Chapter 2

THE PHYSICS OF SOLIDS

Before beginning a general discussion of electronic devices and the more complex aspects of semiconductors and other electronic materials, it is helpful to have an idea of their physics, especially their electronic structure. This chapter provides a partial review of the physics of solids. The nature of materials is determined by the interaction of their valence electrons with their charged nuclei and core electrons. This determines how elements react with each other, what structure the solid prefers, its optoelectronic properties and all other aspects of the material. The following sections describe the general method for understanding and modeling the energies of bands of electronic states in solids. A more detailed discussion of semiconductor bonding is provided in Chapter 5.

2.1 ELECTRONIC BAND STRUCTURES OF SOLIDS

There are two approaches taken when considering how the weakly bound (valence) electrons interact with the positively charged atomic cores (everything about the atom except the valence electrons) and with other valence electrons in a solid. We will consider first the direct approach of solutions to the differential equations that describe the motion of electrons in their simplest form and the consequences of this behavior. This requires many simplifying assumptions but gives a general idea for the least complex problems. The second approach is to follow the electronic orbitals of the atoms as they mix themselves into molecular states and then join to form bonding and antibonding combinations and finally bands of states, which we leave for Chapter 5. To see why it should be possible to mix electronic states in linear combinations, it is useful to consider some direct solutions of the Schrödinger equation, which governs the motion of electrons in an arbitrary potential. As we will see, this is a second-order linear differential equation. It therefore has particular solutions that can be constructed from linear combinations of any set of solutions of the general equation. Consequently, we may suspect that, at least in some cases, a linear combination of atomic orbitals should describe the general problem of a solid. In the following discussion we will assume that the material is a periodic crystal. Some of the results turn out to be applicable in most respects to aperiodic (amorphous) structures as well.

Electrons behave as both waves and particles. The consequences of their wave and particle nature are derived through the formalism of quantum mechanics. The requirement for conservation of energy and momentum forces the electrons to select specific states described by "quantum numbers," analogous to resonant vibrations of a string on a musical instrument. The "resonant" states associated with each set of quantum numbers results in a set of "wave functions" which describe the probability of finding an electron around a given location at a given time. The wave functions of the resonant states are found as follows.

The total energy, E_{tot}, of an electron is the sum of its potential and kinetic energies. In classical terms one could express this relationship as

$$E_{tot} = \mathbf{p}^2/2m + U(\mathbf{r}), \qquad 2.1$$

where $U(\mathbf{r})$ describes the local potential energy of the particle at position $\mathbf{r}$ and the kinetic energy is given by the classical expression $E_{kin} = \mathbf{p}^2/2m$ in which m is the particle mass and $\mathbf{p}$ is the momentum. This equation applies to any classical body. When the object in question is of a scale small enough that quantum mechanical behaviors become dominant, we need to rewrite Equation 2.1 in quantum mechanical terms. Under such conditions, the exact energy (E) at any specific time (t) can not be described to within an accuracy better than $\Delta E \Delta t = \hbar$ where $\hbar$ is Plank's constant divided by 2π. To account for this uncertainty the particle must be described by a probability distribution (its wave function) rather than by indicating a specific position. Its total energy is given by the change in the wave function, $\Psi(\mathbf{r},t)$, per unit time multipled by $i\hbar$, where $i^2=-1$. The momentum of a quantum particle is, likewise, the spatial derivative of the wave function multiplied by $i\hbar$. Based on the classical behavior and using the mathematics of operators, the kinetic energy, $\mathbf{p}^2/2m$ becomes the second derivative of the wave function times $-\hbar^2/2m$. Similarly, the potential energy represents a weighted average potential using Ψ as the weighting function. Substituting these expressions in Equation 2.1 yields:

$$i\hbar \frac{d\Psi}{dt} = -\frac{\hbar^2}{2m} \nabla^2 \Psi + U(\vec{r})\Psi, \qquad 2.2$$

where ∇^2 is referred to as the Laplacian and is the second spatial derivative of the function it operates on (in this case the wave function). Equation 2.2 is the full time-dependent Schrödinger Equation and describes not only the steady-state behavior of an electron but also the way in which the electron changes energy as a function of time. Whenever the potential that the electron experiences does not change with time, the time variable can be separated from the space variable. In this case, the energy of the particle cannot change with time, and the spatial-portion of the Schrödinger Equation becomes:

$$E\,\psi(\mathbf{r}) = (-\hbar^2/2m)\,\nabla^2\,\psi(\mathbf{r}) + U(\mathbf{r})\psi(\mathbf{r}), \qquad 2.3$$

where $\psi(\mathbf{r})$ is the time-independent wave function. The electronic structure of solids is derived by the solution of this equation under the boundary conditions appropriate to the solid being modeled. We will now consider some solutions to Equation 2.3.

2.1.1 Free electrons in solids

The simplest form of Equation 2.3 is the special case of $U(\mathbf{r})=0$, where there is no potential affecting the motion of electrons. For simplicity we will make the further restriction of considering only a one-dimensional problem. In this case,

$$E\psi(x) + \frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} = 0. \qquad 2.4$$

The general solution to this equation, obtained by Fourier transform methods, is a linear combination of two waves moving in the positive and negative x directions:

$$\psi(x) = A_+ e^{ikx} \pm A_- e^{-ikx}, \qquad 2.5$$

where A_+ and A_- are the amplitudes of the two waves, and $k = \frac{2\pi}{\lambda}$ is the wavenumber of the waves (electrons) with wavelength λ. The energies of these waves are determined by substituting Equation 2.5 into 2.4. The second derivative of $\psi(x)$ from Equation 2.5 is $d^2\psi(x)/dx^2 = k^2\psi(x)$. Thus,

$$-\frac{\hbar^2 k^2}{2m}\psi(x) + E\,\psi(x) = 0 \qquad 2.6$$

from which $E = \frac{\hbar^2 k^2}{2m}$. This holds for any A_+ and A_- but requires $A_+^2 + A_-^2 = 1$ for a wave with unit amplitude. The $\psi(x)$ are known as eigenvectors of Equation 2.4 and the energies are the eigenvalues. The momentum of this wave is $\boldsymbol{p} = \hbar\boldsymbol{k}$, thus **k** represents the electron momentum to within a factor of $\hbar$. This energy vs. wavenumber ["E(k)"] relationship is illustrated in Figure 2.1.

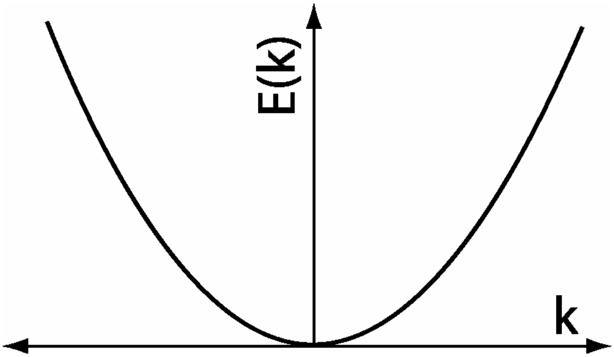

Figure 2.1: The energy vs. momentum diagram for a free electron in the absence of a periodic potential.

2.1.2 Free electrons in a periodic potential

In a solid there is a regular spacing of atoms. In a crystal, this spacing is defined by the translation vectors of the Bravais lattice. (See Chapter 4 for a description of semiconductor crystal lattices.) In an amorphous material the spacing is the average distance from one atom to its nearest neighbors. The result is the same – there is an imposed periodicity on the wave functions. The wave function must have the same value at equivalent positions in the solid. These positions are separated by lattice translation vectors **R**, thus $\psi(\mathbf{r}) = \psi(\mathbf{r}+\mathbf{R})$ in a crystal. For a wave of the form given in Equation 2.5, this imposes an additional condition that $\psi(x)=\psi(x+L_x)$ where L_x is the lattice spacing along the x direction. Likewise, the potential energy of a particle will be periodic (at least locally) such that $U(\mathbf{r}) = U(\mathbf{r}+\mathbf{R})$. The periodicity of ψ requires that:

$$\psi(\vec{r}) = e^{i\vec{k}\bullet\vec{r}} c_k(\vec{r}), \qquad 2.7$$

where $c_{nk}(\mathbf{r})=c_{nk}(\mathbf{r}+\mathbf{R})$ are the Fourier components for wave vector **k** of the wave function $\psi(\mathbf{r})$ [the proof is called "Bloch's Theorem"]. It can further be shown that the electron wave vector is given by **k**=**b**/N where **b** is a reciprocal lattice vector of the crystal lattice (see Chapter 4) and N is the number of unit cells in the [real-space] lattice. Therefore, any change in **k** must be by a unit vector of the reciprocal lattice, as in any diffraction problem. Electron waves in a solid are susceptible to scattering as one would have for x-rays and the problem can be represented with, for example, an Ewald sphere construction as for normal diffraction of x-rays in a periodic crystal.

The primary consequence of Equation 2.7 is that electron wave behaviors are reproduced whenever the wave vector **k** is changed by a translation vector of the reciprocal lattice, 2π/a, where a is the one dimensional lattice constant along a given direction. This means we can replicate Figure 2.1 every 2π/a units along the wave

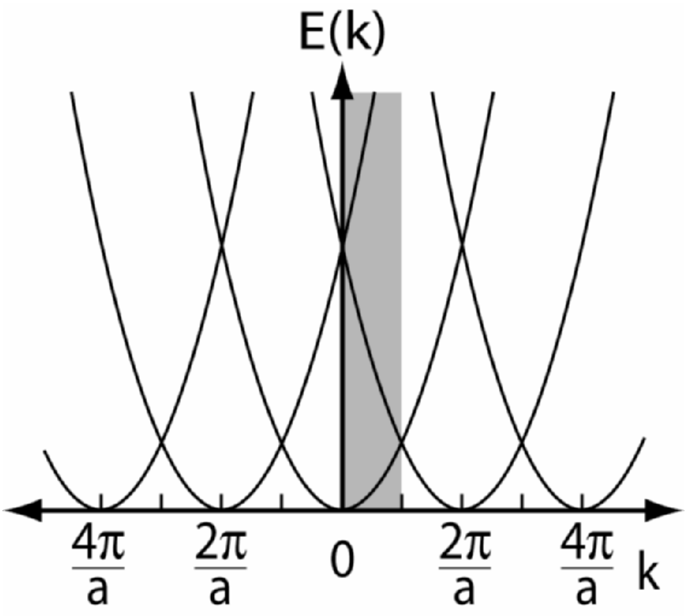

Figure 2.2: The periodic structure of the free electron energy vs. wave number in a periodic solid. The minimum section of the plot needed to provide a complete description of the relationship of E to k is shaded gray.

vector axis as shown in Figure 2.2. Note that if you consider the symmetry of this plot you will find that all of the necessary information is contained within the space between $0 < \mathbf{k} < \pi/a$. Thus, we will represent electron energy vs. wave vector plots in this reduced zone of **k** values hereafter.

2.1.3 Nearly free electrons

A somewhat more realistic picture is the case where U(**r**) is not zero or constant but varies weakly with position. This is the case referred to as "nearly free" electron behavior. The solutions to the Schödinger equation can be constructed from the same set of plane waves we had in Equation 2.7 with the proviso that U(**r**) is not too large. In this case, a general solution to Equation 2.3 is still given, at least approximately, by Equation 2.7 but now the c_{nk} must account for the effect of the periodic potential. This imposes the following constraint on the coefficients:

$$\left[\frac{\hbar^2}{2m}k^2 - E\right]c_k + \sum_{k'\neq k} U_{k'}c_{k'} = 0, \qquad 2.8$$

where the U_k terms are the Fourier components of the potential U for wavevector k. The free electron behavior is what is given inside the square brackets of Equation 2.8. Because the periodic potential is assumed to be small, the individual U_k terms are modest and the second term in Equation 2.8 represents only a minor perturbation on the result. Furthermore, because the c_k terms are Fourier coefficients, the c_k components can be obtained from the Fourier expansion of the U_k in the free electron plane waves. In other words, because both the wave function and the periodic potential can be expanded in the same Fourier terms, the c_k terms are related to the U_k terms via Equation 2.8. To construct a nearly free electron wave function, as modified by the U_k, from purely free electron waves we multiply each possible free electron wave by c_k and add the results to produce the new wave function solution in the presence of the periodic potential, as in Equation 2.7.

The U_k terms serve to mix the free electron plane waves producing interference effects. The closer the plane wave is to the periodicity of the lattice the more strongly it will interact with the crystal and, likewise, the stronger the component of the Fourier transform of U_k. Consider the interaction of two waves with the same or nearly the same wave vectors and energies. Graphically, the interactions occur near the points of intersection of curves in Figure 2.2. When the energy difference between different branches of the E(k) diagram (different curves in Figure 2.2) is large on a scale of the potential energy, then the behavior is essentially free electron like. However, near the intersection of two curves the energies are modified. Approximating the periodic potential with only its first Fourier component, then Equation 2.8 yields two equations for the two curves, which can be represented in matrix form as:

$$\begin{vmatrix} E-E_1 & -U \\ -U^* & E-E_2 \end{vmatrix} = 0\,, \tag{2.9}$$

where $E_1 = \frac{\hbar^2 k_1^2}{2m}$ and $E_2 = \frac{\hbar^2 k_2^2}{2m}$ are the free-electron-like behaviors for the two curves (subscripts 1 and 2) near the meeting point, E is the energy at the meeting point, and U is the first Fourier component of the periodic potential. Note that at the intersection, $k_1=k_2$ so that $E_1=E_2$. This condition has the solution

$$E = \frac{\hbar^2 k^2}{2m} \pm |U|. \tag{2.10}$$

See Figure 2.3 for an illustration of this situation. The result has two implications that are important:

- Waves interact with each other to raise or lower their combined energies. Graphically, when curves on the E(k) diagram intersect they may interfere resulting in local changes in their energies. Note that

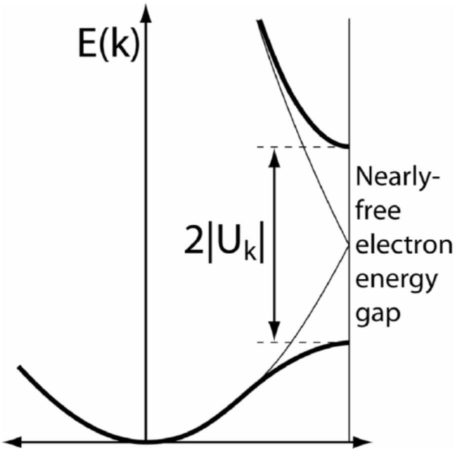

Figure 2.3: A schematic diagram showing the modification of the E(k) relationship near the zone boundary. (Gap is exaggerated to enhance visibility.)

waves which are perpendicular to each other can not interact and the branches of the E(k) diagram cross without modification.

- When the periodicity of the electron wave matches the periodicity of the lattice, the electrons are diffracted by the lattice (the waves interact strongly with the periodic potential and the Fourier component of U is likewise large).

