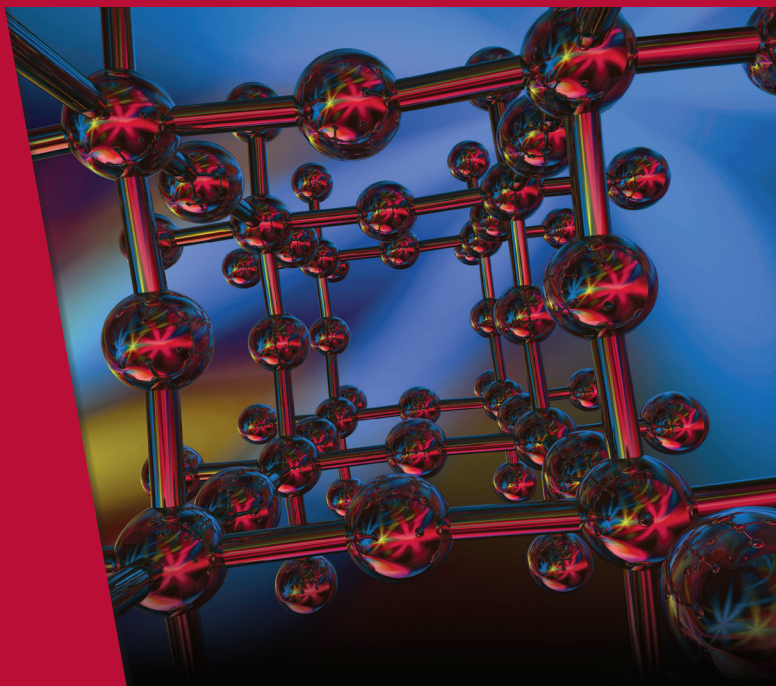
SPIE. FIELD GUIDE

Field Guide to

Solid State Physics



Marek S. Wartak
C. Y. Fong

Field Guide to

Solid State Physics

Marek S. Wartak
C. Y. Fong

SPIE Field Guides
Volume FG43

John E. Greivenkamp, Series Editor

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Names: Wartak, Marek S., author. | Fong, C. Y. (Ching-yao), author.
Title: Field guide to solid state physics / Marek S. Wartak and Ching-Yao Fong.
Other titles: Solid state physics
Description: Bellingham, Washington : SPIE, 2019. | Series: The field guide series | Includes bibliographical references and index.
Identifiers: LCCN 2018037785 (print) | LCCN 2018053727 (ebook) | ISBN 9781510622753 (pdf) | ISBN 9781510622760 (epub) | ISBN 9781510622777 (mobi) | ISBN 9781510622746 (spiral)
Subjects: LCSH: Solid state physics.
Classification: LCC QC176 (ebook) | LCC QC176 .W345 2019 (print) | DDC 530.4/1–dc23
LC record available at https://lccn.loc.gov/2018037785

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: 360.676.3290
Fax: 360.647.1445
Email: Books@spie.org
Web: www.spie.org

Printed in the United States of America.
First printing.
For updates to this book, visit http://spie.org and type “FG43” in the search field.

Introduction to the Series

In 2004, SPIE launched a new book series, the SPIE Field Guides, focused on SPIE's core areas of Optics and Photonics. The idea of these Field Guides is to give concise presentations of the key subtopics of a subject area or discipline, typically covering each subtopic on a single page, using the figures, equations, and brief explanations that summarize the key concepts. The aim is to give readers a handy desk or portable reference that provides basic, essential information about principles, techniques, or phenomena, including definitions and descriptions, key equations, illustrations, application examples, design considerations, and additional resources.

The series has grown to an extensive collection that covers a range of topics from broad fundamental ones to more specialized areas. Community response to the SPIE Field Guides has been exceptional. The concise and easy-to-use format has made these small-format, spiral-bound books essential references for students and researchers. I have been told by some readers that they take their favorite Field Guide with them wherever they go.

We are now pleased and excited to extend the SPIE Field Guides into subjects in general Physics. Each Field Guide will be written to address a core undergraduate Physics topic, or in some cases presented at a first-year graduate level. The Field Guides are not teaching texts, but rather references that condense the textbooks and course notes into the fundamental equations and explanations needed on a routine basis. We truly hope that you enjoy using the Field Guides to Physics.

We are interested in your suggestions for new Field Guide topics as well as what could be added to an individual volume to make these Field Guides more useful to you. Please contact us at **fieldguides@SPIE.org**.

John E. Greivenkamp, *Series Editor*
College of Optical Sciences
The University of Arizona

The Field Guide Series

Look for all of the physics *SPIE Field Guides*:

Classical Mechanics
Opto-Electronics and Photonics
Quantum Mechanics
Radiation Detectors
Solid State Physics

Solid state physics is a branch of physics that deals primarily with the physical properties of periodic condensed matters, especially the electromagnetic, thermodynamic, and structural properties of various systems, such as semiconductors, quantum structures, and superconductors; these properties are the consequences of solids interacting with light and under external fields, etc.

The *Field Guide to Solid State Physics* provides a compact introduction of select topics within the field of condensed-matter physics. For students and engineers alike, the book facilitates an in-depth understanding of physical concepts, as well as their applications, to help them develop new ideas for innovative devices. The topics chosen were influenced by our own areas of interest: single-particle and many-body interactions in the form of quasi-particle and collective excitations. Whenever possible, simple line art illustrates the essential concepts.

Over the last few decades, we have witnessed the significant (and increasing) effect of solid state physics on everyday life. The field is essential for the development of state-of-the-art concepts because it provides effective guidance for designing circuits and new materials for electronic and spintronic devices (it contributed to both the transistor and the semiconductor chip). Over the last ten years, more than half of the Nobel Prizes in physics were awarded to topics relevant to solid state physics.

Special thanks to Scott McNeill (SPIE Press) for a very careful reading of the manuscript and introducing a significant number of improvements.

Marek S. Wartak
Wilfrid Laurier University

C. Y. Fong
University of California, Davis

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Glossary of Symbols

0D	Zero-dimensional
1D	One-dimensional
2D	Two dimensional
3D	Three-dimensional
$\mathbf{A}$	Vector potential
$\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$	Vectors defining unit cell
$\mathbf{B}$	Magnetic field
$\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$	Vectors defining reciprocal lattice
bcc	Body-centered cubic
BCS	Bardeen–Cooper–Schrieffer
BJT	Bipolar junction transistor
BL	Bravais lattice
BZ	Brillouin zone
CB	Conduction band
CB	Coulomb blockade
CBM	Conduction band minimum
c_V	Specific heat at constant volume
D_n	Diffusion coefficient
DFT	Density functional theory
dHvA	de Haas–van Alphen
DOS	Density of states
$\mathbf{E}$	Electric field
E_c	Charging energy
E_F	Fermi energy
E_g	Bandgap energy
F	Free energy
$f_{FD}(\varepsilon)$	Fermi-Dirac distribution function
fcc	Face-centered cubic
FD	Fermi–Dirac
FQHE	Fractional quantum Hall effect
FS	Fermi surface
GaAs	Galium arsenide
Ge	Germanium
GL	Ginzburg–Landau
H	Hamiltonian
hcp	Hexagonal close-packed
HRs	Hund's rules
$I[f]$	Collision integral
$\mathbf{J}$	Total angular momentum

Glossary of Symbols

$\mathbf{k}$	Wavevector
$\mathbf{K}$	Reciprocal lattice vector
KS	Kohn–Sham
$\mathbf{L}$	Total orbital momentum
LDA	Local density approximation
LED	Light-emitting diode
$\mathbf{M}$	Magnetization
m^*	Effective mass
MBE	Molecular beam epitaxy
MOS	Metal–oxide–semiconductor
MOSFET	Metal–oxide–semiconductor field effect transistor
MRI	Magnetic resonance imaging
n	Density of electrons
N_A	Concentrations of acceptors
N_D	Concentrations of donors
n_i	Intrinsic concentration
$n_{\mathbf{q}}(\mathbf{r},t)$	Local density of phonons
NMR	Nuclear magnetic resonance
NN	Nearest neighbor
p_F	Fermi momentum
R_H	Hall coefficient
rf-SQUID	Radio-frequency superconducting quantum interference device
RKKY	Ruderman–Kittel–Kasuya–Yosida
RL	Reciprocal lattice
$\mathbf{S}$	Spin angular momentum operator
SET	Single-electron transistor
Si	Silicon
SOI	Silicon-on-isolator
T	Absolute temperature
T_1	Longitudinal relaxation time
T_2	Transverse relaxation time
T_C	Critical temperature
TB	Tight binding
$u_{\mathbf{k},n}(\mathbf{r})$	Periodic part of the Bloch function
$V(\mathbf{r})$	Periodic potential
(h,k,l)	Miller indices
$\hbar$	Dirac constant

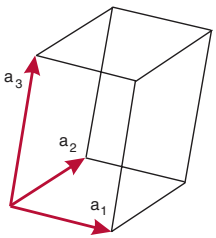
Glossary of Symbols

Δ	Order parameter
$\varepsilon(\omega)$	Electrical permittivity
λ	Wavelength
$\boldsymbol{\mu}$	Magnetic moment
μ_B	Bohr magneton
μ_n	Mobility
$\rho(E)$	Density of states
σ	Electrical conductivity
τ	Average time between collisions
ϕ_0	Flux quantum
Φ	Magnetic flux
χ	Magnetic susceptibility
ψ	Wave function
$\psi_{\mathbf{k},n}(\mathbf{r})$	Bloch function
ω	Frequency
ω_c	Cyclotron frequency
ω_p	Plasma frequency
Ω	Volume of unit cell in real space
$\Omega_{\mathbf{K}}$	Volume of the primitive cell in the reciprocal lattice
QD	Quantum dot
QPC	Quantum point contact
VB	Valence band
VBM	Valence band maximum
VLSI	Very-large-scale integration
WS	Wigner–Seitz

Primitive Cell

A substance is considered to be in a crystalline state (known as a **crystal**) if its constituent atoms are arranged in a regular and repetitive pattern.

Crystals can be described by an object that, when repeated, generates the entire crystal. Such an object can be considered as a "building block" called a **unit cell**. The unit cell is spanned by three independent vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$, which may or may not lie along the Cartesian coordinate axes. It contains only one lattice point, and its volume is



$$\Omega = (\mathbf{a}_1 \times \mathbf{a}_2) \cdot \mathbf{a}_3$$

Its vertices define the lattice points. Each unit cell has the same shape, volume, and atomic arrangements. The choice of unit cell is not unique, and there is an infinite number of possible ways to choose a unit cell. The smallest possible unit cell is called the **primitive cell**. The vectors that define a primitive cell are called **primitive lattice vectors**.

The symmetry of a crystal is determined by the arrangement of atoms in its unit cell. A simple cubic unit cell is the simplest 3D case. Each of its eight corners contains identical atoms. More-complicated cases can have atoms in the center of the faces or at the center of the cube, known as a face-centered or body-centered cubic unit cell, respectively. They are important structures because many metals, ionic solids, and intermetallic compounds crystallize in such structures. Experimental techniques for studying crystal structures include

- x-ray diffraction structural analysis,
- neutron diffraction, and
- electron diffraction.

Bravais Lattice

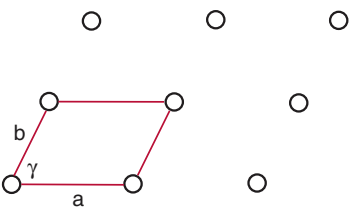
An ideal crystal, i.e., without defects or imperfections, consists of a regular arrangement of atoms. A **Bravais lattice (BL)** is used to characterize such systems. It is a mathematical set of points that are equivalent to each other. The various symmetry operations (see "Elements of Symmetry") possible for a 3D set of points result in 14 different BLs. Each lattice represents a set of points in a space that forms a periodic structure. Each point has exactly the same environment. The **lattice points** are specified by a **lattice vector R** as

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$$

where n_1, n_2, and n_3 are integers, and $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are three independent vectors.

Representative unit cells of the 14 BLs are shown on page 4. An important role is played by a face-centered cubic lattice, which is a structure that appears in many semiconductors.

It is often important to know the arrangement of neighboring atoms, especially the **nearest neighbors (NNs)**. The number of NNs of an atom is known as the **coordination number**. A BL can also be introduced in two dimensions; a 2D lattice can be specified by vectors **a** and **b** and the angle γ (of specific value) between them. Other 2D lattices are square, rectangular, centred rectangular, and hexagonal. Symmetry operations, like those for 3D strustures, can be used to describe 2D structures.



For crystals with many atoms, a building block of atoms, called the basis, is associated with each lattice point. Together, they create the crystal structure.

Elements of Symmetry

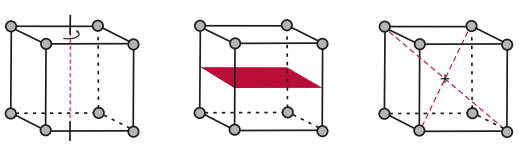
An ideal crystal possesses a certain type of **symmetry**. Crystal symmetry is defined by a group of operations that leave the crystal invariant. Operations that are applied through some fixed points within a unit cell are called **point group operations**.

Operations (coordinate transformations) include reflections, rotations, inversions, and translations by a vector of a fraction of a lattice vector. If the translations are followed by the point group operations, then these operations define a space group. Typically, a unit cell is expected to be invariant under certain transformations, which defines the symmetries possessed by a crystal. In crystals, rotations are only allowed for particular angles $\theta_n = 2\pi/n$, where $n = 2, 3, 4, 6$, consistent with translations of lattice vectors. Such an axis of symmetry is of the n^{th}-fold axis and is denoted as C_n. A cube has three types of symmetrical axes:

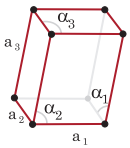
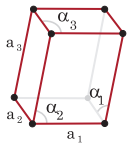
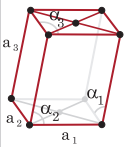
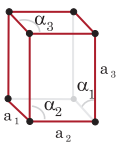
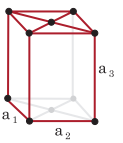
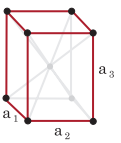
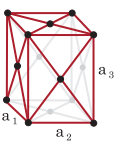
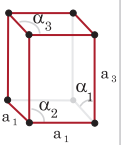
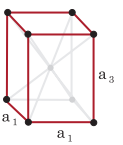
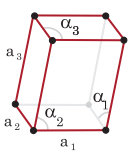
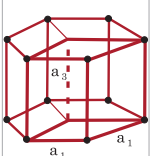
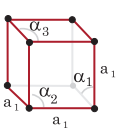
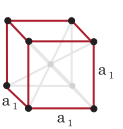
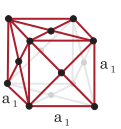
- three C_4 axes that pass through the centers of opposite faces,
- four C_3 axes that are along diagonals, and
- six C_2 axes that pass through the midpoints of opposite edges.

Each axis has counterclockwise and clockwise rotations. The same rule applies to other rotational axes.

Some symmetry operations for the simple cubic structure are illustrated below. The left image shows rotation through an axis, the middle image shows mirror reflection with respect to a plane, and the right image shows inversion through a point. All of these operations pass through the center of the cube. There are a total of 32 point symmetry groups that describe various properties of solid state systems.



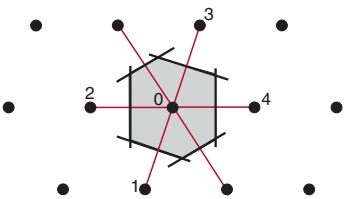
Summary of Bravais Lattices

	Simple	Base-centered	Body-centered	Face-centered
Triclinic				
Monoclinic				
Orthorhombic				
Tetragonal				
Rhombohedral				
Hexagonal				
Cubic				

Wigner–Seitz Cell

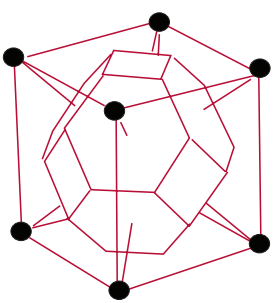
A **Wigner–Seitz (WS) cell** is a special primitive cell that contains one lattice point. The WS cell is defined for a general lattice as the smallest polyhedron bounded by planes that are the perpendicular bisectors joining one lattice point to the others. The WS cell has the smallest possible volume (3D) and area (2D).

The WS cell in the reciprocal lattice defines the first BZ and has the full symmetry of the BL. It is constructed around an arbitrary lattice point, and when translated over all lattice points, it will fill the space of the entire crystal without overlapping.



It is difficult to illustrate the construction of a WS cell in 3D, so the image here depicts the construction of a WS cell in 2D real space.

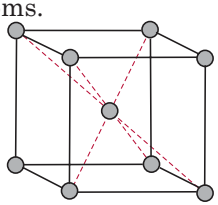
- Select an arbitrary atom 0.
- Determine the NN atoms of the same species as the chosen atom.
- Draw lines from the atom 0 to all NNs, e.g., lines from atom 0 to all atoms 1, 2, 3, 4.
- Draw lines (planes, in 3D) passing through the middle of the lines. The cell enclosed by those lines (planes) is the WS primitive cell.



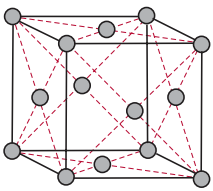
The 3D WS cell for the body-centered cubic lattice shown here is a truncated octahedron with volume $\frac{1}{2}a^3$. The cube around it is a conventional body-centered cubic cell with a lattice point on each vertex and at its center.

Body- and Face-Centered Cubic Structures

Body-centered cubic (bcc) and **face-centered cubic (fcc) structures** are important lattices because they appear in many solids, e.g., Ar, Ag, and Ce (fcc) and Ba, Cr, Fe, K, and Li (bcc). Both structures are modified from the simple cubic system and are shown here with the locations of the atoms.



bcc



fcc

The bcc structure has eight atoms at the corners and one atom at the body center. Assume that $\mathbf{a}\hat{x}$, $\mathbf{a}\hat{y}$, $\mathbf{a}\hat{z}$ are primitive vectors of the simple cubic lattice, where $\hat{x}$, $\hat{y}$, $\hat{z}$ are orthogonal unit vectors. They also define the conventional cell of the bcc structure, which is also determined by the following vectors for this cell (not shown here):

$$\mathbf{a}_1 = \mathbf{a}\hat{x}, \mathbf{a}_2 = \mathbf{a}\hat{y}, \mathbf{a}_3 = \frac{\mathbf{a}}{2}(\hat{x} + \hat{y} + \hat{z})$$

The bcc structure can also be described by the more symmetric set of vectors

$$\mathbf{a}_1 = \frac{\mathbf{a}}{2}(\hat{y} + \hat{z} - \hat{x}), \mathbf{a}_2 = \frac{\mathbf{a}}{2}(\hat{x} + \hat{z} - \hat{y}), \mathbf{a}_3 = \frac{\mathbf{a}}{2}(\hat{x} + \hat{y} - \hat{z})$$

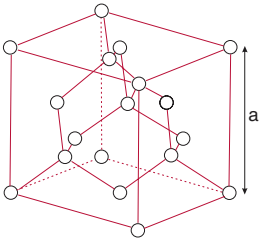
In this structure, the **coordination number**, i.e., the number of NNs, is 8. The bcc structure shown here belongs to the body-centered BLs.

Similarly, a symmetric set of primitive vectors for a fcc lattice is

$$\mathbf{a}_1 = \frac{\mathbf{a}}{2}(\hat{y} + \hat{z}), \mathbf{a}_2 = \frac{\mathbf{a}}{2}(\hat{x} + \hat{z}), \mathbf{a}_3 = \frac{\mathbf{a}}{2}(\hat{x} + \hat{y})$$

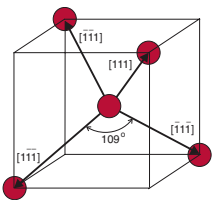
There are only two fcc lattices among the 14 BLs—one in the cubic system, and another in the orthorhombic system.

Diamond Structure



In a **diamond structure** in the conventional fcc lattice (shown here with the locations of atoms and the bonds (lines) between them), all atomic sites are occupied by carbon atoms. Semiconductors such as diamond (C), silicon (Si), germanium (Ge), and grey tin (α-Sn) crystallize in this structure.

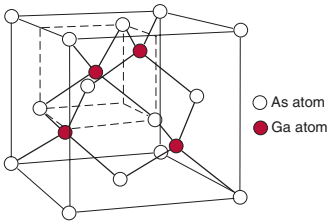
The structure is a combination of two identical interpenetrating fcc lattices. One of the sublattices is shifted along the body diagonal of the cubic cell by one quarter of the length of the diagonal. The diamond structure is thus fcc with a basis containing two identical atoms.



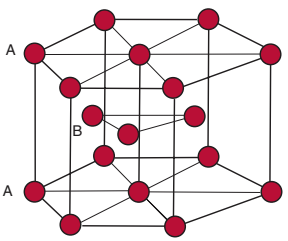
A simple method for constructing a diamond lattice considers it as a fcc structure with an extra atom placed at $\frac{1}{4}\mathbf{a}_1 + \frac{1}{4}\mathbf{a}_2 + \frac{1}{4}\mathbf{a}_3$ from each of the fcc atoms. The basic element of the structure is a tetrahedron where a C atom is at the center, and its four NNs are at the corners of the cube (or vice versa). Each atom forms four bonds with its NNs. Atoms in diamond-type crystals form covalent bonding. The bonding energy is associated with the shared valence electrons between atoms and depends on the relative orientation of atoms.

The atomic arrangement in the diamond structure helps explain its mechanical, chemical, and metallurgical properties. These semiconductor crystals can be cleaved along certain atomic planes to produce excellent planar surfaces, e.g., diamonds used in jewelry. Such surfaces are used as Fabry–Pérot reflectors in semiconductor lasers. Chemical reactions performed with such crystals, such as etching, often occur preferentially in certain directions.

Zincblende and Hexagonal Structures



A **zincblende structure** has the form of a diamond except that the two sublattices are occupied by different types of atoms, e.g., As and Ga, as shown here. It is a fcc structure with a two-atom basis. The structure is typical for semiconductors such as GaAs, ZnS, ZnSe, SiC, and GaP. For example, in ZnS, the Zn atoms occupy positions $(0,0,0)$, $(0,\frac{1}{2},\frac{1}{2})$ and $(\frac{1}{2},0,\frac{1}{2})$, $(\frac{1}{2},\frac{1}{2},0)$, and S atoms occupy positions $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$, $(\frac{1}{4},\frac{3}{4},\frac{3}{4})$ and $(\frac{3}{4},\frac{1}{4},\frac{3}{4})$, $(\frac{3}{4},\frac{3}{4},\frac{1}{4})$. Each atom has four equally distant neighbors of the opposite type located at the corners of a regular tetrahedron. There are four molecules of ZnS per conventional cell. In this compound, the presence of two different kinds of atoms means that the system does not have inversion symmetry.



Another important structure is the **hexagonal close-packed (hcp) structure**. All atoms in the A-plane have an identical environment and can be taken as lattice points. Atoms that form the B-plane "feel" different environmentally and do not lie on lattice points. The unit cell contains a basis of an A atom at (0, 0, 0) and a B atom at $(\frac{2}{3},\frac{1}{3},\frac{1}{2})$. There are two atoms in each primitive unit cell. The close-packed A-planes form two horizontal planes, and the B-planes are halfway between. The fcc structure has a hexagonal packing in the direction along [111], the cube diagonal. However, it is a close packing with three layers ABC instead of two, as in the hcp (e.g., Ti, Mg, Cd, Zn, He).

