# 3

# Inorganic semiconductor nanostructures

## 3.1 INTRODUCTION

The information technology revolution of the previous decades has been based firmly on the development and application of inorganic semiconductors. Silicon forms the basis of the vast majority of electronic devices, whilst compound semiconductors such as gallium arsenide (GaAs) are used for many optoelectronic applications. Conventional devices utilise bulk semiconductors in which charge carriers are free to move in all three spatial directions. However, the formation of a nanostructure, in which the dimensions along one or more directions are reduced below $\sim$10 nm, dramatically modifies the carrier properties, which become governed by the laws of quantum mechanics. Transistors in current-generation microprocessors have feature sizes as small as 50 nm, with a reduction to less than 10 nm predicted by 2016. Such devices will soon enter deep into the nanoscale regime, where their behaviour will no longer follow a simple extrapolation from larger devices. In this case the advent of quantum mechanical behaviour may be seen as a problem to be overcome but, as will be seen in this chapter, inorganic semiconductor nanostructures exhibit a wide range of new and unusual properties, which can be employed to fabricate improved and novel electronic and electro-optical devices.

This chapter is organised in the following manner. After a brief review of basic semiconductor physics, the modified electronic properties of semiconductor nanostructures are considered. The different techniques developed to fabricate inorganic semiconductor nanostructures are then discussed, including a comparison of their relative advantages and disadvantages. Novel physical processes which occur in semiconductor nanostructures are considered and the experimental techniques used

*Nanoscale Science and Technology* Edited by R. W. Kelsall, I. W. Hamley and M. Geoghegan
© 2005 John Wiley & Sons, Ltd

to probe their structural, electronic and optical properties are described. A number of representative applications are discussed and possible future developments are considered. The treatment in this chapter is relatively non-mathematical and the emphasis is on breadth rather than depth. Suggestions for further reading are given, these will hopefully allow the reader to explore a particular topic in greater detail.

## 3.2 OVERVIEW OF RELEVANT SEMICONDUCTOR PHYSICS

### 3.2.1 What is a semiconductor?

Semiconductors behave as insulators at absolute zero temperature ($T = 0$) but at non-zero temperatures ($T > 0$) exhibit a relatively small electrical conductivity, the size of which increases rapidly with increasing temperature. Furthermore, their electrical conductivity can be increased by adding small amounts of certain impurities (dopants) or by illumination with particular wavelengths of light. These properties contrast strongly with those of good conductors (metals), whose electrical conductivity is many orders of magnitude larger, decreases relatively weakly with increasing temperature and, to a good approximation, is not affected by small levels of impurities or illumination.

The electronic band theory of solids has been described in Chapter 1, and Figure 3.1 summarises the main features of the band structure of a semiconductor. At absolute zero temperature all states in the valence band are occupied by electrons and all states in the conduction band are empty. Under these conditions, electrical conduction cannot occur. As the temperature is increased, electrons are excited from the valence band across the band gap $E_g$ into the conduction band. Electrical conduction is now possible via the small number of electrons in the conduction band and the large number of electrons which remain in the valence band, but whose motion is limited because there are only a small number of vacancies. Although electrical conduction in the valence band is due to the movement of the large number of electrons, it is more convenient instead to consider this contribution to the electrical conductivity in terms of the much smaller number of vacancies. These vacancies, which are termed holes, move in the opposite direction to the electrons and hence they behave as carriers of opposite charge sign.

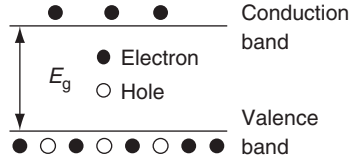


**Figure 3.1** The electronic band structure of a semiconductor. Electrical conduction occurs in the conduction band by a small number of electrons and in the valence band by a small number of vacancies or holes

### 3.2.2 Doping

Electrons and holes created by thermal excitation across the band gap are known as intrinsic carriers. For these carriers the density of electrons in the conduction band $n$ equals the density of holes in the valence band $p$. Although $n$ and $p$ increase rapidly with increasing temperature (resulting in increased electrical conductivity), their absolute values are relatively small. For example, in Si ($E_g = 1.12\,\text{eV}$) $n = p \sim 10^{15}\,\text{cm}^{-3}$ at 300 K, many orders of magnitude less than the density of electrons in the conduction band of a typical conductor ($\sim 10^{22}\,\text{cm}^{-3}$). Consequently, a semiconductor's intrinsic electrical resistance, which is inversely related to its conductivity, is relatively high; potentially a serious problem for many device applications.

Fortunately, the electron or hole densities in a semiconductor can be increased significantly, and in a controllable manner, by the addition of small amounts of certain impurities, a process known as doping. If atoms with one additional valence electron are added to the host semiconductor, such as phosphorus to silicon, the impurity atoms form a series of new states within the forbidden band gap, located slightly below the bottom of the conduction band. At $T = 0$ these states are occupied by the additional electrons but for $T \neq 0$ these electrons may be thermally excited into the conduction band, increasing the free electron density. Because the impurity states are relatively close to the conduction band, this excitation requires relatively little thermal energy, and at moderately high temperatures all the impurity atoms will lose their electrons to the conduction band. Therefore $n$ is increased by an amount approximately equal to the density of impurity atoms, which are known as donors; this process is called n-type doping. Similarly, the introduction of impurity atoms with one fewer valence electron than the host semiconductor, such as boron to silicon, forms a series of levels slightly above the top of the valence band, which are unoccupied at $T = 0$. For $T \neq 0$ electrons may be excited into these states, leaving free holes in the valence band. This is called p-type doping and increases the free hole density by an amount approximately equal to the density of the impurity atoms, which are known as acceptors. Electrons or holes produced by doping are known as extrinsic carriers and for a doped semiconductor $n \neq p$. The concept of doping is summarised schematically in Figure 3.2.

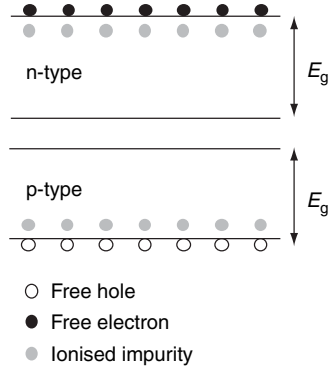


**Figure 3.2** n- and p-type doping. Impurity states are formed just below the bottom of the conduction band and just above the top of the valence band for n- and p-type doping, respectively. The close proximity of these states to the respective band edges means the thermal excitation of extrinsic carriers is much more probable than intrinsic carrier excitation across the band gap

### 3.2.3 The concept of effective mass

Electrons and holes in a semiconductor are not free particles, possessing, in addition to kinetic energy, potential energy due to their electrostatic interaction with the charged ions. Particles with potential energy are considerably more difficult to describe mathematically than free particles, but in a solid this problem can be simplified by using the concept of effective mass. In this model the electrons and holes are treated as free particles by assigning them a modified mass, the effective mass, which combines their potential and kinetic energies into a single kinetic-like energy. The effective mass must be used in all equations describing the dynamical properties of carriers in a solid. The symbol for effective mass is $m^*$ with, in general, a subscript e or h to indicate the effective mass of the electron or hole, respectively. Effective masses are expressed as a multiple of the free electron mass, and holes and electrons typically have different effective masses. As an example, the semiconductor GaAs has an electron effective mass $m_e^* = 0.067m_e$ and a hole effective mass $m_h^* = 0.35m_e0$, where $m_e0$ is the free electron mass.

### 3.2.4 Carrier transport, mobility and electrical conductivity

An externally applied voltage produces an electric field within a semiconductor and this results in an electrostatic force that acts on the charge carriers. This force produces an acceleration and hence motion of the carriers along the field direction; it is this motion which constitutes an electrical current. Carriers will be accelerated by the electric field until they hit an obstacle, at which point their velocity is randomised. Following this collision, the acceleration recommences. This process is shown schematically in Figure 3.3. Although the time between any two collisions is random, there will be a well-defined average scattering time, $\tau$, between collisions and this allows a mean velocity, the carrier drift velocity $v_d$, to be defined. The electrical conductivity is directly proportional to $v_d$, which is a measure of how easily the carriers are able to move through the semiconductor. At low fields $v_d$ is proportional to the size of the electric field, hence a measurement-independent quantity can be obtained by dividing $v_d$ by the field; the resultant quantity is the carrier mobility $\mu$. It can be shown that the mobility is related to the scattering time by $\mu = e\tau/m^*$, where $e$ is the electronic charge.

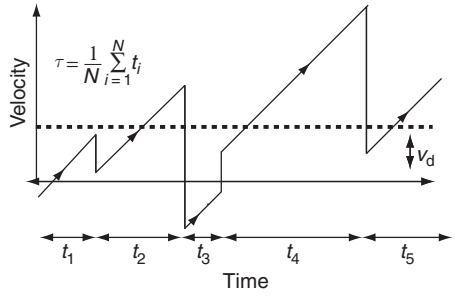


**Figure 3.3** The dynamics of a charge carrier subjected to a constant electric field. A mean scattering time, $\tau$, can be defined by averaging over a large number of events, leading to an average drift velocity $v_d$

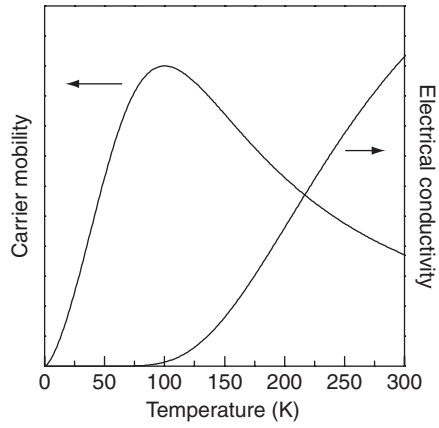


**Figure 3.4** The temperature variation of the electrical mobility and electrical conductivity for a semiconductor

The mean time between collisions is dependent on the nature of the collisions. A carrier travelling through a periodic crystal lattice only experiences a collision if there is a local departure from the crystal periodicity. This can result from the presence of an impurity atom, such as a dopant atom, or by thermal vibrations of the lattice, the quanta of which are termed phonons. Scattering by impurity atoms is important at low temperatures but decreases with increasing temperature. In contrast, scattering by phonons increases with temperature, reflecting the increasing amplitude of the lattice vibrations. The combined effect of these two processes is to give a mobility which, at low temperatures, increases with increasing temperature, followed by a decrease at high temperatures. This behaviour is shown schematically in Figure 3.4. Section 3.6.1 shows that in certain semiconductor nanostructures it is possible to turn off the scattering by impurity atoms, resulting in very high carrier mobilities at low temperatures. The temperature variation of the electrical conductivity is also shown in Figure 3.4. For the case where electrical conduction is dominated by one type of carrier, this is given by Equation 2.9; for the more general case it is necessary to include two terms representing the contributions of both electrons and holes. The electrical conductivity increases monotonically with temperature, the high-temperature decrease in $\mu$ being overwhelmed by the very rapid increase in $n$.

### 3.2.5 Optical properties of semiconductors

The application of semiconductors in electro-optical devices relies on their ability to efficiently emit or detect light. If photons of energy greater than or equal to the band gap are incident on a semiconductor, they may excite an electron from the valence band to the conduction band. In this process the photon is destroyed (absorbed) and an electron and hole are created. In the reverse process an electron in the conduction band may return to the valence band and recombine with a hole; the energy lost by the electron creating a photon. As the energies of the electron and hole will generally be very close to the bottom of the conduction band and the top of the valence band respectively, the emitted photon will have an energy approximately equal to the band gap of the semiconductor.

### 3.2.6 Excitons

The band gap of a semiconductor represents the energy required to create an electron and hole when there is no final interaction between the two carriers. However, the negatively charged electron and positively charged hole may interact to form a hydrogen-atom-like complex in which the two carriers orbit each other, a system known as an exciton. The electrostatic interaction between the electron and hole reduces their energy compared to the non-interacting case, resulting in a series of energy levels just below the conduction band edge. These excitonic states have discrete energies

$$E_n = E_g - E_b/n^2 \qquad (n = 1, 2, 3, \ldots, \infty), \tag{3.1}$$

where for $n \to \infty$ they merge into the continuum states of the conduction band. The binding energy of the exciton $E_b$ is the energy difference between the lowest exciton state ($n = 1$) and the conduction band edge ($n = \infty$). In addition to the states formed below the conduction band edge, there is a modification of the states above the band edge, referred to as the Sommerfeld enhancement. Absorption into excitonic states is possible, and the inset of Figure 3.5 shows how the absorption of a bulk semiconductor is modified by the inclusion of excitonic effects. Although there are an infinite number of exciton states, their absorption strength and separation both decrease rapidly with increasing $n$, and hence experimentally only absorption into the $n = 1$ state is generally observed. Figure 3.5 shows absorption spectra for the semiconductor gallium arsenide (GaAs) at low temperature and room temperature. At low temperature, excitonic effects are clearly visible in the spectrum as an enhanced absorption close to the band gap. However, because the exciton binding energy in GaAs is only 4.2 meV, at room temperature there is sufficient thermal energy ($k_B T = 25$ meV) to ionise the majority of excitons. Hence excitonic effects are absent, or at most extremely weak, in GaAs and

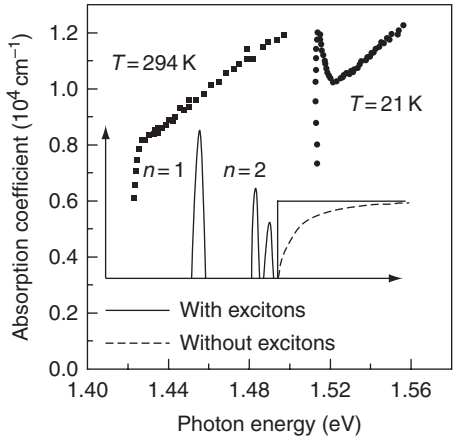


**Figure 3.5** Low-temperature and room temperature absorption spectra of bulk GaAs. The energy shift between the two spectra results from the temperature variation of the band gap. The inset shows the density of states with and without the inclusion of excitonic effects. Reproduced from M. D. Sturge, *Phys. Rev.* **127**, 768 (1962). Copyright 1962 by the American Physical Society

most other bulk semiconductors at room temperature. Section 3.6.6 shows that it is possible to significantly increase the exciton binding energy in a nanostructure, allowing the observation of excitonic effects at higher temperatures.

### 3.2.7 The pn junction

The majority of semiconductor devices are based on the pn junction, which is formed at the interface between two regions, one doped n-type the other p-type. In equilibrium a potential step is formed at the interface which prevents the net movement of electrons from the n-type region into the p-type region, and vice versa for holes. In addition, free carriers are absent from regions either side of the junction, forming a depletion region. A schematic diagram of a pn junction under equilibrium conditions is shown in Figure 3.6(a). If an external voltage of the correct sign is applied (Figure 3.6(b)) the potential step is reduced, allowing electrons and holes to move across the junction, a process known as injection. In a pn junction designed for optical applications, an undoped or intrinsic (i) region may be placed between the n- and p-type regions to form a p-i-n structure. Electrons and holes meet in the intrinsic region, where they recombine to produce photons. A nanostructure may be incorporated within the intrinsic region, providing a convenient and efficient mechanism for injection of electrons and holes into the nanostructure.

The bias condition for current injection is referred to as forward bias. Changing the polarity of the applied voltage produces reverse bias. In this case the potential step is increased and there is negligible current flow. However, electrons and holes created in the intrinsic region by photon absorption may be swept out into the n- and p-type

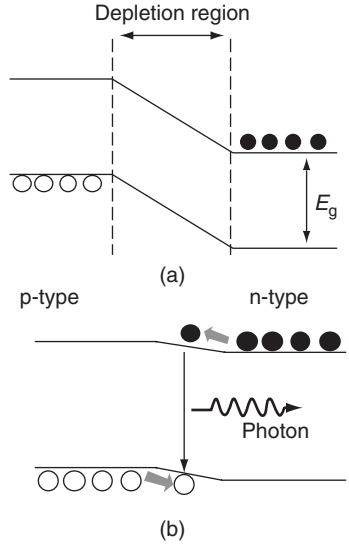


**Figure 3.6** Schematic band diagrams of a pn junction (a) under equilibrium conditions and (b) with an external voltage applied to reduce the potential step, resulting in carrier injection across the junction

regions, respectively, resulting in an electrical current that can be measured by an external circuit. This process allows a semiconductor to act as a photon detector.

### 3.2.8 Phonons

Carriers in a solid may lose or gain energy by emitting or absorbing a phonon. Figure 3.7 shows the phonon dispersion – the frequency-wave vector relationship – calculated for a one-dimensional chain consisting of atoms of two alternating masses. This model provides a good approximation to real semiconductors. Of the two calculated branches, the lower or acoustic branch corresponds to the propagation of sound, and has a frequency which tends to zero for small wave vectors. The upper or optical branch corresponds to phonons which can interact with electromagnetic radiation, and has a frequency which remains non-zero for small wave vectors. For three-dimensional solids each branch consists of three sub-branches, corresponding to the three possible directions of the lattice vibrations with respect to the propagation direction, two transverse and one longitudinal. For GaAs and related semiconductors the strongest carrier–phonon interaction occurs for longitudinal optical (LO) phonons, and it is these phonons that are preferentially emitted as carriers lose energy.

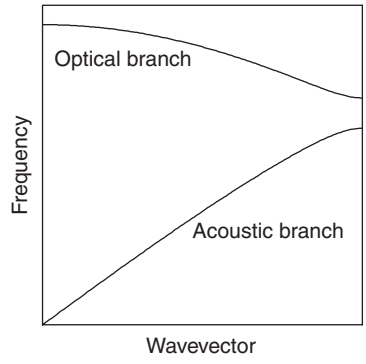


**Figure 3.7** Schematic diagram of the phonon dispersion relationship for a one-dimensional linear chain consisting of atoms of alternating mass

### 3.2.9 Types of semiconductor

The majority of purely electronic devices are based on the elemental semiconductor silicon (Si). However Si has an indirect band gap, with the lowest energy state in the conduction band occurring at a different wavevector to the lowest energy state in the valence band. When an electron and hole in Si recombine, this wavevector difference must be conserved, in addition to energy conservation. A photon is unable to conserve both energy and wavevector, so a second particle, usually a phonon, must be created in addition to the photon. This two-particle, photon plus phonon, recombination process occurs relatively slowly, hence it allows other processes to occur in which a photon is not created.

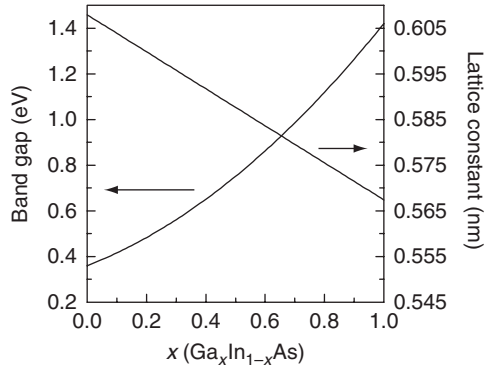


**Figure 3.8** The compositional variation of the lattice constant and band gap of the ternary alloy semiconductor $Ga_xIn_{1-x}As$

For example, the electron may return to the valence band by relaxing via phonon emission through a series of impurity states formed within the band gap, or it may transfer its energy to a second electron which is excited to a higher state in the conduction band. As a result of these non-radiative processes the majority of electrons and holes recombine without the emission of a photon; consequently, the light production efficiency of Si is very poor, making it unsuitable for many electro-optical applications.

Light production efficiency is much greater in direct band gap semiconductors where the recombining electron and hole have the same wavevector, and only a photon is required to satisfy energy conservation. For electro-optical applications, binary semiconductors consisting of elements from columns three and five of the periodic table are typically used, the majority of which have direct band gaps. Examples of III–V semiconductors include gallium arsenide (GaAs), indium phosphide (InP) and gallium nitride (GaN). It is also possible to form semiconductors by combining elements from columns two and six, although these II–VI semiconductors, which include cadmium telluride (CdTe) and zinc selenide (ZnSe), are technologically less important. Furthermore, it is possible to combine two semiconductors to form an alloy semiconductor. For example, InAs and GaAs can be combined to form the ternary semiconductor gallium indium arsenide ($Ga_xIn_{1-x}As$), where the variable $x$ ($0 \leq x \leq 1$) indicates the relative proportions of InAs and GaAs. The properties of an alloy semiconductor are approximately equal to the appropriate weighted average of the constituent semiconductors. Figure 3.8 shows the variation of the band gap and lattice constant of $Ga_xIn_{1-x}As$ as a function of $x$. Both quantities vary smoothly between the values for InAs and GaAs. One important practical application of alloy semiconductors is that a specific band gap can be obtained by a suitable choice of composition.

## 3.3 QUANTUM CONFINEMENT IN SEMICONDUCTOR NANOSTRUCTURES

In a bulk semiconductor, carrier motion is unrestricted along all three spatial directions. However, a nanostructure has one or more of its dimensions reduced to a nanometre

length scale and this produces a quantisation of the carrier energy corresponding to motion along these directions. In this section the nature of this quantisation for nanostructures of different dimensionalities is considered.

### 3.3.1 Quantum confinement in one dimension: quantum wells

Consider initially an isolated, thin semiconductor sheet of thickness $L$. Carrier motion is unrestricted along the two orthogonal directions within the plane of the sheet, but is quantised perpendicular to the plane, forming a one-dimensional quantum well. The resultant quantised energy levels are found by solving the one-dimensional form of the time-independent Schrödinger equation (1.3):

$$-\frac{\hbar^2}{2m^*}\frac{d^2\psi_n(x)}{dx^2} + V(x)\psi_n(x) = E_n\psi_n(x), \tag{3.2}$$

where $V(x)$ is the potential and $\psi_n(x)$ and $E_n$ are the wavefunction and energy of the $n$th confined state. For the present case, $V(x)$ is zero within the semiconductor (which extends from $x = 0$ to $x = L$) and is infinite elsewhere; this is the infinite-depth potential well model. Solving the Schrödinger equation and applying the boundary condition that the wavefunctions must be zero at the edges of the sheet, results in the following energies and wavefunctions:

$$E_n = \frac{h^2n^2}{8m^*L^2} \qquad \psi_n(x) = \sqrt{\frac{2}{L}}\sin\left(\frac{n\pi x}{L}\right) \qquad (n = 1, 2, 3, \ldots, \infty) \tag{3.3}$$

Figure 3.9 shows the energies and wavefunctions for the first three confined states ($n = 1, 2$ and 3) of an infinite-depth potential well.

