SEMICONDUCTORS AND NANOSTRUCTURES

Jeffrey Davis and Chennupatla

CONTENTS

- 2.1 Basic Semiconductor Physics
 - 2.1.1 Band Structure
 - 2.1.2 Conduction Properties
 - 2.1.3 Compound Semiconductors
 - 2.1.4 Phonons
 - 2.1.5 Other Lattice Excitations
- 2.2 Optical Properties of Semiconductors
 - 2.2.1 Excitons
 - 2.2.2 Defect States
 - 2.2.3 Free Carrier Absorption
 - 2.2.4 Optical Emission (Radiative Recombination)
 - 2.2.5 Inelastic Scattering
 - 2.2.6 Optical Properties in a Magnetic Field
 - 2.2.6.1 Zeeman Splitting
 - 2.2.6.2 Diamagnetic Shift
 - 2.2.6.3 Landau Levels
 - 2.2.6.4 Other Magneto-Optical Effects
 - 2.2.7 External Electric Fields
 - 2.2.8 Strain Fields
- 2.3 Low-Dimensional Semiconductor Nanostructures
 - 2.3.1 How They Are Realized
 - 2.3.2 Effects of Reduced Dimensionality
 - 2.3.3 Implications for Optical Properties
 - 2.3.3.1 Electro- and Magneto-Optical Effects
- 2.4 Summary
- References

2.1 BASIC SEMICONDUCTOR PHYSICS

A semiconducting material is defined as one that is electrically insulating at a temperature of absolute zero but that becomes conducting for temperatures below its melting point. The property that first attracted the attention of physicists in the early nineteenth century was the decreasing resistivity with increasing temperature, in stark contrast to metals where the opposite is true (Faraday 1839, Smith 1964, Ashcroft and Mermin 1976, p. 563). Both of these effects occur in semiconductors because it is possible to thermally excite charge carriers from a full energy band that prevents conduction to an unoccupied band that then allows electrons to move freely. The structure of these energy bands is used to explain most of the properties of semiconductors. In this chapter, we therefore begin with a brief introduction to semiconductor band structures and then continue with our discussion on some of the important and interesting properties observed in semiconductors and their nanostructures.

2.1.1 Band Structure

Consider a single silicon atom; the solution to the Schrödinger equation reveals discrete electronic orbitals, with four electrons in the valence shell. This silicon atom is able to covalently bond with four other silicon atoms, which in turn are able to bond with four silicon atoms, and so on, forming a crystal lattice. In a crystal lattice, the electronic energy structure changes. To a first approximation, the atomic orbitals are replaced by filled bonding and empty anti-bonding orbitals, with pairs of electrons localized between the two nuclei they are binding. In the absence of interactions between different pairs of electrons, every electron in the lattice will have an identical set of discrete states. In practice, however, this is a strongly interacting system and the Hamiltonian needs to include the electromagnetic interactions between all the valence electrons. The effect of this is that the Pauli exclusion principle prevents more than two electrons from existing in the same quantum state, lifting the degeneracy and generating an additional state for each additional Si-Si bond, with all states split by an amount that is determined by the strength of the interaction. In other words, a single Si-Si bond will have a single molecular orbital that is doubly degenerate (i.e., it can contain two electrons of opposite spin); a pair of Si-Si bonds will have two molecular orbitals (each doubly degenerate) that are split in energy due to the electromagnetic interaction; three Si-Si bonds will have three molecular orbitals that are split in energy; and so on, until, in a semi-infinite crystal lattice, there are a semi-infinite number of Si-Si bonds, and a semi-infinite number of states, split by a semi-infinitesimal amount, as depicted in Figure 2.1a. An alternative way of looking at the formation of bands is depicted in Figure 2.1b, where the energy structure is shown as a function of atom separation. As the nuclei become closer, the discrete atomic states shift and form bands as the interaction between electrons increases (Allison 1971). In a semi-infinite lattice, rather than a series of discrete states, there is, therefore, a quasi-continuum of states that is referred to as an energy band. In a tetravalent system, such as Si, there are actually four valence bands, corresponding to the four molecular bonding orbitals for each Si atom, and another four "conduction" bands for the corresponding antibonding orbitals. At low temperature, the valence bands are all fully occupied, and the conduction bands are empty. The properties of these bands and splitting between them are determined largely by the type of atoms, the crystal structure, and relevant momentum and angular momentum considerations.

The vast majority of semiconductors are tetravalent; however, trivalent, heptavalent, and hexavalent semiconductors also occur, with corresponding crystal structures. Regardless, the periodic nature of semiconductor crystals allows one to introduce the concepts of the unit cell and reciprocal lattice. The unit cell is defined as a volume of the lattice that is repeated within the crystal, completely filling the space with no overlap, and usually represents the full symmetry of the lattice. In this way, the unit cell, together with the associated lattice vectors

Figure 2.1 (a) As the number of bonds increases, so does the number of discrete bonding and antibonding states. In a crystal lattice, these distinct states form a quasi-continuum, leading to the formation of energy bands. Alternatively in (b), as the separation between atoms decreases, and interactions between their electrons increase, the electron energy transitions from discrete atomic states to bands separated by a band gap. (Adapted from *Electronic Engineering Materials and Devices*, McGraw Hill, London, 1971. With permission.)

(i.e., the vectors that define the unit cell) can be used to describe the entire crystal. The choice of unit cell is not unique, and different unit cells may be defined for different crystals. In addition to the conventional unit cell described earlier, one particularly useful and commonly defined unit cell is the Wigner-Seitz unit cell (Ashcroft and Mermin 1976). The defining feature of this representation is that it defines a primitive unit cell (i.e., it contains exactly one lattice point per cell) and represents the full symmetry of the crystal. The Wigner-Seitz unit cell is shown together with the conventional unit cell for a face-centered cubic lattice in Figure 2.2.

The reciprocal lattice is defined by the set of wave vectors, $\mathbf{k}$, which satisfy the condition of defining a set of plane waves that represent the symmetry of the direct lattice). The reciprocal lattice vectors of the real space Bravais lattice.* Just as we defined a primitive unit cell in the real lattice, an equivalent primitive cell of the reciprocal lattice can be defined. A particular formation of such a primitive cell is referred to as the first Brillouin zone. The reciprocal lattice is determined in the same way as the Wigner-Seitz primitive cell in the real lattice. For more information on these fundamental concepts can be found in any solid state textbook, for example, Ashcroft and Mermin (1976) and Kittel (2005).

* A Bravais lattice is defined as an infinite array of discrete points with an arrangement and orientation that looks exactly the same regardless of which point in the array is viewed from. Where this is not the case, the array contains multiple atoms and can usually be defined, which can then be represented as a Bravais lattice.

Figure 2.2 (a) The conventional unit cell for a face-centered cubic lattice (f.c.c.). (b) The corresponding Wigner-Seitz unit cell (note the cubic lattice depicted here is different from that in (a)). (c) The first Brillouin zone of an f.c.c. crystal, with the points of high symmetry labeled.

The unique definition of the first Brillouin zone allows several points of high symmetry to be defined, as depicted in Figure 2.2c for a face-centered cubic lattice. The most obvious is at the zone center and is given the label Γ ; others are found at the edges of the Brillouin zone at different vertices and plane centers. To explain the purpose of introducing the reciprocal lattice and these high symmetry points, consider again an electron in the crystal lattice. The energy bands described earlier assume stationary electrons; however, one of the key properties of semiconductors (and metals) is that the electrons are not always stationary, and so the kinetic energy (KE) needs to be included. In a crystal, however, this is not simply $KE = \frac{1}{2}mv^2$, with m the mass and v the velocity, as the electrons interact with the nuclei (leading to a change in their effective mass), and the extent of this interaction is dependent on the electron's speed and direction of motion. In order to fully describe the energy of an electron, it then becomes necessary to define the kinetic energy for motion in different directions. This is where the reciprocal lattice and key symmetry points become useful.

The defined points of high symmetry in the reciprocal lattice correspond to wave vectors that define the momentum of electrons in the real crystal lattice. Plotting the energy of the electrons for these points in the reciprocal lattice, and along the lines joining these points, thereby provides a good description of the total energy landscape.

This type of representation is known as a dispersion relation, examples of which are shown in Figure 2.3. The horizontal axis in these figures is labeled by the high symmetry points in the reciprocal lattice. Between these points the position is varied linearly along the line connecting them in the reciprocal lattice. In real space, this can be thought of as changing either the direction or magnitude of the electron wave vector/momentum. The Γ -point is located at the center of the Brillouin zone and corresponds to zero momentum. The regions between the Γ -point and any other point, or between two points on the surface of the Brillouin zone, therefore correspond to increasing the magnitude of the wave vector/momentum in the direction defined by the other high symmetry point, or changing the direction of the wave vector/momentum, respectively.

The complete band structure of semiconductors is usually defined in this way, and many of their properties are described with respect to this representation. For example, the effective mass of electrons in the crystal lattice is determined by the curvature of the relevant band, where a steeper dispersion curve corresponds to a smaller effective mass. Conversely, if the band structure is not known, the effective mass can be determined experimentally, as described in Section 2.2.6, and this helps to define the band structure.

Conduction, thermal, and optical properties are also described within the band structure formalism, and similarly can also be used to experimentally determine aspects of the band

Figure 2.3 (a) The band structure of GaAs (Reprinted figure with permission from Bergstresser, T.K., *Phys. Rev.*, 141, 789, 1966. Copyright (1966) by the American Physical Society.) and (b) the band structure of Si (Reprinted figure with permission from J.R. and Cohen, M.L., *Phys. Rev. B*, 10, 5095, 1974. Copyright (1966) by the American Physical Society.) The horizontal axis represents the magnitude and direction of the wave vector, and in this way identifies the orientation of the wave vector, while between the points the magnitude is varied.

structure. The band structure can also be calculated by solving the Schrödinger equation for the crystal lattice, which is often achieved practically by making use of the $\mathbf{k} \cdot \mathbf{p}$ approximation. For details of this type of calculation, we refer the reader to Haug and Koch (2004).

Perhaps, the single most important property of semiconductors is the band gap. It is worth taking the time to detail exactly what is meant by this term. In the dispersion relation shown in Figure 2.3, it can be seen that the energy gap between the valence band and the conduction band varies as a function of wave vector, and so it is impossible to give a single value for the band gap of a material. The band gap of a material is defined as the energy difference between the highest energy valence band and the lowest minimum of the lowest energy conduction band. In the case of direct gap semiconductors, as shown in Figure 2.3a, this is straightforward, as they both occur at the same point in $\mathbf{k}$ -space (i.e., at the Γ -point of the Brillouin zone, $\mathbf{k} = 0$). In some materials, however, the valence band maximum and the conduction band minimum can be displaced from each other in momentum space, as shown in Figure 2.3b for silicon. In this case, the band gap is still defined as the energy difference between the valence band maximum and conduction band minimum, and is referred to as an indirect band gap.

2.1.2 Conduction Properties

At low temperature and in pure, intrinsic, semiconductors, the valence band is full and the conduction band is empty. In other words, all electrons are bound to their atoms, and in order to move to another atomic site the energy of the band gap would have to be overcome. This is very unlikely at low temperatures, meaning there are no free electrons, and the material is insulating. If, however, an electron can in some way be excited from the valence band, there is a continuum of empty states at the same or similar energy in the conduction band, and energy is required to move throughout the crystal lattice, allowing it to conduct.

In addition, the vacancy created in the valence band can be filled by another valence band electron, thereby moving the vacancy to another site, where another electron can move into it, and so on, thereby allowing the free movement of charge. Rather than thinking of this in terms of many electrons moving one at a time in sequence, it is helpful to think of the vacancy moving. In this way, by exciting the electron from the valence band to the conduction band, a quasiparticle in the valence band, referred to as a hole, is generated. The hole then has the opposite properties to an electron, including a positive charge, and it can be treated as a charge carrier that conducts electricity.