The energy perturbation given in Equation 2.10 technically applies only at the boundary, $k_1=k_2$. As one moves away from the boundary, the magnitude of this interaction decreases quadratically with k. This is the same behavior as at the bottom of the free-electron curve. Another way to view this behavior is that near any extremum (maximum or minimum) of an arbitrary function, a power-law expansion of that function is always quadratic. Thus, near enough to any local maximum or minimum of an E(k) diagram, the behavior of an electron will always appear free-electron like. This provides a partial, if circular, justification of the approximations made above. A rigorous justification is provided by quantum mechanical perturbation theory and may be found in most quantum mechanics texts.

A somewhat more visual representation of the wave functions of the nearly free electron model is shown in Figure 2.4. The wave vector **k** of the electron wave $\psi(\mathbf{k})$ exactly matches the periodicity of the reciprocal lattice at any diffraction condition in a given direction (i.e.: it is a translation vector of the reciprocal lattice). As the wave approaches resonance with the lattice, the electrons interact increasingly strongly with the lattice potential. The magnitude of the lattice potential, U, then becomes

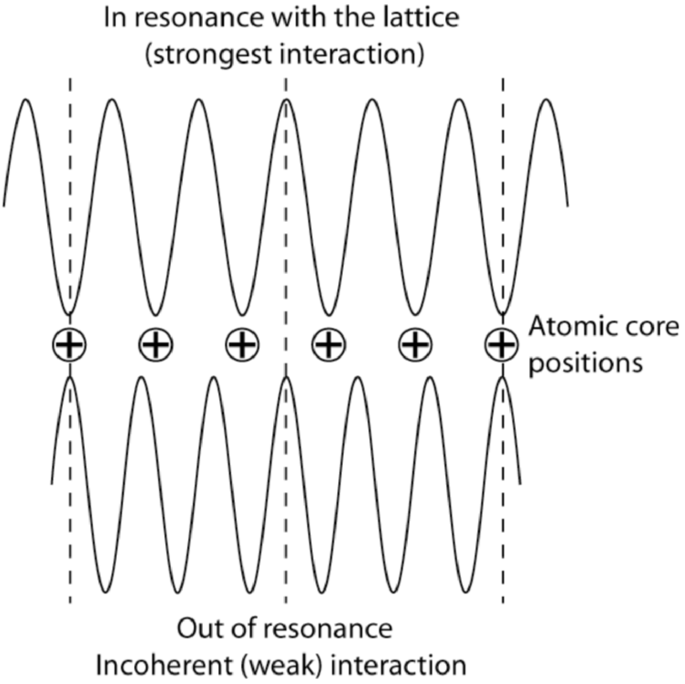

Figure 2.4: A schematic of the amplitudes of the wave functions of two waves. The top wave function shows the resonance of the zone boundary. The bottom wave function has a 90% shorter wavelength. The interaction of the wave with the atomic potential changes the electron energy.

critical. A strong potential affects the energy of the electron waves more, forcing a greater change in wave vector (or momentum) for a given change in kinetic energy. Thus, the parabolic band stretching away from the resonance curves more gradually. Electrons in such a material give the appearance of having a higher mass.

2.1.4 Energy vs. momentum in 3d

So far, we have considered only the nature of electron waves in a simple one-dimensional periodic lattice. Representing the three dimensional behavior of waves is significantly more difficult than representing one-dimensional wave behaviors, although the mathematics is not different. The first problem is to decide what are the wavelengths for diffraction from the lattice in multiple directions. This is solved by the Brillouin zone representation of the lattice.

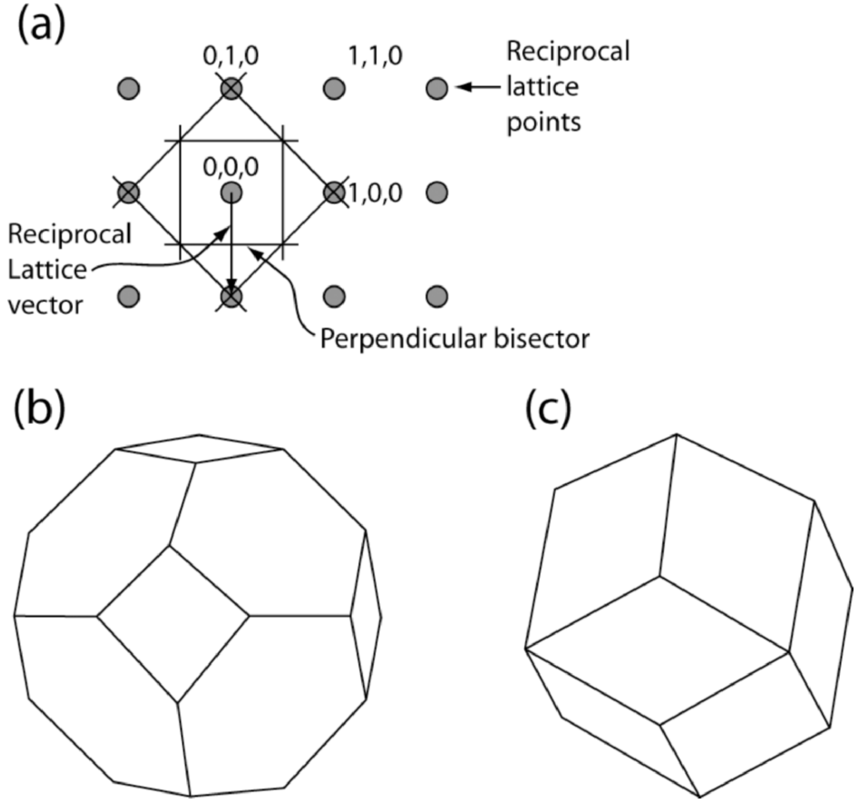

Figure 2.5: (a) The method for constructing the Brillouin zone for a square planar lattice and the first Brillouin zone for (b) a face-centered cubic crystal and (c) a body-centered cubic crystal. For the fcc crystal the diamond-shaped faces of the Brillouin zone are along cube axes, [100]-type directions, while the hexagonal faces are along [111] cube diagonals. For discussion of the [100], [111], and other crystal indices, see Chapter 4.

To construct this representation, one begins by calculating the reciprocal lattice of a given crystal structure (see Section 4.1.2). As described above, the reciprocal lattice gives a picture of the resonant wavelengths of the lattice in given directions. The second step is to determine the volume of reciprocal space closest to a given reciprocal lattice point. This begins by drawing lines (reciprocal lattice vectors) from the point at the origin of reciprocal space to all other points (Figure 2.5a). Each line is then bisected by a plane perpendicular to it. The volume of a given unit cell of reciprocal space is defined by the smallest volume contained within any combination of planes. This volume is referred to as the first Brillouin zone.

Two examples of first Brillouin zone shapes are shown in Figures 2.5b and c. The zone boundary points along primary symmetry directions for each crystal structure are traditionally labeled with letters. The point of zero crystal momentum is at the center of the Brillouin zone and is labeled Γ. The typical representations of energy band structures in three dimensions present the major trends along some of these symmetry directions but not over all of three-dimensional space.

If one carries out the free electron calculation for three dimensions, the complexity of the situation becomes clear, see Figure 2.6. The drawing shows slices through the first Brillouin zone of an fcc crystal along selected symmetry directions. These directions provide a relatively complete view of what the bands are like, although

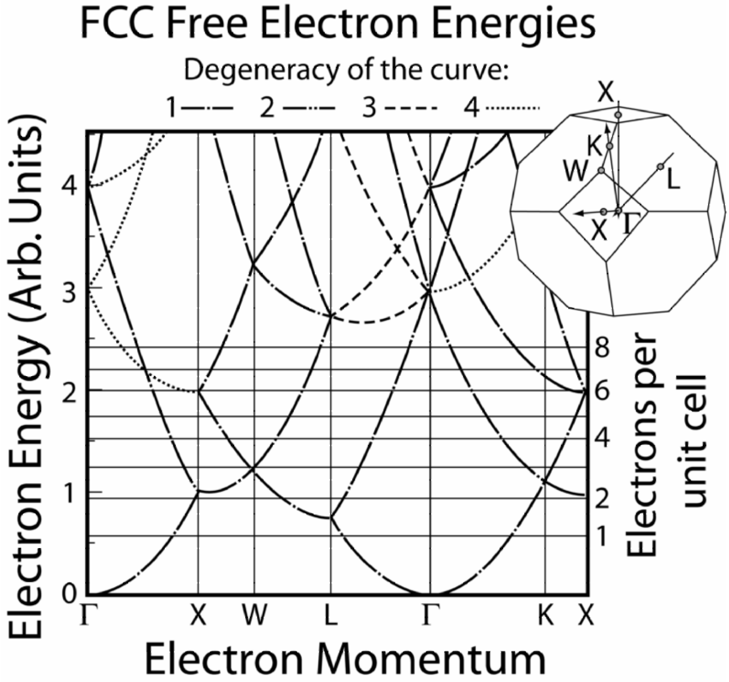

Figure 2.6: Shows the three-dimensional free-electron behavior plotted for a face-centered cubic lattice. Note that the X point is the zone boundary along [100], the L point along [111] and the K point along [110]. Γ is the zero momentum point at the center of the Brillouin zone where the electron wave has no momentum from the perspective of the crystal lattice (it is a standing wave with the lattice periodicity). Figure from Reference 10, Herman, F. in *An Atomistic Approach to the Nature and Properties of Materials* Pask, J.A., Editor New York: Wiley, 1967. Copyright 1967, John Wiley, used by permission.

this does not cover all of three-dimensional reciprocal space. Two typical energy band structures for the real semiconductors GaAs and Ge in three dimensions are shown in Figure 2.7.

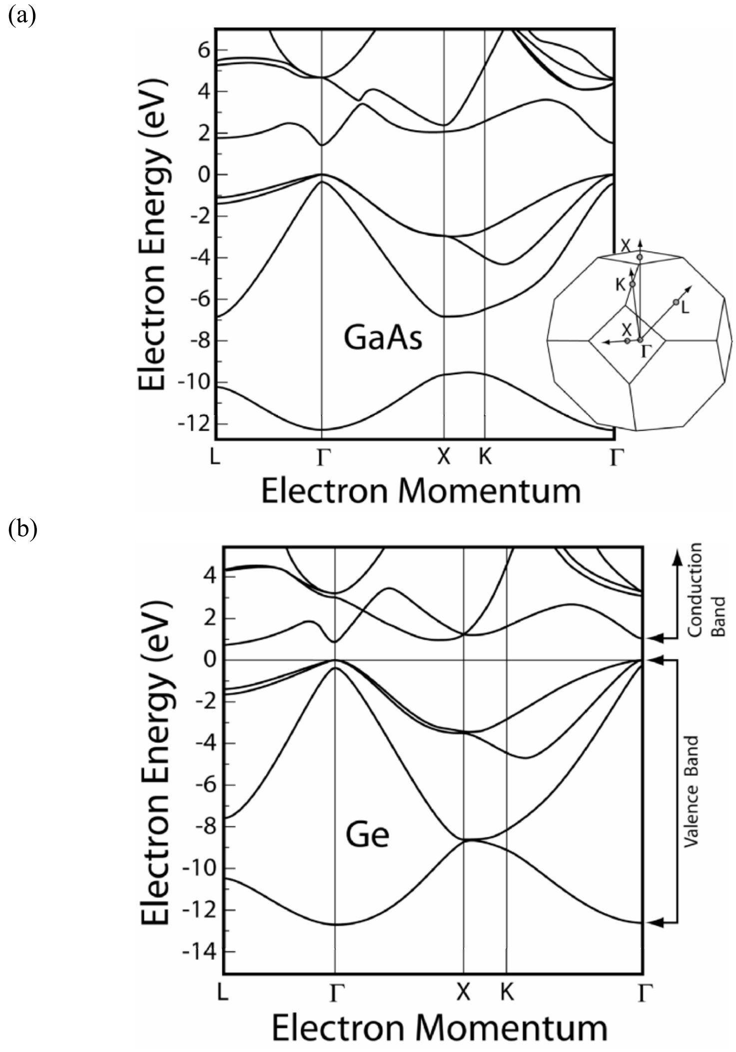

Figure 2.7: Energy vs. momentum diagram in three dimensions for (a) GaAs and (b) Ge. The directions in the Brillouin zone are shown in the inset and the letters indicate the position in momentum space. Figures adapted with permission from Chelikowsky, J.R. and Cohen, M.L. *Phys. Rev. B,* **14**, 556-582 (1976). Copyright 1976, American Physical Society.

The primary points to notice about the three-dimensional structures compared to the one dimensional structure (Figure 2.3.) are as follows:

- There are many branches to the diagrams in three dimensions. Many of them appear free electron like locally for the real materials. Consider, for example, the lowest-energy state in the real materials and the bottom of the nearly-free electron diagram.
- Behaviors along different directions vary because the crystal symmetry and resonant distances change.
- Some branches cross each other without interaction. Others interact, leaving a gap in the states at that point. Branches that do cross without interaction have perpendicular wave functions.
- Electrons fill states from the lowest energy up to a level known as the Fermi energy. In semiconductors, the lowest empty band of states makes up the conduction band. The filled states nearest the conduction band are the valence band and contain the valence electrons. Traditionally, the top of the valence band defines the zero energy, although band structure energies are often plotted relative to the vacuum level.
- A band of states (due to valence s-orbitals as we shall see in Chapter 5) lies just below the valence band in GaAs and other compound semiconductors. When this band does not connect to the upper valence band states it is formally a shallow core level, even though it is broadened by interaction with the other valence electrons. When a connection does occur as in Ge, it is formally part of the valence band. Usually these states are all considered part of the valence band whether there is a gap or not as the separation is small.
- The minimum energy in the conduction band does not necessarily occur at the same momentum as the maximum energy in the valence band. Therefore, an electron at the conduction band minimum may not have the same momentum as the lowest-energy electron vacancy (a hole) in the valence band.
- When the conduction band minimum occurs at the same momentum as the valence band maximum the semiconductor is termed “direct” (Figure 2.7a)
- An “indirect” semiconductor has these extrema at different momenta (Figure 2.7b).

2.1.5 Electrons and holes

In a semiconductor or insulator, there is a gap between the filled and empty electronic states. Charge can be carried either by electrons moving in the normally empty states of the conduction band or by the absence of an electron (a “hole”) moving in the normally filled states of the valence band. If one pictures current flowing in these bands as equivalent to water flowing in pipes, then an insulator or semiconductor can be viewed as two sealed pipes, one completely full of water and one completely empty. Even if one tips the pipes, no water flows. However, either droplets of water in the empty pipe or bubbles in the full pipe, if present, can allow net water transport when the pipes are tilted. The behaviors that apply to electrons also apply to holes (bubbles). Holes are referred to as “quasi-particles” because they do not exist in isolation. They are ghosts of electrons that are only known by the electron's absence in the solid. Nonetheless, in a solid, holes appear to have charge and mass and to be scattered as if they were real Fermions. For our purposes we will treat holes as if they were real particles.