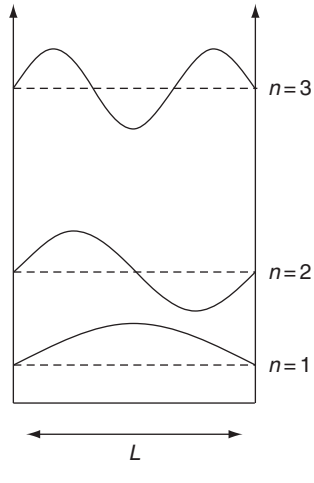


**Figure 3.9** The energies and wavefunctions of the first three confined states of an infinite-depth quantum well

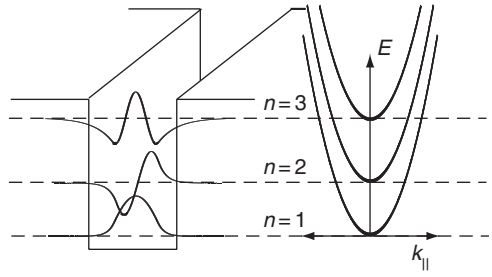


**Figure 3.10** Energies and wavefunctions of the confined states in a finite-depth quantum well (left-hand side). The dispersion resulting from the unrestricted motion of the carriers in the plane of the well is shown to the right

A thin, free-standing semiconductor sheet would possess negligible mechanical strength, and practical quantum wells are formed by sandwiching a thin layer of one semiconductor between two layers of a second, larger band gap semiconductor, which form the barriers. This results in a finite-depth potential well as shown in Figure 3.10. The wavefunctions and energies of the confined states are again determined by the solution of the Schrödinger equation with the appropriate potential, which now remains finite outside of the well. The left-hand side of Figure 3.10 shows energies and wavefunctions for a finite-depth well. In contrast to the infinite-depth well, there are now only a finite number of confined states and the wavefunctions penetrate out of the well and into the barriers. For a finite-depth well it is not possible to obtain analytical forms for the confined energies and the Schrödinger equation must be solved numerically. However, for many applications the energies and wavefunctions of an infinite-depth well can be used as reasonable approximations, particularly for states that lie close to the bottom of the well.

For a semiconductor quantum well both the electron and hole motion normal to the plane will be quantised, resulting in a series of confined energy states in the conduction and valence bands (inset of Figure 3.11). One consequence of this quantum confinement is that the effective band gap of the semiconductor $E_g^{ef}$ is increased from its bulk value by the addition of the electron and hole confinement energies corresponding to the states with $n = 1$:

$$E_g^{ef} = E_g + \frac{h^2}{8m_e^* L^2} + \frac{h^2}{8m_h^* L^2}. \tag{3.4}$$

This effective band gap will determine, for example, the energy of emitted photons, and can be altered by varying the thickness of the well. Figure 3.11 shows an example of this behaviour where the emission spectrum of a structure containing five quantum wells of different widths is shown. Each well emits photons of a different energy; the energy increasing as the width of the well decreases, in agreement with the predictions of Equation (3.4).

Although the carrier energy is quantised for motion normal to the well, within the plane of the well the motion is unrestricted. The total energy of a carrier is given by the sum of the energy due to this unrestricted motion plus the quantisation energy. The in-plane

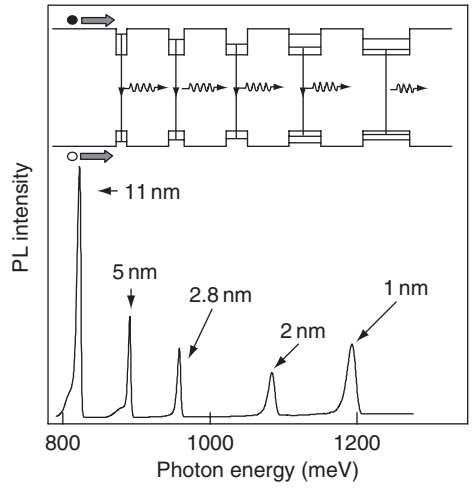


**Figure 3.11** Emission spectrum of a quantum well structure containing five wells of different thicknesses. The wells are $Ga_{0.47}In_{0.53}As$ and the barriers are InP. The inset shows the electronic structure and the nature of the optical transitions

motion is characterised by a wavevector, $k_{\|}$, which corresponds to the combination of the wavevectors for motion along the two mutually orthogonal in-plane directions. If the $z$-axis is taken as being perpendicular to the plane of the well, then the two in-plane directions are $x$ and $y$ and

$$k_{\|} = \sqrt{k_x^2 + k_y^2}. \tag{3.5}$$

From the relationship between momentum, $p = m^* v$, and wave vector, $p = \hbar k$, where $\hbar = h/2\pi$, and the definition of kinetic energy

$$E = \frac{1}{2} m^* v^2 = \frac{p^2}{2m^*}, \tag{3.6}$$

the energy corresponding to in-plane motion can be written in the form

$$E = \frac{\hbar^2 k_{\|}^2}{2m^*}. \tag{3.7}$$

The total energy for a carrier in the $n$th confined state is therefore given by

$$E_{n,k_{\|}} = \frac{h^2 n^2}{8m^* L^2} + \frac{\hbar^2 k_{\|}^2}{2m^*}. \tag{3.8}$$

Since $k_{\|}$ is unrestricted, equation (3.8) gives a continuum of energies for each value of $n$, as shown in the right-hand side of Figure 3.10; these energy bands are known as *subbands*.

### 3.3.2 Quantum confinement in two dimensions: quantum wires

A quantum wire consists of a strip of one semiconductor confined within a second, larger band gap barrier semiconductor. Unrestricted carrier motion is now only possible along the length of the wire and is quantised along the two remaining orthogonal directions. For simple wire shapes (square or rectangular cross sections) it is possible to calculate the quantisation energies for the two directions independently. These two quantisation energies are then added to the energy resulting from the unrestricted motion along the wire. Using the infinite-depth well approximation for the quantised energies, the total energy for a carrier in a quantum wire with $z$ and $y$ dimensions $L_z$ and $L_y$ respectively is

$$E_{n,m,k_x} = \frac{h^2 n^2}{8m^* L_z^2} + \frac{h^2 m^2}{8m^* L_y^2} + \frac{\hbar^2 k_x^2}{2m^*} \qquad (n, m = 1, 2, 3, \ldots). \tag{3.9}$$

The total energy depends on the two quantum numbers $n$ and $m$ and the wave vector for free motion along the wire $k_x$. For each confined state, given by a particular combination of $n$ and $m$, there will be a subband of continuous states resulting from the unrestricted values of $k_x$. In section 3.5 it will be seen that real quantum wires have complex cross sections. This prevents the confined energies from being calculated by separating them into terms corresponding to the two directions perpendicular to the axis of the wire. Instead the confined energies of a quantum wire must be obtained from a numerical solution of the appropriate Schrödinger equation.

### 3.3.3 Quantum confinement in three dimensions: quantum dots

A quantum dot consists of a small region of one semiconductor totally surrounded by a second, larger band gap barrier semiconductor. Carrier motion is now quantised along all three spatial directions and there remains no unrestricted carrier motion. For a simple shape such as a cube or cuboid, confinement for the three spatial directions can be considered separately. In the infinite-depth well approximation, the total energy for a carrier in a cuboid-shaped dot of dimensions $L_z, L_y$ and $L_x$ is a function of three quantum numbers $n$, $m$ and $l$:

$$E_{n,m,l} = \frac{h^2 n^2}{8m^* L_z^2} + \frac{h^2 m^2}{8m^* L_y^2} + \frac{h^2 l^2}{8m^* L_x^2} \qquad (n, m, l = 1, 2, 3, \ldots). \tag{3.10}$$

The energy is now fully quantised and the states are discrete, in a manner similar to those of an atom. The shapes of real quantum dots are more complex than simple cuboids and a calculation of the confined energy levels requires a numerical solution of the relevant Schrödinger equation.

### 3.3.4 Superlattices

It is possible to fabricate a structure consisting of many quantum wells, with each well separated from neighbouring wells by a barrier. If the barriers are sufficiently thick, carriers located in different wells are essentially isolated and the structure behaves identically to a single well, although some properties (e.g., the absorption) will increase linearly with the total number of wells. In this case the structure is known as a multiple quantum well. However, if the thickness of the barriers is reduced, carriers in neighbouring wells may interact via the part of their wavefunctions which penetrates into the barriers. Significant interaction will occur if the wavefunctions (which decay exponentially into the barriers) overlap strongly, requiring relatively thin barriers. For strong interaction, originally identical states in different wells couple together to form a miniband of closely spaced states, a system known as a superlattice. Figure 3.12(a) and (b) show the electronic structures of a multiple quantum well and superlattice, respectively. Figure 3.12(c) demonstrates how the widths of the minibands increase as the barrier width decreases, allowing an increasing interaction between states in different wells. The formation of a superlattice allows carrier motion to occur normal to the plane of the quantum wells, with the resultant structure exhibiting properties intermediate between a bulk semiconductor (3D) and a true quantum well (2D). By extension it is possible to visualise a structure consisting of repeated quantum wires (exhibiting quasi 1D to 2D behaviour) or repeated quantum dots (exhibiting quasi 0D to 1D behaviour).

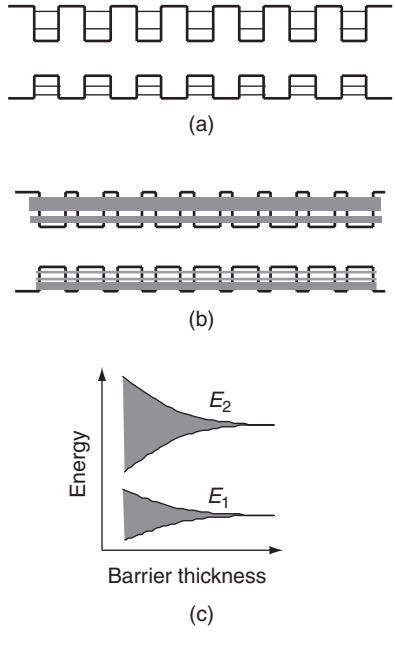


**Figure 3.12** The form of the confined energy states in (a) a multiple quantum well and (b) a superlattice; (c) shows how the discrete states in a multiple quantum well evolve into superlattice minibands as the barrier thickness is reduced

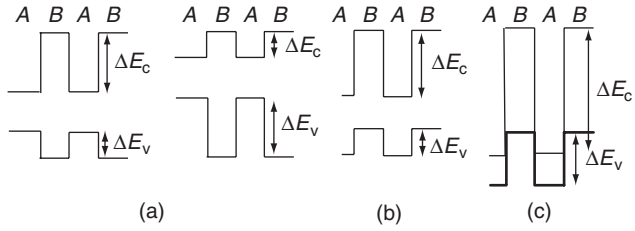


**Figure 3.13** Possible band offsets for a semiconductor quantum well: (a) type I systems with conduction and valence band minimum energies both in semiconductor A; (b) type II system with conduction and valence band minimum energies in A and B, respectively; and (c) an extreme type II system where the minimum energy in the conduction band of semiconductor A lies below the minimum valence band energy of semiconductor B

### 3.3.5 Band offsets

The interface between two different semiconductors is referred to as a heterojunction, and at this point a discontinuity in the energies of both the conduction and valence bands occurs. The numerical sum of these two discontinuities – known as the band offsets: $\Delta E_c$ for the conduction band and $\Delta E_v$ for the valence band – equals the difference between the band gaps of the two semiconductors. However, there are an infinite number of possible combinations of the offsets and Figure 3.13 shows some examples with reference to a quantum well. In these diagrams the electron energy increases when moving vertically up the page, the hole energy increases moving vertically downwards. Hence in both examples shown in Figure 3.13(a) the minimum energy for both electrons and holes occurs in semiconductor *A*. This configuration is known as a type I system and is the one most commonly encountered. In Figure 3.13(b) the minimum energy for electrons occurs in semiconductor *A* but the minimum energy for holes occurs in semiconductor *B*. This configuration is known as a type II system and results in spatially separated electrons and holes. An extreme example of a type II system is shown in Figure 3.13(c). Here the conduction band of semiconductor *A* lies below the valence band of semiconductor *B*, allowing electrons from the latter to transfer to the former. The result is a relatively high density of electrons in the conduction band of semiconductor *A* and the system exhibits semimetallic like properties with a relatively large electrical conductivity.

The magnitudes and signs of the band offsets are important parameters, relevant to a wide range of properties of a nanostructure. However, their accurate theoretical prediction and experimental determination for a given semiconductor combination is relatively difficult.

## 3.4 THE ELECTRONIC DENSITY OF STATES

The concept of the density of states was introduced in Chapter 1, where the density of states for a bulk material was shown to be proportional to $E^{1/2}$. The density of states is strongly affected by reductions in the system dimensionality, because of the corresponding reduction of degrees of freedom in wavevector space. Simple theory, for purely 3D, 2D and 1D systems, gives the density of states to be proportional to $k^{n-2}$, where $n$

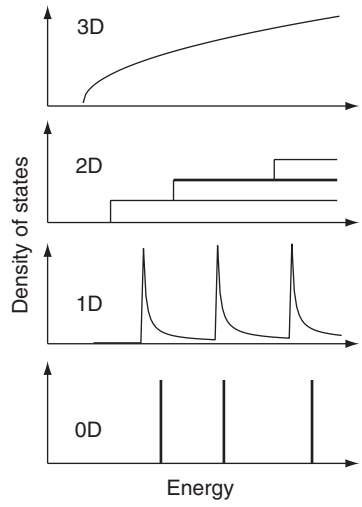


**Figure 3.14** The electronic density of states for a bulk semiconductor (3D), a quantum well (2D), a quantum wire (1D) and a quantum dot (0D)

is the number of dimensions present, and for parabolic bands this translates into a dependence on $E^{(n-2)/2}$. For a quasi low-dimensional system, such as a semiconductor quantum wells or quantum wire, in which a series of confined subbands is formed, the density of states takes the $E^{(n-2)/2}$ dependence in each subband, as shown in Figure 3.14. For a quantum dot, with quantum confinement in all three dimensions, there is no continuous distribution of states and the density of states takes the form of a spectrum of discrete energy values, similar to that found for individual atoms, as shown in the bottom graph on Figure 3.14.

Many of the optical and electronic properties of a nanostructure depend critically on the density of states, hence they exhibit a strong dependence on dimensionality. For example, in electrical transport the density of states determines the number of states available for the motion of the charge carriers, and their scattering time is dependent on the number of available states into which they can be scattered. In addition, the strength of an optical transition is proportional to the density of states at the initial point in the valence band and the final point in the conduction band. This combination is known as the joint density of states and has the same functional form as the individual electron and hole density of states. The energy dependence of the absorption follows the joint density of states and therefore exhibits a very different form for nanostructures of different dimensionality.

## 3.5 FABRICATION TECHNIQUES

In this section the range of techniques developed for the fabrication of semiconductor quantum wells, wires and dots are considered. It begins with established epitaxial techniques capable of producing high-quality quantum wells then moves on to more specialised techniques suitable for fabricating wires and dots. To help compare the

relative advantages and disadvantages of the different techniques, a summary of the requirements for ideal semiconductor nanostructures is first provided.

### 3.5.1 Requirements for an ideal semiconductor nanostructure

The following lists the main requirements for an ideal semiconductor nanostructure. In practice the relative importance of these requirements will be dependent upon on the precise application being considered.

- *Size*: for many applications the majority of electrons and holes should lie in their lowest energy state, implying negligible thermal excitation to higher states. The degree of thermal excitation is determined by the ratio of the energy separation of the confined states and the thermal energy, $k_BT$. At room temperature $k_BT \approx 25\,\text{meV}$ and a rule of thumb is that the level separation should be at least three times this value ($\sim$75 meV). As the spacing between the states is determined by the size of the structure, increasing as the size decreases (Section 3.3), this requirement sets an upper limit on the size of a nanostructure. For electrons in a cubic GaAs quantum dot, a size of less than 15 nm is required. However, below a certain quantum dot size there will be no confined states and this places a lower limit on the dot size.
- *Optical and structural quality*: semiconductors produce light when an electron in the conduction band recombines with a hole in the valence band – a radiative process. However, electron–hole recombination may also occur without the emission of a photon in a non-radiative process. Such processes are enhanced by the presence of certain defects which form states within the band gap. If non-radiative processes become significant then the optical efficiency – the number of photons produced for each injected electron and hole – decreases. For optical applications, nanostructures with low defect numbers are therefore required. Poor structural quality may also degrade the carrier mobility.
- *Uniformity*: devices will typically contain a large number of nanostructures. Ideally each nanostructure should have the same shape, size and composition.
- *Density*: dense arrays of nanostructures are required for many applications.
- *Growth compatibility*: the epitaxial techniques of MBE and MOVPE are used for the mass production of electronic and electro-optical devices. The commercial exploitation of nanostructures will be more likely if they can be fabricated using these techniques.
- *Confinement potential*: the potential wells confining electrons and holes in a nanostructure must be relatively deep. If this is not the case then at high temperatures significant thermal excitation of carriers out of the nanostructure will occur.
- *Electron and/or hole confinement*: for electrical applications it is generally sufficient for either electrons or holes to be trapped or confined within the nanostructure. For electro-optical applications it is necessary for both types of carrier to be confined.
- *p-i-n structures*: the ability to place a nanostructure within the intrinsic region of a p-i-n structure allows the efficient injection or extraction of carriers.

### 3.5.2 The epitaxial growth of quantum wells

The epitaxial techniques of molecular beam epitaxy (MBE) and metallorganic vapour phase epitaxy (MOVPE) are described in Chapter 1. Using these techniques it is possible to deposit semiconductor films as thin as one atomic layer, starting with a suitable substrate material. In this way, quantum well structures can be fabricated with almost perfectly abrupt interfaces. In addition, the doping can be modulated so that only certain layers are doped. MBE and MOVPE are used extensively for the commercial growth of a number of electronic and electro-optical devices, including semiconductor lasers and high-speed transistors. They also form the basis of a number of techniques developed for the fabrication of quantum wires and dots, the most important of which are described in the following sections.

### 3.5.3 Lithography and etching

An obvious fabrication technique for quantum dots or wires is to start with a quantum well, which provides confinement along one direction, and selectively remove material to leave ridges or mesas, forming wires or dots, respectively. Material removal is achieved by the use of electron beam lithography followed by etching. The advantage of this technique is that any desired shape can be produced, although because the electron beam has to be scanned sequentially over the surface, writing large-area patterns is a very slow process. A more serious problem arises from surface damage which results from the etching step. An optically dead surface region is formed, within which the dominant carrier recombination is non-radiative and the importance of which increases as the size of the nanostructure decreases. A suitable choice of semiconductors can minimise this problem (the GaInAs–GaAs system is a common one) but it can never be entirely eliminated. Hence although it is possible to fabricate structures with dimensions as small as $\sim$10 nm, structures with acceptable optical properties are considerably larger ($\gtrsim$50 nm).

### 3.5.4 Cleaved-edge overgrowth

This technique starts with the growth of a quantum well in an MBE reactor. The wafer is then cleaved in situ along a plane normal to the well. The sample is subsequently rotated through 90° and a second quantum well and barrier are deposited on the cleaved surface. This growth sequence is shown in Figure 3.15.

The two quantum wells form a T-shaped structure. At the intersection of the wells, the effective well width is increased slightly, resulting in a reduced potential which traps both electrons and holes. As this potential minimum extends along the intersection of the wells, a quantum wire is formed. The initial growth of multiple wells followed by the overgrowth of the final well allows the formation of a linear array of wires. In addition a second cleave, followed by a further overgrowth step can be used to produce a quantum dot.

The cleaved surface is atomically flat and clean, in contrast to the damaged surfaces formed after etching. As a consequence, cleaved edge overgrowth dots and wires have a high optical quality. With careful optimisation, reasonably deep confinement potentials

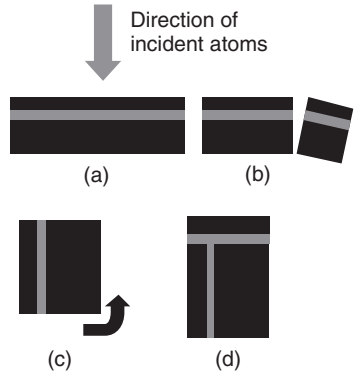


**Figure 3.15** The growth sequence used to produced cleaved edge overgrowth quantum wires: (a) growth of initial quantum well, (b) in situ cleaving, (c) rotation of the structure and (d) growth of a second quantum well on the cleaved surface

are possible; values ~50 meV for the sum of the electron and hole confinement energies have been achieved. However, the separation between the confined states is significantly less than $3k_BT$ and only a single layer of wires can be formed, preventing the fabrication of dense two-dimensional arrays. In addition, the cleaving step is a difficult, non-standard process that requires a significant modification of the MBE reactor.

### 3.5.5 Growth on vicinal substrates

The periodic, crystalline nature of a semiconductor results in flat surfaces only for certain orientations. For the majority of orientations, the surface consists of a periodic series of steps in one or two dimensions. The step periodicity is determined by the orientation of the surface but is typically ~20 nm or less. Although epitaxial growth is generally performed on flat surfaces, growth on stepped or vicinal surfaces provides a technique for the fabrication of quantum wires.

Figure 3.16 shows the main steps in the growth of vicinal quantum wires. Starting with the stepped surface, the semiconductor that will form the wire is deposited by MBE or MOVPE. Under suitable conditions, growth occurs preferentially at the step corners where there is the highest density of unterminated atomic bonds. The growth, consisting of a single atomic layer, proceeds laterally from the corner of the step. When approximately half of the step has been covered, growth is switched to the barrier material which is used to cover the remainder of the step. Growth is then switched back to the initial semiconductor to increase the height of the wire. This growth cycle is repeated until the desired vertical thickness is obtained. Finally the wire is overgrown with a thick layer of the barrier material.

Although very thin wires can be produced using this technique, the growth has to be precisely controlled so that exactly the same fraction of the step is covered during each cycle. In addition, the coverage on different steps may vary and it can be difficult to ensure that the original steps formed on the surface of the substrate are uniform. As a result, quantum wires formed by the vicinal technique tend to exhibit poor uniformity.

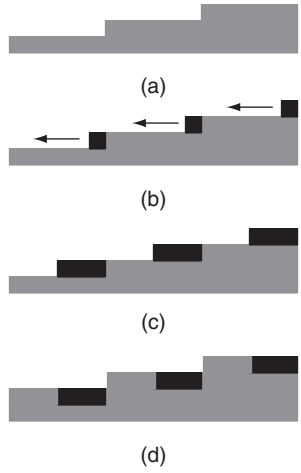


**Figure 3.16** The growth of quantum wires on a vicinal surface: (a) initial stepped surface, (b) initial growth of the wire semiconductor in the corners of the steps, (c) growth proceeds outwards along the steps and (d) second half of step completed with growth of the barrier semiconductor. The figure greatly exaggerates the angle of the surface

### 3.5.6 Strain-induced dots and wires

Applying stress to a semiconductor results in a distortion of the atomic spacing – a strain – which, if of the correct sign, reduces the band gap. If the strain occurs over a small region then a local reduction of the band gap occurs, resulting in the formation of a wire or dot. A strain can be produced by depositing a thin layer of a different material, for example carbon, on the surface of a semiconductor. Because of their different lattice spacings, the materials distort near to their interface in order to fit together. This distortion constitutes a strain, which extends a short distance into the bulk of the semiconductor. If the carbon layer is patterned by lithography and then etched to leave only stripes or mesas, the resulting localised strain fields generate wires or dots in the underlying semiconductor. The remaining isolated regions of carbon are known as stressors. The strain only provides in-plane confinement and so it is necessary to place a quantum well near to the surface to provide confinement along the third direction.

Although this technique involves an etching step, it is only the optically inactive carbon layer that is etched, with the optically active quantum well spatially separated from any surface damaged region. Stressor-induced dots therefore exhibit high optical efficiency. However, the strain field produces only a weak modulation of the band gap, hence the confinement potential is relatively shallow. In addition, fluctuations in the sizes of the stressors results in a distribution of wire and dot sizes. This inhomogeneity is particularly significant in a variant on this technique where the stressors, in the form of small islands, form spontaneously during the deposition of the carbon. Because of the quasi-randomness of this self-assembly processes (Section 3.5.12) there is a relatively large distribution of stressor, and hence underlying quantum dot, size.