In thermal equilibrium, the probability of an electron being excited from the valence band into the conduction band is given by Fermi–Dirac statistics. This introduces the concept of the chemical potential or Fermi level. In semiconductors, the definition of the Fermi level is somewhat different from metals, since it usually lies in the band gap. The Fermi level, or chemical potential, defines the energy at which the probability for occupation is 0.5. At absolute zero, where the valence (conduction) band is completely full (empty), the Fermi level is in the middle of the band gap. At finite temperatures, the Fermi level lies within the order of $k_B T$ (where k_B is the Boltzmann constant and T is the temperature) of the center of the band gap. For intrinsic semiconductors where the band gap is large compared to $k_B T$, the Fermi level remains well within the band gap, and the probability of finding conduction electrons remains low.

In order to promote electrons from the valence band to the conduction band, energy greater than or equal to the band gap needs to be absorbed. For semiconductors, the band gap can range from a few hundred meV (e.g., InSb) up to ~ 6 eV (e.g., AlN) and often corresponds to optical photon energies (1–4 eV), meaning the absorption of visible light can lead to a transient conductivity. This is the basis for many light-detection and light-harvesting devices. For many applications, however, it is also desirable to have permanent conductivity. This can be achieved by doping the crystal lattice with impurities that have more (or fewer) valence electrons than the host semiconductor. For example, if an atom with five valence electrons replaces an atom with four valence electrons (e.g., P in Si), then four of the electrons will simply take the role of those in the original atom, leaving one extra electron. The energy level occupied by this electron is typically located just below the conduction band, as illustrated in Figure 2.4a. It can then be easily promoted thermally into the conduction band, allowing free movement of the charge. This is known as n-type doping and the dopant atoms are known as donors. In other words, the Fermi level is raised into or close to the conduction band, allowing free conduction. On the other hand, if an atom with three valence electrons is included in the lattice (e.g., B in Si), then an empty state is generated just above the valence band, into which an electron from the valence band can be excited, leaving a hole in the valence band that is

Figure 2.4 Schematic showing the incorporation of (a) n-type and (b) p-type dopants, and the excitation that allows them to conduct electrons and holes, respectively.

free to move and conduct electrical current, as shown in Figure 2.4b. This is known as p-type doping, with the dopant atoms known as acceptors, and corresponds to lower energy levels close to the valence band. Most consumer electronics are made up of combinations of intrinsic (pure) and extrinsic (doped) semiconductor regions, which make use of the interface between such regions. For more information on semiconductors and semiconductor devices, we refer the reader to Ashcroft and Mermin (1976), Sze and Kwok (2007), and Seeger (2004).

2.1.3 Compound Semiconductors

To date, we have been assuming that semiconductors are pure homogeneous materials made up of one single type of atom (with the exception of impurities). This need not be the case, however, as it is possible to form ionic crystals that are also semiconducting. Compound semiconductors are typically made by combining group III atoms with group V atoms (e.g., GaAs), or group II with group VI atoms (e.g., ZnO). In this way, the number of valence electrons remains the same as if all atoms were from group IV, and the bonding is tetrahedral. The fact that the nuclei are no longer identical does not affect most properties of the crystal structure and average number of electrons per site that determine semiconductor behavior, and these are unchanged.

One example of a situation in which the local environment and the heterogeneity of the unit cell can play an important role is at boundaries and interfaces. Since a compound semiconductor contains at least two different types of atoms, it is not isotropic in at least one direction, taken in isolation, it will be electrically polarized, as the electric field of the two atoms will almost certainly not be the same. This spontaneous polarization is not seen in an infinite lattice, as the fields from neighboring unit cells counter each other. At a surface, however, there is no neighboring unit cell, and an excess of charge can be induced to an electric field. This type of charge imbalance can also be found at an interface between two different materials, where the spontaneous polarization of the two unit cells does not match. This can be particularly relevant in semiconductor nanostructures.

The major reason compound semiconductors are of particular interest is the range of properties they allow. Take GaAs, for example; by replacing a fraction, x , of the Ga atom with Al (another group III element) to form $\text{Al}_x\text{Ga}_{1-x}\text{As}$, the band gap can be increased or decreased that is dependent on the Al concentration. Alternatively, the band gap can be tuned by replacing Ga with In to form $\text{In}_x\text{Ga}_{1-x}\text{As}$. This ability to tune the band gap provides an extra degree of freedom that is particularly useful when designing semiconductor nanostructures, especially those that emit or absorb light.

2.1.4 Phonons

Much like any other crystalline solid, vibrations within semiconductors are described by classical mechanics, this is similar to the normal modes of vibration. When treated quantum mechanically, however, these “normal modes” have particle-like properties and can be treated as quasiparticles* known as phonons. It then becomes a matter of counting about the number or density of phonons, which increases with increasing temperature.

* A quasiparticle is an elementary excitation that exists only in a solid (e.g., phonons, excitons, etc.) and has properties like position, momentum, and mass that are modified by interactions with the solid. This contrasts with “real” particles like electrons, protons, and photons that can exist in vacuum (Seeger, 2009).

Figure 2.5 (a) Longitudinal acoustic and (b) longitudinal optical phonon modes for a 1D lattice. For acoustic modes the atoms in the unit cell move in phase, while for optical modes they move out of phase.

details regarding the derivation of phonons, see, for example, Ashcroft and Mermin (1976, pp. 437–443) and Ridley (1997).

As an example, consider the simplest case of a one-dimensional (1D) lattice. In this case, there are two general types of vibrations possible, as depicted in Figure 2.5. In the first, (a), the two different atoms that make up the unit cell move in phase, while in the second, (b), they move out of phase. The former case represents acoustic phonons while the latter corresponds to optical phonons, so called because in an ionic crystal this oscillating separation is in fact an oscillating dipole, which makes this type of phonon optically active. Extending this picture to three dimensions is not so straightforward conceptually, but the basic principles remain the same: Acoustic phonons correspond to intercellular interactions (i.e., nuclei within the unit cell move in phase), while optical phonons correspond to intracellular interactions (i.e., nuclei within the unit cell move out of phase) (Ashcroft and Mermin 1976).

One further consideration is the orientation of the vibrational motion relative to the phonon propagation direction. Consistent with the quasiparticle picture, phonons do not exist across the entire lattice simultaneously, but rather are localized, and move through the lattice with a given velocity. This defines the propagation direction and allows the definition of transverse and longitudinal modes. In general, one can define three orthogonal orientations for the vibrational motion: one that is parallel to the propagation direction (longitudinal) (Figure 2.5), and two that are orthogonal to the propagation direction (transverse) (Figure 2.6). This definition of transverse and longitudinal is not always going to be strictly true, as it depends on the direction of propagation relative to the principal axes of the crystal lattice; nonetheless, this terminology remains useful and continues to be used.

A corollary of phonon propagation is that they transfer momentum through the crystal lattice. The momentum of a phonon is manifested in the vibrational motion of the nuclei, and is transferred throughout the lattice as the phonon propagates. The phonon's momentum is one

Figure 2.6 (a) Transverse acoustic and (b) transverse optical phonon modes for a 1D lattice. Again, for acoustic modes the atoms in the unit cell move in phase, while for optical modes they move out of phase.

Figure 2.7 Example of phonon dispersion curve and density of states for Si. with permission from Savrasov, B., *Phys. Rev. B*, 54, 16470, 1996. Copyright American Physical Society.)

of its key properties, and similar to electrons moving throughout the crystal propagating in different directions will have different energies depending on and direction of their wave vector with respect to the principal axes of the lattice. In addition, it is useful to introduce phonon dispersion curves that plot phonon energy against momentum (or wave vector), an example of which is shown in Figure 2.7 (Savrasov et al. 1996). It is evident in this figure that the dispersion of optical phonons is rather flat, while that of the acoustic phonons is strongly dependent on momentum and zero at the $\mathbf{k} = 0$ point. This is typical for most crystals, as is the observation that for a given energy of optical phonons is generally greater than for acoustic phonons, particularly at the zone center. This means that at low temperatures it is possible to “freeze out” optical phonons, allowing the effects of acoustic phonons to be isolated. Optical phonons typically have energies in the range of a few tens of meV, which is comparable to the thermal energy at room temperature (25 meV), thus as the temperature is reduced, the prevalence of optical phonons rapidly decreases. At even lower temperatures, it becomes possible to minimize the optical phonon density, but because the linear dispersion curve goes to zero at the Γ point, there will always be a finite acoustic phonon population at finite temperatures; under certain conditions, however, zero-phonon effects can still be observed (Borri et al. 2001).

A corollary of treating phonons as quasiparticles is that they are able to interact with other particles and quasiparticles. For example, phonons can scatter with electrons, holes, and photons in either elastic or inelastic collisions, and can transfer energy. The effects of these interactions may include alteration of the electrical conductivity, carrier recombination of optically excited carriers, or dephasing of a coherently excited state. Changes to these properties as a function of temperature are therefore usually dominated by phonon interactions, as the phonon density (corresponding to the number of phonons) increases with temperature (Shah 1999).

Interactions of phonons with photons can be divided into two different categories: absorption and Raman scattering. In order for an IR photon to be absorbed

phonon, both energy and momentum must be conserved. From the phonon dispersion curve, it is clear that the only place where a photon will have the same energy and momentum as an acoustic phonon is at $\mathbf{k} = 0$, corresponding to a static electric field. The photon dispersion will however overlap with the optical phonon branches at finite values of E and $\mathbf{k}$. In order for the electromagnetic field of the photon to interact with the phonon modes, the crystal must have a permanent dipole moment, that is, the bonds must be at least a little polar, as will be the case in compound semiconductors, but not monotonic semiconductors such as Si. Furthermore, since the electric field of the photons is transverse, it will interact only with the TO phonon modes. Phonons that satisfy these selection rules are said to be IR active (Klingshirn 2007).

Raman scattering and Brillouin scattering (which is equivalent to Raman scattering, but with acoustic phonons) are discussed in greater detail later in this chapter (and in Chapters 6 and 14). At this point, however, we point out that all phonon modes can potentially be Raman active depending on the crystal structure. The precise selection rules to determine whether a phonon mode is Raman active is somewhat more complicated, requiring group theory and a more detailed analysis of the symmetry of the crystal and corresponding phonon modes. In general, however, if the crystal structure has an inversion center, then the phonon modes will be either IR or Raman active; if the crystal lacks such inversion symmetry, then some phonon modes may be both IR and Raman active. Fox (2001) gives a more detailed summary of the optical properties of phonons.

2.1.5 Other Lattice Excitations

In addition to phonons, there are other forms of lattice interactions that can play an important role in semiconductor physics. We briefly mention two of these here: polarons and magnons.

Polarons are formed by the interaction of the electric field of an excited electron (or hole) with that of the nuclei. This interaction causes a deformation of the lattice in the vicinity of the charge carrier, which leads to a local polarization of the lattice. As the free charge carrier moves throughout the lattice, the deformation and polarization move with it. This combination of the charge carrier and its polarization field is treated as a quasiparticle known as a polaron. One important effect of the local induced field is that it hinders the motion of the electron through the lattice, thus reducing its mobility (Klingshirn 2007). This often leads to a "hopping"-type conductivity when polarons are the dominant charge carriers.

Magnons are quasiparticles that represent quantized spin waves and are completely analogous to phonons. Where phonons represent excitations of the nuclear lattice and carry momentum, magnons represent excitations of the spin lattice and carry angular momentum. Similarly, just as phonons can have an important role in energy loss mechanisms, so too can magnons in magnetic semiconductors (Klingshirn 2007).

2.2 OPTICAL PROPERTIES OF SEMICONDUCTORS

The optical response of semiconductors is one of their most useful characteristics and one of the most important tools for determining their band structure. In discussing the optical response, the most important property is the size of the band gap and whether it is direct or indirect. The presence of a band gap between the full valence band and empty conduction band means that a photon with energy greater than the band gap can be readily absorbed, whereas photons with energy less than the band gap will typically be transmitted. For many semiconductors of interest, the band gap is around the optical region of the electromagnetic spectrum, and so it is not surprising that semiconductors now form the basis for many light emitting and detecting devices.

