

18

Nanostructures

Written by Professor Paul McEuen of Cornell University

IMAGING TECHNIQUES FOR NANOSTRUCTURES	519
Electron microscopy	520
Optical microscopy	521
Scanning tunneling microscopy	523
Atomic force microscopy	526
ELECTRONIC STRUCTURE OF 1D SYSTEMS	528
One-dimensional (1D) subbands	528
Spectroscopy of Van Hove singularities	529
1D metals—Coulomb interactions and lattice couplings	531
ELECTRICAL TRANSPORT IN 1D	533
Conductance quantization and the Landauer formula	533
Two barriers in series-resonant tunneling	536
Incoherent addition and Ohm's law	538
Localization	539
Voltage probes and the Buttiker-Landauer Formalism	540
ELECTRONIC STRUCTURE OF 0D SYSTEMS	545
Quantized energy levels	545
Semiconductor nanocrystals	545
Metallic dots	547
Discrete charge states	549
ELECTRICAL TRANSPORT IN 0D	551
Coulomb oscillations	551
Spin, Mott Insulators, and the Kondo Effect	554
Cooper pairing in superconducting dots	556
VIBRATIONAL AND THERMAL PROPERTIES	557
Quantized vibrational modes	557
Transverse vibrations	559
Heat capacity and thermal transport	561

SUMMARY	562
PROBLEMS	562
1. Carbon nanotube band structure	562
2. Filling subbands	563
3. Breit-Wigner form of a transmission resonance	563
4. Barriers in series and Ohm's law	563
5. Energies of a spherical quantum dot	564
6. Thermal properties in one dimension	564

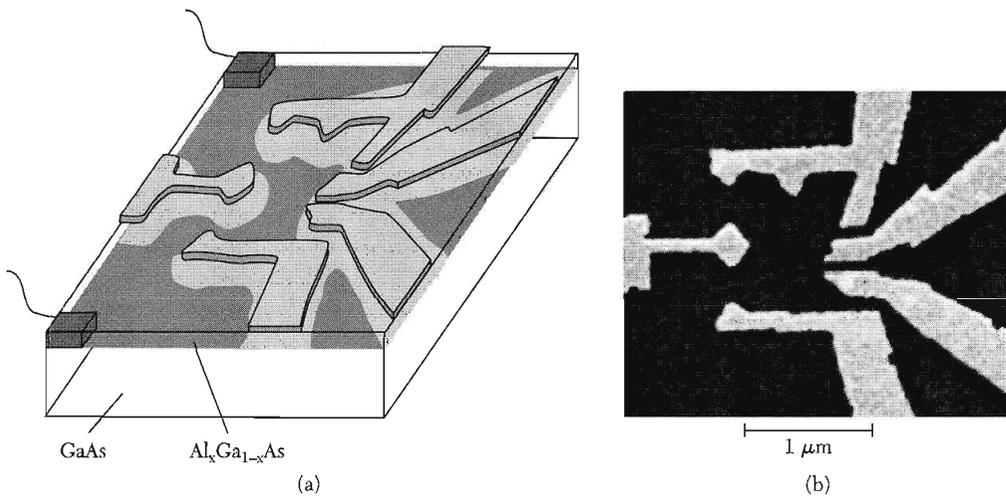


Figure 1 Schematic and scanning electron microscope (SEM) image of a gate electrode pattern on a GaAs/AlGaAs heterostructure used to create a quantum dot of complex shape in the underlying two-dimensional (2D) electron gas. (Courtesy of C. Marcus.)

The previous chapter addressed solids with spatial confinement at the nanometer scale along one direction: surfaces, interfaces, and quantum wells. These systems were effectively two-dimensional, which we define as extended in two directions but of nanometer scale in the third. Only a small number of quantized states—often only one—are occupied in the confined direction. In this chapter we discuss solids confined in either two or three orthogonal directions, creating effectively one-dimensional (1D) or zero-dimensional (0D) **nanostuctures**. Important 1D examples are carbon nanotubes, quantum wires, and conducting polymers. Examples of 0D systems include semiconductor nanocrystals, metal nanoparticles, and lithographically patterned quantum dots. Some examples are shown in Figs. 1 to 3. We will almost exclusively focus on nanostructures that are created from confined periodic solids. Nonperiodic nanostructures are of great interest in other fields, such as molecular assemblies in chemistry and organic macromolecules in biology.

The techniques for the creation of nanostructures can be divided into two broad categories. Top-down approaches use lithographic patterning to structure macroscopic materials at the nanoscale, such as the metallic electrodes on top of a semiconductor heterostructure shown in Fig. 1. Bottom-up approaches utilize growth and self-assembly to build nanostructures from atomic or molecular precursors. A CdSe nanocrystal grown in solution is shown in Fig. 2. It is typically difficult to create structures smaller than 50 nm with

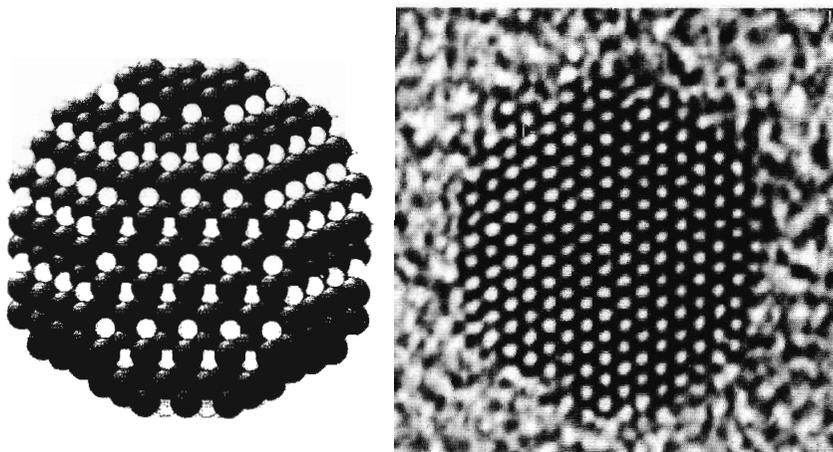


Figure 2 Model and transmission electron microscope (TEM) image of a CdSe nanocrystal. Individual rows of atoms are clearly resolved in the TEM image. (Courtesy of A. P. Alivisatos.)

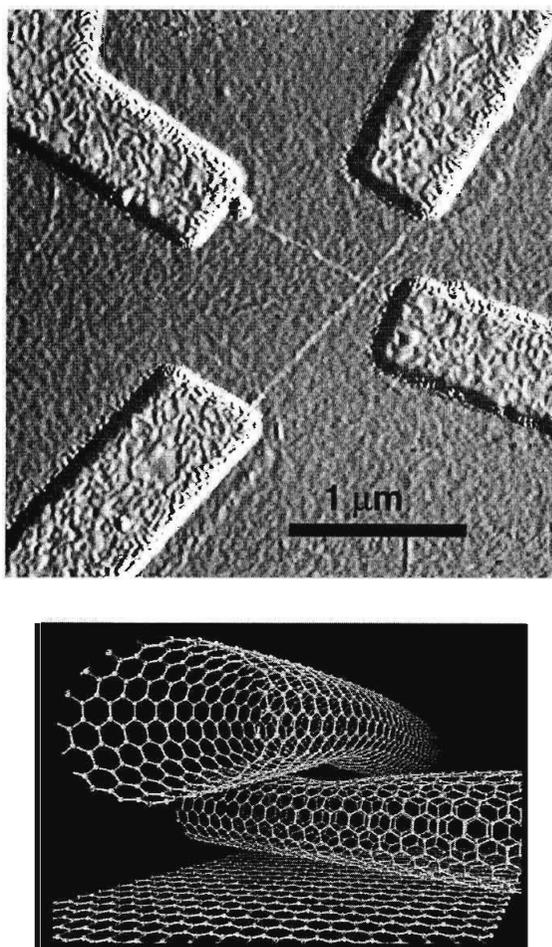


Figure 3 Atomic force microscope (AFM) image of a pair of crossed carbon nanotubes contacted by Au electrodes patterned by electron beam lithography. (Image courtesy of M. S. Fuhrer.) Also shown is a model of the nanotube cross region, showing the honeycomb lattice of the graphene sheets that form the nanotube walls. (Courtesy of P. Avouris.)

top-down techniques, while it is often difficult to create structures larger than 50 nm by bottom-up techniques. A major challenge of nanoscience and technology is to combine these approaches and develop strategies to reliably create complex systems over all length scales, from the molecular to the macroscopic. Figure 3 shows one example, where 100-nm-wide lithographic electrodes make contact to 2-nm-wide carbon nanotubes grown by chemical vapor deposition.

When the extent of a solid is reduced in one or more dimensions, the physical, magnetic, electrical, and optical properties can be dramatically altered. This makes nanostructures a subject of both fundamental and practical interest; their properties can be tailored by controlling their size and shape on

the nanometer scale. One class of effects is related to the large ratio of number of surface atoms to bulk atoms in a nanostructure. For a spherical nanoparticle of radius R composed of atoms with an average spacing a , the ratio is given by

$$N_{\text{surf}}/N \cong 3a/R . \quad (1)$$

For $R = 6a \sim 1$ nm, half of the atoms are on the surface. The large surface area of nanoparticles is advantageous for applications in gas storage, where molecules are adsorbed on the surfaces, or in catalysis, where reactions occur on the surface of the catalyst. It also has dramatic effects on the stability of the nanoparticle. The cohesive energy is dramatically lowered because atoms on the surface are incompletely bonded. Nanoparticles therefore melt at temperatures far below the melting temperature of the corresponding bulk solid.

The fundamental electronic and vibrational excitations of a nanostructure also become quantized, and these excitations determine many of the most important properties of the nanostructured material. These quantization phenomena will be the primary subject of this chapter. Typically, they are important in the 1–100-nanometer size range.

IMAGING TECHNIQUES FOR NANOSTRUCTURES

The development of new techniques to image and probe nanostructures has been essential to the evolution of the field. For periodic 3D structures, the diffraction of electrons or X-rays can be used to determine structure in reciprocal space, which can then be inverted to find the real-space atomic arrangements, as discussed in Chapter 2. For individual nanoscale solids, diffraction is only of limited utility for both fundamental and practical reasons. The solid's small size interrupts the periodicity of the lattice, blurring diffraction peaks, and also produces a very small scattered signal.

Real-space probes that can directly determine the properties of the nanostructure are therefore very valuable. These probes use the interaction of a particle, typically an electron or photon, with the object under study, to create an image. The techniques fall into two major classes, which we will refer to as focal and scanned probe.

In focal microscopy, the probe particle is focused by a series of lenses onto the sample. Figure 4 shows a schematic. The ultimate resolution of the system is limited by the wavelike nature of the particle through the Heisenberg Uncertainty Principle, or, equivalently, diffraction. This smallest feature spacing d that can be resolved is given by

$$d \approx \lambda/2\beta , \quad (2)$$

where λ is the wavelength of the probe particle and $\beta = \sin \theta$ is the numerical aperture defined in Fig. 4. Achieving nanoscale resolution requires using particles with small wavelengths and maximizing the numerical aperture.

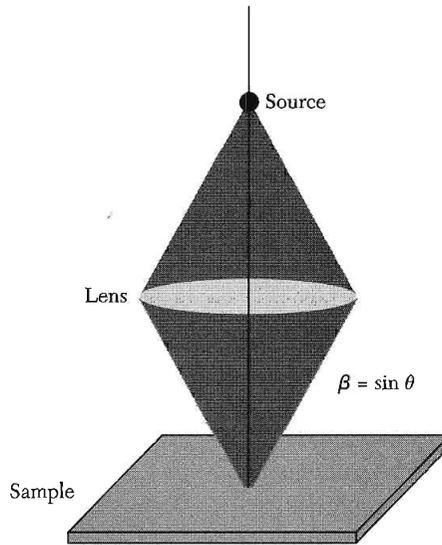


Figure 4 Schematic diagram of a focal microscope. A beam emitted from a source is focused onto the sample by a series of lenses. An equivalent focal system can be used to focus particle/waves emitted from the sample onto a detector.

In scanned probe microscopy, by contrast, a tiny probe is brought close to the sample and scanned over its surface. The resolution of the microscope is determined by the effective range of the interaction between the probe and the structure under study, rather than by the wavelength of the probe particle.

In addition to imaging, scanned and focal probes provide information about the electrical, vibrational, optical, and magnetic properties of individual nanostructures. Of particular importance is the electronic structure, expressed in the density of states. For a finite-sized system, the density of states is a series of delta functions

$$D(\varepsilon) = \sum_j \delta(\varepsilon - \varepsilon_j) , \quad (3)$$

where the sum is taken over all the energy eigenstates of the system. For extended solids, the density of states can be represented by a continuous function, but for a nanostructure the discrete sum form is necessary along the confined directions. This quantized density of states determines many of the most important properties of nanostructures, and it can be directly measured using the techniques described below.

Electron Microscopy

A very powerful focal tool is the electron microscope. A collimated beam of electrons is accelerated by high voltages and focused through a series of electrostatic or magnetic lenses onto the sample under study.

In **transmission electron microscopy**, or TEM, the electron beam travels through the sample and is focused on a detector plate in much the same

way as the image is focused onto the eyepiece of an optical microscope. The ultimate resolving power d is set by the wavelength of the accelerated electrons

$$d = \lambda/2\beta \cong 0.6 \text{ nm}/(\beta\sqrt{V}) , \quad (4)$$

where V is the accelerating voltage (measured in volts). For typical accelerating voltages (100 kV), the theoretical resolving power is therefore subatomic. Other effects, such as imperfections in the lenses, keep the TEM resolution well above this limit, but $d \sim 0.1$ nm has been achieved. Figure 2 shows a TEM image of a semiconductor nanocrystal, where rows of atoms are clearly resolved.

A major limitation of TEM is that the electron beam must penetrate the sample, making it impossible to examine structures on solid substrates. This problem is overcome in the **scanning electron microscope** (SEM). In an SEM, a high-energy (100 V to 100 kV), tightly focused electron beam is scanned over the sample. The number of backscattered electrons and/or the secondary electrons generated by the beam that emerge from the sample depends on the local composition and topography of the sample. These electrons are collected by an electron detector, and an image is formed by plotting this detector signal as a function of the beam location. This powerful technique can be used on most kinds of samples, but it typically has a lower resolution (>1 nm) than the TEM. Figure 1 is an SEM image of metallic electrodes on a GaAs/AlGaAs substrate.

In addition to imaging, the SEM beam can be used to expose an electron-sensitive material and draw small features in a technique known as **electron beam lithography**. The ultimate resolution (<10 nm) is very high, but it is a slow process because the patterns must be drawn pixel by pixel. It is therefore used primarily in research, prototyping, and optical mask fabrication.

Optical Microscopy

The optical microscope is the prototypical focal instrument. Using visible light and a high numerical aperture ($\beta \approx 1$), the highest obtainable resolution is 200–400 nm. For direct imaging, optical microscopy therefore only reaches the edge of the nanoscale realm. However, many of the optical spectroscopies discussed in Chapter 15 have been successfully adapted to study individual nanostructures. These include elastic light scattering, absorption, luminescence, and Raman scattering. Measurements of a single nanostructure, or even a single molecule, are possible if only one is in the field of view of the microscope.

Here we briefly review the emission and absorption of electromagnetic radiation by matter in a manner suitable for applications to nanostructures. Within the electric dipole approximation, Fermi's golden rule gives the transition rate between an initial state i and a higher energy state j due to absorption:

$$w_{i \rightarrow j} = (2\pi/\hbar) \left| \langle j | e\mathbf{E} \cdot \mathbf{r} | i \rangle \right|^2 \delta(\varepsilon_j - \varepsilon_i - \hbar\omega) . \quad (5)$$

Transitions therefore occur between states that have a nonzero dipole matrix element and whose energies differ by the absorbed photon energy $\hbar\omega$. Similarly, the emission rate from state j to i is given by

$$w_{j \rightarrow i} = (2\pi/\hbar) \left| \langle j | e\mathbf{E} \cdot \mathbf{r} | i \rangle \right|^2 \delta(\varepsilon_i - \varepsilon_j + \hbar\omega) + (4\alpha\omega_j^3/c^2) \left| \langle j | \mathbf{r} | i \rangle \right|^2, \quad (6)$$

where $\omega_{ji} = (\varepsilon_j - \varepsilon_i)/\hbar$ and α is the fine structure constant. The first and second terms represent stimulated and spontaneous emission, respectively.

