A Survey of Materials Science

I. Structure

J. W. Morris, Jr.
Department of Materials Science and Engineering
University of California, Berkeley
Fall, 2007
# Table of Contents

## TABLE OF CONTENTS

### CHAPTER 1: INTRODUCTION

1.1 THE THING CALLED MICROSTRUCTURE ................................................. 4  
1.2 WHAT IS MATERIALS SCIENCE? ................................................................. 5  
   1.2.2 Composition and microstructure are independent variables .................. 7  
   1.2.3 Microstructure is controlled by processing ......................................... 7  
1.3 WHY STUDY MATERIALS SCIENCE? ......................................................... 8  
   1.3.1 Necessity ......................................................................................... 8  
   1.3.2 Ambition ......................................................................................... 9  
   1.3.3 Fear ............................................................................................. 10  
1.4 WHY CHANGE FROM GOOD BRONZE TO BAD IRON? ............................... 11

### CHAPTER 2: BONDING

2.1 INTRODUCTION ......................................................................................... 13  
2.2 ELECTRON CONFIGURATIONS IN ATOMS ............................................... 14  
   2.2.1 Electron orbitals in the atom ............................................................ 14  
   2.2.2 The ion core and the valence electrons ............................................ 16  
2.3 THE BOND MODEL ................................................................................. 17  
   2.3.1 Bonding as an electrostatic interaction ............................................ 17  
   2.3.2 The three mechanisms of chemical bonding .................................... 18  
   2.3.3 Shared valence electrons: covalent and metallic bonding ................. 19  
   2.3.4 Ionic bonding ................................................................................. 21  
   2.3.5 Dipole bonding ............................................................................... 22  
2.4 THE BAND MODEL ................................................................................. 24  
   2.4.1 Electron energy levels in atoms ....................................................... 24  
   2.4.2 Electron energy levels in solids ........................................................ 25  
   2.4.3 Core states and valence bands in solids ......................................... 27  
   2.4.4 The band model of conductor type ................................................... 29

### CHAPTER 3: CRYSTAL STRUCTURE

3.1 THE CONCEPT OF MICROSTRUCTURE ...................................................... 32  
3.2 CRYSTALS IN TWO DIMENSIONS ............................................................ 33  
   3.2.1 The crystal lattice ............................................................................ 33  
   3.2.2 Basis vectors and interpenetrating lattices ....................................... 34  
   3.2.3 The two-dimensional "Bravais" lattices .......................................... 36  
   3.2.4 Basis vectors in Bravais lattices ...................................................... 39  
3.3 CRYSTALS IN THREE DIMENSIONS ....................................................... 40  
   3.3.1 Three-dimensional crystal lattices ..................................................... 40  
   3.3.2 The Bravais lattices ........................................................................ 41  
   3.3.3 Distorted cubic cells ...................................................................... 46
### Chapter 4: Defects in Crystals

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>79</td>
</tr>
<tr>
<td>4.2 Point Defects</td>
<td>80</td>
</tr>
<tr>
<td>4.2.1 Intrinsic defects</td>
<td>80</td>
</tr>
<tr>
<td>4.2.2 Extrinsic defects</td>
<td>81</td>
</tr>
<tr>
<td>4.3 Line Defects: Dislocations</td>
<td>85</td>
</tr>
<tr>
<td>4.3.1 The edge dislocation</td>
<td>85</td>
</tr>
<tr>
<td>4.3.2 An edge dislocation in a simple cubic crystal</td>
<td>88</td>
</tr>
<tr>
<td>4.3.3 The Burgers circuit</td>
<td>90</td>
</tr>
<tr>
<td>4.3.4 Motion of an edge dislocation: glide and climb</td>
<td>90</td>
</tr>
<tr>
<td>4.3.5 Screw dislocations</td>
<td>92</td>
</tr>
<tr>
<td>4.3.6 Screw dislocations and plastic deformation</td>
<td>93</td>
</tr>
</tbody>
</table>

**4.3 Points, Directions and Planes in a Crystal**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.1 Points in a crystal</td>
<td>46</td>
</tr>
<tr>
<td>4.3.2 Directions in a crystal</td>
<td>47</td>
</tr>
<tr>
<td>4.3.3 Planes in a crystal</td>
<td>49</td>
</tr>
</tbody>
</table>

**4.5 The Common Crystal Structures**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6 The Close-Packed Structures: FCC and HCP</td>
<td>51</td>
</tr>
<tr>
<td>4.6.1 The FCC and HCP structures</td>
<td>52</td>
</tr>
<tr>
<td>4.6.2 Close-packing in FCC and HCP</td>
<td>53</td>
</tr>
<tr>
<td>4.6.3 Interstitial sites in the FCC and HCP</td>
<td>55</td>
</tr>
<tr>
<td>4.6.4 Properties of materials with the FCC and HCP</td>
<td>57</td>
</tr>
</tbody>
</table>

**4.7 The Body-Centered Cubic Structure**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.7.1 The BCC unit cell</td>
<td>58</td>
</tr>
<tr>
<td>4.7.2 Interstitial voids in the BCC structure</td>
<td>59</td>
</tr>
</tbody>
</table>

**4.8 The Diamond Cubic Structure**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.8.1 The diamond cubic cell</td>
<td>60</td>
</tr>
<tr>
<td>4.8.2 The DC structure as a stacking of planes and interstitials</td>
<td>61</td>
</tr>
<tr>
<td>4.8.3 Properties of elements with the DC structure</td>
<td>61</td>
</tr>
</tbody>
</table>

**4.9 Solid Solutions and Compounds**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.9.1 Solid solutions</td>
<td>62</td>
</tr>
<tr>
<td>4.9.2 Random solutions; pseudo-atoms</td>
<td>62</td>
</tr>
<tr>
<td>4.9.3 Ordered solutions (compounds)</td>
<td>63</td>
</tr>
</tbody>
</table>

**4.10 Substitutional Compounds**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.10.1 Substitutional BCC superstructure: CsCl</td>
<td>65</td>
</tr>
<tr>
<td>4.10.2 Substitutional FCC superstructure: Cu₃Au</td>
<td>66</td>
</tr>
</tbody>
</table>

**4.11 Interstitial Compounds**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.11.1 Octahedral interstices in FCC: NaCl</td>
<td>67</td>
</tr>
<tr>
<td>4.11.2 Tetrahedral interstices in FCC: β-ZnS</td>
<td>68</td>
</tr>
<tr>
<td>4.11.3 Tetrahedral Interstices in HCP: a-ZnS</td>
<td>69</td>
</tr>
<tr>
<td>4.11.4 Tetrahedral interstices in BCC: the A15 structure (Nb₃Sn)</td>
<td>70</td>
</tr>
</tbody>
</table>

**4.12 Understanding Crystal Structure**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.12.1 The structures of elemental solids</td>
<td>71</td>
</tr>
<tr>
<td>4.12.2 The structures of solid solutions</td>
<td>75</td>
</tr>
<tr>
<td>4.12.3 The structures of ordered compounds</td>
<td>75</td>
</tr>
</tbody>
</table>

**4.13 Ordered Solutions and Compounds**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.13.1 Ordered solutions (compounds)</td>
<td>63</td>
</tr>
<tr>
<td>4.13.2 Random solutions; pseudo-atoms</td>
<td>62</td>
</tr>
<tr>
<td>4.13.3 Properties of elements with the DC structure</td>
<td>61</td>
</tr>
</tbody>
</table>
CHAPTER 5: GLASSES

5.1 INTRODUCTION ................................................................. 110
5.2 THE GLASSY STATE ............................................................ 110
5.3 CONFIGURATIONAL GLASSES .............................................. 111
5.4 CHEMICAL GLASSES .......................................................... 113
  5.4.1 Network oxides ............................................................. 113
  5.4.2 Cross-linked polymers ................................................... 115
  5.4.3 Amorphous semiconductors ........................................... 115
  5.4.4 Metallic glasses .............................................................. 116

CHAPTER 6: MACROMOLECULAR SOLIDS ...................................... 118

6.1 INTRODUCTION ................................................................. 118
6.2 POLYMERS ......................................................................... 118
6.3 CHAIN POLYMERS ............................................................... 119
  6.3.1 Types of chain polymers .................................................. 119
  6.3.2 Internal symmetry in chain polymers ................................. 121
  6.3.3 Defects in chain polymers ............................................... 123
  6.3.4 Solid aggregates of chain polymers ..................................... 124
6.4 NETWORK POLYMERS ........................................................ 125
  6.4.1 Network polymers due to multi-directional bonds .................. 126
  6.4.2 Network polymers due to cross-links ................................. 127
  6.4.3 Elastomers .................................................................... 127
6.5 SILICATES ............................................................................. 128
  6.5.1 The SiO\textsubscript{4} tetrahedron ......................................... 128
  6.5.2 Three-dimensional network structures ............................... 129
  6.5.3 Two-dimensional sheet structures ...................................... 130
  6.5.4 One-dimensional chain structures ..................................... 132
6.6 ARTIFICIAL FIBERS ............................................................ 132
  6.6.1 Glass fibers .................................................................... 133
  6.6.2 Graphite fibers ................................................................ 134
  6.6.3 Kevlar fibers ................................................................... 135
  6.6.4 Whiskers ....................................................................... 136
6.7 QUASICRYSTALS ................................................................. 137
Chapter 1: Introduction

1.1 THE THING CALLED MICROSTRUCTURE

I may, perhaps, catch your attention by beginning with a quotation. I have chosen the following passage, which is Winston Churchill's succinct description of the birth of the Iron Age in Britain:

At this point (~ 400 BC] the march of invention brought a new factor upon the [British] scene. Iron was dug and forged. Men armed with iron entered Britain from the continent and killed the men of bronze. At this point we can plainly recognize across the vanished millenniums a fellow-being. A biped capable of slaying another with iron is evidently to modern eyes a man and a brother. It cannot be doubted that for smashing skulls, whether long-headed or round, iron is best.

Winston Churchill was individually responsible for a measurable fraction of the history of his own age and was, perhaps as a consequence, an unusually perceptive commentator on the histories of others. But in this passage he is wrong. I do not refer to his social psychology, which I am not competent to judge. He is wrong in his metallurgy. Modern research has shown that the iron that appeared in Britain at the end of the Late Bronze Age was, in fact, inferior in its salient mechanical properties to the bronze that preceded it. Since iron is also less dense than bronze, this metal was in every respect less suitable for smashing skulls, whether long-headed (Nordic) or round (Mediterranean).

So why change from good bronze to bad iron? I shall return to that point at the end of the chapter. I first want to excuse Sir Winston. His error is neither uncommon nor unreasonable. It would be made by almost anyone who is unschooled in materials science, and by a good many who claim intimate knowledge of the subject. To the average person the properties of a material are uniquely associated with its name, which is usually derived from its dominant chemical constituent or the whim of the company that manufactures it. Almost everyone knows that the Iron Age succeeded the Bronze Age. The idea that "iron" might be inferior to "bronze" is a possibility that a person who does not know metallurgy is unlikely to consider. Even physical scientists of impeccable credentials often assume that the properties of a material are uniquely associated with the atoms that make it up, and that those properties would be thoroughly understood if we only mastered behavior at the atomic level.

This notion is wrong. The material we call iron can be made weak (easily bent) or strong (virtually impossible for a human to bend), ductile (capable of being bent or deformed into complex shapes without fracture) or brittle (easily broken). Examples of all of these manifestations of iron are common today and useful in engineering. It follows
that when we describe a material as iron we have left out something important. In fact, we have left out something essential.

That thing that is missing from the designation, iron, is called microstructure. While the term "iron" describes the nature of the atoms that are present, or at least the dominant atom type, the term "microstructure" describes how those atoms are arranged. Both are necessary to understand the properties of iron. The same is true of any other engineering material. The composition and the microstructure together define the material; they specify what it is and what engineering properties it will have. Either, alone, is insufficient.

1.2 WHAT IS MATERIALS SCIENCE?

1.2.1 A simple definition

Through a straightforward extension of observations like these, the content of Materials Science can be summarized in two sentences:

1. The properties of a material are determined by its composition and its microstructure.

2. The microstructure of a material is determined by its composition and by the processing it has received.

The body of this course is the development of these two sentences. We shall concentrate on the first, since a substantial majority of those who study materials science at the introductory level intend to specialize in some other branch of engineering. They will be concerned with the engineering properties of materials, how these properties can be inferred from the nature of the material, and how they may change from one example of the material to another, but will be less interested in how materials are processed to control those properties.

Consider the technical terms that appear in statements 1 and 2.

The properties of a material determine its response to external stimuli. For the purposes of this course we shall divide the properties into three classes: thermochemical, electromagnetic and mechanical. The thermochemical properties of a material govern its response to thermal and chemical forces, and include, among others, its thermal stability, its conductivity for heat and mass, its reactivity with the environment and its ability to catalyze changes in the environment. The electromagnetic properties of a material govern its response to electromagnetic fields, and include the type and magnitude of its electrical conductivity, its optical characteristics, and its magnetic behavior. The mechanical properties of a material govern its response to mechanical loads, and include its elasticity, its strength (resistance to plastic deformation), and its toughness (resistance to fracture).
The *composition* of a material is its chemical constitution. The compositions of most inorganic solids are conveniently given by listing the weight or atom fractions of the atomic constituents. This is usually done even when the material can also be described as a chemical compound (such as salt, NaCl). The reason is that the molecules of inorganic substances ordinarily lose their identity in the solid in the sense that there is no unique way to identify the specific molecule to which a particular atom belongs. In most organic solids, on the other hand, and in some inorganic ones, the molecules have a strong internal bonding that is preserved in the solid state. In these cases the molecules retain their identity and it is convenient to specify the composition of the solid by the types and relative fractions of the molecules it contains. We shall return to this point in a later chapter.

The *microstructure* of a material is its atomic configuration - the pattern in which the atoms or molecules are distributed. The concept of *microstructure* is one that most students encounter for the first time in the study of materials science. In fact, it is the concentration on microstructure that differentiates materials science from the fundamental sciences of physics and chemistry. To understand the properties of a material one must not only know what chemical species are present, but must also know how they are configured, since different configurations of the same atoms can produce dramatically different responses to the same external stimuli.

While it is relatively straightforward to specify the composition of a material, it is enormously difficult to describe its microstructure in a completely satisfactory way. A cubic centimeter of solid contains on the order of $10^{22}$ atoms. This is a very big number. To appreciate its size, suppose that we had a characterization device that was so fast and so accurate that it could determine the type and location of an atom within a solid in one second (we have no such device). Suppose that we provided such a device to each member of a labor force of some 3 billion individuals (roughly the population of the earth) and set them to unbroken work (hypothetically, we would feed them intravenously with a nutrient solution that is as rich in caffeine as needed to prevent sleep, and procreate them, as we would have to do, by artificial means). This hyper-competent and super-dedicated labor force could identify the type and location of $10^{22}$ atoms in approximately 100,000 years. But even they could not determine the microstructure of a cubic centimeter of solid material. Atoms move in solids; they *diffuse* at a slow, but finite rate, and many millennia before the $10^{22}$nd atom had been typed and located, the first $10^{21}$st or so would be long gone from the position at which they had been found.

It follows that we can never really know what the microstructure of a material is. We describe it with approximations and abstractions, which, experience shows, work reasonably well a good part of the time. But these descriptions are never complete, never exact, and often impossible to express in a clean, quantitative form. It is this central fact that gives materials science the verbal, qualitative character that the practitioners of the more exact physical and engineering sciences often find disquieting.
1.2.2 Composition and microstructure are independent variables

Given that we can never know the exact microstructure of a material, it is worth asking whether we really need to, particularly in light of what you have probably learned in your studies of physics and chemistry. Imagine that you are given a collection of atoms of some particular composition. What happens if you seal them inside a container and leave them alone? Theoretical physics can answer that question, at least in principle. The atoms will eventually settle into a particular configuration, or microstructure, called the equilibrium state. For given values of the temperature and pressure the equilibrium state is uniquely determined by the chemical composition. Since composition and microstructure determine properties, it is sufficient to know the composition of the material to predict the engineering properties of this ultimate equilibrium state. It is then perfectly correct to talk about the properties of iron, or silicon, or iron-carbon alloys.

But engineering materials are solids. While the atoms that make up a solid will ultimately evolve into a unique equilibrium state, just as those in a vapor will, the rate of evolution is slow. It is so slow that there are virtually no solids in existence that have reached their true equilibrium states. The microstructure of a real solid is not determined by its ultimate equilibrium, but by its history. The same collection of atoms can be found in many different microstructures in the solid state and can, therefore, have many different sets of engineering properties. These microstructures are transient in the absolute sense; they inevitably evolve toward equilibrium. But if they are properly constructed they will remain almost unchanged for times that are very long compared to the design lives of the engineering devices in which they are used (a few days to a few thousand years).

1.2.3 Microstructure is controlled by processing

The microstructure of a material depends on how it has been processed, that is, on what has been done to it by nature or by man to fix its place in its evolution toward equilibrium. That is what is meant by the second statement in the general description of materials science. Given a substance with a particular composition, its microstructure reflects what has happened to it, and can be controlled in the engineering sense by intentionally processing it. The vast majority of the materials used in engineering are artificial, engineered materials. They are heated, beaten and reacted with one another to trap them into microstructures that impart useful engineering properties.

The possibility of processing given materials into various microstructures is the foundation of the greater part of human technology. The ultimate equilibrium states of materials are, for most practical purposes, uninteresting. Materials are made interesting by processing them into microstructures that are, ordinarily, far from ultimate equilibrium, but impart the properties that make engineering devices possible.

A few examples may help. Iron in its pure, equilibrium state is a soft metal that is not particularly attractive for use in plows or saws or automobiles or head-bashers. Even the addition of a bit of carbon, the essential step in converting iron to steel, doesn't neces-
sarily help. The equilibrium state of an iron-carbon alloy at ambient temperature and pressure is a simple physical mixture of almost pure iron and pure graphite whose mechanical properties are, if anything, less attractive than those of iron itself. But if a bit of carbon is dissolved in iron and the alloy is heated, deformed, and cooled quickly, it becomes a new material, called *martensitic steel*, that has a useful strength more than ten times that of soft iron. The transmutation of soft iron to hardened steel is not alchemy. The atoms are the same; only their arrangement has been changed.

Silicon in its pure, equilibrium state is a gray, brittle material that is a poor conductor of electricity. But if the right solutes are added in an inhomogeneous, non-equilibrium distribution that has just the right morphology, the chemically doped silicon becomes a microelectronic device of enormous versatility and power.

Carbon in its pure, equilibrium state is called graphite. Graphite is a flaky, relatively inexpensive substance that is used in pencil leads and lubricants. But when this same material is made in the form of long, flat filaments that are rolled on themselves, it becomes graphite fiber, one of the strongest materials known. And if it is held at high temperature and pressure in the presence of the right impurities for a few hundred thousand years, it metamorphoses into gem quality diamond, whose economic value is many orders of magnitude greater than the graphite from which it was formed. (A diamond is *not* forever. Every diamond in every piece of jewelry in the whole wide world will eventually transform into graphite, if some other mechanism doesn't destroy it first. Gem diamonds burned in the Oakland fire, some inside "fire-proof" safes. Try using that scientific fact to negotiate a better price. Good luck!)

To understand why engineering materials behave as they do we must grasp this thing called microstructure. To understand how engineering materials are made we must appreciate how microstructure is manipulated and controlled. *Microstructure* is the new idea you encounter when you go beyond the study of elementary chemistry and physics to the study of materials science. Its meaning will occupy us for the first part of the course; its engineering implications will fill the remainder.

### 1.3 WHY STUDY MATERIALS SCIENCE?

The vast majority of the students who take this course do not intend careers in Materials Science, but in some other branch of engineering. It is, therefore, pertinent to consider why they should care about the material covered here, other than to satisfy a healthy intellectual curiosity about how the world around them works. There are, in fact, a number of important motivations, the most important of which can be gathered under three headings: necessity, ambition and fear.

#### 1.3.1 Necessity

Engineering devices are made out of materials. More than that, engineering devices *are* materials: sets of materials that are interconnected so that their properties can
be used to accomplish some useful purpose. The selection of suitable materials is an integral part of engineering design. In advanced systems, materials selection is often the critical part of design. The ability to understand and select appropriate materials is an increasingly important part of the qualifications of a competent engineer.

The successful engineer of the future will not only understand the material properties that govern the operation of his devices, but will also have some feeling for the ancillary material properties that determine their manufacturability and durability. It is not enough to design a device; it must be possible to manufacture it at a reasonable cost, and use it with confidence for an acceptable service life.

When almost everything was made of wood, stone or steel, and shape, weight and durability were secondary considerations, the subtleties of materials science could be left to the research laboratories and the materials suppliers. That is no longer the case. Modern engineering materials, from alloy steels to GaAs chips, are developed in response to very specific needs in modern devices. Modern devices, from golf clubs to photonic systems, are designed to take advantage of unique properties of particular materials or attractive opportunities in materials processing. Almost every engineer who participates in the design, manufacture or use of advanced engineering devices inevitably must spend a portion of his time playing materials scientist. The conscientious engineer will want to do it well.

1.3.2 Ambition

We live in a time of phenomenal progress in engineering materials. Almost every day brings the announcement of a discovery, invention or improvement that pushes the frontiers of materials properties or the capabilities of materials processing. Almost every one of these announcements provides an opportunity for the perceptive engineer to design a new product or improve an existing one. Some of the advances in materials are spectacular, and lead to whole new classes of products. Integrated circuits, photocopying machines, and superconducting magnets are examples. Most are less striking, but provide a basis for the steady, incremental improvement of engineering devices that has, arguably, had a much greater cumulative impact on modern technology than the spectacular advances that are featured in the popular press.

As a familiar example, the transistor, which, in its solid-state version, is basically just a composite material with useful electrical properties, was invented by scientists at AT&T Bell Laboratories, incorporated into integrated circuits by engineers at Texas Instruments and at Fairchild Semiconductor, and successfully introduced into consumer electronics by engineers at Sony. The latter groups exploited a breakthrough in engineering materials to create a new class of devices. Their wisdom lay in seeing an opportunity and exploiting it, and, at least in the cases of TI and Sony, they arguably profited more from the invention of the transistor than AT&T (Lucent) itself.

It would be a mistake, however, to think that all ventures in the exploitation of new materials are successful. Counterexamples abound. To cite one of the most famous,
in the 1960's several companies in the United States, Europe and Japan developed composite structural materials that utilized graphite fibers as the strengthening element, and offered combinations of strength, rigidity and low density that exceeded all other materials. Engineers at Rolls Royce moved quickly to adopt these for fan blades in the advanced jet engine they had contracted to build for the Lockheed L-1011 jumbo jet. Very shortly thereafter Rolls Royce was bankrupt. While the graphite fiber composites had excellent mechanical properties under normal loads, they had very little resistance to the impact of particles that were directed perpendicular to the axis of the aligned fibers. Objects that were commonly ingested into the engines, such as birds, pebbles and raindrops, impacted the composite fan blades and destroyed them.

While this particular venture into composite materials technology was unsuccessful to a degree that bordered on catastrophe, the very real advantages of graphite-fiber and other structural composites have led to their widespread use in modern aircraft. However, they have been applied with more care and forethought, and still are not used to nearly the extent their developers envisaged.

In order for an engineer to take advantage of the opportunities new materials provide he must know about them, he must recognize their implications for the devices in his own area of specialization, and he must have some feeling for their limitations as well as their strengths. It usually does not take a PhD in Materials Science to do this. But it does take some knowledge of the subject. One must at least be able to read the right literature and ask the appropriate questions.

1.3.3 Fear

Devices cease to function and structures break apart, usually at the most inopportune times. Often the source of the problem is a material failure. Most material failures are merely troublesome, costly to repair and harmful to the reputations of the engineers and the companies responsible for them. But others cause injury or death. We have all seen headline examples of multi-fatality accidents in which airplanes crashed, spacecraft exploded, and structures collapsed because of the mechanical failure of some critical component. Behind the headlines is a steady stream of less spectacular problems that result from the inappropriate selection, processing or use of engineering materials. Some of these failures involve new phenomena that could not reasonably have been foreseen. Some result from simple carelessness or inevitable human error. However, an unfortunately large number are due to the misguided actions of nominally competent engineers who simply did not understand their materials well enough to anticipate that they might fall apart.

While the breakthroughs that produce new materials often spring from ideas that are fundamental and new, the mistakes that cause materials failures tend to be the very same ones, over and over again. Materials that are very strong (resistant to deformation) are almost always brittle (liable to fracture under relatively low loads). Small pre-existing cracks or flaws can cause fracture at loads well below those that could be easily borne in the absence of a flaw. Materials that are subject to cyclic loads are liable to
mechanical fatigue, a process in which internal cracks form and grow monotonically until the material finally breaks in two. When dissimilar metals are in contact in the presence of water one of them is likely to corrode at a very rapid rate. When metals slide in direct contact with one another, as in an unlubricated bearing or gear, they rapidly heat, soften and wear to failure. Exposure to high temperature accelerates microstructural evolution, and may cause a rapid and permanent deterioration in material properties. At low temperatures materials become strong, but may also become very brittle.

This is a partial list of the classic material failure modes a cautious engineer ought to know about, and will know about if he has an elementary knowledge of Materials Science. We shall cover all of them and others in this course. At the end of it you will hopefully have learned to treat engineering materials with a healthy mixture of confidence and trepidation.

1.4 WHY CHANGE FROM GOOD BRONZE TO BAD IRON?

To return to the question with which I began this chapter, there are a number of competing theories, and I have described some of them in previous editions of these notes. However, in keeping with Occam's razor, the most probable reason is also the simplest. They changed to iron because it was cheap.

In this the British were not unique. They were, in fact, repeating history from the cradle of civilization in the Middle East more than a millennium earlier. Every society, from the Hittites forward, changed to iron weaponry as soon as they learned how to make it, despite the fact that the iron they could make was everywhere inferior to good bronze. (When Goliath met David, in the biblical account, he was carried iron weapons but wearing bronze armor. His choices give a pretty good indication of which metal he thought would do the better job of protecting him.)

Early iron was inferior to good bronze, but it wasn't that bad. And it was plentiful and cheap. Given a choice between a thousand soldiers armed with iron and half that number armed with bronze, the wise king invested in iron. In many societies of the period soldiers were expected to provide their own weaponry. Given that he could afford fifty arrows tipped with iron or twenty tipped with bronze, the smart soldier made up his mind very quickly.

Iron is, arguably, the most versatile metal in the periodic table, and metallurgists gradually learned to make tools and weapons of iron that were far superior to any that preceded them. But that came much later. In the early days iron dominated the market because it was available and it was cheap.

If this is the case, can we, in Churchill's words, "plainly recognize across the vanished millenniums a fellow-being?" Most of us will have little trouble doing that. In fact, steel's place in the world market today is largely due to the fact that it is relatively cheap. One can make a better automobile out of more exotic materials, and the owners of
grand prix race cars do that. But most of us will continue to buy cars made primarily of steel and bank the difference in price. A surprisingly large fraction of the materials used in industry are chosen on the simple basis of cost and availability.
Chapter 2: Bonding

The more we stick together,
   Together, together
The more we stick together the happier we'll be
   - anonymous (for good reason)

2.1 INTRODUCTION

The fundamental reason that atoms bond together into molecules and solids is the electrostatic attraction between negatively charged electrons and positively charged nuclei. Since the atoms are relatively close together in a solid, their outer, *valence* electrons are simultaneously attracted to several different nuclei and, therefore, have lower average energy than they would if the atoms were widely separated. The energy that is released when atoms join together to form solids is called the *cohesive energy* or *binding energy* of the solid.

In principle, one can solve the equations of the quantum theory to find the electron states in a molecule or solid and calculate its binding energy. In practice, this can only be done for the simplest molecules. A small volume of a real material contains about $10^{22}$ atoms that interact simultaneously. Even if the equations that govern this vast system of particles could be solved, there is no practical way to write the solution down.

To get around this problem, theorists use approximations that let them focus on a single electron or a single interatomic bond, and construct models for the material as a whole by combining solutions to these simpler problems. Over the years they have refined their theories to a point where it is possible to produce very accurate descriptions of the electron states in solids. However, as is often the case when one develops approximate solutions to complex problems, they have attacked the problem in several qualitatively different ways. While the different approaches give essentially the same answers to the same questions, they are based on models that are, superficially, very different.

The materials scientist tends to jump back and forth between these various models, using whichever provides the most immediate and useful insight into the problem at hand. Since the different models are all known to be useful this is not an unreasonable way to proceed. But it does lead to some terminological confusion. In an attempt to avoid that we shall consider two complementary descriptions of the mechanisms of bonding in solids.

The two models of bonding that are most widely used, and are most fruitful in their qualitative features are those we shall call the *bond model* and the *band model*. The bond model employs a perspective and a terminology that is largely due to Linus Pauling
[L. Pauling, *The Nature of the Chemical Bond*], though it has been developed by many others. It focuses on the interaction between the valence electrons of the atoms that are immediate neighbors of one another, even when those atoms are immersed in an essentially infinite solid. The electrons are placed in atomic orbitals that overlap to form chemical bonds that are localized in space. The qualitative power of this model derives from the fact that the bonds are relatively easy to visualize, particularly when they are saturated in the sense that all the available bonding states are filled with electrons. The model is particularly useful for neutral molecules, in which the bonds are almost always saturated. It also provides a simple physical picture of bonding in most semiconducting and insulating solids. It is less useful for metals, which have inherently unsaturated bonds. And it is awkward to use when one is trying to understand properties such as electrical conductivity, which necessarily depend on the ability of electrons to move through distances that are very long on the atomic scale.

The band model begins from a very different starting point. It ignores the atomic orbitals of the electrons and instead places them in delocalized, molecular orbitals that spread throughout the solid. For the beginning student, the band model has the shortcoming that it explicitly invokes the underlying quantum mechanics and is, therefore, more difficult to master. Moreover, the electron states depend on exactly how the atoms are arranged, and are often difficult to visualize. On the other hand, the band model is the basis for almost all modern treatments of the solid state. At least its terminology should be understood. It provides a particularly useful framework for understanding the important electromagnetic properties of materials.

We shall briefly describe each of these models, beginning with the bond model. First we need a brief review of electron states in atoms.

### 2.2 ELECTRON CONFIGURATIONS IN ATOMS

#### 2.2.1 Electron orbitals in the atom

An atom consists of a nucleus of positively charged protons and neutral neutrons that is surrounded by negatively charged electrons. The number of protons in the atom sets its atomic number and its nuclear charge. In the neutral atom the number of electrons is equal to the number of protons. Since the weight of an electron is very small compared to that of a nuclear particle, the atomic weight is nearly equal to the weight of the nucleus.

The electrons are confined to discrete orbital states whose identity is fixed by the quantum theory, which also asserts, via the Pauli Exclusion Principle, that no more than one electron can occupy a given state. The admissible electron states in an isolated atom are specified by four quantum numbers whose values are represented by the symbols n, l, m and s. Electron states are distinct if they differ in the value of any one of these quantum numbers.
The principal quantum number, \( n \), identifies the primary orbit of the electron. Its possible values are the positive integers. The energy of the electron increases monotonically with the value of \( n \).

The azimuthal quantum number, \( l \), specifies the total angular momentum of the electron. Its value, \( l \), is equal to the number of nodes in the angular part of the electron wave function, which must be an integer such that \( 1 \leq n-1 \). Hence there are \( n-1 \) possible values of \( l \). The electron energy increases monotonically with the magnitude of the angular momentum, and, hence, increases with \( l \). It is common to denote the values of \( l = 0,1,2,3 \) by the lower case letters \( s,p,d,f \); hence the "2p orbital" indicates a state with \( n = 2, l = 1 \).

Since angular momentum is a vector, it has a definite direction in space. When the atom is placed in a field that establishes an axis in space then the direction of the angular momentum vector is quantized so that its component along that axis has one of several possible discrete values. These values are specified by the quantum number, \( m \), which is an integer such that \(-1 \leq m \leq 1 \). There are, hence, \( 2l + 1 \) distinct values of \( m \) for each value of \( l \). Since the most common source of a unique axis in an atom is the internal magnetic field, \( m \) is called the magnetic quantum number.

Finally, each electron has an internal spin. The spin is specified by the quantum number \( s \), whose value is \( \pm 1/2 \). There are, hence, two discrete electron states of different spin associated with each set of values of \( n, l, m \).

The energy of an electron in a free atom is nearly independent of the values of the quantum numbers \( m \) and \( s \). Hence the electronic configuration of an atom in its ground state is ordinarily written in a form that indicates the number of electrons in each occupied orbital with given values of \( n \) and \( l \). The symbol \( 2s^2 \), for example, has the meaning that there are two electrons in the \( s \)-orbital \((l = 0)\) of the second electron shell \((n = 2)\). The electronic configuration of the ground state of the element carbon (atomic number 6) is written \( 1s^22s^22p^2 \), which signifies that there are 2 electrons in the \( 1s \) orbital, 2 in the \( 2s \) orbital, and 2 in the \( 2p \) orbital. The \( 2p \) orbital can hold a total of 6 electrons since there are six distinct values of the pair \((m,s)\) when \( l = 1 \). In carbon only two of these states are occupied. Since the six states are degenerate (have equal energy) the configuration simply notes that two of them are filled without attempting to specify which two (because of the uncertainty principle, that question does not have a simple answer, anyway).