Figure 2.8 (a) The direct transition (I) in an indirect gap material requires for energy than the phonon-assisted transition (II) between the band edges. In the transition, the phonon (dashed line) provides the additional momentum to allow the transition. (b) Similar processes can also occur in direct gap materials such as GaN and lead to the appearance of phonon replica peaks labeled by the dashed lines and the labels indicate the number of phonons involved in the transition. (Reprinted figure with permission from Klingshirn, *Phys. Rev. B*, 54, 2518, 1996. Copyright (1996) by the American Physical Society.)

Knowledge only of the band gap, however, is not enough to fully describe the optical properties. In addition to energy conservation, which can be specified with respect to the photon, momentum and angular momentum must also be conserved.

The momentum carried by a photon ($p = \hbar\mathbf{k} = h/\lambda$) is small compared to that of electrons/holes, and so optical transitions where the change in momentum (or hole wave vector) is close to zero tend to dominate. This is particularly the case for direct gap semiconductors. For indirect gap semiconductors, a valence to conduction band transition with minimum photon energy can occur only if a phonon is absorbed or emitted to facilitate conservation of momentum. This type of phonon-assisted transition dominates for indirect gap materials, such as silicon, causing the optical properties to be increasingly temperature dependent. This type of phonon-assisted transition is shown in Figure 2.8a, where path I represents the direct transition, which requires a photon with energy greater than the band gap, and path II represents the phonon-assisted transition where a photon and phonon are absorbed simultaneously to allow the indirect transition between the band edges, conserving both energy and momentum.

Phonon-assisted transitions can also occur in direct gap semiconductors, leading to additional "phonon-replica" peaks in the absorption/emission spectrum, superimposed on the "zero-phonon" peak by multiples of the phonon energy. The presence of phonon-assisted transitions is significantly enhanced for strongly polar crystals due to the strong interaction of phonons with incident photons (Fox 2001). An emission spectrum for GaN, a very direct gap material, is shown in Figure 2.8b, with five phonon replica peaks present.

Another important consideration in determining the optical properties of semiconductors is the conservation of angular momentum. Photons always carry exactly one unit of angular momentum, and so only transitions that involve a change of angular momentum are allowed. In order to assess the selection rules for optical transitions, we again refer to the band structure. Figure 2.9 shows the three highest energy bands corresponding to the atomic p-orbitals, with orbital angular momentum $l = 1$.

Figure 2.9 The three highest valence bands are shown. The SO band, with $j = 1/2$, is split from the $j = 3/2$ bands due to the spin-orbit interaction. The two $j = 3/2$ bands are also split except at the point of highest symmetry.

degenerate p-like valence band is split by the spin-orbit interaction* into a fourfold degenerate band with total angular momentum $j = 3/2$ and doubly degenerate band with $j = 1/2$. The lowest band, with $j = 1/2$, is referred to as the spin-orbit split-off (SO) band and in most cases does not significantly contribute to the optical properties. At points of high symmetry, the $j = 3/2$ band is fourfold degenerate; however, away from the zone center (i.e., the Γ -point) a splitting occurs, as can be seen in Figure 2.9. The bands are split according to the magnitude of the z -projection of the angular momentum, or magnetic quantum number, m_j , so that the states with $m_j = \pm 1/2$ form one band and those with $m_j = \pm 3/2$ form the other. The dispersion relation of a free particle is parabolic, since $E = p^2/2m$, and the curvature is related to the particle mass. The different curvature of the two $j = 3/2$ bands is interpreted as the effective mass of the holes that are generated when electrons are excited. The band with $m_j = \pm 1/2$, which has greater curvature, is therefore termed the light-hole (LH) band, and that with $m_j = \pm 3/2$, the heavy-hole (HH) band. The physical reason for the different curvatures is that HHs, with a greater magnetic quantum number, and hence greater magnetic moment, interact more strongly with the lattice, giving the appearance of being “heavier” and therefore leading to a shallower dispersion relation.

The angular momentum of the conduction band is determined entirely by the electron spin, as the orbital angular momentum, l , equals zero for the lowest conduction band. The total angular momentum is, therefore, $j = 1/2$ and the band is doubly degenerate with $m_j = \pm 1/2$. If we represent each of these energy bands as discrete states, as shown in Figure 2.10, the allowed transitions can be determined by the requirement for a change of one unit of angular momentum in absorbing a photon. The states involved in these transitions are all pure spin states and require a change in m_j of ± 1 , which means that these pure transitions are circularly polarized, with σ^+ polarized photons adding 1 to m_j and σ^- subtracting 1 from m_j . Excitation with linearly polarized light would, in contrast, excite a linear superposition of the

Figure 2.10 The requirement for conservation of angular momentum limits the transitions from the $j = 3/2$ bands to those indicated, each of which are circularly polarized.

two circularly polarized transitions. This can be a useful tool in determining the semiconductor states and identifying or selecting transitions under various conditions.

In addition to the optically allowed transitions described earlier, there are also transitions that involve the addition of zero (e.g., $m_j = 1/2$ to $m_j = 1/2$) or two (e.g., $m_j = -3/2$ to $m_j = 3/2$) units of angular momentum. These transitions are one-photon forbidden, but two-photon allowed. In two-photon absorption, the angular momentum contribution from the photon is added or subtracted, leading to a change of two or zero units of angular momentum. The other requirement for two-photon absorption is that the sum of the photon energies equals the energy of the transition. These principles can be extended to higher order multiphoton absorptions, and allow access to other “dark” states.

2.2.1 Excitons

To date, we have discussed electrons and holes as free carriers that do not interact. This is not always the case. The treatment of holes as particles with positive charge implies that electrons and holes should interact via the Coulomb potential. This interaction, and the Coulombic attraction can lead to bound electron-hole pairs, for which two cases are typically considered. The first is that of strong electron-hole attraction, where the electron and hole are tightly bound to each other and are located within a single unit cell. This is referred to as a Frenkel exciton, and these are commonly observed in molecular systems; however, in most semiconductors they are not observed, as the interaction between the electron and hole is strongly screened by the dielectric constant. In such cases, the electron and hole are only weakly bound and extend over many unit cells. These excitons are known as Wannier-Mott excitons, and we will focus our attention on them, since this description usually applies to excitons observed in semiconductors (see, e.g., 2001, Klingshirm 2007).

Wannier excitons are analogous to an electron and proton “orbiting” about a common center of mass, as in the hydrogen atom, and many of their properties can be derived from the well-known properties of the hydrogen atom, simply by accounting for changes in effective mass and dielectric constant. This bound pair state is lower in energy than the unbound electron and hole, and the amount of energy required to separate the exciton into its constituent electron and hole is called the exciton binding energy. In common semiconductors, the binding energy is in the range of tens of meV, which means, particularly at low temperatures, that exciton formation is often the dominant excitation.

* The spin and orbital angular momentum of charged particles generate finite magnetic moments that interact with each other, leading to a splitting of otherwise degenerate states. This interaction is referred to as the spin-orbit interaction.

Continuing the analogy with the hydrogen atom, in addition to the ground state of the exciton, there are a series of quantized excited states, with energy given by

$$E_{bn} = \frac{E_{b0}}{n^2}, \quad (2.1)$$

where

E_{b0} is the binding energy of the lowest exciton state
 n is the principal quantum number

The exciton transition energy is then given by the difference between the band gap and the exciton binding energy:

$$E_{xn} = E_g - E_{bn}. \quad (2.2)$$

The Bohr radius, which represents the average separation between the bound electron and hole, varies from ~3 to 35 nm in common bulk semiconductors and is defined as

$$a_0 = \frac{\hbar^2 \epsilon_0}{m_{eff} e^2}. \quad (2.3)$$

To ascertain the role of excitons in the optical properties of semiconductors, consider the oscillator strength, which depends linearly on the overlap of the electron and hole wave functions, and is directly proportional to the absorption magnitude. It is clear that the wave function overlap between bound electrons and holes (i.e., excitons) will be far greater than between free electrons and holes, and therefore that the oscillator strength for excitons will exceed that for free electron-hole pairs. As a result, excitons appear in absorption spectra as a series of sharp, intense peaks decreasing in intensity as energy is increased from the ground state exciton to the band-edge transition energy (which can also be renormalized by the electron-hole Coulomb interaction, discussed in more detail in Chapter 8). It then becomes clear that at temperatures where the thermal energy is less than the exciton binding energy and excitons are abundant, they dominate the optical response of semiconductors (Klingshirm 2007).

Exciton binding energies vary over more than an order of magnitude in common semiconductors, from 4.3 meV in GaAs (corresponding to a temperature of 50 K, below which excitons begin to dominate) to 60 meV for ZnO (corresponding to 696 K, indicating that excitons dominate at and above room temperature) (Jagadish and Pearton 2006). A list of the exciton binding energy, Bohr radius, and band gap for many common bulk semiconductors can be found in Table 2.1. The absorption spectrum of high-quality GaAs at 1.2 K is shown in Figure 2.11, where peaks corresponding to the first three exciton states can be identified.

The optical selection rules that govern excitation of electrons from the valence band to the conduction band also apply for excitons, and to date we have been considering just the bright, optically allowed excitons. The electron and hole states involved in the one-photon forbidden transitions discussed in the previous section can also form excitons. These “dark excitons” have angular momentum zero or two, meaning they cannot be generated in one-photon processes. In addition to generating these dark excitons with multiphoton absorption, they can also be formed from bright excitons following a spin flip of either the electron or hole. Such spin-flip processes can be quite common under certain conditions, and thus present an important relaxation pathway for bright excitons. Furthermore, because dark excitons are not optically active, radiative recombination is closed as a relaxation pathway, and so they tend to have significantly longer lifetimes.

TABLE 2.1

The Band Gap (E_g), Exciton Binding Energy, and Exciton Bohr Radius for Several Common Semiconductors

Material	Band Gap (eV)	Binding Energy (meV)
GaN (Fox 2001)	3.5	23
ZnO (Madelung et al. 2002, Jagadish 2006)	3.4	60
ZnSe (Fox 2001)	2.8	20
CdS (Fox 2001)	2.6	28
CdSe (Fox 2001)	1.8	15
GaAs (Fox 2001)	1.5	4.2
InP (Fox 2001)	1.4	4.8
Si (Smith et al. 1975)	1.1	14.7
GaSb (Fox 2001)	0.8	2.0
Ge (Madelung et al. 2002, Smith et al. 1975)	0.66	3.6
InAs (Madelung et al. 2002)	0.36	1.0

Figure 2.11 The absorption spectrum of high-quality GaAs at 1.2 K. Peaks of the first three exciton states can be seen. (Reprinted from *J. Luminescence*, 30, Fox et al., 154. Copyright 1985, with permission from Elsevier.)

Just as optical transitions between unbound electrons and holes predominate at the center of the Brillouin zone (i.e., $\mathbf{k} = 0$), so too excitons with $\mathbf{k} \sim 0$ dominate because of the minimal momentum carried by photons. Excitons with $\mathbf{k} \neq 0$ are involved in optical transitions where absorption or emission of a phonon is involved, to phonon replicas, as discussed previously. The coupling of excitons and phonons is stronger than that between unbound carriers and phonons because the stronger interaction of excitons enhances the interaction. This means that phonon-assisted exciton transitions are more prevalent than the equivalent transition between unbound electrons and holes. Similarly, phonon-assisted transitions allow the formation of indirect excitons in materials such as Si and Ge. This is in contrast to direct excitons in such materials, which can be formed by excitation high up into the conduction band, but which are rapidly relaxed to the rapid relaxation of the electron. Indirect excitons have a finite momentum

conserved, which leads to reduced relaxation pathways, and an increase in their lifetime from nanoseconds to microseconds, compared to excitons in direct gap materials (Timusk 1976).

Under certain conditions, other bound particles can be detected. These include charged excitons, which consist of either two electrons and one hole, or two holes and one electron, and which can dominate for n- or p-doped semiconductors, respectively. Another excitation that is commonly observed in intrinsic semiconductors is the biexciton, which consists of two electrons and two holes. Biexcitons become increasingly prevalent as the number of carriers injected into the material increases (biexciton luminescence increases quadratically with excitation intensity), and typically have a binding energy of a few meV (Klingshirn 2007).

