BY-NC-SA 2.0*
  - 9.3: Why do cracks weaken a material? - *CC BY-NC-SA 2.0*
  - 9.4: Inglis and the crack tip stress idea - *CC BY-NC-SA 2.0*
  - 9.5: Can we calculate the energy changes? - *CC BY-NC-SA 2.0*
  - 9.6: What about tension? - *CC BY-NC-SA 2.0*
  - 9.7: Another way of expressing the energies - *CC BY-NC-SA 2.0*
  - 9.8: Another way of calculating the energies - *CC BY-NC-SA 2.0*
  - 9.8: Why bother if they are the same? - *CC BY-NC-SA 2.0*
  - 9.9: Coping with a scatter in strength - *CC BY-NC-SA 2.0*
  - 9.10: When does the sample fail completely? - *CC BY-NC-SA 2.0*
  - 9.11: Sub-critical crack growth and R-curves - *CC BY-NC-SA 2.0*
  - 9.12: Summary - *CC BY-NC-SA 2.0*
  - 9.13: Questions - *CC BY-NC-SA 2.0*
- 10: Casting - *CC BY-NC-SA 2.0*
  - 10.1: Introduction - *CC BY-NC-SA 2.0*
  - 10.2: Heat Transfer - *CC BY-NC-SA 2.0*
  - 10.3: Heat Transfer Simulation - *CC BY-NC-SA 2.0*
  - 10.4: Microstructure and Segregation in Castings - *CC BY-NC-SA 2.0*
  - 10.5: Sand Casting - *CC BY-NC-SA 2.0*
  - 10.6: Die Casting - *CC BY-NC-SA 2.0*
  - 10.7: Continuous Casting - *CC BY-NC-SA 2.0*
  - 10.8: Investment Casting - *CC BY-NC-SA 2.0*
  - 10.9: Other Methods of Casting - *CC BY-NC-SA 2.0*
  - 10.10: Summary - *CC BY-NC-SA 2.0*
  - 10.11: Questions - *CC BY-NC-SA 2.0*
- 11: Coating mechanics - *CC BY-NC-SA 2.0*
  - 11.1: Introduction - *CC BY-NC-SA 2.0*
  - 11.2: Misfit Strains - *CC BY-NC-SA 2.0*

- 11.3: Force and Moment Balances - *CC BY-NC-SA 2.0*
- 11.4: The Biaxial Modulus - *CC BY-NC-SA 2.0*
- 11.5: The Stoney Equation - the Thin Coating Limit - *CC BY-NC-SA 2.0*
- 11.6: Spallation (Interfacial Debonding) - *CC BY-NC-SA 2.0*
- 11.7: Summary - *CC BY-NC-SA 2.0*
- 11.8: Questions - *CC BY-NC-SA 2.0*

12: Creep Deformation of Metals - *CC BY-NC-SA 2.0*

- 12.1: Introduction - *CC BY-NC-SA 2.0*
- 12.2: Creep Mechanisms - *CC BY-NC-SA 2.0*
- 12.3: Constitutive Laws for Creep - *CC BY-NC-SA 2.0*
- 12.4: Uniaxial Creep Testing - Practical Basics - *CC BY-NC-SA 2.0*
- 12.5: Multiaxial Creep Testing – The Creeping Coil - *CC BY-NC-SA 2.0*
- 12.6: Multiaxial Creep Testing – Indentation Creep Plastometry - *CC BY-NC-SA 2.0*
- 12.7: Designing for Creep Resistance - Nickel Based Superalloys - *CC BY-NC-SA 2.0*
- 12.8: Summary - *CC BY-NC-SA 2.0*
- 12.9: Questions - *CC BY-NC-SA 2.0*