By summing over all possible states, these relations can be used to calculate the total power $\sigma' E^2$ absorbed from the electromagnetic field and hence the real part of the conductivity:

$$\sigma'(\omega) = (\pi e^2 \omega / V) \sum_{ij} \left| \langle j | \hat{\mathbf{n}} \cdot \mathbf{r} | i \rangle \right|^2 [f(\varepsilon_i) - f(\varepsilon_j)] \delta(\varepsilon_j - \varepsilon_i - \hbar\omega), \quad (7)$$

where $\hat{\mathbf{n}}$ is a unit vector pointing in the direction of the electric field. The absorption is proportional to the joint density of all initial and final states separated by an energy $\hbar\omega$, weighted by the dipole matrix element and the occupation factors of the states. The Fermi functions indicate that absorption only occurs when the initial state i is filled and the final state j is empty.

The above relations show that absorption and emission can be used to probe the electronic energy level spectra of nanostructures. Measurements can readily be performed on macroscopic collections of nominally identical nanostructures, but the effects of inhomogeneous broadening due to the variation in the properties of the individual nanostructures are significant. Furthermore, sometimes only a few or even a single nanostructure is available for measurement. Optical measurements that probe single nanostructures have therefore proven to be particularly valuable.

Figure 5 shows an example of spontaneous emission, or fluorescence, from individual optically excited semiconductor quantum dots. The emission occurs from the lowest energy state in the conduction band to the highest energy state in the valence band. The linewidths of the emission lines from single nanocrystals are very narrow, but they are distributed over a range of energies due to variations in the nanocrystal size, shape, and local environment. Measurements of an ensemble therefore show a broad peak that does not accurately reflect the properties of a single nanocrystal.

In addition to their use in probing nanostructures, optical focal systems are also widely used for microfabrication. In projection photolithography, a pattern on a mask is projected onto a photosensitive resist using optical elements. Following exposure and development of the resist, the pattern is transferred into the material of interest by etching or deposition through the resist stencil. Optical lithography is the basis for the mass-fabrication of microelectronic and micromechanical systems. By using wavelengths into the

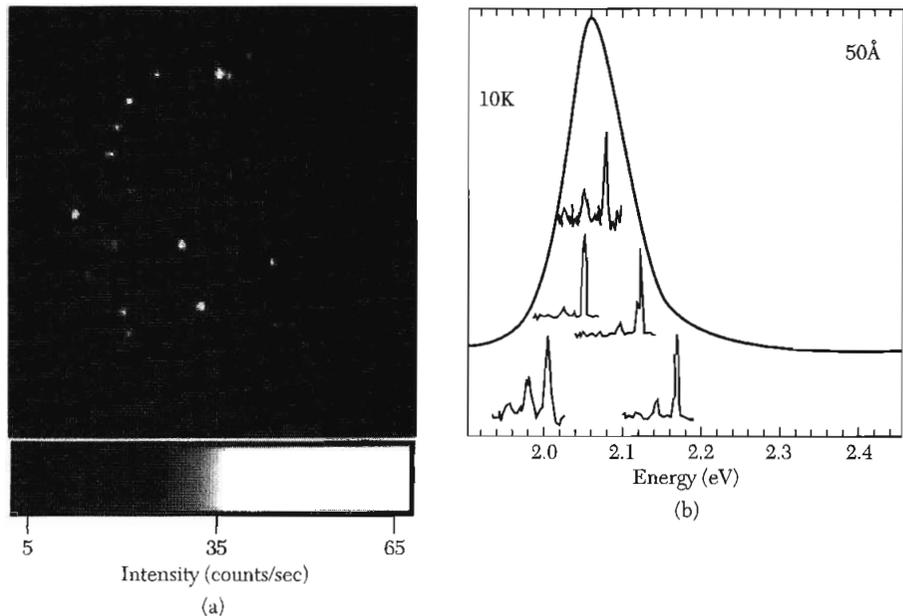


Figure 5 Left: Image of the fluorescence from individual CdSe nanocrystals dilutely distributed on a surface at $T = 10$ K. Right: Spectra of the fluorescence of a number of different individual nanocrystals. In each spectrum, the high-energy peak is the primary transition between the lowest electronic state in the conduction band and the highest energy state in the valence band. The lower energy peaks are associated transitions involving the emission of an LO phonon. Variations in the nanocrystal size and local electronic environment shift the positions of the peaks. The broad peak is the spectrum obtained for an ensemble of nominally identical nanocrystals. (After S. Empedocles et al.)

deep-UV, devices with features of 100 nm are in commercial production. Further improvements using Extreme UV light or even X-rays are possible, but the masks and focusing elements become more and more challenging to fabricate and control.

Scanning Tunneling Microscopy

The most famous scanned probe instrument is the **scanning tunneling microscope** (STM), schematically shown in Fig. 6. Its invention was a breakthrough in the field of nanoscience. In an STM, a sharp metal tip, preferably one with a single atom protruding from the end, is brought to within a nanometer of the conducting sample to be studied. The position of the tip is controlled with picometer precision using piezoelectric materials that expand or contract in response to electrical signals from a control system. A voltage bias V is applied to the sample, and a tunneling current I flowing between the tip and the sample is measured. The current is proportional to \mathfrak{S} , the

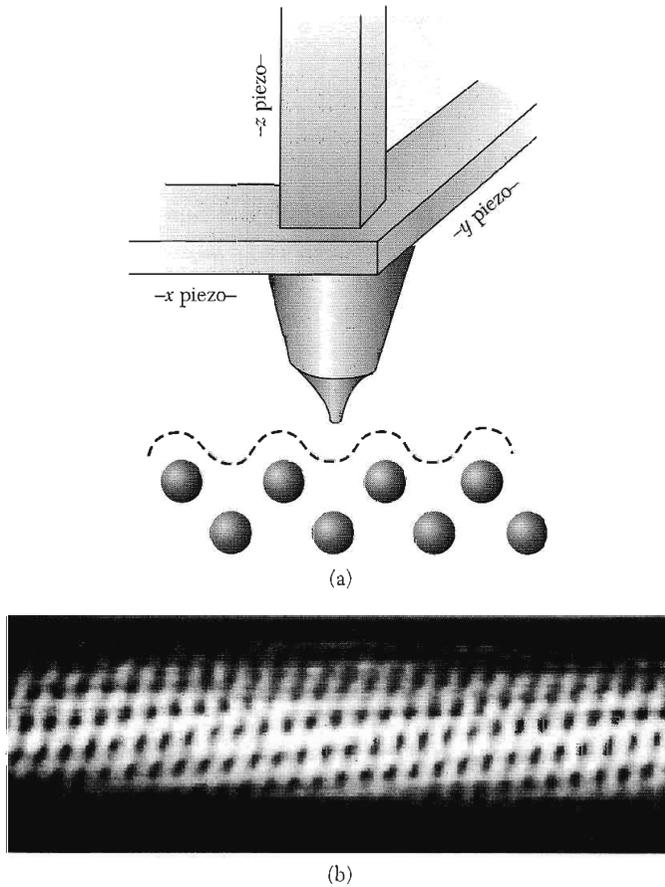


Figure 6 Schematic of a scanning tunneling microscope (STM). When operated in feedback mode, the piezos scan the tip over the sample and maintain a constant tunneling current between the tip and the sample. (Courtesy of D. LePage.) Lower: STM image of a carbon nanotube. (Courtesy of C. Dekker.)

tunneling probability through the gap between the tip and sample. The tunneling probability is exponentially sensitive to the tunneling distance. In the WKB approximation,

$$\mathfrak{S} \propto \exp(-2\sqrt{2m\phi/\hbar^2} z) , \quad (8)$$

where z is the distance between the tip and sample and ϕ is the effective barrier height for tunneling. For typical parameters, a 0.1-nm change in the tip position leads to an order of magnitude change in \mathfrak{S} .

When the STM operates in feedback mode, I is maintained at a constant value by changing the tip height z . The STM thus tracks the surface topography, and very small changes in the height of the surface can be detected

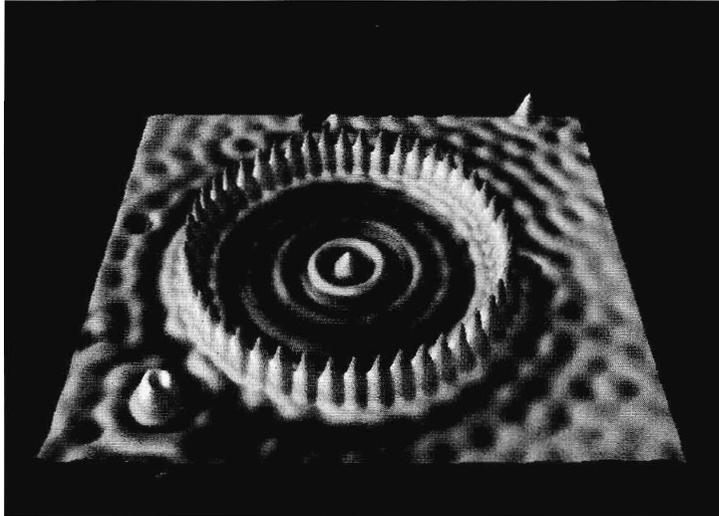


Figure 7 A “quantum corral” of mean radius 7.1 nm was formed by moving 48 Fe atoms on a Cu (111) surface. The Fe atoms scatter the surface state electrons, confining them to the interior of the corral. The rings in the corral are the density distribution of the electrons in the three quantum states of the corral that lie close to the Fermi energy. The atoms were imaged and moved into position by a low-temperature, ultra-high vacuum scanning tunneling microscope. (Image courtesy of D. M. Eigler, IBM Research Division.)

(<1 pm). This is illustrated in Fig. 6, where an STM image of a carbon nanotube is shown. The STM can also be used to manipulate individual atoms on a surface. An example is shown in Fig. 7, where the STM tip is used to construct a “quantum corral” by pushing Fe atoms on a Cu (111) surface into a ring.

The STM tunneling current I as a function of bias V can give spatial and spectroscopic information about the quantum states of a nanostructure. At zero temperature, the derivative of the current with respect to voltage is

$$dI/dV \propto \Im \sum_j |\psi_j(\mathbf{r}_t)|^2 \delta(\varepsilon_F + eV - \varepsilon_j) . \quad (9)$$

It is proportional to the density of states at the tunneling electron energy $\varepsilon_F + eV$, weighted by the electron probability density of those states at the STM tip position \mathbf{r}_t .

For the quantum corral, the electrons in the 2D surface state of Cu are reflected by the Fe atoms, creating a discrete set of states in the interior of the corral. The observed ripples in the image in Fig. 7 are due to the modulations of the probability density $|\psi_j(\mathbf{r}_t)|^2$ of these localized states near the tunneling electron energy. Images at different bias voltages yield the spatial structure of quantized states at different energies.

Atomic Force Microscopy

The **atomic force microscope** (AFM) was developed soon after the STM. It is a much more flexible technique than STM and can be used on both conducting and insulating samples. However, it typically has poorer resolution. An AFM measures the force between the tip and the sample, rather than the tunneling current. A sharp tip is mounted on the end of a millimeter-sized cantilever, as shown in Fig. 8. A force F exerted on the tip by the sample deflects the cantilever by Δz :

$$F = C\Delta z, \quad (10)$$

where C is the force constant of the cantilever. The displacement of the cantilever is measured as a function of tip position, often by using the back of the cantilever as a reflector for a laser beam (Fig. 8). Motion of the reflector changes the path of the laser beam, which is detected using a photodiode array; picometer-scale displacements can easily be measured. Since a typical value of the force constant is $C = 1\text{N/m}$, pN-scale forces can be transduced. Forces well below 1 fN have been measured under special circumstances.

The simplest mode of operation is contact mode, where the tip is dragged along in contact with the surface and the cantilever deflection is measured. This gives a measure of the sample topography, but it can damage the sample. Noncontact or intermittent-contact imaging modes are less invasive, and they also can give information about the long-range forces between the sample and the tip. In these techniques, the cantilever oscillates just above the sample due to an applied driving force of amplitude F_ω near the cantilever resonance

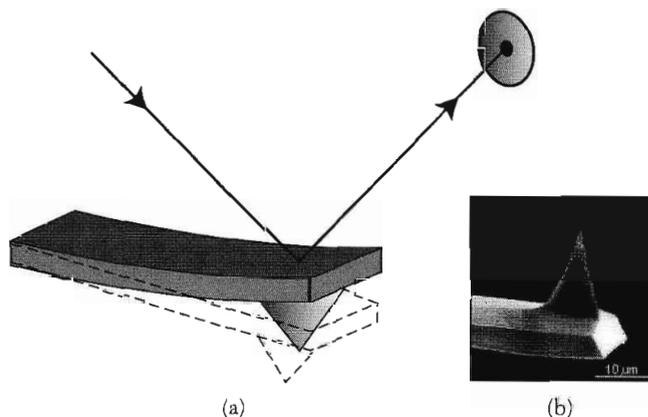


Figure 8 (a) Schematic of an atomic force microscope (AFM). Deflections of the cantilever are measured by a photodetector registering the position of a laser beam that reflects off the top of the cantilever. (Courtesy of Joost Frenken.) Inset: SEM image of an AFM tip. The effective radius of curvature of the tip can be less than 10 nm.

frequency ω_0 . Modeling the cantilever as a driven simple harmonic oscillator, the magnitude of the cantilever response at a frequency ω is given by

$$|z_\omega| = \frac{F_\omega}{C} \frac{\omega_0^2}{[(\omega^2 - \omega_0^2)^2 + (\omega\omega_0/Q)^2]^{1/2}}, \quad (11)$$

where Q , the quality factor of the oscillator, is the ratio of the energy stored in the cantilever to the energy dissipated per cycle. Note that on-resonance, $\omega = \omega_0$, the response is Q times larger than at low frequencies, making the detection of small forces possible.

The parameters characterizing the oscillating cantilever are sensitive to any forces that occur between the tip and the sample. These forces can be van der Waals, electrostatic, magnetic, or many others. The interaction shifts the resonance frequency ω_0 and/or modifies Q . This change is recorded and used to construct an image. For example, in tapping mode imaging, the tip 'taps' the surface during the closest approach of the oscillation cycle, causing both a frequency shift and additional dissipation. The nanotube device shown in Fig. 3 is imaged in tapping mode.

Another important technique is **Magnetic Force Microscopy** (MFM), briefly discussed in Chapter 12. The tip is coated with a magnetic material so that it has a magnetic moment μ normal to the surface of the sample. It then feels a force due to variations in local magnetic fields produced by the sample

$$F(z_0 + \Delta z) = F(z_0) + \left. \frac{\partial F}{\partial z} \right|_{z=z_0} \Delta z = \mu \left. \left(\frac{\partial B}{\partial z} \right) \right|_{z=z_0} + \mu \left. \left(\frac{\partial^2 B}{\partial z^2} \right) \right|_{z=z_0} \Delta z, \quad (12)$$

where z_0 is the tip's equilibrium position and Δz is the displacement during the oscillation. The term $\mu(\partial B/\partial z)$ produces a static deflection of the cantilever, but does not alter the oscillation frequency or the damping. The term $\mu(\partial^2 B/\partial z^2)\Delta z$, on the other hand, has the form of a force constant change δC , since it is linear in the displacement Δz of the cantilever. It therefore shifts the resonance frequency of the cantilever. Monitoring this frequency shift produces an image. Gradients of other local force fields can be similarly measured.

There are many other scanned probe techniques. Near-field scanning optical microscopy (NSOM) creates optical images with a resolution below the diffraction limit by using a scanned subwavelength aperture through which photons 'tunnel'. Scanning capacitance microscopy (SCM) measures capacitance variations between the tip and the sample as a function of position. This ever-growing family of techniques is increasingly used to characterize objects ranging from individual molecules to Si transistors in integrated circuits.

ELECTRONIC STRUCTURE OF 1D SYSTEMS

The quantized electronic states of nanostructures determine their electrical and optical properties, and they influence the physical and chemical properties as well. To describe these states, we take as our starting point the band structure of the bulk material. An effective mass approximation is used for the electronic dispersion of a given band, and the associated wavefunctions are treated as plane waves. These are simplifications; the bands are not always parabolic, and the true eigenstates are Bloch states, not plane waves. However, these assumptions greatly simplify the mathematics and are qualitatively (and often quantitatively) correct. We will also often neglect the Coulomb interactions between electrons. However, there are many cases in the physics of nanostructures where electron–electron interactions cannot be ignored, as discussed later in this chapter.

One-dimensional (1D) Subbands

Consider a nanoscale solid in the geometry of a wire. Its dimensions along the x and y are nanoscale, but it is continuous in z . The energies and eigenstates of such a wire are given by

$$\varepsilon = \varepsilon_{i,j} + \hbar^2 k^2 / 2m ; \quad \psi(x,y,z) = \psi_{i,j}(x,y)e^{ikz} , \quad (13)$$

where i and j are the quantum numbers labeling the eigenstates in the x,y plane and k is the wavevector in the z direction. For the rectangular wire shown in Fig. 9, $\varepsilon_{i,j}$ and $\psi_{i,j}(x,y)$ are just particle-in-a-box energies and eigenstates discussed in Chapter 6.