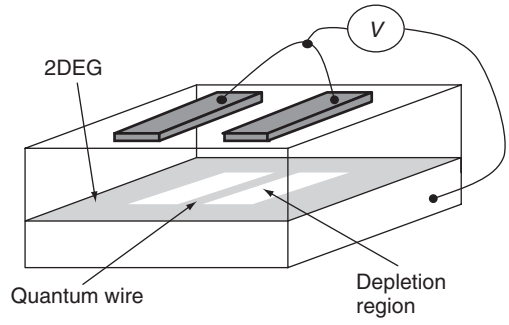


**Figure 3.17** A schematic diagram of a split gate quantum wire. The electrons remaining below the gap between the gates form a quantum wire. A 2DEG remains in the regions away from the gates

### 3.5.7 Electrostatically induced dots and wires

It will be shown in Section 3.6.1 that the technique of modulation doping results in two-dimensional sheets of electrons (or holes) exhibiting very high mobilities at low temperatures. The electron density of a two-dimensional electron sheet or gas (2DEG) can be modified by depositing a metal layer, known as a Schottky gate, on the surface of the semiconductor, and this gate can be patterned using electron beam lithography followed by etching. The application of a negative bias voltage to the gate repels the electrons of the 2DEG from the region immediately below the gate, forming a region depleted of free electrons. If two parallel but spatially separated metal gates (a split gate) are formed on the surface, then biasing these gates depletes the regions below the gates but leaves a long, thin undepleted region directly below the gap between the gates. This is shown schematically in Figure 3.17. This undepleted central region forms a quantum wire, consisting of a one-dimensional strip of free electrons. With increasing gate voltage the depleted regions extend horizontally outwards from the regions directly below the gates, allowing the width of the quantum wire to be varied, a very useful experimental parameter. More complicated structures, in which constrictions are added to the gap between the gates, allow the formation of quantum dots. Further complexity involves the use of gates on both the top and bottom surfaces of the structure, allowing the properties of two parallel and interacting 2DEGs to be varied.

Electrostatically induced nanostructures form clean systems as only the metal is etched, not the semiconductor. Their very high low-temperature carrier mobility makes them excellent structures for studying transport processes in low-dimensional systems. However, the shallow potential minima and small energy-level spacing, a result of their relatively large size, limits their use to cryogenic temperatures. In addition, only electrons or holes are confined in a given structure, hence they are not suitable for optical applications.

### 3.5.8 Quantum well width fluctuations

The width of a quantum well is in general not constant but exhibits spatial fluctuations which form during growth. Potential minima for electrons and holes are formed at

points where the well width lies above its average value, spatially localising the carriers within the plane of the well. With the well providing confinement along the growth direction, the carriers are confined in three dimensions, forming a quantum dot. Although dots formed by well width fluctuations have very good optical properties, their confining potential is very small, as are the spacings between the confined levels. The in-plane size of the dots is difficult to control – the well width fluctuations are essentially random – and the spread of dot sizes is large. Although it is possible to study zero-dimensional physics in these dots, they are not suitable for device applications.

### 3.5.9 Thermally annealed quantum wells

Starting with a GaAs–AlGaAs quantum well, grown using standard epitaxial techniques, a very finely focused laser beam is used to locally heat the structure. This produces a diffusion of Al from the AlGaAs barrier into the GaAs well, resulting in a local increase of the band gap. By scanning the beam along the edges of a square, a potential barrier is produced surrounding the unannealed centre. Electrons and holes within this square are confined by this potential barrier, with confinement along the growth direction provided by the quantum well potential. The carriers are confined in all three directions, forming a quantum dot. Quantum wires may be formed by scanning the laser beam along the edges of a rectangle. Because the size of the focused laser beam is $\sim$1 μm the minimum dot size is fairly large ($\sim$100 nm), resulting in very closely spaced energy levels. In addition, the annealing processes can affect the optical quality of the semiconductor by introducing defects. The technique is relatively slow and hence is not suitable for the production of large arrays of dots or wires. It also requires specialised, non-standard equipment.

### 3.5.10 Semiconductor nanocrystals

Very small semiconductor particles, which act as quantum dots, can be formed in a glass matrix by heating the glass together with a small concentration of a suitable semiconductor. Dots with radii between 1 and about 40 nm are formed; the radius being a function of the temperature and heating time. Although the dots have excellent optical properties, the insulating glass matrix in which they are formed prevents the electrical injection of carriers.

### 3.5.11 Colloidal quantum dots

These II–VI semiconductor quantum dots are formed by injecting organometallic reagents into a hot solvent. Nanoscale crystallites grow in the solution, with sizes in the range 1 to about 10 nm. Subsequent chemical and physical processing may be used to select a subset of the crystallites which display good size uniformity. The dots can be coated with a layer of a wider band gap semiconductor which acts to passivate the surface, reducing non-radiative carrier recombination and hence increasing the optical

efficiency. Organic capping groups are also used to provide additional surface passivation. Colloidal quantum dots exhibit excellent optical properties and can be deposited on to a suitable surface to form close-packed solid structures. Electrical contact to the dots is more difficult, although they can be deposited on silicon wafers that have been prepatterned with metal electrodes. In addition colloidal dots, either incorporated into a polymer film or as a thin film deposited by spin casting, can be formed into layered devices that allow the injection of both electrons and holes.

### 3.5.12 Self-assembly techniques

Using self-assembly techniques, quantum dots or wires form spontaneously under certain epitaxial growth conditions. The resulting structures have high optical quality, and self-assembled quantum dots, in particular, are suitable for a wide range of electro-optical applications. Much of the current interest in the physics of semiconductor nanostructures and the development of device applications is based on self-assembled systems. The underlying fabrication techniques are therefore described in detail in this section.

The first form of self-assembly involves growth on prepatterned substrates. The main steps of this technique for the fabrication of quantum wires can be understood with reference to Figure 3.18. Starting with a flat semiconductor substrate, an array of parallel stripes are formed in a layer of etch resist by optical holography or electron beam irradiation. The structure is then etched in an isotropically acting acid, which attacks different crystal surfaces at different rates, resulting in the formation of an array of parallel V-shaped grooves. The patterned substrate is then cleaned and transferred to a growth reactor.

Quantum wires formed by this technique are typically GaAs with AlGaAs barriers. Initially AlGaAs is deposited, which grows uniformly over the whole structure, sharpening the bottom of the grooves which, after the etching step, have a rounded profile. Next a thin layer of GaAs is deposited which grows preferentially at the bottom of the grooves due to diffusion of Ga atoms from the side walls; the diffusion length for Ga atoms is greater than that of Al atoms. A spatially modulated quantum well is formed, with an increased well thickness at the bottom of the groove. Carriers in this thicker region have a lower energy, so a quantum wire is formed with a crescent-shaped cross section

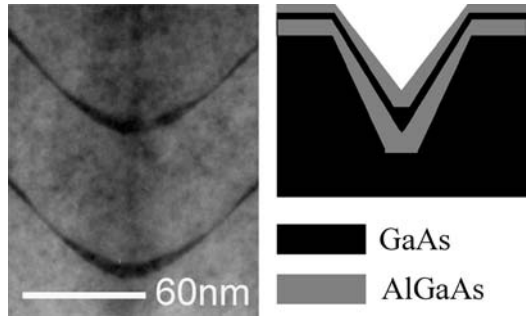


**Figure 3.18** A schematic cross section of a V-groove quantum wire and a transmission electron microscope image of two stacked GaAs–AlGaAs quantum wires. Image courtesy of Dr Gerald Williams, QinetiQ, Malvern

and running along the length of the groove. The shape of the wire's cross section reduces the sharpness of the growth surface at the upper surface of the wire. However, following the wire growth, a second AlGaAs barrier is deposited, which re-establishes the sharpness of the groove. This resharpening permits the growth of further, nominally identical wires. Figure 3.18 also shows a cross-sectional TEM image of a V-groove quantum wire structure containing two wires.

V-groove quantum wires are spatially separated from the original etched surface, hence they exhibit excellent optical quality. By careful optimisation of the growth conditions it is possible to achieve a separation between the wire transitions of 80 meV. However, this energy represents the sum of the electron and hole confinement energies. Because of the larger hole effective mass, the main contribution will be due to the electrons, with the spacing between the confined hole subbands being relatively small ($\sim$10 meV). In addition, the optical spectra of V-groove quantum wires are subject to inhomogeneous broadening resulting from wire width fluctuations; these are probably related to the original groove quality. Furthermore, numerous emission lines are observed, arising from the different spatial regions of the structure. These regions include the quantum wires and quantum wells formed on the sides of the groove (the side-wall wells) and quantum wells formed in the region between the grooves (the top wells). A vertical well is also formed in the AlGaAs at the centre of the groove where the diffusion of Ga to the groove centre produces a Ga rich region. All these regions may capture carriers, reducing the fraction that recombine in the wire (Section 3.7.1). Although the top wells and a fraction of the side wells can be removed by a suitable short wet etch, this requires a further fabrication step followed by a return to the growth reactor to complete a p-i-n structure if required. Alternatively, high-energy ions incident at a shallow angle may be used to degrade the optical efficiency of the top wells and upper section of the side wells.

It is also possible to form quantum dots by a refinement of this technique in which the substrate is initially patterned with two orthogonal arrays of stripes. Following etching, a grid of tetrahedral pits is formed, within which the quantum dots are grown. The resultant dots are lens shaped with height 5 nm, diameter 20 nm and exhibit a typical spacing between optical transitions of $\sim$45 meV.

The second class of self-assembled growth occurs on unpatterned substrates and is driven by strain effects. Hence before this technique can be described, it is necessary to briefly discuss the growth of strained quantum well structures. Quantum wells are generally based on combinations of semiconductors having identical or very similar lattice constants. If a semiconductor is grown on a substrate having a significantly different lattice constant, then initially it will grow in a strained state to allow the atoms at the substrate–epitaxial layer interface to 'fit together'. However, energy is required to strain a material and this energy builds up as the thickness of the epitaxial layer increases. Eventually sufficient energy accumulates to break the atomic bonds of the semiconductor and dislocations – discontinuities of the crystal lattice – form. The thickness of material which can be grown before dislocations start to form is known as the critical thickness. The critical thickness depends on the semiconductor being grown and also the degree of lattice mismatch between this semiconductor and the underlying semiconductor or substrate.

Dislocations provide a very efficient mechanism for non-radiative carrier recombination and a structure which contains dislocations will, in general, exhibit a very poor

optical efficiency. When growing strained semiconductor layers it is therefore very important that the critical thickness is not exceeded.

A good illustration of a strained semiconductor system is provided by $In_xGa_{1-x}As$–GaAs, where the $In_xGa_{1-x}As$ provides the quantum wells and the GaAs provides the substrate and barriers. Because the substrate is much thicker than the epitaxial layers, strain is present only in the $In_xGa_{1-x}As$. As the In composition of the $In_xGa_{1-x}As$ increases, the lattice mismatch between the two semiconductors also increases (Figure 3.8), and hence the $In_xGa_{1-x}As$ becomes increasingly strained. For low In compositions, $x \leq \sim 0.2$, it is possible to grow quantum wells with thicknesses $\sim$10 nm, which is below the critical thickness. However for higher $x$ the critical thickness decreases rapidly, making the growth of quantum wells of reasonable width impossible.

The lattice constants of InAs and GaAs are very different (7%) and therefore the critical thickness for the growth of an InAs layer on GaAs is expected to be very small. Initially InAs grows on GaAs as a highly strained two-dimensional layer. However, continued two-dimensional growth quickly leads to the formation of dislocations for deposition beyond approximately two atomic layers. In contrast, for certain growth conditions it is observed that following the deposition of approximately one atomic layer, the growth mode changes to three-dimensional, leading to the formation of small disconnected islands. These islands, which sit on the original two-dimensional layer, referred to as the wetting layer, act as quantum dots.

The transition from two- to three-dimensional growth is known as the Stranski–Krastanow growth mode and is a consequence of the balance between elastic and surface energies. All surfaces have an associated energy due to their incomplete atomic bonds, and this energy is directly proportional to the area of the surface. The surface area after island formation is greater than the initial flat surface, hence the island configuration has an increased surface energy. However, within the islands, the lattice constant of the semiconductor can start to revert back to its unstrained value as the atoms are free to relax in the in-plane direction, being unconstrained by surrounding material The lattice spacing reverts smoothly back to its unperturbed value with increasing height in the dot and no dislocations are formed. This relaxation results in a reduction of the elastic energy. For systems where the reduction in elastic energy exceeds the increase in surface energy, a transition to three-dimensional growth occurs, as this represents the lowest energy and hence the most favourable state. After formation, the quantum dots are generally overgrown by a suitable barrier semiconductor. Figure 3.19 shows the main steps in the self-assembled growth of quantum dots.

The structure of self-assembled quantum dots depends on the growth conditions, most importantly the temperature of the substrate and the rate at which material is deposited. For InAs dots grown on GaAs a typical base size is between 10–30 nm, height 5–20 nm and density $1 \times 10^9$ to about $1 \times 10^{11}$ cm$^{-2}$. However, values outside this range may be possible by careful control of the growth conditions. Because of their small size the energy separation between the confined levels is relatively large ($\sim$30$-$90 meV for electrons but less for holes due to their larger effective mass) and the confinement potential is relatively deep $\sim$100$-$300 meV, with both electrons and holes being confined. Self-assembled quantum dots contain no dislocations and so exhibit excellent optical properties. Their two-dimensional density is high and can be increased further by growing multiple layers. Self-assembled dots are grown by conventional epitaxial processes and can be readily incorporated in the intrinsic region of a p-i-n structure.

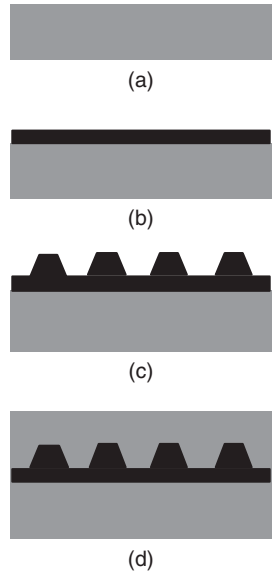


**Figure 3.19** The main steps in the formation of self-assembled quantum dots: (a) initial surface, (b) deposition of the two-dimensional wetting layer, (c) formation of three-dimensional islands once the critical thickness for the 2D to 3D growth transition is reached and (d) overgrowth of the dots

Structural uniformity is reasonable (see below) but because of their small size, relatively small fluctuations in size and/or composition result in large variations in the confined energy levels. Figure 3.20 shows a cross-sectional TEM image of an uncapped InAs self-assembled quantum dot grown on a GaAs substrate. Also shown is an image, recorded along the growth direction, showing the random spatial distribution of the quantum dots. Figure 3.21 shows an AFM image of InAs self-assembled quantum dots.

Although self-assembled quantum dots have been extensively studied, there remains considerable uncertainty about their precise shape. Various shapes have been reported, including pyramids, truncated pyramids, cones and lenses (sections of a sphere). Difficulties associated with structural measurements (Section 3.7.2) may contribute to this reported range of shapes, but it also seems possible that the shape depends on the growth conditions.

Self-assembled quantum dots were first observed for InAs grown on GaAs, where the lattice mismatch is 7%. This and related InGaAs dots also grown on GaAs still form the most mature system. However, self-assembled dots appear to form for most semiconductor combinations with a sufficient degree of lattice mismatch. Examples include InP dots on InGaP, InGaN dots on GaN and Ge dots on Si.

The formation of self-assembled dots is a semi-random process. Dots at different positions start to form at slightly different times, as the amount of deposited material will vary slightly across the growth surface. Defects on the surface may also encourage or inhibit dot formation. As a result, the final shape and size (and possibly composition) will vary slightly from dot to dot, which in turn will result in an an inhomogeneous distribution of the energies of the confined states.

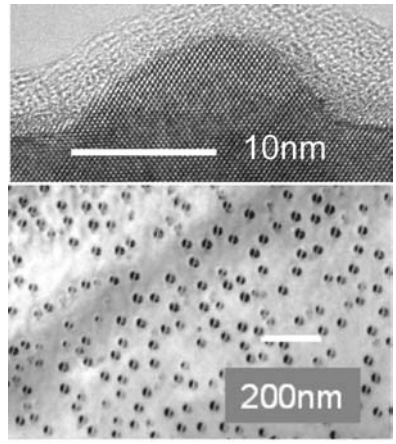


**Figure 3.20** Upper panel: cross-sectional transmission electron microscope image of an uncapped InAs self-assembled quantum dot grown on a GaAs substrate. The speckled material above the dot is glue used to mount the sample. Lower panel: plan view (along the growth axis) TEM image of InAs self-assembled quantum dots grown on a GaAs substrate. In this sample the dots are overgrown with GaAs and the image results from the strain fields produced by the dots. Upper figure reproduced from P. W. Fry, I. E. Itskevich, D. J. Mowbray, M. S. Skolnick, J. J. Finley, J. A. Barker, E. P. O'Reilly, L. R. Wilson, I. A. Larkin, P. A. Maksym, M. Hopkinson, M. Al-Khafaji, J. P. R. David, A. G. Cullis, G. Hill and J. C. Clark, *Phys. Rev. Lett.* **84**, 733 (2000). Copyright 2000 by the American Physical Society

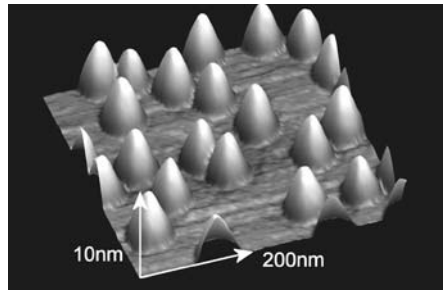


**Figure 3.21** An AFM image of uncapped InAs self-assembled quantum dots grown on a GaAs substrate. Note the different horizontal and vertical scales. Image courtesy of Mark Hopkinson, University of Sheffield

The emission from a single dot has the form of a very sharp line, similar to that observed from an atom. However, many measurements simultaneously probe a large number of dots, referred to as an ensemble. Typically 10 million dots may be probed, each of which will contribute to the measured spectrum. Because each dot emits light at a slightly different energy, the sharp emission from each dot will merge into a broad emission band whose width is related to the degree of dot uniformity. The resulting emission is inhomogeneously broadened. The lower spectrum in Figure 3.22 shows the emission from an ensemble of self-assembled quantum dots, with a line width of 60 meV. The sharp emission, characteristic of a single dot, can be recovered by reducing the number of dots probed. This can be

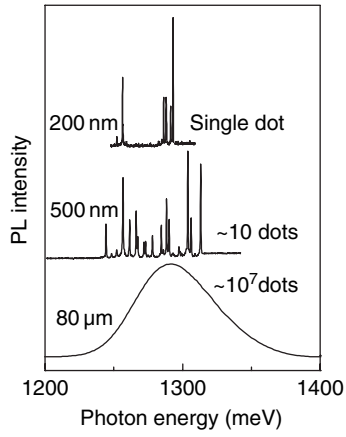


**Figure 3.22** PL spectra of different numbers of self-assembled quantum dots recorded through holes of different sizes formed in an opaque metal mask. For small numbers of dots the sharp emission lines due to individual dots are observed. Data courtesy of Jonathan Finley, University of Sheffield

achieved by depositing an opaque metal mask on the sample surface in which submicron apertures are formed using electron beam lithography, followed by etching. Spectra for two aperture sizes are shown in Figure 3.22. A series of sharp lines, each arising from a different dot are visible, with the number of lines decreasing as the aperture size, and hence dot number below the aperture, decreases. The line width of each emission line is less than 30 μeV, a value limited by the resolution of the measurement system. Higher-resolution measurements have given linewidths as small as 1 μeV. Section 3.6.12 discusses mechanisms that determine this homogeneous line width, which is typically over 1000 times smaller than the ensemble inhomogeneous line width.

The non-uniformity of self-assembled quantum dots, and the resulting inhomogeneous broadening of the optical spectra, is disadvantageous for a number of device applications. For example, the absorption and emission is spread over a broad energy range instead of being concentrated at a single energy. However, there are some applications which make use of this inhomogeneous broadening. Applications of self-assembled quantum dots are discussed in detail in Section 3.8.

Once a layer of self-assembled InAs quantum dots has been deposited and overgrown with GaAs, a flat surface is regained, allowing a second layer of dots to be deposited. This procedure can be repeated, allowing the growth of multiple dot layers. In the first layer, the dot positions are essentially random. However, for multiple layers with relatively thin intermediate GaAs layers a correlation of dot positions is found, with vertical stacks of dots extending across all layers. This vertical alignment occurs as a result of a strain field formed in the GaAs immediately above the dots, produced by the InAs towards the top of the dot reverting back to its unstrained state. The size of this strain field decreases rapidly as the GaAs thickness increases but may be sufficiently large to act as a nucleation site for dots in the subsequent layer. Vertical dot alignment therefore occurs only for very thin GaAs layers ($<$10 nm) where the strain field at the surface remains sufficiently large. For thicker GaAs layers, the strain field is reduced essentially to zero and the dots in the next layer form at random positions. For very closely separated dot layers, dots within a vertical stack are able to interact either by a

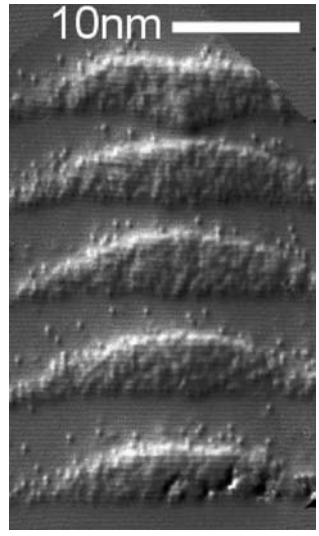


**Figure 3.23** An STM cross-sectional image of the cleaved surface of a structure containing five layers of InAs self-assembled quantum dots with 10 nm thick intermediate GaAs layers. The thin GaAs layers result in a vertical alignment of the quantum dots. Reproduced by permission of the American Institute of Physics from D. M. Bruls, P. M. Koenraad, H. W. M. Salemink, J. H. Wolter, M. Hopkinson and M. S. Skolnick, *Appl. Phys. Lett.* **82**, 3758 (2003)

coupling of the carrier wavefunctions to form a one-dimensional superlattice or by carrier tunnelling between the dots. Figure 3.23 shows an STM image of a structure containing five InAs quantum dot layers, with each layer separated by 10 nm of GaAs.

### 3.5.13 Summary of fabrication techniques

This section has considered the different techniques that have been developed for the fabrication of inorganic semiconductor nanostructures. Each technique has relative advantages and disadvantages which, to some extent, are dependent on the precise application. Although no one technique yet satisfies all of the requirements listed in Section 3.5.1, this has not prevented the use of quantum wells, wires and dots in a number of electronic and electro-optical devices. Example applications are discussed in Section 3.8.

