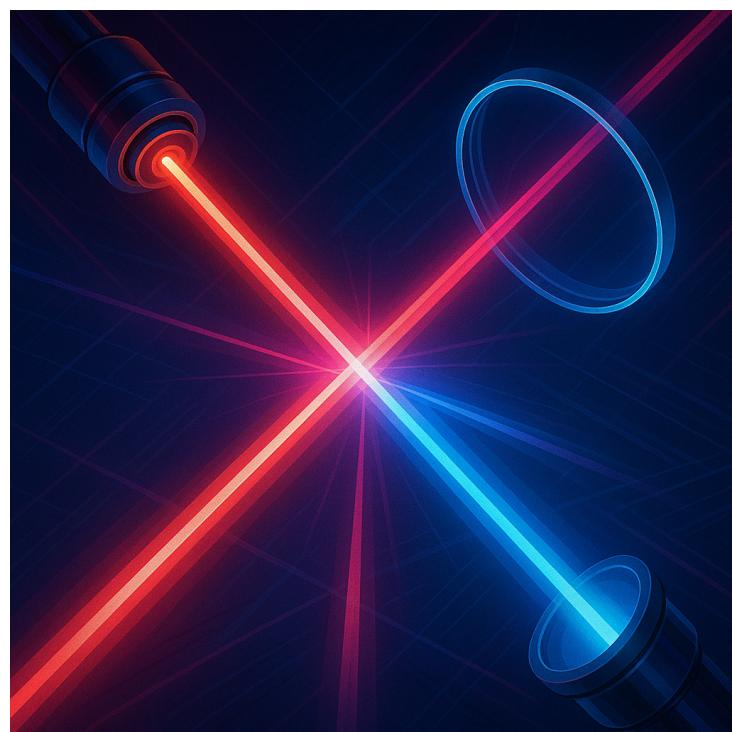


UNIVERSITÀ DI PISA

Notes on Physics of Materials for Photonics

Notes written by Jacopo Martellotto



June 26, 2025

Contents

1 Structural Properties of Semiconductors	5
1.1 Crystal Growth	5
1.1.1 Bulk Crystal Growth	5
1.1.2 MBE: Molecular Beam Epitaxy	6
1.2 Crystal Structure	7
1.2.1 Lattice Types	7
1.2.2 Basic Crystal Structures	9
1.3 Artificial Structures: Superlattices and Quantum Wells	10
1.3.1 Surfaces: Ideal Versus Real	11
1.3.2 Interfaces	11
1.4 Defects in Semiconductors	11
1.4.1 Point Defects	12
1.4.2 Line defects: Dislocations	12
1.5 Strained Heterostructures	13
1.5.1 Coherent and Incoherent Structures	13
1.6 Strain Tensor in a lattice mismatched epitaxy	14
2 Semiconductor Bandstructure	17
2.1 Introduction	17
2.2 Block Theorem and Crystal Momentum	17
2.2.1 Significance of the k -Vector	19
2.3 Metals, Insulators, and Semiconductors	20
2.4 Tight-Binding Method	22
2.4.1 Tight Binding Model for s-Bands	25
2.4.2 Bandstructure of Semiconductors	28
2.5 Spin-Orbit Coupling	29
2.5.1 Symmetry of Bandedge States	31
2.6 Selected Bandstructures	33
2.6.1 Silicon	33
2.6.2 GaAs	34
2.7 Mobile Carriers: Intrinsic Carriers	36
2.8 Doping: Donors and Acceptors	39
2.8.1 Carriers in Doped Semiconductors	41
2.8.2 Mobile Carrier Density and Carrier Freezeout	42
2.8.3 Equilibrium Density of Carriers in Doped Semiconductors	43
2.8.4 Heavily Doped Semiconductors	44
3 Bandstructure Modifications	45
3.1 Bandstructure of Semiconductor Alloys	45
3.1.1 GaAs/AlAs Alloy	48

CONTENTS

3.2	Bandstructure Modifications by Heterostructures	48
3.2.1	Bandstructure in Quantum Wells	50
3.2.2	Valence Bandstructure in Quantum Wells	52
3.3	Strain and Deformation Potential Theory	53
3.3.1	Strained Quantum Wells	55
3.4	Polar Heterostructures	55
4	Lattice Vibrations and Phonons	59
4.1	Introduction	59
4.2	Lattice Vibrations	59
4.2.1	Phonon: Quantization of Lattice Vibrations	63
4.3	Phonon Statistics	64
4.3.1	Conservation Laws in Scattering of Particles involving Phonons	65
4.4	Polar Optical Phonons	65
4.5	Phonons in Heterostructures	67
4.5.1	Confined Optical Phonons	67
4.5.2	Interface Phonons	68
4.6	Phonon Scattering: General Formalism	68
5	Optical Properties of Semiconductor	71
5.1	Maxwell Equations and Vector Potential	71
5.2	Electrons in Electromagnetic Field	74
5.3	Interband Transitions	78
5.3.1	Interband Transitions in Bulk Semiconductors	78
5.3.2	Interband Transitions in Quantum Wells	81
5.4	Indirect Interband Transitions	83
5.5	Intraband Transitions	86
5.5.1	Intraband Transitions in Bulk Semiconductors	86
5.5.2	Intraband Transitions in Quantum Wells	87
5.6	Charge Injection and Radiative Recombination	89
5.6.1	Spontaneous Emission Rate	89
5.6.2	Gain in a Semiconductor	91
5.7	Nonradiative Recombination	93
5.7.1	Charge Injection: Nonradiative Effects	93
5.8	Semiconductor Light Emitters	94
5.8.1	Light Emitting Diode	95
5.8.2	Laser Diode	95
6	Excitonic Effects	101
6.1	Introduction	101
6.2	Excitonic States in Semiconductors	102
7	Paper Exam	107
Bibliography		113

Structural Properties of Semiconductors

1.1 Crystal Growth

1.1.1 Bulk Crystal Growth

Semiconductor technology depends critically upon the availability of high quality substrates with as large a diameter as possible. Bulk crystal growth techniques are used mainly to produce substrates on which devices are eventually fabricated. For some semiconductors like Si and GaAs the bulk crystal growth techniques are highly matured. For the growth of boules from which substrates are obtained, one starts out with a purified form of the elements that are to make up the crystal.

One of the most important techniques that is used is the Czochralski method, which is used to grow single crystal silicon boules. The process involves melting high purity silicon in a crucible and then dipping a seed crystal into the melt. The seed crystal is slowly pulled upwards while rotating, allowing the silicon to crystallize onto the seed. The CZ technique is widely employed for Si, GaAs and InP.

A second bulk crystal growth technique involves a charge of material loaded in quartz container. The charge may be composed of either high quality polycrystalline material or carefully measured quantities of elements which make up a compound crystal. The container called a "boat" is heated till the charge melts and wets the seed crystal.

The seed is then used to crystallize the melt by slowly lowering the boat temperature starting from the seed end.

The easiest approach for the boat technique is to use a horizontal boat. However, the shape of the boule that is produced has a D-shaped form.

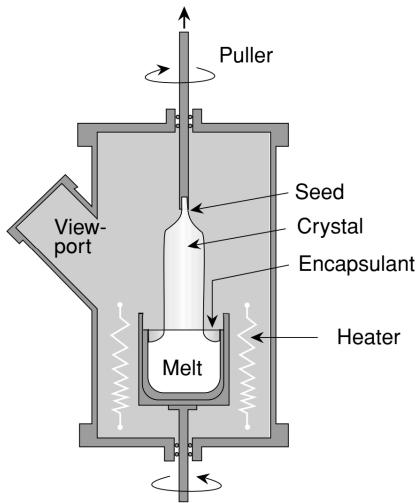


Figure 1.1. Schematic of the Czochralski (CZ) method for growing single crystal silicon boules.

1.1.2 MBE: Molecular Beam Epitaxy

MBE is capable of controlling deposition of submonolayer coverage on a substrate and has become one of the most important epitaxial techniques. Almost every semiconductor has been grown by this technique. MBE is a high vacuum technique in which crucibles containing a variety of elemental charges are placed in the growth chamber.

The elements contained in the crucibles make up the components of the crystal to be grown as well as the dopants that may be used. When a crucible is heated, atoms or molecules of the charge are evaporated and these travel in straight lines to impinge on a heated substrate.

The growth rate in MBE is ~ 1.0 monolayer per second and this slow rate, coupled with shutters placed in front of the crucibles, allows one to switch the composition of the growing crystal with monolayer control. Since no chemical reactions occur in MBE, the growth is the simplest of all epitaxial techniques and is quite controllable. However, since the growth involves high vacuum, leaks can be a major problem. The low background pressure in MBE allows one to use electron beams to monitor the growing crystal.

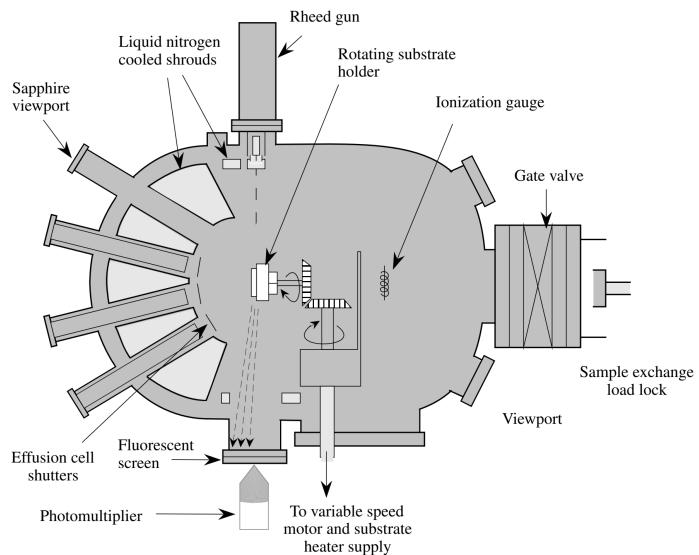


Figure 1.2. Schematic of a molecular beam epitaxy (MBE) system.

1.2 Crystal Structure

Crystals are composed of identical fundamental units, which can be individual atoms or clusters of atoms. In naturally occurring crystals, the crystalline symmetry is inherent and determined by natural processes. However, recent developments in crystal growth techniques now enable scientists to create artificial crystals with altered crystalline structures.

To define a crystal structure, we use two key concepts. First, the lattice is an abstract, periodically arranged set of points in space, where every point has an identical environment. Second, a basis, which is a group of atoms, is attached to each of these lattice points, creating the full crystal structure.

An important property of lattice is the ability to define three vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 , such that any lattice point \mathbf{R}' can be expressed as:

$$\mathbf{R}' = \mathbf{R} + m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3$$

where \mathbf{R} is a lattice point, and m_1 , m_2 , and m_3 are integers. Such a lattice is called **Bravais lattice**. The vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 are called the lattice vectors, and they define the unit cell of the crystal. The unit cell is the smallest repeating unit that can be used to construct the entire crystal structure by translation along the lattice vectors.

$\mathbf{Lattice} + \mathbf{Basis} = \mathbf{Crystal\ Structure}$

The translation vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 are called primitive if the volume of the cell is formed by them is the smallest possible. There is no unique way to choose the primitive vectors. One choice is to pick:

- \mathbf{a}_1 to be the shortest period in the lattice;
- \mathbf{a}_2 to be the shortest period not parallel to \mathbf{a}_1 ;
- \mathbf{a}_3 to be the shortest period not coplanar to \mathbf{a}_1 and \mathbf{a}_2 .

The volume cell enclosed by the primitive vectors is called the primitive cell, and it is the smallest volume that can be used to construct the entire crystal structure by translation along the lattice vectors.

Due to the periodic nature of a lattice, it's helpful to establish the structure's symmetry. This symmetry is characterized by a group of point group operations, which are transformations performed around a specific point. These operations include **rotation**, **reflection**, and **inversion**. The symmetry of a crystal significantly influences its electronic properties.

1.2.1 Lattice Types

The different types of lattice structures found in nature are characterized by their **symmetry groups**, with **rotation** being a crucial one. Lattices can exhibit rotational symmetries of 2π , $\frac{2\pi}{2}$, $\frac{2\pi}{3}$, $\frac{2\pi}{4}$, and $\frac{2\pi}{6}$. These are commonly denoted as 1, 2, 3, 4, and 6, respectively. Other rotation axes, such as $\frac{2\pi}{5}$ or $\frac{2\pi}{7}$, are not permitted because such structures cannot completely fill an infinite space.

In three dimensions, there are 14 distinct types of lattices. These lattice classes are defined by the relationships between their **primitive vectors** \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 , as well as the **angles** α , β , and γ between them. The most general lattice is **triclinic** (where $\alpha \neq \beta \neq \gamma$ and $\mathbf{a}_1 \neq \mathbf{a}_2 \neq \mathbf{a}_3$), and there are 13 additional special lattice types.

We will focus on the **cubic lattice**, which is the structure adopted by all semiconductors. There are three types of cubic lattices: **simple cubic**, **body-centered cubic**, and **face-centered cubic**.

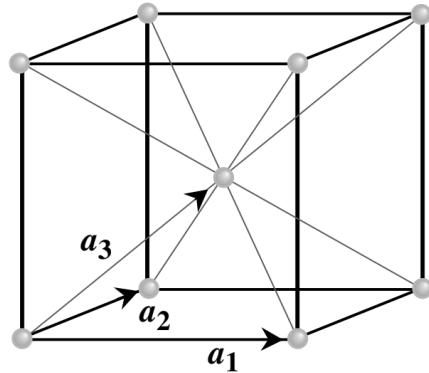


Figure 1.3. Schematic of the cubic lattice structure.

Simple Cubic

The simple cubic lattice is generated by the primitive vectors

$$a\mathbf{x}, a\mathbf{y}, a\mathbf{z}$$

where the $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are unit vectors.

Body-Centered Cubic

The bcc lattice can be generated from the simple cubic structure by placing a lattice point at the center of the cube. If $\hat{\mathbf{x}}, \hat{\mathbf{y}},$ and $\hat{\mathbf{z}}$ are three orthogonal unit vectors, then a set of primitive vectors for the body-centered cubic lattice could be

$$\mathbf{a}_1 = a\hat{\mathbf{x}}, \quad \mathbf{a}_2 = a\hat{\mathbf{y}}, \quad \mathbf{a}_3 = a\hat{\mathbf{z}}$$

A more symmetric set for the bcc lattice is

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

Face-Centered Cubic

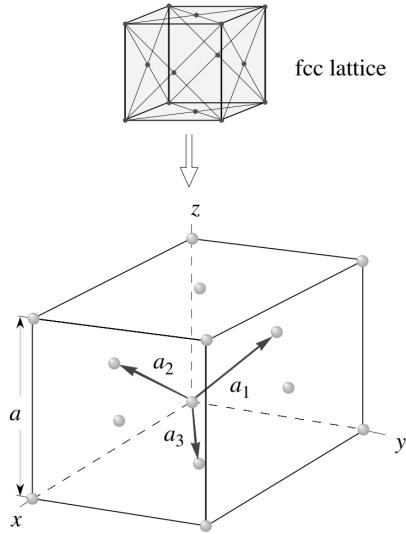


Figure 1.4. Primitive vectors of the face-centered cubic (fcc) lattice.

The face-centered cubic (fcc) Bravais lattice is another fundamental structure commonly found in semiconductors. It can be generated by starting with a simple cubic lattice and adding an extra lattice point at the center of each of the cube's faces. A convenient and symmetric choice of primitive vectors for the fcc lattice is given by:

$$\begin{aligned}\mathbf{a}_1 &= \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{z}}) \\ \mathbf{a}_2 &= \frac{a}{2}(\hat{\mathbf{y}} + \hat{\mathbf{z}}) \\ \mathbf{a}_3 &= \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}})\end{aligned}$$

Both the face-centered cubic and the body-centered cubic Bravais lattices play a crucial role in solid-state physics, as many materials crystallize in these arrangements, with atoms or ions located at each lattice point. Nearly all semiconductors relevant to electronics and optoelectronics adopt the fcc crystal structure.

1.2.2 Basic Crystal Structures

Diamond and Zinc Blende Structures

Most semiconductors relevant to electronics and optoelectronics are based on an underlying face-centered cubic (fcc) lattice. However, unlike simple fcc structures, they feature a basis consisting of two atoms. The positions of these two basis atoms are:

$$(0, 0, 0) \quad \text{and} \quad \left(\frac{a}{4}, \frac{a}{4}, \frac{a}{4}\right)$$

Because each of these atoms occupies its own fcc sublattice, the full crystal structure can be interpreted as two interpenetrating fcc lattices. One of these lattices is displaced from the other by a translation along the body diagonal direction:

$$\frac{a}{4}(1, 1, 1)$$

This type of arrangement gives rise to the well-known diamond and zinc blende structures, depending on the atomic species involved.

Hexagonal Close-Packed (HCP) Structure

The hexagonal close-packed (HCP) structure is another important crystal arrangement found in many materials, including some semiconductors. The HCP lattice can be visualized as a stacking of close-packed planes of atoms, with each plane being arranged in a hexagonal pattern.

The primitive vectors for the HCP lattice can be defined as follows:

$$\begin{aligned}\mathbf{a}_1 &= \frac{a}{2}(\hat{\mathbf{x}} + \hat{\mathbf{y}}) \\ \mathbf{a}_2 &= \frac{a}{2}(-\hat{\mathbf{x}} + \hat{\mathbf{y}}) \\ \mathbf{a}_3 &= c\hat{\mathbf{z}}\end{aligned}$$

Here, c is the height of the hexagonal unit cell, and the relationship between a and c is determined by the packing efficiency of the structure.

The HCP structure is characterized by a coordination number of 12, meaning each atom is in contact with 12 others. This high coordination number contributes to the stability and strength of materials with an HCP crystal structure.

Notation to Denote Planes and Points in a Lattice: Miller Indices

A straightforward convention is used to describe lattice planes, directions, and points within a crystal. For lattice planes, the procedure is as follows:

1. Define the coordinate axes (typically aligned with the primitive vectors).
2. Determine where the plane intersects the x , y , and z axes, using units of the lattice constants.
3. Take the reciprocal of these intercepts and reduce the resulting values to the smallest set of integers.

The resulting set of integers (hkl) is used to denote a family of parallel lattice planes, referred to as a *Miller index*. The notation $\{hkl\}$ indicates a family of crystallographically equivalent planes related by symmetry.

1.3 Artificial Structures: Superlattices and Quantum Wells

The electronic and optical characteristics of semiconductors can be significantly modified by forming heterostructures—combinations of different semiconductor materials. Advanced epitaxial growth techniques such as Molecular Beam Epitaxy (MBE) and Metal-Organic Chemical Vapor Deposition (MOCVD) enable precise control of the chemical composition at the monolayer scale (on the order of ~ 3 Å). These methods have been successfully used to grow a wide range of semiconductor materials, from those with zero bandgap like α -Sn and HgCdTe, to wide bandgap compounds such as ZnSe and CdS. Through heteroepitaxial growth, it is possible to fabricate layered structures with atomic-level precision, allowing intentional variation of the crystal's periodicity along the growth direction. This leads to the creation of *superlattices*, where two or more semiconductor materials—typically labeled A and B—are alternately deposited with respective thicknesses d_A and d_B . The resulting periodicity in the growth direction is $d_A + d_B$. An example is the (GaAs)/(AlAs) superlattice. Remarkably, modern epitaxial techniques have achieved alternating layer thicknesses down to a single monolayer. Although superlattices demonstrate the high precision of these growth methods, the most commonly used heterostructures in practical applications are *quantum wells*. A quantum well consists of a thin layer of a lower bandgap semiconductor embedded between two layers of a higher bandgap material. These structures

harness quantum mechanical effects, which are now essential in the operation of many electronic and optoelectronic devices.

1.3.1 Surfaces: Ideal Versus Real

The structural and electronic behavior of a crystal at its surface can differ significantly from that of the bulk. In the bulk, the crystal structure is determined by the minimization of internal chemical energy, with atoms adopting positions based on interactions with their nearest and next-nearest neighbors. However, at the surface, this balance is disrupted due to the abrupt reduction in the number of neighboring atoms. As a result, the atomic arrangement that was energetically favorable within the bulk may no longer be optimal at the surface. To compensate, atoms at the surface often rearrange themselves into new configurations—known as *surface reconstructions*—that lower the system’s total energy under the new boundary conditions.

1.3.2 Interfaces

Just like surfaces, interfaces play a critical role in the operation of semiconductor devices. In previous sections, we introduced heterostructures and superlattices, both of which inherently involve interfaces between different semiconductor materials. These interfaces are typically of very high quality, exhibiting minimal disruption in bonding—though defects such as dislocations can arise, particularly in strained-layer systems (which will be discussed later). Despite the high crystalline quality, there is often a small degree of interface roughness—on the order of one to two monolayers. This can be attributed to factors such as suboptimal growth conditions or slight timing inaccuracies in the switching between different semiconductor elements during epitaxial growth. While the crystal lattice and its periodic structure are preserved across the interface, some degree of atomic disorder in the chemical composition can be present within the interfacial planes. This microscopic disorder can significantly impact the behavior of electronic and optoelectronic devices. One of the most technologically significant interfaces is that between silicon and silicon dioxide (Si/SiO_2). The performance and scalability of nearly all modern electronic devices—including those that power the consumer electronics revolution—depend heavily on the exceptional quality of this interface. Interestingly, even though silicon and silicon dioxide differ substantially in both lattice constant and crystal structure, the interface is remarkably well-behaved. Typically, the transition region spans just a few monolayers and may include some degree of structural disorder or partial amorphization, leading to fluctuations in chemical composition that are important in determining device characteristics.

1.4 Defects in Semiconductors

In practical semiconductor materials, the presence of defects is inevitable. These imperfections may arise from thermodynamic factors or from unintentional contamination during the crystal growth process. Defects in crystalline semiconductors are generally classified into four categories:

1. *Point defects* — localized disruptions such as vacancies, interstitials, or substitutional atoms;
2. *Line defects* — also known as dislocations, which extend along one dimension;
3. *Planar defects* — including stacking faults or grain boundaries that span across planes;
4. *Volume defects* — such as precipitates or voids that affect larger regions of the material.

These defects can significantly degrade the electrical and optical properties of semiconductor devices and are generally undesirable. Therefore, minimizing their occurrence is a key goal in

semiconductor fabrication. In the following, we provide a brief introduction to the most relevant types of defects.

1.4.1 Point Defects

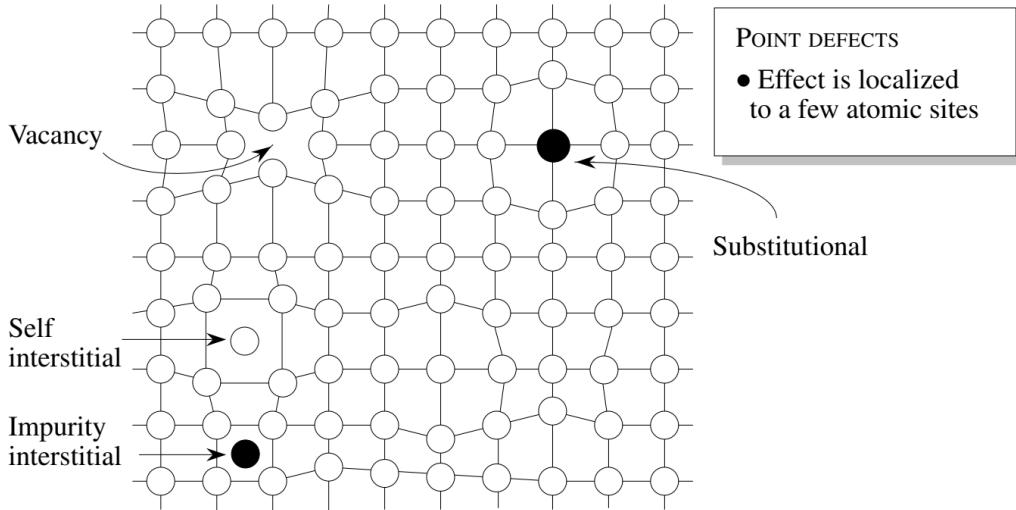


Figure 1.5. A schematic showing some important point defects in a crystal lattice.

Point defects are localized imperfections in a crystal lattice that disrupt the periodic structure only within one or a few unit cells. These types of defects are among the most fundamental and occur in all real crystalline materials. Their presence is governed by thermodynamic principles, and their equilibrium concentration can be estimated using the relation:

$$\frac{N_d}{N_{\text{tot}}} = k \exp\left(-\frac{E_d}{k_B T}\right)$$

where N_d is the defect (e.g., vacancy) concentration, N_{tot} is the total number of atomic sites in the crystal, E_d is the defect formation energy, k is a dimensionless constant typically ranging from 1 to 10 for semiconductors, k_B is Boltzmann's constant, and T is the absolute temperature during crystal growth. For most semiconductors, the formation energy of a vacancy is typically on the order of an electronvolt.

A particularly significant point defect in compound semiconductors, such as GaAs, is the *anti-site* defect. In this case, an atom from one sublattice—for example, a gallium (Ga) atom—occupies a site normally reserved for an arsenic (As) atom, resulting in a defect often denoted as Ga_{As} . These anti-site defects can act as deep traps or recombination centers, ultimately degrading the performance of electronic and optoelectronic devices.

1.4.2 Line defects: Dislocations

Unlike point defects, line defects—commonly referred to as dislocations—extend over many atomic sites and can be described by a line connecting the disrupted region. A typical example of a dislocation occurs when an extra half-plane of atoms is inserted or removed from the crystal, resulting in what is known as an *edge dislocation*. Another form arises from shear deformation or crystal slip, where atomic planes shift and bonds are broken and reformed along a slip line. Dislocations are particularly problematic in the growth of strained heterostructures, where lattice mismatch between different materials can induce mechanical stress. In optoelectronic devices,

the presence of dislocations can lead to non-radiative recombination and significant degradation in device efficiency or even complete failure. For this reason, minimizing and controlling dislocation density is a critical aspect of high-quality semiconductor fabrication.

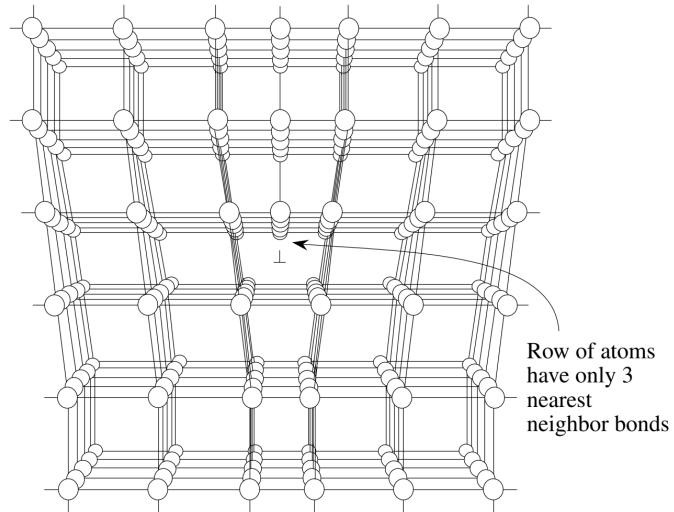


Figure 1.6. A schematic showing the presence of dislocation. The line defect is produced by adding an extra half plane of atoms. At the edge of the extra plane, the atoms have a missing bond.

1.5 Strained Heterostructures

During epitaxial growth, the deposited layer (or overlayer) may possess a lattice constant that differs from that of the underlying substrate. When this occurs, the resulting structure is referred to as *strained epitaxy*. This type of growth introduces strain in the crystal due to the lattice mismatch between the two materials. Strained epitaxy is a key area of research in modern crystal growth, as it enables the engineering of material properties through controlled strain, which can enhance the performance of electronic and optoelectronic devices.

1.5.1 Coherent and Incoherent Structures

Consider a scenario in which an epitaxial overlayer with lattice constant a_L is deposited on a substrate with lattice constant a_S . The lattice mismatch between the two materials introduces strain, which can be quantified as:

$$\varepsilon = \frac{a_S - a_L}{a_L}$$

To understand the implications of this mismatch, imagine depositing a single monolayer of the overlayer onto the substrate. If the overlayer retains its natural lattice constant a_L , then after every $1/\varepsilon$ bonds, a discontinuity arises—either a missing or an extra bond—due to the misalignment. Since the interface is two-dimensional, these discontinuities form rows, giving rise to dislocations. Dislocations are energetically unfavorable because they leave some atoms at the interface improperly bonded, increasing the system's total energy. An alternative scenario is one where the overlayer adjusts its in-plane lattice constant to match that of the substrate. In this case, all interfacial atoms remain bonded, but the overlayer is placed under strain, storing strain energy in the crystal. As the layer grows thicker, this strain energy increases. Whether the system remains coherently strained or begins to introduce dislocations depends on a competition between the elastic strain energy and the energy cost of forming dislocations. For small lattice mismatches (typically $\varepsilon < 0.1$), the overlayer can initially grow coherently, fully strained to the

substrate. However, as thickness increases, it becomes energetically favorable for the system to relieve strain by forming dislocations. This transition is characterized by a so-called *critical thickness*, d_c , which gives a rough estimate of the maximum thickness for coherent growth. A simplified expression for d_c is:

$$d_c \sim \frac{a_S}{2|\varepsilon|}$$

In practice, the exact point at which dislocations begin to form depends on a range of factors, including growth temperature, surface morphology, and the kinetics of dislocation motion. Nonetheless, Equation (1.14) provides a useful guideline for distinguishing between coherent and incoherent growth regimes for a given lattice mismatch.

1.6 Strain Tensor in a lattice mismatched epitaxy

To analyze how strain affects the electronic properties of semiconductors, one must first determine the strain tensor resulting from epitaxial growth. When an epitaxial layer is deposited on a substrate with a slightly different lattice constant, coherent strain can develop if the lattice mismatch is small and growth conditions are carefully controlled. In such cases, the epitaxial layer undergoes *pseudomorphic growth*, where the lattice constant of the layer in directions parallel to the interface is forced to match that of the substrate. As a result of this constraint, the perpendicular lattice constant adjusts according to the Poisson effect: if the layer is compressed in-plane, it expands out-of-plane, and if stretched in-plane, it contracts in the perpendicular direction. For epitaxial growth that proceeds in a layer-by-layer fashion, the strain in the crystal is typically biaxial within the substrate plane and uniaxial in the perpendicular direction. The in-plane (biaxial) strain ε_{\parallel} can be expressed in terms of the lattice constants of the substrate (a_S) and the epitaxial layer (a_L) as:

$$\varepsilon_{\parallel} = \frac{a_S - a_L}{a_L} = \frac{a_S}{a_L} - 1 = \varepsilon$$

This strain plays a fundamental role in modifying the band structure and other physical properties of the material.

The strain tensor ε in a crystal can be expressed in terms of the displacement gradient tensor \mathbf{u} , which describes how the atomic positions deviate from their equilibrium positions. The strain tensor is given by:

$$\varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$$

where i and j denote the Cartesian coordinates (1, 2, 3 corresponding to x , y , and z axes, respectively). The displacement gradient tensor \mathbf{u} is defined as:

$$u_i = x_i - x_i^0$$

where x_i is the current position of the atom and x_i^0 is its equilibrium position. The strain tensor captures how the crystal lattice deforms due to external forces or internal stresses, providing a complete description of the strain state in the material.

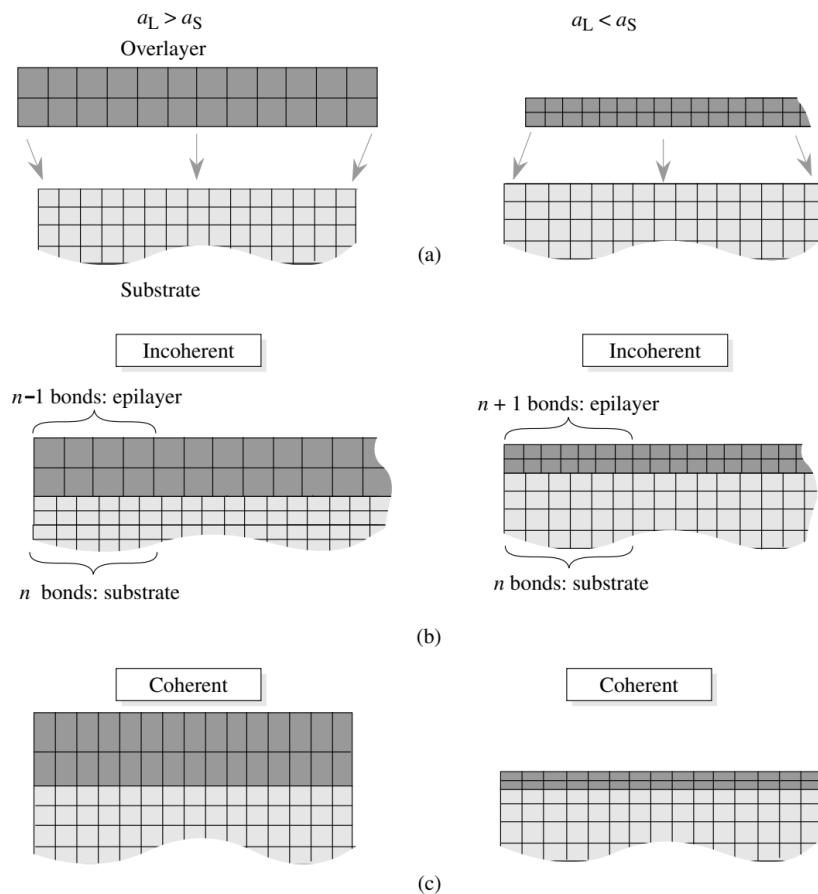


Figure 1.7: (a) Conceptual illustration of an overlayer with a lattice constant different from the substrate, placed without lattice distortion. (b) Dislocations form at the interface at positions where atomic bonding is disrupted due to lattice mismatch. (c) Coherent case where the overlayer is elastically strained to match the substrate, resulting in no dislocations at the interface.

Semiconductor Bandstructure

2.1 Introduction

The behavior of electrons in semiconductors is governed by the solutions to the Schrödinger equation tailored to the crystal environment. These solutions yield the electronic bandstructure, which characterizes the allowed energy states for electrons in the material. Determining the electronic spectrum is, however, a highly complex task. In solids, the presence of a vast number of atoms arranged closely together generates a complicated potential energy landscape for electrons. Furthermore, electron-electron interactions and dynamic lattice vibrations (phonons) introduce additional complexity through time-dependent variations in the potential. To make the problem tractable, the effects of atomic vibrations and electron-electron scattering are initially excluded and later incorporated using perturbation theory. These effects primarily contribute to the scattering of electrons between different quantum states. The analysis of bandstructure becomes significantly more manageable in crystalline solids. In such systems, the electrons move in a spatially periodic potential due to the regular arrangement of atoms. This periodicity leads to solutions that obey Bloch's theorem, which will be addressed in the following section. Realistic methods for computing the bandstructure of semiconductors generally fall into two broad categories:

1. Approaches that capture the full conduction and valence band profiles.
2. Approaches that focus specifically on the bandstructure near the band edges.

2.2 Block Theorem and Crystal Momentum

To analyze the electronic characteristics of a material, it is essential to determine the electron wavefunctions and their corresponding energies within a solid. In this context, we focus solely on crystalline materials. The behavior of electrons in a periodic solid is governed by the Schrödinger equation:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r})$$

Here, $U(\mathbf{r})$ represents the potential energy experienced by the electrons. Owing to the periodic structure of the crystal, this potential exhibits the same periodicity \mathbf{R} as the lattice:

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R})$$

In the case where the background potential is absent, the electron wavefunction within a volume V takes the form $\psi(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}}$, and the corresponding momentum and energy of the electron

are given by:

$$\mathbf{p} = \hbar\mathbf{k}, \quad E = \frac{\hbar^2 k^2}{2m}$$

This wavefunction is delocalized over the entire sample, implying that the probability density $\psi^*\psi$ is uniform throughout space.

When considering a periodic crystal, it is expected that the electron probability distribution remains identical in each unit cell, since all cells are structurally the same. If the potential were disordered, this uniformity would not hold. Given a lattice translation vector \mathbf{R} , we require:

$$|\psi(\mathbf{r})|^2 = |\psi(\mathbf{r} + \mathbf{R})|^2$$

This condition ensures that one cannot distinguish between different unit cells based on electronic probability density. It is important to note that while the probability is periodic, the wavefunction itself need not be. The appropriate form of the wavefunction in a periodic potential is described by Bloch's theorem. According to this theorem, the eigenfunctions of the Schrödinger equation in a periodic lattice are of the form:

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

where $u_{\mathbf{k}}(\mathbf{r})$ is a function that shares the periodicity of the crystal:

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

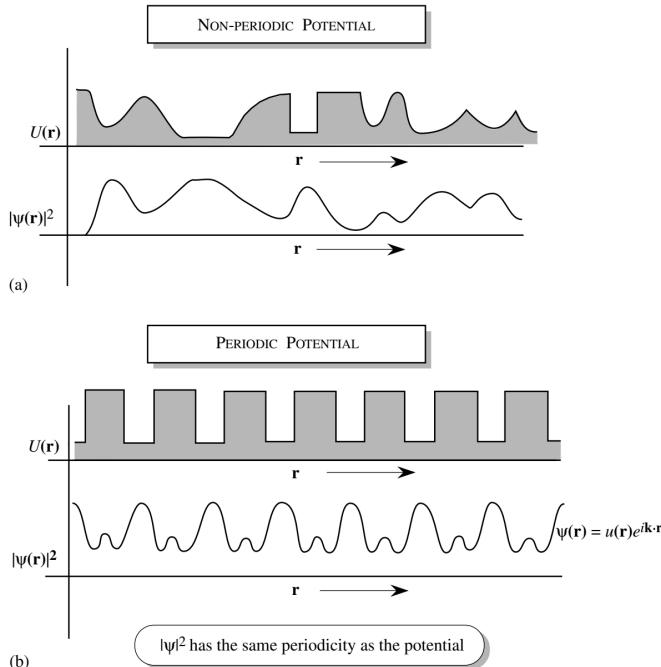


Figure 2.1. (a) Example of an electronic wavefunction in a disordered material. (b) In a periodic potential, $|\psi|^2$ shares the same periodicity, as required by Bloch's theorem.

This implies that the full wavefunction transforms under lattice translations as:

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

The vector \mathbf{k} , known as the wavevector or \mathbf{k} -vector, is fundamental in describing the electronic properties of crystalline solids.

2.2.1 Significance of the k-Vector

One of the key consequences of Bloch's theorem is that, in a perfectly periodic potential provided by the crystal lattice, electrons can propagate through the material without experiencing scattering. In such a scenario, the electron wavefunction, which resembles $\sim e^{i\mathbf{k}\cdot\mathbf{r}}$, describes an extended state that spans the entire crystal. To understand how electrons behave under external influences, we seek to derive an equation of motion that captures their response to external forces. Let \mathbf{F}_{ext} denote an external force acting on the electron, and \mathbf{F}_{int} represent the internal force arising from the atomic lattice. Newton's second law can then be expressed as:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}_{\text{ext}} + \mathbf{F}_{\text{int}}$$

However, this form is not particularly useful, as it involves the internal forces, which are typically complex to evaluate. To describe the electron's motion more effectively, we need an equation that depends solely on the external forces. We now outline a simplified derivation of such an expression, which also helps clarify the physical interpretation of the wavevector \mathbf{k} introduced in Bloch's theorem. Starting from the time-dependent Schrödinger equation, the general solution for electrons in a periodic potential can be written as:

$$\psi(\mathbf{r}, t) = u_{\mathbf{k}}(\mathbf{r})e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$$

where the energy of the electron is connected to the angular frequency ω by the relation:

$$E = \hbar\omega$$

The Bloch function thus corresponds to a plane wave that extends throughout the crystalline solid. To describe a spatially localized electron, we consider a wavepacket composed of Bloch states centered around a specific wavevector \mathbf{k} . The group velocity of this wavepacket is given by:

$$\mathbf{v}_g = \frac{d\omega}{d\mathbf{k}} = \frac{1}{\hbar} \frac{dE}{d\mathbf{k}} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k})$$

In the presence of an electric field \mathbf{F} , the energy gained by the electron over a short time interval δt is:

$$\delta E = -e\mathbf{F} \cdot \mathbf{v}_g \delta t$$

More generally, the change in energy can be expressed as:

$$\delta E = \frac{dE}{d\mathbf{k}} \cdot \delta\mathbf{k} = \hbar\mathbf{v}_g \cdot \delta\mathbf{k}$$

Equating the two expressions for δE , we obtain:

$$\delta\mathbf{k} = -\frac{e\mathbf{F}}{\hbar} \delta t$$

which leads to the equation of motion for the wavevector:

$$\hbar \frac{d\mathbf{k}}{dt} = -e\mathbf{F}$$

Since $-e\mathbf{F}$ is the force acting on the electron, this can be generalized to:

$$\hbar \frac{d\mathbf{k}}{dt} = \mathbf{F}_{\text{ext}}$$

This result mirrors Newton's second law:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}_{\text{ext}}$$

provided we identify $\hbar d\mathbf{k}/dt$ with the momentum of the electron inside the crystal. While $\hbar d\mathbf{k}/dt$ behaves as the momentum in response to external forces, it is not the true mechanical momentum since it incorporates the effects of the internal periodic potential. Instead, $\hbar\mathbf{k}$ is referred to as the *crystal momentum*.

2.3 Metals, Insulators, and Semiconductors

From atomic physics, it is known that bound electrons occupy discrete energy levels, separated by regions where no states exist. In solids, these discrete levels broaden into continuous energy bands, with forbidden gaps—bandgaps—separating them. Once the electronic band structure is known, a key question arises: which of the available states are occupied by electrons, and which remain empty? Two important cases emerge regarding electron occupation: in one scenario, an energy band is fully occupied at 0 K, while the next higher band is separated by an energy gap E_g and remains completely unoccupied. In the second case, the highest energy band that contains electrons is only partially filled. At this stage, a critical concept must be introduced. When an energy band is completely filled, the electrons within it cannot contribute to electrical conduction. This is a direct consequence of the Pauli exclusion principle: since electrons are fermions, they can only transition into unoccupied states. In a filled band, all such states are already occupied, leading to a cancellation of net motion—electrons moving in opposite directions balance each other. As a result, materials with fully filled bands and an energy gap above them exhibit, in principle, infinite resistivity and are classified as insulators or semiconductors. In contrast, if the highest occupied band is only partially filled, available empty states exist within the same band. This allows electrons to accelerate under an external field, enabling electrical conduction. Such materials exhibit low resistivity and are termed metals.