To understand the transport of current it is important to keep in mind the fundamental nature of electrons and holes. Two of the more significant consequences of quantum theory are that electrons and other Fermions have a "spin" quantum number of ±1/2, and that no two spin ±1/2 particles can occupy the same state (can have the same set of values for their quantum numbers) at a time (the Pauli exclusion principle). Thus, filling a set of states begins with the lowest energy and adds electrons into higher and higher energy states as the more desirable levels become filled. This is not the case for integer spin "Bosons" such as photons, deuterium atoms, and helium 4 atoms. Bosons can all fill the same low energy state. In superconductors electrons form pairs having integral spin. These "Cooper" pairs act as Bosons with remarkable effects on the electronic properties of the material. Holes behave as do electrons but prefer states with higher energies (from the perspective of electrons) as do bubbles in water.

Electrons in solids are not scattered by each other without some external mediating force that allows one electron to lose energy or change direction by transfer of that energy or momentum to another electron. The possible transitions are limited by availability of empty and filled states (because electrons are Fermions), and thus so are the scattering events. Clearly then it will be important to know something about the distribution of these filled and empty states. The energy and momentum needed for scattering events is stored in the form of heat in the solid and is exchanged by absorption or emission of phonons (lattice vibrations) and photons. The probability of obtaining a given amount of energy from the remainder of the solid decreases exponentially as the energy needed increases. Likewise, the availability of phonons varies with temperature.

As noted above, at zero Kelvin, electronic states are occupied beginning with the lowest energy state and continuing upward until all electrons in the solid are accounted for. Abruptly, one would go from all states being filled to all being empty in this case. At higher temperatures, heat in the lattice puts random amounts of energy into individual electrons. Some of these electrons may have extremely high energies at times. The probability of finding an electron as a function of energy is given by the Fermi function:

$$f(E) = \frac{1}{1+e^{(E-E_f)/k_B T}}, \qquad 2.11$$

where the electron has an energy E relative to a reference state, the "Fermi energy" E_f. k_B is the Boltzman constant, and T is the temperature in Kelvin. There are several important facts about the Fermi function that are observable with brief inspection. First, when $E=E_f$ then the exponential is always 1 and the function value is 1/2. This is, in fact, the means of defining and determining the value of E_f. The Fermi energy is that energy for which $f(E_f)=0.5$. When E is very different from E_f then Equation 2.11 can be approximated by a Boltzmann distribution:

$$f(E) \approx e^{-(E-E_f)/k_BT} \tag{2.12}$$

It is common to use this distribution in all cases. However, one should bear in mind that this is an approximation and that for some situations its use may not be appropriate. The development of a distribution consistent with the Fermi function requires very little time (typically picoseconds) as electrons rapidly exchange photons and phonons with the remainder of the solid until this distribution is established.

One might ask how one can account for a situation (common in most semiconductors and insulators) in which the Fermi energy is within the energy gap. In this case, there are no states at the Fermi energy itself. The Fermi function remains valid and well defined and the probability of finding an electron drops exponentially moving away from E_f. Whatever remains of the function value above zero for electrons (or below unity for holes) describes the number of electrons/holes that are present in states in the bands at given energies, temperature, and Fermi level. If the distribution of states were the same in the conduction and valence bands and no states existed within the energy gap, then charge neutrality would require that the Fermi energy be near the gap center, as discussed in Section 2.2.1.

To establish the dynamic equilibrium distribution defined by Equation 2.11, electrons are constantly being transferred from the valence band to the conduction band by energy/momentum absorption. This is known as generation of free carriers. Carriers are also constantly falling back into the holes created by the earlier band-to-band transfer. This is recombination. Generation occurs by absorption of light or heat, while recombination releases light, in so-called “radiative” recombination, or heat (non-radiative recombination).

Generation by absorption of energy from the solid is “thermal” generation and from absorption of external light shining on the material is “optical” generation. The thermal generation rate, g_{th}, increases approximately as a Boltzmann distribution with the usual temperature dependence and an activation energy equal to that of the energy gap, E_g, of the material,

$$g_{th} = Ce^{-E_g/k_BT}, \tag{2.13}$$

where C is a proportionality constant. The constant may be determined based on the requirement that recombination must equal thermal generation if Equation 2.11 is to hold. For optical generation,

$$g_{op} = \int_0^\infty \frac{G(\nu)Q(\nu)}{h\nu} d\nu, \tag{2.14}$$

where G(ν) is the optical intensity at frequency ν, and Q(ν) contains the quantum-mechanical conservation rules and other aspects related to the band structure of the material and the optical absorption coefficient. Because optical generation does not depend upon temperature, a carrier distribution determined primarily by optical rather than thermal generation is not described by Equation 2.11.

Recombination depends upon the product of electron and hole densities and a factor giving the probability of a given energy release mechanism (for example, radiative or non-radiative). This probability includes the same quantum-mechanics-determined factors as in Equation 2.13 or 2.14, and incorporates the nature of the band structure. The recombination rate can be expressed as

$$r_{rec} = Q'np, \tag{2.15}$$

where Q’ is the a recombination rate per electron-hole pair including all possible recombination processes. Recombination occurs rapidly in direct-gap semiconductors (Q’ large) and slowly in indirect-gap semiconductors (Q’ small).

The details of the calculation of the factors C, Q, and Q’ above are complex and beyond what we will need here. However, they can be found described in texts on optical generation and recombination. See, for example recommended readings by Bube or Ashcroft and Mermin.

2.1.6 Direct and indirect semiconductors

The “direct” and “indirect” behaviors mentioned above are sufficiently important that they deserve special mention. The critical aspects of the energy band structures of these two types of semiconductor are shown schematically in Figure 2.8. The minimum energy of the conduction band in indirect materials is at a different momentum than that of the maximum energy of the valence band. Electrons in the conduction band rapidly relax to the minimum band energy. Holes equally rapidly move to the maximum energy of the valence band. Therefore, electrons and holes do not normally have the same momentum in an indirect semiconductor while in a direct-gap material these momenta are equal. This has consequences for the minority carrier lifetimes and optical properties of semiconductors.

Optical absorption/emission involves absorption/creation of a photon with a consequent change in energy of an electron, usually resulting in the transfer of that electron to/from the conduction band from/to the valence band. However, because photons have almost no momentum, only vertical transitions on an E(k) band diagram are allowed in purely optical processes. The only alternative to this is the rare case when a phonon is present together with a photon. Phonons have a large momentum (they involve collective motion of massive atoms) and relatively low energy. (See Section 2.3.3.1 for more discussion of phonons.) For practical purposes one can assume that photons lead to vertical transitions among bands on the E(k) diagram while phonons make horizontal transitions (see Figure 2.9.)

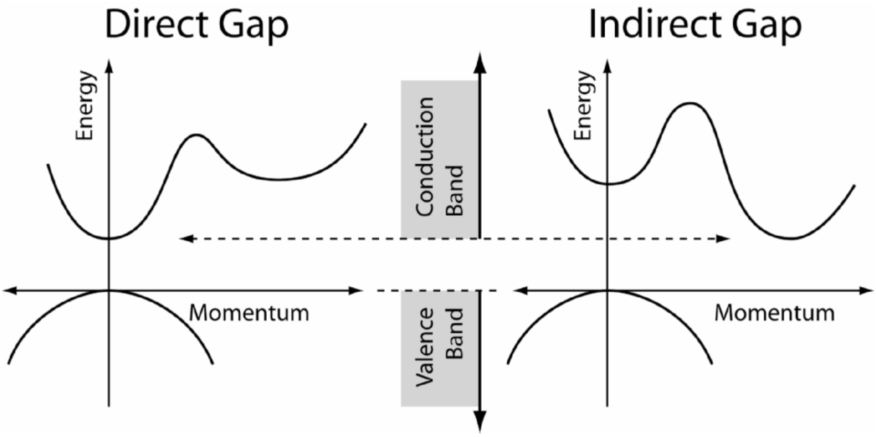

Figure 2.8: A schematic of the difference between a direct gap and an indirect gap semiconductor.

To make a diagonal indirect transition, both an appropriate energy photon and an appropriate momentum phonon must be present together with the electron. This is referred to as a three-body interaction because there are three particles (electron, phonon, and photon) participating. Such collisions are over 1000 times less likely than a simple electron-photon interaction at common temperatures. This means that electrons and holes of different momenta do not recombine rapidly. Typically, electrons and holes in pure direct-gap semiconductors last no more than ~10^{-8} s,

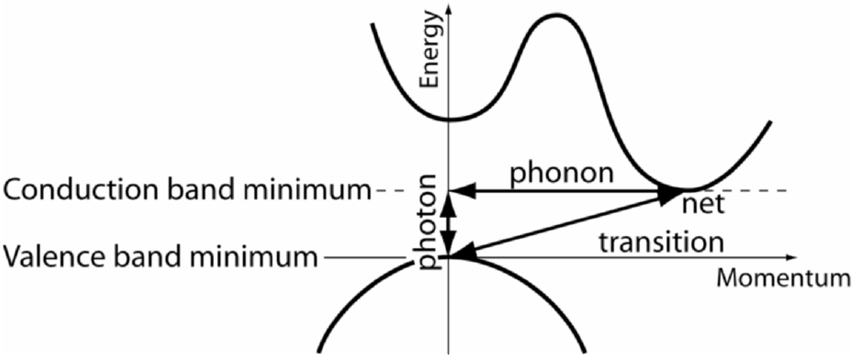

Figure 2.9: A schematic of the transitions produced by a phonon (horizontal) and a photon (vertical) on an E(k) diagram. Indirect transitions require both and thus are less likely.

while in indirect-gap materials the lifetime for free carriers can be 1000 times longer. Furthermore, when recombination does occur in indirect-gap materials, most of the energy is usually emitted in the form of heat in the solid rather than as light. This is why most light-emitting devices use direct-gap materials. Generation of carriers in indirect-gap materials is equally difficult so indirect materials tend to have low optical absorption coefficients below the point where direct-transitions become possible.

Because in indirect-gap semiconductors generation and recombination of carriers are equally difficult, even in indirect-gap materials the distribution of electrons and holes is governed by the Fermi function. Because the Fermi energy can never be farther than half the energy gap from one band edge or the other, the density of carriers in an indirect-gap material is still determined by the *minimum energy gap* in spite of the difference in momenta of the band minima in indirect-gap materials.

The consequences of the need for a phonon to permit an electron and a hole to interact in indirect-gap materials include:

- The maximum recombination time for electrons and holes is much longer (of the order of microseconds) in an indirect-gap material than in a direct gap semiconductor (nanoseconds).
- The probability of a photon emission or absorption event for energies between the minimum gap energy and the minimum direct gap energy (the lowest energy for a vertical transition) is far below that at or above the minimum direct gap energy.

2.1.7 Effective mass

The nearly-free electron model of solids shows that electrons are strongly affected by the lattice in which they move. The interaction of the electron with a solid modifies the relationship between energy and momentum. Recall that the total energy of an electron is given by the sum of kinetic and potential energies. Because the interactions of an electron with the lattice potential and the lattice spacing along a given direction vary from material to material, the change in kinetic energy with wavelength changes from one material to another. In some materials the energy gain is small for a given momentum increase. In others, it is larger for the same added momentum. Given the classical relationship between energy and momentum, $E=\mathbf{p}^2/2m$, the change in proportionality constant between energy and momentum appears as a change in the effective mass of the electron. The exact form of this relationship can be developed from the above energy band relationship as follows.

The energy of a free electron Bloch wave is related to the wave vector as:

$$E = \frac{\hbar^2 k^2}{2m^*} \qquad 2.16$$

where m^* is the effective electron mass. Differentiating both sides with respect to k,

$$\frac{k}{m^*}=\frac{1}{\hbar^2}\frac{dE}{dk} \tag{2.17}$$

and differentiating a second time,

$$\frac{1}{m^*}=\frac{1}{\hbar^2}\frac{d^2E}{dk^2}. \tag{2.18}$$

In other words, the effective mass of an electron Bloch wave is inversely related to the curvature of the E(k) diagram. Strong curvatures mean small effective masses and vice versa. The effective mass is generally expressed in units of electron rest mass in a vacuum, i.e. m^*/m_0.

The effective mass concept is only defined near a band maximum or minimum where the energy can be approximated as a quadratic function of wave vector, k. Note that near a minimum in energy bands the E(k) function is concave up and $1/m^*>0$ while near a maximum, $1/m^*<0$. The negative effective mass near a maximum indicates that charge is being carried by holes rather than electrons.

2.1.8 Density of states

One of the most fundamental properties of a material, which determines many of its properties, is its density of states. This refers to the number of states per unit energy in the band structure. To put this in more visual terms, if one takes a thin horizontal slice through an energy band structure diagram such as those shown in Figure 2.7, the "blackness" of the slice (the amount of band line that occurs in that slice) is the density of states. The density of states for a complex band structure can be computed and is normally developed as part of calculations describing a semiconductor (or other material). However, it is not straightforward to present a simple formula for the density of states of such a real system. We will have to be content with a derivation of the density of states for a free electron.

For a free electron in three dimensions the energy is given by

$$\mathrm{E}=\frac{\hbar^2}{2\mathrm{m}^*}\vec{\mathbf{k}}^2=\frac{\hbar^2}{2\mathrm{m}^*}\left(\mathbf{k}_\mathrm{x}^2+\mathbf{k}_\mathrm{y}^2+\mathbf{k}_\mathrm{z}^2\right) \tag{2.19}$$

where k_x, k_y, and k_z are the components of the wave vector **k** along the three coordinate directions. If one thinks of wave vectors as resonant states of the crystal, the longest wavelength or lowest energy state would be a half wavelength across the entire width of the crystal, L. Higher energy modes are integer multiples of this state. Therefore, the separation of states along any given direction, i, is π/L_i and the wave vector is then $k_i=n_i\pi/L_i$. Therefore, the energy can be rewritten (with $h=2\pi\hbar$) as:

$$E = \frac{h^2}{8mL^2}\left(n_x^2 + n_y^2 + n_z^2\right) = k_F^2 E_0 \qquad 2.20$$

where n_x, n_y and n_z are the indices of the reciprocal lattice points inside a sphere of radius $k_F = \sqrt{n_x^2 + n_y^2 + n_z^2}$. $E_0 = \frac{h^2}{8m^* L^2}$ is the lowest energy state of the system. The number of electrons that can be accommodated in states with energy E or less is:

$$N(E) = 2\left(\frac{1}{8}\right)\left(\frac{4\pi}{3}k_F^3\right) = \frac{\pi}{3}\left(\frac{E}{E_0}\right)^{3/2} \qquad 2.21$$

For any process to occur, an electron must be present in an initial state with energy E_i. The higher the density of states at this energy, the more electrons can participate in the process. The electron is changing energy by an amount ΔE. Therefore, there must be an empty state with energy $E_f = E_i + \Delta E$ to receive it. In addition, momentum must be conserved. The process begins from any initial state E_i such that the rest of the above criteria are met. The overall rate of the process is given by an integral over all initial states i containing electrons.