The ground state electronic configurations of the different elements are found by filling the available electron states in increasing order of energy until all of the electrons in the atom are accounted for. The lowest energy states are the two degenerate states in the \( 1s \) orbital. These are used to complete the configurations of the atoms in the first row of the periodic table: \( 1s^1 \) (H) and \( 1s^2 \) (He). The elements with atomic numbers 3 through 10 fill the second row of the periodic table, and have electron configurations that sequentially occupy the states of the second shell: \( 1s^22s^1 \) (Li), \( 1s^22s^2 \) (Be), \( 1s^22s^22p^1 \) (B), \( 1s^22s^22p^6 \) (Ne). The elements with atomic numbers 11 through 18 fill the third
row of the periodic table. Their electron configurations are determined by adding 3s and 3p electrons to the electron configuration of Ne. The 3s and 3p orbitals are filled in the ground state of Ar, whose configuration terminates with the filled shells 3s\(^2\)3p\(^6\).

The third electron shell also contains a d-orbital (l = 2) with 10 electron states. However, because the energy of an orbital increases with its angular momentum as well as with its principal quantum number, the 3d orbitals have higher energy than the 4s. The fourth row of the periodic table therefore begins with the elements K (...4s\(^1\)) and Ca (...4s\(^2\)). The 3d shell is then filled through the sequence of 10 transition metals Sc (3d\(^1\)4s\(^2\)) to Zn (3d\(^{10}\)4s\(^2\)). The ground state electronic structure of Cu (3d\(^{10}\)4s\(^1\)) is anomalous; because of electron-electron interactions it is energetically favorable to demote an electron from the 4s shell to fill the last empty state in the 3d. The fourth row is then completed by the six elements Ga (4s\(^2\)4p\(^1\)) to Kr (4s\(^2\)4p\(^6\)) that differ in the number of electrons in the 4p state. The fifth row of the periodic table is similar to the fourth. The 5s shell is first filled in the elements Rb and Sr, then the 4d is filled in the sequence of transition metals Y to Cd, and finally the 5p is filled in the elements In to Xe. While there is an 4f-orbital (l = 3), this orbital is not occupied until the La-series of rare earth elements (Ce-Lu) in the sixth row of the periodic table.

### 2.2.2 The ion core and the valence electrons

It is useful to divide the electrons in the free atom into two sets: core electrons and valence electrons. The division is based on the fact that the electrons in filled inner shells and, usually, the electrons in unoccupied d- and f-orbitals are concentrated rather tightly about the nucleus while the electrons in the outermost s- and p-orbitals extend well away from the nucleus into space. When atoms approach one another their outermost s- and p-electrons overlap and interact at internuclear distances that are so large that the inner electron orbitals are only slightly perturbed. Chemical bonding is primarily governed by the outermost, or valence electrons. It is, hence, a useful model to picture the atom as a relatively immutable ion core that contains the nucleus and core electrons, surrounded by a relatively diffuse cloud of valence electrons.

The number of valence electrons is ordinarily the sum of the numbers of electrons in the outermost s- and p-orbitals. For the purpose of understanding chemical bonding at the elementary level it is not necessary to distinguish between s- and p-states in the valence shell. The atomic s- and p-states are strongly perturbed by the interactions that occur during bonding and lose their distinction.

Complications do intrude, however, in deciding the valence of transition metals with unoccupied d-orbitals. The d-electrons may also participate in bonding. This can happen in two ways. First, d-electrons can be promoted into unoccupied sp valence states if the increase in interatomic binding energy outweighs the energy difference between the d- and sp orbitals. As a consequence the transition metals often exhibit multiple valence states in chemical bonding; for example, Ti (3d\(^2\)4s\(^2\)) sometimes bonds as if it had 2 valence electrons and sometimes 4, Fe (3d\(^6\)4s\(^2\)) sometimes has valence 2 and sometimes 3, and Cu (3d\(^{10}\)4s\(^1\)) exhibits valence 1 or 2, depending on the chemistry of the
compound in which it appears. Second, if there are empty d-orbitals the atoms may approach closely enough that these overlap slightly and directly participate in bonding. There is almost always some contribution from d-bonding in solids or chemical compounds that include transition metals.

2.3 THE BOND MODEL

2.3.1 Bonding as an electrostatic interaction

When two atoms approach one another their electrons and nuclei interact electrostatically. The electrons are mobile and, until the atoms are very close to one another, they can always configure themselves so that the net interaction is attractive. Atoms always bond to one another. The differences in the strengths of the bonds that are formed between different atomic species are due to differences in the specific electronic configuration that causes attraction.

![Binding Energy Diagram](image_url)

Fig. 2.1: Schematic drawing of the binding potential between two atoms as a function of their internuclear separation.

In addition to the attractive interactions that draw atoms together, there are also repulsive interactions that prevent a molecule or solid from collapsing on itself. The repulsive interaction is short-range, and is due to the Pauli Exclusion Principle. The filled electron shells in the ion cores cannot overlap one another. If the atoms draw too closely together, the ion cores must distort to keep the electrons apart, at a high cost in energy. The ion cores behave somewhat like rubber balls that deform on contact since they cannot penetrate one another.

The sum of the attractive and repulsive interactions leads to a two-body bonding potential whose qualitative form is shown in Fig. 2.1. Because of the long-range attraction, the potential energy of the atom pair decreases monotonically as they approach one another from a large distance, until they become so close that they sense the repulsive interaction of the ion cores. As the repulsive interaction increases in strength, the energy passes through a minimum at a finite separation, \( r_0 \), and increases as the internuclear
distance is decreased further. Neglecting the effect of atom vibrations, the equilibrium separation of the atoms is \( r_0 \), and the binding energy is the value of the potential, \( \Phi(r_0) \), at the equilibrium separation.

The bonding interaction between two atoms is reasonably well fit by the simple equation,

\[
\Phi(r) = \frac{A}{r^m} - \frac{B}{r^n}
\]

where \( \Phi(r) \) is the potential energy of the bond as a function of the internuclear separation, \( r \), \( A \) and \( B \) are constants, \( n \) is an exponent that governs the attractive, long-range interaction (its value depends on the specific bond type), and \( m > n \) is an exponent that governs the repulsive, short-range interaction. This function is called the Lennard-Jones potential, and has the shape shown in Fig. 2.1. The associated interatomic force,

\[
F = -\frac{d\Phi}{dr} = \frac{mA}{r^{m+1}} - \frac{nB}{r^{n+1}}
\]

vanishes in the limit of infinite separation, is attractive when the atoms are far apart, vanishes at the equilibrium separation, \( r_0 \), is repulsive when the atoms are very close together, and increases without bound as \( r \to 0 \) so that the atoms cannot merge into one another.

We shall find this simple representation of the binding potential useful at several points in the course to clarify the consequences of chemical binding.

**2.3.2 The three mechanisms of chemical bonding**

\[\text{Shared Electron Bond} \quad \text{Ionic bond} \quad \text{Dipole bond}\]

Fig. 2.2: The three basic mechanisms of interatomic bonding

While all interatomic bonding is due to electrostatic interactions between the electrons and the nuclei, it is useful to distinguish three basic mechanisms (Fig. 2.2). These correspond to the three different ways in which the electron clouds on adjacent atoms can
interact. For simplicity we consider a single atom pair, though the same considerations apply to groups of atoms in molecules and solids.

2.3.3 Shared valence electrons: covalent and metallic bonding

In the first mechanism the valence electron clouds interpenetrate so that, effectively, the atoms share electrons. This is the most potent mechanism of bonding. When an electron in an orbital about one nucleus penetrates the electron cloud of the other it feels the electrostatic attraction of both nuclei, and its energy is lowered substantially. However, the Pauli exclusion principle severely restricts the extent to which the electron clouds can penetrate one another. When an electron on one atom penetrates the electron cloud of another its charge density overlaps into the occupied orbitals of that atom. This is forbidden unless the orbitals are only partly filled; atoms can only share electrons if there are empty states in the outer, valence shells of each. The number of empty states in the valence shell determines the number of electrons that can be accommodated, that is, the number of shared electron bonds that can be formed.

The shared electron bond is the most important type of bond in engineering solids. Shared electron bonds are conventionally separated into two categories, covalent and metallic bonding, which lead to qualitatively different properties in the solid state.

**Covalent bonding**

Covalent bonding describes a situation in which all available valence states on the bonding atoms are used by shared electrons; the bonds are said to be **saturated**. The covalent bond is the usual bond type in molecular solids, semiconducting solids, and elemental insulators. Silicon offers a classic example of covalent bonding (Fig. 2.3). Each silicon atom has four valence electrons, and, in the solid, each atom has four nearest neighbors. If each silicon atom shares one electron with each of its four neighbors the valence shell of every atom is saturated.

![Covalent bonding in silicon](Fig. 2.3: Two-dimensional representation of covalent bonding in silicon, and the tetrahedral coordination of Si atoms in three dimensions.)

Covalent bonding tends to be directional; the bonds act as if they pointed in specific directions in space. To understand why this might be so, we can again consider the silicon atom. In the free atom the valence electrons have the configuration $3s^23p^2$. 
The unoccupied orbitals are in the 3p-shell. If atoms with this configuration were to form saturated bonds then some of the bonds would have to involve 3s-electrons on one atom shared with 3p states on the other. This asymmetric situation can be avoided if the four valence electrons are placed in hybrid atomic orbitals that are mixtures of the 3s and 3p states. The symmetric set of mixed s-p orbitals contains four "sp³ hybrid orbitals" that mix part of the s-state with part of each of the three p-states. The geometry of an sp³ hybrid orbital is such that the bulk of the associated charge density is concentrated into an extended lobe that is directed radially outward from the atom. The lobes of the four sp³ hybrid orbitals point toward the vertices of a tetrahedron like that on the right-hand side of Fig. 2.3. The energy of the free atom is raised by reconfiguring the electrons into sp³ hybrid orbitals, since all six of the higher-energy p-orbitals are used in this configuration, while only two of them are used in the 3s²3p² configuration of the free atom. However, the greater extension of these orbitals toward the nuclei of neighboring atoms lowers the energy per atom in the solid by an amount that is more than sufficient to compensate. The bonds in silicon are hence extended and directed in a tetrahedral pattern.

Similar considerations apply to other elements with valence four, and to stoichiometric compounds whose average valence is four, such as GaAs, ZnS, and SiC. Other covalently bonded materials, for example, elemental S and P, use other combinations of orbitals. Because these bonding orbitals are well-defined and saturated, there is always a strong tendency for covalent bonds to form at well-defined angles to one another.

Materials with saturated covalent bonds are inevitably poor conductors of electricity. This behavior can be understood by referring to the two-dimensional drawing in Fig. 2.3. For a solid to carry a current, electrons must be free to move in response to an electric field. But it is very difficult to displace an electron in the saturated bond structure shown in Fig. 2.3. Since each bond is filled by two electrons, the Pauli principle has the consequence that an electron can only jump from one bond to another if a second electron performs an equal and opposite jump to create an empty state to hold it. Hence there is no net flow of electricity. Electrons can only move freely through a covalent solid if they are excited to high energy states that correspond to unoccupied states of the free atom. This excitation is unfavorable since it increases the energy of the solid. If the required excitation energy is large, as it is in the diamond modification of carbon, then the solid is an insulator. If the required excitation energy is smaller, as it is in silicon, the material will conduct small currents at sufficiently high temperature, and is called a semiconductor. Materials with saturated covalent bonds are invariably either semiconductors or insulators.

Covalently bonded solids also tend to be difficult to deform mechanically. Without invoking the detailed mechanism of deformation, it can easily be seen that the shape of a solid can only be changed if the atoms move with respect to one another. Since covalent bonds are directional, it is relatively difficult to produce the local atomic reconfigurations that are required for deformation. The diamond modification of carbon, a covalently bonded material, is the hardest substance known.


**Metallic bonding**

The second category of shared electron bonds is called metallic. Metallic bonds are unsaturated; only some of the available valence states are needed when all valence electrons are shared. A classic example of the metallic bond is the body-centered cubic (BCC) modification of solid sodium (Na). In BCC sodium each atom has a single valence electron, and each atom has eight neighbors. When the valence electrons are fully shared with neighboring atoms many available valence states are left empty. The bonds are unsaturated, and hence metallic. The transition metals have additional empty d-states in the ion core, and can achieve additional bonding through the partial sharing of these.

![Illustration of metallic bonding](image)

**Fig. 2.4:** Illustration of metallic bonding. Valence electrons bind a distribution of ion cores.

Since the number of valence electrons in a metallic solid is much less than the number of available bonding states, it is somewhat meaningless to associate a particular electron with a particular bond. It is common to visualize metallic bonding by imagining the ion cores of the individual atoms suspended in a "sea" of valence electrons that bind them to one another and hold them in place, as illustrated in Fig. 2.4. Because of the relative freedom enjoyed by the bonding electrons, metallic bonding leads to high electrical conductivity. Moreover, the metallic bond is relatively non-directional. The large number of empty bonding states makes it possible for the electron cloud to relax in response to local perturbations in the configuration of atoms. Hence metals are often *ductile*. Metallic gold can be beaten into sheet or drawn into wire; covalent diamond cannot.

### 2.3.4 Ionic bonding

In the second basic mechanism of bonding one or more electrons are transferred from one atom to another so that the neutral atoms are converted into ions with alternate positive and negative charges. The ions then bond electrostatically to one another. This kind of bonding is called ionic. It predominates in molecules and solids that are made up of two or more kinds of atoms with very different ionization potentials, so that the electron transfer is energetically preferred. In pure ionic bonding the transferred charge is just sufficient to strip the valence electrons from one species, called the *cation*, and fill the valence shell of the other, called the *anion*, so that both have filled outer shells.
A familiar example of an ionic solid is table salt, NaCl. Sodium has a single electron in its outer shell. Chlorine has seven electrons in its valence shell. The transfer of an electron from Na to Cl creates the ions Na\(^+\) and Cl\(^-\), both of which have inert-gas electron configurations. In the solid state the Na\(^+\) and Cl\(^-\) ions minimize their energies by surrounding themselves with ions of the opposite type, as illustrated in Fig. 2.5. It is not particularly helpful to picture a NaCl molecule in the solid state; since the ions are not paired there is no unambiguous way to decide which particular ions belong to a given molecule. It is more useful to picture the solid as a stoichiometric mixture of Na\(^+\) and Cl\(^-\) ions.

![Schematic illustration of bonding in NaCl.](image)

Fig. 2.5: Schematic illustration of bonding in NaCl.

Most metal oxides and halides have predominantly ionic bonding. Because the valence electrons are localized and tightly bound to individual ions, an ionically bonded solid is usually an electrical insulator.

Saturated bonds between dissimilar atoms often have a mixed character that is partly covalent and partly ionic. Electrons are shared between neighboring atoms, but the sharing is asymmetric. The shared electron density is biased toward one of the two atoms, creating a net charge separation that is equal to a fraction of an electron charge. The degree of ionicity can be related to the difference in the electronegativity of the atoms, which is a semi-empirical parameter that measures the tendency of an atom to acquire an extra electron. The electronegativity decreases monotonically as one moves from right to left across a row of the periodic table, for example, from Cl to Na. As a consequence the ionic character of a heteronuclear bond increases as the atoms are drawn from successively separated columns of the periodic table (for example, NaCl is more ionic than ZnS). The electronegativity decreases as one moves down a column of the periodic table, essentially because the large number of electrons in a relatively heavy atom creates a diffuse charge distribution that partly shields the valence electrons from the nuclear charge. The ionic character of a heteronuclear bond decreases as the atoms become heavier (for example, ZnO is more ionic than CdS).

### 2.3.5 Dipole bonding

The third basic mechanism of bonding is the dipole bond, which is the dominant source of bonding between atoms of the noble gases and between neutral molecules. The
A dipole bond involves no charge transfer. It has its source in the distribution of charge in the neutral atom or molecule.

Many molecules contain permanent dipoles that are due to the inhomogeneity of the internal charge distribution. A familiar example is the water molecule, H$_2$O. As shown in Fig. 2.6, the configuration of the water molecule is such that the two hydrogen atoms are asymmetric about the oxygen, and lie to one side of it. Since the hydrogen nucleus is positively charged and the electron cloud is displaced toward the oxygen nucleus, the molecule has an inhomogeneous charge distribution, positive on the hydrogen side, negative on the oxygen side, that constitutes a permanent electric dipole. In the solid state the dipoles align and bond together, as illustrated in Fig. 2.2. Organic molecules often contain ions that create permanent internal dipoles, and bond into solids by aligning these dipoles with one another.

![Drawing of water molecule, H$_2$O, with its representation as a dipole.](image)

Local dipoles in many organic molecules are associated with dangling hydrogen atoms. These form local dipoles because the electron distribution about the hydrogen atom is strongly polarized toward the atom to which it is bonded, leaving a nearly bare nucleus on the external side. The local dipole interacts significantly with ions of O, N, and F, which are relatively small and negatively charged, so that they sense the intense local field of the H nucleus and interact with it. This type of bonding is particularly important in biological systems. For example, it binds the double helix of DNA. It is treated as a distinct bonding type (hydrogen bonding) in many texts.

A second kind of dipole bond, called the Van der Waals bond, is always formed between atoms, and is the primary source of bonding in electrically homogeneous molecules and noble gases that have no other bonding mechanism. Van der Waals bonding is due to transient dipole moments. The electrons that orbit the nucleus are in constant motion. When neutral atoms or electrically homogeneous molecules are close to one another, the interaction between the electrons causes their orbital motion to become correlated; they avoid one another to the extent possible. This correlated motion produces aligned, transient dipole moments that have a weak attractive interaction. This interaction is responsible, for example, for the solidification of inert gases at sufficiently low temperature.

Dipole-bonded solids are inevitably electrical insulators since the electrons are localized on atoms or molecules.
The different kinds of bonding occur simultaneously in most molecules and solids. However, one type usually predominates, and it is, therefore, convenient to classify materials as covalent, metallic, ionic or polar according to the dominant bonding mechanism. This classification is usually sufficient to characterize those properties that derive from the bond type. But it must not be applied too rigorously; each type of interaction makes some contribution to bonding in almost every case.

2.4 THE BAND MODEL

As we discussed in Section 2.1, the bond model is useful largely because bonds are easy to visualize. Electrons are localized on atoms or in simple bonds between neighboring atoms, and occupy electron states that are taken to be atomic orbitals or simple combinations of them. These simplifications are, however, achieved at some cost. The bond model is not a convenient one to describe those properties of solids that are strongly influenced by the fact that electrons are relatively free to move, such as electrical conductivity. Nor does it yield simple explanations for many of the bonding properties of solids, such as the fact that many elements and compounds can be found in a number of different atomic configurations in the solid state.

To understand these and other important phenomena, physicists have developed an alternate picture, which we call the band model, that places the electrons in delocalized orbitals that spread throughout the solid. This model is a bit more difficult to understand than the bond model, partly because its quantum mechanical features are closer to the surface. However, the effort is worthwhile. The band model provides a powerful theoretical foundation for understanding the behavior of solids. It is the basis of virtually all of modern solid state physics.

2.4.1 Electron energy levels in atoms

We can describe the qualitative features of the band model and discuss some of its consequences without invoking any complex mathematics. To begin, consider the electron states in an isolated atom. It is useful to plot the energies of the electron states against the potential energy of an electron on a line drawn through the atomic nucleus, as in done in Fig. 2.7. The potential energy is due to the electrostatic interaction between the electron and the nucleus. The electrostatic interaction between an electron of charge, -e, and a nucleus of charge, ze, has the result that an electron that is a distance, r, away from the nucleus has the potential energy

\[ V(r) = -\frac{ze^2}{r} \]  

2.3

The potential energy is related to the work that would have to be done to move a static electron away from the nucleus; the work required to displace the electron from a distance \( r_1 \) to \( r_2 \) is just
W = V(r₂) - V(r₁)

The potential energy is plotted and labeled in Fig. 2.7. It has a singularity at the position of the nucleus, and increases rapidly as the electron separates from it. The bound electron states of the atom have negative energies, E, with the consequence that electrons in these states cannot escape from the atom unless additional energy is added. The energies of the bound electrons are quantized into a sequence of discrete values, or energy levels, such as those indicated by the horizontal lines in the figure. These levels are separated by energy gaps that represent ranges of energy that an atomic electron cannot have. Electrons with positive energies are free to escape from an isolated atom and can have any energy. They may occupy any one of the continuum of unbound electron states indicated in the diagram.

Fig. 2.7: Schematic diagram of the electron energy states in an isolated atom.

2.4.2 Electron energy levels in solids

When the atom is joined to a large number of like atoms in a simple solid the potential and the associated electron energy levels are modified in two ways. These are illustrated in the diagram shown in Fig. 2.8, which is a schematic drawing of the potential and the electron energy levels along a chain of atoms within the solid.

First, when an electron is in the region between two nuclei it is attracted by both of them. The attractive interaction with two or more nuclei is responsible for the relatively low potential in the interatomic regions that is illustrated in Fig. 2.8. The valence electrons, in particular, take advantage of this fact to lower their energies by penetrating into the interatomic regions. The overall decrease in the energies of the valence electrons is the principal source of binding energy in the solid.

Second, the spread of the valence electron density creates a compatibility problem that reduces the binding energy. Once the electrons spread out into the internuclear regions, it is no longer possible to assign a particular electron to a particular atom. The
electrons are mobile and fill the whole solid. However, the Pauli Exclusion Principle still applies. An electron that is in, say, the 3s state in the isolated atom cannot simply move through the solid by successively occupying the 3s states of its neighbors, since these states are already filled. The valence electron states on the atoms must combine together to make new states that are independent of one another, spread through the solid, and hold no more than two electrons (one of each spin). These are the one-electron states of the solid.

![Diagram of electron energy states](image)

Fig. 2.8: Schematic diagram of the potential and the electron energy states along a line of atoms in a solid. The atom at the left lies at the solid surface.

One way to construct the different one-electron states of a solid is to let the valence electrons move through the solid by occupying identical states on the atoms they pass by, but traveling with different momenta. Since the momentum of an electron is part of its quantum state, electrons with different momenta are in different quantum states and, hence, satisfy the Pauli principle. An electron can, of course, be viewed as either a particle or a wave. If it is regarded as a particle, its momentum is written

\[ p = m^* v \]  \hspace{1cm} 2.5

where \( p \) is the magnitude of the electron's momentum, \( v \) is its velocity and \( m^* \) is its effective mass, which may differ from its physical mass because of the interaction with the nuclei and the other electrons. If the electron is regarded as a wave, its momentum can be written

\[ p = \frac{h}{\lambda} = \hbar k \]  \hspace{1cm} 2.6
where \( \lambda \) is the wavelength of the electron state, \( h \) is Planck's constant, \( h = h/2\pi \), and \( k = 2\pi/\lambda \) is the wavenumber. It follows from the Heisenberg Uncertainty Principle that the allowed values of \( k \) are separated from one another by a difference of the order

\[
\Delta k \approx \frac{2\pi}{a}
\]

where \( a \) is the interatomic separation. They vary from zero to the maximum value needed to accommodate two electrons per atom; since each state can hold electrons of either spin, the number of discrete values of \( k \) that are needed to describe the states that arise from a given atomic state is equal to the number of atoms in the solid. The momentum difference between successive states is of the order \( \Delta p \approx 10^{-24} \text{ kg} \cdot \text{m/sec}, \) which is a small number.

As the momentum, \( p \), increases, the kinetic energy, \( T \), increases as well, where

\[
T = \frac{p^2}{2m^*} = \frac{\hbar^2 k^2}{2m^*}
\]

This has the consequence that the discrete energy levels of the isolated atoms broaden into bands of allowed energy states, as illustrated in Fig. 2.8. Each band contains narrowly spaced electron energy levels, whose number is of the order of the number of atoms in the solid. However, these energy levels are so tightly spaced that the bands do not cover all possible energies. Just as there are discrete electron energy states in the free atom, separated by gaps that represent energies no electron can have, the electron states in the solid are gathered into bands that define the ranges of energy electrons can have, separated by energy gaps that define ranges of energy electrons cannot have.

### 2.4.3 Core states and valence bands in solids

The extent to which the atomic levels of the atom are broadened into bands increases with the degree of overlap between the corresponding electron orbitals on adjacent atoms. It is, therefore, useful to divide the atomic levels into two sets: the core states that are localized near the nucleus, and the valence states that spread out from the nucleus and overlap strongly.

The core states on separate atoms do not overlap significantly. Most of their charge density is concentrated near the nucleus where the potential is very nearly the same as it would be in an isolated atom. Hence they tend to retain their identity in the solid. There is always at least some overlap, since in quantum mechanics, unlike classical mechanics, a particle does penetrate into regions where its potential energy exceeds its total energy. The core electrons do tunnel from atom to atom at a finite rate, but since tunneling is difficult, the effective masses, \( m^* \), of the electrons in these bands is very high. Since the kinetic energy changes very little with \( p \), the bands are narrow. The core states are usually described as being tightly bound to the separate atom cores. When
it is necessary to describe their character or behavior we simply invoke the characteristics of the corresponding states of the free atom.

The valence and excited electron states overlap greatly and produce broad bands of states in which electrons move easily through the whole solid (small m*). The valence electrons control bonding and conduction in the solid. They move easily through the solid, and behave very much like free electrons in unbound states.

The development of bands as the electron states come to overlap one another is illustrated in Fig. 2.9, which shows the results of quantum mechanical calculations of the electron energies in sodium and carbon as a function of interatomic separation in the configuration each adopts in the solid state. In both cases the valence orbitals spread significantly and overlap one another in energy, while the core orbitals (illustrated by the 2p orbital of Na) remain narrow even at the equilibrium separation, r₀. The overlap of the valence states can have one of two results, which are illustrated by these two elements. In Na the 3s and 3p bands simply overlap to create a very broad band of allowed energy states at r₀. In carbon the 2s and 2p bands interfere and split, opening a new gap of energies that electrons cannot have when the atoms are spaced by r₀.

![Fig. 2.9: The splitting of atomic states into bands of allowed states as the interatomic spacing decreases in sodium and carbon.](image)

Once the associated bands begin to overlap in energy the atomic states lose their identity. It is not particularly meaningful to distinguish 2s and 2p states in solid carbon; just as in the bond model, the 2s and 2p states are mixed, or hybridized.

Because of the qualitative differences in the bands produced by the core and valence electrons, physicists have found it useful to use two different models to describe them. The core electrons are often treated in the **tight-binding model**, which approximates the electron states in the solid as linear combinations of the electron states on the individual atoms. Even the bands that arise from the partly filled d-electron shells
in the transition metals can be described reasonably well in this way. The valence electrons, on the other hand, are so heavily concentrated in the interstitial spaces between atoms that it is not useful to model them in terms of the atomic states. The most useful model for the valence electrons is the nearly free electron model in which the electrons behave as if they were almost free to move through the solid, but are slightly perturbed by their electrostatic interaction with the ion cores.

We shall not discuss these models in detail here, though we shall return to the nearly free electron model when we discuss the electromagnetic properties of solids. However, it is important to recognize some of the important qualitative implications of the band model for the type of conduction and the nature of interatomic bonding in solids.

2.4.4 The band model of conductor type

Once the available electron states have been identified, the ground state of the solid is obtained by filling the states with electrons in order of increasing energy until all of the available electrons have been used. The electrical character of the solid is determined by the highest electron energy in the ground state.

The band structures of metallic conductors

![Band Structures](image)

Fig. 2.10: Three configurations of the electron energy bands that lead to metallic conduction (the shaded region shows the filled electron states for E < E_F; filled states in overlapping bands are shaded dark). (a) The valence band is only partly filled. (b) The valence and conduction bands overlap. (c) A narrow, partly filled d-band overlaps the valence band.

If the most energetic level that is filled in the ground state of a solid falls in the interior of a band then the energy of this state is called the Fermi energy, E_F, and the material is a metallic conductor. In the interior of a band there are many empty electron states arbitrarily close to E_F, and electrons can easily move in response to an applied electric field. (More precisely, E_F is the energy of the electron state that has probability 1/2 of being filled. In a conductor, this level is very close to the most energetic level that
would be filled if the electrons were placed in the states of lowest possible energy. We shall discuss this point further when we consider the electrical properties of conductors.)

In the band model there are two situations that lead to metallic conduction. They are illustrated schematically in Fig. 2.10, which is a drawing of the band structure like Fig. 2.8, but with the potential energy plot removed for simplicity. The first kind of metal has a valence band that is only partly filled. Since there are two electrons states per energy level, a band necessarily contains an even number of electron states. Elements that have odd valence are, hence, necessarily metallic unless the atoms are gathered into molecular groupings that have an even number of valence electrons. Group IA elements such as Li, Na and K, Group IB elements like Cu, Au and Ag, and Group IIIB elements like Al, Ga and In are all metals. However, the Group IIIB metal boron (B) is a semiconductor in its normal solid configuration. The reason is that boron atoms are associated in the solid in pairs that behave like diatomic molecules whose net valence is even.

The second kind of metal has an even number of valence electrons, but has a partly filled upper band because nominally independent bands overlap. There are two ways in which this happens. The normal valence elements from Groups IIA, IIB, and IVB (white Sn and Pb) are metals essentially because the band of excited states (Fig. 2.8), which is often called the conduction band, overlaps the valence band. Some electrons from the valence band fill lower energy states in the conduction band, and both bands are left partly filled. The transition metals are metallic because a narrow band of partly filled d-states overlaps the valence band at E_F. Both the d-band and the valence band are then only partly filled. While very little electrical current is ordinarily carried by the d-electrons, which are localized near the ion core, the empty states in the d-band provides a reservoir of available electron states that ensures the availability of empty states in the valence band.

*The band structures of semiconductors and insulators*

In many other solids the highest filled energy level in the ground state is precisely at the top of the valence band, as illustrated in Fig. 2.11. These solids cannot easily conduct electricity. The valence electron states are filled and their velocities sum to zero. The nearest empty states are in the band of excited states (called the conduction band) that is separated from the valence band by a finite energy gap, \( E_G \). Materials that have the band structure described in Fig. 2.11 are called intrinsic semiconductors (or insulators). As we shall see at a later point, mobile electrons can also be introduced into the solid by solute atoms or other defects, in which case the conductivity is called extrinsic.

Whatever the value of \( E_G \), a material with a band configuration like that shown in Fig. 2.11 always has some conductivity. Because of thermal activation a few electrons are excited into states in the conduction band at any finite temperature, leaving a corresponding number of vacant states (holes) in the valence band. The electrons in the par-
tially filled states of the conduction band can conduct electricity, and there is also some conductivity in the valence band because of the vacant states there. In an intrinsic semiconductor the number of conduction electrons (n) is equal to the number of holes (p) and is given approximately by the Arrhenius function

\[ n \sim N_0 e^{-\frac{E_G}{2kT}} \]  

2.9

The Fermi energy, \( E_F \), is the energy of a state that has probability 1/2 of being filled by an electron. In an intrinsic semiconductor or insulator it is located in the center of the band gap.

![Band configuration in an intrinsic semiconductor or insulator.](image)

The valence band is filled, the conduction band empty. The Fermi energy is located in the center of the band, dividing the filled and empty states.

The magnitude of the band gap, \( E_G \), determines whether the material is called a semiconductor or insulator. When the band gap, \( E_G \), is large the number of conduction electrons is very small. The conductivity is low and the material is an insulator. As \( E_G \) decreases the number of conduction electrons increases. The conductivity becomes measurable and the material is called a semiconductor.
Chapter 3: Crystal Structure

Without stirring abroad
One can know the whole world
Without looking out of the window
One can see the way of heaven
...The sage knows without having to stir
Identifies without having to see
   - Lao Tzu, "Tao Te Ching"

3.1 THE CONCEPT OF MICROSTRUCTURE

The properties of a material are determined by the atoms or molecules it contains (composition) and by their arrangement (microstructure). The microstructure is the difficult part of this verbal equation. Given the available materials characterization techniques and all plausible extensions of them, there is no hope of specifying where the atoms are located in a solid and there is no practical way of storing that vast amount of information in retrievable form. A practical description of the microstructure of any real material must, therefore, be based on strong simplifying approximations.

The practical description of microstructure begins from the observation that one can locate an arbitrarily large number of atoms when they are arranged in one of two ideal ways. First, let the atoms are distributed in a perfectly regular pattern. Their positions can be specified by locating a few of them for reference and giving the rule that governs the positions of all the others with respect to these. The simplest regular distributions are periodic distributions in which an atom or identical group of atoms is repeated at regular intervals in space. The materials whose atoms or molecules fill space by periodic repetition are called crystalline solids. Most engineering materials fall into this class. Their crystal structures are described by giving the elementary configuration of atoms that is repeated through space and the translation vectors, or lattice vectors, that determine the repeat pattern. Since no real material is a perfect crystal, the microstructure of a crystalline solid is described by noting its basic crystal structure and, to the extent possible, describing the distribution of lattice defects, or places where the ideal pattern is not followed.

Second, let the atoms or molecules be distributed through space in a perfectly random way. While the precise location of each particle is unspecified, the probability that a particular type of particle is located at a particular place is well defined, and often provides an adequate description of the microstructure. The materials that have approximately random distributions of atoms or molecules are called amorphous solids, or glasses. They are mostly materials that are cooled from a liquid or gaseous state so quickly that a nearly random atom distribution is preserved. Polymeric and other organic
solids that are made of large molecules are often amorphous. Inorganic compounds, such as the silica glasses that are used for window panes, can often be made amorphous by rapid cooling from the melt. Many semiconductors and some metal alloys can be made amorphous by solidifying them very rapidly or depositing them from the vapor onto a cold substrate.

The microstructure of an engineering material is described by deciding whether it is more nearly crystalline or more nearly amorphous, stating its crystal structure if it is crystalline, and completing the description of microstructure by noting its defects, or deviations from the ideal state. In the following chapters we shall first describe the crystal structures of simple solids and the defects they contain, and then discuss the structures of amorphous materials.

3.2 CRYSTALS IN TWO DIMENSIONS

First consider periodic distributions of atoms over a plane, or two-dimensional crystals. Most of the important elements of crystal structure are present in two-dimensional crystals, which are easier to draw and visualize than three-dimensional atom arrangements. Two-dimensional crystals are relevant to materials science; the atoms in films on solid surfaces are often arranged in two-dimensional crystalline patterns.