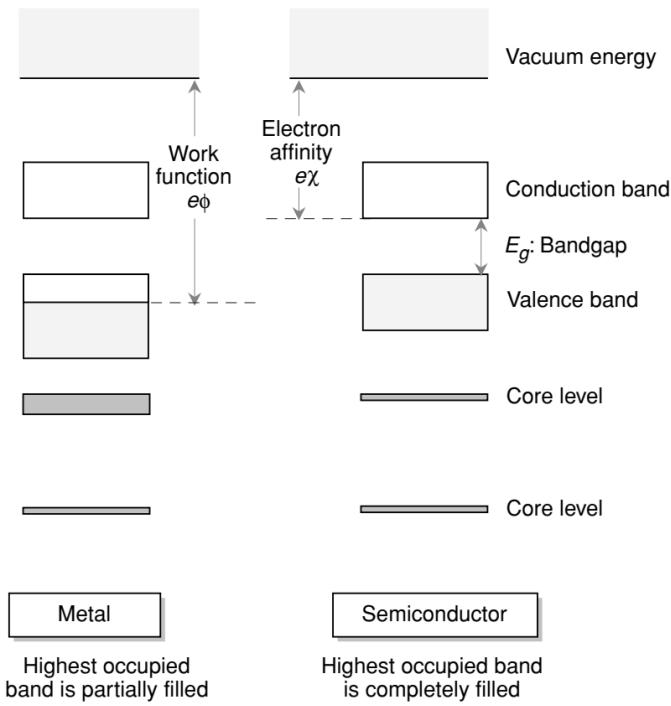


Figure 2.2. Band occupation at 0 K for a metal and a semiconductor. In metals, the top band is partially filled. In semiconductors, the valence band is full and separated from the empty conduction band by the bandgap E_g . Work function and electron affinity are indicated.

The band that is fully occupied by electrons at absolute zero in semiconductors is known as the *valence band*, whereas the higher energy band that remains empty at 0 K is referred to as the *conduction band*. In metals, the energy difference between the vacuum level and the highest occupied electronic state is termed the *work function*. For semiconductors, the energy separating the vacuum level from the bottom of the conduction band is called the *electron affinity*. Metals are

characterized by extremely high electrical conductivity, resulting from the large number of electrons available for conduction. However, this abundance of charge carriers also makes it difficult to significantly modify their conductivity. Semiconductors, in contrast, exhibit zero conductivity at 0 K and relatively low conductivity at finite temperatures. Importantly, their conductivity can be altered dramatically—by several orders of magnitude—which is the foundation of their utility in active electronic components. As previously discussed, semiconductors are defined as materials in which the valence band is completely filled and the conduction band completely empty at 0 K. When the temperature is raised above absolute zero, some electrons gain enough thermal energy to transition from the valence band into the conduction band. This results in the creation of empty states (holes) in the valence band and free electrons in the conduction band. When electrons are thermally excited into the conduction band, the valence band is left with some unoccupied states. Consider the situation illustrated in the scheme below, where an electron with wavevector \mathbf{k}_e is missing from the valence band. In a completely filled valence band, the total sum over all electron wavevectors is zero:

$$\sum_i \mathbf{k}_i = \sum_{\mathbf{k}_e} \mathbf{k}_e + \sum_{-\mathbf{k}_e} (-\mathbf{k}_e) = 0$$

This simply reflects the fact that for every occupied positive wavevector state, there exists a corresponding occupied state with a negative wavevector.

Now, when an electron is missing from a particular state \mathbf{k}_e , the net wavevector of the system becomes:

$$\mathbf{k}_h = -\mathbf{k}_e$$

This unoccupied state is referred to as a *hole*, and the wavevector $-\mathbf{k}_e$ is attributed to it. Although the missing electron originated from the state \mathbf{k}_e , the hole is described as having wavevector $\mathbf{k}_h = -\mathbf{k}_e$, as illustrated. The hole's position corresponds to where the electron would have been.

It is important to emphasize that a hole represents the absence of an electron in the valence band. When the valence band is completely filled, it cannot contribute to electrical conduction. However, once an electron is removed, current can flow. Under the application of an electric field, electrons drift in the direction opposite to the field, which effectively causes the unoccupied (hole) state to move in the direction of the field. Thus, the hole behaves like a positively charged particle.

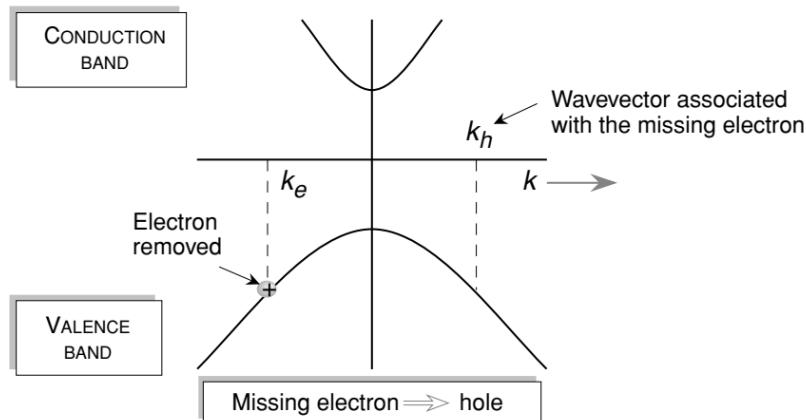


Figure 2.3. Illustration of the wavevector of the missing electron \mathbf{k}_e and the corresponding hole wavevector $-\mathbf{k}_e$.

The response of the hole to electric and magnetic fields \mathbf{F} and \mathbf{B} is described by the following

equation of motion:

$$\frac{d\mathbf{k}_h}{dt} = e [\mathbf{F} + \mathbf{v}_h \times \mathbf{B}]$$

where $\hbar\mathbf{k}_h$ is the crystal momentum of the hole and \mathbf{v}_h is its velocity.

2.4 Tight-Binding Method

Before delving into the properties of various semiconductors, it is helpful to first examine the atomic structure of the constituent elements.

Group IV Semiconductors

- **C:** 1s² 2s² 2p²
- **Si:** 1s² 2s² 2p⁶ 3s² 3p²
- **Ge:** 1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p²

Group III–V Semiconductors

- **Ga:** 1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p¹
- **As:** 1s² 2s² 2p⁶ 3s² 3p⁶ 3d¹⁰ 4s² 4p³

A key observation from these configurations is that the valence electrons in all relevant semiconductor elements occupy either *s*- or *p*-type orbitals. While this is true for isolated atoms, it also holds in crystalline solids: electrons in the valence and conduction bands retain this *s*- or *p*-like orbital character, even though they become delocalized as Bloch states. This feature plays a fundamental role in determining the optical and transport behavior of semiconductors.

Bandstructure and the Tight Binding Method

We now turn to techniques for calculating the bandstructure of materials. As previously noted, there are both comprehensive methods and those valid only near specific energy ranges. We begin with the *tight binding method* (TBM), an empirical approach that uses experimental data to model the bandstructure.

In TBM, atomic orbitals are used as the basis for constructing Bloch functions. The periodic part of the Bloch wavefunction is written as a linear combination of atomic orbitals centered on the crystal lattice sites. If $\phi(\mathbf{r} - \mathbf{R})$ represents an atomic orbital centered at position \mathbf{R} , the Bloch wavefunction can be expressed as:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}_n} \phi_n(\mathbf{r} - \mathbf{R}) \exp(i\mathbf{k} \cdot \mathbf{R}_n)$$

The periodic part of the Block Function is expanded in terms of the atomic-like orbitals of the atoms of the unit cell (index n in the summation).

As mentioned, the outermost electrons in the semiconductor elements are associated with *s*- and *p*-type orbitals. While core electrons do not significantly contribute to band formation, the valence orbitals do. When atoms are brought together to form a solid, these discrete atomic states broaden into energy bands. Although the atomic orbitals are no longer exact eigenstates in the crystal, they serve as an excellent approximate basis for describing the electronic states. This forms the central idea of the tight binding model. We start from the known atomic solutions, which satisfy:

$$H_{\text{at}} \psi_n = E_n \psi_n$$

The goal is to construct approximate Bloch-like states from these atomic solutions:

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

However, these states do not account for the interactions between atoms in the crystal. To include this effect, we consider the total crystal Hamiltonian as:

$$H_{\text{cryst}} = H_{\text{at}} + \Delta U(\mathbf{r})$$

Here, $\Delta U(\mathbf{r})$ represents the additional potential arising from neighboring atoms—this is what perturbs the atomic states and causes band formation.

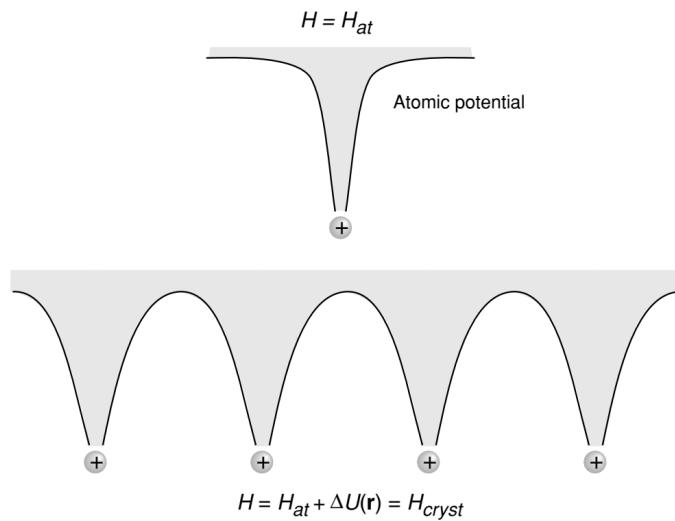


Figure 2.4. The effect of neighboring atoms is to perturb the atomic states, leading to the formation of bands. The perturbation is represented by $\Delta U(\mathbf{r})$.

The new wavefunctions must respect Bloch's theorem and can be written in the general form:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi(\mathbf{r} - \mathbf{R}) \quad (2.1)$$

In this context, the function $\phi(\mathbf{r})$ is not itself an atomic orbital but is instead constructed as a linear combination of atomic eigenfunctions. Specifically, we can express $\phi(\mathbf{r})$ in terms of a finite set of atomic orbitals $\psi_n(\mathbf{r})$ as:

$$\phi(\mathbf{r}) = \sum_{n=1}^N b_n \psi_n(\mathbf{r}) \quad (2.2)$$

Here, N is the number of atomic orbitals included in the basis, and b_n are coefficients to be determined. In the tight binding method, only a limited number N of orbitals are retained for computational tractability.

To determine the coefficients b_n , we substitute the equation above into the time-independent Schrödinger equation and derive a system of N coupled equations by projecting onto the atomic basis set. The Schrödinger equation for the crystal takes the form:

$$H\Psi_{\mathbf{k}} = E(\mathbf{k})\Psi_{\mathbf{k}} \quad (2.3)$$

Substituting Eq.(2.1) and Eq.(2.2) into Eq.(2.3), multiplying from the left by $\psi_m^*(\mathbf{r})$, and integrating over all space gives:

$$\int d^3r \psi_m^*(\mathbf{r}) \left\{ [H_{\text{at}} + \Delta U(\mathbf{r})] \sum_{\mathbf{R},n} b_n e^{i\mathbf{k}\cdot\mathbf{R}} \psi_n(\mathbf{r} - \mathbf{R}) - E(\mathbf{k}) \sum_{\mathbf{R},n} b_n e^{i\mathbf{k}\cdot\mathbf{R}} \psi_n(\mathbf{r} - \mathbf{R}) \right\} = 0 \quad (2.4)$$

Since the atomic orbitals are orthonormal when centered on the same atom, we have:

$$\int d^3r \psi_m^*(\mathbf{r}) \psi_n(\mathbf{r}) = \delta_{mn} \quad (2.5)$$

However, this orthogonality does not generally hold when the orbitals are centered at different lattice sites. That is:

$$\int d^3r \psi_m^*(\mathbf{r}) \psi_n(\mathbf{r} - \mathbf{R}) \neq \delta_{mn} \quad \text{for } \mathbf{R} \neq 0 \quad (2.6)$$

This non-orthogonality must be accounted for in the tight binding formalism and leads to overlap integrals that influence the coupling between neighboring atomic orbitals.

In the summation over lattice vectors, we separate the terms with $\mathbf{R} = 0$ and $\mathbf{R} \neq 0$ to get,

$$\begin{aligned} (E(\mathbf{k}) - E_m) b_m &= \\ &- (E(\mathbf{k}) - E_m) \sum_{n=1}^N \left(\sum_{\mathbf{R} \neq 0} \int \psi_m^*(\mathbf{r}) \psi_n(\mathbf{r} - \mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}} d^3r \right) b_n \\ &+ \sum_{n=1}^N \left(\int \psi_m^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_n(\mathbf{r}) d^3r \right) b_n \\ &+ \sum_{n=1}^N \left(\sum_{\mathbf{R} \neq 0} \int \psi_m^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_n(\mathbf{r} - \mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}} d^3r \right) b_n \end{aligned}$$

It is important to recognize the use of the identity:

$$\int d^3r \psi_m^*(\mathbf{r}) H_{\text{at}} \psi_n(\mathbf{r}) = E_m \delta_{mn} \quad (2.7)$$

where E_m is the energy eigenvalue associated with the atomic orbital ψ_m . Note that these energies E_m may not exactly correspond to the isolated atomic energy levels, since the presence of neighboring atoms in the crystal modifies them. As such, in the tight binding method, these energies are often treated as fitting parameters to match experimental data or more accurate calculations.

An additional approximation typically made in the tight binding framework is:

$$\int d^3r \psi_m^*(\mathbf{r}) \psi_n(\mathbf{r} - \mathbf{R}) e^{i\mathbf{k}\cdot\mathbf{R}} \approx 0 \quad (2.8)$$

This implies that the spatial overlap between atomic orbitals centered on different atoms is negligible, which is valid when the atomic orbitals are strongly localized.

Another important concept is the *on-site matrix element*, defined as:

$$\int d^3r \psi_m^*(\mathbf{r}) H \psi_n(\mathbf{r}) = E_m + \int d^3r \psi_m^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_m(\mathbf{r}) \quad (2.9)$$

For many forms of the potential $\Delta U(\mathbf{r})$, the integral:

$$\int d^3r \psi_m^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_n(\mathbf{r}) \quad (2.10)$$

vanishes when $m \neq n$, which further simplifies the matrix elements.

To obtain the eigenvalues $E(\mathbf{k})$ and corresponding eigenfunctions (with coefficients b_n), we solve a system of N coupled linear equations. These form an $N \times N$ *secular equation*, from which the energy dispersion $E(\mathbf{k})$, or the bandstructure, can be determined. For each wavevector \mathbf{k} , the solution yields N energy levels, corresponding to the bands derived from the N atomic orbitals used in the basis.

To gain intuition about this method, we next examine a simple case involving only *s*-orbitals—a so-called *s-band model*.

2.4.1 Tight Binding Model for s-Bands

In the case of semiconductors, each unit cell typically contains two atoms in the basis. For each atom, it is necessary to include at least the outer shell orbitals: one *s*-orbital and three *p*-orbitals (p_x , p_y , and p_z). As a result, the tight binding secular equation becomes significantly more complex due to the increased number of basis functions.

To gain some intuitive understanding, we now consider a simplified model: a crystal with a single atom per unit cell and only one orbital per atom—the *s*-orbital. In this scenario, there is only one atomic level, and all tight binding coefficients $\{b_n\}$ are zero except for the *s*-level, where $b_s = 1$. This simplification reduces the problem to a single equation.

We obtain:

$$\begin{aligned} E(k) - E_s = & - (E(k) - E_s) \sum_{\mathbf{R} \neq 0} \int d^3r \psi_s^*(\mathbf{r}) \psi_s(\mathbf{r} - \mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}} \\ & + \int d^3r \psi_s^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_s(\mathbf{r}) \\ & + \sum_{\mathbf{R} \neq 0} \int d^3r \psi_s^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_s(\mathbf{r} - \mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}} \end{aligned}$$

We define the following integrals to simplify notation:

$$\alpha(\mathbf{R}) = \int d^3r \psi_s^*(\mathbf{r}) \psi_s(\mathbf{r} - \mathbf{R}) \quad (2.11)$$

$$\beta_s = - \int d^3r \psi_s^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_s(\mathbf{r}) \quad (2.12)$$

HOPPING INTEGRAL

$$\gamma(\mathbf{R}) = - \int d^3r \psi_s^*(\mathbf{r}) \Delta U(\mathbf{r}) \psi_s(\mathbf{r} - \mathbf{R}) \quad (2.13)$$

As discussed earlier, we typically neglect the overlap integral $\alpha(\mathbf{R})$, i.e., we set $\alpha(\mathbf{R}) = 0$, assuming strong localization of atomic orbitals. The energy dispersion relation simplifies to:

$$E(\mathbf{k}) = E_s - \beta - \sum_{\mathbf{R}} \gamma(\mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}} \quad (2.14)$$

The off-site matrix elements $\gamma(\mathbf{R})$ decay rapidly with increasing distance \mathbf{R} , so we consider only the nearest-neighbor interactions. Additionally, due to lattice symmetry, $\gamma(-\mathbf{R}) = \gamma(\mathbf{R})$.

Let us now apply this to a face-centered cubic (fcc) lattice. The twelve nearest neighbors of a lattice site in an fcc crystal are located at:

$$\frac{a}{2}(\pm 1, \pm 1, 0), \quad \frac{a}{2}(\pm 1, 0, \pm 1), \quad \frac{a}{2}(0, \pm 1, \pm 1)$$

Taking into account all twelve neighbors, the energy becomes:

$$E(\mathbf{k}) = E_s - \beta - \gamma \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}$$

$$= E_s - \beta - [4\gamma] \left[\cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) + \cos\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_z a}{2}\right) + \cos\left(\frac{k_z a}{2}\right) \cos\left(\frac{k_x a}{2}\right) \right]$$

This final expression captures **the energy dispersion relation** for an *s*-band in an fcc lattice, considering only nearest-neighbor interactions.

We now analyze the energy band derived above within the context of the first Brillouin zone of the face-centered cubic (fcc) lattice. From the geometry of the Brillouin zone, we observe that there are six equivalent *X*-points and eight equivalent *L*-points, which are critical in determining the behavior of the electronic bandstructure.

By evaluating the energy expression from the last relation obtained along various high-symmetry directions in \mathbf{k} -space, we can obtain the band dispersion throughout the Brillouin zone. The resulting energy bands reflect the symmetry of the lattice and provide insight into the electronic properties of the material.

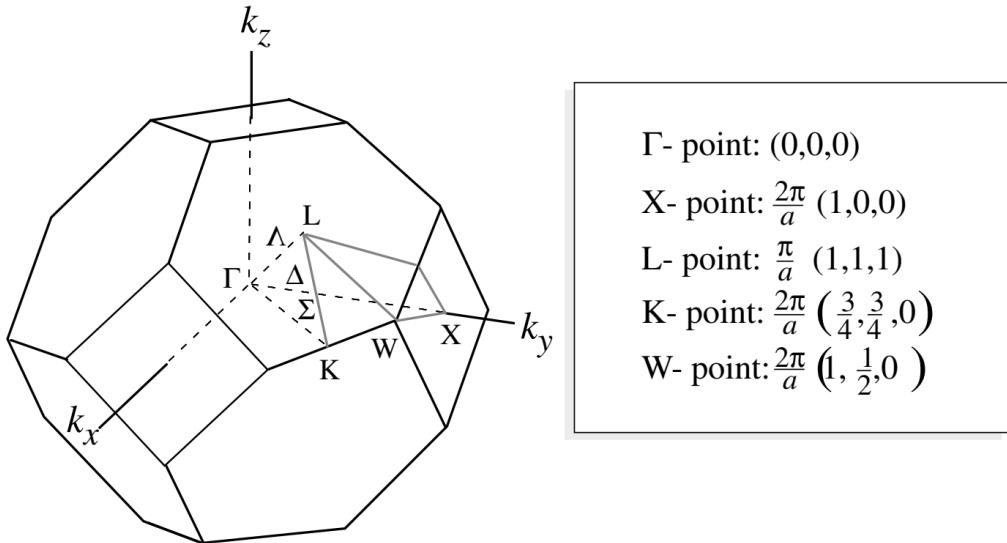


Figure 2.5. Brillouin zones of the face-centered cubic (fcc) lattice. The high-symmetry points Γ , X , and L are indicated, along with their corresponding coordinates in reciprocal space.

To further understand the nature of the bandstructure, we examine the dispersion relation $E(\mathbf{k})$ along key high-symmetry directions in the first Brillouin zone of the fcc lattice. These paths correspond to the directions $\Gamma \rightarrow X$, $\Gamma \rightarrow L$, and $\Gamma \rightarrow K$, and the corresponding energy expressions can be obtained by substituting specific components of \mathbf{k} into Eq. (2.35).

- **Along ΓX :** Set $k_x = \frac{2\pi\alpha}{a}$, $k_y = k_z = 0$, with $0 \leq \alpha \leq 1$. The energy becomes:

$$E(\mathbf{k}) = E_s - \beta - 4\gamma(1 + 2 \cos \pi\alpha)$$

- **Along ΓL :** Set $k_x = k_y = k_z = \frac{2\pi\alpha}{a}$, with $0 \leq \alpha \leq \frac{1}{2}$. The energy expression is:

$$E(\mathbf{k}) = E_s - \beta - 12\gamma \cos^2(\pi\alpha)$$

- **Along ΓK :** Set $k_x = k_y = \frac{2\pi\alpha}{a}$, $k_z = 0$, with $0 \leq \alpha \leq \frac{3}{4}$. The energy becomes:

$$E(\mathbf{k}) = E_s - \beta - 4\gamma \left(\cos^2 \pi\alpha + 2 \cos \left(\frac{\pi\alpha}{2} \right) \right)$$

The bandstructure—i.e., the relation between energy E and wavevector \mathbf{k} —is typically plotted along these high-symmetry paths, as shown in Fig.2.5. In the illustrative case where $\gamma = 1.0 \text{ eV}$ and $E_s + \beta = 0$, the resulting bandwidth is found to be:

$$\Delta E = 16\gamma$$

This shows that greater overlap between neighboring atomic orbitals (i.e., larger γ) leads to wider energy bands. It is particularly useful to analyze the bandstructure near the Γ point, where the wavevector \mathbf{k} is small. By choosing $k_x = k_y = k_z = k/\sqrt{3}$ such that $ka \ll 1$, the cosine terms can be expanded, yielding:

$$E(\mathbf{k}) = E_s - \beta - 12\gamma + \gamma a^2 k^2 \quad (2.15)$$

Comparing this expression with the free electron dispersion relation:

$$E(p) = E_0 + \frac{p^2}{2m}$$

we can rewrite the tight binding result in a similar form:

$$E(\mathbf{k}) = E_s - \beta - 12\gamma + \frac{\hbar^2 k^2}{2m^*}$$

This allows us to define an effective mass m^* for the electrons near the bottom of the band:

$$m^* = \frac{\hbar^2}{2\gamma a^2} \quad (2.16)$$

This effective mass characterizes the curvature of the band near Γ and plays a central role in transport phenomena in semiconductors.

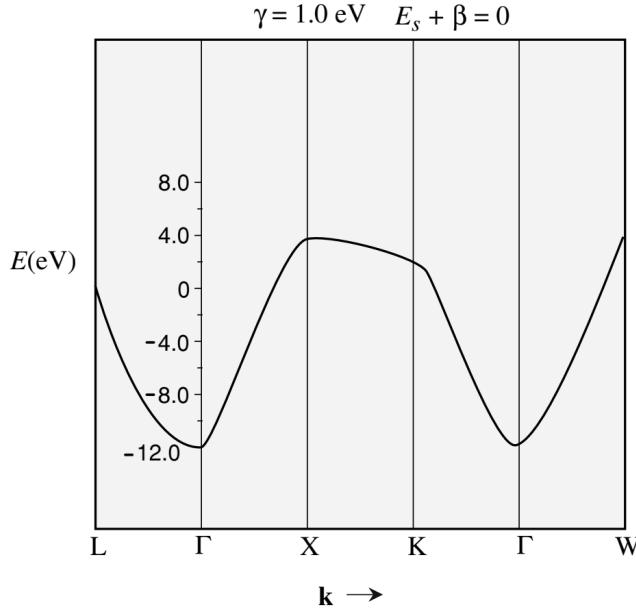


Figure 2.6. Bandstructure of an *s*-band in a face-centered cubic (fcc) lattice. The energy dispersion is plotted along high-symmetry directions in the Brillouin zone, illustrating the relationship between energy E and wavevector \mathbf{k} . The effective mass m^* is defined near the Γ point.

2.4.2 Bandstructure of Semiconductors

We now extend the discussion of TBM to semiconductor materials. As previously mentioned, the outermost electrons in semiconductors are typically described using atomic orbitals of type *s*, p_x , p_y , and p_z . The accuracy of the method can be improved by including additional atomic states, though the minimal useful basis already captures much of the essential physics.

In semiconductor crystals, there are typically two atoms per unit cell. Therefore, to describe the central-cell portion of the Bloch functions, we require a total of eight atomic orbitals—four per atom. We construct the Bloch-like wavefunction as:

$$\Psi(\mathbf{k}, \mathbf{r}) = \sum_{R_i} \sum_{j=1}^2 \sum_{m=1}^4 C_{mj}(\mathbf{k}) \phi_{mj}(\mathbf{r} - \mathbf{r}_j - \mathbf{R}_i) e^{i\mathbf{k}\cdot\mathbf{R}_i} \quad (2.17)$$

Here, \mathbf{R}_i runs over the lattice vectors, j indexes the atoms within each unit cell, m refers to the atomic orbital type (*s*, p_x , p_y , p_z), and ϕ_{mj} are the atomic-like basis functions centered at position \mathbf{r}_j in the unit cell. The coefficients $C_{mj}(\mathbf{k})$ are to be determined by solving the Schrödinger equation.

As done in the simpler *s*-band case, we now formulate the Schrödinger equation as a matrix eigenvalue problem in the form of a secular determinant:

$$|\langle \phi_{mj} | H - E(\mathbf{k}) | \Psi(\mathbf{k}) \rangle| = 0 \quad (2.18)$$

Here, H is the crystal Hamiltonian, and the determinant must vanish for nontrivial solutions to exist.

In the tight binding method (TBM) applied to zinc-blende crystals using an sp^3 basis, the wavefunction is constructed from eight atomic orbitals: one *s*-orbital and three *p*-orbitals (p_x, p_y, p_z)

for each of the two atoms in the Wigner–Seitz cell. This results in a total of eight basis functions per unit cell.

This formulation assumes that the bandstructure exhibits spin degeneracy—that is, each electronic state is doubly degenerate due to the presence of both spin-up and spin-down electrons. If one wishes to include the effects of spin-orbit coupling or any interaction that breaks spin degeneracy, the basis must be expanded accordingly.

In the tight binding approximation, the top of the valence band is predicted to be three-fold degenerate, corresponding to the degeneracy of the p_x , p_y , and p_z orbitals. When spin is taken into account, this degeneracy becomes six-fold, since each orbital state can have both spin-up and spin-down components.

However, experimental observations and more accurate theoretical models show that, in most semiconductors, the top of the valence band is actually two-fold degenerate (four-fold with spin). Additionally, there exists another band—also two-fold degenerate—that lies slightly below the valence band maximum. These features cannot be captured by a purely non-relativistic tight binding model. To accurately reproduce this band splitting, relativistic effects must be incorporated. The interaction responsible for this splitting is known as *spin-orbit coupling*, which lifts part of the degeneracy by coupling the electron's spin and orbital angular momentum.

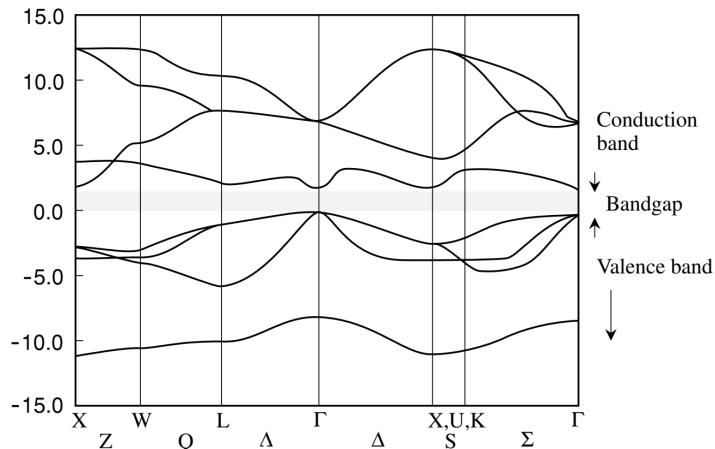


Figure 2.7. Band structure of GaAs obtained using the tight-binding model, excluding spin-orbit interaction. As shown, this approach alone does not provide a precise description of the valence band edge unless spin-orbit effects are taken into account.

2.5 Spin-Orbit Coupling

In nearly all semiconductors, it is observed that the states at the top of the valence band are predominantly derived from p -type atomic orbitals. Consequently, any bandstructure calculation that does not include spin effects tends to produce an inaccurate description of the valence band. The spin degree of freedom allows the electron to interact with the magnetic field generated by its own orbital motion.

An electron occupying a p -orbital possesses a non-zero orbital angular momentum \hbar , which leads to a significant coupling between its spin and orbital motion. Since the valence band maximum

is mainly composed of such p -like states, spin-orbit coupling has a pronounced influence in this energy region.

While the spin-orbit interaction can be calculated fairly accurately for isolated atoms, performing such calculations for crystalline solids is more complex. Therefore, in practical applications, the spin-orbit interaction is often incorporated through an effective model that uses a fitting parameter λ , adjusted to match experimental results. In many materials, the magnitude of this interaction is relatively small, and its effects can be treated perturbatively.

When including the spin-orbit interaction in the tight binding Hamiltonian, the total Hamiltonian takes the form:

$$H = H_{\text{tb}} + H_{\text{so}} \quad (2.19)$$

where H_{tb} is the tight binding Hamiltonian, and H_{so} represents the spin-orbit coupling term. This additional term can induce coupling between states of different spin. A commonly used model for this interaction is:

$$H_{\text{so}} = \lambda \mathbf{L} \cdot \mathbf{S}$$

Here, \mathbf{L} is the orbital angular momentum operator, \mathbf{S} is the spin angular momentum operator, and λ is a material-dependent constant representing the strength of the interaction.

The total angular momentum operator \mathbf{J} is defined as:

$$\mathbf{J} = (\mathbf{L} + \mathbf{S})^2 = \mathbf{L}^2 + \mathbf{S}^2 + 2\mathbf{L} \cdot \mathbf{S} \quad (2.20)$$

Using this, we can derive the expectation value of $\mathbf{L} \cdot \mathbf{S}$ as:

$$\langle \mathbf{L} \cdot \mathbf{S} \rangle = \frac{1}{2} [J^2 - L^2 - S^2]$$

which leads to:

$$\langle \mathbf{L} \cdot \mathbf{S} \rangle = \frac{\hbar^2}{2} [j(j+1) - l(l+1) - s(s+1)] \quad (2.21)$$

Here, j , l , and s are the quantum numbers corresponding to the total, orbital, and spin angular momentum, respectively. This relation provides a straightforward way to estimate the spin-orbit interaction energy—provided the states are characterized by well-defined angular momentum quantum numbers.

However, it is important to note that many states in solids, such as a p^\uparrow state, are not pure eigenstates of \mathbf{J} , but rather are mixed combinations of such states. In these cases, a more detailed analysis is required to fully account for the spin-orbit coupling effects.

The spin-orbit Hamiltonian can be expressed in terms of the quantum numbers associated with total, orbital, and spin angular momentum as:

$$H_{\text{so}} = \frac{\lambda \hbar^2}{2} [j(j+1) - l(l+1) - s(s+1)] \quad (2.22)$$

For electrons in p -type orbitals, we have $l = 1$ and $s = \frac{1}{2}$. The value of j depends on the specific total angular momentum eigenstate, and is determined by the decomposition of the mixed states into pure angular momentum components.

Due to the orthogonality of the total angular momentum eigenstates, many cross terms in matrix element calculations vanish, simplifying the application of the spin-orbit interaction to the bandstructure problem.

The influence of spin-orbit coupling is more clearly illustrated when the states are described in the total angular momentum basis, rather than in the p_x , p_y , and p_z orbital basis. In this representation, the p -orbitals (including spin) correspond to six states of the form $|j, m\rangle$, where

j is the total angular momentum quantum number and m is its projection.

These six states are:

$$|3/2, +3/2\rangle, \quad |3/2, -3/2\rangle, \quad |3/2, +1/2\rangle, \quad |3/2, -1/2\rangle, \quad |1/2, +1/2\rangle, \quad |1/2, -1/2\rangle$$

As follows from the spin-orbit Hamiltonian, this interaction splits the six degenerate p -states into two distinct sets: a quartet with $j = 3/2$, and a doublet with $j = 1/2$. The energy separation between these two groups is denoted by Δ_{so} , known as the spin-orbit splitting.

At the valence band edge of semiconductors, this results in a two-fold degenerate state (four-fold including spin) at the zone center, and a lower-energy two-fold degenerate split-off band. The upper set of degenerate states possesses different effective masses, which gives rise to the so-called *light hole* (LH) and *heavy hole* (HH) bands.

This spin-orbit-induced splitting is a key feature in the accurate description of valence band structure in semiconductors.

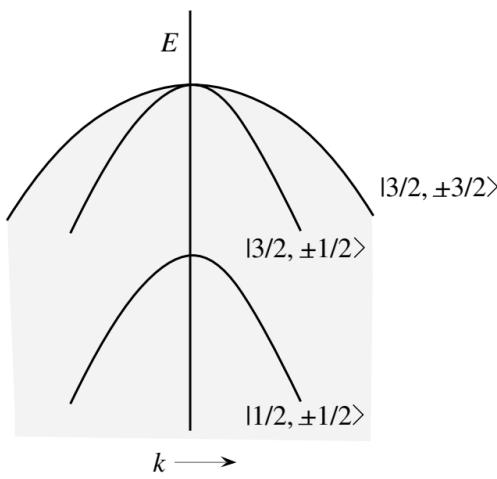


Figure 2.8. The general form of the valence bandstructure after incorporating spin-orbit coupling. At the Brillouin zone center ($\mathbf{k} = 0$), the electronic states exhibit well-defined angular momentum character.

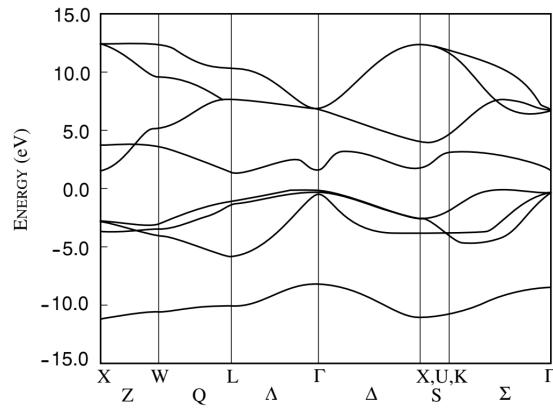


Figure 2.9. Tight-binding bandstructure of GaAs computed with spin-orbit coupling included. The degeneracy between the light-hole, heavy-hole, and split-off bands is lifted as a result of the spin-orbit interaction.

2.5.1 Symmetry of Bandedge States

In direct bandgap semiconductors, the conduction band minimum occurs at the Γ -point. The periodic part of the Bloch wavefunction at this point is spherically symmetric and primarily composed of s -type atomic orbitals. As one moves away from the band edge, the eigenstates acquire increasing contributions from p -type orbitals. However, the dominant s -character near the conduction band edge gives rise to important optical transition selection rules, which will be discussed in more detail later.

In contrast, indirect bandgap semiconductors such as silicon and germanium exhibit conduction band minima away from the Γ -point. In silicon, the minimum lies near the X -point and is six-fold degenerate, while in germanium it occurs at the L -point. The electronic states at these minima are highly anisotropic and involve complex combinations of s - and p -type orbitals (p_x , p_y , and p_z) in their wavefunction composition.

Despite the differences in conduction band behavior, the valence band edge structure in most semiconductors is qualitatively similar. The periodic component of the wavefunction at the

valence band maximum is primarily *p*-type, which enhances the influence of spin-orbit coupling in this region.

In the absence of spin-orbit interaction, the top of the valence band is three-fold degenerate (six-fold when spin degeneracy is included). When spin-orbit coupling is taken into account, this degeneracy is lifted, resulting in a four-fold degenerate manifold at the top of the valence band and a lower two-fold degenerate split-off band. The upper four-fold degenerate states comprise the heavy hole (HH) and light hole (LH) bands, while the lower-lying two-fold degenerate states form the split-off band.

Heavy hole states

$$\boxed{\Phi_{3/2,+3/2} = -\frac{1}{\sqrt{2}} (|p_x\rangle + i|p_y\rangle) |\uparrow\rangle} \quad (2.23)$$

$$\boxed{\Phi_{3/2,-3/2} = \frac{1}{\sqrt{2}} (|p_x\rangle - i|p_y\rangle) |\downarrow\rangle} \quad (2.24)$$

Light hole states

$$\boxed{\Phi_{3/2,+1/2} = -\frac{1}{\sqrt{6}} [(|p_x\rangle + i|p_y\rangle) |\downarrow\rangle - 2|p_z\rangle |\uparrow\rangle]} \quad (2.25)$$

$$\boxed{\Phi_{3/2,-1/2} = \frac{1}{\sqrt{6}} [(|p_x\rangle - i|p_y\rangle) |\uparrow\rangle + 2|p_z\rangle |\downarrow\rangle]} \quad (2.26)$$

Split-off hole states

$$\boxed{\Phi_{1/2,+1/2} = -\frac{1}{\sqrt{3}} [(|p_x\rangle + i|p_y\rangle) |\downarrow\rangle + |p_z\rangle |\uparrow\rangle]} \quad (2.27)$$

$$\boxed{\Phi_{1/2,-1/2} = \frac{1}{\sqrt{3}} [(|p_x\rangle - i|p_y\rangle) |\uparrow\rangle + |p_z\rangle |\downarrow\rangle]} \quad (2.28)$$

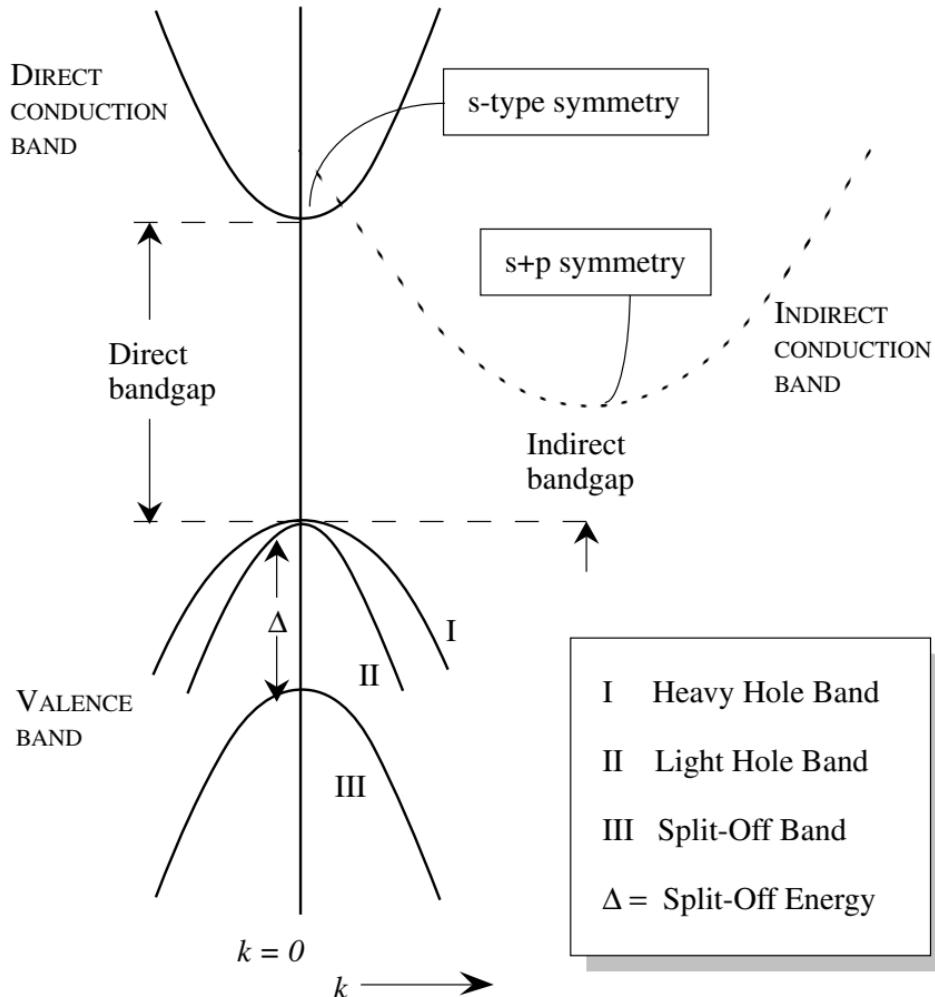


Figure 2.10. Schematic of the valence band and conduction bands in direct and indirect bandgap semiconductors. Solid and dashed lines represent the conduction bands of direct and indirect materials, respectively.

2.6 Selected Bandstructures

2.6.1 Silicon

Silicon forms the foundational material of the modern electronics industry. A key feature of silicon is that it possesses an indirect bandgap, which significantly limits its usefulness in optical applications—particularly for light-emitting devices.

The conduction band minimum in silicon does not occur at the Γ -point, but rather near the X -point in the Brillouin zone, approximately at $(2\pi/a)(0.85, 0, 0)$. There are six equivalent X -points in the Brillouin zone, leading to six degenerate conduction band valleys. As a result, the conduction band edge in silicon exhibits a multi-valley structure.

Near the conduction band edge, the periodic part of the Bloch wavefunctions exhibits a strong mixture of s - and p -type character. Specifically, along the longitudinal direction (aligned with the valley axis, e.g., the x -axis for the [100] valley), the states have contributions from s - and p_x -type orbitals. In the transverse directions (y and z), the wavefunction is composed primarily of p_y - and p_z -like components.

