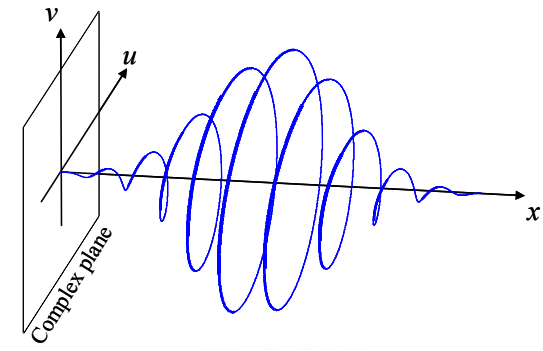
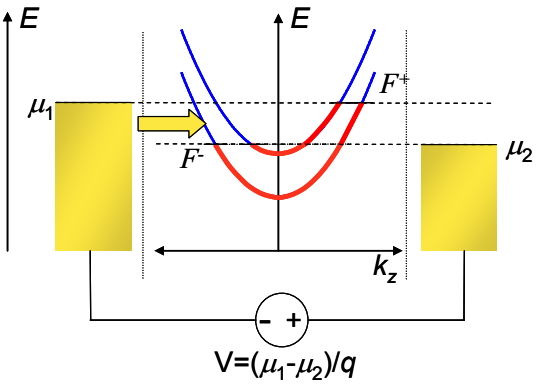
OpenCourseWare

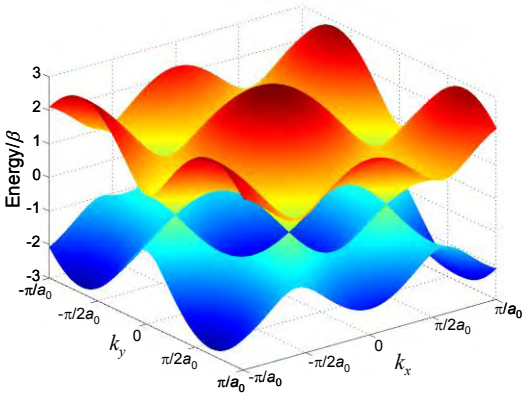
Introduction to Nanoelectronics

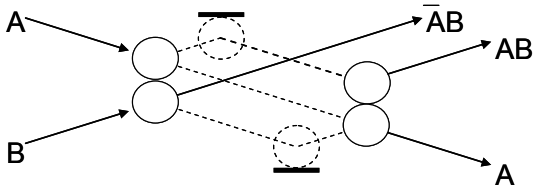
by Marc Baldo

Introduction to Nanoelectronics









Marc Baldo

MIT OpenCourseWare Publication May 2011

Preface to the OpenCourseWare publication

About eight years ago, when I was just starting at MIT, I had the opportunity to attend a workshop on nanoscale devices and molecular electronics. In particular, I remember a presentation by Supriyo Datta from Purdue. He was describing electronic devices from the „bottom up" – starting with quantum mechanical descriptions of atoms and molecules, and ending up with device-scale current-voltage characteristics.

Although I did not understand the details at the time, it was clear to me that this approach promised a new approach to teaching electronic devices to undergraduates. Building from a few basic concepts in quantum mechanics, and a reliance on electric potentials rather than fields, I believe that the „bottom up" approach is simpler and more insightful than conventional approaches to teaching electronic transport. After five years of teaching the material, it is still remarkable to me that one can derive the current-voltage characteristics of a ballistic nanowire field effect transistor within a 45 minute lecture.

This collection of class notes is my attempt to adapt the „bottom up" approach to an undergraduate electrical engineering curriculum. It can serve several roles. For most seniors, the class is intended to provide a thorough analysis of ballistic transistors within a broader summary of the most important device issues in computation. But for those intending to specialize in electronic devices, the class is designed as an introduction to dedicated courses on quantum mechanics, solid state physics, as well as more comprehensive treatments of quantum transport such as those by Supriyo Datta himself. I can recommend both his books[1,2], and the „nanohub" at Purdue University: http://nanohub.org/topics/ElectronicsFromTheBottomUp.

The notes are designed to be self contained. In particular, this class is taught without requiring prior knowledge of quantum mechanics, although I do prefer that the students have prior knowledge of Fourier transforms.

Finally, I decided to share these notes on MIT's OpenCourseWare with the expectation of collaboration. The „bottom up" approach is still relatively novel, and these notes remain largely unpolished, with substantial opportunities for improvement! For those needing to teach a similar topic, I hope that it provides a useful resource, and that in return you can share with me suggestions, corrections and improvements.

Marc Baldo
May 2010, Cambridge, MA

1. Electronic Transport in Mesoscopic Systems, Supriyo Datta, Cambridge University Press, 1995
2. Quantum Transport: Atom to Transistor, Supriyo Datta, Cambridge University Press, 2005

Acknowledgements

These notes draws heavily on prior work by Supriyo Datta, „Electronic Transport in Mesoscopic Systems", Cambridge University Press, 1995 and „Quantum Transport: Atom to Transistor", Cambridge University Press, 2005. I have also made multiple references to the third edition of „Molecular Quantum Mechanics" by Atkins and Friedman, Oxford University Press, 1997.

I would also like to thank Terry Orlando, Phil Reusswig, Priya Jadhav, Jiye Lee, and Benjie Limketkai for helping me teach the class over the years.

My work is dedicated to Suzanne, Adelie, Esme, and Jonathan.

Contents

Introduction

Modern technology is characterized by its emphasis on miniaturization. Perhaps the most striking example is electronics, where remarkable technological progress has come from reductions in the size of transistors, thereby increasing the number of transistors possible per chip.

With more transistors per chip, designers are able to create more sophisticated integrated circuits. Over the last 35 years, engineers have increased the complexity of integrated circuits by more than *five orders of magnitude*. This remarkable achievement has transformed society. Even that most mechanical creature of modern technology, the automobile, now typically contains half its value in electronics.[†]

The industry"s history of steady increases in complexity was noted by Gordon Moore, a co-founder of Intel. The eponymous Moore"s law state*s the complexity of an integrated circuit, with respect to minimum component cost, will double in about 18 months*. Over time the „law" has held up pretty well; see Fig. 1.

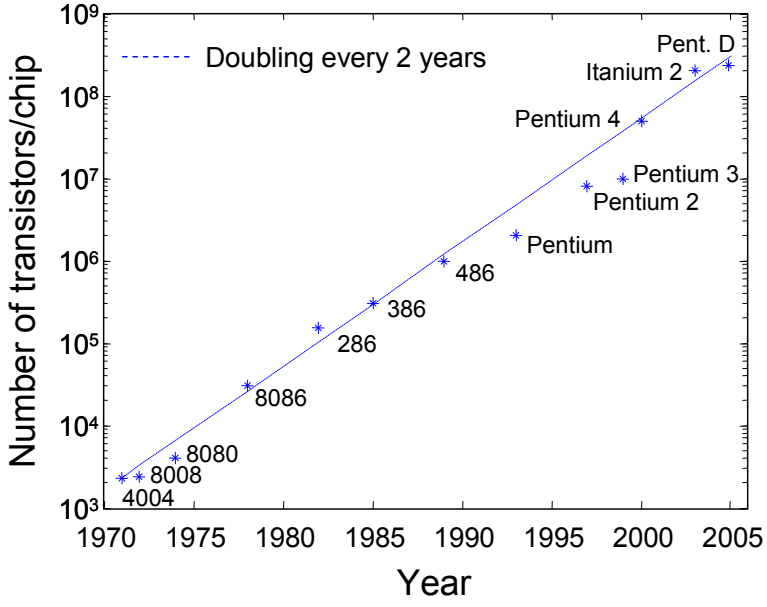


Fig. 1. The number of transistors in Intel processors as a function of time. The trend shows a doubling approximately every two years rather than 18 months as originally predicted by Gordon Moore.

Miniaturization has helped the digital electronics market alone to grow to well over $300 billion per year. The capital investments required of semiconductor manufacturers are substantial, however, and to help reduce risks, in 1992 the Semiconductor Industries

[†] If this seems hard to believe, consider the number of systems controlled electronically in a modern car. The engine computer, the airbags, the anti-skid brakes, etc..

Association began the prediction of major trends in the industry – the International Technology Roadmap for Semiconductors – better known simply as „the roadmap".

The roadmap is updated every few years and is often summarized by a semi-log plot of critical feature sizes in electronic components; see Fig. 2.

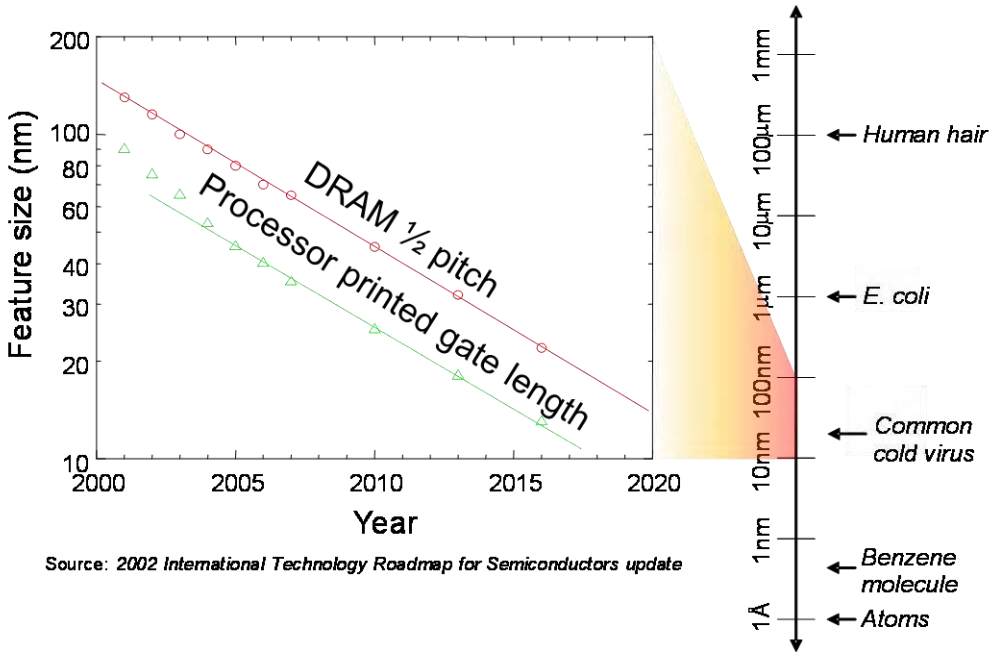


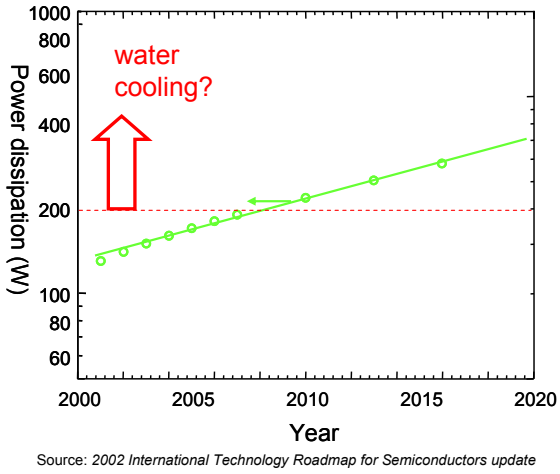
Fig. 2. The semiconductor roadmap predicts that feature sizes will approach 10 nm within 10 years. Data is taken from the 2002 International Technology Roadmap for Semiconductors update.

At the time of writing, the current generation of Intel central processing units (CPUs), the Pentium D, has a gate length of 65 nm. According to the roadmap, feature sizes in CPUs are expected to approach molecular scales (< 10 nm) within 10 years.

But exponential trends cannot continue forever.