3.2.1 The crystal lattice

When an atom or group of atoms is repeated periodically over a plane it is possible to find two non-parallel vectors, $\mathbf{a}_1$ and $\mathbf{a}_2$, that connect identical sites in the repeated atom group. The directions of the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ give the two directions in which the atom group is repeated. Their magnitudes give the repeat periods. If $\mathbf{a}_1$ and $\mathbf{a}_2$ connect identical sites, a vector that is equal to $m\mathbf{a}_1 + n\mathbf{a}_2$, where $m$ and $n$ are any positive or negative integers, also connects identical sites. Identical atom groups surround every point, $\mathbf{r}$, in the plane, that satisfies the relation

$$\mathbf{r} = m\mathbf{a}_1 + n\mathbf{a}_2$$  

where $m$ and $n$ are integers. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ are called the lattice vectors, and the points, $\mathbf{r}$, are lattice points that form a regular periodic grid, or lattice, on the plane. A two-dimensional crystal results when an identical group of atoms appears at every lattice point so that the physical environment of every lattice point is precisely the same.

A two-dimensional crystal of identical atoms is drawn in Fig. 3.1. The repeated group is a single atom, the lattice is specified by the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ in the figure, and an atom appears at every lattice point. As shown in the figure, the lattice divides the plane into elementary blocks, or unit cells, that are parallelograms and are stacked in two dimensions over the plane. An alternate and equivalent description of the crystal structure is obtained by specifying the atom configuration within the unit cell and noting that identical unit cells are repeated periodically to fill the plane.
Note that there is one atom per unit cell in the structure shown in Fig. 3.1. The atom at each corner of the cell is shared by four adjacent cells for a net of one atom per cell. The unit cell of a structure that contains a net of one atom per cell is called a primitive cell.

![Figure 3.1: A two-dimensional crystal. The lattice vectors are drawn. The unit cell is the shaded parallelogram in the figure.](image1)

In Fig. 3.1 the atoms sit on the lattice points. Since the lattice points are determined by eq. 3.1, this means that the origin of the coordinate system was placed at an atom site. This choice is usually convenient, but it is not necessary; the origin can be placed anywhere. Fig. 3.2 shows the same distribution of atoms that appears in Fig. 3.1, referred to a lattice with a different origin. The lattice is identical, but displaced. The unit cell has the same shape but the atom configuration within it is different. It is still true that every lattice point has the same environment, and it is possible to make a one-to-one association between the atoms and the lattice points so that the crystal can be made by placing an atom in an identical position with respect to each lattice point. The shaded unit cell in Fig. 3.2 makes it clear that the cell is primitive; it contains exactly one atom, and a net of one lattice point.

![Figure 3.2: An alternate crystal lattice and unit cell for the atom configuration shown in Fig. 3.1.](image2)

### 3.2.2 Basis vectors and interpenetrating lattices

Complex two-dimensional crystals are generated by placing identical groups of atoms (molecules) at each lattice point. An example is shown in Fig. 3.3, where the repeated group is a two-atom cluster (or diatomic molecule), distributed over a lattice
identical to that shown in Fig. 3.1. The unit cell of the lattice has the same shape and size. It is a parallelogram defined by the vectors $a_1$ and $a_2$, and is primitive since it contains a net of one lattice point. However, the unit cells of the two crystal structures are different. The shaded cell in Fig. 3.3 is non-primitive since it contains two atoms. This result emphasizes the importance of differentiating the crystal lattice and the crystal structure. The lattice is a periodic distribution of points in space that have identical atomic environments. The structure is the actual atom distribution. To specify it, we need both the crystal lattice and the internal configuration of the atom group that is repeated at each lattice point. Alternatively, we can specify the geometry of the unit cell and the arrangement of atoms within it.

![Crystal lattice diagram](image)

Fig. 3.3: A two-dimensional crystal made by repeating a two-atom cluster periodically over the plane. The black dots are the lattice points. The open circles are the atoms. The unit cell is shaded.

![Unit cell diagram](image)

Fig. 3.4: The unit cell of the structure in Fig. 3.2, showing how basis vectors are used to locate the two atoms in the unit cell.

The positions of the atoms in a complex two-dimensional crystal are given by a set of equations of the form

$$r_i = ma_1 + na_2 + \tau_i$$

where the $r_i$ are the vector positions of atoms of the $i$th type in the repeated atom cluster, $m$ and $n$ are integers (positive, negative or zero), and $\tau_i$ is a small vector (called a basis vector) that gives the vector displacement of the $i$th atom of the cluster from a lattice point. The lattice points are automatically generated as the integers $m$ and $n$ range over all possible values. The crystal structure can then be drawn by placing an atom of the
appropriate type at a position displaced by the vector, \( \mathbf{\tau}_i \), from each lattice point, and repeating the procedure for each value of \( \mathbf{\tau}_i \).

A suitable set of basis vectors for the structure in Fig. 3.3 is shown in Fig. 3.4. The lower corner of the unit cell is a lattice point, and the two atoms within the unit cell are located by vector displacements from the lattice point. It should be clear that the basis vectors can be chosen in many equivalent ways, just as the unit cell can be.

A complex two-dimensional crystal can also be described as a set of interpenetrating primitive lattices, as illustrated in Fig. 3.5. Let there be \( n \) atoms in the cluster that is repeated at each lattice point. The \( i^{\text{th}} \) of these is displaced by the vector \( \mathbf{\tau}_i \) from the lattice point, as illustrated in Fig. 3.4. Since there is an identical atom located at \( \mathbf{\tau}_i \) from every lattice point, if we consider only the \( i^{\text{th}} \) atom in every cluster these are distributed on the sites of a lattice that is identical to the basic lattice of the crystal. Since this is true for each of the distinct atoms in the cluster, the whole crystal can be described as a set of \( n \) identical, interpenetrating lattices, each of which defines the positions of one of the atoms of the basic cluster. The identical primitive lattices that give the positions of the \( i^{\text{th}} \) and \( j^{\text{th}} \) atoms in the repeated group are displaced from one another by the vector \( \mathbf{\tau} = \mathbf{\tau}_i - \mathbf{\tau}_j \). The structure of Fig. 3.3 is redrawn in Fig. 3.5 to illustrate how the two atoms in the unit cell are located on the lattice points of identical lattices that interpenetrate and are slightly displaced from one another.

![Diagram](image)

**Fig. 3.5:** The structure of Fig. 3.3 re-drawn to show that the two atoms in the unit cell are located on the lattice sites of two identical, interpenetrating lattices.

### 3.2.3 The two-dimensional "Bravais" lattices

All two-dimensional crystal structures can be described as the repetition of an atom or group of atoms at the lattice points of a two-dimensional primitive lattice. The unit cell of the primitive lattice is a parallelogram whose sides are defined by the lattice vectors, \( \mathbf{a}_1 \) and \( \mathbf{a}_2 \). However, if the sides of the parallelogram have just the right lengths and angles the parallelogram reduces to a simpler figure, such as a rectangle, a square, or a piece of a hexagon. When the unit cell has one of these simpler shapes, it is useful to recognize that fact, since the symmetry of the unit cell often helps to visualize the crystal structure, to understand why the material chooses a particular crystal structure, and to anticipate material properties.
There are four cases in which the unit cell of a two-dimensional primitive lattice has a geometrically simple shape. Together with the parallelogram, these figures define the five two-dimensional lattices, which we shall call the \textit{two-dimensional Bravais lattices} to emphasize their connection with the three-dimensional Bravais lattices that are discussed in a later section. Let $a = |\mathbf{a}_1|$, $b = |\mathbf{a}_2|$ and $\gamma$ be the angle between $\mathbf{a}_1$ and $\mathbf{a}_2$. The five Bravais lattices are determined by the relative values of $a$ and $b$ and the value of $\gamma$.

The three simplest of the two-dimensional Bravais lattices are drawn in Fig. 3.6. When $a \neq b$ and $\gamma \neq 90^\circ$ the unit cell is a parallelogram, and is called an \textit{oblique lattice}. When $\gamma = 90^\circ$ the cell is \textit{rectangular} if $a \neq b$ and \textit{square} if $a = b$. In all three of these cases the unit cell is primitive.

![Fig. 3.6: The unit cells of the oblique, rectangular and square lattices.](image)

A \textit{close-packed hexagonal lattice} is formed when $a = b$ and $\gamma = 60^\circ$ or $120^\circ$. The lattice and its unit cell are drawn in Fig. 3.7.

![Fig. 3.7: The close-packed hexagonal lattice and its primitive cell. The hexagonal cell is outlined on the lattice.](image)

While the primitive cell of the close-packed hexagonal lattice is a parallelogram, the special symmetry of the parallelogram has the consequence that each atom is located at the center of a hexagon of like atoms as indicated by the heavy lines in the figure. For purposes of understanding and visualizing the close-packed hexagonal structure it is usually more convenient to use the hexagon as the unit cell. It is a regular hexagon with lat-
tice points at each of its six vertices and a lattice point at its center. It is a non-primitive cell; since each vertex is shared by three adjacent hexagonal cells, there is a net total of three lattice points per unit cell.

The close-packed hexagonal lattice is the most important of the two-dimensional lattices since it shows the pattern that is formed when spherical atoms are packed together as tightly as possible in a plane. As we shall see, the three-dimensional crystal structures of a great many engineering materials can be visualized as different ways of stacking close-packed planes of atoms that have the hexagonal configuration shown in Fig. 3.7.

The fifth distinct two-dimensional lattice is the face-centered rectangular lattice that is formed when \( a = b \) and \( \gamma \neq 60^\circ, 90^\circ \) or \( 120^\circ \). The lattice is drawn in Fig. 3.8. Since \( a = b \) the primitive unit cell can be replaced by a rectangular cell with a lattice point at the center of its face, as outlined by the heavy lines in the figure. This cell is non-primitive (there are two lattice points per cell) but reveals the simple symmetry of the lattice more directly than does the primitive cell.

![Fig. 3.8: The face-centered rectangular lattice showing the lattice vectors, the primitive cell, and the face-centered rectangular cell.](image)

The restriction that \( \gamma \neq 60^\circ, 90^\circ \) or \( 120^\circ \) in a face-centered rectangular lattice comes from the fact that when \( \gamma = 60^\circ \) or \( 120^\circ \) the lattice is close-packed hexagonal, while when \( \gamma = 90^\circ \) the lattice is square.

Note that the unit cells of all the two-dimensional Bravais lattices have four or six sides. This is a consequence of the fact that the primitive cell of a two-dimensional lattice is always a parallelogram, a four-sided figure. When the edge lengths and angles are just right, as illustrated in Fig. 3.7, the parallelogram generates a lattice that has hexagonal symmetry. However, one cannot create a lattice of unit cells that have 3, 5, or more than 6 sides.

This important result can also be visualized by considering how the unit cells of the lattice fill the plane. Both mathematicians and masons use the term *tiling* to describe the operation of filling a plane with simple geometric figures. It is impossible to "tile" a plane with identical figures that have 3, 5 or more than 6 sides without leaving gaps. Gaps between unit cells are impermissible since the unit cell must locate all of the lattice positions when it is periodically repeated over the plane; it must identify where lattice
points are, and also where they are not. If gaps appear then atoms can be present in the gaps without disturbing the arrangement of atoms within the cells, and the cell does not uniquely determine the atom configuration.

![Diagram of atoms and basis vectors](image)

**Fig. 3.9:** An arrangement of atoms generated by repeating the triangular cell shown at right. The cell does not determine whether there are atoms in the shaded gaps in the pattern.

As an example, the impossibility of a triangular unit cell is illustrated in Fig. 3.9. A stacking of identical triangles leaves gaps in the plane, and a specification of the arrangement of particles within the triangular cell leaves undetermined the existence or configuration of particles in the gaps.

### 3.2.4 Basis vectors in Bravais lattices

In a two-dimensional crystal, an identical atom or group of atoms is located at each point of a two-dimensional primitive lattice that is, necessarily, one of the five two-dimensional Bravais lattices. As described in section 3.2.2, the position of each atom in the repeated group is specified by a vector, \( \tau_i \), that gives the vector displacement of that atom from the lattice point. These vectors are called *basis vectors*. Hence the structure of a two-dimensional crystal is described by giving the Bravais lattice, the set of basis vectors \( \{ \tau \} \) that are associated with each lattice point, and the identity of the atom that is located by each of the basis vectors. If the primitive cell of the Bravais lattice contains more than one atom, the structure is non-primitive, and one must use basis vectors to describe it.

Even monatomic solids can have structures that are inherently non-primitive. The crystal of diatomic molecules drawn in Fig. 3.3 is an obvious case. But there are also patterns of like atoms that appear very regular to the eye, yet still require basis vectors for their description. An example is the *simple hexagonal structure* shown in Fig. 3.10, which is the basic pattern of atom sites in the atom planes that are stacked to make the graphite phase of carbon. The atoms cannot be referred to the hexagonal cell shown in Fig. 3.7 since the lattice sites would not all have the same environment; some would be vacant. A hexagonal net can be drawn that connects the vacant sites at the centers of the hexagons shown in Fig. 3.10. The net is shown in dark lines in the figure. A possible
choice for the unit cell is shown at the right. The cell contains two atoms that are located by two basis vectors. The simple hexagonal structure is generated by repeating this non-primitive cell over the plane.

![Hexagonal arrangement of atoms](image)

...  
Fig. 3.10: A simple hexagonal arrangement of atoms showing the hexagonal unit cell and the basis vectors of the two atoms in the cell.

### 3.3 Crystals in Three Dimensions

#### 3.3.1 Three-dimensional crystal lattices

The essential features of a three-dimensional crystal are identical to those of the two-dimensional crystals we discussed in the previous section. The atoms in a crystal have a periodic distribution in space. An identical atom or group of atoms occupies every point of a three-dimensional primitive lattice. The lattice is generated by three, non-coplanar lattice vectors, \( \mathbf{a}_1, \mathbf{a}_2 \) and \( \mathbf{a}_3 \). The directions of the three lattice vectors give the directions in which the atom group is repeated. The magnitudes of the lattice vectors are the repeat periods. The lattice points of the primitive lattice are located at positions, \( \mathbf{r} \), that satisfy the relation

\[
\mathbf{r} = m\mathbf{a}_1 + n\mathbf{a}_2 + p\mathbf{a}_3
\]

where \( m, n \) and \( p \) are integers (positive, negative or zero). The lattice points fill three-dimensional space. The three lattice vectors have no necessary relation to one another beyond the requirement that they not lie in the same plane.

The primitive cell of the crystal lattice is a parallelepiped (a six-sided figure with parallel sides) whose eight corners are located at the terminal points of the vectors \( \mathbf{0}, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_1+\mathbf{a}_2, \mathbf{a}_1+\mathbf{a}_3, \mathbf{a}_2+\mathbf{a}_3, \) and \( \mathbf{a}_1+\mathbf{a}_2+\mathbf{a}_3 \). An example of a monatomic crystal is shown in Fig. 3.11 with two possible choices of its unit cell, one in which the atoms are placed on the lattice sites and one in which they are slightly displaced from the lattice sites. Both cells are primitive; since each point at the corner of the unit cell is shared
with eight adjacent cells, there is a net of one lattice site per cell, and, hence, one atom per cell.

When the repeated element is a group of atoms rather than a single atom each cell of the crystal lattice contains two or more atoms whose positions are specified by basis vectors from the origin of the cell. The positions of the lattice points in space are still given by equation 3.3. The positions, $\mathbf{r}_i$, of the $i^{th}$ atom in the repeated group are given by the equation

$$\mathbf{r}_i = m\mathbf{a}_1 + n\mathbf{a}_2 + p\mathbf{a}_3 + \mathbf{\tau}_i$$  \quad (3.4)$$

where $m, n$ and $p$ are integers, $\mathbf{\tau}_i$ is the basis vector that locates the $i^{th}$ atom in the group with respect to the lattice point specified by $m, n$ and $p$, and there is one equation of the form 3.4 for each atom in the repeated group. The atoms in the repeated group can also be visualized as filling the points of identical sublattices that are offset from one another by the vectors, $\mathbf{\tau}_i$.

Fig. 3.11: A monatomic three-dimensional crystal with two possible choices of the primitive cell.

### 3.3.2 The Bravais lattices

Since every three-dimensional lattice can be can be referred to a set of three non-coplanar lattice vectors, every three-dimensional lattice has a parallelepipedal primitive cell like that shown in Fig. 3.11. However, just as in the two-dimensional case, the magnitudes and directions of the lattice vectors may be related so that the primitive cell reduces to a simpler figure (such as a cube) or can be redrawn as a non-primitive cell with a relatively simple figure (such as a hexagonal prism). When the unit cell can be chosen to have a simple shape it is usually convenient to do so. Adopting the geometrically simplest cell makes it easier to visualize the atom arrangement and helps to interpret the properties of the crystal.
A detailed analysis of the geometry of the parallelepiped shows that there are thirteen special relations between the lattice vectors that simplify the geometry of the unit cell. Together with the general parallelepiped, these define the fourteen Bravais lattices to which the crystal structures of all materials are referred. The fourteen lattices are divided into seven crystal classes on the basis of the overall geometry of the geometrically simplest cell.

While we shall not prove that there are only fourteen Bravais lattices, it is relatively easy to see that there are at least fourteen, and it is not difficult to construct them. To do that, we note that any two of the three lattice vectors define a plane. The lattice points in the plane form a two-dimensional lattice which must be one of the five two-dimensional Bravais lattices discussed in the previous section. The third lattice vector then translates the plane periodically through space. From this perspective a three-dimensional lattice is simply a periodic stacking of planes that contain identical two-dimensional lattices.

Any two of the three lattice vectors can be taken to define the planar lattice that is repeated in space. Let us choose those two that give the simplest two-dimensional lattice, and use these for the bottom or basal plane of the three-dimensional cell.

![Fig. 3.12: The triclinic primitive cell: a ≠ b ≠ c; Ĉ, β, γ ≠ 90°.](image)

The first crystal class is called triclinic. If the simplest two-dimensional lattice is a general parallelogram, and if it is repeated by translating it by an arbitrary vector, then the simplest unit cell is a parallelepiped with no special symmetry. The cell is called a triclinic cell, and is shown in Fig. 3.12. A set of primitive lattice vectors is drawn in the cell. In keeping with usual notation, the lengths of the three edges of the cell are denoted a, b and c, and the opposite included angles are denoted Ĉ, β and γ (that is, Ĉ is the angle between edges b and c, etc.). In the triclinic class, there is no necessary relation between the edge lengths of the unit cell; in general, a ≠ b ≠ c. The angles, Ĉ, β and γ differ from 90°. There is only one geometrically distinguishable cell in this class: the triclinic primitive cell.

The second crystal class is the monoclinic. The basal (high-symmetry) plane of a monoclinic crystal contains a rectangular, square, hexagonal or face-centered rectangular lattice. The basal plane is repeated in the perpendicular direction by a vector that points
in an arbitrary direction with respect to the plane. The two distinct monoclinic cells are shown in Fig. 3.13, with a primitive set of lattice vectors indicated. If the cell has a rectangular lattice in its basal plane then the simplest monoclinic cell is the *primitive monoclinic cell*. If the basal lattice is square, hexagonal or face-centered rectangular then the basal plane can be drawn as a face-centered rectangle and the simplest cell is a *base-centered monoclinic cell*. This cell is non-primitive; there are two atoms lattice points per cell. The distinguishing feature of the monoclinic crystal is the angle, $\gamma = 90^\circ$. The edge lengths, $a$, $b$ and $c$, and the angles, $\hat{C}$ and $\beta$, have no necessary relation, so long as they do not produce a figure of higher symmetry. In general, $a \neq b \neq c$, and $\hat{C}$ and $\beta \neq 90^\circ$.

![Diagram](a) ![Diagram](b)

**Fig. 3.13:** The monoclinic crystal class: $a \neq b \neq c$; $\gamma = 90^\circ$, $\hat{C}$, $\beta \neq 90^\circ$. (a) The primitive monoclinic cell. (b) The base-centered monoclinic cell.

![Diagram](a) ![Diagram](b) ![Diagram](c) ![Diagram](d)

**Fig. 3.14:** The orthorhombic system: $a \neq b \neq c$, $\hat{C} = \beta = \gamma = 90^\circ$. (a) Primitive cell; (b) base-centered cell; (c) body-centered cell; (d) face-centered cell. Primitive lattice vectors are drawn and the primitive cell of the face-centered lattice is indicated.

The third distinguishable crystal class is the *orthorhombic*, which is the class to which the monoclinic reduces when the symmetric basal plane is repeated in the perpendicular direction so that $\hat{C} = \beta = \gamma = 90^\circ$. There are four distinct unit cells of the orthorhombic class. They are illustrated in Fig. 3.14. The first two are the primitive and base-centered orthorhombic cells, which are analogs of the monoclinic cells shown in Fig.
3.13. The third is the *body-centered orthorhombic* cell that is formed when the basal lattice is rectangular and the third lattice vector is one-half the body diagonal of the orthorhombic cell so that alternate basal lattices lie directly above one another. The fourth is the *face-centered orthorhombic* cell that is formed when the basal lattice is a face-centered rectangle and the third lattice vector carries the face-centered point into the face of the orthorhombic cell so that alternate basal lattices lie directly above one another. The body-centered lattice contains two lattice points per cell; the face-centered lattice contains four.

![Body-Centered Orthorhombic Cell](image)

**Fig. 3.15:** The tetragonal crystal class: \( a = b \neq c; \hat{C} = \beta = \gamma = 90^\circ \). (a) The primitive tetragonal cell; (b) the body-centered tetragonal cell.

The fourth distinguishable crystal class is the *tetragonal*, which is the special case of the orthorhombic when \( a = b \) (the basal lattice is square). There are only two distinguishable members of the tetragonal class: the *primitive tetragonal* and the *body-centered tetragonal* cells, which are special cases of the primitive and body-centered orthorhombic cells. While base-centered and face-centered tetragonal lattices can be drawn, the former can be made into a primitive tetragonal lattice and the latter into a body-centered tetragonal lattice by simply redefining the edges of the cell. The two tetragonal lattices are shown in Fig. 3.15.

![Tetragonal Cell](image)

**Fig. 3.16:** The cubic system: \( a = b = c; \hat{C} = \beta = \gamma = 90^\circ \). (a) Simple (primitive) cubic cell; (b) body-centered cubic (BCC); (c) face-centered cubic (FCC).
The fifth crystal class is the *cubic*, which is the special case of the orthorhombic when \( a = b = c \) so that the cell is a cube. There are three distinguishable cubic cells: simple cubic, body-centered cubic and face-centered cubic. They are special examples of the corresponding orthorhombic cells. However, there is no base-centered cubic cell. The base-centered orthorhombic cell reduces to a tetragonal primitive cell when \( a = b = c \); since the lattice configuration in the base is different from that in the sides it does not have cubic symmetry. The three unit cells of the cubic system are drawn in Fig. 3.16. As in the orthorhombic case, the *body-centered cubic* cell has two lattice points per cell and the *face-centered cubic* cell has four.

![Diagram of cubic crystal structure](image)

**Fig. 3.17:** The rhombohedral class: \( a = b = c; \hat{C} = \beta = \gamma \neq 60^\circ \) or \( 90^\circ \). The primitive rhombohedral cell is shown, along with its derivation through a diagonal stretch of the FCC cell.

The sixth distinguishable crystal class is the *rhombohedral*, which can be derived from the face-centered lattice by the simple procedure that is indicated in Fig. 3.17. Let the lattice vectors shown in Fig. 3.16c be used to define the primitive cell of the FCC structure. The geometry of the primitive cell is like that shown in Fig. 3.17. Now let the FCC cell be stretched or contracted along the cube diagonal so that the cubic symmetry is lost. Since the three lattice vectors are symmetrically disposed about the cube diagonal, they remain equal in length after the stretch. The angles between them change, but also remain equal. The result is the *primitive rhombohedral cell*, the only member of the rhombohedral class, which has \( a = b = c \) and \( \hat{C} = \beta = \gamma \neq 60^\circ \) or \( 90^\circ \) (the former generates the FCC cell, the latter the simple cubic cell).

The seventh and final crystal class is the *hexagonal*. This crystal class also contains only one lattice, which is obtained when the basal lattice is hexagonal and the third lattice vector is perpendicular to the basal plane. Hence \( a = b, \) \( c \) is arbitrary, \( \hat{C} = \beta = 90^\circ, \gamma = 120^\circ \). Both the primitive cell and the larger, hexagonal cell of the hexagonal class are shown in Fig. 3.18. The full hexagonal cell is usually used to describe materials with hexagonal structure since it clearly shows the symmetry of the pattern. The non-primitive hexagonal cell contains a net of three lattice sites. The hexagonal cell suffers from the disadvantage that it cannot be generated by three vectors which makes its
geometry difficult to treat mathematically. It is usually better to refer to the primitive cell for the mathematical description of a hexagonal crystal.

![Hexagonal cell diagram](image)

**Fig. 3.18:** The hexagonal cell: $a = b$, $c$ arbitrary, $\beta = 90^\circ$, $\gamma = 120^\circ$. The primitive cell is outlined in dark and embedded in the larger regular hexagonal cell.

### 3.3.3 Distorted cubic cells

As we shall see, the structures of the important engineering solids tend to be cubic, largely because deviations from cubic symmetry introduce anisotropy in the engineering properties that make them less useful for many applications. When engineering materials are found in structures that do not have cubic symmetry this is often because some chemical or physical perturbation has disturbed a basically cubic atom arrangement. It is, therefore, often useful to visualize the non-cubic structures as simple distortions of the cubic ones. If one begins with a cubic cell and stretches it along a cube diagonal the result is a rhombohedral cell, as illustrated in Fig. 3.17. Stretching along a cube edge creates a tetragonal cell of the type illustrated in Fig. 3.15. If the cubic cell is stretched by different amounts along two of its edges it becomes orthorhombic, of the type illustrated in Fig. 3.14.

### 3.4 POINTS, DIRECTIONS AND PLANES IN A CRYSTAL

To describe atom arrangements in crystals it is useful to have notations that designate points, directions, and planes within the unit cell and its crystal environment. The most convenient system employs crystal axes, $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_3$ that lie along the three edges of the unit cell of the Bravais lattice of the crystal. When the Bravais lattice is primitive, the crystal axes are the same as the lattice vectors, $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$, but when the cell is non-primitive they are different. For example, when a material has a cubic crystal structure we ordinarily choose the crystal axes to be equal, orthogonal vectors along the edges of the cubic unit cell, whether or not the cell is primitive. When the structure is orthorhombic the crystal axes are orthogonal, but unequal in length. When the structure is hexagonal the crystal axes are the same as the lattice vectors (Fig. 3.18): the axes $\mathbf{e}_1$ and $\mathbf{e}_2$ are equal in length, but make an angle of $120^\circ$, the vector $\mathbf{e}_3$ may be any length, but is...
orthogonal to the other two. (Many texts use a four-axis system to describe the hexagonal crystal. The fourth axis, actually, a third axis in the basal plane of the hexagon, is redundant and serves no useful purpose. We shall not discuss the "hexagonal notation" in this course.)

### 3.4.1 Points in a crystal

A particular point in a crystal is uniquely identified by its coordinates \((x,y,z)\) in the coordinate system that is defined by the origin and the crystal axes. The origin is connected to the point \((x,y,z)\) by the vector \(\mathbf{r} = xe_1 + ye_2 + ze_3\). Hence the point \((0,0,1)\) lies at the end of the vector \(e_3\); in a cubic structure this is the corner of the unit cell that is vertically up from the origin. The point \((\frac{1}{2},\frac{1}{2},\frac{1}{2})\) is located in the center of the unit cell, and the point \((\frac{1}{2},\frac{1}{2},0)\) lies in the center of the basal face of the unit cell, as shown for the cubic case in Fig. 3.19.

![Cubic unit cell](image)

**Fig. 3.19:** A cubic unit cell showing the crystal axes. Three positions within the cell are indicated and their coordinates given.

Points that are equivalent by the symmetry of the unit cell have a similar geometric environment and define positions at which atoms would ordinarily have physically similar behavior. We shall hence occasionally use the nomenclature "points of the type \((x,y,z)\)", to refer to the point \((x,y,z)\) and all other points that are similarly situated within the unit cell. In a cubic crystal the phrase "points of the type \((1,0,0)\)" means points that are located at cell corners, while "points of the type \((\frac{1}{2},\frac{1}{2},0)\)" means points that are located at the centers of cube faces. We shall use the symbol \(\{x,y,z\}\) to indicate points of a given type.

### 3.4.2 Directions in a crystal

It is often necessary to designate directions in a crystal in order to describe its physical behavior. For example, the force required to stretch a macroscopic crystal by a certain amount depends on its direction with respect to the unit cell. We specify a direction in the crystal by the simplest vector that takes that direction from the origin. The vector
\[ \mathbf{r} = h\mathbf{e}_1 + k\mathbf{e}_2 + l\mathbf{e}_3 \]

connects the origin to the point \((h,k,l)\) and has a direction that is uniquely specified by the components \((h,k,l)\). To simplify the notation the direction is given by simply enclosing the components in brackets: \([hkl]\). The crystal axes are understood. Hence the \([001]\) direction in a cubic crystal points along the vertical edge of the cube, the \([111]\) direction points along the diagonal of the cube, and the \([110]\) direction lies along the diagonal of the basal plane of the cube, as shown in Fig. 3.20.

It does not matter that the direction vectors \([hkl]\) shown in Fig. 3.20 have different lengths; the symbol \([hkl]\) is only intended to specify a direction, and the direction of a vector is independent of its length. It is, therefore, customary to choose \(h,k\) and \(l\) to be the smallest integers that designate a vector in the required direction. To draw a direction \([hkl]\) in the unit cell we simply note that any vectors whose indices \([hkl]\) have given ratios lie in the same direction. If we divide through by the largest of the three numbers, the resulting indices are the coordinates of the point at which a vector in the direction \([hkl]\) that begins from the origin exits the unit cell, and we can draw the direction by simply connecting the origin to that point. For example, the direction \([123]\) is also the direction of the vector \([\frac{12}{33}, 1]\), which connects the origin to the point \((\frac{1}{3}, \frac{2}{3}, 1)\) in the unit cell. Similarly, if the direction is given as a vector in the unit cell that connects the origin to a given point, \((x,y,z)\), where at least one of \(x, y, z\) is a fraction, one can find the integers \([hkl]\) that specify the direction by simply expressing \(x, y\) and \(z\) as fractions with a common denominator and clearing the fractions by multiplying through by the common denominator. For example, the direction that connects the origin to the point \((\frac{1}{2}, \frac{1}{2}, 1)\) in the unit cell is the \([125]\) direction.

![Fig. 3.20: A unit cube showing three crystallographic directions.](image)

Since a vector may have negative components, one or more of the indices \([hkl]\) may be negative. In this case we use a compact notation that places the negative sign above the index. For example, the direction \([100]\) points along the \(\mathbf{e}_1\) axis in the negative direction.

Because of the symmetry of the crystal it is often physically meaningless to distinguish between different directions that have a similar geometric identity. For example, a cubic crystal responds in the same way when it is stretched in a direction that parallels an edge of its unit cell, whichever edge is stretched. The directions that have the same geometric identity in the unit cell are those that can be converted into one another by
performing geometric operations that turn the unit cell into a replica of itself. These directions are said to be "related by symmetry".

The symbol \(<\text{hkl}\>\) is used to designate the family of directions that includes [hkl], that is, all directions that are related by symmetry to the direction [hkl]. In a cubic crystal that has a primitive cell all directions that are obtained by permuting the order of the numbers [hkl], such as [khl], and all that are obtained by taking the negative of any number, such as [fikl], are related by symmetry. The symbol \(<\text{hkl}\>\) is used to designate this whole group of related directions. Hence the symbol \(<\text{100}\>\) designates all of the directions that lie along cube edges; the symbol \(<\text{110}\>\) designates all of the directions that are diagonal to the faces of the cube. In less symmetric structures the family \(<\text{hkl}\>\) is more restricted.

### 3.4.3 Planes in a crystal

It is finally necessary to have a notation that identifies planes in a crystal. A plane is designated by the symbol (hkl), where \(h\), \(k\) and \(l\) are integers that are called the Miller indices of the plane. The Miller indices of a crystal plane are found by following a recipe that is illustrated in Fig. 3.21. Let a plane cut the three coordinate axes at the points \(a\), \(b\) and \(c\). Take the reciprocals of these numbers, \(\frac{1}{a} \), \(\frac{1}{b} \), \(\frac{1}{c} \), and convert these into a set of three integers (assuming this is possible) by finding a common denominator and multiplying through by that denominator. The resulting integers, \(h\), \(k\) and \(l\), are the Miller indices of the plane, and are enclosed in parentheses, (hkl). For example, the cube face plane shown in Fig. 3.21 cuts the three coordinate axes at 1, \(\infty\), \(\infty\), and is, therefore, the (100) plane. The second plane shown in the figure cuts the axes at \(\frac{2}{3} \), \(\frac{1}{2} \), 1, and is, therefore, the (342) plane.

If the Miller indices differ in sign only one of them need be taken negative since the indices can be multiplied by (-1) in clearing the fraction. It is then customary to place the minus sign above the index, as in the case of the directional indices. Hence the (342) plane intercepts the axes at \(-\frac{2}{3} \), \(\frac{1}{2} \), 1. It is always possible to find integer values for the Miller indices when the intercepts \(a\), \(b\), \(c\) are rational numbers. If any one of them is not a rational number, the plane is said to be an irrational plane.