Close to the band edge, the energy dispersion relation can be approximated by an ellipsoidal form. For a given valley aligned along the [100] direction, the energy as a function of wavevector \mathbf{k} is given by:

$$E(\mathbf{k}) = \frac{\hbar^2 k_x^2}{2m_l^*} + \frac{\hbar^2}{2m_t^*} (k_y^2 + k_z^2) \quad (2.29)$$

Here, m_l^* is the effective mass along the longitudinal axis of the ellipsoid (parallel to the valley direction), and m_t^* is the effective mass in the transverse directions. This anisotropic mass behavior reflects the ellipsoidal constant-energy surfaces around each conduction band minimum.

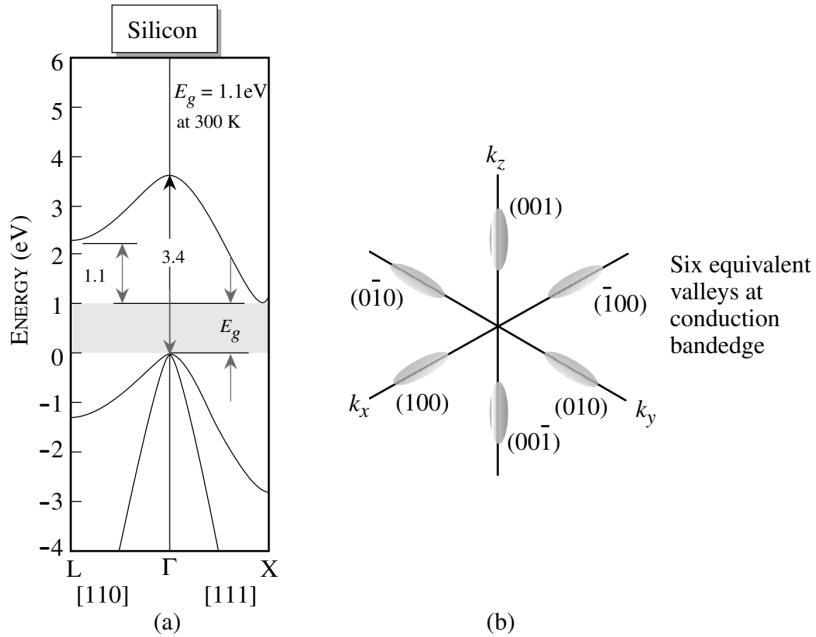


Figure 2.11.

2.6.2 GaAs

The bandstructure near the band edge in GaAs is characterized by a direct bandgap located at the Γ -point. This direct bandgap is one of the key reasons for the technological relevance of GaAs, as it enables excellent optical properties and high electron mobility in the conduction band.

Near the conduction band minimum at Γ , the energy-momentum relationship can be approximated by the simple parabolic form:

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

with an effective mass $m^* = 0.067m_0$, where m_0 is the free electron mass.

A more accurate representation, especially at higher energies, is provided by the nonparabolic dispersion relation:

$$E(1 + \alpha E) = \frac{\hbar^2 k^2}{2m^*}$$

where the nonparabolicity parameter α has a value of approximately 0.67 eV^{-1} .

For high electric field transport, it is important to consider higher energy conduction band valleys. Above the Γ -valley lie the L -valleys. There are eight equivalent L -points in the Brillouin zone, but due to equivalence under reciprocal lattice translations, only four of these are distinct. The energy separation between the Γ minimum and the L -valleys is approximately $\Delta E_{\Gamma L} = 0.29 \text{ eV}$. The effective mass in the L -valley is significantly larger than at Γ , with:

$$m_L^* \approx 0.25 m_0 \quad (2.30)$$

This mass difference plays a crucial role in high-field transport behavior, including the onset of negative differential resistance. At still higher energies lies the X -valley, separated from Γ by approximately $\Delta E_{\Gamma X} \approx 0.58$ eV. The effective mass in the X -valley is also large:

$$m_X^* \approx 0.6 m_0 \quad (2.31)$$

Under high electric fields, electrons can be transferred from the Γ -valley into the higher-energy L and X valleys, making these regions of the conduction band particularly important for transport properties.

The valence band in GaAs exhibits the standard structure consisting of heavy hole (HH), light hole (LH), and split-off (SO) bands. Due to the relatively large spin-orbit splitting in GaAs, the SO band lies significantly lower in energy and generally does not contribute to electronic or optoelectronic processes under typical operating conditions.

The valence band structure can be described using the Kohn-Luttinger parameters for GaAs, which are:

$$\gamma_1 = 6.85$$

$$\gamma_2 = 2.10$$

$$\gamma_3 = 2.90$$

These parameters yield the following density-of-states effective masses for holes:

$$m_{\text{HH}}^* = 0.45 m_0$$

$$m_{\text{LH}}^* = 0.08 m_0$$

As with silicon, the hole effective masses in GaAs are highly anisotropic, and the $E(\mathbf{k})$ relation for holes reflects this anisotropy in both curvature and carrier dynamics.

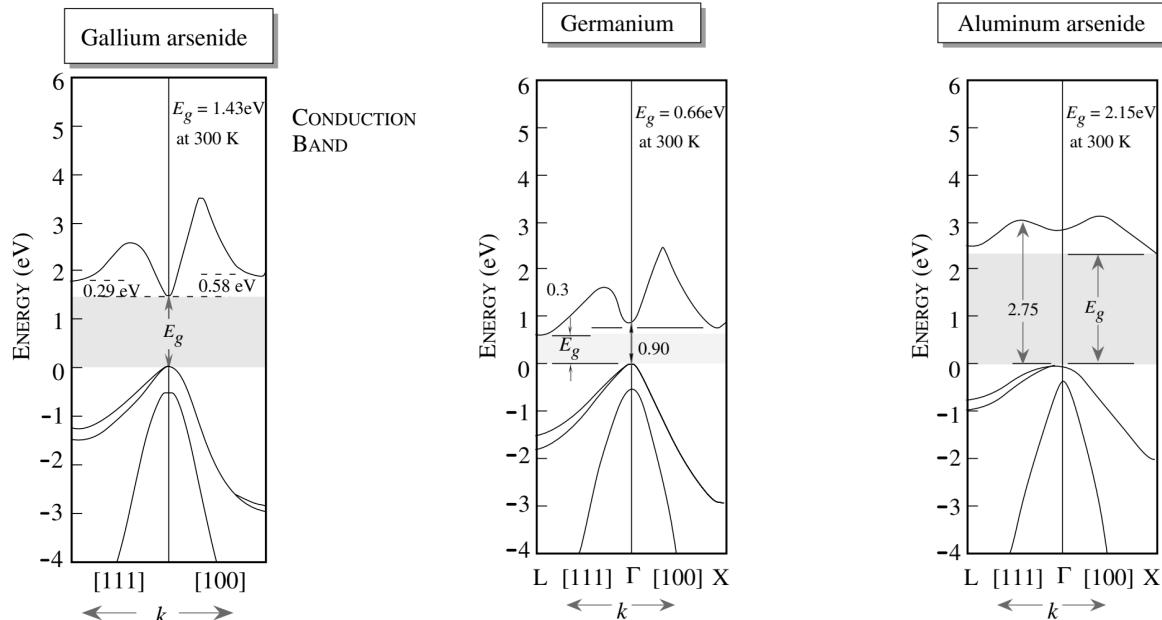


Figure 2.12.

Figure 2.13.

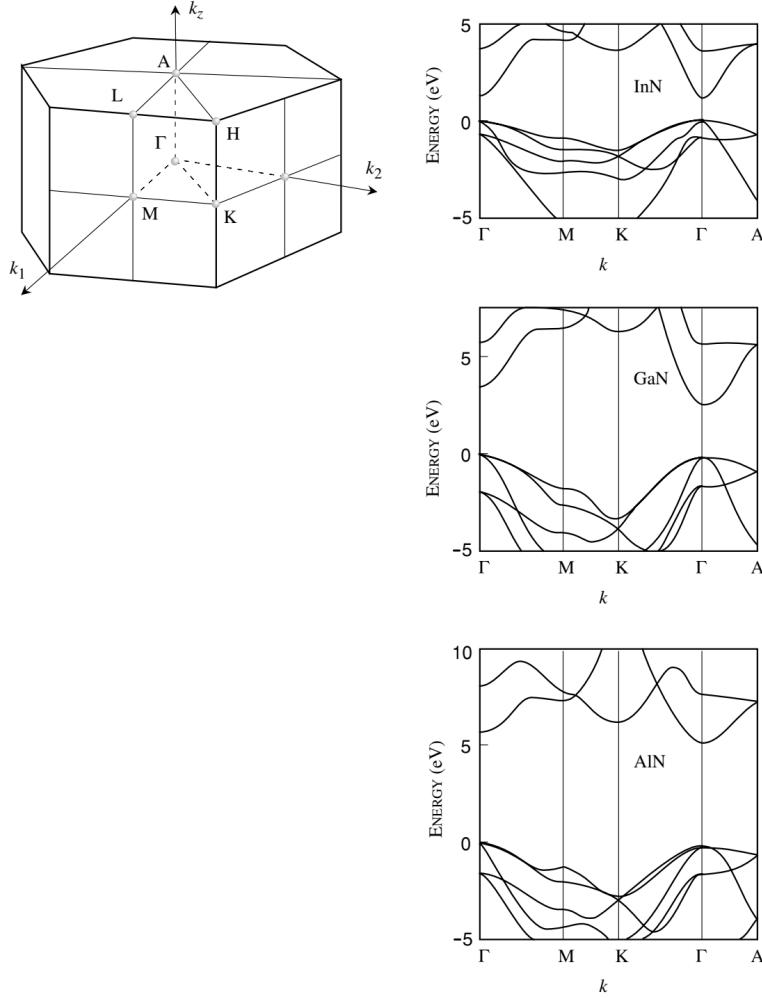


Figure 2.14. Bandstructure of InN, GaN, and AlN. Also shown is the Brillouin zone.

2.7 Mobile Carriers: Intrinsic Carriers

In metals, the density of mobile charge carriers is extremely high, typically on the order of $\sim 10^{23} \text{ cm}^{-3}$. In contrast, a semiconductor with a completely filled valence band and an empty conduction band does not conduct current under equilibrium conditions.

However, if electrons are excited from the valence band into the conduction band—either thermally or through doping—then two types of mobile charge carriers can exist. The electrons promoted to the conduction band contribute to conduction, and the absence of electrons (i.e., holes) left behind in the valence band also behave as positively charged mobile carriers.

If the concentration of conduction band electrons is denoted by n , and the concentration of holes in the valence band is p , then the total mobile carrier concentration in the material is given by:

$$n_{\text{total}} = n + p \quad (2.32)$$

In a three-dimensional semiconductor, the density of states (DOS) in the conduction band is given by:

$$N(E) = \frac{\sqrt{2} (m_{\text{dos}}^*)^{3/2} (E - E_c)^{1/2}}{\pi^2 \hbar^3} \quad (2.33)$$

where:

- $N(E)$ is the density of available electronic states per unit volume per unit energy,
- m_{dos}^* is the effective density of states mass for the conduction band,
- E_c is the conduction band edge.

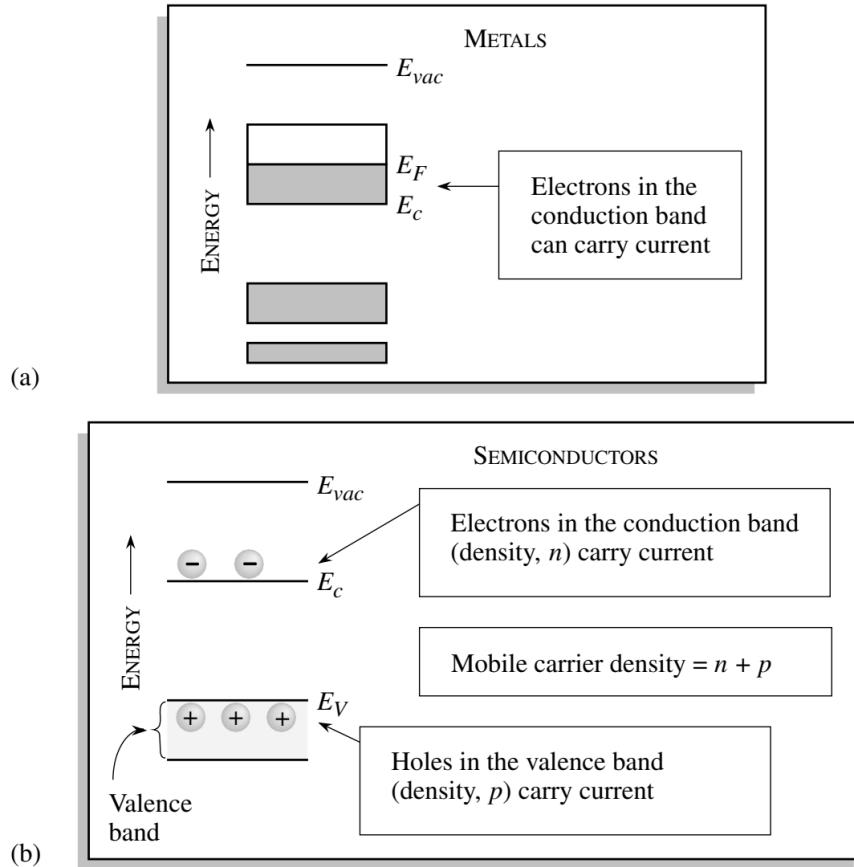


Figure 2.15. (a) Schematic of allowed energy bands in a metal. Electrons in the highest partially filled band contribute to electrical conduction. (b) Energy band diagram of a typical semiconductor, where current is carried by electrons in the conduction band and holes in the valence band.

In direct bandgap semiconductors, the density of states effective mass m_{dos}^* for the conduction band is simply equal to the electron effective mass at the band minimum. However, in indirect bandgap semiconductors, m_{dos}^* is defined as the geometric mean of the effective masses along the three principal axes:

$$m_{\text{dos}}^* = (m_1^* m_2^* m_3^*)^{1/3} \quad (2.34)$$

where m_1^*, m_2^*, m_3^* are the effective masses along the principal crystallographic directions. For silicon, which has six equivalent X -valleys, this becomes:

$$m_{\text{dos}}^* = 6 (m_t^2 m_l)^{1/3} \quad (2.35)$$

where m_l and m_t are the longitudinal and transverse effective masses of the conduction band ellipsoids, respectively.

For the valence band, which consists of heavy hole (HH) and light hole (LH) bands, the effective

density of states mass can be approximated by:

$$m_{\text{dos}}^* = \left(m_{\text{hh}}^{*3/2} + m_{\text{lh}}^{*3/2} \right)^{2/3} \quad (2.36)$$

In intrinsic semiconductors, conduction band electrons originate from the valence band, and thus the electron and hole densities are equal:

$$n = p = n_i = p_i$$

The electron concentration in the conduction band is given by the integral:

$$n = \int_{E_c}^{\infty} N_e(E) f(E) dE \quad (2.37)$$

Using the expression for the density of states and the Fermi-Dirac distribution, this becomes:

$$n = \frac{1}{2\pi^2} \left(\frac{2m_e^*}{\hbar^2} \right)^{3/2} \int_{E_c}^{\infty} \frac{(E - E_c)^{1/2}}{1 + \exp\left(\frac{E - E_F}{k_B T}\right)} dE \quad (2.38)$$

In the non-degenerate limit, where $(E - E_F) \gg k_B T$, we can approximate the Fermi-Dirac function by the Boltzmann distribution, and the integral simplifies to:

$$n = N_c \exp\left(\frac{E_F - E_c}{k_B T}\right) \quad (2.39)$$

where the effective density of states in the conduction band is:

$$N_c = 2 \left(\frac{m_e^* k_B T}{2\pi\hbar^2} \right)^{3/2} \quad (2.40)$$

A similar result holds for the hole concentration:

$$p = N_v \exp\left(\frac{E_v - E_F}{k_B T}\right) \quad (2.41)$$

where:

$$N_v = 2 \left(\frac{m_h^* k_B T}{2\pi\hbar^2} \right)^{3/2} \quad (2.42)$$

The product np is independent of the Fermi level and depends only on temperature and intrinsic material properties:

$$np = N_c N_v \exp\left(\frac{-E_g}{k_B T}\right) \quad (2.43)$$

This result is known as the *law of mass action*. If n increases, p must decrease accordingly to maintain this constant product, and vice versa.

For an intrinsic semiconductor, where $n = p = n_i$, we can express the intrinsic carrier concentration as:

$$n_i = p_i = 2 \left(\frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e^* m_h^*)^{3/4} \exp\left(\frac{-E_g}{2k_B T}\right) \quad (2.44)$$

Additionally, the intrinsic Fermi level E_i is approximately given by:

$$E_i = \frac{E_c + E_v}{2} + \frac{3}{4} k_B T \ln\left(\frac{m_h^*}{m_e^*}\right) \quad (2.45)$$

In computing the density of states masses m_e^* and m_h^* , one must account for the number of equivalent conduction band valleys and the contributions from both heavy and light hole bands in the valence band.

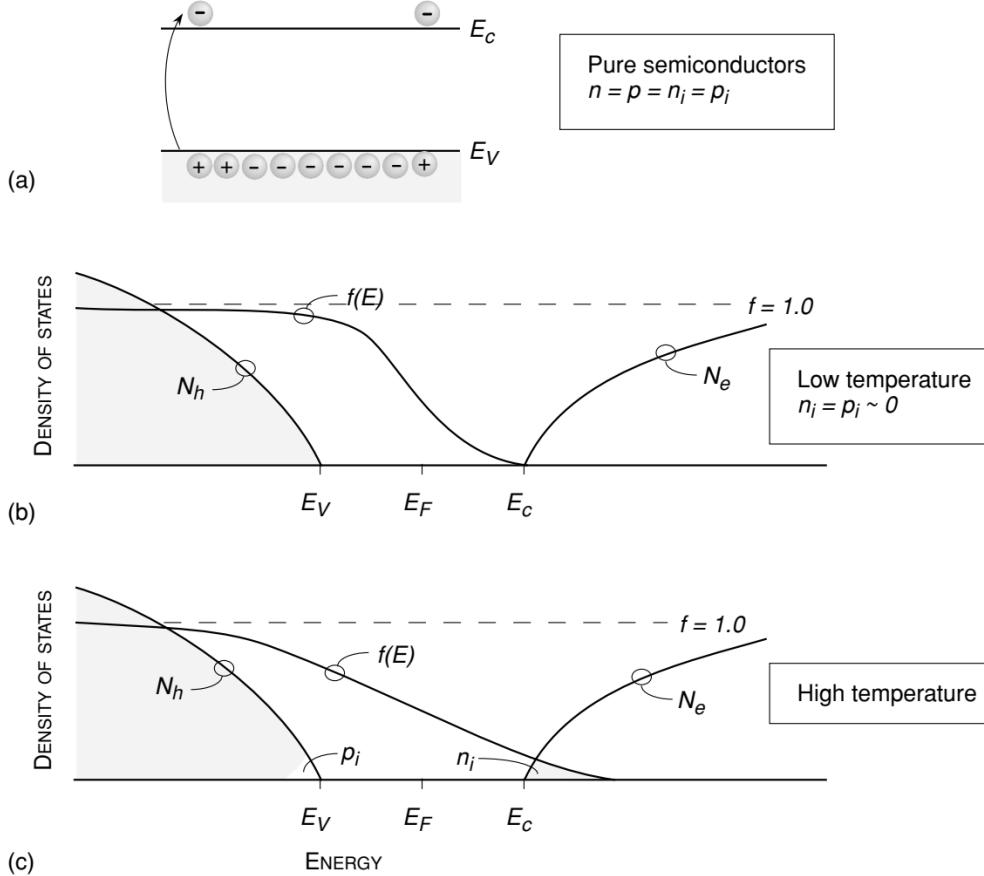


Figure 2.16. (a) Illustration of equal electron and hole concentrations in an intrinsic (pure) semiconductor. (b) Density of states and Fermi-Dirac distribution at low temperature. (c) Density of states and Fermi function at high temperature, where intrinsic carrier concentrations n_i and p_i increase significantly.

2.8 Doping: Donors and Acceptors

There are two primary types of dopants in semiconductors: *donors*, which contribute electrons to the conduction band, and *acceptors*, which accept electrons from the valence band, thereby creating holes. To understand how donor and acceptor states arise, consider a donor atom substituted into a crystal lattice.

For example, in silicon, a typical donor is a group V (pentavalent) atom that replaces a Si atom. Four of the donor's valence electrons form covalent bonds with neighboring Si atoms, just as in a normal Si lattice. The fifth valence electron, however, is left loosely bound. It experiences an attractive force from the now positively charged donor ion, which has a net charge of $+e$. This interaction is Coulombic in nature but is screened by the dielectric constant ϵ of the semiconductor:

$$U(r) = -\frac{e^2}{4\pi\epsilon r} \quad (2.46)$$

where $\epsilon = \epsilon_0\epsilon_r$, the product of vacuum permittivity and the relative dielectric constant of the material.

This scenario closely resembles the hydrogen atom problem, but with two modifications: the electron has an effective mass m^* rather than the free electron mass, and the Coulomb potential is reduced by the dielectric screening. The ground-state energy of the donor-bound electron is

then given by:

$$E_d = E_c - \frac{m^* e^4}{2(4\pi\epsilon)^2 \hbar^2} = E_c - 13.6 \text{ eV} \left(\frac{m^*}{m_0}\right) \left(\frac{1}{\epsilon_r^2}\right) \quad (2.47)$$

where E_c is the conduction band edge, and m_0 is the free electron mass. In contrast to the hydrogen atom, where the energy level is referenced to the vacuum, in semiconductors, the donor energy is referenced from the conduction band edge. The effective mass m^* used in this calculation is the conductivity effective mass m_σ^* , which reflects how electrons respond to external fields. This mass also governs donor binding energies and charge transport.

For direct bandgap semiconductors such as GaAs, the conductivity mass is simply the electron effective mass at the conduction band minimum. In indirect semiconductors like silicon, the conductivity mass is given by:

$$m_\sigma^* = 3 \left(\frac{1}{m_l^*} + \frac{2}{m_t^*}\right)^{-1} \quad (2.48)$$

where m_l^* and m_t^* are the longitudinal and transverse effective masses of the conduction band valleys.

According to this simplified model, the donor binding energy depends only on the properties

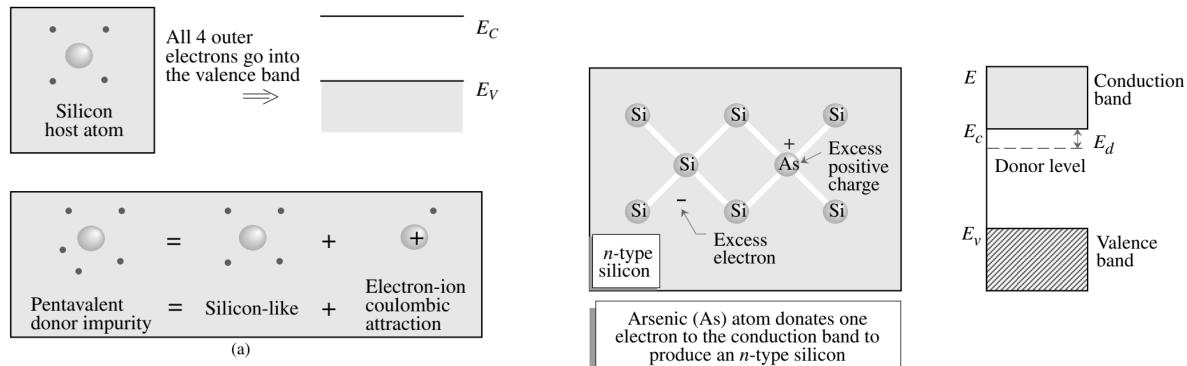


Figure 2.17. Schematic illustrating the donor model in semiconductors. The problem is approached by considering the host atom plus a Coulomb interaction. Silicon contributes four valence electrons per atom. A donor atom provides five electrons, four of which fill the valence band, while the fifth can be thermally excited into the conduction band.

of the host material—namely, its dielectric constant ϵ_r and effective mass m^* —and not on the specific dopant species. Using this model, typical donor binding energies are:

- Germanium (Ge): $\sim 0.006 \text{ eV}$
- Silicon (Si): $\sim 0.025 \text{ eV}$
- Gallium Arsenide (GaAs): $\sim 0.007 \text{ eV}$

In practice, small deviations from these values occur due to simplifications in the model. The discrepancy arises from local modifications in the potential near the impurity site—known as the *central cell correction*—which slightly perturbs the donor energy levels beyond what is predicted by the hydrogenic approximation.

Acceptors represent another class of intentional impurities. Like donors, they introduce localized energy levels in the bandgap. An acceptor is neutral when unoccupied and negatively charged when it captures an electron from the valence band, leaving behind a mobile hole. These levels

are created by substituting host atoms with impurities that have one fewer valence electron. For example, group III elements act as acceptors in group IV semiconductors such as Si and Ge, while Si can act as an acceptor in GaAs if it replaces an As atom.

Thus, donors and acceptors serve to modulate the free carrier concentrations in semiconductors, enabling control over electronic and optoelectronic behavior.

2.8.1 Carriers in Doped Semiconductors

As previously discussed, introducing a donor atom into a semiconductor can, in principle, contribute an extra electron to the conduction band. However, whether this electron becomes a free carrier or remains bound to the donor site depends on several factors: the donor binding energy, the donor concentration, and the temperature.

At very low temperatures, donor electrons typically remain bound to their parent atoms. This phenomenon, known as *carrier freezeout*, results in minimal free carrier concentration and hence poor conductivity. As the temperature increases, thermal energy becomes sufficient to ionize donor electrons, which then occupy states in the conduction band and act as mobile charge carriers. These free electrons significantly affect the electrical conductivity of the material. The donor site, now lacking an electron, becomes a positively charged ion.

A similar process occurs for acceptor atoms: when an acceptor captures an electron from the valence band, it becomes negatively charged and creates a mobile hole in the valence band.

In general, as discussed before, the relationship between the electron concentration n and the Fermi level E_F is given by:

$$n = \int_{E_c}^{\infty} N(E) f(E) dE$$

This equation must usually be solved numerically. However, a useful analytical approximation is provided by the *Joyce–Dixon* formula. According to this approximation, the Fermi level is related to the carrier concentration as follows:

$$E_F = E_c + k_B T \ln \left(\frac{n}{N_c} \right) + \frac{k_B T}{\sqrt{8}} \left(\frac{n}{N_c} \right)^{1/2} \quad (2.49)$$

Alternatively, for hole concentration p , the expression becomes:

$$E_F = E_v - k_B T \ln \left(\frac{p}{N_v} \right) - \frac{k_B T}{\sqrt{8}} \left(\frac{p}{N_v} \right)^{1/2} \quad (2.50)$$

The effective density of states in the conduction band is given by:

$$N_c = 2 \left(\frac{m_e^* k_B T}{2\pi\hbar^2} \right)^{3/2}$$

and similarly, the valence band effective density of states is:

$$N_v = 2 \left(\frac{m_h^* k_B T}{2\pi\hbar^2} \right)^{3/2}$$

These expressions allow one to estimate the Fermi level E_F for a known carrier concentration n , or vice versa, by iterative methods. If the correction terms ($n/\sqrt{8}N_c$) or ($p/\sqrt{8}N_v$) are neglected, the equations reduce to the standard Boltzmann approximation:

$$n \approx N_c \exp \left(\frac{E_F - E_c}{k_B T} \right) \quad (2.51)$$

and

$$p \approx N_v \exp \left(\frac{E_v - E_F}{k_B T} \right) \quad (2.52)$$

This simplified form is valid under non-degenerate conditions, where the Fermi level is several $k_B T$ away from the band edges.

2.8.2 Mobile Carrier Density and Carrier Freezeout

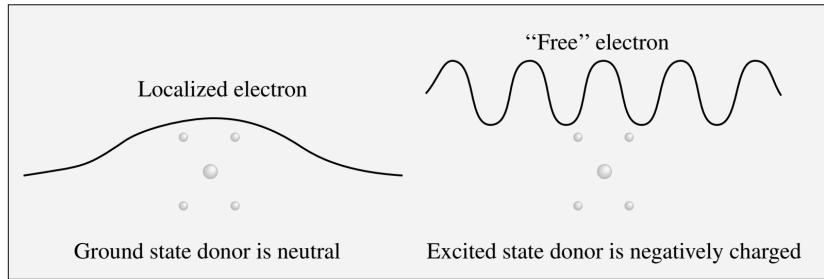


Figure 2.19. An electron bound to a donor does not participate in conduction. Only when the donor is ionized does the electron become free and contribute to electrical transport.

In the ground state of a donor atom embedded in a semiconductor, the excess electron introduced by the donor is localized near the donor nucleus. Since this bound state does not contribute to electrical conductivity, such electrons are not useful for modifying the electronic properties of the material.

At very low temperatures, donor electrons remain bound to the donor atoms—a phenomenon known as *carrier freeze-out*. As the temperature increases, thermal excitation allows the donor electrons to be ionized into the conduction band, where they become mobile and contribute to current transport. The donor atom, once ionized, becomes positively charged and acts as a scattering center for conduction electrons. The detailed analysis of scattering effects will be addressed in a later section.

The electron–donor system may exist in the following possible configurations:

1. the donor electron is ionized and free in the conduction band;
2. one electron is bound to the donor atom (with spin-up or spin-down);
3. two electrons attempt to occupy the donor site.

Although the third case is theoretically possible, the strong Coulombic repulsion between two electrons on the same donor site makes such double occupation energetically prohibitive. Therefore, a donor can bind at most one electron at a time, with either spin orientation.

Due to the two available spin states for the bound electron, the donor occupation statistics follow Fermi–Dirac statistics with a factor of two in the numerator. The probability that a donor level at energy E_d is occupied by an electron is given by:

$$f_d = \frac{1}{1 + \frac{1}{2} \exp\left(\frac{E_d - E_F}{k_B T}\right)} \quad (2.53)$$

where E_F is the Fermi level, k_B is Boltzmann's constant, and T is the temperature. The number density of electrons bound to donor sites is therefore:

$$n_d = N_d \cdot f_d = \frac{N_d}{1 + \frac{1}{2} \exp\left(\frac{E_d - E_F}{k_B T}\right)} \quad (2.54)$$

For $(E_d - E_F) \gg k_B T$, this simplifies to the Boltzmann approximation:

$$n_d \approx 2N_d \exp\left(-\frac{E_d - E_F}{k_B T}\right) \quad (2.55)$$

Similarly, for acceptor levels, the probability that a hole is trapped at an acceptor level is given by:

$$p_a = \frac{N_a}{1 + \frac{1}{4} \exp\left(\frac{E_F - E_a}{k_B T}\right)} \quad (2.56)$$

Here, E_a is the acceptor level energy, and the factor of $\frac{1}{4}$ arises from the four-fold degeneracy associated with the doubly degenerate valence band and spin.

These expressions allow us to estimate the ionization state of donor and acceptor levels as a function of temperature and Fermi energy.

2.8.3 Equilibrium Density of Carriers in Doped Semiconductors

In a doped semiconductor, we may introduce donor atoms, acceptor atoms, or both. To determine the electron and hole concentrations under such conditions, we must consider the occupation functions of free carriers and dopant states. The key constraint that must be satisfied is the *charge neutrality condition*, given by:

$$n_c + n_d = N_d - N_a + p_v + p_a \quad (2.57)$$

where:

- n_c is the density of free electrons in the conduction band,
- n_d is the density of electrons bound to donor atoms,
- p_v is the density of free holes in the valence band,
- p_a is the density of holes bound to acceptor atoms.

The electron and hole densities depend on the Fermi energy E_F , and hence the neutrality condition, combined with the Fermi–Dirac occupation probabilities, allows us to determine E_F at a given temperature. In the general case, this requires numerical methods: a trial Fermi level is chosen and adjusted iteratively until the neutrality equation is satisfied.

However, in the case of low doping levels—where carrier concentrations are small—the Boltzmann approximation can be applied, allowing for an analytical treatment. Under this approximation, and neglecting the unity in the denominator of the Fermi function, the free electron concentration is given by:

$$n = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) \quad (2.58)$$

For electrons bound to donors, we have:

$$n_d = \frac{N_d}{1 + \frac{1}{2} \exp\left(\frac{E_d - E_F}{k_B T}\right)} \approx \frac{N_d}{2} \exp\left(\frac{E_d - E_F}{k_B T}\right) \quad (2.59)$$

Combining these, the total electron concentration becomes:

$$n = n + n_d = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) + \frac{N_d}{2} \exp\left(\frac{E_d - E_F}{k_B T}\right) \quad (2.60)$$

Alternatively, this can be rearranged as:

$$\frac{n_d}{n_d + n_c} = \frac{1}{1 + \frac{N_c}{2N_d} \exp\left(-\frac{E_c - E_d}{k_B T}\right)} \quad (2.61)$$

For acceptor doping, a similar analysis leads to:

$$\frac{p_a}{p + p_a} = \frac{1}{1 + \frac{N_v}{4N_a} \exp\left(-\frac{E_a - E_v}{k_B T}\right)} \quad (2.62)$$

where:

- N_c , N_v are the effective density of states in the conduction and valence bands, respectively,
- N_d , N_a are the donor and acceptor concentrations,
- E_c , E_v , E_d , and E_a are the band edge and dopant energy levels.

These relations provide a way to estimate carrier concentrations and dopant ionization under low doping conditions.

In the Fig.2.19 we show how free electron density varies with temperature in a n-type silicon sample. As temperature increases, the fraction of “ionized” donors starts to increase until all of the donors are ionized and the free carrier density is equal to the donor density. This region is called the saturation region. Eventually, as the temperature is further raised, the carrier density starts to increase because of the intrinsic carrier density exceeding the donor density. At low temperatures the electrons are bound to the donors. This is the freezeout regime. Semiconductor devices usually operate in the saturation region where the mobile carrier density is essentially independent of temperature and is approximately equal to the doping density. Semiconductor devices cannot operate in the high temperature intrinsic regime, since it is not possible to control intrinsic carrier density by applying an external bias. Thus devices cannot be “shut off ” due to the leakage current from intrinsic carriers. High temperature electronics require large bandgap semiconductors for which the upper temperature limit is high.

2.8.4 Heavily Doped Semiconductors

In the theory discussed so far, we have made several important assumptions that are valid only under low doping conditions:

1. The bandstructure of the host crystal is assumed to remain unperturbed, such that the bandedge states are still well described by simple parabolic bands.
2. Dopants are treated as isolated and non-interacting, with their potential modeled as a simple Coulombic potential.

These assumptions break down as the doping level increases. When the average spacing between impurity atoms becomes comparable to $\sim 100 \text{ \AA}$, the potential experienced by an impurity electron is significantly influenced by neighboring impurities. This situation is analogous to the transition from isolated atomic levels to energy bands in a crystal. When atoms are far apart, discrete energy levels exist; however, as their separation decreases to a few angstroms, these levels broaden into bands due to interatomic interactions.

Similarly, at high doping levels, impurity states can overlap and form *impurity bands*. In this regime, several additional physical effects arise, altering the electrical and optical properties of the semiconductor. These effects must be considered for a proper understanding of heavily doped semiconductors.

Bandstructure Modifications

3.1 Bandstructure of Semiconductor Alloys

One of the simplest and most effective methods to alter the electronic properties of a material or to engineer entirely new properties is through the formation of alloys. Alloying has long been employed across various material classes—not only semiconductors, but also metals and insulators—as a technique to tailor physical and chemical characteristics.

In the context of semiconductors, the motivation to create alloys is typically driven by two primary goals:

1. *Tuning the bandgap*: This is particularly important in the design of light-emitting and light-detecting devices, where the bandgap dictates the energy (and thus the wavelength) of the absorbed or emitted photons.
2. *Lattice constant engineering*: Alloying allows for the synthesis of materials with desired lattice parameters, enabling lattice matching (or deliberate mismatch) with available substrates. For instance, the InGaAs alloy is frequently used because it can be lattice-matched to InP substrates, which are commercially available.

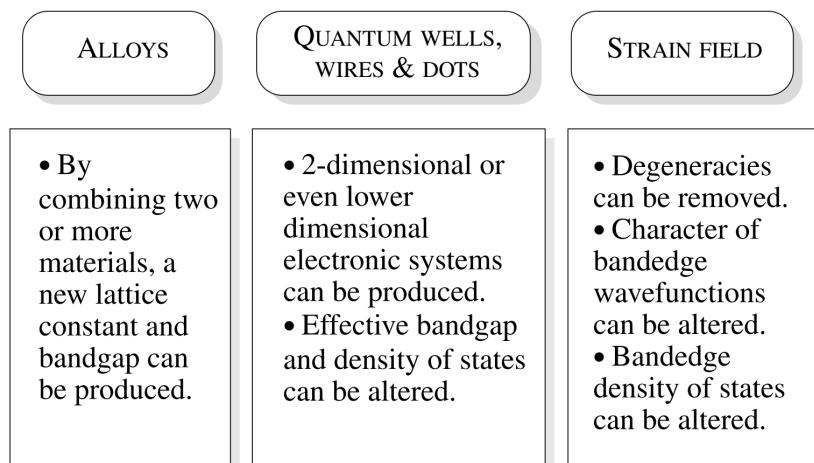


Figure 3.1. Different approaches to modify Bandstructure of semiconductors.

When an alloy of the form A_xB_{1-x} is produced via random mixing of two components A and B (a concept that can be generalized to multicomponent systems), the lattice constant of the

resulting alloy is often estimated using Vegard's law:

$$a_{\text{alloy}} = x a_A + (1 - x) a_B$$

This empirical relationship assumes a linear interpolation between the lattice constants a_A and a_B of the pure materials and is valid for random alloys where no phase separation occurs and both components share the same crystal structure. The term "random" in this context refers to the statistical distribution of atoms within the alloy. Specifically, in a random alloy, the probability that a given atomic site is occupied by atom A is x , while the probability that it is occupied by atom B is $1 - x$. However, several distinct atomic arrangements are theoretically possible when forming such an alloy:

1. *Phase separation*: Atoms of type A cluster together in one region, while atoms of type B cluster in another, forming distinct domains of the pure components.
2. *Random distribution*: Each atomic site is occupied according to the probability x for A and $1 - x$ for B , independent of neighboring sites.
3. *Ordered structure*: Atoms A and B arrange themselves in a well-defined periodic pattern, forming a superlattice.

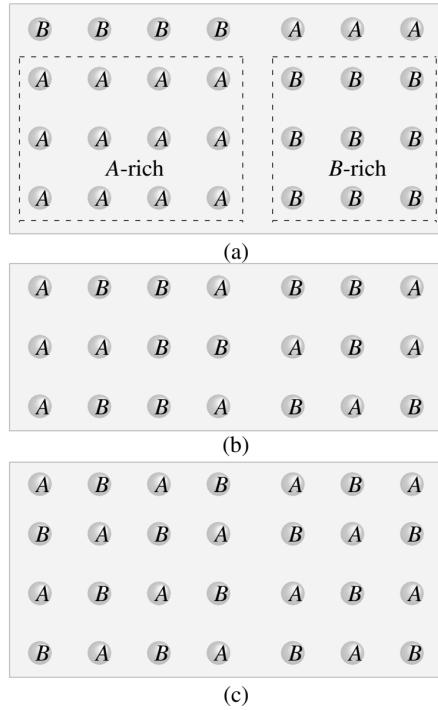


Figure 3.2. A schematic example of (a) a clustered, (b) a random, and (c) an ordered alloy structure.

In practice, most semiconductor alloys used in modern electronic and optoelectronic applications are synthesized with the goal of producing statistically random distributions of constituent atoms. Such random alloys provide the desired combination of bandgap tunability and structural compatibility with various substrates.

Alloys typically retain a well-defined crystal structure; however, the random occupation of lattice sites by different atomic species eliminates the strict periodicity of the background potential—except in the case of ordered alloys or superlattices. As a result, Bloch's theorem no longer applies, and the electronic wavefunctions can no longer be represented as simple traveling waves. Instead, they become more complex, exhibiting spatially varying probability distributions. To

make progress in understanding the bandstructure of such disordered systems, a commonly employed approximation is the *virtual crystal approximation* (VCA). This approach replaces the spatially random potential experienced by the electrons with an averaged periodic potential:

$$U_{\text{av}}(\mathbf{r}) = xU_A(\mathbf{r}) + (1 - x)U_B(\mathbf{r}) \quad (3.1)$$

Here, U_A and U_B are the atomic potentials of the two constituent species, and x represents the concentration of species B . The VCA allows one to treat the alloy as if it were a periodic crystal with an effective potential. In practical implementations—such as within the tight-binding method—this corresponds to averaging the matrix elements of the Hamiltonian. For direct bandgap semiconductors, where the conduction and valence band edges occur at the Γ -point (i.e., $\mathbf{k} = 0$), the bandgap is also approximated by a linear interpolation:

$$E_g^{\text{alloy}} = xE_g^A + (1 - x)E_g^B \quad (3.2)$$

However, in many real materials, this linear relationship is not sufficient due to a phenomenon known as *bandgap bowing*. This deviation arises from the increasing disorder and complex interactions introduced by alloying. A more accurate empirical expression for the bandgap is:

$$E_g^{\text{alloy}} = a + bx + cx^2 \quad (3.3)$$

where a , b , and c are material-specific constants, and c is known as the bowing parameter. In addition to bandgap behavior, the effective masses at the band edges also vary with composition. Within the VCA, the effective mass of the alloy can be approximated by:

$$\frac{1}{m_{\text{alloy}}^*} = \frac{x}{m_A^*} + \frac{1 - x}{m_B^*} \quad (3.4)$$

This follows from the expression for the energy dispersion near the band edge, assuming parabolic bands:

$$E_{\text{alloy}}(k) = x \frac{\hbar^2 k^2}{2m_A^*} + (1 - x) \frac{\hbar^2 k^2}{2m_B^*} \quad (3.5)$$

Several important alloy systems have played critical roles in advancing semiconductor electronics and optoelectronics. These will be briefly discussed in the following sections.