The dispersion relation consists of a series of 1D subbands, each corresponding to a different transverse energy state $\varepsilon_{i,j}$. The total density of electronic states $D(\varepsilon)$ is the sum of the density of states of the individual subbands:

$$D(\varepsilon) = \sum_{i,j} D_{i,j}(\varepsilon) , \quad (14)$$

where $D_{i,j}(\varepsilon)$ is given by

$$D_{i,j}(\varepsilon) = \frac{dN_{i,j}}{dk} \frac{dk}{d\varepsilon} = (2)(2) \frac{L}{2\pi} \left[\frac{m}{2\hbar^2(\varepsilon - \varepsilon_{i,j})} \right]^{1/2} = \frac{4L}{\hbar v_{i,j}} \quad \text{for } \varepsilon > \varepsilon_{i,j} \quad (15)$$

$$= 0 \quad \text{for } \varepsilon < \varepsilon_{i,j} .$$

The first factor of two in the middle expression is due to spin degeneracy and the second from including both positive and negative values of k . In the right expression, $v_{i,j}$ is the velocity of the electron in the i,j subband with kinetic energy $\varepsilon - \varepsilon_{i,j}$. Note that the density of states diverges as $(\varepsilon - \varepsilon_{i,j})^{-1/2}$ at each subband threshold. These are called **van Hove singularities**. This behavior stands in contrast to three dimensions, where $D(\varepsilon)$ goes to zero at low energies

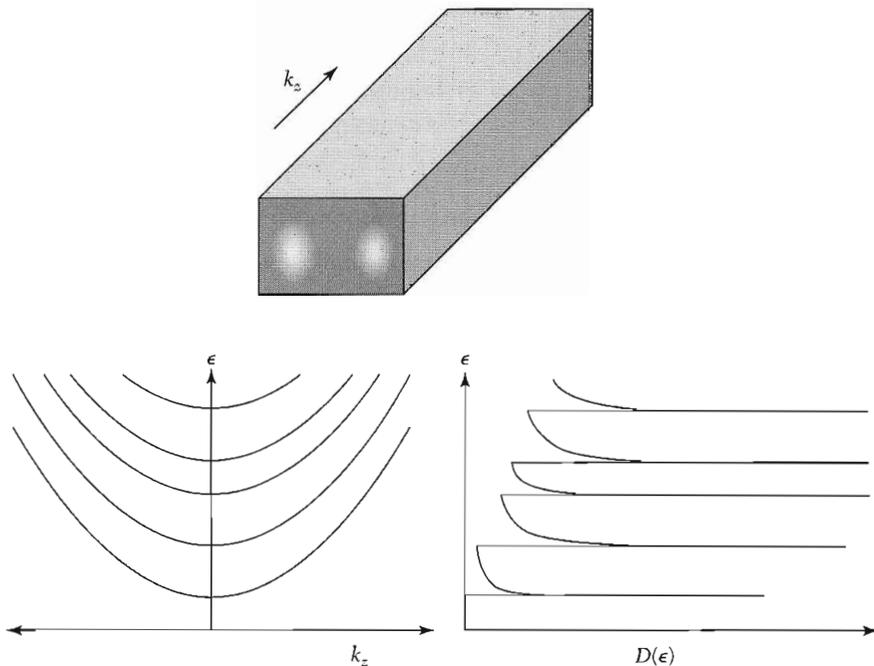


Figure 9 Schematic of a rectangular quasi-one-dimensional wire, along with the dispersion relations and the density of states of the 1D subbands. The peaks in the density of states at the subband thresholds are called Van Hove singularities. The probability density for the $i = 2, j = 1$ state is shown as a gray scale on the cross section of the wire.

(Chapter 6), and two dimensions, where $D(\varepsilon)$ steps up a constant value at the bottom of each 2D subband (Prob. 17.3)

Spectroscopy of Van Hove Singularities

The Van Hove singularities described by (15) affect the electrical and optical properties of 1D systems. Here, we discuss the case of a semiconducting carbon nanotube, whose band structure is calculated in Prob. 1 and shown in Fig. 10a. Van Hove singularities are seen in scanning tunneling spectroscopy, as shown in Fig. 10b. Peaks in the differential conductance, which is proportional to the density of states by (9), are observed at bias voltages corresponding to the energies of these singularities.

The optical absorption and emission of semiconducting nanotubes are also dominated by these singularities, since they depend on the initial and final density of states by (5)–(7). Figure 10c shows the photoluminescence intensity of a collection of carbon nanotubes as a function of the wavelength of the exciting and emitted light. The absorption of the incident light is enhanced when the energy matches $\varepsilon_{c2} - \varepsilon_{v2}$, the energy difference between the 2nd Van Hove singularities in the conduction and valence bands. The electrons and holes relax quickly to the bottom of the first subband, where they recombine,

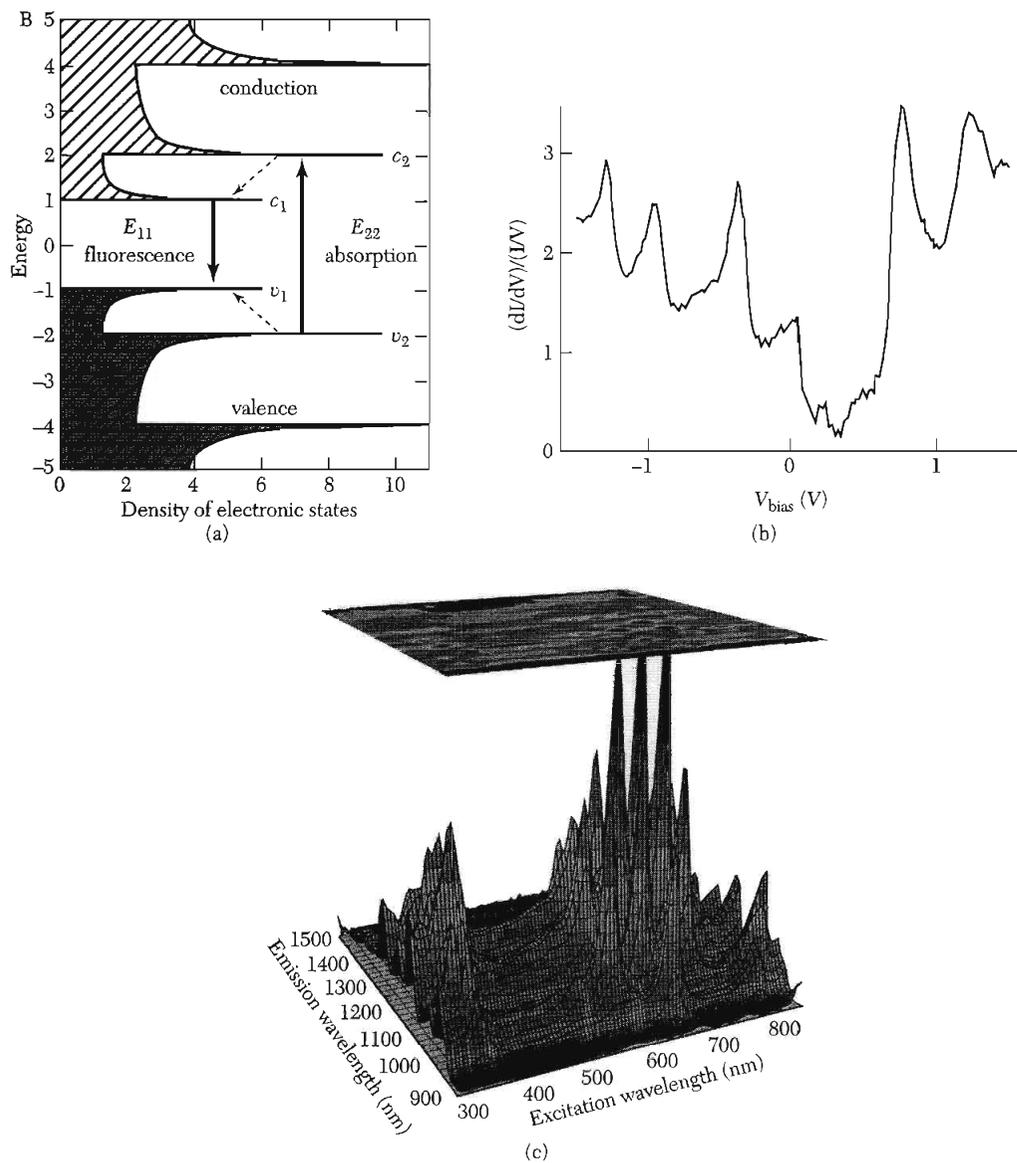


Figure 10 (a) The density of states for a semiconducting carbon nanotube as a function of energy. The Van Hove singularities are seen in the STM tunneling spectra of a nanotube shown in (b). In (c), the emission intensity is plotted as a function of the emission wavelength and the excitation wavelength. Peaks in the intensity are observed when the absorption and emission energies correspond to those shown in the diagram (a). Different peaks correspond to nanotubes with different radii and chirality. [After Bachilo et al. (a and c) and C. Dekker (b).]

producing luminescence with energy $\varepsilon_{c1} - \varepsilon_{v1}$. Peaks are therefore observed when the emitted and absorbed light simultaneously match the energies between the 1st and 2nd van Hove singularities, respectively. Different peaks in the emission intensity in the plot correspond to nanotubes of varying diameters and chiralities.

1D Metals—Coulomb Interactions and Lattice Couplings

In a quasi-one-dimensional metal, electrons fill up individual 1D subbands, with the Fermi energy and the total number of subbands occupied determined by the electron density. For a strictly 1D metal, there is only one (spin-degenerate) subband occupied. In this case, if there are n_{1D} carriers per unit length:

$$n_{1D} = 2k_F/\pi. \quad (16)$$

The Fermi surface of a 1D metal consists of just two points, at $+k_F$ and $-k_F$, as shown in Fig. 11. This is quite different from the Fermi surfaces in 3D and 2D free-electron metals, which consist of a sphere and a circle, respectively. Two consequences of this unusual Fermi surface are discussed below.

Coulomb interactions cause scattering among electrons near the Fermi energy. For 3D metals, scattering is strongly suppressed near ε_F by the restrictions of energy/momentum conservation combined with the Pauli exclusion principle. At an energy ε measured relative to ε_F , $1/\tau_{ee} \approx (1/\tau_0) (\varepsilon/\varepsilon_F)^2$, where $1/\tau_0$ is the

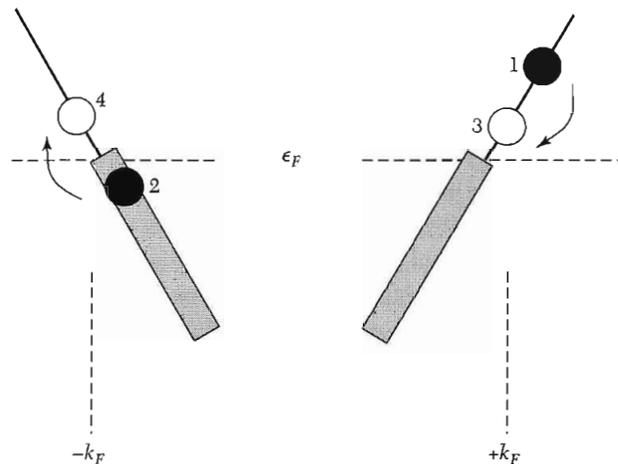


Figure 11 Electronic structure of a 1D metal near the Fermi energy. The Fermi surface consists of two points at $\pm k_F$. The scattering of electrons from filled states 1 and 2 to empty states 3 and 4 conserves energy as long as the energy difference is the same between 1 and 3 and between 2 and 4. Momentum is simultaneously conserved because the energy is locally linear in k .

classical scattering rate. By the uncertainty principle, this produces an uncertainty in the energy of the electron:

$$\delta\epsilon(3D) \approx \hbar/\tau_{ee} \sim (\hbar/\tau_0)(\epsilon/\epsilon_F)^2 . \quad (17)$$

As the energy becomes small (measured relative to ϵ_F), the uncertainty in the energy goes to zero as the second power of ϵ . The uncertainty $\delta\epsilon$ is therefore guaranteed to be small in comparison to ϵ sufficiently close to ϵ_F . This ensures that the quasiparticles near the Fermi surface are well defined.

The case of one dimension is shown in Fig. 11. Energy and momentum conservation are equivalent in this case since for small ϵ the energy is locally linear in the momentum change $\Delta k = k - k_F$:

$$\epsilon \equiv (\hbar^2 k_F / m) \Delta k . \quad (18)$$

Referring to Fig. 11, energy conservation requires that for an electron in state 1 at energy ϵ to scatter to state 3, a simultaneous scattering of an electron from state 2 to 4 must occur. The only restriction is that the final state energy 3 be positive and less than ϵ . This gives a reduction factor $1/\tau_{ee} \sim (1/\tau_0)(\epsilon/\epsilon_F)$, and, from the uncertainty principle,

$$\delta\epsilon(1D) \approx \hbar/\tau_{ee} \sim (\hbar/\tau_0)(\epsilon/\epsilon_F) . \quad (19)$$

Since the uncertainty is linear in ϵ , there is no guarantee that $\delta\epsilon$ will be smaller than ϵ as $\epsilon \rightarrow 0$. The fundamental assumption behind Fermi liquid theory, that weakly interacting quasiparticles exist as $\epsilon \rightarrow 0$, is therefore not guaranteed in 1D. In fact, the ground state of the interacting 1D electron gas is believed *not* to be a Fermi liquid, but rather a Luttinger liquid whose low-energy excitations are collective in nature. The excitations are more analogous to phonons or plasmons—a collective motion of many objects—than isolated electrons moving independently of their neighbors. This collective nature has a number of effects. For example, tunneling into a 1D metal is suppressed at low energies because the tunneling electron must excite the collective modes. In spite of this issue, the independent electron picture remains a useful approximation for the 1D electron gas. It has been successful in describing most, but not all, experiments on real 1D systems, and will be adopted below.

A second unusual property of 1D metals is that they are unstable to perturbations at a wavevector $2k_F$. For example, a distortion of the lattice at this wavevector will open up a bandgap in the electronic spectrum, converting the metal to an insulator. This is the Peierls instability treated in detail in Chapter 14. This effect is particularly important in 1D conducting polymers such as polyacetylene (Fig. 12). It has one conduction electron per carbon atom spacing a , and therefore from (16) $k_F = \pi/2a$. Without any distortions, polyacetylene would have a half-filled band and be a metal. However, a lattice distortion

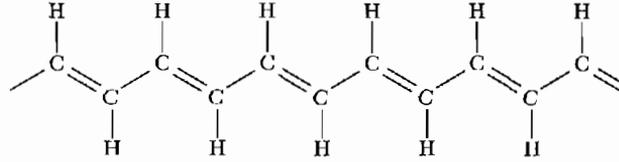


Figure 12 Structure of polyacetylene. Due to the Peierls distortion, the lattice is dimerized, with carbon atoms joined by double bonds in the diagram closer together than those linked by single bonds. The Peierls distortion opens a semiconducting gap of approximately 1.5 eV.

at $2k_F = \pi/a$, corresponding to a wavelength $2a$, opens up a gap at the Fermi energy. This corresponds to a dimerization of the lattice. This dimerization produces alternating single and double bonds along the chain and turns polyacetylene into a semiconductor with a bandgap of 1.5 eV.

Polyacetylene and related semiconducting polymers can be made into field-effect transistors, light-emitting diodes, and other semiconducting devices. They can also be doped chemically, producing metallic behavior with conductivities comparable to traditional metals. However, they retain the mechanical flexibility and ease of processing characteristic of polymers. Their discovery has led to a revolution in flexible plastic electronics.

The Peierls distortion is large in polymers because their backbone consists of a single atomic chain, which can easily distort. Other 1D systems such as nanotubes and nanowires are much stiffer, and the Peierls transition is not observed at experimentally relevant temperatures.

ELECTRICAL TRANSPORT IN 1D

Conductance Quantization and the Landauer Formula

A 1D channel has a finite current-carrying capacity for a given voltage applied across its ends. It therefore has a finite conductance even if there is no scattering in the wire. Consider a wire with one subband occupied connecting two larger reservoirs with a voltage difference V between them, as shown in Fig. 13. The right-going states will be populated up to an electrochemical potential μ_1 and left-going states will be populated up to μ_2 , where $\mu_1 - \mu_2 = qV$ and $q = -e$ for electrons and $+e$ for holes. The net current flowing through the channel due to the excess right-moving carrier density Δn is then

$$I = \Delta n q v = \frac{D_R(\varepsilon) q V}{L} q v = \frac{2}{h v} v q^2 V = \frac{2e^2}{h} V, \quad (20)$$

where $D_R(\varepsilon)$, the density of states of right-moving carriers, is $1/2$ the total density of states given in (15).