## 3.6 PHYSICAL PROCESSES IN SEMICONDUCTOR NANOSTRUCTURES

### 3.6.1 Modulation doping

As discussed in Section 3.2.4, the low-temperature carrier mobility of a bulk semiconductor is limited by scattering from impurities. This mechanism is particularly efficient

in doped semiconductors where the scattering results from the charged dopant atoms. Consequently, the low-temperature carrier mobility in a bulk doped semiconductor is very low. However, in a semiconductor nanostructure it is possible to spatially separate the dopant atoms and the resulting free carriers. This significantly reduces impurity scattering, resulting in an extremely high carrier mobility at low temperatures. A spatial separation of the dopant atoms and free carriers is achieved through remote or modulation doping, as shown schematically for n-type doping of a quantum well structure in Figure 3.24(a). In this example the donor atoms are placed only in the wider band gap barrier material, the quantum well remains undoped. This is achieved during MBE growth by only opening the shutter in front of the cell containing the dopant atoms during growth of the barriers. In MOVPE the gas carrying the dopant atoms is similarly switched. Following thermal excitation of the electrons from the donors into the conduction band of the wider band gap semiconductor, these free electrons transfer into the lower-energy quantum well states, resulting in a spatial separation of the free electrons and the charged donor atoms. The confined electrons in the quantum well are said to form a two-dimensional electron gas (2DEG); a two-dimensional hole gas can similarly be formed by doping the barriers p-type. The non-zero charge present in both the barriers and the well (the total charge in the structure remains zero but there are equal and opposite charges in the well and barriers) results in an electrostatic bending of the conduction and valence band edges, as indicated in Figure 3.24(b). This band bending allows the formation of a modulation doping induced 2DEG at a single interface (a single heterojunction) between two different semiconductors, as shown in Figure 3.24(c). Here the combined effects of the conduction band offset and the band bending result in the formation of a triangular-shaped potential well that provides confinement of the electrons.

In a modulation-doped structure the barrier region immediately adjacent to the well is generally undoped, forming a spacer layer which further separates the charged dopant atoms and the free carriers. By optimising the width of this spacer layer and the structural uniformity of the interface, and by minimising unintentional background impurities, it is possible to achieve an extremely high carrier mobility at low temperatures. Figure 3.25 compares the temperature variation of the electron mobility of standard bulk GaAs, a very clean bulk specimen of GaAs and a series of GaAs–AlGaAs single heterojunctions. At high temperatures, where mobility is limited by phonon scattering, the mobilities of the different structures are very similar. At low temperatures the mobility in bulk GaAs is increased in the cleaner material, where a lower impurity

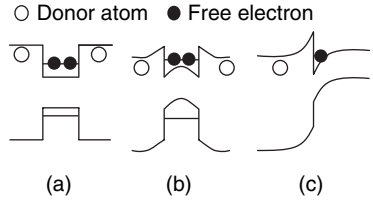


**Figure 3.24** (a) n-type modulation doping of a quantum well showing the transfer of electrons from the barriers to the well; (b) the band edge profile of a modulation-doped quantum well, including the effects of band bending which results from the non-zero space charges in the well and barriers; (c) the production of a 2DEG in a modulation-doped single heterojunction

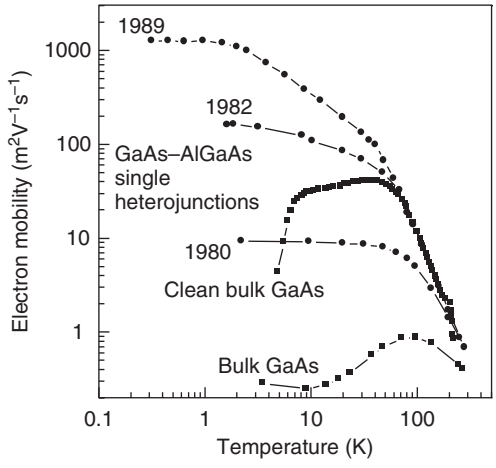


**Figure 3.25** The temperature dependence of the electron mobility of two bulk GaAs samples having different purity, plus a series of n-type modulation-doped GaAs–AlGaAs single heterostructures. The labels for the heterostructures give the year of growth and demonstrate the improvement of the low-temperature mobility with time. Data reproduced by permission of the American Institute of Physics from L. Pfeiffer, K. W. West, H. L. Stormer and K. W. Baldwin, *Appl. Phys. Lett.* **55**, 1888 (1989) and C. R. Stanley, M. C. Holland, A. H. Kean, M. B. Stanaway, R. T. Grimes and J. M. Chamberlin, *Appl. Phys. Lett.* **58**, 478 (1991)

density reduces the charged impurity scattering. However, the absence of doping results in a low carrier density and, as a consequence, a low electrical conductivity. In contrast, modulation doping results in high free carrier densities and a low-temperature mobility more than two orders of magnitude higher than that of the clean bulk GaAs sample. The data for the different heterojunctions presented in Figure 3.25 demonstrates how the low-temperature mobility has increased over time, reflecting optimisation of the structure, the use of purer source materials and improved cleanliness of the MBE growth reactor. By the end of the 1980s a low-temperature mobility of $1200\,m^2V^{-1}s^{-1}$ had been achieved for a GaAs–AlGaAs single heterojunction. Since then progress has been less rapid, suggesting that the limits of material purity and interface structural perfection are being approached. The current state of the art is $3000\,m^2V^{-1}s^{-1}$ for a modulation-doped 30 nm wide GaAs–AlGaAs quantum well.

The ability to produce 2DEGs exhibiting an extremely high mobility has allowed the observation of a range of interesting and novel physical processes, a number of which are discussed in the following sections. In addition, modulation doping can be used to provide the channel of field effect transistors (FETs), where it is particularly useful for high-frequency applications. FETs that incorporate modulation doping are known as high electron mobility transistors (HEMTs) or modulation-doped field effect transistors (MODFETs). Although modulation doping provides only a minor enhancement of the room temperature carrier mobility, it produces free carriers that are confined within a two-dimensional sheet, in contrast to a layer of non-zero thickness produced by conventional doping. This precise positioning of the carriers in the channel results in FETs which exhibit improved linear characteristics and, for reasons that are still unclear, lower noise.

### 3.6.2 The quantum Hall effect

The Hall effect is a standard characterisation technique that can be used to determine both the density and type of majority carrier in a semiconductor. The inset to Figure 3.26 shows a schematic diagram of a Hall measurement, which can be applied to either a bulk semiconductor or a suitable nanostructure. An external voltage source causes a current $I_x$ to flow along the bar, resulting in a current density $J_x$. On application of a magnetic field $B_z$, applied normal to the plane of the sample, a lateral electric field $E_y$ is produced, which appears as a voltage $V_y$ measured across the sample. The quantity $E_y/B_zJ_x$ is known as the Hall coefficient, $R_H$, and for a bulk sample in which transport is dominated by one type of carrier (electrons or holes) $R_H = 1/ne$, where $n$ is the free carrier density. The magnitude and sign of $R_H$ allow determination of the free carrier density and the type of majority carrier, respectively.

Experimentally the electric field along the sample, $E_x$, can also be determined by measuring $V_x$ as shown in Figure 3.26. This allows two resistivities to be defined and measured:

$$\rho_{xx} = \frac{E_x}{J_x}, \quad \rho_{xy} = \frac{E_y}{J_x}. \tag{3.11}$$

For a bulk semiconductor $R_H = E_y/B_zJ_x$, hence $\rho_{xy} = R_HB_z$. $\rho_{xy}$ therefore increases linearly with increasing magnetic field, whilst $\rho_{xx}$ remains constant. However, for a two-dimensional system, a very different behaviour is observed, as shown in Figure 3.26. In this case, although $\rho_{xy}$ increases with increasing field, it does so in a step-like manner. In addition, $\rho_{xx}$ oscillates between zero and non-zero values, with zeros occurring at fields where $\rho_{xy}$ forms a plateau. This surprising behaviour is known as the quantum

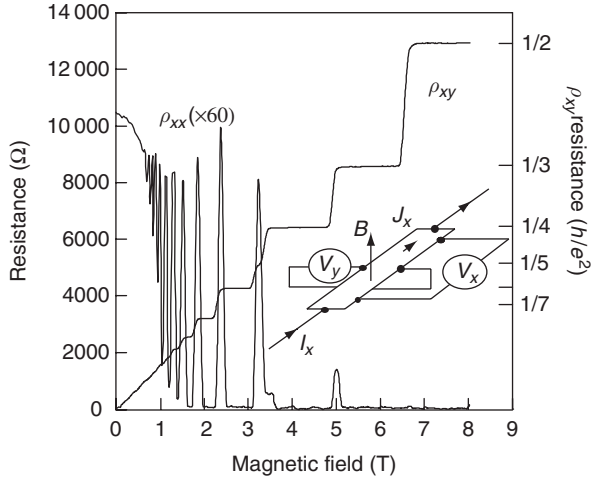


**Figure 3.26** The integer quantum Hall effect as measured for a 2DEG in a GaAs–AlGaAs single heterojunction at a temperature of 50 mK. The inset shows the experimental geometry used for Hall effect measurements. Data reproduced from M. A. Paalanen, D. C. Tsui and A. C. Gossard, *Phys. Rev. B* **25**, 5566 (1982). Copyright 1982 by the American Physical Society

Hall effect and was discovered in 1980 by Klaus von Klitzing, a discovery for which he was awarded the 1985 Nobel physics prize. The quantum Hall effect arises as a result of the form of the density of states in a two-dimensional system in a magnetic field. This corresponds to that of a fully quantised system, with quantisation in one direction resulting from the physical structure of the sample and in the remaining two directions by the magnetic field. A full discussion of the physics underlying the quantum Hall effect, including the importance of disorder, is beyond the scope of this book. Suggestions for further reading are given at the end of this chapter.

An important practical application of the quantum Hall effect arises from the plateau values of $\rho_{xy}$. It can be shown that these are given by

$$\rho_{xy} = \frac{1}{j}\frac{h}{e^2} = \frac{25813}{j}\Omega, \tag{3.12}$$

where $j$ is an integer whose value decreases with increasing magnetic field. $\rho_{xy}$, which is independent of the sample, can be measured to very high accuracy and is now used as the basis for the resistance standard and also to calculate the fine structure constant $\alpha = \mu_0 ce^2/2h$, where the permeability of free space, $\mu_0$, and the speed of light, $c$, are defined quantities.

The parameter $j$ in Equation (3.12) is known as the filling factor and denotes the number of different states occupied by the free carriers. The degeneracy (the maximum number of electrons or holes a given state can contain) of the states formed in a magnetic field increases with increasing field; consequently, for a constant total electron number, the number of states occupied, and hence $j$, decreases with increasing field. The quantum Hall effect discussed so far, and shown in Figure 3.26, occurs for integer values of $j$ and is therefore known as the integer quantum Hall effect. However, in samples with very high carrier mobilities, plateaus in $\rho_{xy}$ and minima in $\rho_{xx}$ are also observed for fractional values of $j$, giving rise to the fractional quantum Hall effect. The discovery and theoretical interpretation of the fractional quantum Hall effect, which results from the free carriers behaving collectively rather than as single particles, led to the award of the 1998 Nobel physics prize to Stormer, Tsui and Laughlin.

### 3.6.3 Resonant tunnelling

Quantum mechanical tunnelling, in which a particle passes through a classically forbidden region, is the mechanism by which α-particles escape from the nucleus during α-decay and electrons escape from a solid in thermionic emission. In both cases the tunnelling probability is a very sensitive function of the energy of the particle and the thickness and height of the potential barrier. Carrier tunnelling can also be observed in nanostructures where a tunnelling barrier is formed by sandwiching a thin layer of a wide band gap semiconductor between layers of a smaller band gap semiconductor. Incorporated into a suitable device, this allows the behaviour of electrons incident on the barrier to be studied. The energy of the electrons is determined by the voltage applied to the device, and the probability of tunnelling through the barrier is reflected by the magnitude of the current that flows. Of greater interest, however, is the case of

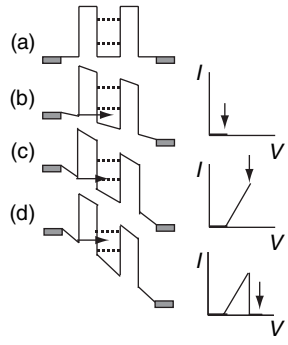


**Figure 3.27** Schematic conduction band diagram of a double-barrier resonant tunnelling structure. The band structure is shown for different applied voltages and with the corresponding current: (a) no applied voltage, (b) voltage below the first resonance, (c) in resonance with the lowest energy state in the quantum well, and (d) above the first resonance

two barriers separated by a thin quantum well, a double-barrier resonant tunnelling structure (DBRTS). A schematic diagram of a DBRTS is shown in Figure 3.27. Quantised energy levels are formed in the quantum well, as described in Section 3.3.1. A DBRTS is generally grown between two doped layers (n-type in Figure 3.27) which provide reservoirs of carriers.

Figure 3.27 shows a DBRTS for various applied voltages. For the sign of voltage shown, electrons travel from left to right. Electrons are first incident on the leftmost barrier, through which they attempt to tunnel into the well, followed by tunnelling out of the well via the second barrier. At low voltages, condition (b), the electron energy following tunnelling into the well is below that of the lowest confined state. Hence there are no available states in the well and it acts as a further barrier. For this condition (b) the two barriers plus the well act as one effective thick barrier, as a consequence the tunnelling probability, and hence the current, is very low. As the voltage is increased to condition (c), the energy of the electrons tunnelling through the first barrier comes into resonance with the lowest state in the well. The effective barrier width is now reduced and it becomes much easier for the electrons to pass through the structure. As a result, the current increases significantly. For a further increase in voltage, condition (d), the resonance condition is lost and the current decreases. Although for condition (d) the energy of the tunnelling electrons coincides with higher energy states in the quantum well subband, these states correspond to non-zero in-plane motion (Section 3.3.1). The tunnelling electrons are moving parallel to the growth direction only, so tunnelling into these subband states is not possible. However, for higher applied voltages, additional resonances may be observed, corresponding to higher confined states. Figure 3.27 also shows the expected current–voltage characteristic of a DBRTS, indicating the relationship between specific points on the characteristic and the different bias conditions of the structure. Figure 3.28 shows experimental results obtained for a DBRTS consisting of a 20 nm GaAs quantum well confined between 8.5 nm AlGaAs barriers. Resonances with five confined quantum well states are observed. Beyond each resonance a DBRTS exhibits a negative differential resistance, a region where the current decreases as the applied voltage is increased. This characteristic has a number of applications, including the generation and mixing of microwave signals.

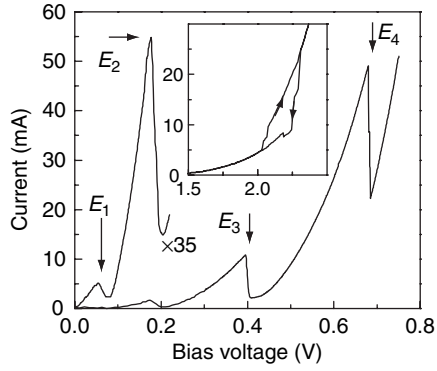


**Figure 3.28** Current–voltage relationship for a double-barrier resonant tunnelling structure consisting of a 20 nm GaAs quantum well with 8.5 nm $Al_{0.4}Ga_{0.6}As$ barriers. Resonances with four confined well states are observed. The inset shows the hysteresis observed for an asymmetrical device with 8.5 and 13 nm $Al_{0.33}Ga_{0.67}As$ barriers and a 7.5 nm $In_{0.11}Ga_{0.89}As$ quantum well. Data courtesy of Philip Buckle and Wendy Tagg, University of Sheffield

DBRTS devices may also exhibit hysteresis in their current–voltage characteristics, particularly when the two barriers have unequal thicknesses. A thinner first barrier allows carriers to tunnel easily into the well whilst a thicker second barrier impedes escape, resulting in charge build-up in the well. This charge build-up modifies the voltage dropped across the initial part of the structure and maintains the resonance condition to higher voltages than would occur in an empty well. This broadened resonance is only observed as the voltage is increased, allowing charge to accumulate in the well. When the voltage is finally taken above the resonance condition, the well empties. If the voltage is now decreased, a narrower resonance is observed as there is now no charge accumulation. For such a structure the current follows a path that is dependent upon the direction in which the voltage is swept; the current–voltage characteristics exhibit hysteresis. The inset to Figure 3.28 shows the characteristics of an asymmetrical DBRTS with 8.5 and 13 nm thick $Al_{0.33}Ga_{0.67}As$ barriers and a 7.5 nm $In_{0.11}Ga_{0.89}As$ quantum well.

### 3.6.4 Charging effects

A charge carrier in a quantum dot is highly spatially localised. If a quantum dot already contains one or more carriers, then significant energy is required to add an additional carrier, as a result of the work that must be done against the repulsive electrostatic force between like charges. This charging energy modifies the energies of the dot states relative to their energies in the uncharged system.

The inset to Figure 3.29 shows the conduction band profile of a structure consisting of a quantum dot placed close to a reservoir of free electrons. Applying a voltage to a metal contact on the surface of the structure allows the energy of the dot to be varied with respect to the reservoir. If a given energy level in the dot is below the energy of the reservoir, then electrons will tunnel from the reservoir into the dot level. Alternatively, if the energy level is above the reservoir, then the level will be unoccupied. Hence by

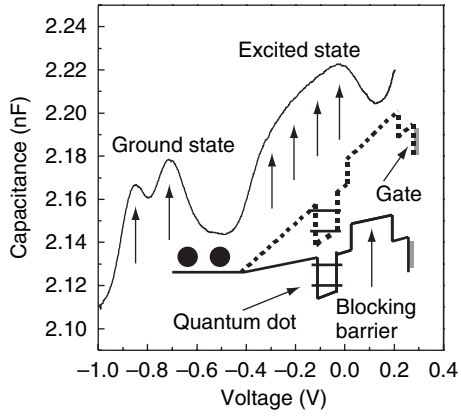


**Figure 3.29** Capacitance–voltage profile of approximately 2 million self-assembled quantum dots showing features as successive electrons are added to the dots. The vertical arrows indicate the positions of the features corresponding to the loading of electrons into the ground state (two features) and the excited state (four unresolved features). The inset shows the conduction band profile for two bias voltages. Data reproduced by permission of EDP Sciences from M. Fricke, A. Lorke, J. P. Kotthaus, G. Medeiros-Ribeiro and P. M. Petroff, *Europhys. Lett.* **36**, 197 (1996)

varying the gate voltage, the dot states can be sequentially filled with electrons. This filling can be monitored by measuring the capacitance of the device, which will exhibit a characteristic feature each time an additional electron is added to the dot.

Figure 3.29 shows the capacitance of a device containing approximately 2 million self-assembled quantum dots. These dots have two confined electron levels: the lowest level (ground state) is able to hold two electrons (degeneracy 2) with the excited level able to hold four electrons (degeneracy 4). In the absence of charging effects, only two features would be observed in the capacitance trace, one at the voltage corresponding to the filling of the ground state, the other when the voltage reaches the value required for electrons to transfer into the excited state. However, once one electron has been loaded into the ground state, charging effects result in an additional energy, hence a higher voltage, being required to add the second electron. This leads to two distinct capacitance features corresponding to the filling of the ground state. Similarly, four distinct features are expected as electrons are loaded into the excited state, although in the present case inhomogeneous broadening prevents them being individually resolved. This charging behaviour is known as Coulomb blockade and is observed experimentally when the charging energy exceeds the thermal energy $k_B T$.

Coulomb blockade effects may also be observed in transport processes where carriers tunnel through a quantum dot. Suitable dots may be formed electrostatically using split gates to define the dot and to provide tunnelling barriers between the dot and two 2DEGs which act as carrier reservoirs. The device is analogous to the resonant tunnelling structures considered in Section 3.6.3 but with the quantum well replaced by a quantum dot. An additional gate electrode allows the energy of the dot to be varied with respect to the carrier reservoirs. The relatively large dot size ($>100$ nm) results in Coulomb charging energies – given by $e^2/2C$ where $C$ is the dot capacitance – much larger than the confinement energies. The Coulomb charging energies therefore

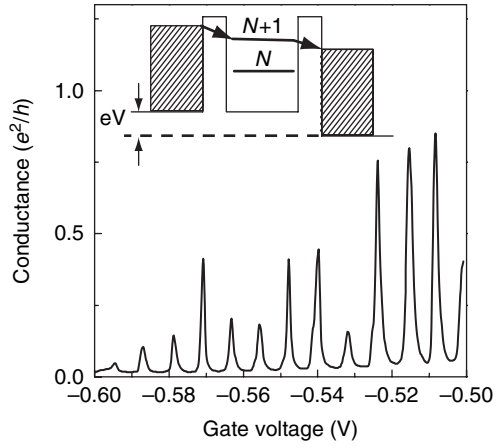


**Figure 3.30** Coulomb blockade oscillations for an electrostatically defined quantum dot measured at a temperature of 10 mK. The inset shows the transport of a single electron through the structure. Data reproduced from L. P. Kouwenhoven, N. C. van der Vaart, A. T. Johnson, W. Kool, C. J. P. M. Harmans, J. G. Williamson, A. A M. Staring and C. T. Foxon, *Zeitschrift für Physik B* **85**, 367 (1991). Copyright 1991 Springer-Verlag

dominate the energetics of the system. The inset to Figure 3.30 shows a schematic diagram of the structure with a small bias voltage applied between the left and right reservoirs. The dot initially contains $N$ electrons, resulting in a dot energy indicated by the lower horizontal line. An additional electron can tunnel into the dot from the left-hand reservoir but this increases the dot energy by an amount equal to the charging energy. This process is therefore only energetically possible if the energy of the dot with $N+1$ electrons lies below the maximum energy of the electrons in the left-hand reservoir. Tunnelling of this additional electron into the right-hand reservoir may subsequently occur, but only if the $N+1$ dot energy lies above the maximum energy of this reservoir, as the electron can only tunnel into an empty state. If these two conditions are satisfied, requiring that the dot energy for $N+1$ electrons lies between the energy maxima of the two reservoirs, a sequential flow of single electrons through the structure occurs. For this condition the system exhibits a non-zero conductance. As the gate voltage varies the dot energy, rigidly shifting the different confined states up or down, the condition for sequential tunnelling will be successively satisfied for different values of $N$, and a series of conductance peaks will be observed. An example is shown in Figure 3.30 for a dot of radius 300 nm. This large dot size results in a large capacitance and a correspondingly small charging energy (0.6 meV for the present example). Hence measurements must be performed at very low temperatures in order to satisfy the condition $e^2/2C \gg k_B T$. Practical applications of Coulomb blockade are described in Section 8.7.

### 3.6.5 Ballistic carrier transport

The carrier transport considered in Section 3.2.4 is controlled by a series of random scattering events. However the high carrier mobilities which can be obtained by the use

of modulation doping correspond to very long distances between successive scattering events, distances that can significantly exceed the dimensions of a nanostructure. Under these conditions a carrier can pass through the structure without experiencing a scattering event, a process known as ballistic transport. Ballistic transport conserves the phase of the charge carriers and leads to a number of novel phenomena, two of which will now be described.

When carriers travel ballistically along a quantum wire there is no dependence of the resultant current on the energy of the carriers. This behaviour results from a cancellation between the energy dependence of their velocity ($v = \sqrt{2E/m^*}$) and the density of states, which in one dimension varies as $1/\sqrt{E}$ (Section 3.4). For each occupied subband a conductance equal to $2e^2/h$ is obtained, a behaviour known as quantised conductance. If the number of occupied subbands is varied then the conductance of the wire will exhibit a step-like behaviour, with each step corresponding to a conductance change of $2e^2/h$. Quantum conductance is readily observable in electrostatically induced quantum wires (Section 3.5.7). The gate voltage determines the width of the wire, which in turn controls the energy spacing between the subbands. For a given carrier density, reducing the subband spacing results in the population of a greater number of subbands, and hence increased conductance. Figure 3.31 shows quantum conductance in a 400 nm long, electrostatically induced quantum wire. The structure of the device is shown in the inset. Such measurements are generally performed at very low temperatures to obtain the very high mobilities required for ballistic transport conditions. In contrast to the highly accurate values observed for $\rho_{xy}$ in the quantum Hall effect, which are independent of the structure and quality of the device, the quantised conductance values of a quantum wire are very sensitive to any potential fluctuations, which may result in scattering events. This sensitivity prevents the use of quantum conductance as a resistance standard.