2.2.2 Defect States

There are many different types of defects that can be incorporated in semiconductor materials. Some may be deliberately included to change the band gap, or introduce free carriers, as discussed previously. Others may arise unintentionally and possess properties that prohibit proper device operation.

In general, defects can be classified as point defects, which usually involve isolated atoms or sites in the crystal lattice, or line defects, which include dislocations in the crystal lattice. We concentrate on the former, as they tend to have a greater effect on the optical properties.

Point defects can be further classified according to the nature of the fault. The most common types of point defects are vacancies, created by a missing atom; interstitials, created by an atom occupying an interstitial site; and substitutional defects, where an atom is replaced by an atom of a different type. Here we briefly discuss the optical properties of such defects, but for a complete review of the electrical, thermal, and optical properties of defects and their prevalence, we refer the reader to references Lannoo and Bourgoin (1981), Watts (1977), Pantelides (1978), and Klingshirn (2007).

Each of the point defects mentioned earlier can act as either a donor or acceptor (depending on the bulk crystal structure, and the atom/s involved), and it is frequently this property, rather than the specific type of defect, that determines the effect on the optical properties. One of the greatest effects occurs when these defects are able to bind an exciton (i.e., the bound electron and hole are both localized at the defect). The binding energy for such a system is typically greatest for a neutral acceptor,* followed by a neutral donor and then by an ionized donor, while ionized acceptors are not usually able to bind excitons (Klingshirn 2007). These bound excitons have zero degrees of translational freedom, and so interact less with charge and lattice excitations, leaving few available broadening mechanisms. Such states appear as very sharp peaks in luminescence and absorption spectra, redshifted from the free exciton peak by the defect binding energy. The magnitude of this binding energy is typically less than 10 meV, which means that defect-bound excitons can be observed only at low temperatures, where the thermal energy is insufficient to promote escape from the defect.

Another corollary of the lack of translational degrees of freedom for excitons bound to defect states is that the requirements for momentum conservation are reduced.[†] This means that these excitons can easily couple to acoustic phonons in direct gap semiconductors, leading to acoustic wings in absorption and emission spectra. Similarly, for indirect semiconductors, defect states can be observed without the participation of phonons.

* A neutral acceptor (donor) is a p-type (n-type) defect that has not given up its extra hole (electron), while an ionized acceptor (donor) is one that has.

[†] The Heisenberg uncertainty principle states that $\Delta x \Delta p > \hbar/2$, and so the accurately known position (i.e., Δx is small) leads to a large uncertainty in momentum, Δp . This then means that the strict requirement for conservation of momentum can be relaxed.

The effects of defect states have so far been discussed with respect to their structures, and indeed, the spectra of defects can be rather complicated, and can involve processes involving excited bound states, multiple defects, free-to-bound recombination of defects. There exists, however, a range of defects that are not and instead assist non-radiative recombination processes, thereby quenching luminescence. Dislocations are a prime example of this type of defect, where the electron is transferred to the lattice. Another example is so-called deep centers (as opposed to the “shallow” defect states discussed to date) where the defect state lies in the middle of the band gap. In this case, they can interact with both the conduction and valence bands and lead to rapid non-radiative recombination (Pantelides 1986, Klingshirn 2007). In some cases, however, they can act as radiative centers, and lead to absorption and emission from the band-edge transition (Pantelides 1986). The prevalence of these “deep centers” increases as the band gap increases, and so they are more common in wide gap semiconductors such as ZnO and GaN (Reshchikov and Morkoc 2005), and in insulators.

2.2.3 Free Carrier Absorption

When additional carriers introduced by defects can be thermally ionized, the free carrier concentration increases, and their interaction with light fields becomes very similar to free carrier absorption. This is particularly relevant for n- or p-doped materials and small-gap semiconductors, where defect states tend to be very shallow and ionize easily. To understand these interactions, we treat the material as a simple metal and analyze the properties on the basis of the free electron model, as discussed in Chapters 1 and 3. The main difference between metals and free carriers in semiconductors is that the carrier concentration can be changed in semiconductors by dopant density, and typically is significantly less than the free carrier concentration in metals. As a result, the plasma frequency (Chapters 1 and 3), which determines many of the optical properties, is in the infrared range for doped semiconductors, compared to the visible range for metals.

In semiconductors, since either electrons or holes can be the majority carriers, free carrier absorption can contribute to the optical properties in different ways. When electrons are the majority carriers, free carrier absorption is typically due to intraband transitions, and typically involves the involvement of a phonon in order to conserve momentum, as shown in Figure 2.12a. In this case, free carrier absorption can be used as an important tool for measuring phonon scattering rates and interaction strengths (Yu and Cardona 2005). When holes are the majority carrier, interband transitions are more common, involving transitions between SO, LH, and HH bands, as shown in Figure 2.12b.

Figure 2.12 (a) Free carrier transitions in the conduction band are limited to intraband transitions and usually require the involvement of a phonon. (b) Free carrier transitions can be inter-band transitions, between SO, LH, and HH bands.

Using this process, the ratios of the effective mass at the top of the valence bands can be determined by comparing the absorption spectra for transitions between different bands (Fan 1967, Yu and Cardona 2005).

One final difference between the optical properties of free carriers in semiconductors and metals is that the plasma resonances can be much sharper in semiconductors. The reason for this is that for frequencies above the plasma frequency the photon energy is often less than the band gap in semiconductors, whereas in metals, it may still be sufficient to allow inter-band transitions, thereby broadening the plasma edge.

2.2.4 Optical Emission (Radiative Recombination)

To this point, we have concentrated predominantly on the interaction of semiconductors with an incoming light field. What happens after this initial interaction is, however, at least as important (Shah 1999). Many applications of semiconductors utilize the recombination of electrons and holes and subsequent emission of light, for example, as luminescence in LEDs, or stimulated emission in lasers. The dynamic relaxation processes that can occur between excitation and recombination, together with the processes that can affect both radiative and non-radiative recombination dynamics are discussed in detail in later chapters (e.g., Chapter 8). In this section, we limit the discussion to a brief description of the interactions between semiconductors and outgoing light fields.

Emission of light from a semiconductor is referred to generally as luminescence, occurring when an electron and hole recombine, most often through an electronic transition from the conduction band to the valence band. This transition usually occurs from a state close to the conduction band minimum to a state close to the valence band maximum, regardless of where in the band structure the carriers are initially generated. The reason for this is that carriers generated at higher energies undergo intraband relaxation, or cooling, at a much faster rate than they radiatively recombine (Chapter 8, Shah 1999). As a result, emission spectroscopies are a sensitive probe of low energy states. This type of band-edge emission is particularly dominant in direct gap semiconductors; in indirect gap semiconductors, however, it requires the participation of phonons and is therefore much less efficient than in direct gap semiconductors, as is the case for absorption.

Depending on how the electrons and holes are initially created in the sample, this type of radiative recombination is referred to as photoluminescence, for carriers excited optically (Chapter 5); thermoluminescence, for carriers excited thermally; electroluminescence, for carriers injected into the sample by an external current; or cathodoluminescence, for carriers generated by electron bombardment.

Regardless of the manner of excitation, the emission of light from semiconductors can generally be treated as the inverse process to absorption. In the case of absorption, a semiclassical approach is normally sufficient. If we follow this same procedure, however, we find that in order for emission to occur, an electromagnetic wave needs to be present (Yu and Cardona 2005). This process is known as stimulated emission; however, our experience tells us that it is also possible for emission to occur in the absence of an electromagnetic wave. To explain this, a full quantum mechanical treatment of the light field is required, which identifies the probability of emission as being proportional to $1 + N_p$, which is clearly nonzero even when N_p , the photon number, is zero. In this case, the "spontaneous emission" can be thought of as being stimulated by the "zero-point amplitude" of photons,* which is independent of N_p , while "stimulated emission" corresponds to the part proportional to N_p .

* "Zero-point" energy is a fundamental outcome of quantum mechanics that identifies the lowest energy state as still having a finite energy. This is evident even in the simplest case of a quantum harmonic oscillator, where the ground state is higher in energy than the potential minimum.

Figure 2.13 The strong interaction of an exciton and photon leads to the polaritons, which have a band structure that is generated by combining the exciton and photon with an avoided crossing at their intersection.

As with any rate equation, the rate of spontaneous emission is dependent on the population of initial states (i.e., the density of electrons in the conduction band) and the density of unoccupied final states (i.e., the density of holes in the valence band) (Chapter 8). An additional consideration is the overlap of the carriers in both reciprocal and real space, recombination typically occurs between carriers at the center of the Brillouin zone where the change in momentum is small. In real space, the recombination rate is proportional to the overlap of the electron and hole wave functions. For free electron-hole recombination, the overlap is averaged over the entire crystal and is dependent on the concentration of electrons and holes. For excitons, the electron-hole overlap is far greater, and so the recombination rate is faster. With regard to luminescence from excitons, however, one further consideration is required, as identified by Toyozawa (1959). To accurately reproduce experimental luminescence spectra, it is essential to include the strong coupling between exciton and photon. This coupling combines to create a polariton, as indicated in the band structure in Figure 2.13. The dispersion curve of the lower polariton does not have a minimum, and so one might not expect a sharp emission peak. However, by considering the lifetimes of polaritons at different wave vectors along the dispersion curve, it can be shown that a bottleneck occurs near the transverse exciton. This leads to an asymmetric peak that adequately reproduces experiments. A detailed discussion of the exciton polariton spectrum is beyond the scope of this chapter, and the reader is referred to Toyozawa (1959) and Yu and Cardona (2005).

2.2.5 Inelastic Scattering

The final type of interaction between semiconductors and optical light fields discussed here involves the inelastic scattering of an incident light field, resulting in a scattered photon shifted in energy. Normally, we think of light scattering off a static defect or

it is also possible for light to scatter from a transient change in the susceptibility of the material, such as those generated by phonons. This type of inelastic scattering between photons and phonons is referred to as Raman scattering, or in the specific case of acoustic phonons, Brillouin scattering.

The interaction of a light field with a phonon, or more specifically with the modulated susceptibility induced by the phonon, can generate a photon with energy that has been either increased or decreased by the phonon energy. The two Raman signals are then observed at $\omega - \nu$ corresponding to Stokes scattering, and at $\omega + \nu$ corresponding to anti-Stokes scattering, where ω is the frequency of the light and ν is the frequency of the phonon.

In addition to conserving energy, the Raman scattering process is also required to conserve momentum. Where photons are incident on the material, the magnitude of the change in momentum is limited to less than twice the photon momentum (Chapter 6). This means that for optical photons, only phonons close to the zone center, that is, those with little momentum, can scatter in this way with the light field. For acoustic phonons, this also limits the energy of the phonons as, unlike optical phonons, the dispersion curve increases almost linearly with increasing momentum. A useful implication of this is that the velocity of acoustic phonons can be determined by careful analysis of Brillouin scattering (Benedek and Fritsch 1966).

The requirement that only phonons with little momentum can participate in Raman scattering is relaxed when two-phonon Raman scattering processes are considered. In this case, the two phonons can provide roughly equal and opposite momentum, thereby allowing momentum to be conserved, even though the photon momentum still only changes a small amount. This means that two-phonon Raman scattering can involve all phonons, independent of momentum, and provides a means for measuring the phonon density of states (Uchinokura et al. 1974).

2.2.6 Optical Properties in a Magnetic Field

A free electron in a magnetic field will rotate about the magnetic field with a frequency given by $\omega_c = eB/m_{\text{eff}}$, where ω_c is the cyclotron frequency and m_{eff} is the effective mass (Ashcroft and Mermin 1976). In semiconductors, where there are free electrons in the conduction band, or free holes in the valence band, this is also the case. Indeed experiments making use of this phenomenon are used to determine the effective mass of electrons and holes by measuring the cyclotron frequency using resonant microwaves in a manner similar to nuclear magnetic resonance (NMR) (Yu and Cardona 2005).

In addition to this common effect, there are three main effects of an external magnetic field that can be seen in the optical response of semiconductors: Zeeman splitting, diamagnetic shift, and the formation of Landau levels. A detailed understanding of these effects allows the experimentalist to measure various properties of semiconductor materials.