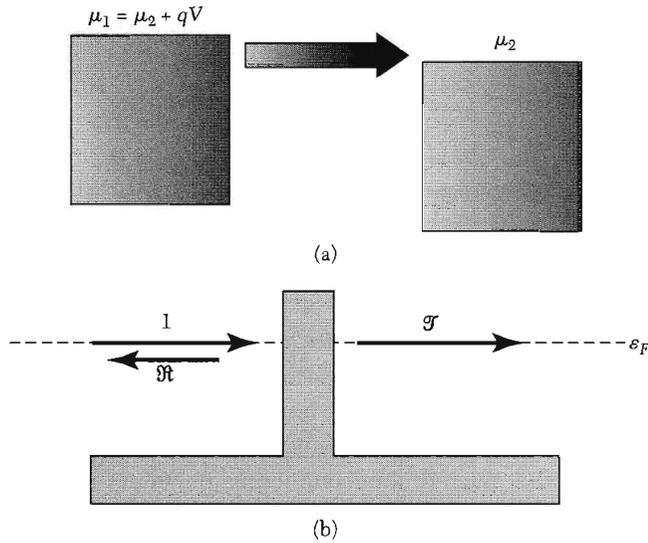


Figure 13 (a) The net current propagating between two reservoirs for an applied bias voltage difference $V_1 - V_2$. (b) Schematic representation of transmission probability \mathfrak{T} and reflection probability \mathfrak{R} from a barrier in the channel, where $\mathfrak{T} + \mathfrak{R} = 1$.

Remarkably, in 1D the velocity exactly cancels with the density of states to create a current that depends only on the voltages and fundamental constants. The two-terminal conductance I/V and resistance V/I are then

$$G_Q = 2e^2/h ; \quad R_Q = h/2e^2 = 12.906 \text{ k}\Omega . \quad (21)$$

A perfectly transmitting one-dimensional channel has a finite conductance whose value is the ratio of fundamental constants. This is called the **conductance quantum** G_Q ; its inverse is called the **resistance quantum** R_Q . While derived here in the effective mass approximation, it is true for a 1D band of arbitrary dispersion.

The quantization of conductance is dramatically illustrated in the data in Fig. 14. A short quasi-1D channel is formed between two regions of a 2D electron gas in a GaAs/AlGaAs heterostructure. As the carrier density of the channel is increased, the conductance increases in discrete steps of height $2e^2/h$. Each step corresponds to the occupancy of an additional 1D subband in the wire. Conductance quantization is also observed in atomic-scale bridges between macroscopic metals.

If the channel is not perfectly conducting, the overall conductance is the quantum of conductance times the probability $\mathfrak{T}(\epsilon_F)$ for electron transmission through the channel (Fig. 13):

$$G(\epsilon_F) = (2e^2/h)\mathfrak{T}(\epsilon_F) . \quad (22)$$

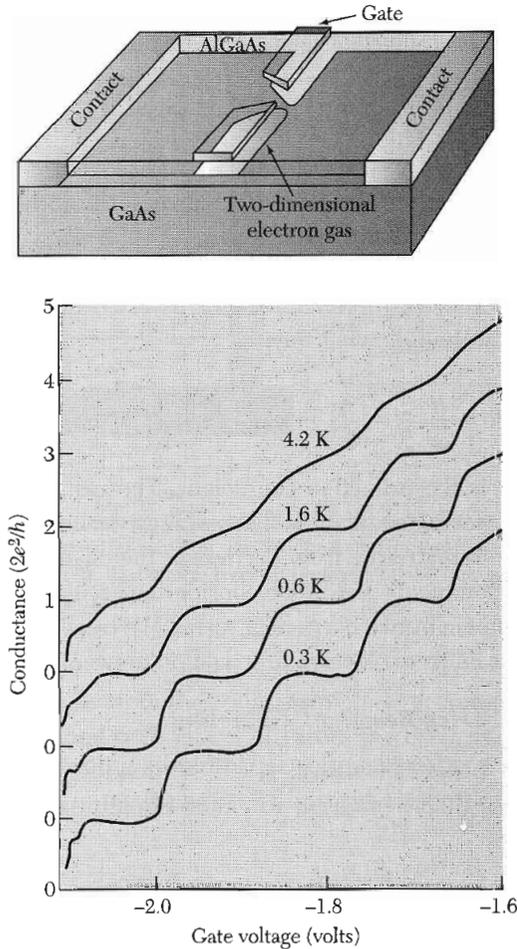


Figure 14 Conductance quantization in a short channel electrostatically defined in a GaAs/AlGaAs heterostructure at different temperatures. A negative gate voltage V_g applied to the metallic gates on the surface of the sample depletes the carriers in the underlying two-dimensional electron gas, creating a narrow channel. The channel is fully depleted of carriers at $V_g = -2.1\text{V}$. Individual 1D subbands become occupied with increasing V_g , with each new subband adding a conductance of $2e^2/h$. (Courtesy of H. Van Houten and C. Beenakker.)

This equation is often called the **Landauer Formula**. For a quasi-1D system with multiple channels, we sum over the contributions of each channel, since conductances in parallel add:

$$\mathfrak{S}(\varepsilon_F) = \sum_{i,j} \mathfrak{S}_{i,j}(\varepsilon_F) , \quad (23)$$

where i, j label the transverse eigenstates. For example, for N perfectly transmitted channels in parallel, $\mathfrak{S} = N$, as for the data in Fig. 14.

At finite temperatures or biases, the Fermi-Dirac energy distributions f of the electrons in the left and right leads must be taken into account:

$$I(\mathcal{E}_F, V, T) = (2e/h) \int_{-\infty}^{\infty} d\mathcal{E} [f_L(\mathcal{E} - eV) - f_R(\mathcal{E})] \mathfrak{S}(\mathcal{E}) . \quad (24)$$

The net current is simply the difference between the left- and right-moving currents, integrated over all energies.

The Landauer formula (22) directly relates the resistance of a system to the transmission properties of the channel. Let us rewrite the resistance for the one-channel case in the following way:

$$R = \frac{h}{2e^2} \frac{1}{\mathfrak{S}} = \frac{h}{2e^2} \frac{\mathfrak{S} + (1 - \mathfrak{S})}{\mathfrak{S}} = \frac{h}{2e^2} + \frac{h}{2e^2} \frac{\mathfrak{R}}{\mathfrak{S}} , \quad (25)$$

where $\mathfrak{R} = 1 - \mathfrak{S}$ is the reflection coefficient. The resistance of the device is the sum of the first term, the quantized contact resistance, and the second term, the resistance due to scattering from barriers in the channel. The latter term is zero for a perfect conductor. Below we consider an application of the Landauer formula to the problem of two barriers in series. We treat this in both the coherent and incoherent limits of electron propagation between the barriers.

Two Barriers in Series-Resonant Tunneling

Consider two barriers in series separated by a distance L , with transmission/reflection amplitudes t_1, r_1 and t_2, r_2 , as shown in Fig. 15. These amplitudes are complex:

$$t_j = |t_j| e^{i\varphi_j} ; \quad r_j = |r_j| e^{i\varphi_j} . \quad (26)$$

To calculate the transmission probability \mathfrak{S} through the entire double barrier structure, we need the corresponding transmission amplitude. For an incident wave from the left whose amplitude is 1, the amplitudes defined in Fig. 15 are given by

$$a = t_1 + r_1 b ; \quad b = ar_2 e^{i\varphi} ; \quad c = at_2 e^{i\varphi/2} , \quad (27)$$

where $\varphi = 2kL$ is the phase that an electron with kinetic energy $\hbar^2 k^2 / 2m$ accumulates propagating the distance $2L$ on a round trip between the barriers. Combining these to solve for the transmitted amplitude yields:

$$c = \frac{t_1 t_2 e^{i\varphi/2}}{1 - r_1 r_2 e^{i\varphi}} . \quad (28)$$

The transmission probability through the double barrier is then

$$\mathfrak{S} = |c|^2 = \frac{|t_1|^2 |t_2|^2}{1 + |r_1|^2 |r_2|^2 - 2|r_1||r_2|\cos(\varphi^*)} \quad (29)$$

where $\varphi^* = 2kL + \varphi_{r1} + \varphi_{r2}$.

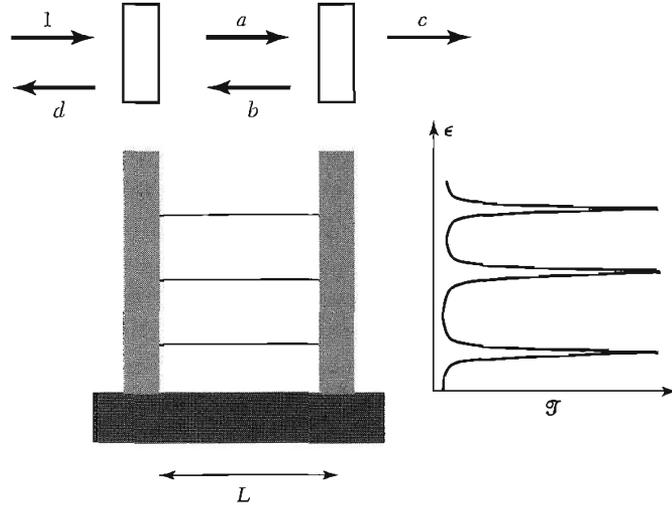


Figure 15 Resonant tunneling through two identical barriers in series separated by a length L . The upper diagram shows the transmission amplitudes between and outside the barriers for a unity amplitude incident wave. The transmission resonances at the energies of the quasibound states between the barriers are shown.

This is plotted in Fig. 15. Note the round-trip phase accumulation φ^* includes the phase shifts associated with reflections from the barriers.

The transmission probability (29) is greatly enhanced when $\cos(\varphi^*)$ approaches unity, because the denominator becomes small. This occurs for the resonance condition

$$\varphi^* = 2kL + \varphi_{r1} + \varphi_{r2} = 2\pi n \quad (30)$$

where n is an integer. This is a general property of waves, and is due to the constructive interference of many pathways through the sample. This can be easily seen by rewriting (28) using the series expansion $1/(1-x) = \sum_{m=0}^{\infty} x^m$:

$$c = t_1 t_2 e^{i\varphi/2} / (1 - r_1 r_2 e^{i\varphi}) = t_1 t_2 e^{i\varphi/2} [1 + r_1 r_2 e^{i\varphi} + (r_1 r_2 e^{i\varphi})^2 + \dots] \quad (31)$$

The m th order in the expansion corresponds to a path with m round trips between the barriers. On resonance, these paths add in phase to yield a strongly enhanced transmission.

Consider the special case where the barriers are the same: $t_1 = t_2$. We then have

$$\Im(\varphi^* = 2\pi n) = |t_1|^4 (1 - |r_1|^2)^{-2} = 1 \quad (32)$$

The transmission on resonance through a symmetric double-barrier structure is 1, even if the transmission through each of the individual barriers is small.

This is called **resonant tunneling**. Off resonance, the denominator of (29) is of order unity for opaque barriers, and the transmission is roughly the product of the transmission coefficients of each of the two barriers in series: $\mathfrak{T} \sim |t_1|^2 |t_2|^2$.

The resonance condition $\varphi^* = 2\pi n$ corresponds to the energies of the quasibound electronic states confined between the two barriers. For very opaque walls, this is just the particle-in-a-box quantization condition: $kL = \pi n$. We derived the resonant tunneling condition for a one-dimensional case, but it is a general result. The transmission through a confined electron system is strongly enhanced at energies corresponding to the bound-state energy levels of the confined electrons. This is also evident from the STM tunneling expression (9); quasibound states produce peaks in the differential conductance.

For the case of opaque barriers, $|t_1|^2, |t_2|^2 \ll 1$, the cosine term in the denominator of (29) can be expanded, as shown in Prob. 3, yielding the familiar Breit-Wigner form for a resonance:

$$\mathfrak{T}(\varepsilon) = \frac{4\Gamma_1\Gamma_2}{(\Gamma_1 + \Gamma_2)^2 + 4(\varepsilon - \varepsilon_n)^2} \quad \text{where} \quad \Gamma_j = \frac{\Delta\varepsilon}{2\pi} |t_j|^2. \quad (33)$$

The resonances are thus Lorentzian peaks with a width in energy of $\Gamma = \Gamma_1 + \Gamma_2$ determined by the energy level spacing $\Delta\varepsilon$ and the transmission probabilities through the two barriers. This is just the uncertainty principle broadening of the level due to the finite lifetime of the double-barrier bound state.

Incoherent Addition and Ohm's Law

If we instead treat the electron classically, we add probabilities rather than amplitudes. This is valid if the electron effectively loses track of its phase between the barriers due to, for example, inelastic scattering from phonons. This corresponds to replacing (27) by

$$|a|^2 = |t_1|^2 + |r_1|^2 |b|^2; \quad |b|^2 = |a|^2 |r_2|^2; \quad |c|^2 = |a|^2 |t_2|^2. \quad (34)$$

This gives

$$\mathfrak{T} = \frac{|t_1|^2 |t_2|^2}{1 - |r_1|^2 |r_2|^2}. \quad (35)$$

Some elementary manipulations (Prob. 4) yield

$$R = (h/2e^2)(1 + |r_1|^2/|t_1|^2 + |r_2|^2/|t_2|^2). \quad (36)$$

The resistance is just the sum of the quantized contact resistance and the intrinsic resistances of the individual barriers (see Eq. 25). This is Ohm's law—resistors in series add. It is valid if interference effects can be neglected.

Equation (36) allows us to connect to the Drude formula. Consider a process that gives a backscattering rate $1/\tau_b$. This backscattering could result from either an elastic scattering process such as impurity scattering or from an inelastic

process such as phonon scattering. For propagation over a small distance dL , the reflection probability $d\mathfrak{R}$ ($\ll 1$) is

$$d\mathfrak{R} = \frac{1}{\tau_b} \frac{dL}{v_F} = \frac{dL}{\ell_b} . \quad (37)$$

This gives a contribution to the resistance, yielding a resistivity:

$$\rho_{1D} = dR/dL = (h/2e^2)/\ell_b . \quad (38)$$

This is equal to the 1D Drude resistance $\sigma_{1D}^{-1} = (n_{1D}e^2\tau/m)^{-1}$, as shown in Prob. 4. Ignoring interference effects, the resistances of individual segments add ohmically, giving

$$R = R_Q + (h/2e^2)(L/\ell_b) . \quad (39)$$

Localization

Now consider when two barriers are connected in series, but coherence is *not* neglected. However, we average over all possible phases, corresponding to an average over different energies. From Eq. (29), the average resistance is

$$\langle R \rangle = \frac{h}{2e^2} \frac{1 + |r_1|^2|r_2|^2 - 2|r_1||r_2|\langle \cos\varphi^* \rangle}{|t_1|^2|t_2|^2} = \frac{h}{2e^2} \frac{1 + |r_1|^2|r_2|^2}{|t_1|^2|t_2|^2} . \quad (40)$$

Notably, the phase-averaged resistance (40) is *larger* than the resistance in the incoherent limit (36).

To understand the scaling with length associated with (40), consider a long conductor of length L consisting of a series of only elastic (phase-preserving) scatterers characterized by an elastic backscattering length ℓ_e . Assume that the conductor has a large resistance $\langle R \rangle$, so that $\mathfrak{R} \approx 1$ and $\mathfrak{S} \ll 1$. For a small additional length dL , there will be an additional reflection and transmission $d\mathfrak{R} = dL/\ell_e$, as in (37), and $d\mathfrak{S} = 1 - d\mathfrak{R}$. Combining these according to the prescription of (40), and assuming that $d\mathfrak{R} \ll 1$, gives

$$\langle R + dR \rangle = \frac{h}{2e^2} \frac{1 + \mathfrak{R}d\mathfrak{R}}{\mathfrak{S}(1 - d\mathfrak{R})} \approx \langle R \rangle \left(1 + \frac{2dL}{\ell_e} \right) , \quad (41)$$

or equivalently,

$$\langle dR \rangle = \langle R \rangle (2dL/\ell_e) . \quad (42)$$

Separating variables and integrating both sides of the equation yields

$$\langle R \rangle = (h/2e^2) \exp(2L/\ell_e) . \quad (43)$$

Remarkably, the resistance grows exponentially with the length of the sample, rather than linearly as in an ohmic conductor. This behavior is a result of localization. Due to quantum interference among the states scattered by disorder, the states become localized on a size scale $\xi \sim \ell_e$, where ξ is called the localization

length. There are no extended states that traverse the entire length of the conductor, so the resistance is exponentially large. A similar result holds for quasi-1D systems, but with a localization length $\xi \sim N\ell_e$, where N is the number of 1D subbands occupied.

