Metal-Insulator-Semiconductor and Metal-Insulator-Metal Structures. Part VI. Quantum-Effect Devices. Electron Bands and Densities of States

Alexander Gabovich, **KPI**,

Lecture 13

Tunnel diode

Tunneling of electrons through a potential barrier is an effect predicted by quantum mechanics that gives the electrons a finite probability of passing through the barrier, as opposed to the electrons needing an energy greater than the barrier potential energy to overcome it.

To illustrate this effect, let us take an infinite potential well and introduce a finite potential barrier in it (Figure 10.1A). The wave function of an electron in this potential well can be calculated using numerical simulations (see Problems 1.3 and 1.4). Let us focus on the lowest or ground-state energy level. In the absence of a potential barrier the lowest energy of an electron can be found using Equation 1.1.11: $E = \frac{\pi^2 \hbar^2}{2ma^2}$. For a well width of 50 nm the corresponding lowest energy value is approximately 0.15 meV. Let us introduce a potential barrier 40 mV in height and 2 nm in width inside the potential well. According to classical mechanics an electron confined in the left-hand side of the potential well does not possess enough energy to overcome the 40-mV potential barrier and venture into the right part of the well. If the calculation is made using quantum mechanics, on the other hand, one finds that there is a non-zero probability of finding the electron at the right of the potential barrier, as shown in Figure 10.1B.

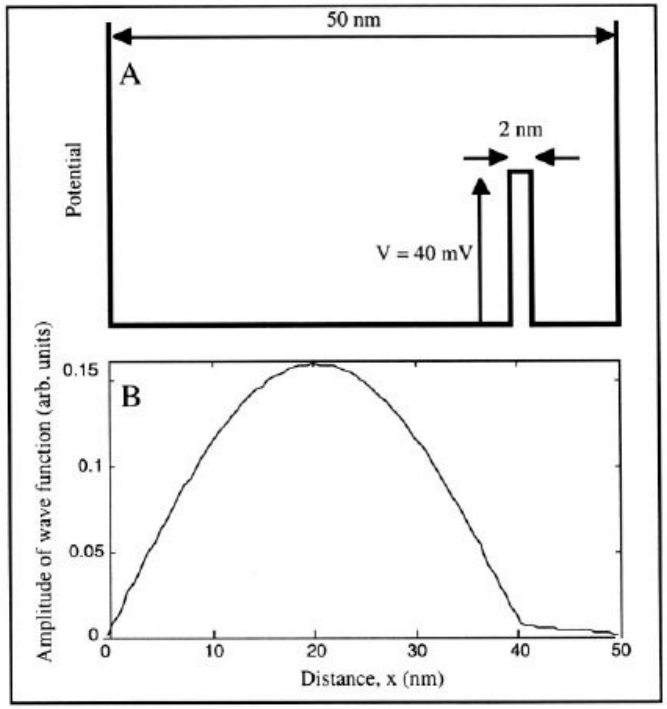


Figure 10.1: A: Infinite potential well with a potential barrier inside it; B: Corresponding lowest-energy wave function.

Tunnel diode

In a more general sense, tunneling through a potential barrier can be characterized by a transmission coefficient which represents the probability of an electron passing through the barrier. The value of this transmission coefficient depends on the shape of the barrier (rectangular, triangular, etc.), on its width and its height. The thinner and the lower the barrier, the higher the transmission coefficient. In the particular case of a rectangular barrier, the transmission coefficient, T, is given by:

$$T = \frac{1}{1 + \frac{1}{4}\frac{V^2}{E(V-E)}\sinh^2\left(\frac{a}{\hbar}\sqrt{2m(V-E)}\right)}$$

where a and V are the width and the height of the potential barrier, respectively, and E is the energy of the electron ($E<V$).[2]

A tunnel diode is a PN junction where both P- and N-type regions are degenerately doped. As a result, the Fermi level in the N-type material is above the minimum of the conduction band and the Fermi level in the P-type material is below the maximum of the valence band. The doping concentrations are so high that the width of the space-charge region at the junction is extremely thin (Equation 4.2.12), and usually measures less than 10 nm.

As in any PN junction the existence of a space-charge region gives rise to a potential barrier. This barrier height is noted Φ_o which is a function of the doping concentrations according to Equation 4.2.9. The barrier prevents electrons from diffusing from the N-type region into the P-type material and vice-versa. Φ_o is relatively large because of the doping levels, but the width of the barrier is very small ($\leq$ 10 nm).

Tunnel diode

In order for electrons to tunnel through the potential barrier certain conditions must be met:

1- The energy of the electron must be conserved. In terms of an energy band diagram representation, this condition means that an electron tunneling from the N-type region into the P-type region must do so in a horizontal trajectory (Figure 10.2B).

2- There must be occupied states on the side of the junction that emits electrons.

3- There must be empty permitted states on the side of the junction which receives the electrons. Because of condition (1), these states must have the same energy as the states defined in (2).

4- The potential barrier height must be low enough and its width must be small enough for tunneling to take place.

The electron current from the N-type conduction band into the P-type valence band is given by:

$$I_{c \to v} = A \int F_c(E)\, n_c(E)\, T_t\, [1 - F_v(E)]\, n_v(E)\, dE \qquad (10.1.1)$$

where A is the area of the diode, $F_c(E)$ and $F_v(E)$ are the Fermi-Dirac distribution functions in the N-type conduction band and the P-type valence band, respectively, $n_c(E)$ and $n_v(E)$ are the density of states in the conduction and valence band, and T_t is the tunneling probability of an electron. This probability depends essentially on the width of the potential barrier, and it is independent of the direction of the electron (left to right or right to left).

Tunnel diode

The positive sign of the current is due to the fact that electrons carry a negative charge and flow in the negative x-direction (Figure 10.2). The current due to the electron flow from the N-type conduction band into the P-type valence band is equal to:

$$I_{v\to c} = -A \int F_v(E)\, n_v(E)\, T_t\, [1 - F_c(E)]\, n_c(E)\, dE \qquad (10.1.2)$$

The total current is obtained by adding 10.1.1 and 10.1.2:

$$I_t = I_{c\to v} + I_{v\to c} = A \int T_t\, [F_c(E) - F_v(E)]\, n_v(E)\, n_c(E)\, dE \qquad (10.1.3)$$

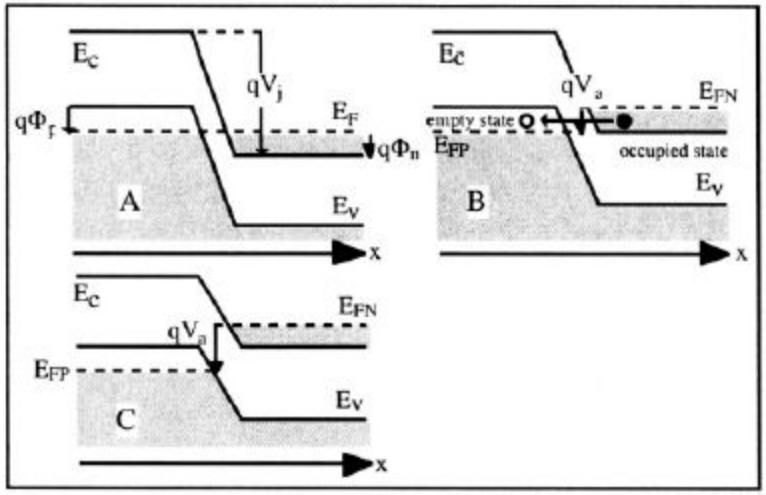


Figure 10.2: Energy band diagram for different increasing forward bias values. A: Zero applied bias; B: Maximum tunneling current; C: Tunneling current has vanished. The shaded areas represent states filled by electrons. [3] V_j is the built-in junction potential and V_a is the external applied voltage.

Tunnel diode

Calculating the tunnel current is relatively complex. We will only describe qualitatively what happens using the energy band diagrams of Figure 10.2.

A: Let us start with a zero applied bias. In that case $F_v(E)$ and $F_c(E)$ are equal because the Fermi level, E_F, is unique, and the tunneling current is equal to zero, according to Equation 10.1.3 - Figure 10.3.A.

B: If a forward bias, V_a, is applied the quasi-Fermi level and the energy bands in the N-type region move up with respect to the P-type region. As a result there are empty states in the P-side valence band which have the same energy as occupied states in the N-side conduction band. This condition allows for a tunneling current $I_{c \to v}$ to take place. This current increases with increased applied bias, V_a, until a maximum is reached. The maximum current occurs when the number of states in the N-conduction band having the same energy as empty states in the P-valence band is maximum (Figure 10.3.B).

C: If the applied bias, V_a, is further increased the number of empty valence states having the same energy as occupied conduction states decreases until the tunneling current eventually vanishes. A "valley" point of the I-V characteristics is reached when tunneling ceases (Figure 10.3.C).

D: In addition to the band-to-band tunneling current a "regular" PN junction current flows through the diode. As the forward bias is increased the current will increase again, as in a regular PN junction diode (Figure 10.3.D). In the part of the curve between the peak and the valley the tunnel diode has a negative resistance characteristics ($R = dV/dI < 0$).

Tunnel diode

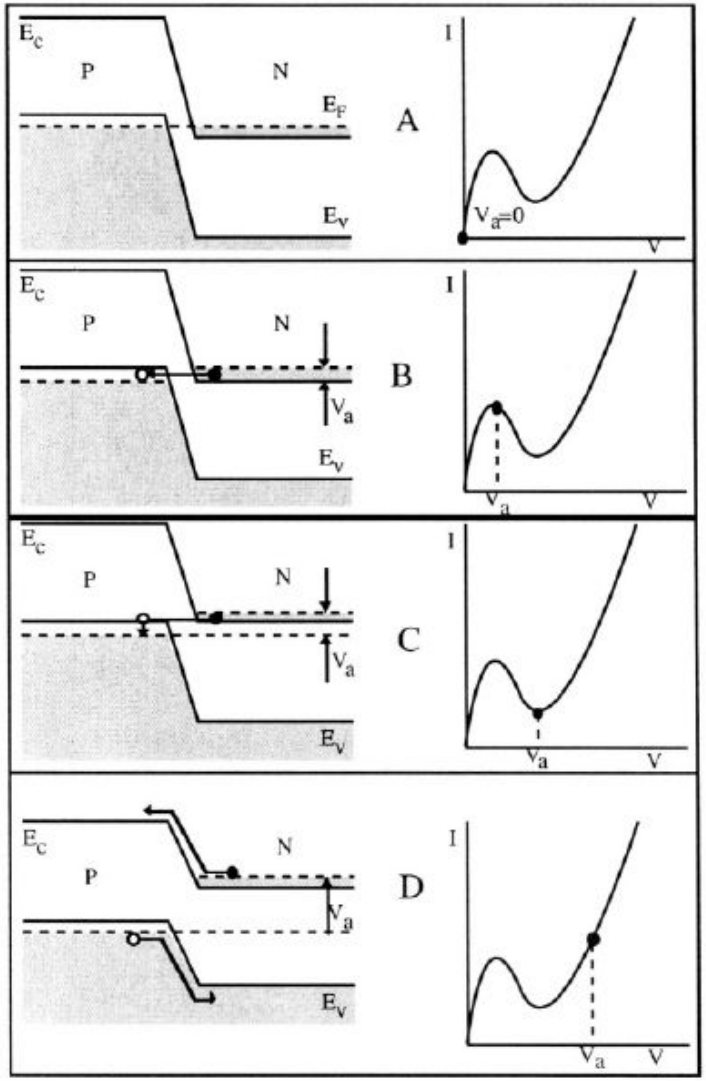


Figure 10.3: Energy band diagram I-V characteristics of a tunnel diode; A: V_a=0; B: peak tunneling current; C: valley current where tunneling ceases; D: "regular" PN junction diffusion current.

Tunnel diode

The tunnel diode is associated with the quantum tunneling phenomena.[5] The tunneling time across the device is very short, permitting its use well into the millimeter-wave region. Because of its mature technology, the tunnel diode is used in special low-power microwave applications, such as local oscillators and frequency-locking circuits.

A tunnel diode consists of a simple p–n junction in which both the p- and n-sides are degenerate (i.e., very heavily doped with impurities). Figure 1 shows a typical static current-voltage characteristic of a tunnel diode under four different bias conditions. The I-V characteristic is the result of two current components: tunneling current and thermal current.

When no voltage is applied to the diode, it is in thermal equilibrium ($V = 0$). Because of the high dopings, the depletion region is very narrow and the tunneling distance d is quite small (5–10 nm). The dopings also cause the Fermi levels to be located within the allowed bands. The amount of degeneracy, qV_p and qV_n, shown at the far left of Fig. 1, is typically 50–200 meV.

When a forward bias is applied, there exists a band of energy states that is occupied on the n side and a corresponding band of energy states that is available and unoccupied on the p side. The electrons can tunnel from the n-side to the p-side. When the applied bias equals approximately $(V_p + V_n)/3$, the tunneling current reaches its peak value I_P and the corresponding voltage is called the peak voltage V_P. When the forward voltage is further increased, there are fewer available unoccupied states on the p-side ($V_P < V < V_V$, where V_V is the valley voltage) and the current decreases. Eventually, the band is "uncrossed," and at this point the tunneling current can no longer flow. With still further voltage increase, the normal thermal current will flow (for $V > V_V$).

From this discussion we expect that in the forward direction the tunneling current increases from zero to a peak current I_P as the voltage increases. With a further increase in voltage, the current then decreases to zero when $V = V_n + V_p$, where V is the applied forward voltage. The decreasing portion after the peak current in Fig. 1 is the negative differential resistance region. The values of the peak current I_P and the valley current I_V determine the magnitude of the negative resistance. For this reason their ratio I_P/I_V is used as a figure of merit for the tunnel diode.

Tunnel diode

An empirical form for the *I-V* characteristics is given by

$$I = I_P\left(\frac{V}{V_P}\right)\exp\left(1-\frac{V}{V_P}\right)+I_0\exp\left(\frac{qV}{kT}\right), \tag{1}$$

where the first term is the tunnel current and I_P and V_P are the peak current and peak voltage, respectively, as shown in Fig. 1. The second term is the normal thermal current. The negative differential resistance can be obtained from the first term in Eq. 1:

$$R=\left(\frac{dI}{dV}\right)^{-1}=-\left[\left(\frac{V}{V_P}-1\right)\frac{I_P}{V_P}\exp\left(1-\frac{V}{V_P}\right)\right]^{-1}. \tag{2}$$

Figure 2 shows a comparison of the typical current-voltage characteristics of Ge, GaSb, and GaAs tunnel diodes at room temperature. The current ratios of I_P/I_V are 8:1 for Ge and 12:1 for GaSb and GaAs. Because of its smaller effective mass (0.042 m_0) and small bandgap (0.72 eV), the GaSb tunnel diode has the largest negative resistance of the three devices.

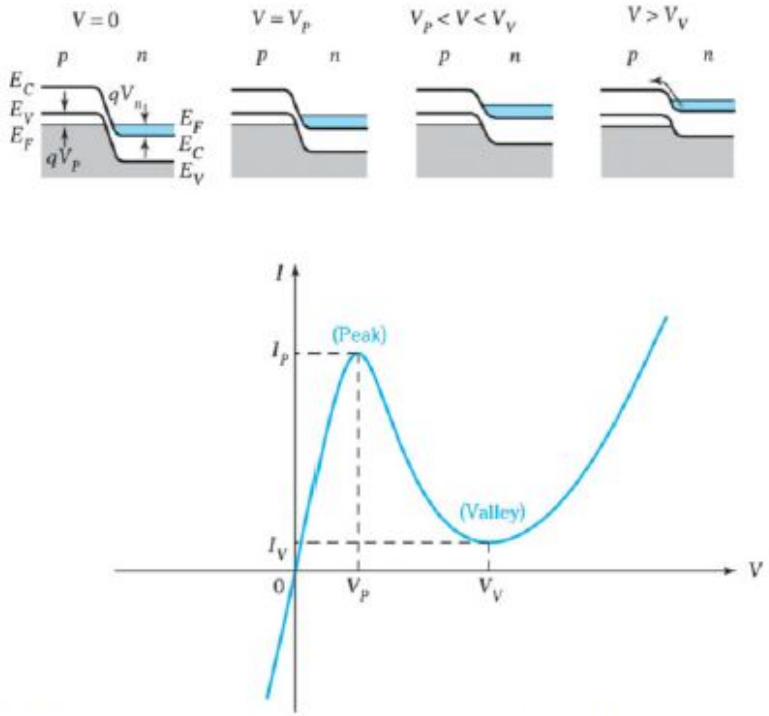


Fig. 1 Static current-voltage characteristics of a typical tunnel diode. I_P and V_P are the peak current and peak voltage, respectively. I_V and V_V are the valley current and valley voltage, respectively. The upper figures show the band diagrams of the device at different bias voltages.