Reciprocal Lattice Space

Reciprocal lattice (RL) space, often called **k**-space, is used to represent wave propagation processes in a solid; its dimensionality coincides with that of a wavevector. Consider a periodic function $f(\mathbf{r})$ in a crystal, i.e., obeying conditions $f(\mathbf{r} + \mathbf{R}) = f(\mathbf{r})$, where $\mathbf{R}$ is a lattice vector. The function can be expanded in a Fourier series

$$f(\mathbf{r}) = \sum_{\mathbf{K}} A_{\mathbf{K}} e^{i\mathbf{K}\cdot\mathbf{r}}$$

where $A_{\mathbf{K}} = \frac{1}{\Omega}\int_{\Omega} f(\mathbf{r}) e^{-i\mathbf{K}\cdot\mathbf{r}} d\mathbf{r}$, and Ω is the volume of a unit cell. Vectors $\mathbf{K}$ have dimensions of 1/length and are therefore called RL vectors. They can be expressed as

$$\mathbf{K} = h_1\mathbf{b}_1 + h_2\mathbf{b}_2 + h_3\mathbf{b}_3$$

where h_1, h_2, and h_3 are integers. Vectors $\mathbf{b}_i$ are basis vectors in **k**-space and are given in terms of basis vectors in direct space as

$$\mathbf{b}_1 = \frac{2\pi}{\Omega}(\mathbf{a}_2 \times \mathbf{a}_3), \quad \mathbf{b}_2 = \frac{2\pi}{\Omega}(\mathbf{a}_3 \times \mathbf{a}_1), \quad \mathbf{b}_3 = \frac{2\pi}{\Omega}(\mathbf{a}_1 \times \mathbf{a}_2)$$

Both set of vectors have the following property:

$$\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$$

Vectors $\mathbf{b}_i$ formally define a RL, which is a mathematical construct important in the study of quantum properties of solids. They are determined by vectors defining the unit cell, i.e., vectors $\mathbf{a}_j$. Therefore, a RL has a definite structure determined by the crystal. For example, if the BL is a bcc, its RL is face centered. The volume of the primitive cell in the reciprocal lattice is

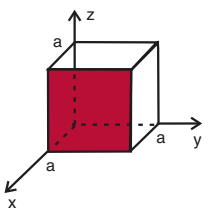
$$\Omega_{\mathbf{K}} = \mathbf{b}_1 \cdot (\mathbf{b}_2 \times \mathbf{b}_3) = \frac{(2\pi)^3}{\Omega}$$

where Ω is the volume of the primitive cell in the related real space.

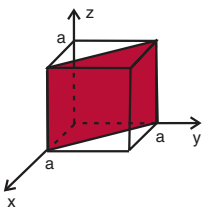
Miller Indices

In order to specify directions and planes in 3D crystals, e.g., when determining the crystal structure, one must specify a set of periodic planes. In 3D space, there are many possibilities to form such a set of parallel planes that act as reflectors (as illustrated by the Bragg reflections), but the **Miller indices** consist of a set of three integers that conveniently label crystal planes and also identify crystallographic directions. The planes are specified by (*hkl*). The Miller indices of a crystal plane are determined as follows:

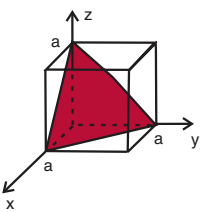
1. Find the intercepts with a plane on the $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ axes in units of the respective lattice constant. The intercepts are labeled as (l_1, l_2, l_3).
2. Take the reciprocals of the intercepts $(\frac{1}{l_1}, \frac{1}{l_2}, \frac{1}{l_3})$.
3. Convert them to integers by multiplying by the smallest constant that makes them integers (*hkl*).



(100)



(110)



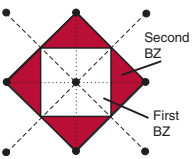
(111)

The Miller indices for the most popular orientations of cubic systems are shown here. For example, (100) is the plane perpendicular to $\mathbf{a}_1$.

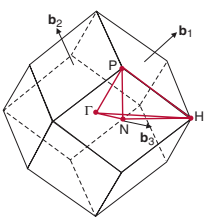
Other crystals, such as the trigonal and hexagonal, use four Miller indices. Convention states that the first three indices refer to three coplanar directions in the basal plane, and the fourth indicates the uniaxial axis, or *c* axis.

Brillouin Zones

A Fourier series can be carried out to introduce the unit cell in **k**-space and define the **Brillouin zone (BZ)**; a 2D case is considered here for simplicity. Construction of the BZ follows the procedure of the Wigner–Seitz construction in direct space. The first BZ is established by selecting an arbitrary lattice point in the reciprocal space and connecting it by lines to its NN points. The lines are then bisected by the perpendicular planes. The interior enclosed by those planes forms the first BZ. The second BZ is created by connecting the selected lattice point to the next NN, repeating the previous construction (bisection) and subtracting it from the first BZ.



The vector in the first BZ is connected to the vector in the other BZ by a reciprocal lattice vector **K**. Let the vector in the first BZ be **k**, then **k**′ in the other BZ is given by $\mathbf{k}' = \mathbf{k} + \mathbf{K}$. The **k**-vectors in the first BZ are the most important vectors for a periodic condensed matter. Each of these vectors can be associated with a wave in real space. The wave associated with any **K**-vectors has a wavelength shorter than the lattice constant. An example of the first BZ (in 3D) of the bcc lattice, shown here, is known as a regular rhombic dodecahedron. The red lines are the irreducible part of the first BZ. The standard labeling of the symmetry **k**-points is indicated, and the center point is labeled as Γ.



Typically, band-structure calculations give the band energies along the lines from the center to one of these **k**-points or between these points. The concept of the BZ is also used to treat amorphous materials, which, like glasses, have no periodic structure. It is impossible to define a unit cell. However, it can be modeled by using a large unit cell with randomly arranged atoms inside. Only the $\mathbf{k} = 0$ point has a physical meaning.

Bands and Molecular Bonds

Electronic properties, including the optical properties of a solid, can be characterized by using the so-called one-electron model, even though it can involve interactions with many electrons, or by having an orbital occupied by an electron interact with electrons that occupy other orbitals and atomic potentials of neighboring atoms under the periodic boundary condition. The energies of the electrons are thus broadened. With periodic boundary conditions, **bands** are formed instead of energy levels, as in isolated atoms. Two scenarios apply to band formation.

- With **tight binding**, atomic levels and orbitals are the essential physical ingredients. When the atoms are infinitely distant, their s- and p-states are degenerate, respectively.

  As the atoms are brought together, the s- and p-orbitals of one atom overlap with other s- and p-orbitals of its neighboring atoms, respectively, with possibly different quantum numbers. The overlaps lift the respective degenerate s- and p-states, and thus the energy spectra of the states are broadened and form bands.

  When the atoms are brought closer, the s- and p-states overlap. These s- and p-states can form states with mixed s- and p-characteristics, called **s-p hybridization**. Depending on the atoms, it is also possible to have s-d and d-p hybridizations when the atoms have d-states.

- With **free electrons**, an electron is presumed to not be initially subjected to any periodic potential. Its energy consists of the kinetic energy only, expressed as $\hbar^2\mathbf{k}^2/(2m)$. Note that energy is a function of $\mathbf{k}$ (1D), the momentum of the electron. The energy is simply a parabola in $\mathbf{k}$-space. When a weak periodic potential is introduced, the BZs appear, and at BZ boundaries the potential breaks the parabolic energy curves and creates gaps. The states between gaps form bands.

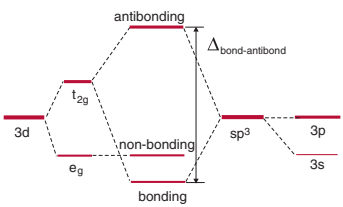
Molecular Orbitals

The essential difference between molecules and bands is that molecules do not have periodicity. Electrons in the molecules interact with nuclei and among themselves, forming **molecular orbitals** via three approaches:

Tight binding: Since a molecule lacks periodic structure, tight binding can show how the neighboring atomic orbitals overlap to form molecular orbitals.

One-electron model: The Hamiltonian for a molecule is specified by one-electron kinetic energy, the nuclei–electron interaction, and the effective electron–electron interaction acting on one electron. The last interaction can be treated in two ways:

- By the overlapping of the atomic orbitals centered at the NN atomic pairs. They can be fitted to some experimental results or calculated with the considered atomic orbitals.
- By first-principles self-consistent calculations.

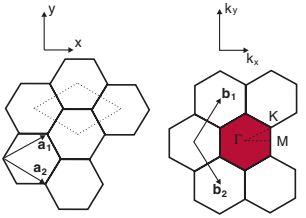


Bonding and antibonding orbitals are formed when overlaps occur between any atomic orbital and its neighboring orbitals. The lower energy of the molecular orbital is called the **bonding state**, whereas the higher-energy state is called the **antibonding state**. For the ground state of a molecule, the bonding state is occupied, but the antibonding state is not. If different atomic orbitals—such as the s-orbital of one atom and the p-state of the neighboring atom—overlap, the resultant states are called **hybridized states**.

Many-electron model: A hydrogen molecule (H_2) has two electrons. A molecule is a many-electron problem; electrons are fermions and obey the Pauli principle.

Graphene Structure

Graphene is a single sheet of hexagonally arranged carbon atoms in a periodic honeycomb structure. The unit cell (left) of graphene and the first BZ (right) are shown here. The depicted basic lattice vectors are



$$\mathbf{a}_1 = \left(\sqrt{3}\mathbf{a}_x + \mathbf{a}_y\right)a/2 \text{ and } \mathbf{a}_2 = \left(\sqrt{3}\mathbf{a}_x - \mathbf{a}_y\right)a/2$$

where $a = \sqrt{3}b$, and $b = 0.142$ nm is the interatomic distance between carbon atoms in graphene. The high-symmetry points in the BZ are

$$\Gamma = (0,0), \quad \mathbf{K} = \left(\frac{2\pi}{\sqrt{3}a}, \frac{2\pi}{3a}\right), \quad \mathbf{M} = \left(\frac{2\pi}{\sqrt{3}a}, 0\right)$$

The resulting 2D dispersion relation in the vicinity of a particular electron state is approximately $E(k_x, k_y) = \pm\gamma_0 w(k_x, k_y)$, where $\gamma_0 \approx 3$ eV, and

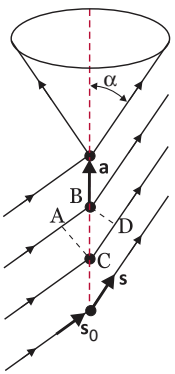
$$w(k_x, k_y) = \sqrt{1 + 4\cos\left(\frac{\sqrt{3}k_x a}{2}\right)\cos\left(\frac{k_y a}{2}\right) + 4\cos^2\left(\frac{k_y a}{2}\right)}$$

A unique feature of a graphene is the band at one of the BZ boundaries: the **K**-point, where the top of the valence bands and the bottom of the conduction bands meet. The dispersion near the **K**-point is linear in both bands. Therefore, the band structure demonstrates the dispersion of a Dirac electron.

It is now possible to stack graphene in the direction perpendicular to the layer. The layers can be held together by the **van der Waals interaction** $\sim(1/r_{ij}^6)$; their origin is the overlap of the p_z states of the carbon atoms. The interaction, in fact, happens in many layered materials, such as transition metal diselenides.

von Laue Method

The internal structure of crystals can be examined by performing diffraction experiments using electromagnetic waves of the appropriate wavelength comparable to the characteristic dimension within the crystal, which is a typical distance between the atoms. Ordinary light is not suitable as its wavelength is too large, but x-rays with wavelengths of a few angstroms are good candidates.



The practical method of analysis was developed by von Laue. The figure here illustrates a 1D crystal illuminated by x-rays that are scattered by atoms. The rays propagate in all directions. The next task is to analyze the rays scattered at an angle α defined with respect to the axis determined by scattering atoms. The distance between atoms is a, and vectors $\mathbf{s}_0$ and $\mathbf{s}$ represent the directions of the incoming and scattered radiation, respectively. The optical path difference is (assuming the refractive index is 1) $\Delta = CD - AB$. When $\Delta = n\lambda$ (where n is a natural number), the scattered radiation is amplified along the cone. The condition for constructive interference is $a(s - s_0) = h\lambda$. For a 3D crystal and three crystallographic axes, the above formula is generalized and leads to the three Laue equations:

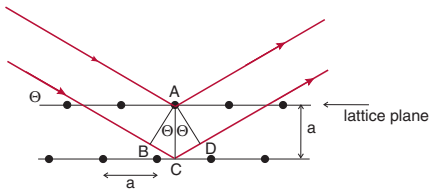
$$a_1(s - s_0) = h\lambda$$
$$a_2(s - s_0) = k\lambda$$
$$a_3(s - s_0) = l\lambda$$

where $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are translation vectors, and h, k, and l are natural numbers (Miller indices). The interpretation of experiments is based on the assumption that atoms in the crystal are periodically distributed. The largest intensity of the diffracted x-rays happens when the rays scattered on three crystallographic axes are combined.

Bragg's Law

The Laue equations lead to **Bragg's law**, which is a simple representation of x-ray diffraction in crystals, by squaring the equations and adding them together. The formula for the square of cosines can be used with the equation $\frac{1}{d^2} = \frac{h^2}{a^2} + \frac{k^2}{b^2} + \frac{l^2}{c^2}$ to produce the Bragg equation:

$$2a \sin \theta = n\lambda$$



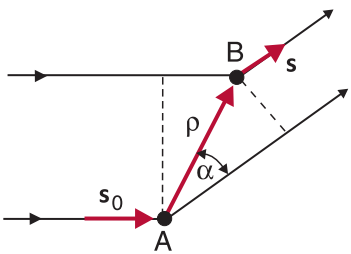
The Bragg equation can also be found by considering x-ray diffraction by the planes of atoms. The Bragg condition is obtained for constructive interference between the x-rays from different crystal planes. The expression states that if the path difference, i.e., the distance $BC + CD$, between two x-ray beams equals the integer multiples of the wavelength, a constructive interference results. There are two issues with this simple description:

1. It is unknown how the incident x-ray interacts with the atoms in the planes to produce the reflected x-rays. Note that the electron generates the reflected x-rays. When an x-ray is incident to a plane of atoms, the light mass of electrons causes the electrons to oscillate in the opposite direction of the electric field of the x-ray. This oscillation with respect to the heavy mass nucleus induces a dipole, which in turn radiates an x-ray. The wavefront defines the diffracted x-rays.
2. The Miller indices of the planes in Bragg's law are also unknown. The sample must be cut to prepare it for an x-ray measurement. The normal of the surface is determined by the rotation method (not discussed here). The intercepts of a plane along the basis vectors of the unit cell can be found from the normal and the unit cell. From the intercepts, the Miller indices can be easily determined.

von Laue Conditions for X-Ray Diffraction

von Laue suggested that

1. The crystal should be treated as arrays of points in a macroscopically small but microscopically large region.
2. The incident x-ray should be treated as a plane wave.
3. The scattered x-ray should be treated as a spherical wave radiated from a certain point in the array.



Consider the phase shift between the rays scattered by two atoms A and B. Assume that atom A is located at the center of a coordinate system, and the location of atom B is determined by vector $\boldsymbol{\rho}$, which can be expressed in terms of the coordinates of point $B(u,v,w)$ as $\boldsymbol{\rho} = u\mathbf{a}_1 + v\mathbf{a}_2 + w\mathbf{a}_3$. The path difference between the two rays is $\Delta = \rho(\mathbf{s} - \mathbf{s}_0)$, where $\mathbf{s}_0$ and $\mathbf{s}$ are the incoming and scattered rays, respectively. Using the above relations, the phase shift of the interfering rays is

$$\varphi = \frac{2\pi}{\lambda}[u\mathbf{a}_1 \cdot (\mathbf{s} - \mathbf{s}_0) + v\mathbf{a}_2 \cdot (\mathbf{s} - \mathbf{s}_0) + w\mathbf{a}_3 \cdot (\mathbf{s} - \mathbf{s}_0)]$$

After applying the Laue relations, phase shift can be expressed as $\varphi = \frac{2\pi}{\lambda}(uh + vk + wl)$. The total amplitude of all of the scattered rays on atoms is

$$F_{hkl} = \sum_{j=1}^{N} f_j \exp[2\pi i(u_j h + v_j k + w_j l)]$$

where f_j is the scattering amplitude on j^{th} atom, and u_j, v_j, and w_j are coordinates of the j^{th} atom. The intensity equals $I_{hkl} \sim |F_{hkl}|^2$, and thus its value for different values of hkl allows for the location of atoms in the unit cell. Atoms at the given positions generate unique intensities that can be experimentally observed.

Free-Electron Approximation

Various approximations describe electrons in metals. The omission of electron–ion interactions is known as the **free-electron approximation**. For simple metals, nuclei can be considered to provide a smeared and positively uniform background that creates a constant potential that acts on electrons. Additionally, electrons are assumed to not interact among themselves (hence the term "free"). A constant potential can be set to zero. An electron is assumed to be located within a cube of volume $V = L^3$, and the whole space is filled with such boxes. Periodic boundary conditions are imposed. Thus, the following Schrödinger equation describes an electron:

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) = \varepsilon\psi(\mathbf{r})$$

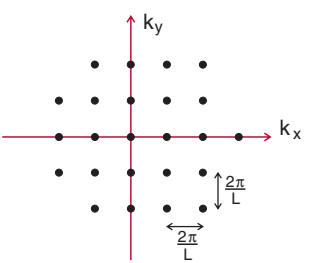
with periodic boundary conditions on the wave function

$$\psi(x+L,y,z) = \psi(x,y,z), \quad \psi(x,y+L,z) = \psi(x,y,z),$$
$$\text{and } \psi(x,y,z+L) = \psi(x,y,z)$$

The normalized solution of the Schrödinger equation is $\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}}e^{i\mathbf{k}\cdot\mathbf{r}}, (\Omega = L^3)$ with the eigenvalue $\varepsilon_{\boldsymbol{k}} = \frac{\hbar^2k^2}{2m}$. Vector $\mathbf{k}$ is the eigenvalue of the linear momentum of the electron $\hat{\mathbf{p}}$ as one has the relation $\hat{\mathbf{p}}\psi(\mathbf{r}) = \hbar\mathbf{k}\psi(\mathbf{r})$. The boundary conditions of the wave function determine the allowed values of the $\mathbf{k}$-vector to be

$$k_x = \frac{2\pi}{L}n_x, k_y = \frac{2\pi}{L}n_y, k_z = \frac{2\pi}{L}n_z$$

where n_x, n_y, and n_z take the values of 0, ±1, ±2, ... The allowed values of the $\mathbf{k}$-points in two dimensions are shown here. The number of allowed $\mathbf{k}$-values per unit volume of $\mathbf{k}$-space (density of $\mathbf{k}$-points) is $\Omega/(2\pi)^3$.



Bloch Theorem and Function

The crystal potential is periodic, i.e., it obeys the relation $V(\mathbf{r}) = V(\mathbf{r} + \mathbf{R})$, where $\mathbf{R}$ is a lattice vector. Electrons moving in a periodic potential are known as **Bloch electrons**, which are described by the one-electron Schrödinger equation

$$H\psi_{\mathbf{k},n}(\mathbf{r}) = \varepsilon_{\mathbf{k}}\psi_{\mathbf{k},n}(\mathbf{r})$$

where the Hamiltonian contains the kinetic energy and a periodic potential. It has the form

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})$$

The eigenstates $\psi_{\mathbf{k},n}(\mathbf{r})$ (**Bloch functions**) of the above Hamiltonian have the form of a plane wave times a function, which has the periodicity of the lattice (**Bloch theorem**)

$$\psi_{\mathbf{k},n}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k},n}(\mathbf{r})$$

where $u_{\mathbf{k},n}(\mathbf{r})$ is a periodic function:

$$u_{\mathbf{k},n}(\mathbf{r}) = u_{\mathbf{k},n}(\mathbf{r} + \mathbf{R})$$

Bloch functions are typically supplemented by the periodic boundary conditions (known as the **Born–von Karman conditions**), which impose conditions on the allowed values of $\mathbf{k}$ and can be expressed as

$$\psi_{\mathbf{k},n}(\mathbf{r} + N_i\mathbf{a}_i) = \psi_{\mathbf{k},n}(\mathbf{r}),$$

where $\mathbf{a}_i$ are appropriate vectors, and N_i are integers that determine the number of unit cells in the i^{th} direction. Boundary conditions imply that $\exp(iN_i\mathbf{k}\cdot\mathbf{a}_i) = 1$. The allowed wavevectors of the Bloch function can be expressed in the form

$$\mathbf{k} = \sum_{i=1}^{3}\frac{m_i}{N_i}\mathbf{b}_i$$

where m_i are integers, and $\mathbf{b}_i$ are the primitive reciprocal lattice vectors. From the above, one determines the volume in $\mathbf{k}$-space occupied by each state, which is

$$\Delta_{\mathbf{k}} = \frac{\mathbf{b}_1}{N_1}\cdot\left(\frac{\mathbf{b}_2}{N_2}\times\frac{\mathbf{b}_3}{N_3}\right)$$

Nearly-Free-Electron Model

The effect of the periodic potential of an electron is analyzed based on the Schrödinger equation and by expanding the periodic potential $V(\mathbf{r})$ and Bloch wave function $\psi_{n\mathbf{k}}(\mathbf{r})$ in a Fourier series as $V(\mathbf{r}) = \sum_G V_G \exp(i\mathbf{G}\cdot\mathbf{r})$ and $\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}'} C_{n,\mathbf{k}-\mathbf{G}'} \exp(i(\mathbf{k}-\mathbf{G}')\cdot\mathbf{r})$. After some manipulations,

$$(E - E_{\mathbf{k}-\mathbf{G}'})C_{n,\mathbf{G}'} = \sum_{\mathbf{G}} V_{\mathbf{G}-\mathbf{G}'} C_{n,\mathbf{G}'}$$

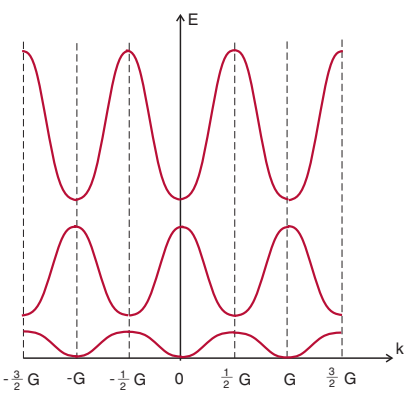
To illustrate how to reach these solutions, only the cases where $\mathbf{G}' = 0$ and $\mathbf{G}' = \mathbf{G}$ are considered. The following system determines the energies:

$$(E - E^0_{\mathbf{k}})C_0 = V_0 C_0 + V_{\mathbf{G}} C_{\mathbf{G}}$$
$$(E - E^0_{\mathbf{k}-\mathbf{G}})C_{\mathbf{G}} = V_{-\mathbf{G}} C_0 + V_0 C_{\mathbf{G}}$$

To determine the eigenstate E, evaluate a determinant of the above system to get

$$E_{\mathbf{k}\pm} = \frac{1}{2}(E^0_{\mathbf{k}} - E^0_{\mathbf{k}-\mathbf{G}}) \pm \frac{1}{2}\left[(E^0_{\mathbf{k}} - E^0_{\mathbf{k}-\mathbf{G}})^2 + 4|V_{\mathbf{G}}|^2\right]^{1/2}$$

We also used the fact that $V_{-\mathbf{G}} = V_{\mathbf{G}}{}^*$. The energy for a 1D crystal is illustrated here with the so-called **repeated zone representation**. The main difference from the free-electron model is a deviation from the parabolic dispersion relation at the BZ boundaries and the formation of energy bands.