Already, in CPUs there are glimmers of the fundamental barriers that are approaching at smaller length scales. It has become increasingly difficult to dissipate the heat generated by a CPU running at high speed. The more transistors we pack into a chip, the greater the power *density* that we must dissipate. At the time of writing, the power density of modern CPUs is approximately 150 W/cm^2; see Fig. 3. For perspective, note that the power density at the surface of the sun is approximately 6000 W/cm^2. The sun radiates this power by heating itself to 6000 K. But we must maintain our CPUs at approximately room temperature. The heat load of CPUs has pushed fan forced convection coolers to the limits of practicality. Beyond air cooling is water cooling, which at greater expense may be capable of removing several hundred Watts from a 1cm^2 sized chip. Beyond water cooling, there is no known solution.

Introduction



Source: *2002 International Technology Roadmap for Semiconductors update*

Asus StarIce
Length: 5.5", Weight: 2lb
Noise: 63 dB full power

TunicTower
Length: 6", Weight:
Noise: 49 dB full po

Fig. 3. Expected trends in CPU power dissipation according to the roadmap.

Power dissipation is the most visible problem confronting the electronics industry today. But as electronic devices approach the molecular scale, our traditional understanding of electronic devices will also need revision. Classical models for device behavior must be abandoned. For example, in Fig. 4, we show that many electrons in modern transistors electrons travel „ballistically" – they do not collide with any component of the silicon channel. Such ballistic devices cannot be analyzed using conventional transistor models. To prepare for the next generation of electronic devices, this class teaches the theory of current, voltage and resistance from atoms up.

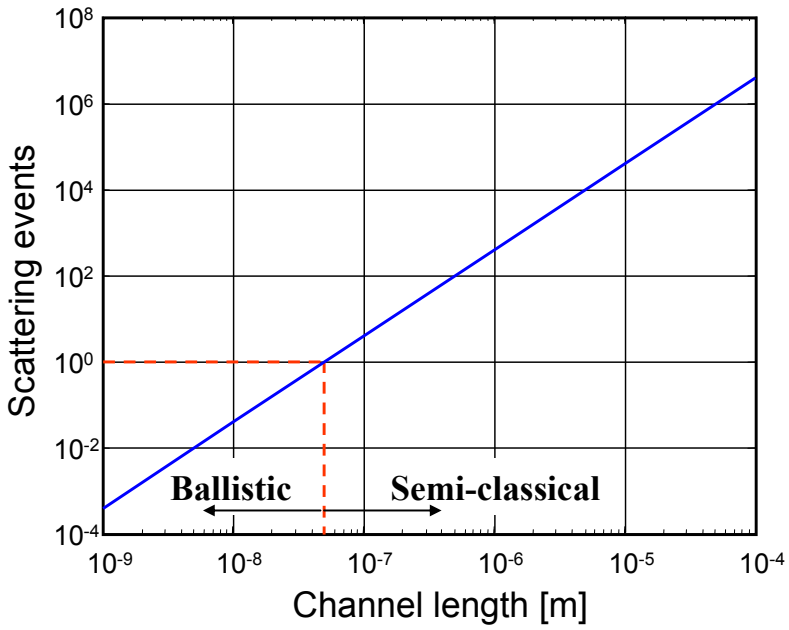


Fig. 4. The expected number of electron scattering events in a silicon field effect transistor as a function of the channel length. The threshold of ballistic operation occurs for channel lengths of approximately 50nm.

In Part 1, „The Quantum Particle", we will introduce the means to describe electrons in nanodevices. In early transistors, electrons can be treated purely as point particles. But in nanoelectronics the position, energy and momentum of an electron must be described probabilistically. We will also need to consider the wave-like properties of electrons, and we will include phase information in descriptions of the electron; see Fig. 5. The mathematics we will use is similar to what you have already seen in signal processing classes. In this class we will assume knowledge of Fourier transforms.

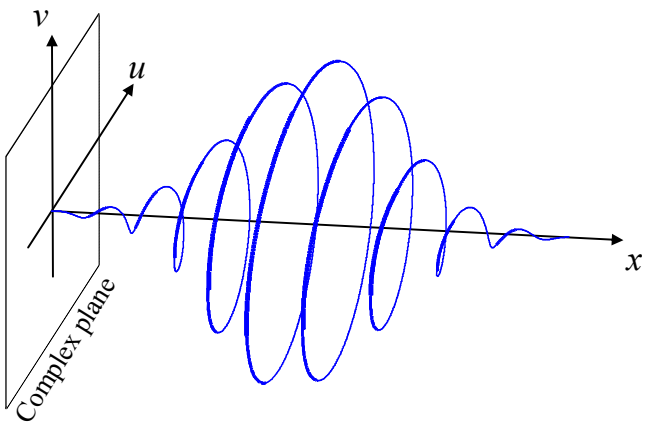


Fig. 5. A representation of an electron known as a wavepacket. The position of the electron is described in 1 dimension, and its probability density is a Gaussian. The complex plane contains the phase information.

Part 1 will also introduce the basics of Quantum Mechanics. We will solve for the energy of an electron within an attractive box-shaped potential known as a „square well". In Part 2, „The Quantum Particle in a Box", we apply the solution to this square well problem, and introduce the simplest model of an electron in a conductor – the so-called particle in a box model. The conductor is modeled as a homogeneous box. We will also introduce an important concept: the density of states and learn how to count electrons in conductors. We will perform this calculation for „quantum dots", which are point particles also known as 0-dimensional conductors, „quantum wires" which are ideal 1-dimensional conductors, „quantum wells" (2-dimensional conductors) and conventional 3-dimensional bulk materials.

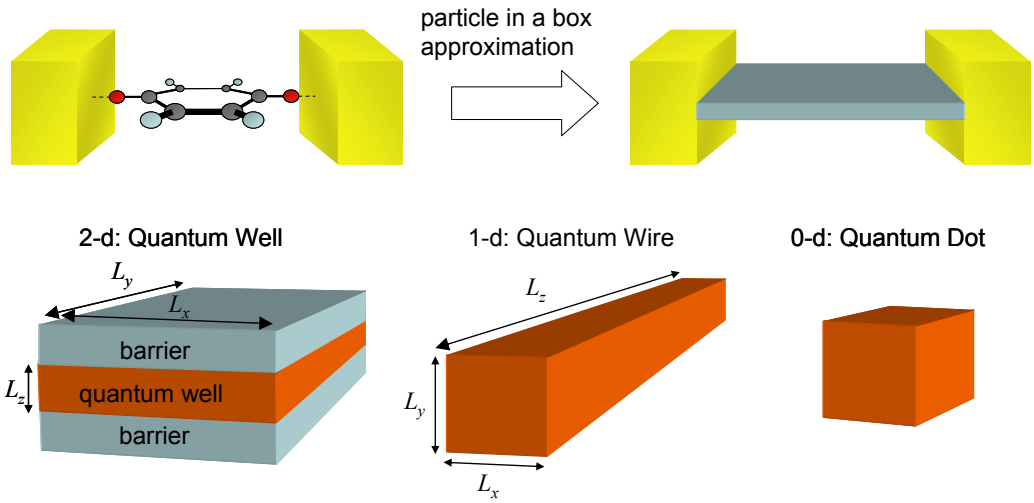


Fig. 6. The particle in a box approximation and conductors of different dimensionality.

Introduction

In Part 3, „Two Terminal Quantum Dot Devices" we will consider current flow through the 0-dimensional conductors. The mathematics that we will use is very simple, but Part 3 provides the foundation for all the description of nano transistors later in the class.

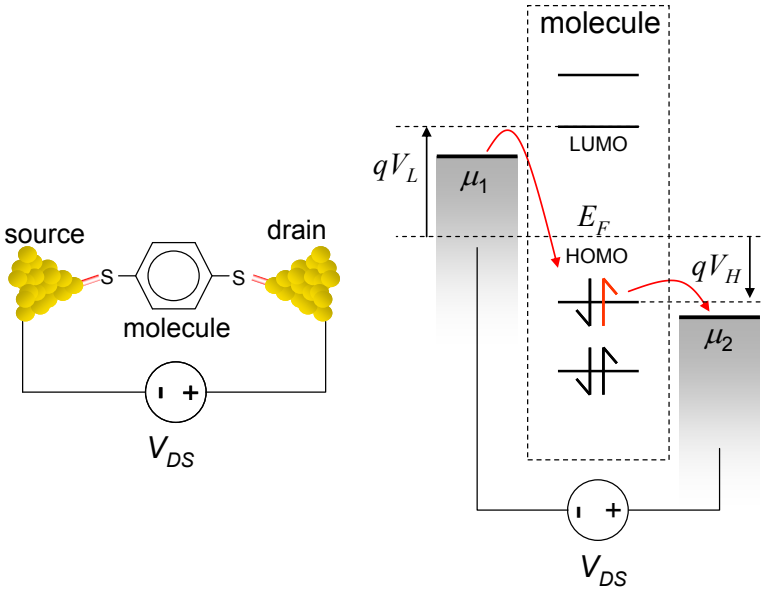


Fig. 7. A two terminal device with a molecular conductor. Under bias in this molecule, electrons flow through the highest occupied molecular orbital or HOMO.

Part 4, „Two Terminal Quantum Wire Devices", explains conduction through nanowires. We will introduce „ballistic" transport – where the electron does not collide with any component of the conductor. At short enough length scales all conduction is ballistic, and the understanding ballistic transport is the key objective of this class. We will demonstrate that for nanowires conductance is quantized. In fact, the resistance of a nanowire of vanishingly small cross section can be no less than 12.9 kΩ. We will also explain why this resistance is independent of the length of the nanowire! Finally, we will explain the origin of Ohm"s law and „classical" models of charge transport.

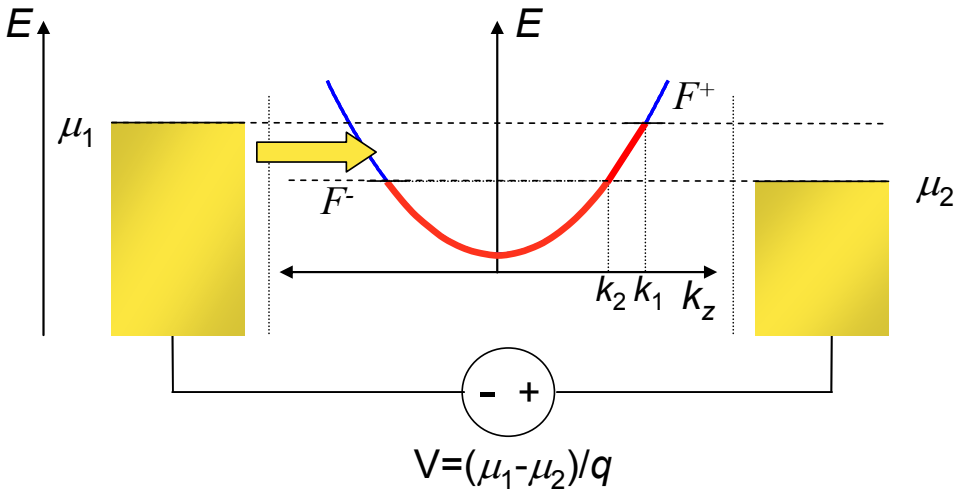


Fig. 8. From Part 4, this is a diagram explaining charge conduction through a nanowire. The left contact is injecting electrons. The resistance of the wire is calculated to be 12.9 kΩ independent of the length of the wire.

In Part 5, „Field Effect Transistors" we will develop the theory for this most important application. We will look at transistors of different dimensions and compare the performance of ballistic and conventional field effect transistors. We will demonstrate that conventional models of transistors fail at the nanoscale.