Tunnel diode

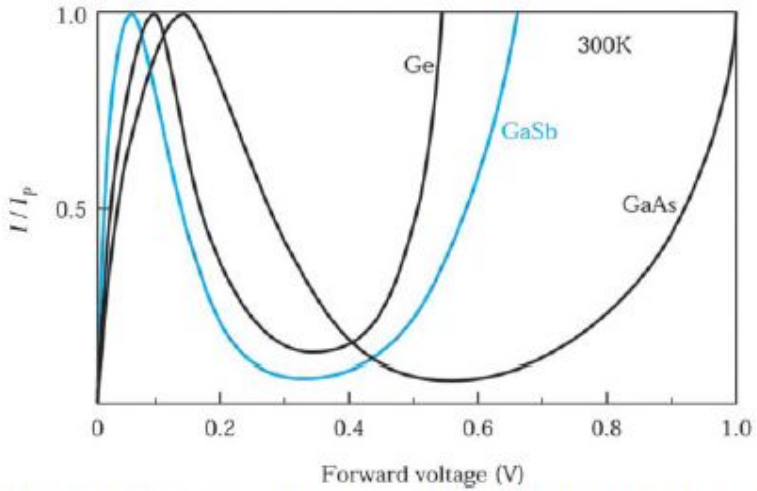


Fig. 2 **Typical current-voltage characteristics of Ge, GaSb, and GaAs tunnel diodes at room temperature.**

Leo Esaki

Low-dimensional devices

In a low-dimensional device carriers are no longer moving in a three-dimensional crystal, but they are confined within a two-, one- or zero-dimensional space. This is realized by fabricating devices where carriers are confined within a thin crystal, such as a quantum wire, or in a low-dimensional potential well, such as a quantum-well device.

In the case of a three-dimensional (3D) crystal the density of allowed states in an energy band is a square root function of the energy, as demonstrated in Section 1.1.8 and shown in Figure 10.4.

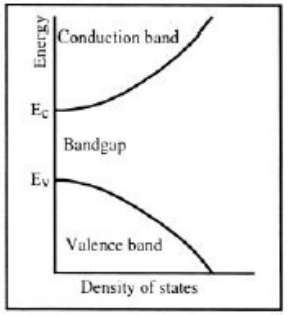


Figure 10.4: Density of states in the conduction and valence band near bandgap in a 3D semiconductor.

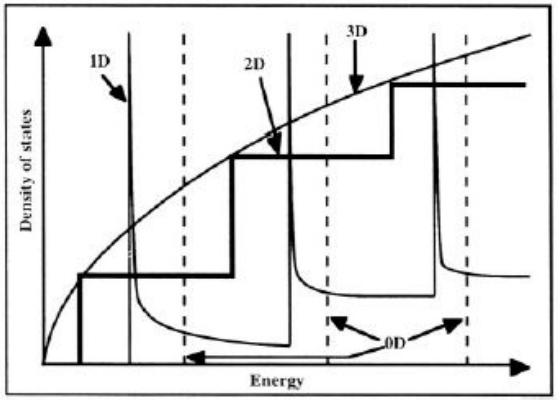


Figure 10.5: Density of states in zero- (0D), one- (1D), two- (2D) and three-dimensional (3D) crystals. [4]

Low-dimensional devices

In the case of low-dimensional structures the energy bands, and in particular the distribution of permitted states, is quite different from that of a 3D crystal (Figure 10.5). In a zero-dimensional (0D) crystal (also called "quantum dot") the permitted energy levels are discrete. In a one-dimensional (1D) crystal (also called "quantum wire") they are basically also discrete, but tend to spread out between the "quantized" levels. In a two-dimensional (2D) crystal the density of states is a staircase function of the energy. Figure 10.6 shows the different geometries (3, 2, 1 and 0-D samples) which correspond to the densities of stated in Figure 10.5.

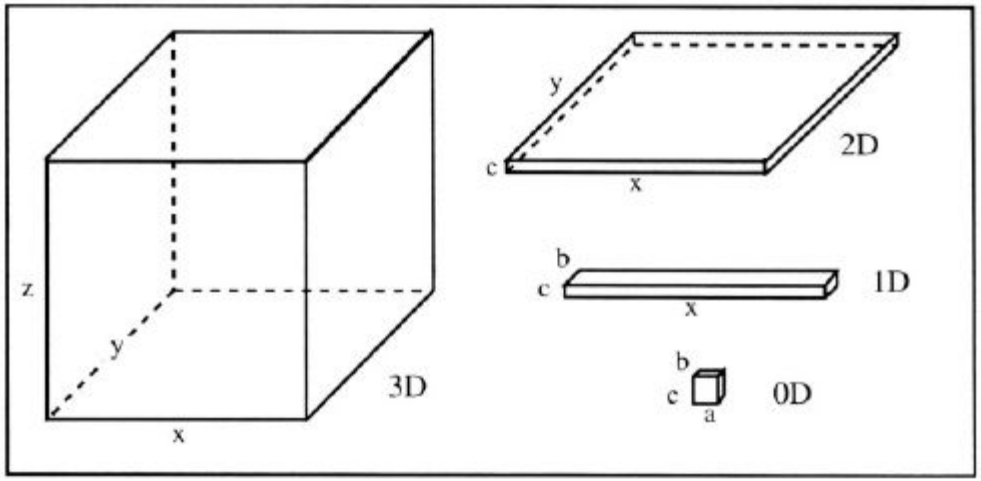


Figure 10.6: Geometry of 3D, 2D, 1D and 0D samples. x, y, and z represent spatial directions and a, b, and c represent small dimensions in the x, y, and z direction, respectively. [5,6,7]

Low-dimensional devices. Energy bands

The energy band calculations are based on the time-independent Schrödinger equation:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(x,y,z) + V(x,y,z)\,\Psi(x,y,z) = E\,\Psi(x,y,z) \qquad (10.2.1)$$

which can be re-written, if $r = (x,y,z)$:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(r) + V(r)\,\Psi(r) = E\,\Psi(r) \qquad (10.2.2)$$

We have solved this equation in Section 1.1.3 using the Krönig-Penney model. In the case of a *three-dimensional crystal* we have seen that near the bottom of the conduction band the energy of the electron as a function of the *k*-vector is parabolic, and behaves approximately as a free electron. In that case the periodic potential variation in the crystal can be neglected and one obtains:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi(r) = E\,\Psi(r)$$

The solution to the latter equation is $\Psi(r) = A\,exp(jkr)$, from which the energy can be found:

$$\frac{\hbar^2k^2}{2m}\Psi(r) = E\,\Psi(r) \Rightarrow \frac{\hbar^2k^2}{2m} = E_k \qquad (10.2.3)$$

The k-vectors for a 3D sample can be found by imposing the Born-von Karman boundary conditions (Expression 1.1.13): $\Psi(x,y,z) = \Psi(x+NL,y,z)$, $\Psi(x,y,z) = \Psi(x,y+NL,z)$ *and* $\Psi(x,y,z) = \Psi(x,y,z+NL)$ where L is the size (length) of the crystal unit cell, and N is the number of such cells in each direction of space. If the crystal has a cubic lattice and has a cubic shape, each dimension of the crystal is equal to NL and one obtains:

$$k_x = \frac{2\pi}{NL}n_x,\ k_y = \frac{2\pi}{NL}n_y,\ k_z = \frac{2\pi}{NL}n_z \qquad (10.2.4)$$

The unit volume in k-space corresponding to each permitted k value is:

$$3D\ unit\ volume = \left(\frac{2\pi}{NL}\right)^3 = \frac{8\pi^3}{V} \qquad (10.2.5)$$

where V is the crystal volume.

Low-dimensional devices. Energy bands

Using equation (1.1.31) we obtain the values of the permitted wave number:

$$k = \frac{2\pi n}{NL} \quad (n=0,\pm1,\pm2,\pm3,\ldots,\pm(N-1)/2,\ +N/2)$$

where N is the number of crystal cells (about 10^{22} per cubic centimeter). The number of permitted k values is, therefore, very large, and one can consider that k does not vary in a discrete manner, but in a continuous way. Finally the permitted energy levels in a three-dimensional crystal are given by:

$$E_k = \frac{\hbar^2}{2m}\left[\left(\frac{2\pi n_x}{NL}\right)^2 + \left(\frac{2\pi n_y}{NL}\right)^2 + \left(\frac{2\pi n_z}{NL}\right)^2\right] \qquad (10.2.6)$$

If we now reduce the size of crystal in the in the z-direction to a very small value, c, we obtain a *two-dimensional crystal* (Figure 10.6). The wave functions in the z-direction are confined within an infinite potential well having a width, c, which is equal to the sample thickness. In the z-direction the wave function is finite inside the crystal and it is equal to zero outside it. Using the technique of separation of variables the wave function can be written as the product of two wave functions:

$$\Psi(x,y,z) = A\ exp(jkr)\ \Phi_c(z) \qquad (10.2.7)$$

with $r = (x,y)$. In the z-direction the electron behaves like a "particle in a box" in an infinite potential well of width c. From Section 1.1.1.2 we know that the equation to be solved is:

$$-\frac{\hbar^2}{2m}\frac{d^2\Phi_c(z)}{dz^2} = E\ \Phi_c(z) \qquad (10.2.8)$$

and that its solution is:

$$\Phi_c(z) = A\ exp(jkz) + B\ exp(-jkz) \qquad (10.2.9)$$

Low-dimensional devices. Energy bands

Using the boundary conditions of vanishing wave function at the sides of the crystal $\Phi_c(0) = 0$ and $\Phi_c(c) = 0$ we obtain:

$$\Phi_c(z) = C \sin\left(\frac{\pi n_z z}{c}\right)$$

with $n_z = 1, 2, 3, \ldots$ (10.2.10)

The energy values in the z-direction can then be extracted:

$$-\frac{\hbar^2}{2m}\frac{d^2\left(C \sin\left(\frac{\pi n_z z}{c}\right)\right)}{dz^2} = E_z\, C \sin\left(\frac{\pi n_z z}{c}\right) \Rightarrow E_z = \frac{\hbar^2}{2m}\left(\frac{\pi n_z}{c}\right)^2 \quad (10.2.11)$$

The permitted energy levels (eigenvalues) for the electrons in the crystal can be obtained by summing the energy levels in the z-direction and the energy levels for $r = (x,y)$:

$$E_{ck} = \frac{\hbar^2}{2m}\left[\left(\frac{2\pi n_x}{NL}\right)^2 + \left(\frac{2\pi n_y}{NL}\right)^2 + \left(\frac{\pi n_z}{c}\right)^2\right]$$

which can also be written:

$$E_{ck} = E_c + \frac{\hbar^2}{2m}\left[\left(\frac{2\pi n_x}{NL}\right)^2 + \left(\frac{2\pi n_y}{NL}\right)^2\right] \quad (10.2.12)$$

where: $E_c = \frac{\hbar^2}{2m}\left(\frac{\pi n_z}{c}\right)^2$. The volume of the crystal is $V = c(NL)^2$. The 2D unit volume in k-space corresponding to each permitted k value in the sample is:

$$2D\ unit\ volume = \left(\frac{2\pi}{NL}\right)^2 = \frac{4\pi^2 c}{V} \quad (10.2.13)$$

Low-dimensional devices. Energy bands

The permitted energy values are obtained by adding the energy levels which are a function of k_x and k_y and a series of discrete energy levels produced by the wave function confinement in the z-direction. For each discrete energy level resulting from the confinement, $E_c(n_z)$, there exists a 2D energy band corresponding to the possible k_x and k_y values. Such an energy band is called an energy subband (Figure 10.7). It is worth noting that the minimum energy of the electron, which was equal to zero in the three-dimensional case (when $n_x=n_y=n_z=0$ in (10.2.6)) is now equal to:

$$\frac{\hbar^2}{2m}\left(\frac{\pi}{c}\right)^2 \neq 0 \text{ (for } n_z = 1).$$

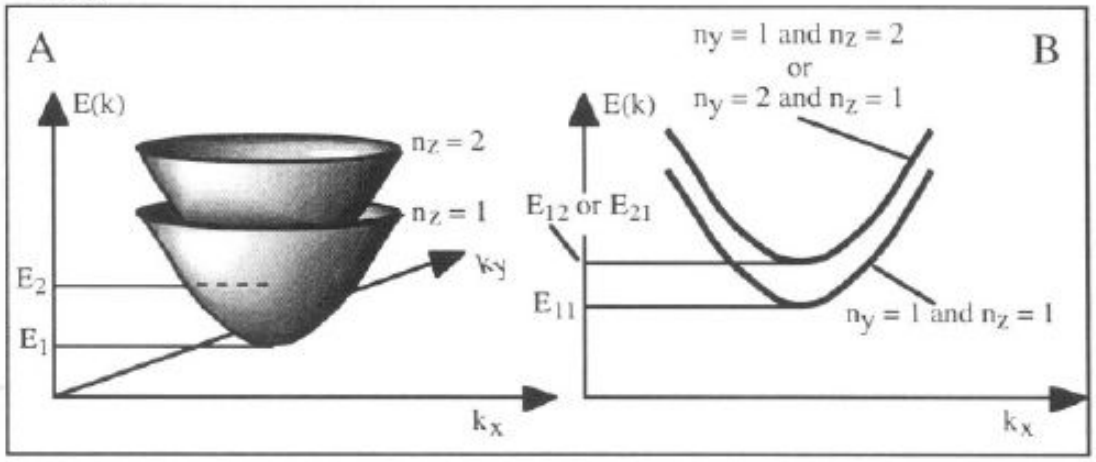


Figure 10.7: Energy vs. wave vector or wave number in a (A) two-dimensional and (B) one-dimensional semiconductor. The width and height of the 1D crystal are taken equal ($b=c$). Two subbands are shown for each sample.