Here the factor of two is because each state has two possible electron spins and the 1/8 is because we must take only positive values of n_x, n_y, and n_z. The density of these states g(E) per unit volume of reciprocal space in an energy interval dE is given by (1/V)(dN/dE), where $V=L^3$ is the crystal volume (for a cube-shaped solid), or:

$$g(E) = \frac{1}{V}\frac{dN}{dE} = \frac{\pi}{2L^3}E_0^{-3/2}E^{1/2} = \frac{\pi}{2}\left(\frac{8m^*}{h^2}\right)^{3/2}E^{1/2} \qquad 2.22$$

This density of states will apply to any band extremum where the band can be approximated at least locally with a quadratic dependence of energy on momentum (free-electron-like). Since all band edges will have this general behavior, Equation 2.22 provides an approximate picture of the number of states per unit energy near a band edge.

The importance of the density of states may be found in the calculation of the rate of any process in a solid, from scattering of an electron off a defect or another electron to absorption and emission of light. The rate of such a process is given in its most general form by “Fermi’s Golden Rule”. Mathematically, the rate of a process (**H**) moving an electron from state ψ_i to state ψ_f may be written in symbols as follows:

$$r_{i\to f} = \int \big(g(E_i)f(E_i)\big)\big(g(E_f)(1-f(E_f))\big)\langle\psi_i\,|\,\mathbf{H}\,|\,\psi_f\rangle\,\delta\big(E_i - E_f + \Delta E\big)\,dE_i, \qquad 2.23$$

where E_i and E_f are the initial and final state energies, ΔE is the change in energy of the process, g_i and g_f are the density of states at energies E_i and E_f, $f(E_i)$ and $f(E_f)$ are the corresponding Fermi functions, and $\delta(E_i-E_f+\Delta E)$ is a Kroniker delta function which enforces conservation of energy. The expression $<\psi_i|\mathbf{H}|\psi_f>$ is a mathematical function which enforces the conservation of momentum of the electrons and other quantum mechanical selection rules in process **H**. In words this formula states the following.

Equation 2.23 is used in calculating quantities such as $Q(\nu)$ in Equation 2.14 and many other values. It is fundamental to detailed analysis of the physics of solids. Therefore, the density of states is an essential element of understanding the general behaviors of materials. We will encounter it regularly in Chapter 3 where the effective density of states at the band edge is fundamental to analysis of the operation of electronic devices.

2.2 INTRINSIC AND EXTRINSIC SEMICONDUCTORS

Semiconductors are called semiconductors because their ability to conduct electricity is neither very great nor very small. Indeed, the conductivity of a typical semiconductor can be controlled by temperature to such an extent that most can appear to be insulators at sufficiently low temperatures and metals at sufficiently high temperatures. Impurities can also play a role. A typical conductivity behavior for a semiconductor containing added impurities is shown schematically in Figure 2.10. A semiconductor free of significant impurities is termed "intrinsic" while those doped with impurities are termed "extrinsic". Both of these behaviors are essential to the operation of microelectronic devices.

2.2.1 Intrinsic semiconductors

An intrinsic semiconductor has no impurities and the number of electrons, n, in the conduction band exactly matches the number of holes, p, in the valence band, n=p. The number of electrons or holes is named the intrinsic carrier concentration n_i. This carrier concentration is given by the probability that a state at energy E is filled (as given by the Fermi function, Equation 2.11) multiplied by the density of states at that energy (Equation 2.22 for free electrons) integrated over all energies at or above the conduction band edge:

$$n_i = \int_{E_C}^{\infty} f(E)g(E)dE \qquad 2.24$$

Substituting from Equations 2.11 and 2.22, Equation 2.24 becomes:

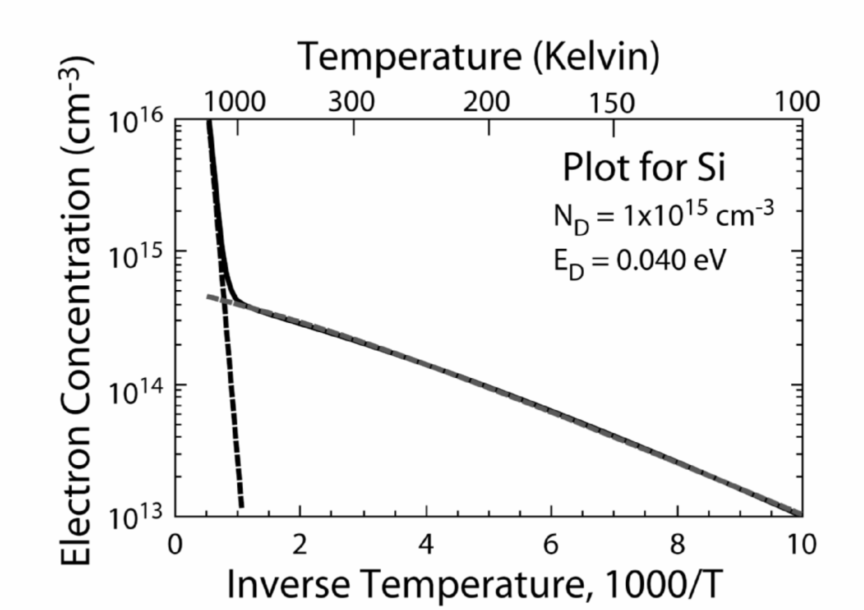

Figure 2.10: A plot of the electron concentration in a piece of silicon doped with 10^{15} cm^{-3} donor atoms having an ionization energy of 0.04 eV. The steep slope at high temperature (low inverse temperature) corresponds to the intrinsic behavior for carriers crossing the energy gap. The lower temperature behavior occurs in the presence of the 10^{15} cm^{-3} electron donors. The slopes of the two curves correspond to the 1.1 eV energy gap and the 0.04 eV donor ionization energies, respectively.

$$n_i = \frac{\pi}{2}\left(\frac{8m_e}{h^2}\right)^{3/2}\int_{E_C}^{\infty}\frac{E^{1/2}dE}{1+e^{(E-E_F)/k_BT}} \qquad 2.25$$

This equation can be simplified by noting that the band edge energy is typically far from the Fermi energy in units of k_BT, in an intrinsic material. Therefore, the Boltzmann approximation (Equation 2.12) can be used. With the further definition:

$$N_c = 2\left(\frac{2\pi m_e^* k_B T}{h^2}\right)^{3/2}, \qquad 2.26$$

which is the effective density of states at the conduction band edge, after some algebra, we find:

$$n_i = N_C e^{-(E_C-E_F)/k_BT} \qquad 2.27$$

We can circumvent the issue of the Fermi energy by noting that a similar definition can be made based n_i=p, and with a similar definition for the effective valence-band density of states, N_V:

$$n_i = N_V e^{-(E_F - E_V)/k_B T}, \tag{2.28}$$

from which, multiplying Equations 2.27 and 2.28,

$$n_i^2 = N_C N_V e^{-(E_C - E_V)/k_B T}. \tag{2.29}$$

Taking the square-root gives a value for n_i which does not depend upon E_F:

$$n_i = \sqrt{N_C N_V} e^{-E_{gap}/2 k_B T} \tag{2.30}$$

The most important point to note about Equation 2.30 is that n_i depends exponentially with temperature on *half* of the semiconductor energy gap. Narrow-gap semiconductors will have large intrinsic carrier concentrations while wide-gap materials will have fewer mobile carriers at a given temperature.

As a final point, note that if one knows that n=p=n_i, then from Equations 2.28 and 2.30, the Fermi energy of an instrinsic semiconductor (also known as the intrinsic energy E_i) can be derived. With some algebra, this can be shown to be:

$$\mathrm{E_F = E_i = \frac{E_{gap}}{2} + \frac{3k_BT}{4} \ln\left(\frac{m_h^*}{m_e^*}\right)}. \tag{2.31}$$

In other words, the Fermi energy is near the middle of the energy gap and deviates from the exact center by a factor that depends logarithmically on the ratio of effective masses of the two bands and linearly on temperature.

2.2.2 Extrinsic semiconductors

When impurities are added to a semiconductor the bonding pattern in that semiconductor is modified. This will be discussed extensively in Chapter 7. In the mean time, it is useful to have a general idea of the effect of impurities on carrier concentrations. Impurities that are added intentionally to control the carrier concentrations are called “dopants”. When a dopant has the same basic electronic structure as the atom it replaces (for example, partially-filled “s” and “p” orbitals for dopants in Si), the bonding behavior is substantially unchanged. As we will see, this usually results only in small changes in bonding and primarily affects only one band edge region. If the atom then has one more or one fewer electron that the atom it replaces, this state has an extra electron or hole in it, not present in the host semiconductor. The extra electron or hole can often escape the impurity atom and move freely through the semiconductor causing a change in conductivity.

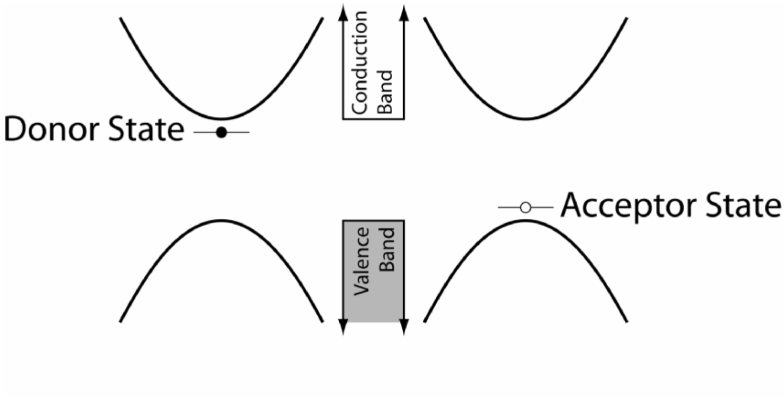

Figure 2.11: Momentum space diagrams of dopant states in the energy gap of a direct semiconductor. Note that the dopant states are somewhat extended in momentum space (they are not points) because they are somewhat localized in real space. Usually one represents the bands as ranges of energies where states are present, as shown at the center of the figure and momentum is largely ignored.

A single impurity atom in a direct gap semiconductor might produce a dopant state such as those shown in Figure 2.11. If the atom has one fewer electron than the host, it is likely to accept an electron from the relatively electron-rich semiconductor. Consequently, it is referred to as an acceptor dopant. If the atom has one extra electron it is generally a donor.

Electrons or holes in these dopant states can be easily released into the bands if the energy of the state is within a small number of thermal energy units (i.e. k_BT) of the appropriate band edge (the valence band for acceptors, the conduction band for donors). The probability of finding a carrier is calculated based on the Fermi function and the separation of the dopant state from the band edge. Thus, for donors:

$$n = \frac{N_D}{e^{(E_C - E_D)/k_BT} + 1} \approx N_D e^{-(E_C - E_D)/k_BT}. \qquad 2.32$$

A similar equation can be defined for holes,

$$p = \frac{N_A}{e^{(E_A - E_V)/k_BT} + 1} \approx N_A e^{-(E_A - E_V)/k_BT}. \qquad 2.33$$

In these equations, E_D and E_A are the donor and acceptor energies, E_C and E_V are the conduction and valence band edge energies, and N_D and N_A are the concentrations of donors and acceptors, respectively. The energy differences are the energies necessary to transfer a carrier from the dopant state to the band edge. These equations assume that the Fermi energy is at E_A or E_D, the temperature is moderate, and that the doping concentration is low.

It is typical to assume that all dopant atoms are ionized in a semiconductor at room temperature. However, this is not the case. From Equations 2.32 or 2.33, when the dopant level lies k_BT from the band edge, only 27% of dopant atoms are typically ionized. When the energy difference is half of k_BT, only 38% are ionized. Because it is uncommon to find isolated dopant states much closer than k_BT from the band edges, even the best dopants are rarely fully ionized. This becomes significant when working with compensated semiconductors where both donors and acceptors are present together. In these materials donors will spontaneously transfer electrons to acceptors with lower energy states, leading to full ionization of either donors or acceptors, whichever is the less common.

When one adds dopant to a pure semiconductor the Fermi level shifts, approaching the doping state energy. A brief examination of Equation 2.27 shows that if the Fermi energy rises in the band gap, the concentration of holes in the valence band must shrink. While the intrinsic condition n=p no longer holds, the condition

$$np=n_i^2 \qquad 2.34$$

is still true. Therefore, doping a semiconductor with an electron donor (making it "n-type") increases the concentration of free electrons and decreases the concentration of free holes by the same factor. In this case, electrons are called "majority carriers" and holes are "minority carriers". The same behaviors hold when holes are the majority carrier due to acceptor doping.

2.3 PROPERTIES AND THE BAND STRUCTURE

Our objective in designing materials is to understand what determines their useful properties and to change them such that they better meet our needs. The preceding sections have described some of the basic physics that underlies useful properties. This section describes in more detail how these properties arise. In later portions of the book we will see how specific properties are obtained by design in specific real materials. While real phenomena are more complex than the simple models presented here, real behaviors can generally be understood based on the concepts outlined.

2.3.1 Resistance, capacitance, and inductance

Resistance, R, capacitance, C, and inductance, L, are all observable phenomena used in circuit elements. They result from the motion of electrons in solids and their interaction with each other and with the atoms surrounding them. Each also depends upon the geometry of the circuit element producing the effect. The basic materials properties, however, are dependent only upon the electronic structure in that material and not on the geometry. These underlying properties are the resistivity, ρ, of a conducting material in a resistor, the dielectric constant, ε, of the insulating material in a capacitor, and the permeability, μ_B, in an inductor. The relationships are as follows.

$$R = \frac{\rho \ell}{A}, \tag{2.35}$$

where ℓ is the length of the conductor and A is its cross-sectional area. The resistivity, ρ, of a conductor is inversely related to its conductivity, σ as $\rho = 1/\sigma$. Likewise, C and ε are related through

$$C = \frac{\varepsilon A}{d}. \tag{2.36}$$

Here A is the area of the capacitor plates and d is their separation or the thickness of the insulating (dielectric) material between them.

Much of the interaction of electrons is through their electric and magnetic fields. The latter is responsible for the phenomenon of inductance. It results from the flow of current around a loop or coil of wire. The magnetic field produced depends upon the permeability of the material inside the wire coil. The greater the energy stored in the magnetic field, the higher the inductance of the coil. The inductance of a solenoid depends upon the permeability as:

$$L = \mu_B \frac{N^2 A}{\ell}, \tag{2.37}$$

where N is the number of turns of wire in the coil, A is the coil area, and ℓ is the coil length.

As we will see below, resistivity is directly related to the mobility of charges in the material, the dielectric response to its polarizability, and the permeability to the magnetic moment of the constituent atoms.

2.3.1.1 Mobility and electrical conductivity

Conduction in solids can occur by motion of electrons or charged atoms (ions). Most of this section is devoted to electronic conduction but a brief mention of ionic conduction is in order.

Ionic conduction occurs by diffusion ionized atoms, generally at high temperatures, although we will see it again in organic materials at room temperature. Inorganic solids with strong ionic bonding and high ionic diffusivity conduct charge well at sufficient temperatures. These materials usually include dopant ions that induce formation of charged vacancies on the same sublattice to compensate for dopant charge. Either the vacancies or the charged ions may then diffuse, carrying current.