Tight Binding

In explaining the details of band structure in solids, one evaluates the energy states of an electron using either the nearly-free-electron model (for weak potential) or the **tight-binding (TB) approximation** in the strong potential limit. This approximation uses an almost-atomic-like description of electronic states. Atomic states obey the equation

$$H_{at}\phi_n^{at}(\mathbf{r}) = E_n^{at}\phi_n^{at}(\mathbf{r})$$

where H_{at} is a Hamiltonian that contains the linear combination of atomic potentials. In this approach, a solid is regarded as a collection of weakly interacting neutral atoms. Electron wave functions are mostly localized around the atomic site. The overlap of atomic wave functions can change the picture of the isolated atoms. The corrections to the atomic case are given as $H = H_{at} + \Delta V(\mathbf{r})$. The Schrödinger equation becomes $H\psi_{n\mathbf{k}}(\mathbf{r}) = E_{n\mathbf{k}}\psi_{n\mathbf{k}}(\mathbf{r})$. The wave function is constructed as a linear combination of localized functions of the form

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\phi_n^{at}(\mathbf{r} - \mathbf{R})$$

where $\phi_n^{at}(\mathbf{r} - \mathbf{R})$ is a wave function of the isolated atom, and $\mathbf{R}$ are the positions of the atoms in the crystal. The energy bands are generated by solving the Schrödinger equation and after manipulation can be expressed as

$$E_{n\mathbf{k}} = E_m^{at} - \alpha - \sum_{\mathbf{R}\neq 0}\beta_{\mathbf{R}}e^{i\mathbf{k}\cdot\mathbf{R}}$$

with the following definitions:

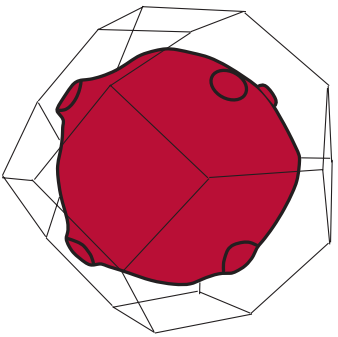
$$\alpha = -\int \psi_m^{*at}(\mathbf{r})\Delta V(\mathbf{r})\psi_n^{at}(\mathbf{r})d\mathbf{r}$$

$$\beta = -\int \psi_m^{*at}(\mathbf{r})\Delta V(\mathbf{r})\psi_n^{at}(\mathbf{r} - \mathbf{R})d\mathbf{r}$$

The constant α corresponds to on-site energy, whereas β is the overlapping integral and determines the widths of the energy bands.

Fermi Surface

A surface in the **k**-space of constant energy, which at $T = 0$ separates the filled electronic states from the unoccupied states, is known as the **Fermi surface (FS)**. It is a fundamental concept in understanding the behavior of electrons in metals. For example, the velocities of electrons have directions normal to the FS, as determined by $V = (\partial\varepsilon/\partial\mathbf{k})_{FS}$. The FS has periodicity in **k**-space as $\varepsilon_F = \varepsilon(\mathbf{k}) = \varepsilon(\mathbf{k} + \mathbf{b})$, where **b** is a reciprocal vector, and ε_F is the Fermi energy. Electrons with ε_F will have a wavevector that equals $\mathbf{k}_F$. For free electrons, the value of the Fermi momentum vector $\mathbf{k}_F$ is the same in all directions in the reciprocal space, and the FS becomes a sphere. Inside a real solid that is a 3D object, the FS can cross several bands, which causes its deformation, depending on the shape of the bands and number of electrons there. There is also an effect of Bragg reflection that occurs at the boundaries of the BZ. The FS of a real metal (e.g., silver) in the first BZ is shown here.



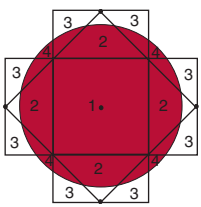
In the case when a spherical FS crosses the boundary of the BZ, it will be modified at different BZ boundaries. When the FS at high zones is mapped to the first zone, it can split into separate regions. If all of the regions are located within the first zone in **k**-space, it is called a closed FS. If the extended zone scheme is used, the FS extends continuously through **k**-space and is called an open FS. In the case of a free- and single-electron model (which means no potential and electron–electron interaction), the FS will be the region of the Fermi sphere within the first BZ due to the half-filled band.

Fermi Surface in 2D: Harrison's Method

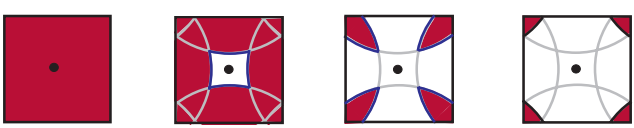
Plots of energy contours in **k**-space are the E-**k** (energy–wavevector) relations in 3D space. In 3D crystals, the E-**k** relation is described by constructing surfaces of constant energy. In 2D, the surfaces are curves, one of which corresponds to the Fermi energy (the highest energy of the occupied states). Harrison recognized that the distortion and/or discontinuity of the FS arise when the FS is in close proximity to the BZ boundaries, where a crystal symmetry controls the band shape.

The construction of the FS of free electrons suggested by Harrison, known as **Harrison's method**, works in 3D; the image here illustrates the general concept in the simpler, 2D case. The method works in **k**-space by starting with the given density of electrons n and determining the radius of the FS in the free-electron model as

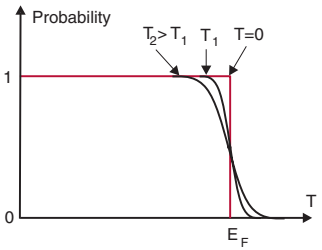
$$p_F = \hbar(3\pi^2 n)^{1/3}$$



Shown here is the location of the FS circle centered at the origin and covering the first four repeated BZs for a square lattice in 2D. The sections of the FS that belong to the same BZ can be separated from the parts that belong to other BZs, mapping the FS using the reduced zone scheme. The numbers indicate the BZs. Parts of the FS belong to the first BZ as well as to the second, third, and fourth BZ. Pieces of the FS in the first four BZs are shown mapped to the first BZ. The shaded areas represent occupied electron states. The first band is completely filled by electrons, and the portion of the FS in band 2 contains holes, whereas the FS in bands 3 and 4 contains both electrons and holes. The energies in the higher-number zones are also high.



Fermi–Dirac Distribution



The statistical behavior of electrons (i.e., fermions) at nonzero temperatures is described by the **Fermi–Dirac (FD) distribution** function. The FD distribution function is the occupation probability for an electron to occupy an energy state in **k**-space at temperature T. At $T = 0$ (shown here), all electrons occupy states with energies smaller than some energy, known as the Fermi energy E_F. States with larger energies are vacant. The FD distribution function is given by

$$f_{FD}(\varepsilon) = \frac{1}{\exp\left(\frac{\varepsilon - E_F}{k_B T}\right) + 1}$$

where k_B is the Boltzmann constant, and T is the absolute temperature. Note that for all temperatures $f_{FD}(E_F) = 0.5$, except $T = 0$ K. After summing all states of the electrons in the system,

$$\sum_n f_{FD}^{(n)}(E, T) = N$$

where N is the total number of electrons per unit volume. At 0 K, the FD distribution is a step function with all electrons below the Fermi energy. At temperature T, the probability to finding an electron at E_F is 0.5. At higher temperatures, more electrons are excited above the Fermi level. At very high temperatures, the distribution approaches the classical Maxwell–Boltzmann distribution function

$$f_{MB}(E) = \exp(-(E - \mu)/(k_B T))$$

where μ is the chemical potential; for FD statistics, $\mu = E_F$ at $T = 0$. In metals, the Fermi energy is on the order of 5 eV, which is much larger than the room-temperature energy 0.025 eV.

Density of States

In many solid state problems, one needs to evaluate the following expression

$$\frac{1}{V}\sum_{\mathbf{k}} F(E(\mathbf{k}))$$

for an arbitrary function $F(E(\mathbf{k}))$. A simple trick using the delta function gives

$$\frac{1}{V}\sum_{\boldsymbol{k}} F(E(\boldsymbol{k})) = \frac{1}{V}\sum_{\boldsymbol{k}} \int dE F(E(\boldsymbol{k}))\delta(E - E(\boldsymbol{k})) \equiv \int dE F(E(\boldsymbol{k}))\rho(E)$$

Here we have introduced the **density of states (DOS)** $\rho(E)$, defined as (V being volume)

$$\rho(E) = \frac{1}{V}\sum_{\boldsymbol{k}} \delta(E - E(\boldsymbol{k}))$$

The DOS determines the number of quantum states $dN(E)$ in a small interval, the dE around energy E, and $dN(E) = \rho(E)dE$. For example, to evaluate the DOS for free electrons in three dimensions, the energy of free electrons $\varepsilon(\mathbf{k})$ depends on the magnitude of $\mathbf{k}$ but not on its direction, and it has the form $\varepsilon(\mathbf{k}) = \hbar^2 k^2/2m$. In this case, the calculation of $\rho(E)$ can be performed in the spherical coordinates as

$$\begin{aligned}\rho_{3D}(E) &= \frac{1}{8\pi^3}\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta \int_0^{\infty} k^2 dk \delta(E - E(k)) \\ &= \frac{1}{8\pi^3} 4\pi \int_0^{\infty} k^2 dk \delta\left(E - \frac{\hbar^2 k^2}{2m}\right)\end{aligned}$$

To evaluate the integral, change the variables $\hbar^2 k^2/2m = E'$ and have

$$\begin{aligned}\rho_{3D}(E) &= \frac{1}{2\pi^2}\int_0^{\infty} dE' \frac{m}{\hbar^2}\sqrt{\frac{2m}{\hbar^2}E'}\delta(E - E') \\ &= \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}(E)^{1/2}\end{aligned}$$

which shows parabolic dependence versus energy. The energy dependence of the DOS for various low-dimensional systems shows different behavior.

Electronic Heat Capacity

The determination of a **specific heat** is one of the simplest experiments, and the example here evaluates the specific heat of a free-electron system by summing the energies of electrons with the appropriate distribution function. The specific heat at a constant volume is

$$c_V = \frac{1}{V}\left(\frac{\partial E}{\partial T}\right)_V$$

where E is the energy of an electron system, given by

$$E = 2\frac{V}{(2\pi\hbar)^3}\int d\mathbf{p}\ \varepsilon(\mathbf{p})f(\mathbf{p})$$

Here, $f(\mathbf{p})$ is the FD distribution function. A 3D isotropic electron system is assumed, and the sum is replaced by an integral as

$$\sum_{\mathbf{k}} \cdots \rightarrow \frac{V}{(2\pi)^3}\int d\mathbf{k} = \int_0^\infty d\varepsilon\rho(\varepsilon)$$

where $\rho(\varepsilon)$ is the DOS. The energy is obtained by substituting for the FD distribution function:

$$E = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2} H$$

where

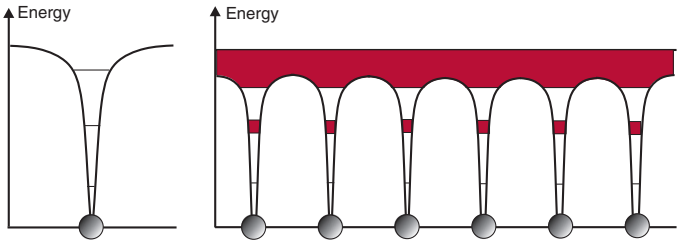
$$H = \int_0^\infty d\varepsilon \frac{\varepsilon^{3/2}}{1+\exp\left(\frac{\varepsilon-\varepsilon_F}{k_B T}\right)}$$

In the low-temperature limit $k_B T << \varepsilon_F$, function H is evaluated to be

$$H \approx \frac{2\varepsilon_{F0}^{5/2}}{5}\left[1+\frac{5}{12}\left(\frac{\pi k_B T}{\varepsilon_F}\right)^2\right]$$

It can be concluded from the previous expansion that at low temperatures the specific heat due to the electrons at a constant volume is $c_v = \text{const} \times T$.

Kronig–Penney Model



The potential of an ion can be approximated as shown here or by a rectangular version that allows for an analytical approach. The latter introduces rectangular wells with a width b surrounded by barriers with a thickness a. The Bloch theorem indicates that a solution for such potential has the form $\psi_k(x + n(a + b)) = e^{i\mathbf{k}n(a+b)}\psi_k(x)$, where $\mathbf{k}$ is a wavevector that also contains a band index. The Schrödinger equation can be used with the continuities of a wave function and its derivative to produce

$$\frac{\beta^2 - \alpha^2}{2\alpha\beta}\sinh(\beta b)\sin(\alpha a) + \cosh(\beta b)\cos(\alpha a) = \cos[\mathrm{k}(a + b)]$$

where $\alpha^2 = (2mE)/\hbar^2$, and $\beta^2 = (2m(V_0 - E))/\hbar^2$. V_0 is the depth of the well. **Kronig–Penney** made a simplifying assumption that $V_0 \to \infty$ and $b \to 0$ but keep $V_0 b$ finite. With the previous assumptions, potential wells are very deep, have very small widths, and can be replaced by periodic delta functions. The equation becomes

$$P\,\sin(\alpha a)/\alpha a + \cos(\alpha a) = \cos(ka)$$

where $P = (mV_0 b)/\hbar^2$. The equation is solved graphically, and the results are as follows:

1. For each allowed value of the function αa on the left side, there is a value of $\mathbf{k}a$; thus, a relation between α (or energy) and wavevector $\mathbf{k}$ can be established.
2. Relation E versus $\mathbf{k}$ defines the allowed energy bands and the forbidden bands.

Effective Mass

Computed energy bands of solids show a complicated dependence of wavevectors for both the conduction and valence bands. Near the bottom of the conduction band, the (approximate) energy shows spherical or ellipsoidal dependence. For the spherical case,

$$E(k) = \frac{\hbar^2 k^2}{2m^*}$$

The relation, which is an appropriate model for the Γ minimum (special point) of the conduction band, represents a parabolic dependence of energy on the wavevector. It describes a band with a spherical surface of equal energy and an isotropic **effective mass** m^*, which can be very different from a free-electron mass m_0. The modification of the electron's mass from the m_0 value is attributed to interactions with other electrons and ions. In the more complicated ellipsoidal case, the energy is

$$E(k) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y} + \frac{k_z^2}{m_z}\right)$$

which describes bands with ellipsoidal surfaces of equal energy and can represent the L and X minima (special points that have small groups) of the conduction band. Here, the effective mass depends on the direction of the wavevector, and the tensor effective mass can be introduced by differentiating the band structure at any **k**-point

$$\frac{1}{m_{ij}^*} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k_i \partial k_j}$$

Here, $i, j = x, y, z$. More realistic models can include the effect of nonparabolicity in the E versus **k** relation by introducing the nonparabolicity constant α and have $E(1 + \alpha E) = (\hbar^2 k^2)/2m^*$. Thus, the effective mass depends on the energy.

Born–Oppenheimer Approximation

Since the mass of the atomic nuclei M_a is almost 2000 times heavier than the mass of the electron, Born and Oppenheimer claimed that the electrons should be able to adjust instantaneously to the positions of the nuclei and proposed that (i) the wave function of a solid is composed of a product of the nuclei wave function and the electronic wave function, and (ii) the time-dependent part of the Schrödinger equation for a solid is associated with the nuclei wave function. The approximation separates the Schrödinger equation for the solid into three parts:

Electronic

This portion includes (i) the sum of the kinetic energy of each electron, (ii) the nuclei–electron interaction with the nuclei at their equilibrium positions at $T = 0$ K, and (iii) the electron–electron interaction to determine the total energy of the electronic system. This energy depends on the positions of the nuclei. Electronic energies can be formulated at the equilibrium positions of the nuclei.

Nuclei

This part consists of the kinetic energies of the nuclei and the nuclei–nuclei interactions. The positions of the nuclei are not restricted to their equilibrium positions because of their time variation. This many-nuclei equation can be used to determine the vibrational modes and the phonon spectra of the nuclei.

Nuclei–electron coupling

This part involves the energy of the electronic system, which is parameterized by the equilibrium positions of the nuclei and the changes of the electronic part by the motions of the nuclei, and vice versa. The parameterized part can be modified to include the displacements of the nuclei near their equilibrium positions. The electron–phonon coupling is described by taking the derivatives of the energy with respect to the positions and describing the displacements by the phonon modes. This is a way to model the electron–phonon coupling.

Hartree Approximation

The electronic part of the Schrödinger equation governs a solid system with many electrons, and it is impossible to solve without making an approximation. One approach treats each electron as an independent entity. The **Hartree approximation** suggests a variational scheme with an approximate-trial, many-electron wave function as the product of the wave function for each independent electron. In this way, the many-electron problem is reduced to a one-electron problem.

The many-electron Hamiltonian involves the sum of kinetic energies, the sum of the electrons interacting with the nuclei, and the electronic pairwise Coulomb repulsion. The periodicity of a solid allows one to consider the atoms and electrons in a unit cell. The Hamiltonian is simplified by specifying the atomic arrangement in a unit cell and summing all unit cells. The Hartree approximation minimizes the Hamiltonian with respect to each electron wave function.

The trade-off is that the one-electron Schrödinger equation should be solved self-consistently by integrating the charge density to produce the total number of electrons per unit cell.

The Hartree method is useful as a starting point for other methods, but it is incorrect as the wave function is not antisymmetrized because electrons are Fermions.

Slater Determinant

The **Slater determinant** is the simplest type of fermionic wave function: a many-electron wave function within the independent particle concept that satisfies the Pauli principle. The wave function is antisymmetric in the exchange of any two particles. For an N-electron system, the Slater determinant is

$$\psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{r}_1) & \chi_2(\mathbf{r}_1) & \cdots & \chi_N(\mathbf{r}_1) \\ \chi_1(\mathbf{r}_2) & \chi_2(\mathbf{r}_2) & & \chi_N(\mathbf{r}_2) \\ \vdots & & & \vdots \\ \chi_1(\mathbf{r}_N) & \chi_2(\mathbf{r}_N) & \cdots & \chi_N(\mathbf{r}_N) \end{vmatrix}$$

where $\chi_i(\mathbf{r}_j)$ is a wave function of an individual i^{th} electron at position $\mathbf{r}_j$. When two rows or two columns of the determinant wave function are the same, the wave function is zero; therefore, the Slater determinant wave function obeys the Pauli principle.

The Slater determinant is used in the Hartree–Fock theory as an approximation of the electronic wave function. Minimizing the expectation value of the energy using the Slater determinant with respect to a wave function leads to the Hartree–Fock equation

$$-\frac{\hbar^2}{2m}\nabla^2\chi_i(\mathbf{r}) + U^{ion}(\mathbf{r})\chi_i(\mathbf{r}) + U^{el}(\mathbf{r})\chi_i(\mathbf{r}) - \sum_j \int d\mathbf{r}' \frac{e^2}{|\mathbf{r}-\mathbf{r}'|}\chi_j^*(\mathbf{r}')\chi_i(\mathbf{r}')\chi_j(\mathbf{r}) = \varepsilon_i\chi_i(\mathbf{r})$$

where $U^{ion}(\mathbf{r}) = -Ze^2\sum_{\mathbf{R}}\frac{1}{|\mathbf{r}-\mathbf{R}|}$ is the potential of the ions, and $U^{el}(\mathbf{r}) = -e\int d\mathbf{r}'\rho(\mathbf{r}')\frac{1}{|\mathbf{r}-\mathbf{r}'|}$ is the potential energy acting on the given electron with $\rho(\mathbf{r}) = -e\sum_i|\chi_i(\mathbf{r})|^2$ as the total electronic charge density. The third term is the exchange energy.

Hartree–Fock Approximation

The cohesive energies calculated by the Hartree approximation are always higher than the measured results. The physical origin of the difference is due to the approximation, which does not take into the account the fermionic nature of the electrons. They should obey the Pauli principle—no two electrons can occupy the same state. Hartree and Fock independently used the Slater determinant as a trial wave function to minimize the many-electron Hamiltonian. The use of the antisymmetrized wave function to invoke the Pauli principle is called the **Hartree–Fock approximation**. The resulting one-electron Hamiltonian within the Hartree–Fock approximation (the Hartree–Fock equation) has an extra term due to the antisymmetrization of the trial wave function, i.e., the Slater determinant. The extra term is given by

$$-\sum_{s_j \| s} \int d\mathbf{r}' \frac{\chi_j^*(\mathbf{r})\chi_i(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|}\chi_j(\mathbf{r})$$

In the summation, $\mathbf{s}_j \| \mathbf{s}$ indicates the spin of the j^{th} state parallel to the spin of the electron involved. $\mathbf{s}$ is the spin of the i^{th} particle. In the integrand, the $\mathbf{r}'$ dependence is in the wave function of the j^{th} state χ_j^* as well as in the state of the interested electron, and the $\mathbf{r}$ dependence is in the wave function of the j^{th} state χ_j. In the extra term given here, $\chi_i(\mathbf{r}')$ and $\chi_{ij}(\mathbf{r})$ are interchanged as compared to the direct term. The summation and the integral are called the exchange term (see "Slater Determinant"). Physically, it means that electrons with spin orientations parallel to the spin orientation of the given electron should never make contact due to the Pauli principle.

The negative sign in front of the exchange term is required because the Hartree term counts toward the Coulomb repulsion for electrons that are never close to the electron of interest. Therefore, this contribution should be subtracted from the direct term.

Density Functional Theory

Density functional theory (DFT) lays the foundation for addressing the many-electron problem in condensed matter systems, and the concept is now the backbone of electronic structure calculations. The basic premise is that the electronic density $\rho(\mathbf{r})$ is the fundamental physical quantity of a many-electron system.

Hohenberg and Kohn proved two theorems:

1. $\rho(\mathbf{r})$ is uniquely related to the external potential. The proof of the theorem is by contradiction.
2. The ground-state electronic charge density is determined by minimizing the energy functional $E[\rho(\mathbf{r})]$ with the condition of $\int \rho(\mathbf{r}) d^3r = N$, where N is the total number of electrons per unit volume in the system of interest. The proof of this theorem is performed by applying the variational principle method to $E[\rho(\mathbf{r})]$ with respect to $\rho(\mathbf{r})$. $E[\rho(\mathbf{r})]$ consists of the many-electron kinetic energy functional, the electron–external potential functional, and the electron–electron interaction energy functional. This last functional involves the direct Coulomb interaction and the exchange correlations. The inclusion of correlations in addition to the direct Coulomb and exchange interactions are important features of DFT and make it more powerful conceptually than the Hartree–Fock approach.

However, DFT does not provide any practical calculational schemes. By incorporating DFT and making use of the Kohn–Sham equations, many existing calculational methods (such as the pseudopotential, linearized augmented plane wave, linear combination of muffin-tin orbitals, and tight-binding methods) become potent tools for many-electron problems in quantum chemistry, condensed-matter physics, and even biological systems. These methods can predict electronic, vibrational, magnetic properties, and new structures from first principles.

Kohn–Sham Equations

To apply the concept of DFT to calculate the electronic properties of condensed-matter systems, **Kohn–Sham (KS)** suggested the following scheme to reduce the many-electron problem to a one-electron problem:

1. Adding and subtracting a sum of noninteracting, electronic kinetic energy terms.
2. Separating the electron–electron interaction functional into a direct Coulomb repulsion functional and the exchange-correlation energy functional, which is difficult to treat.
3. Grouping the many-electron kinetic energy functional and the subtracted part of the sum of the noninteracting kinetic energy into the exchange-correlation energy functional.
4. Approximating the exchange-correlation energy functional to be dependent on the local charge density. This approximation is called the **local density approximation (LDA)**.

One of the KS equations is

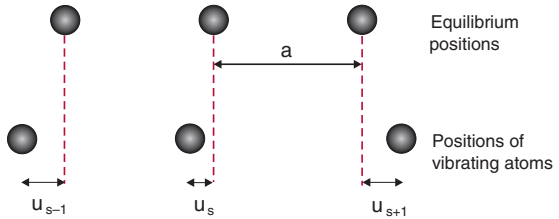
$$\left\{-\frac{1}{2}\nabla^2 + V_{KS}(\mathbf{r})\right\}\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r})$$

where the potential $V_{KS}(\mathbf{r}) = V_n + V_H + V_{XC}$. The first term involves the interaction between an electron and the collection of atomic nuclei; the second is the Hartree term that represents the Coulomb forces; and the third is the exchange-correlation energy.

This KS equation is a one-electron equation similar to the Hartree and Hartree–Fock equations but with the exchange-correlation approximated by the LDA. Due to the presence of the electronic charge density in the LDA, the equation should be solved self-consistently with the sum of the densities of the occupied states. These equations have been applied to many existing schemes to calculate electronic properties such as tight binding, linearized and augmented plane waves, linear combinations of muffin-tin orbits, and pseudopotential methods. Note that DFT is characterized by the two Hohenberg–Kohn theorems.

Atomic Linear Chain Vibrations: Monatomic Case

The positions of vibrating atoms in a 1D chain are shown here. If the displacements are perpendicular to the chain, the vibration is called a **transverse mode**. In general, there are one longitudinal mode and two transverse modes.