2.2.6.1 Zeeman Splitting

Excitons, electrons, and holes all possess spin and in some cases orbital angular momentum. This means that they, as charge carriers, also have a finite magnetic dipole moment, proportional to the z -projection of the angular momentum, m_j . When an external magnetic field is applied to such a dipole moment, its energy is shifted by an amount proportional to $g\mu_B \mathbf{B} \cdot \mathbf{j}$ ($=g\mu_B B m_j$), where g is the Lande g -factor, μ_B is the Bohr magneton, $\mathbf{B}$ is the magnetic field, and $\mathbf{j}$ is the angular momentum vector, with the scalar product, $\mathbf{B} \cdot \mathbf{j}$, replaced by the magnitude of the magnetic field and m_j , the projection of the angular momentum in the direction of $\mathbf{B}$. This shift is known as the Zeeman shift and causes otherwise degenerate levels to become split by an amount linearly dependent on the applied field strength. Figure 2.14

Figure 2.14 The Zeeman splitting proportional to m_j is shown, with the solid lines the original states and the dashed lines the Zeeman shifted states. The splitting between the HH and LH bands is larger than that of the LH and conduction band states; however, the change in transition energies is given by $\pm g\mu_B B$.

shows the splitting of the HH, LH, and conduction band states that were present in Figure 2.10 and the allowed transitions between them. From this figure, it is clear that the two allowed HH transitions are split by the same amount as the two allowed LH transitions. The reason for this becomes more evident by considering the total angular momentum of the excitons, as shown in Figure 2.15. In this case, the ground state is the zero exciton state and the optically allowed excited states are the one exciton states with angular momentum projections of ± 1 . In this representation, the optically allowed HH-excitons both have angular momentum vectors ± 1 , and both are therefore split by the same amount in an applied magnetic field. The exciton states with $m_j = 0$, ± 2 are not optically allowed and correspond to the forbidden transitions that require a change of zero or two units of angular momentum, respectively.

Figure 2.15 By switching to the exciton representation, it becomes clear that the HH and LH excitons are split by the same amount by the Zeeman effect. The lower levels represent the zero exciton ground state, and the dark excitons are also shown for completeness.

2.2.6.2 Diamagnetic Shift

The application of a magnetic field to an electron or hole causes it to rotate in an orbit, thereby giving it additional angular momentum. Similarly, the application of a magnetic field to an exciton causes a change in the relative motion of the electron and hole, and a change in the overall angular momentum. If we treat the exciton states as atomic orbital-like states, then this deformation can be thought of as adding a component of a different state, the amount of which is proportional to the field strength. For example, an exciton in a purely s-like state ($l = 0$) will develop some p-like character ($l = 1$) in a magnetic field. This causes a change in the total angular momentum of the exciton, with magnitude linearly dependent on the field strength. The orientation of the angular momentum is such that the exciton develops a magnetic dipole moment in the opposite direction to the applied field, as predicted by Lenz's rule. The magnitude of this magnetic dipole is then linearly dependent on field strength. The energy of such a magnetic dipole, μ , in an external magnetic field, $\mathbf{B}$, is given by $\mu \cdot \mathbf{B}$, that is, it is linearly dependent on the field strength and on the magnitude of the dipole moment, as observed in the Zeeman effect. Since the magnitude of the field-induced dipole moment is also linearly dependent on the magnetic field strength, the energy of an exciton in an applied magnetic field increases quadratically with the field strength. This is known as a diamagnetic shift.

2.2.6.3 Landau Levels

The presence of the cyclotron orbits discussed at the beginning of this section can be predicted from classical mechanics. However, a quantum mechanical treatment of this interaction between free carriers and an external magnetic field reveals additional behavior.

Solving the Hamiltonian for an electron in a magnetic field reveals that parallel to the field the wave function is unchanged and the electron is free. Perpendicular to the field, however, the Hamiltonian resembles that for a simple harmonic oscillator and gives eigenvalues:

$$E = \left(n + \frac{1}{2}\right) \hbar \omega_c \quad \text{with } n = 0, 1, \dots, \quad (2.4)$$

where ω_c is the cyclotron frequency ($\omega_c = eB/m_{\text{eff}}$). As a result, the carrier is free to move in the direction parallel to the field, but confined perpendicular to it into concentric circles, which correspond to the different states of the harmonic oscillator, the so-called Landau levels. This confinement in two dimensions generates a quasi-1D system, similar to a quantum wire, with the energies of the Landau levels given by

$$E(n, \mathbf{k}_{\parallel}) = E_0 + \left(n + \frac{1}{2}\right) \hbar \omega_c + \frac{\hbar^2 \mathbf{k}_{\parallel}^2}{2m_{\text{eff}}}, \quad (2.5)$$

where

n is the Landau quantum number

E_0 is the state energy in the absence of the magnetic field

$\mathbf{k}_{\parallel}$ is the wave vector parallel to the applied magnetic field

Equation 2.5 shows that Landau levels are equally spaced in energy, and the spacing varies linearly with field strength, since $\omega_c \propto B$. The radius of the lowest Landau level is given by $l_b = [\hbar c / (eB)]^{1/2}$ and is closely related to the density of states for each level, which resembles that of the 1D systems discussed in the next section. Transitions between Landau levels are allowed for a change of $\Delta n_{\parallel} = \pm 1$, which allows the levels to be identified as long as their spacing is greater than their broadening. The splitting of the Landau levels arises as a result of the

contrast to Zeeman splitting, which occurs as a result of the quantization of momentum intrinsic to the carrier/s, and its interaction with the applied mag

2.2.6.4 Other Magneto-Optical Effects

In addition to the three main effects discussed earlier, under certain conditions optical effects can be observed, and we mention two of these here.

The Faraday effect arises when an external magnetic field causes circular polarization with different orientations (i.e., σ^+ and σ^-) to propagate at different speeds, a change in the phase of the rotating electric field. For linearly polarized light, a rotation of the polarization axis.

A closely related effect is the magneto-optical Kerr effect; the only real difference between the Faraday and Kerr effects is that the Kerr effect occurs in reflection, while the Faraday effect occurs in transmission. A corollary of this difference is that in the Kerr effect directionality is imposed, that of the plane of reflection. This means that the geometries that can be established between the incoming light field, the plane of reflection, and the magnetic field direction, each leading to slightly different changes in polarization (Kato et al. 2004, Klingshirn 2007). These effects are often used to study the polarization of spins in semiconductors, which is particularly significant for spintronic devices, as detailed in Chapter 13 and in references Kikkawa and Awschalom (2007), Atatüre et al. (2007), and Gupta et al. (1999).

2.2.7 External Electric Fields

The application of an electric field to an isolated atom causes a reduction of ionization energy, which leads to a field-dependent splitting of otherwise degenerate levels. This is known as the Stark effect and is the analog of the Zeeman effect for magnetic fields. In semiconductors, a similar effect is seen for excitons. The origin of this can be seen by examining the potential in the z -direction, with and without an applied electric field, as shown in Figure 2.16. Without an applied electric field, the potential well responsible for exciton formation is shown (Figure 2.16a), with the $n = 1$ wave function also represented. The application of an electric field in the z -direction places a constant gradient on the potential, which distorts the exciton wave function and shifts the exciton energy by an amount that is quadratically dependent on field strength. In practice, however, it is difficult to observe this effect in bulk semiconductors because the magnitude needs to be larger than the speed of the carrier.

Figure 2.16 The potential profile of an atom or exciton (a) is altered by the application of an electric field (b). As a result of the electric field, the energy of the localized state is shifted.

exciton transition. In most semiconductors, this means fields greater than 10^6 Vm^{-1} would be required (Klingshirn 2007). An additional effect of such strong fields, however, is to broaden, or destroy, the excitonic resonance. From Figure 2.16b, it can also be seen that the application of an electric field makes it possible for the electron to tunnel out of the Coulomb barrier, ionizing the exciton. Increasing the field increases the probability of this occurring, and the presence of a significant tunneling probability leads to further broadening of the transition. The width of the exciton transition is also enhanced due to increased collisions with high energy free carriers that can be accelerated by a strong electric field. The Stark effect has however been observed for excitons in bulk semiconductors, using modulation spectroscopies with weak electric fields (Cardona 1969). The full derivation of the Stark shift in semiconductors can be found in Haug and Koch (2004).

Another effect of an applied electric field that is observed in bulk semiconductors, in the absence of excitons, is the Franz–Keldysh effect. Once again, the derivation of this result can be found in Haug and Koch (2004) where it is revealed that the perturbation introduced by the applied field leads to the solutions of the Schrödinger equation being Airy functions rather than plane waves. The effect of this is that rather than a sharp cutoff at the band edge, an exponential low energy tail is present in the absorption spectrum. The extent of this tail increases with increasing field strength, and can be thought of as a photon-assisted, field-induced tunneling of an electron across the band gap. At energies above the band edge, the Airy functions are oscillatory, and lead to oscillations in the absorption spectrum, known as Franz–Keldysh oscillations. Where excitons are dominant, however, the Franz–Keldysh effects are typically modified by large excitonic effects (Haug and Koch 2004).

Two further effects of external electric fields are analogous to the magneto-optical Faraday and Kerr effects. These are known as the electro-optical Kerr effect and the Pockels effect. In these processes, a birefringence* is introduced, which can then rotate the polarization of an electromagnetic field. In the Pockels effect, a crystal without inversion symmetry is required, and the amount of birefringence, and hence amount of rotation, is linearly proportional to the field strength. In contrast, the birefringence introduced in the DC-Kerr effect (which is present in all materials to some extent) varies quadratically with field strength (Klingshirn 2007). Both the Kerr effect and Pockels effect are used extensively as tools in optical spectroscopy to control and modulate pulses of light due to their fast response times.

2.2.8 Strain Fields

In addition to magnetic and electric fields, external mechanical strain fields can be applied that may alter the optical properties of crystalline semiconductors. Physically, a strain field causes a shifting of the nuclei from their equilibrium positions in the crystal lattice. Depending on the direction of the strain, this can lead to a shift and/or splitting of peaks in the absorption spectrum. In order to quantify the effect of a strain field, one typically refers to the deformation potential, which is the amount of energy the band extrema move for a given deformation. For the conduction band and valence band, it is given by $a \times dE_{cv}/dx$, where a is the lattice constant and dE_{cv} is the change in the band-edge energy caused by a length variation of the unit cell by an amount dx .

The deformation energy is typically defined for strain applied in different directions relative to the key directions in the unit cell. Where the strain is applied in a direction such that it alters the symmetry of the unit cell, a splitting of degenerate states occurs in addition to the shift caused by altering the lattice constant. For example, the HH and LH bands are

Figure 2.17 In the unperturbed system (a), the symmetry of the polar bond polarization is zero. When the system is compressed (b), the system loses this net polarization across the unit cell results.

degenerate at $k = 0$ in unstrained bulk materials (Figure 2.9); this degeneracy is lifted by mechanical strain on the lattice. As a result of this splitting, it becomes possible to excite the two in optical experiments where transitions near $k = 0$ dominate, allowing the identification of specific spin states in the conduction band using circularly polarized light, as identified in Figure 2.10.

In compound semiconductors, a strain field also has the potential to induce spontaneous polarization. In an effect similar to the spontaneous polarization present in ionic crystals, a strain field that alters the symmetry of the crystal structure can also alter the electric field across the material. This is represented schematically in Figure 2.17 where the strain field reduces the electric fields that previously canceled each other out. This is the basis for the piezoelectric effect in materials.

The role played by strain fields has become increasingly important as semiconductor nanostructures have been developed. These nanostructures typically rely on interfaces between different semiconductor materials, which often have different lattice constants, thus introducing strain at the interfaces. The following section is devoted to such structures and their optical properties.