Fast ionic conductors are used as solid electrolytes in fuel cells and sensors. The search for fast ion conductors operating near room temperature is a matter of current electronic materials research. The practical small-scale application of some types of fuel cells as replacements for batteries may hinge on success in this search.

Much more common and more relevant to most electronic applications is electronic conductivity. When a material is placed in an electric field, **E**, a current density, **J**, given by **J**=σ**E**, is induced. Here σ is the electrical conductivity:

$$\sigma = q(\mu_p p - \mu_n n)\,, \tag{2.38}$$

where q is the moving carrier charge, μ_n and μ_p are the electron and hole mobilities, and n and p are the electron and hole mobile carrier concentrations, respectively. Both the number of carriers and their mobility depend upon the energy band structure of the solid.

The number of carriers moving in a metal depends upon the density of carriers near the Fermi energy. To accelerate a carrier there must also be an empty state to move into with higher energy. The majority of carriers in the material come from or accelerate into states within ~k_BT of the Fermi energy. The effective density of mobile carriers in a metal is therefore approximately the number within k_BT of the Fermi energy, of the order of one per atom or $>>10^{22}$ cm^{-3}.

When an electron is placed in an electric field, it accelerates continuously until it bumps into something. The acceleration, **a**, is given by **a** = q**E**/m where m is the mass of the moving carrier. If the average time between collisions is τ, the "drift" velocity achieved during the acceleration time is $\mathbf{v}_d = \mathbf{a}\tau$. Classically one would expect $\mathbf{v}_d = \mathbf{a}\tau/2$ but if one conducts a detailed calculation doing the average over all trajectories after true times, not average times, then the velocity is increased by a factor of two. Substituting from the formula relating **a** with **E**,

$$\mathbf{v}_d = \frac{q\tau}{m^*}\mathbf{E}. \tag{2.39}$$

The quantity $q\tau/m^*$ is the carrier mobility, μ_n or μ_p.

The two variable components of mobility are the mean time between collisions and the effective mass of the moving charge. The mean time between scattering events depends upon how many things there are to bump into in the solid and how often each gives rise to a collision. The three primary scattering centers are other electrons, the atoms of the solid, and defects in a crystal. The first two are generally determined by fundamental properties of the material, while the latter is controlled by the density of imperfections. Defect cross-sections (the probability that a carrier passing through a given area containing the defect will be scattered) are the calculated quantities determining how effectively a defect will scatter a carrier.

Scattering of electrons off other electrons may be estimated from Fermi's Golden Rule (Equation 2.23). In words, one finds that the more electrons there are to scatter the moving charge and the larger the number of states the moving carrier can scatter into, the higher the scattering rate and the shorter the mean time between scattering events. These factors can be related in metals to the density of states at the Fermi energy, $g(E_f)$. When $g(E_f)$ is high there are many carriers to scatter from with energies just below E_f and many states to scatter into just above E_f. Metals such as Mo have low conductivities because they have high densities of states around E_f. Metals such as Cu, Ag, and Au are ideal compromises where there are high enough carrier densities for good conduction but low enough densities of states to keep the mean scattering times long. In metals, the mobility of electrons is generally low (less than 1 $cm^2V^{-1}s^{-1}$). In some semiconductor multilayer structures at low temperatures, by contrast, mobilities may exceed 10^6 cm^2 $V^{-1}s^{-1}$. In spite of this high mobility, the conductivity of these structures is modest because the carrier density is relatively low. Organic conductors tend to have both low mobilities ($<10^{-2}$ cm^2 $V^{-1}s^{-1}$) and low carrier concentrations. Thus, they are very poor conductors in most cases.

Scattering off atoms in the solid can be due to the presence of anomalously charged elements (dopants) that disrupt electrical periodicity. Scattering can also result from transfer of kinetic energy to or from the lattice in the form of phonons.

The details of scattering phenomena are beyond the scope of this book. Additional information may be found in the suggested readings.

As we saw in Equation 2.39, scattering is only part of the picture. Changes in carrier effective mass can also have an effect. Scattering rates often change by orders of magnitude as temperature or defect density change, while effective mass only varies by up to one order of magnitude. Therefore, the effective mass is usually less important than the scattering rate effect. Nonetheless, effective mass differences contribute to determining intrinsic mobility differences from one semiconductor to another and differences between electron and hole conduction within a given semiconductor. The effective mass is relatively easily determined, once one has a calculated band structure, and can be measured experimentally with good accuracy. Therefore, it is generally much better known than the scattering terms. Furthermore, the effective mass is a constant for a given material while the scattering rate is not.

The mobility can be connected to the diffusivity of a carrier. Diffusion depends upon the distance a carrier travels between collisions, as does mobility. Einstein developed a quantitative relation between the two quantities, showing that the mobility multiplied by the thermal energy (k_BT) was equal to the charge multiplied by the diffusivity. For example,

$$\mu_n k_B T = qD_n \qquad 2.40$$

for electrons. This relationship is most useful in balancing the diffusion current of electrons or holes in a concentration gradient against a drift current due to an electric field and is essential to understanding the behavior of diodes.

2.3.1.2 Dielectric constant, piezoelectric response, and permeability

The dielectric constant and permeability of materials are directly related to their ability to be polarized or magnetized, respectively. In general terms, polarizability is related to the magnitude of the motion of charge in a material resulting from the application of an external electric field. In such a field, electrons would naturally tend to move. Likewise, positive and negative ions would move and in opposite directions. The electrons and ions cannot move far without generating a large electric field of their own. Their motion results in the accumulation of negative charge in the material at the positive end of the external electric field (see Figure 2.12) and positive charge at the negative end, and reduces or cancels the applied field within the material. Thus, one converts the single long dipole associated with the external electric field into two smaller dipoles at each end of the dielectric. This increases the field strength outside and decreases it inside the dielectric. The increased field strength around the edges of the dielectric increases charge accumulation in surrounding materials, resulting in more capacitance. A similar phenomenon occurs in magnetic materials. Rather than charge accumulation, atomic magnetic moments align resulting in cancellation of the magnetic field within the material and increased magnetic fields around the edges. We will return to magnetic response of materials at the end of this section. For now we will consider the dielectric response in more detail.

The optical and dielectric properties of semiconductors are direct results of the dielectric constant, $\varepsilon=\varepsilon_0\varepsilon_r$ (where $\varepsilon_0=8.85\text{x}10^{-14}$ F cm^{-1} is the permittivity of free space and ε_r is the relative dielectric constant of the material in question), or the index of

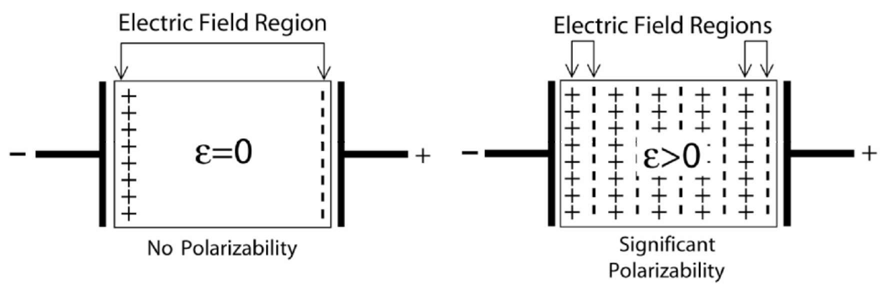

Figure 2.12: A schematic diagram showing the response of a dielectric material to an applied electric field.

refraction, $\eta = \sqrt{\varepsilon}$. To see how this relationship occurs, note that the redistribution of charge resulting from the applied field, E, results in a "displacement" field, D within a material of:

$$D = E + 4\pi P = \varepsilon E\,, \tag{2.41}$$

where P is the polarizability of the material. The dielectric constant, in turn, can be written in terms of the polarizability of the solid as:

$$\varepsilon = (1 + 4\pi P / E). \tag{2.42}$$

The total polarizability is most conveniently represented in a crystalline material in terms of the polarization, p, of a unit cell of the lattice of volume *v*, as:

$$P = \frac{p}{v}. \tag{2.43}$$

The polarization of the unit cell depends upon the applied field. More field produces more polarization. The useful the field-independent materials property is the polarizability per unit cell, α, giving $p = \alpha E_{local}$. Here E_{local} is the electric field inside the material that produces the polarization of the unit cell.

The polarizability, α, is due to two contributions, atomic polarizability and displacement polarizability. The atomic polarizability is caused by the motion of the *electron* cloud in the material with respect to the ionic cores of the atoms on lattice sites. The displacement polarizability is similar but involves motions of charged *atoms* with respect to one another. Consequently, it requires that there be two or more types of atoms in the lattice having a net polarity or difference in electric charge (i.e. at least partial ionic bonding) and results from motion of the negative *atoms* with respect to the positive *atoms*. These two polarization responses are shown schematically in Figure 2.13.

The polarization behavior resulting from atomic polarizability can be estimated by representing the electron cloud as a negatively-charged shell connected to a positive core by springs (Figure 2.14). The very heavy mass of the atomic core relative to the electron cloud means that the electrons can be assumed to be the only particles moving. The spring constant in the model is related to an observable resonant frequency ω_0, which is typically consistent with the frequency of high-energy photons. Thus, for all normal electronic devices, we can assume that the atomic polarizability is constant. However, in optical applications this polarizability can sometimes vary significantly as one passes through the resonant frequencies of the system. Atomic polarizabilities resulting from the displacement of the electron cloud with respect to the atom cores are of the order of 10^{-24} cm^{-2} V^{-1}.

The displacement polarizability results from motion of heavy atoms with respect to each other. These particles, being much more massive than electrons, respond to fields much more slowly. The same general model of masses connected by springs

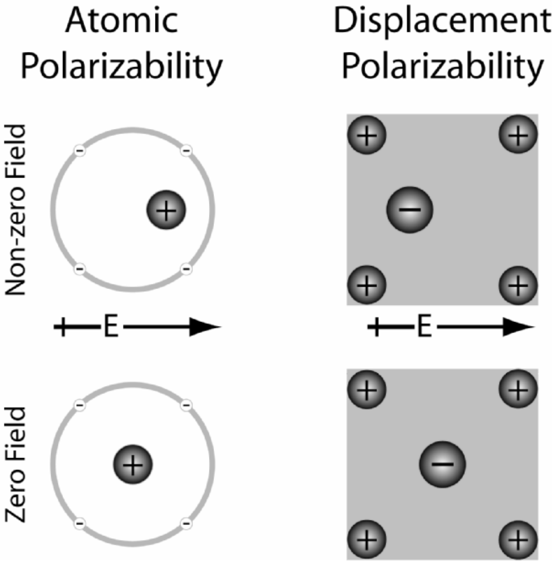

Figure 2.13: A schematic diagram showing the two major components to polarizability. In atomic polarizability, the electron cloud is displaced with respect to the remaining positive components of the atoms (the protons and the non-valence electrons). In displacement polarizability, atoms of different charge (positive or negative) with respect to their average charge.

can be applied, resulting in a harmonic oscillator response driven by any oscillations in the applied field. Such oscillations can induce very large changes in dielectric response with frequency. The displacement polarizability resulting from this behavior is

$$\alpha_{displacement} = \frac{Z^2 q^2}{(\overline{\omega}^2 - \omega^2)}\left(\frac{1}{M^+} + \frac{1}{M^-}\right), \qquad 2.44$$

where M^+ and M^- are the masses of the positively and negatively-charged ions, Z is the average ionic charge in units of electron charge (q), ω is the frequency of the applied electric field, and $\overline{\omega}$ is the resonant frequency of the system. Clearly, when $\omega \sim \overline{\omega}$ the system has a very large polarizability and consequently near this frequency there is a large change in dielectric constant. For normal materials the resonant frequency is near the "Debye frequency" which is related to the vibrational modes of the solid (see Section 2.3.3.). The resonant energy of these vibrations is typically 10-100 meV, or of the order of k_BT at room temperature. Thus, the displacement polarizability of partially ionic crystals becomes frequency-dependent for photons in the infrared portion of the optical spectrum. The low-frequency dielectric constants

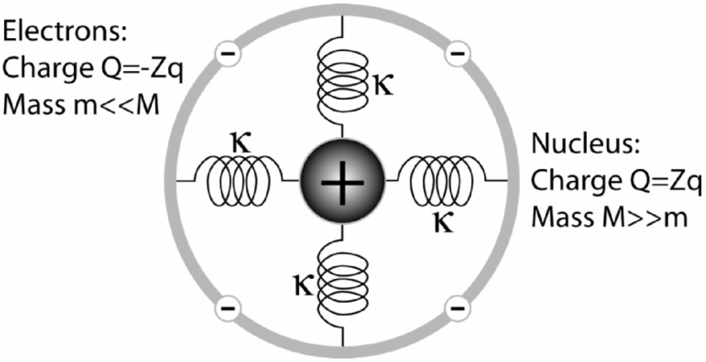

Figure 2.14: A model for picturing the response of the electron cloud of a solid to an applied electric field as a simple harmonic oscillator system.

are essentially the static dielectric constants resulting from the nearly constant atomic and displacement polarizabilities for fixed electric fields or very low frequencies.

The above considerations lead to the following general trends in dielectric response of materials.

The atomic polarizability:

- increases with the number of valence electrons;
- increases with decreasing binding energy for valence electrons around atoms (this changes the "spring constant" and lowers the resonant frequency).

The displacement polarizability:

- requires charge transfer among atoms in the unit cell (atomic solids and those without charge transfer show no displacement polarizability) so more ionic materials have a larger polarizability;
- larger displacements and "softer" crystals increase displacement dielectric constant.

Piezoelectric response is related to ionic displacement dielectric response. In a heteropolar (partially ionic) material that lacks a center of inversion symmetry, displacement of atoms of one polarity with respect to atoms of another polarity results in a change in shape of the material. A relationship between shape and applied electric field is termed a "piezoelectric" response. When the unit cell of the lattice includes inversion symmetry such a displacement moves charge but does not change the shape. Consequently, such materials are not piezoelectric. An example of how a material can lack an inversion center is found in all zincblende-structure materials. In these materials, a cation and anion lie at opposite ends of each bond and the structure is not symmetric around this bond. Furthermore, all bond pairs are oriented in the same way. Therefore, the structure lacks an inversion symmetry. Piezoelectric behavior can work either way – an applied field induces a shape change, but a shape change can produce an electric field. Most compound semiconductors exhibit a piezoelectric response. Typical magnitudes of this response for some common semiconductors are presented in Table 2.1. The permeability of a material is tied to its ability to be magnetized as dielectric response is tied to the ability to be polarized.