The inset to Figure 3.32 shows a structure where a quantum wire splits into two wires, which subsequently rejoin after having enclosed an area $A$. Under ballistic

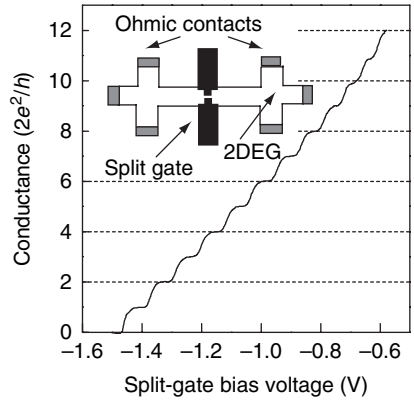


**Figure 3.31** Quantised conductance steps in a 400 nm long electrostatically defined quantum wire measured at a temperature of 17 mK. The wire is produced by a split gate formed on the surface of a modulation-doped GaAs–AlGaAs heterostructure. The inset shows the form of the electrical contacts. Data reproduced by permission of the American Institute of Physics from A. R. Hamilton, J. E. F. Frost, C. G. Smith, M. J. Kelly, E. H. Linfield, C. J. B. Ford, D. A. Ritchie, G. A. C. Jones, M. Pepper, D. G. Hasko and H. Ahmed, *Appl. Phys. Lett.* **60**, 2782 (1992)

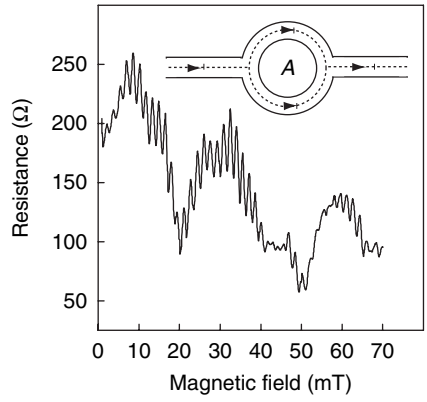


**Figure 3.32** The Aharonov–Bohm effect in a 1.8 μm diameter ring, measured at a temperature of 280 mK. The inset shows the geometry of the structure. Data and figure reproduced from G. Timp, P. M. Mankiewich, P. deVegvar, R. Behringer, J. E. Cunningham, R. E. Howard, H. U. Baranger and J. K. Jain, *Phys. Rev. B* **39**, 6227 (1989). Copyright 1989 by the American Physical Society

transport conditions the wavefunction of an electron incident on the loop will split into two components which, upon recombining at the far side of the loop, will interfere. This process requires that the phase of the electron wavefunction is conserved as it transits the structure. If a magnetic field is now applied normal to the plane of the loop, the wavefunctions acquire or lose an additional phase, depending on the sense in which they traverse the loop. The phase difference between the two paths increases by $2\pi$ when the magnetic flux through the loop, given by the area multiplied by the field ($=BA$) changes by $h/e$. As the magnetic field increases, the system oscillates between conditions of constructive interference (corresponding to a high conductance) and destructive interference (corresponding to low conductance). The change in field, $\Delta B$, between two successive maxima (or minima) is given by the condition $\Delta BA = h/e$, resulting in the conductance of the system oscillating periodically with the field. An example of this behaviour, known as the Aharonov–Bohm effect, is shown in Figure 3.32 for a loop of diameter 1.8 μm, formed from the 2DEG of a GaAs–AlGaAs single heterostructure by electron beam lithography.

### 3.6.6 Interband absorption in semiconductor nanostructures

As discussed in Section 3.2.5, a semiconductor can absorb a photon in a process where an electron is promoted between the valence and conduction bands. The strength of this absorption is proportional to the density of states in both bands – the joint density of states. The joint density of states has a form similar to the individual density of states and is therefore a strong function of the dimensionality of the system. In addition, the absorption will be modified by the quantised energy levels of a nanostructure, resulting in a number of different energy transitions occurring between the confined hole and electron states.

A further modification arises from the influence of excitonic effects. In a nanostructure the electron and hole are prevented from moving apart in one or more directions and, as a result, their average separation is decreased and the exciton binding energy is increased. For an ideal two-dimensional system, the exciton binding energy is increased by a factor of four compared to the bulk value. However, in a real quantum well the electron and hole wavefunctions penetrate into the barriers, and the ideal two-dimensional case is never achieved, although up to an approximately twofold enhancement is possible. A larger binding energy decreases the probability of exciton ionisation at high temperatures and, as a consequence, stronger excitonic effects are observed in a nanostructure at room temperature than in a comparable bulk semiconductor.

Figure 3.33 shows the absorption spectrum of a 40-period GaAs multiple quantum well structure with wells of width 7.6 nm and AlAs barriers. The gross form of the spectrum consists of a series of steps, representing the density of states of a two-dimensional system (Section 3.4), with an excitonic enhancement at the onset of each step. A number of transitions are observed between the $m$th confined hole state and $n$th confined electron state. These transitions are subject to selection rules that result in transitions between identical electron and hole index states ($m = n$) being the most intense. Three orders of such transitions are observed in the spectrum of Figure 3.33. Weaker transitions occur when the index changes by an even number ($|n - m| = 2, 4$, etc.). Transitions where the index changes by an odd number ($|n - m| = 1, 3$, etc.) are forbidden by parity considerations. The spectrum is further complicated by the presence of two different valence bands, known as the heavy and light hole bands, which result in two distinct series of confined valence band states. Transitions are possible from both series to the conduction band.

Figure 3.34 compares the room temperature absorption of bulk GaAs, and a GaAs multiple quantum well structure with 10 nm wide wells. The enhancement of the excitonic strength in the quantum well, due to the increased exciton binding energy, is

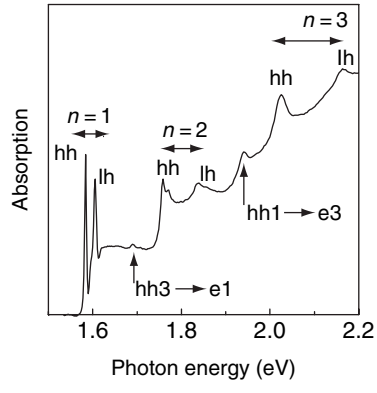


**Figure 3.33** Low-temperature absorption spectrum of a 40-period GaAs–AlAs multiple quantum well structure with 7.6 nm wide wells. The most intense features result from transitions between the $n$th ($n = 1, 2, 3$) confined light hole (lh) and heavy (hh) hole states and identical index electron states. In addition, two weaker transitions are observed between the first and third heavy hole and electron states (hh3 → e1 and hh1 → e3). Data reproduced by permission of Taylor and Francis Ltd from A. M. Fox, *Contemp. Phys.* **37**, 111 (1996)

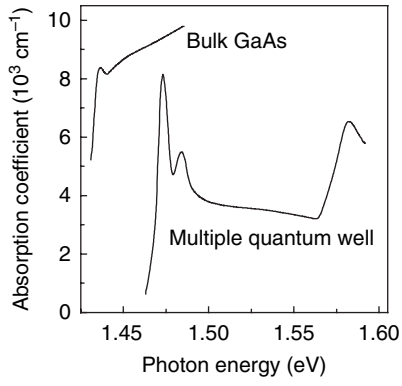


**Figure 3.34** Comparison of the room temperature absorption spectra of bulk GaAs and a 77-period GaAs–$Al_{0.28}Ga_{0.72}As$ multiple quantum well with 10 nm wide wells. Data reproduced by permission of the American Institute of Physics from D. A. B. Miller, D. S. Chemla, D. J. Eilenberger, P. W. Smith, A. C. Gossard and W. T. Tsang, *Appl. Phys. Lett.* **41**, 679 (1982)

clearly visible. The presence of an exciton provides a sharp onset of the band edge absorption, and this has a number of practical applications, for example in optical modulators. Nanostructures allow excitonic effects to be exploited more readily as these effects persist to considerably higher temperatures than in bulk semiconductors.

The absorption spectrum of a quantum wire should be similar to that of a quantum well, but with a further enhancement of the exciton binding energy provided by the additional quantum confinement. There will also be a modification due to the $1/\sqrt{E}$ form of the density of states. However, to date the quality of available quantum wires is not sufficient to observe these effects clearly, with spatial fluctuations of the wire cross section leading to significant inhomogeneous broadening of the absorption spectrum.

The energy levels of a quantum dot are already discrete, hence excitonic effects are less obvious in zero-dimensional systems. In this case they reduce the energies of the optical transitions compared to those that would occur between single-particle, non-interacting states. The absorption spectrum of a quantum dot resembles that of an atom, consisting of a number of discrete absorption lines between which there is zero absorption.

### 3.6.7 Intraband absorption in semiconductor nanostructures

The interband absorption discussed in the previous section occurs between the valence and conduction bands. However, in nanostructures, absorption between confined electron (or hole) levels can also occur, a process known as intraband absorption. The inset to Figure 3.35 shows intraband absorption for the conduction band of a quantum well. Electrons are required in the initial state and these are generally provided by doping. Photon absorption involves the excitation of electrons between the $n = 1$ and higher confined electron states. In some structures excitation between confined well states and the unconfined states of the barrier is also possible. The energy separation between the confined states is typically $\sim$10–200 meV, corresponding to wavelengths $\sim$6–120 μm

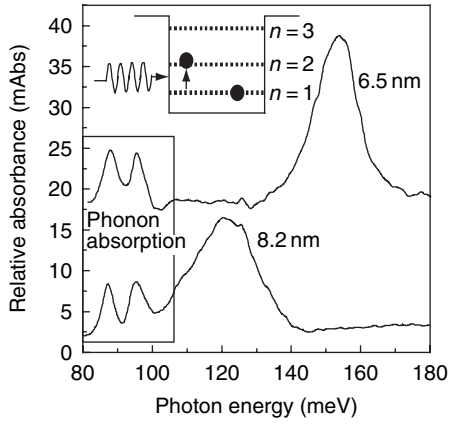


**Figure 3.35** Intraband absorption spectra of two 50-period GaAs multiple quantum well structures with wells of width 6.5 and 8.2 nm. Transitions occur between the $n = 1$ and $n = 2$ confined electron levels. The inset shows a schematic diagram of the absorption process. Data reproduced by permission of the American Institute of Physics from L. C. West and S. J. Eglash, *Appl. Phys. Lett.* **46**, 1156 (1985)

and resulting in absorption in the infra-red region of the electromagnetic spectrum. The energies of the transitions can be varied over this wide range by altering the well width, allowing the absorption to be tuned to a specific wavelength. Intraband absorption spectra are shown in Figure 3.35 for two quantum wells of widths 6.5 and 8.2 nm. With decreasing well width, the separation between the $n = 1$ and $n = 2$ states increases and the absorption shifts to higher energy. There are no excitonic effects associated with intraband absorption, because only one type of carrier is involved; excitons require both an electron and a hole, which are only present in interband processes.

Intraband absorption is also observed between the confined states of quantum wires and dots. For quantum wells, one important selection rule for intraband absorption requires that the incident radiation has an electric field component normal to the plane containing the wells. As a consequence, for light incident along the growth direction, which is the most convenient experimental geometry, intraband absorption does not occur. Intraband absorption in a quantum well therefore requires the use of less convenient geometries, including light incident on the edge of the structure or normally incident light that is bent into the structure by a diffraction grating deposited on the surface. In contrast, for quantum dots this selection rule is modified and intraband absorption is possible for normally incident light. This difference is particularly advantageous for the practical application of intraband absorption in infra-red detectors.

### 3.6.8 Light emission processes in nanostructures

Electrons and holes can be created in a semiconductor either optically (Section 3.2.5), with incident photons of energy greater than the band gap, or by electrical injection in a pn junction (Section 3.2.7). The electrons and holes are typically created with excess energies above their respective band edges. However, the time required to lose this excess energy is

generally much shorter than the electron–hole recombination time, consequently the electron and holes relax to their respective band edges before recombining to emit a photon. Emission therefore occurs at an energy corresponding to the band gap of the structure, with a small distribution due to the thermal energies of the electrons and holes. The influence of rapid carrier relaxation is demonstrated in the emission spectrum of a structure containing quantum wells of five different widths (Figure 3.11). Only emission corresponding to the lowest-energy transition of each well is observed, even though the wider wells contain a number of confined states.

Higher-energy transitions in a nanostructure can be observed in emission if the density of electrons and holes is sufficiently large that the underlying electron and hole states are populated. This can occur under high-excitation conditions where the lower energy states become fully occupied and carriers are prevented from relaxing into these states by the Pauli exclusion principle. Figure 3.36 shows emission spectra of an ensemble of self-assembled quantum dots for different optical excitation powers. At low powers the average number of electrons and holes in each dot is very small, and consequently only the lowest-energy, ground state transition is observed. However with increasing power the ground state, which has a degeneracy of two, is fully occupied and emission from higher-energy, excited states is observed.

As discussed in Section 3.5.12, fluctuations in the size, shape and composition of quantum dots result in the significant inhomogeneous broadening of optical spectra recorded for large numbers of dots. Only by probing a small number of dots can the predicted very sharp emission be observed. Figure 3.37 shows emission spectra obtained from a single self-assembled InAs quantum dot as a function of the incident laser power.

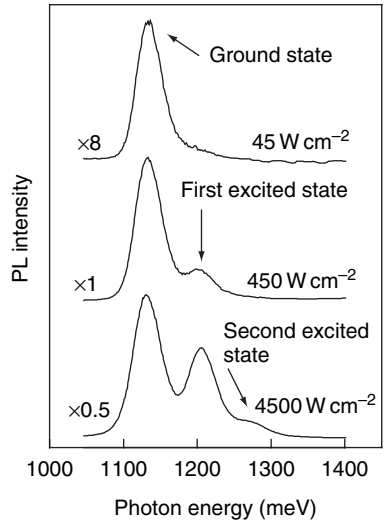


**Figure 3.36** Emission spectra of an ensemble of InAs self-assembled quantum dots for three different laser power densities. At the highest power, emission from three different transitions is observed. The numbers by each spectra indicate the relative intensity scale factors. Data reproduced from M. J. Steer, D. J. Mowbray, W. R. Tribe, M. S. Skolnick, M. D. Sturge, M. Hopkinson, A. G. Cullis, C. R. Whitehouse and R. Murray, *Phys. Rev. B* **54**, 17738 (1996). Copyright 1996 by the American Physical Society

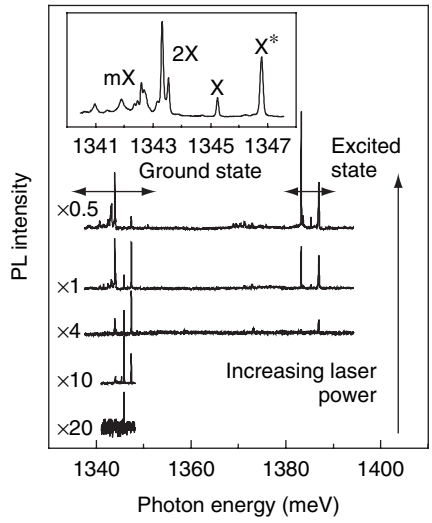


**Figure 3.37** Emission spectra of a single InAs self-assembled quantum dot as a function of laser power. The inset shows emission from the ground state in greater detail. Emission lines are observed corresponding to an exciton recombining in an otherwise empty dot (X), the dot occupied by one additional exciton (the biexciton 2X), the dot occupied by $> 1$ additional excitons (the mX lines) and one additional hole (a charged exciton $X^*$). Data reproduced from J. J. Finley, A. D. Ashmore, A. Lemaître, D. J. Mowbray, M. S. Skolnick, I. E. Itskevich, P. A. Maksym, M. Hopkinson and T. F. Krauss, *Phys. Rev. B*, **63**, 073307 (2001). Copyright 2001 by the American Physical Society

At low powers a single, very sharp line is observed, arising from the recombination of an electron and hole from their respective ground states. At high powers these states are fully occupied and carriers are forced to occupy higher-energy excited states from which emission is then observed.

An added complication in the spectra of Figure 3.37 is that, at high excitation powers, multiple emission lines are observed for the ground state and excited state transitions. This is shown more clearly in the inset to Figure 3.37, which depicts the ground state emission for high laser power. These multiple emission lines result from interactions between the carriers confined within the dot. If the dot is occupied by a single electron and hole, one exciton, then emission will occur at a particular energy. If, however, the dot is occupied by two electrons and two holes, two excitons, then the energy of the first recombining electron and hole will be perturbed by the Coulomb interactions between the two excitons, and the emission will be slightly shifted in energy. For a dot initially occupied by three excitons, the recombination of the first electron and hole will be further perturbed. Lines corresponding to different recombination processes are observed in the same spectrum because the carrier population of the dot fluctuates during the time required to record the emission spectra, typically a few seconds. The recombination of an electron and hole in an otherwise empty dot is known as exciton recombination (X), that of an electron and hole in a dot occupied by an additional electron and hole as a biexciton (2X). The inset to Figure 3.37 also shows a line labelled $X^*$, arising from exciton recombination in the presence of just an additional hole. This

process is possible if the carrier capture probability of the dot is different for electrons and holes.

It is also possible to observe emission when carriers make an intraband transition between the quantised electron or hole states of a nanostructure. However, this emission is generally relatively weak as there are other competing processes, including the emission of one or more phonons. An important application of intraband emission is in a new and increasingly important class of laser, the quantum cascade laser, which is discussed in detail in Section 3.8.2.

### 3.6.9 The phonon bottleneck in quantum dots

In both bulk semiconductors and nanostructures, electrons and holes may lose any excess energy by emitting a series of phonons. Figure 3.38(a) shows a typical case for electrons in a quantum well. The electrons initially have zero energy as they are at the conduction band edge of the barrier, but on transferring into the well they are left with an excess energy. Associated with each confined well state is a continuum of states, resulting from the in-plane motion (Section 3.3.1), and this allows the electron to lose energy by emitting a sequence of phonons, as shown in the figure. For III–V semiconductors the carriers interact most strongly with longitudinal optical (LO) phonons and it is these phonons which are emitted as the carriers lose energy. A typical LO phonon energy is $\sim$30 meV and a typical time to emit a single LO phonon is $\sim$150 fs. Once the carriers reach an energy less than one LO phonon energy from the band edge, further LO phonon emission becomes impossible and the final energy is lost by the emission of low-energy acoustic phonons, a slower process compared to LO phonon emission.

Figure 3.38(a) also applies to bulk semiconductors and quantum wires, both of which have a continuum of states. However, the situation is very different for a quantum dot, where the energy levels are discrete. Here emission by a series of LO phonons is only possible if the spacing between the energy levels equals the LO phonon energy, an unlikely coincidence. Hence carrier relaxation in a quantum dot must occur by an alternative, slower process. Possibilities include the emission of multiple phonons, for example an LO plus an acoustic phonon, or an Auger process where the energy released by one carrier as it relaxes is transferred to a second carrier, which is excited to a higher

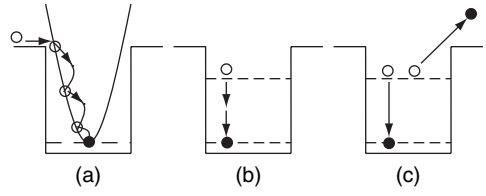


**Figure 3.38** (a) Electron relaxation in a quantum well. The electron is injected into the well with excess energy but the continuum of well states allows this energy to be lost by the emission of a sequence of LO phonons. In a quantum dot, the separation between the discrete, confined states does not generally match the LO phonon energy and the electron relaxes by (b) simultaneously emitting two different phonons or (c) by transferring energy to a second electron that is excited into the continuum states of the barrier

energy state, for example the continuum states associated with the barrier. These processes are shown schematically in Figure 3.38(b) and (c). The slow carrier relaxation predicted to result from the discrete energy levels of a quantum dot is known as the phonon bottleneck. In severe cases this slow relaxation may affect the performance of quantum dot devices, particularly for high-speed applications. Experimental studies of carrier relaxation mechanisms are discussed in Section 3.7.1.

### 3.6.10 The quantum confined Stark effect

Applying an electric field to a semiconductor causes the band edges to tilt along the field direction (inset of Figure 3.39(a)). Although spatially separated, states in the conduction band are now closer in energy to states in the valence band, resulting in absorption occurring below the band gap energy. This process is known as the Franz–Keldysh effect and can be used as the basis for an electro-optical modulator for light tuned to an energy slightly below the band gap. A practical modulator requires a high on/off contrast ratio, which is possible if there is a steep absorption onset at the band gap. In a bulk semiconductor the $\sqrt{E}$ variation of the absorption provides only a weak onset, although this is enhanced by the presence of an exciton. However, the application of only a relatively weak electric field pulls apart the

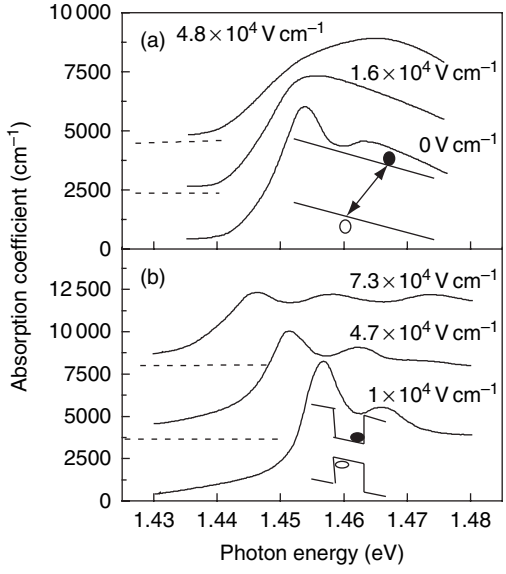


**Figure 3.39** Quantum well absorption spectra for (a) electric fields applied in the plane of the well, equivalent to a bulk semiconductor, and (b) along the growth axis. For spectra along the growth axis, the potential wells prevent ionisation of the excitons, so they are observed in the spectra to considerably higher fields. The insets show the band edge profiles for the two cases. Data reproduced from D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood and C. A. Burrus, *Phys. Rev. B* **32**, 1043 (1985). Copyright 1985 by the American Physical Society

electron and hole, ionising the exciton. With increasing field, excitonic effects are therefore quickly lost from the absorption spectra of a bulk semiconductor, as demonstrated in Figure 3.39(a), which shows absorption spectra as a function of electric field applied in the plane of a quantum well, equivalent to applying a field to a bulk semiconductor. In contrast, when an electric field is applied normal to the plane of a quantum well, the potential wells prevent the separation of the electron and hole, inhibiting the ionisation of the exciton (inset of Figure 3.39(b)). Excitonic effects are observed to considerably higher fields, as shown in Figure 3.39(b). The field-induced shift of the transition energy in a low-dimensional structure is known as the quantum-confined Stark effect and has practical applications in electro-optical modulators. Although initially observed in quantum wells, the quantum-confined Stark effect is also observed in quantum wires and dots.