2.3 LOW-DIMENSIONAL SEMICONDUCTOR NANOSTRUCTURES

2.3.1 How They Are Realized

Nanostructures are defined as materials that have at least one dimension small enough to be comparable to the de Broglie wavelength of the charge carriers. In semiconductors, they are usually not isolated structures, as is the case with metal nanoparticles (Chapter 3), but rather nanoscale regions of one type of semiconductor surrounded by a different semiconductor. The two different regions will usually have different band gaps, which will lead to a potential barrier for electrons and/or holes. An example of this is shown schematically in one dimension in Figure 2.18. If the band gap of the barrier material is larger than that of the nanostructure, creating a potential well for electrons and/or holes, as well for electrons in the conduction band and holes in the valence band. This confinement of electrons and holes will be confined to this region, and will, when this region is small compared to the de Broglie wavelength, exhibit properties that are well described by the well-known quantum mechanical problem of a “particle in a box.” In practice, however, we are dealing with charge carriers that are confined in one (quantum wells [QWs]), two (quantum wires [QWRs]), or three (quantum dots [QDs]) dimensions, but which behave as free particles in the other two or zero dimensions, respectively, as represented in Figure 2.19. We will come back to the topic of carrier confinement in each dimensionality in due course, but first there are some general considerations to take into account.

The type of confinement depicted in Figure 2.18 is characteristic of a type-II heterostructure, where both electrons and holes are confined in the same region. That is, the

* Birefringence is a property of some materials whereby the velocity of an electromagnetic wave is different for different orthogonal polarization orientations.

Figure 2.18 The different band gaps of materials A and B lead to confinement of electrons in the conduction band and holes in the valence band in this example of a type-I nanostructure.

Figure 2.19 The confinement in one, two, and three dimensions shown leads to the formation of quantum wells, quantum wires, and quantum dots, respectively.

as an example, the conduction band of material B is lower in energy, and the valence band higher, than those of material A. This is not necessarily the case whenever two different materials form a nanostructure, as the alignment of the bands is also important and is not simply dependent on the respective band gaps. The factor that controls the relative band alignment of the two different regions is the electron affinity (i.e., how strongly the nuclei attract excess electrons). Depending on the choice of materials, it then becomes possible to create type-II and type-III nanostructures, as depicted in Figure 2.20, where both valence band and conduction band are greater in B than A, confining electrons and holes to opposite sides of the barrier. Nanostructures that will confine just electrons or just holes can then be designed. The difference in type-III nanostructures is that the valence band of material B is higher in energy than the conduction band of material A. This leads to electrons from the valence band of material B relaxing into the conduction band of A without excitation, thereby confining intrinsic electrons to material A and holes to material B. For the remainder of this chapter, however, we focus on type-I nanostructures and confinement of both electrons and holes, as they are much more common.

Making such nanostructures is not as straightforward as simply placing the different materials together, and has only become possible over the past two decades due to great advances in epitaxial growth technologies, particularly molecular beam epitaxy (MBE) (Tsao 1993)

Figure 2.20 Type-II and type-III confinements lead to the localization of electrons in different regions of the nanostructure. Type-III is a special case of type-II, where the valence band of material B is lower than the valence band of material A.

and metal-organic chemical vapor deposition (MOCVD) (Herman et al. 2000). To tune the band gap of compound semiconductors by varying the concentration of constituents is essential to control the potential profiles and ensuing properties of nanostructures.

The biggest challenge in growing semiconductor nanostructures is to ensure that the interfaces between different materials are of high quality. If this is not the case, there will be a high number of defect and interface states that will dominate the electronic properties, and the desired confinement effects will be obscured. One of the most important considerations is to select materials that have similar lattice constants (i.e., the size of the atoms in the two materials is similar), as a large lattice mismatch at the interface leads to dislocations and a high concentration of defects. For thin films, it is, however, possible to grow high-quality interfaces even when the lattice constants are not perfectly matched. This is achievable because a thin layer of one material can be grown with a lattice constant different from its natural value and instead matches that of the material onto which it is deposited. This thin layer is then under strain, but depending on the extent of the strain and thickness of the layer, it can form an interface free of dislocations, allowing quantum wells (QWs) to be grown.

The strain present at hetero-interfaces in epitaxial layers is in some instances a key factor in growing nanostructures of a given dimensionality. The Stranski–Krastanow growth mode, for example, utilizes the strain (and other material properties) to form quantum dots. As the growth changes from being 2D to 3D, this process generates islands (quantum dots) on top of a thin (1–2 monolayers) 2D “wetting layer” (Stranski and Krastanow 1999; Seeger 2004).

The selective growth of quantum wells, quantum wires, and quantum dots is a complex process, and depends on a wide variety of growth and material parameters. Now it is possible to grow many of these structures controllably and repeatedly. For more details of these growth techniques, we refer the reader to Zhang et al. (2007).

2.3.2 Effects of Reduced Dimensionality

The particle in a box problem is one of the first concepts usually encountered in quantum mechanics, and the solutions of the Schrödinger equation are well known.

Figure 2.21 Confining a particle in an infinite potential well leads to the formation of quantized energy levels (as given by Equation 2.7) and wave functions that go to zero at the barrier (as given by Equation 2.6).

derived. The effects of electron and/or hole confinement are quantitatively similar to this problem, and so we begin with a brief review of the particle in a box problem.

Consider a 1D potential well, as depicted in Figure 2.21, with infinitely high barriers. It is straightforward to show that the solution to the Schrödinger equation $\mathbf{H}\Psi = E\Psi$ gives

$$\Psi = \sin\left(\frac{n\pi x}{L_x}\right), \quad (2.6)$$

$$E = \frac{(\hbar n \pi)^2}{2mL_x^2}, \quad n = 1, 2, 3, \dots, \quad (2.7)$$

where

L_x is the width of the well

m is the mass of the particle

n is an integer that represents the principal quantum number

It is evident from Equation 2.7 that the energy becomes quantized and proportional to the square of the principal quantum number, as indicated in Figure 2.21 where the first four states are shown for the particle in a 1D box.

This corresponds to 1D confinement in QWs, and this problem can be extended relatively easily to 2D and 3D confinements, corresponding to quantum wires and dots, respectively. One of the biggest differences between this case and that of realistic confinement in semiconductor nanostructures is that the barriers are not infinitely high. This means that the wave functions do not go exactly to zero at the barriers, and the probability of the carriers tunneling out of the potential well becomes finite, as does the probability of escaping from the well due to thermal or other excitations.

TABLE 2.2

Number of Carriers and Density of States for Infinite 1D, 2D, and 3D

Dimensionality of Infinite Lattice	Density of States	Integrated Density
1 (quantum wire)	$D(E) = (L/\pi)(2m/\hbar^2)^{1/2} E^{-1/2}$	$N(E) = (2L/\pi)(2m/\hbar^2)^{1/2} E^{1/2}$
2 (quantum well)	$D(E) = (A/2\pi)(2m/\hbar^2)$	$N(E) = (A/2\pi)(2m/\hbar^2) E$
3 (bulk)	$D(E) = (V/2\pi^2)(2m/\hbar^2)^{3/2} E^{1/2}$	$N(E) = (V/3\pi^2)(2m/\hbar^2)^{3/2} E^{3/2}$

L , A , and V are the length, area, and volume of the unit cell, respectively.

Additionally, in real semiconductor nanostructures, along with the confinement in one, two, or three dimensions, there is a semi-infinite system in the other two, one, or zero dimensions, respectively. Hence, to fully define the states of quantum wells and wires, we must also consider the properties of carriers in 2D and 1D systems.

The different dependence on wave vector of the energy of carriers in one, two, or three dimensions can be conceptualized by considering the density of carriers. For a given number of carriers, the density in a 1D system will be proportional to $1/L$, in two dimensions proportional to πk^2 , and in three dimensions to $4\pi k^3/3$. By using these relationships with the fact that the energy $E = \hbar^2 k^2/(2m)$, the density of states $D(E)$ (effectively gives the number of states at a given energy), and the integrated density of states up to a given energy E , $N(E)$, for each dimensionality can be obtained, as shown in Table 2.2 (Poole et al. 2003).

In considering carriers in semiconductor nanostructures, it then becomes necessary to combine the solutions for the particle in a box problem with the solutions for the free electron system of reduced dimensionality. Figure 2.22 shows the number of carriers as a function of energy for each dimension of confinement. It can be seen that the density of states that for quantum dots, the confinement in three dimensions leads to truly discrete energy levels, similar to atomic energy levels. In contrast, confinement in two dimensions shows a step-like density of states, with discrete subbands corresponding to the discretization caused by confinement in one dimension. In one dimension, the flat density of states between steps due to the kinetic energy of the carrier in the other two dimensions (as revealed by $D(E)$ for the 2D system shown in Figure 2.22) is determined by the kinetic energy in a 1D system. QDs, on the other hand, consist of truly discrete peaks, as in a 0D system there is no kinetic energy.

Also shown in Figure 2.22 is a representation of the band structure for each dimension of confinement. In bulk systems, the energy band changes as a function of wave vector in three dimensions and is represented on a dispersion curve, as discussed in Chapter 3. In a quantum well, the confinement in one dimension means that the corresponding energy is ill-defined, and the symmetry of the system is now dependent more on the other two dimensions than the crystal lattice. As a result, the energy for each QW state varies as $k_x^2 + k_y^2$, which corresponds to the kinetic energy of the carriers in two dimensions. Quantum wires are similar; the wave vector is ill-defined in two dimensions, and the energy varies as k_x^2 . Quantum dots then present the other extreme, where the energy is ill-defined in all three dimensions, and there is no kinetic energy, leading to truly discrete energy states.

The band structures depicted in Figure 2.22 assume perfect nanostructures. In practice this is rarely achieved. Not only are the shapes rarely uniform, but there is also strain induced at the interfaces, caused by lattice mismatch. The effect of

Figure 2.22 The number of carriers, density of states, and band structure are shown for bulk semiconductors, quantum wells, quantum wires, and quantum dots.

alter the band structure, meaning that the confinement potential profile no longer appears as a square well, but can be rather more complicated. The specific effect of the strain can be as varied as the size and shape of the QW, QWr, or QD. This is particularly the case for QDs, where many different shapes and orientations are possible. For a more detailed discussion of the effects of strain in QDs, we refer the reader to Yoffe (2001).

The properties of semiconductor nanostructures can then be well described by reference to a combination of the density of states and the dispersion relations or band structures depicted in Figure 2.22. In the following section, we pay particular attention to the optical properties of such nanostructures, and the implications of confinement in one, two, or three dimensions, but first we discuss briefly some of the other effects of reduced dimensionality.

One of the most important considerations for electrons and holes in bulk semiconductors is the extent and nature of their interactions with other quasiparticles. These include carrier-carrier and carrier-phonon interactions and can be responsible for altering many of the properties that make semiconductors so useful. In bulk systems, electrons and holes are free to move in all directions and interact with many other carriers and phonons. When carriers are confined, however, their ability to interact with phonons and other carriers is greatly reduced.

As a result, for example, in QDs at low temperatures and low carrier densities, long coherence times are observed (Meier et al. 2007), a useful and necessary requirement for quantum information processing.

On the other hand, some types of interactions may actually be increased by the confinement. Because the wave vector is ill-defined in one, two, or three dimensions for carriers in nanostructures, the requirement to conserve momentum can be relaxed, which is only required in the unconfined directions. This can be useful, for example, in carrier mixing experiments (Chapter 10) or detrimental due to increased scattering rates and lifetimes.

2.3.3 Implications for Optical Properties

Most of the optical properties in bulk semiconductors are defined by the crystal symmetry, and the ensuing band structure. For semiconductor nanostructures, the significance of the crystal symmetry is reduced, and the shape and symmetry of the nanostructure become important. This is evidenced by the density of states and band structures depicted in Figure 2.22. As a result, the optical properties of nanostructures can vary greatly. In this section, we discuss some of the general effects of confinement on the optical properties and some considerations that only become important at low dimensions.