![Diagram of cubic crystal with planes](image)

Fig. 3.21: Cubic crystal with the (100) and (342) planes shaded.
Given the recipe I have provided, the Miller indices may seem a rather arbitrary way to identify a crystal plane. However, they have three simple properties that make them particularly useful.

First, the plane (hkl) is perpendicular to the direction [hkl]. While I shall not prove this result, it is obvious in the case of the (100) plane, which designates a cube face perpendicular to the [100] direction. This is a particularly useful property. In mathematics, it is customary to identify a plane by specifying a normal vector, \( \mathbf{n} \); the plane contains all points that can be reached from the base of the normal vector by moving perpendicular to it. The Miller indices of a plane immediately give us the direction of its normal vector.

Second, the Miller indices of all planes that are parallel and equidistant are the same. For example, the plane that intersects the axes at the points 2, 0, 0 is a face plane that is twice as far from the origin as the (100) plane drawn in Fig. 3.21. It also has the Miller indices (100). The plane that is parallel to the plane (342) drawn in Fig. 3.21, but twice as far from the origin has intercepts \( \frac{4}{3} \), 1, 2. Its Miller indices are also (342). It follows that the Miller indices actually designate an infinite set of parallel planes that are equidistant from one another. These planes are equivalent crystallographically since they differ only through the choice of the location of the origin, which is arbitrary. It is, therefore, useful that they have the same Miller indices.

Third, when the structure is cubic, which is the most common case in engineering, the normal distance between parallel planes that have the same Miller indices is given in terms of the Miller indices by the simple formula

\[
 d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}}
\]

where \( a \) is the length of the side of the unit cell. The proof of this relation will not be given here, but is contained in standard crystallographic texts. A somewhat similar, but more complicated relation holds for non-cubic structures.

![Fig. 3.22: Drawing showing examples of \{100\}, \{110\} and \{111\} planes in a cube.](image)
As in the case of points and directions in a crystal, planes with different indices may be physically equivalent by the symmetry of the unit cell. The set of symmetrically equivalent planes is called a family of planes. The family of planes that includes \( \{hkl\} \) is denoted by the symbol \( \{hkl\} \).

The crystal structures we shall most often consider are cubic. In a cubic structure it may be important that a particular plane is a cube face rather than a cube diagonal, but is usually unimportant which particular face it is; the faces can be interchanged by a simple rotation of the cube. More generally, the symmetry of the cube relates all planes whose Miller indices are permutations or negatives of one another. In a crystal with full cubic symmetry the symbol \( \{hkl\} \) represents the family of planes that is obtained from \( \{hkl\} \) by permuting the indices or replacing any one of them by its negative. For example, the \( \{100\} \) planes are the cube faces. The crystal planes we shall refer to most often are the \( \{100\} \) planes, which are cube faces, the \( \{110\} \) planes, which are perpendicular to the face diagonals, and the \( \{111\} \) planes, which are perpendicular to the body diagonals. Examples of these three planes are drawn in Fig. 3.22.

### 3.5 THE COMMON CRYSTAL STRUCTURES

In the following sections I shall describe the crystal structures that are most common in engineering materials. For this purpose it is useful to divide materials into four classes: elemental solids, whose atom sites are filled by atoms of the same type, solid solutions, whose atom sites are filled by more than one kind of atom in an approximately random way, ordered compounds, whose atom sites are filled by atoms of two or more types in a regular way, and molecular solids, whose atom sites are occupied by molecules that retain their identity in the solid.

While an infinite number of structures are possible, a great many of the solids of engineering interest have one of a few simple crystal structures, or have structures that are straightforward derivatives of these. The common crystal structures of the elemental solids and solid solutions are the face-centered cubic (FCC), hexagonal close-packed (HCP), body-centered cubic (BCC), and diamond cubic (DC). The common crystal structures of ordered compounds are based on these. They include substitutional compounds that are obtained by substituting atoms on the sites of the elemental structures in a regular way, and interstitial compounds that are obtained by placing atoms in the empty sites between atoms of the elemental structures. The prototype substitutional structures that we shall describe include the Cu₃Au and the CsCl structures. The prototype interstitial structures include the NaCl, the \( \cdot \) and \( \beta \)-ZnS, and the Nb₃Sn (A15) structures. A substantial fraction of all the engineering structures of interest can be related to these prototypes.

For the moment I shall simply describe the crystal structures. It is also important to have some feeling for why materials should take on the structures that they do. This is a difficult problem that remains only partly solved. Many structures can, however, be un-
derstood in terms of the simple bonding models that were described in Chapter 2, and I shall discuss them from this perspective toward the end of this chapter.

3.6 THE CLOSE-PACKED STRUCTURES: FCC AND HCP

3.6.1 The FCC and HCP structures

Almost all of the elemental solids and alloys that are important in engineering have one of four crystal structures: face-centered cubic (FCC), hexagonal close-packed (HCP), body-centered cubic (BCC) and diamond cubic (DC). It is useful to begin with the FCC and HCP structures since, despite the obvious difference between their unit cells, they are very closely related to one another. They are the two close-packed structures that represent the two different ways in which identical, spherical atoms can be packed together as tightly as possible.

![FCC and HCP structures](image)

Fig. 3.23: The unit cells of the face-centered cubic (FCC) and hexagonal close-packed (HCP) crystal structures. The Bravais cell in HCP is indicated by dark lines.

The unit cells of the face-centered cubic (FCC) and the hexagonal close-packed (HCP) structures are drawn in Fig. 3.23. The FCC lattice is one of the cubic Bravais lattices diagrammed in Fig. 3.16. In the FCC structure like atoms are placed on each of the FCC lattice sites. Since each of the corner atoms is shared by a total of eight cells in the FCC crystal and each of the face atoms is shared by two cells there are 4 atoms per unit cell.

The unit cell of the HCP structure is a hexagonal prism with atoms at its corners and atoms at the centers of its upper and lower faces. There are also three atoms in a triangular configuration in a plane at half the height of the unit cell; each of these atoms is located directly above the center of an equilateral triangle of atoms in the basal plane, as shown in the figure. Since each of the corner atoms in the hexagon is shared by six adjacent cells and each of the face atoms is shared by two cells, the total number of atoms per unit cell in the hexagonal close-packed structure is 6.
The HCP lattice is inherently non-primitive. The unit cell of the hexagonal Bravais lattice is indicated by the heavy lines in Fig. 3.23. There are two atoms per lattice point of the Bravais lattice. The basis vector of the HCP lattice is indicated by the dark arrow in Fig. 3.23.

### 3.6.2 Close-packing in FCC and HCP

While the FCC and HCP cells look quite different as they are drawn in Fig. 3.23, the two structures have an important physical similarity. The FCC and HCP structures are the two alternate ways of packing spheres of identical size as densely as possible. For this reason they are called the close-packed structures.

To see this, recognize that if spheres of the same size are packed in a planar array as tightly as possible they adopt the hexagonal pattern shown in Fig. 3.24. The pattern is drawn twice in Fig. 3.24. In the first drawing we show only the spheres, but in the second the centers of the spheres are labeled with the symbol A, and the interstices between the spheres are labeled alternately with the symbols B and C.

![Hexagonal packing](image)

Fig. 3.24: The hexagonal configuration of spheres packed as tightly as possible into a plane. The spheres and interstices are labeled in the re-drawn figure (b).

Now let a second layer of spheres be placed on top of the first, or A-layer. These spheres will naturally sit on the interstices of the A-layer, that is, on the sites labeled B or C in Fig. 3.24b. But the B and C-sites cannot both be filled in the second layer of a close-packed elemental solid. As shown in Fig. 3.25, if a sphere occupies a site of type B then its size prevents spheres from appearing on adjacent sites of type C. Hence only the B or C-sites are filled in the second layer. Assume the B-sites are filled, as in Fig. 3.25. Since the choice of which sites are B-type and which C-type is completely arbitrary, the two methods of stacking identical spheres in two close-packed layers, AB and AC, are indistinguishable.
Fig. 3.25: Segment of the configuration of two layers of close-packed spheres in which the second layer lies on the B-sites of the first.

When the third layer of spheres is added, however, there are two independent choices that lead to two distinct structures. Each of the spheres in the third layer sits on a tripod formed by three adjacent spheres in B-sites in the second layer. However, the interstices in the B-layer are, alternately, A and C-sites, just as B and C-sites alternated in the A-layer. The third layer of spheres may be placed either on A-sites or on C-sites, but not on both. Hence the three-layer stacking is either ABA, as shown in Fig. 3.26a, or ABC, as shown in Fig. 3.26b.

Fig. 3.26: The two possible types of three-layer stacking of equal-sized spheres: (a) ABA (the A-sites in the third plane obscure those in the first plane); (b) ABC.

A comparison of Fig. 3.26a and the HCP unit cell drawn in Fig. 3.23 shows that the ABA packing gives precisely the HCP unit cell. Since the unit cell is repeated indefinitely in the crystal, the hexagonal close-packed structure is a stacking of close-packed planes in the sequence ...ABABAB... .
3.6.3 **Interstitial sites in the FCC and HCP structures**

In addition to the atom positions in a crystal structure the *interstitial sites*, or *voids* in the structure, are also important. These interstitial sites are empty spaces in the structure that additional atoms may fill. The interstitial voids are likely positions for small impurity atoms, which are therefore called *interstitial impurities*. Moreover, the interstitial sites may be filled by sets of atoms in an ordered way, giving rise to *interstitial compounds* based on the simple elemental lattices. As we shall see, the diamond cubic structure and several of the common crystal structures of binary compounds can be regarded as interstitial structures based on FCC.

There are two voids in the FCC structure that have significant size; these are the interstitial sites that are important in interstitial solutions and compounds based on FCC. The largest of these is the *octahedral void*, whose center is equidistant from six neighboring atoms of the FCC structure. The octahedral void that is easiest to visualize is located in the center of the FCC cell, and is drawn in Fig. 3.28a. The center of the cell is equidistant from the atoms in the centers of the six faces. These six atoms are the corners of an octahedron, which is drawn in the figure. Considering the atoms in adjacent cells shows that there are also octahedral voids at the center of each edge of the FCC cell.

**Fig. 3.27**: View of the FCC cell along its diagonal showing the cube axes. The opaque atom is at the upper corner. Darker atoms lie in successively lower (111) planes. The atom at the far corner is obscured.
Since the voids on the edge are shared equally by four cells, the total number of octahedral voids per cell in the FCC structure is 4, or one per atom.

![Diagram](image1)

**Fig. 3.28:** (a) The octahedral void (●) in the center of the FCC cell, showing the octahedron formed by its six neighbor atoms. (b) The location of octahedral voids (●) in the FCC unit cell.

The second important set of interstitial voids in FCC is the set of *tetrahedral voids*. The center of a tetrahedral void is equidistant from four atoms that form a tetrahedron. An example is shown in Fig. 3.29a. The four atoms that coordinate the tetrahedral void in Fig. 3.29a are sited on four of the eight corners of a small cube which is an octet of the unit cell. This cube is drawn in the figure. The center of the tetrahedral void is at the center of the cube, equidistant from its corners, and hence occupies the position \( \left( \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) \). By symmetry, there are tetrahedral voids similarly situated with respect to each of the corners of the unit cube as shown in Fig. 3.29b. There are, hence, eight tetrahedral voids per FCC unit cell, or two per atom.

![Diagram](image2)

**Fig. 3.29:** (a) The center of the tetrahedral void (●) at \( \left( \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \right) \), in the center of an octet of the unit cube. (b) The eight tetrahedral voids in the FCC unit cell.

It is also useful to locate the tetrahedral and octahedral voids in the ABC stacking of close-packed planes that generates the FCC structure (Fig. 3.26-27). The tetrahedral voids are sites that are equidistant from four atoms. An atom in an A-layer neighbors three B-atoms in a triangular configuration in the B-plane immediately above, and three
C-atoms in the plane immediately below. Hence the spaces immediately above and immediately below it are tetrahedral voids. The centers of these are displaced from the centers of the A-atoms by the vector $\frac{1}{4}[111]$, and are, hence, distant by $3/4$ of the interplanar spacing. Since similar reasoning holds for atoms in B and C-layers, it follows that every atom in the stacking of close-packed planes has tetrahedral voids just above and just beneath it for a total of two per atom.

The octahedral voids are the sites that are equidistant from six atoms. If the neighboring planes in the stacking are B and C-planes, as in Fig. 3.26b, then the site that lies half-way between the empty A-sites in the two planes is equidistant from three B-sites and three C-sites, and is, therefore, the center of an octahedral void. Hence the octahedral voids are located immediately above the interstices in a plane that are not used in the succeeding plane, and are centered at $1/2$ of the interplanar spacing. Since the number of these is equal to the number of sites in the plane, there is one octahedral void per atom.

I shall leave it as an exercise to compute the size of the octahedral and tetrahedral voids in the FCC structure. The geometry of a tetrahedron is difficult to treat directly, but becomes simple when the tetrahedron is inscribed in a cube, as it is in the construction in Fig. 3.29a. The octahedral void is somewhat larger than the tetrahedral.

The positions and sizes of the octahedral and tetrahedral voids in the HCP structure can be inferred from the fact that it is a stacking of close-packed planes in an ABABA pattern. As we have already discussed for the case of the FCC structure, tetrahedral voids are located immediately above and immediately below each atom, with their centers displaced by $3/4$ of the interplanar spacing. Octahedral voids are located along the line of unused C-sites, with their centers midway between the close-packed planes. Since the geometry of a void is determined by the adjacent atom planes the sizes of the voids in HCP are the same as in FCC.

### 3.6.4 Properties of materials with the FCC and HCP structures

With the exception of the inert gases, elements that crystallize in the FCC and HCP structures are metals. This property follows from the nature of bonding: since each atom has twelve neighbors it is not possible to form saturated bonds. The FCC and HCP metals also tend to be malleable (easily deformed). This property is suggested by the fact that atoms in the close-packed structures bond as if they were spheres, which suggests in turn they their energy is not strongly changed by the local distortions of the lattice that occur during deformation.

However, because hexagonal close-packing produces a unique direction in the crystal lattice (the vertical direction in the unit cell) materials that crystallize in the HCP structure are anisotropic in many of their macroscopic properties, such as electrical conductivity, and show an especially strong anisotropy in their mechanical properties. Largely for this reason materials with hexagonal structures are much less commonly used in engineering than those with the more isotropic cubic structures.
3.7 THE BODY-CENTERED CUBIC STRUCTURE

3.7.1 The BCC unit cell

The unit cell of the body-centered cubic structure (BCC) is drawn in Fig. 3.30. The unit cell is a cube with atoms at each of its eight corners and an additional atom in its center. There are, therefore, two atoms per unit cell in the BCC structure.

![Diagram of BCC unit cell]

Fig. 3.30: The unit cell of the BCC crystal structure.

As in the case of FCC, the BCC structure is obtained by placing atoms on the points of a Bravais lattice, the BCC lattice shown in Fig. 3.16b. It follows that one can draw a primitive unit cell for the BCC structure. The most convenient choice defines the primitive cell by the three vectors \( \mathbf{a}_1 = \frac{1}{2} [111] \), \( \mathbf{a}_2 = \frac{1}{2} [1\bar{1}1] \), \( \mathbf{a}_3 = \frac{1}{2} [11\bar{1}] \), which are equal in magnitude. (Note that this choice of the primitive cell differs from that shown in Fig. 3.16b.)

The BCC structure is not quite close-packed. Given a BCC structure of spherical atoms that pack as tightly as possible, they touch along the body diagonals, the \(<111>\) directions. The planes of closest packing are the \{110\} planes. Each atom has 8 nearest neighbors (as opposed to 12 in the FCC and HCP structures), but has six additional neighbors that are only slightly further away (the six second-near neighbors of the atom at the body center are the body-centered atoms in each of the six adjacent cells).

All of the elements that crystallize in the BCC structure are metallic conductors; the eight-fold coordination in the nearest neighbor shell does not permit saturated bonds. Two very different classes of metals crystallize in the BCC structure. The first class includes simple metals such as Na, K, Rb and Ca. These elements have simple metallic bonding, but the precise form of the bonding function is such that the slightly more open BCC structure with its larger total number of near neighbors in the first two coordination shells is preferred, particularly at high temperature. The second class includes transition metals such as Ti, Mo, W and Fe. In these materials there is at least some contribution to the bonding from the d-electrons that occupy unfilled orbitals in the ion core. The symmetry of d-bonding promotes the BCC structure.

Page 58
3.7.2 Interstitial voids in the BCC structure

The interstitial voids in the BCC structure are important both as sites for solute atoms and as atom positions in interstitial compounds based on BCC. The large voids in the BCC structure are classified as octahedral and tetrahedral as in the close-packed structures. However, in the BCC structure the octahedra and tetrahedra are distorted. As we shall see at a later point, the imperfect geometry of the interstitial voids has important consequences for the properties of BCC materials that contain solutes in interstitial sites.

Fig. 3.31: A drawing of "octahedral" voids (●) at face-center and edge-center positions in the BCC unit cell.

There is an interstitial void of the octahedral type in the BCC structure at the center of each face and at the center of each edge. The atomic coordination around these voids is shown in Fig. 3.31. It should be obvious from the figure that the void is asymmetric. The octahedral void at the face center is defined by two body-centered atoms at a distance \( a/2 \), where \( a \) is the edge length of the unit cell, and also by four atoms at the corners of the unit cell at the greater distance \( a\sqrt{2} \). The void hence has a short axis along the \( \mathbf{a}_1 \), or x-direction. The void shown in Fig. 3.31 at the edge center has the same geometry, but its short axis lies in the \( \mathbf{a}_3 \), or z-direction.

Since there is an octahedral void at the center of each face and each edge in the BCC unit cell there are 6 per unit cell, or 3 per atom. It is sometimes convenient to designate these by the axis of the short dimension. The symbol \( O_z \) indicates an octahedral void whose center is closest to the atoms on the \( \mathbf{e}_3 \), or z-axis; \( O_x \) and \( O_y \) are defined in a similar way. There are 2 voids of each type in the cell. Hence there is an \( O_x, O_y \) and \( O_z \) void associated with each atom in the BCC structure.

The geometry of a tetrahedral void in the BCC structure is shown in Fig. 3.32. Each face of the BCC unit cell contains the centers of four tetrahedral voids. They are located at positions of the type \( \{0 \frac{1}{2} \frac{1}{4} \} \), which are displaced by the distance \( \frac{a}{4} \) from the center of the face toward one of the four edges that bound it. There are 12 tetrahedral voids per unit cell, or 6 per atom. While the center of the tetrahedral void is equidistant from the four atoms that define it, these atoms are not equidistant from one another. Hence the figure is not a true tetrahedron.
3.8 THE DIAMOND CUBIC STRUCTURE

3.8.1 The diamond cubic cell

The diamond cubic structure is an example of an ordered interstitial structure based on FCC. Its unit cell is drawn in Fig. 3.33. The unit cell contains atoms on the FCC positions, and also has atoms in four of the eight tetrahedral voids.

Fig. 3.32: A "tetrahedral" void in the BCC structure.

Fig. 3.33: The unit cell of the diamond cubic structure. For clarification the atoms in the tetrahedral voids are white while the FCC atoms are shaded.

Fig. 3.34: The DC unit cell redrawn to show the successive close-packed (111) planes and the $\frac{1}{4}[111]$ vectors connecting atoms in these planes to the atoms that fill tetrahedral voids.
The atom positions in the diamond cubic structure do not form a Bravais lattice, but a lattice with a basis. The underlying Bravais lattice is FCC. Atoms are located at each point of an FCC lattice, and also at every point that is displaced from an FCC lattice point by the basis vector, \( \mathbf{\tau} = \frac{1}{4} [111] \), which is the vector that connects an FCC position to the center of the closest tetrahedral void in the [111] direction. Fig. 3.33 is re-drawn in Fig. 3.34 to show this connection.

### 3.8.2 The DC structure as a stacking of planes and interstitials

The pattern in which the tetrahedral voids are occupied in the DC structure is such that each atom, whether in a void or on the FCC lattice, has exactly four equidistant neighbors in a tetrahedral coordination around it. It is not easy to see this by examining the DC unit cell. A better understanding can be achieved by recalling how the tetrahedral voids are configured when the FCC structure is analyzed as a stacking of close-packed planes.

There is a tetrahedral void immediately above and immediately below each atom in the (..ABC..) FCC stacking of close-packed planes. The diamond cubic structure is obtained from the FCC by placing atoms in all of the upper or all of the lower voids, but not both. If we use a small letter (a) to indicate a filled tetrahedral void immediately above or below an FCC site (A) then the stacking in the diamond cubic structure is either ...AaBbCc..., or, equivalently, ...aAbBcC... . Making the first choice, an atom in a B site, for example, has four equidistant nearest neighbors: the atom in the b-site immediately above it, and the three atoms in the a-sites just below it. Similarly, a b-atom has four nearest neighbors: the atom in the B-site immediately below it, and the three atoms in the C-sites just above it. Since the lattice and interstitial atoms in the DC structure are the same, it does not matter which we call the lattice atoms and which the interstitials. Interchanging the designation of lattice and interstitial atoms simply replaces the ...AaBbCc... pattern by the ...aAbBcC... pattern, which is equivalent.

While the atoms in the A planes (and those in the a-planes) of the diamond cubic structure have a hexagonal close-packed configuration, they cannot be visualized as close-packed planes of spheres. If the atoms are modeled as spheres they are held apart by the spheres of like size that sit in the tetrahedral interstitial sites. We leave it as an exercise to find the volume per atom when spheres are packed in a diamond cubic structure. The result will show that the DC structure is very open relative to the FCC, HCP or BCC packing of spheres.

### 3.8.3 Properties of elements with the DC structure

The most striking feature of the diamond cubic structure is its four-fold coordination. The elements that crystallize in this structure are Group IVA elements, carbon (in its high pressure form, diamond), silicon, germanium and tin (in its low-temperature form, gray tin), which form saturated covalent bonds by sharing their four valence electrons with four equivalent neighbors. The diamond cubic structure is important in engineering primarily because it is the structure of the important elemental
semiconductors, Si and Ge. Gray Sn is also a semiconductor, but since it is only stable at very cold temperature it is not commonly used. (Gray tin has a negative significance in engineering in that it is the cause of "tin pest"; metallic Sn disintegrates into powder when cooled below its transformation temperature. The "pest, if my sources are correct, struck the army of Napoleon during his 1812 invasion of Russia with most embarrassing consequences, but that is another story.) The diamond form of carbon is an insulator. Its bonding is essentially the same as that in Si and Ge, but the band gap in diamond is so large (equivalently, the covalent bonds are so strong) that its electrical conductivity is very small.

Because of the strong preference for four-fold coordination, materials that have the diamond cubic structure resist distortions that change the local atomic configuration and, therefore, tend to be mechanically hard and brittle. Diamond is the hardest substance known. On the other hand, it is relatively easy to form diamond into faceted gemstones because it is brittle, and fractures readily along well-defined crystallographic planes. (Gemologists talk about "cutting" diamond, but diamond is, in fact, virtually impossible to cut. They fashion it by "cleaving" it, that is, by fracturing it along particular crystallographic planes.)

### 3.9 Solid Solutions and Compounds

#### 3.9.1 Solid solutions

No real material is chemically pure. In addition to the dominant atom type, the solvent or matrix specie, other species are present as intentional additions (soluties) or inadvertent additions (impurities). The solutes and impurities occupy one of two types of sites in the solid, depending on their size and chemical compatibility with the solvent. If they are similar in size and chemical character they replace solvent atoms on the crystal lattice, and are hence called substitutional solutes. If they are much smaller in size or very different in chemical character they may occupy the empty spaces between solvent atoms on the crystal lattice. These empty spaces are called interstitial sites, and solutes or impurities that preferentially fill them are called interstitial solutes. When the solutes are distributed in a more or less random way through the crystal the material is called a solid solution, or alloy. When the solutes fill an ordered distribution of atom sites the material is called an ordered solution, or compound.

#### 3.9.2 Random solutions; pseudo-atoms

Fig. 3.35 is a drawing of a close-packed hexagonal array of atoms that contains both substitutional and interstitial solutes in an irregular pattern. In the strict sense the solutes destroy the periodicity of the lattice. While there are atoms at the nodes of a close-packed hexagonal lattice the presence of substitutional and interstitial solutes has the consequence that the points of the grid do not have the same environment, and hence do not make up a true crystal lattice. The repeat period, or unit cell, is the size of the whole crystal. To describe the atom configuration it is, strictly, necessary to state what
kind of atom is located at each site of the close-packed hexagonal grid and where interstitial atoms are located.

However, when the solute concentration is small or the solute specie is easily accommodated the solute atoms are distributed in a more or less random way over the sites they occupy. In fact, their positions change with time; thermal agitation has the consequence that they move from site to site. Every point of the grid of atom sites is statistically equivalent to every other in the sense that the probability that it is occupied by a given type of atom at any given time is the same. We describe random solid solutions of this type as crystals of pseudo-atoms, which are imaginary atoms that average the properties of the atoms that are actually present. After the atoms are replaced by statistically averaged pseudo-atoms every point of the grid has the same environment, and it is a proper crystal lattice.

![Diagram](image)

Fig. 3.35: (a) A close-packed hexagonal crystal with substitutional and interstitial solutes. (b) A close-packed hexagonal crystal of identical pseudo-atoms (only the substitutional atoms are shown).

As an example the substitutional part of the solution shown in Fig. 3.35a is redrawn in Fig. 3.35b as a close-packed hexagonal structure of pseudo-atoms that average the properties of the substitutional species. To complete the description the interstitial sites indicated in Fig. 3.35a should also be filled with pseudo-atoms that reflect the statistical probability that an atom occupies a particular interstitial site. The pseudo-atoms make up a well-defined crystal. The description of the material as a crystal of pseudo-atoms does not give the actual configuration of atoms, but it does reproduce the statistical distribution (the probability that an atom of given type is located at a given position at any given time), which is an adequate description of the atomic arrangement for most purposes.

**3.9.3 Ordered solutions (compounds)**

When the concentration of solute is large enough that the solute atoms interact significantly with one another they often form ordered solutions in which the solute atoms are distributed in a regular way over the sites they occupy. In an ordered solution the solute and solvent species are present in a definite proportion (to within defects in the
packing). An ordered solid solution of species B in A has a definite chemical formula, \( A_xB_y \), which gives the stoichiometric proportions of the two species when the order is perfect. Because of the formal similarity between the stoichiometric formula of an ordered solid solution and that of an isolated molecule in chemistry, ordered solid solutions are often called \textit{compounds}. However, ordered solutions differ from chemical compounds in that the atoms are distributed over the sites of a continuous lattice. It is not ordinarily useful to distinguish individual \textit{molecules}, \( A_xB_y \), in the solid since there is no unique way to identify the molecule to which particular atom ought to be assigned.

Two examples of two-dimensional ordered solutions are given in Fig. 3.36. Both have the stoichiometric formula \( A_3B \) (the A atoms are light, the B atoms are shaded in the figure), and both are obtained by distributing B atoms in a regular way over the sites of a close-packed hexagonal background lattice. In fact, both of these ordering patterns are found on atom planes in the three-dimensional crystal structures of important compounds of the type \( A_3B \).

![Diagram](image)

**Fig. 3.36:** Two substitutionally ordered structures of the type \( A_3B \) on a close-packed hexagonal lattice: (a) hexagonal superstructure; (b) rectangular superstructure. The unit cells are outlined.

The ordering of the solute changes the crystal structure of the material. At the very least it increases the size of the unit cell. It often changes the type of the unit cell as well. For example, the ordering pattern of Fig. 3.36a changes the two-dimensional unit cell from the primitive hexagonal cell that would pertain if the crystal were disordered (Fig. 3.35) into the larger hexagonal cell that is outlined in the figure. The crystal structure remains hexagonal, but the unit cell now contains a net of four atoms, three of type A and one of type B. The ordering pattern of Fig. 3.36b changes the Bravais lattice, which is hexagonal if the solution is random, but is rectangular when it is ordered in the pattern shown.

The overall crystal lattice of an ordered solution is determined by the ordering pattern of the solute atoms, which often obscures the simple underlying symmetry of the grid of atom positions. For example, to understand the behavior of a material that has the atom pattern shown in Fig. 3.36b it is often much more relevant that the underlying pattern of atom positions is hexagonal close-packed than that the B atoms are ordered in a rectangular pattern. In many materials the structure of the solution changes from
disordered to ordered as the temperature is lowered; the \textit{ordering reaction} is simply a reconfiguration of the solute onto selected atom sites of the parent, disordered lattice. For this reason it is common and useful to give both the ordered and disordered structures of an ordered compound, that is, to state both the \textit{parent structure} of atom positions, irrespective of atom type, and the \textit{superstructure} that is developed when the atoms order onto selected sites.

For example, the structure in Fig. 3.36a can be described as a hexagonal ordering of the $A_3B$ type on a hexagonal close-packed lattice, while the structure in Fig. 3.36b is a rectangular ordering of the $A_3B$ type on a hexagonal close-packed lattice.

We use this method of description to treat the three-dimensional structures of compounds. To describe ordered compounds we almost always refer them to the parent structure of a hypothetical solid solution in which substitutional and interstitial species are randomly distributed over the sites they occupy in the ordered frame. The atomic order then creates a \textit{superstructure} of the disordered \textit{parent structure}. This method leads to a simple interpretation of a wide variety of important material structures that are otherwise difficult to visualize. It should be kept in mind, however, that the actual lattice of the final compound is determined by the periodicity of the crystal in its final, ordered state. The Bravais lattice of the ordered compound may have a significant different symmetry from that of the parent structure, which may cause qualitative differences in its engineering properties.

The crystal structures of many of the multi-component solids that are important in engineering are simple modifications of the FCC, BCC or HCP structures. This includes the classic crystal structures of binary compounds, such as those of the prototype structures we describe in the following sections. It also includes the structures of many oxides and complex ordered compounds, although in this case the simplicity of the basic structure is often difficult to see.

\section*{3.10 \textbf{SUBSTITUTIONAL COMPOUNDS}}

In this section we describe the CsCl and Cu$_3$Au structures as common examples of substitutional compounds.

\subsection*{3.10.1 Substitutional BCC superstructure: CsCl}

The prototype example of a substitutional compound based on BCC is the crystal structure of CsCl, whose unit cell is drawn in Fig. 3.37. In CsCl the body centers of the unit cell are occupied by one kind of atom (let us assume Cs) while the corner sites are filled by the other (Cl). Since there is one body-centered and a net of one corner site per unit cell of the BCC structure the CsCl structure has the correct stoichiometry. Since the body center and corner sites in the BCC structure are entirely equivalent (they can be converted into one another by translating the origin to the center of the unit cell) it does not matter which atom we take to occupy the center and which the corner sites.
Note that while the parent structure of CsCl is BCC, the ordered superstructure has a simple cubic Bravais lattice. As drawn, the B atoms are located at the sites of the Bravais lattice while the positions of the A atoms are given by a basis vector, \( \mathbf{\tau} = \frac{1}{2} [111] \), that connects the body-centered site to the origin of the unit cell.

![Unit cell of an AB compound that has the CsCl structure.](image)

**Fig. 3.37:** The unit cell of an AB compound that has the CsCl structure.

The compounds that have the CsCl structure include ionic solids, such as CsCl itself, and intermetallic compounds, such as \( \beta' \)-brass (CuZn). Because of the eight-fold coordination these compounds are necessarily metallic or ionic in their bonding; the configuration is inconsistent with saturated covalent bonds. Since all of the nearest neighbors of an A-site are B-sites, the CsCl structure is a natural one for binary ionic compounds like CsCl.

In the case of the intermetallic compounds one can actually observe the separation of elements that creates substitutional order on the BCC parent lattice. A classic example occurs in Cu-Zn (brass). When a liquid solution of equal amounts of Cu and Zn is solidified it crystallizes into a BCC structure (\( \beta' \)-brass) in which Cu and Zn atoms are randomly distributed over the sites of the BCC lattice. However, if this structure is cooled to a temperature below about 460 °C the Cu and Zn atoms spontaneously arrange themselves (undergo an *ordering reaction*) that creates the CsCl structure.

### 3.10.2 Substitutional FCC superstructure: Cu$_3$Au

The Cu$_3$Au, or L1$_2$ structure is a particular ordering of atoms on the FCC lattice that is drawn in Fig. 3.38. The Cu atoms occupy the face centers of the FCC unit cell while the Au atoms sit on the corners. Since there are 3 face-centered atoms per unit cell and 1 corner atom, the structure has atom sites in the correct proportion to give the stoichiometric formula A$_3$B.

The Bravais lattice of the Cu$_3$Au superstructure is, again, simple cubic. As it is drawn the Au atoms are located on the sites of the simple cubic lattice, and the Cu atoms can be located by the three basis vectors, \( \mathbf{\tau}_1 = \frac{1}{2} [110], \mathbf{\tau}_2 = \frac{1}{2} [101], \mathbf{\tau}_3 = \frac{1}{2} [011] \) from each lattice site.
Fig. 3.38: The unit cell of the Cu₃Au structure.

Binary compounds that have the Cu₃Au structure are metallic; this structure is found in many intermetallic compounds. Strong ionicity is unlikely since the like ions that occupied the Cu positions in the structure would be nearest neighbors of one another, and their mutual repulsion would cause the structure to have a relatively high energy. However, Cu₃Au-type ordering is also found in many polyatomic compounds that bond ionically. In these compounds ions are distributed both over a parent FCC lattice that is ordered in the Cu₃Au pattern, and over the octahedral interstitial positions, where a Cu₃Au pattern also appears on the FCC arrangement of octahedral interstitial sites. An example is the perovskite structure, which is the structure of a number of important ferroelectric oxides, and is the basic structure from which the crystal structures of several of the oxide superconductors are derived.