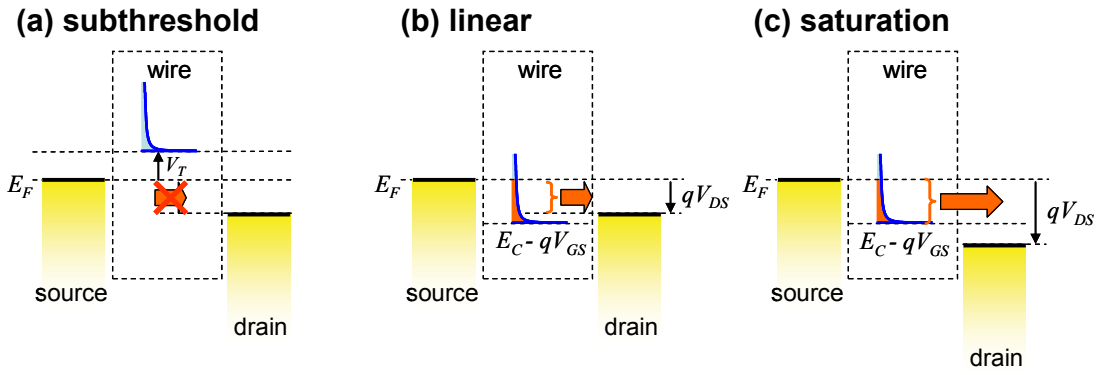


Fig. 9. Three modes of operation in a nanowire field effect transistor. In subthreshold only thermally excited electrons from the source can occupy states in the channel. In the linear regime, the number of available states for conduction in the channel increases with source-drain bias. In saturation, the number of available channel states is independent of the source-drain bias, but dependent on the gate bias.

Part 6, „The Electronic Structure of Materials" returns to the problem of calculating the density of states and expands upon the simple particle-in-a-box model. We will consider the electronic properties of single molecules, and periodic materials. Archetypical 1- and 2-dimensional materials will be calculated, including polyacetylene, graphene and carbon nanotubes. Finally, we will explain energy band formation and the origin of metals, insulators and semiconductors.

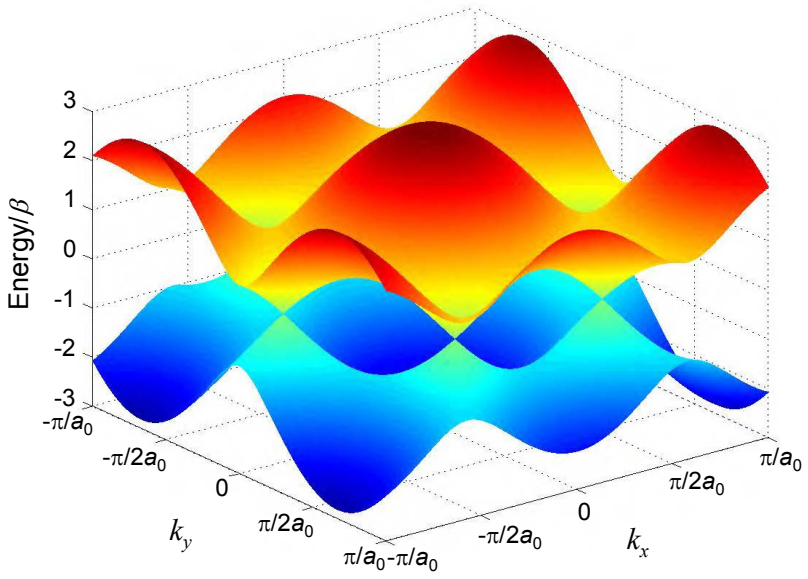
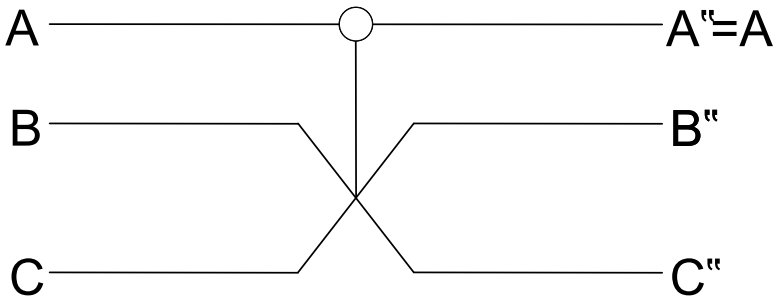


Fig. 10. This is the bandstructure of graphene. There are two surfaces that touch in 6 discrete point corresponding to 3 different electron transport directions within a sheet of graphene. Along these directions graphene behaves like a metal. Along the other conduction directions it behaves as an insulator.

Introduction

The class concludes with Part 7, „Fundamental Limits in Computation". We will take a step back and consider the big picture of electronics. We will revisit the power dissipation problem, and discuss possible fundamental thermodynamic limits to computation. We will also introduce and briefly analyze concepts for dissipation-less „reversible" computing.



A	B	C	A"	B"	C"
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	0	0	1
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	0
1	1	1	1	1	1

Fig. 11. The Fredkin gate is a reversible logic element which may be used as the building block for arbitrary logic circuits. The signal A may be used to swap signals B and C. Because the input and output of the gate contain the same number of bits, no information is lost, and hence the gate is, in principle, dissipation-less. After Feynman.

Part 1. The Quantum Particle

This class is concerned with the propagation of electrons in conductors.

Here in Part 1, we will begin by introducing the tools from quantum mechanics that we will need to describe electrons. We will introduce probabilistic descriptions of the key physical properties: position, momentum, time and energy. In the next part we will consider electrons in the simplest possible model of a conductor – a box – *i.e.* we will ignore atoms and assume that the material is perfectly homogeneous.

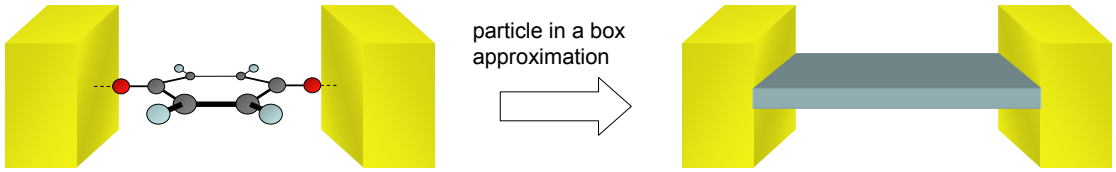


Fig. 1.1. The „particle in a box" takes a complex structure like a molecule and approximates it by a homogeneous box. All details, such as atoms, are ignored.

This model of electrons in conductors is known as „the particle in a box". It is surprisingly useful, and later in the class we will employ it to describe the behavior of modern transistors.

But first we will need a way to describe electrons. It is often convenient to imagine electrons as little projectiles pushed around by an electric field. And in many cases, this classical model yields a fairly accurate description of electronic devices.

But not always. In nanoscale devices especially, electrons are better described as waves.

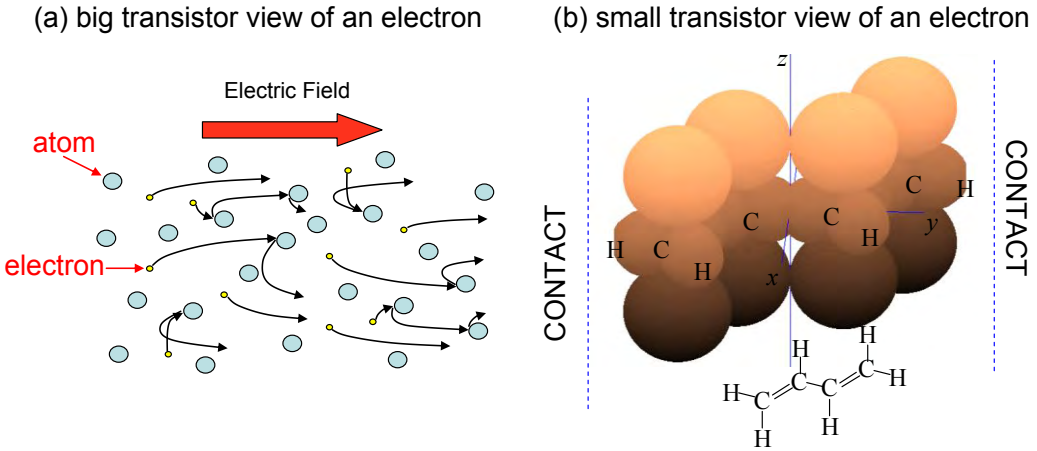


Fig. 1.2. Two representations of electrons in a solid. In **(a)** the electrons are represented as hard little spheres, propelled by the electric field, and bouncing off atoms. In **(b)** we draw an approximate representation of the molecule 1,3-butadiene positioned between contacts. Now the electrons are represented by probability clouds.

Part 1. The Quantum Particle

Waves in electronics

Consider a beam of electrons propagating through a small hole - a very crude model for electrons moving from a contact into a nanoscale conductor. We might expect that the electrons would continue in straight lines after passing through the hole. But if the hole is small enough (the dimensions of a nanoscale transistor, for example), then the electrons are observed to diffract. This clearly demonstrates that we must consider the wave properties of electrons in nanoelectronic devices.

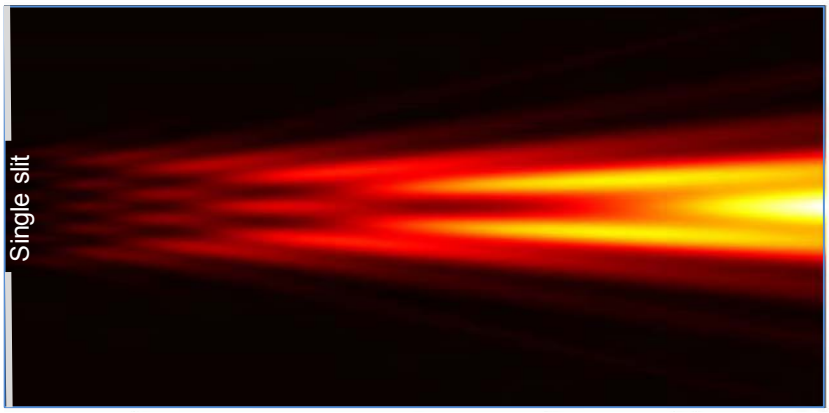


Fig. 1.3. A simulation of electron diffraction through a single slit. This experiment is analyzed in Problem 1.

The diffraction pattern shown above is obtained by assuming each point inside the single slit is a source of expanding waves; see Problem 1. An easier example to analyze is the double slit experiment, in which we assume there are only two sources of expanding waves. Like the single slit example, the result of a double slit experiment is consistent with electrons behaving like waves.

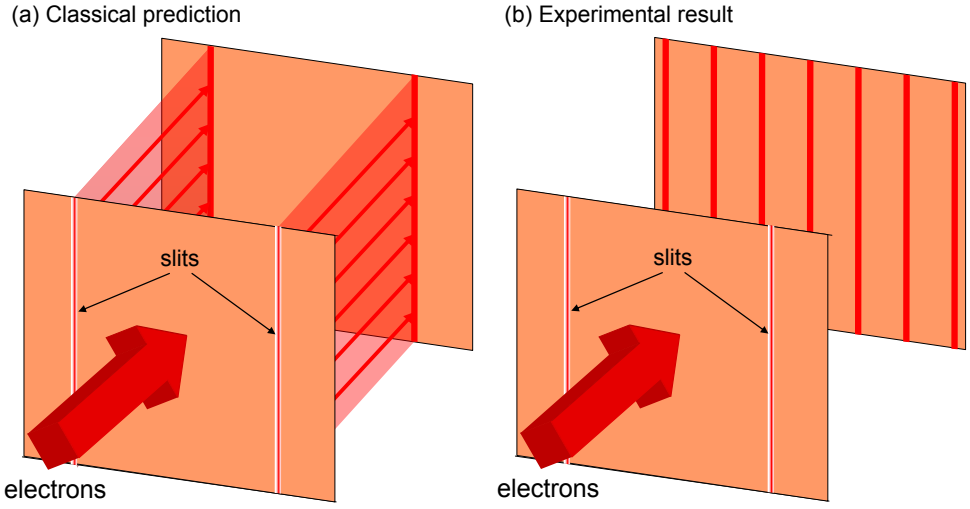


Fig. 1.4. Classically, we would predict that electrons passing through slits in a screen should continue in straight lines, forming an exact image of the slits on the rear screen. In practice, however, a series of lines is formed on the rear screen, suggesting that the electrons have been somehow deflected by the slits.