Displacement of an atom (ion) at position s is described by u_s, which is along the chain. Dynamic equations governing the displacements take the form of

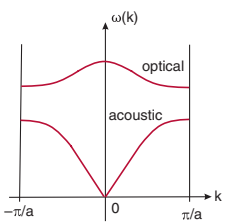
$$M\frac{du_s^2}{dt^2} = C(u_{s+1} + u_{s-1} - 2u_s)$$

where M is the mass of an atom. Only one spring constant C is used. The mode of the vibration is called a **longitudinal mode**; atoms vibrate along the direction of the sound wave propagating along the chain. Harmonic time variation and the following spatial dependence are assumed:

$$u_{s\pm1} = ue^{i\omega t}e^{iq(s\pm1)a}$$

After substituting the assumed solution into the equation of motion by using trigonometry with q in the first BZ, the resulting dispersion relation is

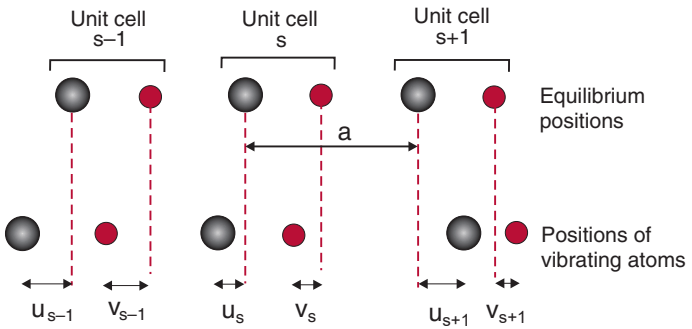
$$\omega = \sqrt{\frac{4C}{M}}\left|\sin\frac{1}{2}qa\right|$$



The plot of the dispersion shown here is known as an **acoustic branch**. The branch with higher energy, known as the optical branch, will be discussed next. It exists when there are more than one atom per unit cell.

Vibrations with Two Atoms per Unit Cell

The monatomic case can be extended to describe vibrations with two atoms per unit cell.



Vibrations in such a system are described by two coupled equations (one for each subsystem):

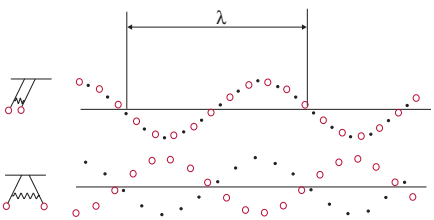
$$M_1 \frac{du_s^2}{dt^2} = C(v_s + v_{s-1} - 2u_s), \quad M_2 \frac{dv_s^2}{dt^2} = C(u_{s+1} + u_{s-1} - 2v_s)$$

These equations can be solved similar to the monatomic case to obtain the following dispersion relation:

$$\omega^2 = C\left(\frac{1}{M_1} + \frac{1}{M_2}\right) \pm C\sqrt{\left[\left(\frac{1}{M_1} + \frac{1}{M_2}\right)^2 - \frac{4\sin^2 ka}{M_1 M_2}\right]}$$

When two atoms in the unit cell vibrate in phase, the solution is known as an **acoustic branch**. It is also possible for two atoms in a unit cell to displace relative to each other (the vibration is called an optical branch).

A distinction between the optical and acoustic waves in a linear diatomic lattice is illustrated here analogous to two coupled mechanical pendulums.



Phonons

If the vibrations are coherent, the diplacements of neighboring atoms have a phase coherence, and such vibrational waves are **normal modes**. The quantized normal modes are called **phonons**, which are bosons. Each phonon posseses energy $E = \hbar\omega$ and momentum $\mathbf{p} = \hbar\mathbf{k}$, where $\mathbf{k}$ is a quasi-wavevector and is restricted to the first BZ, and ω is the frequency of the corresponding normal vibration, which is a periodic function of $\mathbf{k}$.

In the harmonic approximation, the energy of phonons in a crystal is expressed as a sum of the energies of independent oscillators

$$E = E_0 + \sum_{k\alpha} \hbar\omega_\alpha(\boldsymbol{k}) N_\alpha(\boldsymbol{k})$$

where E_0 is the energy of the zero-point vibrations, subscript α denotes the branches of phonons, and $\omega_\alpha(\mathbf{k})$ determines dispersion. In thermodynamic equilibrium, the average number of phonons is determined by the Bose–Einstein distribution

$$N_\alpha(\mathbf{k}) = \left[\exp\left(\frac{\hbar\omega_\alpha(\mathbf{k})}{k_B T}\right) - 1\right]^{-1}$$

An accounting of the anharmonic effects in crystals describes phonon–phonon interactions that are weak but play a fundamental role in the relaxation processes that determine the spatial and temporal attenuation of phonons. The wavelength of a normal mode can be related to the momentum of a phonon: k (in one dimension) or $|\mathbf{k}|$ (in three dimensions). For the 3D case at low frequencies, the three acoustic branches have linear dispersions (energy versus momentum relations) with slopes as the sound speeds. To probe the dispersion relations, neutron scattering is used to determine the dispersion. Infrared light can also be used to probe $|\mathbf{k}| = 0$ phonons due to the fact that its wavelength is at least 10^3 times longer then the length of a unit cell.

Debye Model and Approximation

The measurement of the specific heat of a crystal at a low temperature shows T^3 behavior due to phonons. This can be determined using the Bose–Einstein distribution function and the linear relationship between the frequencies of phonons and their momenta: the well-known **Debye model**. Specifically, the dispersion relation of the acoustic branches for small values of q is $\omega = \upsilon \cdot q$, where υ is the sound speed, and q is the phonon momentum.

The development of the Debye model of the specific heat of crystals C_V starts with the definition

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V$$

Here, the total energy U of phonons at temperature T is the sum of the energies of all phonon modes

$$U = \sum_{\boldsymbol{q},p} < n_{\boldsymbol{q},p} > \hbar\omega_{\boldsymbol{q},p}$$

where p is the polarization, and $< n_{\mathbf{q},p} >$ is the thermal equilibrium occupancy given by the Planckian distribution function. To evaluate the sum, the sum is replaced by an integral as

$$\sum_{\mathbf{q},p} \ldots \rightarrow \sum_{p} \int d\omega_{\mathbf{q},p} D_p(\omega_{\mathbf{q},p}) \ldots$$

where $d\omega_{\mathbf{q},p} D_p(\omega_{\mathbf{q},p})$ is the number of modes with polarization p in the frequency range $(\omega,\ \omega + d\omega)$. The upper limit of the frequency integral is approximated by θ_D, the Debye temperature, which is the essence of the **Debye approximation**. After substituting the expression for a Planckian distribution, the specific heat becomes

$$C_V = \frac{T^3}{2\pi^2\hbar^3 k_B^{-4}} \int_0^{\theta_D/T} \frac{x^4 e^x}{(e^x - 1)^2}\, dx$$

In the low-temperature limit, $\theta_D/T \rightarrow \infty$, and the expression can be evaluated with the result $C_V \sim T^3$.

Anharmonic Effects and Thermal Expansion

The harmonic approximation is a second-order derivative of the potential with respect to atomic displacements. It describes non-interacting phonons but not many experimental facts, such as the volume expansion of a solid when it is heated.

Hooke's law is an acceptable approximation when considering small oscillations about the equilibrium position, but if the vibrations go beyond a very small amplitude, higher-order terms must be included. To account for experimental facts, one needs to take higher-order expansion of the interatomic potential with respect to the atomic displacements, known as **anharmonic effects**, which occur in real crystals and account for the nonsymmetric potential.

The commonly considered anharmonic effects are the third- and fourth-order anharmonicities

$$V^{(3)} = (1/3!) \sum_{l,j,k} \partial^3 V / (\partial x_l \partial x_j \partial x_k)$$

$$V^{(4)} = (1/4!) \sum_{l,j,k,m} \partial^4 V / (\partial x_l \partial x_j \partial x_k \partial x_m)$$

where x_j is the atomic displacement component. Anharmonicity accounts for both elastic and inelastic collisions of phonons leading to their scattering and decay. The above anharmonic terms are responsible for the thermal conductivity due to the phonon–phonon scattering processes, volume expansions, change in the dielectric function of the crystal with temperature, etc. In the harmonic approximation, the average position of the atoms does not change no matter the size of the amplitude of thermal vibrations. When considering anharmonic terms, the average position of an atom depends on the amplitude and therefore the temperature.

The theory of anharmonicity is a typical many-body problem in solid state systems, related to fundamental questions such as the ability to form two-phonon states and create solitons.

Thermal Conduction Due to Phonons

Thermal conduction in a solid can be viewed as the change of local density of phonons $n_{\mathbf{q}}(\mathbf{r},t)$ (which depends on the local temperature), where $\mathbf{q}$ is the momentum of a phonon. It also includes the polarization labeling of the longitudinal or transverse mode. Then, by applying the Boltzmann equation,

$$\frac{\partial n_{\mathbf{q}}(\mathbf{r},t)}{\partial t} = \frac{\partial n_{\mathbf{q}}(\mathbf{r},t)}{\partial \boldsymbol{r}} \cdot \frac{\partial \mathbf{r}}{\partial t} = \frac{\partial n_{\mathbf{q}}(\mathbf{r},t)}{\partial \mathbf{r}} \cdot \mathbf{v}_{\mathbf{q}}$$

where $\mathbf{v}_{\mathbf{q}}$ is the velocity $[\partial \mathbf{r}/\partial t]$ of the phonon. Since the thermal conductance is measured when a sample is placed in contact with reservoirs at both ends with different temperatures (high and low), $\partial n_{\mathbf{q}}(\mathbf{r},t)/\partial \mathbf{r}$ should be expressed as a temperature gradient of $n_{\mathbf{q}}^{(0)}$, i.e., the local-equilibrium Bose–Einstein distribution function:

$$\frac{\partial n_{\mathbf{q}}(\mathbf{r},t)}{\partial \mathbf{r}} = -\frac{\partial n_{\boldsymbol{q}}^{(0)}(\mathbf{r},t)}{\partial T} \cdot \frac{\partial T}{\partial \mathbf{r}} = -\nabla T \frac{\partial n_{\mathbf{q}}^{(0)}(\mathbf{r},t)}{\partial T}$$

Here, the minus sign is due of the change from high to low temperature. ∇T is the change of temperature in the direction from high- to low-temperature ends. Using the relaxation-time τ approximation, the change of $n_{\mathbf{q}}$ during τ is

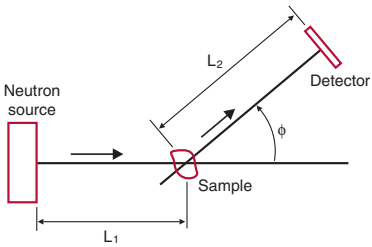
$$\Delta n_{\mathbf{q}}(\mathbf{r},t) = \frac{\partial n_{\mathbf{q}}^{(0)}(\mathbf{r},t)}{\partial t} \cdot \tau$$

and the thermal current is

$$\mathbf{I} = \sum_{\mathbf{q}} \Delta n_{\mathbf{q}} \hbar \omega \mathbf{v}_{\mathbf{q}} = -\kappa^{T} \nabla T$$

The term in the middle is the change of the local distribution function multiplied by the associated energy and the velocity of the contributing phonons. In the last term, κ^{T} is the conductivity tensor, which explains why thermal current may not have the same direction as the temperature gradient.

Experimental Determination of Phonon Spectra



Detailed **phonon spectra** can only be determined by the neutron-scattering method. Neutrons are produced by nuclear fission, and their energies are ~2 MeV. They then pass through a thick material (not shown here) to reach energies ~0.1 eV, suitable for diffraction experiments. Neutrons hit the crystal (the "sample" in the figure), and the scattering occurs at a Bragg's angle. The time-of-flight method is used to measure the energy of the diffracted neutrons and their momentum by placing the detector at a certain angle ϕ with respect to the forward direction. Interpretation of the experimental results uses the energy-conservation law and the momentum-conservation law:

$$\mathbf{k}' = \mathbf{k} + \mathbf{K}$$

where $\mathbf{k}'$ and $\mathbf{k}$ are the wavevectors of the scattered and incident neutrons, respectively, and $\mathbf{K}$ is the reciprocal lattice vector. If the scattering is inelastic and involves either the absorption or emission of a phonon, the momentum-conservation law becomes $\mathbf{k}' = \mathbf{k} \pm \mathbf{q} + \mathbf{K}$, where $\pm\,\mathbf{q}$ corresponds to the creation or absorption of a phonon with momentum $\mathbf{q}$:

$$E = E' \pm \hbar\omega_{\mathbf{q}}, \quad \mathbf{p}' = \mathbf{p} \pm \hbar\mathbf{q} \pm \hbar\mathbf{K}$$

where $\hbar\omega_{\mathbf{q}}$ is the energy of the phonon, $\hbar\mathbf{q}$ is its momentum and E' and E are the energies of the scattered and incident neutrons, respectively. The first equation applies to energy conservation, whereas the second applies to momentum.

With knowledge of the crystal structure and its orientation with respect to the incident neutrons, the wavevectors $\mathbf{k}$ and $\mathbf{k}'$ of neutrons corresponding to the intensity peak can be evaluated, and hence $\mathbf{q}$ can be determined.

Drift–Diffusion Model

The **drift–diffusion model** of electrical transport plays a major role in describing the current flow in semiconductors. It takes into account contributions from electrons and holes. This page concentrates on the former.

Under an external voltage applied to the semiconductor material, electrons in the conduction bands tend to move toward the positive electrode. The average velocity of electrons $\mathbf{v}_d$ is known as the **drift velocity**, and it contributes to the current as $J_n = qn\mathbf{v}_d = qn\mu_n E$, where μ_n is the mobility of electrons, and E is the electric field.

Diffusion is another mechanism that contributes to the flow of the current, even in the absence of external fields. It exists when the carrier concentration is not constant throughout the semiconductor, i.e., when there is a gradient of carrier concentration. Diffusion contributes to the current as $J_n = qD_n\nabla n$, where D_n is the diffusion coefficient. A full description of the transport in semiconductors also requires continuity equations for electrons and holes:

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla \cdot J_n = \frac{\partial p}{\partial t} + \frac{1}{q}\nabla \cdot J_p = -R(\psi, n, p)$$

where ψ is the electrostatic potential, defined as $E = -\nabla\psi$, and R is the net generation–recombination rate of electron–hole pairs per unit volume. The contributions to the electron and hole currents from drift and diffusion are

$$J_n = qn\mu_n E + qD_n\nabla_n, \quad J_p = qp\mu_p E - qD_p\nabla_p$$

The above system of equations is supplemented by the Poisson equation $-\nabla \cdot (\varepsilon\nabla\psi) = q(p - n - C)$, where C is the net impurity concentration, which is essentially independent of time. The diffusion coefficients and mobilities are related by the Einstein relations for electrons and holes:

$$\frac{D_n}{\mu_n} = \frac{k_B T}{q}, \quad \frac{D_p}{\mu_p} = \frac{k_B T}{q}$$

Boltzmann Equation

A powerful formalism for treating various transport properties in crystals is based on the **Boltzmann equation**. In this formalism, the central role is played by distribution function $f = f(t,\mathbf{r},\mathbf{p})$ describing an electron (more precisely, a quasiparticle) at a position $\mathbf{r}$ with momentum $\mathbf{p}$ and operating in 6D phase space $(\mathbf{r},\mathbf{p})$. The total variation of f in time due to collisions is $(df)/(dt) = I[f]$. The right side has the so-called collision integral describing the effect of collisions. Its detailed form is determined independently. Since function f depends on t, $\mathbf{r}$, and $\mathbf{p}$, its time variation is expressed as

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial \mathbf{r}}\frac{d\mathbf{r}}{dt} + \frac{\partial f}{\partial \mathbf{p}}\frac{d\mathbf{p}}{dt}$$

From the definition, $\mathbf{v} = (d\mathbf{r})/(dt)$, where $\mathbf{v}$ is the velocity of an electron, and $\mathbf{F} = (d\mathbf{p})/(dt)$ is the total force acting on an electron. With the above identifications, the practical form of the Boltzmann equation becomes

$$\frac{\partial f}{\partial t} + \mathbf{v}\frac{\partial f}{\partial \mathbf{r}} + \mathbf{F}\frac{\partial f}{\partial \mathbf{p}} = I[f]$$

The above approximation is valid if one can simultaneously determine the location and momentum of an electron. However, the Heisenberg principle prevents this, and thus a wave packet must be created to describe an electron. Therefore, the description based on the Boltzmann equation is valid in the following limits:

$$p >> \Delta p \sim \frac{\hbar}{\Delta r} >> \frac{\hbar}{l} \text{ and } T >> \frac{\hbar}{\tau_\xi}$$

where l is the mean free path of an electron, T is the temperature, and τ_ξ is the energy relaxation time.

Collision Integral

An explicit form of the **collision integral** is necessary to solve the Boltzmann kinetic equation, which involves the scattering of electrons. Elastic scattering is described by a collision integral that incorporates the Pauli principle as

$$I[f] = \sum_{\mathbf{p}'} \begin{bmatrix} S(\mathbf{p}', \mathbf{p}) f(\mathbf{p}')(1 - f(\mathbf{p})) - \\ S(\mathbf{p}, \mathbf{p}') f(\mathbf{p})(1 - f(\mathbf{p}')) \end{bmatrix} \delta(\varepsilon(\mathbf{p}) - \varepsilon(\mathbf{p}'))$$

Further progress is obtained by assuming that kernels are symmetric, i.e., $S(\mathbf{p}',\mathbf{p}) = S(\mathbf{p},\mathbf{p}')$. The relation

$$f(\mathbf{p}')(1 - f(\mathbf{p})) - f(\mathbf{p})(1 - f(\mathbf{p}')) = f(\mathbf{p}') - f(\mathbf{p})$$

is also involved. Next, the summation is replaced by an integral, and the external perturbation is assumed to be small, i.e., $f(\mathbf{p}) = f_0(\mathbf{p}) + f_1(\mathbf{p})$, with $|f_1| << |f_0|$. The collision integral becomes ($d\Omega$ is the solid angle)

$$I[f] = \int \frac{d\Omega'}{4\pi} W(\alpha)[f_1(\mathbf{p}') - f_1(\mathbf{p})]$$

Here, $W(\alpha)$ is proportional to the kernel and contains all other factors resulting from the change of variables, and α is the angle between $\mathbf{p}$ and $\mathbf{p}'$. For an external perturbation due to the electric field, $f_1 = g\mathbf{p} \cdot \mathbf{E}$, where $g(\varepsilon)$ is some function of the electron's energy. In this case, evaluation of the integral gives

$$I[f] = -\frac{f_1}{\tau}$$

where $\tau^{-1} = \int W(\alpha)(1 - \cos\theta)(d\Omega/4\pi)$ is the average time between collisions (also called the relaxation time). A combination of the above results produces the Boltzmann kinetic equation in the relaxation-time approximation:

$$\frac{\partial f}{\partial t} + \mathbf{v}\frac{\partial f}{\partial \mathbf{r}} + \mathbf{F}\frac{\partial f}{\partial \mathbf{p}} = -\frac{f - f_0}{\tau}$$

Electrical Conductivity

The elementary theory of the **electrical conductivity** of metals was initiated by Drude and Lorentz based on classical physics. It was Sommerfeld who first analyzed electrical conductivity by applying the Boltzmann transport equation and Fermi–Dirac statistics. The Boltzmann equation is used here within the relaxation-time approximation to evaluate the electrical conductivity of metals by assuming a steady state and uniformity in space:

$$q\mathbf{E}\frac{\partial f}{\partial \mathbf{p}} = -\frac{f_1}{\tau}$$

where q is the electronic charge. After linearizing $f, f = f_0 + f_1$, where f_1 is proportional to the electric field $\mathbf{E}$, and $|f_1| << |f_0|$. The result is

$$q\mathbf{E}\frac{\partial f_0}{\partial \mathbf{p}} = -\frac{f_1}{\tau}$$

and the solution is $f_1 = -\tau q\mathbf{E} \cdot \mathbf{v}(\partial f_0/\partial \varepsilon)$, where

$$\frac{\partial f_0}{\partial \mathbf{p}} = \frac{\partial f_0}{\partial \varepsilon}\frac{\partial \varepsilon}{\partial \mathbf{p}} = \mathbf{v}\frac{\partial f_0}{\partial \varepsilon} = -\mathbf{v}\delta(\varepsilon - \varepsilon_F)$$

The electrical current is defined as

$$\mathbf{j} = 2q\int \frac{d\mathbf{p}}{(2\pi\hbar)^3}\mathbf{v}f(\mathbf{p})$$

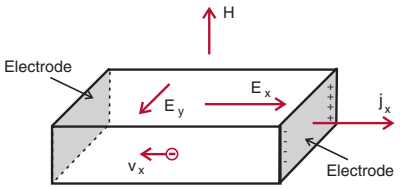
Substituting the above solution with the definition of electrical conductivity, i.e., $\mathbf{j} = \sigma\mathbf{E}$, and evaluating the integral produces

$$\sigma = ne^2\tau_F/m$$

where τ_F is the relaxation time at the Fermi energy.

This approach is known as **Sommerfeld theory**, and it provides essentially the same result for electrical conductivity as that obtained by Lorentz.

Hall Effect



When a metal or semiconductor is placed in a magnetic field **H** (along the z axis) and a current density **j** is passed through it along the x direction, a transverse electric field is created that generates voltage along the y direction in a process known as the **classical Hall effect**. The following equation describes the flow of the current:

$$\frac{d\mathbf{p}}{dt} = -q\left(\mathbf{E} + \frac{\mathbf{p}}{m} \times \mathbf{H}\right) - \frac{\mathbf{p}}{\tau}$$

For a steady state, i.e., $(d\mathbf{p})/(dt) = 0$, in the experiment, $j_y = 0$. With this simplification, the result is

$$E_y = -\frac{\omega_c \tau}{\sigma_0} j_x = -\frac{H}{ne} j_x$$

where $\sigma_0 = ne^2\tau/m$ is the DC conductivity, and $\omega_c = eH/m$ is the cyclotron frequency. The Hall coefficient is introduced as

$$R_H = \frac{E_y}{j_x H} = -\frac{1}{ne}$$

R_H describes the nature of charges (positive "holes" or negative "electrons"), which is important information when working with semiconductors. It also gives the concentration of carriers.

Besides the above fundamental aspects, the Hall effect finds many applications in technology, such as Hall probes (to measure magnetic fields or generate electricity) or Hall sensors (to determine the speed of rotations).

Landau Levels

The quantum-mechanical problem of electron motion in a uniform magnetic field was analyzed by Landau in 1930. The Schrödinger equation for a free electron in a magnetic field is

$$\frac{\hbar^2}{2m}(\mathbf{p} - e\mathbf{A})^2\psi = E\psi$$

The magnetic field **B** can be related to the vector potential **A** with a Landau gauge by setting **A** = [0, *Bx*, 0]. The wave function is assumed to have the form

$$\psi = e^{ik_y y}e^{ik_z z}X(x)$$

Thus,

$$-\frac{\hbar^2}{2m}\frac{d^2X(x)}{dx^2} + \frac{1}{2m}(eBx - \hbar k_y)^2X(x) = \left(E - \frac{\hbar^2 k_z^2}{2m}\right)X(x)$$

This is the Schrödinger equation for a harmonic oscillator. Its solution gives the energy spectrum of an electron to be

$$E_n = \hbar\omega_c\left(n + \frac{1}{2}\right) + \frac{\hbar^2 k_z^2}{2m}$$

where $\omega_c = |e|B/m$ is known as the cyclotron frequency, and $n = 0,1,2,\ldots$ (this approach neglects electron spin). The above relation demonstrates that the electron motion transverse to the magnetic field is quantized, whereas along the magnetic field it behaves as a free particle. The quantized values of energy are known as **Landau levels**. To see such quantization in the experiment, the electron must make several revolutions during cyclotron motion between two consecutive collisions. After comparing the energy of the n^{th} Landau level with the energy of a free electron, the result is

$$\frac{\hbar^2}{2m}\left(k_x^2 + k_y^2\right) = \hbar\omega_c\left(n + \frac{1}{2}\right)$$

which defines a series of cylinders parallel to the magnetic field. The n^{th} cylinder has an area of

$$A_n = \pi\left(k_x^2 + k_y^2\right) = \left(n + \frac{1}{2}\right)\frac{2\pi eB}{\hbar}$$

An electron can jump from a cylinder n to $n+1$ by absorbing a photon.