At very low temperatures, only coherent scattering processes occur and the resistance is exponentially large by (43). At finite temperatures, electrons retain their phase memory only over the phase coherence length ℓ_ϕ due to their interaction with other degrees of freedom such as phonons or electrons. This length typically is a power-law function of temperature, $\ell_\phi = AT^{-\alpha}$, since the number of electronic and vibrational excitations present is a power law in T . The resistance of each phase coherent segment can be approximated by (43) with ℓ_ϕ replacing L . The resistance of each phase coherent segment decreases rapidly with increasing temperature (as the exponential of a power law in T). This dramatically decreases the overall resistance, which is the (incoherent) series combination of L/ℓ_ϕ such phase coherent sections. At a sufficiently high temperature where $\ell_\phi \leq \ell_e$, all phase coherence is lost between scattering events and the ohmic expression (36) is applicable.

A related issue is the nature of the electronic states in 2D and 3D systems in the presence of disorder. In 2D, it is believed that, for noninteracting electrons, all states are also localized by disorder. In 3D, on the other hand, a critical amount of disorder is required to localize the states. The subject of localization continues to be of great fundamental interest and controversy, particularly when the effects of coulomb interactions between the electrons are included.

Voltage Probes and the Büttiker-Landauer Formalism

In many measurements, more than two probes are connected to a conductor. Some are used as voltage probes (which draw no net current from the sample) and others as current probes, as shown in Fig. 16. Büttiker extended the Landauer formalism to deal with this multiprobe case. Define $\mathfrak{G}^{(n,m)}$ as the total transmission probability for an electron leaving contact m to arrive at contact n , including the contributions from all the 1D channels. For a current probe n with N_n channels, the electrochemical potential of the contact is fixed by an applied voltage, and the net current that flows through the contact is

$$I_n = (2e^2/h)(N_n V_n - \sum_m \mathfrak{G}^{(n,m)} V_m) . \quad (44)$$

This is just the current flowing out of the contact minus the currents flowing in that originated from each of the other contacts. Note that $N_n = \sum_m \mathfrak{G}^{(n,m)}$, which can be easily obtained from (44) by considering the equilibrium case where all the voltages are equal and all the currents are zero.

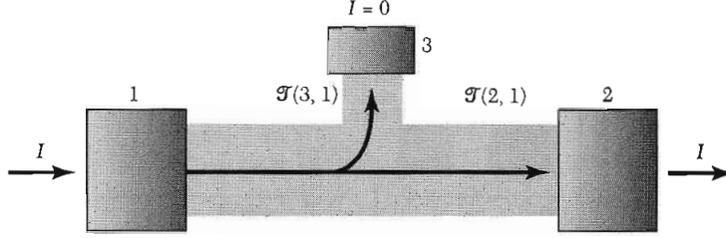


Figure 16 Schematic representation of a multiterminal conductor. Contacts 1 and 2 are current probes; contact 3 is a voltage probe. The transmission probability from contact 1 to 2 and from 1 to 3 is schematically indicated.

For a voltage probe, the potential V_n adjusts itself so that no net current flows ($I_n = 0$):

$$V_n = \frac{\sum_{m \neq n} \mathfrak{T}^{(n,m)} V_m}{\sum_{m \neq n} \mathfrak{T}^{(n,m)}} . \quad (45)$$

The electrochemical potential measured by the probe is the weighted average of the electrochemical potentials of the different contacts, where the weighting coefficients are the transmission probabilities.

Equations (44–45) have a number of surprising consequences. Since the measured currents and voltages depend on $\mathfrak{T}^{(n,m)}$, the details of the path that an electron takes in traversing the sample influences the resistance. A voltage probe can disturb the paths, and the measured voltage can in turn be affected by transmission through all parts of the sample. Below we present three examples that illustrate these properties.

Consider a voltage probe connected to the center of an otherwise ballistic 1D conductor, as shown in Fig. 16. Assume that electrons leaving from probe 1 either arrive at probe 2 or 3, but none are directly backscattered. The voltage read by probe 3 is then

$$V_3 = \frac{\mathfrak{T}^{(3,1)} V}{\mathfrak{T}^{(3,1)} + \mathfrak{T}^{(3,2)}} = \frac{V}{2} , \quad (46)$$

where for the last step we assumed that the voltage probe couples symmetrically to the left and right moving channels, $\mathfrak{T}^{(3,1)} = \mathfrak{T}^{(3,2)}$. The voltage measured in the channel is just the average of the voltage of the two contacts.

The current flowing out of contact 1 is given by:

$$I = (2e^2/h)(V - \mathfrak{T}^{(1,3)} V_3) = (2e^2/h)V(1 - \frac{1}{2}\mathfrak{T}^{(1,3)}) , \quad (47)$$

where (46) has been employed in the second step. Note that the presence of the voltage probe decreases the transmission below the unity value of a perfect

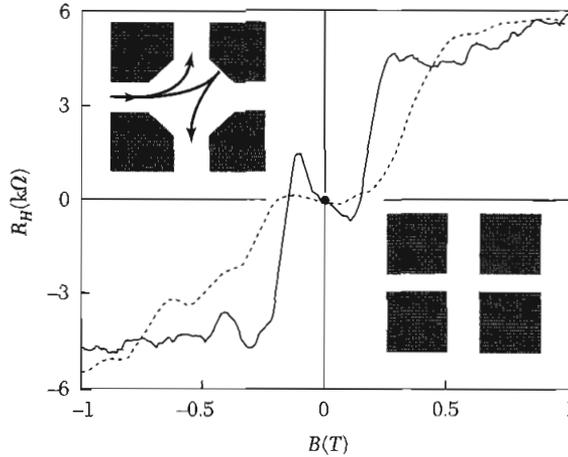


Figure 17 Four-terminal Hall resistance measurements of submicron junctions of different shapes. In the junction shown schematically in the upper left, the Hall resistance is negative at small B and positive at large B . The reason is indicated in the diagram; at small B , the electrons bounce off the wall into the “wrong” probe. (After C. Ford et al.)

channel. Some of the electrons scatter into the voltage probe, are re-emitted, and then return to contact 1. This shows that voltage probes are in general invasive; they influence what they measure unless they only couple very weakly to the system.

Figure 17 shows a measurement of the Hall resistance of two nanoscale crosses patterned in a high mobility 2D electron gas whose geometries are shown in the insets. The junction region is ballistic, meaning that there is no scattering from disorder, only from the sample walls. The measured Hall resistance is not of the form $B/n_s e$, where n_s is the sheet carrier concentration, as expected for a macroscopic 2D electron gas, but has instead a number of notable features. Most surprisingly, the Hall voltage is of the opposite sign at low B compared to high B for the sample shown in the upper left inset. This can be easily understood from the shape of the classical electron paths sketched on the figure. At high B , the Lorentz force preferentially deflects the electron into the upper electrode, giving the expected sign of the Hall voltage. At low B , however, the electron bounces off the boundary of the conductor and arrives at the lower electrode, reversing the sign of the measured Hall voltage. For a small multiprobe conductor, the resistance is a measure of the electron trajectories through the sample and not simply related to intrinsic material properties like the electron density.

Equations (44–45) can be used to treat arbitrarily complex microscopic (or even macroscopic) conductors. It has been widely used to describe measurements on small disordered metal samples at low temperatures as a function of magnetic field B . These samples have many transverse channels and contain impurities. The elastic scattering length ℓ_e is less than the sample dimensions,

but the phase coherence length ℓ_φ is greater. Electrons therefore propagate diffusively, but phase-coherently, through the sample. This is called the **mesoscopic** regime. In a semiclassical picture, the transmission amplitude between two probes n and m corresponds to the sum of many different classical paths through the sample:

$$(CGS) \quad t^{(m,n)} \propto \sum_j a_j \exp\left(\frac{i}{\hbar} \int_\ell^m (\mathbf{p} - e\mathbf{A}/c) \cdot d\mathbf{l}\right);$$

$$(SI) \quad t^{(m,n)} \propto \sum_j a_j \exp\left(\frac{i}{\hbar} \int_\ell^m (\mathbf{p} - e\mathbf{A}) \cdot d\mathbf{l}\right). \quad (48)$$

Note that the phase associated with each path amplitude a_j contains a contribution from the magnetic vector potential \mathbf{A} , as described in Appendix G. Since $\mathfrak{T}^{(n,m)} = |t^{(n,m)}|^2$, quantum interference among different conducting pathways through the sample modulates the transmission.

An interesting example is shown in Fig. 18. On the left, the four-terminal resistance of a nanoscale metallic wire is shown. Aperiodic fluctuations are seen in the conductance versus magnetic field B . These fluctuations are due to modulations of the interference between the many diffusive paths linking the contacts. Since there are many paths, the result is an essentially random variation. These modulations are referred to as conductance fluctuations.

When an additional loop is added that is outside the region between the contacts, as shown on the right of Fig. 18, the conductivity G qualitatively changes. A periodic modulation with magnetic field is seen. This is due to the **Aharonov-Bohm effect**. The vector potential modulates the quantum interference between those electron paths that encircle the ring and those that do not. For simplicity, consider the interference between just two such paths with transmission amplitudes a_1 and a_2 (Fig. 18) in the absence of a magnetic field. At finite B ,

$$(CGS) \quad |a_1 + a_2 \exp[(ie/hc) \oint_{loop} \mathbf{A} \cdot d\mathbf{l}]|^2 = |a_1|^2 + |a_2|^2 + 2|a_1||a_2| \cos[2\pi\Phi/(hc/e)];$$

$$(SI) \quad |a_1 + a_2 \exp[(ie/\hbar) \oint_{loop} \mathbf{A} \cdot d\mathbf{l}]|^2 = |a_1|^2 + |a_2|^2 + 2|a_1||a_2| \cos[2\pi\Phi/(h/e)]. \quad (49)$$

In the last step we have employed Stokes's theorem, where Φ is the magnetic flux going through the loop and hc/e is the magnetic flux quantum (h/e in SI). With increasing flux, the transmission through the wire oscillates with the period of one flux quantum. This effect is closely related to the superconducting flux quantization discussed in Chapter 10, except that the charge appearing in the flux quantum here is e , not $2e$, since the carriers here are electrons, not Cooper pairs.

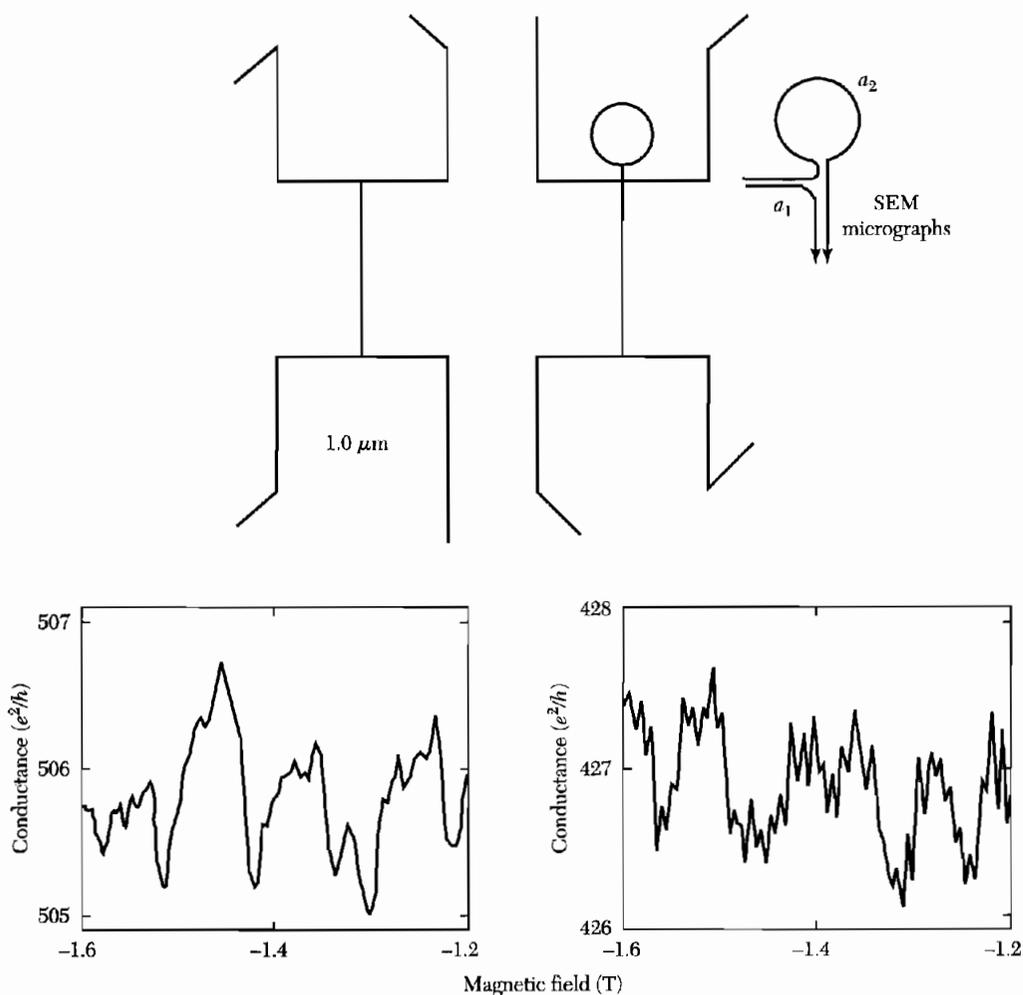


Figure 18 Upper: SEM micrographs of two vertical Au wires with current and voltage probes attached. In the device on the right, an extra loop has been added outside the region between the probes. The diagram to the right shows two paths, one that encircles the ring and one that does not. Lower left: Conductance versus magnetic field for the left sample. Aperiodic conductance fluctuations are seen due to quantum interference between the conducting paths through the sample. Lower right: Periodic oscillations are observed associated with the Aharonov-Bohm effect for paths enclosing the loop nominally outside the region between the contacts, showing the non-local nature of diffusive coherent transport in mesoscopic systems. (After R. Webb.)

It is remarkable that the addition of a loop outside of the region between the voltage contacts changes the measured properties. Resistances in the mesoscopic regime are nonlocal. Electrons coherently diffuse throughout the entire sample while journeying between contacts, and their phase remembers the journey.

ELECTRONIC STRUCTURE OF 0D SYSTEMS

Quantized Energy Levels

A system of electrons fully confined in all three dimensions will have discrete charge and electronic states, as do atoms and molecules. They are often called artificial atoms or **quantum dots** to reflect the importance of quantization phenomena on their properties.

As a simple example, consider an electron in a spherical potential well. Due to the spherical symmetry, the Hamiltonian separates into angular and radial parts, giving eigenstates and eigenenergies:

$$\varepsilon_{n,l,m} = \varepsilon_{n,l} ; \quad \psi(r,\theta,\phi) = Y_{l,m}(\theta,\phi)R_{n,l}(r) , \quad (50)$$

where $Y_{l,m}(\theta,\phi)$ are the spherical harmonics and $R_{n,l}(r)$ are the radial wavefunctions. The energy levels and radial wavefunctions depend on the details of the particular confining potential. For an infinite spherical well, where $V = 0$ for $r < R$ and is infinite otherwise,

$$\varepsilon_{n,l} = \hbar^2 \beta_{n,l}^2 / (2m^* R^2) ,$$

$$R_{n,l}(r) = j_l(\beta_{n,l} r/R) , \quad r < R . \quad (51)$$

The function $j_l(x)$ is the l th spherical Bessel function and the coefficient $\beta_{n,l}$ is the n th zero of $j_l(x)$. For example, $\beta_{0,0} = \pi$ (1S), $\beta_{0,1} = 4.5$ (1P), $\beta_{0,2} = 5.8$ (1D), $\beta_{1,0} = 2\pi$ (2S), and $\beta_{1,1} = 7.7$ (2P). The labels in parentheses are the atomic notations for the states, which have the usual degeneracies associated with spin and angular momentum orientation.

Semiconductor Nanocrystals

A semiconductor nanocrystal such as the one shown in Fig. 2 can, to a good approximation, be described by the spherical model given above. Both the electron states in the conduction band and the hole states in the valence band are quantized. For a CdSe nanoparticle, the conduction band effective mass $m_e^* = 0.13 m$, and the electron energy levels are $\varepsilon_{n,l} = (2.9 \text{ eV}/R^2) (\beta_{n,l}/\beta_{0,0})^2$, where R , the radius of the nanoparticle, is expressed in nanometers. For $R = 2$ nm, the spacing between the lowest two energy levels is $\varepsilon_{0,1} - \varepsilon_{0,0} = 0.76$ eV.

The 1S electron state increases in energy with decreasing R , while the 1S hole state decreases. The bandgap therefore grows and can be tuned over a wide range by changing R . This is shown in Fig. 19, where the absorption spectra of CdSe nanocrystals of different sizes are presented. For the smallest radii, the threshold for absorption shifts by nearly 1 eV from its bulk value. A similar shift is seen in the emission spectrum. The optical spectra of nanocrystals can be tuned continuously across the visible spectrum, making them useful in applications from fluorescent labeling to light-emitting diodes.