Low-dimensional devices. Energy bands

In the case of a *one-dimensional crystal* the dimensions in both the y and z directions of the sample are very small as shown in Figure 10.7. The width of the crystal is noted "b" and its height is noted "c". The wave functions are now confined in both the y and z directions. Using the technique of separation of variables the wave function can be written as the product of two wave functions:

$$\Psi(x,y,z) = A\, exp(jkx)\, \Phi_{bc}(y,z) \qquad (10.2.14)$$

The wave function in the directions of confinement, $\Phi_{bc}(y,z)$, corresponds to that of a particle in two-dimensional infinite potential of width "b" and height "c". The wave function can be found using the Schrödinger equation adapted to this particular geometry:

$$-\frac{\hbar^2}{2m}\left(\frac{d^2\Phi_{bc}(y,z)}{dy^2}+\frac{d^2\Phi_{bc}(y,z)}{dz^2}\right) = E\,\Phi_{bc}(y,z) \qquad (10.2.15)$$

which has the solution:

$$\begin{aligned}\Phi_{bc}(y,z) &= \Phi_b(y)\,\Phi_c(z)\\ \Phi_b(y) &= A\,exp(jky) + B\,exp(-jky)\\ \Phi_c(z) &= C\,exp(jkz) + D\,exp(-jkz)\end{aligned} \qquad (10.2.16)$$

Using the boundary conditions $\Phi_b(0) = 0$, $\Phi_c(0) = 0$, $\Phi_b(b) = 0$ and $\Phi_c(c) = 0$, one obtains:

$$\Phi_b(y) = F\sin\left(\frac{\pi n_y y}{b}\right) \text{ and } \Phi_c(z) = G\sin\left(\frac{\pi n_z z}{c}\right)$$

$$\text{with } n_y = 1, 2, 3,\ldots \text{ and } n_z = 1, 2, 3,\ldots \qquad (10.2.17)$$

Low-dimensional devices. Energy bands

The permitted energy levels can then be found:

$$-\frac{\hbar^2}{2m}\frac{d^2\left(F\sin\left(\frac{\pi n_y y}{b}\right)\right)}{dy^2} = E_y F\sin\left(\frac{\pi n_y y}{b}\right) \Rightarrow E_y = \frac{\hbar^2}{2m}\left(\frac{\pi n_y}{b}\right)^2$$

$$-\frac{\hbar^2}{2m}\frac{d^2\left(G\sin\left(\frac{\pi n_z z}{c}\right)\right)}{dz^2} = E_z G\sin\left(\frac{\pi n_z z}{c}\right) \Rightarrow E_z = \frac{\hbar^2}{2m}\left(\frac{\pi n_z}{c}\right)^2 \quad (10.2.18)$$

The permitted energy levels for the electrons in the crystal can therefore be obtained by summing the energy levels in the x, y and z directions:

$$E_{bck} = \frac{\hbar^2}{2m}\left[\left(\frac{2\pi n_x}{NL}\right)^2 + \left(\frac{\pi n_y}{b}\right)^2 + \left(\frac{\pi n_z}{c}\right)^2\right]$$

or:

$$E_{bck} = E_{bc} + \frac{\hbar^2}{2m}\left[\left(\frac{2\pi n_x}{NL}\right)^2\right] \quad (10.2.19)$$

where $E_{bc} = \frac{\hbar^2}{2m}\left[\left(\frac{\pi n_y}{b}\right)^2 + \left(\frac{\pi n_z}{c}\right)^2\right]$. The 1D unit volume in k-space corresponding to each permitted k value is:

$$1D\ unit\ volume = \frac{2\pi}{NL} = \frac{2\pi\, b\, c}{V} \quad (10.2.20)$$

Low-dimensional devices. Energy bands

The permitted energy values are thus obtained by adding the energy levels which are a function of k_x (which vary in a continuous manner) and a series of discrete energy levels produced by the wave function confinement in the y and z directions. The discrete energy levels resulting from the confinement, $E_{bc}(n_y,n_z)$, are the minima of energy subbands. The other energy values in each subband are obtained by adding $E_{bc}(n_y,n_z)$ to the energies corresponding to k_x values (Figure 10.7.B). It is worth noting that the minimum energy of the electron, which was equal to zero in the three-dimensional case (when $n_x=n_y=n_z=0$ in Equation 10.2.6) is now equal to $\frac{\hbar^2}{2m}\left(\left(\frac{\pi}{b}\right)^2+\left(\frac{\pi}{c}\right)^2\right)\neq 0$ (for $n_y = n_z = 1$).

In the case of a *zero-dimensional crystal* the dimensions in all x, y and z directions are very small. The length, width and height of the crystal are noted "a", "b" and "c". The wave function is now confined in the x, y and z direction. Using the technique of separation of variables the wave function can be written as the product of separate wave functions:

$$\Psi(x,y,z) = \Phi_a(x)\,\Phi_b(y)\,\Phi_c(z) \qquad (10.2.21)$$

The wave function $\Psi(x,y,z)$ can be found by solving the Schrödinger equation in a three-dimensional potential well:

$$-\frac{\hbar^2}{2m}\left(\frac{d^2\Psi(x,y,z)}{dy^2}+\frac{d^2\Psi(x,y,z)}{dy^2}+\frac{d^2\Psi(x,y,z)}{dz^2}\right)=E\,\Psi(x,y,z) \qquad (10.2.22)$$

which has the solution:

$$\begin{aligned}\Psi(x,y,z) &= \Phi_a(x)\,\Phi_b(y)\,\Phi_c(z)\\ \Phi_a(x) &= A\,exp(jkx) + B\,exp(-jkx)\\ \Phi_b(y) &= C\,exp(jky) + D\,exp(-jky)\\ \Phi_c(z) &= E\,exp(jkz) + F\,exp(-jkz)\end{aligned} \qquad (10.2.23)$$

Low-dimensional devices. Energy bands

Applying the following boundary conditions $\Phi_a(0) = 0$, $\Phi_b(0) = 0$, $\Phi_c(0) = 0$, $\Phi_a(a) = 0$, $\Phi_b(b) = 0$ and $\Phi_c(c) = 0$, one obtains:

$$\Phi_a(x) = G \sin\left(\frac{\pi n_x x}{a}\right),\ \Phi_b(y) = H \sin\left(\frac{\pi n_y y}{b}\right) \text{ and } \Phi_c(z) = I \sin\left(\frac{\pi n_z z}{c}\right)$$

where n_x, n_y and n_z can take on values 1, 2, 3,... (10.2.24)

The energy eigenvalues in the different directions are:

$$-\frac{\hbar^2}{2m}\frac{d^2\left(G \sin\left(\frac{\pi n_x x}{a}\right)\right)}{dx^2} = E_x\, G \sin\left(\frac{\pi n_x x}{a}\right) \Rightarrow E_x = \frac{\hbar^2}{2m}\left(\frac{\pi n_x}{a}\right)^2$$

$$-\frac{\hbar^2}{2m}\frac{d^2\left(H \sin\left(\frac{\pi n_y y}{b}\right)\right)}{dy^2} = E_y\, H \sin\left(\frac{\pi n_y y}{b}\right) \Rightarrow E_y = \frac{\hbar^2}{2m}\left(\frac{\pi n_y}{b}\right)^2$$

$$-\frac{\hbar^2}{2m}\frac{d^2\left(I \sin\left(\frac{\pi n_z z}{c}\right)\right)}{dz^2} = E_z\, I \sin\left(\frac{\pi n_z z}{c}\right) \Rightarrow E_z = \frac{\hbar^2}{2m}\left(\frac{\pi n_z}{c}\right)^2 \quad (10.2.25)$$

where the constants G, H and I have been determined by applying the boundary conditions. The electron energy values are obtained by summing the three latter equations, which yields:

$$E_{abc} = \frac{\hbar^2}{2m}\left[\left(\frac{\pi n_x}{a}\right)^2 + \left(\frac{\pi n_y}{b}\right)^2 + \left(\frac{\pi n_z}{c}\right)^2\right] \quad (10.2.26)$$

The permitted energy levels are thus a succession of discrete levels produced by the confinement in the three-dimensional potential well. The minimum energy value (when $n_x = n_y = n_z = 1$) is equal to

$$\frac{\hbar^2}{2m}\left(\left(\frac{\pi}{a}\right)^2 + \left(\frac{\pi}{b}\right)^2 + \left(\frac{\pi}{c}\right)^2\right) \neq 0\,.$$

Low-dimensional devices. Electron densities of states

In a *three-dimensional crystal* the volume of a lattice unit cell is equal to L^3 and the volume V of the crystal is equal to $V = (NL)^3$. The unit volume corresponding to each permitted state (*i.e.* to each k value) is equal to $\left(\frac{2\pi}{NL}\right)^3 = \frac{8\pi^3}{V}$ (Relationship 10.2.5).

Using a similar approach to that of Section 1.1.8 we will now consider a *sphere* in k-space which contains all the wave vectors corresponding to the electrons having an energy below a given maximum value. To each wave vector, $k \leq k_{max}$, correspond two electrons by virtue of the Pauli exclusion principle. The number of electrons is thus given by:

$$n = 2\left(\frac{4\pi}{3}k_{max}^3\right)\frac{V}{8\pi^3} \qquad (10.2.27)$$

and, in a unit volume ($V=1$):

$$n = 2\left(\frac{4\pi}{3}k_{max}^3\right)\left(\frac{1}{2\pi}\right)^3 \qquad (10.2.28)$$

The latter relationship enables us to link k_{max} to the electron concentration: $k_{max} = (3\pi^2 n)^{1/3}$.

The density of states is defined by $\rho = dn/dE$. We will use the symbol ρ for the density of states instead of $n(E)$, which was used in Equation 1.1.48 to avoid confusion between the number of electrons, n, and the density of states.

Using the following relationships we can relate the density of states, ρ, to energy values:

$$\frac{dn}{dk} = \frac{d}{dk}\left[2\left(\frac{4\pi}{3}k^3\right)\left(\frac{1}{2\pi}\right)^3\right] = 2\frac{4\pi k^2}{(2\pi)^3}$$

$$E = E_k = \frac{\hbar^2 k^2}{2m} \Rightarrow k = \left(\frac{2m}{\hbar^2}E_k\right)^{1/2}$$

$$\frac{dk}{dE} = \left(\frac{2m}{\hbar^2}\right)^{1/2}\frac{E_k^{-1/2}}{2} \qquad (10.2.29)$$

Finally we obtain the density of states as a function of E:

$$\rho = \frac{dn}{dE} = \frac{dn}{dk}\frac{dk}{dE} = \frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}E_k^{1/2} \qquad (10.2.30)$$

Thus, the density of states near a band extremum, such as the minimum of the conduction band, varies as the square root of the energy.

Low-dimensional devices. Electron densities of states

In a *two-dimensional crystal* confined in the z-direction the 2D volume of a lattice unit cell is equal to L^2 and the volume of the crystal is equal to $V = c(NL)^2$. The unit volume corresponding to each permitted state (*i.e.* to each k value) is equal to $\left(\frac{2\pi}{NL}\right)^2 = \frac{4\pi^2 c}{V}$ (Relationship 10.2.13). Using a similar approach to that of Section 1.1.7 we now have to consider a *circle* in k-space which contains all the wave vectors corresponding to the electrons having an energy below a given maximum value. To each wave vector, $k \leq k_{max}$, correspond two electrons by virtue of the Pauli exclusion principle. The number of electrons is thus given by:

$$n = 2\left(\pi k_{max}^2\right)\frac{V}{4\pi^2 c}$$

and, in a unit volume ($V=1$):

$$n = 2\left(\pi k_{max}^2\right)\left(\frac{1}{2\pi}\right)^2 \frac{1}{c} \qquad (10.2.31)$$

The latter relationship enables us to link k_{max} to the electron concentration: $k_{max} = (2\pi nc)^{1/2}$. The density of states in a subband is defined by: $\rho = dn/dE$. Thus we find:

$$\frac{dn}{dk} = \frac{d}{dk}\left[2\left(\pi k^2\right)\left(\frac{1}{2\pi}\right)^2 \frac{1}{c}\right] = \frac{k}{\pi c}$$

$$E = E_{ck} = E_c + \frac{\hbar^2 k^2}{2m} \Rightarrow \frac{dE}{dk} = \frac{\hbar^2 k}{m} \Rightarrow \frac{dk}{dE} = \frac{m}{\hbar^2 k}$$

$$\rho = \frac{dn}{dE} = \frac{dn}{dk}\frac{dk}{dE} = \frac{m}{\hbar^2 \pi}\frac{1}{c} \qquad (10.2.32)$$

Low-dimensional devices. Electron densities of states

Thus, the density of states near a subband extremum, such as the minimum of the conduction band, is constant and independent of the energy. However, one has to take into account that there are several subbands. The total number of electrons is obtained by adding the number of electrons in the different subbands:

$$n(E) = \int \rho dE = \frac{1}{c} \frac{m}{\hbar^2 \pi} \sum_c \theta(E-E_c) \qquad (10.2.33)$$

where the function θ is defined as:

$$\theta(E-E_c) = 0 \text{ if } E < E_c \text{ and } \theta(E-E_c) = E-E_z \text{ if } E \geq E_c.$$

In a *one-dimensional crystal* the 1D volume of a lattice unit cell is equal to L and the volume of the crystal is equal to $V = b\ c\ NL$. The unit volume corresponding to each permitted state (*i.e.* to each k value) is equal to $\left(\frac{2\pi}{NL}\right)$ $= \frac{2\pi\ b\ c}{V}$ (Relationship 10.2.20). Using a similar approach to that of Section 1.1.8 we now have to consider a *line segment* in k-space which contains all the wave vectors corresponding to the electrons having an energy below a given maximum value. The length of this segment is $2k_{max}$. To each wave vector, $k \leq k_{max}$, correspond two electrons by virtue of the Pauli exclusion principle. The number of electrons is thus given by:

$$n = 4\ k_{max} \frac{V}{2\pi\ b\ c}$$

and, in a unit volume ($V=1$): $$n = 2 \frac{k_{max}}{\pi} \frac{1}{b\ c} \qquad (10.2.34)$$

The latter relationship enables us to link k_{max} to the electron concentration: $k_{max} = \frac{\pi n b c}{2}$.

Low-dimensional devices. Electron densities of states

The density of states in a subband is defined by: $\rho = dn/dE$. Thus we find:

$$\frac{dn}{dk} = \frac{d}{dk}\left[\frac{2k}{\pi}\frac{1}{b}\frac{1}{c}\right] = \frac{2}{\pi}\frac{1}{b}\frac{1}{c}$$

Using $E = E_{bck} = E_{bc} + \frac{\hbar^2 k^2}{2m} \Rightarrow k = \left(\frac{2m}{\hbar^2}(E_{bck} - E_{bc})\right)^{1/2}$

we find $$\frac{dk}{dE} = \left(\frac{2m}{\hbar^2}\right)^{1/2} \frac{(E_{bck} - E_{bc})^{-1/2}}{2}$$

and thus:

$$\rho = \frac{dn}{dE} = \frac{dn}{dk}\frac{dk}{dE} = \frac{1}{\pi}\left(\frac{2m}{\hbar^2}\right)^{1/2} (E_{bck} - E_{bc})^{-1/2}\frac{1}{b}\frac{1}{c} \qquad (10.2.35)$$

where E_{bck} is a continuous function of k in the x direction and a discrete function in the y and z directions.

Thus, the density of states near a subband extremum, such as the minimum of the conduction band, now varies as an inverse square root function of the energy as a function of k. Again, one has to take into account that there are several subbands corresponding to the discretization in the y and z directions. The total number of electrons is obtained by adding the number of electrons in the different subbands:

$$n(E) = \int \rho dE = \frac{2}{\pi}\left(\frac{2m}{\hbar^2}\right)^{1/2}\frac{1}{b}\frac{1}{c}\sum \theta(E_{bck} - E_{bc})(E_{bck} - E_{bc})^{1/2} \qquad (10.2.36)$$

where the function θ is defined as:

$\theta(E_{bck} - E_{bc}) = 0$ if $E_{bck} - E_{bc} < 0$ and $\theta(E_{bck} - E_{bc}) = 1$ if $(E_{bck} - E_{bc}) \geq 0$.

Low-dimensional devices. Electron densities of states

The density of states for a 1D and 2D and 3D semiconductor sample with specified dimensions is shown in Figures 10.8 to 10.11.

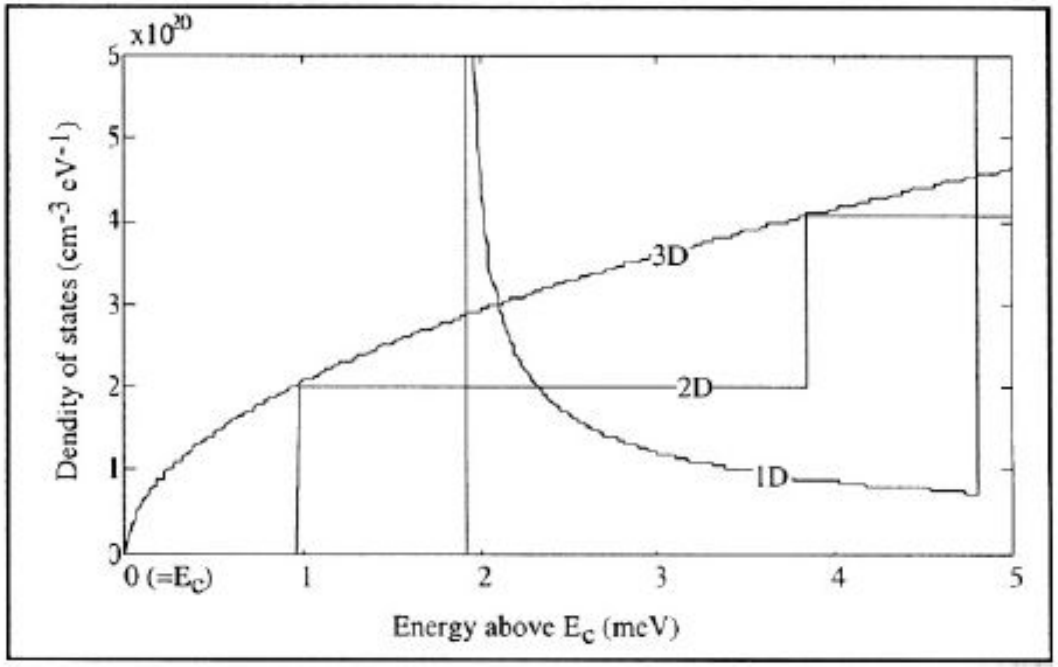


Figure 10.8: Density of states in the conduction band, as a function of energy, in silicon 1D, 2D, and 3D crystals. In the 2D sample the crystal height, c, is 20 nm, and in the 1D crystal the height, c, and the width, b, of the sample, are both equal to 20 nm.