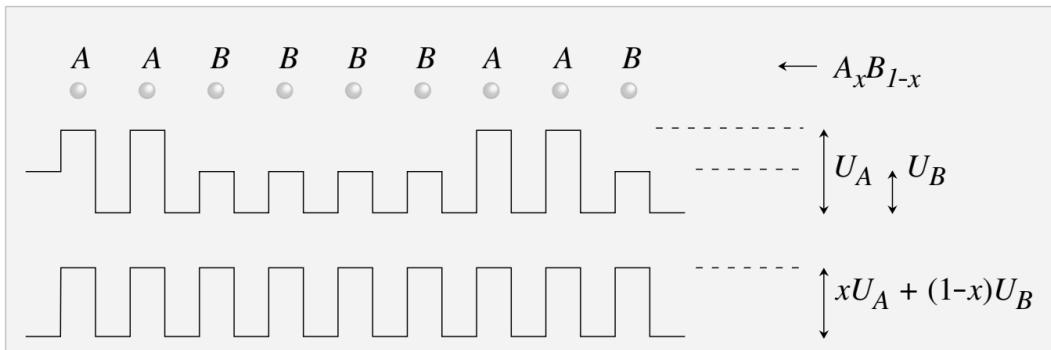


Figure 3.3. Motivation for the virtual crystal approximation (VCA). The top part illustrates a lattice with A- and B-type atoms, each associated with localized potentials U_A and U_B . In the VCA, extended states experience an effective periodic potential given by the weighted average $xU_A + (1 - x)U_B$.

3.1.1 GaAs/AlAs Alloy

The AlGaAs system is among the most important and widely studied semiconductor alloy systems. A key advantage of this system is that AlAs and GaAs are nearly lattice matched, allowing the alloy $\text{Al}_x\text{Ga}_{1-x}\text{As}$ to be epitaxially grown on GaAs substrates without the buildup of significant strain energy. This lattice compatibility makes AlGaAs highly suitable for the fabrication of high-speed electronic and optoelectronic devices.

In electronic applications, AlGaAs plays a central role in the development of modulation-doped field-effect transistors (MODFETs), while in optoelectronics it is used in a variety of devices such as modulators, photodetectors, and laser diodes.

One particularly noteworthy property of the AlGaAs alloy system is the compositional dependence of its electronic band structure, specifically the nature of its bandgap. As the aluminum content increases, the conduction band minimum transitions from the Γ -valley to the X -valley, resulting in a switch from a direct to an indirect bandgap.

This transition typically occurs when the aluminum mole fraction exceeds approximately 35%. Consequently, for most optoelectronic device applications that require efficient radiative recombination (i.e., direct bandgap), the aluminum composition is chosen to be below this threshold. This compositional dependence of the conduction band valleys is critical in determining the optical and electronic properties of AlGaAs-based devices and must be carefully accounted for in device design and material synthesis.

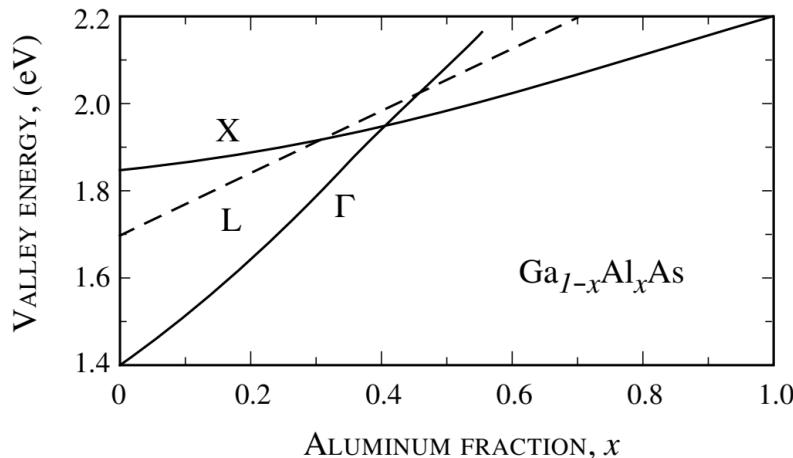


Figure 3.4. The variation of conduction band valleys in AlGaAs as a function of composition at 300 K.

3.2 Bandstructure Modifications by Heterostructures

It is possible to vary the chemical composition of a semiconductor structure along the growth direction using advanced epitaxial techniques such as molecular beam epitaxy (MBE) or metal-organic chemical vapor deposition (MOCVD). By introducing compositionally distinct layers, one can fabricate heterostructures capable of confining electronic states, thereby creating lower-dimensional systems. These systems are now widely utilized in high-performance optoelectronic and electronic devices.

Considerable effort has been devoted to extending confinement from one dimension (quantum wells) to two-dimensional (quantum wires) and three-dimensional (quantum dots) systems. While confinement in one dimension is conceptually straightforward and technologically mature, achieving confinement in additional directions often requires complex and challenging epitaxial or post-growth processing techniques.

As previously discussed, strained epitaxy can lead to the formation of self-assembled quantum dots, which serve as quasi-zero-dimensional (0D) systems. These have enabled the development of novel device architectures. Although our focus will remain primarily on quantum wells, we will also highlight key considerations related to quantum dots.

When two semiconductors are joined to form a heterostructure, one of the most critical questions is how the band edges of the two materials align at the interface. Suppose material *A* has a bandgap E_g^A , conduction band edge E_c^A , and valence band edge E_v^A , and material *B* has a bandgap E_g^B , with corresponding band edges E_c^B and E_v^B . Several distinct band alignments may arise depending on the specific materials used.

In semiconductor physics, it is common to use electron affinity (or work function) to estimate how the conduction or valence bands align. However, this approach often fails for real heterojunctions due to complex interfacial effects such as charge transfer and atomic bonding. As a result, accurate determination of band lineups usually relies on experimental measurements, though theoretical predictions can often capture general trends. Three main types of band alignments are typically observed:

- **Type I (Straddling gap):** Both the conduction band minimum and valence band maximum of the narrower-gap material lie within the bandgap of the wider-gap material. This allows both electrons and holes to be confined in the same region. Common examples include GaAs/AlGaAs, InGaAs/InP, and GaN/AlGaN systems.
- **Type II (Staggered gap):** The conduction band minimum lies in one material while the valence band maximum lies in the other. Although this configuration may lead to a small effective bandgap, the spatial separation of carriers weakens optical transitions. Antimonide-based systems such as InSb and GaSb often exhibit type II alignment. The InAs/GaSb system is a well-known example.
- **Broken gap:** In some cases, the conduction band of one material lies below the valence band of the other. These are sometimes referred to as broken gap type II heterostructures.

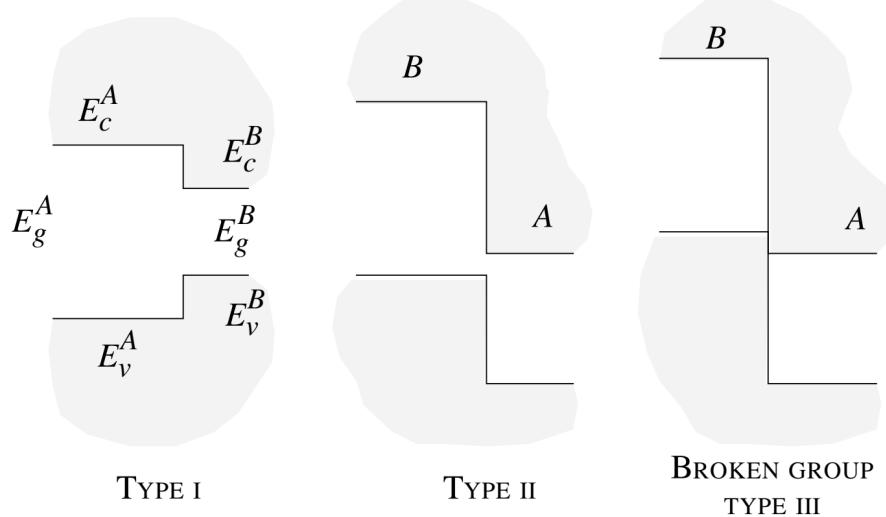


Figure 3.5. Various possible bandedge lineups in heterostructure.

Once the band alignment is determined, the next step is to describe the electronic states in the heterostructure. A widely used method for this purpose is the $k \cdot p$ approach. In its simplest form, the Schrödinger equation for a carrier in a heterostructure can be approximated as:

$$\left[-\frac{\hbar^2}{2m_0} \nabla^2 + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}) \Rightarrow \left[-\frac{\hbar^2}{2m^*} \nabla^2 + E_{\text{edge}} \right] \phi = E\phi \quad (3.6)$$

Here, the atomistic potential $V(\mathbf{r})$ is replaced by the band-edge energy E_{edge} , and the complex effect of the crystal environment is incorporated into the effective mass m^* . This simplification forms the basis for the envelope function approximation commonly used in modeling quantum wells and other heterostructures.

3.2.1 Bandstructure in Quantum Wells

To calculate the bandstructure in quantum wells, it is essential to understand the nature of the energy levels and wavefunctions near the edges of the bandgap, as these determine the electronic states in the quantum well. We consider the case where the well region is composed of a direct bandgap semiconductor. In such systems, the conduction band states are primarily of s -type character, while the valence band states are p -type. Although the formalism can be extended to other material systems, we focus here on a basic quantum well structure.

The Schrödinger equation describing the electron states in the quantum well, within the effective mass approximation, is given by:

$$\left[-\frac{\hbar^2}{2m^*} \nabla^2 + V(z) \right] \Psi = E\Psi \quad (3.7)$$

Here, m^* is the effective mass of the electron, and $V(z)$ is the quantum well potential, which depends only on the growth direction z . The total wavefunction can be separated into its in-plane and out-of-plane components:

$$\Psi(x, y, z) = e^{i(k_x \cdot x + k_y \cdot y)} f(z) \quad (3.8)$$

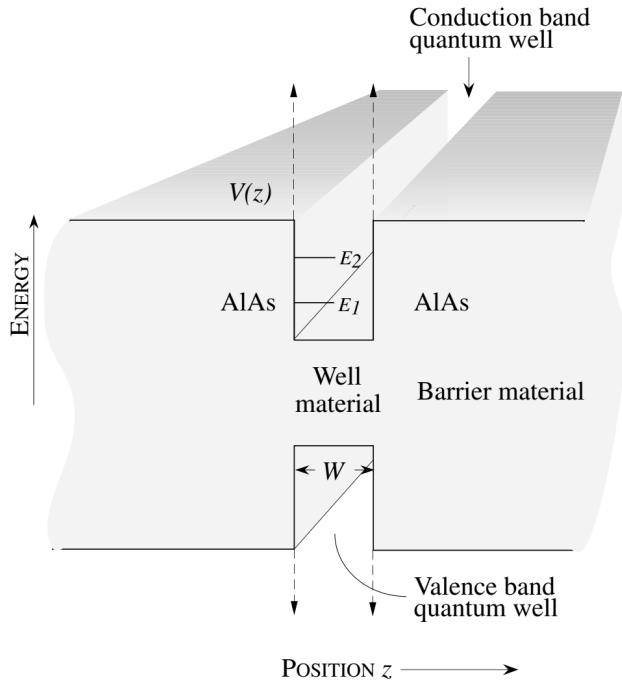


Figure 3.6. A quantum well formed for the electron and holes in a heterostructure.

Substituting this into the Schrödinger equation yields a one-dimensional equation for $f(z)$:

$$\left[-\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial z^2} + V(z) \right] f(z) = E_n f(z) \quad (3.9)$$

This one-dimensional quantum well problem is standard and can be found in many undergraduate quantum mechanics texts. Assuming infinite potential barriers, the wavefunctions are:

$$f_n(z) = \begin{cases} \cos\left(\frac{n\pi z}{W}\right), & \text{if } n \text{ is even} \\ \sin\left(\frac{n\pi z}{W}\right), & \text{if } n \text{ is odd} \end{cases}$$

where W is the well width. The corresponding quantized energy levels are:

$$E_n = \frac{\pi^2 \hbar^2 n^2}{2m^* W^2} \quad (3.10)$$

Including the in-plane kinetic energy contribution, the total energy of the electron becomes:

$$E(k) = E_n + \frac{\hbar^2 k_{\parallel}^2}{2m^*} \quad (3.11)$$

This results in subbands within the conduction and valence bands. If the potential barrier V is finite, the wavefunction decays exponentially into the barrier region. The solution inside the well remains sinusoidal, but matching boundary conditions leads to the transcendental equations:

$$\alpha \tan\left(\frac{\alpha W}{2}\right) = \beta \quad \text{or} \quad \alpha \cot\left(\frac{\alpha W}{2}\right) = -\beta \quad (3.12)$$

where

$$\alpha = \sqrt{\frac{2m^* E}{\hbar^2}}, \quad \beta = \sqrt{\frac{2m^*(V_c - E)}{\hbar^2}} \quad (3.13)$$

These equations can be solved numerically to obtain the quantized energy levels E_1, E_2, E_3, \dots , each corresponding to a subband in the quantum well.

In the valence band, both heavy hole and light hole subbands exist, and their detailed dispersion relations are affected by strain and confinement. The resulting subband structure plays a significant role in determining the optical and transport behavior of heterostructures.

An important consequence of the subband formation is its impact on the density of states (DOS), which influences both optical absorption and carrier dynamics.

For a quantum well, the DOS in the conduction band is:

$$N(E) = \sum_i \frac{m_i^*}{\pi \hbar^2} \sigma(E - E_i) \quad (3.14)$$

where σ is the Heaviside step function and E_i are the subband energies.

For the valence band, including both heavy and light hole contributions, the DOS becomes:

$$N(E) = \sum_i \sum_{j=1}^2 \frac{m_j^*}{\pi \hbar^2} \sigma(E_{ij} - E) \quad (3.15)$$

Here, $j = 1$ denotes heavy holes and $j = 2$ denotes light holes, and m_{ij}^* are the respective effective masses.

In the simplified model described here, the conduction band state is assumed to be purely *s*-type, and a basic effective mass approximation is used. More sophisticated calculations can incorporate full-band models, such as eight-band $k \cdot p$ theory, which captures coupling between conduction and valence bands. For unstrained structures, such enhancements do not significantly alter the conduction subband energies. However, for strained heterostructures, a more complete bandstructure model becomes necessary.

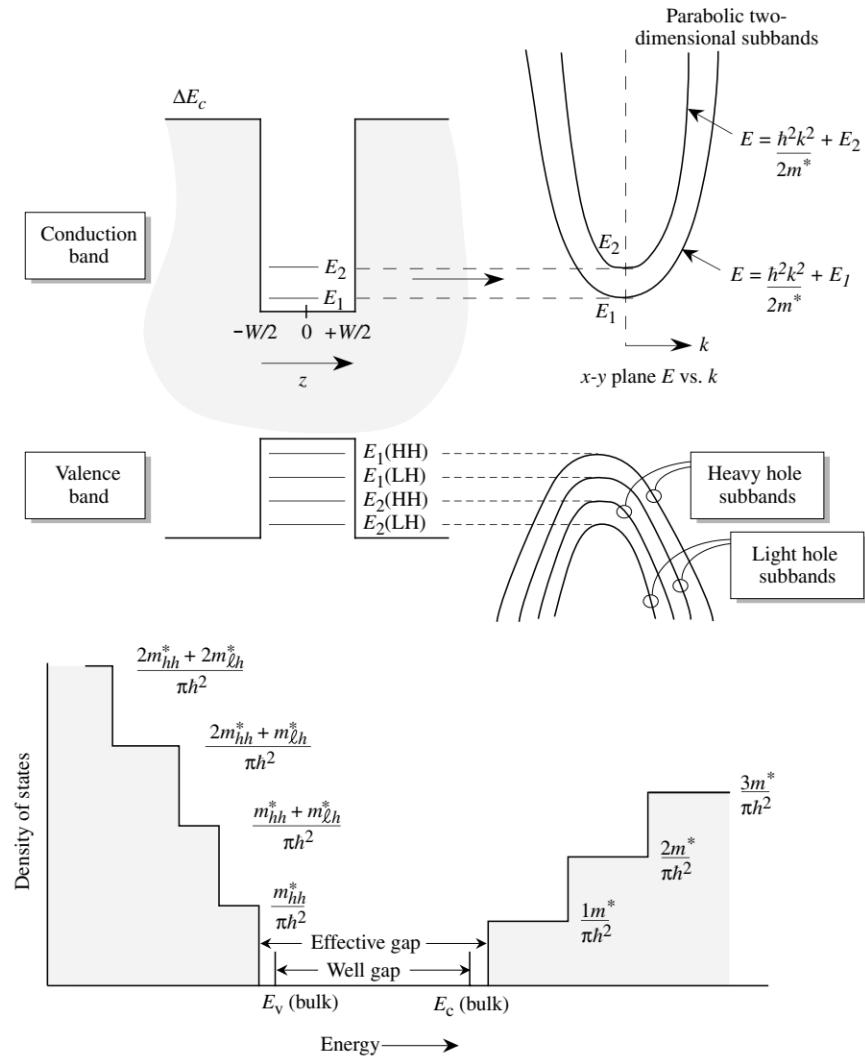


Figure 3.7. Density of states in a 3-, 2-, and 1-dimensional system with a parabolic energy momentum relations.

3.2.2 Valence Bandstructure in Quantum Wells

The description of quantum well bandstructure and density of states outlined earlier is quite accurate for electron states. This is because, in direct bandgap semiconductors, the conduction band is derived predominantly from an s -type orbital, and can therefore be treated as a single non-degenerate band.

However, the valence band states arise from p -type orbitals, leading to the formation of heavy-hole (HH) and light-hole (LH) bands. While the HH states (with total angular momentum $j = 3/2$, $m_j = \pm 3/2$) and LH states ($j = 3/2$, $m_j = \pm 1/2$) are pure and uncoupled at $k = 0$, they exhibit strong mixing as one moves away from the Brillouin zone center.

For the purpose of determining subband energy levels, the starting energies can still be calculated separately for HH and LH bands, similarly to the electron case. However, the in-plane ($k_x - k_y$) dispersion of hole states is only approximately given by a simple parabolic relation:

$$E(\mathbf{k}) = E_{n,i} + \frac{\hbar^2 k^2}{2m_i^*} \quad (3.16)$$

where i denotes either the HH or LH band. A more accurate description of the valence band structure is obtained by solving the Kohn–Luttinger form of the Schrödinger equation:

$$[H + V_p(z)]\Psi = E\Psi \quad (3.17)$$

Here, H_p is the Kohn–Luttinger Hamiltonian, which in the case of HH and LH coupling becomes a 4×4 matrix differential operator. Treating the out-of-plane momentum as an operator ($k_z = -i\partial/\partial z$), we obtain the following forms for the Hamiltonian matrix elements:

$$H_{\text{hh}} = -\frac{\hbar^2}{2m_0} \left[(\gamma_1 + \gamma_2)(k_x^2 + k_y^2) - (\gamma_1 - 2\gamma_2) \frac{\partial^2}{\partial z^2} \right] + V_p(z) \quad (3.18)$$

$$H_{\text{lh}} = -\frac{\hbar^2}{2m_0} \left[(\gamma_1 - \gamma_2)(k_x^2 + k_y^2) - (\gamma_1 + 2\gamma_2) \frac{\partial^2}{\partial z^2} \right] + V_p(z) \quad (3.19)$$

$$c = \sqrt{3} \frac{\hbar^2}{2m_0} [\gamma_2(k_x^2 - k_y^2) - 2i\gamma_3 k_x k_y] \quad (3.20)$$

$$b = -i\sqrt{3} \frac{\hbar^2}{m_0} \gamma_3 (-k_y - ik_x) \frac{\partial}{\partial z} \quad (3.21)$$

In these expressions, $V(z)$ represents the quantum well confinement potential, and γ_1 , γ_2 , and γ_3 are the Luttinger parameters describing the anisotropic hole mass.

When biaxial strain is present, additional terms must be included in the Hamiltonian to account for strain-induced modifications in the band structure. These will be introduced in later sections. The general form of the hole wavefunction can be written as:

$$\Psi_h^m(\mathbf{k}_\parallel, z) = \sum_v g_v^m(z) U^v(\mathbf{r}) e^{i\mathbf{k}_\parallel \cdot \boldsymbol{\rho}} \quad (3.22)$$

Here, $g_v^m(z)$ are envelope functions that depend on the confinement potential along the growth direction, $U^v(\mathbf{r})$ are the bulk valence band Bloch functions, m is the subband index, and v labels the spinor components associated with the total angular momentum $j = 3/2$ states.

This formalism provides a more complete and accurate picture of the valence band substructure in quantum wells, particularly when dealing with hole transport, optical transitions, and strain-induced effects.

3.3 Strain and Deformation Potential Theory

The influence of strain on the electronic and optical properties of semiconductors has been studied extensively over several decades. These investigations have been critical for developing a deeper theoretical understanding of semiconductor bandstructure, particularly by identifying the symmetry properties of energy states involved in optically observed transitions.

In addition, strain studies have provided valuable insights into the determination of deformation potentials, which play a fundamental role in describing electronic transport under the influence of lattice scattering. Traditionally, strain was introduced into semiconductors through external mechanical means, often employing sophisticated tools such as diamond anvil cells. These experiments were primarily focused on elucidating the fundamental physics of semiconductors.

With the development of strained heteroepitaxy, it has become possible to incorporate strain directly during the epitaxial growth process. Remarkably, built-in strains of a few percent can now be achieved simply by depositing a film onto a substrate with a mismatched lattice constant, as discussed in prior sections.

Once the strain tensor is known, the deformation potential theory can be employed to compute the effect of strain on the electronic states throughout the Brillouin zone. The perturbation

Hamiltonian due to strain is introduced and analyzed within the framework of first-order perturbation theory. It is generally expressed as:

$$H_{ij} = \sum_{\alpha\beta} D_{ij}^{\alpha\beta} \varepsilon_{\alpha\beta} \quad (3.23)$$

Here, $D_{ij}^{\alpha\beta}$ are the matrix elements of the deformation potential tensor D_{ij} , and $\varepsilon_{\alpha\beta}$ are the components of the strain tensor. The deformation potential operator D_{ij} transforms as a second-rank tensor under the crystal's symmetry operations.

This formalism provides a powerful and practical framework for quantifying the influence of strain on semiconductor electronic structure, and is widely used in the analysis

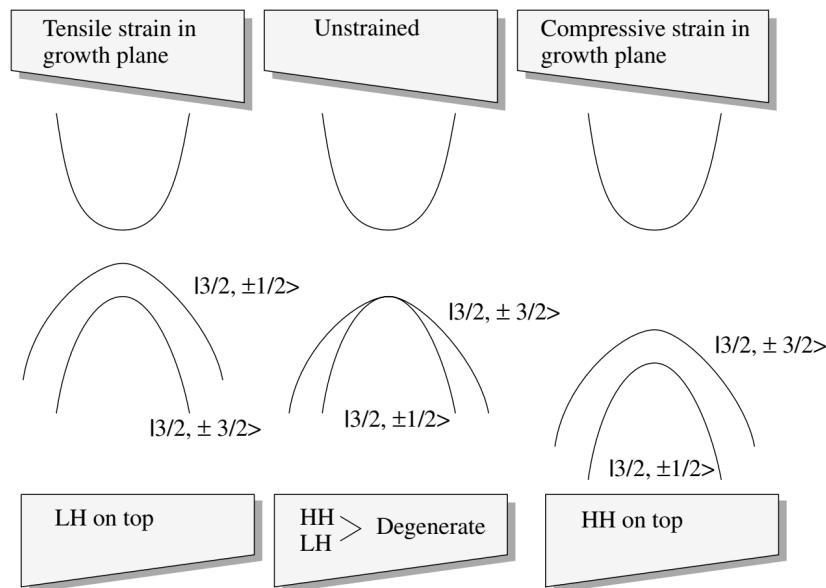


Figure 3.8. The effect of strain on the bandstructure of a semiconductor. The conduction band minimum (CBM) and valence band maximum (VBM) shift in response to applied strain.

Strain introduced in a semiconductor layer grown along the (001) crystallographic direction can significantly modify the band structure, particularly near the band edges. In the case of a direct bandgap semiconductor, the conduction band edge shifts upward or downward relative to its unstrained position, depending on the nature of the strain. However, since the conduction band minimum is typically a non-degenerate state, no splitting occurs under strain.

In contrast, the valence band edge is degenerate in the unstrained bulk system. Even in unstrained quantum wells, quantum confinement alone can lift this degeneracy between heavy-hole (HH) and light-hole (LH) states. However, the resulting energy splitting is usually modest, typically on the order of 10–15 meV.

When biaxial strain is applied, the impact on the valence band becomes much more pronounced. Under biaxial compressive strain, the bandgap increases and the HH–LH degeneracy is significantly lifted. The energy separation between these states can reach values as large as 100 meV, making strain a powerful tool for engineering the valence band density of states.

Specifically, under biaxial compressive strain, the heavy-hole state shifts above the light-hole state. Conversely, under biaxial tensile strain, the light-hole state lies higher in energy than the heavy-hole state. This strain-induced splitting is particularly important for controlling optical transitions and enhancing the performance of optoelectronic devices.

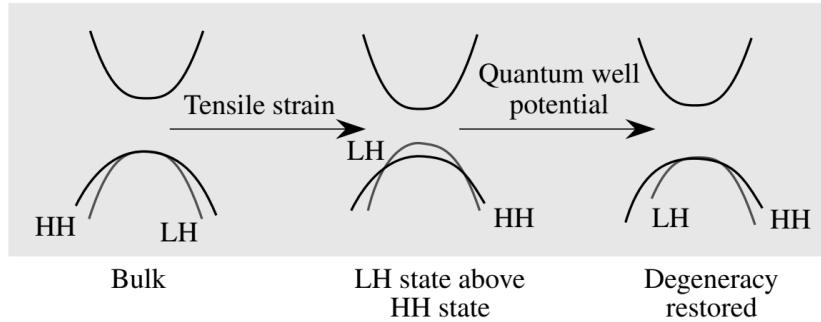


Figure 3.9. Effect of biaxial tensile strain and quantum confinement on conduction and valence band states. With appropriate tensile strain, heavy-hole and light-hole states can become degenerate at $\mathbf{k} = 0$ within a quantum well structure.

3.3.1 Strained Quantum Wells

Due to limitations imposed by critical thickness, most strained layers grown by epitaxy are implemented in the form of quantum wells. In these structures, both the effective bandgap and the carrier effective masses are strongly influenced by the presence of strain.

As previously discussed, strain—particularly biaxial strain—has a pronounced effect on the valence band edge states. One of the most significant consequences is the reduction of the band-edge density-of-states effective mass. In some cases, the incorporation of strain can lead to a reduction in this mass by nearly a factor of three. It is evident from the discussion that strain can be a powerful tool for tailoring the electronic bandstructure of semiconductors. The magnitude of strain considered here is readily achievable through strained-layer epitaxy. Importantly, the induced splitting between heavy-hole and light-hole states in the valence band can reach values on the order of 100 meV. Such splitting is accompanied by significant modifications in the band curvature, which directly influence the effective masses and density of states.

It is worth noting that achieving similar levels of uniaxial strain through external mechanical means is extremely challenging. The ability to incorporate such strain epitaxially provides a unique and practical avenue for bandstructure engineering.

An important question is whether the strain-induced modifications translate into observable changes in physical properties. This topic will be explored in greater depth in later chapters dealing with transport and optical behavior. As will be shown, the effects are indeed substantial and play a key role in the performance of strained semiconductor devices.

3.4 Polar Heterostructures

We have previously discussed that in certain materials, a net polarization can arise due to various mechanisms such as spontaneous polarization, the piezoelectric effect, or ferroelectricity. In face-centered cubic (fcc) semiconductors, spontaneous polarization does not occur; however, strain-induced effects can still lead to a net polarization.

One important consequence of polarization differences in semiconductor heterostructures is the emergence of significant built-in electric fields at interfaces. For example, in a polar heterostructure, the interface electric field can be expressed as:

$$F = \frac{P}{\epsilon} \quad (3.24)$$

where P is the net polarization and ϵ is the dielectric constant of the material. This fixed interfacial charge can be harnessed to induce built-in electric fields, generate band bending, and create

mobile charge carriers. With proper structural design, this interfacial charge can effectively serve the role of a delta-doping sheet.

In zinc blende semiconductors, the piezoelectric effect is the dominant mechanism for polarization. The polarization induced by strain is given in general by:

$$P_i = \sum_{k,l} e_{ikl} \varepsilon_{kl} \quad (3.25)$$

According to Nye (1957), in zinc blende structures only one independent piezoelectric coefficient exists. In the reduced index notation—where $xx \Rightarrow 1$, $yy \Rightarrow 2$, $zz \Rightarrow 3$, $yz \Rightarrow 4$, $zx \Rightarrow 5$, and $xy \Rightarrow 6$ —the only non-zero components of the piezoelectric tensor are:

$$e_{14} = e_{25} = e_{36} \quad (3.26)$$

This implies that only shear strain contributes to a finite piezoelectric polarization. As previously discussed, in (100) growth the strain tensor is diagonal, and hence, no piezoelectric polarization arises under such epitaxial conditions. However, when growth is performed along directions such as (111), off-diagonal components of the strain tensor become significant. This results in a strong dipole moment across the quantum well, inducing an internal electric field. The corresponding field is given by:

$$F = \frac{\sqrt{3}e_{14}\varepsilon_{xy}}{\varepsilon_s} \quad V/m \quad (3.27)$$

Here, e_{14} is the piezoelectric coefficient (typically on the order of 0.1 C/m^2), ε_{xy} is the off-diagonal strain component, and ε_s is the dielectric constant. A strain magnitude of approximately 1% can generate electric fields as high as 10^5 V/cm , highlighting the importance of piezoelectric effects in strained quantum wells.

In wurtzite structures (e.g., InN, GaN, AlN), epitaxial growth is typically conducted along the c -axis, i.e., in the (0001) or ($\bar{0}001$) direction. For a sufficiently thick substrate, the in-plane strain components are given by:

$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{a_3}{a_0} - 1 \quad (3.28)$$

where a_s is the lattice constant of the substrate and a_0 is that of the unstrained layer. The strain along the growth direction is determined by:

$$\varepsilon_{zz} = -2 \frac{c_{13}}{c_{33}} \left(\frac{a_s}{a_0} - 1 \right) \quad (3.29)$$

These strains generate a piezoelectric polarization in the strained layer. For wurtzite crystals, the total polarization due to strain is given by:

$$P_{pz} = e_{33}\varepsilon_{zz} + e_{31}(\varepsilon_{xx} + \varepsilon_{yy}) \quad (3.30)$$

where e_{33} and e_{31} are piezoelectric constants. The resulting polarization field is directed along the (0001) axis, corresponding to the Ga-face direction in the crystal, following the convention that places Ga atoms on the upper site of each bilayer.

The electric field induced by this polarization can be expressed simply as:

$$F = \frac{P}{\varepsilon_s} \quad (3.31)$$

Typical material constants for InGaN systems include $c_{13} \approx 109 \text{ GPa}$ and $c_{33} \approx 355 \text{ GPa}$, yielding a ratio $2c_{13}/c_{33} \approx 0.6$. The piezoelectric and spontaneous polarizations play a major role in determining the band alignment and carrier behavior in group-III nitride heterostructures.

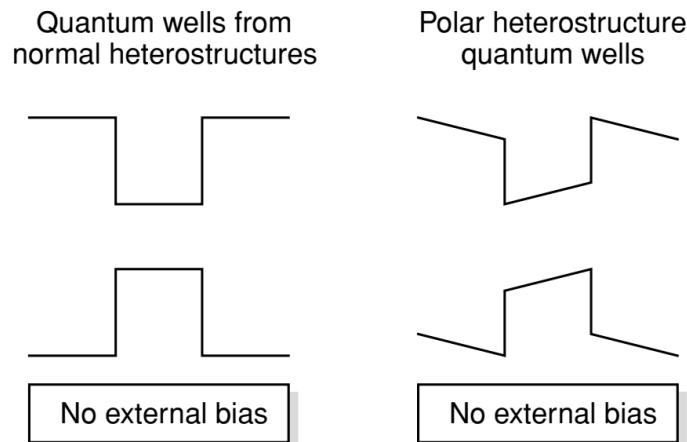


Figure 3.10. A comparison of (undoped) nonpolar and polar quantum well band profiles in the absence of any external electric field.

Lattice Vibrations and Phonons

4.1 Introduction

In a crystalline material, atoms are not fixed at rigid lattice sites but vibrate around their equilibrium positions. One of the most important scattering mechanisms for mobile carriers in semiconductors arises from these lattice vibrations.

In the earlier discussion on bandstructure, we assumed that the background potential is perfectly periodic and time-independent. However, in real materials, the ions that constitute the crystal lattice are dynamic and undergo thermal vibrations, which become more pronounced with increasing temperature.

Scattering results from the disturbances in the periodic potential caused by lattice vibrations. Before addressing how these vibrations lead to scattering, it is essential to first understand the fundamental properties of lattice vibrations themselves. Once these properties are established, we can then analyze how the resulting potential fluctuations contribute to carrier scattering in semiconductors.

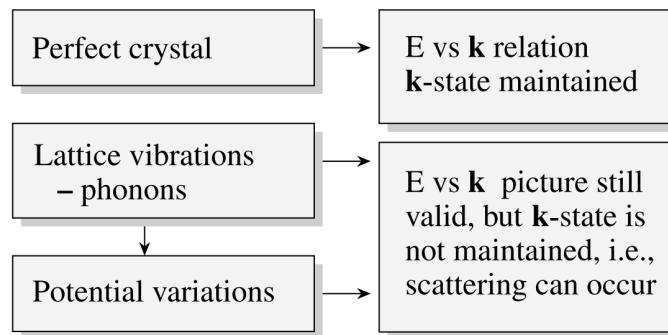


Figure 4.1. Electron scattering arises from imperfections in the crystal, including lattice vibrations (phonons) and local potential variations, which disrupt coherent transport.

4.2 Lattice Vibrations

The reason a particular crystal structure is chosen by a material is fundamentally related to the minimization of the system's total energy. As atoms approach each other to form a crystal, they experience both attractive and repulsive interactions. The attractive forces can arise from

various mechanisms, such as Van der Waals interactions (due to induced dipole moments when electron clouds are perturbed by nearby atoms), ionic bonding (involving electron transfer between atoms), and covalent bonding (where electrons are shared between atoms).

When atoms are brought very close, however, there is a strong repulsive interaction primarily due to the Pauli exclusion principle, which prevents electrons on neighboring atoms from occupying the same space.

As a result, the total potential energy of the system as a function of interatomic separation exhibits a minimum at some equilibrium spacing R_0 , where the system achieves mechanical stability.

In general, we can expand the crystal binding energy around the equilibrium spacing R_0 as a Taylor series:

$$U(R) = U(R_0) + \frac{dU}{dR} \Big|_{R_0} \Delta R + \frac{1}{2} \frac{d^2U}{dR^2} \Big|_{R_0} (\Delta R)^2 + \dots \quad (4.1)$$

where $\Delta R = R - R_0$ is the deviation from equilibrium.

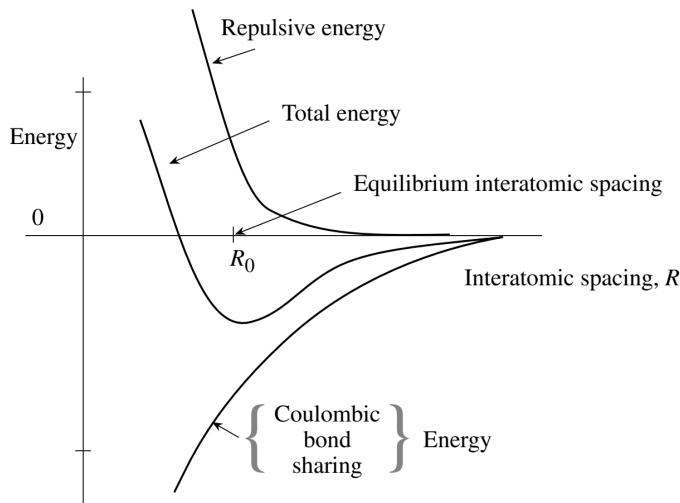


Figure 4.2. General relationship between binding energy and atomic separation in a crystal. In semiconductors, long-range attraction typically arises from electrostatic interactions or covalent bond energy associated with electron sharing.

The second term in the expansion is zero since R_0 is the equilibrium interatomic separation. Retaining terms up to second order in ΔR , we obtain what is known as the *harmonic approximation*:

$$U(R) = U(R_0) + \frac{1}{2} C(\Delta R)^2 \quad (4.2)$$

where the force constant C is given by:

$$C = \frac{\partial^2 U}{\partial R^2} \Big|_{R_0} \quad (4.3)$$

The corresponding restoring force is:

$$F = -C\Delta R \quad (4.4)$$

Due to this restoring force, the atoms in the crystal vibrate similarly to a particle attached to a spring. To analyze such vibrations in semiconductors, consider a diatomic linear lattice (two atoms per basis). The atoms occupy equilibrium positions and vibrate around them. Assuming

nearest-neighbor interactions only, and letting M_1 and M_2 be the atomic masses, and u_s , v_s be the displacements of the two atoms in the s -th unit cell, we write the equations of motion:

$$M_1 \frac{d^2 u_s}{dt^2} = C(v_s + v_{s-1} - 2u_s) \quad (4.5)$$

$$M_2 \frac{d^2 v_s}{dt^2} = C(u_s + u_{s+1} - 2v_s) \quad (4.6)$$

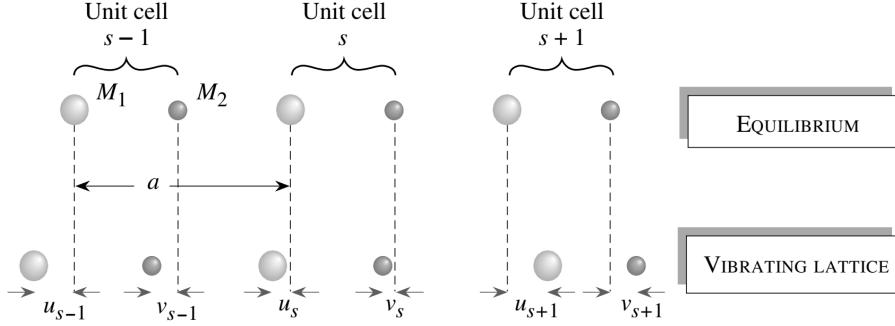


Figure 4.3. Lattice vibrations in a crystal with two atoms per unit cell, having masses M_1 and M_2 , connected by a force constant C between adjacent atomic planes.

We look for traveling wave solutions with alternating amplitudes:

$$u_s = u e^{iska} e^{-i\omega t} \quad (4.7)$$

$$v_s = v e^{iska} e^{-i\omega t} \quad (4.8)$$

Here, a is the distance between identical planes of atoms (i.e., the lattice periodicity). Substituting into the equations of motion gives:

$$-\omega^2 M_1 u = Cv[1 + e^{-ika}] - 2Cu \quad (4.9)$$

$$-\omega^2 M_2 v = Cu[1 + e^{ika}] - 2Cv \quad (4.10)$$

These form a set of coupled eigenvalue equations, which can be represented in matrix form:

$$\begin{pmatrix} 2C - \omega^2 M_1 & -C(1 + e^{-ika}) \\ -C(1 + e^{ika}) & 2C - \omega^2 M_2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0$$

The determinant of this system must vanish:

$$\begin{vmatrix} 2C - \omega^2 M_1 & -C(1 + e^{-ika}) \\ -C(1 + e^{ika}) & 2C - \omega^2 M_2 \end{vmatrix} = 0$$

Expanding and simplifying:

$$M_1 M_2 \omega^4 - 2C(M_1 + M_2)\omega^2 + 2C^2(1 - \cos ka) = 0 \quad (4.11)$$

Solving this quadratic in ω^2 , we obtain:

$$\omega^2 = \frac{2C(M_1 + M_2) \pm \sqrt{4C^2(M_1 + M_2)^2 - 8C^2(1 - \cos ka)M_1 M_2}}{2M_1 M_2} \quad (4.12)$$

Let us consider two limiting cases: - For small k (long-wavelength limit), we obtain:

$$\omega_{\text{acoustic}} \approx \sqrt{\frac{C/2}{M_1 + M_2}} ka \quad (4.13)$$

$$\omega_{\text{optical}} \approx \sqrt{\frac{2C}{\mu}} \quad \text{with } \mu = \frac{M_1 M_2}{M_1 + M_2} \quad (4.14)$$

- At the Brillouin zone boundary $k = \pi/a$, the frequencies are:

$$\omega_1 = \sqrt{\frac{2C}{M_1}} \quad (4.15)$$

$$\omega_2 = \sqrt{\frac{2C}{M_2}} \quad (4.16)$$

This analysis yields two distinct branches of lattice vibrations:

- The *acoustic branch*, where $\omega \rightarrow 0$ as $k \rightarrow 0$, representing sound wave propagation.
- The *optical branch*, where ω remains finite as $k \rightarrow 0$, corresponding to out-of-phase motion of the two atoms.

The acoustic mode has a linear dispersion relation near $k = 0$, and the group velocity (sound speed) is given by:

$$v_s = \left. \frac{d\omega}{dk} \right|_{k \rightarrow 0} = a \sqrt{\frac{C}{M}}, \quad \text{with } M = \frac{M_1 + M_2}{2} \quad (4.17)$$

It is important to examine the eigenfunctions (i.e., the displacement amplitudes u_s) corresponding to the optical and acoustic branches of the dispersion relation. For $k = 0$, in the case of the *optical branch*, the angular frequency is given by:

$$\omega^2 = 2C \left(\frac{1}{M_1} + \frac{1}{M_2} \right) \quad (4.18)$$

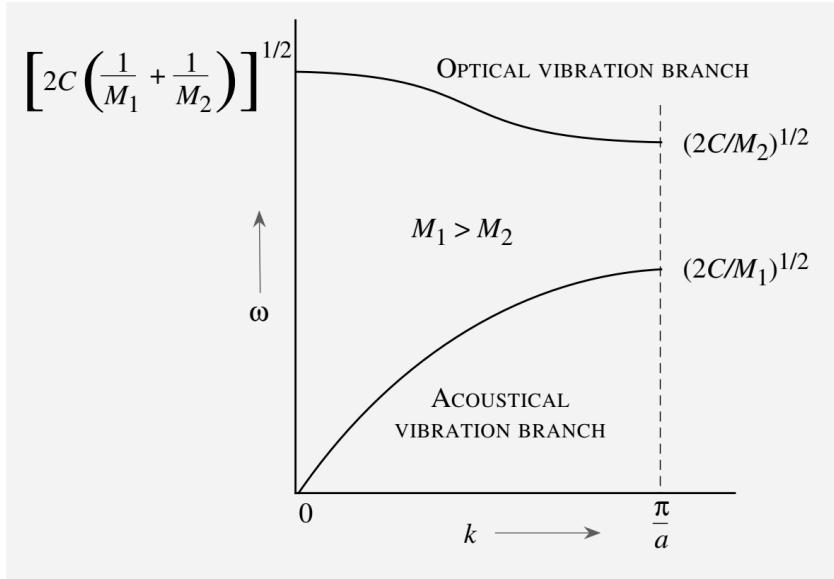


Figure 4.4.