3.11 INTERSTITIAL COMPOUNDS

We have already discussed the octahedral and tetrahedral voids in the FCC, HCP and BCC structures. The crystal structures of many compounds that are important in engineering are derived from these elemental structures by filling the interstitial voids in an ordered way. We shall discuss four examples that can serve as prototypes: the NaCl and β-ZnS structures, which are based on FCC, the Č-ZnS structure, which is based on HCP, and the Nb₃Sn (A15) structure, which is based on BCC.

It is possible to understand many, though not all, of the crystal structures that are important in engineering by recognizing that they can be constructed by distributing atoms in an ordered way over the sites of a basic FCC, HCP or BCC lattice and the tetrahedral and octahedral voids it contains.

3.11.1 Octahedral interstices in FCC: NaCl

The unit cell of the NaCl structure is drawn in Fig. 3.39. One set of atoms (assume Na) sits on the sites of the FCC lattice. The other (assume Cl) fills the octahedral voids. Since there is one octahedral void per atom in a close-packed structure the stoichiometric composition is NaCl. The octahedral voids in FCC also form an FCC lattice. This can be seen by simply displacing the origin of the unit cell by the vector \( \frac{1}{2} [100] \) to move it from the corner to the center of the octahedral void on the forward edge.
of the unit cell. It follows that it is irrelevant which atom of an AB compound is assigned to the FCC lattice and which to the octahedral voids in drawing the structure.

![FCC and octahedral structure](image)

\[ = \text{FCC atoms (Na)} \]
\[ = \text{octahedral atoms (Cl)} \]

... Fig. 3.39: The NaCl structure. Note that it is irrelevant whether Na or Cl is assigned to the FCC sites.

The two types of atoms in the NaCl structure strictly alternate; the nearest neighbors of an FCC atom are octahedral atoms, the nearest neighbors of an octahedral atom are on FCC sites. The NaCl structure is hence a natural one for equi-valent ionic compounds. There are many examples among the halides, such as NaCl, and simple oxides, such as MgO. Because of the eight-fold coordination of the nearest neighbor shell, binary compounds with the NaCl structure cannot have saturated covalent bonds. However, they can be metallic. Some of the semiconducting III-V compounds transform to the NaCl structure under high pressure and become metallic. The structures of the superconducting oxides are also based on NaCl; the perovskite structure of these oxides is an NaCl structure that is modified by L12-type ordering on both the lattices of the FCC and the octahedral atoms to provide sites for four components.

### 3.11.2 Tetrahedral interstices in FCC: β-ZnS

The β-ZnS, or sphalerite structure is a tetrahedrally coordinated FCC interstitial structure. We have, in fact, already studied it, since it is the binary analog of the diamond cubic (DC) structure. One component (assume Zn) occupies the sites of the FCC lattice while the other (assume S) fills 1/2 of the tetrahedral voids in the same pattern used in diamond so that all atoms have precisely 4 equidistant neighbors in tetrahedral coordination. The unit cell of the β-ZnS structure is drawn in Fig. 3.40. It stacking pattern is AaBbCc, where the ABC sites are occupied by one atom and the abc sites by the other. Since AaBbCc becomes the equivalent sequence aAbBcC if the origin is displaced by the vector \( \frac{1}{4} [111] \) (as we discussed in connection with the diamond structure) it does not matter which specie is located on the FCC lattice sites and which in the tetrahedral voids.
Fig. 3.40: The β-ZnS structure. One kind of atom (shaded) fills the sites of an FCC lattice while the other (white) fills one-half of the tetrahedral voids.

Since each atom in the β-ZnS structure has four nearest neighbors in tetrahedral coordination the β-ZnS structure is a natural structure for covalently bonded binary compounds. It is important in engineering because it is the structure of the important III-V semiconductors, such as GaAs and InSb, and II-VI semiconductors, such as ZnS and CdS (though both of these structures are also found in the Č-ZnS structure). Light elements that crystallize in the β-ZnS structure tend to be insulators that strongly resemble diamond in their properties, such as BN and SiC. Since all of the neighbors of an FCC atom in the β-ZnS structure are tetrahedral atoms, and vice versa, this structure is also a natural one for equi-valent ionic compounds. It is, however, somewhat difficult to decide which compounds of the β-ZnS type should be regarded as ionic; the bonding is almost always partly covalent.

### 3.11.3 Tetrahedral Interstices in HCP: â-ZnS

Given the similarity between the atom packing and void configurations in the FCC and HCP structures, one might expect to find interstitial compounds based on the HCP structure as well. There is, in fact, an alternate hexagonal structure of diamond, a mineral called lonsdaleite, but this structure is uncommon and specific to carbon. The hexagonal analog to the β-ZnS structure is common. It is called the wurtzite, or Č-ZnS structure. It is a binary compound that is constructed by placing one kind of atom (assume Zn) on the HCP lattice sites and filling alternate tetrahedral voids with the other (assume S). The stacking sequence is, hence, AaBbAaBb, or, equivalently, aAbBaAbB. The unit cell of the Č-ZnS structure is shown in Fig. 3.41. It is derived from the HCP unit cell in Fig. 3.23 by placing atoms of the second type in the tetrahedral voids that lie above each of the atoms in the HCP unit cell.
Fig. 3.41: The Ć-ZnS structure. The open circles shown atoms on the HCP sites, the shaded ones show atoms in one-half of the tetrahedral voids.

The most common materials that exhibit the Ć-ZnS structure are the I-VII compound AgI, the II-VI compounds BeO, ZnO, ZnS and CdS, the III-V compound AlN, and the IV-IV compound SiC. Several of these, ZnS, CdS and SiC, are found in both the Ć-ZnS and β-ZnS structures. The Ć-ZnS structure tends to appear in compounds that are slightly more ionic in their bonding than those that form the β-ZnS; it can be shown that the Ć-ZnS structure yields a slightly lower energy when the species are ions. For much the same reason, Ć-ZnS compounds tend to be electrical insulators with large band gaps (CdS is an exception; its band gap is relatively small).

3.11.4 Tetrahedral interstices in BCC: the A15 structure (Nb₃Sn)

There are also interstitial compounds that are based on BCC, and utilize either the octahedral or tetrahedral voids. As an example, we shall describe one of the more important, the A15 structure that is taken on by a number of compounds with the stoichiometric formula A₃B. The most interesting of these are the A₁₅ superconductors, Nb₃Sn, Nb₃Al, Nb₃Ge, V₃Ga, which are the superconducting materials currently used in very high field superconducting magnets.

The A15 structure is drawn in Fig. 3.45. The minor element, B(Sn), fills the BCC sites while the major element, A(Nb), fills one-half of the tetrahedral voids. Since there are six tetrahedral voids per atom in BCC this filling yields the correct stoichiometry. The tetrahedral voids are filled in a pattern that preserves the cubic symmetry of the BCC cell. There are four tetrahedral voids in each face of BCC. In the A15 structure two of the voids in each face are occupied, in a pattern such that the filled tetrahedral voids form linear chains of atoms parallel to the three cube edges (the three <100> directions). Since there are equal numbers of chains along the three axes, the net effect of adding the tetrahedral atoms is to expand the cell without distorting it.

The separated chains of tetrahedral atoms in the A15 structure are important to its superconductivity; the chains provide easy pathways for current flow. Since equivalent chains parallel each cube edge, the conductivity is isotropic (the same in every direction).
The A15 superconducting compounds offer combinations of high superconducting transition temperature, high current capacity and isotropic conductivity that make them attractive for use in high field superconducting magnets.

![Diagram of A15 structure](image)

**Fig. 3.45:** The A15 structure of A₃B compounds such as Nb₃Sn. The minor element, B(Sn), fills the BCC sites. The major element, A(Nb), fills one-half of the tetrahedral voids in BCC.

### 3.12 UNDERSTANDING CRYSTAL STRUCTURE

Since the properties of a material depend strongly on its crystal structure, it would be useful to have simple rules that predict the structure a material of given composition will adopt. Unfortunately, while there are such rules, they are either very approximate or mathematically complex. Even given the powerful tools of modern theoretical physics, it is not easy to predict crystal structures accurately. The basic reason that the problem is hard, at least for solids that have shared-electron bonds (metallic or covalent) is that these solids care surprisingly little which crystal structure they have.

To a good approximation the cohesive energy of a solid can be written in the form

\[ E = E_0(v) + E_1(\{R\}) \]

where \( E_0(v) \), the dominant contribution, depends only on the atomic volume, \( v \). The influence of the specific configuration of atoms is gathered in the second term, \( E_1(\{R\}) \), where \( \{R\} \) is the configuration, the set of position vectors that specify the arrangement of atoms at a given value of the atomic volume, \( v \). The configurational energy, \( E_1 \), is ordinarily much smaller than \( E_0 \). To predict the preferred structure of a solid from first principles one must calculate the cohesive energy to very high accuracy since the choice of structure is based on differences in the value of the second term, which is a small correction to a relative large number. It can be done for some elemental solids. Marvin Cohen and his colleagues in the Department of Physics here at Berkeley have accurately calculated the structures of several light elements from first principles. But that is not an exercise I would recommend to the uncommitted.
However, there are qualitative trends in the structures of elements and compounds that permit us to understand why many materials have the crystal structures they do. These trends can be described in terms of three variables: the average number of valence electrons per atom, the relative sizes of the atoms or ions that are present, and the difference in electronegativity (tendency to attract an electron) between them. The electron/atom ratio can often be used to predict or understand the structures of elemental solids and substitutional solid solutions and compounds. The size difference affects the possibility of forming substitutional solid solutions, the preferred interstitial site in interstitial compounds, and the preferred coordination in ionic solids. The electronegativity difference is an indication of the ionic contribution to bonding.

3.12.1 The structures of elemental solids

A glance at the periodic table shows that there are systematic trends in structure with valence, particularly if we focus our attention on the lighter elements and ignore those with complex structures we have not studied. The Group IA elements Li and Na are HCP at low temperature. The Group IIA elements Be and Mg and the IIB elements Zn and Cd are also HCP. The Group IIIA element Al is FCC. The Group IVA elements C(diamond), Si, Ge and Sn are DC. The transition metals also follow a rough sequence with valence, changing from HCP (Sc and V) to BCC (V and Cr) to FCC (Fe(high temperature), Ni, Cu) as the total number of outer-shell electrons (d+s+p) increases.

Recent research in a branch of the quantum theory of solids that is called pseudo-potential theory has produced reasonably accurate models that suggest why these trends should appear. The model that applies to the non-transition elements is relatively simple. A brief description may help clarify why elements choose the structures they do.

Fig. 3.46: The characteristic shape of the two-body interaction potential in a valence solid. The parameter, R, ranges over several atom spacings.
In elemental solids of the normal valence type the configurational energy, $E_1(\{\mathbf{R}\})$, can be written to a good approximation as a sum of binary interactions between the atoms,

$$E_1(\{\mathbf{R}\}) = \frac{1}{2} \sum_{\mathbf{R}, \mathbf{R}'} \Phi(|\mathbf{R} - \mathbf{R}'|)$$  \hspace{1cm} (3.8)

where the potential, $\Phi$, depends only on the magnitude of the distance, $\mathbf{R} = |\mathbf{R} - \mathbf{R}'|$, between the atom at $\mathbf{R}$ and the one at $\mathbf{R}'$, and the whole set of atom positions, $\{\mathbf{R}\}$, is constrained so that the mean atomic volume is fixed. The configurational interaction, $\Phi$, is relatively long-range and oscillatory. Its form is sketched in Fig. 3.46 below. The precise form of the potential, $\Phi$, varies from one element to another. However, the variation is primarily due to two differences: the valence, $z$, and the atomic volume, $v$, which increases with the atomic number. The dominant part of the variation with atomic volume is a relatively simple one. In fact, if we measure the distance, $\mathbf{R}$, in units of $R_0 = v^{1/3}$, which is the effective radius of the atom, we can eliminate most of the dependence of the potential on the atomic volume. With this choice of units, elements that have the same valence have very nearly the same potential. In particular, the maxima and minima of the potential are located at almost the same values of $R^* = R/R_0$.

To use this potential to estimate the relative energy of a particular structure, note that if a particular atom is taken as a reference, the atoms around it can be divided into sets so that there are $n_i$ atoms in the $i^{th}$ neighbor shell at a distance, $R_i$. This construction is illustrated in Fig. 3.47 for a hexagonal close-packed structure in two dimensions. In this case there are 6 atoms in the first neighbor shell at $R_1$, 6 in the second neighbor shell at $R_2$, and 6 in the third neighbor shell at $R_3$. Using this construction eq. 3.8 can be rewritten

$$E_1(\{\mathbf{R}\}) = \frac{1}{2} \sum_k n_k \Phi(R_k)$$  \hspace{1cm} (3.9)

where $n_k$ is the number of atoms in the $k^{th}$ shell around a reference atom and $R_k$ is the radius of the $k^{th}$ neighbor shell. If we change the radius ($R_0$) of the atom then all of the distances, $R_k$, scale with the change; the dimensionless distance to the $k^{th}$ shell, $R_k = R_k/R_0$, depends on the crystal structure only.

It follows from Fig. 3.46 and equation 3.9 that the energy of a particular structure is relatively low if the neighbors are concentrated at radii for which $\Phi(R)$ is near one of its minima, and is relatively high if they are concentrated near maxima of $\Phi(R)$. By simultaneously plotting the distributions of neighbors for several different candidate structures it is often visually apparent which will have the least energy, and, therefore, will be the preferred structure.
Fig. 3.47: A two-dimensional hexagonal close-packed arrangement of atoms illustrating the first three neighbor shells.

This procedure has been used with considerable success to predict not only the structures of simple metals, but also the diamond structure of Group IV semiconductors, and even the complex structures of such elements as phosphorous. The theory reveals three particularly important features of bonding and crystal structure. First, it shows why there should be a systematic trend of structure with valence; valence is the principal variable that determines the shape of the bonding function. It also explains why the structure is not uniquely determined by the valence. The valence does not completely determine the function, $\Phi(R)$, and minor differences for elements with the same valence may cause them to have different structures, particularly if the energies of the different structures are close. Second, the model shows why many solids choose structures that are not close-packed, such as the BCC and the diamond cubic structures. If the valence is such that the potential would be high at the nearest-neighbor shell of a close-packed structure, the solid can lower its energy by adopting a more open structure than locates its neighbors at distances where the potential is relatively low. Third, the theory demonstrates that the distinction we have drawn between covalent and metallic bonding is not a very meaningful one. Both "covalent" and "metallic" solids from shared-electron bonds that lead to bonding functions that have the same qualitative form (Fig. 3.46).

Similar theoretical treatments have been constructed for the transition metals. They are complicated by the fact that the bonding interaction of the d-orbitals is both short-range and directional. A rough model considers bonding to be the sum of short-range d-bonds and a long-range, sp bonding that has a form like that in Fig. 3.46. Both depend strongly on the number of electrons in the outer shell, which determines the number and configuration of d-electron bonds and the valence of the sp-bonds. An important qualitative feature of this work explains why elements near the center of the transition metal series tend to have BCC structures. Bonding by d-electrons is most important for these elements, since they have partly filled d-shells. The shape of the atomic d-orbitals is such that d-electron bonds are strongest when the structure is BCC.
3.12.2 The structures of solid solutions

Unless the atoms in a solid solution have similar sizes and electronegativities it is unlikely that the solution will exist outside a narrow range of compositions. If the atomic sizes are very different then the solute distorts the lattice. To relieve the distortion the solution either separates into its components or forms a compound that has a structure that is compatible with the size difference. In either case the range of solubility is limited. A useful general rule states that concentrated solid solutions cannot exist unless the size difference between the atoms is less than 15%. A large difference in electronegativity causes a charge transfer between unlike atoms that ordinarily causes them to bond to one another. If unlike atoms preferentially bond, the solid forms an ordered compound to maximize the number of bonds, and the mutual solubility is limited. Since the electronegativity varies with the column in the periodic table, from a minimum value for the alkali metals (Group IA) to a maximum for the halides (Group VIIA), elements that are widely separated in the periodic table do not ordinarily form solid solutions.

When two elements do form a substitutional solid solution, its crystal structure can often be guessed from its mean valence (average number of valence electrons per atom). The reason is that a random solid solution behaves very much like an elemental solid of "pseudo-atoms" that average the properties of its constituents. The preferred structures of these pseudo-elements are determined by valence for the same reasons the structures of elemental solids are.

However, the average valence of a solid solution takes non-integral values and varies with composition. Since the preferred structure varies with the valence, it is common for the structure of a solid solution to change with its composition. The classic examples are solutions of the Group IB elements, Cu, Ag and Au with a variety of solutes. The Group IB elements are FCC in the pure state and have effective valence, \( z = 1 \). Many of their solid solutions change to BCC at an electron/atom ratio, \( z = 1.5 \). The critical electron/atom ratio is independent of the valence of the solute; BCC-based structures are observed for CuZn, CuAl and CuSn, all of which have \( z = 1.5 \). When \( z \) reaches 1.61 the solutions adopt a complex structure that is known as the \( \gamma \)-brass structure. When \( z = 1.75 \) they become HCP. In these cases as well, the change in structure is determined by the average valence and is independent of the actual valence of the solute. The rules that give the structures of these solutions as functions of valence are known as the \textit{Hume-Rothery rules}, after their discoverer. Similar relations pertain to many other systems.

3.12.3 The structures of ordered compounds

For simplicity we consider binary compounds. There are two kinds: substitutional compounds, in which the two elements are distributed in an ordered way over the sites of a parent lattice, and interstitial compounds, in which one element establishes a lattice frame and the other fills an ordered arrangement of interstitial sites in that frame. In both cases there are two structural decisions to be made: the choice of the basic structure and
the choice of the ordered arrangement that distributes the elements with respect to that structure.

When the compound contains more than two elements the ordered structure is often a complex one that contains ordered distributions on both the substitutional and interstitial sites.

**Substitutional compounds**

The elements that form substitutional compounds are similar in size and, ordinarily, not too different in electronegativity. In many cases the basic structure can be easily guessed, since it is the preferred structure of the pseudo-atom that averages the valence of the two elements. For example, the compound GaAs joins a valence 3 element (Ga) and a valence 5 element (As). The compound has average valence, \( z=4 \), and an average atomic number equal to that of the Group IV element, Ge. Since Ge is diamond cubic, it is not surprising that the structure of GaAs is \( \beta \)-ZnS, which is the binary analog of the DC structure. As a second example, at high temperature CuZn has a BCC structure, in keeping with the Hume-Rothery rule for its valence, \( z=1.5 \). At lower temperature it orders into the CsCl structure, which is an ordered version of BCC, and has the properties that it maintains the stoichiometry, CuZn, and maximizes the number of Cu-Zn bonds.

In other cases the ordering pattern is less obvious, even when the basic structure is known. An example is the Cu-Au binary system. At high temperature Cu and Au form FCC solid solutions at all compositions (the average of Cu and Au is Ag, which is FCC). At low temperature the Cu-Au solution orders, but only into one of three compounds: Cu\(_3\)Au, CuAu, or Au\(_3\)Cu. No other ordered compounds occur. It is relatively easy to see why the compounds Cu\(_3\)Au and Au\(_3\)Cu should appear on this list. They both have the Cu\(_3\)Au structure (Fig. 3.38), which has just the right properties: it is an ordering of the FCC structure, and it maximizes the number of Cu-Au bonds for this composition and structure. The CuAu structure, which we have not studied, has the same properties. However, we could find an ordered structure that has these properties for any composition, and we would like to know why these structures do not appear. There are theories that address this subject in some detail; basically, they show that when the composition differs from that of the structures we have identified, it is energetically more favorable for the solution to decompose into a mixture of two structures than to adopt some other ordered pattern. For example, the energy of a mixture of the Cu\(_3\)Au and CuAu structures has a lower energy than the probable energy of a homogeneously ordered arrangement at any composition between 0.25 and 0.5 Au. The theories that identify the ordered structures and their probable energies are elegant and fascinating, but they are beyond the scope of this course.

In the examples we have examined so far, the basic structure of the ordered compound is determined by the "pseudo-atom" that averages its constituents. In some cases, however, the interaction between the elements is so strong that it dominates over the average and changes the structure. An example is ZnO. The "average" of Zn and O is Si,
which crystallizes in the diamond cubic structure. However, ZnO has the hexagonal, Č-ZnS structure rather than the β-ZnS structure that is derived form the diamond cubic. The apparent reason is that there is a strong ionic component to the bonding of ZnO, and the Č-ZnS structure has a configuration that is slightly more favorable for ionic bonding.

**Interstitial compounds**

There are two common types of interstitial compounds: ionic compounds in which the interstitial ions interact electrostatically with ions on the primary lattice, and intermetallic compounds in which the interstitials interact primarily through the elastic distortions they impose on the parent lattice.

An important example of the latter are the compounds of iron with carbon and nitrogen, which are used to strengthen steel. These compounds have ordered structures that minimize the elastic energy of the overlapping distortions of the interstitial atoms. There are good theories of these structures, but they are beyond the scope of this course.

The ionic interstitial compounds have structures that are more easily explained. Since the attractive interaction between unlike charges decreases with distance, an ionic compound prefers a structure in which the ions are as closely packed as possible, under the constraint that the nearest neighbors of an ion are all ions of opposite charge. The ionic structures are such that the cations (which are the smaller ions because they have lost valence electrons) sit in interstitial sites that have only anions on lattice sites as neighbors. The structure that provides the most favorable packing depends on the relative size of the ions. Ideally, the ion should just fill the interstitial site. It should not have a radius smaller than that required to contact its anion neighbors, since it should prevent these from overlapping one another.

In the simplest common case the ions have the same valence and r*, the ratio of the cation to anion radius, is less than 1. When r* is close to 1 the preferred structure is the CsCl structure, in which each cation is coordinated by eight anions. The CsCl structure is the most densely packed structure of similarly sized ions that can be used; the close-packed structures, FCC and HCP, cannot be used because the nearest neighbors of an ion in the close-packed structures are also nearest neighbors of one another, and repel strongly. If r* falls below 0.732 the corner atoms of the CsCl cell impinge on one another. At this point the preferred structure becomes the NaCl structure, in which the cations fill octahedral interstitial sites with six neighbors. This structure is preferred until r* falls below 0.414, at which point the six neighbors of the octahedral void impinge. At this point the preferred structure is Č-ZnS with the cations sited in tetrahedral voids (the β-ZnS structure is also possible, but has a slightly higher energy when the bonding is strongly ionic). The anions that coordinate the tetrahedral void impinge when r* falls below 0.225. At this point the preferred coordination becomes one in which the cation is coordinated by a triangle of anions. This planar configuration is not found in any of the structures we have studied.
Fig. 3.48: The preferred structures of a binary ionic compound for various values of the cation/anion radius ratio, r*.

Many important ionic compounds contain three or more components and have very complex structures. Nonetheless, the simple coordination rules tend to be preserved. The number of anions that surround and coordinate a cation is usually determined by the radius ratio, and decreases from 8 to 6 to 4 as r* varies through the sequence given in Fig. 3.48.
Chapter 4: Defects in Crystals

...Perfection's a gift of
  The gods, few can boast they possess it - and most
Of you, my dears, don't.
  - Ovid, The Art of Love

4.1 INTRODUCTION

A perfect crystal is an idealization; there is no such thing in nature. Atom arrangements in real materials do not follow perfect crystalline patterns. Nonetheless, most of the materials that are useful in engineering are crystalline to a very good approximation. There is fundamental physical reason for this. The preferred structures of solids at low temperature are those that minimize the energy. The low-energy atomic configurations are almost invariably crystalline since the regular pattern of the crystal lattice repeats whatever local configuration is most favorable for bonding. There is also a fundamental physical reason why the crystal is imperfect. While a perfect crystalline structure may be preferred energetically, at least in the limit of low temperature, atoms are relatively immobile in solids and it is, therefore, difficult to eliminate whatever imperfections are introduced into the crystal during its growth, processing or use.

The fact that real materials are not perfect crystals is critical to materials engineering. If materials were perfect crystals then their properties would be dictated by their composition and crystal structure alone, and would be very restricted in their values and their variety. It is the possibility of making imperfectly crystalline materials that permits materials scientists to tailor material properties into the diverse combinations that modern engineering devices require. As we shall see repeatedly in the body of this course, the most important features of the microstructure of an engineering material are the crystalline defects that are manipulated to control its behavior.

In this Chapter we shall identify and describe the various defects that are found in crystalline solids. We shall hint at their consequences to indicate why they are worth studying, but shall defer a detailed discussion of their behavioral role until we discuss particular classes of engineering properties.

It is useful to classify crystal lattice defects by their dimension. The 0-dimensional defects affect isolated sites in the crystal structure, and are hence called point defects. An example is a solute or impurity atom, which alters the crystal pattern at a single point. The 1-dimensional defects are called dislocations. They are lines along which the crystal pattern is broken. The 2-dimensional defects are surfaces, such as the external surface and the grain boundaries along which distinct crystallites are joined.
together. The 3-dimensional defects change the crystal pattern over a finite volume. They include precipitates, which are small volumes of different crystal structure, and also include large voids or inclusions of second-phase particles.

### 4.2 POINT DEFECTS

A point defect disturbs the crystal pattern at an isolated site. It is useful to distinguish intrinsic defects, which can appear in a pure material, from extrinsic defects, which are caused by solute or impurity atoms.

#### 4.2.1 Intrinsic defects

An intrinsic defect is formed when an atom is missing from a position that ought to be filled in the crystal, creating a vacancy, or when an atom occupies an interstitial site where no atom would ordinarily appear, causing an interstitialcy. The two types of intrinsic point defects are shown in Fig. 4.1.

![Fig. 4.1: Illustration of a vacancy and an interstitial in a two-dimensional hexagonal lattice.](image)

Because the interstitial sites in most crystalline solids are small (or have an unfavorable bonding configuration, as, for example, in the diamond lattice) interstitialcies are high-energy defects that are relatively uncommon. Vacancies, on the other hand, are present in a significant concentration in all crystalline materials. Their most pronounced effect is to govern the migration of atoms on the crystal lattice (solid state diffusion). In order for an atom to move easily from one crystal lattice site to another the target site must be vacant. As we shall see, the rate of diffusion on the crystal lattice is largely governed by the concentration of vacancies.

Ordered compounds can have more complex intrinsic defects. In most compounds the different species are charged to at least some degree. An intrinsic defect destroys the local charge balance, which must be restored in some way. The compound defects that preserve charge are easiest to visualize in binary ionic solids like NaCl. An isolated vacancy in an ionic solid creates an excess charge. The excess charge can be
compensated by a paired vacancy on the sublattice of the other specie; for example, the excess charge associated with a Na vacancy is balanced if there is a Cl vacancy nearby. A neutral defect that involves paired vacancies on the cation and anion sublattices is called a Schottky defect. Alternatively, the charge imbalance caused by the vacancy can be corrected by adding an interstitial of the same specie; a Na vacancy is compensated by a Na interstitial. A neutral defect that is made up of a paired vacancy and interstitial is called a Frenkel defect. In compounds whose atoms are less strongly ionized it is energetically possible for species to exchange sites, so that an A-atom appears on the B-sublattice or vice versa. This type of point defect is called an anti-site defect, and is fairly common in semiconducting compounds such as GaAs.

4.2.2 Extrinsic defects

The extrinsic point defects are foreign atoms, which are called solutes if they are intentionally added to the material and are called impurities if they are not. The foreign atom may occupy a lattice sites, in which case it is called a substitutional solute (or impurity) or it may fill an interstitial site, in which case it is called an interstitial solute. Since the interstitial sites are relatively small, the type of the solute is largely determined by its size. Small atoms, such as hydrogen, carbon and nitrogen are often found in interstitial sites. Larger atoms are usually substitutional.

More complex extrinsic defects appear in compounds. If the valence of a substitutional defect in an ionic solid differs from that of the lattice ion then the excess charge is often compensated by a paired vacancy or interstitial. For example, when Mg\(^{++}\) ions are substituted for Na\(^+\) in NaCl they tend to be paired with vacancies on the Na sublattice to maintain local charge neutrality. In semiconductors substitutional atoms with the wrong valence acts as electron donors or acceptors, as described below.

Extrinsic point defects affect almost all engineering properties, but they are particularly important in semiconducting crystals, where extrinsic defects are used to control electrical properties, and in structural metals and alloys, where extrinsic defects are added to increase mechanical strength. While these properties will be discussed later in the course, it is perhaps useful to identify the characteristics of the point defects that affect them.

Donors and acceptors in semiconductors

Point defects are intentionally added to semiconductors to control the type and concentration of charge carriers. Consider, for example, boron (valence 3) as a substitutional solute in elemental silicon. The saturated covalent bonds in silicon are shown schematically in Fig. 4.2a, and depend on the availability of four valence electrons per silicon atom. Since the bonds are saturated, silicon has very low conductivity in its pure state; pure silicon can only conduct electricity when electrons are excited into high-energy electron states. If boron is added, as in Fig. 4.2b, a valence electron is missing from the immediate environment of the boron atom, causing a hole in the bonding pattern. Electrons can then move from bond to bond by exchanging with the hole. The
exchange requires some energy to separate the hole from the boron ion core, but this energy is small compared to that required to excite an electron from a Si-Si bond into a high-energy state. The room-temperature conductivity of Si increases significantly when a small amount of B is added. Electron-deficient solutes like boron that cause holes in the configuration of bonding electrons are called *acceptors*.

![Diagrams](image-url)

**Fig. 4.2:** (a) Tetrahedral bonding configuration in Si. (b) Bonding around a B solute, showing a hole (□). (c) Bonding around a P solute, showing an electron (e) in a loose excited orbital.

The conductivity also rises when a solute with an excess of electrons is added to a semiconductor with saturated bonds. For example, let phosphorous (valence 5) be added to Si, as in Fig. 4.2c. The 5 valence electrons of P are sufficient to fill the local covalent bonds with one electron left over. This electron can only go into an excited state, and orbits about the P ion core somewhat as shown in the figure. It requires a relatively small energy increment to free this electron from the P core, in which case it can transport current by moving through the lattice. The conductivity of Si rises dramatically if a small amount of P is added. Electron-excess solutes such as P in Si are called *donors*.

Semiconductors whose electrical properties are controlled by electrically active solutes are called *extrinsic semiconductors*. Almost all of the semiconductors that are used in engineering devices are extrinsic.

*S solution hardening in structural materials*

The addition of solute atoms almost always increases the mechanical strength of a solid. The phenomenon is called *solution hardening*. It is due to the fact that the solute atom is always a bit too large or a bit too small to fit perfectly into the crystal lattice site it is supposed to occupy, and distorts the crystal lattice in its attempt to fit as well as possible. As we shall see later, this distortion impedes the motion of the linear defects (dislocations) that are responsible for plastic deformation and, consequently, hardens the crystal. The distortion due to a substitutional solute is relatively small, though the associated hardening may be large enough to be useful in the engineering sense. The distortions due to interstitial atoms such as carbon and nitrogen are normally much greater because of the small size of the interstitial void in which they must fit. The hardening effect of interstitial solutes is large and technologically important; for example, high strength structural steels are alloys of Fe and C.
There is a simple crystallographic reason why interstitial solutes such as C are particularly effective in strengthening BCC metals such as Fe. The carbon atoms occupy octahedral interstitial sites in the BCC structure. Since an atom in an octahedral void in BCC is closer to two of its neighbors that to the other four, it causes an asymmetric distortion of the lattice. As shown in Fig. 4.3, the octahedron is stretched along its short axis, which is the $a_3$, or $z$-axis in the case shown in the figure. The asymmetric distortion of the interstitial site increases its interaction with the dislocations that cause plastic deformation and promotes hardening. In FCC alloys the interstitial sites are symmetric, and the lattice distortion is isotropic. Interstitial solutes are effective in hardening FCC alloys, but are less effective than in BCC alloys.

![Fig. 4.3: The local distortion of the BCC lattice by an interstitial atom in an $O_z$ void.](image)

**Random solid solutions**

At low concentrations it is usually reasonable to assume that the solutes are randomly distributed over the sites they occupy, in which case the material is called a random solid solution. The basic properties of a random substitutional solid solution can often be estimated by treating the material as if it were made up of pseudo-atoms that have the average properties of the components of the solution. In keeping with this idea, most of the properties of random substitutional solid solutions are simply proportional to the solute concentration. We shall see many examples in the following. A particular example is *Vegard's Law*, which asserts that when the solute concentration in a cubic solid solution is small its lattice parameter changes linearly with the concentration:

$$\frac{a - a_0}{a_0} = Kx$$  \hspace{1cm} 4.1$$

where $a_0$ is the lattice parameter of the crystal in pure form, $x$ is the atom fraction of solute and $K$ is a constant. The constant, $K$, is usually much larger for an interstitial solute because the solute atom must be fit into the small interstitial hole.
Tetragonal distortion by octahedral interstitials in BCC

At larger solute concentrations the solute atoms interact and influence one another, and their actual configuration in the crystal becomes important. An important and instructive example occurs in the case of octahedral interstitials in BCC materials, and is responsible for the technologically important fact that steel (Fe-C) that has a carbon content greater than a few tenths of an atom percent has a body-centered tetragonal rather than a BCC structure when it cooled rapidly (quenched) from elevated temperature. The body-centered tetragonal structure (BCT) differs from the BCC in that it is elongated along one of the three crystal axes (assume the \( a_3 \), or z-axis), as shown in Fig. 4.4.

![Diagram of tetragonal distortion](image)

Fig. 4.4: Tetragonal distortion of a cube. The distortion creates a tetragonal figure with an increased value of \(|a_3|\).