One of the most fundamental effects of reduced dimensionality is that the density of states is reduced. For QDs, it is clear that the size will be reduced since the wave function is confined in three dimensions. For QWs, however, it is not so clear; the confinement in the well compresses the wave function in this direction, which also leads to a reduction in the radius in the plane of the well which is smaller than the 3D value (this may seem counterintuitive, but stems from the fact that the wave function tries to conserve its spherical shape and is clearly revealed in the complete treatment of the problem (Haug and Pashynin 1998)). The size of the exciton is thereby reduced both perpendicular and parallel to the direction of confinement. An important corollary of the reduced size of electron and hole wave functions in semiconductor nanostructures is that there is greater overlap between them. This in turn means that the oscillator strength is enhanced, and brighter transitions are observed. This makes semiconductor nanostructures especially attractive for light-emitting and absorbing applications. One of the primary advantages of increased overlap and oscillator strength is that the radiative lifetime is reduced for each dimension of confinement.

Where electrons and holes are confined in the same region (i.e., in type-I semiconductors), the exciton binding energy is, in most cases,* also increased over that in the bulk. The increased binding energy results from the increased confinement, which means that the carriers have fewer dimensions in which to move, and so require greater energy to separate. The degree of confinement, the greater the exciton binding energy. As a result, excitons dominate the optical properties, even more so than for bulk semiconductors, and the discussion will revolve around excitons.

Another general effect of confinement is that the reduced dimensionality of the wave vector in the direction of confinement to become ill-defined. This has several implications for the optical properties, particularly for processes where conservation of momentum prevents them from occurring. In particular, absorption and emission in direct gap semiconductor materials become more probable, and indeed the

* This is not always the case, as, for example, when an internal electric field is present in the nanostructure, it may induce electron-hole separation.

effort around the world to develop nanostructures in silicon-based semiconductors, with the aim of being able to integrate optically active devices with current, highly evolved silicon processing technologies. Additionally, the collapse of requirements for momentum conservation also alters the interactions with phonons, which in some cases can lead to enhanced exciton-LO phonon interactions that can result in sidebands.

In addition to these very general effects of confinement, the optical spectrum of nanostructures also changes, and in the coming pages we discuss these effects for each dimension in turn.

For quantum wells, the density of states appears as a step-like function, with each step corresponding to different quantum states in the direction of confinement. One might initially expect that the absorption and emission spectrum would then closely replicate this density of states. In practice, however, much like in bulk semiconductors, emission is predominantly from the band edge due to rapid intraband relaxation. Furthermore, the excitonic transition is seen below the band edge and tends to dominate due to the greater oscillator strength, especially at low temperatures. Figure 2.23 shows the absorption spectrum for GaAs as it changes from a bulk to a 2D system. In bulk GaAs, the spectrum shows the exciton peak, followed by the featureless absorption band at higher energies. As the thickness of the GaAs layer is reduced, discrete peaks are seen corresponding to excitons at the edge of each step in the density of states. The observation of these confinement effects first occurs for layer thicknesses less than twice the Bohr radius ($a_B = 13.6$ nm in GaAs). As the width is reduced further, the $n = 1$ transition shifts to higher energies, as expected from Equation 2.7 (which shows that the

Figure 2.23 The transition from bulk GaAs to a thin GaAs QW shows the effects of the transition from bulk to quantum-confined systems (Gossard 1983, Gobel and Ploog 1990). The Bohr radius in bulk GaAs is 13.6 nm; however, this value changes as the extent of confinement changes. [Reprinted with kind permission from Springer Science+Business Media: Klingshirn, C.F., *Semiconductor Optics*, 2nd edn., Springer-Verlag, Berlin, Germany, 2007.]

energy is inversely proportional to well width), and the spacing between levels increases. This blue shift of the $n = 1$ peak as compared to the band edge of the bulk material is referred to as the confinement potential. The ability to change the confinement potential, hence the transition energy, by changing the size of the nanostructures is similar to QDs and QWs. This provides a means of tuning the transition energy with the material content and is one property that makes nanostructures so important for optoelectronic applications.

Another noticeable feature that appears as the well width decreases is the splitting of the exciton peak into LH and HH excitons. This splitting is due to the removal of degeneracy of the HH and LH valence bands at $\mathbf{k} = 0$, where most optical transitions occur in gap materials. This occurs as the symmetry of the system becomes more dependent on the nanostructure shape and less on the crystal lattice; as a result, the symmetry of the bands decreases with decreasing well width, thereby increasing the splitting of the bands.

It can also be seen in Figure 2.23 that as the well width decreases, the width of the excitonic peaks increases. This is due largely to inhomogeneous broadening, or due to variations in the thickness of the well. In a bulk semiconductor, the linewidth is determined predominantly by the quantity and nature of defects. In high-quality samples, the linewidth can be very small. In quantum wells, where the well width is an important feature, fluctuations in the well width lead to fluctuations in the transition energy, and an increase in inhomogeneous broadening. The strong dependence of the transition energy on the well width can then be explained by considering, for example, the change in energy caused by fluctuations of 0.5 nm in a 5 nm well compared to the same size fluctuations in a 20 nm well. The relative change to the width is much greater in the narrower well, and since the energy is inversely proportional to well width, the absolute variation in energy (due to inhomogeneous broadening) caused by the 0.5 nm fluctuations in well width is four times greater in the narrower well. The spectral width of the exciton is therefore greater for narrower wells. Broadening due to defects in QWs is also increased for narrower wells for the same reason. The inhomogeneous linewidth seen for QWs is, therefore, greatly dependent on the quality of the sample, particularly the interfaces.

Quantum wires are confined in two dimensions, and while their size in the third dimension is ideally uniform, this is rarely the case in practice. Not only are there fluctuations in the wire diameter leading to inhomogeneous broadening, but the cross section of these wires is rarely perfectly circular, or even square. The extent of these fluctuations is highly dependent on the growth technique, and can drastically alter the optical properties (Zhu et al. 2006). Some growth techniques will give wires that are rectangular, with confinement in two dimensions greater than in the other, while other techniques can give triangular cross sections. The specific optical properties of each of these types of quantum wires subsequently vary greatly, and we will not discuss any specific case. Instead, we refer the reader to Zhang et al. (2007) and the references therein, and suggest that the general properties of absorption and emission spectra can be understood from the properties of quantum wells and quantum dots discussed here.

Quantum dots can be made by many different methods, take many different shapes, and possess vastly different properties. Three of the most common types of semiconductor quantum dots are colloidal QDs, which tend to be close to spherical in shape and have no surface potential barriers; interfacial QDs, which are formed by monolayer fluctuations in quantum wells and can be flat with relatively small barriers; and Stranski-Krastanow self-assembled QDs, which are typically pyramidal in shape, with intermediate barrier heights. The specific properties of types of QDs possess the same general properties to a large extent, and with some exceptions, we shall treat them as such here.

The density of states for QDs, as shown in Figure 2.22, suggests that they should exhibit narrow, discrete, atomic-like peaks in the absorption/emission spectrum. In practice, however, one typically looks at an ensemble of dots, and the observed linewidths are significantly increased due to large inhomogeneous broadening. The extent of the broadening can be explained by extension of the discussion on the broadening of QW spectral peaks. Confinement in three dimensions means that the confinement energy is high, and even the slightest change in the size or shape of the dots will lead to a large shift in energy. Much work has been done to refine growth techniques to reduce the distribution of size and shape of QDs, but inhomogeneous broadening remains a significant issue in QD ensembles, with spectral widths anywhere from 10 nm to >100 nm depending on the band gap and growth technique. If, however, one is able to study an isolated QD through, for example, the spatially resolved techniques discussed in Part IV of this book, sharp atomic-like peaks are indeed observed. In Stranski–Krastanow self-assembled QDs, the wetting layer may also contribute to the optical spectrum, acting effectively as a narrow quantum well and appearing slightly blueshifted from the QD peaks.

In quantum dots, the behavior of the system is highly dependent on the symmetry of the dot, and very little on the crystal structure, as evidenced primarily by the large inhomogeneous broadening. The degeneracy of the HH and LH bands at $\mathbf{k} = 0$ is usually lifted as a result of the asymmetry of QDs by an amount greater than that discussed for QWs. In interfacial and self-assembled QDs, which tend to be highly asymmetric, the HH–LH splitting is usually very large, to the extent that the HH band becomes the highest occupied state and dominates the optical properties.

It was discussed earlier that the allowed optical transitions from the HH band can be divided into two circularly polarized transitions, σ^+ and σ^- . It has, however, been shown that in quantum dots that are elongated in one dimension, the two degenerate HH exciton states become mixed, and a splitting between linearly polarized transitions is induced by the anisotropy, as indicated in Figure 2.24 (Tartakovskii et al. 2004). This presents a clear example of how the properties of nanostructures, and quantum dots in particular, are strongly dependent on the size and shape of the nanostructure, more so than the crystal structure.

In addition to transitions between valence band and conduction band states, the discretization and formation of subbands presents another possible type of transition: intraband or inter-subband transitions. Intraband absorption is an important process, particularly for infrared photodetectors, and typically requires free electrons in the conduction band in the initial state, which are generally provided by doping. Excitation to higher confined electron states or

Figure 2.24 In asymmetric quantum dots, angular momentum ceases to be a good quantum number, and mixed HH exciton states are formed, split by the anisotropy energy Δ_{cm} , typically

unconfined states of the barrier is then typically in the range of 10–200 meV and by altering the size of the nanostructure. These processes can occur in QWs, QDs, and QDs, however, the selection rules state that the electric field must be parallel to the direction. For quantum dots, this is not a problem; for quantum wells, however, excitation incident normal to the QW plane, as would usually be convenient, is light incident on the edge of the structure, or bent into the QW by a diffraction on the surface. For this reason, QDs are expected to have an enhanced infrared performance over their QW counterparts. However, in reality, the volume of the dot plays an important role in terms of infrared absorption (Kelsall et al. 2005).

To date, much of the discussion has revolved around the effects of confinement on absorption, and while many of these properties can be applied equally to emission, several processes that only become important for emission. Many of these processes involve the excited carriers are initially generated, both in real space and in energy states, regardless of how they got there. In general, there are three possible locations where carriers can be generated: in the barrier; in high energy, but still confined states within the nanostructure; or in the lowest excited states of the nanostructure.

Where excited carriers are initially in the barrier region of the nanostructure, the probability of capture into the QW, QWr, or QD is dependent on the carriers relaxing into or being captured by the nanostructure. The dynamics of these processes depends on many factors, including the spatial proximity to the quantum structure and carrier mobility, the energy difference between the barrier and confined states, and the availability of phonon modes for the relaxation process, among others, as discussed in Chapter 8. Under these conditions, it is possible that carriers capture electrons before holes or vice versa, which can further complicate the dynamics and also reveal specific electron or hole processes.

When carriers are in higher excited levels in the nanostructure, either because they were generated there or because they were captured into these subbands from the barrier, they will typically relax to the lowest subband before recombining radiatively. This relaxation is typically mediated by phonons, and the dynamics are often revealed by Raman spectroscopies, as discussed in Chapter 6.

In QWs and QWr, inter-subband relaxation typically occurs rapidly, because of energy and momentum conservation requirements of the phonon-induced transitions. This can be explained by again referring to the band structures depicted in Figure 2.22. In quantum wells and wires, the transitions are truly between bands, and so there are a large number of possible energy and momentum combinations that will facilitate the inter-subband relaxation. As a result, there will be many phonon modes available to participate in the relaxation process.

In QDs, however, despite the removal of any momentum conservation requirements, the subbands are actually discrete sub-states, and require phonons of exactly the right energy to facilitate the inter-subband relaxation. If there are no phonon modes available to assist in the relaxation into the lowest subband is drastically slowed, and the subband relaxation efficiency reduced (Yoffe 2001). This phenomenon, which can be beneficial for some applications, depending on the desired application, is known as the “phonon bottleneck,” and while the logic may be infallible, in practice this is not always the case. For example, in self-assembled QDs, the presence of a wetting layer provides access to a high density of nearby electronic states, and the expected lengthening of the lifetime is not observed. Experiments have often reported reduced emission intensity in QDs have often been reported as evidence of a phonon bottleneck, which has in turn been refuted and attributed to other non-radiative relaxation pathways, due to surface effects and/or Auger recombination (Yoffe 2001). Recent experiments identifying the extended lifetime of excited carriers in QDs

the phonon bottleneck (Urayama et al. 2001). In these cases, it seems clear that a phonon bottleneck is present, suggesting that the presence or absence of this effect is highly sample dependent.