Cyclotron Resonance

If a static magnetic field **B** is applied to a metallic (or doped semiconductor) sample, the conduction electrons move in a circular path perpendicular to the direction of the magnetic field with the cyclotron frequency. The frequency is determined by balancing the Lorentz force $e\mathbf{v} \cdot \mathbf{B}$ and the centrifugal force $m^* v^2/r$, given by $\omega_c = eB/m^*$, which is the cyclotron frequency, and m^* is the effective mass.

Additionally, a perpendicular electric field is applied at microwave frequencies; resonant absorption occurs when its frequency equals the cyclotron frequency.

A main feature of **cyclotron resonance** is the possibility of determining parameters directly through the measurement of the absorption frequency.

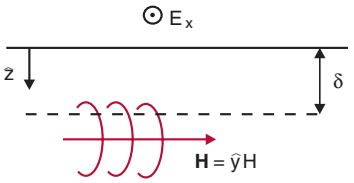
Electrons move in fixed orbits with an energy that equals $\hbar\omega_c n$, where n is a positive integer that labels the Landau levels. The ground-state absorption is associated with the change of electronic energy in $\hbar\omega_c$. It results in a direct determination of effective mass m^* through a measurement of the absorption frequency.

Cyclotron resonance occurs at microwave frequencies. The main condition for observing it is that the electrons must complete an orbit before collisions, which results in the condition $\omega\tau > 1$, where τ is the scattering time. This condition can only be met at very low temperatures.

Cyclotron-resonance experiments are used to determine the extrema of Fermi surfaces.

Related to cyclotron resonance are cyclotron waves that propagate in metals. They propagate perpendicular to the static magnetic field with a frequency in the vicinity of the cyclotron resonance and its harmonics. Cyclotron waves propagate with velocities comparable to the velocities of electrons. They do not interact with phonons.

Azbel–Kaner Cyclotron Resonance



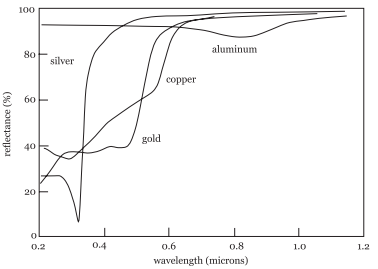
When a semiconductor sample is placed in a large, static magnetic field and a rf electromagnetic field, sharp peaks in the energy absorption are observed when the frequency coincides with the **cyclotron resonance** frequency ω_c. In metals, compared to semiconductors, the observation of cyclotron resonance is very difficult because the required high-frequency electromagnetic field can only penetrate into metals to a very small depth, known as the **skin depth**. To avoid this problem, Azbel and Kaner suggested the configuration shown here. A strong magnetic field is applied parallel to the metal surface, and the electric field is in the perpendicular direction. Electrons undergo cyclotronic motion. Each time electrons enter skin depth δ, they are accelerated by the rf field. If the radio frequency coincides with the cyclotron frequency, $\omega = n\hbar\omega_c$ (n is an integer.). Skin depth δ is much smaller than the radius of electron's orbit, and therefore electrons see the microwave field for a much shorter time compared to the orbit period.

Because the Fermi energy of a metal is very large compared to the microwave energy $\hbar\omega$, only the electrons that are close to the Fermi surface can participate in the resonance. Therefore, the method is used to measure cyclotron frequencies in metals and to study the properties of the Fermi surfaces.

In order to avoid electron–phonon scattering and to have a large scattering time $\tau \sim 10^{-9}$ s^{-1}, Azbel–Kaner cyclotron-resonance experiments are performed at low temperatures with very pure samples.

Optical Properties of Metals

The **optical properties of metals** are mostly determined by electrons. Metals show high reflectivity of light, which is manifested by their shiny surface. They reflect light in the infrared and visible regions, and up to a certain frequency in the ultraviolet (known as the **plasma frequency**, referring to the density fluctuations of all conduction electrons. A quantum of plasma oscillation, known as a plasmon, has an energy $\hbar\omega_p$). All four metals plotted here show nearly 100% reflectance at low frequencies.



The optical response of metals mainly originates from the conduction electrons. The earliest realistic model for a metal was proposed by Drude. The Drude model of free electrons in metals assumes that they behave like a classical gas. There is no interaction between the electrons except for scattering. The average scattering interval time is τ, and the motion of electrons in an electric field is thus described by

$$m\frac{d^2x}{dt^2}+\frac{m}{\tau}\frac{dx}{dt}=-eE$$

where x is the electronic displacement. The above equation leads to the permittivity of a electron gas system:

$$\varepsilon(\omega)=1-\frac{\omega_p^2\tau^2}{1+\omega^2\tau^2}+i\frac{\omega_p^2\tau}{\omega(1+\omega^2\tau^2)}$$

where $\omega_p=\sqrt{ne^2/\varepsilon_0 m}$ is the plasma frequency.

de Haas–van Alphen Effect

Several methods are used to experimentally determine a Fermi surface, among them cyclotron resonance and the **de Haas–van Alphen (dHvA) effect**, which manifests as sharp oscillations of magnetization (or magnetic susceptibility) as a function of the inverse of the magnetic field. The frequency of these oscillations provides direct information about the cross-sectional area of the Fermi surface.

The dHvA effect originates from quantizing effects that the magnetic field has on the electron motion if the orbit is closed. A description of the effect starts from the Bohr–Sommerfeld quantization rule

$$\int p \cdot dq = \left(n + \frac{1}{2}\right) h$$

where p and q are the momentum and position, respectively, n is an integer, and h is Planck's constant. Stokes' theorem is then used to derive the relation for the area S of the electron's orbit in $\mathbf{k}$-space as

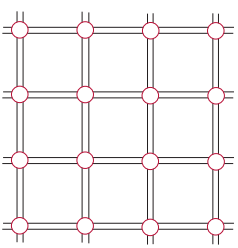
$$S = \frac{2\pi e B}{\hbar}\left(n + \frac{1}{2}\right)$$

As the magnetic field increases, the radius of each cylinder associated with the Landau level increases and eventually becomes larger than that of the cross-section of the Fermi surface. Electrons at the highest level are then dispersed to the Landau levels with lower energies as more (Landau) states become available. The free energy decreases, as does the DOS at the Fermi energy. The behavior is a periodic function of the magnetic field. These oscillations of the free energy and DOS act as a function of the magnetic field. These so-called "quantum oscillations" include more than the oscillations in the magnetization (the dHvA effect); similar oscillations of the magnetoresistance are known as the Shubnikov–de Haas effect. The frequency of the dHvA oscillations is $f = \hbar S_F / (2\pi e)$, from which the cross-section of the Fermi surface S_F is determined for various orientations of the magnetic field.

Bonding Model of Semiconductors

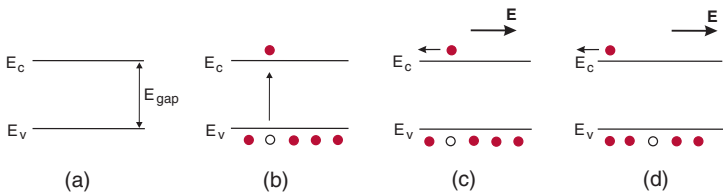
The basic property that distinguishes a **semiconductor** from a metal is the appearance of an energy gap. At zero temperature, it is defined as the difference between occupied bands (the valence bands) and unoccupied bands (the conduction bands). The value of a typical gap is 1.0 eV. Another physical property that distinguishes a semiconductor from a metal is the value of resistivity. In a semiconductor, a typical value is 100 Ω-cm, as compared to the value of copper 10^{-6} Ω-cm.

The most popular semiconductors are silicon (Si) and germanium (Ge). They belong to the group-IV elements in the periodic table and crystallize in a diamond structure. Another popular semiconductor compound, gallium-arsenate (GaAs), is a III-V semiconductor. The latter has a zincblende structure, which is a variation of the diamond structure with two identical atoms that are replaced by atoms with different atomic species. The tetrahedron around the center atom, formed by its four neighbors (shown on page 7), is the unique feature of the diamond structure. Every Si atom has four identical neighbors that interact via covalent bonding. It consists of four sp^3 hybrid orbitals, which have electron distributions directed toward the four corners of a tetrahedron. The 3D tetrahedral structure is often represented in a simplified form on a flat surface, as shown here; each circle represents the core of a semiconductor atom formed with electrons in inner orbits, and each line represents a shared valence electrons. This picture shows how easily a missing atom, point defect, or broken atom–atom bond could free an electron.



Intrinsic Semiconductors

A semiconductor with a negligibly small number of impurities is called an **intrinsic (pure) semiconductor**. The conductivities of pure crystalline Si or GaAs are very small even at room temperature and are impractical for device applications.



In intrinsic semiconductors, carriers are created (mostly) by thermal energy or by the breaking of covalent bonds. Thermal excitation creates one electron in the conduction band and one hole in the valence band, as shown in (b). Therefore, the concentration of electrons and holes are the same and equal the intrinsic concentration n_i, $n = p = n_i$:

$$n_i = \frac{(2\pi k_B T)^{3/2}}{4\pi^3 \hbar^3} \left(m_e^* m_h^*\right)^{3/4} \exp\left(\frac{E_g}{2k_B T}\right)$$

Here, m_e^* and m_h^* are the effective mass of electrons and holes, respectively. In Si, $n_i = 1.5 \times 10^{10}\text{cm}^{-3}$, and in Ge, $n_i = 2 \times 10^{13}\text{cm}^{-3}$ at room temperature.

The movements of an electron and hole under external electric field are illustrated in (c) and (d).

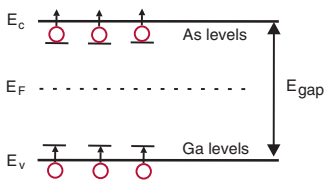
The Fermi energy E_F of an intrinsic semiconductor is located approximately at the middle of the bandgap E_g:

$$E_F = -\frac{E_g}{2} + \frac{3}{4} k_B T \ln\left(\frac{m_h^*}{m_e^*}\right)$$

Extrinsic Semiconductors

The addition of impurities provides semiconductors with specified properties important for device applications. Doping is the most effective process when host atoms are replaced by other atoms in a controlled way. The resulting semiconductors are called **extrinsic semiconductors**.

There are two types of doping: n-type and p-type, characterized by doping concentrations N_D and N_A, known respectively as the concentrations of donors and acceptors. In extrinsic semiconductors, the values of N_D and N_A exceed n_i, known as intrinsic concentration. The main property of an extrinsic semiconductor is the distinct difference between the concentrations of electrons and holes. If the dominant carriers are electrons, the material is called an **n-type semiconductor**; if holes dominate, the material is called a **p-type semiconductor**. In the former, the host atoms are replaced by atoms with a larger valence. Since electrons are associated with the doped atoms, their energy levels fall within the gap of the host semiconductors and are located just below the conduction band edge. For heavy doping, the repulsive Coulomb interaction forces the doping levels to merge into the conduction band.

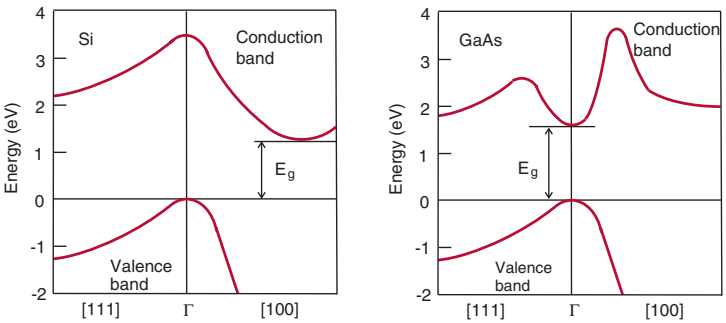


In p-type materials, the doped atoms have a valence smaller than that of the host atoms. Their energy levels are located just above the valence band edge. Dopants from group V, i.e. P, As, and Sb (known as donors), donate electrons, whereas those from group III, i.e., B, Al, Ga, In, are known as acceptors. If the semiconductor contains only donors and the temperature is high enough to ionize electrons associated to impurities, the Fermi level is

$$E_F = k_B T \ln\left(\frac{N_D}{N_c}\right), \text{ where } N_c = \frac{(2\pi m_e^* k_B T)^{3/2}}{4\pi^3 \hbar^3}$$

Band Structure of Semiconductors

The **band structure** of the group-IV semiconductors, such as Si, shows that the valence band maximum (VBM) does not happen at the same **k**-point in the BZ as the conduction band minimum (CBM), i.e., **indirect-gap semiconductors**. In most of the materials from the III–V group, such as GaAs, and the II–VI semiconductor materials, the VBM and CBM happen at the same **k**-point, i.e., **direct-gap semiconductors**.



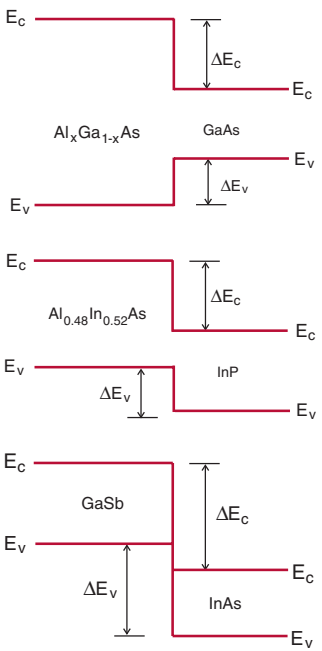
Special **k**-points in the band structure are labeled with different letters, such as Γ, L, and X. The Γ-point appears at the center of the first BZ, and the others occur at the BZ boundaries. Each one has its own symmetry: the Γ-point has the symmetry of the crystal, and the other **k**-points have symmetries of their own small groups. Compatibility relations connect the bands at special **k**-points and at **k**-points along the lines that connect the special **k**-points.

When an electron gains energy, e.g., with temperature, and moves from the valence to conduction band, it leaves an empty "space" in the valence band known as a hole. The concept of holes in semiconductors comes from Dirac's theory of a relativistic electron. In semiconductors, holes are considered as real particles, like electrons.

Semiconductor Heterostructures

A **heterostructure** consists of two or more semiconductor materials that have different energy gaps. Since the values of the energy gaps differ, the conduction and valence band edges cannot be simultaneously continuous across the interface. The energy differences between the conduction and valence band edges at the interface are usually called the conduction-band and valence-band discontinuity, respectively.

- In a type-I structure (top), which has a straddled band lineup, the conduction and valence band edges of the material with a smaller bandgap are located within the bandgap of a material with a larger bandgap, e.g., AlGaAs-GaAs. The sum of the edge discontinuities equals the energy gap difference. In this structure, electrons and holes are localized in the lower-bandgap material.
- In a type-II heterostructure (middle), the discontinuities have different signs.



- In a type-III heterostructure (bottom), the band structure is such that the top of the valence band in one material lies above the conduction band minimum of the other material.
- In type-II and -III structures, staggered lineups are created wherein electrons and holes are localized in different materials.

Bloch Oscillations

Bloch oscillations are named after Felix Bloch, who showed theoretically that a wave packet associated with an electron and peaked at a momentum k undergoes periodic oscillations in the momentum space under an electric field. The effect, in principle, can be observed in all periodic systems. If k in the crystal increases and hits the boundary of the first BZ, the electron wraps its trajectory to the opposite side of the BZ due to the Bragg reflection. If it is assumed to remain in the same band, an electron will perform periodic motion.

Consider a 1D periodic lattice of period a. A particle of mass m and charge q in the presence of electric field E along the x direction is described by a Hamiltonian

$$H = \frac{1}{2m}(p + qEt)^2 + U(x)$$

The eigenstates of this Hamiltonian have the form

$$\psi(x,t) = \psi_{k,n}(x)e^{-iqEtx/\hbar}$$

where $\psi_{k,n}(x)$ is a Bloch function. The eigenstates obey periodic boundary conditions, i.e.,

$$\psi(x+L,t) = \psi(x,t) = \psi(x,t)e^{-i(k-qEt/\hbar)L}$$

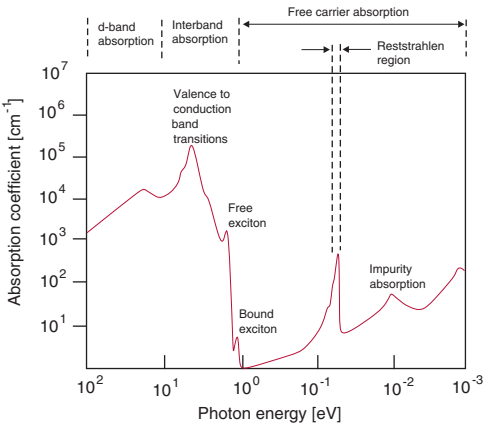
where $L = N \cdot a$, and N is a natural number. The fulfillment of the boundary conditions requires $k - (qEt/\hbar) = (n/N)(2\pi/a)$. A period of oscillations can be determined by starting with the equation $v_x = (1/\hbar)(\partial E/\partial k_x)$. The $E - k$ relation is sinusoidal, i.e., $E = A\cos ak$, so the speed can be integrated to obtain

$$x(t) = \int v(k(t))dt = -\frac{A}{qE}\cos\left(\frac{aqE}{\hbar}t\right)$$

Therefore, the period of Bloch oscillations is $T = (2\pi\hbar)/(aqE)$. Bloch oscillations have been observed in cold atoms and ultra-small Josephson junctions.

Optical Absorption in Semiconductors

Optical absorption in semiconductors (an important physical property) is characterized by a refractive index n and absorption coefficient α, which are related to the real and imaginary parts of the relative permittivity. They are dependent on the wavelength λ of the electromagnetic radiation (known as dispersion), as shown here with various possible absorption processes. (The photon energy increases from right to left).



1. Band-edge transitions, and the band-to-band absorption of photons, are associated with the excitation of an electron from the valence bands to the conduction bands. Near the band edge, the absorption coefficient is $\alpha(h\upsilon) = A(h\upsilon - E_g)^{1/2}$, where E_g is the direct band-gap, and A is a constant. This absorption coefficient can be derived by calculating the probability of a transition between states with different energies.
2. Exciton absorption peaks are close to the fundamental absorption edge and are observed at low temperatures.
3. **Reststrahlen**, or lattice absorption, is the process that occurs when the radiation is absorbed by the vibrations of the crystal ions.
4. Free-carrier absorption occurs.

Optical Properties of Semiconductors: Experiments

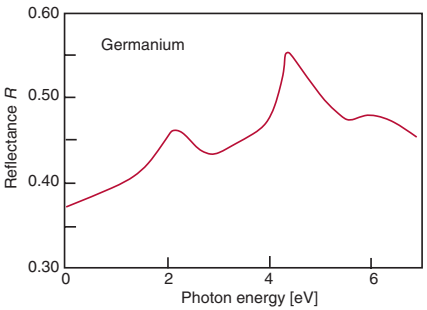
Various methods have been developed to experimentally measure the optical constants of semiconductors:

- **Reflectance measurements** are conducted on freshly cleaved crystals. Light in the optical frequency range is incident near normal to the cleaved surface. The detector is placed in front of the sample to measure the reflectivity.
- In **absorption measurements**, the detector is located behind the sample, which is a thin film.

The main parameters are the refractive index and the extinction coefficient. They are defined via the speed of light in a given medium v as $v = c/N = c/(n - i\kappa)$, where N is the complex index of refraction, n is the ordinary index of refraction, and κ is the extinction coefficient.

Reflectance R is defined as the ratio of the reflected and incident light intensities. It is related to the above parameters as

$$R = \frac{(n-1)^2 + \kappa^2}{(n+1)^2 + \kappa^2}$$

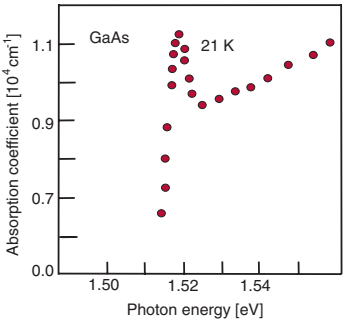


A typical result of the spectral dependence of the reflectance of germanium is shown here. The refractive index can be expressed as an empirical expression (**Sellmeier dispersion equation**)

$$n^2 - 1 = \frac{G_1\lambda^2}{\lambda^2 - \lambda_1^2} + \frac{G_2\lambda^2}{\lambda^2 - \lambda_2^2} + \frac{G_3\lambda^2}{\lambda^2 - \lambda_3^2}$$

The coefficients G_i are tabulated for popular semiconductors and their compounds.

Excitons



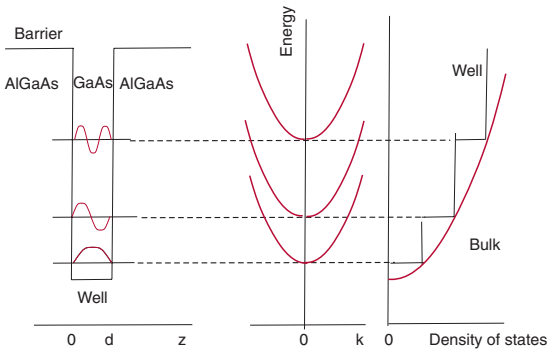
When an electron and a hole are created sufficiently close to each other, the Coulomb interaction can become significant, producing a situation where the electron and hole orbit each other or their centers of mass and form an electron–hole pair called an **exciton**. Excitons are observed in the absorption spectra of semiconductors and insulators just below the fundamental absorption edge. Typically, excitons have been studied in two limiting cases:

1. For strong electron–hole attraction, e.g., molecular crystals, the electron and the hole are tightly bound to each other within the same or nearest-neighbor unit cells (**Frenkel excitons**).
2. In most semiconductors and insulators, the Coulomb interaction is strongly screened by the valence electrons via the large dielectric constant, i.e., electrons and holes are only weakly bound (**Wannier–Mott excitons**).

When a light is used to probe the fundamental gap of an insulator, the electron-absorbing light cannot be excited to the minimum of the conduction band. The dielectric constant for an insulator is small due to the tightly bound electrons in the insulator. Consequently, the excited electron experiences an attractive Coulomb potential of a hole, left in the valence band, of $-e^2/(\varepsilon r)$ (Gaussian units), where e is the value of an electronic charge, ε is the dielectric constant, and r is the distance between the excited electron and the hole left behind. The exciton peak in the absorption spectra is shown here, and the energy of an exciton is localized in the gap and near the bottom of the conduction band.

Quantum Wells

Modern growth techniques, such as **molecular beam epitaxy (MBE)**, can grow thin layers of a semiconductor on the order of 100 Å, e.g., a layer of GaAs sandwiched between two AlGaAs materials. The larger bandgap of AlGaAs creates a potential that confines carriers (electrons and holes) within the GaAs layer. The figure here shows only the confining potential for the electrons; a similar (inverted) potential exists for the holes.



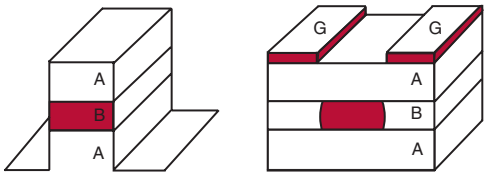
A **quantum well** creates a potential that confines a particle (such as an electron) to one dimension. The energy spectrum depends on three parameters: two continuous k_x, k_y, and the discrete quantum number n:

$$E = \varepsilon_n + \frac{\hbar^2}{2m}(k_x^2 + k_y^2)$$

A particle can move freely only in the x-y plane, i.e., within the plane of the quantum well. Motion along the z axis is quantized. The figure shows quantized energy levels along the z axis and free (parabolic) dispersions within the plane, i.e., along k_x and k_y. The rightmost figure depicts the DOS for 2D and bulk, i.e., large, samples.

There are many applications of quantum wells, including semiconductor lasers and light-emitting diodes and modulators. These devices show superior characteristics when compared to bulk devices.

Quantum Wires



Quantum wires are created when the motion of the carriers within a quantum well is further restricted along one direction only. They can be fabricated by imposing an additional confinement on 2D electron gas formed by a quantum well. The figure shows a heterostructure formed by two materials. Two possible methods are illustrated: (left) etching and (right) split-gating, where additional confinement is imposed by an electrostatic potential applied to a metallic split gate placed on top of the heterostructure.