The source of the BCT structure lies in the asymmetry of the octahedral void that is the preferred site for carbon. Since there are equal numbers of \( O_x \), \( O_y \) and \( O_z \) octahedra in the BCC lattice, a random distribution of octahedral interstitials causes, on the average, an equal distortion of the three crystal axes and simply expands the BCC lattice. However, because of the asymmetric distortion, octahedral interstitials interact with one another and, if they are present in any significant concentration, preferentially occupy one of the three types of octahedral sites so that their short axes are aligned. When this happens there is a net long-range distortion of the crystal lattice in the direction of the short axis of the interstitial void that changes the structure from BCC to BCT.

When pure Fe is cooled from high temperature its structure changes from FCC to BCC. Carbon is more soluble in the FCC phase. If an FCC Fe-C alloy with a moderate carbon content is cooled very rapidly (quenched) then the carbon is trapped in the BCC product. The carbon interstitials preferentially adopt aligned octahedral sites, with the consequence that the quenched alloy, which is called martensite, has a body-centered tetragonal (BCT) structure. It is a very strong material with many important uses.

Decomposition and ordering of solid solutions

Very few species are mutually soluble in all proportions. The solubility range is ordinarily limited by the preferential interaction of the solute atoms, and is limited in one of two ways. If the solute atoms bond preferentially to one another they tend to become
associated in the solid solution, or cluster. As the concentration of solute increases the tendency to cluster becomes more pronounced, until the solution spontaneously decomposes into a mixture of two solutions, or phases, one of which is rich in the solute and one in the solvent. The point at which this decomposition occurs defines the solubility limit of the solid solution. On the other hand, if the solute atoms bond preferentially to the solvent then they tend to adopt ordered configurations in which solute and solvent alternate in a regular crystal pattern. When the solute concentration becomes great enough the solution decomposes into a mixture of a solution that is rich in solvent and an ordered compound that contains nearly stoichiometric proportions of the two atom types.

Note that the definition of a point defect in an ordered compound of species A and B is different from that in a solid solution of B in A. A point defect in a compound is a deviation from the perfectly ordered state of the compound structure. For example, in a solid solution of Au in Cu each Au atom constitutes a point defect, since the pure Cu crystal is the reference. In the intermetallic compound Cu₃Au, on the other hand, an Au atom is a point defect only if it occupies a Cu position in the Cu₃Au structure (in which case it is an anti-site defect), while a Cu atom is a point defect if it occupies an Au site. This re-emphasizes the important point that a "crystal defect" is a deviation from the configuration the crystal would have if it were perfectly ordered, and has no meaning until the reference state, the perfectly ordered crystal, is defined.

4.3 LINE DEFECTS: DISLOCATIONS

Dislocations are linear defects; they are lines through the crystal along which crystallographic registry is lost. Their principle role in the microstructure is to control the yield strength and subsequent plastic deformation of crystalline solids at ordinary temperatures. Dislocations also participate in the growth of crystals and in the structures of interfaces between crystals. They act as electrical defects in optical materials and semiconductors, in which they are almost always undesirable.

The concept of a dislocation in a solid was introduced by Volterra in the nineteenth century. However, it was not until much later that their relevance to the deformation of crystals was recognized; the notion of a dislocation as the carrier of plastic deformation did not appear until 1934. Since the 1950's it has been possible to observe and study dislocations directly using such techniques as transmission electron microscopy and x-ray topography. While dislocations influence many aspects of physical behavior, they are studied almost exclusively in Materials Science. Most of you are encountering the concept of a dislocation for the first time.

4.3.1 The edge dislocation

The simplest way to grasp the idea of a dislocation is to imagine how you might go about creating one. We begin by making an edge dislocation, which is the easiest type of dislocation to visualize in a crystal, and follow the recipe laid down by Volterra.
Consider the solid body that is drawn in Fig. 4.5. For the present purpose it does not matter whether the body is crystalline; it may be easier to imagine that it is rubber.

![Diagram of a block of material with a planar cut](image)

**Fig. 4.5:** A block of material with a planar cut indicated by the shading.

To create an edge dislocation in this body we first make a planar cut part way through it, as illustrated by the shaded region in the figure. We then fix the part of the body below the cut, and apply a force to the body above the cut that tends to displace it in the direction of the cut, as illustrated in Fig. 4.6. The upper part slides, or *slips* over the lower by the vector distance \( \mathbf{b} \), which is the relative displacement of the two lips of the cut. The plane of the cut, where slip occurs, is called the *slip plane*. The cut is finite and constrained at its end, so material accumulates there. The end of the cut, or equivalently, the boundary of the planar region of slip, is a linear discontinuity in the material. The situation after the slip is shown in Fig. 4.6.

![Diagram of the upper part slipped by vector b](image)

**Fig. 4.6:** The upper part of the body has been slipped by the vector \( \mathbf{b} \) over the shaded area. The terminal line is a discontinuity, marked by the heavy line.

Now suppose that we have a mechanism for welding the material back together which is so efficient that it is impossible to tell after the fact that the weld was ever made. If we match the material above the shaded plane to that below so that there is no physical discontinuity across the plane and weld the lips together, the shading disappears since it is impossible to tell that the material was ever separated. However, matching the material across the plane of slip requires that excess material be gathered at the line at which slip terminates. This line is, therefore, a linear defect in the material. It is called an *edge dislocation*. It is an isolated defect, as shown in Fig. 4.7 since after the material has been
re-welded there is no unique way to determine how the dislocation was created. For example, the dislocation would be exactly the same if the material to the right of it on the slip plane were slipped by the vector \(-\mathbf{b}\) on the same plane. Only the exterior step indicates the origin of the dislocation, and this may be removed, or may have pre-existed the formation of the dislocation.

However it was created, the edge dislocation in Fig. 4.7 has the property that it defines an element of slip, \(\mathbf{b}\), where the vector \(\mathbf{b}\) is called the Burgers vector of the dislocation. We can always identify the slip plane of a dislocation like that shown in Fig. 4.7. It is the plane that contains both the Burgers vector, \(\mathbf{b}\), and the line of the dislocation. However the dislocation actually came to its present position, its net effect is that the material above the slip plane to the left of the dislocation (in the direction of \(-\mathbf{b}\)) has been displaced by \(\mathbf{b}\) relative to that below the slip plane. The dislocation is a linear defect whose location is defined by its line and whose nature is characterized by its Burgers vector, \(\mathbf{b}\). In the case shown in Fig. 4.7 the Burgers vector is perpendicular to the dislocation line. This perpendicularity is characteristic of an edge dislocation.

![Diagram of edge dislocation](image.png)

**Fig. 4.7:** Isolated edge dislocation after the cut surface has been rejoined.

If a dislocation moves the area that has been slipped grows or shrinks accordingly. Imagine that the dislocation is initially created at the left edge of the slip plane in Fig. 4.7, and is then gradually moved to the right edge. Applying the construction in Fig. 4.6 to the initial and final positions of the dislocation, it follows that the motion of the dislocation through the body causes the whole volume of material above the slip plane to be displaced by the vector \(\mathbf{b}\) with respect to that below it, as shown in Fig. 4.8.

![Diagram of dislocation movement](image2.png)
Fig. 4.8: Final state of the body after an edge dislocation with Burgers' vector \( \mathbf{b} \) has crossed the whole of the slip plane shown.

Fig. 4.8 illustrates the connection between the motion of a dislocation (in this case, an edge dislocation) and the plastic, or permanent deformation of a material. As we shall discuss in more detail when we consider the mechanical properties of materials, plastic deformation changes the shape of a body without changing its volume, since the total number of atoms and the crystal structure remain the same. A change in shape that occurs at constant volume can always be represented geometrically as the sum of elementary deformations of a type known as \textit{simple shear}. A simple shear is the kind of deformation that deforms a cube into a parallelogram; it changes the angles between initially perpendicular directions in the cube. The shear due to the passage of an edge dislocation is illustrated in Fig. 4.9. While the dislocation translates the top of the crystal rigidly over the bottom to create a discrete step, the Burgers' vector has atomic dimensions, so the step is invisible. Macroscopic deformation is the sum of the slip caused by many dislocations.

Fig. 4.9: Figure (b) is obtained from (a) by a simple shear of the top over the bottom. Figure (c) shows how the same shear can be caused by an edge dislocation.

4.3.2 An edge dislocation in a simple cubic crystal

The procedure that was used to create the edge dislocation that appears in Fig. 4.7 made no reference to the structure of the solid, and can be used to form an edge dislocation in any material. However, when the material is crystalline the ordered pattern of atoms restricts the values that the Burgers vector, \( \mathbf{b} \), can have. The restriction is introduced by the welding step that is used to change the configuration shown in Fig. 4.6 to that in Fig. 4.7. The welding must be so perfect that it is impossible to tell that the two surfaces were ever separated. If the solid is crystalline this can only be true if the crystal structure is continuous across the slip plane after the weld is made. It follows that the relative displacement across the slip plane must equal a lattice vector so that atoms can re-bond without changing their local atomic configurations. Since the relative displacement is equal to the Burgers vector, \( \mathbf{b} \), of the dislocation, \( \mathbf{b} \) must be a lattice vector. If the dislocation is an edge dislocation, \( \mathbf{b} \) must also be perpendicular to the dislocation line.
The geometry of an edge dislocation is relatively easy to visualize when the crystal has a simple cubic crystal structure. The atomic configuration around the dislocation line is more complicated in real crystal structures, but it is not necessary to deal with that complexity to understand the behavior of dislocations at the level we shall need in this course. Whenever we need to consider the crystallography of the dislocation we shall assume that the crystal structure is simple cubic.

An edge dislocation in a simple cubic structure is drawn in Fig. 4.10, which shows both a two-dimensional view and a three-dimensional section along the dislocation line. The dislocation can be created by making a cut in the crystal on the dashed plane that terminates at the dislocation line, displacing the material above the cut plane to the left of the dislocation by one lattice spacing, and allowing the atoms to re-bond across the slip plane. This recipe recreates the simple cubic unit cell everywhere except on the dislocation line itself (ignoring the small elastic distortion of the cells that border the dislocation line). Hence the Burgers vector, \( \mathbf{b} \), of the dislocation that is drawn in the figure is \( \mathbf{b} = \mathbf{a}[100] \), where \( \mathbf{a} \) is a vector along the edge of the cubic unit cell.

![Diagram of dislocation](image)

**Fig. 4.10:** An edge dislocation in a simple cubic structure. The dotted plane is the slip plane.

The process that creates the edge dislocation shown in Fig. 4.10 leaves one extra vertical half-plane of atoms above the slip plane. This extra half-plane terminates at the dislocation line, and is compressed there, as shown in the figure. The distortion at the dislocation line is local. The simple cubic arrangement of atoms is essentially restored a few atom spacings away from the dislocation line. The influence of the dislocation on the atomic configuration rapidly decays into a small displacement that decreases in magnitude with the inverse cube of the distance from the dislocation line. The local distortion near the dislocation line (or dislocation core) is indicated in the figure.

In principle, the Burgers vector of a crystal dislocation can be any lattice vector; for example, it is geometrically possible for an edge dislocation to be the termination of any integral number of lattice planes. In reality, however, the Burgers vector is almost invariably equal to the shortest lattice vector in the crystal. The reason is that the energy per unit length of dislocation line, which is called the *line energy*, or, in a slightly
different context, the *line tension* of the dislocation, increases with the square of the magnitude of \( \mathbf{b}, |\mathbf{b}|^2 \). (While we shall not prove this, it is obvious from Fig. 4.10 that the local distortion of the crystal would increase dramatically if two or more extra half-planes terminated at the dislocation line.) Let the Burgers vector, \( \mathbf{b} \), be the vector sum of smaller lattice vectors. \( \mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2 \). Unless \( \mathbf{b}_1 \) and \( \mathbf{b}_2 \) are perpendicular, \(|(\mathbf{b}_1 + \mathbf{b}_2)|^2 > |\mathbf{b}_1|^2 + |\mathbf{b}_2|^2| \), and the dislocation can decrease its energy by splitting into two or more dislocations that have smaller Burgers vectors.

### 4.3.3 The Burgers circuit

While it is always possible to find the Burgers vector, \( \mathbf{b} \), of a dislocation by determining the slip that would be required to make it, this is often inconvenient. A simpler method uses a geometric construction known as the Burgers circuit.

To construct the Burgers circuit, choose a direction for the dislocation line and draw a clockwise closed circuit in the perfect crystal by taking unit steps along the lattice vectors. An example is drawn for a \{100\} plane in a simple cubic crystal in Fig. 4.11. If the same circuit is drawn so that it encloses a dislocation, it fails to close. The vector (from the starting position) that is required to complete the circuit is the Burgers vector, \( \mathbf{b} \), of the dislocation, and measures the net displacement experienced by an imaginary observer who completes a circuit around the dislocation that would be closed in a perfect crystal.

![Burgers circuit](image)

**Fig. 4.11:** A Burgers circuit closes in a \{100\} plane of a cubic crystal, but fails to close by the Burgers vector, \( \mathbf{b} \), when the same circuit encloses an edge dislocation.

### 4.3.4 Motion of an edge dislocation: glide and climb

The reason that dislocations control the plastic deformation of crystalline solids is that it is relatively easy to move dislocations to produce shear deformation of the sort that is pictured in Fig. 3.9. It would be enormously difficult to shear a crystal by forcing the glide of rigid planes of atoms over one another; one would have to force the simultaneous reconfiguration of every crystal bond that crossed the slip plane. The same result is more easily achieved by moving dislocations stepwise through the crystal. Stepwise
dislocation motion requires a much smaller force since each elementary step can be accomplished by reconfiguring only the bonds that neighbor the dislocation line.

The stepwise motion of an edge dislocation in a simple cubic crystal is illustrated in Fig. 4.12. In order for the dislocation to move one lattice spacing to the right it is only necessary to break the bond indicated by the long dash in 4.12a and establish the bond indicated by the short dash. The new configuration is shown in Fig. 4.12b. Of course one bond must be broken for each plane through which the dislocation threads, so a significant force is still required. But the force is small compared to that required to slip the upper part of the crystal as a rigid body. If the dislocation moves through the crystal in a sequence of individual steps like that shown in Fig. 4.12 it causes a net slip of the material above its plane of motion by the Burgers vector, \( \mathbf{b} \), and hence causes a rigid displacement of the whole upper part of the crystal.

![Fig. 4.12: Glide of an edge dislocation.](image)

The type of motion that is illustrated in Fig. 4.12 is called dislocation glide, and is relatively easy to accomplish. However, an edge dislocation cannot glide in an arbitrary direction. It can only glide in a particular plane, the slip plane or glide plane, which contains both the Burgers vector and the dislocation line.

![Fig. 4.13: Climb of an edge dislocation.](image)

When an edge dislocation moves out of its glide plane its motion is called climb. The climb of a dislocation is difficult at ordinary temperatures since it requires that atoms be absorbed on or liberated from the extra half-plane of atoms that defines the dislocation
line. The climb of an edge dislocation is illustrated in Fig. 4.13. The mechanism is slightly different depending on whether the dislocation moves up, which contracts the extra half-plane, or down, which extends it.

If the dislocation climbs up atoms must be liberated from the edge of the extra half-plane. Since the number of atoms is conserved, this requires the absorption of vacancies from the lattice. One vacancy is needed per plane the dislocation threads. If the dislocation climbs down it must add atoms to the extra half-plane, and can only do this by liberating one vacancy per plane into the matrix, as shown in Fig. 4.13b. Both processes are difficult except at high temperature when, as we shall see, the equilibrium concentration of vacancies is high and the exchange of vacancies and atoms is relatively easy.

Because of the difficulty of climb at ordinary temperatures the plastic deformation of real crystals tends to occur through the motion of dislocations on well-defined planes that are the glide planes of the active dislocations. Under a microscope one can often see discrete slip steps on the surface of a crystal that has been deformed. These result from the glide of many dislocations on closely spaced, parallel planes. At high temperature climb becomes possible and the slip planes are less well-defined. When this happens the strength of the crystal (its resistance to plastic deformation) decreases dramatically. For this reason most solids are relatively soft at high temperature.

### 4.3.5 Screw dislocations

Our discussion to this point has dealt only with edge dislocations, that is, dislocations in which the Burgers vector is perpendicular to the dislocation line. Dislocations in real crystals rarely have a pure edge character. Their Burgers vectors lie at various angles to their line directions. In the extreme case the Burgers vector is parallel to the dislocation line, which is the characteristic of a screw dislocation. A screw dislocation is difficult to visualize in a crystal, but can be created by a method suggested by Volterra that closely resembles the way the edge dislocation was formed. A screw dislocation of the general Volterra type is shown in Fig. 4.14.

![Fig. 4.14: A method for forming a screw dislocation in a solid.](image-url)
To introduce a screw dislocation we slice the solid part-way through in the direction of its width, as shown in Fig. 4.14. But instead of displacing the material above the cut toward the dislocation line we displace it by a vector, \( \mathbf{b} \), that lies parallel to the dislocation line, as shown in the figure. The direction of the force required to do this is also indicated. The material is then re-welded so that it is continuous across the plane of slip. The residual distortion is concentrated at the dislocation line, which then constitutes an isolated linear defect.

If the body shown in Fig. 4.14 is crystalline then the cut surface shown in the figure is a plane of atoms. In order for the crystal to be continuous across the slip plane after it is rejoined, the displacement, \( \mathbf{b} \), must be such that this plane of atoms joins continuously onto a crystallographically identical plane. It follows that a closed circuit (Burgers circuit) that encloses a screw dislocation not only fails to close, but produces a net translation by \( \mathbf{b} \) along the dislocation line, where \( \mathbf{b} \) is a lattice vector. A circuit that starts on one plane of atoms finishes on another a distance \( \mathbf{b} \) below. Continuing the circuit causes a displacement by \( \mathbf{b} \) at each revolution, without the circuit ever leaving the atom plane. The effect of a screw dislocation is to join a set of parallel atom planes so that they become a single plane like one that would be created by extending a plane outward from the thread of a screw. Hence the name: screw dislocation.

As in the case of an edge dislocation the line energy of a screw dislocation is proportional to the square of its Burgers vector. Hence the Burgers vector of the screw dislocation is ordinarily the smallest lattice vector that is compatible with the direction of its line.

### 4.3.6 Screw dislocations and plastic deformation

A screw dislocation differs from an edge not only in its geometry but in the way it accomplishes plastic deformation. The most important qualitative differences concern its direction of motion under an applied force and its relative freedom of movement.

Figure 4.14 suggests the connection between slip and dislocation motion for a screw dislocation. As the screw dislocation is displaced through the width of the body the material above its plane is slipped in the direction of the Burgers vector, hence along the length of the body. It follows that the longitudinal force shown in the figure acts to drive the screw dislocation sideways. If a screw dislocation is passed through the full width of the body it causes the shear shown in Fig. 3.8, which is the same as that caused by the passage of an equivalent edge dislocation through the length.

In contrast to an edge dislocation, a screw dislocation can glide in any plane. Since the Burgers vector lies parallel to the dislocation line both are in any plane that contains the dislocation line, and the screw dislocation can move in any direction perpendicular to its line.
4.3.7 Dislocations in real materials: mixed dislocations

Dislocations in real materials are most commonly neither pure edge nor pure screw in their character, but are mixed dislocations whose Burgers vectors lie at an intermediate angle to the local direction of the dislocation line. Because of the way in which they interact with other elements of the microstructure (which we shall discuss at a later point) dislocation lines are ordinarily curved. Since the dislocation bounds a region that has been slipped by the Burgers vector, \( \mathbf{b} \), the Burgers vector is the same at every point on the dislocation line. Hence the character of a curved dislocation changes continuously along its length. For this reason it is often most useful to think of a dislocation as the boundary of a surface over which the crystal has slipped rather than as a defect with a particular local atomic configuration.

![Dislocation loop in a crystal](image)

Fig. 4.15: A dislocation loop in a crystal.

This point is illustrated by the dislocation loop shown in Fig. 4.15. Dislocation loops are relatively common features of the microstructures of structural alloys. The loop shown in the figure is created by a process that is equivalent to a slip of the cylinder of material above the loop by the vector \( \mathbf{b} \). The dislocation loop is the boundary of the surface on which slip has occurred. The character of the dislocation changes continuously around the loop. The dislocation has pure edge character at the extremities of the loop along the length of the body, and has pure screw character at the extremities along the width.

The qualitative nature of the force on a curved dislocation, such as the loop in Fig. 4.15, can also be understood by regarding it as the boundary of the slipped region. If the force on the body is oriented to shear the body by the vector \( \mathbf{b} \) then the dislocation will tend to move so that it increases the area of slip; the loop will expand. If the force acts to shear the body in the direction \(-\mathbf{b}\), the dislocation will tend to move so that it shrinks the slipped area; the loop will contract.

Whether edge, screw or mixed in character, all dislocations have three important properties. First, they bound surfaces that divide the crystal so that the part above the surface is slipped by the vector, \( \mathbf{b} \), with respect to that below. Second, the slip on the surface bounded by the dislocation is constant and equal to a lattice vector. Hence the Burgers vector, \( \mathbf{b} \), is a lattice vector and has the same value at every point on the dislocation line. Third, since the slip is constant, a dislocation cannot simply end within a
material; there is always a boundary between a slipped and an unslipped area. The dislocation can terminate at a free surface, as in Fig. 4.7 and 4.14, it can close on itself, as in Fig. 4.15, or it can end at a junction with other dislocations that bound surfaces over which the magnitude or direction of slip is different.

4.3.8 Partial dislocations

In the original concept of Volterra, the Burgers' vector of a crystal dislocation is a translation vector of the crystal, that is, a vector that connects atom positions so that the crystal can be rejoined perfectly across its glide plane. Such a dislocation is called a total dislocation. The element of slip it induces carries an atom in the glide plane from one atom position to another. Only total dislocations can be true two-dimensional defects. If the Burgers' vector were not a translation vector of the crystal the slipped surfaces would not weld perfectly together across the glide plane and a planar, surface defect would be left behind whenever the dislocation moved.

While most crystal dislocations are total dislocations when viewed from sufficiently far away, it is not uncommon to find them dissociated locally into a configuration that can be described as two parallel partial dislocations connected by a planar defect that is called a stacking fault in the crystal. The prototypic example is found in FCC crystals.

Fig. 4.16: Illustration of the slip of close-packed planes. The dashed arrow shows slip by a total dislocation, \( \mathbf{b} = \frac{1}{2} \{ \bar{1}01 \} \), which carries B→B. The solid arrows show the competing process of successive slip by two partial dislocations. The shaded atom in the cell is the intermediate, C-site position reached after partial slip.

The common dislocation in the FCC structure is the dislocation that causes close-packed planes to slip over one another. The close-packed planes are \{111\} planes, and the Burgers' vectors of the dislocations that induce slip on these planes are the \( \frac{1}{2} <110> \) vectors that connect atoms to their nearest neighbors in the \{111\} planes. Fig. 4.16 illustrates the path by which an FCC crystal is slipped by one atom spacing in the (111) plane. The slip vector is \( \frac{1}{2} \{ \bar{1}01 \} \), and is the slip caused by a dislocation with Burgers' vector \( \mathbf{b} = \frac{1}{2} \{ \bar{1}01 \} \). Assuming that the glide plane of the dislocation lies between (111) planes of A
and B-type atom sites, the element of slip carries an atom from one B-site to another, as shown in the figure. For clarity, the figure is drawn twice. The drawing on the left shows the displacement in a stacking of close-packed planes (A,B). The drawing on the right places the A and B planes in an FCC unit cell.

Direct slip from one B position to another is relatively difficult, as one can easily show by trying to slide two close-packed arrays of balls over one another. The atoms (balls) in the B-plane must ride over the A atoms, as illustrated in the figure. It is much easier to accomplish the slip in two sequential steps, as indicated in the figure, that circumvent the A atoms. The B atoms are first slipped into C positions, then moved from C back to B again. The slip can be accomplished by the sequential passage of two dislocations with the Burgers’ vectors \( b_1 \) and \( b_2 \). However, \( b_1 \) and \( b_2 \) are not lattice vectors; they are examples of partial dislocations. Their sum is the total dislocation, \( b \).

![Diagram of partial dislocations](image)

**Fig. 4.17:** Configuration of atoms in the (111) plane showing the partial and total slip vectors.

To find the Burgers’ vectors of the partial dislocations let the close-packed plane be the (111) plane, as shown in the right-hand figure in 4.16. The atom shift is to the shaded atom position, followed by a shift to the final atom position in the face center. The (111) plane of B-sites in the unit cell in Fig. 4.16 is redrawn in Fig. 4.17. As shown in this figure, the Burgers’ vector of the total dislocation is \( b = \frac{1}{2} [101] \). The first partial slip is from the B site at the corner of the unit cell to the nearest C-site in the plane. This slip is 1/3 of the vector connecting B-sites along the height of the triangular configuration of atoms in the plane, which is \( \frac{1}{2} [1\bar{1}2] \). Hence

\[
b_1 = \frac{1}{6} [1\bar{1}2] \quad 4.2
\]

The second partial slip is from the C-site to the final B-site position. As shown in the figure, this vector is parallel to the vector \( \frac{1}{2} [211] \), and is 1/3 of its length. Hence

\[
b_2 = \frac{1}{6} [211] \quad 4.3
\]

The vector sum of the two partial dislocation vectors is
\[ \mathbf{b}_1 + \mathbf{b}_2 = \frac{1}{6} [\mathbf{1} \mathbf{1} 2] + \frac{1}{6} [\mathbf{2} \mathbf{1} 1] = \frac{1}{2} [\mathbf{1} 0 1] = \mathbf{b} \]  

4.4

Splitting the total dislocation, \( \mathbf{b} \), into the partials, \( \mathbf{b}_1 \) and \( \mathbf{b}_2 \), not only facilitates slip, but also lowers the total energy. The energy per unit length of dislocation line is proportional to \( |\mathbf{b}|^2 \). Since

\[ |\mathbf{b}|^2 > |\mathbf{b}_1|^2 + |\mathbf{b}_2|^2 \]

4.5

the energy of a parallel pair of partial dislocations, \( \mathbf{b}_1 \) and \( \mathbf{b}_2 \), is less than that of the total dislocation, \( \mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2 \).

![Diagram of dislocation](image)

**Fig. 4.18:** The total dislocation, \( \mathbf{b} \), divided into partial dislocations, \( \mathbf{b}_1 \) and \( \mathbf{b}_2 \) separated by a stacking fault.

However, to understand the configuration that appears when the dislocation is split we must recognize that eq. 4.5 gives only part of the energy of the pair of partials. Let the total dislocation divide, as illustrated in Fig. 4.18. In the area between the partials, which is shaded in the figure, the region above the slip plane has been slipped by the vector \( \mathbf{b}_1 \), but has not yet experiences the slip \( \mathbf{b}_2 \). Let the slip plane lie between A and B-planes, as illustrated in Fig. 4.16. The partial slip \( \mathbf{b}_1 \) carries the atoms in B-sites into C-sites, and causes corresponding changes C \( \rightarrow \) A and A \( \rightarrow \) B in successively higher planes. The FCC crystal has the stacking ..ABCABC.. But the stacking encountered in a direction perpendicular to the glide plane that passes through the shaded area between the partials is ..ABCAJCABC.., where the slash indicates the glide plane. The partial dislocation perturbs the sequence of close-packed planes; it introduces a planar defect called a *stacking fault* on the glide plane between the two partials.

A stacking fault has a positive energy, \( \sigma_s \), per unit area. The separation of the two partials, \( \mathbf{b}_1 \) and \( \mathbf{b}_2 \), creates a stacking fault whose area increases as the partials move apart. There is, therefore, an equilibrium separation between the partial dislocations. They separate to decrease the energy per length of dislocation, but are attracted by the need to minimize the stacking fault energy. The balance of these two effects determines the equilibrium separation, which is of the order of 5-500\( \text{Å} \) in typical FCC crystals.
Similar considerations apply to HCP and diamond cubic crystals. Total dislocations in the close-packed planes tend to divide into partials separated by ribbons of stacking fault. In BCC crystals the division of total dislocations into partials is much less pronounced. Separated partials cannot ordinarily be resolved in BCC, even through high resolution electron microscopy. Their presence is inferred from features of the mechanical behavior that are most easily explained by assuming that total dislocations dissociate slightly.

4.4 TWO-DIMENSIONAL DEFECTS: INTERFACES

The two-dimensional defects that appear in crystals can be usefully divided into three types: free surfaces, which are the external surfaces at which the solid terminates at a vapor or liquid, intercrystalline boundaries, which separate grains or distinct phases within the solid, and internal defects that disrupt the crystalline pattern over a surface within a crystal. All of these defects have two important characteristics. First, since they are surfaces in a crystal they have particular atomic structures that depend on orientation. Second, they have a positive energy. The energy per unit area is ordinarily equal to the surface tension, \( \sigma \), of the interface.

4.4.1 Free surfaces

The interface between a crystalline solid and a vapor or liquid governs the interaction between the two phases, and influences the behavior of the solid in many important ways that we shall discuss later in the course. It also affects the shape of the solid. The shape and properties of the free surface are determined by a combination of its structure and its energy. These two parameters are related to one another.

Fig. 4.19: Solid surface showing segregated solutes and adsorbates.
The crystal structure of the solid extends up to the free surface. The structure of the surface is derived from the pattern of atomic packing on the crystal plane that lies parallel to the surface, and the behavior of the surface can often be inferred from that pattern. However, the structure of the surface always differs at least a bit from that of a crystal plane that parallels it within the bulk crystal. Even when the crystal sits in a perfect vacuum, the atoms in the free surface are bonded on only one side and the spacing and the configuration of the atoms in the first few planes at the interface adjusts to accommodate the asymmetry of the bonding in the best possible way. The structure of a real interface is further complicated by bonding interactions across the interface and by chemical changes in the interfacial plane. The dangling bonds at the interface provide favorable sites for the adsorption of atoms that do not fit well into the bulk lattice. The interface is often enriched in solute species (surfaceants). Oxygen, sulfur and phosphorous are common surfactants in engineering solids.

Fig. 4.20: A surface of overall orientation \((hkl)\) made up of close-packed segments in steps.

The interfacial tension of the free surface depends on its structure, and is determined, in large part, by the number of bonds that are broken because the crystal is not continued on the far side of the interface. As a consequence the interfacial tension, \(\sigma\), is a function of the orientation of the surface. In a typical metal, the interfacial tension is a relatively weak function of the orientation. Because metallic bonding is relatively non-directional, the interfacial tension of a typical metal tends to vary with the density of packing in the interface and is minimized for surface orientations that parallel close-packed planes (\(\{111\}\) in FCC, \(\{110\}\) in BCC). Covalent solids have fewer neighbors, and more directional bonds, and are much more strongly anisotropic in their surface tension.

The external surfaces of metallic and covalent solids tend to parallel the planes of closest atomic packing, as illustrated in Fig. 4.19. This tendency also influences the atomic structures of interfacial planes that have other orientations. An interface that does not parallel a close-packed plane often has a structure in which flat, close-packed segments form ledges of atomic height that give a stepped structure with the proper overall orientation, as illustrated in Fig. 4.20. While the stepped morphology increases
the surface area, the overall surface tension is lower because of the relatively low surface tension of the close-packed steps.

![Diagram showing NaCl structure](image)

*Fig. 4.21:* The NaCl structure, showing the electrical neutrality of the \{100\} plane and the preferential charge on the \{111\} plane.

The interfacial tension of a surface in an ionic solid is strongly affected by the distribution of electric charge. In strongly ionic materials (for example, NaCl) the electrically neutral surfaces (\{100\} in NaCl) have much lower interfacial tensions than the charged surfaces (such as \{111\} (Fig. 4.21)). The relative interfacial tensions of charged and uncharged surfaces can, of course, be altered by the addition of surfactants, which are often preferentially attracted to the charged surfaces.

*The shapes of crystals*

The preferred shape of a crystal that is bounded by free surfaces is the one that minimizes its total interfacial energy. The macroscopic surface of the crystal can always be treated as if it were made up of facets, the \(k\)th one of which has area, \(A_k\), and interfacial tension, \(\sigma_k\). The energy is, then,

\[
E = \sum_k \sigma_k A_k
\]

When the crystal is a simple metal the surface tension is nearly isotropic. The preferred shape is nearly spherical, since a spherical shape minimizes the surface area for given volume. When the crystal is ionic or strongly covalent there are crystal planes whose tensions are, relatively, very low. To minimize the total interfacial energy, ionic and covalent solids tend to assume polygonal shapes with large, flat facets parallel to the low-energy planes.

This is the physical reason that compound solids and minerals often look crystalline, while metals usually do not. The natural macroscopic shapes of such compounds are polygonal (crystalline) because of the strong anisotropy of the surface tension. You can readily verify, for example, that the grains of common table salt have a cubic shape; this shape reflects the low energy of the neutral \{100\} surfaces in the NaCl structure. Many common minerals have hexagonal structures in which the prismatic planes (the
vertical planes in the hexagonal unit cell) are electrically neutral. These materials tend to grow as well-shaped hexagonal prisms. Simple metals, on the other hand, are more isotropic in their interfacial tensions, and tend to form crystals that are cylindrical or spherical in their macroscopic shapes.

The shapes of crystals can be altered and controlled by adding selective chemical surfactants that modify the surface tensions of particular planes. For example, NaCl can be made to crystallize in an irregular particle with mainly {111} faces by adding a suitable surfactant.

**Solid-liquid interfaces**

The interface between a crystalline solid and a vapor is usually sharp. The surface is a crystal plane or stepped layer of planes. Interfaces between solids and liquids, on the other hand, are often diffuse. There is a high density of atoms in the liquid near the solid surface, and these may have a local arrangement that does not differ that greatly from the crystalline pattern in the solid. In these cases the tension of the solid-liquid interface is a weak function of surface orientation, and the solid adopts a diffuse interfacial structure like that shown in Fig. 4.22.