Low-dimensional devices. Electron densities of states

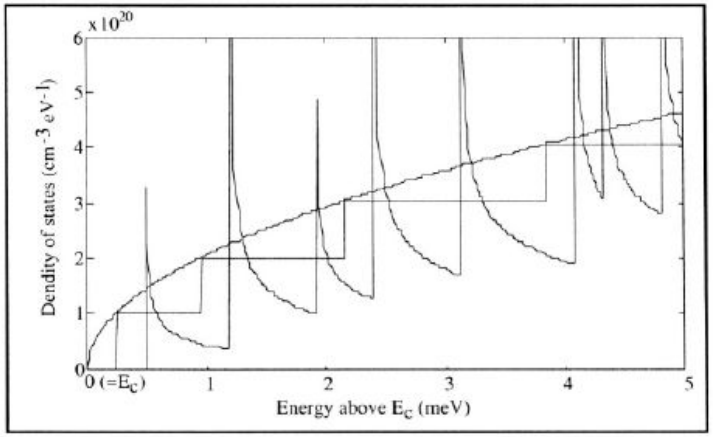


Figure 10.9: Density of states in the conduction band, as a function of energy, in silicon 1D, 2D, and 3D crystals. In the 2D sample the crystal height, c, is 40 nm, and in the 1D crystal the height, c, and the width, b, of the sample, are both equal to 40 nm.

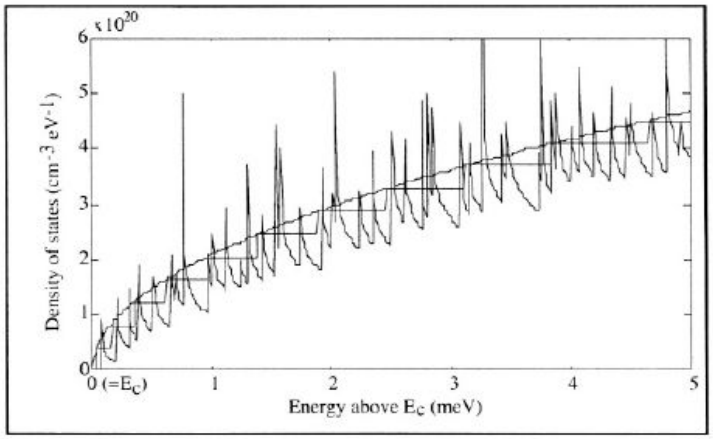


Figure 10.10: Density of states in the conduction band, as a function of energy, in silicon 1D, 2D, and 3D crystals. In the 2D sample the crystal height, c, is 100 nm, and in the 1D crystal the height, c, and the width, b, of the sample, are both equal to 100 nm. The dimensions b and c are now large enough for both the 1D and 2D distributions to "follow" the 3D curve.

Low-dimensional devices. Electron densities of states

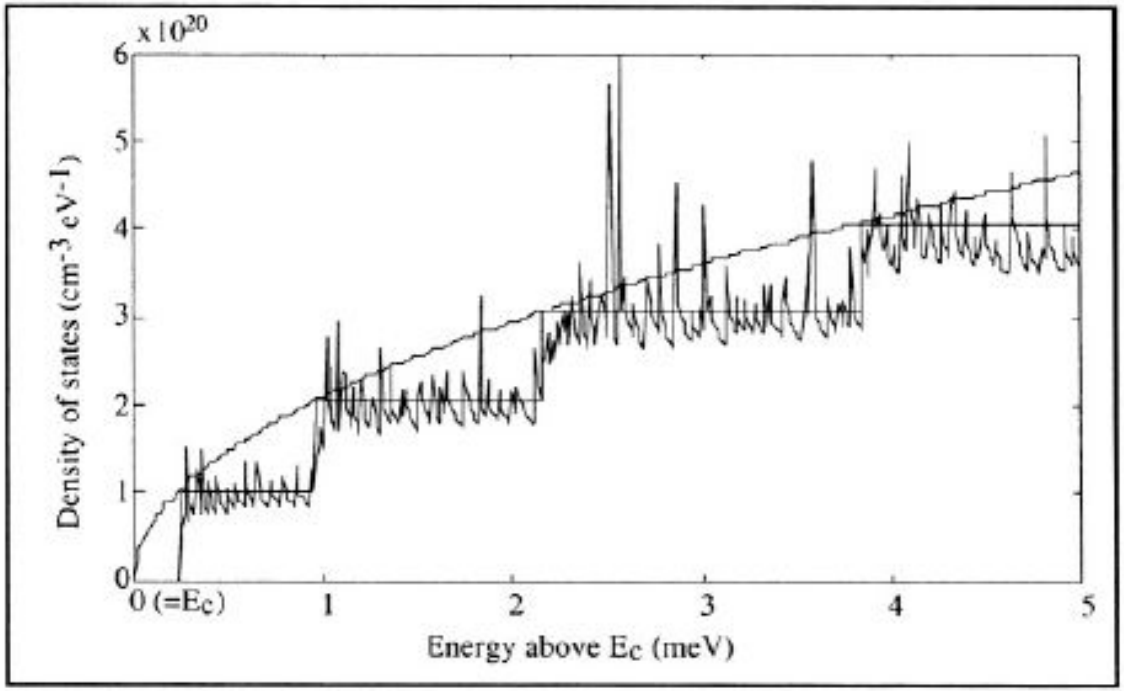


Figure 10.11: Density of states in the conduction band, as a function of energy, in silicon 1D, 2D, and 3D crystals. In the 2D sample the crystal height, c, is 40 nm, and in the 1D crystal the height, c, and the width, b, of the sample, are equal to 40 and 400 nm, respectively. The width, b, is now large enough for the 1D distribution to "follow" the 2D curve.

Quantum dots. Electron densities of states

Neglecting the periodic potential existing in solids, we can imagine a zero dimensional solid in which electron is confined in a three dimensional potential box with extremely small (<100 nm) length, breadth and height as a 0D solid. This will have the discrete energy levels as discussed above with density of states given as

$$D(E) = \frac{dN}{dE} = \sum_{\varepsilon_i} \delta(E - \varepsilon_i) \tag{1.45}$$

where ε_i are discrete energy levels and δ is Dirac function. The density of states as a function of energy would appear as illustrated in Fig. 1.9.

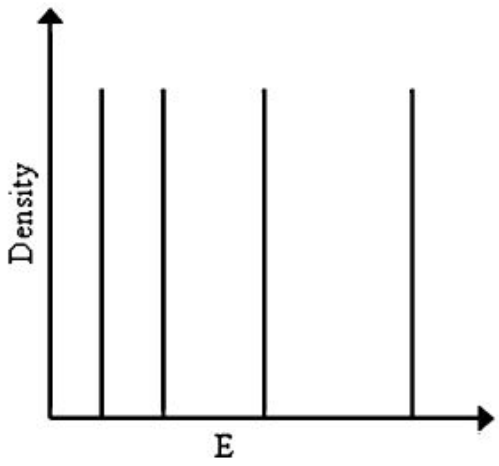


Fig. 1.9 Density of states for a particle in a zero dimensional solid

Quantum dot. What is it?

The presence of a discrete energy spectrum distinguishes quantum dots from all other solid-state systems and has caused them to be called "artificial atoms." However, the atom–quantum dot analogy should not be carried too far: Unlike electrons in an isolated atom, carriers in semiconductor quantum dots—which contain from a few thousand to tens of thousands of atoms arranged in a nearly defect-free three-dimensional crystal lattice—interact strongly with lattice vibrations and could be strongly influenced by defect, surface, or interface states.

One of the most important consequences of strong carrier confinement in quantum dots is the prominent role of many-particle effects. Coulomb interactions between carriers control the quantum dot charging and carrier recombination dynamics.

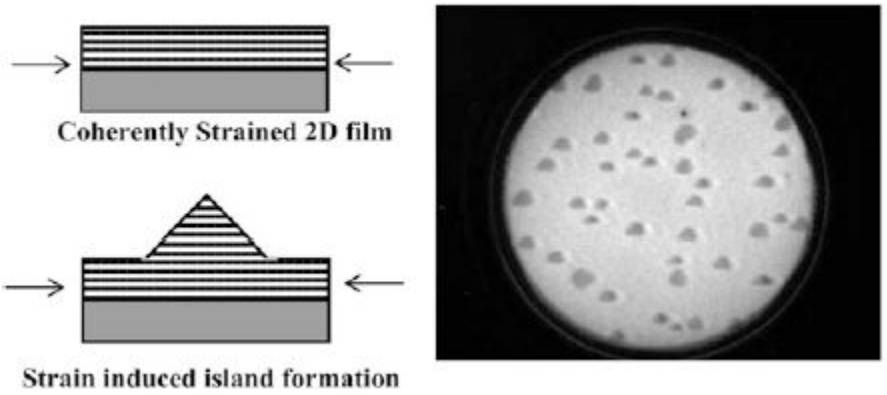
Self-assembled quantum dots have smaller sizes and stronger confinement potentials than lithographically defined quantum dots and therefore permit the study of different quantization regimes. Unlike lithographically defined nanostructures, self-assembled quantum dots can be easily fabricated and readily analyzed using optical spectroscopy and measurements of their transport properties. Based on experimental results, it seems likely that self-assembled quantum dots will play a key role in the emerging fields of single-particle electronics and photonics.

Quantum dots. Growth

Self Assembly in Inorganic Materials

It is possible to spontaneously create the quantum dots for example of germanium (Ge) on silicon (Si) or indium arsenide (InAs) on gallium aresenide (GaAs). The origin of self assembly is strain induced. Germanium and silicon have only 4 % lattice mismatch. Therefore Ge can be deposited epitaxially on Si single crystal upto 3–4 monolayers. Although grown (hetero)epitaxially, the layers of deposited Ge are highly strained (coherently i.e. without any defects or dislocations). When further deposition takes place, the lattice strain caused by depositing Ge on Si with different lattice constants cannot be accommodated.

This results in spontaneous formation of nanosized islands or quantum dots. However the temperature of the substrate has to be >350 °C during deposition or post-deposition annealing is required. Figure 6.6 schematically illustrates the growth mechanism as well as an electron microscopy image of germanium islands on Si (111) surface. The size of the islands depends upon the growth temperature as well as the substrate plane on which it grows.



Growth mechanism of Ge on Si and photograph showing island formation (Field of view 10 μm)

Quantum dots. Growth

Growing quantum dots, rings, and lattices

Fabrication of self-assembled quantum dots begins with some form of atomic deposition onto the surface of a semiconductor substrate, where the deposited material is chosen to have a smaller bandgap than the substrate. During the deposition process, epitaxial islands spontaneously form for energetic reasons on the crystal surface. These islands are then made into quantum dots by covering them with another semiconductor layer having a larger bandgap than the islands.

In practice, in-situ growth techniques such as molecular beam epitaxy or metalorganic chemical-vapor deposition are used to obtain the requisite ultraclean conditions and exquisite control of deposition parameters. With these techniques, a flux of atoms (for example, gallium, indium, or arsenic) is sent onto an ultraclean gallium arsenide surface held at high temperature. After diffusing across the clean reconstructed surface, the atoms arrange themselves, starting from step edges, to form a continuous epitaxial layer. The surface, interfacial, and elastic energies of the epitaxial film change during the film deposition process and the atomic arrangement on the surface develops so as to minimize the sum of these energies. The elastic strain energy of the film grows quadratically with the film thickness, and if the epitaxial film material has a lattice parameter even a few percent different from that of the substrate, nanometer-sized islands can form on the surface to minimize the total energy.[4] This film relaxation is elastic, so no defects are introduced in the island formation process.

Quantum dots. Growth

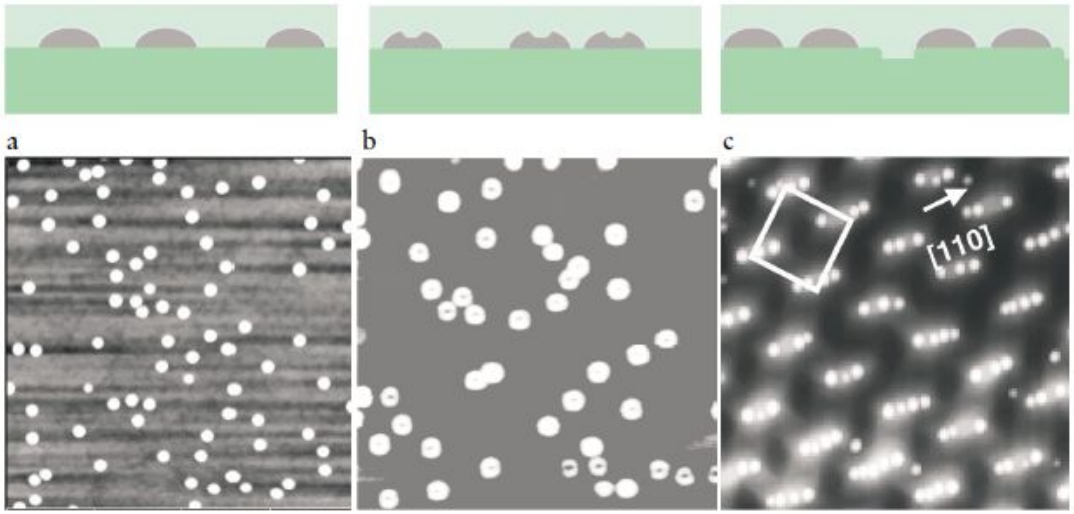


FIGURE 1. ATOMIC FORCE MICROGRAPH images (bottom) of islands epitaxially grown on a gallium arsenide-(001) substrate, and schematics (top) of the profile of the islands showing their shape and distribution. The islands are transformed into quantum dots or rings by covering them with a GaAs epitaxial film (represented by the light shading on the top drawings). The lateral size of each AFM image is 1 μm, and the quantum dot density is between 5×10^9 cm^{-2} and 2×10^{10} cm^{-2}. (a) Indium arsenide islands are randomly nucleated on the surface. The islands are shaped like truncated pyramids with a base of 30–40 nm and a height of 4–8 nm. (b) $In_xGa_{1-x}As$ ring-shaped islands are randomly distributed on the substrate. The depressed center of the ring can be seen as a small black dot in the image. (c) InAs islands on a GaAs substrate form a two-dimensional lattice with a unit cell (indicated in white) containing three to four islands per lattice point. The lattice is formed by patterning the substrate before nucleation with a periodic mesa pattern incorporating localized stress centers under the mesas. Nucleation takes place preferentially on top of the mesas in order to minimize the film energy on the surface and relax local elastic stresses. The number of islands in the basis, the lattice period, and the orientation of the two-dimensional lattice of islands can all be adjusted by varying the InAs flux and the size, orientation, and period of the mesa lattice.