Review of Classical Waves

A wave is a periodic oscillation. It is convenient to describe waves using complex numbers. For example consider the function

$$\psi(x) = e^{ik_0 x} \tag{1.1}$$

where x is position, and k_0 is a constant known as the wavenumber. This function is plotted in Fig. 1.5 on the complex plane as a function of position, x. The phase of the function

$$\phi = k_0 x \tag{1.2}$$

is the angle on the complex plane.

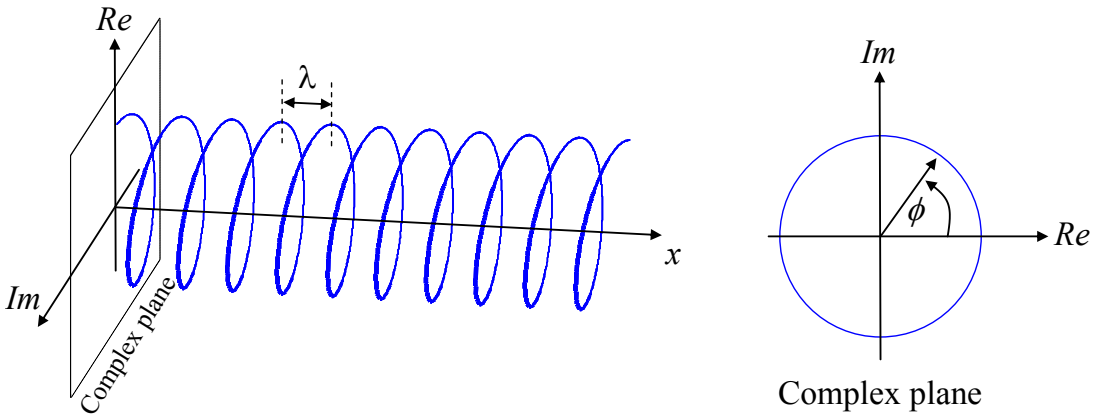


Fig. 1.5. A standing wave with its phase plotted on the complex plane.

The wavelength is defined as the distance between spatial repetitions of the oscillation. This corresponds to a phase change of 2π. From Eqns. (1.1) and (1.2) we get

$$k_0 = \frac{2\pi}{\lambda} \tag{1.3}$$

This wave is independent of time, and is known as a standing wave. But we could define a function whose phase varies with time:

$$\psi(t) = e^{-i\omega_0 t} \tag{1.4}$$

Here t is time, and ω is the angular frequency. We define the period, T, as the time between repetitions of the oscillation

$$\omega_0 = \frac{2\pi}{T}. \tag{1.5}$$

Plane waves

We can combine time and spatial phase oscillations to make a traveling wave. For example

$$\psi(x,t) = e^{i(k_0 x - \omega_0 t)} \tag{1.6}$$

We define the intensity of the wave as

$$|\psi|^2 = \psi^* \psi \tag{1.7}$$

Where ψ^* is the complex conjugate of ψ. Since the intensity of this wave is uniform everywhere $|\psi|^2 = 1$ it is known as a *plane wave*.

A plane wave has at least four dimensions (real amplitude, imaginary amplitude, x, and t) so it is not so easy to plot. Instead, in Fig. 1.6 we plot planes of a given phase. These planes move through space at the phase velocity, v_p, of the wave. For example, consider the plane corresponding to $\phi = 0$.

$$k_0 x - \omega_0 t = 0 \tag{1.8}$$

Now,

$$v_p = \frac{dx}{dt} = \frac{\omega_0}{k_0} \tag{1.9}$$

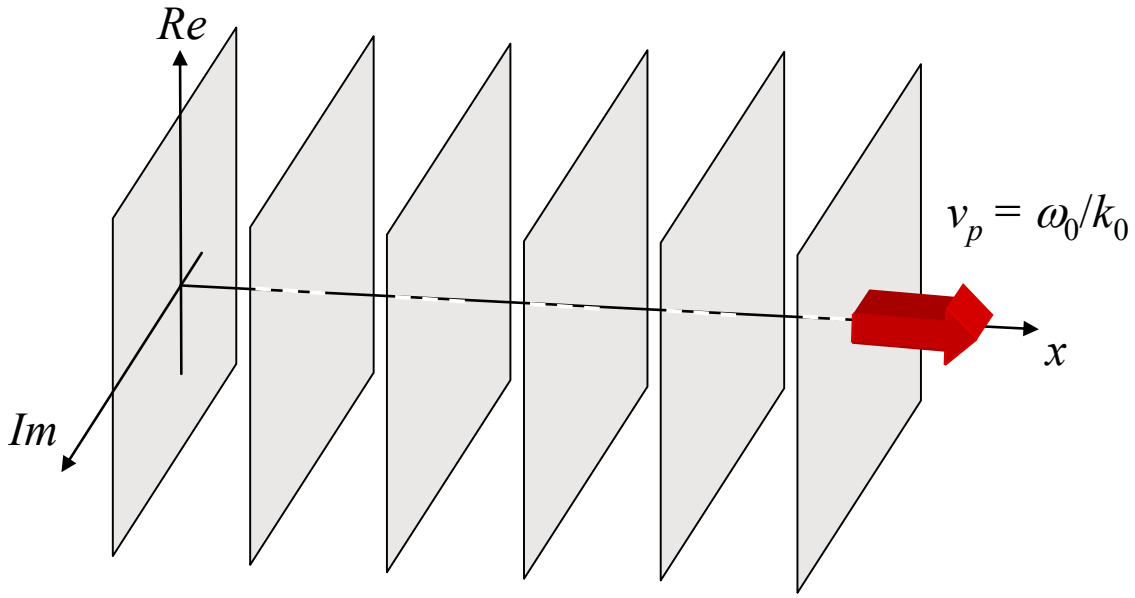


Fig. 1.6. In a plane wave planes of constant phase move through space at the phase velocity.

The double slit experiment

We now have the tools to model the double slit experiment described above. Far from the double slit, the electrons emanating from each slit look like plane waves; see Fig. 1.7, where s is the separation between the slits and L is the distance to the viewing screen.

At the viewing screen we have

$$\psi(x = L, t) = A\exp\left[i\left(k_0 r_1 - \omega_0 t\right)\right] + A\exp\left[i\left(k_0 r_2 - \omega_0 t\right)\right] \tag{1.10}$$

The intensity at the screen is

$$\begin{aligned}|\psi|^2 &= \left\{A\exp\left[-i\omega_0 t\right]\left(\exp\left[ik_0 r_1\right]+\exp\left[ik_0 r_2\right]\right)\right\}\left\{A\exp\left[-i\omega_0 t\right]\left(\exp\left[ik_0 r_1\right]+\exp\left[ik_0 r_2\right]\right)\right\}^* \\ &= |A|^2 + |A|^2\exp\left[i\left(k_0 r_2 - k_0 r_1\right)\right] + |A|^2\exp\left[i\left(k_0 r_1 - k_0 r_2\right)\right] + |A|^2 \\ &= 2|A|^2\left(1+\cos\left(k\left(r_2 - r_1\right)\right)\right)\end{aligned} \tag{1.11}$$

where A is a constant determined by the intensity of the electron wave. Now from Fig. 1.7:

$$\begin{aligned} r_1^2 &= L^2 + \left(s/2 - y\right)^2 \\ r_2^2 &= L^2 + \left(s/2 + y\right)^2 \end{aligned} \tag{1.12}$$

Now, if $y << s/2$ we can neglect the y^2 term:

$$\begin{aligned} r_1^2 &\approx L^2 + \left(s/2\right)^2 - sy \\ r_2^2 &\approx L^2 + \left(s/2\right)^2 + sy \end{aligned} \tag{1.13}$$

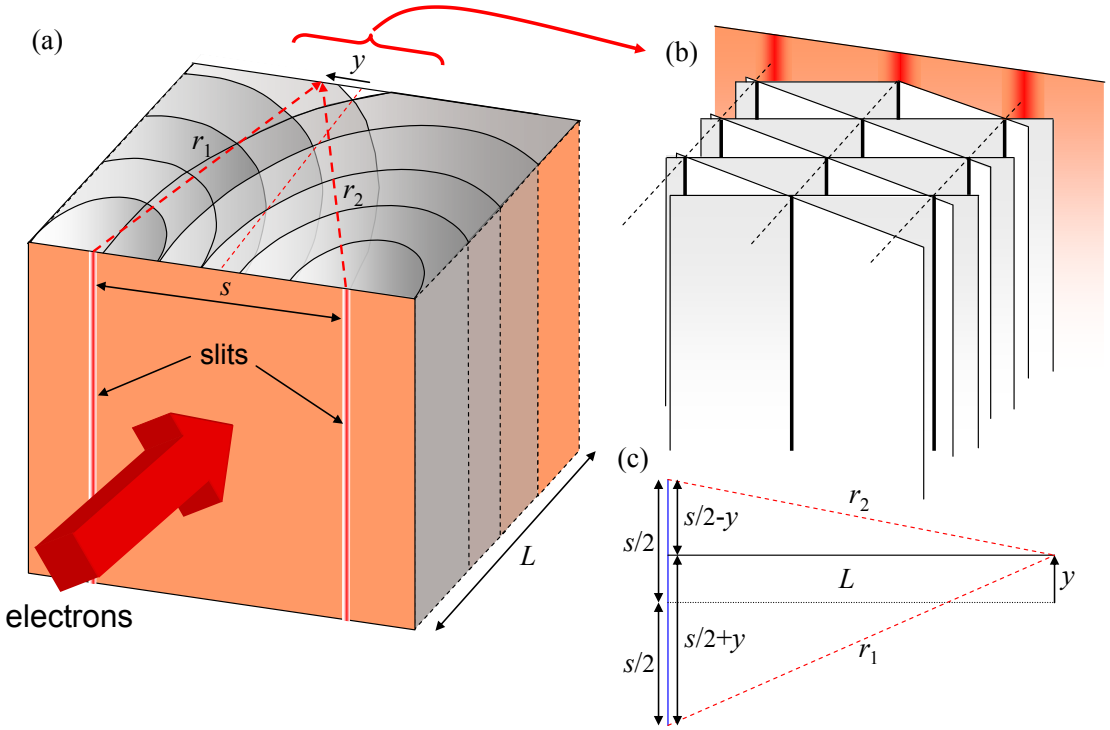


Fig. 1.7. Far from the double slit, the electrons from each slit can be described by plane waves. Where the planes of constant phase collide, a bright line corresponding to a high intensity of electrons is observed.

Then,

$$r_1 \approx \sqrt{L^2 + (s/2)^2}\left(1 - \frac{1}{2}\frac{sy}{L^2 + (s/2)^2}\right)$$
$$r_2 \approx \sqrt{L^2 + (s/2)^2}\left(1 + \frac{1}{2}\frac{sy}{L^2 + (s/2)^2}\right) \quad (1.14)$$

Next, if $L >> s/2$

$$r_1 \approx L - \frac{1}{2}\frac{sy}{L}$$
$$r_2 \approx L + \frac{1}{2}\frac{sy}{L} \quad (1.15)$$

Thus, $r_2 - r_1 = \frac{sy}{L}$, and

$$|\psi|^2 = 2|A|^2\left(1 + \cos\left(ks\frac{y}{L}\right)\right) \quad (1.16)$$

At the screen, *constructive* interference between the plane waves from each slit yields a regular array of bright lines, corresponding to a high intensity of electrons. In between each pair of bright lines, is a dark band where the plane waves interfere *destructively*, i.e. the waves are π radians out of phase with one another.