![Possible configuration of a diffuse interface.](image)

**Fig. 4.22:** Possible configuration of a diffuse interface.

The possibility of forming a diffuse interface is important in the solidification of solids from the melt and the growth of crystals from the melt. Atoms of the liquid can be easily incorporated into a diffuse interface, with the consequence that a solid that has such an interface can solidify easily and rapidly.

### 4.4.2 Interfaces between crystals

The most important internal interfaces in a typical solid are those that separate the individual crystallites it contains. Almost all crystalline solids are polycrystalline. They are composed of many small crystallites (grains) of the primary constituent, and may also contain several distinct constituents (phases) of different composition and crystal structure. The interfaces that separate grains are called grain boundaries. Those that separate phases are called two-phase interfaces.
Grain boundaries

The structure of a grain boundary depends on the misorientation of the crystal grains that it separates. When the misorientation is small the boundary, which is then called a low-angle boundary, is a reasonably simple planar array of dislocations. When the misorientation is larger the boundary structure is more complicated, and it is often useful to picture the boundary as simply a region of disorder between the two crystal grains.

![Schematic drawing of polygranular grain structure showing grain boundaries and two-phase interfaces.](image)

Fig. 4.23: Schematic drawing of polygranular grain structure showing grain boundaries and two-phase interfaces.

The structure of a low-angle tilt boundary is shown in Fig. 4.24. A tilt boundary is a vertical array of parallel edge dislocations. When a tilt boundary is introduced into a single crystal it creates a misorientation by the tilt angle, \( \theta \), between the volumes of crystal on either side. If the dislocations in the boundary have Burgers' vector, \( \mathbf{b} \), and are separated by a vertical distance, \( d \), then the tilt angle, \( \theta \), is given by

\[
\theta = \tan^{-1} \left[ \frac{|\mathbf{b}|}{d} \right]
\]

Fig. 4.24: A low-angle tilt boundary as a vertical array of parallel edge dislocations.

A second type of low-angle boundary that can also be generated by a simple array of dislocations is a twist boundary. If two sets of parallel screw dislocations are arrayed in a perpendicular pattern in a plane the result is a twist in the boundary plane. The crystal on one side of the plane is rotated with respect to that on the other by a twist angle.
about an axis perpendicular to the plane of the boundary. A twist boundary is not simple to draw, and we shall not attempt to do so here.

The structures of high-angle grain boundaries can be crudely understood by simply matching two crystals that are misoriented by the angle of the boundary. As illustrated in Fig. 4.25, this process ordinarily leads to a relatively poor match of crystal planes along the boundary. While there are relaxations in the boundary that reduce its disorder somewhat, a typical high-angle boundary is relatively disordered, with a high density empty spaces and open channels. High-angle grain boundaries provide easy paths for atom diffusion in the solid state. They also provide preferential sites for solutes and impurities that do not fit very well in the bulk crystal, and adsorb as surfactants on the boundary.

![High-angle grain boundary](image)

Fig. 4.25: A possible configuration of a high-angle grain boundary.

There are, however, certain high-angle boundaries whose misorientations provide a relatively good geometric match between the adjacent crystals. These are called coincidence boundaries. Their study is beyond the scope of this course.

The equilibrium shape of a grain

![Hexagonal grain structure](image)

Fig. 4.26: Ideal hexagonal configuration of a two-dimensional grain structure with constant $\sigma_B$. 
The interfacial tension of a grain boundary is often a relatively weak function of its orientation; for most purposes we can assume that it has the constant value, $\sigma_B$. When this is true a local equilibrium is achieved when grain boundaries meet in three-grain junctions in a configuration in which the angle between adjacent boundaries is $120^\circ$. When the grain structure is two-dimensional, as it is in a thin film one grain thick, then local equilibrium is achieved when the grains are hexagonal platelets, as in Fig. 4.26.

When the grain structure is three-dimensional a microstructure that is very near to local equilibrium can be built up of fourteen-sided polyhedra that are known as Thomson tetrakaidecahedra. The tetrakaidecahedron is shown in Fig. 4.27. It looks very complex but, in fact, has a very simple construction. If one takes a body centered cubic array of points and constructs planes that bisect the lines connecting a given point to its first and second nearest neighbors, the inner envelope of these planes is the tetrakaidecahedron. The tetrakaidecahedron is just the Wigner-Seitz cell of the BCC structure. It follows that tetrakaidecahedra stack in a body-centered cubic array. The BCC array of tetrakaidecahedra is sometimes useful as a model of a three-dimensional microstructure. However, its accuracy is limited; real polygranular solids are built up of irregularly shaped grains that have a distribution of sizes and an irregular pattern in space (as in Fig. 4.22). The grains in real microstructures are never quite in minimum-energy configurations. A somewhat better model of a real microstructure is a three-dimensional foam of the type that forms on beverages that are familiar to undergraduates.

![Fig. 4.27: (a) Thomson tetrakaidecahedron inscribed in a cube. (b) Packing of tetrakaidecahedra into a BCC array.](image)

The positive surface tension of the grain boundary has the consequence that the energy of a polygranular material increases in proportion to the total grain boundary area. Large grains have a smaller surface energy per unit volume than small ones. If a polygranular solid is held at high temperature where the atoms are relatively mobile then grains can grow or shrink by adding or losing atoms to adjacent grains. The large grains grow at the expense of the smaller ones to decrease the energy, which causes the mean grain size to increase with time.

The grain structure is usually characterized by its mean grain size, which is measured as the mean grain diameter. The mean grain size provides an incomplete
characterization of the grain structure; both the distribution of grain sizes and the grain shape are also important. These aspects of the grain structure are more difficult to measure quantitatively. In practice a characterization of the grain structure usually consists of a value for the mean grain size and a set of photomicrographs that present the geometrical features of the grain structure in visual form.

Two-phase interfaces

Two-phase interfaces have many of the properties of grain boundaries, including a specific structure and interfacial energy. The most important structural characteristic of a two-phase interface is its coherency. If it is possible to order the phases on both sides of the interface on the same crystal lattice the interface is said to be coherent. A two-dimensional example is given in Fig. 4.28. Coherent interfaces are common in real systems. When small precipitates form in the interior of a solid the precipitate phase that forms first usually involves only a slightly different atomic ordering on the lattice of the parent structure and has a coherent interface with it. In the modern science of thin-film electronics multi-layer structures are often deposited atom layer by atom layer from the vapor phase, to create a heterostructure of atom layers of different composition that are coherent with one another.

Fig. 4.28: Coherent, ordered precipitate in a two-dimensional crystal.

On the other hand, since two different crystalline phases almost never fit perfectly together, both are distorted if they are joined by a coherent interface. In the case of a small precipitate particle the lattice distortion that is caused by the coherent interface increases with the particle size, and provides an increasing driving force to destroy continuity at the interface. When a second-phase particle grows to large size it almost always loses coherency with the surrounding matrix. The most common mechanism causing the loss of coherency is the accumulation of dislocations at the interface.

The surface tension of a coherent two-phase interface is highly anisotropic since the two phases only fit well when they have a simple crystallographic relation to one another. The volumetric distortion of the precipitate also depends strongly on its shape and crystallographic fit. As a consequence, coherent second-phase particles such as small precipitates tend to have characteristic shapes and interfaces that lie almost entirely in a
particular crystallographic plane of the parent crystal, which is known as the **habit plane**
of the precipitate. The well-defined habit is lost when the precipitate is incoherent.

The positive interfacial tension of second-phase particles has the consequence that
the energy per unit volume decreases as the particle grows. Given a distribution of
precipitate sizes and a temperature high enough to permit atomic rearrangements the
larger particles grow at the expense of the smaller ones to decrease the energy, and the
second-phase particle distribution **coarsens**.

### 4.4.3 Interfaces within crystals

The surface defects that commonly appear in the interior of a crystal are of two
types: **stacking faults** and **antiphase boundaries**. We have already encountered the
stacking fault, which is a defect in the sequence of planes of atom positions in the crystal.
An antiphase boundary is a defect in the sequence of atom distributions over the planes.
Anti-phase boundaries can only occur in ordered compounds.

**Stacking faults**

The simplest examples of stacking faults (and the only ones that we shall study) are
found in close-packed structures. Using the notation introduced in the previous
chapter, a stacking of close-packed planes in the order ...ABCABC... generates the FCC
structure. A stacking fault occurs wherever this pattern is broken. For example, in a
crystal with the local stacking sequence ...ABCABC... the extra B plane in the middle
constitutes a stacking fault.

It is sometimes useful to distinguish three kinds of stacking fault. An **extrinsic**
stacking fault is one that can be created by inserting an extra plane of atoms into the
structure. In the ...ABC||ABC... pattern given above the plane marked || is an extrinsic
stacking fault. Extrinsic stacking faults appear, for example, after an irradiation
treatment creates many interstitialcies that subsequently gather together on close-packed
planes.

An **intrinsic** stacking fault is one that can be created by removing a plane of
atoms. In the sequence ...ABC||BCABC.. the defect marked || corresponds to a missing
A-plane, and is hence an intrinsic stacking fault. An intrinsic stacking fault is created by
the condensation of vacancies onto a close-packed plane. It is also created by a partial
slip of the crystal to the right of the fault that carries A-sites to B-sites, B-sites to C-sites,
and C-sites to A-sites. A partial slip of this type is caused by the motion of a partial
dislocation of the type discussed at the end of section 4.3.

The third kind of stacking fault is called a **twin boundary**, and is a surface that
separates two volumes of crystal that are mirror images of one another, as illustrated in
Fig. 4.28. In the sequence ...ABCAB|C|BACBA... the plane marked C is a twin
boundary; on either side the crystal has a perfect FCC structure, but the two FCC lattices
are mirror images of one another across the twin plane. Twin boundaries are most often
introduced during the growth of a crystal. If the crystal grows by adding planes to the right and if an extrinsic fault (B) occurs just to the right of ICl, the energy of the fault is usually minimized if the next plane is added at A, and so on, to maximize the fraction of proper FCC packing in the crystal. The formation of twin boundaries is also important in the plastic deformation of some materials, as we shall discuss later in the course.

Fig. 4.29: The relation of crystal planes across a twin boundary.

Note that if a stacking fault is repeated with perfect periodicity it does not constitute a fault, but creates a new crystal structure. For example, the sequences ABCAB, ABCABCAB, and ABABABAB are regular mixtures of FCC and HCP packing that generate new structures if they are repeated indefinitely. Such mixed structures occur in nature. They are usually found in materials that exhibit both FCC and HCP-type structures, which indicates that the stacking fault energy is relatively small. Mixed structures are particularly common in SiC, ZnS and CdS, all of which have alternate Č-ZnS (FCC) and β-ZnS (HCP) structures. The mixed structures are called polytypes, and can have very long repeat periods.

**Antiphase boundaries**

The internal defects known as antiphase boundaries only occur in ordered compounds. An example is given in Fig. 4.30, which illustrates a {110} antiphase boundary in a compound with the CsCl structure.

An ordered compound can always be regarded as a stacking of parallel planes, each of which has a particular substitutional pattern of the atomic species. Fig. 4.30a, for example, illustrates how the CsCl structure can be made by stacking {110} planes that alternate in the specific arrangement of the two types of atoms. An antiphase boundary is created when a plane with the wrong pattern appears, and subsequent planes follow it in their patterns. In Fig. 4.29b the third {110} plane has its atom positions shifted by \( \frac{1}{2} \langle 110 \rangle \) from the expected pattern, and is, hence, an antiphase defect. An antiphase boundary resembles an intrinsic stacking fault in that it can be created either by omitting one plane in the stacking pattern, or by a slip displacement at the antiphase plane. Mechanical slip is an important source of antiphase defects. Note that the antiphase defect in Fig. 4.30b is precisely that which would be created if a dislocation of Burgers vector \( \frac{1}{2} [110] \) passed through the plane. Since \( \frac{1}{2} [110] \) is a vector connecting nearest neighbor atoms in the closest-packed planes of the BCC structure, it is a common Burgers vector for dislocations in BCC.
Fig. 4.30: (a) The CsCl structure drawn as a stacking of \{110\} planes.  
(b) An antiphase boundary in the CsCl structure made by a displacement of the type \frac{1}{2} \langle110\rangle in the third plane.

4.5 **VOLUME DEFECTS**

Volume defects in crystals are three-dimensional aggregates of atoms or vacancies. It is common to divide them into four classes in an imprecise classification that is based on a combination of the size and effect of the particle. The four categories are: *precipitates*, which are a fraction of a micron in size and decorate the crystal; *second-phase particles* or *dispersants*, which vary in size from a fraction of a micron to the normal grain size (10-100\(\mu\)m), but are intentionally introduced into the microstructure; *inclusions*, which vary in size from a few microns to macroscopic dimensions, and are relatively large, undesirable particles that entered the system as dirt or formed by precipitation; and *voids*, which are holes in the solid formed by trapped gases or by the accumulation of vacancies.

Precipitates are small particles that are introduced into the matrix by solid state reactions. While precipitates are used for several purposes, their most common purpose is to increase the strength of structural alloys by acting as obstacles to the motion of dislocations. Their efficiency in doing this depends on their size, their internal properties, and their distribution through the lattice. However, their role in the microstructure is to modify the behavior of the matrix rather than to act as separate phases in their own right.

Dispersants are larger particles that behave as a second phase as well as influencing the behavior of the primary phase. They may be large precipitates, grains, or polygranular particles distributed through the microstructure. When a microstructure contains dispersants such properties as mechanical strength and electrical conductivity are some average of the properties of the dispersant phase and the parent.

Inclusions are foreign particles or large precipitate particles. They are usually undesirable constituents in the microstructure. For example, inclusions have a deleterious effect on the useful strength of structural alloys since they are preferential sites for
failure. They are also often harmful in microelectronic devices since they disturb the geometry of the device by interfering in manufacturing, or alter its electrical properties by introducing undesirable properties of their own.

Voids (or pores) are caused by gases that are trapped during solidification or by vacancy condensation in the solid state. They are almost always undesirable defects. Their principal effect is to decrease mechanical strength and promote fracture at small loads.
Chapter 5: Glasses

There's a race of men that don't fit in
A race that can't stand still
So they break the hearts of kith and kin
And they roam the world at will
- Robert W. Service, "The Men That Don't Fit In"

5.1 INTRODUCTION

As discussed at the beginning of Chapter 3, there are two idealized cases in which one can easily describe the positions of an arbitrarily large number of atoms or molecules: when they are distributed in a perfectly regular way, as they are in a crystalline solid, and when they are distributed in a random way, as they are in an amorphous solid or glass. In this chapter we discuss the microstructure of glasses.

The most familiar glasses are polymeric or silicate solids, whose structures will be discussed more fully in Chapter 6. However, in recent years it has been shown that many materials can be made to have approximately amorphous atomic arrangements. These often have unusual properties and, as a consequence, are the subject of increasing interest in materials research. Important examples include amorphous silicon and compound semiconductor films that have useful electrical or optical properties, amorphous metals that have unusual mechanical or magnetic properties, and amorphous coatings of several types that provide protection against corrosion or mechanical wear.

Most of the useful properties of amorphous materials derive from their isotropy (the fact that their properties are independent of direction), their fine-scale homogeneity, and their formability. Isotropy is important, for example, in optical glasses which must transmit light identically in all directions to maintain an undistorted image. Homogeneity is important, for example, in coatings that protect against corrosion, since material inhomogeneities create electrochemical couples that cause corrosion. Formability is important in plastics and silica glasses that are made into inexpensive containers.

5.2 THE GLASSY STATE

There is no known case in which a random distribution of atoms is the most stable configuration of a solid. Given the kinetic opportunity any solid will crystallize into an ordered configuration. It follows that a glassy structure is a non-equilibrium one.

The glassy structure is usually obtained by quenching the material (cooling it rapidly) from a liquid or gaseous state in which it is disordered. If the cooling is sufficiently fast the atoms or molecules are trapped in an irregular configuration that they do
not have sufficient thermal energy to escape. Glassy solids are made by deposition from the vapor onto a cold substrate or by quenching from the melt. Polymeric solids and silicates crystallize slowly; in these materials "quenching" may mean a reasonably slow cool from the molten state. Typical metals crystallize very quickly; in the processing of amorphous metals "quenching" usually means an extremely rapid cooling. This can be accomplished, for example, by pouring a thin film of the molten metal onto a chilled metal substrate in a process that is known as splat quenching or melt spinning, depending on precisely how it is carried out.

In principle any material can be made in an amorphous microstructure by cooling it sufficiently fast. In practice, it has proven very difficult to achieve amorphous states in simple solids such as the pure metals. Even if it were possible to make an amorphous structure of, say, copper, the achievement would probably be technologically uninteresting since at normal temperatures the amorphous material would crystallize rapidly. The amorphous solids that are of interest are those that can be made in the glassy state and retained in that state long enough to be processed and used in engineering devices.

The systems that form useful glasses can be divided into two classes. The first class includes molecular solids such as the long-chain polymers whose complex molecular geometry makes crystallization difficult. We shall call these configurational glasses since their amorphous structure is due to the physical difficulty of reconfiguring them into a crystalline pattern. The second class includes solutions and compounds, such as the silicates, whose local bonding inhibits crystallization. We shall call these chemical glasses.

### 5.3 CONFIGURATIONAL GLASSES

The prototype materials that are glass-formers because of their complex molecular geometry are the long-chain polymers, whose structures we shall discuss in the following chapter. Long-chain polymers are linear molecules that contain hundreds to thousands of atoms. They may be imagined as long, flexible strands that have a relatively weak polar interaction with one another.

In the liquid state the molecules are disorganized and tangled, in a configuration that resembles spaghetti in boiling water. Thermal agitation keeps the chains separated and permits them to slide over one another so that the material flows as a liquid under relatively small loads. As the liquid polymer is cooled the disordered arrangement is retained, but the intensity of thermal agitation decreases so that it is more difficult for the chains to move past one another. This is reflected in an increase in the viscosity, or resistance to flow, as the temperature drops. At a sufficiently low temperature, which is called the glass transition temperature, the chains are effectively trapped into disordered configurations that are maintained over time. The material begins to behave as a solid. The viscosity increases dramatically at the glass transition point.
The thermal behavior of a typical glass-forming polymer is illustrated in Fig. 5.1, which is a plot of volume against temperature. If the material were cooled so slowly that it could always achieve its most stable configuration it would crystallize, with a discontinuous change in volume, at its melting temperature $T_m$. If it is cooled too quickly for crystallization to occur it remains liquid, with increasing viscosity, until the lower temperature, $T_g$, the glass-transition temperature. At the glass transition the molecular configuration is frozen in whatever arrangement it happens to have. Hence there is no change in the specific volume. The discontinuity that indicates the glass transition is in the slope of the $v$-$T$ curve. Below $T_g$ the molecules in the frozen configuration effectively touch one another; the slope of the $v$-$T$ curve, the coefficient of thermal expansion, is discontinuous there.

![Graph](image)

Fig. 5.1: Plot of specific volume ($v$) against temperature ($T$) for a material that can form a glass or crystal.

The glass transition temperature, $T_g$, is not a true material property. It depends on the precise configuration of the molecules in the liquid, and is, hence, lowered by any change in the liquid that makes that arrangement more regular. Decreasing the cooling rate generally decreases $T_g$ because the chains have time to achieve a partial alignment. $T_g$ is also suppressed if the viscous liquid is subjected to a process that partially aligns the polymer chains, such as passing it through a die. If the cooling is sufficiently slow or the chains are mechanically aligned then the material may form a mixed state in which small crystalline regions sit in a glassy matrix.

The more complicated the geometry of a long-chain polymer the more easily it can be made and retained as a glass. Glass-formation is promoted by virtually any geometric irregularity that makes it more difficult to arrange the molecules in a crystalline pattern. Bulky sidegroups on the polymer chain, irregular placement of sidegroups, chemical defects, and cross-links between polymer strands all complicate the geometry of the polymer and promote the formation and retention of the glassy structure. Glass formation can also be enhanced by making a solution or mixture of different polymers or by mixing inorganic fillers into the polymer, since both processes inhibit local alignment into a crystalline pattern.
Most configurational glasses are thermoplastic; they soften and become very easy to deform and shape when they are heated to a temperature just above $T_g$. At this temperature the material is a very viscous liquid. Its high viscosity has the consequence that it flows very slowly under gravity or small mechanical loads, and hence retains its shape in normal handling, at least for a short period of time. However, the viscous glass deforms easily under larger mechanical loads and can be formed into complex shapes. Note that in order for a material to be thermoplastic the glass transition must be reversible. The material must not crystallize spontaneously on heating. Chain polymers of complex geometry are very difficult to crystallize, and hence have this property.

5.4 CHEMICAL GLASSES

The second class of glass-forming materials includes those that are difficult to crystallize because of the complexity of bonding. There are several distinct types of these chemical glasses that exhibit different structural characteristics and thermal behavior.

5.4.1 Network oxides

The prototype example of a chemical glass is silica glass, whose structure is described in the following chapter as an important example of a network oxide. Like many such oxides, silicon dioxide (silica) has a relatively open structure. The silica structures are based on the $\text{SiO}_4^{-4}$ silicate tetrahedron that is shown in Fig. 5.2. A silicon atom is tetrahedrally coordinated by four oxygen atoms. Each of the oxygen atoms has an unsatisfied bond that can be used to join it to the Si atom in the center of another tetrahedron, in which case the oxygen atom is effectively a member of two adjacent tetrahedra that are joined at their corners. If all the oxygen atoms are bonded in this way the solid has the stoichiometric formula $\text{SiO}_2$. Alternatively, the oxygen atom may bond to some other chemical species, for example, a cationic impurity such as Na or Mg.

![Fig. 5.2: The SiO$_4$ tetrahedron.](image)

In the three-dimensional crystalline form of SiO$_2$, quartz, the silicate tetrahedra are joined at their corners in a perfectly regular pattern to create a regular three-dimensional network (hence the name, network oxide). However, this is a very open
structure, and is consequently difficult to form. Moreover, there are several distinct crystal modifications of quartz that differ in the exact spatial arrangement of the silicate tetrahedra, which suggests that the tetrahedra can be reoriented with respect to one another without a substantial energy change.

![Schematic illustration of the disordered structure of pure silica glass. The symbols (O) represent O atoms, (•) represent Si.](image1)

Fig. 5.3: Schematic illustration of the disordered structure of pure silica glass. The symbols (O) represent O atoms, (•) represent Si.

In the liquid the tetrahedra tend to be preserved in a local sense, but the tetrahedra are rotated about the Si-O-Si bond to create an irregular packing that has a higher density. Even when silica is chemically pure (SiO$_2$) the irregular structure is preserved when the melt is cooled quickly. Bonds must be broken and restored to establish the crystal structure. This is relatively difficult to do, and the thermodynamic driving force that impels it is not great since the ordered and disordered structures have very similar local configurations. The result is that quenched silica forms a glass that has a disordered structure somewhat like that illustrated in two dimensions in Fig. 5.3.

![Illustration of a glassy structure formed when ionic species (cross-hatched) disrupt the bonding configuration in silica.](image2)

Fig. 5.4: Illustration of a glassy structure formed when ionic species (cross-hatched) disrupt the bonding configuration in silica.

The formation of glassy structures in silica is promoted by the addition of glass-forming additives. These include other oxides, such as PbO$_2$, that participate in the silica network with a slightly different bonding configuration, and ionic species, such as Na, that replace Si by completing the oxygen bonds to yield a structure somewhat like that drawn in two dimensions in Fig. 5.4. The glass-forming additives disrupt the pattern of
silicate tetrahedra and must be rejected to form a quartz crystal. When the material is cooled quickly they are trapped in the solid and prevent crystallization.

Silica glasses have a glass-transition behavior very much like that of a configurational glass. Fig. 5.1 applies to the silica glasses as well. They are generally thermoplastic. At temperatures just above Tg it becomes relatively easy to reorient silica tetrahedra and break and re-establish local bonds, so the material behaves as a very viscous liquid. If the material contains a sufficient concentration of glass-forming additives it is reasonably stable against crystallization and can hence be heated, formed and cooled without losing the glassy state.

5.4.2 Cross-linked polymers

A second type of chemical glass includes the cross-linked polymers, which we shall discuss further in Chapter 6. These materials are made of polymeric molecules that are chemically bonded to one another by cross-link bonds, in which an atom or group of atoms simultaneously bonds to more than one polymer chain. The cross-links usually form in a random pattern when the polymer is reacted with a bonding agent. The cross-links set the relative positions of the molecules. While it would be energetically favorable for the cross-links to reconfigure in a periodic arrangement, this would require that the cross-links be broken and re-established, which is a negligibly slow process at moderate temperature if the cross-linking bonds are stable. Hence cross-linked polymers are frozen into an irregular distribution through space.

Since a transition to the liquid state would also require that the cross-link bonds be broken, typical cross-linked polymers do not melt. They are thermosetting polymers that remain solid until they decompose chemically. Cross-linked polymers are, hence, very difficult to form into any shape other than that which they assumed in the chemical reaction or deposition process that created them.

5.4.3 Amorphous semiconductors

A third type of chemical glass is represented by the amorphous structures of the semiconducting elements, such as Si, and the associated semiconducting compounds. Most semiconductors are amorphous when they are formed by deposition from the vapor onto a cold substrate. The atoms essentially stick where they strike the cold surface. Each atom finds the best available bonding configuration with its immediate neighbors. In this way a glassy structure with a quasi-tetrahedral pattern is built up. Since reconfiguration of the atoms is inhibited by relatively stable bonds to near neighbors, the amorphous structure can be retained at temperatures near room temperature for very long times. Amorphous semiconductors not thermoplastic; they crystallize on heating.

The amorphous semiconductors have electrical and optical properties that fall in an intermediate ground between the properties of semiconductors and those of metals. They have many applications in microelectronics and in optical devices such as
photocopying machines and solar cells. We shall discuss some of these later in the course.

5.4.4 Metallic glasses

A fourth type of glass whose amorphous state is due to complex chemical bonding includes most of the metallic glasses, which are actually alloys of metals with metalloid species such as boron, carbon, phosphorous and others. The metals crystallize in simple structures that have nearly close packing to maximize non-directional metallic bonding. The metalloids, on the other hand, seek covalent bonds with a particular geometric configuration. Because of the incompatibility in bonding between the metals and metalloids, the solubility of metalloid impurities in crystalline metals is quite low. However, they have significant solubility in the more open and irregular structure of the liquid phase, and can be trapped into the solid by quenching it.

Metalloid solutes distort the crystalline structure of the metal, as illustrated by the particular example shown in Fig. 5.5. If they are trapped in sufficient concentration by quenching, the distortion of the crystal lattice is so great that crystallization does not occur and the quenched metal has a glassy structure. It is usually necessary to quench very rapidly to retain the glass-forming elements in the matrix. Amorphous metals are usually made by rapid deposition of films on cold substrates or rapid quenching of thin sheets between high-conductivity copper rollers that act as heat sinks.

Fig. 5.5: Distortion of a hexagonal packing of atoms by substitutional impurities that prefer three-fold coordination.

Metallic glasses can also be made by direct implantation of metalloid impurities into the metal. In this method an energetic beam of atoms or ions of the metalloid species is directed onto the metal surface. Because of their high kinetic energy, the metalloid atoms penetrate into the metal and become trapped in its lattice, causing severe local distortions. If a sufficiently high concentration of the metalloid is built up in this way, the metal lattice becomes so badly distorted that it loses its crystalline pattern and becomes glassy. Implantation is a particularly useful way to produce amorphous surface layers on structural metals.

The common metallic glasses are not thermoplastic. They tend to crystallize on heating, usually after rejecting the glass-forming solutes into precipitate phases. However, many of them are sufficiently stable to be technologically useful as amorphous
sheets or coatings. Important applications include soft magnets for electrical generators and wear and corrosion-resistant coatings for structural metals.
Chapter 6: Macromolecular Solids

I've got one word of advice for you, kid:
"plastics"
- helpful inebriate in the film "The Graduate"

6.1 INTRODUCTION

In this chapter we discuss a class of solids whose microstructures tend to fall in the gray area between the crystalline solid and the glass. In these solids, virtually every atom has a well-ordered atomic environment that constitutes a molecular building block. The building blocks may be repeated through space in a network that forms a well-defined crystal lattice, they may be assembled into sheets or chains, or they may be simply jumbled together to form a glass. Since the most general characteristic of the solid is the presence of a definable molecular unit that is repeated in some pattern, we shall call these materials macromolecular solids.

The class of macromolecular solids includes polymeric solids, which range from biological materials to commercial plastics, silicate materials, including many natural minerals and synthetic glasses, and artificial fibers, which are used in optical devices and structural composites. It also includes a newly discovered, intermediate type of structure called a quasicrystal. Each type of macromolecular solid is obtained by assembling a repeatable molecular element into a macromolecule. In some cases, including most polymeric solids, the macromolecule then becomes the building block for a crystalline or glassy arrangement of molecules. In other cases, including the three-dimensional network silicates and the quasicrystals, the macromolecule is the solid.

6.2 POLYMERS

As the name implies, a polymer is a molecule that is made by the repetition of a basic unit, or monomer. Almost all useful polymers employ carbon-carbon bonds to form a chain with a carbon atom backbone (a few useful polymers are based on silicon chains; these are known as silicones). The prototypic polymer sequence is that which generates the molecules methane (CH_{4}), ethane (C_{2}H_{6}), propane (C_{3}H_{8}) and polyethylene (C_{n}H_{n+2}), whose structures are shown in the two-dimensional representations given in Fig. 6.1. The carbon-carbon chain can be repeated almost indefinitely; common commercial polymers have 10^{4} - 10^{6} carbon atoms in their backbone chains.
To understand the structures of polymeric solids, and, in fact, of molecular solids in general, it is useful to address them on two levels: the internal structure of the molecule and the configuration of the molecules in the solid. In the polymeric solids much of the complexity is in the structure of the polymer itself. Given that the hydrogen atoms in the polyethylene chain in Fig. 6.1 can be replaced by other atoms or groups of atoms, and that the carbon chain itself can be altered chemically in a number of ways, the variety of possible polymers is almost endless. A detailed treatment of polymer chemistry is beyond the scope of this course. The discussion given here is confined to a few details that are helpful in understanding the microstructures and properties of polymeric materials.

Fig. 6.1: Sequence of carbon-chain polymers based on the C-H bond.

It is useful to divide polymeric solids into two classes: *chain polymers*, which are linear molecules (with possible side branches) that are joined by polar bonds to form molecular solids, and *network polymers*, in which individual molecules are joined together by molecular bonds to create three-dimensional networks.

### 6.3 Chain Polymers

#### 6.3.1 Types of chain polymers

Not all organic compounds can polymerize. In order for a molecule to form a simple chain polymer it must be capable of bonding to replicas of itself at both its ends. It must, therefore, be able to provide unsaturated bonds at both its ends. Molecules that have this property are said to be *bifunctional* in that they can attach functional groups at either end. There are two simple types of bifunctional groups that lead to two types of polymers, called *addition polymers* and *condensation polymers*. (A third type, *ring-opening polymers*, will not be treated here.)

*Addition polymers*

The simplest bifunctional molecules are those that contain unsaturated carbon bonds, which appear as double bonds between adjacent carbon atoms. Many of the most important chain polymers are formed by polymerizing simple molecules that have the generic structure shown in Fig. 6.2. It is possible to join a molecule to either side of the molecule shown in Fig. 6.2 by replacing the double bond between the carbon atoms by a
single bond; hence the molecule is bifunctional. A polymer that is made by sequential addition to the sides of a bifunctional molecule is called an addition polymer.

In Fig. 6.2 R represents an atom or a functional group of atoms. When R is a hydrogen atom the molecule is ethylene, and the polymer is polyethylene, which is drawn in Fig. 6.1. When R is a chlorine atom the molecule is vinyl chloride, and the polymer is polyvinyl chloride (PVC). When R is a hydroxyl group (OH) the molecule is vinyl alcohol, and the polymer is polyvinyl alcohol. When R is a CH₃ group the molecule is propylene, and the polymer is polypropylene. When R is a benzene ring the molecule is styrene, and the polymer is polystyrene.

\[
\begin{align*}
H & \quad H \\
| & \quad | \\
C & \quad C \\
| & \quad | \\
H & \quad R
\end{align*}
\]

Fig. 6.2: The structure of a simple monomer that is the basis for many important polymers.

Slightly more complicated bifunctional molecules have one or more of the hydrogen atoms in Fig. 6.2 replaced by other atoms or functional groups. Two examples are drawn in Fig. 6.3. Note that methyl methacrylate (Fig. 6.3b) can polymerize via the C = C bond, but also has the double bond C = O which can react with other molecules. Molecules that have reactive sites in addition to the bonds needed for polymerization are called multifunctional.

\[
\begin{align*}
F & \quad F & \quad H \quad CH₃ \\
| & \quad | & \quad | \\
C & \quad C & \quad C \\
| & \quad | & \quad | \\
F & \quad F & \quad H \quad C = O \\
& & \quad O - CH₃
\end{align*}
\]

Fig. 6.3: Examples of monomers that create addition polymers: (a) tetrafluoroethylene (teflon); (b) methyl methacrylate (plexiglass or Lucite).

Condensation polymers

A monomer that does not contain a double bond in its central chain cannot simply add copies of itself. It must polymerize through a chemical reaction in which bonds are broken and reactions products (usually water, H₂O, or ammonia, NH₃) are released. This kind of reaction is called a condensation reaction, and the resulting polymers are called condensation polymers.
An important example of a condensation polymer is the fiber, nylon, which is formed by a reaction between hexamethylenimine-diamine \((\text{NH}_2(\text{CH}_2)_6\text{NH}_2)\) and adipic acid \((\text{COOH}(\text{CH}_2)_4\text{COOH})\). The reaction is diagrammed in Fig. 6.4. The organic acids have end groups of the form \(\text{COOH}\) while amines have the end group \(\text{NH}_2\). Neither type of molecule will ordinarily polymerize since \(\text{COOH}\) and \(\text{NH}_2\) groups do not interact easily with themselves. However, the \(\text{COOH}\) and \(\text{NH}_2\) groups react readily with one another to release a water molecule. In the case of nylon the reaction produces a bifunctional molecule that has an acid group at one end and amine group at the other, and can hence add replicas of itself indefinitely. A water molecule is released whenever monomers are added or polymer chains joined.