Quantum dots. Growth

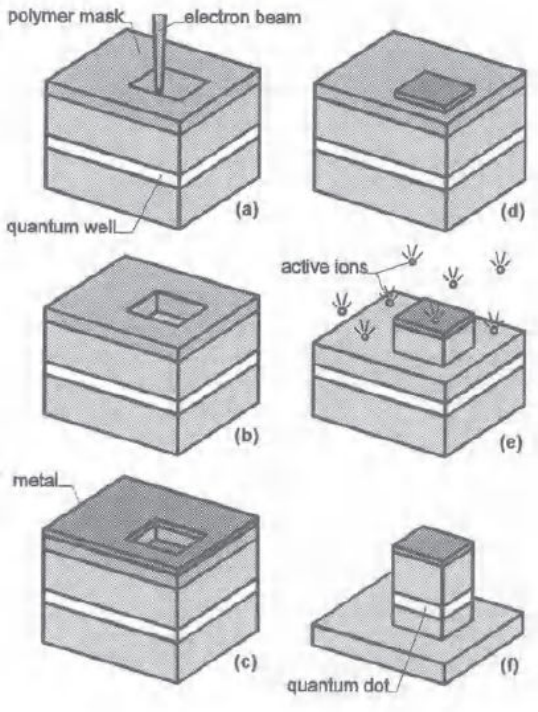


Fig. 2.1a–f. Process of quantum dot etching [117]

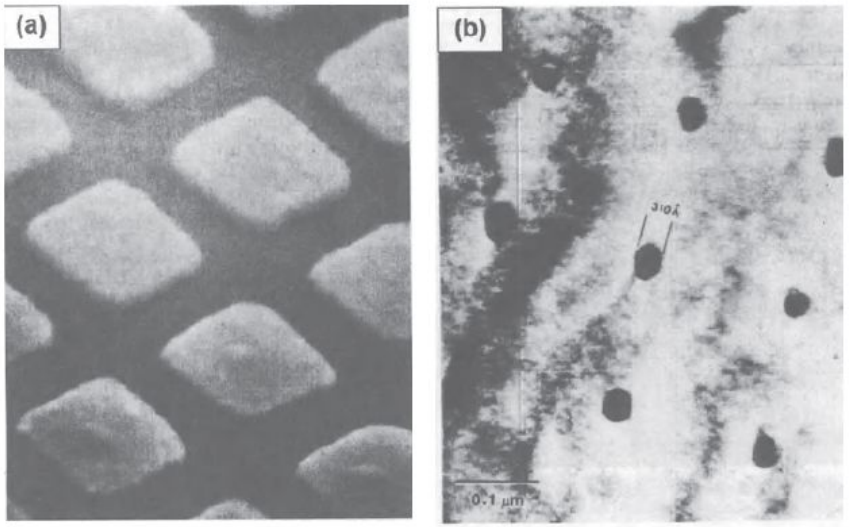


Fig. 2.3. Etched quantum dots: (**a**) diameter of 200 nm, GaAs/AlGaAs well, electron scanning microscope picture [125]; (**b**) diameter of 30 nm, InGaAs/InP well, transmission electron microscope picture [128]

Quantum dots. Growth

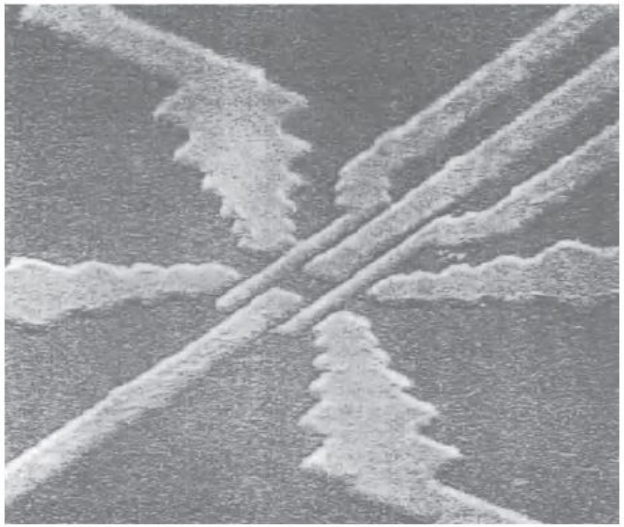


Fig. 2.4. Quantum dot on the intersection of electrodes; four internal electrodes localize the electrons, and four external ones serve as contacts for the electrons tunneling to and from the dot [117]

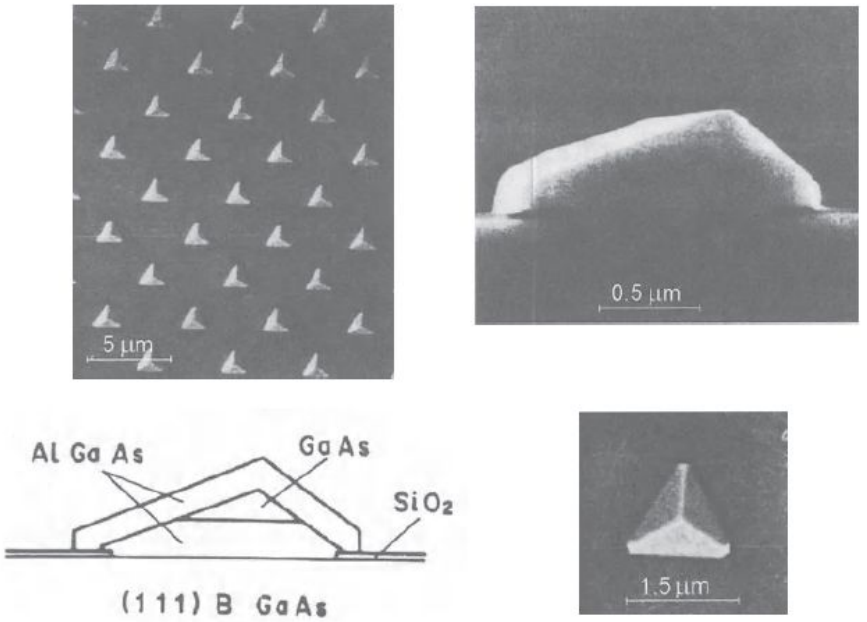


Fig. 2.7. Quantum dots created on the surface of GaAs in selective MOCVD growth (scanning electron microscope pictures); width of the electron localization area at the top of the pyramid is about 100 nm [45]

Quantum-dot field-effect device

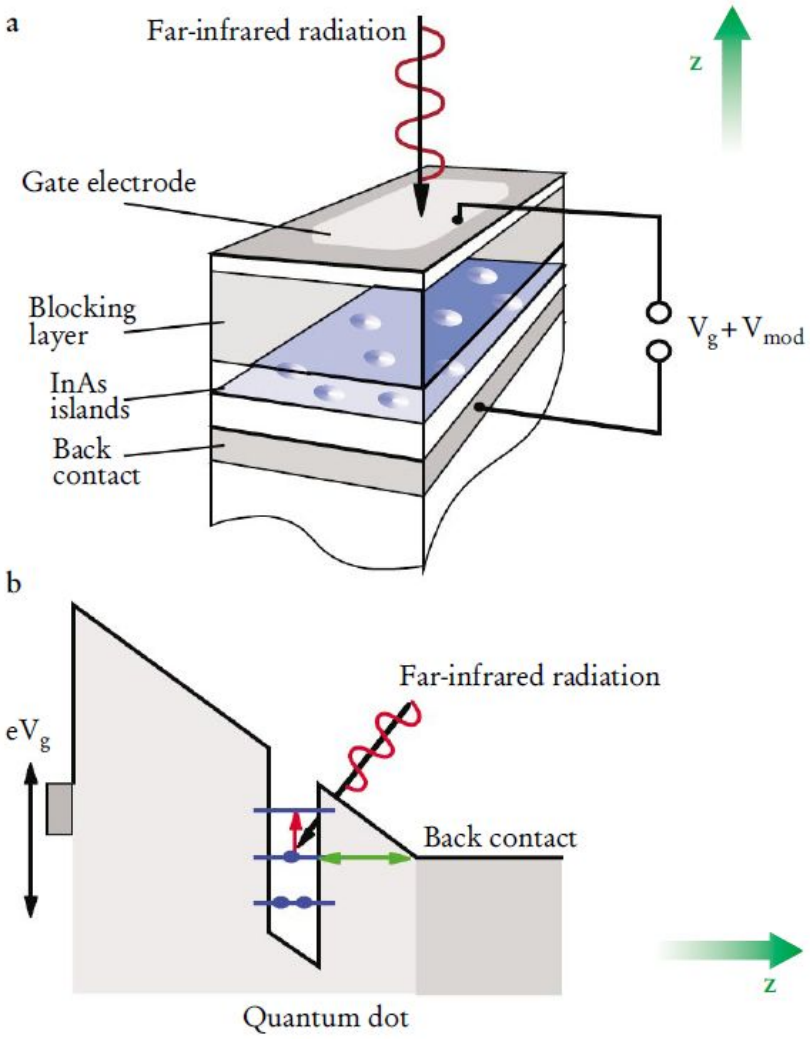


FIGURE 2. QUANTUM-DOT FIELD-EFFECT DEVICE used to probe the electronic states of self-assembled quantum dots and quantum rings. (a) The device consists of a capacitor-like structure between two parallel conducting plates; the self-assembled islands are embedded, as indicated, in an internal layer. The lower plate (back contact) is part of the grown semiconductor structure and consists of a highly (silicon) doped GaAs layer. The top plate is a thin, transparent, metal layer evaporated on top of the sample. The field inside the capacitor structure, and thus the energetic alignment between the islands and the back contact, can be tuned by an external voltage V_g applied between the back contact and the top gate. Because the back contact is located only some tens of nanometers away from the islands, electrons can readily tunnel back and forth between the two if an electronic level in the islands is aligned with the chemical potential of the back contact. When a small AC voltage V_{mod} is added to V_g, the resulting shift of charge back and forth between the back contact and the dots will manifest itself by an increased capacitive signal in the external circuit, which can be used for capacitance spectroscopy. (b) A schematic conduction band diagram shows that electrons enter the quantum dots by tunneling from the back gate through the GaAs barrier (the triangular section above the double-headed green arrow). The allowed dipole transitions of the many-particle system are measured using infrared spectroscopy, as indicated: The lower energy level is populated from the back contact using an applied bias (green arrows), then an incident infrared photon induces a transition between two quantum dot energy levels (red arrow), and the transition energy is measured by infrared absorption spectroscopy.

Capacitance spectroscopy using quantum dots

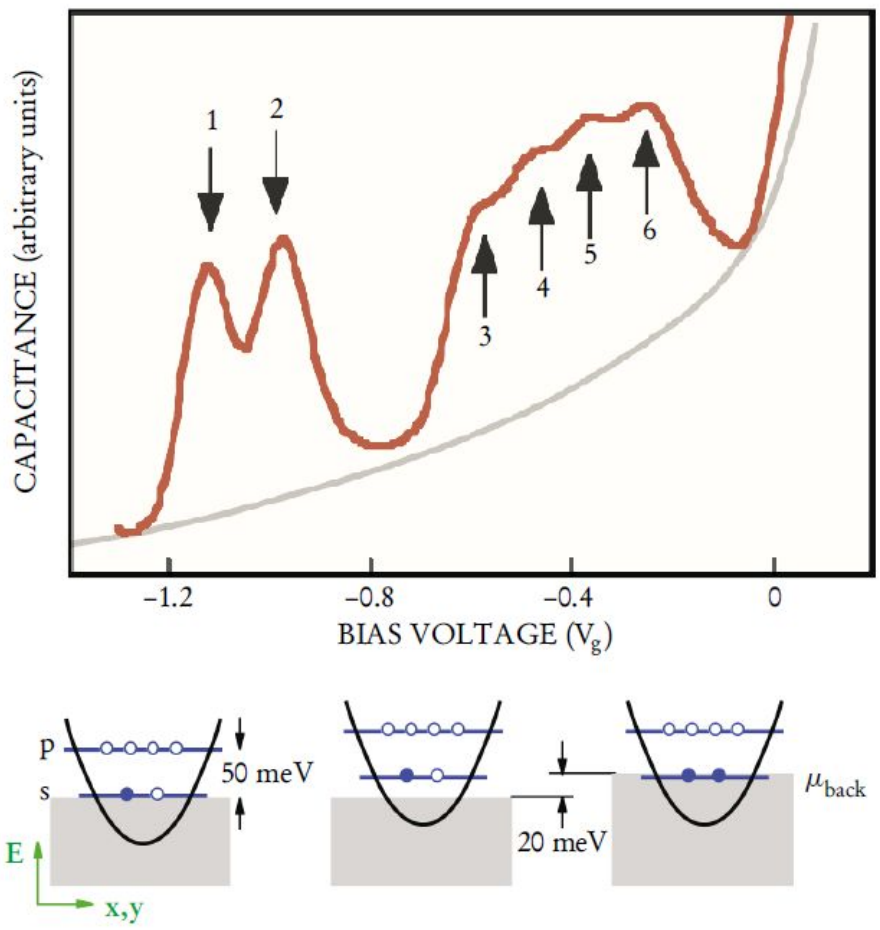


FIGURE 3. CAPACITANCE SPECTROSCOPY reveals quantum-dot electron occupancy and ground-state energies in the field-effect device shown in figure 2. Above, measured capacitance is plotted as a function of bias voltage V_g. Numbered arrows indicate the peaks that occur where an additional electron enters the dot. Below, the schematic band diagrams show the changes in the effective confining potential (black curve) and in the lowest energy levels (connected blue circles, filled in for occupied levels) as electrons are added to the quantum dot. The Fermi level μ_{back} of the back gate, which is proportional to the bias voltage, is at the top of the dark shading in the diagrams. Although the lowest energy state is doubly spin degenerate, electron–electron interaction makes it harder to load the second electron into a quantum dot than the first. The energy levels when the first electron enters the dot are shown at bottom left; Coulomb interaction raises the energy of the two-electron state by almost 20 meV compared to the one-electron state (the so-called Coulomb blockade), as shown at bottom center. To load the second electron the back gate potential must be raised (bottom right). Adding a third electron would require both the Coulomb energy and the quantization energy of about 50 meV, so there is a gap in voltage between the second and the third peak in the charging characteristic (top). To convert from bias voltages to energies, a scaling factor (about 7 for the present structure) is used, which can be derived from a simple evaluation of the capacitative energy at the quantum dot layer.

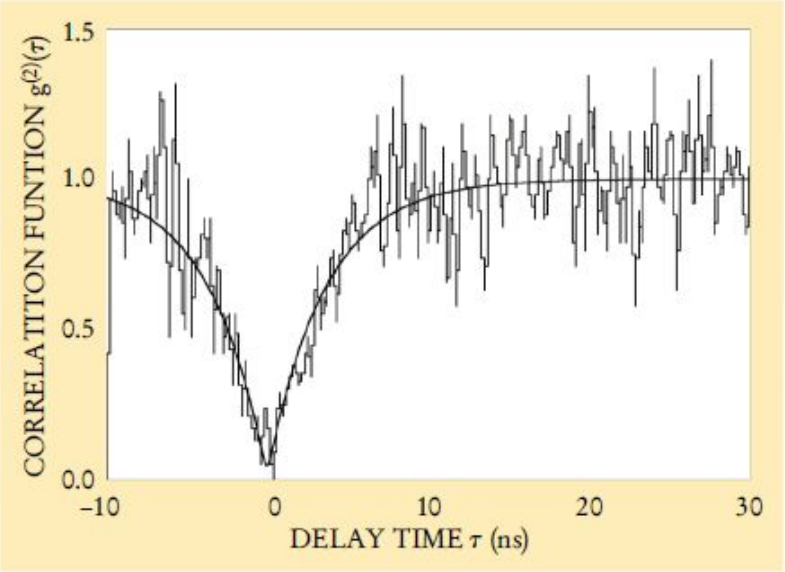
Photon correlations. The quantum nature of light

Since the pioneering experiments by Robert Hanbury-Brown and Richard Q. Twiss, photon correlation measurements have become a major tool in quantum optics and the spectroscopy of single-quantum systems.[11] Such experiments measure the likelihood that, given an initial photon detection event at time t, a second photon will be detected at time $t + \tau$. The experimental setup consists of a 50/50 beamsplitter and two single-photon counting avalanche photodiodes (SPADs), each of which generates a voltage pulse upon detection of a single photon. The pulses from the two SPADs are used to start and stop a time-to-amplitude converter, which translates the time delay between the start and stop pulses into a voltage amplitude. The result is a count of the number of pairs of photons $n(\tau)$ that have an arrival-time separation of τ. In the limit where the reciprocal of the average counting rate is much longer than the monitored time range, $n(\tau)$ is proportional to the normalized intensity correlation function $g^{(2)}(\tau)$.[11]

For a coherent light source, such as a single-mode laser, $g^{(2)}(\tau) = 1$. Physically, this means that detection of a photon at time $\tau = 0$ gives us no information about the succeeding photon-detection event; that is, the photons that make up a coherent light beam are completely uncorrelated. If, on the other hand, a light source satisfies either $g^{(2)}(0) < g^{(2)}(\tau)$ or $g^{(2)}(0) < 1$, this indicates that two photons are unlikely to appear simultaneously at the detectors. The first observation (in 1977) of this "photon antibunching" in fluorescence from a single atom has been generally regarded as the first proof of the quantum nature of light, since these correlations cannot be explained without quantizing Maxwell's equations.[11]

Fluorescence from a single anharmonic quantum emitter has the property that $g^{(2)}(0) \approx 0$. This strong quantum correlation among photons is easily washed away with an increasing number of emitters. We can therefore consider strong photon antibunching ($g^{(2)}(0) < 0.5$) as direct evidence that the source of the radiation field is a single-emitter quantum system. In contrast, observation of sharp emission peaks in photoluminescence spectra is not sufficient to rule out the presence of two or more identical emitters.