Table 2.1: **Typical Piezoelectric Stress Coefficients for Selected Materials**

Material	**e_{31} (C/m^2)**	**e_{33} (C/m^2)**	**Material**	**e_{31} (C/m^2)**	**e_{33} (C/m^2)**
Calculated values from Reference [1]					
ZnO (h)	-0.51	0.89	AlN (h)	-0.6	-0.47
BeO (h)	-0.02	0.02	GaN (h)	-0.49	-0.84
			InN(h)	-0.57	-0.88
CdTe (c)	-0.01	0.03	AlAs (c)	0.01	-0.01
ZnS (c)	-0.05	0.10	GaAs (c)	0.06	-0.12
ZnSe (c)	-0.02	0.04	InAs (c)	0.01	-0.03
AlP (c)	-0.02	0.04	AlSb (c)	0.02	-0.04
GaP (c)	0.03	-0.07	GaSb (c)	0.06	-0.12
InP (c)	-0.02	0.04	InSb (c)	0.03	-0.06
Experimental values from Reference [2]					
ZnO (h)	-0.46 †	1.27	GaAs (c)	-0.16 *	
AlN (h)	-0.48 †	1.55	CdS (h)	-0.21 †	0.44
a-SiC (h)	0.08 †	0.2	$Sr_{.5}Ba_{.5}Nb_2O_6$	5.19 †	9.81
			$LiNbO_3$	3.64 †	1.65

Note: The change in polarization δP_i along direction i is given by

$$\delta P_i = \sum_j e_{ij}\varepsilon_j$$ where ε_j is the strain along direction j.

h: hexagonal c: cubic † e_{15} value * e_{14} value

The magnetic response of a material depends upon the arrangement of electrons in the atoms of that material. Electrons have a "spin" of $\hbar/2$ which is related to their intrinsic magnetic moment of $q\hbar/2m$. When the spin vectors of the electrons in an atom align, the atom has a higher magnetic moment. The ability of electron spins to align is limited by quantum state availability. Half-filled states allow any combination of spins and high magnetic response, while completely filled states allow only one combination of antiparallel spins. When the spins of separate atoms align parallel to each other the system develops a net magnetization in local domains. A sufficiently large applied field can align the magnetic fields of the domains with the external field resulting in a net magnetization. If this field persists after removal of the applied field the material is a ferromagnet.

When the spins within atoms or molecules do not align spontaneously from atom/molecule to atom/molecule but where such alignment can be induced by an applied field, the material is called a paramagnet. In the least magnetic materials, diamagnets, the magnetic moments within an atom or molecule cancel exactly. In these materials, only small magnetic responses can be produced. Finally, it is possible for spins among atoms to spontaneously align opposite to one another. This behavior is called antiferromagnetic and resists magnetization. The permeability of ferromagnets is of the order of hundreds to hundreds of thousands, paramagnets have values greater than unity by 10-10,000 parts per million, while diamagnets have permeabilities smaller than unity by one to 200 parts per million.

2.3.2 Optical properties

The absorption, transmission, and reflection of light all depend critically on the density of states and Fermi energy in a material. Transparent materials have a gap in the density of states around the Fermi energy such that there are no states available into which an electron absorbing a photon (of small-enough energy) could move. A schematic of a typical density of states in a metal is shown schematically in Figure 2.15. A large density of electrons and states available near E_f, allows electrons to move in response to electromagnetic waves causing their reflection. If this occurs uniformly across photon energies, the material appears silvery or white. An exceptionally large density of states at specific energies can produce selective absorption, and consequently the material appears colored (as in Cu or Au). When the material has an energy gap (or very low density of states) between filled and empty levels, it allows light with energy below the gap energy to pass through the material. Transparent materials have very large energy gaps. Examples of transparent materials include silicon dioxide (as in window glass) and aluminum oxide (which is colored by impurities to make ruby and several other gemstones).

The reverse process, emission of light, can also occur when a material has a large density of electrons in a band with high energy (for example the conduction band) and holes at a lower energy (such as in the valence band). This unstable situation is resolved when electrons lose energy to fill the holes. The liberated energy may be emitted in the form of light.

The dielectric constant described above, and in particular the optical dielectric constant resulting from atomic polarizability, has a strong effect on optical properties including reflectivity of the material. Using this effect, even transparent materials can affect reflectivity when coated on another material. This is the basis of antireflection coatings on lens surfaces and window glass. Refraction effects resulting from the change optical dielectric properties of materials across an interface are also fundamental to the operation of lenses, optical fibers, and similar devices.

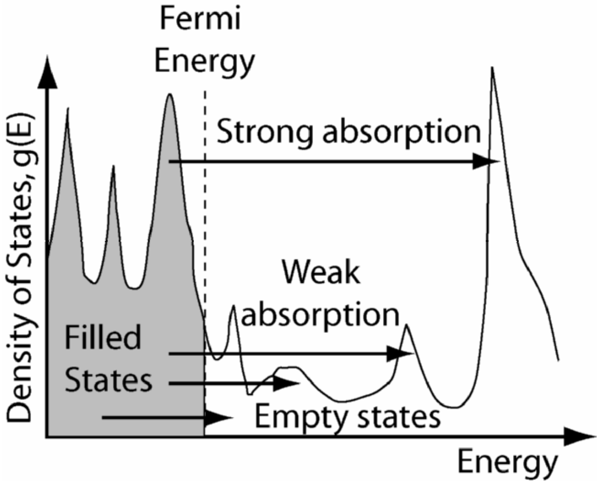

Figure 2.15: A schematic representation of the density of states of a metal which would lead to strong absorption in a limited range of energies and weak absorption at other energies.

2.3.3 Thermal properties

There are a variety of thermal properties of materials such as heat capacity, thermal conductivity, thermal expansion coefficient, and many more. For purposes of this text we will focus on the thermal conductivity as it is critical to many electronic devices. The reader is referred to the suggested readings for details of the full spectrum of thermal properties and for details not described here. The thermal conductivity, κ, of a material results from transport of energy via electrons or via lattice vibrations (phonons). The total thermal conductivity can thus be written simply as:

$$\kappa = \kappa_{phonons} + \kappa_{electrons} . \qquad 2.45$$

Typical thermal conductivities for a range of materials are given in Table 2.2. Let us consider the two contributions to thermal conductivity separately.

2.3.3.1 Heat conduction by phonons

"Phonon" is the name given to a cooperative vibration of atoms in a crystalline lattice much as a wave at sea is a cooperative motion of many water molecules. We saw above that phonons are essential to making transitions between minimum-energy states of indirect-gap semiconductors. They are also the primary reservoir of thermal energy in most solids and therefore contribute to thermal conduction. For these and other reasons, it is important to know more about phonons.

Table 2.2: **Thermal Conductivities of Selected Materials**

Material	**K_{Th} (W/cm K)**	**Material**	**K_{Th} (W/cm K)**	**Material**	**K_{Th} (W/cm K)**
C [3]	20	AlN [5]	2.85	AlAs [4]	0.08
Si [3]	1.56	GaN [5]	1.3	GaAs [4]	0.54
Ge [3]	0.6	InN [5]	0.45	InAs [4]	0.26
β-SiC [3]	5	AlP [4]	0.9	AlSb [4]	0.56
ZnO [6]	1.1	GaP [4]	1.1	GaSb [4]	0.33
Al_2O_3 [7]	0.39	InP [4]	0.7	InSb [4]	0.18

Source citation numbers for values are given parentheses.

One can calculate energy-momentum relationships for phonons just as was done above for electrons. The equations are not the same but the ideas are similar. They are the solutions for the wave equation for atoms rather than electrons in a regular periodic solid. Phonons are sound waves with a wavelength of the order of the lattice constant in the solid. The number of phonon modes is limited by the physical dimensions of the solid and the number of atoms that make it up, and well-defined energy momentum dispersion relations exist for phonons as for electrons. A typical phonon dispersion relation for a one-dimensional solid consisting of alternating heavy and light atoms is shown in Figure 2.16.

When the atoms all tend to move together in one direction or the other over many atom spacings the vibration wavelength is long (small k) and the energy is low. These modes are called acoustic phonons because audible sound vibrations have wavelengths many times the interatomic spacing. When the wavelength approaches the lattice spacing the resonance with the lattice causes a gap in the states, exactly as in electron band structures. In solids where there are two or more distinguishable atoms, not necessarily of different chemistry, additional states occur with short wavelengths. These are referred to as “optical” phonons and result in the upper branch of the dispersion relation in Figure 2.16. Note that the energy is linearly related to momentum for low energy acoustic phonons. Thus, sound waves traveling in a solid have energies that are linearly related to their momenta, as one would expect. As with electrons, the situation in three dimensions is much more complex than the simple one-dimensional behavior shown in Figure 2.16, but the same general results follow.

Full three dimensional phonon dispersion curves are shown for Si and GaN in Figure 2.17.[5] The frequency of atoms moving in solids can be $\sim 10^{13}$ s^{-1} for optical phonons in some solids. Among other points, this frequency defines the upper end of the range of attempt frequencies for atomic transport processes such as diffusion and for many reactions.

We can see from Figures 2.16 and 2.17 that lattice vibrations can exist with a range of energies and momenta. The energy of the average phonon in a solid at equilibrium is the average thermal energy of an atom in that solid, k_BT. Therefore, one can estimate the distribution of phonons based on the mode energies (see Fig 2.17) and the temperature of the solid. Phonons travel at the speed of sound. This is roughly constant over a wide range of temperatures. To conduct heat by phonon motion, the phonons scatter in the solid and come to equilibrium with the local temperature of the lattice. These phonons can then spread out through the solid as a wave moves across the ocean, carrying heat until they scatter again.

The phonon (lattice vibration) contribution to thermal conductivity is well approximated by:

$$\kappa_{phonons} = \frac{1}{3} C_V \bar{v} \Lambda \qquad 2.46$$

where C_V is the heat capacity of the solid at constant volume, $\bar{v}$ is the average speed of sound, and Λ is the mean free path of phonons between scattering events. The mean free path for phonons decreases as the number of phonons present in the solid increases due to a higher scattering probability.

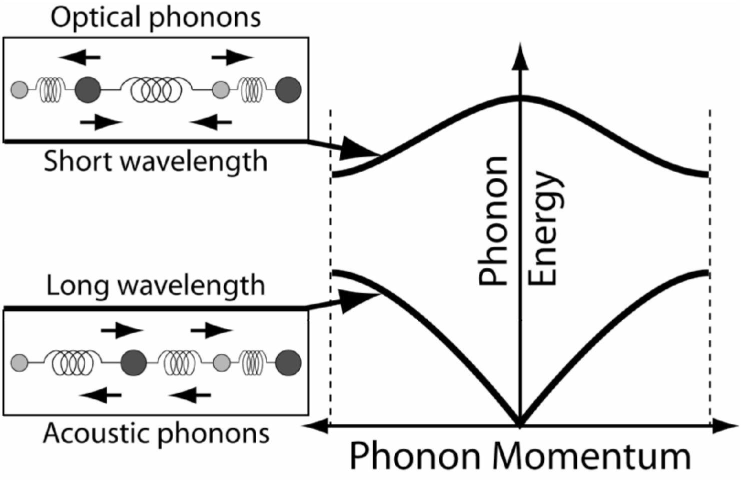

Figure 2.16: A typical phonon dispersion relation for a one-dimensional lattice of balls connected by springs. On the long-wavelength "acoustic" branch atoms move as groups in one direction or another with the direction varying over relatively long distances. For the higher-energy "optical" branch atoms move in opposite directions over very short distances.

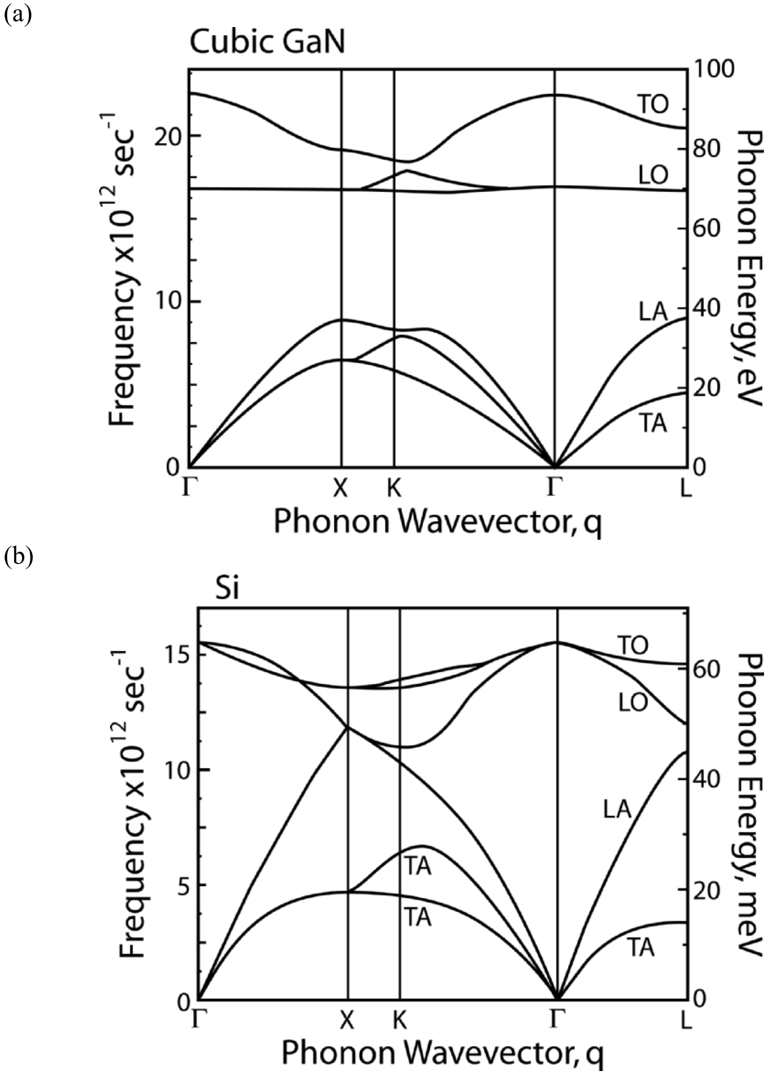

Figure 2.17: The phonon dispersion relations for (a) GaN and (b) Si. TA, LA, LO, and TO refer to transverse acoustic, longitudinal acoustic, longitudinal optical and transverse optical phonons, respectively. Each of these represents a particular vibrational mode. Longitudinal modes run along bonds as in Figure 2.16, while for transverse modes the vibration velocity is perpendicular to the bonds. There are two transverse modes because there are two axes perpendicular to a bond direction. Figures after Levinshtein, Rumyantsev, Sergey, and Shur, Reference [5], p. 27 and 184, respectively. This material is used by permission of John Wiley & Sons Inc.

It can be shown that the phonon mean free path in a typical solid can be written approximately as:

$$\Lambda \approx \Lambda_0 e^{\Theta/2T} \qquad 2.47$$

where Θ is called the Debye temperature. At low temperatures the heat capacity is also related to the Debye temperature through the Debye approximation:

$$C_v \approx cNk_B\left(\frac{T}{\Theta}\right)^3, \qquad 2.48$$

where c is a constant. Thus, there are phonon-related contributions to the thermal conductivity given in Equation 2.46 that enter through Λ and through C_V.