The spacing between the bright lines at the viewing screen is

$$\frac{2\pi}{\lambda}s\frac{y}{L} = 2\pi \quad (1.17)$$

Rearranging,

$$y = \frac{L\lambda}{s} \quad (1.18)$$

Interpretation of the double slit experiment

It is notable that the fringe pattern is independent of intensity. Thus, the interference effect should be observed even if just a single electron is fired at the slits at a time. For example, in Fig. 1.8 we show the buildup of the fringe pattern from consecutive electrons. The only conclusion is that the electron – which we are used to thinking of as a particle - also has wave properties.

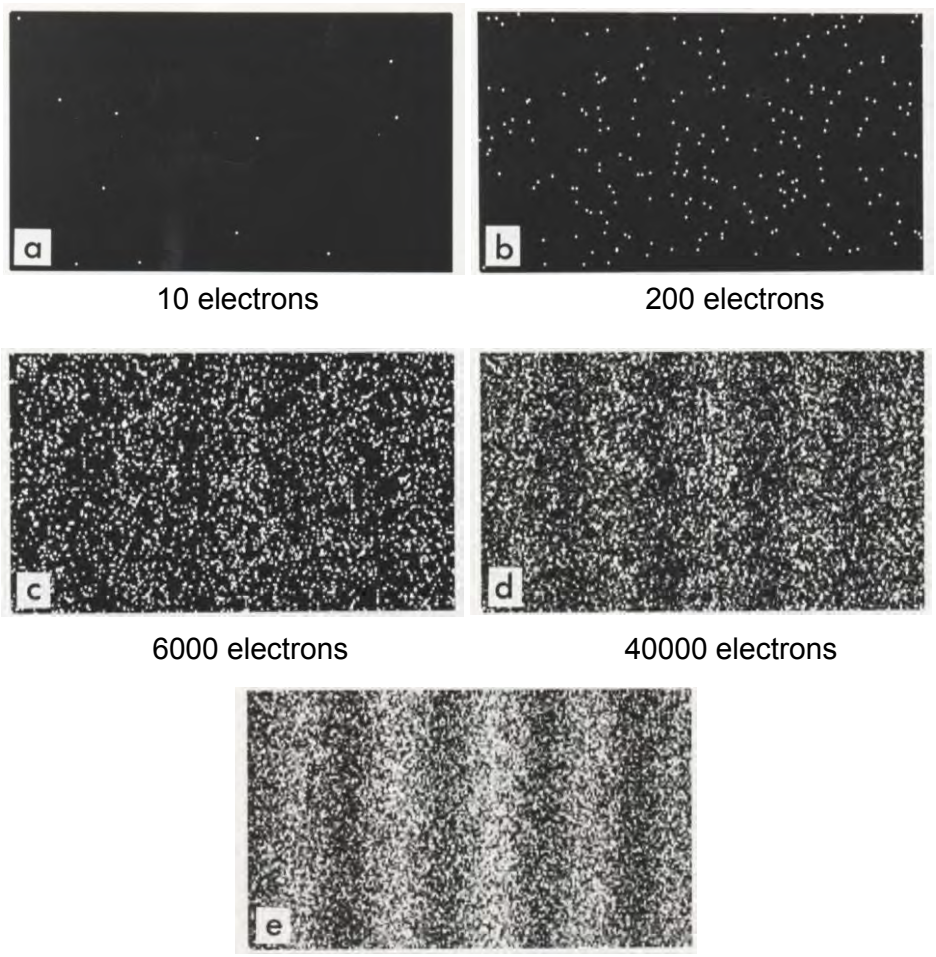


Fig. 1.8. The cumulative electron distribution after passage through a double slit. Just a single electron is present in the apparatus at any time. From A. Tanamura, *et al*. Am. J. Physics **57**, 117 (1989).

Part 1. The Quantum Particle

The Wavefunction

The wave-like properties of electrons are an example of the „wave-particle duality". Indeed, in the early 20th century, quantum mechanics revealed that a combination of wave and particle properties is a general property of everything at the size scale of an electron.

Without addressing the broader implications of this unusual observation, we will simply note that our purposes require a suitable mathematical description for the electron that can describe both its particle and wave-like properties. Following the conventions of quantum mechanics, we will define a function known as the wavefunction, $\psi(x,t)$, to describe the electron. It is typically a complex function and it has the important property that its magnitude squared is the probability density of the electron at a given position and time.

$$P(x,t) = \left|\psi(x,t)\right|^2 = \psi^*(x,t)\psi(x,t) \tag{1.19}$$

If the wavefunction is to describe a single electron, then the sum of its probability density over all space must be 1.

$$\int_{-\infty}^{+\infty} P(x,t)\,dx = 1 \tag{1.20}$$

In this case we say that the wavefunction is *normalized* such that the probability density sums to unity.

Frequency domain and *k*-space descriptions of waves

Consider the wavefunction

$$\psi(t) = a\,e^{-i\omega_0 t} \tag{1.21}$$

which describes a wave with amplitude a, intensity $|a|^2$, and phase oscillating in time at angular frequency ω_0. This wave carries two pieces of information, its amplitude and angular frequency.† Describing the wave in terms of a and ω_0 is known as the frequency domain description. In Fig. 1.9, we plot the wavefunction in both the time and frequency domains.

In the frequency domain, the wavefunction is described by a delta function at ω_0. Tools for the exact conversion between time and frequency domains will be presented in the next section. Note that, by convention we use a capitalized function (A instead of ψ) to represent the wavefunction in the frequency domain. Note also that the convention in quantum mechanics is to use a negative sign in the phase when representing the angular frequency $+\omega_0$. This is convenient for describing plane waves of the form $e^{i(kx-\omega t)}$. But it is exactly opposite to the usual convention in signal analysis (i.e. 6.003). In general, when you see i instead of j for the square root of -1, use this convention in the time and frequency domains.

---

† Note the information in any constant phase offset, ϕ, as in $\psi = \exp\left[i\omega_0 t + i\phi\right]$ can be contained in the amplitude prefactor, i.e. $\psi = a\exp\left[i\omega_0 t\right]$, where $a = \exp\left[i\phi\right]$.

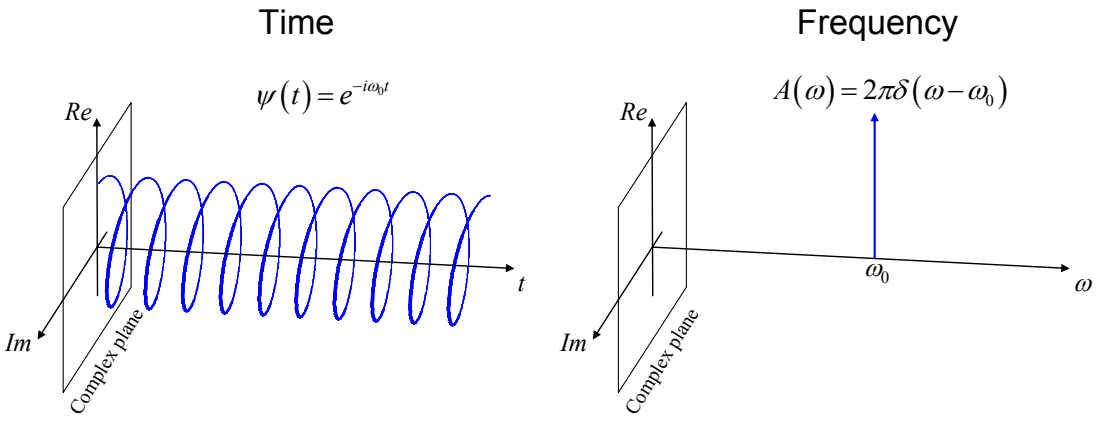


Fig. 1.9. Representations of the angular frequency ω_0 in time and frequency domains.

Similarly, consider the wavefunction

$$\psi(x) = a\, e^{ik_0 x} \tag{1.22}$$

which describes a wave with amplitude a, intensity $|a|^2$, and phase oscillating in *space* with spatial frequency or wavenumber, k_0. Again, this wave carries two pieces of information, its amplitude and wavenumber. We can describe this wave in terms of its spatial frequencies in *k-space*, the equivalent of the frequency domain for spatially oscillating waves. In Fig. 1.10, we plot the wavefunction in real space and k-space.

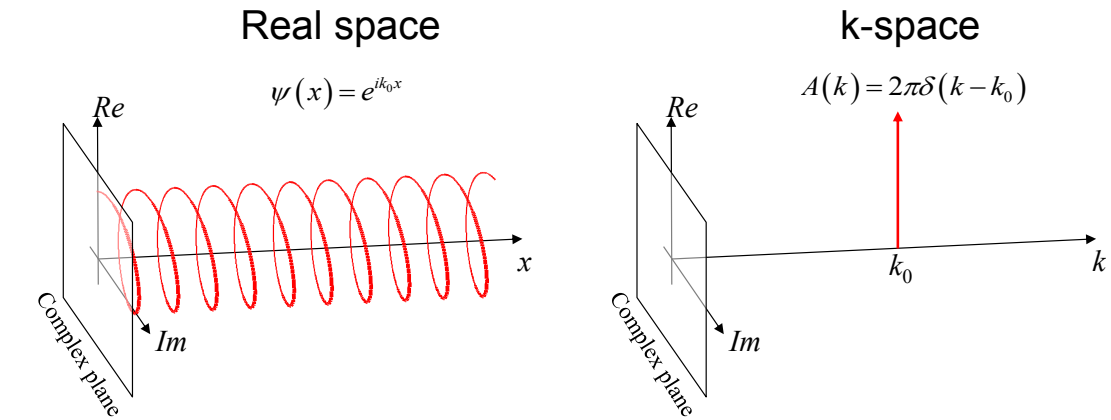


Fig. 1.10. Representations of wavenumber k_0 in real space and k-space.

Next, let"s consider the wavefunction

$$A(\omega) = a\, e^{i\omega t_0} \tag{1.23}$$

which describes a wave with amplitude a, intensity $|a|^2$, and phase oscillating in the frequency domain with period $2\pi / t_0$. This wave carries two pieces of information, its amplitude and the time t_0. In Fig. 1.11, we plot the wavefunction in both the time and frequency domains.

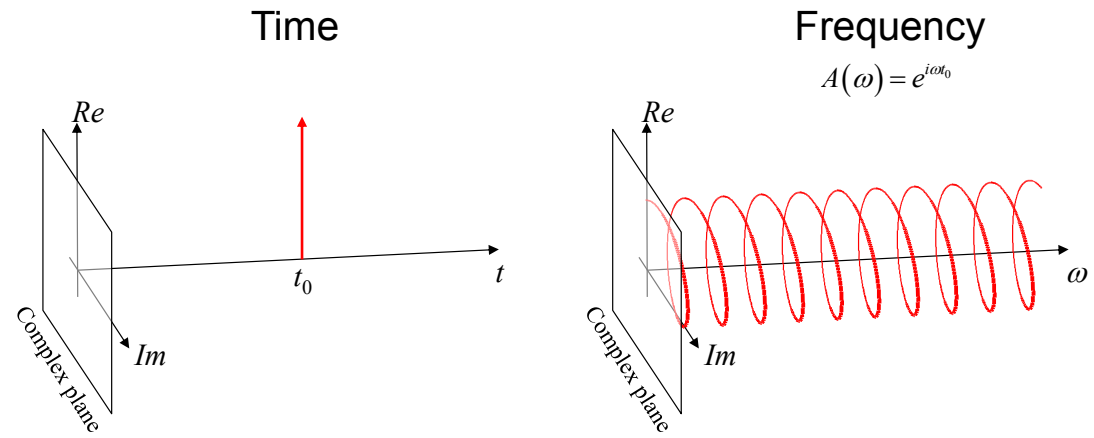


Fig. 1.11. Representations of time t_0 in time and frequency domains.