![Reaction Diagram](image)

Fig. 6.4: The formation of an AB group from the reaction of an amine and an acid. The product can generate a regular copolymer.

Many industrially important polymers are condensation polymers, including polyesters, phenol-formaldehyde, and urea-formaldehyde resins. Most of the macromolecules that govern the behavior of biological systems are also condensation polymers, including cellulose and the nucleic acids, DNA and RNA.

### 6.3.2 Internal symmetry in chain polymers

Polyethylene (Fig. 6.1) is a completely symmetric chain polymer. The monomers are identical to one another, and the side groups are indistinguishable. However, polymers can be asymmetric in either sense. If the side groups of the monomer are not identical it is necessary to specify the configuration of the side groups to describe the polymer. If more than one kind of monomer is included in the chain the polymer is called a copolymer, and the particular sequence of monomers must be given to describe its structure.

*Asymmetric monomers*

The simplest kind of asymmetric monomer is illustrated in Figs. 6.2 and 6.3(b); different molecular side-groups are bonded to the two sides of a carbon atom. Polymers that contain identical numbers of asymmetric monomers may have different properties because of differences in the way the side-groups are placed. In this sense an asymmetric polymer resembles a crystalline solid with interstitial solutes; it is necessary to specify the state of order of the interstitials.
Three orientational patterns are reasonably common and have been given
descriptive names. A polymer is called *isotactic* when successive monomers have the
same orientation; the polymer derived from the molecule in Fig. 6.2 is isotactic when the
group R is always on the same side of the carbon chain, as shown in Fig. 6.5a. When the
position of the side group (R) regularly alternates sides of the carbon chain, as in Fig.
6.5b, the polymer is said to be *syndiotactic*. When the position of the side group is
random with respect to the carbon chain the polymer is said to be *atactic*.

![Diagram](image)

**Fig. 6.5:** Relative orientations of asymmetric monomers: (a) isotactic; (b)
syndiotactic; (c) atactic (random).

Because of their ordered configurations isotactic and syndiotactic polymers tend
to be straight or helical chains that can form strong polar bonds to one another by
bringing appropriate side-groups into registry. These ordered polymers are often easy to
crystallize in the solid state. Atactic polymers have irregular chains that bond more
weakly and do not readily crystallize.

A second kind of asymmetric monomer has only a single side-group. These
monomers necessarily appear in pairs in the polymer chain, and are joined by a carbon-
carbon double bond (Fig. 6.6). The unbalanced side groups may be on the same side of
the chain or on opposite sides. If they are on the same side the polymer is said to have a
cis-configuration, while if they are opposed they are said to be in a trans-configuration.

![Diagram](image)

**Fig. 6.6:** The structures of (a) cis-polyisoprene (rubber) and (b) trans-
polyisoprene (gutta percha).

A classic example that illustrates the importance of the side-group orientation at a
double bond is the contrast between cis-polyisoprene (natural rubber) and trans-polyiso-
prene (gutta percha). Their respective configurations are shown in Fig. 6.6. The CH₃
and H sidegroups in the cis-configuration interfere geometrically, which causes the
polymer chain to bend at the double bond and makes the rubber molecule kinked and
irregular in shape. Rubber polymers do not pack well together; as a consequence natural
rubber is soft and elastic. In contrast, the sidegroups in the trans-configuration
accommodate one another more readily, so the gutta percha polymer chain is relatively straight. Gutta percha crystallizes into a solid that is rigid compared to natural rubber.

**Copolymers**

A polymer that contains two or more chemically distinct monomers in the polymer chain is called a copolymer. A copolymer is the polymeric analog of the substitutional solid solution on a crystal lattice.

Let the two basic elements of a chain copolymer be the molecular groups A and B. A *random copolymer* of composition \( x \) is made by randomly selecting a fraction, \( x \), of the A-groups in the polymer and replacing them by B-groups. This makes, in effect, a one-dimensional random solution. If the B-groups are in ordered sites along the chain the configuration is called a *regular copolymer*. A regular copolymer is, effectively, a one-dimensional ordered compound.

...AAABABAABAAAABAB.. ..ABABABABAB.. (a) (b) ...

Fig. 6.7: (a) Random copolymer of monomers A and B. (b) Regular copolymer.

Many important chain polymers are regular copolymers since it is often necessary to join two distinct molecules to obtain a basic unit that can polymerize by adding replicas of itself. An example is nylon, which was discussed above and is illustrated in Fig. 6.4. It is a copolymer of an organic acid, adipic acid, and an amine, hexamethylenamine-diamine. Note that nylon is, necessarily, a regular copolymer. Only the acid group can join on the amine side, and only the amine group can join on the acid side.

**6.3.3 Defects in chain polymers**

The next level of complexity in a chain polymer is its defect structure. Chain polymers are basically one-dimensional objects, and hence can contain 0-dimensional point defects or 1-dimensional line defects. The point defects in polymer chains lead to structures that resemble solid solutions and ordered compounds. The line defects are topologically like volume defects in three dimensional solids.

The *intrinsic defect* in a polymer chain is an improperly oriented monomer in an isotactic or syndiotactic polymer. Virtually all other point defects can be regarded as extrinsic in that they can be made by substituting a different atom group for a monomer or other elementary segment of the chain. If many equivalent defects are added the polymer is a *copolymer*, as described above.

Long-chain polymers can also be formed by joining short chains of different polymers. The result is called a *block copolymer*. The added B-group forms a linear defect
in this case, and the polymer is the one-dimensional analog of a two-phase mixture or composite material in three dimensions.

\[ \text{AAAAAAABBBBBBA}} \text{AAAAAAAAABBBBBBA}} \text{AAAAAAAAAAAAAA} \]

(a)

\begin{align*}
\text{AAAAAAABBBBBBA}} & \quad \text{AAAAAAABBBBBBA}} \\
\text{AAAAAA} & \quad \text{AAAAAA} \\
\text{AAAAA} & \quad \text{AAAAAA} \\
\end{align*}

(b) (c)

Fig. 6.8: (a) Block copolymer; (b) branched polymer; (c) graft copolymer.

Linear defects also occur in chemically homogeneous polymers when a chain segment attaches to the main backbone of the polymer and branches away from it. Such a polymer is called a \textit{branched} polymer. (When the branched segments are very short it is sometimes called a \textit{feathered} polymer.) When the branch is a segment of a different polymer the result is called a \textit{graft copolymer}. Examples are shown in Fig. 6.8.

### 6.3.4 Solid aggregates of chain polymers

The packing of chain polymers into three-dimensional molecular solids may be crystalline (ordered) or amorphous (disordered). As in all solids the most suitable crystal structure provides the lowest energy. However, the crystallization of a chain polymer is inhibited and often prevented by its physical length and chemical complexity. The tendency of a polymer to crystallize depends on its geometric regularity and on the strength of the bonding between adjacent polymer chains. Hence isotactic and syndiotactic polymers crystallize more readily than atactic ones, trans-configured polymers crystallize more readily than cis-configured ones, and polymers that contain regularly spaced dipoles interact more strongly and crystallize more readily than those that are electrically homogeneous. Irregular branches on the polymer chains inhibit crystallization.

When a long-chain polymer crystallizes it is, for geometric reasons, almost impossible for it to achieve the most stable possible configuration, which is ordinarily a parallel ordering of the polymer chains. Instead, the chain forms a semi-crystalline pattern, partly by aligning with its neighbors, and partly by looping back on itself in a regular way. The former leads to what is called a \textit{fringed micelle} in which different chains parallel one another along part of their length and are disordered elsewhere to create a partially crystallized medium. The latter creates what is called a \textit{folded chain}. The two configurations are illustrated in Fig. 6.9. Whatever the internal structure,
regions of crystal order within a polymer are imperfect and separated by diffuse walls in which the configuration is more nearly random. Even highly crystallized polymers are rarely more than 90% crystalline.

![Crystal patterns](image)

(a) (b)

Fig. 6.9: Crystal patterns in a long-chain polymer: (a) fringed micelles; (b) folded chains.

When polymer crystals are grown from the melt they tend to form grain-like structures called *spherulites*. Spherulites grow as spheres radially outward from a center. Polymer chains are added along the periphery of the sphere so the chains tend to lie in a tangential direction. Spherulites contain dense distributions of small crystallites that are oriented radially. The spherulites cause very interesting optical effects in transparent polymers, which is one reason for the interest in them.

Long-chain polymers are often mixed in the solid state with other polymers, chemical additives or inorganic particles that are intended to improve strength, formability or cost. The most common additives are clay and talc, which are included to improve formability, carbon particles and glass fibers, which are added to increase strength, and smaller organic molecules called *plasticizers*, which are added to improve polymer flow by inhibiting crystallization and minimizing binding between polymers. All of these additives are well-mixed to form fine dispersions or molecular solutions.

The crystallization of a polymer generally increases its strength and rigidity with respect to the disordered state. It also dramatically changes the behavior on heating or cooling. When an amorphous chain polymer is heated or cooled it undergoes an essentially continuous transition between the liquid and solid states, that is, it behaves like *glass*. Such polymers are *thermoplastic*; they can be easily formed into complex shapes by heating them to a temperature at which they are soft. They harden on cooling without appreciable geometric distortion since there is no volumetric discontinuity in the transition to the more solid state. Crystalline polymers, on the other hand, melt discontinuously on heating and crystallize discontinuously on cooling. They are relatively strong, but are not thermoplastic and are, hence, more difficult to form.

### 6.4 NETWORK POLYMERS
A second class of polymeric materials includes the network polymers that form three-dimensional networks rather than linear chains. In a sense, these polymers are three-dimensional macromolecules. However, it is often reasonable to picture them as long-chain polymers that are chemically bonded to one another. Because the chains are tied to one another by chemical bonds they cannot move easily with respect to one another, even at high temperature. Consequently, these polymers do not soften on heating. They retain their structure and most of their strength until the temperature is high enough to break the bonds between the chains or functional groups, at which point the polymer usually decomposes chemically. They are called thermosetting polymers because the reactions that establish the three-dimensional network of bonds usually occur at high temperature, and set the polymer into a permanent configuration.

The molecules that form network polymers are multifunctional; they contain more active sites than are necessary to make the polymer chain, and the remainder are available to create three-dimensional bonds. The network polymers can be roughly divided into two types: those that are inherently three-dimensional in the sense that their bonds branch in all directions, and those that contain chains that bond to one another by discrete bridging bonds called cross-links.

### 6.4.1 Network polymers due to multi-directional bonds

A classic example of a multi-directional bonded network polymer is phenol-formaldehyde, or bakelite. It is based on the phenol group, which contains a benzene ring of six carbon atoms that has five C-H bonds at angles of 60° to one another in the plane and one external bond to an OH group. Two phenol groups can be bonded together by condensation reaction with formaldehyde, OCH₂, to form an H₂O molecule and a CH₂ bridging bond. The reaction is illustrated in Fig. 6.10. Such bonds can form easily on three C-H sites at 120° to one another on the benzene ring, so the reaction naturally leads to a network polymer.

![Diagram of phenol-formaldehyde condensation reaction](image)

... 

Fig. 6.10: The polymerization reaction that produces the network polymer phenol-formaldehyde (bakelite). Multiple reactions create a network of bonds from each benzene ring, as shown at right.
6.4.2 Network polymers due to cross-links

A second way to create a network polymer is to establish chemical bonds between neighboring chains at multifunctional sites. These cross-links may be regularly spaced, but are more often distributed in an approximately random way so that the polymer chains are frozen into an amorphous configuration.

The classic example of a network polymer that forms by cross-linking is vulcanized rubber. The vulcanization reaction is induced by heating natural rubber in the presence of sulfur (you probably learned in grammar school that Charles Goodyear "accidentally" discovered this reaction and was subsequently immortalized in the name of the Goodyear Tire and Rubber Company). The reaction is diagrammed in Fig. 6.11. Sulfur atoms break the C=C bonds on the polyisoprene chains of natural rubber to create bridging bonds to neighboring chains. Note the importance of the cis-configuration of natural rubber at the double bond, which creates an open space that allows chemical access to the double bond.

![Diagram of vulcanized rubber with sulfur atoms cross-linking the chains.]

Fig. 6.11: The vulcanization of rubber. Sulfur atoms react with the C=C double bonds on adjacent chains to create cross-links.

6.4.3 Elastomers

Polymers that form network bonds or dense distributions of cross-links are useful as rigid structural materials. Polymers that have highly kinked chains in the isolated state, as rubber does, can also be made into very flexible materials by introducing a relatively low density of cross-links. These materials are known as elastomers. They are useful because they stretch under relatively light mechanical loads, but recover their initial configurations as soon as the load is released.

![Diagram of an elastomer structure showing the flexibility.]

...
Fig. 6.12: Schematic drawing of cross-linked polymer chains in an elastomeric material.

Vulcanized rubber is the classic example of an elastomer. The cis-configuration at the C=C bond causes the polymer chain to bend or kink; the individual polymer chains resemble coiled springs. The periodic cross-links between chains create a network that holds the chains in a particular configuration in the absence of a mechanical load. The chains straighten under mechanical load so that the material stretches just as a spring does. Since the stretching is primarily due to the straightening of the chains it involves a relatively small distortion of the interatomic bonds and can be accomplished by a relatively small mechanical load. When the load is removed the network relaxes into its original configuration. The polymer configuration that leads to elastomeric behavior is drawn in an idealized way in Fig. 6.12.

### 6.5 Silicates

A second class of macromolecular materials that is important in nature and in engineering is the silicates, which are the building blocks for many of the minerals of which the world is made. The silicates are inorganic but, like the polymers, arrange themselves into solids in many different ways. Their chemical and structural flexibility is responsible for the rich variety of rocks, clays and minerals that are found in nature. In engineering, the silicates are the prototype ceramic materials, which have since prehistory been the materials of choice for pots, plates and ornamental objects, and are also used widely in modern engineering devices as electrical insulators and dielectrics, optically transparent media, and structural materials that retain strength and chemical stability at very high temperatures.

Many oxides and other inorganic compounds appear in natural minerals and are used in modern ceramics. However, we shall confine this discussion to the structures of the silicates. They are common in nature and are useful structural prototypes for other oxide minerals and ceramics.

#### 6.5.1 The SiO$_4$ tetrahedron

The basic structural element of the silicates is the SiO$_4$ tetrahedron that is shown in Fig. 6.13. It contains a central silicon atom surrounded by four oxygen atoms in a tetrahedral configuration. Since silicon has valence 4 it can saturate its valence shell by forming covalent bonds with the 4 oxygen atoms. However, oxygen has valence 6, so after sharing an electron with silicon each oxygen atom is still deficient by one electron in its valence shell.

The possibility of making many different structures from a distribution of SiO$_4$ tetrahedra is due to the fact that there are two very different ways to complete the oxygen bond. The first is illustrated on the left-hand side of Fig. 6.14. If a second silica tetrahedron is joined at a corner occupied by oxygen then the oxygen atom is simultaneously a
member of both tetrahedra. Its bonding is completed by a covalent bond to the silicon atom at the center of the second tetrahedron. In fact, it is not necessary that the second atom in the bond be silicon. Many atoms, particularly including aluminum, iron and lead, form geometrically repeatable coordinated complexes with oxygen with the general structural formula $M_xO_y$. An oxygen atom in the silica tetrahedron may complete its bonding by simultaneously becoming a member of the oxide shell of any of these to create a mixed oxide.

![Diagram of SiO4 tetrahedron](image)

**Fig. 6.13:** The SiO$_4$ tetrahedron.

The second way to complete the oxygen bond is to associate it with a cation that donates an electron to fill its outer shell. If the ion carries a single positive charge then it seals the tetrahedron at the corner, as shown at the extreme right in Fig. 6.14. If the ion is divalent then it can form an ionic bridging bond between two discrete tetrahedra, as shown in the center of Fig. 6.14.

![Diagram of bonding of silicate tetrahedra](image)

**Fig. 6.14:** Illustration of bonding of silicate tetrahedra: shared corners, bridging ionic bonds by divalent cations, and single bonds to univalent cations.

### 6.5.2 Three-dimensional network structures

If silicate tetrahedra bond to one another by sharing corners then the tetrahedra form a three-dimensional network in space, as illustrated in Fig. 6.15. Since every oxygen in the network is a member of two tetrahedra the stoichiometric formula is SiO$_2$. In the lowest-energy configuration the tetrahedra are stacked in a regular way so that the material is crystalline, and is called quartz.
Crystalline quartz takes on several different structures as the temperature and pressure are changed. The network structure of the silicate tetrahedra is a very open one, and adjacent tetrahedra can re-orient by rotation about the Si-O-Si bond without distorting it. As a consequence there are several different ways of arranging the tetrahedra in space that differ very little in energy, and provide alternate crystal structures.

The openness of the network structure also has the consequence that the energy of a regular crystalline arrangement of the tetrahedra is not greatly different from that of a disordered arrangement in which the tetrahedra have an irregular distribution and fit onto one another by rotation about the Si-O-Si bond. This is the structure of silica glass, which was discussed in the previous chapter. The formation of a glass is strongly promoted by the addition of ionic solutes such as Na, Mg or Pb to the silica melt. These form non-directional ionic bonds with adjacent tetrahedra which disturb the local configuration of the tetrahedra and makes it more difficult for them to achieve a crystalline pattern. Silica glasses are thermoplastic.

It is also possible to form mixed oxide structures by replacing the silica tetrahedra by a coordinated oxygen complex of some other metal, $\text{M}_x\text{O}_y$. If the placement is done in an ordered way the mixed oxide is crystalline. An important example is aluminum oxide, which forms the network oxide alumina, $\text{Al}_2\text{O}_3$, and mixes with silica to form the aluminosilicates. Aluminosilicates are among the most common minerals of the earth. Aluminosilicates can be made to have complex network structures. Synthetic aluminosilicates include the zeolites, whose network frames are tailored to provide channels and cavities that make them useful as ion exchange resins and chemical catalysts. Mixed oxides with iron are also common in nature.

The distortion of the three-dimensional lattice frame by mixed oxides promotes the formation of glass. Lead oxide is a common additive in glass.

6.5.3 Two-dimensional sheet structures

If the $\text{SiO}_4$ tetrahedron is stood on the triangular base made by three of its oxygen atoms, the fourth apex of the tetrahedron points out of the basal plane. The tetrahedra can
join by sharing only those corners that are in the basal plane to form a continuous sheet, as shown in Fig. 6.16. The unsatisfied oxygen bonds point out of the plane of the sheet. The unsatisfied bonds can be completed by adding a layers of ionic species above and below the plane as needed, or by adding a plane of a second oxide whose layer also contains unsatisfied bonds. Both types of minerals appear in nature.

A common sheet structure has two of these sheets configured to form a sandwich in which the unsatisfied oxygen bonds all point toward the midplane of the sandwich. The oxygen bonds are then completed by ionic species that fill the midplane of the sandwich, as shown in Fig. 6.17. There are no dangling bonds on the outside of the sandwich. Hence adjacent sandwich composite layers are only weakly bound to one another and separate easily. This is the elementary structure of mica and of simple clays.

Typical clay minerals that are used to make pottery consist of platelets that have neutral sandwich structures like that drawn in Fig. 6.17. The structure is responsible for their formability. To enhance their formability the clays are mixed with water. The water molecules penetrate into the interstices between the platelets of the clay, separating and lubricating them so that they glide easily over one another. The wet clay is then plastic, and can be molded into useful shapes.

![Figure 6.16](image_url)

**Fig. 6.16:** A sheet of SiO$_4$ tetrahedra made by tetrahedra that share three corners in a plane. The fourth corner of the tetrahedron has an unsatisfied bond pointing out of the plane of the figure.

If the clay is dried directly after forming it shrinks as water is lost and accommodates the shrinkage by cracking along the easy planes of fracture that separate the neutral sandwiches of the clay. Hence one cannot make a useful ceramic object by simply molding clay. To complete the manufacture of a ceramic the molded clay is fired in a furnace. The beneficial consequence of firing comes from the fact that the two-dimensional silicate structure that makes clay is less stable than the three-dimensional network structures that can be made of the same materials. Raising the temperature by firing drives out the water and makes it kinetically possible for the sheets to reconfigure in three-dimensional patterns. Effectively, firing changes clay into stone.
Fig. 6.17: The structure of a neutral sandwich of silicate planes. The two layers of the sandwich are bonded together with positive ions.

6.5.4 One-dimensional chain structures

If only two of the corners of each tetrahedron are shared the tetrahedra can be arranged in a linear chain with the unsatisfied oxygen bonds on the periphery. It is also common for silicates to form double chains by sharing appropriate bonds in the basal plane of the tetrahedron as diagrammed in Fig. 6.18. This is the basic structure of the mineral asbestos, for example. If the outer bonds are satisfied by ionic species (primarily Mg$^{++}$ in asbestos), the result is a neutral chain structure that is very strong and stable. Chain silicates like asbestos form fibers can be made into cloth that provides excellent thermal insulation. Unfortunately, the fibers are also irritating to the human lung. The problem of asbestosis was virtually unknown until recently, but has largely curtailed the use of asbestos in insulation.

Fig. 6.18: The structure of a double chain of silicate tetrahedra as found, for example, in asbestos.

6.6 ARTIFICIAL FIBERS

A macromolecular structure that is increasingly important in modern materials science is the macromolecular fiber, which is a long, thin cylinder or thread. Natural fibers have been spun and woven into cloth for millennia. Artificial fibers have been made for almost as long; thread made of finely drawn gold wire is found in very ancient sites. However, the last few decades have seen an explosion in the creation and engineering use of artificial fibers. Engineers have learned to spin continuous fibers of exceptional strength that are used not only as fabrics, but also as continuous reinforce-
ments for structural materials that range from reinforced rubber used in automobile and truck tires to ultralight, stiff composite materials that make up a significant fraction of the structure of modern aircraft. The fiber-reinforced materials are called *structural composites*.

The traditional textile fibers are strands of natural materials, based primarily on the cellulose polymer that makes up the fibrous structure of woods and plants. Short cellulose fibers are aligned and twisted into thread whose mechanical strength is primarily due to the physical interference of the twisted and tangled strands. The traditional metal fibers were ductile metal drawn into thin threads. In recent years the natural fibers have been supplemented by aligned, synthetic polymers, such as nylon, dacron and the polyesters. The metal fibers have been largely replaced by continuous strands of glass, graphite or polymer.

The synthetic fibers that have excited the greatest interest in modern engineering are those whose exceptional strengths make them useful for the reinforcement of low density structural materials. Their interesting microstructures are worth noting. Four examples that represent distinct types of artificial fibers are discussed below: glass fibers, graphite fibers, organic fibers of the kevlar type, and microcrystalline whiskers.

### 6.6.1 Glass fibers

The simplest of the modern fibers are drawn silica glass. They are made by extruding semi-molten silica through a die to produce a continuous glass filament. Glass fibers are inexpensive, lightweight, and relatively strong. They are embedded in polymer matrices to create *fiberglass*.

Fiberglass is used so widely that its structural strength is almost automatically accepted. However, on reflection the structural importance of glass fiber is counterintuitive. In almost all familiar situations glass is catastrophically brittle. One does not live in glass houses or (excepting Wonder Woman) fly in glass airplanes. So how can glass fibers be strong?

The reason, which we shall study in some detail later in the course, is that glass is inherently strong, because of the strong Si-O bonds that hold the atoms together, but is ordinarily very brittle, because of its sensitivity to small cracks or flaws that propagate to fracture under small loads. The secret to making reliable glass fiber is to draw the fiber so that it has an essentially flaw-free surface. The drawing operation also imparts a slight compressive stress in the surface that acts to close any small flaws that may appear. When flaws are not present the natural strength of the glass emerges. Glass fibers can be made to have exceptionally high strengths. Their mechanical reliability is further enhanced by using them as reinforcements in ductile, plastic matrices. When a glass fiber breaks, the ductile matrix helps to prevent the fracture from propagating through the material.
In recent years glass fibers have become as important for their optical properties as for their structural strength. Properly made fibers function as efficient optical fibers that channel optical signals with very little distortion. They are widely used in optical communications. A substantial fraction of all telephone traffic is now carried through glass fiber. Optical glass fibers are also used as light pipes in medical and other applications, where their ability to channel light is used to view otherwise inaccessible places, such as the recesses of the human body. We shall study the optical properties of glass fibers at a later point.

6.6.2 Graphite fibers

A second example of the use of a "weak" material to make a strong macromolecular fiber is graphite fiber. We have already mentioned the graphite structure of carbon. It consists of stacked planes in which carbon atoms are strongly bound to one another in the open, hexagonal pattern shown in Fig. 6.19. However, the bonding between successive planes is relatively weak. The planes slide easily over one another, and graphite deforms readily under small loads.

The structure of graphite fiber is illustrated in Fig. 6.19. To make the fiber the graphite is reacted so that very long, continuous planes of graphite form in a configuration in which they are rolled into cylinders. When the fiber is pulled along its axis the load is borne by the graphite plane. To deform or break the fiber the load must be sufficient to break the strong bonds in the graphite plane. As a consequence, graphite fibers are among the strongest materials known.

![Graphite fiber structure](image-url)

Fig. 6.19: Graphite fiber. The hexagonal planes of graphite are wrapped so that the fiber axis lies in the strongly bonded plane.

Since graphite is a relatively open structure of a light element, graphite fibers have very low density. Graphite-reinforced plastics, called graphite composites, combine exceptionally low density with exceptionally high strength and stiffness in the direction of the fibers, and are used in such high-tech devices as aircraft, golf clubs and tennis rackets where these properties are particularly useful.
However, the graphite composites suffer from the draw-back that they are relatively weak when loaded perpendicular to the fiber axis. This problem can be partly overcome by cross-laminating the fibers so that fibers are oriented in two or more directions. But this compromises the strength and stiffness of the composite. Graphite composites are used primarily in applications where the direction of the maximum load is fixed, as it is in many aircraft parts.

6.6.3 Kevlar fibers

Strong fibers can also be made of polymers that are carefully aligned so that the strong axis of the polymer chain lies along the fiber. On of the most successful of the fibers of this type is called kevlar, and was developed by the DuPont company some years ago. (The name, kevlar, is a DuPont trademark.)

While not all the details of kevlar fibers are known, they are essentially based on an aromatic polyamide called poly(paraphenylene terephthalamide). The name may be unpronounceable, but the polymer is not that hard to understand. It is a copolymer that is formed by a condensation reaction that is essentially the same as the reaction that forms nylon (Fig. 6.4). The structural formula of the copolymer unit can be written \((\text{NH})-\text{R}-(\text{NH})-(\text{CO})-\text{R'}-(\text{CO})\), just as in nylon. The difference is that in kevlar the central molecular groups, which are labeled R and R' in Fig. 6.4, are identical benzene rings. The benzene rings make the polymer rigid. Any attempt to bend the polymer is opposed by the geometric interference of the benzene rings.

![Diagram of a polymer sheet in kevlar](image)

Fig. 6.20: The planar configuration of a polymer sheet in kevlar.

The molecules of kevlar can be aligned so that the oxygen atoms on the C=O group of a polymer oppose hydrogen atoms in the N-H group of its immediate neighbor. A dense array of polar, hydrogen bonds is then formed between the adjacent molecules. By repeating this alignment the molecules form sheets, as shown in Fig. 6.20, that are essentially two-dimensional crystals with strong bonding in the plane. However, the
saturation of bonds in the plane has the consequence that neighboring planes are weakly bound to one another.

The microstructure of kevlar fiber is derived from the two-dimensional kevlar crystal through two additional steps, which are illustrated in Fig. 6.21. First, the crystal planes are configured in regularly pleated sheets in which planar segments that are several hundred atom groups long are joined together. The angular deviation between the adjacent planar segments is only about 10°, but the pleated sheets still have some of the character of an accordion fold. The pleated sheets are then arranged radially into a cylindrical fiber with the fiber axis in the plane of the sheet.

While kevlar fibers are not as strong or as stiff as graphite fibers, they are strong, and they are much more easily deformed. The pleating of the individual sheets permits some elastic extension prior to failure. When the fiber is pulled along its axis the sheet stretches like an accordion before the load is applied directly to the bonds in the sheet plane. Moreover, since the sheets are weakly bound to one another they can reconfigure by sliding over one another, so the fibers have some ductility. When a kevlar fiber is stretched to failure in the axial direction it gradually pulls apart in a fibrous manner rather than fracturing as a glass or graphite fiber would. When a kevlar fiber is bent it buckles into a permanent plastic set rather than breaking.

![Fig. 6.21: Pleated sheets of kevlar, as shown at left, are aligned radially to produce the cylindrical fiber shown at right. The deflection at the pleat (≈ 10°) is exaggerated for illustration.](image)

The combination of high strength and ductility that is offered by polymer fibers of the kevlar type make them particularly suitable for applications that require a combination of strength and flexibility. Kevlar fibers are used as reinforcing cords for automobile tires and as fabric for bullet-proof vests, among many other applications.

### 6.6.4 Whiskers

A final kind of short fiber that is widely used as a reinforcement is the whisker, which is an almost perfect crystal with a needle or thin columnar shape. Whiskers are formed by crystallizing a solid under conditions of high supersaturation, such as by deposition from the vapor or precipitation from a highly supersaturated liquid or solid solu-
tion. The useful properties of whiskers are based on their geometry, which is that of a short fiber, and their crystal perfection.

Most crystalline solids are at least moderately defective. No matter how carefully the crystal is grown it contains a distribution of dislocations which are generated by thermal stresses or irregularities in the deposition of successive planes. However, crystals that grow as thin needles or columns under high supersaturation often contain very few defects, and these may spontaneously migrate out of the crystal during or shortly after its growth. Such crystals are called whiskers because they are small and needle-like in shape, and tend to cover the surface on which they grow very much like a three-day growth covers a man's face. They have very high strength since they contain no dislocations that can induce plastic deformation under small loads.

Whiskers are invariably short, no more than a few millimeters in length, and thin. They hence cannot be used for continuous fiber reinforcement, but are useful as reinforcing dispersants. The whiskers that are most commonly used in engineering today are SiC whiskers that are dispersed through relative soft aluminum alloys. Because of the low density and high strength and stiffness of the SiC whiskers the resulting composite materials are light-weight, strong and stiff, and are attractive for use in aircraft. However, since the whiskers also limit the ductility of the material and make it more liable to fracture, their use in critical structural components is limited.

6.7 QUASICRYSTALS

A final class of macromolecular solids includes a new class of materials, called quasicrystals, that has only been discovered in the last few years. The quasicrystals are intermetallic compounds, most frequently based on Mn or Al, that have regular three-dimensional atomic configurations that are not periodic and, hence, cannot be referred to a crystal lattice.

To understand why quasicrystals might exist, recall that only certain patterns of points can be repeated in a crystal lattice. For example, the unit cell of a two-dimensional crystal can have four or six sides, but cannot have three or five, since identical triangles or pentagons cannot be multiplied indefinitely to fill a plane. Suppose that one has an atom or molecule whose bonding is such that it would prefer to have five neighbors. Since a pentagonal cell is impossible it cannot do this in a crystalline array. It can adopt a crystal structure in which each lattice point contains a group of atoms that have five-fold coordination, since one can create a crystal by placing pentagons on the points of a Bravais lattice, but at least some of the atoms on the boundaries of these molecular groups necessarily have non-pentagonal configurations.

However, one can create almost pentagonal configurations of atoms in a plane if one drops the constraint that the configuration be periodic. Beginning with a pentagonal configuration at the origin, one adds additional atoms in a roughly spiral pattern about the origin, to create a macromolecule in which each atom is coordinated by five neighbors.
The resulting configuration is not crystalline; the macromolecule has a unique origin, and no two points in it have completely identical environments. But it can be made.

Mathematicians have studied such non-repeating patterns for many years. They are examples of two-dimensional patterns called Penrose tilings, after the British mathematician Roger Penrose, who studied the various ways of filling a plane with patterns of tiles. While the mathematical analysis of Penrose tilings is most developed for the two-dimensional case, three-dimensional examples have also been identified in which one fills three dimensions with solid bricks of relatively simple shape that form an aperiodic pattern with a unique origin.

It was recently recognized that there are intermetallic compounds that can be formed in structures that seem aperiodic and can be described in terms of Penrose tilings. They are called quasicrystals since their arrangements are regular and predictable, but aperiodic. The most thoroughly investigated are compounds of Mn. Manganese has a predilection for ten-neighbor coordinations that cause it and its compounds to have complicated crystal structures. Certain of its compounds seem to form in quasicrystalline patterns. Compounds of Al and other metals have also been observed in apparently quasicrystalline structures.

At present the quasicrystals are of scientific interest only. They invariably have very poor mechanical properties, and no useful electromagnetic properties have been identified. Their main engineering interest is in their avoidance, since their appearance, for example, in Al-Li aerospace alloys, causes a loss of ductility and toughness. Even their existence remains somewhat controversial. Since much of the scientific evidence that establishes the aperiodic structures of quasicrystals can also be interpreted by assuming a highly defective crystal structure, some knowledgeable crystallographers (following the late Linus Pauling) deny that they are real. However, there is no inherent reason why quasicrystals cannot form; they are simply macromolecules made by accreting atoms on a seed whose symmetry is incompatible with periodicity. The bulk of the available evidence suggests that at least some of the reported examples are real.