13: Crystallinity in polymers - *CC BY-NC-SA 2.0*

- 13.1: Introduction - *CC BY-NC-SA 2.0*
- 13.2: Crystalline and Amorphous Polymers - *CC BY-NC-SA 2.0*
- 13.3: Spherulites and Optical Properties - *CC BY-NC-SA 2.0*
- 13.4: Polymer Stress-Strain Curve - *CC BY-NC-SA 2.0*
- 13.5: Tensile Testing of Polyethylene - *CC BY-NC-SA 2.0*
- 13.6: Summary - *CC BY-NC-SA 2.0*
- 13.7: Questions - *CC BY-NC-SA 2.0*

14: Crystallographic Texture - *CC BY-NC-SA 2.0*

- 14.1: Introduction - *CC BY-NC-SA 2.0*
- 14.2: Measurement of Texture - *CC BY-NC-SA 2.0*
- 14.3: Representing Texture - *CC BY-NC-SA 2.0*
- 14.4: Origins of Texture - *CC BY-NC-SA 2.0*
- 14.5: Effect of Texture - *CC BY-NC-SA 2.0*
- 14.6: Summary - *CC BY-NC-SA 2.0*
- 14.7: Questions - *CC BY-NC-SA 2.0*

15: Crystallography - *CC BY-NC-SA 2.0*

- 15.1: Introduction - *CC BY-NC-SA 2.0*
- 15.2: Lattices - *CC BY-NC-SA 2.0*
- 15.3: Unit Cell - *CC BY-NC-SA 2.0*
- 15.4: Lattice Geometry - *CC BY-NC-SA 2.0*
- 15.5: Crystal Structure - *CC BY-NC-SA 2.0*
- 15.6: Close Packing and Packing Efficiency - *CC BY-NC-SA 2.0*
- 15.7: Symmetry - *CC BY-NC-SA 2.0*
- 15.8: Combining symmetry - *CC BY-NC-SA 2.0*
- 15.9: Crystal Systems - *CC BY-NC-SA 2.0*
- 15.10: Summary - *CC BY-NC-SA 2.0*
- 15.11: Questions - *CC BY-NC-SA 2.0*

16: Deformation of honeycombs and foams - *CC BY-NC-SA 2.0*

- 16.1: Introduction - *CC BY-NC-SA 2.0*
- 16.2: Compression of a Honeycomb: Experimental - *CC BY-NC-SA 2.0*
- 16.3: Elastic Behaviour - *CC BY-NC-SA 2.0*
- 16.4: Yielding and Plateau Behaviour - *CC BY-NC-SA 2.0*
- 16.5: Densification - *CC BY-NC-SA 2.0*
- 16.6: Other Porous Structures - *CC BY-NC-SA 2.0*
- 16.7: Porous Structures in Bending - *CC BY-NC-SA 2.0*
- 16.8: Summary - *CC BY-NC-SA 2.0*
- 16.9: Questions - *CC BY-NC-SA 2.0*

17: Introduction to Deformation Processes - *CC BY-NC-SA 2.0*

- 17.1: Introduction - *CC BY-NC-SA 2.0*
- 17.2: Rolling - *CC BY-NC-SA 2.0*
- 17.3: Forging - *CC BY-NC-SA 2.0*
- 17.4: Extrusion - *CC BY-NC-SA 2.0*
- 17.5: Drawing - *CC BY-NC-SA 2.0*
- 17.6: Other Processes - *CC BY-NC-SA 2.0*
- 17.7: Summary - *CC BY-NC-SA 2.0*
- 17.8: Questions - *CC BY-NC-SA 2.0*

18: Dielectric materials - *CC BY-NC-SA 2.0*

- 18.1: Introduction - *CC BY-NC-SA 2.0*
- 18.2: Electric Dipole - *CC BY-NC-SA 2.0*
- 18.3: Polarisation Mechanisms - *CC BY-NC-SA 2.0*
- 18.4: Capacitors - *CC BY-NC-SA 2.0*
- 18.5: The Dielectric Constant - *CC BY-NC-SA 2.0*
- 18.6: Variation of the Dielectric Constant in Alternating Fields - *CC BY-NC-SA 2.0*
- 18.7: Effect of Structure on the Dielectric Constant - *CC BY-NC-SA 2.0*
- 18.8: Effect of Temperature on the Dielectric Constant - *CC BY-NC-SA 2.0*
- 18.9: Loss in Dielectrics - *CC BY-NC-SA 2.0*
- 18.10: Dielectric Breakdown - *CC BY-NC-SA 2.0*
- 18.11: Applications of Dielectrics - *CC BY-NC-SA 2.0*
- 18.12: Summary - *CC BY-NC-SA 2.0*