Quantum wires are analyzed by solving a Schrödinger equation with the appropriate potential. Assuming that the wire extends along the x axis and that the potential restricts the movement along the y and z axes, the energies of a particle are

$$E = \varepsilon_l + \varepsilon_m + \frac{\hbar^2}{2m^*} k_x^2$$

Here, l and m are quantum numbers. For infinite potential barriers along the y and z directions, the discrete energies become

$$\varepsilon_l + \varepsilon_m = \frac{\hbar^2 \pi^2}{2m^* L_y^2} l^2 + \frac{\hbar^2 \pi^2}{2m^* L_z^2} m^2$$

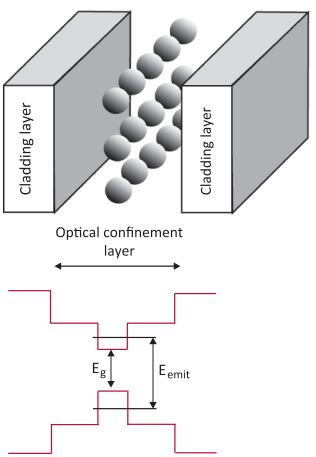
Quantum wires can be n- or p-doped, and they can be used to make electronic devices. Other applications include nanowire transistors. Both theoretically and experimentally, nanowire **metal-oxide-semiconductor (MOS)** transistors can be fabricated at the wafer level using **silicon-on-insulator (SOI)** substrates, and these novel devices have shown electrical properties that are comparable or even superior to standard transistors.

Quantum Dots

When quantum confinement occurs in all three dimensions, a quantum box or **quantum dot (QD)** is formed, which is a zero-dimensional (0D) system as the confinement of electron motion exists in all three directions. QDs are often called quantum atoms.

The length of a QD must be small in all three directions, i.e., L_x, L_y, $L_z \sim \lambda_F$, where $\lambda_F = 2\pi/k_F$ is the Fermi wavelength. In such a case, energy is quantized in all three directions: $E = \varepsilon_n + \varepsilon_m + \varepsilon_l$. QDs can be fabricated using self-assembly techniques, where lattice mismatch plays an important role. The effect is similar to the creation of water droplets on a glass. For example, the mismatch (which causes strain) between the Ge and Si lattice constants forms 3D islands when Ge is deposited epitaxially on Si. The shape of the islands depends on the number of deposited Ge monolayers.

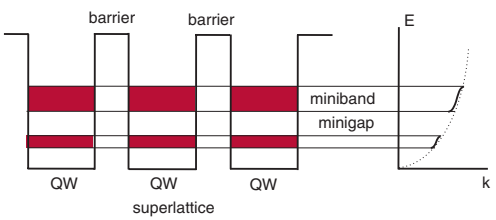
One possible application of QDs is semiconductor lasers, where QDs act as the active medium for light emission. The top figure depicts spherical QDs embedded in a semiconductor matrix. The bottom image shows the band diagram for valence and conduction bands for one dot. Materials selected for QDs must have a smaller bandgap than the surrounding medium.



High-density QD structures can be fabricated using various self-organized growth schemes as well as etching and regrowth. For semiconductor lasers based on QDs, several improved parameters have been observed, including modulation bandwidth, lasing threshold, and relative intensity noise.

Superlattices

An infinite sequence of identical quantum wells separated by identical barriers is known as a **superlattice**, and it is therefore a periodic structure fabricated from two different semiconductors that should have the same lattice constants to avoid defect states at the boundaries. They should also have different fundamental bandgaps to form quantum wells and barriers. The typical layer width is an order of magnitude larger than the lattice constants. The AlAs and GaAs superlattices were the first to be studied because their lattice constants matched. The BZ of the periodic superlattice direction is very small, and the 1D bands in that direction form subbands in the well regions. If the gaps of two semiconductors differ significantly and the Fermi energies of two semiconductors at the middle of the gaps are aligned, then the valence band of the large-gap semiconductor serves as a well for p-doped carriers in this region and the other serves as a barrier. If the smaller-gap semiconductor is n-doped, then its conduction band provides a well and the semiconductor with the large gap acts as a barrier. A superlattice can also be considered as a multiple-quantum-well system, shown here with the formation of minibands and minigaps.

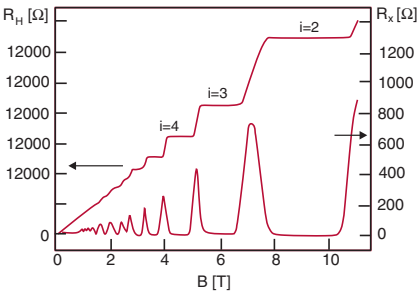


Superlattices are typically formed by introducing a periodic variation of donor or acceptor impurities or by periodic variation of the material composition during crystal growth.

An important application of superlattices is the quantum cascade laser, which uses intrasubband transitions instead of interband transitions, as in traditional lasers. It is usually used in the mid- to far-infrared portion of the electromagnetic spectrum.

Quantum Hall Effects

Two **quantum Hall effects** were discovered in 2D systems under large magnetic fields: the integer quantum Hall effect and the **fractional quantum Hall effect (FQHE)**.



The integer quantum Hall effect consists of a quantization of the Hall resistance R_H (no longer linear in a magnetic field, as in the classical Hall effect), which shows plateaus at particular values of the magnetic field. At each plateau there is a vanishing longitudinal resistance R_x. The Hall resistance is expressed in terms of a universal constant as

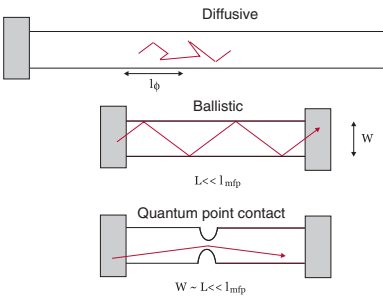
$$R_H = \left(\frac{h}{e^2}\right)\frac{1}{n}$$

where n is an integer. It occurs at low temperatures, so the thermal energy is much smaller than the spacing between Landau levels, i.e., $k_B T \ll \hbar\omega_c$. The quantum Hall effect is simply a quantization of the Hall resistance when Landau levels expand with a magnetic field. The effect does not depend on the properties of the sample, including impurities, and is therefore used as the resistance standard.

In high-quality samples (e.g., high mobility) at very low temperatures in large magnetic fields, a FQHE was also observed. The original discovery corresponded to $n = 1/3$. The FQHE is due to strong electronic correlations when the Landau level is only partially filled and there are strong Coulomb interactions. Later, it was theorized that the FQHE can be described using anions with non-Abelian statistics; these particles can play an important role in quantum computers.

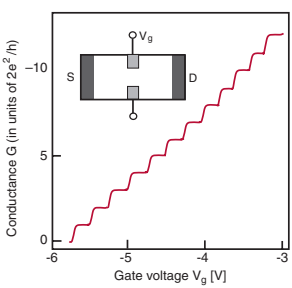
Quantum Point Contacts

With quantum wires, a different type of transport associated with the electronic motion can be expected. The illustration here depicts the trajectories of electrons in the diffusive and ballistic regimes. Here, W is the width of the constriction, L is the length of the wire, l_ϕ is the phase coherence length, and l_{mfp} is the mean free path. When the length of the wire is longer than the mean free path, an electron undergoes many scattering events, and the transport in the wire is diffusive. If the length of the wire becomes comparable to the mean free path, the transport is ballistic, which occurs when the width is comparable to the length and much smaller than the mean free path, corresponding to a **quantum point contact (QPC)**.

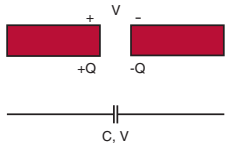


In a QPC, the constriction is very narrow, and therefore the electronic energies form 1D subbands. The conductance of such systems depends on a number of 1D bands in the channel. The resistance is determined by using the Landauer formula $R = h/(2e^2N) = (12.9/N)k\Omega$, where N is the number of conduction channels (subbands).

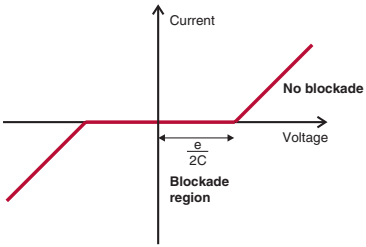
The measurements of conductance versus gate voltage shown here were obtained at temperature $T = 60$ mK. The inset depicts a so-called split-gate device. Gate voltage V_g controls the number of subbands open for electron transport. As the QPC narrows, N decreases, and therefore the conductance decreases in integral units of $G_0 = 2e^2/h$.



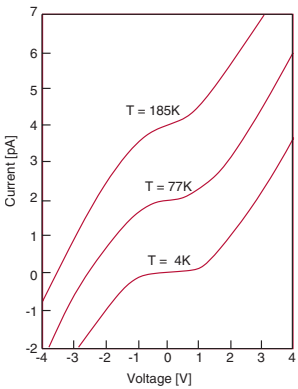
Coulomb Blockade



The movement of an electron from one electrode to another in a very small capacitor has a significant influence on the electrostatic potential and thus on the ground state of the system. Such an effect is generally referred to as the **Coulomb blockade (CB)**. Due to the CB effect, an electron cannot pass between the electrodes unless the energy in the system equals the Coulomb energy $e^2/2C$, i.e., the energy required to add an elementary charge to the capacitor. This effect can only be observed in a nano-scale capacitor and at very low temperatures. In the former, the CB manifests in the I–V characteristics. The image depicts a typical current–voltage characteristic (at $T = 0$).



The formula for Coulomb energy can be used to estimate the required value of the capacitance to see the effect. For the capacitance $C = 1 \times 10^{-13}$ F, the needed voltage to detect tunneling is ~0.7 μV. However, for a much smaller capacitor $C = 1 \times 10^{-19}$ F, the required voltage is ~0.7 V.



The temperature should be low enough to see the effect; specifically, the charging energy should be much greater than the thermal energy, i.e.,

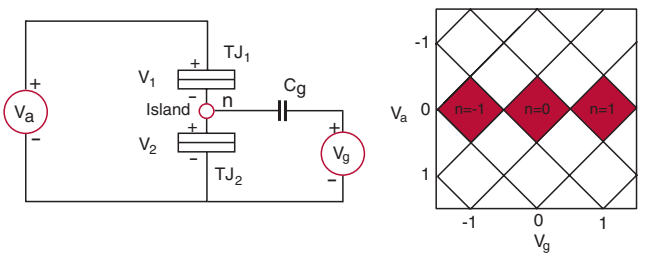
$$E_c = \frac{e^2}{2C} \gg \frac{1}{2} k_B T$$

For capacitance $C = 5 \times 10^{-17}$ F, the estimate is $T \ll 33$ K. Typical experimental results are shown here.

Single-Electron Transistor

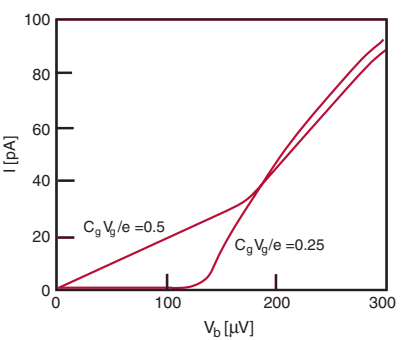
A **single-electron transistor (SET)** consists of a small metallic island connected to two leads by tunnel junctions. A third lead connected to a capacitor provides a gate voltage to the island. The leads probe the charging effects on the island. Only the case where the leads and the island are normal metals (not superconducting) is addressed here.

If the charging energy $E_C = e^2/(2C_\Sigma)$ is much larger than the thermal energy k_BT and the junction resistance is larger than the quantum resistance $h/e^2 = 25.8k\Omega$, then the charge on the island is quantized in units of e. Here, e is the charge of one electron, and C_Σ is the total capacitance of the island consisting of the gate capacitance and the intrinsic capacitances of the top and bottom tunneling junctions, respectively.

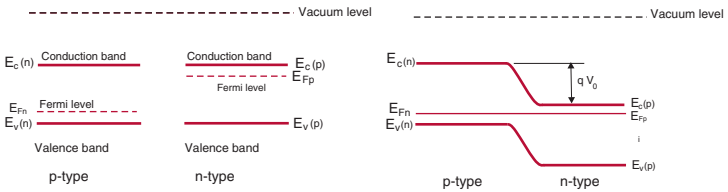


The left figure illustrates a SET where n is the number of excess electrons on the island, and the right is a stability diagram. Due to the Coulomb blockade, the current is zero in the shaded areas. The CB effect can be removed by tuning the potential V_g to the gate electrode.

The current–voltage characteristic of a SET depends on the gate voltage. This last figure shows the I–V characteristics for two values of gate voltage. The lower curve indicates the CB effect of tunneling, which is suppressed in the other curve.

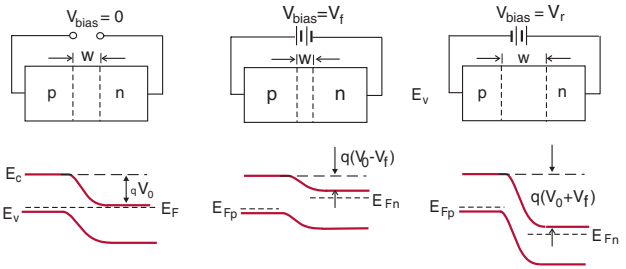


p-n Junction

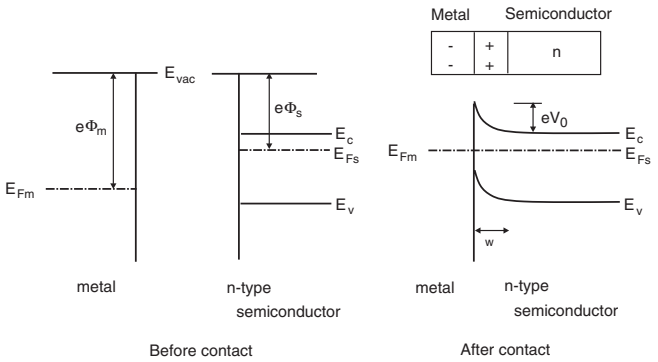


Shown here are energy diagrams before and after a **p-n junction** is made with n- and p-doped semiconductors. The Fermi levels E_F on both sides are not equal initially. Recall that E_F is related to the density of electrons; the unequal E_F causes a concentration gradient. The configuration after contact is not stable. Electrons of the n-type will be driven to the left by the concentration gradient. Some of the holes are driven to the right. As the holes drift to the right, they can recombine with electrons in the n-region and leave new positive charges in the p-region. These positive charges exert repulsive forces on holes moving to the right and stop the drifting. A similar situation happens to the electrons. Finally, an equilibrium condition can be established with no flow of current.

The density of the current is proportional to the gradient of E_F. Since there is no current flowing, E_F remains constant across the junction. With the external voltage, a p-n junction can be forward or reverse biased. Two quasi-Fermi levels are introduced in such cases, and their difference is proportional to the external bias.



Schottky Junction

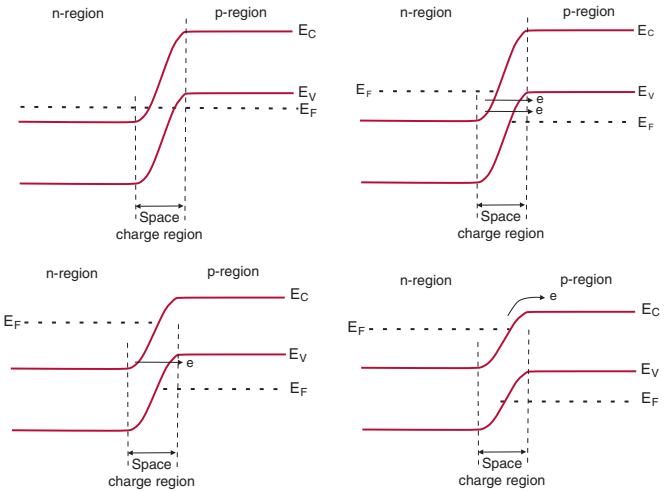


A metal–semiconductor contact can be formed either as a **Schottky junction** or an ohmic junction. The former behaves similar to a p-n junction, whereas the latter behaves like a resistor. The difference is mainly determined by the band alignment between the metal and the semiconductor. The figure shows a Schottky junction formed between a metal and an n-type semiconductor: (left) schematic of energy levels for the isolated materials and (right) after making contact. The work function Φ_m of a metal is the energy required to remove an electron from the Fermi level to vacuum outside the metal. In a similar way, the work function Φ_s of a semiconductor is defined. Since the Fermi energy of the semiconductor is higher than that of a metal, after contact is made a flow of electrons from a semiconductor to a metal forms a depletion region of width w in the semiconductor near the interface. As the charges in the depletion region of a semiconductor accumulate, the same amount of opposite-sign charges appears in a metal. The transfer of electrons stops when the Fermi levels in the metal and semiconductor are the same; the system is then in equilibrium. The built-in potential V_0 is thus created:

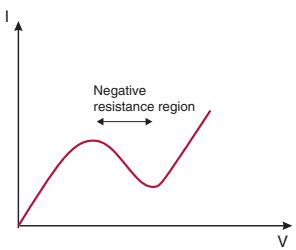
$$V_0 = |\Phi_m - \Phi_s|$$

Potential V_0 is determined by assuming that the depletion region exists only in a semiconductor.

Tunnel Diode



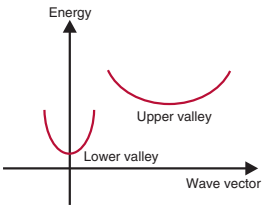
The structure shown in the top-left graph depicts heavily doped n- and p-type semiconductors and a thin barrier region between them. The respective Fermi energies are equal. The other graphs show the band alignment for the increasing forward-bias voltage. Electrons in the conduction band of the n-region energetically align with the holes in the valence band of p-region (top right). The tunneling of electrons from the n-region to the p-region results in the flow of a current. An increase in the bias voltage causes more electrons to tunnel. The resulting current reaches its maximum value when all electrons are aligned with holes. Further increases in the bias voltage cause the current to decrease because of the smaller alignment of electrons with holes (see below).



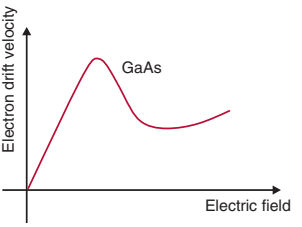
After the minimum value is reached and the bias voltage continues to increase, electrons are no longer energetically aligned with holes, tunneling diminishes, and diffusion contributes mostly to the current. The resulting I–V characteristic shows a region of negative differential resistance.

Gunn Effect

A Gunn diode operates based on the **Gunn effect**, which is the so-called transferred electron device. The effect manifests in heavily n-doped GaAs, wherein electrons populate the conduction band minimum around the center of the BZ. When a GaAs sample is connected to a battery, the Fermi energy E_F is raised, and when the E_F overlaps with the lowest conduction band at the X-point of the BZ (zone edge of the [100] direction), electrons are pushed from the center of the zone to the X-point (known as the L-valley). Electrons are transferred from one valley to another, both in the conduction band. The effective mass of an electron at point X (upper valley) is larger than the one at the center of the BZ. Mobility is large in the lower valley and very small in the upper valley. The two valleys are separated by a small energy gap (~0.3 eV).

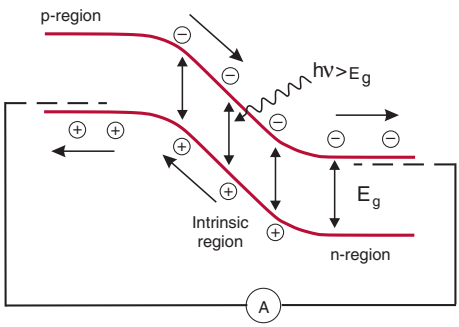


The externally applied electric field accelerates electrons. When they acquire sufficient energy, they will move into the low-mobility upper valley and slow down. A plot of the electron drift velocity versus the electric field is shown here. This dependence has a region where the differential mobility $\mu_d = dv/dE$ is negative, which in turn produces a region of negative differential conductivity $\sigma_d = en\mu_d$, which is a feature of the Gunn effect.



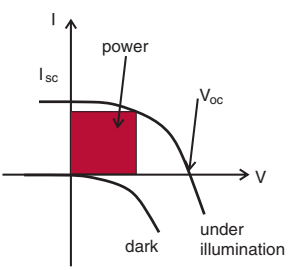
The mechanism was first suggested by Ridley and Watkins in 1961 and independently by Hilsum in 1962. Gunn was the first to observe microwave oscillations of a current in GaAs and InP in 1962. They were related to instabilities caused by the negative differential conductance.

Solar Cells



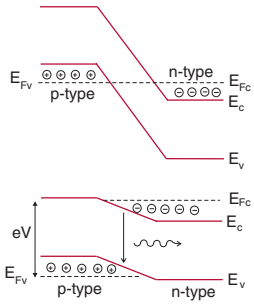
The Sun supplies more solar energy to the Earth in one hour than is used globally in one year. Thus, it is very important to invent devices that efficiently convert solar energy into electricity: **solar cells**. The simplest solar cell consists of a single p-i-n junction formed by an appropriately doped semiconductor. It produces potential difference and thus acts as a battery. The principle of operation is shown above. Photons of frequency ν with energy $\hbar\nu > E_g$, where E_g is the bandgap energy, generate electron–hole (e-h) pairs. Electrons go downward to the right, and holes go upward to the left. They contribute to the current passing through the load. Each generated electron increases the charge in the n-region by $-e$; similarly, each hole makes the p-region more positive by $+e$. Thus, a potential difference is created between metallic electrodes.

A semiconductor can only convert photons efficiently into current with energies that equal the bandgap. Photons with energies smaller than the bandgap are not absorbed, and photons with higher energies reduce their energies to the bandgap energy by thermalization of the photogenerated carriers, a process that involves losses.



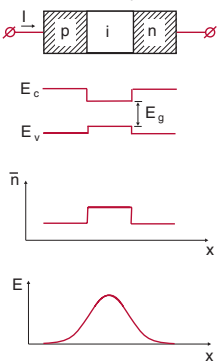
Solar cells are characterized by three important parameters: the short-circuit current I_{sc}, the open-circuit voltage V_{oc}, and the maximum power, represented by a shadowed square area.

Light-Emitting Diodes



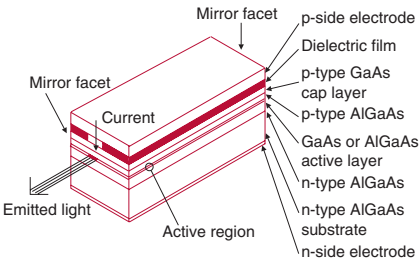
A **light-emitting diode (LED)** is essentially a p-n semiconductor junction that consists of heavily doped p- and n-type semiconductors. A **homojunction** (i.e., made of the same semiconductors) is shown here. The top image illustrates a situation with zero external voltage. The Fermi levels in the p- and n-regions are the same. The circles with − and + designate electrons and holes, respectively. The bottom image shows a forward bias junction with external voltage V. The Fermi levels are separated. The applied voltage raises the potential energy of the n-side and lowers that of the p-side. With a sufficient voltage, electrons and holes will acquire enough energy to move into the central junction region and recombine, thus generating photons. Such a system emits incoherent light when forward biased. The radiated wavelength is $\lambda = 1.24/E_g$, where E_g is the bandgap energy expressed in eV, and the wavelength is in micrometers. Different materials have different bandgap energies and thus emit radiation of different wavelengths.

The above homojunction LED is not a very efficient source as it does not confine carriers and photons very well. A better structure is created using materials with different bandgaps. A double-heterojunction structure, shown here, provides excellent confinement and thus strong interaction between carriers and electromagnetic radiation. Electrons and holes are confined to the central region, where they interact with an electromagnetic field. A profile of the refractive index, with a value slightly larger in the central region, acts as a guide for the electromagnetic field E.



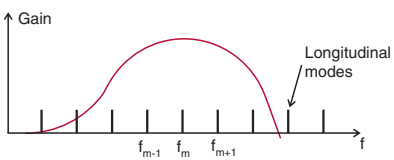
Semiconductor Lasers

Semiconductor laser diodes are similar to LEDs, i.e., they have comparable band diagrams. The main difference is that a laser requires an optical gain for stimulated amplification of the emitted photons and a resonant cavity for optical feedback. These are typically not required for LEDs, which output incoherent light.