### 3.6.11 Non-linear effects

The presence of a high density of additional carriers destroys an exciton. This results from a number of physical effects, including the screening of the Coulomb interaction between the electron and hole by the additional carriers and, more importantly, that these carriers occupy the states from which the exciton is formed – Pauli exclusion blocking. Figure 3.40 shows quantum well absorption spectra for the unperturbed state and soon after a very short laser pulse has created a large density of electrons and holes. The excitonic features are quenched by the optically created carriers. This behaviour provides the possibility for an all-optical modulator, where the carriers created by an intense laser beam switch a second weaker beam that is tuned to one of the exciton features. Although similar effects are observed in bulk semiconductors, the incident power required to 'switch off' an exciton in a quantum well is found to be significantly

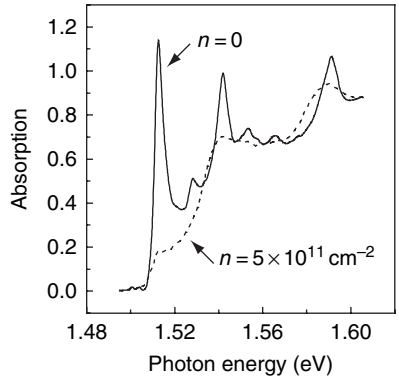


**Figure 3.40** Absorption spectra of a 156-period GaAs–$Al_{0.3}Ga_{0.7}As$ multiple quantum well with 20.5 nm wide wells. Spectra are shown for the unperturbed system ($n = 0$) and 100 ps after a pulsed laser has created a carrier density of $n \approx 5 \times 10^{11}$ cm$^{-2}$ in each well, quenching the exciton features. Data reproduced by permission of Elsevier from C. V. Shank, R. L. Fork, R. Yen, J. Shah, B. I. Greene, A. C. Gossard and C. Weisbuch, *Solid Stat. Commun.* **47**, 981 (1983). Copyright 1983

less than in a bulk semiconductor. In addition, excitons are also present in quantum wells at room temperature, an essential requirement for practical devices. A further manifestation of the ability to control the strength of an exciton is that the fractional absorption experienced by light tuned to one of the exciton energies is dependent on the intensity of that light. With increasing power, larger numbers of carriers are excited and the exciton strength is reduced, decreasing the absorption. The resultant variable absorption is a non-linear effect and has a number of practical applications in optical systems.

### 3.6.12 Coherence and dephasing processes

When an exciton is optically excited in a semiconductor there is a well-defined phase relationship between the exciton wavefunction and the optical field. The field and exciton are coherent. Over time, various scattering processes will randomly alter the exciton phase, and the phase relationship with the field will be lost. This is known as dephasing. Alternatively, if a collection of excitons are created, they will initially have the same phase but this coherence will decay at a rate determined by the dephasing time. The most important dephasing mechanisms in semiconductors are the scattering of the exciton by other excitons, free carriers or phonons. These mechanisms are very efficient, hence dephasing times are generally very short. Typical values for a quantum well are 10 ps at 4 K, decreasing to 0.1 ps at room temperature. Recently there has been increased interest in dephasing processes as a result of the possible use of excitons for the basic elements of quantum computers; so-called qubits. For this application a system is required that retains its phase over a time interval sufficiently long to allow a number of logical operations to be performed. Quantum dots are of particular interest for this application as the spatial localisation of excitons within a dot should prevent interaction with other excitons, and the limited number of confined energy levels may reduce phonon scattering. Measurements on self-assembled quantum dots at low temperatures reveal very long dephasing times, with a value of $\sim$1 ns as the temperature approaches absolute zero. However, the dephasing time decreases with increasing temperature, and by room temperature it has a value similar to that found for a quantum well. The dephasing time of the exciton in a quantum dot also determines the homogeneous line width of the emission via the uncertainty relationship $\tau\Delta E = 2\hbar$, where $\tau$ is the dephasing time and $\Delta E$ is the linewidth. A dephasing time of 1 ns equates to a very narrow linewidth of $\sim 1\mu$eV.

## 3.7 THE CHARACTERISATION OF SEMICONDUCTOR NANOSTRUCTURES

In this section the main experimental techniques that can be used to probe the structural, electronic and optical properties of inorganic semiconductor nanostructures are reviewed. Emphasis is placed on the information provided by the techniques, and their relative advantages and disadvantages. Additional background material, relevant to many of the techniques, is given in Chapter 2.

### 3.7.1 Optical and electrical characterisation

The most commonly applied optical characterisation technique is photoluminescence (PL). PL involves creating electrons and holes by illuminating the structure with photons of sufficient energy, generally using light from a laser as described in Section 2.7.1.2. Typically photon energies are used such that absorption occurs in the barriers, as this produces a relatively high density of electrons and holes and therefore results in a strong PL signal. After creation the carriers diffuse spatially and are captured and localised in different parts of the structure. This is followed by relaxation as the carriers lose any excess kinetic energy, generally reaching the lowest possible energy states before recombining to emit a photon. The energies of the emitted photons are determined using a spectrometer and suitable detector.

The carrier transport efficiency between different parts of a nanostructure affects which regions contribute to the PL. In quantum wells and self-assembled quantum dots there is a very rapid transfer of carriers from the barriers into the lower energy states provided by the wells or dots. As a result, emission from the barriers, and from the wetting layer in quantum dot structures, is generally not observed. However, in nanostructures where the different regions are spatially well separated, efficient carrier transport may not be possible, and a number of regions may contribute to the PL. For example, in V-groove quantum wire structures (Section 3.5.12), the quantum wires, the quantum wells formed on the sides of the grooves and between the grooves, and the bulk GaAs may all give distinct emission, as shown in Figure 3.41. This behaviour complicates the interpretation of the emission spectrum and is a serious disadvantage for optical devices where emission at a single energy is generally required.

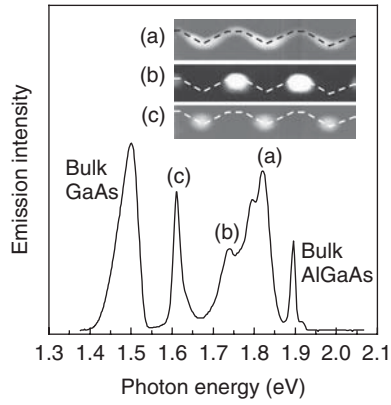


**Figure 3.41** Cathodoluminescence (CL) spectrum of a V-groove quantum wire structure showing emission from the different regions of the structure. Scanning CL images are shown recorded for the detection of emission lines (a), (b) and (c). The forms of these images are consistent with line (a) arising from carrier recombination in the quantum wells formed on the side walls of the grooves, (b) quantum wells formed on the planar regions between the grooves, and (c) the quantum wires at the bottom of the grooves. The dashed lines in the images indicate the approximate position of the pre-growth surface. Data and images courtesy of Dr Gerald Williams, QinetiQ, Malvern

In conventional PL measurements the exciting laser beam is focused to a spot of diameter $\sim$100 μm. Because of the high density of dots or wires in a typical structure, this results in the simultaneous measurement of many nanostructures. For example, V-groove quantum wires are typically spaced by $\sim$1 μm and self-assembled quantum dots may have a density $\sim 10^{11}\,\mathrm{cm}^{-2}$. For these densities a spot of diameter 100 μm excites 100 wires or 10 million dots. Consequently, the PL spectra are inhomogeneously broadened due to unavoidable fluctuations in the wire or dot shape, size and composition. Although the magnitude of this broadening provides information on the homogeneity of the nanostructures, it prevents the study of processes that occur on comparable or smaller energy scales, such as the perturbation of the emission energy of a quantum dot as additional excitons are added (Section 3.6.8).

In order to study physical processes occurring on an energy scale smaller than the inhomogeneous broadening, it is necessary to probe individual dots or wires. This can be achieved through reducing the size of the focused laser beam, either by using a large-aperture microscope objective, for which a diffraction-limited spot size of $\sim$1 μm is possible, or by using a scanning near-field optical microscope (SNOM) that allows the diffraction limit to be circumvented. The principle and operation of a SNOM are discussed in Sections 2.4.1 and 9.1.1. Both techniques allow single quantum wires to be studied. However, the tendency of carriers to diffuse away from the illuminated area, coupled with the typically large quantum dot densities, necessitates the use of additional steps to probe single quantum dots. This generally involves the reduction of the dot density, achieved by modifying the growth conditions, followed by the physical isolation of a single dot, either by etching submicron mesas or forming small holes in an otherwise opaque metal surface mask. Examples of single quantum dot spectroscopy are shown in Figure 3.22, where the spectra are recorded through different sized holes formed in a metal mask, and Figure 3.37, which shows spectra of a single quantum dot isolated in a 200 nm diameter mesa.

Carriers may also be created electrically by placing a nanostructure in a p-i-n device. In this case the process of light emission is known as electroluminescence (EL). EL has the advantage that the rate of carrier injection is uniform across the area studied, in contrast to PL, where the incident laser beam has a non-uniform, Gaussian profile. EL from a single quantum dot may be observed by exciting a relatively large number of dots and selecting the emission from a single dot by using a small aperture in a metal mask, which also forms one of the contacts. EL can also be excited using current injection from a scanning tunnelling microscope (STM). This method offers high spatial resolution and is therefore particularly suitable for the study of single quantum dots.

A third mechanism for the excitation of luminescence is a beam of high-energy electrons, such as found in an electron microscope: this is the cathodoluminescence (CL) technique described in Section 2.7.3.2. Although the nominal spatial resolution provided by the electron beam is degraded by carrier diffusion, a careful choice of beam voltage and beam current makes it possible to observe the emission from a single quantum wire and a few quantum dots. One powerful application of CL is in identifying the origin of the different features present in the emission spectra of complex structures; for example V-groove quantum wires. Initially the emission spectrum for excitation of a large area is recorded. The system is then set to detect photons corresponding to one of the emission features, and the electron beam is raster scanned over the sample. Regions of the structure responsible for the selected emission appear bright in the resultant image, allowing their position and shape, and hence their origin, to be determined. This procedure

is then repeated for the different emission features. Scanning CL images from the cleaved edge of a V-groove quantum wire structure are shown in Figure 3.41. Detecting photons corresponding to line (c) results in emission located at the bottom of the grooves, and therefore originating from the quantum wires. Lines (a) and (b) result in emission from the side walls of the grooves and the flat surfaces between the grooves, consistent with emission from the side and top quantum wells, respectively. These wells have different thicknesses, hence they emit at different energies, a result of the dependence of the GaAs growth rate on surface orientation.

Carrier relaxation is generally much faster than radiative recombination (Section 3.6.9); hence only the lowest-energy, ground state transition is observed in emission. Emission from higher energy states can be observed by increasing the carrier excitation or injection rate so that the population of carriers in the ground state is sufficient to block relaxation from the excited states. An example is shown for quantum dots in Figure 3.36. Although high carrier injection allows the excited states to be observed, the system is highly occupied and the interaction between the carriers may significantly perturb the energies of the transitions. In addition, emission techniques do not allow the true relative strengths of the transitions to be determined, as the emission intensity is dependent not only on the intrinsic transition strength but also on the carrier populations in the initial and final states.

A determination of the unperturbed energies and the strengths of optical transitions therefore requires the use of an absorption technique. However, it is very difficult to measure the direct absorption of a single nanostructure because only a very small fraction of the light is absorbed in comparison to the majority of the light that simply passes through the sample. Measuring this small change in transmitted light against the large background is technically very difficult. For quantum wells it is possible to use multiple-well structures to increase the absorption to a measurable level, and for colloidal and nanocrystal dots it is possible to produce films or solutions containing the dots of sufficient thickness to allow direct absorption measurements. However, absorption measurements of epitaxially grown wires and dots are more difficult because, as they occupy only a relatively small fraction of the cross-sectional area, their intrinsic absorption is very low. The absorption spectrum of a single layer of self-assembled quantum dots can be measured, but this requires the use of very sensitive and expensive commercial systems. Even with such systems the very small absorption signal requires the use of extremely long integration times, typically a few hours. By focusing a laser beam to a spot size $<1$ μm, using a large-aperture microscope objective, it is possible to measure the absorption spectrum of a single self-assembled quantum dot, whose physical cross section is now a reasonable fraction of the laser beam area. However, the absorption is still too low to allow the direct measurement of the absorption. Consequently, a modulation technique consisting of an oscillating electric field that varies the transition energies via the quantum-confined Stark effect is used. The resulting spectra correspond to the first derivative of the absorption.

Because of these experimental difficulties, absorption studies of quantum wires and dots generally make use of an absorption-related technique; photocurrent (PC) spectroscopy and photoluminescence excitation (PLE) are the main examples. In PC, incident photons create electrons and holes in a nanostructure, which is placed in the intrinsic region of a p-i-n device. Under suitable conditions the carriers are able to escape from the nanostructure before recombining, giving rise to a current that can be measured by an

external circuit. As the energy of the incident photons is varied, a change in absorption will alter the number of carriers created, and hence the magnitude of the photocurrent. In PLE the intensity of the photoluminescence is monitored as the energy of the exciting photons is varied. A change in the absorption alters the number of carriers created, and hence the intensity of the photoluminescence. Both PC and PLE measure a small signal against a zero background; although the majority of incident light still passes through the structure, this does not contribute to the measured signal. PC and PLE are therefore referred to as background-less techniques and are particularly suited to the study of nanostructures that absorb only very weakly. However, both techniques involve an additional step beyond photon absorption. In PC the photo-created carriers must escape from the nanostructure and in PLE they must relax to the transition being monitored. If the probability of these processes is not constant, but depends on the initial energy of the carriers, then the form of the resultant spectra will not reflect the true absorption.

A further important class of optical characterisation techniques is time-resolved spectroscopy. Here a structure is excited by a very short pulse of light, and the subsequent temporal changes in its properties are determined as carriers recombine or relax in energy. The pulse length and time resolution of the measurement system are typically in the range 1 ns to 0.1 ps; the precise value is dependent on the physical processes being studied. Time-resolved spectroscopy has been used to study a range of carrier processes in semiconductor nanostructures, including the rate at which carriers are captured from the barriers into the nanostructure, the rate at which carriers relax between confined levels, dephasing times and recombination times. Figure 3.42 shows results from a study of carrier relaxation mechanisms in InGaAs self-assembled quantum dots. Electrons and holes are initially created in the GaAs barriers by 1.5 ps pulses

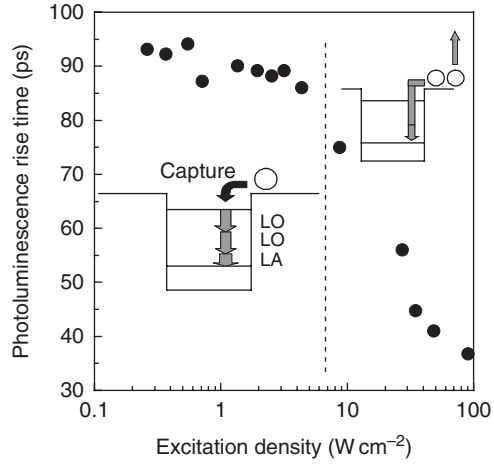


**Figure 3.42** The rise time of the PL from InGaAs self-assembled QDs measured as a function of laser power density. The rise time indicates the time required for the carriers to be captured by the dots followed by energy relaxation to the ground state. Two possible relaxation processes are shown schematically. The vertical dashed line indicates the laser power corresponding to the creation of an average of one electron and hole per quantum dot. Data from B. Ohnesorge, M. Albrecht, J. Oshinowo, A. Forchel and Y. Arakawa, *Phys. Rev. B* **54**, 11532 (1996). Copyright 1996 by the American Physical Society

of light. The PL from the ground state of the quantum dots is detected and the rise time of this emission provides an indication of the time taken for carriers to be captured into the dots, followed by relaxation to their ground state. In Figure 3.42 the photoluminescence rise time is plotted as a function of the laser power, which is used to vary the density of carriers created in the structure. For low powers, equivalent to less than one electron and hole per dot, the rise time is relatively long, and reflects the slow relaxation of carriers between the discrete states of the dots by the emission of multiple phonons. This is a manifestation of the phonon bottleneck (Section 3.6.9). This slow rise time is retained until the laser power becomes sufficient to excite an average of one electron and hole per dot, indicated by the vertical dashed line in the figure. Above this power the rise time starts to decrease, reflecting carrier capture and relaxation by a much faster process in which the energy lost by one carrier is transferred to a second carrier. This mechanism, known as an Auger process, requires an electron or hole in addition to the relaxing carriers, hence it only becomes possible above an average electron–hole number of one per dot. Further increase in the number of carriers per dot increases the efficiency of this process, resulting in the continuous decrease of the rise time observed at high laser powers. The insets to Figure 3.42 show the carrier relaxation processes relevant to the low and high carrier density regimes.

Electrical measurements in their basic form involve the determination of currents and voltages from which sample resistances and electrical conductivity can be obtained. With the addition of a magnetic field, the integer and fractional quantum Hall effects can be studied, and measurements at very low fields allow a determination of carrier mobility. Two-dimensional carrier densities are determined from the period of the $\rho_{xx}$ oscillations in the integer quantum Hall regime; this is also known as the Shubnikov–de Haas effect. Information on the dominant carrier scattering mechanism may be obtained by studying the temperature variation of the carrier mobility.

The capacitance of a quantum dot structure is affected by the charge state of the dots and this provides a method for probing the number of electrons or holes which have been loaded into a dot (Figure 3.29 and discussion in Section 3.6.4). A refinement of this technique is deep-level transient spectroscopy (DLTS), which measures the temporal evolution of the capacitance following the application of a voltage pulse. If the size and sign of the pulse are chosen such that excess carriers are loaded onto the dot, then a transient change in the capacitance will occur. Following removal of the pulse, the capacitance will revert back to its original value as the excess carriers escape from the dot. By measuring the rate at which the capacitance recovers it is therefore possible to determine the carrier escape rate, and measurements as a function of temperature allow the height of the potential barrier confining the carriers to be determined.

### 3.7.2 Structural characterisation

A number of techniques are available for the direct study of the physical structure of quantum wells and superlattices. These include: X-ray diffraction, which can determine periodicity, layer uniformity and in some cases composition; and transmission electron microscopy, which can determine layer thicknesses and compositions and which enables

the nature of interfaces to be directly imaged. X-ray studies of quantum wires and dots are more limited, although shallow incidence X-ray diffraction has been used to determine the distribution of indium in self-assembled quantum dots. Transmission electron microscopy has been applied extensively to the study of quantum dots and wires, and examples of cross-sectional images are presented in Figure 3.18 and Figure 3.20. Plan-view images recorded along the growth axis, and for a geometry sensitive to strain, reveal the spatial distribution of self-assembled quantum dots (Figure 3.20), although because the strain field extends beyond the dots, their size and shape is not directly imaged. A complication arising with cross-sectional images of quantum dots is that for dot shapes other than cuboidal the apparent shape depends on where the dot is sectioned during sample preparation. This complication also applies to compositional studies.

Compositional information is provided by electron energy loss spectroscopy (EELS) and energy-dispersive X-ray spectroscopy (EDX), techniques which are described in Section 2.7.3. For quantum wells and wires, which have a uniform cross section through the thinned specimen, it is possible to obtain absolute compositions from both EELS and EDX by first calibrating the system with samples of a known composition. However, for quantum dots the uncertainty in how the dot has been sectioned makes it difficult to achieve absolute compositional determination. Figure 3.43 shows an EELS image of the gallium distribution in an InAs self-assembled quantum dot grown within a GaAs matrix. In addition, the two traces show EDX line scans of the indium distribution along directions parallel to the growth axis and passing through either just the wetting layer or both the dot and wetting layer. The profile of the latter trace is broader, reflecting the additional indium-containing region of the quantum dot.

Atomic force microscopy (AFM) is routinely used to study nanostructures as no complicated sample preparation is required. AFM allows the shape, size and distribution of self-assembled quantum dots to be determined (Figure 3.21) but these dots must

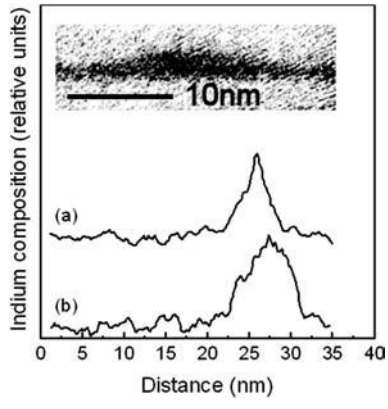


**Figure 3.43** EDX profiles of a self-assembled InAs quantum dot structure recorded for line scans (a) passing through the wetting layer and (b) passing through the wetting layer and a quantum dot. Also shown is an EELS image of the Ga composition in an InAs self-assembled quantum dot structure with GaAs barriers. Data reproduced by permission of the Institute of Physics from M. A. Al-Khafaji, A. G. Cullis, M. Hopkinson, L. R. Wilson, S. R. Parnell, D. J. Mowbray and M. S. Skolnick, *Inst. Phys. Conf. Ser.* **161**, 585 (1999)

be situated on the sample surface, requiring the growth to be terminated immediately after the dots have been formed. In contrast, dots for optical and other applications must be covered by a relatively thick 'protective' layer to prevent the strong non-radiative carrier recombination which otherwise occurs at a free surface. As there is some evidence that material diffuses into and out of the dots during the growth of the capping layer, altering their shape, size and composition, it is possible that dot parameters determined by AFM may be different to those of covered dots.

AFM may also be used to determine alloy compositions in aluminium-containing nanostructures. This requires the nanostructure to be cleaved and exposed to air, causing the aluminium-containing materials to oxidise and bow outwards slightly from the surface. The degree of bowing is proportional to the local aluminium content and can be measured by an AFM. By calibrating the system with samples of known composition it is possible to produce a two-dimensional image of the aluminium composition.

Structural information may also be obtained from scanning tunnelling microscopy (STM). Figure 3.23 shows an STM image of a cleaved sample containing a stack of five InAs self-assembled quantum dots. High tunnelling current corresponds to regions of high indium composition, a result of a smaller local band gap and the increased strain which causes the cleaved surface to bow out slightly. A simulation of these two contributions to the tunnelling current allows a determination of the indium distribution across the cleaved surface.

## 3.8 APPLICATIONS OF SEMICONDUCTOR NANOSTRUCTURES

A number of practical applications of semiconductor nanostructures have been briefly discussed previously and in this section some of these are considered in greater detail. There is insufficient space to consider all current or potential future applications, and the aim is to provide examples of the more important ones and to demonstrate the breadth of possibilities provided by inorganic semiconductor nanostructures.

### 3.8.1 Injection lasers

Semiconductor injection lasers are physically small, convert electrical energy into light with high efficiency, have very long operating lifetimes, and can be switched on and off (i.e., modulated), extremely rapidly. This makes them very suitable for many applications, including data transmission and storage, printing and medical uses. However, initial devices, which were based on bulk, three-dimensional semiconductors, were far from ideal. Their threshold current, the current that must be applied for the laser to operate, was relatively high and increased rapidly with device temperature. This section shows that the use of nanostructures in the emitting region of a laser can result in significant performance improvements.

A laser is an optical oscillator and therefore contains a gain mechanism by which light is amplified. Gain is the inverse of absorption and in a semiconductor it occurs

when there are a large number of electrons in the conduction band and a large number of holes in the valence band, a condition known as a population inversion. A population inversion is achieved in a forward-biased p-i-n structure where large numbers of electrons and holes are injected into the intrinsic region. For the system to oscillate or lase, the total gain must balance the total loss. This requires the creation of sufficient densities of electrons and holes, hence the injection of a corresponding current, to achieve the required gain. The gain is proportional to the density of occupied states: hence, because the density of states in a bulk semiconductor increases with energy (Section 3.4), the gain will increase with increasing carrier density as higher energy states are successively filled. In a bulk system it is therefore necessary to 'fill' the density of states to a point where sufficient gain is reached, with the carriers at lower energies effectively wasted. This process is shown schematically in Figure 3.44(a). These wasted carriers result in a relatively large threshold current.