Substituting this value into the equation of motion, we obtain the relation between the amplitudes of the atomic displacements:

$$u = -\frac{M_2}{M_1}v \quad (4.19)$$

In this case, the two atoms vibrate against each other, such that their center of mass remains stationary. This describes an out-of-phase motion, which is characteristic of optical phonons. In contrast, for the *acoustic branch*, in the long-wavelength limit ($k \rightarrow 0$), the solution leads to:

$$u = v \quad (4.20)$$

This corresponds to both atoms moving in phase, i.e., the entire lattice unit oscillates as a whole, which is characteristic of sound wave propagation. Furthermore, for each wavevector k , there exist three possible vibrational modes:

- One **longitudinal mode**, where atomic displacements are parallel to the direction of wave propagation.
- Two **transverse modes**, where atomic displacements are perpendicular to the direction of wave propagation.

In general, these modes have different frequencies because the restoring forces differ for longitudinal and transverse motions.

In ionic crystals such as GaAs, the *optical vibrations* induce oscillating dipole moments due to the displacement of positive and negative ions relative to each other. These polarization fields interact with the vibrations, especially in the longitudinal mode. This interaction introduces an additional long-range restoring force for the longitudinal optical (LO) phonons, increasing their frequency relative to the transverse optical (TO) phonons.

As a result, in such materials, the LO phonon frequency is greater than the TO phonon frequency at the zone center ($k = 0$). This splitting is a fundamental feature of polar optical phonons in ionic semiconductors.

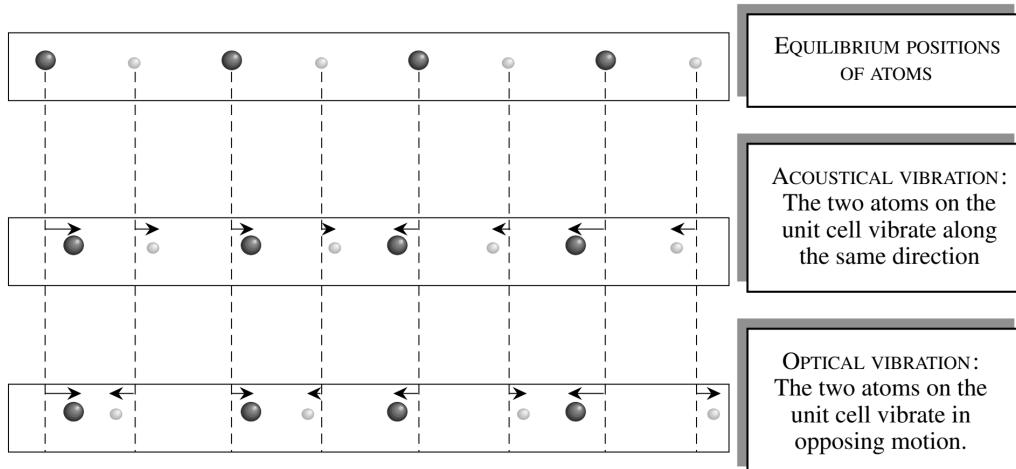


Figure 4.5.Difference between as acoustical mode an optical mode.

4.2.1 Phonon: Quantization of Lattice Vibrations

In the previous discussions, we evaluated the dispersion relation ω vs. k for a set of coupled harmonic oscillator equations. If we now consider a single harmonic oscillator problem in quantum mechanics with the Hamiltonian:

$$H = \frac{P^2}{2m} + \frac{1}{2}Cx^2 \quad (4.21)$$

the energy of the vibrating particle is quantized and given by:

$$\epsilon_n = \left(n + \frac{1}{2}\right)\hbar\omega, \quad n = 0, 1, 2, \dots \quad (4.22)$$

where $\omega = \sqrt{C/m}$ is the classical frequency of oscillation. In classical physics, the oscillator energy can vary continuously with the amplitude of vibration. However, in quantum mechanics, the oscillator has a minimum energy of $\hbar\omega/2$ and its energy increases in discrete steps of $\hbar\omega$. In the framework of second quantization, the integer n represents the number of "particles" or quanta of vibrational energy in the system. These quanta are called *phonons*. A phonon is the

quantized mode of lattice vibration, analogous to how a photon is the quantum of light. For a single oscillator, the frequency ω is fixed. However, in a crystal made up of many coupled atoms (i.e., coupled oscillators), the vibrational frequencies span a range. These are indexed by a wavevector k , leading to a dispersion relation $\omega(k)$. The quantized energy for a given wavevector k is:

$$\epsilon_k = \left(n_{\mathbf{k}} + \frac{1}{2} \right) \hbar \omega_{\mathbf{k}} \quad (4.23)$$

where $n_{\mathbf{k}}$ is the number of phonons in the mode with wavevector \mathbf{k} and frequency $\omega_{\mathbf{k}}$.

To determine how many phonons occupy a given mode, we will later define the appropriate statistical distribution for phonons. The allowed wavevectors k in a crystal with N unit cells and periodic boundary conditions are given by:

$$|\mathbf{k}| = \frac{2\pi n}{Na}, \quad n = 0, \pm 1, \dots, \pm \left(\frac{N}{2} - 1 \right) \quad (4.24)$$

This quantization implies there are a total of $3N$ vibrational modes in the crystal: N longitudinal and $2N$ transverse modes. Each mode can have an integer occupation number n_k determined by phonon statistics.

4.3 Phonon Statistics

We have briefly introduced the concept that lattice vibrations can be quantized in terms of particles known as k -phonons. To determine how many phonons occupy a specific vibrational mode with angular frequency ω at a given temperature T , one must consider the relevant statistical distribution. Phonons, being bosonic in nature, are governed by Bose–Einstein statistics under conditions of thermal equilibrium.

The average phonon number in a mode with frequency ω is expressed as:

$$\langle n_{\omega} \rangle = \frac{1}{\exp\left(\frac{\hbar\omega}{k_B T}\right) - 1} \quad (4.25)$$

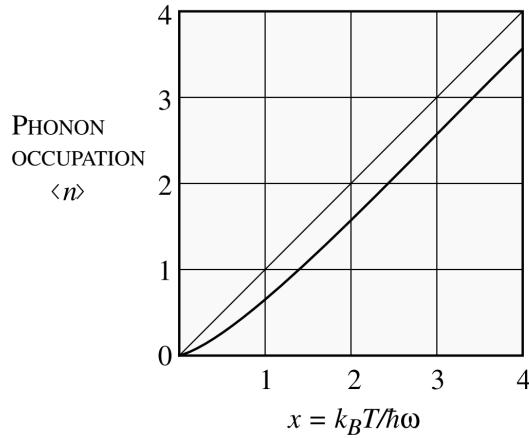


Figure 4.6. Bose–Einstein distribution.

Unlike electrons, whose occupancy follows the Fermi–Dirac distribution and is restricted to at most one particle per state, bosons such as phonons are permitted to share quantum states. Consequently, the occupation number for phonons can exceed unity. As temperature increases, lattice vibrations become more intense, leading to a higher average phonon population, $\langle n_{\omega} \rangle$. It is worth highlighting that at lower temperatures, the occupation of optical phonons remains minimal due to their relatively high energy at any wavevector k . In contrast, acoustic phonons,

which can have very small $\hbar\omega$ values at low k , maintain significant occupancy even at low temperatures. Therefore, optical phonons contribute less significantly at low temperatures. In the limiting case where $\hbar\omega \ll k_B T$, the phonon occupation number simplifies to:

$$\langle n \rangle \approx \frac{k_B T}{\hbar\omega} \quad (4.26)$$

The total energy associated with the lattice vibrations, neglecting the zero-point energy, is given by summing over all wavevectors k and polarizations ρ :

$$U = \sum_{\mathbf{k}, \rho} \langle n_{\mathbf{k}, \rho} \rangle \hbar\omega_{\mathbf{k}, \rho} \quad (4.27)$$

Here, \mathbf{k} represents the phonon wavevector and ρ denotes the polarization index of each vibrational mode.

4.3.1 Conservation Laws in Scattering of Particles involving Phonons

A phonon characterized by a wave vector \mathbf{k} can interact with other particles such as electrons or photons, imparting momentum as though it carries a value $\hbar\mathbf{k}$. It is important to recall that for electrons, the quantity $\hbar\mathbf{k}$ (related to the wave vector) represents the relevant momentum, rather than the actual mechanical momentum of the particle. Although phonons are treated as if they possess momentum, they do not carry any real physical momentum. This is because lattice vibrations involve atoms oscillating about their equilibrium positions in opposite directions, meaning the overall motion of the crystal remains unchanged. Thus, the net physical momentum associated with lattice vibrations is zero.

This can also be demonstrated mathematically. The total momentum of the crystal can be expressed as:

$$\mathbf{P} = M \frac{d}{dt} \sum_s \mathbf{u}_s \quad (4.28)$$

where M is the total mass and \mathbf{u}_s denotes the displacement of atom s . Given the form of the vibrational solution, this summation evaluates to zero, which aligns with the physical expectation that the center of mass of the crystal does not move.

When examining electron-phonon interactions, it is found that first-order phonon scattering processes must obey conservation laws. Specifically, the conservation of crystal momentum and energy leads to the following relations:

$$\mathbf{k}_i = \mathbf{k}_f \pm \mathbf{q} \quad (4.29)$$

$$\mathbf{E}_i = \mathbf{E}_f \pm \hbar\omega_{\mathbf{q}} \quad (4.30)$$

Here, \mathbf{k}_i and \mathbf{k}_f represent the initial and final wave vectors of the electron, \mathbf{q} is the phonon wave vector, and \mathbf{E}_i , \mathbf{E}_f , and $\hbar\omega_{\mathbf{q}}$ correspond to the respective energies involved. For more intricate scattering events involving multiple phonons, these conservation laws are extended accordingly to account for the additional complexity.

4.4 Polar Optical Phonons

In previous discussions concerning optical phonons, we have not accounted for the fact that, in certain semiconductors, the constituent atoms possess net charges—specifically, positive and negative ions such as cations and anions. This ionic characteristic is not present in group IV elemental semiconductors like Si, Ge, or C. However, in compound semiconductors, the ionic nature gives rise to an additional restoring force due to long-range polarization fields that accompany lattice vibrations.

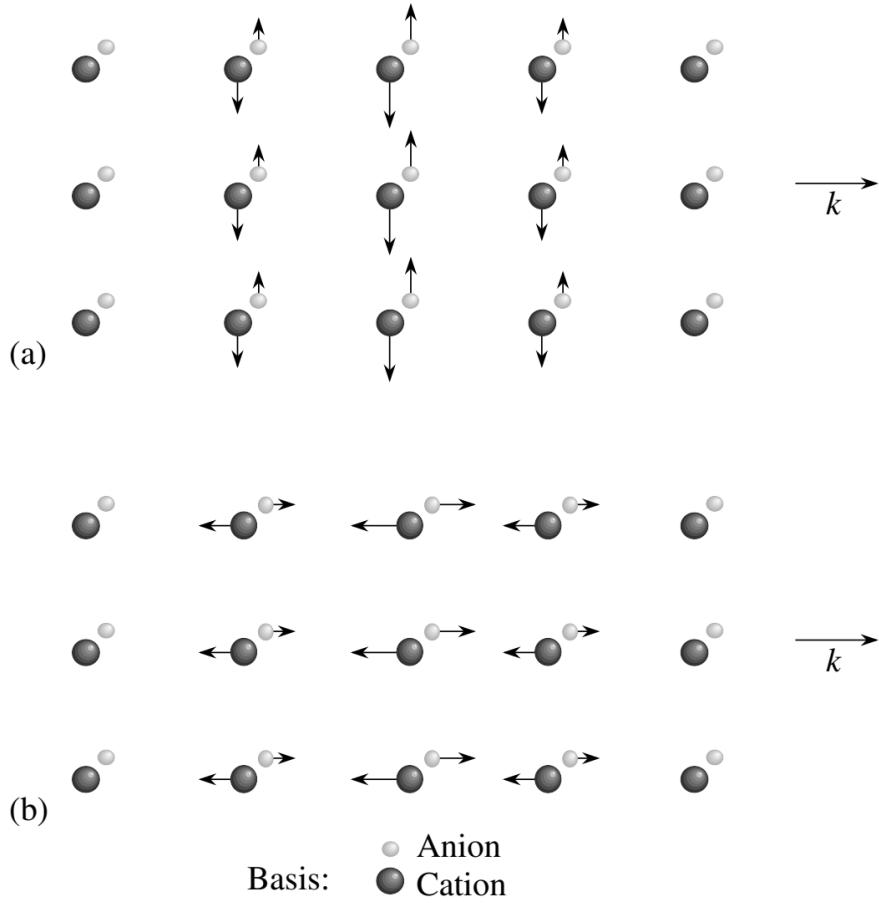


Figure 4.7. Optical vibrations in an ionic crystal.(a) Transverse optical modes do not induce macroscopic polarization. (b) Longitudinal optical modes generate long-range electric fields due to polarization effects.

To examine this phenomenon, we consider two conceptual approaches. The first, although highly simplified, effectively illustrates the essential physical concepts. We work in terms of the relative displacement vector

$$\mathbf{u}_r = \mathbf{u} - \mathbf{v} \quad (4.31)$$

which denotes the relative motion between the sublattices of positively and negatively charged ions. For transverse vibrational modes, which do not generate macroscopic polarization, the equation of motion takes the form

$$M\ddot{\mathbf{u}}_r + M\omega_t^2\mathbf{u}_r = 0 \quad (4.32)$$

where M is the reduced mass of the two atoms, and ω_t is the transverse optical phonon frequency. In contrast, longitudinal vibrational modes induce an electric field as a result of polarization. The equation of motion in this case incorporates both the mechanical restoring force and the force due to the induced electric field, yielding

$$-\omega_l^2 M \mathbf{u}_r = -\omega_t^2 M \mathbf{u}_r + \mathbf{F}_i e^* \quad (4.33)$$

Here, \mathbf{F} represents the internal electric field generated by polarization, and e^* is an effective charge per ion. Assuming n unit cells per unit volume, the polarization is given by

$$\mathbf{P} = n e^* \mathbf{u}_r \quad (4.34)$$

The associated electric field in a vacuum, using ϵ_0 as the permittivity of free space, is

$$\mathbf{F} = -\frac{\mathbf{P}}{\epsilon_0} = -\frac{ne^*\mathbf{u}_r}{\epsilon_0} \quad (4.35)$$

Substituting this into the equation of motion leads to a corrected expression for the longitudinal optical phonon frequency:

$$\omega_l^2 = \omega_t^2 + \frac{ne^{*2}}{\epsilon_0 M} \quad (4.36)$$

If the material possesses a static relative dielectric constant ϵ_r such that $\epsilon = \epsilon_r \epsilon_0$, the equation becomes

$$\omega_l^2 = \omega_t^2 + \frac{n\epsilon_{rel}e^{*2}}{M\epsilon_0} \quad (4.37)$$

The effective charge e^* will be addressed further in the context of polar optical phonon scattering. It should be noted that the longitudinal optical phonon frequency exceeds that of the transverse phonons near the Brillouin zone center ($k \approx 0$). In group IV elemental semiconductors, no such frequency splitting exists at $k = 0$ due to the absence of ionicity. However, in III-V compound semiconductors, the difference becomes significant because of their polar nature.

An important observation is that optical phonons exhibit minimal dispersion near $k = 0$; their energy bands are nearly flat, unlike those of acoustic phonons. Accordingly, in future discussions on electron-phonon scattering, we will assume that optical phonons are dispersionless.

4.5 Phonons in Heterostructures

It is evident that the behavior of phonons bears a strong resemblance to that of electrons, as both phenomena involve solving differential equations within a periodic potential. In the context of heterostructures, the concept of electronic band structure has already been explored, particularly in relation to systems such as quantum wells and superlattices. These same structural configurations exert a notable influence on phonon dispersion, much like their effect on electronic spectra.

The distinction between the electronic and phononic cases, however, lies primarily in the characteristic length scales involved. For electrons, the band offsets in semiconductors give rise to heterostructure effects when the quantum well dimensions approach the electron de Broglie wavelength, typically on the order of 100 Å. In contrast, for phonons, the relevant length scales are considerably shorter—on the order of a few monolayers.

Consequently, there exist phonon modes that are uniquely associated with interfaces and with superlattices of narrow periodicity. These interface-related phonon modes can play a significant role in determining the thermal and vibrational properties of the heterostructure.

4.5.1 Confined Optical Phonons

In superlattices with narrow periodicity, phonon modes can become confined within individual layers, in much the same way that electron wavefunctions are confined within a quantum well. A notable example of this occurs in systems composed of alternating layers of GaAs and AlAs. The optical phonon dispersion relations of these two materials do not intersect, implying that optical modes originating in one material are unable to propagate into the adjacent layer. As a result, these phonon modes are spatially confined to their respective layers.

This confinement leads to a situation analogous to a particle in an infinite square potential well. The permitted wave vectors for such confined modes are quantized according to the relation

$$|k_n| = \frac{n}{2d}, \quad \text{for } n = 1, 2, 3, \dots \quad (4.38)$$

where d denotes the thickness of the individual GaAs or AlAs layer. Remarkably, the energies of these confined phonon modes have been observed to closely follow the dispersion curves of the corresponding bulk materials. That is, the function $E(k)$ remains nearly unchanged whether evaluated at the continuous k values characteristic of bulk GaAs or at the discrete k values dictated by quantization within the confined layer.

4.5.2 Interface Phonons

In semiconductor heterostructures, there also exist vibrational modes within the optical phonon frequency range that are characterized by having their maximum amplitude localized at the interfaces between two distinct materials. These so-called interface modes are of particular interest because they are accompanied by macroscopic electric fields. Such modes can be effectively understood using a simplified electrostatic model. The discontinuity in material properties across the interface—such as dielectric constants and ionic character—gives rise to polarization effects, which in turn generate electric fields associated with the vibrational modes.

4.6 Phonon Scattering: General Formalism

In our treatment of electronic bandstructure, we have assumed a static, time-independent periodic potential as the background. This assumption enables the derivation of electronic bands. In contrast, lattice vibrations introduce a time-dependent aspect to the crystal potential, leading to the quantized vibrational excitations known as phonons and their associated dispersion relations. The possibility of separating, at first order, the electronic behavior from lattice dynamics relies on the validity of the adiabatic approximation. This approximation is applicable to systems where the time dependence of the Hamiltonian evolves slowly enough that, at each instant, the system can be regarded as effectively time-independent. While phonon frequencies fall in the terahertz range—on the order of 10^{12} Hz—this still qualifies as “slow” compared to electronic time scales due to the large difference in mass between electrons and nuclei. As such, the adiabatic approximation proves to be highly effective, allowing us to proceed in two distinct stages:

1. Determine the electronic states in a perfect lattice, yielding the electronic bandstructure.
2. Analyze the interactions between electrons and lattice vibrations, or phonon scattering.

In previous sections, it has been shown that in semiconductors with two atoms per basis, both acoustic and optical phonon modes arise. Near the Brillouin zone center ($k = 0$), acoustic phonons exhibit a linear relationship between ω and k , while optical phonons are nearly dispersionless. Each of these types of phonons exists in both longitudinal and transverse polarization modes. The motion of atoms within the lattice produces strain, which, according to deformation potential theory, leads to perturbations in the electronic energy levels. This relationship has already been explored in the context of strain-induced modifications to bandstructure.

Generally, the interaction between electrons and phonons depends on the atomic displacement \mathbf{u} . The resulting perturbation potential U takes different forms depending on the type of phonon:

$$\begin{aligned} \text{Acoustic phonons: } U &\sim D \frac{\partial u}{\partial x} \\ \text{Optical phonons (non-polar): } U &\sim D_0 u \\ \text{Polar optical phonons: } U &\sim e^* u \end{aligned}$$

Here, e^* is an effective charge per ion.

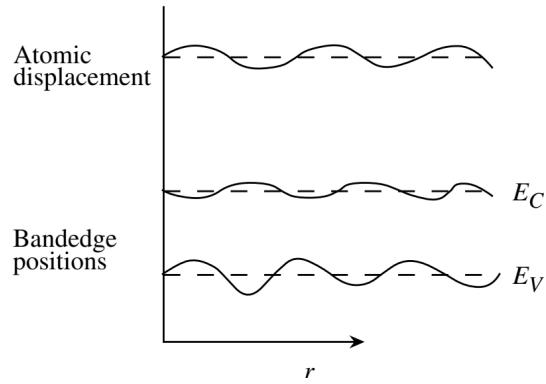


Figure 4.8. Effect of atomic displacements on bandedge energy levels in real space.

To describe the scattering of electrons by lattice vibrations, we invoke the framework of second quantization. In this formalism, classical vibrational modes are quantized and described by phonons—quanta of vibrational energy. A harmonic oscillator with frequency ω has quantized energy levels given by

$$E = \left(n + \frac{1}{2} \right) \hbar\omega \quad (4.39)$$

where n is the number of phonon quanta. The phonon number operator is defined as $n = a^\dagger a$, with a^\dagger and a being the phonon creation and annihilation operators, respectively. These operators satisfy the following relations:

$$a|n\rangle = \sqrt{n}|n-1\rangle, \quad a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle \quad (4.40)$$

The displacement and momentum operators in terms of a and a^\dagger are:

$$u = \sqrt{\frac{\hbar}{2M\omega}}(a + a^\dagger), \quad p = i\sqrt{\frac{M\hbar\omega}{2}}(a^\dagger - a)$$

These expressions allow us to understand electron-phonon interactions in terms of phonon creation (emission) and annihilation (absorption) processes. In the quantum description, the emission of a phonon corresponds to the application of the creation operator, while absorption corresponds to the action of the annihilation operator.

Since the lattice vibrational system consists of many coupled oscillators, the physical displacement \mathbf{u} representing atomic motion must be expressed in terms of normal coordinates—modes in which the vibrations are uncoupled. This transformation enables each mode to be treated as an independent harmonic oscillator. The displacement field can thus be written as

$$\mathbf{u} = \frac{1}{\sqrt{N}} \sum_{\mathbf{q}} [\theta_{\mathbf{q}b} \mathbf{b}_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{R}} + c.c.] \quad (4.41)$$

where q is the phonon wavevector, b indexes the phonon branches, θ_b is the normal mode amplitude, and N is the number of unit cells. This formalism provides the foundation for analyzing phonon-mediated scattering processes in semiconductors.

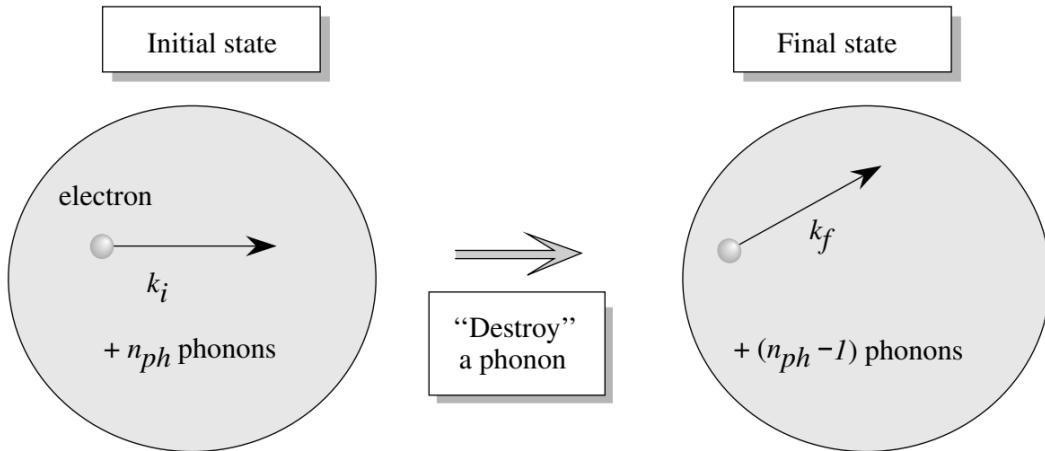
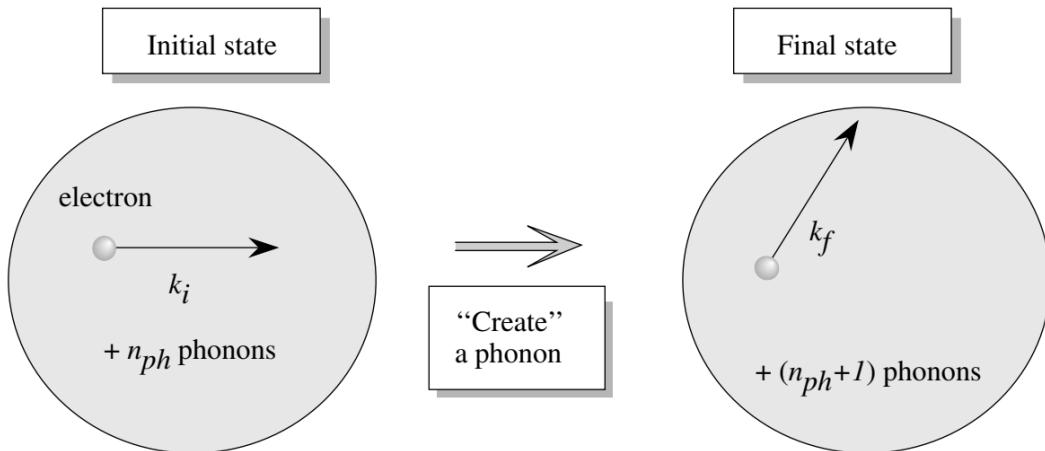
ABSORPTION:**EMISSION:**

Figure 4.9. (a) Schematic of a phonon absorption process, where a phonon is annihilated and the electron's energy and momentum change accordingly. (b) Schematic of a phonon emission event, where a phonon is generated as the electron loses energy.

The two fundamental phonon-related processes involved in electron–phonon interactions correspond to the annihilation and creation of phonons. These are commonly referred to as **phonon absorption** and **phonon emission**, respectively.

In the quantum mechanical treatment of lattice vibrations, the interaction Hamiltonian includes two terms within the brackets. The first term represents the removal of a phonon from the initial state during the scattering process—this corresponds to **phonon absorption**. The second term accounts for the addition of a phonon to the final state—this is the **phonon emission** process. A key distinction between the two processes arises from the phonon occupation number. Phonon absorption requires that a phonon be present initially; thus, if the initial phonon population is zero, the probability of absorption vanishes. In contrast, the emission process involves a factor of $(n + 1)$, meaning that emission can occur even in the absence of initial phonons, due to spontaneous emission.

At thermal equilibrium, the average phonon occupation number follows Bose–Einstein statistics.

Optical Properties of Semiconductor

5.1 Maxwell Equations and Vector Potential

The properties of electromagnetic fields in a medium are described by the four Maxwell equations. Apart from the electric field \mathbf{F} , magnetic field \mathbf{B} , and velocity of light, the effects of the material are represented by the dielectric constant ϵ , permeability μ (assumed $\mu = \mu_0$), and electrical conductivity. We start with the four Maxwell equations:

$$\nabla \times \mathbf{F} + \frac{\partial \mathbf{B}}{\partial t} = 0, \quad \nabla \times \mathbf{H} - \frac{\partial \mathbf{D}}{\partial t} = \mathbf{J}, \quad \nabla \cdot \mathbf{D} = \rho, \quad \nabla \cdot \mathbf{B} = 0 \quad (5.1)$$

Here, $\mathbf{D} = \epsilon \mathbf{F}$, $\mathbf{B} = \mu \mathbf{H}$, and \mathbf{J} , ρ are current and charge densities, respectively. In electron-photon interactions, it is convenient to use the scalar and vector potentials ϕ and \mathbf{A} , defined by:

$$\mathbf{F} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla \phi, \quad \mathbf{B} = \nabla \times \mathbf{A} \quad (5.2)$$

These definitions automatically satisfy the first and fourth Maxwell equations. The potentials ϕ and \mathbf{A} are not unique, and can be transformed using the gauge transformation:

$$\mathbf{A}' = \mathbf{A} + \nabla \chi, \quad \phi' = \phi - \frac{\partial \chi}{\partial t} \quad (5.3)$$

These transformations do not affect the physical fields \mathbf{F} and \mathbf{B} .

Rewriting the second and third Maxwell equations in terms of ϕ and \mathbf{A} , we get:

$$\frac{1}{\mu_0} \nabla \times (\nabla \times \mathbf{A}) + \epsilon \frac{\partial^2 \mathbf{A}}{\partial t^2} + \epsilon \nabla \left(\frac{\partial \phi}{\partial t} \right) = \mathbf{J} \quad (5.4)$$

$$\frac{\partial}{\partial t} (\nabla \cdot \mathbf{A}) + \nabla^2 \phi = -\frac{\rho}{\epsilon} \quad (5.5)$$

Using the identity:

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla (\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (5.6)$$

We obtain:

$$-\nabla^2 \mathbf{A} + \epsilon \mu_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} + \nabla \left(\nabla \cdot \mathbf{A} + \epsilon \frac{\partial \phi}{\partial t} \right) = \mu_0 \mathbf{J} \quad (5.7)$$

We now impose the Lorentz gauge condition:

$$\nabla \cdot \mathbf{A} + \epsilon \frac{\partial \phi}{\partial t} = 0 \quad (5.8)$$

This simplifies the Maxwell equations to:

$$\nabla^2 \mathbf{A} - \epsilon\mu_0 \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu_0 \mathbf{J}, \quad \nabla^2 \phi - \epsilon\mu_0 \frac{\partial^2 \phi}{\partial t^2} = -\frac{\rho}{\epsilon} \quad (5.9)$$

This form is especially useful for generalization to relativistic electrodynamics. In the **Coulomb (radiation) gauge**, where $\mathbf{J} = 0$ and $\rho = 0$, we may set $\phi = 0$ and $\nabla \cdot \mathbf{A} = 0$. The vector potential becomes a solution of the wave equation, typically:

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 \left[e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)} + \text{c.c.} \right] \quad (5.10)$$

Where the wavevector \mathbf{k} and angular frequency ω satisfy:

$$|\mathbf{k}|^2 = \epsilon\mu_0\omega^2 \quad (5.11)$$

The corresponding electric and magnetic fields are:

$$\mathbf{F} = -\frac{\partial \mathbf{A}}{\partial t} = 2\omega \mathbf{A}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t) \quad (5.12)$$

$$\mathbf{B} = \nabla \times \mathbf{A} = -2\mathbf{k} \times \mathbf{A}_0 \sin(\mathbf{k} \cdot \mathbf{r} - \omega t) \quad (5.13)$$

The Poynting vector representing the optical power is:

$$\mathbf{S} = (\mathbf{F} \times \mathbf{H}) \quad (5.14)$$

$$= \frac{4vk^2}{\mu_0} |A_0|^2 \sin^2(\mathbf{k} \cdot \mathbf{r} - \omega t) \hat{k} \quad (5.15)$$

The average power carried by the electromagnetic wave in the medium, where $v = \frac{c}{\sqrt{\epsilon}}$ is the speed of light in the material and \hat{k} is the propagation direction unit vector, can be expressed as:

$$\langle \mathbf{S} \rangle_{\text{time}} = \hat{k} \frac{2vk^2 |\mathbf{A}_0|^2}{\mu_0} \quad (5.16)$$

By rewriting the wave vector magnitude $|\mathbf{k}|$ in terms of the angular frequency and the wave velocity, i.e.,

$$|\mathbf{k}| = \frac{\omega}{v} \quad (5.17)$$

we can substitute into the expression for the time-averaged Poynting vector to obtain:

$$\langle \mathbf{S} \rangle_{\text{time}} = 2v\epsilon\mu_0\omega^2 |\mathbf{A}_0|^2 \hat{k} \quad (5.18)$$

The energy stored per unit volume in the electromagnetic field can be related to the Poynting vector and the wave velocity as

$$\left| \frac{S}{v} \right| = \frac{2\epsilon\omega^2 |\mathbf{A}_0|^2}{c^2} \quad (5.19)$$

Assuming a photon population of n_{ph} , the energy density in a volume V becomes

$$\frac{n_{\text{ph}}\hbar\omega}{V} \quad (5.20)$$

Equating the two energy densities, we can isolate the amplitude of the vector potential:

$$|\mathbf{A}_0|^2 = \frac{n_{\text{ph}}\hbar}{2\epsilon\omega V} \quad (5.21)$$

With these foundational relations established, we now consider wave propagation in a medium. By substituting $\mathbf{J} = \sigma\mathbf{F}$ into Maxwell's equations and eliminating the magnetic field, we obtain the wave equation for the electric field:

$$\nabla^2\mathbf{F} = \epsilon\mu_0\frac{\partial^2\mathbf{F}}{\partial t^2} + \sigma\mu_0\frac{\partial\mathbf{F}}{\partial t} \quad (5.22)$$

This equation describes a damped wave. A general solution takes the form

$$\mathbf{F} = \mathbf{F}_0 \exp\{i(\mathbf{k} \cdot \mathbf{r} - \omega t)\} \quad (5.23)$$

Substituting this into the wave equation yields the relation for the wavevector magnitude:

$$-k^2 = -\epsilon\mu_0\omega^2 - \sigma\mu_0i\omega \quad (5.24)$$

Or more compactly,

$$k = \frac{\omega}{c} \left(\tilde{\epsilon} + \frac{\sigma\mu_0 i}{\omega} \right)^{1/2} \quad (5.25)$$

Here, $\tilde{\epsilon}$ represents the relative dielectric constant. In vacuum, where $\sigma = 0$ and $\tilde{\epsilon} = 1$, the wavevector simplifies to

$$k = \frac{\omega}{c} \quad (5.26)$$

Within a material, the propagation speed is affected by a complex refractive index n_r , defined by

$$n_r = \left(\tilde{\epsilon} + \frac{\sigma\mu_0 i}{\omega} \right)^{1/2} \quad (5.27)$$

We can express n_r in terms of its real and imaginary parts as

$$n_r = n'_r + i n''_r \quad (5.28)$$

Substituting this into the expression for k , we obtain

$$k = \frac{n'_r \omega}{c} + i \frac{n''_r \omega}{c} \quad (5.29)$$

The electric field wave for propagation in the $+z$ direction can be written as

$$\mathbf{F} = \mathbf{F}_0 \exp \left\{ i\omega \left(\frac{n'_r z}{c} - t \right) \right\} \exp \left(\frac{-n''_r \omega z}{c} \right) \quad (5.30)$$

Here, the phase velocity is reduced by a factor n'_r , making it c/n'_r , and the wave amplitude decays exponentially as $\exp(-2\pi n''_r / n'_r)$ per wavelength. This decay in amplitude is a result of absorption of the electromagnetic energy.

The absorption coefficient α , which characterizes the intensity loss, is given by

$$\alpha = \frac{2n''_r \omega}{c} \quad (5.31)$$

In the special case where there is no absorption, the real part of the refractive index n'_r equals the total refractive index n_r , and $n''_r = 0$.

The absorption coefficient is measurable and provides direct information about the imaginary component n''_r of the refractive index.

The interaction of the medium with the electromagnetic field is fully described by a complex dielectric constant or equivalently a complex refractive index. The real and imaginary parts of the dielectric constant $\tilde{\epsilon}$ are related to the refractive index components as follows:

$$\tilde{\epsilon}_1 = n'^2_r - n''^2_r, \quad \tilde{\epsilon}_2 = 2n'_r n''_r \quad (5.32)$$

These expressions relate the real ($\tilde{\epsilon}_1$) and imaginary ($\tilde{\epsilon}_2$) parts of the relative permittivity to the components of the complex refractive index.

A very significant relationship exists between the real and imaginary parts of the refractive index based on causality—this is known as the Kramers–Kronig relation. It connects the frequency-dependent refractive index to the absorption coefficient via

$$n(\omega_0) - 1 = \frac{c}{\pi} \mathcal{P} \int_0^\infty \frac{\alpha(\omega)}{\omega^2 - \omega_0^2} d\omega \quad (5.33)$$

This formulation is powerful as it enables the calculation of the refractive index if the absorption coefficient is known. In cases where the absorption spectrum varies under external influences (e.g., electric fields), changes in the refractive index can be predicted accordingly.

5.2 Electrons in Electromagnetic Field

Electrons, possessing a negative charge, interact with electric and magnetic fields originating from electromagnetic radiation. The electric field exerts a force equal to the product of the charge and the field itself, whereas the magnetic field contributes through the Lorentz force, which involves the cross product between the electron's velocity and the magnetic field.

This interaction energy defines the interaction Hamiltonian, which plays a central role in electron scattering processes. The total Hamiltonian for a charge e in the presence of an electromagnetic field is

$$H = \frac{1}{2m_0}(\mathbf{p} - e\mathbf{A})^2 + e\phi + V(\mathbf{r}) \quad (5.34)$$

Expanding the square and rearranging terms:

$$H = \frac{\mathbf{p}^2}{2m_0} - \frac{e}{2m_0}(\mathbf{p} \cdot \mathbf{A} + \mathbf{A} \cdot \mathbf{p}) + \frac{e^2}{2m_0}\mathbf{A}^2 + e\phi + V(\mathbf{r}) \quad (5.35)$$

In this expression, \mathbf{A} denotes the vector potential and $V(\mathbf{r})$ represents any potential due to the crystal lattice.

Within the framework of quantum mechanics, the momentum operator \mathbf{p} is defined differentially. Therefore, the operator form of $\mathbf{p} \cdot \mathbf{A}$ becomes

$$\frac{e}{2m_0}\mathbf{p} \cdot \mathbf{A} = \frac{e}{2m_0}\mathbf{A} \cdot \mathbf{p} - \frac{ie\hbar}{2m_0}\nabla \cdot \mathbf{A} \quad (5.36)$$

Substituting this result into the Hamiltonian yields

$$H = \frac{\mathbf{p}^2}{2m_0} - \frac{e}{m_0}\mathbf{A} \cdot \mathbf{p} + \frac{ie\hbar}{2m_0}\nabla \cdot \mathbf{A} + \frac{e^2}{2m_0}\mathbf{A}^2 + e\phi + V(\mathbf{r}) \quad (5.37)$$

To analyze the influence of electromagnetic radiation on an electron, we apply perturbation theory. In the quantum mechanical description of radiation, the electromagnetic field is represented using creation and annihilation operators, akin to those used in harmonic oscillator problems—similar to the formalism previously applied to phonon interactions.

The time-dependent Schrödinger equation describing this system is then:

$$i\hbar \frac{\partial \psi}{\partial t} = \left[-\frac{\hbar^2}{2m_0}\nabla^2 + \frac{ie\hbar}{m_0c}\mathbf{A} \cdot \nabla + \frac{ie\hbar}{2m_0}(\nabla \cdot \mathbf{A}) + \frac{e^2}{2m_0}\mathbf{A}^2 + e\phi + V(\mathbf{r}) \right] \psi \quad (5.38)$$

We adopt the radiation gauge, where $\nabla \cdot \mathbf{A} = \phi = 0$, and use time-dependent perturbation theory to compute scattering rates using Fermi's golden rule. Assuming the optical power is modest, the vector potential \mathbf{A} remains small, so we can estimate:

$$\left| \frac{ie\hbar}{m_0}\mathbf{A} \cdot \nabla \right| : \left| \frac{\hbar^2}{2m_0}\nabla^2 \right| \approx \left| \frac{e^2 A^2}{2m_0^2} \right| : \left| \frac{ie\hbar}{m_0}\mathbf{A} \cdot \nabla \right| \approx \frac{eA}{p} \quad (5.39)$$

For an optical intensity of 1 W/cm^2 , the corresponding photon density at 1 eV energy is approximately 10^9 cm^{-3} . With an electron velocity of 10^6 cm/s , we find:

$$\frac{eA}{p} \sim 10^{-5} \quad (5.40)$$

This implies that even for a beam as intense as 1 MW/cm^2 , the quantity eA/p remains sufficiently small for perturbation theory to be applicable. Therefore, we will only consider terms linear in \mathbf{A} .

The time-dependent Schrödinger equation now simplifies to:

$$i\hbar \frac{\partial \psi}{\partial t} = (H_0 + H')\psi \quad (5.41)$$

where the unperturbed Hamiltonian is:

$$H_0 = -\frac{\hbar^2}{2m_0} \nabla^2 + V(\mathbf{r}) \quad (5.42)$$

and the perturbation is given by:

$$H' = \frac{ie\hbar}{m_0} \mathbf{A} \cdot \nabla \quad (5.43)$$

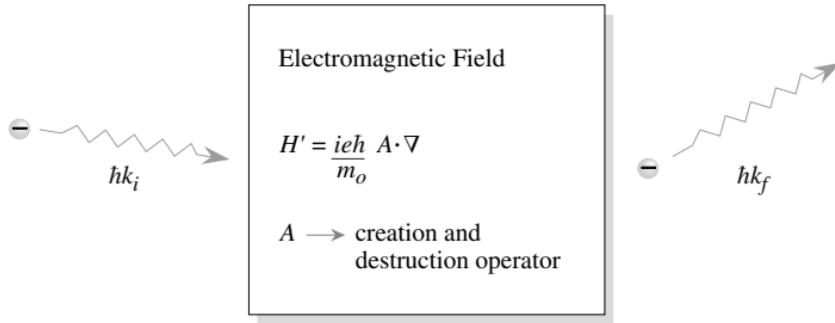


Figure 5.1. Schematic of the scattering process of an electron by the electromagnetic field.

The scattering process arises from this perturbation, where the electromagnetic field induces transitions in the electron states. In the second quantization framework, \mathbf{A} is quantized using photon creation and annihilation operators b^\dagger and b , analogous to the harmonic oscillator treatment.