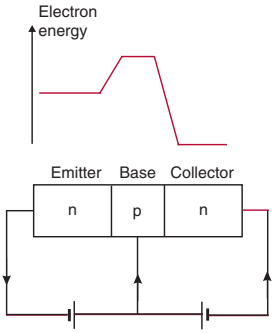


Lasers operate under forward bias. Carriers injected into the active region recombine there, causing spontaneous emission of photons. If the electrical current is larger than the so-called threshold value, then many carriers participate in the stimulated emission and generate a large gain. When the gain exceeds losses, laser oscillations occur. As a result, a laser generates coherent light with a narrow linewidth. Most semiconductor lasers are edge emitters, as shown above. They are just a bit more complicated than a p-n junction. The middle contains a material with a smaller bandgap energy. One of the typical materials used is AlGaAs-GaAs.

The **resonant cavity** of a laser is formed by two end mirrors that are created by cleaving the front and back faces of the semiconductor along parallel crystalline planes. The reflectance of the AlGaAs–air interface is ~32%, which produces sufficient feedback for oscillations. Typical cavity lengths are around 300 μm. The locations of the resonances or longitudinal modes of the cavity are illustrated here, along with the gain spectrum. The laser operates in the mode that has the largest gain.



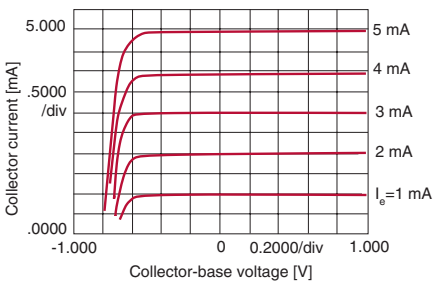
Bipolar Junction Transistor



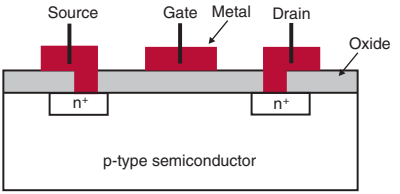
A **transistor** is a solid state analogue of a triode vacuum tube. The main action controls the flow of charges between two electrodes by varying the potential at the gate electrode.

A **bipolar junction transistor (BJT)**, shown here with an energy diagram, consists of two p-n junctions formed by, e.g., n-p-n regions. The thin region between the n-regions is called a **base**. The outer regions with the same type of conductivity (e.g., n-type) are called the **emitter** and the **collector**. In a BJT, small variations of the electric current or voltage cause significant variations in the current flowing between the emitter and the collector. The BJT can be designed as a homostructure or heterostructure, or with graded-band semiconductors.

The application of forward bias to the base (p-type) with respect to the emitter (n-type) reduces the barrier potential for electrons, allowing them to diffuse across the base–emitter junction. Since the base is very thin, electrons diffuse further across it into the collector. Thus, the base bias voltage efficiently controls the flow of the current between the emitter and the collector. The resulting current–voltage characteristics of a n-p-n Si BJT are shown here. Depending on their function, BJTs come in various types. The basic materials used in transistors are Si, Ge, GaAs, and InP. The major function that BJT provides is serving as a generator and/or amplifier.

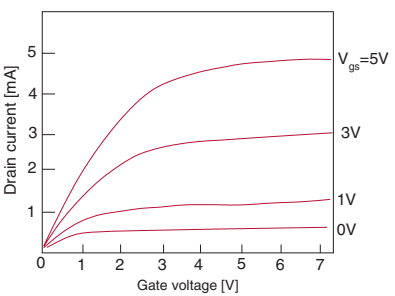


MOSFET



The first **metal–oxide–semiconductor field effect transistor (MOSFET)** was fabricated in 1960 and soon became the basic building block of very-large-scale integrated (VLSI) circuits. The device shown here is a metal–insulator–semiconductor planar configuration with an oxide as an insulator. The flow of carriers comes from the emitting electrode, known as the **source**, to the collecting electrode, the **drain**. The flow is controlled by the electrode in the middle, known as the **gate**. A small change of voltage at the gate modulates the width of the region under the gate, thus causing large changes of the value of a current flowing in the source–drain circuit (field effect). The limiting values of the source–drain current are determined by the geometry of the device and by doping. The frequency characteristics of the device (especially its upper operating frequency) are determined by the gate length and mobility of the carriers in the material.

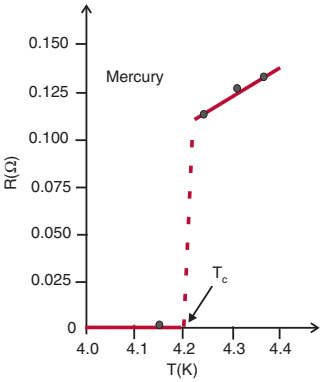
Typical characteristics, i.e., drain current versus gate voltage, for a few values of gate voltage are shown here. The gate voltage varies the conductance of the channel. The effect can be interpreted as a gate-controlled potential barrier.



MOSFETs are not limited to VLSI circuits. They play an important role in power-electronics circuits, as switching devices, and in microwave applications.

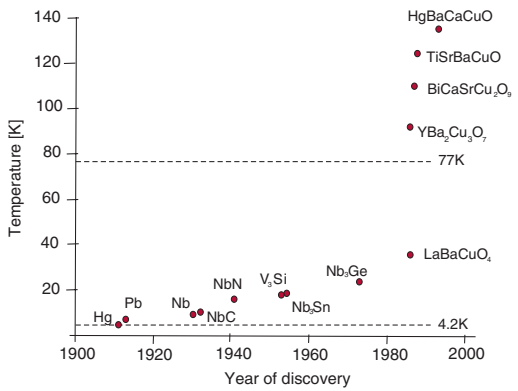
Zero Resistivity

Zero resistivity is the spectacular manifestation of a sample's superconductivity. It was discovered in 1911 by Kamerlingh Onnes, who measured the resistivity of mercury as a function of temperature and discovered that it suddenly drops to zero. The temperature at which that happens is known as the **critical temperature** T_C.

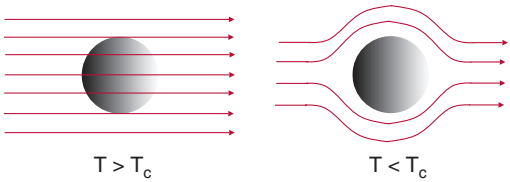


Element	Critical Temperature T_C (K)
Zinc (Zn)	0.85
Aluminum (Al)	1.18
Indium (In)	3.4
Tin (Sn)	3.7
Mercury (Hg)	4.15
Lead (Pb)	7.2

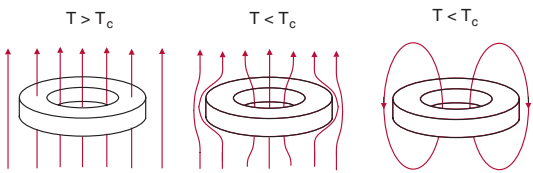
Historical changes in the critical temperature are shown below. The sudden increase around 1986 indicates the appearance of high-temperature superconductors.



Meissner Effect



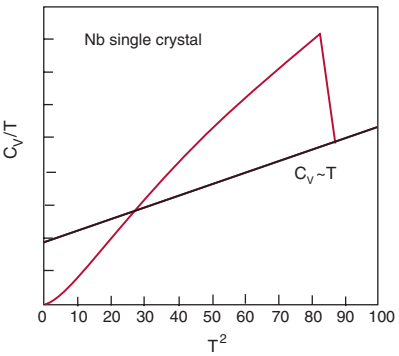
The expulsion of a magnetic field from a spherical superconductor is called the **Meissner effect** and is one of the fundamental properties of a sample in the superconducting state. In the normal state (left image), i.e., above the critical temperature, magnetic field lines pass through the material. When $T < T_C$ (right image), the magnetic field is expelled from the sample. It can only penetrate a superconducting material up to the so-called penetration depth λ, which is ∼17 nm.



Now consider a magnetic field inside and around a superconducting ring. The field inside the ring is trapped when the ring becomes superconductive, rather than expelled as for a superconductor without a hole. In the superconducting ring there is no decay of the flowing current (known as a persistent current). The current flows so long as it is kept below T_C. The circulating current generates magnetic flux (right image).

Zero resistivity and the Meissner effect define a superconductor. The Meissner effect causes some remarkable phenomena such as levitation, where a magnet floats above a superconductor (or vice versa). Such behavior allows for the frictionless movement of one of the objects on top of the other and is the principle behind ultra-high-speed trains.

Specific Heat



In a normal metal, the measured **specific heat** C_V for a typical superconductor above the critical temperature is $C_V = \gamma T + AT^3$. The first term, which has a linear dependence on temperature, represents the contribution from electrons, with γ known as the **Sommerfeld constant**. The AT^3 term represents the contribution of the lattice vibrations and originates from Debye's theory of a specific heat of solids.

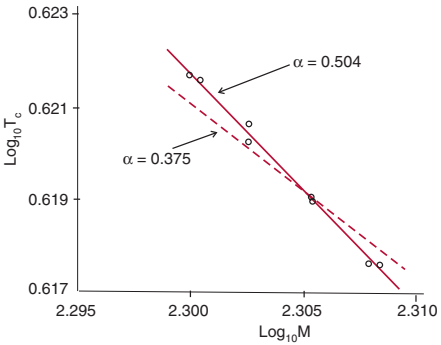
In the superconducting state, the experimental data of the specific heat shows a discontinuity at critical temperature T_C and then decreases slowly as the temperature is lowered:

$$C_V = AT^3 + B \exp\left(-\frac{\Delta}{k_B T}\right)$$

which means that the linear term, originating from electrons, is replaced by a new term that shows an exponential dependence on temperature. The exponential behavior suggests that when the metal becomes a superconductor, an energy gap characterized by Δ is involved and that thermal excitations cross the gap between the ground state and the excited state.

The experimental data on the specific heat played an important role in the development of the theory of superconductivity by indicating that any theory should account for the existence of an energy gap. It also shows that the normal–superconducting transition is a second-order phase transition because there is no change in latent heat and only the discontinuity in the specific heat can be noted.

Isotope Effect



The **isotope effect** was discovered in 1950 by Maxwell (and independently by Reynolds), who found that the critical temperature T_C in Pb is a function of the isotope mass. The experimental evidence indicates that the critical temperature T_C and the average isotopic mass M are connected by the relation $T_C M^{\alpha} = \text{constant}$, where the exponent $\alpha \approx 0.5$. This result can be understood by observing that vibrations of ions in crystals depend on their masses. Since quanta of these vibrations are known as phonons, it can be concluded that phonons play some role in the mechanism of superconductivity.

Consider the vibration of a diatomic molecule. Its vibration frequency is

$$f = \frac{1}{2\pi}\left(\frac{k}{\mu}\right)^{1/2}$$

where k is a force constant between two atoms, and μ is the reduced mass $1/\mu = 1/m_1 + 1/m_2$. After substituting various isotopes, k does not change, so the effect comes from the reduced mass μ. The values of α in some cases are much smaller, and for Ru and Zr, $\alpha = 0$.

The isotope effect indicates that the lattice plays an important role in the mechanism of superconductivity or, more precisely, that superconductivity is caused by an interaction between the electrons in a metal and the vibrations of the crystal lattice.

Two-Fluid Model and Order Parameter

Onnes was the first to liquefy helium and also discover the phenomenon of superconductivity in 1911. In 1937, **superfluidity** was discovered by Pyotr Kapitsa (and independently by John F. Allen). There is a connection between the frictionless flow of liquid helium below T_C and the flow of superconducting current. This fact was examined in 1938 by Fritz London, who suggested that superfluidity and superconductivity should be understood from the same point of view. The analogies can explain the emergence of the so-called **two-fluid model**. The focus here is on superconductors wherein the superfluid flow is associated with charged particles.

The two-fluid model assumes that there are two component fluids in a superconductor: the normal fluid and the superfluid. The superelectrons (at the Fermi energy) condense into a superfluid state (condensate), which extends over the entire sample and can move as a whole without friction. The superelectrons can flow without friction through the sample. At nonzero temperatures, some electrons will transition from the condensate and form a weakly interacting gas of excitations (normal fluid) that also extends over the entire sample. Normal electrons are subject to Ohm's law, just as in normal metals.

Normal and superconducting components can be described within the Drude model by the following equations for current densities:

$$\mathbf{J}_s = -en_s\mathbf{v}_s, \quad \mathbf{J}_n = -en_n\mathbf{v}_n$$

The Newton equation $m(d\mathbf{v}_s/dt) = -e\mathbf{E}$ is used for the superfluid component. The so-called **first London equation** combines the last relation with the expression for superfluid current density:

$$\frac{d\mathbf{J}_s}{dt} = \frac{e^2 n_s}{m}\mathbf{E}$$

London Equation

London developed a theory based on classical electrodynamics to explain the Meissner effect. To illustrate his theory, assume that superconducting electrons are described by

$$\psi(\mathbf{r}) = \sqrt{\rho(\mathbf{r})}\ \exp\ (i\phi(\mathbf{r}))$$

where $\rho(\mathbf{r})$ is the density of superconducting electrons, and $\phi(\mathbf{r})$ is the phase of the condensate. After substituting into the quantum-mechanical expression for the current

$$\mathbf{J}_s = \frac{ie\hbar}{2m}(\psi^*\nabla\psi - \psi\nabla\psi^*)$$

in the presence of a magnetic field $\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}$, where $\mathbf{A}$ is the vector potential, the result is

$$\mathbf{J}_s = \frac{n_s e^2}{m_e}\mathbf{A}$$

Apply $\nabla\times$ to this relation to obtain the **London equation**:

$$\nabla \times \mathbf{J}_s = \frac{e^2 n_s}{m_e}\mathbf{B}$$

where n_s is the density of superconducting electrons. It expresses the link between an external magnetic field in a superconductor with the current. The combination of the London equation with Ampere's law $\nabla \times \mathrm{B} = \mu_0 \mathrm{J}_s$ produces the following differential equation for the spatial variation of the magnetic field inside the sample:

$$\nabla^2\mathbf{B} = \frac{\mu_0 q_s^2 n_s}{m_s}\mathbf{B} = \frac{1}{\lambda^2}\mathbf{B}$$

where λ is the London penetration depth. In order to analyze the distribution of a field close to the surface of a superconductor, consider the 1D case where the x axis is perpendicular to the interface between the superconductor and the air, and vector $\mathbf{B}$ is parallel to the interface and perpendicular to the x axis. In such a situation, the solution to the above equation gives a field that decays exponentially as $B(x) \sim \exp(-x/\lambda)$. The parameter λ gives the penetration depth of a magnetic field inside a superconductor.

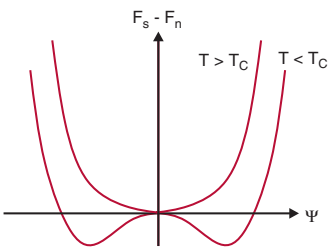
Ginzburg–Landau Theory

The **Ginzburg–Landau (GL) theory** of superconductivity is based on the following assumptions:

- a complex-order parameter in the superconducting state ψ goes to zero at the phase transition, and
- the free energy F can be expanded in powers of ψ.

The Taylor expansion of the free energy in powers of the order parameter (without currents and magnetic fields) is

$$F_s = F_n + \alpha(T)|\psi|^2 + \frac{1}{2}\beta(T)|\psi|^4 + \ldots$$



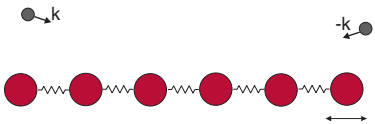
Physically, the order parameter is related to the local density of superconducting electrons as $n_s = |\psi|^2$, with $\alpha(T)$ and $\beta(T)$ as parameters, depending on the temperature. The equilibrium condition $\partial F_s/\partial|\psi|^2 = 0$ leads to $|\psi|^2 = -\alpha(T)/\beta(T)$. When ψ has a spatial variation and if there is an external magnetic field, kinetic energy $(1/2m^*)\hat{\mathbf{p}}^2$ is added, where $\hat{\mathbf{p}}$ is the canonical momentum operator $\hat{\mathbf{p}} = -i\hbar\nabla - e^*\mathbf{A}(\mathbf{r})$, and $\mathbf{A}(\mathbf{r})$ is the vector potential. In such a case, the free energy becomes

$$F = \int \left\{ F_0 + \frac{1}{2m^*}|[-i\hbar\nabla - e^*\mathbf{A}(\boldsymbol{r})]\psi(\mathbf{r})|^2 \right\}$$

It contains two unknown parameters m^* and e^*, which are used to determine various properties of superconductors, such as the current density:

$$\mathbf{J}_s = -\frac{\partial F_s}{\partial \mathbf{A}} = -\frac{ie^*\hbar}{2m^*}(\psi^*\nabla\psi - \psi\nabla\psi^*) - \frac{(2e^*)^2}{m^*}|\psi|^2\mathbf{A}$$

Cooper Pair (Electron–Phonon Interaction)



Due to the measured isotope effect on the transition temperature of a superconductor, the importance of the interaction between electrons and vibrational modes has been recognized. In 1956, Cooper analyzed the interaction between a pair of electrons. The **Cooper pair** model is illustrated by two electrons in a 1D lattice interacting with an ionic chain. The figure here depicts an electron that comes from the left with momentum $\hbar\mathbf{k}$. It attracts the positively charged ion and causes the ion to move. The collective motion of the ions creates a phonon. The ions move back and forth about their respective equilibrium positions, as shown by the arrow below the right-side ion. If another electron approaches the right side of the ionic chain and its closest ion (the right-side ion) moves in the "proper phase," i.e., toward the approaching electron, there is a net attraction between the electron and the ion on the right side. Since the vibration motion of the ions is induced by the left-side electron, there is a net attraction between the electrons. This is the paring concept that explains how two electrons can attract each other through the mediation of a phonon. The interaction is not strong due to the phase requirement for the ion moving toward the electron. The effective electron–electron interaction potential is

$$V = -4\pi e^2/(|\varepsilon_k - \varepsilon_{k'}|^2 - (\hbar\omega_q)^2)$$

where ε_k and $\varepsilon_{k'}$ are energies of electrons, and ω_q is the phonon frequency. An attractive potential between the two electrons requires $|\varepsilon_k - \varepsilon_{k'}| < \hbar\omega_q$. The conditions to form a Cooper pair are therefore:

1. $\mathbf{k}' = -\mathbf{k}$ to reduce Coulomb repulsion because the Fourier transform of the Coulomb interaction is $4\pi e^2/|\mathbf{k} - \mathbf{k}'|^2$.
2. Opposite spin orientations of two electrons are required to keep those electrons close and maintain their phase coherence.

Cooper Problem

As a result of the electron–phonon interaction, the net interaction of two electrons can be attractive and form a bound pair. The pairing conditions are

1. The momenta of the electrons that form a pair should be in opposite directions.
2. The spins of the electrons should be opposite.

Cooper considered two identical electrons, constrained to a spin singlet state with a center of mass at rest. The orbital wave function of such a system is

$$\psi_{\text{orb}}(\mathbf{r}_1, \mathbf{r}_2) = \psi_{\text{orb}}(\mathbf{r}_1 - \mathbf{r}_2) = \sum_{\mathbf{k}} c_{\mathbf{k}} e^{i\mathbf{k}(\mathbf{r}_1 - \mathbf{r}_2)}$$

Electrons interact through a potential $V(\mathbf{r})$ in the presence of a filled Fermi sea. The problem is described by the following Schrödinger equation:

$$-\frac{\hbar}{2m}(\nabla_1^2 + \nabla_2^2)\psi(\mathbf{r}_1, \mathbf{r}_2) + V(|\mathbf{r}_1 - \boldsymbol{r}_2|)\psi = (E + 2\varepsilon_F)\psi$$

After defining the energy eigenvalue E of this problem relative to $2\varepsilon_F$, where ε_F is the Fermi energy,

$$E = -2\varepsilon_c \frac{1}{\exp(2/N(0)V_0) - 1} \approx -2\varepsilon_c e^{2/N(0)V_0}$$

where $\varepsilon_c \ll \varepsilon_F$ is a cut-off energy, and $N(0)$ is the DOS at ε_F. The coherent condensate of the Cooper pairs within a metal is described by a complex wave function (known as the order parameter) $\psi = |\psi|e^{i\theta}$ with phase θ, which depends on the position.

Cooper pairs have the binding energy $2\Delta \ll \varepsilon_F$ (Δ is the energy gap) and are characterized by a correlation radius, the so-called coherence length $= 10^{-8}$ m (larger than the average electron–electron distance, which is on the order of 10^{-9} m). Therefore, the wave functions of the Cooper pairs strongly overlap. The formation of Cooper pairs in a charged-electron Fermi liquid in a metal creates superconductivity.

BCS Theory

The **Bardeen–Cooper–Schrieffer (BCS) theory** of superconductivity states that the Cooper pairs act more like bosons than fermions. At $T = 0$, a pair function is introduced, which is a combination of occupied and unoccupied pair states, and $u_{\mathbf{k}}^2 + \upsilon_{\mathbf{k}}^2 = 1$. The ground state of the system is then represented as the product of pair functions

$$|\Psi_{BCS}\rangle_k = \prod_k (u_k|1\rangle_k + \upsilon_k|0\rangle_k)$$

The central parameter of the BCS approach is the order parameter Δ_k, determined by minimizing the energy $E = <H>$, which produces an equation obeyed by the order parameter. By assuming $\Delta_k = \Delta$,

$$1 = \frac{g}{V}\sum_k \frac{1}{2\sqrt{\xi_k^2 + \Delta^2}} \tanh \frac{\sqrt{\xi_k^2 + \Delta^2}}{2k_B T}.$$

with $V_{\mathbf{kk}'} = -g$, where g is a constant. Here, $V_{\mathbf{kk}'}$ is the effective interaction between electrons with wavevectors $\mathbf{k}$ and $-\mathbf{k}$ and spins $\uparrow$ and $\downarrow$. Excitation energies for electrons near the Fermi energy are $\xi_k = \varepsilon_{\mathbf{k}} - \mu$ (μ is the chemical potential).

The previous gap equation produces

$$\Delta(T) = 3.2k_B T_C \left(1 - \frac{T}{T_C}\right)^{1/2}$$

and an expression for the transition temperature

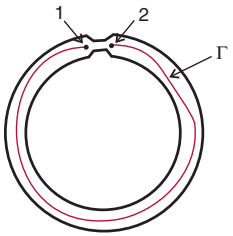
$$T_C \approx 1.14\hbar\omega_D \exp\left(-\frac{1}{N(0)g}\right)$$

The BCS theory also explains specific heat experiments. At T_C, the specific heat is

$$c_s - c_n \approx \frac{4mp_F}{7\xi(3)\hbar^3} T_C$$

where $\xi(3)$ is the Riemann zeta function.

Flux Quantization



An interesting effect predicted by the GL theory (and later verified experimentally) is the quantization of a magnetic flux. To establish **flux quantization**, consider a thick superconducting ring (transversal dimension ≫ Meissner penetration depth) with a weak link. Superconducting electrons in the ring are described by the "wave function" of the condensate

$$\psi(\mathbf{r}) = \sqrt{\rho(\mathbf{r})}\exp(i\theta(\mathbf{r}))$$

where $\rho(\mathbf{r})$ is the density of the Cooper pairs. The expression for the supercurrent flowing in the ring is

$$\mathbf{j}_s = \frac{ie\hbar}{2m}(\psi^*\nabla\psi - \psi\nabla\psi^*)$$

In the presence of a vector potential $\mathbf{A}$, the above expression for the wave function of the condensate can be used to obtain the supercurrent:

$$\mathbf{j}_s = \frac{e\hbar}{m}\left(\nabla\theta - \frac{2\pi}{\phi_0}\mathbf{A}\right)\rho(\mathbf{r})$$

where $\phi_0 = h/(2e)$ is the flux quantum. Now consider the line integral of the supercurrent over contour Γ. The integral is further replaced by the surface integral. Inside the superconductor, the supercurrent vanishes, and

$$\int_\Gamma \nabla\theta d\mathbf{l} = \int_{\text{surface}} \nabla\theta d\mathbf{l} + \Delta\theta = 2\pi n + \Delta\theta$$

where $\Delta\theta$ is the phase difference across a weak link, and phase θ is defined by modulo 2π. A combination of these results produces the **fluxoid quantization condition**

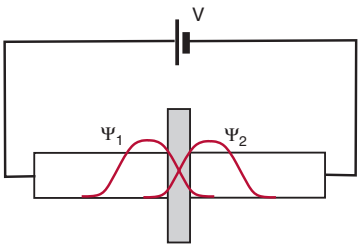
$$\Phi + \frac{\phi_0}{2\pi}\Delta\theta = n\phi_0$$

where the magnetic flux is defined as

$$\Phi = \int_{\text{surface}} (\nabla \times \mathbf{A})d\mathbf{s}$$

Josephson Effect

The **Josephson effect** occurs in a variety of situations where two superconductors are separated by a "weak link," which can be an insulating region, normal metal, or narrow constriction. In such situations there exists "normal tunneling" of single electrons. (Josephson showed that there is also tunneling of Cooper pairs.)