- 18.13: Questions - *CC BY-NC-SA 2.0*
- 19: Diffraction and imaging - *CC BY-NC-SA 2.0*
  - 19.1: Introduction - *CC BY-NC-SA 2.0*
  - 19.2: Diffraction Patterns - *CC BY-NC-SA 2.0*
  - 19.3: Two-dimensional Diffraction - *CC BY-NC-SA 2.0*
  - 19.4: Image Formation - *CC BY-NC-SA 2.0*
  - 19.5: Information Obtained from Diffraction Pattern - *CC BY-NC-SA 2.0*
  - 19.6: Applications of the Theory of Optical Diffraction and Imaging - *CC BY-NC-SA 2.0*
  - 19.7: Summary - *CC BY-NC-SA 2.0*
  - 19.8: Questions - *CC BY-NC-SA 2.0*
- 20: Diffusion - *CC BY-NC-SA 2.0*
  - 20.1: Fick's First Law of Diffusion - *CC BY-NC-SA 2.0*
  - 20.2: Fick's Second Law of Diffusion - *CC BY-NC-SA 2.0*
  - 20.3: Applications of Diffusion - *CC BY-NC-SA 2.0*
  - 20.4: Interdiffusion - *CC BY-NC-SA 2.0*
  - 20.5: Microstructural Effects - *CC BY-NC-SA 2.0*
  - 20.6: Temperature Effects - *CC BY-NC-SA 2.0*
- 21: Dislocation Energetics and Mobility - *CC BY-NC-SA 2.0*
  - 21.1: Introduction - *CC BY-NC-SA 2.0*
  - 21.2: Making a Dislocation - *CC BY-NC-SA 2.0*
  - 21.3: Join the Crystals to Form the Dislocation - *CC BY-NC-SA 2.0*
  - 21.4: Dislocation Width - *CC BY-NC-SA 2.0*
  - 21.5: Form of the Displacement - *CC BY-NC-SA 2.0*
  - 21.6: Change in the Misfit Energy of a Dislocation as it Moves - *CC BY-NC-SA 2.0*
  - 21.7: Peierls Energy - *CC BY-NC-SA 2.0*
  - 21.8: What is the Peierls Stress? - *CC BY-NC-SA 2.0*
  - 21.9: Determining the Peierls Stress - *CC BY-NC-SA 2.0*
  - 21.10: Lattice Resistance - *CC BY-NC-SA 2.0*
  - 21.11: Uses and Limitations of the Atomistic Model - *CC BY-NC-SA 2.0*
  - 21.12: Summary - *CC BY-NC-SA 2.0*
  - 21.13: Questions - *CC BY-NC-SA 2.0*
- 22: Introduction to Dislocations - *CC BY-NC-SA 2.0*
  - 22.1: Introduction - *CC BY-NC-SA 2.0*
  - 22.2: Dislocations in 2D - *CC BY-NC-SA 2.0*
  - 22.3: Bubble Raft - *CC BY-NC-SA 2.0*
  - 22.4: Dislocation Glide - *CC BY-NC-SA 2.0*
  - 22.5: Dislocation Motion - *CC BY-NC-SA 2.0*