2.3.3.1 Electro- and Magneto-Optical Effects

In addition to altering optical effects, carrier confinement also alters some of the electro- and magneto-optical effects discussed previously. The most common of these effects caused by confinement is the quantum-confined Stark effect (QCSE) (Miller et al. 1984), which can be present in nanostructures of all dimensionality, but which for simplicity we will discuss in terms of a 1D quantum well.

As discussed previously (Figure 2.16), the application of an electric field induces a “tilt” on the band structure, and for a quantum well, the effects can be quite significant. There are two main effects associated with the QCSE, both of which can be identified in Figure 2.25. Firstly, the tilt imparted on the quantum well causes one side to be lower, and the other higher. As a result, the lowest level in the conduction band is lowered, and the highest state in the valence band is raised. The transition between these states is thus redshifted and in some cases beyond the transition energy of the corresponding bulk material. The other major effect is that the electron wave function is pushed to one side of the well, while the hole wave function is pushed to the other. The effect of this is to reduce the exciton binding energy, reduce the oscillator strength, and increase the radiative lifetime.

In some semiconductor nanostructures, there can be an internal electric field present without application of an external field. In these cases, the electric field is caused by the difference in spontaneous polarization and/or piezoelectric polarization at the interface of the nanostructure and the barrier. The shape of the potential in these materials is then always sloped, and QCSEs are always present, as is the case, for example, in ZnO-based quantum wells (Davis and Jagadish 2009) and GaN-based quantum wells (Leroux et al. 1998).

One effect of an applied magnetic field, discussed previously, is that by creating Landau levels, the field is able to reduce the dimensionality of a bulk semiconductor to a pseudo 1D system. This effect is enhanced in quantum wells, as it is able to reduce the dimensionality from a 2D system to a 0D system. The effect is much the same as going from quantum wells to quantum dots, with reductions in the interactions with the environment, and increases in coherence times; by varying the field strength the effect of this change can then be studied systematically (Davis et al. 2007).

Figure 2.25 The application of an external electric field leads to the quantum-confined Stark effect, which redshifts the transition energy and pushes the electron and hole to opposite sides of the potential well.

One final effect of an applied magnetic field on nanostructures is in reference to asymmetric QDs discussed earlier (i.e., where the states become mixed and transitions linearly polarized). In this case, the interaction of the carriers with the field is dependent on angular momentum, and the previously mixed states are forced back to pure spin states once again. As a result, the optically allowed transitions once again become circularly polarized and are split by the Zeeman effect discussed earlier.

2.4 SUMMARY

The material covered in this chapter could easily be expanded to fill an entire book or several volumes. Indeed, there are several excellent references devoted entirely to semiconductor physics, nanostructures, and their optical properties. As a result, this chapter is neither fully inclusive nor include all the details of this subject. In this chapter, we discuss semiconductors and their nanostructures, we do however hope to have provided a brief introduction to many and diverse properties of semiconductors as a basis for understanding the data from the experimental techniques described in the subsequent chapters. For those seeking further information, we refer the reader to the many excellent references mentioned throughout this chapter.

REFERENCES

- Allison, J. 1971. *Electronic Engineering Materials and Devices*. London U.K.: Macmillan.
- Ashcroft, N.W. and Mermin, N.D. 1976. *Solid State Physics*. New York: Thomson.
- Atatüre, M., Dreiser, J., Badolato, A., and Imamoglu, A. 2007. Observation of the quantum-confined Stark effect from a single confined spin. *Nature Physics*. 3: 101.
- Bauer, E. 1958. Phänomenologische theorie der kristallabscheidung an ooberflächen. *Zeitschrift für Kristallographie*. 110: 372–394.
- Benedek, G.B. and Fritsch, K. 1966. Brillouin scattering in cubic crystals. *Physica*. 32: 647.
- Borri, P., Langbein, W., Schneider, S. et al. Ultralong dephasing time in InGaAs quantum wells. *Physical Review Letters*. 87: 157401.
- Cardona, M. 1969. *Modulation Spectroscopy*. New York: Academic Press.
- Chelikowsky, J.R. and Cohen, M.L. 1974. Electronic structure of silicon. *Physica*. 40: 5095.
- Cohen, M.L. and Bergstresser, T.K. 1966. Band structures and pseudopotential calculations for fourteen semiconductors of the diamond and zinc-blende structures. *Physica*. 32: 789.
- Davis, J.A., Wathen, J.J., Blanchet, V., and Phillips, R.T. 2007. Time-resolved photoluminescence spectroscopy of excitons in a single quantum well. *Physical Review B*. 75: 045408.
- Davis, J.A. and Jagadish, C. 2009. Ultrafast spectroscopy of ZnO/ZnMgO quantum wells. *Optics and Photonics Review*. 3: 85.
- Fan, H.Y. 1967. Effects of free carriers on the optical properties. *Optical Properties of Semiconductors and Semimetals*. New York: Academic Press.
- Faraday, M. 1839. *Experimental Researches in Electricity*. London, U.K.: Taylor.
- Fehrenbach, G.W., Schafer, W., and Ulbrich, R.G. 1985. Excitonic versus plasma emission in highly excited gallium arsenide. *Journal of Luminescence*. 30: 154.
- Fox, A.M. 2001. *Optical Properties of Solids*. New York: Oxford University Press.
- Gobel, E.O. and Ploog, K. 1990. Fabrication and optical properties of semiconductor quantum wells and superlattices. *Progress in Quantum Electronics*. 14: 289.
- Gossard, A.C. 1983. *Thin Films, Preparation and Properties*. New York: Academic Press.
- Gupta, J.A., Awschalom, D.D., Peng, X., and Alivisatos, A.P. 1999. Spin coherence in semiconductor quantum dots. *Physical Review B*. 59: R10421.

- Haug, H. and Koch, S.W. 2004. *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 4th edn. Singapore: World Scientific Publishing.
- Herman, M.A., Richter, W., and Sitter, H. 2004. *Epitaxy: Physical Foundation and Technical Implementation*. Berlin, Germany: Springer.
- Jagadish, C. and Pearton, S.J. 2006. *Zinc Oxide Bulk, Thin Films and Nanostructures*. Oxford, U.K.: Elsevier.
- Kato, Y.K., Myers, R.C., Gossard, A.C., and Awschalom, D.D. 2004. Observation of the spin hall effect in semiconductors. *Science*. 306: 1910.
- Kelsall, R.W., Hamley, I.W., and Geoghegan, M. 2005. *Nanoscale Science and Technology*. Chichester, U.K.: John Wiley & Sons.
- Kikkawa, J.M. and Awschalom, D.D. 1998. Resonant spin amplification in n-type GaAs. *Physical Review Letters*. 80: 4313.
- Kittel, C. 2005. *Solid State Physics*, 8th edn. Hoboken, NJ: John Wiley & Sons.
- Klingshirn, C.F. 2007. *Semiconductor Optics*, 2nd edn. Berlin, Germany: Springer-Verlag.
- Kovalev, D., Averbouch, B., Volm, D., Meyer, B.K., Amano, H., and Akasaki, I. 1996. Free exciton emission in GaN. *Physical Review B*. 54: 2518.
- Lannoo, M. and Bourgoin, J. 1981. *Point Defects in Semiconductors I, Theoretical Aspects*, Springer Series on Solid-State Science, Vol. 22. Berlin, Germany: Springer.
- Leroux, M., Grandjean, N., Laugt, M. et al. 1998. Quantum confined stark effect due to built-in internal polarization fields in (Al,Ga)N/GaN quantum wells. *Physical Review B*. 58: 13371.
- Madelung, O., Rössler, U., and M. Schulz. 2002. *Group IV Elements, IV-IV and III-V Compounds. Part B—Electronic, Transport, Optical and Other Properties*. Berlin, Germany: Springer-Verlag.
- Meier, T., Thomas, P., and Koch, S.W. 2007. *Coherent Semiconductor Optics*. Berlin, Germany: Springer-Verlag.
- Miller, D.A.B., Chemla, D.S., Damen, T.C. et al. 1984. Band-edge electroabsorption in quantum well structures: The quantum-confined stark effect. *Physical Review Letters*. 53: 2173.
- Pantelides, S. 1978. The electronic structure of impurity and defect states in semiconductors. *Reviews of Modern Physics*. 50: 797.
- Pantelides, S. 1986. *Deep Centre in Semiconductors, A State of the Art Approach*. New York: Gordon and Breach.
- Poole, C.P. Jr. and Owens, F.J. 2003. *Introduction to Nanotechnology*. Hoboken, NJ: John Wiley & Sons.
- Prasankumar, R.P., Upadhyay, P.C., and Taylor, A.J. 2009. Ultrafast carrier dynamics in semiconductor nanowires. *Physics Status Solidi (B)*. 246: 1973.
- Reshchikov, M.A. and H. Morkoc. 2005. Luminescence properties of defects in nanostructures. *Journal of Applied Physics*. 97: 061301.
- Ridley, B.K. 1997. *Electrons and Phonons in Semiconductor Multilayers*. Cambridge, U.K.: Cambridge University Press.
- Savrasov, B. 1996. Linear-response theory and lattice dynamics: A muffin-tin-orbital approach. *Physical Review B*. 54: 16470.
- Seeger, K. 2004. *Semiconductor Physics: An Introduction*, 9th edn. Vienna, Austria: Springer-Verlag.
- Shah, J. 1999. *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*, 2nd edn. Berlin, Germany: Springer-Verlag.
- Smith, D.L., Pan, D.S., and McGill, T.C. 1975. Impact ionization of excitons in Ge and Si. *Physical Review B*. 12: 4360.
- Smith, R.A. 1964. *Semiconductors*. Cambridge, U.K.: Cambridge University Press.
- Stranski, I.N. and Krastanow, L. 1939. Abhandlungen der mathematisch-naturwissenschaftlichen klasse. *Akademie der Wissenschaften und der Literatur in Mainz*. 146: 797.
- Sze, S.M. and Kwok, K. Ng. 2007. *Physics of Semiconductor Devices*. Hoboken, NJ: John Wiley & Sons.
- Tartakovskii, A.I., Cahill, J., Makhonin, M.N. et al. 2004. Dynamics of coherent and incoherent spin polarizations in ensembles of quantum dots. *Physical Review Letters*. 93: 057401.
- Timusk, T. 1976. Far-infrared absorption study of exciton ionization in germanium. *Physical Review B*. 13: 3511.
- Tsao, J.Y. 1993. *Materials Fundamentals of Molecular Beam Epitaxy*. San Diego: Academic Press.
- Toyozawa, Y. 1959. On the dynamical behavior of an exciton. *Progress of Theoretical Physics*. Supplement 12: 111.
- Urayama, J., Norris, T.B., Singh, J., and Bhattacharya, P. 2001. Observation of phonon-assisted relaxation in quantum dot electronic relaxation. *Physical Review Letters*. 86: 4930.
- Uchinokura, K., Sekine, T., and Matsuura, E. 1974. Critical-point analysis of the Raman spectrum of silicon. *Journal of Physics and Chemistry of Solids*. 35: 1035.
- Watts, R.K. 1977. *Point Defects in Crystals*. New York: Wiley-Interscience.
- Yoffe, A.D. 2001. Semiconductor quantum dots and related systems: Electronic structure and related properties of low dimensional systems. *Advances in Physics*. 40: 1.
- Yu, P.Y. and Cardona, M. 2005. *Fundamentals of Semiconductors*, 3rd edn. Berlin, Germany: Springer-Verlag.
- Zhang, L., Fang, X., and Ye, C. 2007. *Controlled Growth of Nanomaterials*. Berlin, Germany: Springer-Verlag.
- Zhu, Q., Pelucchi, E., Dalessi, S., Leifer, K., Dupertuis M.A., and Kapon, E. 2002. Photoluminescence, quantum confinement and carrier capture in self-ordered pyramidal GaInAs quantum dots. *Nano Letters*. 2: 1036.