Photon correlation measurements have recently been carried out in both colloidal and self-assembled quantum dots.[13] The figure shows the measured intensity correlation function of the fluorescence from the fundamental exciton line of a single self-assembled quantum dot. The dip at zero time delay is direct evidence that the fluorescence comes from a single anharmonic quantum emitter. The time scale on which the correlation function changes from $g^{(2)}(0) \approx 0$ to $g^{(2)}(\tau) \approx 1$ is ultimately determined by the single-exciton recombination time. Thus, in addition to confirming the prior photoluminescence-based identification of single quantum-dot exciton lines, these measurements are also useful as a spectroscopic tool for measuring quantum-dot recombination times without requiring mode-locked laser sources for excitation.



The quantum nature of light revealed by quantum dots

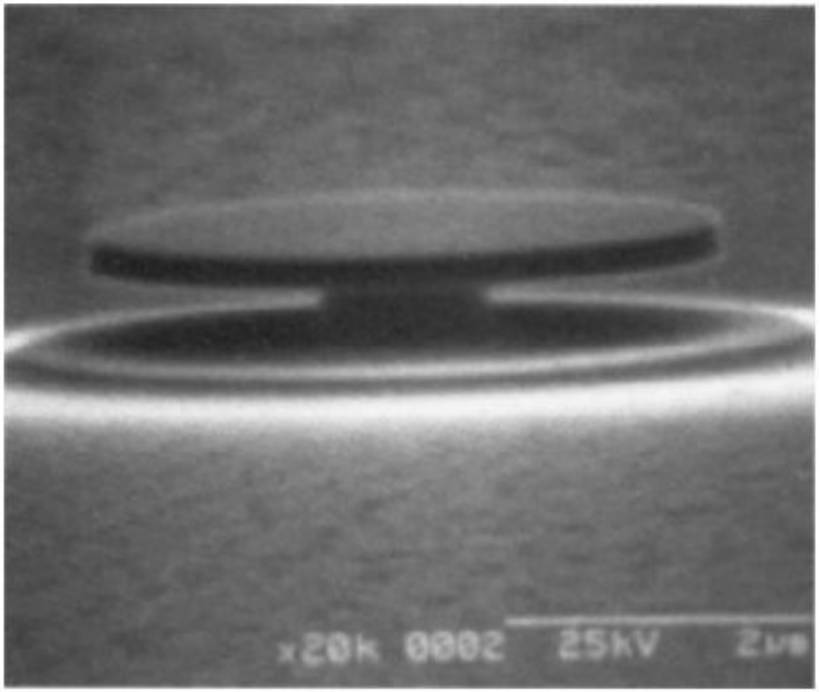


Fig. 1. The microdisk structure, which consists of a 5-μm-diameter disk and a 0.5-μm post. The GaAs disk area that supports high-quality factor WGMs is 200 nm thick and contains InAs quantum dots.

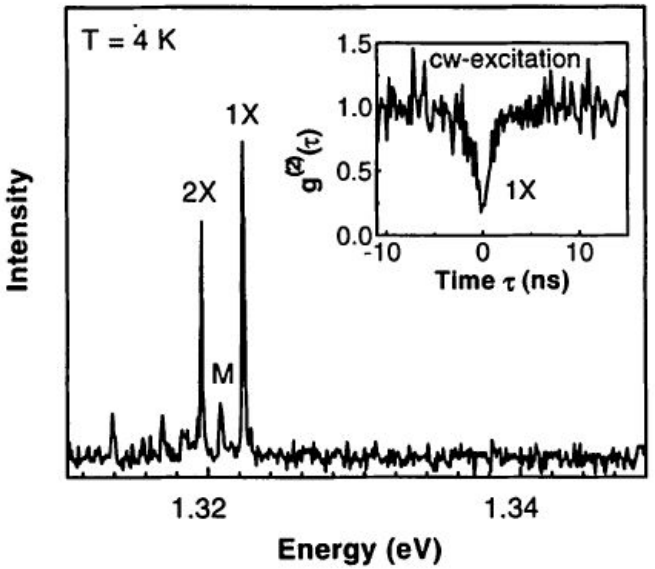


Fig. 2. Photoluminescence spectrum of a single InAs quantum dot embedded in a 5-μm-diameter microdisk. Contributions from the excitonic ground state transition (1X), higher excited states [for example, biexciton (2X)], and a WGM (M) are visible. (Inset) Measured normalized cw correlation function $g^{(2)}(\tau)$ of the single quantum dot 1X transition. The time bin is 195 ps and the excitation power is 160 W/cm^2.

The quantum nature of light revealed by quantum dots

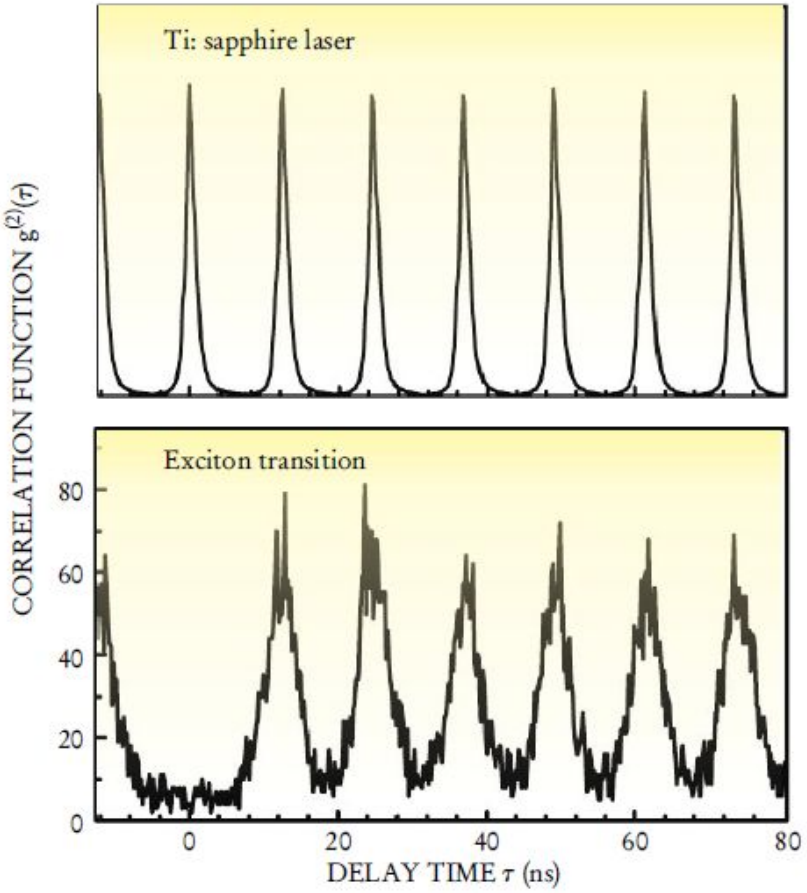


FIGURE 6. INTENSITY CORRELATION FUNCTION for a quantum-dot single-photon source based on pulsed laser excitation of a single quantum dot embedded in a microdisk. Plotted are the correlation function for the titanium-sapphire excitation laser (top) and quantum-dot ground-state emission under excitation conditions (bottom), as a function of the time delay τ between the photon arrival times (see the box on page 51). The absence of a peak at $\tau = 0$ in the emission correlation function shows that none of the pulses contains more than one photon. (Adapted from ref. 15.)

The quantum-dot memory device

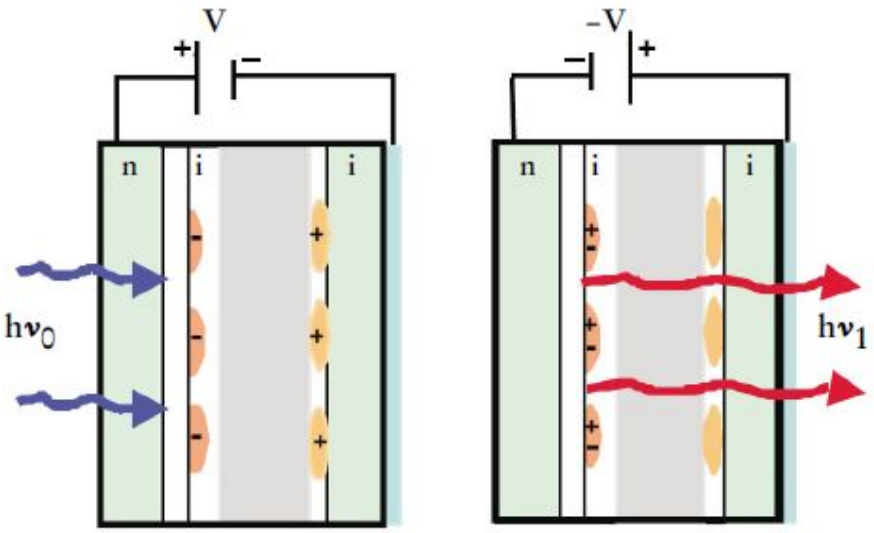


FIGURE 7. A QUANTUM-DOT MEMORY DEVICE based on controlled charge storage. Two strain-coupled quantum-dot layers are fabricated within a field-effect transistor with an n^+ gallium arsenide-type doped back gate and a gold semitransparent Schottky front gate. The quantum-dot layers (light green) are separated by a thin aluminum arsenide layer (gray) that permits very fast (about 0.5 ps) electron transfer. (a) For the write cycle, a photon-induced exciton is dissociated by the internal field into an electron and hole pair, which are stored, respectively, in an indium arsenide quantum dot and in a nearby strain-induced GaAs quantum dot. (b) The information is read optically by applying a positive voltage pulse to the front Schottky contact, which drives the hole from the strain-induced quantum dot into the quantum dot containing the electron. The exciton is regenerated and recombines to emit a photon, which is then observed using standard photon-detection techniques.

Light emission from quantum dots

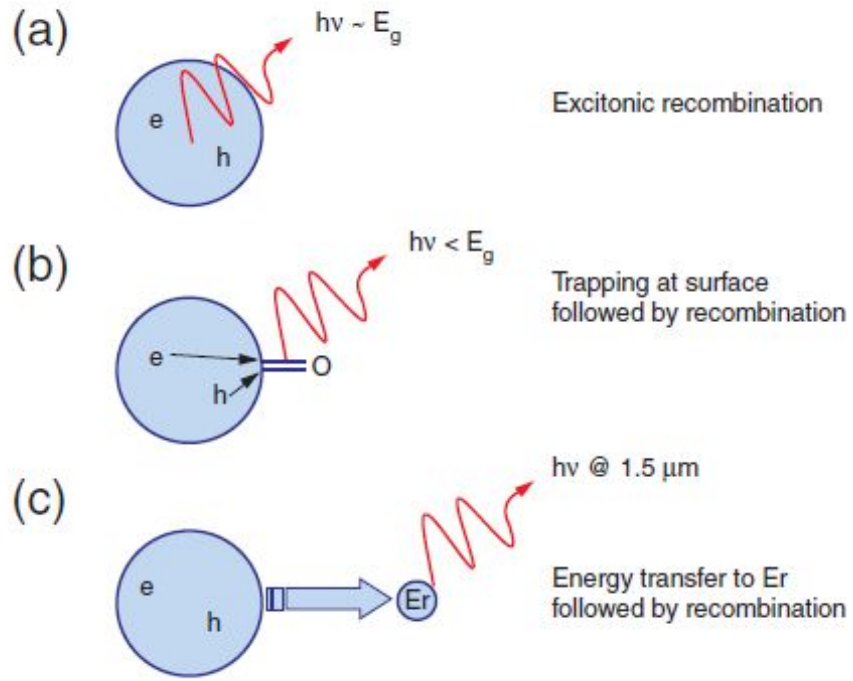


Fig. 2 Possible mechanisms that can lead to radiative light emission in Si QDs. (a) Recombination occurs between an electron in the conduction band and a hole in the valence band. The emitted photon has an energy equal to the band gap energy. (b) Recombination involves an electron, a hole, or both electron and hole trapped in a surface state. In this case, the emitted photon energy is less than the band gap energy since the surface levels lie within the band gap. (c) The exciton generated within the nanocrystal is quickly captured by Er atoms located in close proximity and photons are emitted at a wavelength of 1.5 µm.

Light emission from quantum dots. Sharp atomic-like lines due to the exciton effects

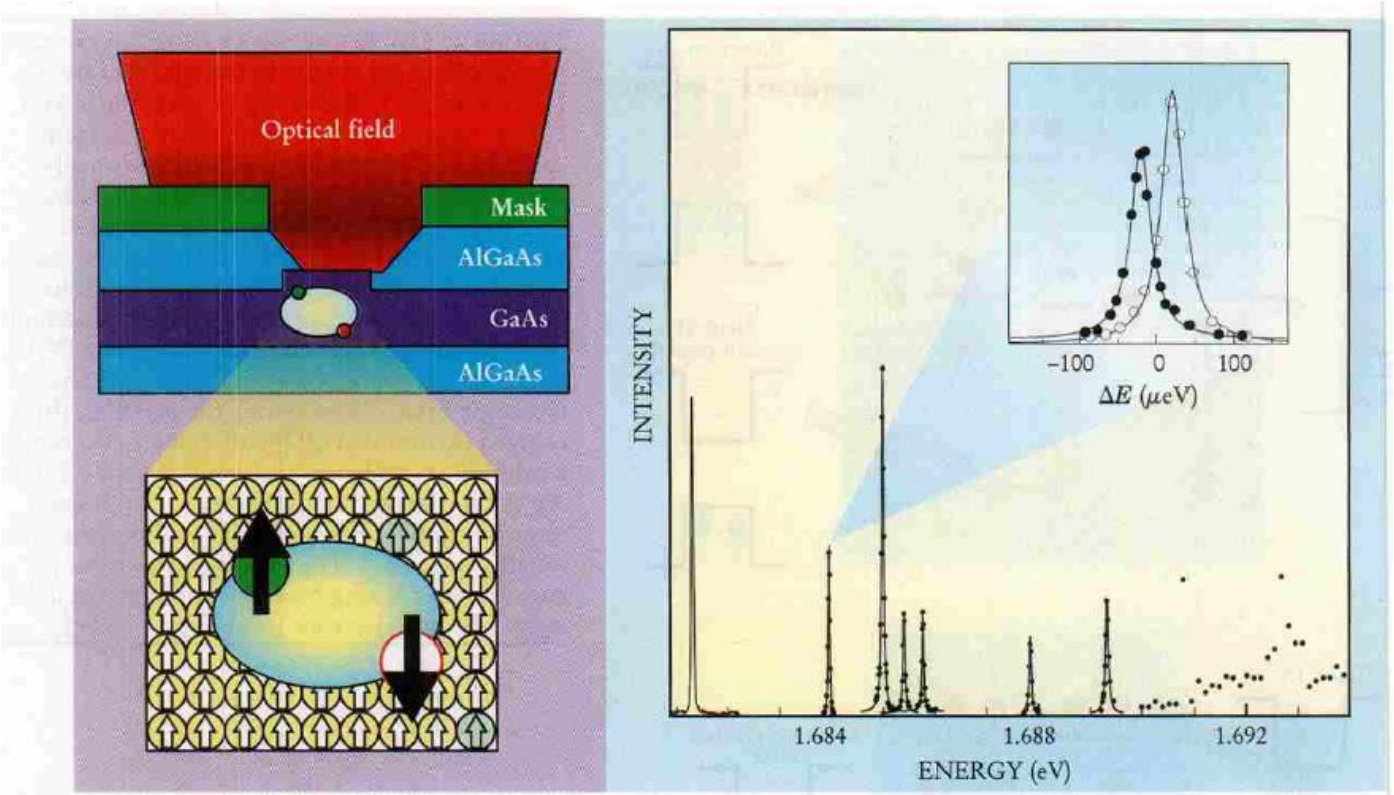


FIGURE 1. A SINGLE QUANTUM DOT can be selectively excited, detected, or both using high spatial and spectral resolution. The spectrum on the right shows extremely sharp discrete excitation resonances reminiscent of an atomic spectrum. These sharp spectral lines arise from single exciton states that differ in their orbital wavefunctions. Substates give rise to fine structure, which usually shows up as a doublet, as shown in the inset. (Adapted from ref. 14.)