To establish the main relations, consider two superconductors separated by a thin insulator (thickness of ~10 nm). Both superconductors are characterized by the macroscopic quantum-mechanical wave functions ψ_1 and ψ_2. The dynamics of the two superconductors are described by the following coupled Schrödinger equations:

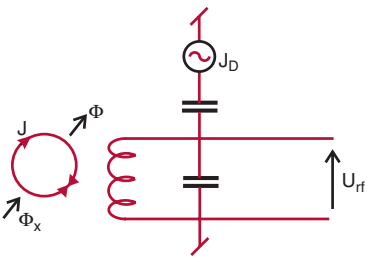
$$i\hbar\frac{\partial\psi_1}{\partial t} = E_1\psi_1 + K\psi_2, \quad i\hbar\frac{\partial\psi_2}{\partial t} = E_2\psi_2 + K\psi_1$$

where $\psi_{1,2}$ are the many-body wave functions for the Cooper pair states in superconductors 1 and 2, and K is the coupling constant due to tunneling. After writing the wave functions as $\psi_{1,2} = |\psi_{1,2}|\exp(i\theta_{1,2})$, substituting into above equations, and assuming identical superconductors,

$$\frac{d\theta}{dt} = \frac{2\,\mathrm{eV}}{\hbar}$$

Here, $\Delta\theta = \theta_2 - \theta_1$, and $E_2 - E_1 = 2$ eV. The current passing through the Josephson junction can also be evaluated with the result $I = I_0 \sin\Delta\theta$, which depends on the phase difference across the junction. I_0 is called the **critical current**, and it is a parameter characteristic of the junction.

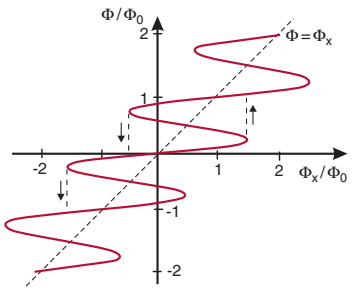
rf-SQUID



The **superconducting quantum interference device (SQUID)** is a very sensitive instrument used for the non-destructive measurement of magnetic fields. A radio-frequency-based version (called a **rf-SQUID** or one-junction device, shown here) consists of a superconducting ring interrupted by a weak link that, in particular, can be a Josephson junction. The ring is coupled with the resonant circuit, which is driven by an external current source J_D. The response of the SQUID to the external magnetic Φ_x is contained in the voltage U_{rf}.

The response current can be determined using the Josephson equation for the current and fluxoid quantization condition

$$\Phi + \frac{\Phi_0}{2\pi}\Delta\theta = n\Phi_0$$



where n is a natural number, $\Delta\theta$ is the phase difference across the Josephson junction, and $\Phi_0 = h/(2e)$ is the flux quantum. The supercurrent becomes

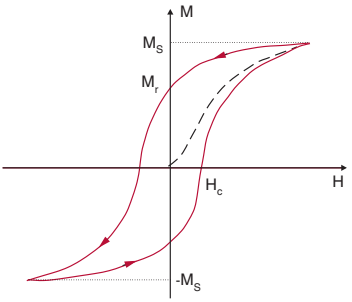
$$I = -I_0 \sin\left(2\pi\frac{\Phi}{\Phi_0}\right)$$

which indicates that the current output as a function of the magnetic field (flux) is oscillatory. The dependence of a total flux versus external flux in a hysteretic mode is shown here. Transitions between states are indicated.

SQUIDs allow for the measurement of magnetic fields as small as 10^{-15} Tesla, which is some 10^{-11} times smaller than the Earth's natural magnetic field.

Traditional Magnetism

From the classical point of view, **magnetism** relates to magnetic moments and their behavior under external magnetic fields. There are various physical mechanisms and types of magnetism that operate at different length scales, such as exchange or Ruderman–Kittel–Kasuya–Yosida (RKKY).



A fundamental property of magnetic substances is **magnetization**. Its origin is linked to the properties of electrons. Magnetization can be intrinsic or induced by an external magnetic field. Therefore, magnetic materials can be divided into two groups: one that shows magnetic properties, i.e., has a nonzero magnetic moment after the application of an external magnetic field, and one that possesses a spontaneous magnetic moment, even without an external magnetic field. **Paramagnetic** and **diamagnetic** materials belong to the first group, whereas **ferromagnets** and **antiferromagnets** are in the second group. When an external magnetic field is applied to a ferromagnetic material, the change of magnetization shows hysteretic behavior. A **hysteresis loop** with fundamental points in the **M-H** plane includes saturation magnetization $\mathbf{M}_s$, remnant magnetization $\mathbf{M}_r$, and coercive field $\mathbf{H}_c$.

A strong response to a magnetic field exists in materials whose atoms have magnetic moments, and these magnetic moments interact amongst themselves. The magnetic moment of an electron is caused by its intrinsic properties of spin and electric charge. Opposite to localized magnetic moments, there is the magnetism of delocalized electrons, which uses the terminology of itinerant magnetism. Spontaneous magnetism appears due to the phase transition when either the temperature or pressure is changed.

Concept of Spin

There is experimental evidence that particles possess an internal degree of freedom that behaves like an angular momentum and is termed **spin**. Spin is a physical, observable trait of a particle for which there is no classical analog. The symbol **S** denotes the spin angular momentum operator. Its Cartesian components are Hermitian operators that satisfy the following commutation relations:

$$[S_x, S_y] = i\hbar S_z, \quad [S_y, S_z] = i\hbar S_x, \quad [S_z, S_x] = i\hbar S_y$$

Particles with integer spin ($s = 0, 1, 2, \ldots$), in units of $\hbar$, are called bosons, whereas those with half-odd-integer spin ($s = 1/2, \ 3/2, \ \ldots$) are called fermions. Electrons are fermions with spin $(1/2)\hbar$.

Spin plays an important role in determining the behavior of particles with spin in a magnetic field. For that, consider $E(B)$ to be the energy of an electronic system in the presence of a magnetic field **B**. The magnetic moment is defined as $\mu = -(\partial E/\partial \mathbf{B})$. The magnetic susceptibility χ (in SI units) is defined as $\chi = M/B$, where M is the magnetization of a system, i.e., magnetic moment per unit volume.

The interaction energy of the spin with magnetic field **B** is

$$\Delta E(S) = -\frac{e}{m_0}\mathbf{S} \cdot \mathbf{B}$$

It is introduced analogous to an electron with angular momentum **L** moving in a closed orbit inside magnetic field **B**. That energy is

$$\Delta E(L) = -\frac{e}{2m_0}\mathbf{L} \cdot \mathbf{B}$$

The Bohr magneton $\mu_B = (e\hbar)/(2m_0)$ is a useful measure of the magnetic energy.

Heisenberg Hamiltonian

The exchange interaction occurs because the wave function of two electrons must be antisymmetric under the exchange of the electron coordinates and spin, i.e., $\psi(\mathbf{r}_1, \mathbf{s}_1; \mathbf{r}_2, \mathbf{s}_2) = -\psi(\mathbf{r}_2, \mathbf{s}_2; \mathbf{r}_1, \mathbf{s}_1)$. The interaction between spins can be expressed in terms of spin operators (or classical variables) in a model known as the **Heisenberg model**, which has spin variables given by three-component real-valued vectors and is described by the following Hamiltonian:

$$H = -\sum_{i \neq j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j$$

A typical situation assumes that J_{ij} is nonzero only for the nearest neighbors, i.e.,

$$J_{ij} = \begin{cases} J & \text{for } i,j = < i,j > \\ 0 & \text{all other cases} \end{cases}$$

For homogeneous ferromagnetic interactions, i.e., where $J_{ij} = J > 0$ for all interactions, the directions of all spins become the same in the ground state. For $J_{ij} < 0$, one has antiferromagnetic interactions.

In the Heisenberg model, the finite-temperature phase transition does not occur in one or two dimensions. The long-range order of continuous spins at finite temperatures is forbidden (**Mermin–Wagner theorem**). Thus, the finite-temperature phase transition caused by the spontaneous breaking of continuous symmetry does not occur. Conversely, when the spatial dimension is more than two, the phase transition occurs at a finite temperature.

The **Heisenberg Hamiltonian** is the starting point for many calculations of the magnetic materials, such as the Weiss molecular field. Weiss suggested (before quantum mechanics was invented) that a spontaneous magnetization of iron is due to the alignment of the atomic magnetic moments and had postulated the existence of a molecular field proportional to the magnetization to explain this alignment.

Ground-State Parameters of Ions: Hund's Rules

At the fundamental level, the magnetic phenomena of solid state systems depend on the magnetic moments of the constituent atoms, which can be created by the electron's motion in an orbit around the nucleus.

Electrons fill atomic shells in such a way as to minimize the energy of the atom. An atom can be treated as a "black box" object and characterized by some number of phenomenological parameters. **L** and **S** are the total orbital and spin angular momenta of the atom, respectively. The momenta interact via the spin–orbit interaction and produce the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$.

In most cases, the ground state and appropriate values of **L**, **S**, and **J** are determined by **Hund's rules (HRs)**, determined to analyze atomic spectra:

1. Electrons occupy orbits to maximize the value of **S** allowed by the Pauli exclusion principle, i.e., the most possible electron spins must be aligned in parallel to minimize the Coulomb interaction between electrons. In this way electrons are kept as far apart as possible.
2. Electrons occupy orbits to maximize the orbital quantum number **L**, consistent with rule 1. Thus they maximize the alignment of orbital angular momenta. In order to minimize their interactions, electrons should orbit in the same direction as much as possible.
3. The total angular momentum **J** is given as $\mathbf{J} = \mathbf{L} - \mathbf{S}$ for a shell that is filled less than half and as $\mathbf{J} = \mathbf{L} + \mathbf{S}$ for orbits filled more than half.
4. If all atoms remain in their ground states, the HRs can estimate magnetic moments. The rules work well to determine the specific ground state of atoms but cannot specify their excited states.

These rules can be broken. For example, rule 4 is often broken in transition metal ions.

Ferromagnetism

One of the traditional types of magnetism is **ferromagnetism**, where the magnetism is finite even without an external magnetic field. Such ferromagnetic materials exhibit one type of magnetic moment (e.g., atomic moments of a single element) and also strong exchange interactions that orient those moments in parallel. As a result, spontaneous magnetization exists below the Curie temperature T_C (even without an external magnetic field). Above T_C, spontaneous magnetization vanishes, and the system behaves like a paramagnet; susceptibility depends on the temperature and is described according to the Curie–Weiss law.

The theoretical description of ferromagnetic metals is based on two main approaches:

- d-electron ferromagnets use the model of delocalized electrons (known as band magnetism), and
- f-electron ferromagnets use the Heisenberg model (localized electrons).

The classical theory of ferromagnetism was developed by Weiss and based on molecular field approximation. In his approximation, the effect on a given spin i from all other spins in the solid is considered as the mean effective field:

$$E_i = -2\sum_j J_{ij}\mathbf{S}_i \cdot \langle \mathbf{S}_j \rangle - g\mu_B \mathbf{S}_i \cdot \mathbf{B} = -\mathbf{S}_i \cdot \mathbf{B}_i^{eff}$$

where $<\mathbf{S}_j>$ is the average value of the other spins in the system.

Ferromagnetism was observed in three of the 3d transition metals (Fe, Co, Ni), in six of the 4f transition metals (Gd, Tb, Dy, Ho, Er, Tm), in a series of alloys, and in intermetallic compounds. These systems are classified according to the electronic structure of the atoms.

Diamagnetism and Paramagnetism

Diamagnetism, which is inherent to all substances, occurs when a material shows magnetization antiparallel to an applied magnetic field. It is a very weak effect that can be explained using Faraday's law: an increasing magnetic field accelerates electrons within the material. Electrons contribute to an electrical current flowing according to Lenz's law in the direction that reduces the strength of the applied field. The current persists even if the applied field has a constant value. The resulting magnetization $\mathbf{M}$ (magnetic moment per unit volume) is in the direction opposite to the field. Diamagnetism is described by the magnetic susceptibility χ, defined as $\mathbf{M} = \chi\mathbf{H}$. Typical values are $\chi \sim -10^{-6}$, i.e., small and negative. In metals and semiconductors, diamagnetism results due to quantization of an electron's orbit in an external magnetic field (known as Landau diamagnetism). In those systems, magnetic susceptibility χ becomes nonlinear and is a periodic function of the magnetic field.

Paramagnetism is the opposite effect to diamagnetism. Paramagnetic materials do not possess nonzero total magnetic moment without an external magnetic field. In a magnetic field, they show a nonzero magnetic moment parallel to the external field. There are several types of paramagnetism:

- Langevin (orientation) paramagnetism occurs when the orientation of magnetic moments of particles in the magnetic field is unrelated to their motion.
- Pauli paramagnetism exists in a degenerate Fermi liquid due to an unequal number of fermions with different values of spin orientation. The magnetic susceptibility does not depend on temperature.
- Van Vleck (polarization) paramagnetism is related to quantum mechanical corrections that are due to the admixture of excited states with nonzero magnetic moments into the ground (nonmagnetic) state of the atoms.

Paramagnetism is used to obtain extremely low temperatures (adiabatic demagnetization).

Bloch Equation and Relaxation Times

The precession of magnetization $\mathbf{M} = [M_x, M_y, M_z]$ in a magnetic field $\mathbf{B}$ in the presence of energy dissipation is described by the **Bloch equation**

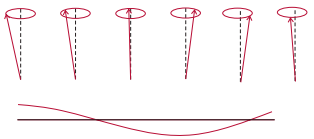
$$\frac{d}{dt}\mathbf{M} = \gamma \mathbf{M} \times \mathbf{B} + \frac{M_x \hat{\mathbf{i}} + M_y \hat{\mathbf{j}}}{T_2} - \frac{M_z - M_0}{T_1}\hat{\mathbf{k}}$$

where M_0 is the equilibrium value of M_z; $\hat{\mathbf{i}}, \hat{\mathbf{j}}, \hat{\mathbf{k}}$ are unit vectors in the laboratory coordinate system; T_1 is the longitudinal relaxation time (spin–lattice relaxation); and T_2 is the perpendicular relaxation time (spin–spin relaxation). Typically, $T_2 \ll T_1$. Time T_1 describes the change of a component of the magnetic moment parallel to the field $\mathbf{B}$, whereas T_2 illustrates the change of a component perpendicular to $\mathbf{B}$. In a magnetic field, magnetic dipoles precess around field $\mathbf{B}$ with the frequency (known as the Larmor frequency) $\omega = (eg_J B)/(2me)$. During the precession, the z-component of magnetization remains constant.

The principle is known as **nuclear magnetic resonance (NMR)**. The precession of magnetization forms the basis of several methods used to determine magnetic structures. To the general public, the effect is known as **magnetic resonance imaging (MRI)**, which has an impact on medical and biological studies, as well as chemical and materials research.

In a typical experiment, the sample of study is placed in a large constant magnetic field, and simultaneously a rf electromagnetic field is applied. An increase in absorption is observed with the resonance. For most ions in a magnetic field ~0.3 T, the Larmor frequency is on the order of 10 GHz (microwaves), whereas for protons in a magnetic field of 2 T, the frequency is ~50 MHz (short wavelengths). Therefore, the microwave techniques cannot be used to analyze living organisms.

Spin Waves



The propagating disturbance in a ferromagnet at low temperature is known as a **spin wave**. The figure shows one propagating perpendicular to spontaneous magnetization. The dispersion relation is

$$\hbar\omega = 4JS[1 - \cos(ka)]$$

where J is the NN exchange interaction, and S is the constant value of all spins in the perfectly aligned state. Here, a is the separation between spins, and k is a wavevector.

The existence of spin waves is due to the strong electrostatic exchange interactions, weak spin–orbit interaction, and dipole–dipole interaction between magnetic atoms. The interactions shift the intrinsic magnetic moments (spins) of atoms from their equilibrium orientations, which results in their propagation in the form of waves–normal modes.

The above classical version of spin waves has a quantum equivalent. The quantized energy of a mode of wavevector k has a form typical for a harmonic oscillator and is

$$E = \left(n + \frac{1}{2}\right)\hbar\omega(k)$$

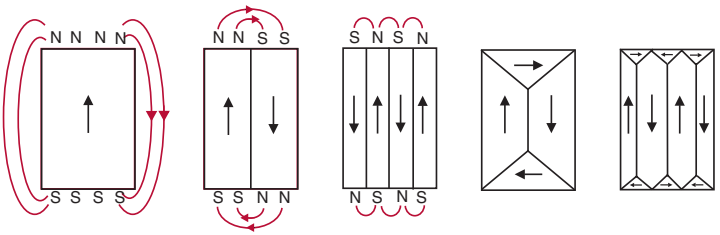
The quantized excitation of a spin wave is known as a **magnon**, which, like a phonon, is a boson. Its contribution to the heat capacity varies as $T^{3/2}$ at low temperatures.

Spin waves are excited by the application of special converters such as antennas, strip-line waveguides, or by a technique of parametric excitation when a single pumping photon breaks into two spin waves.

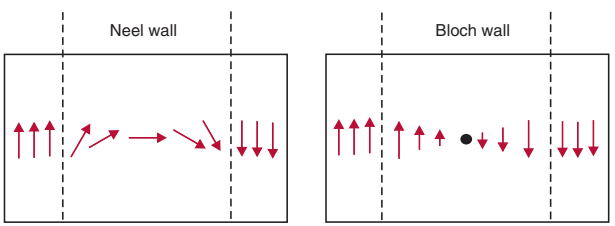
Spin waves are also detected in nonmagnetic materials. The oscillations of the spin density of conduction electrons are due to the exchange interaction. The experimental observation of spin waves in nonmagnetic metals under an external magnetic field has provided evidence that an electron gas exhibits properties of Fermi liquids.

Magnetic Domains

In ferro- and ferrimagnetic materials, a strong and short-range exchange interaction competes with a weak but long-range dipole–dipole interaction. The former tries to order ferromagnetically neighboring moments, whereas the latter tries to destroy the ferromagnetic order. As a result of the competition, the sample is divided into regions, known as **magnetic domains**, in which magnetic moments are ordered ferromagnetically. Between the domains are so-called domain walls (**Bloch walls**) where magnetic moments change orientation.



All spins in the leftmost figure are parallel to each other, and the whole ferromagnetic material exists as a single domain. A very powerful magnet is thus formed. The second image shows the single domain divided into two equal domains with different spin directions. Here, a single-pole magnet is replaced by two magnets with opposite polarities so that the magnetic field is cancelled and does not spread much in space; the magnetic energy is effectively reduced. More complicated situations are shown in the remaining figures. The following figure shows the structure of the Bloch and Neel walls separating the domains, with different spin behaviors.



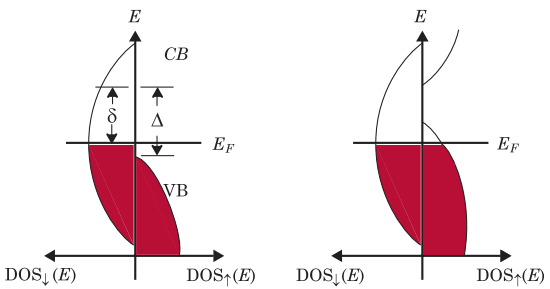
Half-Metals

A **half-metal** (NiMnSb) was the first one predicted in 1983 by de Groot et al. using first-principles calculations. A special feature of a half-metal is that the states of one of its spin channels are intersected by the Fermi energy, E_F, while the other spin channel exhibits semiconducting behavior with the E_F falling in the gap. Because most electrons with one spin type contribute to the conducting states at the E_F, half-metals are ideal for spintronic devices.

The well-known **Julliere formula** can be used to determine the spin polarization P of a sample:

$$P = \frac{N_- - N_+}{N_- + N_+}$$

where N_- is the downspin DOS at E_F, and N_+ is the corresponding DOS for the upspin states.



The left figure shows the DOS of a half-metal. The E_F intersects the states of the metallic channel (the left half); in the right half, E_F is in the gap defined by the valence band (VB) and the conduction band (CB). $P = 1$ for a half-metal. The E_F in the right figure intersects both spin channels, showing the metallic behaviors of a ferromagnetic metal.

Due to the difficulty of growing high-quality samples, the experimental determination of the half-metallicity for NiMnSb is still unresolved. The material proved experimentally to be a half-metal is CrO_2, which has a rutile structure.

Bibliography

Ashcroft, N. W. and N. D. Mermin, *Solid State Physics*, Holt, New York (1976).

Cohen, M. L. and S. G. Louie, *Fundamentals of Condensed Matter Physics*, University of Cambridge Press, Cambridge, UK (2016).

Hook, J. R. and H. E. Hall, *Solid State Physics*, 2nd ed., John Wiley & Sons, New York (2013).

Kittel, C., *Introduction to Solid State Physics*, 8th ed., John Wiley & Sons, New York (2013).

Rogalski, M. S. and S. B. Palmer, *Solid State Physics*, Gordon and Breach Science Publishers, Amsterdam (2000).

Schmool, D. S., *Solid State Physics: From the Material Properties of Solids to Nanotechnologies*, Mercury Learning and Information, Dulles, VA (2017).

Singh, R. J., *Solid State Physics*, Pearson India, Delhi (2011).

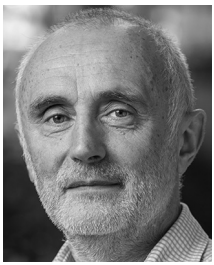
Snoke, D. W., *Solid State Physics: Essential Concepts*, Addison-Wesley, San Francisco (2009).

Index

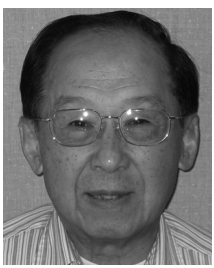
Index

Index

Index



Marek S. Wartak is a Professor in the Department of Physics and Computer Science at Wilfrid Laurier University (Waterloo, Ontario). He has over 30 years of experience in solid state physics, semiconductor physics, photonics and optoelectronics, analytical methods, modeling, and computer-aided design tools. He has worked at Nortel Networks, National Research Council in Canada, California, and Europe. He has led numerous projects and collaborations in modeling and simulations for electronic devices, and active and passive optoelectronics and photonics devices. Those simulations ranged from materials to devices. He was named the 2005–2006 University Research Professor. He is a Senior Member of IEEE, Electron Device Society, and OSA. He is the author of *Computational Photonics: An Introduction with Matlab* (Cambridge University Press, 2013), which has also been translated into Chinese. He teaches solid state physics regularly.



C. Y. Fong is a Distinguished Professor of Physics at the University of California, Davis. He is a Fellow of the American Physical Society and the Institute of Physics, UK. He was elected as Chair of the American Physical Society, California-Nevada Section, in 2007, and he was selected as an Outstanding Referee by the American Physical Society in 2008. He has served as Chair of the Selection Committee for the Nicholas Metropolis Prize of the American Physical Society (2016). Fong has edited two books— *Properties of Impurity States in Superlattice Semiconductors* and *Topics in Computational Materials Science*—and he currently serves as one of the editors of the "Series on Spintronics" (World Scientific, Inc.). He is the author of two books: *Half Metallic Materials and Their Properties* and *Recent Progress in Silicon-based Spintronic Materials*.