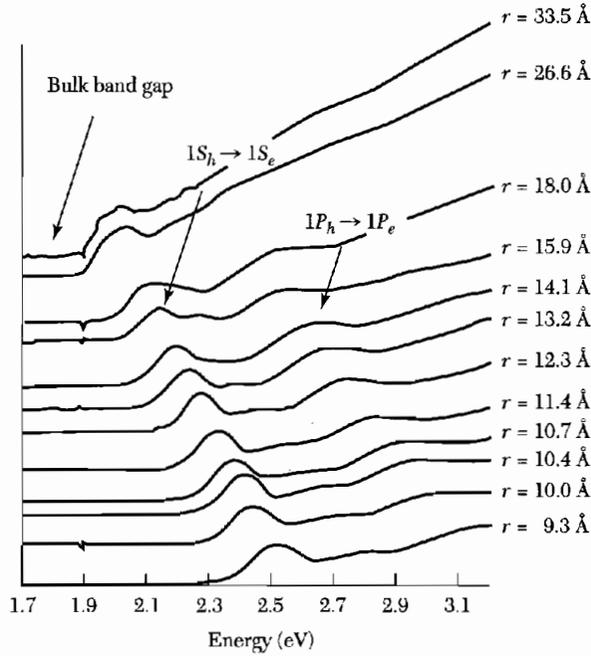


Figure 19 Optical absorption spectra for a series of CdSe nanocrystal samples of different average radii. The lowest transition energy in the smallest nanocrystal sample is shifted by nearly 1 eV from the bulk bandgap. Two dominant transitions are labeled. (Courtesy of A. P. Alivisatos.)

The absorption intensity in nanocrystals becomes concentrated at the specific frequencies corresponding to the transitions between discrete states, as described by (7). An important result for the integrated absorption can be obtained from the Kramers-Kronig relations discussed in Chapter 15. From Eq. (15.11b), we have:

$$\sigma''(\omega) = -\frac{2\omega}{\pi} P \int_0^{\infty} \frac{\sigma'(s)}{s^2 - \omega^2} ds . \quad (52)$$

At very high frequencies, $\omega \rightarrow \infty$, the electron's response is identical to that of a free electron. By (15.20),

$$\sigma''(\omega) = ne^2/m\omega . \quad (53)$$

In addition, as $\omega \rightarrow \infty$, the frequency s in the denominator of (52) can be neglected. Combining (52) and (53) then gives

$$\int_0^{\infty} \sigma'(s) ds = \pi ne^2/2m . \quad (54)$$

A bulk semiconductor and a nanocrystal therefore have the same overall absorption per unit volume when integrated over all frequencies. It is distributed very differently, however. The absorption spectrum of macroscopic

semiconductor is continuous, but in a nanocrystal it consists of a series of discrete transitions with very high absorption intensity at the transition frequencies. These strong transitions at particular frequencies have motivated researchers to create lasers that work on the quantized electronic transitions of quantum dots.

Metallic Dots

For small spherical metallic dots, such as alkali-metal clusters created in an atomic beam, the electrons in the conduction band fill up the quantized energy levels described by (50), as shown in Fig. 20a. These quantized levels affect the electrical and optical properties, and can even influence the stability of the dot. Small clusters can be analyzed by mass spectroscopy to determine the number of atoms in the cluster (Fig. 20b). Since there is one conduction electron per atom in an alkali metal, this is also the number of electrons in the conduction band. Large abundances are seen at certain “magic numbers” of atoms in the cluster. These result from the enhanced stability for clusters with filled electronic shells. For example, the 8-atom cluster peak corresponds to the filling of the 1S ($n = 1, \ell = 0$) and the 1P ($n = 1, \ell = 1$) shells. These filled-shell clusters are analogous to chemically stable filled-shell atoms (the noble gases).

For larger or irregularly shaped metallic dots, the shell structure is destroyed. The level spacing becomes small in comparison to the shifts in the levels due to shape imperfections, faceting of the crystals, or disorder. While the details of the level spectrum are difficult to predict, the average level spacing at the Fermi energy can be estimated using (6.21) as

$$\Delta\varepsilon = 1/D(\varepsilon_F) = 2\varepsilon_F/3N . \quad (55)$$

For a spherical Au nanoparticle with $R = 2$ nm, the average level spacing is $\Delta\varepsilon \sim 2$ meV. This is much smaller than the spacing between the lowest states in the CdSe nanocrystal conduction band (0.76 eV) calculated earlier. Energy-level quantization effects are much more important in semiconductor dots than they are in metallic ones. This is because the energy-level spacing is larger for low-lying states in a 3D potential well and also because the electron effective mass for semiconductors is typically smaller than for metals.

The optical properties of a small metallic dot are typically dominated by its **surface plasmon resonance**. The polarizability of a sphere is, from (16.11),

$$\text{(CGS)} \quad P = \frac{\chi E_0}{1 + 4\pi\chi/3} ; \quad \text{(SI)} \quad P = \frac{\chi\varepsilon_0 E_0}{1 + \chi/3} , \quad (56)$$

where E_0 is the external electric field and χ is the electronic susceptibility. This relation was presented in Chap. 16 for the static case, but it applies at high frequencies as long as the dot is small enough for retardation effects to be

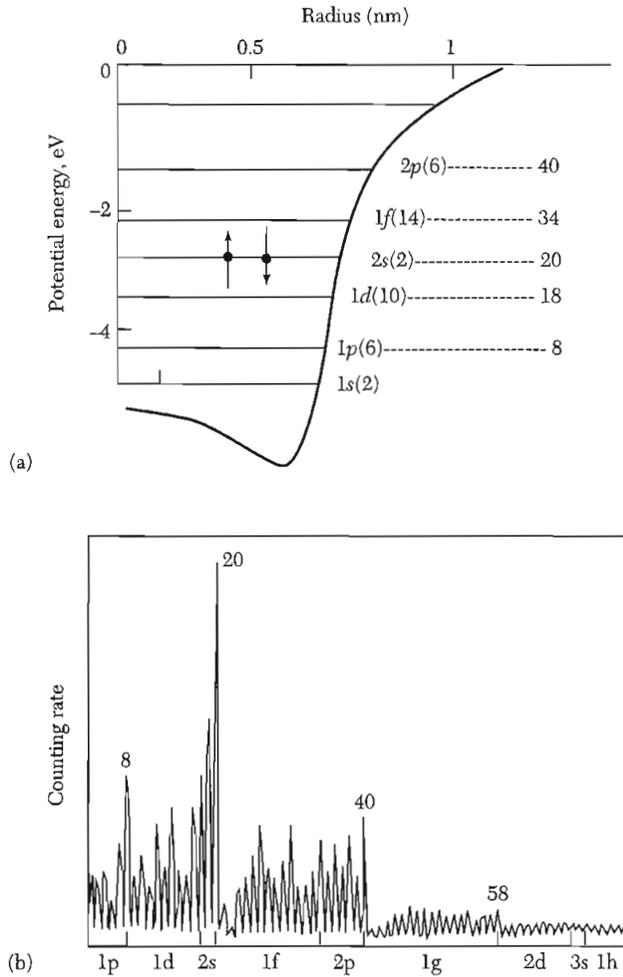


Figure 20 (a) Energy level diagram for the states in a small spherical alkali metallic cluster. The numbers at right of the diagram show the number of electrons required to fill successive electronic shells. (b) Abundance spectrum of Na clusters, showing high intensities for clusters with completely filled electronic shells. (After W. A. de Heer et al.)

absent. Modeling the carriers in the dot as a lossless free electron gas, the susceptibility is, from (14.6),

$$\text{(CGS)} \quad \chi(\omega) = -ne^2/m\omega^2 ; \quad \text{(SI)} \quad \chi(\omega) = -ne^2/m\varepsilon_0\omega^2 . \quad (57)$$

Combining (56) and (57) gives

$$\text{(CGS)} \quad P(\omega) = \frac{E_0}{1-4\pi ne^2/3m\omega^2} = \frac{E_0}{1-\omega_p^2/3\omega^2} ;$$

$$\text{(SI)} \quad P(\omega) = \frac{\varepsilon_0 E_0}{1-ne^2/3m\varepsilon_0\omega^2} = \frac{\varepsilon_0 E_0}{1-\omega_p^2/3\omega^2} , \quad (58)$$

where ω_p is the plasma frequency of the bulk metal. The polarization diverges at a frequency:

$$\omega_{sp} = \omega_p / \sqrt{3} . \quad (59)$$

This is the surface plasmon resonance frequency for a sphere. It shifts the bulk plasma resonance for metals like Au and Ag from the UV into the visible portion of the spectrum. The result (59) is independent of particle size. In reality, however, the optical properties do depend somewhat on size due to retardation effects at larger radii and losses and intraband transitions at smaller radii.

Liquids or glasses containing metallic nanoparticles are often brilliantly colored due to absorption by the surface plasmon resonance. They have been used for hundreds of years in stained glasses. Other optical applications of metallic nanoparticles make use of the large electric field just outside the nanoparticle near resonance. In techniques like surface enhanced Raman scattering (SERS) or second harmonic generation (SHG), weak optical processes in nanostructures near the surface of the nanoparticle become measurable due to the locally high electric fields.

Discrete Charge States

If a quantum dot is relatively isolated electrically from its environment, it has a set of well-defined charge states, like an atom or molecule. Each successive charge state corresponds to the addition of one more electron to the dot. Because of the coulomb repulsion between electrons, the energy difference between successive charge states can be very large. Within the Thomas-Fermi approximation (28), the electrochemical potential for adding the $(N + 1)$ th electron to a dot containing N electrons is given by:

$$\mu_{N+1} = \varepsilon_{N+1} - e\varphi = \varepsilon_{N+1} + NU - \alpha eV_g , \quad (60)$$

where U is the coulomb interaction energy between any two electrons on the dot, often called the **charging energy**. The dimensionless number α is the rate at which a voltage V_g applied to a nearby metal, typically referred to as the gate (see Fig. 21), shifts the electrostatic potential φ of the dot.

In general, U will vary for different electronic states in the dot, but we assume here it is a constant, as in a classical metal. In this case, we can describe the electrostatics and interactions in terms of capacitances:

$$U = e^2/C \quad \text{and} \quad \alpha = C_g/C , \quad (61)$$

where C is the total electrostatic capacitance of the dot and C_g is the capacitance between the dot and the gate. The quantity e/C is the electrostatic potential shift of the dot when one electron is added.

If the dot is in weak electrical contact with a metallic reservoir, electrons will tunnel onto the dot until the electrochemical potential for adding another electron exceeds the electrochemical potential μ of the reservoir (Fig. 21).

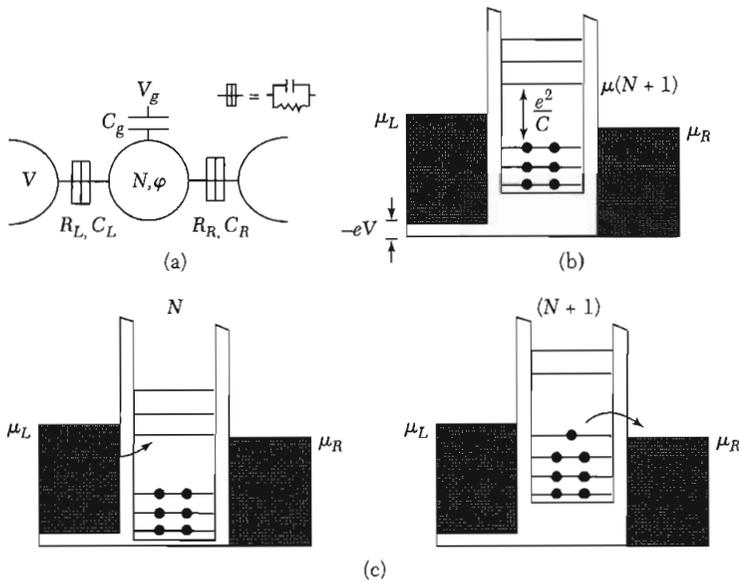


Figure 21 (a) Schematic illustration of a quantum dot in tunnel contact with two metallic reservoirs and capacitively coupled to a gate. Main: Energy level diagrams illustrating the coulomb blockade. In (b) the gate voltage is such that the dot is stable with N electrons, so no current flows. In (c) the blockade is lifted when the electrochemical potential is lowered into the window between the potentials in the leads, allowing successive charging and discharging of the dot and a net current flow.

This sets the equilibrium occupancy N of the dot. The charge state can be changed using the gate voltage V_g . The additional gate voltage ΔV_g required to add one more electron from a reservoir of fixed μ is, from (60),

$$\Delta V_g = (1/ae)(\varepsilon_{N+1} - \varepsilon_N + e^2/C) . \quad (62)$$

Adding an extra electron to the dot requires enough energy to fill up the next single-particle state and also enough energy to overcome the charging energy.

The charging energy U depends on both the size of the dot and the local electrostatic environment. Nearby metals or dielectrics will screen the coulomb interaction and reduce the charging energy. In general, U must be calculated for the specific geometry. As a simple model, consider a spherical dot of radius R surrounded by a spherical metal shell of radius $R + d$. This shell screens the coulomb interaction between electrons on the dot. An elementary application of Gauss's law (Problem 5) gives the capacitance and therefore the charging energy:

$$\text{(CGS)} \quad U = \frac{e^2}{\varepsilon R} \frac{d}{R + d} ; \quad \text{(SI)} \quad U = \frac{e^2}{4\pi\varepsilon_0\varepsilon R} \frac{d}{R + d} . \quad (63)$$

For $R = 2$ nm, $d = 1$ nm, and $\varepsilon = 1$, the charging energy is $e^2/C = 0.24$ eV. This far exceeds $k_B T \approx 0.026$ eV at room temperature, indicating that thermal

fluctuations in the charge of the dot will be strongly suppressed. It is comparable to the energy level spacing (0.76 eV) between the lowest two states in a 2-nm-radius CdSe dot. In contrast, it is much larger than the level spacing (2 meV) for a 2-nm-radius metallic dot. The addition energy of a metallic dot is therefore dominated by the charging energy, but in a semiconductor dot the charging energy and the level spacing are of comparable importance.

Charging effects are destroyed if the tunneling rate between the dot and the electrodes is too rapid. The charge resides on the dot for a time scale of order $\delta t = RC$, where R is the resistance for tunneling to the electrodes. By the uncertainty principle, the energy level will be broadened by

$$\delta\varepsilon \approx h/\delta t = h/RC = (e^2/C)(h/e^2)/R . \quad (64)$$

The uncertainty in the energy of the electron becomes comparable to the charging energy when $R \sim h/e^2$. For resistances below this value, quantum fluctuations due to the uncertainty principle smear out the coulomb charging effects. The conditions for well-defined charge states of a quantum dot are then

$$R \gg h/e^2 \quad \text{and} \quad e^2/C \gg k_B T . \quad (65)$$

ELECTRICAL TRANSPORT IN 0D

Coulomb Oscillations

At temperatures $T < (U + \Delta\varepsilon)/k_B$, the charging energy U and the level spacing $\Delta\varepsilon$ control the flow of electrons through a quantum dot, as shown in Fig. 21. Transport through the dot is suppressed where the Fermi levels of the leads lie between the electrochemical potential for the N and $N + 1$ charge states (Fig. 21b). This is called the **Coulomb blockade**. Current can only flow when $\mu_e(N + 1)$ is lowered to lie between the Fermi levels of the left and right leads. Then an electron can hop on the dot from the left electrode and off the dot to the right electrode, resulting in current flow (Fig. 21c). This process repeats with increasing V_g for each new charge state. This leads to so-called Coulomb oscillations in the conductance as a function of V_g , shown in Fig. 22. If $U \gg k_B T$, these peaks can be very sharp. The spacing between the Coulomb peaks is determined by (62).

Coulomb oscillations are first and foremost a result of charge quantization. They will occur if $U \gg k_B T$ even if the single particle level spacing is very small, $\Delta\varepsilon \ll k_B T$. This is often the case in metallic quantum dots. A device showing Coulomb oscillations is called a **single electron transistor** (SET), since it turns on and off periodically as the occupancy of the dot is changed by e . This effect is quite remarkable, and can be used as an ultrasensitive electrometer. It detects electric fields much as a SQUID (Chap. 10) detects magnetic fields. One is based on the quantization of charge, the other on the quantization of flux.