Light emission from quantum dots. Single-photon sources

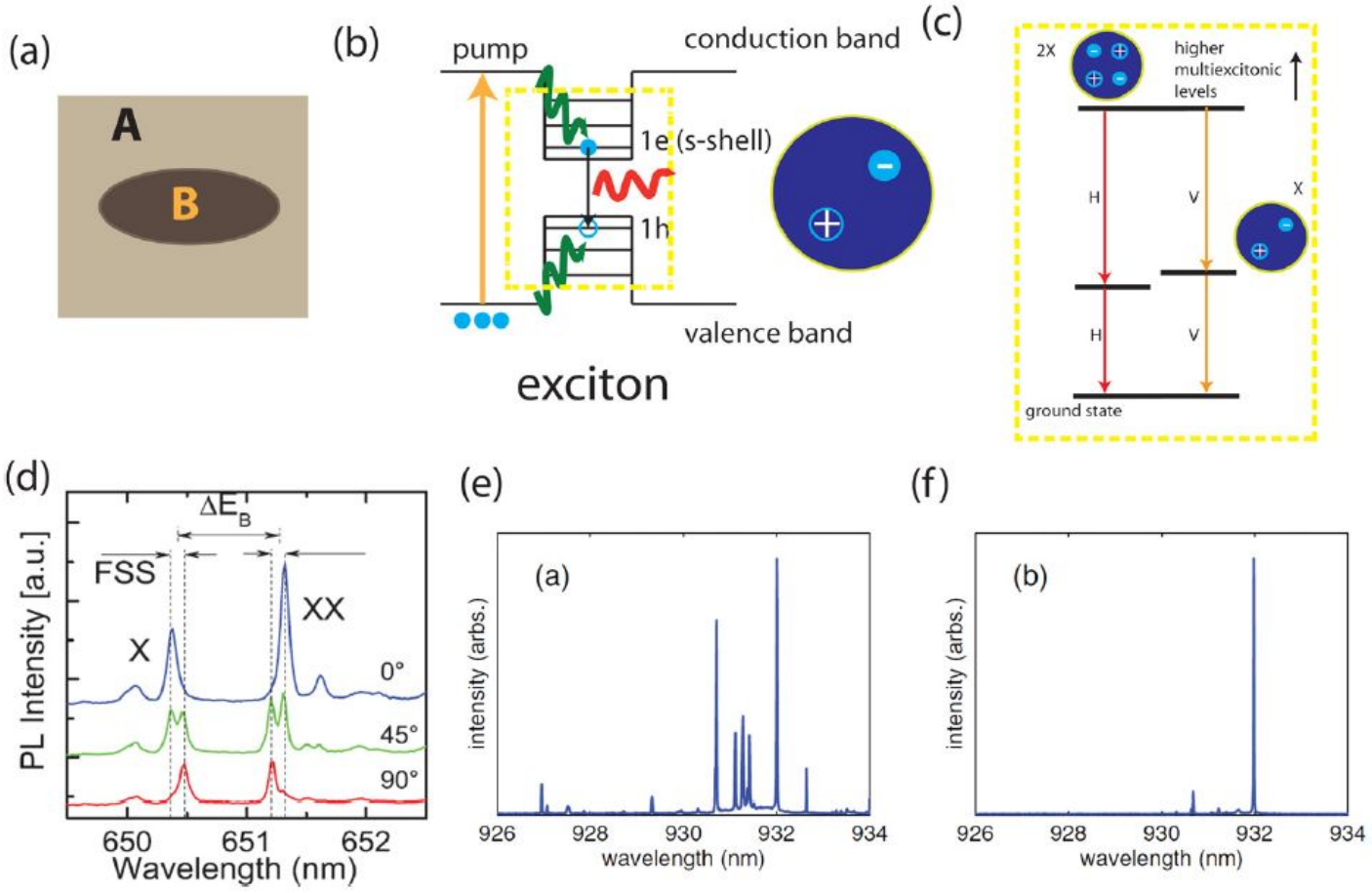


Figure 2. (*a*) A QD consists of a small, nanoscale island of a lower band gap semiconductor (B) embedded in a higher band gap semiconductor (A). (*b*) 1D diagram of the electronic structure of the QD. Incoherent pumping and emission from the exciton state are shown. (*c*) The level structure and fine structure splitting present for the biexciton and exciton. (*d*) Fine structure splitting present in InP/InGaP QD. Spectrum shown for polarizations at 0°, 45° and 90°. The spectrum of InAs/GaAs QD under (*e*) above band and (*f*) resonant excitation. In (*f*), the excitation laser is tuned to the higher order transition inside a QD, while in (*e*), the excitation laser frequency is above the GaAs band gap. (*d*) reproduced with permission from [69]. Copyright 2012 American Institute of Physics. (*e*)–(*f*) reproduced from [70].

Light emission from quantum dots. Single-photon sources. Hanbury Brown and Twiss effect

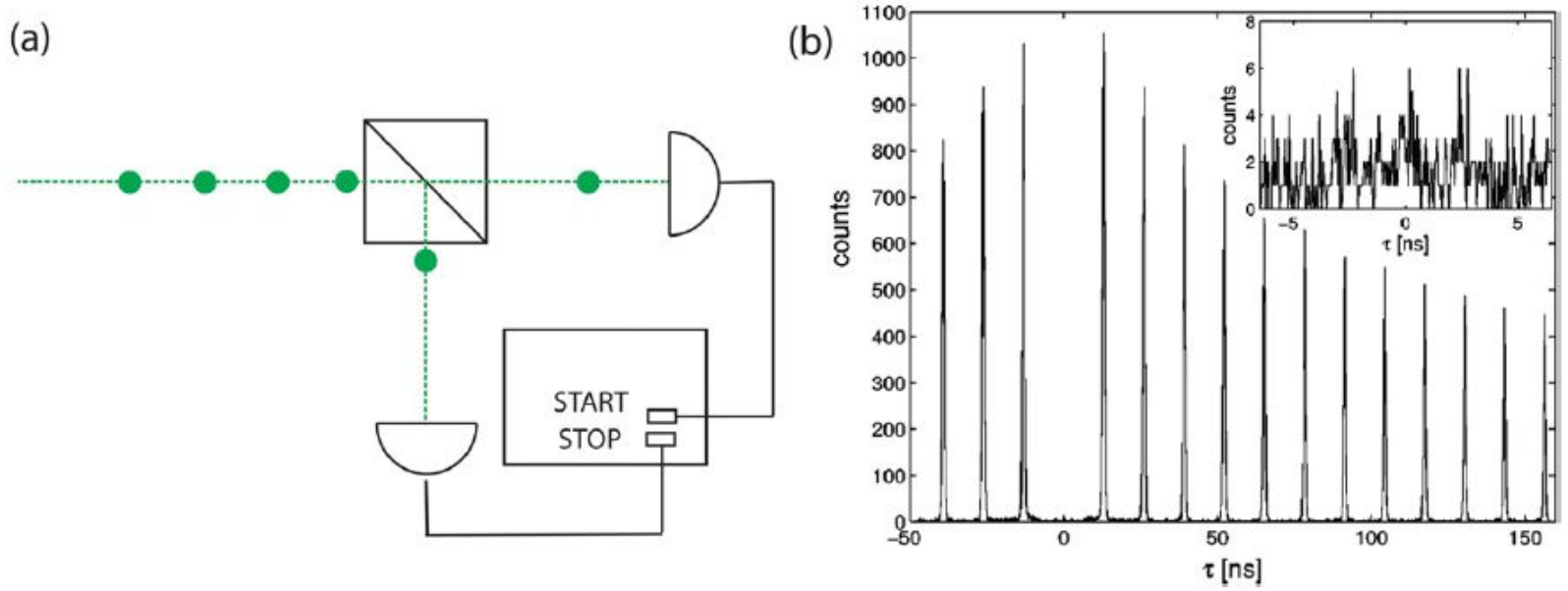


Figure 5. (*a*) HBT measurement setup. (*b*) Pulsed $g^{(2)}(\tau)$ measurement. The missing central peak indicates single-photon emission. Inset shows missing signal at $\tau = 0$. Reproduced with permission from [162]. Copyright 2003 American Institute of Physics.

Light emission from quantum dots. Single-photon sources

A powerful idea in quantum dot physics is to excite a quantum dot with a laser pulse intense enough that the probability of the dot capturing at least one exciton is very close to one [52]. A highly excited quantum dot decays by emitting a series of photons, which have different energies because of the renormalization through the Coulomb interaction. This enables the final photon to be selected by spectral filtering, and this photon can be used for communication.

This idea has now been explored experimentally and, at least at low temperatures, photons on demand have been generated from individual quantum dots [53, 54]. This is often called non-classical light because the temporal ordering of the photons is highly dissimilar from classical light described by Poisson statistics.

The quantum-dot infrared photodetector (QDIP)

The success of quantum well structures for IR detection applications has stimulated the development of QDIPs. In general, QDIPs are similar to QWIPs but with the quantum wells replaced by quantum dots, which have size confinement in all spatial directions.

Fig. 2 shows the schematic layers of a QWIP and a QDIP. In both cases, the detection mechanism is based on the intraband photoexcitation of electrons from confined states in the conduction band wells or dots into the continuum. The emitted electrons drift towards the collector in the electric field provided by the applied bias, and photocurrent is created. It is assumed, that the potential profile at the conduction band edge along the growth direction for both structure have a similar shape as shown in Fig. 2(b).

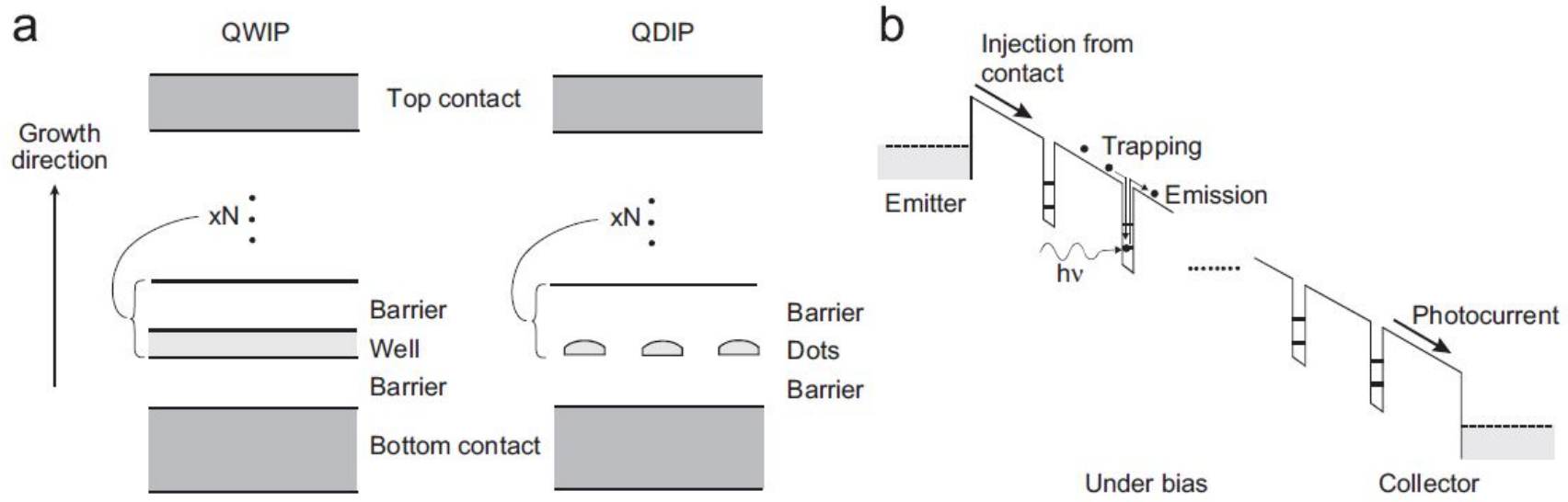


Fig. 2. Schematic layers of QWIP and QDIP (a) and potential profile for both structures under bias (b). For QDIP, influence of wetting layer is neglected (after Ref. [17]).

The quantum-dot infrared photodetector (QDIP)

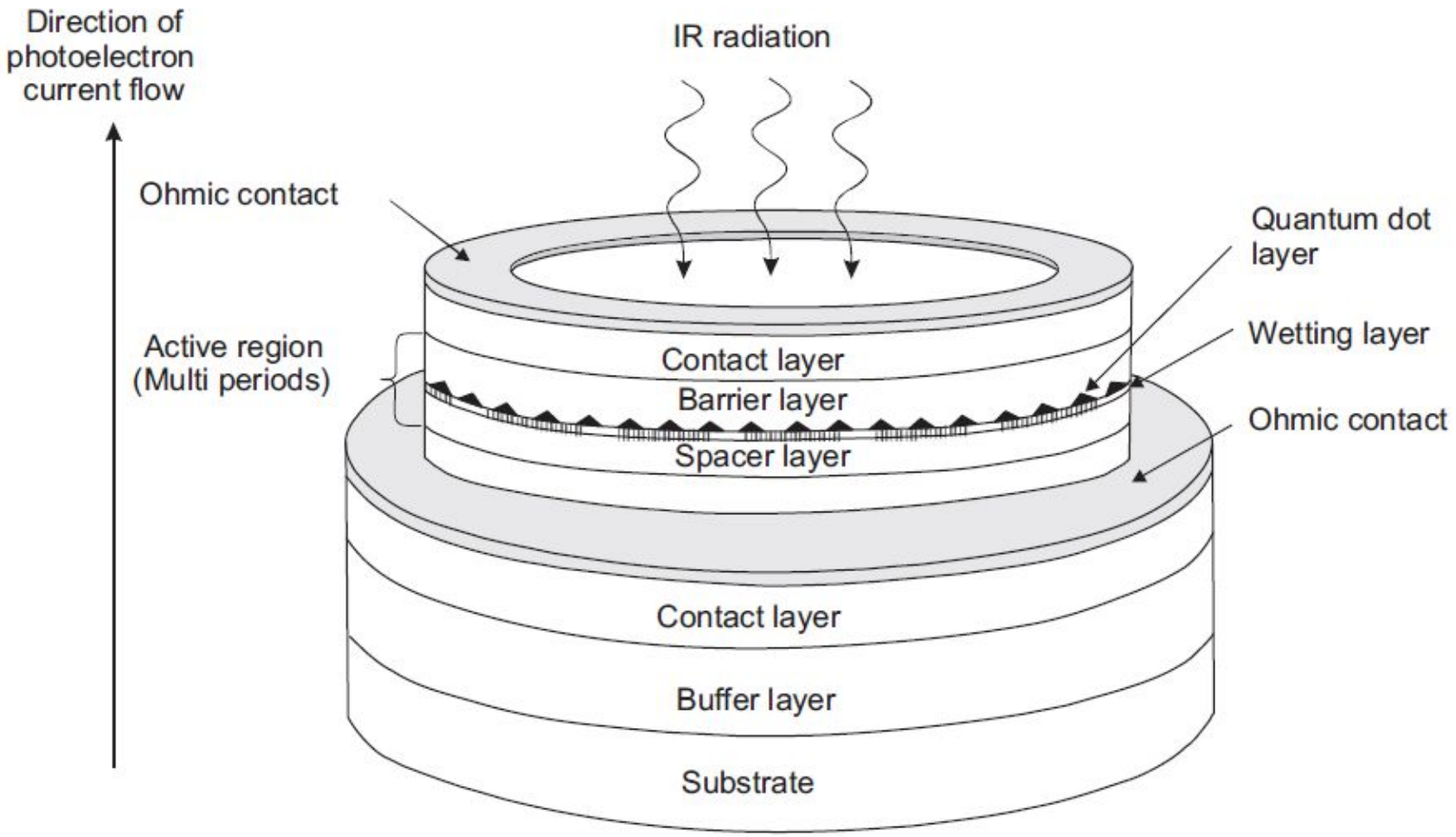


Fig. 3. Schematic diagram of conventional quantum-dot detector structure.