  - 22.6: Dislocation in 3D - *CC BY-NC-SA 2.0*
  - 22.7: Observing Dislocations - *CC BY-NC-SA 2.0*
  - 22.8: Summary - *CC BY-NC-SA 2.0*
  - 22.9: Questions - *CC BY-NC-SA 2.0*
- 23: Elasticity in Biological Materials - *CC BY-NC-SA 2.0*
  - 23.1: Introduction - *CC BY-NC-SA 2.0*
  - 23.2: Hookean Elasticity - *CC BY-NC-SA 2.0*
  - 23.3: S-shaped Curves - *CC BY-NC-SA 2.0*
  - 23.4: Aneurysms - *CC BY-NC-SA 2.0*
  - 23.5: Demonstration of Aneurysms - *CC BY-NC-SA 2.0*
  - 23.6: J-Shaped Curves - *CC BY-NC-SA 2.0*
  - 23.7: Viscoelasticity and Hysteresis - *CC BY-NC-SA 2.0*
  - 23.8: Summary - *CC BY-NC-SA 2.0*
  - 23.9: Questions - *CC BY-NC-SA 2.0*
- 24: Electromigration - *CC BY-NC-SA 2.0*
  - 24.1: Introduction - *CC BY-NC-SA 2.0*
  - 24.2: The Theory of Electromigration - *CC BY-NC-SA 2.0*
  - 24.3: Electromigration Damage - *CC BY-NC-SA 2.0*
  - 24.4: Flux Divergence - *CC BY-NC-SA 2.0*
  - 24.5: Avoiding Electromigration Problems - *CC BY-NC-SA 2.0*
  - 24.6: Alternatives to Aluminium Metallization - *CC BY-NC-SA 2.0*
  - 24.7: Summary - *CC BY-NC-SA 2.0*
  - 24.8: Questions - *CC BY-NC-SA 2.0*
- 25: Ellingham Diagrams - *CC BY-NC-SA 2.0*
  - 25.1: Introduction - *CC BY-NC-SA 2.0*
  - 25.2: Thermodynamics - *CC BY-NC-SA 2.0*
  - 25.3: Chemical Reactions - *CC BY-NC-SA 2.0*
  - 25.4: The Equilibrium Constant - *CC BY-NC-SA 2.0*
  - 25.5: The Ellingham Diagram - *CC BY-NC-SA 2.0*
  - 25.6: Application - *CC BY-NC-SA 2.0*
  - 25.7: Partial Pressure of Reacting Gas - *CC BY-NC-SA 2.0*
  - 25.8: Reading p_O2 from the Ellingham Diagram - *CC BY-NC-SA 2.0*
  - 25.9: Other Gas Mixtures - *CC BY-NC-SA 2.0*
  - 25.10: Reducing Agents - *CC BY-NC-SA 2.0*
  - 25.11: Other Atmospheres and Non-Standard States - *CC BY-NC-SA 2.0*
  - 25.12: Summary - *CC BY-NC-SA 2.0*
  - 25.13: The Interactive Ellingham Diagram - *CC BY-NC-SA 2.0*
  - 25.14: Questions - *CC BY-NC-SA 2.0*
- 26: Expitaxial Growth - *CC BY-NC-SA 2.0*
  - 26.1: Introduction - *CC BY-NC-SA 2.0*