According to Fermi's golden rule, the transition rate from an initial state $|i\rangle$ to a final state $|f\rangle$ is:

$$W(i) = \frac{2\pi}{\hbar} \sum_f |\langle f | H' | i \rangle|^2 \delta(E_f - E_i \mp \hbar\omega) \quad (5.44)$$

Here, the upper sign corresponds to photon absorption and the lower sign to emission.

The initial and final states are described by the electron's momentum and the photon number.

In the case of photon **absorption**, the states are:

$$\begin{aligned} |i\rangle &= |k_i, n_{\text{ph}}\rangle \\ |f\rangle &= |k_f, n_{\text{ph}} - 1\rangle \end{aligned}$$

Emission:

$$\begin{aligned} |i\rangle &= |k_i, n_{\text{ph}}\rangle \\ |f\rangle &= |k_f, n_{\text{ph}} + 1\rangle \end{aligned}$$

Here, k_i and k_f denote the initial and final wavevectors of the electron, respectively, and n_{ph} is the initial photon occupation number. The vector potential A_0 is defined as:

$$A_0 = \sqrt{\frac{\hbar}{2\omega\epsilon V}} (b^\dagger + b) \quad (5.45)$$

In this expression, b^\dagger and b are the photon creation and annihilation operators. This form is chosen to be consistent with the magnitude $|A_0|^2$ used previously.

When calculating matrix elements, we consider the photon operators separately. Let \mathbf{a} represent the polarization unit vector. Focusing on the matrix element due to the photon part only:
Absorption:

$$\begin{aligned} \langle f | \mathbf{A} \cdot \nabla | i \rangle &\Rightarrow \sqrt{\frac{\hbar}{2\omega\epsilon}} \langle k_f, n_{ph} - 1 | (b^\dagger + b) \mathbf{a} \cdot \nabla | k_i, n_{ph} \rangle \\ &= \sqrt{\frac{\hbar}{2\omega\epsilon}} \sqrt{n_{ph}} \langle k_f | \mathbf{a} \cdot \nabla | k_i \rangle \end{aligned}$$

Emission:

$$\begin{aligned} \langle f | \mathbf{A} \cdot \nabla | i \rangle &\Rightarrow \sqrt{\frac{\hbar}{2\omega\epsilon}} \langle k_f, n_{ph} + 1 | (b^\dagger + b) \mathbf{a} \cdot \nabla | k_i, n_{ph} \rangle \\ &= \sqrt{\frac{\hbar}{2\omega\epsilon}} \sqrt{n_{ph} + 1} \langle k_f | \mathbf{a} \cdot \nabla | k_i \rangle \end{aligned}$$

It is crucial to notice the difference in prefactors: $\sqrt{n_{ph}}$ appears for absorption, while $\sqrt{n_{ph} + 1}$ appears for emission. This distinction will play a significant role in later discussions. To proceed with calculating the rate for photon absorption—where a photon of energy $\hbar\omega$ and momentum $\hbar k_{ph}$ is absorbed by an electron—we will need to sum over all possible electron states which can allow such a process to occur.

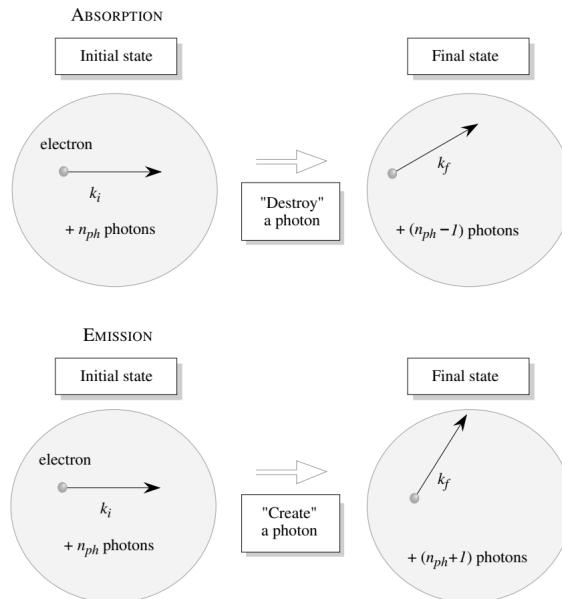


Figure 5.2. (a) Schematic of a photon absorption process, where a photon is annihilated and the electron's energy and momentum change. (b) Photon emission process, where a photon is generated as the electron transitions to a lower energy state.

The process of photon absorption involves the transfer of photon energy $\hbar\omega$ to an electron. To compute the full scattering rate, we integrate over the final electron density of states:

$$W_{\text{abs}} = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \left(\frac{\hbar n_{\text{ph}}}{2\omega\epsilon} \right) \sum_{\text{final states}} \left| \int \psi_f^*(\mathbf{a} \cdot \mathbf{p}) e^{i\mathbf{k}_{\text{ph}} \cdot \mathbf{r}} \psi_i d^3 r \right|^2 \delta(E_i - E_f + \hbar\omega) \quad (5.46)$$

This expression includes a sum over all possible electronic final states for photon absorption, ensuring energy conservation. Alternatively, one might be interested in electron scattering from an initial state with momentum k_i to a final state with momentum k_f , summing instead over photon states that enable this transition.

For photon emission, where an electron in state $\hbar k_i$ emits a photon and transitions to $\hbar k_f$, the corresponding rate is:

$$W_{\text{em}} = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \left(\frac{\hbar(n_{\text{ph}} + 1)}{2\omega\epsilon} \right) \sum_{\text{final states}} \left| \int \psi_f^*(\mathbf{a} \cdot \mathbf{p}) e^{-i\mathbf{k}_{\text{ph}} \cdot \mathbf{r}} \psi_i d^3 r \right|^2 \delta(E_i - E_f - \hbar\omega) \quad (5.47)$$

This emission rate can be broken into two components: one for stimulated emission and one for spontaneous emission:

$$W_{\text{em}} = W_{\text{st}} + W_{\text{spon}} \quad (5.48)$$

where

$$W_{\text{st}} = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar n_{\text{ph}}}{2\omega\epsilon} \sum_{\text{final states}} \left| \int \psi_f^* e^{-i\mathbf{k}_{\text{ph}} \cdot \mathbf{r}} (\mathbf{a} \cdot \mathbf{p}) \psi_i d^3 r \right|^2 \delta(E_i - E_f - \hbar\omega) \quad (5.49)$$

$$W_{\text{spon}} = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar}{2\omega\epsilon} \sum_{\text{final states}} \left| \int \psi_f^* e^{-i\mathbf{k}_{\text{ph}} \cdot \mathbf{r}} \psi_i d^3 r \right|^2 \delta(E_i - E_f - \hbar\omega) \quad (5.50)$$

Stimulated emission arises from the presence of photons already in the system, and the emitted photons are phase coherent with the initial ones. *Spontaneous emission*, on the other hand, originates from quantum vacuum fluctuations ($n_{\text{ph}} = 0$) and results in photons that are phase-incoherent.

This distinction is fundamental in understanding the operational differences between devices such as LEDs and laser diodes.

When examining semiconductor states, we observe that the photon momentum $\hbar k_{\text{ph}}$, for typical photon energies between 0.1 and 2.0 eV, is much smaller than the electron momentum. Therefore, momentum conservation requires that:

$$k_i = k_f \quad (5.51)$$

In first-order perturbation theory, transitions triggered by photon absorption or emission are represented as "vertical" processes in the energy–momentum (E - k) diagram. This is particularly evident in interband transitions. When the photon momentum k_{ph} is small enough to be neglected, the approximation is referred to as the dipole approximation. Under this approximation, the momentum matrix element (the integral within the absolute value bars in the relevant transition rate equations) simplifies significantly. We denote this matrix element as p_{if} .

Let us now consider the case where both the initial and final states follow the Bloch form. In the dipole approximation, the momentum matrix element becomes:

$$p_{if} = -i\hbar \int \psi_{\mathbf{k}_f \ell'}^* \nabla \psi_{\mathbf{k}_i \ell} d^3 r \quad (5.52)$$

We define the initial and final Bloch states as follows:

$$\begin{aligned} |i\rangle &= \psi_{\mathbf{k}_i, \ell} = e^{i\mathbf{k}_i \cdot \mathbf{r}} u_{\mathbf{k}_i, \ell}, \\ |f\rangle &= \psi_{\mathbf{k}_f, \ell'} = e^{i\mathbf{k}_f \cdot \mathbf{r}} u_{\mathbf{k}_f, \ell'}, \end{aligned}$$

where $u_{\mathbf{k}\ell}$ denotes the cell-periodic part of the Bloch wavefunction, and ℓ, ℓ' are band indices. By applying the gradient operator to the initial state, the momentum matrix element simplifies into the following form:

$$p_{if} = \hbar \mathbf{k}_i \int \psi_{\mathbf{k}_f \ell'}^* \psi_{\mathbf{k}_i \ell} d^3r - i\hbar \int u_{\mathbf{k}_f \ell'}^* (\nabla u_{\mathbf{k}_i \ell}) e^{i(\mathbf{k}_i - \mathbf{k}_f) \cdot \mathbf{r}} d^3r \quad (5.53)$$

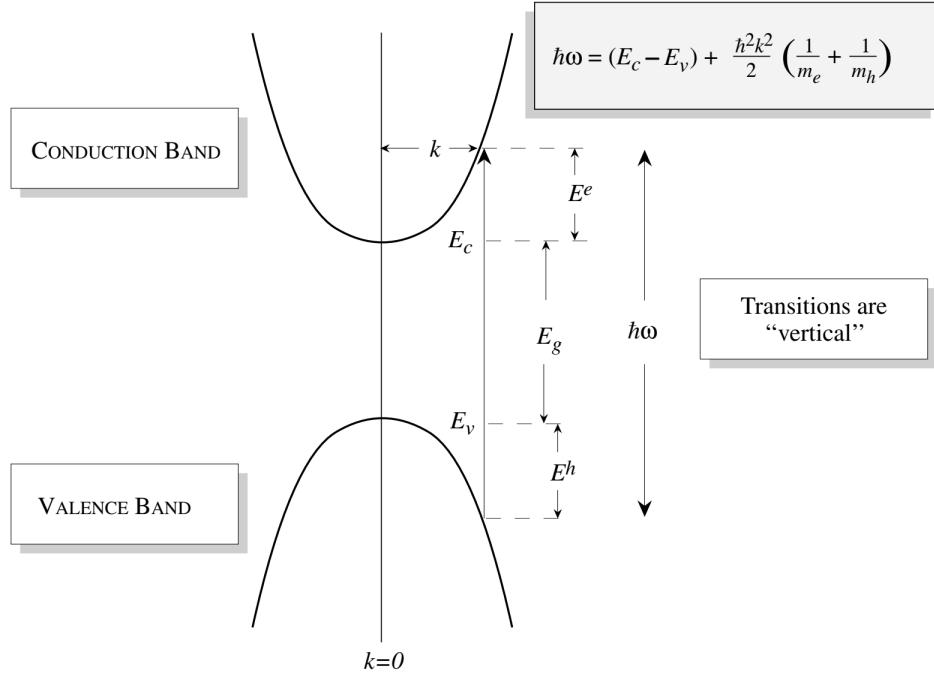


Figure 5.3: Electron and hole energy levels at vertical \mathbf{k} -values. The transition energies are set by the photon energy and the effective masses of the carriers. Due to the negligible photon momentum, these transitions occur vertically in \mathbf{k} -space.

5.3 Interband Transitions

5.3.1 Interband Transitions in Bulk Semiconductors

We now examine the selection rules governing band-to-band transitions in direct-gap semiconductors. For instance, zincblende materials such as GaAs, InAs, and InP. We use the properties of the conduction and valence band central cell functions to simplify the momentum matrix elements. The first term in the momentum matrix expression vanishes due to orthogonality of Bloch states, and we are left with a nonzero contribution only when:

$$\mathbf{k}_i - \mathbf{k}_f = 0 \quad (5.54)$$

This condition implies that the transitions are “vertical” in \mathbf{k} -space, and the matrix element for interband transitions reduces to:

$$\langle u_{c\mathbf{k}} | \mathbf{p}_a | u_{v\mathbf{k}} \rangle \quad (5.55)$$

where $u_{c\mathbf{k}}$ and $u_{v\mathbf{k}}$ represent the central cell wavefunctions for the conduction and valence bands, respectively.

Near the band edge, we approximate $u_{c\mathbf{k}}$ and $u_{v\mathbf{k}}$ by their values at the Brillouin zone center. Specifically:

$$u_{c0} = |s\rangle \quad (5.56)$$

where $|s\rangle$ denotes a spherically symmetric conduction band state.

The valence band states are given by:

$$\textbf{Heavy Hole states: } |3/2, 3/2\rangle = \frac{-1}{\sqrt{2}}(|p_x\rangle + i|p_y\rangle)|\uparrow\rangle \quad (5.57)$$

$$|3/2, -3/2\rangle = \frac{1}{\sqrt{2}}(|p_x\rangle - i|p_y\rangle)|\downarrow\rangle \quad (5.58)$$

$$\textbf{Light Hole states: } |3/2, 1/2\rangle = \frac{-1}{\sqrt{6}}[|p_x\rangle + i|p_y\rangle)|\downarrow\rangle - 2|p_z\rangle|\uparrow\rangle] \quad (5.59)$$

$$|3/2, -1/2\rangle = \frac{1}{\sqrt{6}}[|p_x\rangle - i|p_y\rangle)|\uparrow\rangle + 2|p_z\rangle|\downarrow\rangle] \quad (5.60)$$

By symmetry, only matrix elements of the form:

$$\langle p_x|p_x|s\rangle = \langle p_y|p_y|s\rangle = \langle p_z|p_z|s\rangle = p_{cv} \quad (5.61)$$

are nonzero. Hence, for band-to-band transitions, the allowed nonzero matrix elements are:

$$\langle \text{HH}|p_x|s\rangle = \langle \text{HH}|p_y|s\rangle = \frac{1}{\sqrt{2}}\langle p_x|p_x|s\rangle \quad (5.62)$$

$$\langle \text{LH}|p_x|s\rangle = \langle \text{LH}|p_y|s\rangle = \frac{1}{\sqrt{6}}\langle p_x|p_x|s\rangle \quad (5.63)$$

Additionally, we have:

$$\langle \text{LH}|p_z|s\rangle = \frac{2}{\sqrt{6}}\langle p_x|p_x|s\rangle = \frac{2}{\sqrt{6}}p_{cv} \quad (5.64)$$

It is worth noting that:

$$\langle \text{HH}|p_z|s\rangle = 0 \quad (5.65)$$

It is useful to examine how the squared matrix elements depend on the polarization direction of incident light. In quantum wells, this effect is enhanced due to the heavy-hole (HH) and light-hole (LH) splitting.

z-polarized light:

HH \rightarrow c-band: No coupling

$$\text{LH} \rightarrow \text{c-band: } |\mathbf{p}_{if}|^2 = \frac{2}{3}|\langle p_x|p_x|s\rangle|^2$$

x-polarized light:

$$\text{HH} \rightarrow \text{c-band: } |\mathbf{p}_{if}|^2 = \frac{1}{2}|\langle p_x|p_x|s\rangle|^2 \quad (5.66)$$

$$\text{LH} \rightarrow \text{c-band: } |\mathbf{p}_{if}|^2 = \frac{1}{6}|\langle p_x|p_x|s\rangle|^2 \quad (5.67)$$

y-polarized light:

$$\text{HH} \rightarrow \text{c-band: } |\mathbf{p}_{if}|^2 = \frac{1}{2}|\langle p_x|p_x|s\rangle|^2$$

$$\text{LH} \rightarrow \text{c-band: } |\mathbf{p}_{if}|^2 = \frac{1}{6}|\langle p_x|p_x|s\rangle|^2$$

This shows that z-polarized light does not interact with heavy-hole states. At $\mathbf{k} = 0$, states are pure; further from the zone center, HH and LH states mix. x- and y-polarized light couple more strongly with HH than LH states. This selection rule behavior is crucial in designing lasers and

photodetectors.

To simplify notation, we define:

$$E_p = \frac{2}{m_0} |\langle p_x | p_x | s \rangle|^2 \quad (5.68)$$

For direct transitions using the parabolic band approximation, we have:

$$\hbar\omega - E_g = \frac{\hbar^2 k^2}{2} \left(\frac{1}{m_e^*} + \frac{1}{m_h^*} \right) = \frac{\hbar^2 k^2}{2m_r^*} \quad (5.69)$$

Here, m_r^* is the reduced mass of the electron-hole pair. The reduced joint density of states is given by:

$$N_{cv}(\hbar\omega) = \frac{\sqrt{2}(m_r^*)^{3/2}(\hbar\omega - E_g)^{1/2}}{\pi^2 \hbar^3} \quad (5.70)$$

Combining this with the previous expression for absorption rate, we get:

$$W_{abs} = \frac{\pi e^2 \hbar n_{ph}}{\epsilon m_0^2 \omega} |\mathbf{a} \cdot \mathbf{p}|_{cv}^2 N_{cv}(\hbar\omega) \quad (5.71)$$

In bulk semiconductors, for unpolarized light, the absorption rate becomes (from previous expressions):

$$W_{abs} = \frac{\pi e^2 \hbar n_{ph}}{2\epsilon m_0 \hbar \omega} \left(\frac{2p_{cv}^2}{m_0} \right) \frac{2}{3} N_{cv}(\hbar\omega) \quad (5.72)$$

Before comparing absorption and recombination in bulk and quantum wells, we briefly relate the absorption rate to the absorption coefficient. Considering photons traveling along the x -axis, we write the photon density continuity equation as:

$$\frac{dn_{ph}}{dt} = \frac{\partial n_{ph}}{\partial t} \Big|_{vol} + \frac{\partial(vn_{ph})}{\partial x} \quad (5.73)$$

The first term on the right-hand side is due to absorption, and the second to spatial variation in photon current. In steady-state, the photon density can be modeled as:

$$n_{ph}(x) = n_0 \exp(-\alpha x) \quad (5.74)$$

which defines the absorption coefficient α . Also, since:

$$\frac{\partial n_{ph}}{\partial t} = W_{abs} \quad (5.75)$$

In steady-state we have:

$$W_{abs} = \alpha v n_{ph} \quad (5.76)$$

Solving for α , the absorption coefficient is given by:

$$\alpha = \frac{W_{abs}}{v n_{ph}} \quad (5.77)$$

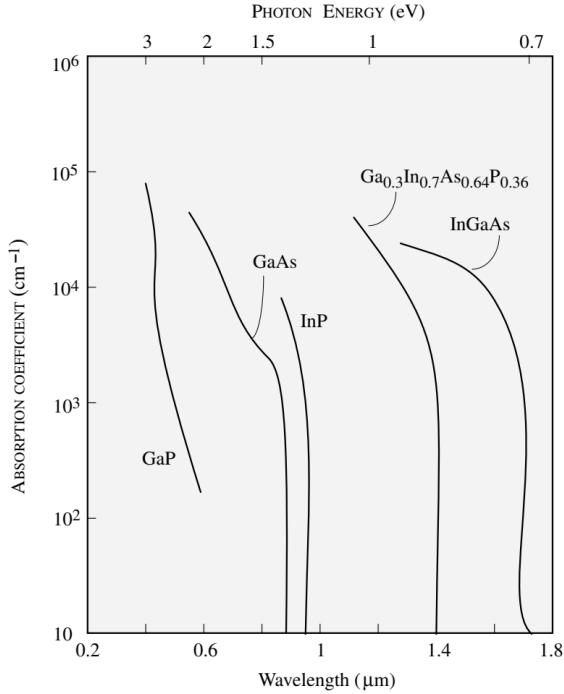


Figure 5.4. Absorption coefficient for several semiconductors.

For the total emission rate, the expression becomes:

$$W_{\text{em}} = \frac{\pi e^2 \hbar}{m_0^2 \hbar \omega \epsilon} (n_{ph} + 1) |\mathbf{a} \cdot \mathbf{p}_{if}|^2 \rho_a(\hbar \omega) \quad (5.78)$$

where ρ_a is the photon density of states for the polarization vector \mathbf{a} . The total photon density of states, accounting for two transverse polarization modes per \mathbf{k} -value, is given by:

$$\rho(\hbar \omega) = \frac{2\omega^2}{2\pi^2 \hbar v^3} \quad (5.79)$$

This expression applies to photons emitted in three-dimensional space. In the case where $n_{ph} = 0$, the emission rate reduces to the spontaneous emission rate W_{spon} . Its inverse defines the electron-hole recombination time τ_o . The quantity τ_o represents the average time required for an electron in a given state \mathbf{k} to recombine with an available hole in the same state.

5.3.2 Interband Transitions in Quantum Wells

The formalism developed so far can be straightforwardly adapted to quantum well structures. In this context, the central cell functions are mostly unaffected by the confining potential. The main differences are the modified wavefunctions—confined in the well—and the step-like density of states characteristic of 2D parabolic bands.

The electron and hole wavefunctions in the quantum well, using the envelope function approximation, are written as:

$$\psi_c^n = \frac{1}{\sqrt{AW}} e^{i\mathbf{k}_c \cdot \mathbf{r}} g_c^n(z) u_{c\mathbf{k}_c}^n, \quad \psi_v^m = \frac{1}{\sqrt{AW}} e^{i\mathbf{k}_v \cdot \mathbf{r}} \sum_{\nu} g_v^{\nu m}(z) u_{v\mathbf{k}_h}^{\nu} \quad (5.80)$$

Here, W is the well width, and A is the in-plane area. The envelope functions $g_c^n(z)$ and $g_v^{\nu m}(z)$ describe the confinement in the well. The momentum matrix element transitions from the 3D

form to a 2D quasi-dimensional form:

$$p_{if}^{3D} = \frac{1}{V} \int e^{i(\mathbf{k}_e - \mathbf{k}_h) \cdot \mathbf{r}} \langle u_v^\nu | p_a | u_c \rangle d^3 r \Rightarrow p_{if}^{2D} = \frac{1}{AW} \sum_\nu \langle g_v^{\nu m} | g_c^n \rangle \int e^{i(\mathbf{k}_e - \mathbf{k}_h) \cdot \mathbf{r}} \langle u_v^{\nu m} | p_a | u_c \rangle d^2 \rho \quad (5.81)$$

The overlap $\langle g_v^{\nu m} | g_c^n \rangle$ reflects the interaction between envelope functions in the z -direction. For symmetric wells, it often satisfies:

$$\sum_\nu \langle g_v^{\nu m} | g_c^n \rangle \approx \delta_{nm} \quad (5.82)$$

The final state density of states (originally from Eqn. 9.38) becomes the 2D reduced density of states:

$$\frac{N_{cv}^{2D}(\hbar\omega)}{W} = \frac{m_r^*}{\pi\hbar^2 W} \sum_{nm} \langle g_v^m | g_c^n \rangle \theta(E_{nm} - \hbar\omega) \quad (5.83)$$

where the total transition energy is

$$\mathbf{E}_{nm} = \mathbf{E}_{gap} + \mathbf{E}_c^n + \mathbf{E}_v^m \quad (5.84)$$

Here, m_r denotes the reduced mass of the electron-hole system. The symbol θ refers to the Heaviside step function, which controls the onset of optical transitions.

$$\alpha(\hbar\omega) = \frac{\pi e^2 \hbar}{m_0^2 c n_r \epsilon_0} \cdot \frac{1}{\hbar\omega} |\mathbf{a} \cdot \mathbf{p}_{if}|^2 \frac{N_{2D}(\hbar\omega)}{W} \sum_{n,m} f_{nm} \theta(E_{nm} - \hbar\omega) \quad (5.85)$$

The overlap factor f_{nm} is defined by:

$$f_{nm} = \left| \sum_\nu \langle g_v^{\nu m} | g_c^n \rangle \right|^2 \quad (5.86)$$

To gain physical insight into this expression, one can analyze a widely studied III-V semiconductor such as GaAs. Consider a quantum well structure composed of a 100 Å GaAs layer embedded in Al_{0.3}Ga_{0.7}As barriers. For bulk GaAs, the absorption coefficient begins at the photon energy $\hbar\omega = E_g$, starting from zero and increasing according to $(\hbar\omega - E_g)^{1/2}$. At higher photon energies, the absorption shows a $1/\hbar\omega$ dependence, which becomes relevant when the density of states no longer maintains a parabolic character.

Due to the degeneracy of heavy-hole (HH) and light-hole (LH) states in the bulk material, polarization effects do not influence the absorption coefficient near the band edge. However, in quantum well structures, the situation changes significantly. The absorption characteristics diverge from the bulk case, primarily due to modifications in the density of states function in low-dimensional systems. Moreover, the splitting of HH and LH levels leads to a strong polarization dependence in the absorption process, as discussed earlier.

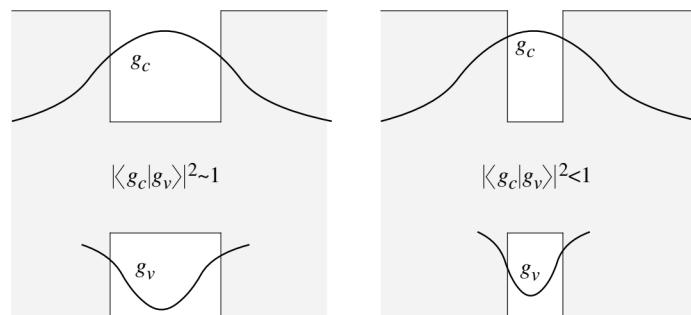


Figure 5.5. Schematic showing how quantum well width affects electron and hole wavefunctions and their overlap integral. As the well becomes very narrow, the overlap between electron and hole wavefunctions decreases.

In the context of quantum wells, the $1/W$ scaling of the absorption coefficient presents an interesting but potentially misleading aspect. This dependence originates from the assumption that the carrier wavefunctions are confined within a region of size comparable to the well width, W . According to this assumption, reducing the well width would indefinitely enhance the absorption coefficient. However, this conclusion is not entirely accurate.

When the quantum well becomes sufficiently narrow, the spatial extent of the electron and hole wavefunctions exceeds the well boundaries, meaning the carriers are no longer strictly confined to the well region. As a result, the spatial domain relevant for the electron wavefunction extends beyond W . Additionally, the discrepancy in effective mass between electrons and holes causes the electron wavefunction to spread more significantly, especially in wider wells. This leads to a reduced overlap between the electron and hole states, which in turn lowers the absorption coefficient from its idealized value of unity.

For typical semiconductor systems, an optimal well width is found to lie in the range of approximately 80–100 Å, where the trade-off between confinement and wavefunction overlap yields the most effective absorption.

5.4 Indirect Interband Transitions

In certain semiconductors like Si, Ge, and AlAs—commonly referred to as indirect bandgap materials—optical transitions near the band edge require the participation of both photons and phonons to satisfy momentum conservation in k -space. Consequently, optical absorption in materials such as silicon and germanium is significantly weaker compared to that in direct bandgap materials like GaAs.

In a typical indirect interband transition, the process occurs through a second-order mechanism. Initially, the electron interacts with a photon and is excited to a virtual intermediate state within the direct band structure, maintaining momentum conservation. Subsequently, the electron is scattered into the indirect conduction band minimum through phonon interaction. Although momentum is conserved during this intermediate state, energy is not. This is because the intermediate step is a virtual process, where the energy mismatch is permissible due to the time-energy uncertainty principle, which relaxes strict energy conservation temporarily.

Ultimately, the entire two-step interaction does conserve energy. The transition probability is still governed by Fermi's golden rule, albeit with a second-order perturbative matrix element. The expression for the transition rate is given by:

$$W_{if} = \frac{2\pi}{\hbar} \int \left| \sum_n \frac{\langle f | H_{\text{per}} | n \rangle \langle n | H_{\text{per}} | i \rangle}{E_i - E_n} \right|^2 \delta(E_f - E_i) \frac{d^3 k}{(2\pi)^3} \quad (5.87)$$

In the case of indirect transitions, the intermediate state is denoted by $|n\rangle$, and the total perturbation is composed of the sum of the photon and phonon interactions:

$$H_{\text{per}} = H_{\text{ph}} + H_{\text{ep}} \quad (5.88)$$

Here, H_{ph} represents the electron-photon interaction, while H_{ep} corresponds to the electron-phonon interaction. Both can contribute to the transition rate. Among the two interaction paths, the photon-initiated processes typically dominate due to a smaller energy denominator (closer to the direct bandgap) compared to phonon-first processes, which involve larger energy gaps such as $|\mathbf{E}_{v\Gamma} - \mathbf{E}_{vX}|$ or $|\mathbf{E}_{v\Gamma} - \mathbf{E}_{vL}|$.

Accordingly, the scattering rate for a given \mathbf{k} is given by:

$$W_{if}(\mathbf{k}) = \frac{2\pi}{\hbar} \int_f \left\{ |M_{\text{em}}|^2 + |M_{\text{abs}}|^2 \right\} \delta(E_f - E_i) \frac{d^3 k}{(2\pi)^3} \quad (5.89)$$

The matrix elements M_{em} and M_{abs} describe second-order interactions in which a photon is absorbed followed by either phonon emission or phonon absorption. Note that the photon energy

$\hbar\omega$ is smaller than the direct gap, so the intermediate state is virtual and energy conservation is not strictly required.

These second-order matrix elements take the form:

$$M_{\text{abs}} = \frac{\left| \langle c, \mathbf{k} + \mathbf{q} | H_{\text{ep}}^{\text{abs}} | c, \mathbf{k} \rangle \right|^2 \left| \langle c, \mathbf{k} | H_{\text{ph}}^{\text{abs}} | v, \mathbf{k} \rangle \right|^2}{(E_{g\Gamma} - \hbar\omega)^2} \quad (5.90)$$

$$M_{\text{em}} = \frac{\left| \langle c, \mathbf{k} - \mathbf{q} | H_{\text{ep}}^{\text{em}} | c, \mathbf{k} \rangle \right|^2 \left| \langle c, \mathbf{k} | H_{\text{ph}}^{\text{em}} | v, \mathbf{k} \rangle \right|^2}{(E_{g\Gamma} - \hbar\omega)^2}$$

Phonon scattering arises mainly from optical intervalley phonons, with a deformation potential interaction described by:

$$M_q^2 = \frac{\hbar D_{ij}^2}{2\rho V \omega_{ij}} \begin{cases} n(\omega_{ij}) & (\text{absorption}) \\ n(\omega_{ij}) + 1 & (\text{emission}) \end{cases} \quad (5.91)$$

Here, D_{ij} is the deformation potential constant, ρ is the material density, ω_{ij} is the phonon frequency for the intervalley transition between the Γ point and the final valley, and V is the normalization volume. Because this is an indirect process, the overall transition probability is typically suppressed by a factor:

$$\frac{M_q^2}{(E_{g\Gamma} - \hbar\omega)^2} \quad (5.92)$$

This suppression factor is commonly in the range of 10^{-2} to 10^{-3} , and exhibits a dependence on temperature via the phonon occupation number $n(\omega_{ij})$. Additionally, the indirect nature of the transition introduces a distribution of final states due to phonon scattering, and the rate must be integrated over all possible final states.

Thus, the total scattering rate becomes:

$$W_{ij}(\mathbf{k}) = \frac{2\pi}{\hbar} \frac{M_{\text{ph}}^2}{(E_{g\Gamma} - \hbar\omega)^2} \frac{\hbar D_{ij}^2}{2\rho\omega_{ij}} J_\nu \times [n(\omega_{ij})N_c(E_1 + \hbar\omega_{ij}) + (n(\omega_{ij}) + 1)N_c(E_1 - \hbar\omega_{ij})] \quad (5.93)$$

where J_ν is the number of equivalent final valleys, N_c is the density of final states in the conduction band, and the energy E_1 is given by:

$$E_1 = \hbar\omega - E_{gK'} - E_{\mathbf{k}} \quad (5.94)$$

Here, $E_{gK'}$ is the energy of the indirect bandgap, and $E_{\mathbf{k}}$ is the initial kinetic energy of the electron measured from the valence band top.

To compute the total absorption coefficient, the rate $W_{ij}(\mathbf{k})$ must be summed over all initial states \mathbf{k} capable of absorbing a photon of energy $\hbar\omega$. The integration limits for $E_{\mathbf{k}}$ depend on whether a phonon is absorbed or emitted:

$$E_{k,\text{max}} = \hbar\omega - E_{gK'} + \hbar\omega_{ij} \quad (\text{phonon absorption})$$

$$E_{k,\text{max}} = \hbar\omega - E_{gK'} - \hbar\omega_{ij} \quad (\text{phonon emission})$$

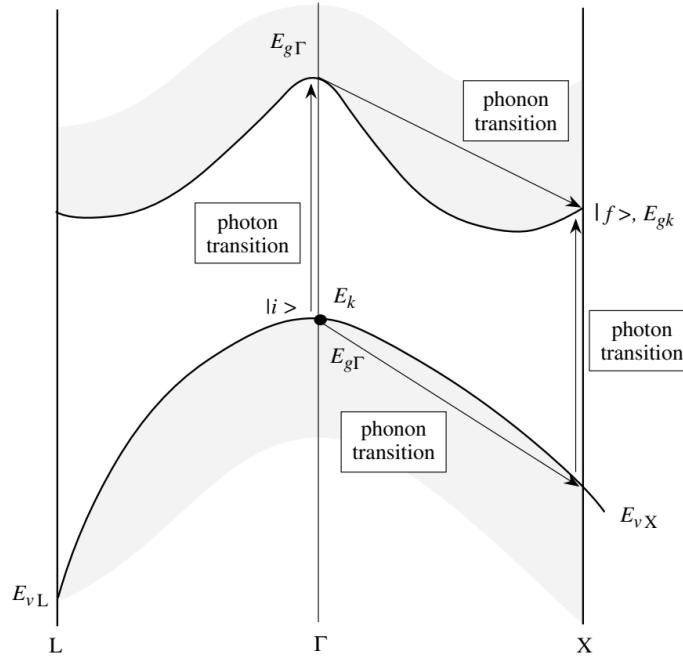


Figure 5.6: Two processes illustrating how an electron can transition from state $|i\rangle$ to state $|f\rangle$ through the combined action of a photon and a phonon. The photon energy does not need to match the vertical transition energy, as the intermediate state is virtual and not physically occupied by the electron.

To obtain the total transition rate, the quantity $W(\mathbf{k})$ is integrated over all initial valence states. Assuming a parabolic band structure, this becomes a straightforward integration over energy from 0 to $E_{k,\max}$, weighted by the density of initial states. For a single spin direction, the density of states in the valence band is given by $N_v(E_{\mathbf{k}})$, and we multiply the rate by $2VN_v(E_{\mathbf{k}}) dE_{\mathbf{k}}$. Performing the integration, we find the absorption rate for photons of energy $\hbar\omega$ to be:

$$W_{\text{abs}}(\hbar\omega) = \frac{M_{\text{ph}}^2 D_{ij}^2 J_\nu(m_c m_v)^{3/2}}{8\pi^2(E_{g\Gamma} - \hbar\omega)^2 \hbar^6 \rho \omega_{ij}} \times \left[n(\omega_{ij}) (\hbar\omega - E_{gK'} + \hbar\omega_{ij})^2 + (n(\omega_{ij}) + 1) (\hbar\omega - E_{gK'} - \hbar\omega_{ij})^2 \right] \quad (5.95)$$

Here, M_{ph}^2 is the photon-related matrix element, expressed as:

$$M_{\text{ph}}^2 = \frac{e^2 \hbar n_{\text{ph}}}{2m_0^2 \epsilon \omega} |\mathbf{a} \cdot \mathbf{p}_{if}|^2 \quad (5.96)$$

The absorption coefficient α is proportional to the transition rate divided by the photon flux, and is thus given by:

$$\alpha(\hbar\omega) = \frac{W_{\text{abs}}}{n_{\text{ph}} v_{\text{ph}}} \quad (5.97)$$

Once the photon energy exceeds the absorption threshold, the coefficient begins to increase with energy as $(\hbar\omega - E_{\text{th}})^2$. This contrasts with direct transitions in direct-gap semiconductors, where the increase follows a $(\hbar\omega - E_g)^{1/2}$ dependence.

In practical measurements, materials like silicon and germanium exhibit a much lower absorption coefficient near the band edge compared to direct-gap materials such as GaAs. However, once the photon energy surpasses the direct transition threshold, the absorption rises steeply due to the opening of direct transition channels.

It is important to note that in this formulation, only phonon-assisted absorption processes are considered. Nonetheless, other mechanisms—such as scattering by impurities, lattice defects, or alloy disorder—can also contribute to absorption in indirect semiconductors. Interestingly, materials with lower crystalline quality may sometimes show enhanced absorption compared to high-purity indirect semiconductors.

In amorphous materials such as amorphous silicon (a-Si), the absence of periodicity eliminates the need to conserve crystal momentum, since \mathbf{k} is no longer a good quantum number. This lack of \mathbf{k} -selection allows broader absorption, as transitions are not constrained by momentum conservation.

5.5 Intraband Transitions

Intraband optical transitions refer to electronic transitions that occur within the same energy band, either the conduction band or the valence band. In bulk semiconductors, these transitions cannot occur as first-order (i.e., single-photon or vertical) processes in momentum space due to the requirement for conservation of both energy and momentum. As a result, first-order intraband transitions are forbidden in bulk materials under normal conditions.

However, in quantum well systems, the situation differs significantly due to quantum confinement effects. The confinement leads to the formation of discrete energy levels or subbands within each band. Transitions between these subbands—referred to as inter-subband transitions—become allowed even though they still occur within the same main energy band.

These inter-subband transitions are a type of intraband process and are of particular interest in long-wavelength optoelectronic applications, such as in infrared detectors and quantum cascade lasers. The ability to engineer subband structures via quantum confinement makes quantum wells a valuable platform for exploiting such optical transitions.

5.5.1 Intraband Transitions in Bulk Semiconductors

In bulk semiconductors, intraband transitions require the involvement of a phonon or another scattering mechanism—such as ionized impurities, defects, or alloy disorder—to conserve momentum. These transitions are fundamentally second-order processes, analogous to those encountered in indirect interband absorption.

This type of absorption, often referred to as free carrier absorption, plays a crucial role in laser operation, especially because it contributes to optical losses in the cladding layers of the laser structure.

The behavior of free carrier absorption can be effectively modeled using the Drude theory of electrical transport. Under this framework, the incident optical field is treated as a time-varying electric field that causes conduction electrons to oscillate within the energy band. In the absence of scattering, these oscillations are lossless, resulting in no net energy absorption from the electromagnetic wave.

However, in real materials, various scattering mechanisms are always present. These interactions cause the electrons to dissipate part of the energy they absorb by emitting phonons, thus enabling net absorption of energy from the optical field. The frequency dependence of the absorption coefficient is then governed by:

$$\alpha(\hbar\omega) \propto \frac{1}{\omega^2} \propto \frac{1}{\mu} \quad (5.97)$$

where ω is the angular frequency of the incident radiation, and μ is the carrier mobility. This relationship indicates that the absorption decreases with increasing photon frequency and is also inversely proportional to the mobility. Hence, materials with high carrier mobility—where scattering is minimal—exhibit very low levels of free carrier absorption.

5.5.2 Intraband Transitions in Quantum Wells

Quantum well structures can significantly modify the electronic properties of semiconductors due to confinement effects. In such systems, the electronic wavefunctions are no longer plane waves along the growth (typically z) direction, allowing for intraband transitions that would otherwise be forbidden in bulk materials. These inter-subband transitions are enabled by quantum confinement and can be selectively accessed depending on the polarization of incident light.

Because the energy separation between subbands is highly sensitive to the well width, inter-subband transitions can be tuned for specific applications, especially in the far-infrared and terahertz spectral regions. This tunability makes quantum wells particularly valuable in detector technologies.

Consider a quantum well structure grown along the z -axis. Due to the confinement in this direction, the envelope wavefunctions for different subbands can be separated into transverse and longitudinal parts. For instance, the first two subband states can be written as:

$$\begin{aligned}\psi^1(\mathbf{k}, z) &= g^1(z)e^{i\mathbf{k}\cdot\boldsymbol{\rho}}u_{n\mathbf{k}}^1(\mathbf{r}) \\ \psi^2(\mathbf{k}, z) &= g^2(z)e^{i\mathbf{k}\cdot\boldsymbol{\rho}}u_{n\mathbf{k}}^2(\mathbf{r})\end{aligned}\quad (5.98)$$

where $g^1(z)$ and $g^2(z)$ represent the envelope functions along the growth direction, $\boldsymbol{\rho}$ is the in-plane coordinate, and $u_{n\mathbf{k}}^i(\mathbf{r})$ are the Bloch functions. These expressions reflect the separation of variables commonly used in effective mass approximations for quantum wells.

In analyzing optical transitions, especially vertical (in \mathbf{k} -space) ones, we observe that g^1 and g^2 are orthogonal. Furthermore, the central cell functions u^1 and u^2 are approximately identical for the different subbands, especially in the conduction band. Under this approximation, the matrix element for momentum transfer due to interaction with light becomes:

$$\mathbf{p}_{if} = -\frac{i\hbar}{W} \int g^{2*}(z)e^{-i\mathbf{k}\cdot\boldsymbol{\rho}} \mathbf{a} \cdot \nabla g^1(z)e^{i\mathbf{k}\cdot\boldsymbol{\rho}} d^2\rho dz \quad (5.99)$$

Here, W is the width of the quantum well, and \mathbf{a} is the polarization vector of the incident light. As in three-dimensional semiconductor systems, the momentum matrix element for inter-subband transitions vanishes if the light polarization vector (or the gradient operator in the integrand) lies within the in-plane $\boldsymbol{\rho}$ -direction. For this reason, transitions involving x - or y -polarized light are forbidden due to symmetry, and the matrix element evaluates to zero. This holds unless strong band mixing exists, such as in valence band transitions, where the central cell functions may differ.