Finally, consider the wavefunction

$$A(k) = a\, e^{-ikx_0} \tag{1.24}$$

which describes a wave with amplitude a, intensity $|a|^2$, and phase oscillating in k-space with period $2\pi/x_0$. This waves carries two pieces of information, its amplitude and the position x_0. In Fig. 1.12, we plot the wavefunction in both real space and k-space.

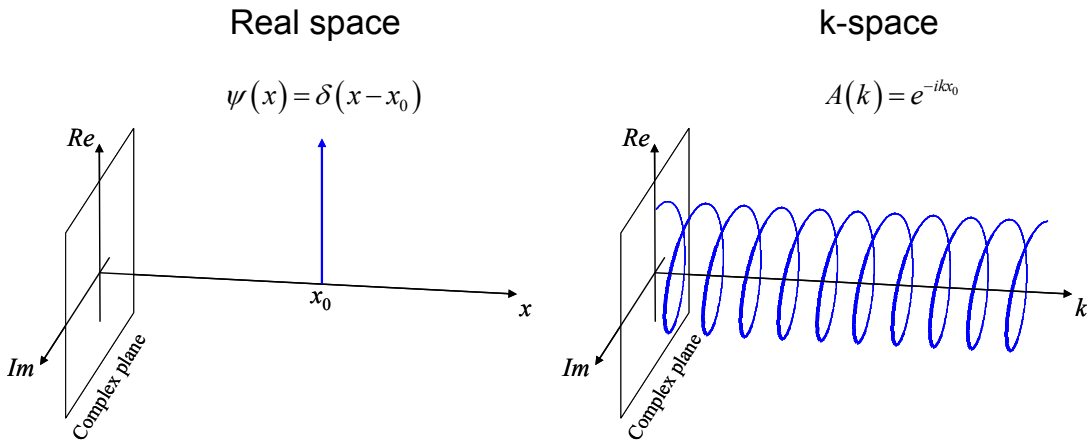


Fig. 1.12. Representations of position x_0 in real space and k-space.

Note that, by convention we use a capitalized function (A instead of ψ) to represent the wavefunction in the k-space domain.

Observe in Fig. 1.9-Fig. 1.12 that a precise definition of both the position in time and the angular frequency of a wave is impossible. A wavefunction with angular frequency of precisely ω_0 is uniformly distributed over all time. Similarly, a wavefunction associated with a precise time t_0 contains all angular frequencies.

In real and k-space we also cannot precisely define both the wavenumber and the position. A wavefunction with a wavenumber of precisely k_0 is uniformly distributed over

all space. Similarly, a wavefunction localized at a precise position x_0 contains all wavenumbers.

Linear combinations of waves

Next, we consider the combinations of different complex exponential functions. For example, in Fig. 1.13 we plot a wavefunction that could describe an electron that equiprobable at position x_1 and position x_2. The k-space representation is simply the *superposition* of two complex exponential functions corresponding to x_1 and x_2.[†]

$$\psi(x) = c_1\delta(x-x_1) + c_2\delta(x-x_2) \quad \Leftrightarrow \quad A(\omega) = c_1 e^{-ikx_1} + c_2 e^{-ikx_2} \tag{1.25}$$

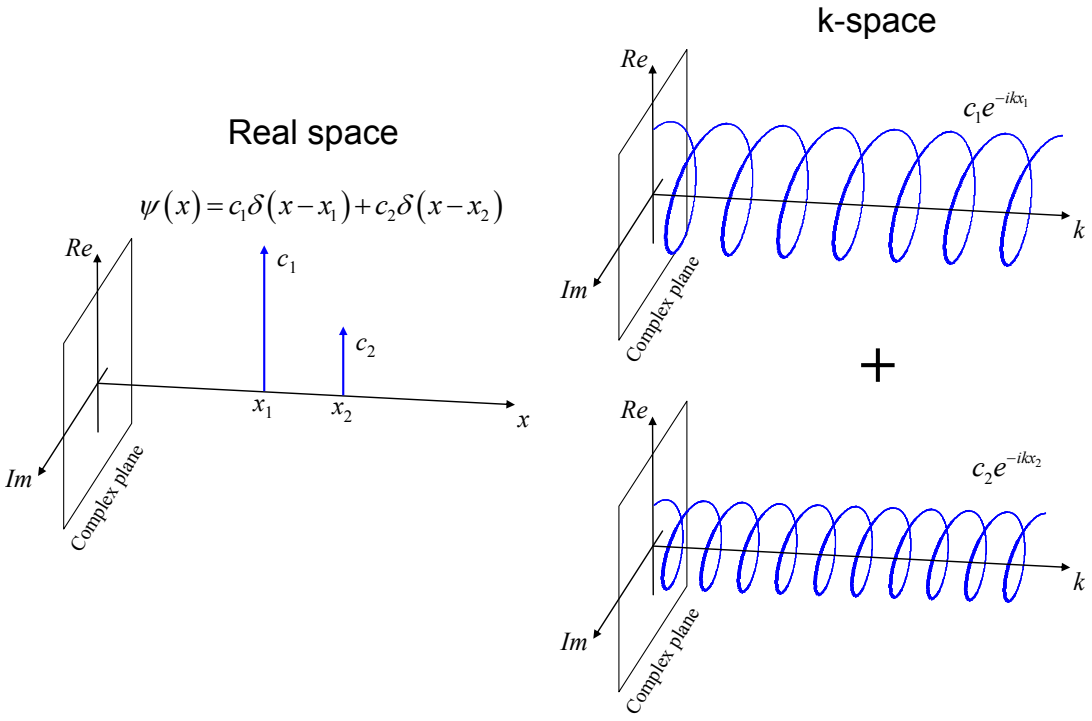


Fig. 1.13. The k-space wavefunction corresponding to two positions x_1 and x_2 is simply the superposition of the k-space representations of $\delta(x\text{-}x_1)$ and $\delta(x\text{-}x_2)$.

We can also generalize to an arbitrary distribution of positions, $\psi(x)$. If $\psi(x)$ describes an electron, for example, the probability that the electron is located at position x is $|\psi(x)|^2$. Thus, in k-space the electron is described by the sum of complex exponentials e^{-ikx} each oscillating in k-space and weighted by amplitude $\psi(x)$.

$$A(k) = \int_{-\infty}^{+\infty} \psi(x) e^{-ikx} dx \tag{1.26}$$

You may recognize this from 6.003 as a Fourier transform. Similarly, the inverse transform is

[†] Note that this wave function is not actually normalizable.

$$\psi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} A(k)e^{ikx}dk\,. \tag{1.27}$$

To convert between time and angular frequency, use

$$A(\omega) = \int_{-\infty}^{+\infty} \psi(t)e^{i\omega t}dt \tag{1.28}$$

and

$$\psi(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} A(\omega)e^{-i\omega t}d\omega\,. \tag{1.29}$$

Note that the factors of $1/2\pi$ are present each time you integrate with respect to k or ω. Note also that when converting between complex exponentials and delta functions, the following identity is useful:

$$2\pi\delta(u) = \int_{-\infty}^{+\infty} \exp[iux]dx\,. \tag{1.30}$$

Wave packets and uncertainty

We now have two ways to describe an electron. We could describe it as a plane wave, with precisely defined wavenumber and angular frequency:

$$\psi(x,t) = e^{i(k_0x-\omega_0t)}\,. \tag{1.31}$$

But as we have seen, the intensity/probability density of the plane wave is uniform over all space (and all time). Thus, the position of the electron is perfectly uncertain – it is probability distribution is uniform everywhere in the entire universe. Consequently, a plane wave is not usually a good description for an electron.

On the other hand, we could describe the electron as an idealized point particle

$$\psi(x,t) = \delta(x-x_0, t-t_0) \tag{1.32}$$

existing at a precisely defined position and time. But the probability density of the point particle is uniform over all of k-space and the frequency domain. We will see in the next section that this means the energy and momentum of the electron is perfectly uncertain, i.e. arbitrarily large electron energies and momenta are possible.

The only alternative is to accept an imprecise description of the electron in both real space and k-space, time and the frequency domain. A localized oscillation in both representations is called a *wave packet*. A common wavepacket shape is the Gaussian.
For example, instead of the delta function, we could describe the electron‟s position as

$$\psi(x) = a\exp\left[-\frac{1}{4}\left(\frac{x}{\sigma_x}\right)^2\right]. \tag{1.33}$$

This function was chosen such that the probability distribution of the electron is

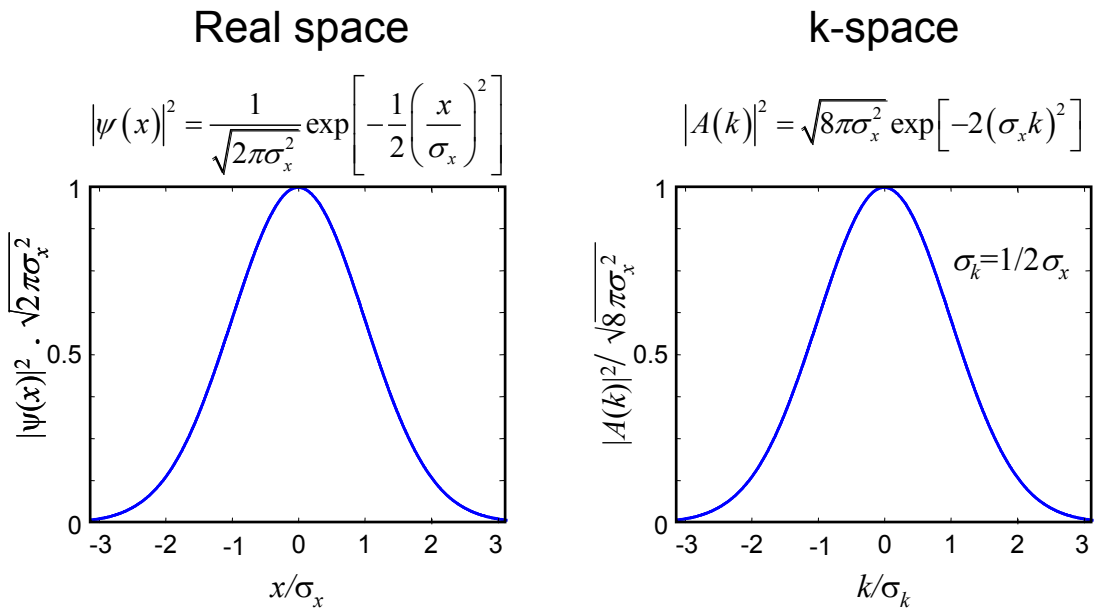


Fig. 1.14. A normalized Gaussian. The Gaussian has the same shape in real space and *k*-space. Note that a Gaussian approximates a delta function in the limit $\sigma_x \to 0$.

$$|\psi(x)|^2 = |a|^2 \exp\left[-\frac{1}{2}\left(\frac{x}{\sigma_x}\right)^2\right] \tag{1.34}$$

where σ_x is the standard deviation. σ_x measures the width of the Gaussian and is often thought of as the *uncertainty* in the position of the electron. The constant, a, is determined by normalizing the probability density over all space, i.e., integrating Eq. (1.34) over x, we get

$$a = \left(2\pi\sigma_x^2\right)^{-1/4} \tag{1.35}$$

Strictly, the uncertainty of a given quantity is defined by

$$\begin{aligned}\sigma^2 &= \left\langle \left(x - \langle x\rangle\right)^2 \right\rangle \\ &= \left\langle x^2 - 2x\langle x\rangle + \langle x\rangle^2 \right\rangle\end{aligned} \tag{1.36}$$

where $\langle x\rangle$ signifies the average or expectation value of x. Because $\langle x\rangle$ is a constant Eq. (1.36) may be simplified:

$$\begin{aligned}\sigma^2 &= \left\langle x^2\right\rangle - 2\langle x\rangle\langle x\rangle + \langle x\rangle^2 \\ &= \left\langle x^2\right\rangle - \langle x\rangle^2\end{aligned} \tag{1.37}$$