- 26.2: The Simulation and its Limitations - *CC BY-NC-SA 2.0*
- 26.3: Summary - *CC BY-NC-SA 2.0*
- 26.4: Questions - *CC BY-NC-SA 2.0*

- 27: Examination of a Manufactured Article - *CC BY-NC-SA 2.0*
  - 27.1: Introduction - *CC BY-NC-SA 2.0*
  - 27.2: Dismantling the Article - *CC BY-NC-SA 2.0*
  - 27.3: Classes of Materials - *CC BY-NC-SA 2.0*
  - 27.4: Metals - *CC BY-NC-SA 2.0*
  - 27.5: Polymers - *CC BY-NC-SA 2.0*
  - 27.6: Ceramics and Composites - *CC BY-NC-SA 2.0*
  - 27.7: Example Article - *CC BY-NC-SA 2.0*
  - 27.8: Summary - *CC BY-NC-SA 2.0*
  - 27.9: Questions - *CC BY-NC-SA 2.0*
- 28: Ferroelectric Materials - *CC BY-NC-SA 2.0*
  - 28.1: Introduction - *CC BY-NC-SA 2.0*
  - 28.2: The Dipole Moment - *CC BY-NC-SA 2.0*
  - 28.3: Polarisation - *CC BY-NC-SA 2.0*
  - 28.4: Switching Polarisation - *CC BY-NC-SA 2.0*
  - 28.5: Measurement of Polarisation - *CC BY-NC-SA 2.0*
  - 28.6: Fabrication of a KNO_3 Ferroelectric Capacitor - *CC BY-NC-SA 2.0*
  - 28.7: Temperature Dependence of the Hysteresis Loop - *CC BY-NC-SA 2.0*
  - 28.8: Barium Titanate - *CC BY-NC-SA 2.0*
  - 28.9: Order of Phase Transitions - *CC BY-NC-SA 2.0*
  - 28.10: Ferroelectrics - Why? - *CC BY-NC-SA 2.0*
  - 28.11: Summary - *CC BY-NC-SA 2.0*
  - 28.12: Questions - *CC BY-NC-SA 2.0*
- 29: Ferromagnetic Materials - *CC BY-NC-SA 2.0*
  - 29.1: Introduction - *CC BY-NC-SA 2.0*
  - 29.2: Types of Magnetism - *CC BY-NC-SA 2.0*
  - 29.3: Why are some materials magnetic? - *CC BY-NC-SA 2.0*
  - 29.4: Why are Only Some Materials Ferromagnetic - *CC BY-NC-SA 2.0*
  - 29.5: Curie-Weiss Law - *CC BY-NC-SA 2.0*
  - 29.6: Domains - *CC BY-NC-SA 2.0*
  - 29.7: Domain Walls - *CC BY-NC-SA 2.0*
  - 29.8: Hysteresis - *CC BY-NC-SA 2.0*
  - 29.9: Anisotropy - *CC BY-NC-SA 2.0*
  - 29.11: Questions - *CC BY-NC-SA 2.0*
  - 29.10: Summary - *CC BY-NC-SA 2.0*
- 30: Finite Element Method - *CC BY-NC-SA 2.0*
  - 30.1: Introduction - *CC BY-NC-SA 2.0*
  - 30.2: Nodes, Elements, Degrees of Freedom and Boundary Conditions - *CC BY-NC-SA 2.0*
  - 30.3: Direct Stiffness Method and the Global Stiffness Matrix - *CC BY-NC-SA 2.0*
  - 30.4: Enforcing Boundary Conditions - *CC BY-NC-SA 2.0*
  - 30.5: Interpolation/Basis/Shape Functions - *CC BY-NC-SA 2.0*
  - 30.6: 1D First Order Shape Functions - *CC BY-NC-SA 2.0*
  - 30.7: 1D Second Order Shapes and Functions - *CC BY-NC-SA 2.0*
  - 30.8: Typical steps during FEM modelling - *CC BY-NC-SA 2.0*
  - 30.9: Summary - *CC BY-NC-SA 2.0*
  - 30.10: Questions - *CC BY-NC-SA 2.0*
- 31: Fuel Cells - *CC BY-NC-SA 2.0*
  - 31.1: Introduction - *CC BY-NC-SA 2.0*
  - 31.2: The Principle - *CC BY-NC-SA 2.0*
  - 31.4: History of The Technology - *CC BY-NC-SA 2.0*
  - 31.4: Types of fuel cells - *CC BY-NC-SA 2.0*
  - 31.5: Solid Oxide Fuel Cells (SOFCs) - *CC BY-NC-SA 2.0*
  - 31.6: Electrolyte - *CC BY-NC-SA 2.0*
  - 31.7: Electrode Materials - *CC BY-NC-SA 2.0*
  - 31.8: Interconnection - *CC BY-NC-SA 2.0*
  - 31.9: Thermodynamics - *CC BY-NC-SA 2.0*
  - 31.22: Questions - *CC BY-NC-SA 2.0*
  - 31.10: System and Outlook - *CC BY-NC-SA 2.0*
  - 31.11: Molten carbonate fuel cells (MCFCs) - *CC BY-NC-SA 2.0*
  - 31.12: Electrolyte - *CC BY-NC-SA 2.0*
  - 31.13: Electrochemistry - *CC BY-NC-SA 2.0*
  - 31.14: Electrodes - *CC BY-NC-SA 2.0*
  - 31.15: Proton Exchange Membran Fuel Cells (PEMFCs) - *CC BY-NC-SA 2.0*
  - 31.16: Membrane - *CC BY-NC-SA 2.0*
  - 31.17: Electrodes and Membrane-Electrode Assembly - *CC BY-NC-SA 2.0*
  - 31.18: Efficiency and Reaction Conditions - *CC BY-NC-SA 2.0*
  - 31.19: Fuelling Requirements - *CC BY-NC-SA 2.0*
  - 31.20: Case Study: Fuelling Practicalities - *CC BY-NC-SA 2.0*
  - 31.21: Summary - *CC BY-NC-SA 2.0*
- 32: The Glass Transition in Polymers - *CC BY-NC-SA 2.0*
  - 32.1: Introduction - *CC BY-NC-SA 2.0*
  - 32.2: Theory 1 - *CC BY-NC-SA 2.0*
  - 32.3: Theory 2 - *CC BY-NC-SA 2.0*