The heat capacity can be calculated exactly if one knows the relationship between energy and momentum for a vibrational mode and how many such modes exist in the material. Such a calculation can be done for a crystalline solid with results such as in Figure 2.17. A full discussion of the phonon modes in solids is beyond that needed for our purposes but may be found in many solid state physics texts.

At higher temperatures and often by room temperature, the heat capacity is simply determined by vibrations of individual atoms. Each atom has three independent vibrational modes, one for each direction in space, and each mode has an average energy k_BT. The heat capacity is, by definition, $\partial E/\partial T$ from which:

$$C_V = 3Nk_B. \qquad 2.49$$

N is the number of particles vibrating per unit volume (the atomic density) and the factor of three results from the three independent coordinates of vibration direction.

In insulators, there are no free electrons and consequently only phonons transmit heat. The maximum thermal conductivity of such a material can be calculated using the formulae above. In most cases, however, this turns out to be a relatively low value and most electrical insulators are relatively good thermal insulators. The notable exception is diamond with the highest thermal conductivity of any material. Because it is an electrical insulator, diamond conducts heat via phonons alone.

In common applications, electrical insulators are also used as heat insulators. Thus, one is concerned with the minimizing, rather than the maximizing thermal conductivity. The best thermal insulators are amorphous materials, which have lower thermal conductivities because phonons cannot exist over extended ranges. Essentially the scattering distance for phonons becomes one atomic spacing. In this case the transmission of heat in the material becomes a process of diffusion of energy among loosely coupled harmonic oscillators. In the classical (high temperature) regime, the minimum thermal conductivity has been estimated to be [8]:

$$\kappa_{\min} = \frac{k_B N^{2/3} v}{2.48} . \tag{2.50}$$

The value for SiO_2 based on this formula is ~0.01 W $cm^{-1}K^{-1}$, which is close to the measured value. Thus, amorphous silica (glass) is a nearly perfect thermal insulator. The only improvement would be to decrease the density of the glass, as gases and vacuum are better thermal insulators than any solid can be. The heat shielding tiles on the space shuttle, for example, make use of underdense fiberous silica to achieve an exceptionally low thermal conductivity.

2.3.3.2 Heat conduction by electrons

As with phonons electrons come to thermal equilibrium with the lattice when they scatter. Electrons easily move through the lattice of a conductive solid with relatively long mean-free path lengths. Therefore, they transport heat well. As with phonons, the electron contribution to thermal conductivity is:

$$\kappa_{electrons} = \frac{1}{3} C_{elec} v_{elect} \Lambda_{elec} \tag{2.51}$$

where C_{elec}, v_{elec}, and Λ_{elec} are the heat capacity, mean velocity, and mean free path of electrons near the Fermi energy in the solid. One can carry out calculations to show the mathematical contributions to the various terms in Equation 2.51. In the end, one can derive an equation for the thermal conductivity of electrons in terms of the temperature, T, and the electrical conductivity, σ, of the material:

$$\kappa_{elec} = L\sigma T , \tag{2.52}$$

which is known as the Wiedemann-Franz law. L is a constant. For a free electron solid L~2.45×10^{-8} WΩ K^{-2}. In real metals, L ranges from ~2.2 to 3.0 $\times 10^{-8}$ WΩ K^{-2}. In general, the thermal conductivity of electrons is much greater than for phonons.

2.4 QUANTUM WELLS AND CONFINED CARRIERS

In Section 2.1. we examined solution to the Schrödinger equation in free space and in a periodic potential and found families of wave functions having a simple, quadratic relationship between wave energy and momentum over wide ranges of energy values. Because there are a very large number of such states in a solid, it is normally impossible to distinguish one individual state from another. In this Section, we will examine, briefly, the impact of artificial potential structures on the wave solutions. For a complete discussion of such effects, the reader is referred to books on quantum mechanics.

Artificial-potential-barrier structures appear in a number of important microelectronic device applications. The best known is probably the laser diode. These ubiquitous devices are deceptive in their outward simplicity – a single small chip of material

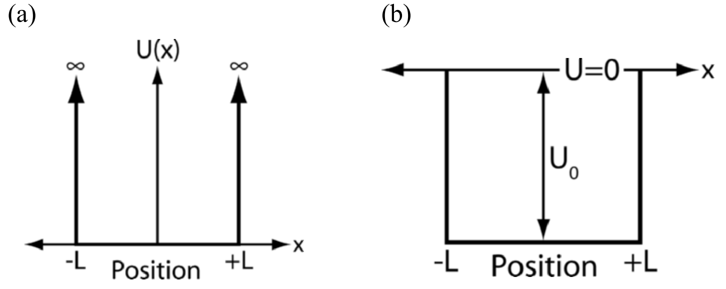

Figure 2.18: (a) a potential well with infinite walls. (b) a well of depth, U_0. Both wells have a width of 2L.

connected by two wires to a switch and batteries, or to a power supply. The trick is in the atomic-scale design of the chip of material. Laser diodes consist of a series of layers that are designed to collect electrons and holes. [A more detailed description of laser diodes may be found in Chapter 3.] Recombination of the free carriers in the trap leads to exceptionally efficient laser light emission. The trap is a potential well resulting from the sandwiching of a narrow-gap semiconductor between two layers of wider-gap material. Discontinuities in the valence and conduction band edges produce a lower-potential region where the carriers tend to collect. Because these wells are of the size scale of an electron wave function, the quantum nature of the system determines its properties. Hence, such structures are known as quantum wells. We will now examine the solutions of Schrödinger's equation in and near such wells.

The traditional beginning is to consider a potential well with infinite side walls [see Figure 2.18(a)]. The potential function in a one-dimensional version of the time-independent Schrödinger's equation (Equation 2.3) then becomes $U=0$ for $|x| < L$ and $U=\infty$ for $|x| \geq L$. The solution to Equation 2.3 for $|x| \geq L$ must be $\psi(x) = 0$. For $|x| < L$ the equation is (in one dimension):

$$\frac{d^2\psi}{dx^2} + \kappa^2\psi = 0\,, \tag{2.53}$$

where

$$\kappa^2 = \frac{2mE}{\hbar^2}\,. \tag{2.54}$$

The solutions to Equation 2.53 are proportional to $\sin(\kappa x)$ and $\cos(\kappa x)$. Since the wave function is zero outside the well and since the wave function inside and outside must match at the boundaries, the values of κ are forced to be $\kappa L = n\pi$, where $n=1,2,3,\ldots$ for the $\sin(\kappa x)$ solution and $\kappa L = (n-1/2)\pi$ for $\cos(\kappa x)$ to produce nodes at

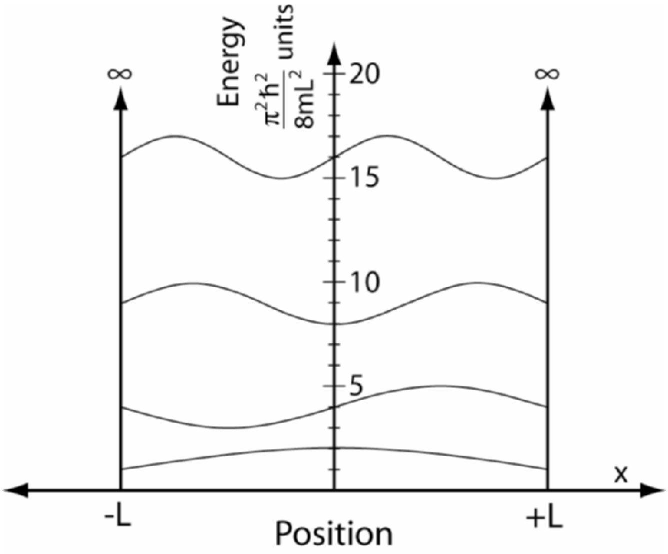

Figure 2.19: Electron states in an infinite-depth quantum well. Note that these are simply harmonics of the fundamental mode with nodes at the edges of the well.

the walls of the well. The allowed energies are then determined from Equation 2.54. Solutions for the infinite well are of the form shown in Figure 2.19.

The situation shown in Figure 2.18 (a) is, of course, not realistic. Infinite potential barriers cannot be produced. A more realistic solution would involve a finite well such as shown in Figure 2.18 (b). Let us consider the solutions of the Schrödinger equation for particles with energies below the energy of the edges of the well. In this case again U=0 for |x|≥L and U=-U$_0$ for |x|<L. Inside the well and away from the edges, the solutions are roughly the same as one would find in the infinite well with somewhat relaxed boundary conditions. The real part of the most general solution to the Schrödinger Equation for the finite potential well is:

$$\begin{aligned} \psi(x) &= Ce^{\kappa x} && x < -L \\ \psi(x) &= A\cos(kx) + B\sin(kx) && -L < x < +L \\ \psi(x) &= De^{-\kappa x} && x > +L \end{aligned} \qquad 2.55$$

The two exponential relationships describe the decay of the wave function outside of the well, while the middle relationship in 2.55 describes the wave function within the well. As in the infinite well, the wave function and its first derivative must be continuous across the well boundaries. The boundary conditions yield four equations (two boundaries, two matching conditions) in four unknowns (A,B,C,D), yielding:

$$\kappa = k\frac{A\sin(kL) - B\cos(kL)}{A\cos(kL) + B\sin(kL)}$$

2.56

$$\kappa = k\frac{A\sin(kL) + B\cos(kL)}{A\cos(kL) - B\sin(kL)}$$

Additional equations for C and D can be developed but will not be reproduced here to save space. Furthermore, one can show that:

$$k^2 = -\frac{2m}{\hbar^2}(U_0 + |E|) > 0$$

$$\kappa = k\tan(kL) \qquad \text{Even solutions} \qquad 2.57$$

$$\kappa = -k\cot(kL) \qquad \text{Odd solutions}$$

The second two formulas in Equation 2.57 are from boundary condition matching. The first part of Equation 2.57 is the bound state requirement and can be rewritten:

$$|E| = -U_0 + \frac{\hbar^2 k^2}{2m} \qquad 2.58$$

This is just a statement that the bound state energy ($\hbar^2$ k^2/2m) lies below the well boundary and is harmonic-oscillator-like. If one defines the energy to be zero at the bottom of the well, then the solutions become simple free-electron waves within the well as in Equation 2.16. This is not surprising as the middle relationship in Equation 2.55 for the wave function within the well is a free electron wave behavior. Because the bound states must be standing waves in the well (must have a maximum or a zero at the well center), the values of k are linked to the well half-width, L. This, finally, leads to the constraint:

$$\sqrt{\frac{2mU_0}{k} - 1} = \begin{cases} \tan(kL) & \text{Even solutions} \\ -\cot(kL) & \text{Odd solutions} \end{cases} \qquad 2.59$$

Note that Equation 2.59 provides a connection between the wavelength, the well width, and the well depth. Therefore, the well dimensions determine both the wavelength and energy of the states within the well. A lot of algebra and discussion has been skipped in writing these solutions. If you want more details, this problem is treated in any basic quantum mechanics text. The form of these solutions is shown in Figure 2.20.

To reiterate the important points from these equations:

- Equation 2.59 shows that the possible values of the wavelength, k are coupled to the well half width, L, and to the depth of the well.

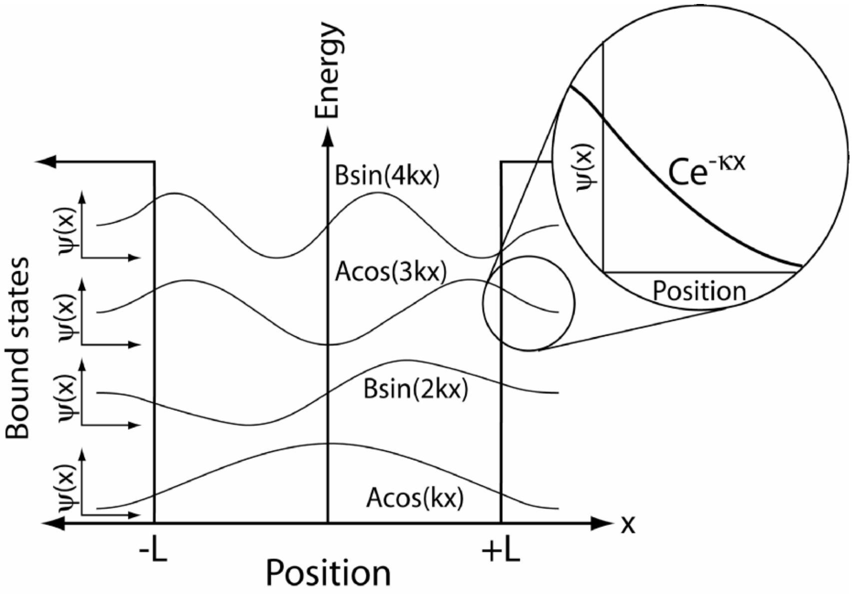

Figure 2.20: Solutions to the Schrödinger equation given by 2.55 and subject to the additional conditions of Equations 2.56-2.59.

- Deeper wells produce more bound states with slightly shorter wavelengths (as the wave function is less inclined to penetrate the well sides).

- Wider wells produce more bound states for a given well depth.

- There are very few states in the well and these have discrete energies.

- There are no states at the bottom of the well as would be judged from the band edges. Therefore, the band gap of the well is greater than the minimum gap of the semiconductor from which it is produced.

- The wave function decays exponentially in the barriers with a decay length κ that increases with decreasing well depth.

- The wave function penetrates the surrounding barriers a significant distance. This gives rise to coupling between adjacent quantum wells.