The situation depicted in Figure 3.44(a), with only the lowest energy states occupied, corresponds to absolute zero temperature. At non-zero temperatures, carriers will be thermally excited to higher states, as shown in Figure 3.44(b). This further wastes carriers by exciting them into states not involved with the lasing process, a particularly serious problem in a bulk laser where the form of the density of states results in a large number of states at higher energies. As the temperature is increased, an increasing fraction of the carriers are thermally excited out of the lasing states and in Figure 3.44(b) the maximum gain is now below that required for lasing action. Consequently, a higher current must be applied to achieve the required gain. This mechanism is the reason why the threshold current of a laser increases with increasing temperature.

The modification of the density of states in a nanostructure overcomes, to various extents, the limitations of bulk semiconductor lasers. In particular, the density of states is increased at low energies with respect to higher energies, which results in both a decrease in the absolute threshold current and also its sensitivity to temperature. In the ultimate limit of a quantum dot with only one confined electron state and one confined hole state, all of the carriers must have the same energy at all temperatures – there are no states available for thermal excitation – and the laser should exhibit a very low and temperature-insensitive threshold current.

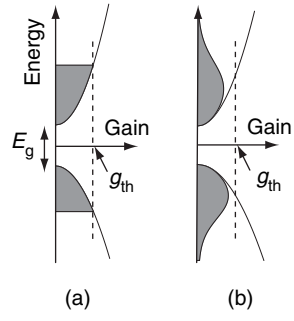


**Figure 3.44** Electron and hole distributions in the conduction and valence bands of a bulk semiconductor laser at (a) absolute zero temperature and (b) non-zero temperature. In (b) thermal excitation of carriers to higher energy states has reduced the maximum gain below the value, $g_{th}$, required for laser action

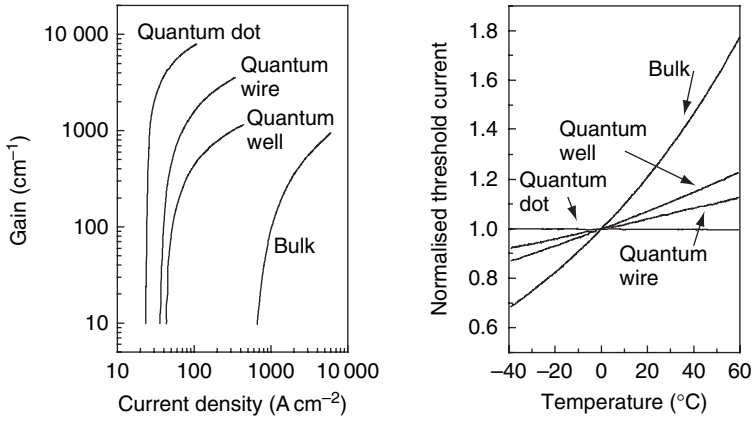


**Figure 3.45** Left-hand panel: calculated gain variation with current density for lasers having different dimensionalities. For a given current density, the gain increases as the dimensionality decreases. Data reproduced by permission of the American Institute of Physics from Y. Arakawa and H. Sakaki, *Appl. Phys. Lett.* **40**, 939 (1982). Right-hand panel: calculated temperature variation of the threshold current density for bulk, quantum well, quantum wire and quantum dot lasers. Decreasing the dimensionality increases the temperature stability. Data reproduced from M. Asada, Y. Miyamoto and Y. Suematsu, *IEEE J. Quant. Electron.* **QE-22**, 1915 (1986). Copyright 1986 IEEE

The potential advantages arising from the use of nanostructures in a laser were first predicted from calculations performed for idealised systems. Figure 3.45 shows how the gain of systems having different dimensionalities develops with increasing current; it also shows the dependence of the threshold current on temperature. A given gain is reached at successively lower currents as the dimensionality is decreased, with threshold current densities of 1050, 380, 140 and 45 A cm$^{-2}$ calculated for bulk, quantum well, quantum wire and quantum dot lasers, respectively.

The temperature dependence of the threshold current density of a semiconductor laser can be described by an equation of the form

$$J_{\text{th}} = J_{\text{th}}^{0} \exp(T/T_0), \tag{3.13}$$

where $J_{\text{th}}^{0}$ is the threshold current density at 0 °C and $T_0$ is a parameter that determines the temperature sensitivity; a larger $T_0$ corresponds to a lower sensitivity. For the idealised lasers considered in Figure 3.45, $T_0$ values of 104, 285, 481 °C and infinity are calculated, demonstrating a reduced sensitivity with the progression bulk→well→wire→dot.

Semiconductor lasers based on quantum wells are now used extensively for many applications, including CD and DVD data storage and fibre-optic data transmission. Quantum wire lasers have been fabricated, although their performance has been limited by the lack of systems with suitable energy spacings between the confined subbands. Greater progress has been made with quantum dot lasers based on self-assembled dots. At room temperature, threshold current densities are approximately one-third the value of comparable quantum well lasers, and the stability of the threshold current is improved by approximately a factor of two. That these improvements are not as great as predicted by the calculations summarised in Figure 3.45 is due to the departure of self-assembled dots

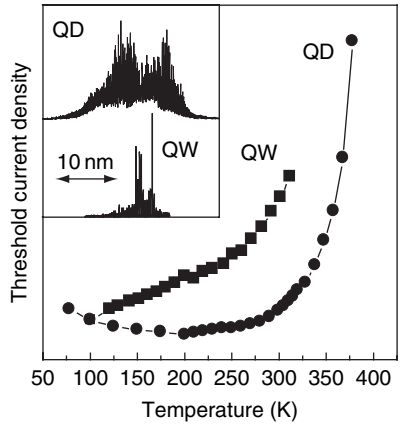


**Figure 3.46** Comparison of the temperature dependence of the threshold current density of a self-assembled quantum dot laser (circles) and a quantum well laser (squares). Because of the different device designs, the threshold current densities are normalised to their 100 K values. The inset compares emission spectra of a quantum well (QW) laser and a quantum dot (QD) laser. Quantum well laser data reproduced from T. Higashi, S. J. Sweeney, A. F. Phillips, A. R. Adams, E. P. O'Reilly, T. Uchida and T. Fujii, *IEEE J. Select. Topics Quant. Electron.* **5**, 413–419 (1999). Copyright 1999 IEEE. Quantum dot laser data courtesy of Ian Sellers, University of Sheffield

from idealised systems. In particular, self-assembled dots have a number of confined levels into which carriers may be thermally excited, as well as the surrounding barrier material, resulting in a finite $T_0$. In addition, there may be non-radiative processes and the dot ensemble is inhomogeneously broadened. Figure 3.46 shows a comparison of the temperature dependence of the threshold current densities of a self-assembled quantum dot laser and a quantum well laser. At low temperatures the expected temperature stability of the quantum dot laser is achieved and below 200 K the quantum dot laser is significantly more stable than the quantum well laser. However, above 200 K the threshold current density of the quantum dot laser increases due to the mechanisms previously discussed, and by 300 K the quantum well and dot lasers exhibit very similar temperature sensitivity. One of the key goals in quantum dot laser development is to extend the temperature-insensitive regime to room temperature and beyond.

In addition to their low threshold current density and high temperature stability, quantum dot lasers offer a number of additional advantages over lasers of higher dimensionality. Their maximum modulation frequency should be higher, although this may be compromised if carriers are unable to relax sufficiently rapidly to the states from which lasing occurs – the phonon bottleneck. Once carriers have been captured into a quantum dot, their subsequent motion is restricted unlike the case, for example, of a quantum well laser where carriers captured into the well are still free to move within the plane of the well. This carrier localisation prevents the diffusion of carriers to non-radiative centres which may be formed on the surface of the device or within the device by, for example, radiation damage. Quantum dots are therefore particularly suitable for the fabrication of very small lasers, with a high area-to-volume ratio, and lasers for use in harsh radiation environments. Carrier localisation also prevents carriers in different dots from directly interacting and subsets of dots with similar emission energies may act as independent lasers. The dot

ensemble may therefore behave as a collection of sub-lasers, with lasing occurring over a significant fraction of the inhomogeneous line width. Although it is a disadvantage for many applications where a single lasing energy is required, a broad emission allows the fabrication of a tunable laser by placing the system in an external cavity, allowing selection of one particular frequency. Figure 3.46 compares quantum dot and quantum well laser spectra; the increased width of the quantum dot spectra is clearly visible. Finally, quantum dots allow emission at new wavelengths to be obtained. GaInAs quantum well lasers grown on GaAs substrates are limited to $<1.2\,\mu m$ by the critical thickness of this strained system (Section 3.5.12). However, InAs quantum dots allow the fabrication of lasers operating in the important $1.3\,\mu m$ telecommunications band, with some prospects for devices operating in the related $1.55\,\mu m$ band. Current 1.3 and $1.55\,\mu m$ lasers require the use of quantum well lasers grown on less technologically convenient InP substrates.

The most common type of semiconductor laser is based on a transverse geometry where light propagates in the plane defined by the quantum well or dots. In this geometry the mirrors at the ends of the optical cavity are formed by cleaving the semiconductor along a particular crystal direction to give atomically flat surfaces. The refractive index difference between air and the semiconductor produces mirrors with a reflectivity of typically $\sim 30\%$. Although this reflectivity is relatively low, and represents a large loss for photons within the cavity, the gain of the system can be increased to compensate for this loss by increasing the length of the cavity. Typical cavity lengths are $\sim 1\,mm$. Although used for many applications, transverse lasers have a number of disadvantages: the cleaving step is difficult and adds significantly to fabrication costs, the different dimensions normal to and in the plane of the laser result in an output beam with an elliptical cross section, and it is not possible to fabricate the two-dimensional arrays which may be required for optical interconnects in future generations of microprocessors. These problems may be overcome by using a geometry in which the light propagates normal to the plane, to give a vertical cavity surface-emitting laser (VCSEL). In this geometry the available gain is much smaller than for a transverse laser as the light passes only once through a layer of quantum dots or a quantum well, instead of along the layer. Consequently, the losses must be reduced significantly and a mirror reflectivity of $\sim 99.9\%$ is required, a value that cannot be achieved with a simple cleaved semiconductor–air interface. Instead the mirrors are formed by depositing alternating layers of two semiconductors having different refractive indexes. The resultant Bragg stack (Section 3.8.8) has a reflectivity that increases with increasing layer number, allowing a reflectivity $>99.9\%$ to be obtained with typically $\sim 20-30$ layer pairs. Although the growth sequence of a VCSEL is more complex than that of a transverse laser, the subsequent fabrication is greatly simplified, with the devices formed by lithography and etching to produce circular mesas; there is no cleaving step. Dense two-dimensional arrays can be formed, with the circular cross section of the devices resulting in symmetrical output beams.

## 3.8.2 Quantum cascade lasers

The lasers discussed in the previous section are based on interband transitions, with lasing occurring between states in the valence and conduction bands, and are classed as bipolar because their operation requires both electrons and holes. Interband lasers

based on various semiconductor combinations are able to cover a spectral range extending from the near infrared through to the violet, ~2.0−0.4 μm. However, lasers operating in the infra-red region between ~2−100 μm are more difficult to obtain, due to a lack of semiconductors with sufficiently small band gaps. This spectral region is technologically important as it contains many molecular absorption bands. Quantitative detection of a specific gas is possible if a laser can be tuned to one of the gas absorption bands.

Recently a new type of semiconductor laser has been developed, the quantum cascade laser, which provides emission in the infrared spectral region. This is an intraband, unipolar device that relies only on electrons, which make transitions between confined quantum well conduction band states. Figure 3.47 shows a typical band structure of a quantum cascade laser. The lasing transition occurs between quantum well levels 3 and 2, the separation of which is typically in the range of ~12−350 meV, corresponding to wavelengths ~100−3.5 μm. The lasing wavelength can be varied by selecting a suitable well width. Electrons are injected into the upper lasing level 3 by tunnelling through the left-hand barrier, B, before relaxing to the lower level 2 by emitting a photon. Level 1 is required to efficiently remove electrons from the lower laser level, as a build-up of electrons in this level can prevent the attainment of a population inversion, which requires a greater number of electrons in level 3 than level 2. The gain provided by a single stage is generally not sufficient to overcome the losses of the system, and structures based on typically 25 coupled stages are used, although devices with as many as 100 stages have been reported. Electrons are transported between stages via the miniband of a superlattice, which is designed to inject electrons having the correct energy into the next stage. In this scheme the electrons cascade down through the multiple stages of the structure.

Quantum cascade lasers based on a number of different designs have been operated successfully, with emission wavelengths from 3.5 to 106 μm. However operation at room temperature has only been achieved for the more limited range of 4.5 to 16 μm, and only at 6 and 9.1 μm has continuous room temperature operation been demonstrated; all other devices operate in pulsed mode. Room temperature continuous operation at only two wavelengths reflects efforts to optimise the relevant devices, and similar operation over approximately the entire range 6 to 9.1 μm should be possible. In general, continuous

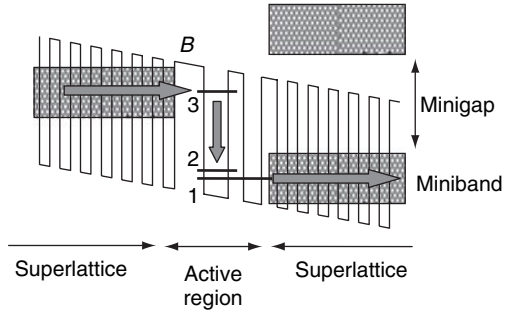


**Figure 3.47** The band structure of a quantum cascade laser. Electrons are transported between stages via the miniband of a superlattice, before tunnelling through barrier B into the upper lasing state, state 3. The minigap formed between the minibands of the superlattice prevents electrons tunnelling directly out of state 3. Courtesy of John Cockburn, University of Sheffield

operation at room temperature is made difficult by significant heat generation, a result of the large threshold current densities, which are typically up to two orders of magnitude higher than those of interband lasers. This inefficiency arises because electrons are able to relax between the lasing levels by emitting a phonon instead of a photon, resulting in a large fraction of the electrons being wasted. In addition, electrons may be thermally excited out of the quantum well from the upper lasing state. However, despite their high threshold current densities, quantum cascade lasers are being investigated for a number of applications, including gas monitoring, free space communication systems and medical imaging.

### 3.8.3 Single-photon sources

It is possible to transmit information using the polarisation state of single photons. Because performing a measurement will disturb the polarisation, attempts to break into the system are readily detectable. This property forms the basis of an intrinsically secure quantum cryptography system, the security being guaranteed by the laws of quantum mechanics. The key component on the transmission side of a quantum cryptography system is the production of a stream of regularly spaced single photons. This can be achieving by using a highly attenuated pulsed laser. However, because of the statistical nature of the number of photons in each pulse, it is necessary to attenuate the laser considerably below an average of one photon per pulse. At higher levels, a significant number of pulses will contain two or more photons, which may allow the integrity of the system to be broken. A maximum of one photon per 100 pulses is typically used, which significantly reduces the data transmission rate.

An alternative method for producing a regular sequence of single photons is based on the emission from a single quantum dot. As discussed in Section 3.6.8, the energy of a photon emitted from a dot containing two excitons is slightly different from that containing a single exciton, a result of the interaction between the charged carriers which perturbs the energy of the system. Hence a dot initially loaded with, for example, five excitons will emit five photons, each having a slightly different energy, as the excitons recombine. By filtering the emission so that only the photon due to the recombination of the final exciton is selected, only one photon will be produced each time the dot is loaded with excitons, irrespective of the initial number of excitons. Single-photon sources have been demonstrated based on the filtered emission from a single self-assembled InAs quantum dot that is periodically loaded with excitons, either optically with a pulsed laser or electrically by placing the dot within a p-i-n structure and applying a pulsed bias voltage. Although the dot produces one photon per cycle, the efficiency with which these photons can escape from the structure is low, due to reflection losses at the air–semiconductor interface. However, the number of external photons per pulse is comparable to that achieved using an attenuated laser and may be increased by the use of a photonic structure (Section 3.8.8). An additional problem is that the selection of one of the emission lines is only possible if it is separated from the neighbouring lines by an energy greater than its line width. At helium temperatures this condition is satisfied but at higher temperatures the emission line width is homogeneously broadened by interaction with phonons, and this broadening limits the use of InAs-based dots to below $\sim$80 K. Higher-temperature operation may be possible with

the use of II–VI dots, which have a larger separation between the different emission lines, although ideally the emission energy should be matched to the 1.3 or 1.55 μm transmission maxima of current optical fibres.

### 3.8.4 Biological tagging

Free-standing colloidal quantum dots (Section 3.5.11) have been developed as a means of tagging biological materials. Different sized dots emit light at different energies or colours when excited with a suitable laser. By preparing a number of dot sizes, various combinations of these sizes can be placed in protective latex beads. When illuminated the beads produce a combination of colours that is characteristic of the dots contained. This produces an optical bar code which can be read with a suitable detection system, allowing the identification of beads containing different dot combinations. The latex beads can be attached to biological materials, allowing the reactions between different materials to be studied or the transport of a particular material through an organism to be monitored. In comparison to traditional tagging that uses organic dyes, colloidal quantum dots offer a greater number of different colours, all of which can be excited with the same laser, in contrast to organic dyes where different dyes require different wavelength excitation. In addition, the emission intensity of colloidal quantum dots is more stable than that of organic dyes following prolonged excitation.

### 3.8.5 Optical memories

Charge storage in a quantum dot has the potential to provide high-density memory systems, with the possibility of both optical and electrical reading and writing. These systems may provide an alternative to the purely electronic random access memory used in present-day computers. Self-assembled quantum dots may have area densities as high as $1 \times 10^{11}\,\mathrm{cm}^{-2}$, a figure that can be increased by at least a factor of ten by using multiple dot layers. A storage density of $1 \times 10^{12}\,\mathrm{bits/cm}^2$ considerably exceeds that of conventional electronic memories. For memory applications the inhomogeneous broadening of the dots is an advantage, as potentially it allows individual dots to be addressed by the use of a laser tuned to their specific emission energy.

If both an electron and a hole are trapped in a quantum dot, they will recombine radiatively on a timescale of $\sim$1 ns, far too short for a practical memory system. A realistic quantum dot memory device therefore requires the storage of only electrons or holes. Figure 3.48 shows the band structure of a prototype quantum dot memory device based on InAs self-assembled quantum dots. Electrons and holes are created by the direct optical excitation of the dots. An applied electric field causes electrons to rapidly escape from the dots, but hole escape is prevented by the large barrier placed to the left of the dot layer. An alternative scheme is to use type II quantum dots; e.g., GaSb dots grown on GaAs, which result in the confinement of one carrier type only. The resultant net positive charge on the dots of Figure 3.48 repels a fraction of the holes from a two-dimensional hole gas (2DHG) placed parallel to the quantum dot layer. This reduced 2DHG density increases its resistance, a change which is monitored by an external

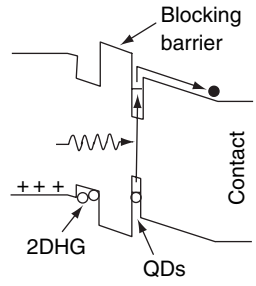


**Figure 3.48** Band structure of a quantum dot memory device

electrical circuit. The device is reset by applying a voltage pulse to a surface gate, causing electrons to flow into the dots, where they recombine with the stored holes. At a temperature of 145 K it is possible to store charge in the dots for in excess of 8 h. This storage time decreases at higher temperatures where the holes may be thermally excited out of the dots, although charge storage up to ~200 K has been observed. In current devices a significant number of dots must be excited as the influence of a single charged dot on the conductivity of the 2DHG is too small to measure. This indicates a significant problem with these devices in that although writing to a single dot appears possible, methods for reading and resetting single dots are unclear. At high temperatures, in addition to the loss of carriers from the dots by thermal excitation, the homogeneous broadening of the dot transitions appears to prevent optical writing to individual dots.

### 3.8.6 Impact of nanotechnology on conventional electronics

The information technology revolution has been driven by the rapid increase in computing power, exemplified by Moore's law which predicts that the number of transistors in microprocessors increases exponentially with time. The initial form of Moore's law, based on integrated circuit development in the mid 1960s, predicted a doubling of transistor number every year. A decade later this was redefined to a doubling every 24 months to account for increased circuit complexity. Figure 3.49 shows a plot of transistor number versus year, demonstrating that Moore's law has held for the previous 30 years or so, with a doubling of transistor number every 26 months. However, increasing the number of transistors on a chip necessitates a commensurate reduction in component size, which also allows the operating frequency to be increased. Over the period covered by Figure 3.49, operating frequencies have increased from ~1 MHz to ~3 GHz. Transistor dimensions will soon enter the nanoscale regime, where their fabrication and operation will cease to follow behaviour scaled down from larger sizes. This section discusses the problems that will be encountered in attempting to continue Moore's law over the next two decades and considers possible solutions, including the application of nanotechnology concepts discussed earlier in this chapter.

The transistors in a microprocessor are of the metal oxide semiconductor field effect transistor (MOSFET) type, the generic structure of which is shown in Figure 3.50. Current is carried between the drain and source contacts by free carriers in the channel:

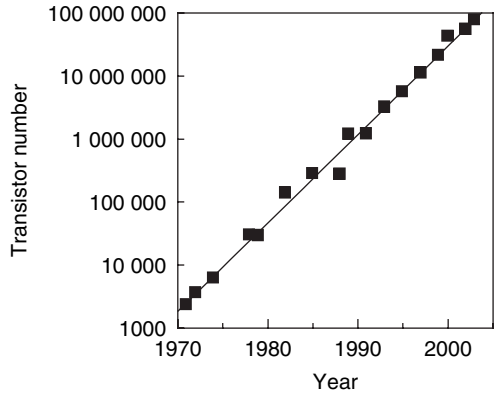


**Figure 3.49** Increase in microprocessor transistor number since 1970. The solid line is a fit to the data and indicates a doubling in number every 26 months

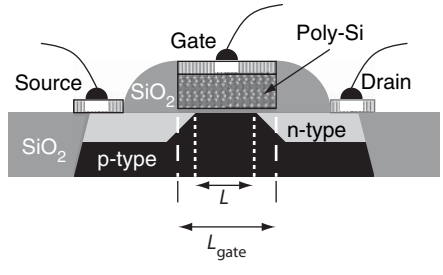


**Figure 3.50** Schematic structure of an n-type metal oxide semiconductor field effect transistor (MOSFET)

either electrons in an n-type device (NMOS) or holes in a p-type device (PMOS). A third contact, the gate, is placed above the channel, with a thin $SiO_2$ insulating layer preventing the flow of current between the gate and the channel. The doping type in the channel immediately below the gate is opposite to the doping type of the source and drain regions. However, application of a suitable gate voltage repels the majority carriers and attracts minority carriers to the channel–insulator interface region, allowing current conduction to occur between the source and the drain. By varying the gate voltage, this conduction path can be switched on or off, giving the basic switching action of the transistor. Neighbouring devices are electrically isolated from each other by insulating $SiO_2$ regions. Si provides the ideal semiconductor for MOSFET construction as its natural oxide ($SiO_2$) is highly insulating, it can form very thin layers and the $Si-SiO_2$ interface has a high structural quality, ensuring that carrier mobilities are not significantly degraded by interface roughness. In practice the metal gate of a MOSFET is replaced with polysilicon to reduce the size of the switching voltage.