We will leave it as an exercise to show that for the Gaussian probability density:

$$\sigma^2 = \int x^2 |\psi(x)|^2 \, dx = \sigma_x^2 \tag{1.38}$$

Thus, the Gaussian is a convenient choice to describe a wavepacket because it has a readily defined uncertainty.
In *k*-space, the electron is also described by a Gaussian (this is another of the convenient properties of this function). Application of the Fourier transform in Eq. (1.26) and some algebra gives

$$A(k) = \left(8\pi\sigma_x^2\right)^{1/4} \exp\left[-\sigma_x^2 k^2\right] \tag{1.39}$$

The probability distribution in *k*-space is

$$\left|A(k)\right|^2 = \left(8\pi\sigma_x^2\right)^{1/2} \exp\left[-2\sigma_x^2 k^2\right] \tag{1.40}$$

Thus, the uncertainty in *k*-space is

$$\frac{1}{2\sigma_k^2} \equiv 2\sigma_x^2 \tag{1.41}$$

The product of the uncertainties in real and k-space is

$$\left|\sigma_x \sigma_k\right| = \frac{1}{2}. \tag{1.42}$$

The product $|\sigma_x||\sigma_k| \geq |\sigma_x \sigma_k|$. Thus,

$$|\sigma_x||\sigma_k| \geq \frac{1}{2}. \tag{1.43}$$

Examples of wavepackets

A typical Gaussian wavepacket is shown in Fig. 1.15 in both its real space and *k*-space representations. Initially the probability distribution is centered at $x = 0$ and $k = 0$. If we shift the wavepacket in k-space to an average value $\langle k \rangle = k_0$, this is equivalent to multiplying by a phase factor $\exp[ik_0x]$ in real space. Similarly, shifting the center of the wavepacket in real space to $\langle x \rangle = x_0$ is equivalent to multiplying the *k*-space representation by a phase factor $\exp[-ikx_0]$.

Real coordinates (x,t)	$\rightleftharpoons$	**Inverse coordinates** (k,ω)
shift by x_0	$\rightleftharpoons$	$\times \exp[-ikx_0]$
$\times \exp[ik_0x]$	$\rightleftharpoons$	shift by k_0
shift by t_0	$\rightleftharpoons$	$\times \exp[i\omega t_0]$
$\times \exp[-i\omega_0 t]$	$\rightleftharpoons$	shift by ω_0

Table 1.1. A summary of shift transformations in real and inverse coordinates.

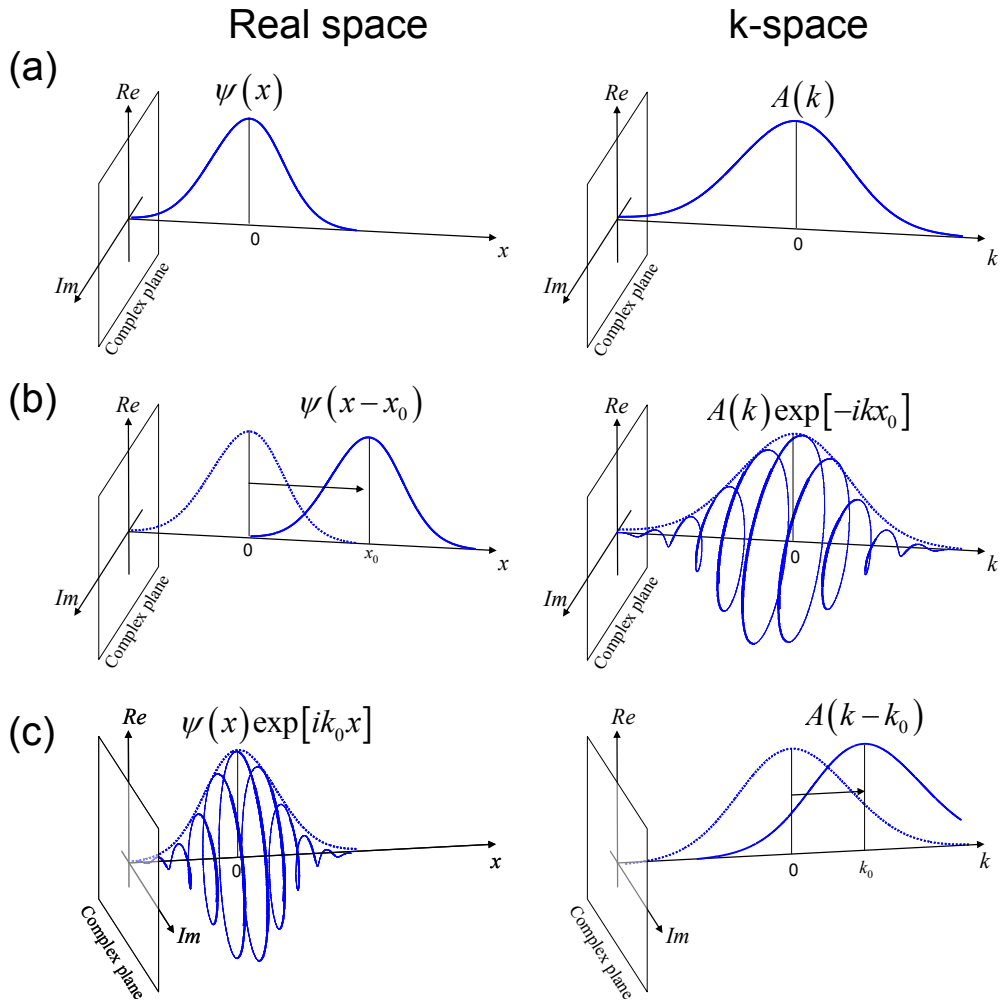


Fig. 1.15. Three Gaussian wavepackets. In **(a)** the average position and wavenumber of the packet is x=0 and k=0, respectively. In (b) the average position has been shifted to $\langle x\rangle = x_0$. In **(c)** the average wavenumber has been shifted to $\langle k\rangle = k_0$.

Expectation values of position

Given that $P(x)$ is the probability density of the electron at position x, we can determine the average, or *expectation value* of x from

$$\langle x\rangle = \frac{\int_{-\infty}^{+\infty} xP(x)\,dx}{\int_{-\infty}^{+\infty} P(x)\,dx} \tag{1.44}$$

Of course if the wavefunction is normalized then the denominator is 1.

We could also write this in terms of the wavefunction

$$\langle x \rangle = \frac{\int_{-\infty}^{+\infty} x|\psi(x)|^2 \, dx}{\int_{-\infty}^{+\infty} |\psi(x)|^2 \, dx} \tag{1.45}$$

Where once again if the wavefunction is normalized then the denominator is 1.

Since $|\psi(x)|^2 = \psi(x)^* \psi(x)$,

$$\langle x \rangle = \frac{\int_{-\infty}^{+\infty} \psi(x)^* x \psi(x) \, dx}{\int_{-\infty}^{+\infty} \psi(x)^* \psi(x) \, dx} \tag{1.46}$$

Bra and Ket Notation

Also known as Dirac notation, Bra and Ket notation is a convenient shorthand for the integrals above.

The wavefunction is represented by a Ket:

$$\psi(x) \rightarrow |\psi\rangle \tag{1.47}$$

The complex conjugate is represented by a Bra:

$$\psi^*(x) \rightarrow \langle\psi| \tag{1.48}$$

Together, the bracket $\langle\psi|\psi\rangle$ (hence Bra and Ket) symbolizes an integration over all space:

$$\int_{-\infty}^{+\infty} \psi^*(x)\psi(x) \, dx \rightarrow \langle\psi|\psi\rangle \tag{1.49}$$

Thus, in short form the expectation value of x is

$$\langle x \rangle = \frac{\langle\psi|x|\psi\rangle}{\langle\psi|\psi\rangle} \tag{1.50}$$

Parseval's Theroem

It is often convenient to normalize a wavepacket in k space. To do so, we can apply Parseval"s theorem.

Let"s consider the bracket of two functions, $f(x)$ and $g(x)$ with Fourier transform pairs $F(k)$ and $G(k)$, respectively..

$$\langle f|g\rangle = \int_{-\infty}^{\infty} f(x)^* g(x) \, dx \tag{1.51}$$

Now, replacing the functions by their Fourier transforms yields

$$\int_{-\infty}^{\infty} f(x)^* g(x)\,dx = \int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} F(k')e^{-ik'x}dk'\right]^*\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} G(k)e^{-ikx}dk\right]dx \tag{1.52}$$

Rearranging the order of integration gives

$$\begin{aligned}&\int_{-\infty}^{\infty}\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} F(k')e^{-ik'x}dk'\right]^*\left[\frac{1}{2\pi}\int_{-\infty}^{\infty} G(k)e^{-ikx}dk\right]dx\\ &\qquad=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(k')^* G(k)\frac{1}{2\pi}e^{-i(k-k')x}dxdk'dk\end{aligned} \tag{1.53}$$

From Eq. (1.30) the integration over the complex exponential yields a delta function

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(k')^* G(k)\frac{1}{2\pi}e^{-i(k-k')x}dxdk'dk = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(k')^* G(k)\delta(k-k')dk'dk \tag{1.54}$$

Thus,

$$\int_{-\infty}^{\infty} f(x)^* g(x)\,dx = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(k)^* G(k)dk \tag{1.55}$$

It follows that if a wavefunction is normalized in real space, it is also normalized in *k*-space, i.e.,

$$\langle\psi|\psi\rangle = \langle A|A\rangle \tag{1.56}$$

where

$$\langle A|A\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} A(k)^* A(k)dk \tag{1.57}$$

Expectation values of *k* and ω

The expectation value of *k* is obtained by integrating the wavefunction over all *k*. This must be performed in *k*-space.

$$\langle k\rangle = \frac{\frac{1}{2\pi}\int_{-\infty}^{+\infty} A(k)^* kA(k)dk}{\frac{1}{2\pi}\int_{-\infty}^{+\infty} A(k)^* A(k)dk} = \frac{\langle A|k|A\rangle}{\langle A|A\rangle} \tag{1.58}$$

From the Inverse Fourier transform in *k*-space

$$\psi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} A(k)e^{ikx}dk \tag{1.59}$$

note that

$$-i\frac{d}{dx}\psi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} kA(k)e^{ikx}dk \tag{1.60}$$

Thus, we have the following Fourier transform pair:

$$-i\frac{d}{dx}\psi(x) \Leftrightarrow k\,A(k) \tag{1.61}$$

It follows that[†]

$$\langle k \rangle = \frac{\langle A | k | A \rangle}{\langle A | A \rangle} = \frac{\langle \psi | -i \frac{d}{dx} | \psi \rangle}{\langle \psi | \psi \rangle} \tag{1.62}$$

Similarly, from the Inverse Fourier transform in the frequency domain

$$\psi(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} A(\omega) e^{-i\omega t} d\omega \tag{1.63}$$

We can derive the Fourier transform pair:

$$i \frac{d}{dt} \psi(t) \Leftrightarrow \omega A(\omega) \tag{1.64}$$

It follows that

$$\langle \omega \rangle = \frac{\langle A | \omega | A \rangle}{\langle A | A \rangle} = \frac{\langle \psi | i \frac{d}{dt} | \psi \rangle}{\langle \psi | \psi \rangle} . \tag{1.65}$$

We define two *operators*

$$\hat{k} = -i \frac{d}{dx} \tag{1.66}$$

and

$$\hat{\omega} = i \frac{d}{dt} . \tag{1.67}$$

Operators only act on functions to the right. To signify this difference we mark them with a caret.