When light is polarized along the growth axis z , the momentum matrix element for a transition between two subbands becomes:

$$\mathbf{p}_{if} = -\frac{i\hbar}{W} \int g^{2*}(z) \hat{z} \frac{\partial}{\partial z} g^1(z) dz \quad (5.100)$$

If the envelope functions $g^1(z)$ and $g^2(z)$ have even and odd parity, respectively, then the product $g^2(z) \cdot \partial_z g^1(z)$ has overall odd parity, and the integral is non-zero. This configuration leads to a non-vanishing momentum matrix element under z -polarized excitation, which may be approximated as:

$$|\mathbf{p}_{if}| \approx \frac{\hbar}{W} \quad (5.101)$$

This approximation remains valid as long as both the ground and excited states are well confined. If the excited state becomes delocalized, a more exact computation of the matrix element is required. In the simple parabolic band approximation, the conduction subbands are assumed to be parallel, with an energy shift corresponding to the subband separation $E_2 - E_1$. Consequently, the joint density of states appears as a delta function at the transition energy. Incorporating

Fermi-Dirac statistics, the absorption rate is given by:

$$W_{\text{abs}} = \frac{\pi e^2 n_{\text{ph}}}{m_0^2 \omega \epsilon} \cdot \frac{1}{W} \sum_f |\mathbf{p}_{if}|^2 \delta(E_2 - E_1 - \hbar\omega) f(E_1) [1 - f(E_2)] \quad (5.102)$$

Assuming that the higher subband is initially unoccupied, the final-state occupation $f(E_2)$ becomes zero, and the sum over final states reduces to:

$$\sum_{\text{2nd subband}} \delta(E_2 - E_1 - \hbar\omega) f(E_1) = N_1 \quad (5.103)$$

where N_1 is the electron population in the first (lowest) subband.

The delta function implies an infinite absorption at resonance under the idealized parabolic band approximation. However, in realistic scenarios, this singularity is mitigated by nonparabolic effects and scattering mechanisms. Due to the finite lifetime of electron states, spectral broadening occurs, as dictated by the time-energy uncertainty principle.

If this broadening is modeled with a Gaussian profile, the two-dimensional density of states becomes:

$$N(E) = \frac{N_1}{\sqrt{1.44\pi\sigma}} \exp\left(-\frac{(E - E_{12})^2}{1.44\sigma^2}\right) \quad (5.104)$$

where E_{12} is the transition energy between subbands and σ is the linewidth of the transition.

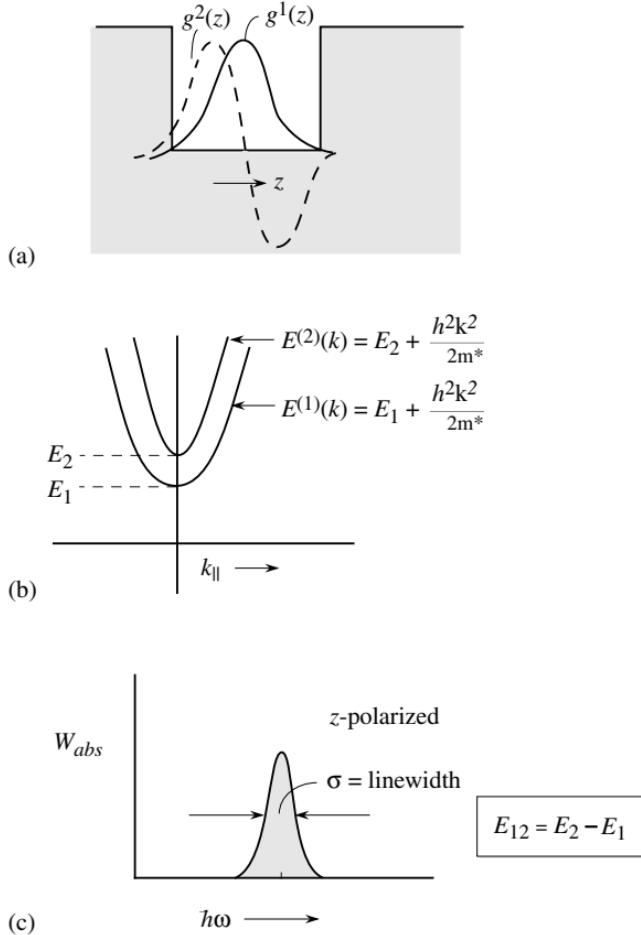


Figure 5.7. Schematic illustration of: (a) envelope functions for two quantized energy levels in a quantum well, (b) the corresponding subband structure, and (c) the absorption spectrum for z -polarized light in the quantum well.

The absorption coefficient increases rapidly as the quantum well width decreases. However, for very narrow wells, the assumption that electrons remain fully confined may no longer hold. For transitions with a linewidth of only 1–2 meV, the absorption coefficient can reach values on the order of 10^4 cm^{-1} , making these transitions particularly attractive for use in photodetectors and optical modulators.

It is important to note that for z -polarized light in a quantum well with confinement along the z -direction, the light propagates in the plane of the substrate. As a result, vertical incidence does not lead to absorption. While valence band quantum wells exhibit band mixing that enables vertical light absorption, conduction band quantum wells lack this feature. This limitation can be problematic for photodetector applications.

To overcome this in imaging arrays, mirrors are often integrated into the detector structure to redirect vertically incident light so that it enters the device along the x - y plane.

5.6 Charge Injection and Radiative Recombination

5.6.1 Spontaneous Emission Rate

When electrons and holes are injected into the conduction and valence bands of a semiconductor, they can recombine radiatively, resulting in photon emission. In the absence of photons already present in the optical cavity (i.e., when the photon density $n_{\text{ph}} = 0$), this process corresponds to spontaneous emission. The characteristic spontaneous emission time in direct-gap materials is typically on the order of 0.5 ns, assuming both an electron and a hole occupy the same \mathbf{k} -state in their respective bands.

However, under realistic conditions, the actual recombination rate must account for the occupation probabilities of electrons and holes at each \mathbf{k} -point. Thus, one integrates over the full range of electronic states, weighted by the respective Fermi-Dirac distributions. The total spontaneous emission rate per unit volume (in $\text{cm}^{-3}\text{s}^{-1}$) is given by:

$$R_{\text{spon}} = \frac{2}{3} \int d(\hbar\omega) \frac{2e^2 n_r \hbar\omega}{m_0^2 c^3 \hbar^2} \left[\int \frac{d^3 k}{(2\pi)^3} |\mathbf{p}_{if}|^2 \delta(E^e(\mathbf{k}) - E^h(\mathbf{k}) - \hbar\omega) f_e(E^e(\mathbf{k})) f_h(E^h(\mathbf{k})) \right] \quad (5.105)$$

The outer integral over $d(\hbar\omega)$ accounts for all photon energies, while the inner $d^3 k$ integral sums over momentum states. The prefactor $2/3$ arises from averaging over random photon polarizations, assuming isotropic emission.

In quantum well structures, the formulation is adjusted to reflect the reduced dimensionality. Specifically, the three-dimensional density of states is replaced by the corresponding two-dimensional form. The spontaneous emission rate per unit area (in $\text{cm}^{-2}\text{s}^{-1}$) becomes:

$$R_{\text{spon}} = \frac{2}{3} \int d(\hbar\omega) \frac{2e^2 n_r \hbar\omega}{m_0^2 c^3 \hbar^2} \sum_{n,m} \left[\int \frac{d^2 k}{(2\pi)^2} |\mathbf{p}_{if}|^2 \delta(E_n^e(\mathbf{k}) - E_m^h(\mathbf{k}) - \hbar\omega) f_e(E_n^e(\mathbf{k})) f_h(E_m^h(\mathbf{k})) \right] \quad (5.106)$$

Here, n and m index the discrete electron and hole subbands in the quantum well.

Spontaneous emission is also commonly described in terms of the radiative lifetime τ_o , defined as the inverse of the spontaneous emission rate. Under this definition:

$$R_{\text{spon}} = \frac{1}{\tau_o} \int d(\hbar\omega) N_{\text{cv}} \{f^e(E^e)\} \{f^h(E^h)\} \quad (5.107)$$

In this expression, N_{cv} denotes the joint density of states for electron-hole recombination, and the integral represents the overlap of electron and hole occupation functions across energies matching the transition condition $E^e - E^h = \hbar\omega$. The spontaneous recombination rate plays a central role in both electronic and optoelectronic device performance. It is therefore useful to examine its behavior under a few common operating conditions. We focus here on electron-hole recombination under three distinct regimes:

i) Minority Carrier Injection

When the electron density n is much greater than the hole density p , and the semiconductor is heavily doped, we can assume that the occupation probability for electrons $f^e(E^e) \approx 1$. In this case, the recombination rate becomes dependent only on the hole distribution:

$$R_{\text{spon}} \approx \frac{1}{\tau_o} \int d(\hbar\omega) N_{\text{cv}} f^h(E^h) \approx \frac{1}{\tau_o} \int d(\hbar\omega) N_h f^h(E^h) \left(\frac{m_r^*}{m_h^*} \right)^{3/2} \quad (5.108)$$

This leads to:

$$R_{\text{spon}} \approx \frac{1}{\tau_o} \left(\frac{m_r^*}{m_h^*} \right)^{3/2} p \quad (5.109)$$

Thus, the spontaneous recombination rate is directly proportional to the minority carrier concentration (holes, in this case).

ii) Strong Injection

Under conditions of high-level injection where both electrons and holes are introduced in significant numbers, the occupation probabilities f^e and f^h approach unity over the relevant energy range. Consequently, the recombination rate is limited by the lower of the two carrier densities:

$$R_{\text{spon}} = \frac{n}{\tau_o} = \frac{p}{\tau_o} \quad (5.110)$$

iii) Weak Injection

In the low-injection regime, the Fermi-Dirac distributions can be approximated by Boltzmann statistics. In this case:

$$f^e \cdot f^h \approx \exp \left(-\frac{E_c - E_{Fn}}{k_B T} \right) \exp \left(-\frac{E_{Fp} - E_v}{k_B T} \right) \exp \left(-\frac{\hbar\omega - E_g}{k_B T} \right) \quad (5.111)$$

This leads to a modified recombination rate:

$$R_{\text{spon}} = \frac{1}{2\tau_o} \left(\frac{2\pi\hbar^2 m_r^*}{k_B T m_e^* m_h^*} \right)^{3/2} np \quad (5.112)$$

To account for excess carriers, we express total carrier densities as the sum of equilibrium and excess components:

$$n = n_0 + \Delta n, \quad p = p_0 + \Delta n \quad (5.113)$$

Substituting into the recombination expression yields:

$$R_{\text{spon}} \approx \frac{1}{2\tau_o} \left(\frac{2\pi\hbar^2 m_r^*}{k_B T m_e^* m_h^*} \right)^{3/2} (\Delta n p_0 + \Delta n n_0) \quad (5.114)$$

If we assume $\Delta n = \Delta p$, we can define the recombination rate per excess carrier as:

$$\frac{1}{\tau_r} = \frac{R_{\text{spon}}}{\Delta n} = \frac{1}{2\tau_o} \left(\frac{2\pi\hbar^2 m_r^*}{k_B T m_e^* m_h^*} \right)^{3/2} (n_0 + p_0) \quad (5.115)$$

At low levels of carrier injection, the recombination time τ_r significantly exceeds the spontaneous emission lifetime τ_o . This occurs because, under such conditions, the probability that an electron finds a hole to recombine with is small.

iv) Inversion Condition

A particularly relevant regime arises when the sum of the electron and hole occupation probabilities satisfies $f^e + f^h = 1$. This situation corresponds to population inversion, a critical condition in which emission and absorption processes occur with equal probability. Assuming symmetric carrier distributions with $f^e \approx f^h = 1/2$, the spontaneous recombination rate simplifies to:

$$R_{\text{spon}} \approx \frac{n}{4\tau_o} \approx \frac{p}{4\tau_o} \quad (5.116)$$

This leads to an effective radiative lifetime of approximately $4\tau_o$, which serves as a useful estimate when determining the threshold current of a semiconductor laser.

The gain and recombination dynamics explored here are fundamental to the design and analysis of both electronic and optoelectronic devices. From the above considerations, it becomes evident that the recombination lifetime associated with a single excess carrier can often be described by the general expression:

$$\tau_r = \frac{\Delta n}{R_{\text{spon}}} \quad (5.117)$$

In the cases of minority carrier injection or under strong injection conditions, $\tau_r \approx \tau_o$. In general, the spontaneous recombination rate R_{spon} exhibits a strong dependence on carrier concentration, which in turn affects the recombination time τ_r .

Empirical measurements show that τ_r can vary over several orders of magnitude depending on the injection level. For example, in gallium arsenide (GaAs), the radiative lifetime can range from microseconds at low injection to nanoseconds or less under high-injection conditions.

5.6.2 Gain in a Semiconductor

When both electrons and holes are injected into a semiconductor, as is the case in light-emitting devices, they may recombine radiatively, emitting photons. Under certain conditions, the rate of photon emission may exceed the rate of absorption, resulting in a net amplification of the optical field. This difference between stimulated emission and absorption defines the optical gain of the material.

In this regime, the gain is expressed as the difference between the emission coefficient and the absorption coefficient. A positive gain implies that a propagating optical wave will experience amplification rather than attenuation.

For parabolic energy bands, the gain as a function of photon energy $\hbar\omega$ can be expressed as:

$$g(\hbar\omega) = \frac{\pi e^2 \hbar}{n_r c m_r^2 \epsilon_0(\hbar\omega)} |\mathbf{a} \cdot \mathbf{p}_{if}|^2 N_{cv}(\hbar\omega) \left[f^e(E^e) - (1 - f^h(E^h)) \right] \quad (5.118)$$

This expression generalizes the form of the absorption coefficient by including the competition between emission and absorption processes. The square bracket term reflects the difference between the probability of photon emission, which is proportional to $f^e \cdot f^h$, and that of photon absorption, proportional to $(1-f^e)(1-f^h)$. When this difference is positive, the medium exhibits net optical gain.

The electron and hole energies are assumed to follow parabolic dispersions, such that:

$$\hbar\omega - E_g = \frac{\hbar^2 k^2}{2m_r^*} \quad (5.119)$$

Here, E_g is the bandgap energy and m_r^* is the reduced effective mass. This relation provides the link between photon energy and the carrier wavevector k in the joint density of states.

The carrier energies in the conduction and valence bands are described using parabolic band

approximations. For electrons and holes involved in optical transitions, the energies are given by:

$$E^e = E_c + \frac{\hbar^2 k^2}{2m_e^*} = E_c + \frac{m_r^*}{m_e^*}(\hbar\omega - E_g) \quad (5.120)$$

$$E^h = E_v - \frac{\hbar^2 k^2}{2m_h^*} = E_v - \frac{m_r^*}{m_h^*}(\hbar\omega - E_g) \quad (5.121)$$

In equilibrium, the carrier distribution is determined by the Fermi level E_F . However, under injection conditions—such as in LEDs or semiconductor lasers—electrons and holes are injected into their respective bands, leading to the formation of quasi-Fermi levels E_{Fn} and E_{Fp} .

At high injection levels, these quasi-Fermi levels may penetrate deep into the conduction and valence bands. When no electrons occupy the conduction band and no holes are present in the valence band (i.e., $f^e(E^e) = 0$ and $f^h(E^h) = 0$), the material exhibits pure absorption, and the gain is simply the negative of the absorption coefficient: $-\alpha(\hbar\omega)$.

A transition to optical gain occurs when the condition:

$$f^e(E^e) > 1 - f^h(E^h) \quad (5.122)$$

is satisfied. This regime corresponds to what is known as *inversion*, where stimulated emission dominates over absorption.

In the presence of gain, the intensity of light traveling through the material increases exponentially with distance. This spatial dependence of optical intensity is described by:

$$I(z) = I_0 \exp(gz) \quad (5.123)$$

where I_0 is the initial intensity and g is the optical gain coefficient. This exponential growth contrasts with the usual case of absorption, where intensity decays as light propagates. The phenomenon of gain is the fundamental principle behind semiconductor laser operation.

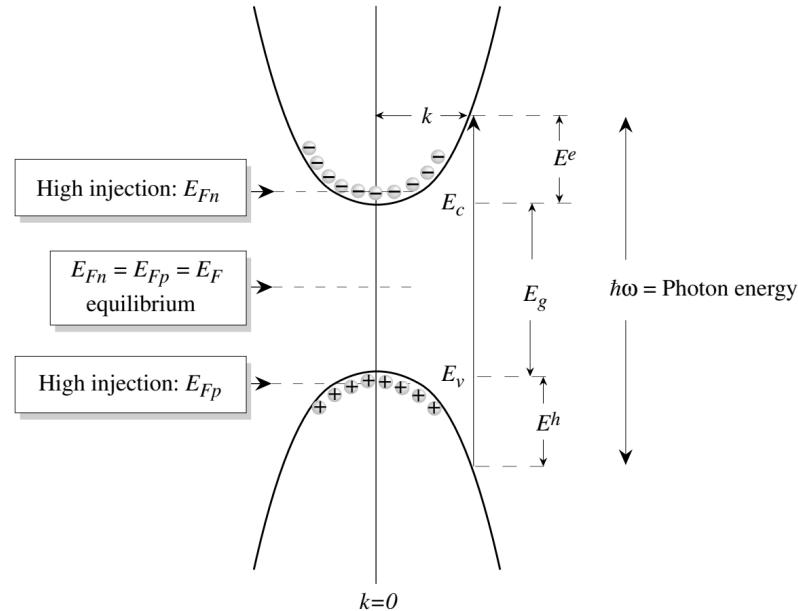


Figure 5.8. Under high injection, the quasi-Fermi levels move deeper into the conduction and valence bands, indicating increased carrier densities. The optical behavior of the semiconductor is influenced by the carrier energies, which depend on the incident photon energy and the effective masses. These factors can lead to population inversion.

5.7 Nonradiative Recombination

5.7.1 Charge Injection: Nonradiative Effects

In an ideal, defect-free semiconductor, Bloch's theorem dictates that electronic states exist only within the allowed energy bands, and no localized states are present within the bandgap. However, in real materials, defects and impurities—either intentional (e.g., dopants) or unintentional (e.g., vacancies or native defects)—introduce localized electronic states in the bandgap region. These states arise from chemical impurities or structural imperfections, leading to energy levels capable of capturing carriers.

While conduction and valence band states are spatially extended, defect-induced states are spatially localized and can act as traps. Electrons moving freely through the conduction band may become captured by such traps. Once trapped, these carriers may recombine with holes from the valence band without emitting a photon, resulting in a nonradiative recombination process. This mechanism competes with radiative recombination and becomes especially significant when defect density is high.

Let us consider a nonradiative recombination mechanism involving a midgap trap with density N_t . If an empty trap is characterized by a capture cross-section σ , and v_{th} is the thermal velocity of the carrier, then the average nonradiative lifetime for electrons and holes, denoted $\tau_{nr}(e)$ and $\tau_{nr}(p)$, are given by:

$$\tau_{nr}(e) = \frac{1}{N_t v_{th} \sigma_n}, \quad \tau_{nr}(p) = \frac{1}{N_t v_{th} \sigma_p} \quad (5.124)$$

Here, σ_n and σ_p are the electron and hole capture cross-sections, respectively. These expressions form the basis of the Shockley-Read-Hall (SRH) theory of nonradiative recombination. To simplify the treatment, we consider the following assumptions:

- i) Equal lifetimes for electrons and holes: $\tau_{nr}(e) = \tau_{nr}(p) = \tau_{nr}$,
- ii) The trap energy level E_t lies at midgap: $E_t = E_{Fi}$,
- iii) High-injection condition: $np \gg n_i^2$, where n_i is the intrinsic carrier concentration.

Under these assumptions, the nonradiative recombination rate becomes:

$$R_t = \frac{np}{\tau_{nr}(n + p)} \quad (5.125)$$

This expression illustrates the dependence of the nonradiative recombination rate on both carrier densities and trap properties. It becomes especially relevant in materials with significant defect populations or under high-level injection conditions.

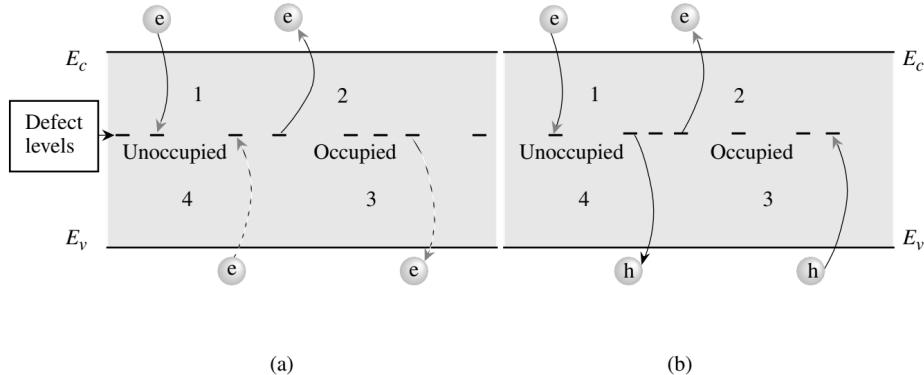


Figure 5.9. Illustration of trapping and recombination processes involving deep levels in the bandgap (dashed line). (a) Processes 1 and 2 depict electron trapping and emission, while 3 and 4 correspond to hole trapping and emission. (b) Electron-hole recombination via deep levels is also shown.

5.8 Semiconductor Light Emitters

Radiative recombination forms the physical foundation for semiconductor-based light-emitting devices such as light-emitting diodes (LEDs) and laser diodes (LDs). These devices are pivotal in modern technology, particularly in display systems and optical communication.

LEDs function through the mechanism of spontaneous emission, where electrons and holes recombine and emit photons randomly in phase and direction. In contrast, laser diodes rely on stimulated emission, a process that is contingent on the photon population already present in the active region of the device.

As previously described, stimulated emission occurs when an incoming photon induces an excited electron to recombine with a hole, emitting a new photon that is coherent with the stimulating one—i.e., it shares the same phase, frequency, and direction. This results in the generation of coherent light, which is the hallmark of laser operation.

In spontaneous emission, however, the photons emitted by electron-hole recombination events are incoherent, meaning they do not maintain a fixed phase relationship with each other. This fundamental difference between spontaneous and stimulated emission underpins the distinct optical characteristics of LEDs and LDs.

This section provides an overview of these two important device classes and the physical principles governing their operation.

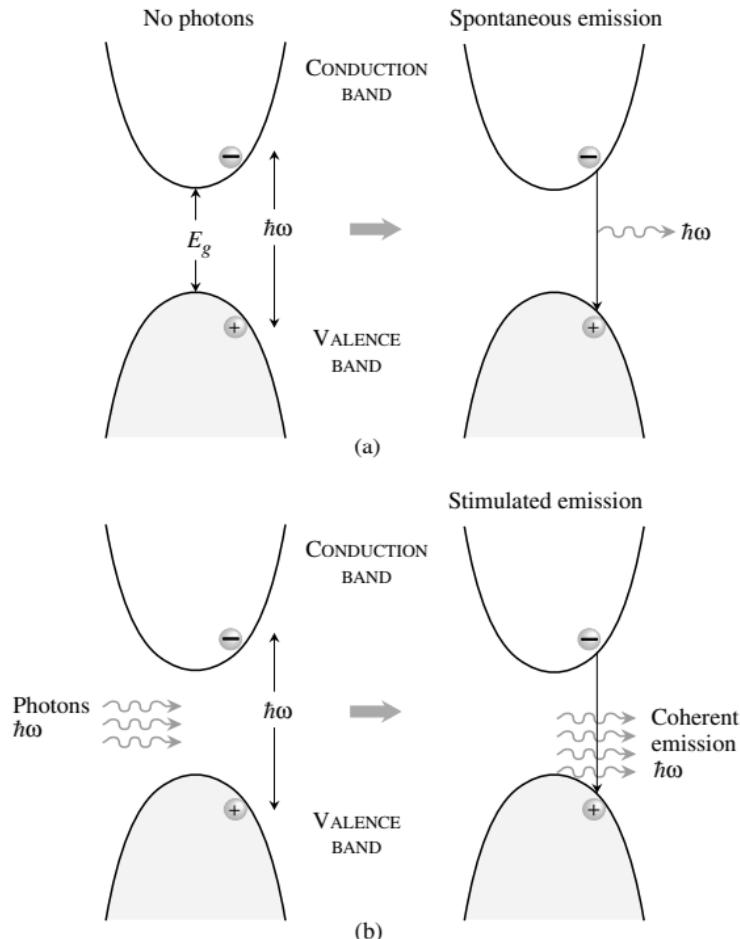


Figure 5.10. (a) Spontaneous emission: in the absence of external photons, electron-hole recombination generates photons with random phase. (b) Stimulated emission: in the presence of incident photons, recombination results in photons that are phase-coherent with the incoming ones.

5.8.1 Light Emitting Diode

The light-emitting diode (LED) operates as a forward-biased p - n junction diode. Under forward bias, minority carriers—electrons from the n -type side and holes from the p -type side—are injected across the junction. These carriers recombine in the active region, either radiatively, producing photons, or non-radiatively, dissipating energy as heat or via defects.

To optimize light output, the device must be engineered to maximize the probability of radiative recombination over non-radiative pathways. This is achieved by careful structural design and doping profiles.

During operation, electrons flow from the n -side into the p -side, while holes are injected from the p -side into the n -side. Ideally, recombination—and thus photon emission—should occur near the top surface of the device. Emission deep within the structure is undesirable, as photons generated in these regions have a higher likelihood of being reabsorbed before escaping.

To mitigate this issue, LEDs are typically designed so that the top layer is p -type. In such configurations, electrons (the majority carriers in the n -type material) dominate the injection current. Since radiative efficiency is enhanced when electron injection dominates, the diode is fabricated with asymmetric doping—heavily doped on the p -side and more lightly doped on the n -side. This ensures that most of the recombination and subsequent light emission occurs close to the emitting surface, improving external quantum efficiency.

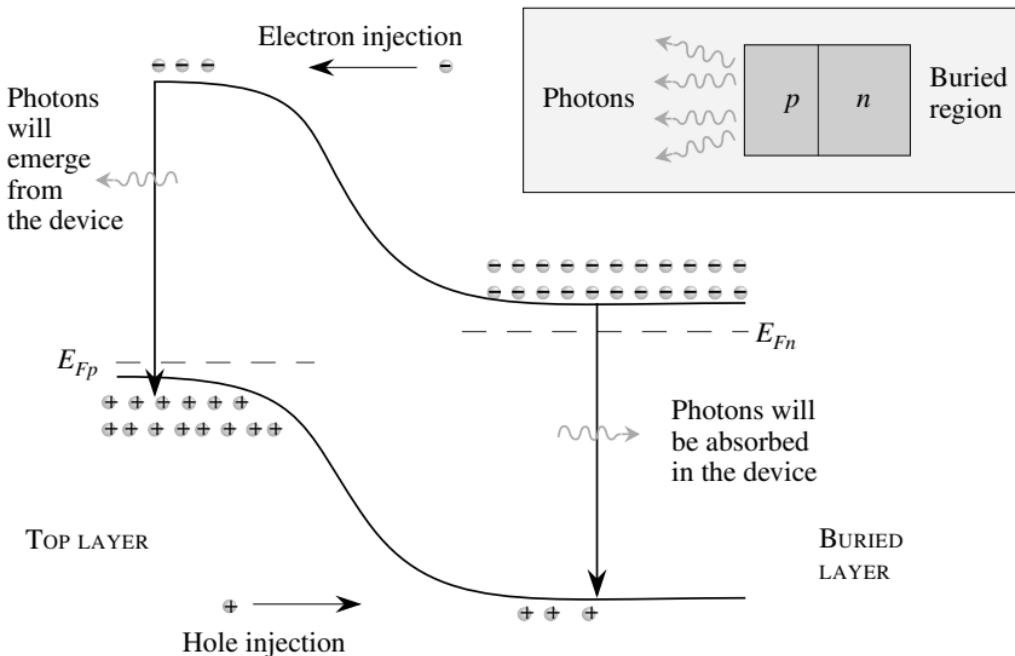


Figure 5.11. In a forward-biased p - n junction, both electrons and holes are injected across the junction. Holes injected into the buried n -region generate photons deep within the device, which are unlikely to escape. In contrast, electrons injected into the p -region recombine near the surface, producing photons with a higher probability of emerging from the LED.

5.8.2 Laser Diode

The laser diode, like the LED, relies on a forward-biased p - n junction to inject electrons and holes into the active region. However, unlike LEDs, laser diodes are engineered to incorporate an optical cavity that confines and guides the generated photons.

This optical cavity functions as a resonator, allowing photons to reflect back and forth multiple times. Because only a fraction of the emitted photons escapes the cavity during each round-trip,

the internal photon density increases, enabling the conditions necessary for stimulated emission and optical amplification.

A commonly employed resonant structure in semiconductor lasers is the Fabry-Pérot cavity. This cavity typically consists of two parallel reflective surfaces—often cleaved crystal facets or deposited mirrors—that define the longitudinal boundary of the laser. These reflective interfaces ensure that only specific optical wavelengths form standing wave patterns, or resonant modes, within the cavity.

The condition for resonance requires that the cavity length L accommodate an integer number of half-wavelengths of the light, given by:

$$L = \frac{q\lambda}{2} \quad (5.126)$$

where q is an integer, and λ is the wavelength of light inside the semiconductor material. The wavelength λ within the material is related to the free-space wavelength λ_0 by the refractive index n_r of the cavity medium:

$$\lambda = \frac{\lambda_0}{n_r} \quad (5.127)$$

In practice, the Fabry-Pérot cavity includes two polished and reflective facets on opposite ends of the diode structure, which reflect most of the light back into the cavity. The remaining surfaces are often intentionally roughened or coated to prevent unwanted reflections, ensuring directional emission and suppressing lateral modes.

This resonant structure allows the laser diode to support coherent light emission by sustaining specific longitudinal modes, which meet the resonance condition. The combination of carrier injection, optical feedback, and gain results in the lasing action that defines the operation of semiconductor lasers.

When a planar heterostructure is employed to confine the optical mode in the z -direction, the governing equation for the optical field becomes:

$$\frac{d^2F_k(z)}{dz^2} + \left(\frac{\epsilon(z)\omega^2}{c^2} - k^2 \right) F_k(z) = 0 \quad (5.128)$$

In this expression, F represents the electric field associated with the optical wave, $\epsilon(z)$ is the spatially varying dielectric constant, and k is the transverse wavevector. The spatial variation of $\epsilon(z)$ is engineered to confine the optical mode along the z -axis. This confinement is typically achieved by embedding the active region within cladding layers made of materials with a larger bandgap.

A critical figure of merit for such a laser structure is the optical confinement factor Γ , which quantifies the fraction of the optical field energy residing in the active region where electron-hole recombination and photon generation occur. It is defined as:

$$\Gamma = \frac{\int_{\text{active region}} |F(z)|^2 dz}{\int |F(z)|^2 dz} \quad (5.129)$$

In conventional double-heterostructure lasers based on bulk materials, the confinement factor Γ approaches unity when the active region thickness is on the order of $\sim 1 \mu\text{m}$. However, in advanced quantum well laser designs, Γ can be as low as 1%. Despite this reduced confinement, quantum well lasers often outperform bulk devices due to their enhanced electronic properties, such as a higher density of states arising from the two-dimensional nature of quantum wells.

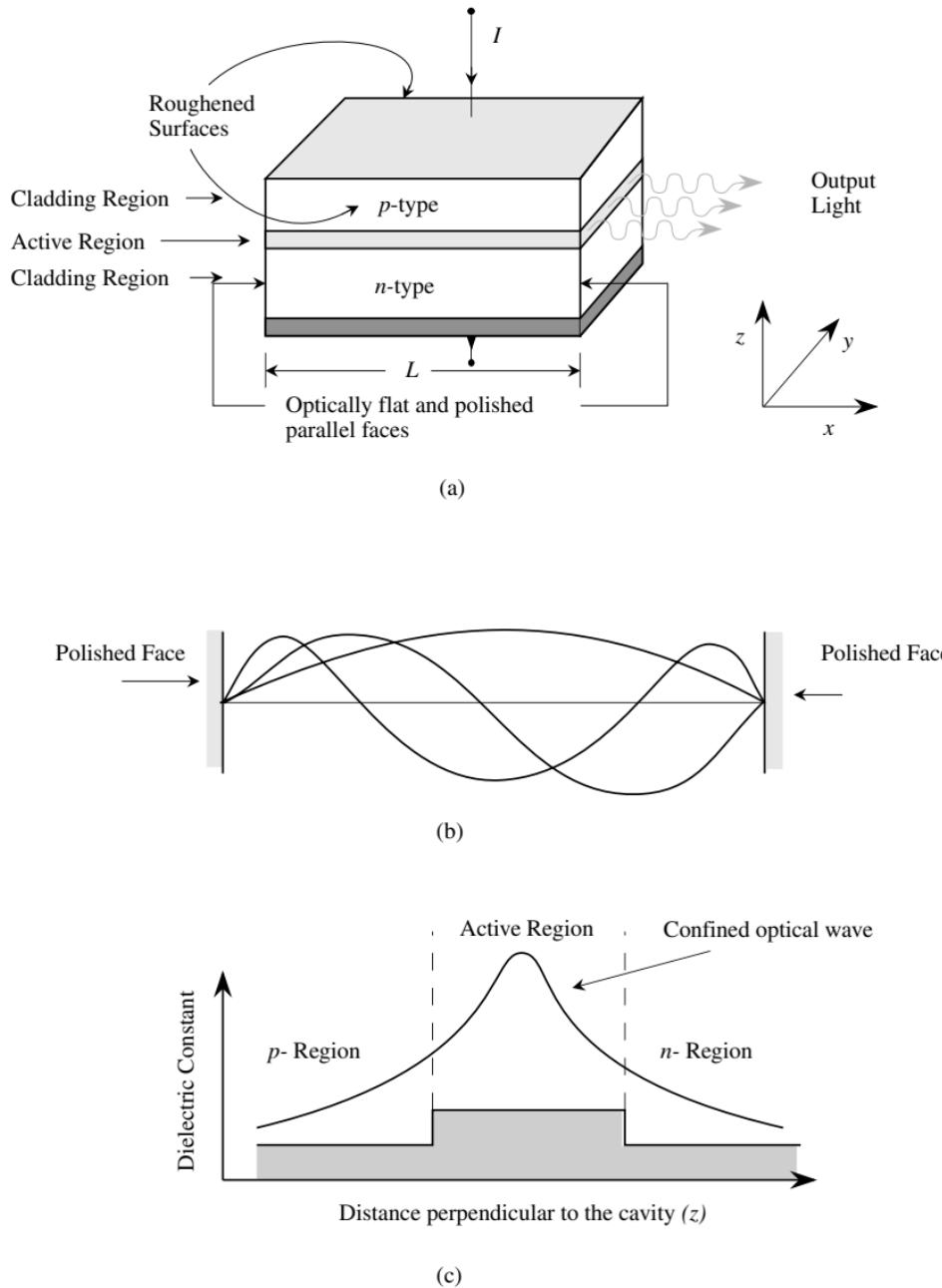


Figure 5.12. (a) Schematic of a laser structure showing the optical cavity and mirrors that confine photons. The active region may be a simple double heterostructure or a more complex quantum well design. (b) Stationary cavity modes, established by the mirror-induced resonances. (c) Optical confinement resulting from spatial variations in the dielectric constant.

Optical Absorption, Loss and Gain

When excess carriers are injected into the active region of a semiconductor laser—either by forward biasing a *p-n* junction or through optical pumping—the optical gain in the medium transitions from negative to positive over a certain energy range. This behavior is typically illustrated in gain versus injection current plots.

The net optical gain is generated solely within the active region where electron-hole recombination occurs. This region is often extremely thin, especially in quantum well lasers. For such structures,

the relevant parameter is the *cavity gain*, defined as:

$$\text{Cavity gain} = g(\hbar\omega) \Gamma \quad (5.130)$$

where $g(\hbar\omega)$ is the material gain as a function of photon energy, and Γ is the optical confinement factor, which represents the fraction of the optical field overlapping with the gain region.

In double heterostructure lasers based on bulk semiconductors, Γ is typically close to unity. In contrast, in quantum well lasers, Γ may be as low as 0.01 due to the extremely thin active region. Nevertheless, quantum well lasers can achieve substantial cavity gain because the material gain g in such structures is significantly enhanced for a given injection level, owing to their two-dimensional carrier confinement.

For laser action to begin, the cavity gain must exceed the optical losses in the system. These losses, often grouped under the term α_{loss} , arise from two primary sources:

- i) Absorption of photons in passive regions of the structure, such as the cladding layers and electrical contacts;
- ii) Photon escape losses at the output facets of the cavity.

The dominant contribution to α_{loss} typically comes from free carrier absorption, which is a second-order interaction involving mobile charge carriers. In high-purity materials with minimal doping and low defect density, this loss can be as low as 10 cm^{-1} . It is important to note that cavity loss is sensitive to both the level of doping and the quality of the crystal structure. Another significant source of photon loss arises from the escape of photons through the cavity mirrors. This form of loss is termed the *cavity loss*, and it is expressed as:

$$\alpha_{\text{cavity}} = -\frac{1}{L} \ln R \quad (5.131)$$

where L is the cavity length, and R is the reflectivity of the mirror surfaces defining the resonant cavity. In the case of a semiconductor-air interface, the reflectivity R can be approximated by the Fresnel formula:

$$R = \left(\frac{n_r - 1}{n_r + 1} \right)^2 \quad (5.132)$$

where n_r is the refractive index of the semiconductor material. It follows that minimizing optical loss requires maximizing the reflectivity R , which can be achieved using high-quality dielectric coatings or Bragg reflectors. Moreover, since the loss depends logarithmically on R , even small increases in reflectivity can lead to substantial reductions in α_{cavity} . This makes mirror design a critical factor in the efficiency of laser diodes, particularly in applications demanding low threshold currents and high output power.

Laser Below and Above Threshold

In this section, we analyze the light output as a function of injected current density in a laser diode. This corresponds to the light output in the lasing mode. When compared with an LED, there is a key difference: the laser diode exhibits an abrupt change in light output behavior around a specific threshold condition. This threshold condition occurs when the cavity gain overcomes the cavity loss for some photon energy, defined as

$$\Gamma g(\hbar\omega) = \alpha_{\text{loss}} - \frac{\ln R}{L} \quad (5.133)$$

In high-quality lasers, $\alpha_{\text{loss}} \sim 10 \text{ cm}^{-1}$, and reflection losses may contribute a similar amount. Another useful definition in laser operation is the condition of transparency, which occurs when the optical gain is zero:

$$\Gamma g(\hbar\omega) = 0 \quad (5.134)$$

Under forward bias, electrons and holes are injected into the active region, where they recombine and emit photons. Initially, at low injection levels, the number of carriers is insufficient to produce significant gain, and emitted photons are either absorbed or lost. As the bias increases, carrier injection increases until the gain reaches the threshold needed to balance the losses, allowing photon buildup in the cavity. Beyond this threshold, stimulated emission dominates spontaneous emission, resulting in a sharp increase in light output.

Let n_{th} be the carrier density at threshold. Then the threshold current density due to radiative recombination is given by

$$J_r(th) = \frac{en_{th}d_{\text{las}}}{\tau_r} = \frac{en_{th}(2D)}{\tau_r} \quad (5.135)$$

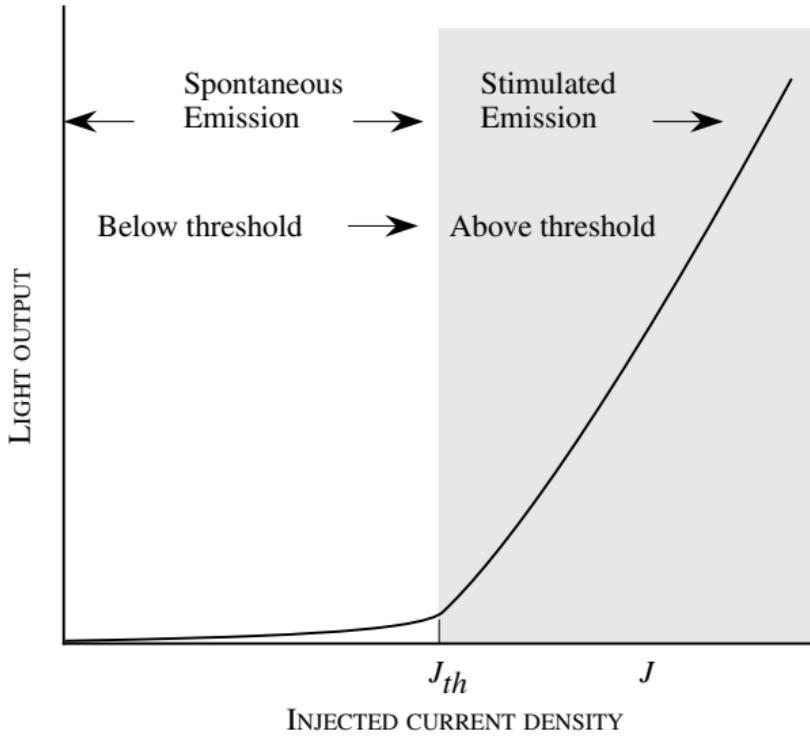


Figure 5.13.

where d_{las} is the thickness of the active region and $n(2D)$ is the areal carrier density. The radiative lifetime is τ_r , which under lasing conditions is given by:

$$\tau_r \sim \frac{\tau_0}{4} \quad (5.136)$$

In addition to the radiative current, we also have a non-radiative current J_{nr} . For Auger processes, we have:

$$J_{nr} = eF n^3 d_{\text{las}} \quad (5.137)$$

The total threshold current density is then:

$$J_{th} = \frac{en_{th}(2D)}{\tau_r} + eF n_{th}^3 d_{\text{las}} \quad (5.138)$$

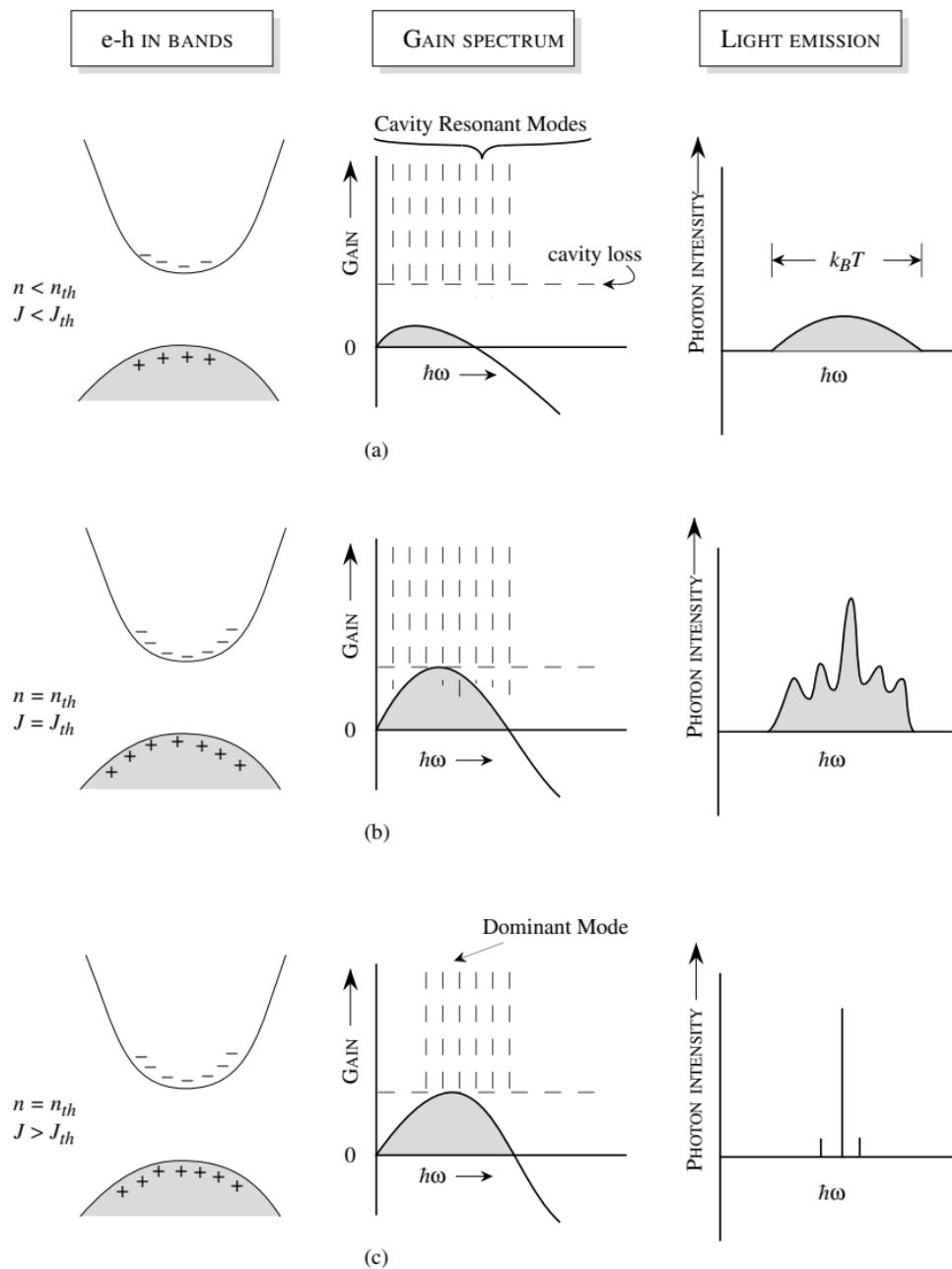


Figure 5.14. (a) Laser operating below threshold: gain is insufficient to overcome cavity losses, and emission is broad, similar to an LED. (b) At threshold: a few resonant modes begin to dominate the emission spectrum. (c) Above threshold: the gain spectrum remains unchanged, but stimulated emission causes a single mode to dominate the output.