The size of a MOSFET is defined by the gate length, and it is this parameter that has been continuously reduced to allow increasing numbers of faster transistors to be fabricated within a given circuit area. Table 3.1 shows the progress required in gate length reduction if Moore's law is to hold until 2018.

**Table 3.1** The reduction in microprocessor transistor sized required to continue Moore's law until 2018. Each generation is classified in terms of the technology node number, which specifies the half-pitch distance of related dynamic random access memories (DRAMs), defined as the smallest separation between lines in the first metal layer. The major nodes are in bold. The data is from the International Roadmap for Semiconductors, 2003

| | **2004** | **2005** | **2006** | **2007** | **2008** | **2009** | **2010** | **2013** | **2016** | **2018** |
|---|---|---|---|---|---|---|---|---|---|---|
| Node number (nm) | **90** | 80 | 70 | **65** | 57 | 50 | **45** | **32** | **22** | 18 |
| Transistor gate length in resist (nm) | 53 | 45 | 40 | 35 | 32 | 28 | 25 | 18 | 13 | 10 |

The first problem encountered in fabricating ever smaller devices is their lithographic definition. As a result of diffraction effects, the minimum feature size that can be imaged in the resist is

$$k\frac{\lambda}{\mathrm{NA}}, \tag{3.14}$$

where $\lambda$ is the wavelength of the radiation, NA is the numerical aperture of the optical system and $k$ (typically 0.5–0.9) is the technology constant that accounts for non-ideal behaviour of the system. It is difficult to increase the value of the numerical aperture significantly above unity, so for conventional lithography the minimum feature size is typically the same order as the radiation wavelength. Some reduction of the minimum feature size for a given wavelength is possible via refinements of the lithographic process. Examples include the use of phase-shifting techniques, off-axis illumination and immersion lithography. In the first modification a phase difference of $\pi$ is introduced for light passing through adjacent features in the mask. Consequently, light diffracted into the region between the features interferes destructively, increasing the contrast and hence resolution of the features. In immersion lithography a liquid in contact with the photoresist decreases the effective wavelength of the incident light, thereby decreasing the minimum feature size that can be formed. For a combination of phase shifting and off-axis illumination it is possible to obtain a $k$ value of 0.3. The use of immersion lithography provides a further decrease equal to the refractive index of the liquid (1.43 for water).

Current commercial high-volume lithography is based on deep ultraviolet (DUV) light produced by excimer lasers, with shorter wavelengths used for successive nodes; examples are 248 and 193 nm for the 130 and 90 nm nodes, respectively. DUV light at 157 nm may be required for the 65 nm node, although recent advances in immersion lithography suggest that 193 nm immersion lithography will provide the dominant technology for the 65 nm node and may be extendable to the 45 nm node. 157 nm immersion lithography may be applicable down to almost the 32 nm node but this seems likely to form the practical limit of DUV lithography; new technologies will be required for future nodes. One possibility is extreme ultraviolet (EUV) lithography which employs wavelengths as short as 13 nm; EUV lithography is predicted to be implemented for the 32 nm node and should be extendable down to the 22 nm node (and possibly the 16 nm node). However, there are many difficult problems with this technology, including generation of the EUV radiation (possibilities include synchrotrons, which are very large and expensive, or laser

plasma sources, which are not yet a mature technology for reflective optics), the requirement that EUV is absorbed by all materials, preventing the use of conventional transmissive optics, and that the optics must be flat to within $\sim$0.1 nm to reduce aberrations to an acceptable level. The ultimate limit of photon-based lithography is X-ray proximity lithography which uses wavelengths of $\sim$1 nm. In contrast to DUV and EUV lithography, where the final image size is obtained from a larger mask size by focusing optics, the absence of suitable X-ray optics requires the use of a mask:image ratio of 1:1. Here the mask is placed slightly above the surface of the semiconductor wafer, projecting an image directly on to the photoresist. Because of the separation between the mask and the wafer, the minimum defined feature size is significantly greater than the radiation wavelength, although features of order 10 nm may be possible. Problems that need to be overcome with X-ray lithography include the generation of X-rays having sufficient intensity, the production of the 1:1 scale masks and the prevention of mask damage by the high-energy photons.

An alternative to photon lithography is electron beam (e-beam) lithography. The electron wavelength is controlled by the accelerating potential, resulting in the possibility of defining feature sizes below 10 nm. Conventional e-beam systems use a serial approach in which the pattern is written by a single electron beam scanned over the surface. However, this approach is too slow for mass volume production. Instead systems consisting of multiple beams or the projection of a broad beam through a suitable mask are being developed, although they are still a long way from commercial systems.

As the feature size becomes ever smaller, fluctuations due to the polymer unit size of the resist become significant. Such fluctuations, which are transferred to the fabricated devices, may impair carrier mobilities, requiring the development of new resists. The ultimate fabrication limit appears to be provided by AFM and related techniques, which allow individual atoms to be positioned on a surface. However, these techniques are relatively slow and it is difficult to see how they can be scaled up to allow mass volume production.

Even if increasingly smaller size MOSFETs can be fabricated, their electrical properties will eventually deviate significantly from those of larger devices. MOSFET technology has generally followed a constant field scaling, in which the physical dimensions and operating voltage are reduced by the same factor, with the substrate doping increased by a similar factor to maintain an unchanged electric field pattern within the channel. This scaling also requires the gate oxide thickness $d_{ox}$ to be reduced, and the relationship $L_{gate} \approx 45 d_{ox}$ is typically used. Although $SiO_2$ thicknesses of 1.5 nm or less can be formed (1.2 nm is used for the present 90 nm node), below $\sim$1.5 nm significant quantum mechanical tunnelling of carriers through the gate oxide occurs, and this leakage current adds to the power consumption of the system. A solution to this problem requires materials with a higher dielectric permittivity than $SiO_2$, allowing thicker gate insulating layers to be used. A short-term solution is provided by the use of SiON, which should provide oxide thicknesses equivalent to slightly less than 1 nm of $SiO_2$. In the longer term, more exotic materials are needed, such as $ZrO_2$ and $HfO_2$, although more development is required before commercial production is possible. Similar high-permittivity materials will also be required for future generations of dynamic random access memories (DRAMs). These devices are based on the storage of charge by arrays of tiny capacitors. As the capacitor size decreases, the use of a higher-permittivity material allows a given capacitance, and hence stored charge, to be maintained.

As the channel length is reduced, the channel potential distribution becomes strongly distorted in the vicinity of the drain region, due to the drain voltage. In addition, charge may leak between the drain and the source in the region away from the gate–channel interface, increasing the current in the off state of the transistor and greatly adding to the power consumption of the device. Modified structures are required to overcome these effects. Possibilities include devices with an underlying $SiO_2$ layer – Si on insulator (SOI) – or with material directly under the channel removed – silicon on nothing (SON). In addition, devices with dual gates and with either a horizontal or vertical channel are being investigated. The vertical channel structure has the advantage that the channel length is defined by an epitaxial growth step, rather than lithography, although the subsequent positioning of the dual gates is very difficult. Higher operating frequencies may also be obtained by increasing the mobility of carriers in the channel. This can be achieved by the use of strained Si or SiGe, both of which have a higher carrier mobility than unstrained Si. However, the use of these materials requires an additional, non-standard step in the fabrication process.

Other problems that will eventually arise as transistor sizes are reduced include limits to doping densities, resulting from the solubility limits of the dopant atoms, and the quantisation of energy levels and charge. Power consumption, which increases with increasing frequency but is also affected by current leakage, is a serious problem to be overcome. Current microprocessors generate heat with a power density equivalent to that of a hotplate. By 2010, at current trends, the power density will be comparable to that produced by a rocket nozzle! Power dissipation problems are worse still for devices fabricated on $SiO_2$ due to this material's poor thermal conductivity. The ultimate size limit of inorganic semiconductor devices may be single-electron transistors, discussed in the next section. Further size reductions are likely to require radically different approaches, such as transistors based on single molecules (Section 8.8.1.2).

Finally, as transistor switching frequencies increase, the speed with which signals propagate between different parts of a microprocessor must also increase. The present generation of microprocessors use copper interconnects, which give reduced propagation delays compared to aluminium which has been traditionally used. Surrounding the interconnect wires with insulators having a lower relative permittivity than the currently used fluorine-doped $SiO_2$ will reduce propagation delays further. In the longer term, critical direct electrical connections are likely to be replaced with optical or radio frequency interconnects.

At the time of writing (early 2004) state-of-the-art production is switching to the 90 nm node, based on a combination of 248 and 193 nm DUV lithography. This technology, which includes copper interconnects with low relative permittivity dielectrics, strained silicon and 1.2 nm thick gate oxide layers, is used for the production of microprocessors with 77 million transistors of gate lengths $\sim$50 nm and operating frequencies of 3.4 GHz. The same technology has been used to produce 512 Mbit static random access memory (SRAM) chips with 330 million transistors. At a research level, transistors with gate lengths as small as 10 nm have been fabricated, although these are of a conventional structure, so for gate lengths below $\sim$30 nm their performance is degraded due to the reasons discussed earlier. MOSFET operating frequencies in excess of 1 THz (1000 GHz) have been reported. Intel has recently announced a new transistor design that incorporates a $ZrO_2$ high-permittivity gate oxide, a thin Si channel on $SiO_2$ to reduce drain–source leakage, and modified drain and source contact regions to minimise series resistances. It is envisaged that this structure will eventually be suitable

for mass production of devices operating above 1 THz. Although these devices are still some way from mass production, they demonstrate that Si-based electronics will provide the dominant technology for the foreseeable future. If the trends described in the International Roadmap for Semiconductors 2003 are followed, then by 2016 microprocessors will contain approximately 9 billion transistors with 10 nm gate lengths and operating at a frequency of 28 GHz.

### 3.8.7 Coulomb blockade devices

In section 3.6.4 we described how the discrete nature of electronic charge, coupled with the small capacitance of a quantum dot, results in the ability to add and remove carriers being a function of the charged state of the dot. For the electrostatically defined Coulomb blockade device shown in Figure 3.30, the current between the reservoirs is controlled by the gate voltage in a manner similar to transistor action. However, in this device, transistor action occurs for the transport of only one electron at a time, although the number of electrons held on the dot may be significantly greater than one. The device can therefore be thought of as a single-electron transistor and, as such, represents the ultimate limit in transistor scaling given that a single electron represents the smallest possible unit of charge. In addition, Coulomb blockade also provides the possibility for single-electron memory cells. A number of schemes have been proposed but, in the simplest form, writing involves the transfer of an additional electron to the dot, with reading relying on the modification of the gate voltage by the presence of the additional electron.

Coulomb blockade effects may be observed in any conducting system of suitably small size, although silicon-based devices are required for compatibility with existing electronics. The observation of Coulomb blockade requires a charging energy considerably greater than the thermal energy, which for silicon-based devices operating at room temperature equates to a dot size of less than 10 nm. This size is well below what is achievable with present lithographic techniques, so devices fabricated in this way can only operate at low temperatures. Alternative fabrication possibilities include dots formed by thickness fluctuations in a thin silicon layer after treatment with an alkali-based solution, small silicon crystallites in a polysilicon layer and germanium self-assembled quantum dots. The first two techniques have both demonstrated room temperature memory operation, although the inherent randomness of the dot formation may make scaling up to large arrays difficult.

For the electrostatically defined dot shown schematically in the inset to Figure 3.30 it is possible to separately control the heights of the two tunnelling barriers via their defining gate voltages. If the height of the left-hand barrier is initially set low, with the height of the right-hand barrier set high, a single electron may tunnel on to the dot but is prevented from leaving the dot. In addition, a second electron is prevented from tunnelling on to the dot because the first electron raises the potential of the dot. If the heights of the barriers are now reversed, the electron can tunnel out of the dot, with the overall effect of these two steps being the transfer of one electron between the reservoirs. This process can be repeated continuously by applying AC voltages, with a suitable phase shift, to the gates. If the applied frequency is $f$ then $f$ electrons per second will move between the reservoirs, giving a current $I = fe$. Such a device has potential as a current standard.

### 3.8.8 Photonic structures

The nanostructures discussed so far modify the electronic properties of the underlying semiconductor. However, in structures or devices that produce or detect light there is also interest in modifying the properties of the photons by creating a photonic structure. The simplest photonic structure uses a one-dimensional optical cavity to confine photons, and this can be combined with a quantum well so that both photons and electrons are confined. An optical cavity is created with two high-reflectivity parallel mirrors, separated by a multiple of the wavelength of the photons to be confined. In a semiconductor system the mirrors are formed by growing a Bragg stack, which consists of a repeated sequence of alternating semiconductor layers. The refractive index change between the semiconductors results in only a small reflectivity, however this is enhanced by the repeated nature of the stack if the thickness of each layer equals a quarter of the wavelength of the photons. By using a Bragg stack consisting of $\gtrsim$20 layers, a reflectivity $\gtrsim$99% is possible. A similar structure is used to form vertical cavity surface-emitting laser (VCSEL) devices (Section 3.8.1).

The inset to Figure 3.51 shows a schematic diagram of a one-dimensional optical cavity, known as a microcavity. Two Bragg stacks and a GaAs layer of thickness equal to the photon wavelength form the optical cavity, which confines photons travelling along the growth direction. A quantum well placed at the centre of the cavity provides confinement of the excitons. The curves in Figure 3.51 show the unperturbed energies of the confined exciton and photon as a function of temperature, which is used to vary their relative energy. The energy of the exciton, which is related to the band gap of the semiconductor, is relatively temperature dependent; the energy of the photon, given by the thickness of the cavity, is less temperature dependent. At resonance the exciton and photon couple together to form a state known as a polariton. This coupling is demonstrated by the experimental data points, which deviate from the calculated, unperturbed

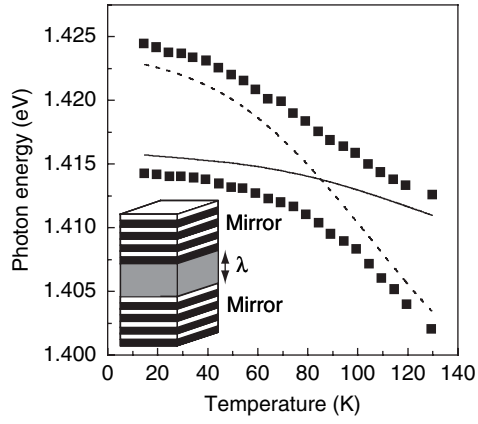


**Figure 3.51** Solid symbols show the temperature dependence of the energies of the quantum well exciton and cavity mode (photon) of a GaAs–AlGaAs microcavity, which exhibit an anti-crossing around 90 K. The lines show the calculated exciton (dashed line) and cavity (solid line) energies in the absence of coupling. The inset shows the physical structure of the microcavity. Data courtesy of Adam Armitage, University of Sheffield

energies to produce an anti-crossing of the exciton and photon states. This mixed exciton–photon polariton state has a number of novel physical properties that result from an in-plane wave-vector energy dependence which is very different from that of the uncoupled photons or excitons. Possible applications include extremely low threshold lasers. Additional applications result from the use of microcavities with very small diameters ($\lesssim 1\,\mu m$). These allow the radiative lifetime of the exciton to be decreased and also increase the fraction of photons that are able to escape from the device. Microcavities containing single quantum dots may provide the efficiency required for practical single-photon sources, as discussed in Section 3.8.3.

Confinement of photons in more than one dimension may be achieved by the use of a periodic refractive index array. A two-dimensional structure can be created by the use of electron beam or DUV lithography to define a periodic pattern on a semiconductor wafer, which is subsequently etched to form a series of pillars and holes. The periodic modulation of the refractive index produces a photonic band structure, with bands of allowed photon states separated by band gaps, in a similar manner to the electronic band structure of a solid which arises from the periodic arrangement of the atoms. Figure 3.52 shows images of a two-dimensional photonic structure that consists of periodic arrays of holes and a series of unpatterned regions, which act as ultrasmall cavities within which photons are confined by the surrounding periodic structure. Photonic structures allow the direction of light propagation to be controlled (bending around corners is possible), permit the control of spontaneously emitted light (e.g.,

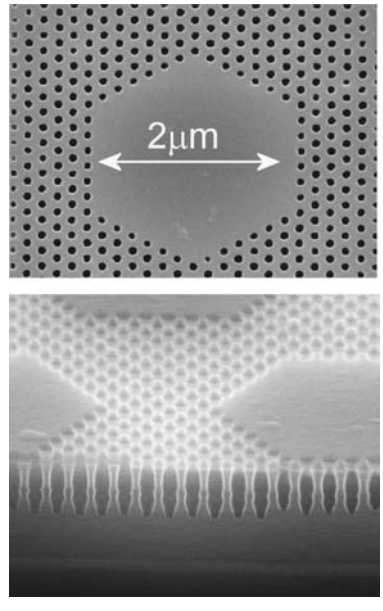


**Figure 3.52** Two-dimensional photonic structure consisting of periodic arrays of holes and a series of unpatterned regions which act as ultra-small cavities within which photons are confined by the surrounding photonic structure. The upper image is recorded via the upper surface of the structure, the lower image is of a cleaved edge. Data courtesy of Alan Bristow, University of Sheffield

suppressing it if it occurs within a band gap), provide strong non-linear effects, modify the dispersion of light (including reducing the propagation speed), and allow the fabrication of very small lasers.

The structure shown in Figure 3.52, an example of a micro laser, provides photon confinement along the growth direction by using a layer of a large refractive index semiconductor sandwiched between lower refractive index layers, forming a planar waveguide. The extension to true three-dimensional periodic structures is desirable, but is difficult to achieve using lithographic techniques. Although photonic structures may be fabricated in a range of materials, semiconductor-based structures can be directly integrated with conventional electro-optical devices, and quantum dots may be incorporated directly into the structure to provide efficient light-emitting centres.

## 3.9 SUMMARY AND OUTLOOK

The importance of inorganic semiconductor nanostructures arises from the opportunities they provide for the controllable modification of the electronic and optical properties of the underlying semiconductors. Consequently, they are of interest for both the study of novel physical phenomena in reduced dimensionality systems, and for a wide range of electronic and electro-optical device applications. Quantum wells are now used in the majority of semiconductor lasers and modulation doping is used in specialist transistors, particularly those required for high-frequency or low-noise applications. The quantum Hall effect provides a resistance standard. It is now possible to purchase lasers based on self-assembled quantum dots and also colloidal quantum dots for biological tagging. Quantum cascade lasers are likely to be used in commercial systems within the next few years.

However, many applications require further developments, particularly in fabrication techniques. For example, it is desirable to increase energy level separations for high-temperature applications, particularly for holes which generally have a larger effective mass. This requires a reduction in the size of nanostructures. Greater control of the self-assembly technique is required to obtain improved uniformity and to be able to position quantum dots at specific positions, possibly by growing on prepatterned surfaces. For some applications the presence of the wetting layer is undesirable and techniques for its removal need to be developed. The ability to fabricate nanostructures emitting or absorbing light at a range of specific wavelengths, from the infrared through the visible to ultraviolet spectral regions, will open up a number of applications. Over the past few years, increasing progress has been made in the study of single nanostructures, and this should lead to a number of applications based on individual quantum dots, including single-electron transistors and memories, and single-photon sources. There is considerable interest in using the spin properties of electrons rather than their charge, a field referred to as spintronics. Nanostructures allow the spins of electrons and holes to be manipulated, and mechanisms which relax the spin may be inhibited in quantum dots, resulting in extremely long spin coherence times. Integration with other areas of nanotechnology is likely to occur, such as the inclusion of a small number of magnetic ions in a quantum dot to give a magneto-optical zero-dimensional structure, and the combination of Si-based electronics with polymer or biological systems. Finally,

a convergence will occur between traditional technologies used for the manufacture of microprocessors, memories and related electronic systems and recently developed nanotechnologies, driven by the continued reduction of device dimensions in state-of-the-art electronic circuits.

## BIBLIOGRAPHY

Basic semiconductor properties are covered in many solid-state physics textbooks, for example *Introduction to Solid State Physics* (7th edn) by C. Kittel (Wiley, Chichester, 1996) and *Solid State Physics* by J. R. Hook and H. E. Hall (Wiley, Chichester, 1991). Textbooks dealing specifically with semiconductors include *Fundamentals of Semiconductors* by P. Y. Yu and M. Cardona (Springer, Berlin, 2001) and *The Physics of Semiconductors with Applications to Optoelectronic Devices* by K. F. Brennan (CUP, Cambridge, 1999). Brennan's book also covers the operation of a number of conventional semiconductor electronic and electro-optical devices, as do *Semiconductor Devices: Basic Principles* by J. Singh (Wiley, Chichester, 2001) and *Semiconductor Optoelectronic Devices* (2nd edn) by P. Bhattacharya (Prentice Hall, New York, 1997). Two recent textbooks, *Band Theory and Electronic Properties of Solids* by J. Singleton (OUP, Oxford, 2001) and *Optical Properties of Solids* by A. M. Fox (OUP, Oxford, 2001), provide good coverage of the electronic and optical properties of bulk semiconductors and semiconductor nanostructures. Singleton's book contains a clear discussion of the quantum Hall effect. *Semiconductor Optics* by C. F. Klingshirn (Springer, Berlin, 1997) provides a comprehensive discussion of the optical properties of semiconductors with a more limited discussion of quantum well structures.

A number of books deal specifically with various aspects of inorganic semiconductor nanostructures. General texts include *Physics of Semiconductors and their Heterostructures* by J. Singh (McGraw-Hill, New York, 1993), *Quantum Semiconductor Structures: Fundamentals and Applications* by C. Weisbuch and B. Vinter (Academic Press, London, 1991) and *Low-Dimensional Semiconductors: Materials, Physics, Technology, Devices* by M. J. Kelly (OUP, Oxford, 1995). Kelly's book contains a good discussion of a range of device applications. The self-assembly technique for the fabrication of quantum dots is covered in considerable depth in *Quantum Dot Heterostructures* by D. Bimberg, M. Grundmann and N. N. Ledentsov (Wiley, Chichester, 1999) and *Heterojunction Band Discontinuities: Physics and Device Applications* edited by F. Capasso and G. Margaritondo (North-Holland, Amsterdam, 1987) covers the calculation and measurement of band offsets, in addition to the general properties and applications of quantum well systems. The *Physics of Low-Dimensional Semiconductors: An Introduction* by J. H. Davies (CUP, Cambridge, 1998), *Wave Mechanics Applied to Semiconductor Heterostructures* by G. Bastard (Halsted Press, Paris, 1988) and *Quantum Wells, Wires and Dots: Theoretical and Computational Physics* by P Harrison (Wiley, Chichester, 1999) all provide a mathematically based treatment of inorganic semiconductor nanostructures.

*Nanoelectronics and Information Technology: Advanced Electronic Materials and Devices* edited by R. Waser (Wiley VCH, Weinheim, 2003) covers the present status of silicon MOSFETs, memory devices and lithographic techniques, and discusses possible future advances and alternative, non-silicon-based technologies. Information concerning

future trends in microprocessor development can be obtained from the International Technology Roadmap for Semiconductors. This is updated yearly and the current edition can be found at the Semiconductor Manufacturing Technology (SEMATECH) website, www.sematech.org/. Finally, the silicon research section of the Intel website, www.intel.-com/research/silicon/, contains many articles relevant to current and future microprocessor development.