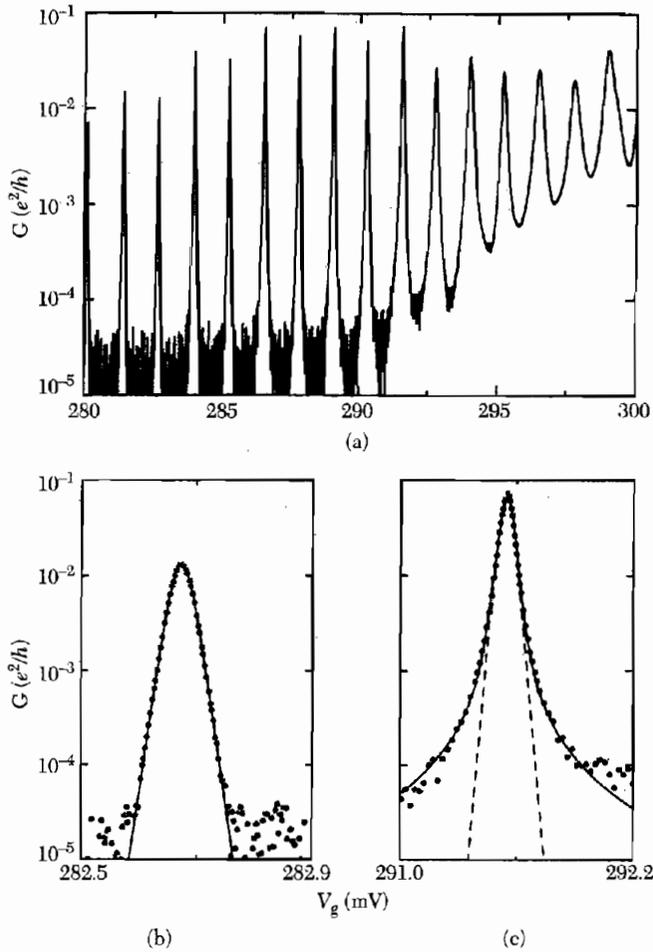


Figure 22 Conductance oscillations versus gate voltage V_g measured in a quantum dot formed in a gated GaAs/AlGaAs heterostructure at $T = 0.1$ K. The data are plotted on a log scale. As the gate voltage increases, the barriers become more transparent and the peaks get broader. The lineshape of the peak in (b) is determined by thermal broadening alone, while the one in (c) also reflects the intrinsic Breit-Wigner lineshape. (Adapted from Foxman et al.)

SETs can also be used to make single electron turnstiles and pumps. Oscillating voltages at a frequency f applied to the gates of a properly designed quantum dot system can shuttle a single electron through dot per cycle of the oscillation. This results in quantized current flowing through the dot:

$$I = ef. \quad (66)$$

Such devices are under investigation as current standards in metrology.

For the quantum dot in Fig. 22, the level spacing $\Delta\varepsilon \gg k_B T$. The N th Coulomb oscillation then corresponds to resonant tunneling through a single

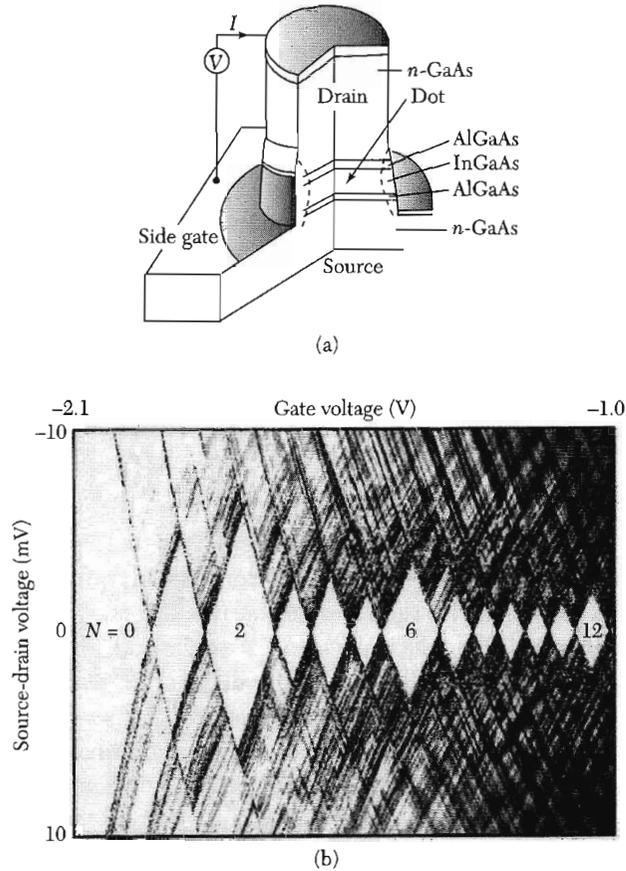


Figure 23 (a) Schematic of a 2D circular quantum dot formed in a GaAs/AlGaAs heterostructure. (b) The differential conductance dI/dV as a function of both gate voltage and source drain bias, plotted as a gray scale. The white diamond regions correspond to different charge states of the dot. A larger charging energy is observed for $N = 2$ and 6 electrons on the dot, corresponding to filled electronic shells. The additional lines on the diagram correspond to excited energy levels of the dot. (Courtesy of L. Kouwenhoven.)

quantized energy level ε_N . The coulomb oscillations in this case are analogous to the theoretical resonant tunneling peaks described by (29) and shown in Fig. 15. A crucial difference from (29) is that the positions of the coulomb peaks are determined by both the level spacing and the coulomb charging energy (62). The lower right panel of Fig. 22 shows a fit of one coulomb peak to the Breit-Wigner form for resonant tunneling (33).

The I - V characteristics of a quantum dot are in general complex, reflecting the interplay of the charging energy, the excited state level spacing, and the source-drain bias voltage. In Fig. 23, measurements of the first few electrons added to a small, 2D circular dot are shown. The differential conductance dI/dV is represented using a gray scale as both the gate voltage and the

source drain bias are varied. Each of the lines seen corresponds to tunneling through an individual quantum state of the dot. The white diamonds along the V_g axis, indicating $dI/dV = 0$, correspond to the Coulomb blockade. Each successive diamond corresponds to another electron on the dot. The point at which different diamonds touch along the V_g axis are the Coulomb oscillations where the charge state of the dot changes. The height of the diamond corresponds to $eV_{\max} = e^2/C + \Delta\varepsilon$, the maximum voltage that can be applied without current flowing in a given charge state. The diamonds corresponding to $N = 2$ and $N = 6$ are noticeably larger than neighboring diamonds, indicating a larger addition energy for adding the third and seventh electron to the dot.

This dot can be effectively modeled by a 2D harmonic oscillator confining potential $U(x, y) = \frac{1}{2}m\omega^2(x^2 + y^2)$, with energy levels:

$$\varepsilon_{ij} = (i + j + 1)\hbar\omega \quad (67)$$

where i and j are nonnegative integers. This level spectrum can be used in conjunction with (62) to find the addition energies that determine the sizes of the diamonds in the figure. The first electron fills the spin-degenerate ground state energy level ε_{00} . The second electron fills the same quantum state, but with opposite spin, at a gate voltage $\Delta V_{g2} = U/\alpha e$ after the first electron is added. The third electron fills one of the degenerate states ε_{01} or ε_{10} after a $\Delta V_{g3} = (U + \hbar\omega)/\alpha e$. The next three electrons fill the rest of these states, each spaced in gate voltage by $U/\alpha e$. The seventh electron fills one of the degenerate states ε_{11} , ε_{20} , or ε_{02} , after a gate voltage $\Delta V_{g7} = (U + \hbar\omega)/\alpha e$. This simple model correctly predicts the larger addition energies for the 3rd and 7th electrons seen in the experiment. The addition energy is larger by the level spacing when the extra electron is added to a new energy level above a filled electronic shell ($N = 2$ and $N = 6$).

Spin, Mott Insulators, and the Kondo Effect

Consider a quantum dot that is occupied by an odd number of electrons in the blockaded region, as shown in Fig. 24. The highest single particle energy level of the dot is doubly degenerate and an electron can either reside in a spin-up or spin-down state. The addition of a second electron of opposite spin is allowed by the Pauli exclusion principle but is energetically prohibited by the coulomb interaction between the electrons. This is analogous to the Mott Insulator where a half-filled band is insulating because coulomb interactions prohibit double-occupancy of the lattice sites by electrons.

The dot therefore has a spin- $\frac{1}{2}$ magnetic moment with two degenerate configurations, spin up and spin down, in the absence of coupling to the leads. If coupling to the leads is included, however, this degeneracy is lifted at low temperatures. The ground state is a quantum superposition of the two spin configurations, with transitions between them accomplished by a virtual

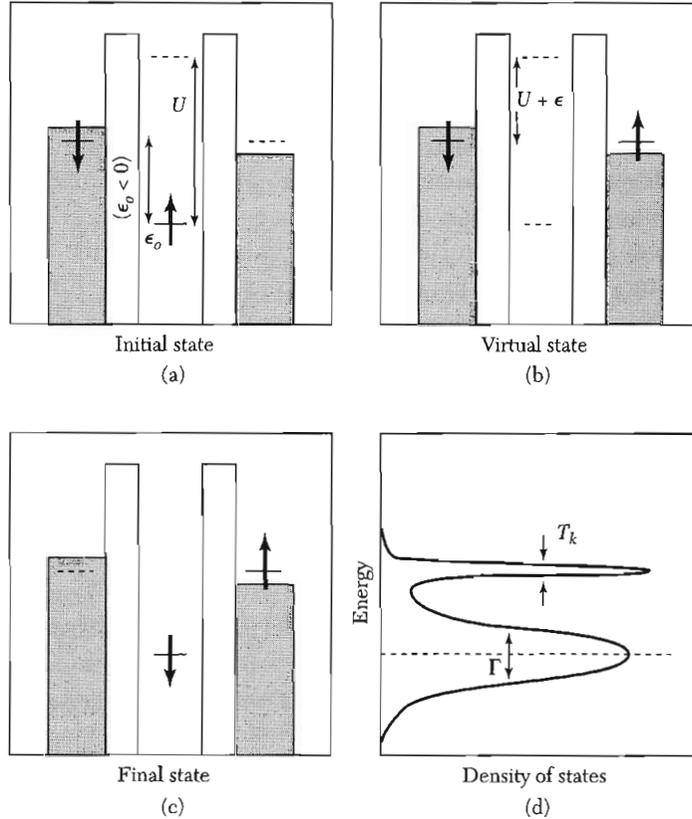


Figure 24 The Kondo effect in a quantum dot. For an unpaired spin on the dot, a virtual process (b) can occur that converts the spin up (a) to the spin down (c) state and transfers an electron from one side of the dot to the other. The ground state of the system is a coherent superposition on the initial and final states shown, creating a spin singlet between the spin on the dot and the spins in the leads. This is called the Kondo effect, and produces a narrow peak of width $\sim k_B T_K$ in the density of states at ϵ_F in addition to the original broadened level of width Γ , as shown in (d).

intermediate state involving an exchange of electrons with the leads, as is illustrated in Fig. 24. This is known as the **Kondo effect**. The local moment pairs with electrons in the metallic electrodes to create a spin singlet. This occurs below a temperature known as the Kondo temperature T_K :

$$T_K = \frac{1}{2}(\Gamma U)^{1/2} \exp[\pi \epsilon_0 (\epsilon_0 + U) / (\Gamma U)] \quad (68)$$

where Γ is the level width defined in (33) and ϵ_0 is indicated in Fig. 24 ($\epsilon_0 < 0$). A peak in the density of states of the dot of width $k_B T_K$ appears at the Fermi energy due to the admixture of states in the electrodes at an energy ϵ_F . The Kondo temperature is very small unless the coupling Γ to the leads is large, since the process involves a virtual intermediate state.

Because the Kondo effect involves exchange of electrons with the leads, it causes transmission through the dot even in the blocked region, as illustrated in Fig. 24a–c. For symmetric barriers and $T \ll T_K$, the transmission coefficient through the Kondo resonance can be unity, just as for resonant tunneling. This effect has been seen in transport through quantum dots and in STM measurements of magnetic impurities on metal surfaces. The Kondo effect was first observed in metals containing magnetic impurities. The formation of a spin singlet between the magnetic impurities and the conduction electrons enhances the scattering of the electrons. This will be discussed further in Chap. 22.

Cooper Pairing in Superconducting Dots

In a small metallic dot made of a superconductor, there is an interesting competition between single electron charging and the Cooper pairing of electrons. With an odd number of electrons residing on the dot, there is

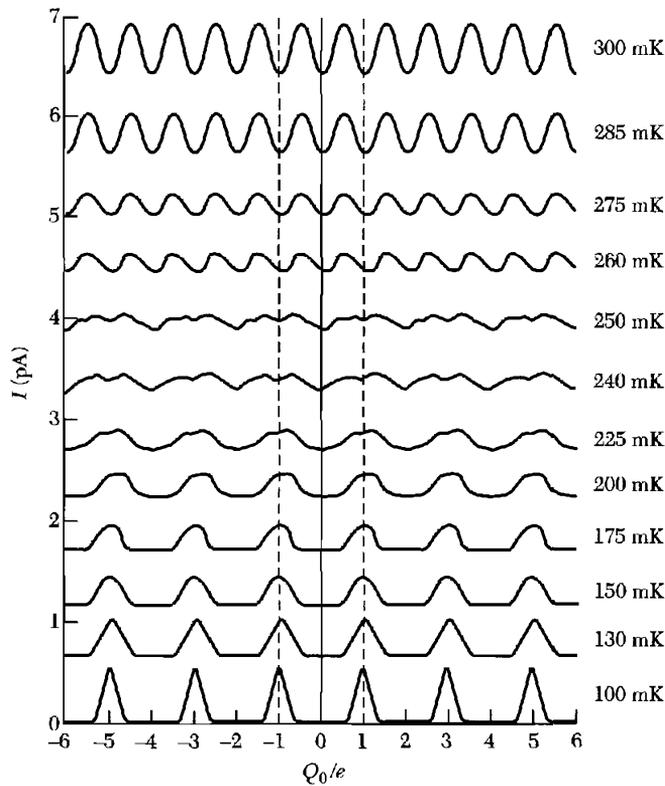


Figure 25 Measurement of coulomb oscillations in a superconducting metallic dot with decreasing temperature. A crossover from e -periodic oscillations to $2e$ -periodic oscillations is seen as the temperature is lowered due to the Cooper pairing of electrons on the dot. (After M. Tinkham, J. M. Hergenrother, and J. G. Lu.)

necessarily an unpaired electron. If the Cooper pair binding energy 2Δ is larger than the charging energy U , it is energetically favorable to add an electron to the dot, paying the energy U in order to gain the pairing energy 2Δ . The odd-charge states are thus energetically unfavorable. Electrons will be added to the dot in Cooper pairs, and the Coulomb oscillations will be $2e$ -periodic. This is shown in Fig. 25. This is a remarkable manifestation of Cooper pairing.

VIBRATIONAL AND THERMAL PROPERTIES

To treat the vibrational properties of nanostructures, we will begin from a continuum description of the elastic properties. This is analogous to employing the band structure as the starting point to describe the electronic properties. It is a good approximation for all but the smallest of nanostructures.

In general, the components of stress and strain in a solid are related by a matrix. A stress along one axis will produce a strain along that axis, but it will also produce strains along other axes. For example, a cube stretched along one axis will typically contract somewhat along the orthogonal axes. To simplify the discussion below, we will ignore the off-diagonal elements and treat the stress-strain matrix as diagonal and isotropic. In other words, strains will only occur along the direction of the stress and the magnitude will be independent of the axis direction. For a more complete treatment, we refer the reader to advanced texts on mechanics.

Quantized Vibrational Modes

Just as the electronic degrees of freedom are quantized, the vibrational frequencies become discrete in a 1D or 0D solid. The continuous low-frequency modes associated with the acoustic modes, $\omega = v_s K$, become instead a series of discrete frequencies ω_j . The exact frequencies and wavevectors depend in detail on the shape and boundary conditions of the solid.