The final point above deserves some specific attention. Two quantum wells that are close enough to each other for the tails of their wave functions to overlap significantly become coupled together. The interaction of their wave functions alters the energies of states in the wells. This interaction produces bonding-like and antibonding-like pairs of states, one raised in energy, the other lowered. Adding a third quantum well adds a third state in each quantum well. An infinite number of coupled wells would produce a band of states in each well associated with each quantum state for a single well. The widths of the bands are directly related to the well-to-well coupling. Well coupling depends upon the depth of the well and exponentially upon the distance between the wells. This situation is shown schematically in Figure 2.21. [9]

It may be helpful to consider the example in Figure 2.21 in more detail. The discussion below is after Holonyak.[9] Suppose the wells shown have a depth U_0=200 meV, a well width 2L = 4 nm, and a well separation s = 4 nm. This is typical of a series of quantum wells produced by growth of 4 nm thick $Al_{0.19}Ga_{0.81}As$ layers alternating with 4 nm thick GaAs layers in a superlattice. The GaAs layers produce

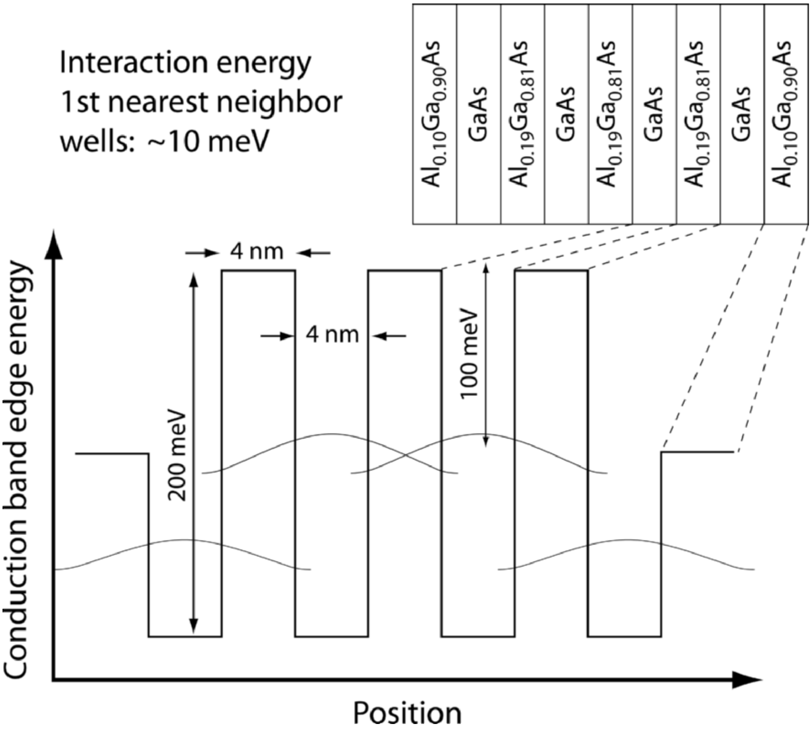

Figure 2.21: An example semiconductor superlattice structure resulting in quantum wells as shown for the conduction band edge. Similar but shallower wells occur in the valence band. The bound states in the wells are also indicated. These states overlap producing splitting and multiple levels in each well (not shown in this sketch). (Courtesy N. Holonyak [9])

the wells and the $Al_{0.19}Ga_{0.81}As$ provides the barriers. The states in the quantum wells couple with each other such that the first-nearest neighbor wells have two states rather than the single state that a single well would have. These states change their energies by ~10 meV relative to the single well. Second nearest neighbor interactions when three or more wells are present cause additional splitting with a further change of energy of ~4.5 meV, and so on. The result is a series of states in the quantum well at the center of the superlattice with energies as shown in Figure 2.22. The more quantum wells that are coupled together the more states occur in each well and the closer together these states are. Wells toward the edges of the superlattice have fewer wells to interact with and, consequently, fewer bound states.

There are several points to notice about Figure 2.22. First, the states increase in energy more than they decrease because as a state drops in the quantum well, the energy barrier separating those states from their neighbors rises. Consequently, the states decay more rapidly in the barrier and their interaction is weaker from well to well. For an infinite number of wells one gets a band of states with a width of

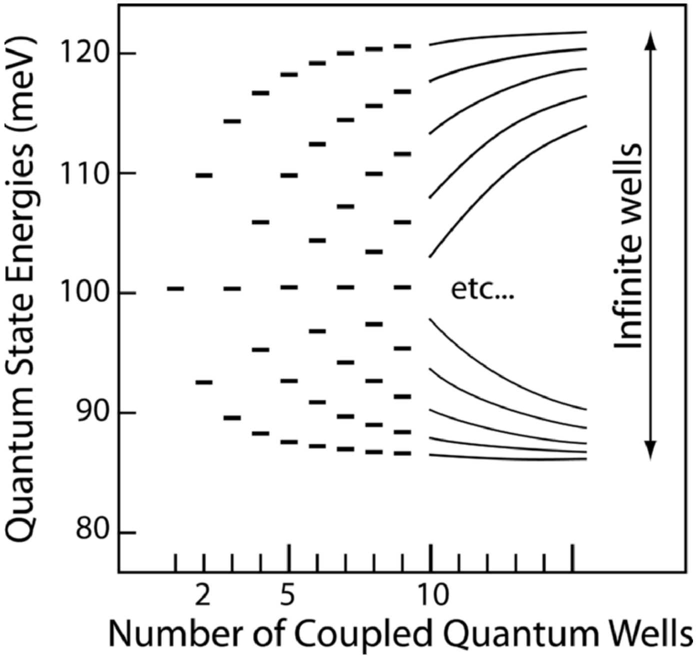

Figure 2.22: A plot of the number of states and their energies in a quantum well which interacts with the specified number of other identical wells. The plot assumes the geometry and energies of Figure 2.21. An essentially infinite superlattice produces a band of states as indicated. (Courtesy N. Holonyak [9])

roughly four times the first-nearest-neighbor interaction energy and with a sinusoidal variation in energy with momentum. Second, for an odd number of quantum wells there is always a state with a binding energy of 100 meV. Third, the result illustrated in Figure 2.22 is for a single bound state. If there had been two states in the well, each would produce a series of states as more wells were added, giving rise to two minibands per state in the isolated well. Such would have been the case for wider wells separated by the same barrier layers. The situation shown in Figures 2.21 and 2.22 is for the conduction band in the example superlattice. A similar set of states would occur in the corresponding wells in the valence band. Because approximately 70% of the band offset (see Chapters 5 and 6 for details) is accommodated in the conduction band, all of the behaviors in Figures 2.12 and 2.22 would hold for the valence band but all barriers would be smaller, leading to more overlap between states and more broadening of the valence band quantum well minibands.

Each state in the quantum well can accommodate electrons or holes and consequently can give rise to absorption and emission of light. In laser diodes we take advantage of single quantum wells to trap electrons and holes to enhance recombination. It is important to use only single quantum wells or widely-separated wells to prevent broadening of the states as in Figure 2.22. Such broadening would give rise to a range of emission wavelengths. In a laser one wants a single emission wavelength (single mode laser) if possible.

2.5 SUMMARY POINTS

- Electrons are Fermions, thus the probability of finding an electron in a given state is given by the Fermi function (Eq. 2.11.).
- The Schrödinger equation solutions describe the electron waves allowed. For a free electron (no potential energy) such solutions result in a quadratic relationship between electron momentum or wave vector and electron energy. In a periodic solid this relationship repeats with a wave vector inversely proportional to the lattice period.
- When a periodic potential is present, energy gaps develop in the quadratic energy/momentum relationship, proportional to the strength of the periodic potential.
- The energy/momentum relationship in three dimensions can include maxima and minima in the energy of given branches. This can lead to indirect or direct energy gaps. This has implications for the optical and electronic properties of the semiconductor.
- The effective mass of the electron is modified from its rest mass in vacuum by the presence of a periodic lattice. The effective mass is inversely related to band curvature in energy/momentum space.
- Density of states per unit volume describes the number of states in a small energy increment around a given energy. The probability of an electron absorbing energy or momentum is generally directly related to the density of states.
- The density of states around any band extremum will be quadratic in energy with respect to the energy of the band edge.
- The density of free carriers in a pure and perfect (intrinsic) semiconductor increases exponentially with temperature with a characteristic energy of half the energy gap.
- The number of holes and electrons exactly matches in an intrinsic material.
- In a doped semiconductor the number of holes and electrons does not balance. The number of majority carriers depends upon the number of dopant atoms and exponentially on the energy separating the impurity state from the appropriate band edge and the temperature. When more than one dopant atom is present, compensation can lower the number of majority carriers.
- Electrical conductivity depends upon the product of carrier density and mobility. Mobility depends upon effective mass and mean time between collisions of electrons causing scattering.
- Dielectric constant depends upon material polarizability. The polarizability consists of atomic polarizability (increases with number of valence electrons and decreasing electron binding energy) and displacement polarizability (increases with polarity of material and magnitude of possible displacements).

- Thermal conductivity depends upon lattice vibration and electronic contributions. The lattice contribution depends upon atomic density while the electronic contribution depends upon conductivity and temperature.
- Lattice vibrations can be described as "phonon" quasiparticles, which carry momentum and heat energy in the lattice. The phonon dispersion relation provides a linear relation between energy and momentum at moderate phonon momenta.
- Quantum wells have states in them whose energy is determined by the well depth and width but is always above the bottom of the well in energy (for electrons).
- Coupled quantum wells result in multiple states in the wells. When enough wells are coupled these become minibands of states in the wells.

2.6 HOMEWORK

1: Energy Band Diagrams

Consider the energy vs. momentum relationships (band diagrams) shown in Figure 2.7.

a) What range of energies does the valence band of GaAs span based on Figure 2.7.?
b) Sketch the most important features of the E(k) diagram for Ge and indicate the minimum energy gap.
c) Which semiconductor (GaAs or Ge) has the indirect energy gap? How do you know?
d) The effective masses of the carriers are given by: $1/m^* = (d^2E/dk^2)$. Which semiconductor, Ge or GaAs has the higher electron effective mass?
e) Sketch the density of states near the top of the valence band. What is the functional form of this curve?
f) If the number of electrons and holes are equal in an intrinsic semiconductor, why is the Fermi Level not exactly in the middle of the energy gap? (Eq. 2.31)

2: Conductivities

a) Could the thermal conductivity of SiO_2 be made lower by changing the material (for example by crystallizing it) but assuming one is not allowed to use alloying, density reduction, etc…?
b) Why is the electrical conductivity of Ag much higher than the electrical conductivity of Mo. Which would you expect to have a higher thermal conductivity? Explain briefly.
c) Explain briefly why it is necessary to have two distinguishable atoms in a crystal unit cell in order to observe the optical branch of the phonon dispersion relationship.

3: Consider phosphorous as an impurity in silicon.

a) Would you expect it to be an electron donor or an acceptor? Explain why.
b) If the ionization energy (the difference in energy between the phosphorous state and the band edge) is 35 meV, calculate the fraction of the phosphorous atoms ionized at (i) 80 K, (ii) 300 K, (iii) 400 K? You may assume that the Fermi energy is at the phosphorous state energy and that electrons escaping into the band need only reach the band edge rather than integrating the product of the Fermi function and the density of states throughout the conduction band.
c) If there are 10^{17} cm^{-3} phosphorous atoms in a sample of Si, calculate the minority carrier concentration at 300K. You may need the following: $N_c=2.8\times10^{19}$ cm^{-3}, $N_v=1.04\times10^{19}$cm^{-3}, and $E_{gap}(300K) = 1.12$ eV for Si.

4: Consider the solutions for the Schrödinger Equation for a quantum well.
   a) Show that Equations 2.55 and 2.56 satisfy the boundary conditions requiring matching of the wavefunctions and their slopes across the boundaries.
   b) Calculate the energies and number of states in quantum wells 85 meV deep and 4 nm wide. Assume the hole effective mass for GaAs. (These are the states that would occur in the valence bands corresponding to the conduction band states in Figure 2.22.)

5: Think about the periodic potential, U(x) discussed in Chapter 2 with which electrons interact.
   a) What is the effect of increasing the periodic potential U on the dispersion of bands in a solid? (Hint: consider Equation 2.10 and Figure 2.3.)
   b) Refer to the free electron dispersion relationship in Figure 2.6 with those for GaAs and Ge. Based on the behaviors where the lower portion of the valence band intercepts the L point (which is in planes most directly bisecting interatomic bonds in these materials) at the Brillouin zone boundary, estimate the magnitude of the first Fourier Coefficient of the atomic potential for these materials. The point in question intercepts the Y axis of Figure 2.6 at roughly 0.8 arbitrary units.

6: Consider the density of states discussion in Section 2.1.8. Between the Brillouin zone boundaries the E(k) relationship turns from concave up to concave down. In this region there is a small area where the energy is roughly linearly related to momentum, E proportional to k.
   a) In this region, derive a formula for the density of states similar to the approach used in developing Equation 2.22.
   b) What is the effective mass of an electron in this region of the energy-momentum relationship in terms of E and k? (Hint: follow the approach by which Equations 2.17 and 2.18 were developed.)

2.7 SUGGESTED READINGS & REFERENCES

Suggested Readings:

Arfken, George, *Mathematical Methods for Physicists.* New York: Academic Press, 1970.

Ashcroft, Neil W. and Mermin, N. David, *Solid State Physics.* Philadelphia, PA: Saunders College, 1976.

Bube, Richard H, *Electrons in Solids: an Introductory Survey*. Boston: Academic Press, 1992.

Cahill, David G, and Pohl, R.O., Lattice Vibrations and Heat Transport in Crystals and Glasses, Ann. Rev. Phys. Chem. 1988; 39: 93-121.

Chelikowsky, J.R. and Cohen, M.L. Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. Phys. Rev. B 1976; 14: 556-582.

Cullity, B.D., *Elements of X-ray Diffraction*. Reading PA: Addison-Wesley, 1978.

Ferry, David K., *Semiconductors.* New York: Macmillan, 1991.

Harrison, Walter A., *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond.* San Francisco: Freeman, 1980.

Herman, F., *An Atomistic Approach to the Nature and Properties of Materials* Pask, J.A., Editor New York: Wiley, 1967.

Kittel, Charles, *Introduction to Solid State Physics.* New York: Wiley, 1996.

References:

[1] Bernardini, F.; Fiorentini, V.; Vanderbilt, D.; "Spontaneous polarization and piezoelectric constants of III-V nitrides." *Physical Review B,* 1997; 56: R10024-7.

[2] Gualtieri, J.G.; Kosinski, J.A.; Ballato, A.; "Piezoelectric materials for acoustic wave applications." *IEEE Transactions on Ultrasonics Ferroelectrics & Frequency Control,* 1994; 41:53-9.

[3] Eberl, K.; Schmidt, O.G.; and Duschl, R.; "Structural Properties of SiC and SiGeC Alloy Layers on Si," in *Properties of Silicon Germanium and SiGe: Carbon*. Erich Kasper and Klara Lyutovich, eds., London, INSPEC, 2000.

[4] Neuberger, M., *Handbook of Electronic Materials, III-V Semiconducting Compounds*. New York, Plenum, 1971.

[5] Levinshtein, Michael E., Rumyantsev, Sergey L., and Shur, Michael S., *Properties of Advanced Semiconductor Materials*. New York, John Wiley & Sons, 2001.

[6] Florescu, D.I.; Mourokh, L.G.; Pollak, F.H.; Look, D.C.; Cantwell, G.; Li, X.; "High spatial resolution thermal conductivity of bulk ZnO (0001)." *J. Appl. Phys.,* 2002; 91: 890-2.

[7] Kinoshita, H.; Otani, S.; Kamiyama, S.; Amano, H.; Akasaki, I.; Suda, J.; Matsunami, H.; "Zirconium diboride (0001) as an electrically conductive lattice-matched substrate for gallium nitride." *Jpn. J. Appl. Phys.,* 2001; 40: L1280-2.

[8] Cahill, David G.; and Pohl, R.O.; "Lattice Vibrations and Heat Transport in Crystals and Glasses." *Ann. Rev. Phys. Chem.,* 1988; 39: 93-121.

[9] Holonyak, N.; University of Illinois, Department of Electrical and Computer Engineering, private communication.

[10] Herman, F., *An Atomistic Approach to the Nature and Properties of Materials* Pask, J.A., Editor New York: Wiley, 1967.

[11] Chelikowsky, J.R. and Cohen, M.L.; Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. Phys. Rev. B 1976; 14: 556-582.