Excitonic Effects

6.1 Introduction

The bandstructure and optical properties of semiconductors discussed so far have been based on the assumption that the valence band is completely filled with electrons while the conduction band remains empty. Under this assumption, the effects of carriers—electrons in the conduction band and holes in the valence band—enter only through the occupation probabilities, without altering the underlying bandstructure.

In practice, however, there exists a Coulombic interaction between electrons and other charged carriers, such as additional electrons or holes. These interactions can lead to significant modifications in material properties. Let us now consider a situation in which a single electron occupies the conduction band, while a corresponding hole is present in the valence band. This configuration introduces a direct Coulomb attraction between the electron and hole, altering the Hamiltonian of the system.

The resultant electron–hole pair, bound by this Coulomb interaction, forms a new quasiparticle known as an exciton. The formation of an exciton necessitates a re-evaluation of the electronic bandstructure to incorporate the effects of this binding.

The study of excitonic transitions in quantum wells has emerged as a major area of research, driven by both fundamental physics and technological applications. In quantum confined systems, the exciton binding energy is significantly enhanced compared to bulk materials. Additionally, the strength of optical transitions is improved, leading to the observation of sharp resonance features in the optical spectra of quantum wells. These excitonic resonances are highly tunable using optical or electronic methods, making them especially valuable in device applications.

Modulation of optical properties is a critical component in the development of advanced optoelectronic technologies. Although several types of external stimuli—such as electric fields, magnetic fields, and mechanical strain—can influence the optical response of a material, only modulation by electric or electromagnetic fields can achieve high-speed performance. In bulk semiconductors, the modulation capabilities are generally limited. However, quantum wells exhibit markedly enhanced modulation behavior, which makes them ideal candidates for integrated photonic and optoelectronic systems.

In this chapter, we will explore the physical principles underlying various approaches to optical modulation, with a particular focus on quantum-confined systems.

6.2 Excitonic States in Semiconductors

The electron–hole pair forms a bound state that can be described by an envelope function. This envelope function, representing the spatial distribution of the bound state, can be derived by treating the Coulombic interaction between the electron and hole as a perturbation.

There are two major categories of excitons, distinguished by the spatial extent of their envelope functions. When the envelope function is confined to only a few unit cells, the exciton is referred to as a Frenkel exciton. Owing to their highly localized nature, the Heisenberg uncertainty principle implies that such excitons require a treatment that incorporates the full electronic bandstructure of the material.

These localized excitons contrast with another class—commonly encountered in semiconductors—where the envelope function is much more extended. Each class of exciton gives rise to different optical and transport properties and must be described using appropriate theoretical frameworks.

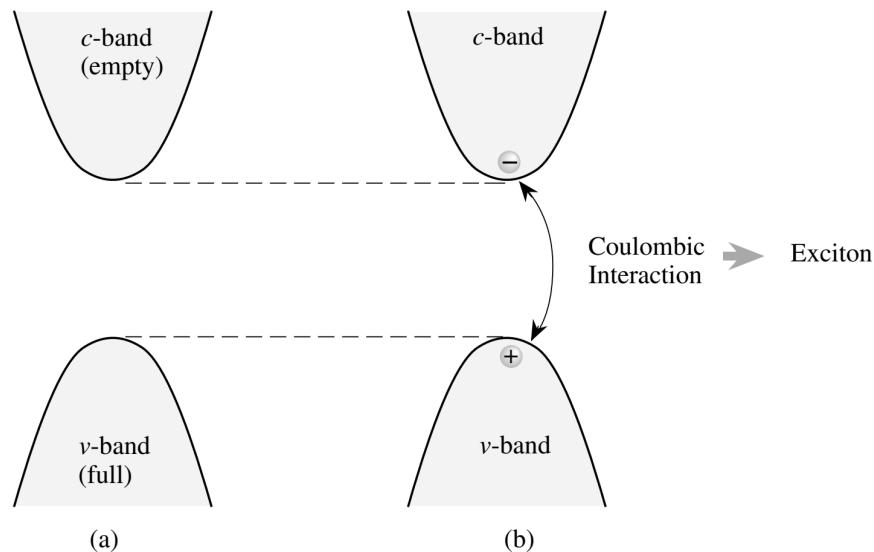


Figure 6.1. (a) Bandstructure under the independent electron approximation. (b) Modification of the bandstructure due to Coulomb interaction between an electron and a hole.

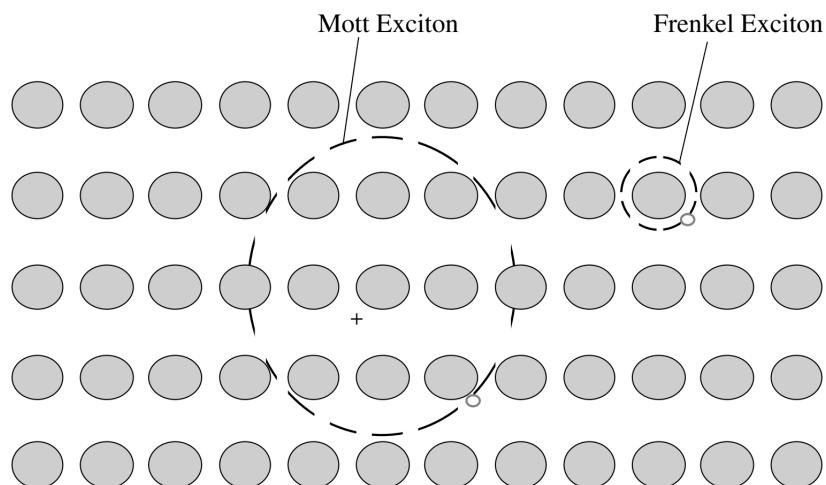


Figure 6.2. Conceptual illustration of the periodic envelope functions of Frenkel and Mott excitons. The Frenkel exciton is localized over a few unit cells, while the Mott exciton extends across many unit cells.

Such excitons, which arise when the electron and hole form a spatially extended bound state, are referred to as Mott excitons. These excitons are central to the excitonic physics in semiconductors. Their properties can be described using the effective mass approximation, whereby the electron–hole interaction is incorporated into a modified Schrödinger equation of the form:

$$\left[-\frac{\hbar^2}{2m_e^*} \nabla_e^2 - \frac{\hbar^2}{2m_h^*} \nabla_h^2 - \frac{e^2}{4\pi\varepsilon|\mathbf{r}_e - \mathbf{r}_h|} \psi_{ex} \right] = E\psi_{ex} \quad (6.1)$$

Here, m_e^* and m_h^* are the effective masses of the electron and hole, respectively, and $|\mathbf{r}_e - \mathbf{r}_h|$ denotes the distance between the electron and hole, which determines the strength of their Coulomb interaction. E_{ex} is the energy of the exciton. This formulation presents a two-body problem, which can be simplified by transforming to center-of-mass and relative coordinates:

$$\mathbf{r} = \mathbf{r}_e - \mathbf{r}_h, \quad \mathbf{R} = \frac{m_e^*\mathbf{r}_e + m_h^*\mathbf{r}_h}{m_e^* + m_h^*} \quad (6.2)$$

and corresponding wavevectors:

$$\mathbf{k} = \frac{m_e^*\mathbf{k}_e + m_h^*\mathbf{k}_h}{m_e^* + m_h^*}, \quad \mathbf{K} = \mathbf{k}_e + \mathbf{k}_h \quad (6.3)$$

The total Hamiltonian can then be written as:

$$H = \frac{\hbar^2 K^2}{2(m_e^* + m_h^*)} + \left(\frac{\hbar^2 k^2}{2m_r^*} - \frac{e^2}{4\pi\varepsilon|\mathbf{r}|} \right) \quad (6.4)$$

The Hamiltonian consists of two parts: the first term describes the center-of-mass motion, which leads to a plane-wave solution:

$$\psi_{cm}(\mathbf{R}) = e^{i\mathbf{K}\cdot\mathbf{R}} \quad (6.5)$$

The second term describes the relative motion of the electron–hole pair. Its solution satisfies:

$$\left[-\frac{\hbar^2 k^2}{2m_r^*} - \frac{e^2}{4\pi\varepsilon|\mathbf{r}|} \right] F(\mathbf{r}) = E_n F(\mathbf{r}) \quad (6.6)$$

This equation is analogous to the hydrogen atom problem, and its solutions $F(\mathbf{r})$ are hydrogen-like wavefunctions. The full exciton wavefunction is then:

$$\psi_{n,\mathbf{K}_{ex}}(\mathbf{r}_e, \mathbf{r}_h) = e^{i\mathbf{K}_{ex}\cdot\mathbf{R}} F_n(\mathbf{r}) \phi_c(\mathbf{r}_e) \phi_v(\mathbf{r}_h) \quad (6.7)$$

Here, ϕ_c and ϕ_v represent the Bloch functions of the conduction and valence band edge states, respectively. The corresponding excitonic energy levels are given by:

$$E_{n,\mathbf{K}_{ex}} = E_n + \frac{\hbar^2}{2(m_e^* + m_h^*)} K_{ex}^2 \quad (6.8)$$

where the eigenvalues E_n are:

$$E_n = -\frac{m_r^* e^4}{2(4\pi\varepsilon)^2 \hbar^2 n^2}, \quad n = 1, 2, 3, \dots \quad (6.9)$$

The quantity E_n is measured relative to the conduction band minimum, i.e., the excitonic energy levels appear just below the bandgap. Typical binding energies for excitons in semiconductors range from approximately 2 to 6 meV.

This modified dispersion relation, which includes the Coulomb interaction, differs from the conventional energy–wavevector (E vs. k) relationship. Since the exciton is a composite particle, the appropriate quantum number is not the electron or hole momentum individually, but the total

crystal momentum \mathbf{K} of the electron–hole pair. When the Coulomb interaction is neglected, the bound-state energy levels vanish, and one recovers the usual free electron–hole dispersion relation.

A key consequence of excitonic formation is that, unlike in free-carrier models where the joint density of states begins at the band edge, the presence of excitons introduces a density of states below the bandgap. Nevertheless, due to momentum conservation, not all exciton states will couple effectively to photons.

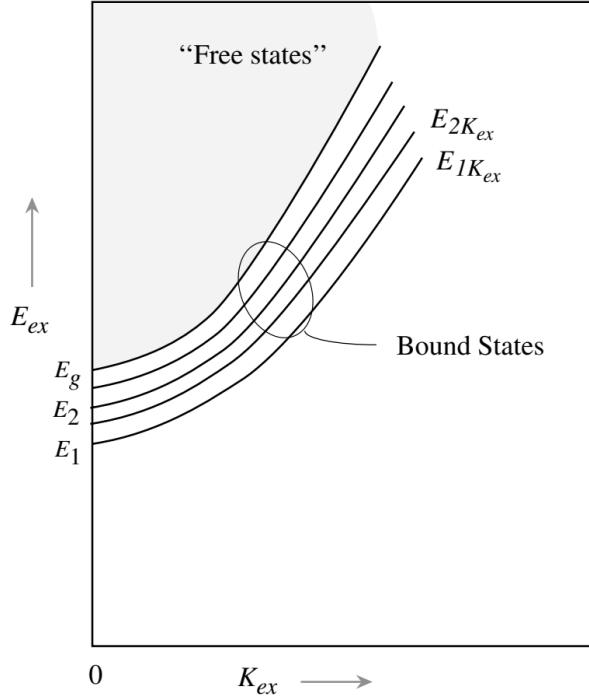


Figure 6.3. Dispersion curves for the electron–hole system in the exciton framework

To analyze the absorption spectra of excitonic transitions in semiconductors, it is useful to examine the problem in greater detail. As previously discussed, the independent-electron model yields the conduction and valence band states. When electron–hole pairs are introduced, their mutual Coulomb interaction can be treated as a perturbation. The resulting excitonic wavefunction can then be expressed as a linear combination of the independent electron basis functions.

The general form of the excitonic Hamiltonian is given by:

$$H = H_0 + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \quad (6.10)$$

Here, H_0 is the unperturbed Hamiltonian representing the independent electron bandstructure, and the second term accounts for the Coulomb interactions between different electrons. The factor of $\frac{1}{2}$ prevents double counting. Indices i and j label individual particles.

Because the full Hamiltonian maintains the symmetry of the underlying crystal lattice, the Bloch theorem remains applicable. Therefore, the total excitonic wavefunction must satisfy:

$$\psi_{\text{ex}}(\mathbf{r}_1 + \mathbf{R}, \mathbf{r}_2 + \mathbf{R}, \dots) = e^{i\mathbf{K}\cdot\mathbf{R}} \psi_{\text{ex}}(\mathbf{r}_1, \mathbf{r}_2, \dots) \quad (6.11)$$

where \mathbf{R} is a lattice vector and \mathbf{K} is the total crystal momentum of the exciton. The exciton state can be written in terms of a basis function $\Phi_{c,\mathbf{k}_e,S_e;v,\mathbf{k}_h,S_h}$, which represents a conduction band state where an electron, with momentum \mathbf{k}_e and spin S_e in the conduction band, and a

valence band hole with momentum \mathbf{k}_h and spin S_h . The difference $\mathbf{k}_e - \mathbf{k}_h$ is equal to the total momentum \mathbf{K} of the exciton.

The exciton wavefunction is expressed as:

$$\psi_{\text{ex}}^{n,\ell,m} = \sum_{\mathbf{k}} A_{n,\ell,m}(\mathbf{k}) \Phi_{c,\mathbf{k}+\mathbf{K}_{\text{ex}}/2,S_e;v,\mathbf{k}-\mathbf{K}_{\text{ex}}/2,S_h}^{n,\ell,m} \quad (6.12)$$

Here, n , ℓ , and m are quantum numbers associated with the exciton state, and $A_{n,\ell,m}(\mathbf{k})$ are expansion coefficients. Since we are dealing with envelope functions that are spatially extended (on the order of 100 Å), the expansion coefficients are expected to be sharply localized in momentum space.

We can define the Fourier transform of $A_{n,\ell,m}(\mathbf{k})$ as the real-space envelope function:

$$F_{n,\ell,m}(\mathbf{r}) = \sum_{\mathbf{k}} A_{n,\ell,m}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \quad (6.13)$$

This function satisfies a Schrödinger-like equation, analogous to the hydrogen atom, but modified for the semiconductor environment:

$$\left[E_{cv}(-i\nabla, \mathbf{K}_{\text{ex}}) - \frac{e^2}{4\pi\epsilon\mathbf{r}} \right] F_{n,\ell,m}(\mathbf{r}) = E_{\text{ex}} F_{n,\ell,m}(\mathbf{r}) \quad (6.14)$$

Here, $E_{cv}(-i\nabla, \mathbf{K}_{\text{ex}})$ results from expanding the difference $E_c(\mathbf{k} + \mathbf{K}_{\text{ex}}/2) - E_v(\mathbf{k} - \mathbf{K}_{\text{ex}}/2)$ in powers of \mathbf{k} and substituting $\mathbf{k} \rightarrow -i\nabla$. Exchange interactions are typically negligible and are therefore omitted. At low carrier densities, the dielectric constant ϵ can be approximated by its static value.

For a simple parabolic band structure, the exciton binding energy levels are:

$$E_n^{\text{ex}} = E_g - \frac{m_r^* e^4}{2(4\pi\epsilon)^2 \hbar^2 n^2} = E_g - \frac{R_{\text{ex}}}{n^2} \quad (6.15)$$

where m_r^* is the reduced mass of the electron–hole pair and R_{ex} is the exciton Rydberg. The kinetic energy associated with the center-of-mass motion of the exciton should be added to this expression for the total energy.

The corresponding envelope functions are hydrogen-like. For example, the ground state wavefunction is:

$$F_{100}(\mathbf{r}) = \frac{1}{\sqrt{\pi a_{\text{ex}}^3}} e^{-r/a_{\text{ex}}} \quad (6.16)$$

with the exciton Bohr radius given by:

$$a_{\text{ex}} = \frac{\epsilon m_0}{\epsilon_0 m_r^*} a_B, \quad \text{where } a_B = 0.529 \text{ Å} \quad (6.17)$$

For most semiconductors, a_{ex} is on the order of 100 Å, indicating that the exciton is spread over many unit cells. This justifies the use of the effective mass approximation in the description of such extended excitonic states.

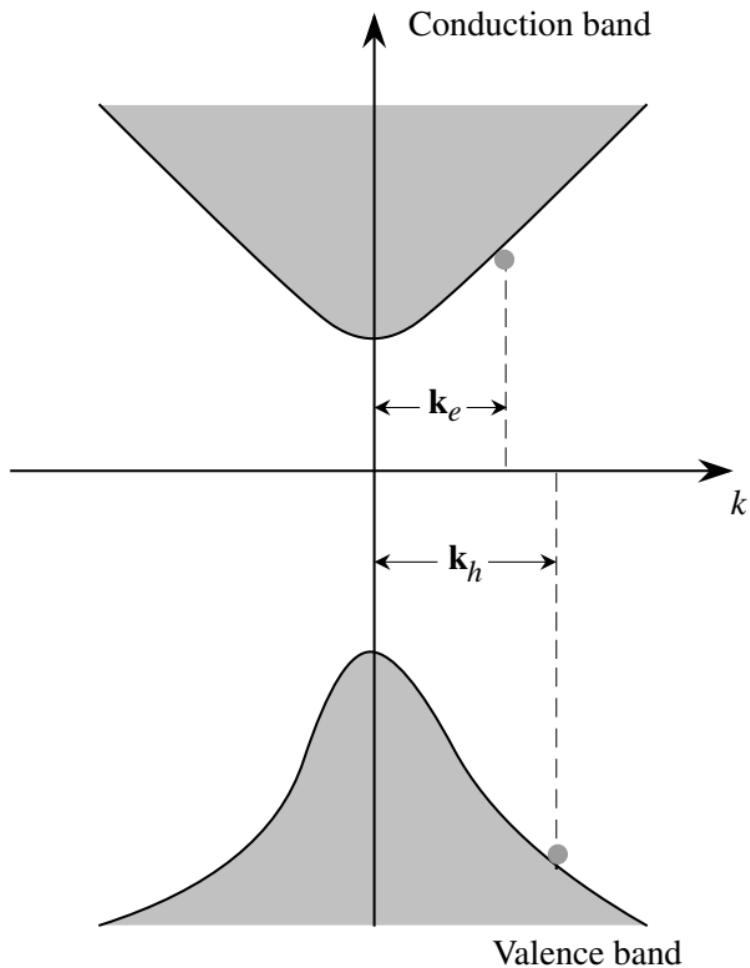


Figure 6.4. Schematic picture of an exciton in the Bloch representation.

Paper Exam

Properties of InAsSbN quantum well laser diodes operating at 2 μm wavelength region grown on InP substrates

Yuichi Kawamura^{a,b,*}, Naohisa Inoue^a

^aResearch Institute for Advanced Science and Technology, Osaka Prefecture University, 1-2 Gakuen-cho, Sakai, Naka-ku, Osaka 599-8570, Japan

^bCREST, Japan Science and Technology Agency, Kawaguchi Center Building, 4-1-8, Hon-cho, Kawaguchi-shi, Saitama 332-0012, Japan

Available online 7 February 2007

Abstract

Effects of rapid thermal annealing (RTA) on emission properties of InAsSbN quantum well laser diodes grown on InP substrates were studied. It was found that a marked enhancement of electroluminescence (EL) intensity about one order of magnitude and a blue-shift of the EL peak energy was observed at low temperature upon RTA. On the other hand, no blue-shift of the EL peak energy was observed near at room temperature. The observed EL peak blue-shift at low temperature can be explained by decrease in localized levels formed by nitrogen introduction in the InAsSbN quantum wells. Laser operation was obtained for both diodes with and without RTA. It was found that threshold current density J_{th} of the InAsSbN laser diode with RTA reduces compared to that without RTA. The lasing wavelength of the diode before RTA is 2.31 μm at 190 K, while that of the diode after RTA is 2.28 μm at 210 K. The T_0 values are 60 K for both diodes.

© 2007 Elsevier B.V. All rights reserved.

PACS: 78.20.-e

Keywords: A3. Molecular beam epitaxy; A3. Quantum wells; B2. Semiconducting III-V materials

1. Introduction

Semiconductor lasers operating at a 2-μm-wavelength region are highly desirable for many applications, such as pollution monitoring, chemical gas analysis and medical diagnostics. Thus far, most of these lasers have been developed for GaSb-based material systems, such as InGaAsSb/AlGaAsSb quantum well lasers [1,2]. It is very convenient to obtain 2-μm-wavelength lasers using InP-based material systems, since the technology for InP-based lasers has been well developed during the research on 1.3–1.5 μm lasers for optical communication systems. Room temperature operation up to 2.2 μm wavelength was reported for InAs strained quantum well lasers grown on InP substrates [3]. In order to get longer wavelength

operation, laser diodes using dilute nitride III-V compound semiconductors, such as InAsN, InGaAsN and InGaAsSbN quantum well lasers grown on InP substrates were also reported [4–6]. Recently, we reported fabrication of InAsSbN quantum well laser diodes grown on InP substrates by molecular beam epitaxy (MBE), where 2.51 μm electroluminescence (EL) at room temperature and 2.31 μm laser operation at 190 K were obtained without annealing [7]. In this paper, effects of rapid thermal annealing (RTA) on emission properties of InAsSbN quantum well laser diodes were studied. It was found that a marked enhancement of EL intensity about one order of magnitude and a blue-shift of the EL peak energy was observed at low temperature upon RTA. However, at room temperature, no blue-shift of the EL peak energy was observed and the EL intensity is almost unchanged. Laser operation was obtained for both diodes with and without RTA. It was found that J_{th} of the diode with RTA decreases compared to that without RTA. Laser operation up to 210 K was obtained for the diode upon

*Corresponding author. Research Institute for Advanced Science and Technology, Osaka Prefecture University, 1-2 Gakuen-cho, Sakai, Naka-ku, Osaka 599-8570, Japan

E-mail address: kwmr@riast.osakafu-u.ac.jp (Y. Kawamura).

RTA with the lasing wavelength of $2.28\text{ }\mu\text{m}$ and the threshold current density of 4.2 kA/cm^2 .

2. Experimental procedure

The InAsSbN quantum well laser diodes were grown on Sn-doped n-type (100) InP substrates by plasma-assisted MBE. The schematic layer structure is shown in Fig. 1. It was composed of a Si-doped n-type lower $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ cladding layer (thickness: $1\text{ }\mu\text{m}$, carrier density: $1 \times 10^{18}\text{ cm}^{-3}$), a Si-doped n-type lower $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}(2.5\text{ nm})/\text{In}_{0.52}\text{Al}_{0.48}\text{As}(2.5\text{ nm})$ superlattice (SL) optical confinement layer (thickness: 150 nm , carrier density: $5 \times 10^{17}\text{ cm}^{-3}$, effective band-gap: 1.05 eV), an undoped lower $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ optical confinement layer (thickness: 100 nm), undoped 3 nm $\text{InAs}_{0.846}\text{Sb}_{0.14}\text{N}_{0.014}$ double quantum well active layers separated by a 10 nm $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}_{0.986}\text{N}_{0.014}$ barrier layer, an undoped upper $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ optical confinement layer (thickness: 100 nm), a Be-doped p-type upper $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}(2.5\text{ nm})/\text{In}_{0.52}\text{Al}_{0.48}\text{As}(2.5\text{ nm})$ SL optical confinement layer (thickness: 150 nm , carrier density: $5 \times 10^{17}\text{ cm}^{-3}$), a Be-doped p-type upper $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ cladding layer (thickness: $2\text{ }\mu\text{m}$, carrier density: $1 \times 10^{18}\text{ cm}^{-3}$), and a Be-doped p-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ cap layer (thickness: 300 nm , carrier density: $2 \times 10^{18}\text{ cm}^{-3}$). All the layers, except the strained InAsSbN/InGaAsN double quantum well active layer (the effective compressive strain is 1.32%), were lattice matched to the InP substrates. The growth temperature was $505\text{ }^\circ\text{C}$, which was monitored using a calibrated infrared pyrometer. In, Ga, and Al metals were used as group III beam sources, while tetramers As_4 and Sb_4 were used as group IV beam sources. Nitrogen was supplied using an electron

cyclotron resonance (ECR) plasma source. The nitrogen gas flow rate was 0.7 sccm with an ECR power of 40 W . The background pressure during growth was about $9 \times 10^{-3}\text{ Pa}$. The InAsSbN quantum well laser diodes had a broad-stripe structure of $100\text{ }\mu\text{m}$ width. The cavity length was about $500\text{ }\mu\text{m}$. AuGeNi and AuZnNi were used as n-electrodes and p-electrodes, respectively. We compared two kinds of the InAsSbN quantum well laser diodes with and without RTA, which was carried out at $600\text{ }^\circ\text{C}$ for 30 s in a nitrogen cover-gas environment.

3. Results and discussion

Fig. 2 shows the EL spectra of the InAsSbN quantum well laser diodes at 10 K before and after RTA. The injection current is 50 mA . EL measurements were carried out using a standard lock-in amplifier technique. EL was detected by a cooled InSb photo-detector. It is known from this figure that upon RTA, a marked enhancement of EL intensity about one order of magnitude is obtained at 10 K . This EL intensity enhancement is probably due to decrease in non-radiative recombination centers related to nitrogen doping in to the InAsSbN active layers. In addition, the EL peak energy shows a blue-shift ($\sim 15\text{ meV}$) upon RTA. It was reported that in the case of InGaAsN layers, a rearrangement of the N nearest-neighbor environments from Ga-rich to In-rich causes a band gap blue-shift [8]. However, in the InAsSbN layers studied here, N is always bounded with In atoms. Therefore, the observed peak blue shift is considered to be induced another mechanisms. Another possibility is improvement of uniformity of nitrogen atom distribution. In fact, the narrowing of the

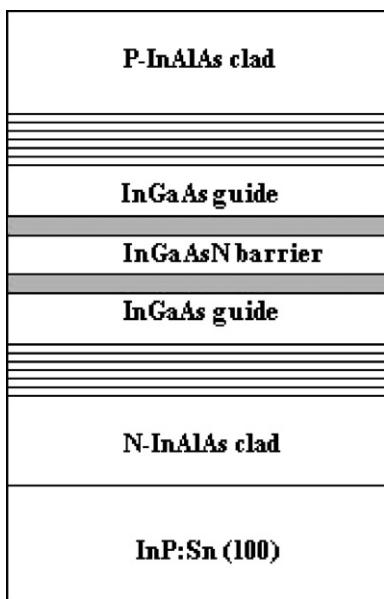


Fig. 1. Schematic layer structure of the InAsSbN quantum well laser diodes.

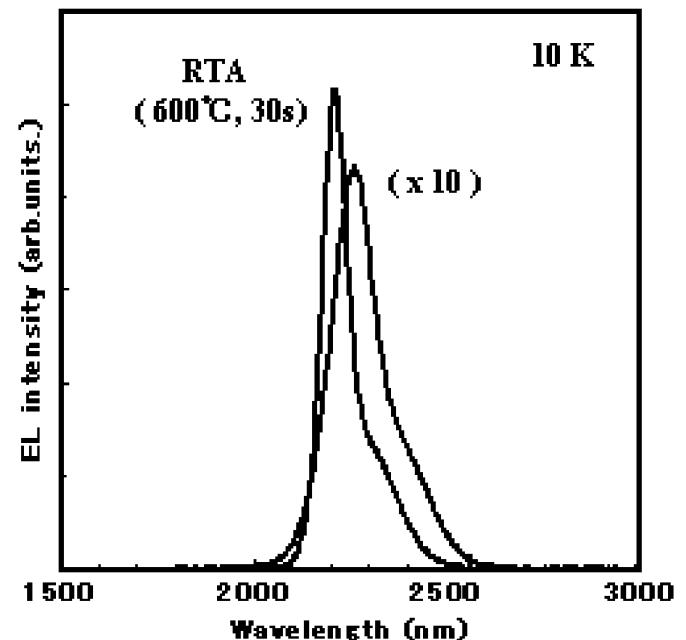


Fig. 2. EL spectra of the InAsSbN quantum well laser diodes with and without RTA at 10 K .

spectral half-width can be explained by this effect. However, no change of the EL spectrum at room temperature can not be explained by the improvement of uniformity of the nitrogen atom distribution.

On the other hand, it was reported that defect-related relatively deep localized levels were formed in the InGaAsN layer on GaAs in the concomitant presence of both In and N [9]. Probably, defect-related localized levels are formed in the InAsSbN quantum well layers, and RTA at 600 °C reduces these localized levels. The decrease in the localized levels causes the observed blue-shift of the EL peak energy at low temperature. In addition, it is noted that there is a shoulder structure at lower energy side of the EL spectra even after RTA. The origin of this shoulder structure is not clear at present. Probably, this shoulder structure is related to phase separation between Sb-rich region and As-rich region in the InAsSbN layer, although further studies are still necessary.

In order to clarify the mechanism of the effect of RTA on the EL spectrum, the temperature dependence of EL peak energy was measured in the range from 10 to 300 K, which is shown in Fig. 3. It is clearly seen that the peak energy difference between the diode with RTA and without RTA is about 15 meV at 10 K. However, with increasing temperature, the peak energy difference becomes small, and almost the same value at room temperature for both diodes. If the peak blue-shift at low temperature is caused by a band gap change, the peak blue-shift should be observed even at room temperature. Therefore, the result shown in Fig. 3 indicates that the peak energy shift at low temperature is not caused by a band-gap change, but is caused by a decrease of localized levels upon RTA, as is already explained.

Fig. 4 shows the temperature dependence of the EL peak intensity of the InAsSbN diodes before and after RTA. It

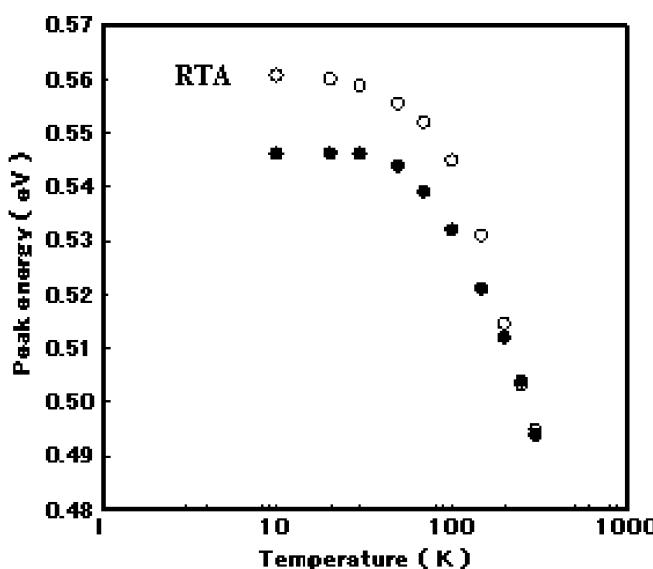


Fig. 3. Temperature dependence of the EL peak energy of the InAsSbN quantum well laser diodes with and without RTA.

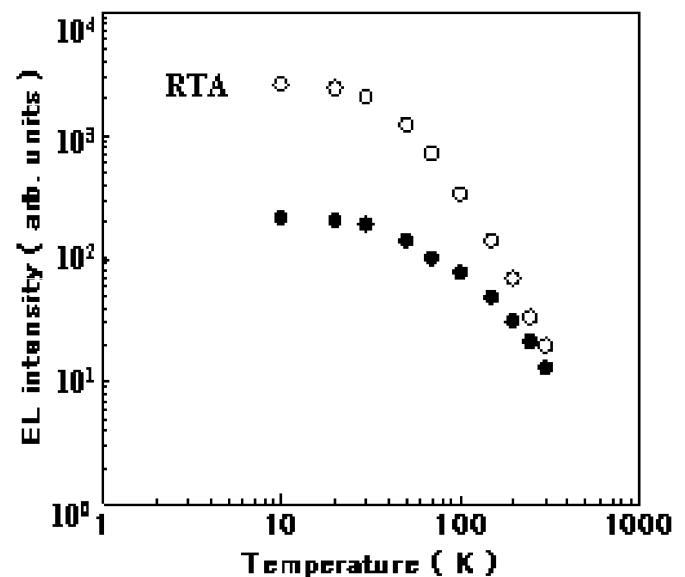


Fig. 4. Temperature dependence of the EL peak intensity of the InAsSbN quantum well laser diodes with and without RTA.

was known from the figure that the EL intensity of the diode with RTA is about one order of magnitude larger than that without RTA at low temperature. However, near room temperature, the EL intensity becomes comparable for both diodes. This result also suggests that the EL enhancement of the diode with RTA at low temperature is related to the decrease of localized centers formed by nitrogen incorporation. Upon RTA, these localized centers diminished, which is consistent with the result shown in Fig. 3.

Laser operation was achieved for both diodes with and without RTA. Figs. 5(a) and (b) show the lasing spectrum of the laser diode before and after RTA under a pulsed condition. The lasing wavelength of the diode before RTA is 2.31 μm at 190 K with the threshold current density J_{th} of 4.2 kA/cm², while that of the diode after RTA is 2.28 μm at 210 K with the J_{th} of 4.4 kA/cm² (2.26 μm at 190 K with the J_{th} of 3.2 kA/cm²). The slight blue-shift of the lasing wavelength upon RTA is considered to correspond to the blue shift of the EL peak energy at low temperature upon RTA. Lasing experiments at higher temperature region could not be carried out, because the maximum supplied current is limited to 2 A for the pulse generator used here. If a narrower stripe electrode structure is used (for example, a SiO₂ stripe structure), higher temperature operation will be possible in these lasers.

Fig. 6 shows the temperature dependence of the threshold current density in the temperature range from 10 to 210 K. The T_0 value estimated in the temperature range from 50 to 210 K is 60 K for both diodes. The reduction in threshold current density for InAsSbN diode with RTA is probably due to the decrease in the non-radiative recombination centers upon RTA. Room temperature operation will be possible by the improvements of the laser structure as well as optimization of RTA conditions.

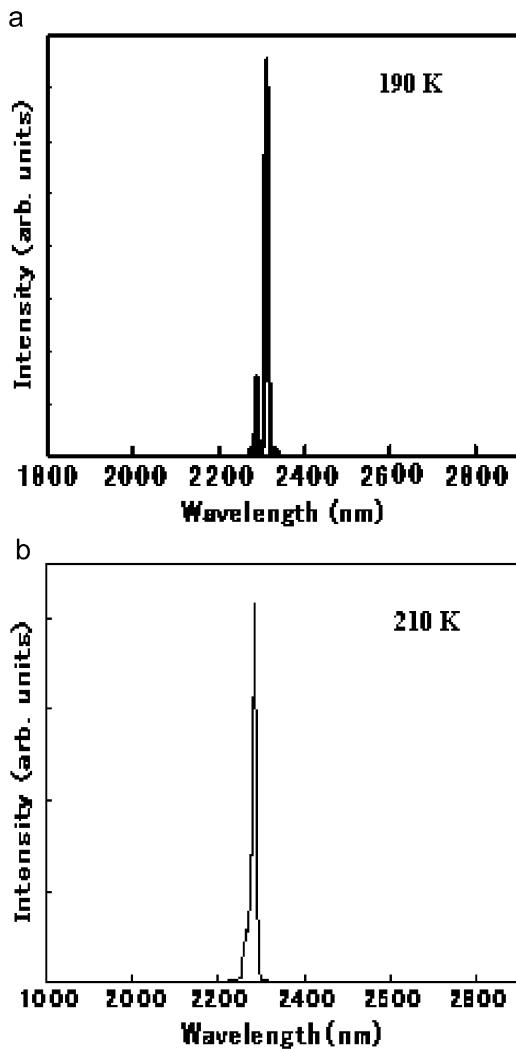


Fig. 5. (a) Lasing spectrum of the InAsSbN quantum well laser diode without RTA at 190 K. (b) Lasing spectrum of the InAsSbN quantum well laser diode without RTA at 210 K.

4. Conclusion

In conclusion, InAsSbN quantum well laser diodes operating at 2 μm wavelength region grown on InP substrates were fabricated by MBE and its emission properties were studied based on RTA effects. It was found that a marked enhancement of EL intensity about one order of magnitude and a blue-shift of the EL peak energy was observed at low temperature upon RTA at 600 °C for 30 s. However, no blue-shift of the EL peak

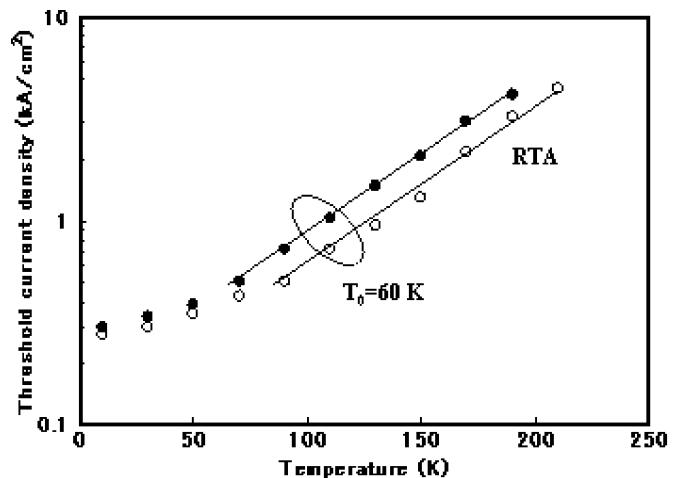


Fig. 6. Temperature dependence of the threshold current density in the temperature range from 10 to 210 K.

energy was observed and the EL intensity is almost unchanged at room temperature. These results suggest that the observed peak blue-shift is not caused by a band gap change, but caused by a reduction of localized levels formed by nitrogen doping upon RTA. Laser operation was obtained for both diodes with and without RTA. It was found that threshold current density J_{th} of the InAsSbN laser diode with RTA reduces compared to that without RTA. The lasing wavelength of the diode before RTA is 2.31 μm at 190 K, while that of the diode after RTA is 2.28 μm at 210 K. The T_0 values are 60 K for both diodes.

References

- [1] J.G. Kim, L. Shterengas, R.U. Martinelli, G.L. Belenky, D.Z. Garbuзов, W.K. Chan, *Appl. Phys. Lett.* 81 (2002) 3146.
- [2] W. Li, J.B. Heroux, H. Shao, W.I. Wang, *Appl. Phys. Lett.* 84 (2004) 2016.
- [3] J.-S. Wang, H.-H. Lin, L.-W. Sung, *IEEE J. Quantum Electron.* 34 (1998) 1959.
- [4] D.-K. Shih, H.-H. Lin, Y.H. Lin, *Electron. Lett.* 37 (2001) 1343.
- [5] K. Kohler, J. Wagner, P. Ganser, D. Serries, T. Geppert, M. Maier, L. Kirste, *J. Phys.: Condens. Matter* 16 (2004) S2995.
- [6] Y. Kawamura, T. Nakagawa, N. Inoue, *Jpn. J. Appl. Phys.* 43 (2004) L1320.
- [7] Y. Kawamura, T. Nakagawa, N. Inoue, *Jpn. J. Appl. Phys.* 44 (35) (2005) L1112.
- [8] P.J. Klar, H. Gruning, J. Koch, S. Schafer, K. Volz, W. Storz, W. Heimbroudt, *Phys. Rev. B* 64 (2001) 121203.
- [9] M.-A. Pinault, E. Toumié, *Appl. Phys. Lett.* 78 (11) (2001) 1562.

Bibliography

- [1] J. Singh, *Electronic and Optoelectronic Properties of Semiconductor Structures*. Cambridge University Press, 2007.