The quantum-dot infrared photodetector (QDIP)

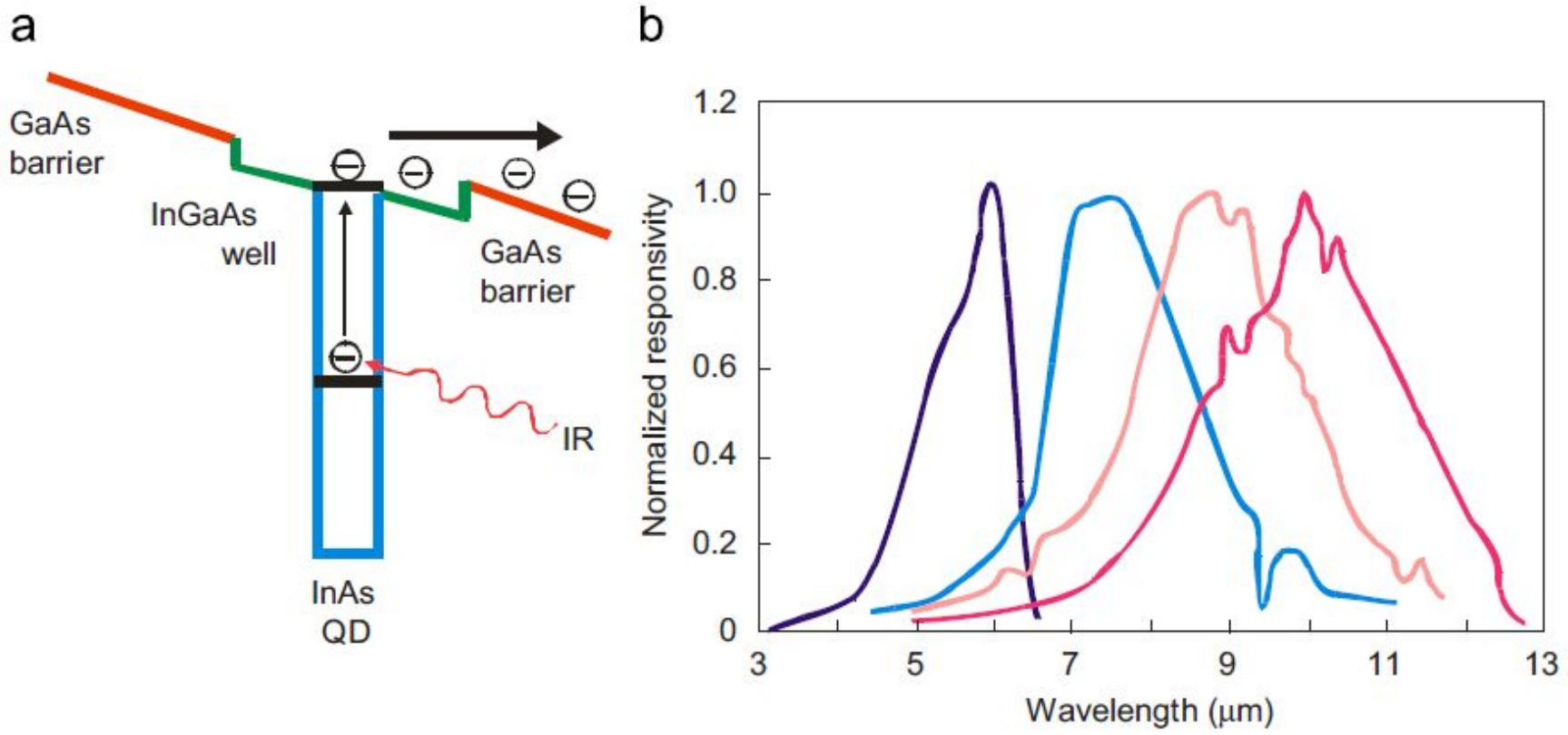


Fig. 4. DWELL infrared detector: (a) the operation mechanism, (b) experimentally measured spectral tunability by varying well width from 55 to 100 Å (after Ref. [21]).

The quantum-dot for solar cells. Size-dependent light absorption spectra

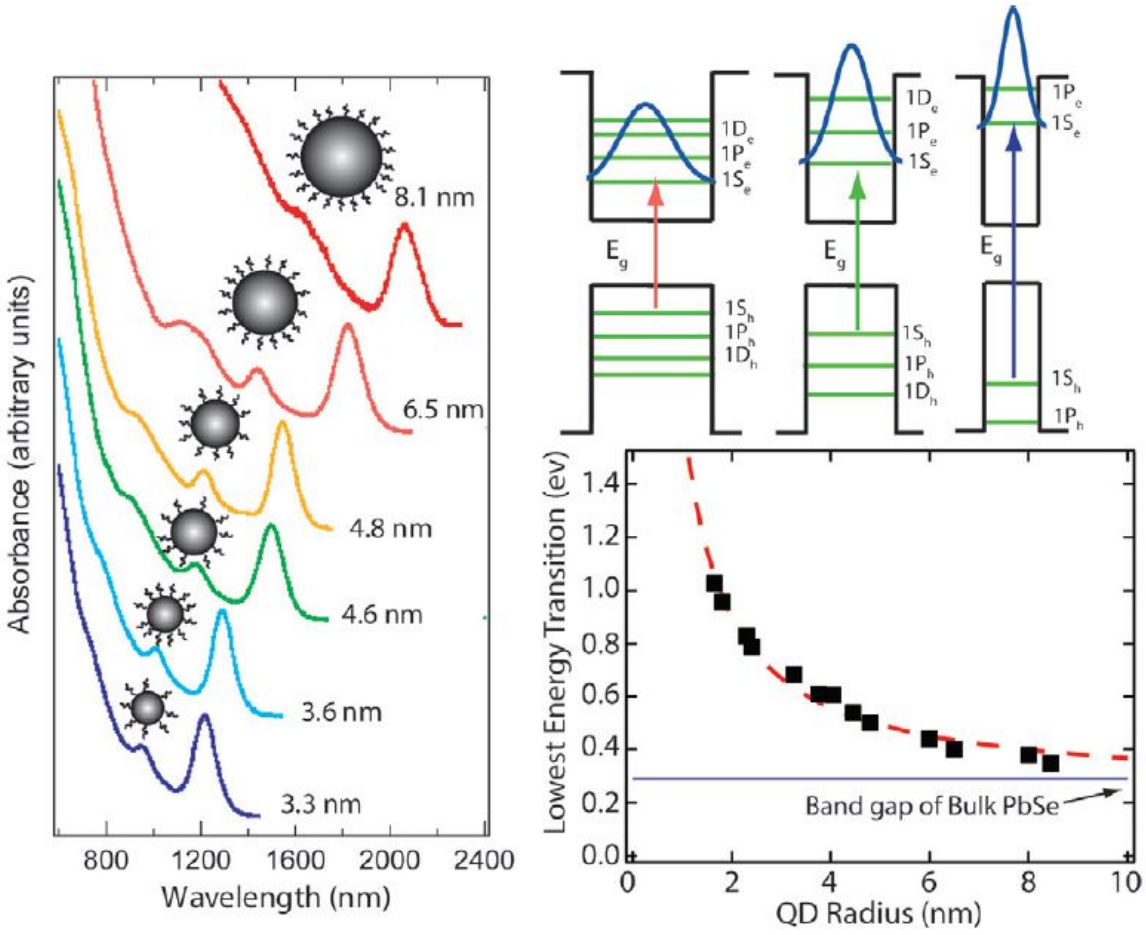


Fig. 5 Absorption spectra of PbSe QDs spanning the range from 3.3 nm to 8.1 nm show strong quantum-confined 1st exciton shifted absorption. Several discrete transitions are observable and represent discrete excitonic transitions. Bottom right shows that for PbSe QDsEg varies approximately as 1/r. Top right depicts the increased quantum confinement that the 1st exciton experiences as the particle size decreases.

Coulomb blockade in quantum dots

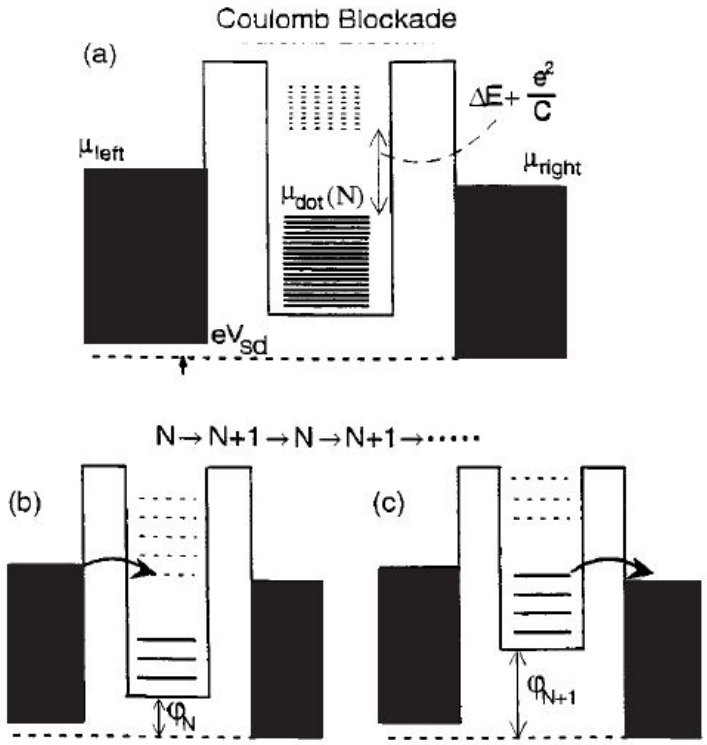


Figure 4. Potential landscape through a quantum dot. The states in the contacts are filled up to the electrochemical potentials μ_{left} and μ_{right}, which are related by the external voltage $V_{sd} = (\mu_{left} - \mu_{right})/e$. The discrete single-particle states in the dot are filled with N electrons up to $\mu_{dot}(N)$. The addition of one electron to the dot raises $\mu_{dot}(N)$ (i.e. the highest solid curve) to $\mu_{dot}(N+1)$ (i.e. the lowest dashed curve). In (a) this addition is blocked at low temperatures. In (b) and (c) the addition is allowed since here $\mu_{dot}(N+1)$ is aligned with the reservoir potentials μ_{left} and μ_{right} by means of the gate voltage. (b) and (c) show two parts of the sequential tunnelling process at the same gate voltage. (b) shows the situation with N and (c) with $N+1$ electrons on the dots.

Coulomb blockade in quantum dots

The electrochemical potential of the dot is defined as $\mu_{\rm dot}(N) \equiv U(N) - U(N-1)$. Electrons can flow from left to right when $\mu_{\rm dot}$ is between the potentials, $\mu_{\rm left}$ and $\mu_{\rm right}$, of the leads (with $eV_{\rm sd} = \mu_{\rm left} - \mu_{\rm right}$), i.e. $\mu_{\rm left} > \mu_{\rm dot}(N) > \mu_{\rm right}$ (figure 4). For small voltages, $V_{\rm sd} \approx 0$, the Nth Coulomb peak is a direct measure of the lowest possible energy state of an N-electron dot, i.e. the GS electrochemical potential $\mu_{\rm dot}(N)$. From equation (1) we obtain

$$\mu_{\rm dot}(N) = (N - N_{\rm o} - 1/2)E_{\rm c} - e(C_{\rm g}/C)V_{\rm g} + E_N. \tag{2}$$

The addition energy is given by

$$\begin{aligned}\Delta\mu(N) &= \mu_{\rm dot}(N+1) - \mu_{\rm dot}(N) = U(N+1) - 2U(N) + U(N-1) \\ &= E_{\rm c} + E_{N+1} - E_N = e^2/C + \Delta E,\end{aligned} \tag{3}$$

with E_N being the topmost filled single-particle state for an N electron dot. The related atomic energies are defined as $A = U(N) - U(N+1)$ for the electron affinity and $I = U(N-1) - U(N)$ for the ionization energy [11]. Their relation to the addition energy is $\Delta\mu(N) = I - A$.

The electrochemical potential is changed linearly by the gate voltage with the proportionality factor $\alpha = (C_{\rm g}/C)$ (equation (2)). The α-factor also relates the peak spacing in the gate voltage to the addition energy: $\Delta\mu(N) = e\alpha(V_{\rm g}^{N+1} - V_{\rm g}^{N})$ where $V_{\rm g}^{N}$ and $V_{\rm g}^{N+1}$ are the gate voltages of the Nth and $(N+1)$th Coulomb peaks, respectively.

Electron tunneling via excited states in quantum dots

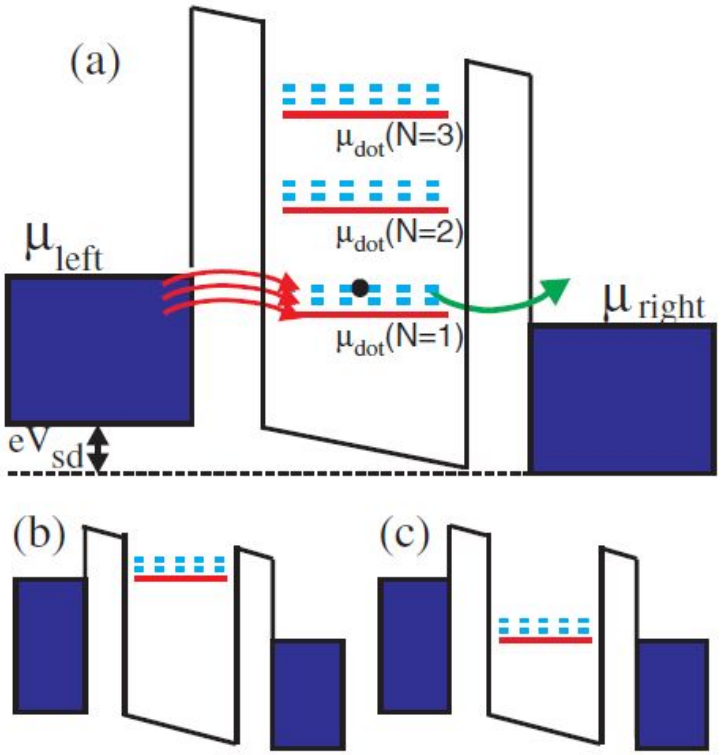


Figure 11. Schematic energy diagram to illustrate tunnelling via ES. The voltage between the source and drain contacts, V_{sd}, opens an energy window, $eV_{sd} = \mu_{left} - \mu_{right}$, between the occupied states in the left and empty states in the right electrodes. Electrons in this window can contribute to the current. (*a*) In this case eV_{sd} is large enough that tunnelling can occur either via the GS or one of the two ESs. (*b*) This alignment of states occurs at a gate voltage corresponding to the lower edge of an excitation stripe (e.g. figure 13). The alignment shown in (*c*) occurs at a gate voltage corresponding to the upper edge of an excitation stripe.