- 32.4: Measurement of Tg - *CC BY-NC-SA 2.0*
- 32.5: Demostrations - *CC BY-NC-SA 2.0*
- 32.6: Summary - *CC BY-NC-SA 2.0*
- 32.7: Questions - *CC BY-NC-SA 2.0*

33: Granular Materials - *CC BY-NC-SA 2.0*

- 33.1: Introduction - *CC BY-NC-SA 2.0*
- 33.2: Terzaghi's Effective Stress Principle - *CC BY-NC-SA 2.0*
- 33.3: Ground Settlement and Centrifuge Modelling Experiments - *CC BY-NC-SA 2.0*
- 33.4: Angle of Repose - *CC BY-NC-SA 2.0*
- 33.5: Principle of Normality - *CC BY-NC-SA 2.0*
- 33.6: Yielding of Disturbed Saturated Soils in Triaxial Stress Tests - *CC BY-NC-SA 2.0*
- 33.7: Original Cam-Clay Model - *CC BY-NC-SA 2.0*
- 33.8: Liquefaction - *CC BY-NC-SA 2.0*
- 33.9: Wet Sand Drying Underfoot - *CC BY-NC-SA 2.0*
- 33.10: Summary - *CC BY-NC-SA 2.0*
- 33.11: Questions - *CC BY-NC-SA 2.0*

34: Indexing Electron Diffraction Patterns - *CC BY-NC-SA 2.0*

- 34.1: Introduction - *CC BY-NC-SA 2.0*
- 34.2: Mathematics - *CC BY-NC-SA 2.0*
- 34.3: Laue Zones - *CC BY-NC-SA 2.0*
- 34.4: Kikuchi Lines - *CC BY-NC-SA 2.0*
- 34.5: Using Polycrystalline Materials in the TEM - *CC BY-NC-SA 2.0*
- 34.6: Convergent Beam Electron Diffraction (CBED) - *CC BY-NC-SA 2.0*
- 34.7: Using Other Methods in Conjunction with Electron Diffraction - *CC BY-NC-SA 2.0*
- 34.8: Summary - *CC BY-NC-SA 2.0*
- 34.9: Questions - *CC BY-NC-SA 2.0*

35: The Jominy End Quench Test - *CC BY-NC-SA 2.0*

- 35.1: Introduction - *Undeclared*
- 35.2: Hardenability - *Undeclared*
- 35.3: Jominy Test - *Undeclared*
- 35.4: Uses of Jominy Data - *Undeclared*
- 35.5: Example Test Data - *Undeclared*
- 35.6: Simulations - *Undeclared*
- 35.7: Summary - *Undeclared*
- 35.8: Questions - *Undeclared*

Back Matter - *CC BY-NC-SA 2.0*

- Index - *CC BY-NC-SA 2.0*
- Glossary - *Undeclared*
- Detailed Licensing - *Undeclared*