An illustrative example is the vibrations around the circumference of a thin cylinder of radius R and thickness $h \ll R$, as shown in Fig. 26. In Fig. 26a,

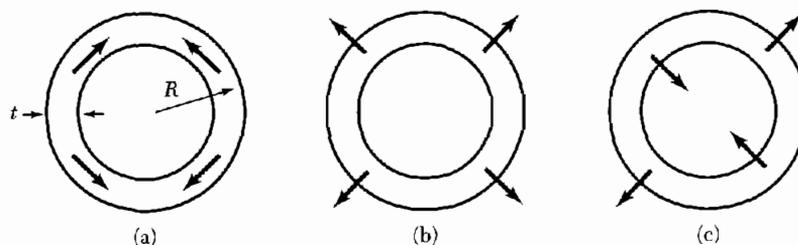


Figure 26 Fundamental vibrational modes of a thin-walled cylinder. Image (a) is a longitudinal compressional mode, (b) is the radial breathing mode (RBM), and (c) is a transverse mode.

a quantized longitudinal acoustic mode is schematically illustrated. The allowed frequencies can be found by applying periodic boundary conditions around the circumference of the cylinder:

$$K_j = j/R ; \quad \omega_{Lj} = (v_L/R)j ; \quad j = 1, 2, \dots \quad (69)$$

Another mode, called the **radial breathing mode** (RBM), is shown in Fig. 26b. The radius of the cylinder uniformly expands and contracts, producing circumferential tension and compression. From elasticity theory, the elastic energy associated with a strain e for an isotropic medium is given by

$$U_{\text{tot}} = \frac{1}{2} \int_V Y e^2 dV , \quad (70)$$

where Y is the elastic (Young's) modulus. The strain in the cylinder for a radius change dr is $e = dr/R$, yielding

$$U_{\text{tot}} = \frac{YV}{2R^2} (dr)^2 , \quad (71)$$

where V is the volume of the cylinder. Equation (71) has the form of a Hooke's law spring energy, where the spring constant is given by YV/R^2 . The vibrational frequency is then

$$\omega_{\text{RBM}} = (YV/MR^2)^{1/2} = (Y/\rho)^{1/2}/R = v_L/R , \quad (72)$$

where in the last step we have defined a longitudinal sound velocity $v_L = \sqrt{Y/\rho}$.

The final class of quantized modes around the circumference, corresponding to transverse acoustic modes, are shown in Fig. 26c. Their wavevectors and frequencies are given by

$$K_j = j/R ; \quad \omega_{Tj} \cong \frac{v_L \hbar j^2}{\sqrt{12}R^2} ; \quad j = 1, 2, \dots \quad (73)$$

Note the frequency scales like K_j^2 . The origin of this behavior will be discussed further below.

Quantized vibrational modes can be measured in a variety of ways. One technique widely used to probe the vibrational structure of individual nanoscale objects is Raman spectroscopy (see Chap. 15). Raman spectroscopy of single nanotubes is shown in Fig. 27. For a nanotube, $v_L = 21$ km/s, and the energy of the radial breathing mode is, from (72),

$$\hbar\omega_{\text{RBM}} = 14 \text{ meV}/R[\text{nm}] .$$

The measured values are in good agreement with this expression. As a result, measurements of the RBM can be used as a diagnostic to infer the radius of a nanotube.

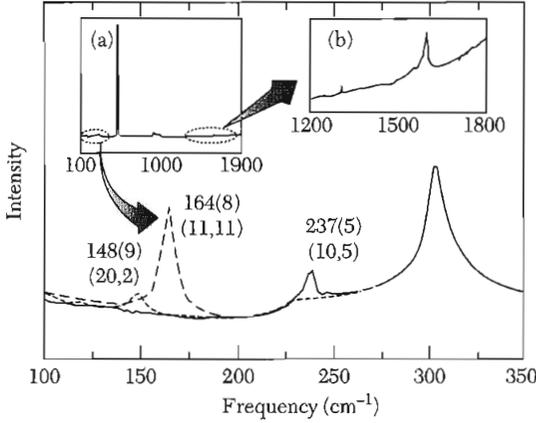


Figure 27 Raman spectra of individual carbon nanotubes. The radial breathing mode frequencies are labeled in the main panel, along with the structural assignments (n, m) . Note: $160 \text{ cm}^{-1} \cong 20 \text{ meV}$. (After A. Jorio et al.)

Transverse Vibrations

We now address the phonons propagating in the direction of the axis of a long, thin object, such as the cylinder discussed in the previous section or a thin, solid beam (Fig. 28). The longitudinal phonons are similar to the 3D case, with a dispersion $\omega = v_L K$, where K is the continuous wavevector. However, there is a fundamental modification of the transverse phonons at wavelengths longer than the thickness h of the beam. Instead of shearing, as in a bulk transverse phonon, the solid bends, as shown in Fig. 28a. This is the classical problem of transverse flexural waves on a beam. The energy of bending comes from the longitudinal compression/stretching of the solid along the inner/outer arcs of the bend. The linear wavevector dependence $\omega_T = v_T K$ characteristic of a bulk solid is changed to dispersion quadratic in K , as we show below.

Consider a transverse standing wave on a solid rectangular beam of thickness h , width w , and length L whose displacement is given by $y(z, t) = y_0 \cos(Kz - \omega t)$. The strain at a given point inside the beam is given by the local curvature and the distance t from the center line of the beam (Fig. 28a):

$$e = -(\partial^2 y / \partial z^2) t = K^2 y t . \quad (74)$$

The total energy associated with this strain is again given by (70):

$$U_{\text{tot}} = (wY/2) \int_0^L \int_{-h/2}^{h/2} (K^2 y t)^2 dt dz = YVK^4 h^2 \langle y^2 \rangle / 24 , \quad (75)$$

where $\langle y^2 \rangle$ is averaged over one period of the oscillation. This again gives an effective spring constant and, from steps analogous to (70–72), an oscillation frequency:

$$\omega_T = v_L h K^2 / \sqrt{12} . \quad (76)$$

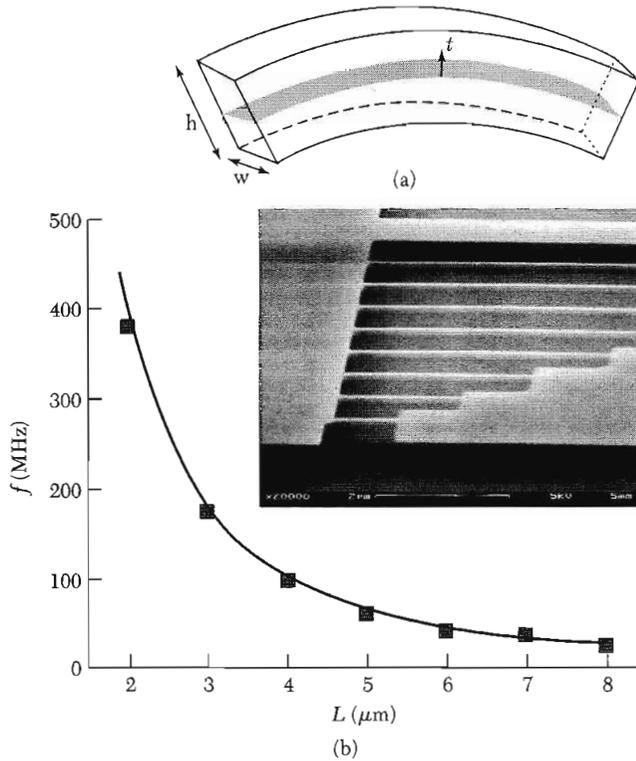


Figure 28 (a) Stresses in a bent beam, showing the inner portion under compression while the outer portion is under tension. (b) SEM micrograph of a series of suspended Si beams of varying lengths L and the measured resonance frequency as a function of L . The line is a fit to the functional form $f = B/L^2$, where B is a constant. (After D. W. Carr et al.)

Note that the frequency depends on the longitudinal sound velocity and not the transverse sound velocity, since the mode is now essentially compressional in nature. It is no longer linear in K since the effective restoring force grows stronger with increasing curvature, i.e., increasing K . In contrast, the torsional mode, corresponding to a twist of the beam along its length, retains its shear character and, $\omega_{\text{twist}} \propto K$.

Transverse vibrational modes described by (76) are frequently observed in microscale and nanoscale beams. A set of nanoscale beams constructed in Si using electron beam lithography and etching is shown in Fig. 28b. The frequencies associated with fundamental resonance $K_1 = 2\pi/L$ of these beams scale as $1/L^2$ (Fig. 28c), as expected from (76).

Note that the modification of the dispersion relation (76) for long wavelength transverse modes is not restricted to nanoscale systems. The only requirement is that the system be in the geometry of a thin beam or slab with transverse dimension h smaller than the wavelength, i.e., $Kh \ll 1$. For example, AFM cantilevers operated in the noncontact mode, as discussed above,

are well described by this relation. The dispersion relation (76) is also related to the $\omega \sim K^2$ dependence seen in (73) for the class of modes shown in Fig. 26c. Both describe the transverse flexural vibrations, one of a beam and the other of a thin shell.

A revolution is underway in the fabrication of small, complex mechanical structures using techniques adapted from microelectronic processing. The beams shown in Fig. 28b are a simple example. These structures can be integrated with electronic devices, creating microelectromechanical systems (MEMs) and nanoelectromechanical systems (NEMs). They are being explored for a variety of applications, including sensing, data storage, and signal processing.

Heat Capacity and Thermal Transport

The above relations indicate that the quantized vibrational mode energies are typically less than $k_B T$ at room temperature except in the very smallest structures. Modes along the confined directions will thus be thermally excited at room temperature. As a result, the lattice thermal properties of nanostructures will be similar to their bulk counterparts. In particular, the lattice heat capacity and thermal conductivity will be proportional to T^3 , as for a 3D solid (Chap. 5).

At low temperatures, however, vibrational excitations of frequency ω in the confined directions will freeze out when $T < \hbar\omega/k_B$. In the case of a long, thin structure, the system will behave as 1D thermal system at sufficiently low temperatures, with a set of 1D phonon subbands analogous to the 1D electronic subbands shown in Fig. 9. A calculation analogous to the one performed in Chap. 5 for a 3D solid yields the heat capacity per 1D acoustic phonon subband with a dispersion $\omega = vk$ (Prob. 6):

$$C_V^{(1D)} = 2\pi^2 L k_B^2 T / 3\hbar v . \quad (77)$$

The thermal conductance G_{th} of the wire is defined as the ratio of the net energy flow through the wire divided by the temperature difference ΔT between its ends. The thermal conductance per 1D phonon subband $G_{th}^{(1D)}$ is

$$G_{th}^{(1D)} = (\pi^2 k_B^2 T / 3\hbar) \mathfrak{S} . \quad (78)$$

This result is derived in Prob. 6 using an approach analogous to the one employed to derive the Landauer formula for conductance of a 1D channel. Note \mathfrak{S} is now the transmission probability for phonons through the structure. For a transverse flexural mode, where $\omega \propto K^2$, the result (77) is modified, but (78) is the same.

Both (77) and (78) are linear in temperature, in contrast to the T^3 result for 3D. The difference reflects the number of modes with energies $\hbar\omega < k_B T$, or equivalently, with wavevectors $K < k_B T / \hbar v$. The number of modes scales like K^D , where D is the dimensionality, producing T^3 in 3D and T in 1D.

Note that in the case of perfectly transmitted phonons through the channel, the thermal conductance (78) is determined only by fundamental constants and the absolute temperature. This result is analogous to the quantized electronic conductance (21) of a 1D channel, which was independent of the electron velocity in the channel. Both the ballistic thermal conductance (78) and the 1D form of the heat capacity (77) have been observed in experiments on narrow wires at very low temperatures.

SUMMARY

- Real space probes can give atomic-scale images of nanostructures.
- The density of states of a 1D subband, $D(E) = 4L/hv$, diverges at the subband thresholds. These are called van Hove singularities.
- The electrical conductance of a 1D system is given by the Landauer formula, $G = (2e^2/h)\mathfrak{T}$, where \mathfrak{T} is the transmission coefficient through the sample.
- The conductance of a quasi-1D system can be strongly influenced by quantum interference among the electron paths traversing the sample, leading to resonant tunneling, localization, and the Aharonov-Bohm effect.
- The optical properties of a quantum dot can be tuned by changing its size and hence its quantized energy levels.
- Adding an extra charge e to a quantum dot requires an additional electrochemical potential given by $U + \Delta\varepsilon$, where U is the charging energy and $\Delta\varepsilon$ is the level spacing.
- The vibrational modes of a nanometer-scale object are quantized.

Problems

1. **Carbon nanotube band structure.** Figure 29 shows the graphene lattice with the primitive lattice translation vectors of length $a = 0.246$ nm, along with the first Brillouin zone. (a) Find the set of reciprocal lattice vectors \mathbf{G} associated with the lattice. (b) Find the length of the vectors \mathbf{K} and \mathbf{K}' shown in Fig. 29 in terms of a .

For energies near the Fermi energy and wavevectors near the K point, the 2D band structure can be approximated as

$$\varepsilon = \pm \hbar v_F |\Delta \mathbf{k}| \quad \Delta \mathbf{k} = \mathbf{k} - \mathbf{K} ,$$

where $v_F = 8 \times 10^5$ m/s. A similar approximation holds near the K' point. Consider a tube rolled up along the x -axis with a circumference na . By applying periodic boundary conditions along the rolled up direction, the dispersions of the 1D subbands near the K point can be found. (c) Show that, if n is divisible by 3, there exists a “massless” subband whose energy is linear in Δk_y . Sketch this subband. These nanotubes are 1D metals. (d) If n is not divisible by 3, the subband structure is that

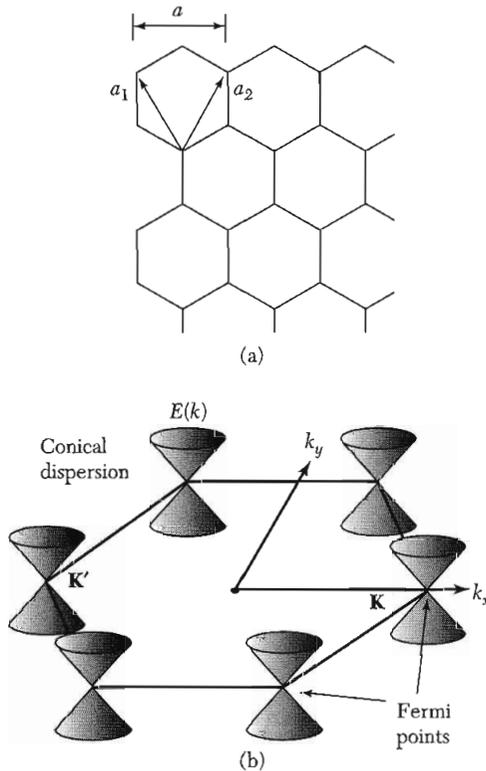


Figure 29 (a) The graphene lattice and (b) the first Brillouin zone of the graphene lattice showing the conical dispersion of the energies near the \mathbf{K} and \mathbf{K}' points.

as shown in Fig. 10. For the case of $n = 10$, find the magnitude of the semiconducting bandgap ε_{11} in eV and show that $\varepsilon_{22}/\varepsilon_{11} = 2$. (e) Again for the $n = 10$ case, show that the dispersion relation of the lowest electron subband is of the form of a relativistic particle, $\varepsilon^2 = (m^*c^2)^2 + (pc)^2$, where v_F plays the role of the speed of light, and find the ratio of effective mass m^* to the free electron mass m .

2. **Filling subbands.** For electrons in a square GaAs wire of width 20 nm, find the linear electron density at which the $n_x = 2$, $n_y = 2$ subband is first populated in equilibrium at $T = 0$. Assume an infinite confining potential at the wire boundary.
3. **Breit-Wigner form of a transmission resonance.** The purpose of this problem is to derive (33) from (29). (a) By expanding the cosine for small phase differences away from resonance, $\delta\varphi = \varphi^* - 2\pi n$, find a simplified form of (29) involving only $|t_1|^2$, $|t_2|^2$, and $\delta\varphi$. (b) Show that, for states in a 1D box, the following relation holds between small changes in the phases and small changes in the energy: $\delta\varepsilon/\Delta\varepsilon = \delta\varphi/2\pi$, where $\Delta\varepsilon$ is the level spacing. (c) Combine (a) and (b) to obtain (33).
4. **Barriers in series and Ohm's Law.** (a) Derive (36) from (35). (b) Show that the 1D Drude conductivity $\sigma_{1D} = n_{1D}e^2\tau/m$ can be written as $\sigma_{1D} = (2e^2/h)\ell_B$. (Note: The momentum relaxation rate and the backscattering rates are related as $1/\tau = 2/\tau_B$ because the former corresponds to the relaxation from \mathbf{p} to 0 while the latter corresponds to relaxation from \mathbf{p} to $-\mathbf{p}$.)

5. **Energies of a spherical quantum dot.** (a) Derive the formula (63) for the charging energy. (b) Show that, for $d \ll R$, the result is the same as that obtained using the parallel plate capacitor result, $C = \epsilon\epsilon_0 A/d$. (c) For the case of an isolated dot, $d \rightarrow \infty$, find the ratio of the charging energy to lowest quantized energy level. Express your answer in terms of the radius R of the dot and the effective Bohr radius a_B^* .
6. **Thermal properties in 1D.** (a) Derive the formula (77) for the low temperature heat capacity of a single 1D phonon mode within the Debye approximation. (b) Derive the relation for the thermal conductance (78) of a 1D phonon mode between two reservoirs by calculating the energy flow out of one reservoir at a temperature T_1 and subtracting the energy flow from the other reservoir at a temperature T_2 . Use an approach analogous to that used to obtain (20) and (24) for the electrical conductance.