We could also define the (somewhat trivial) position operator

$$\hat{x} = x . \tag{1.68}$$

The Commutator

One must be careful to observe the correct order of operators. For example,

$$\hat{x} \hat{k} \neq \hat{k} \hat{x} \tag{1.69}$$

but

$$\hat{x} \hat{\omega} = \hat{\omega} \hat{x} . \tag{1.70}$$

In quantum mechanics we define the commutator:

$$[\hat{q}, \hat{r}] = \hat{q} \hat{r} - \hat{r} \hat{q} \tag{1.71}$$

We find that the operators $\hat{x}$ and $\hat{\omega}$ commute because $[\hat{x}, \hat{\omega}] = 0$.

Considering the operators $\hat{x}$ and $\hat{k}$:

$$\left[\hat{x}, \hat{k}\right] = -ix \frac{d}{dx} + i \frac{d}{dx} x \tag{1.72}$$

[†] We have applied Parseval"s theorem; see the Problem Sets.

To simplify this further we need to operate on some function, $f(x)$:

$$\begin{aligned}\left[\hat{x},\hat{k}\right]f(x) &= -ix\frac{df}{dx} + i\frac{d}{dx}(xf) \\ &= -ix\frac{df}{dx} + if\frac{dx}{dx} + ix\frac{df}{dx} \\ &= if\end{aligned} \tag{1.73}$$

Thus, the operators $\hat{x}$ and $\hat{k}$ do not commute, i.e.

$$\left[\hat{x},\hat{k}\right] = i \tag{1.74}$$

Although we used Fourier transforms, Eq. (1.43) can also be derived from the relation (1.74) for the non-commuting operators operators $\hat{x}$ and $\hat{k}$. It follows that all operators that do not commute are subject to a similar limit on the product of their uncertainties. We shall see in the next section that this limit is known as „the uncertainty principle".

Part 1. The Quantum Particle

Momentum and Energy

Two key experiments revolutionized science at the turn of the 20th century. Both experiments involve the interaction of light and electrons. We have already seen that electrons are best described by wavepackets. Similarly, light is carried by a wavepacket called a photon. The first phenomenon, the photoelectric effect, was explained by assuming that a photon's energy is proportional to its frequency. The second phenomenon, the Compton effect, was explained by proposing that photons carry momentum. That light should possess particle properties such as momentum was completely unexpected prior to the advent of quantum mechanics.

(i) The Photoelectric Effect

It is not easy to pull electrons out of a solid. They are bound by their attraction to positive nuclei. But if we give an electron in a solid enough energy, we can overcome the binding energy and liberate an electron. The minimum energy required is known as the work function, W.

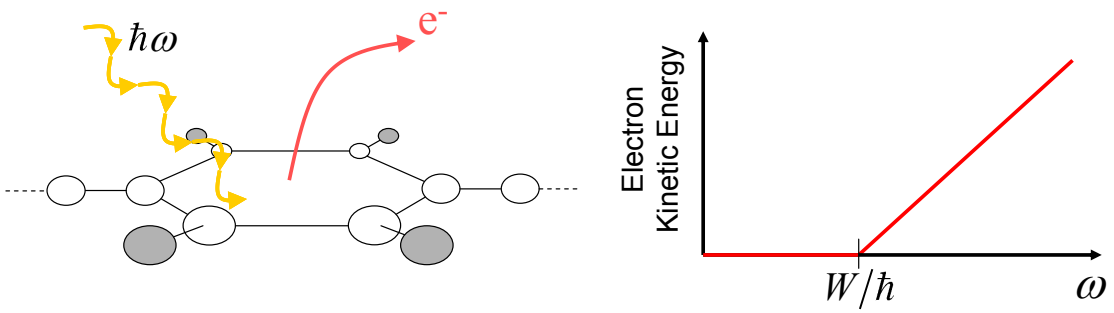


Fig. 1.16. The photoelectric effect: above a critical frequency, light can liberate electrons from a solid.

By bombarding metal surfaces with light, it was observed that electrons could be liberated only if the frequency of the light exceeded a critical value. Above the minimum frequency, electrons were liberated with greater kinetic energy.

Einstein explained the photoelectric effect by postulating that, in a photon, the energy was proportional to the frequency:

$$E = \hbar\omega \tag{1.75}$$

Where $h = 6.62\times10^{-34}$ Js is Planck's constant, and $\hbar$ is shorthand for $\hbar = h/2\pi$. Note the units for Planck's constant – energy × time. This is useful to remember when checking that your quantum calculations make sense.

Thus, the kinetic energy of the emitted electrons is given by

$$electron\ kinetic\ energy = \hbar\omega - W. \tag{1.76}$$

This technique is still used to probe the energy structure of materials

Note that we will typically express the energy of electrons in „electron Volts (eV)". The SI unit for energy, the Joule, is typically much too large for convenient discussion of electron energies. A more convenient unit is the energy required to move a single electron through a potential difference of 1V. Thus, 1 eV = q J, where q is the charge on an electron ($q \sim 1.602\times10^{-19}$ C).

(ii) The Compton Effect

If a photon collides with an electron, the wavelength and trajectory of the photon is observed to change. After the collision the scattered photon is red shifted, i.e. its frequency is reduced and its wavelength extended. The trajectory and wavelength of the photon can be calculated by assuming that the photon carries momentum:

$$p = \hbar k \,, \tag{1.77}$$

where the wavenumber k is related to frequency by

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c} \,, \tag{1.78}$$

where c is the speed of light.

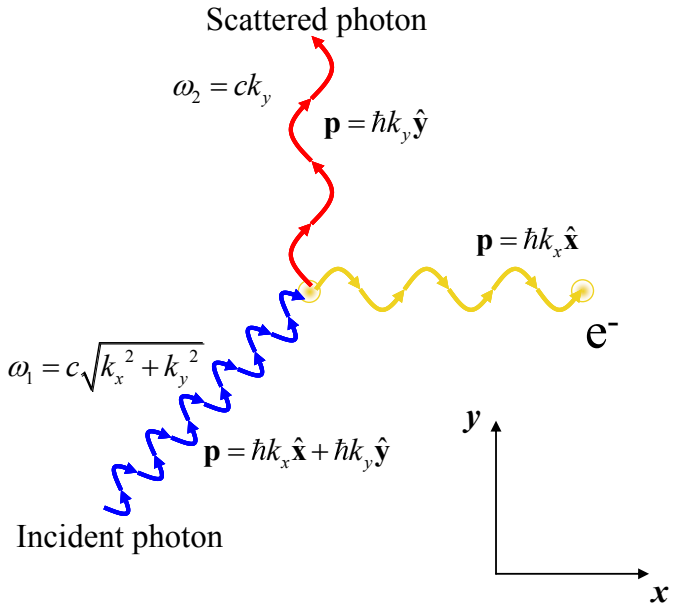


Fig. 1.17. The wavelength shift of light after collision with an electron is consistent with a transfer of momentum from a photon to the electron. The loss of photon energy is reflected in a red shift of its frequency.

These two relations: $E = \hbar\omega$ and $p = \hbar k$ are strictly true only for plane waves with precisely defined values of ω and k. Otherwise we must employ operators for momentum and energy. Based on the operators we defined earlier for k and ω, we define operators for momentum

$$\hat{p} = -i\hbar \frac{d}{dx} \tag{1.79}$$

and energy

$$\hat{E} = i\hbar \frac{d}{dt} \tag{1.80}$$

Recall that each operator acts on the function to its right; and that $\hat{p}\psi$ is not necessarily equal to $\psi \hat{p}$.

The Uncertainty Principle

Now that we see that k is simply related to momentum, and ω is simply related to energy, we can revisit the uncertainty relation of Eq. (1.43)

$$|\sigma_x||\sigma_k| \geq \frac{1}{2}, \tag{1.81}$$

which after multiplication by $\hbar$ becomes

$$\Delta p.\Delta x \geq \frac{\hbar}{2}. \tag{1.82}$$

This is the celebrated Heisenberg uncertainty relation. It states that we can never know both position and momentum exactly.

For example, we have seen from our Fourier transform pairs that to know position exactly means that in k-space the wavefunction is $\Psi(k) = \exp[-ikx_0]$. Since $|\Psi(k)|^2 = 1$ all values of k, and hence all values of momentum are equiprobable. Thus, momentum is perfectly undefined if position is perfectly defined.

Uncertainty in Energy and Time

For the time/frequency domain, we take Fourier transforms and write

$$\Delta E.\Delta t \geq \frac{\hbar}{2}. \tag{1.83}$$

However, time is treated differently to position, momentum and energy. There is no operator for time. Rather it is best thought of as a parameter. But the expression of Eq. (1.83) still holds when Δt is interpreted as a lifetime.

Application of the Uncertainty Principle

The uncertainty principle is not usually significant in every day life. For example, if the uncertainty in momentum of a 200g billiard ball traveling at a velocity of 1m/s is 1%, we can in principle know its position to $\Delta x = (\hbar/2)/(0.2/100) = 3 \times 10^{-32}$ m.

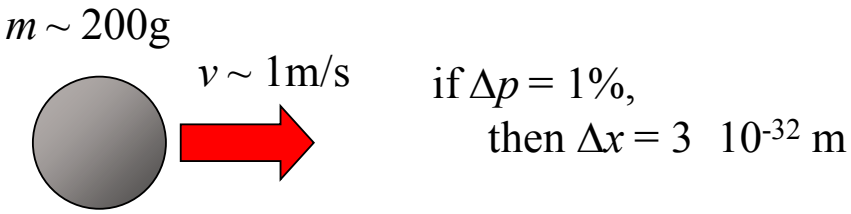


Fig. 1.18. The uncertainty principle is not very relevant to everyday objects.

In nanoelectronics, however, the uncertainty principle can play a role.

For example, consider a very thin wire through which electrons pass one at a time. The current in the wire is related to the transit time of each electron by

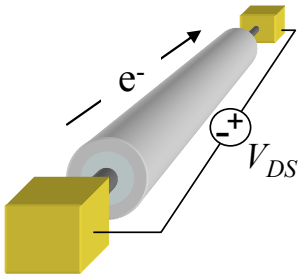


Fig. 1.19. A nanowire that passes one electron at a time.

$$I = \frac{q}{\tau}, \tag{1.84}$$

where q is the charge of a single electron.

To obtain a current of $I = 0.1$ mA in the wire the transit time of each electron must be

$$\tau = \frac{q}{I} \approx 1.6 fs, \tag{1.85}$$

The transit time is the time that electron exists within the wire. Some electrons may travel through the wire faster, and some slower, but we can approximate the uncertainty in the electron‟s lifetime, $\Delta t = \tau = 1.6$ fs.[†]

From Eq. (1.83) we find that $\Delta E = 0.2$ eV. Thus, the uncertainty in the energy of the electron is equivalent to a random potential of approximately 0.2 V.[§] As we shall see, such effects fundamentally limit the switching characteristics of nano transistors.

Schrödinger's Wave Equation

The energy of our electron can be broken into two parts, kinetic and potential. We could write this as

$$total\ energy = kinetic\ energy + potential\ energy \tag{1.86}$$

Now kinetic energy is related to momentum by

$$kinetic\ energy = \frac{1}{2}mv^2 = \frac{p^2}{2m} \tag{1.87}$$

Thus, using our operators, we could write

$$\hat{E}\psi(x,t) = \frac{\hat{p}^2}{2m}\psi(x,t) + \hat{V}\psi(x,t) \tag{1.88}$$

[†] Another way to think about this is to consider the addition of an electron to the nanowire. If current is to flow, that electron must be able to move from the wire to the contact. The rate at which it can do this (i.e. its lifetime on the wire) limits the transit time of an electron and hence the current that can flow in the wire.
[§] Recall that modern transistors operate at voltages ~ 1V. So this uncertainty is substantial.

Where $\hat{V}$ is the potential energy operator.

$$\hat{V} = V(x,t). \tag{1.89}$$

We can rewrite Eq. (1.88) in even simpler form by defining the so called Hamiltonian operator

$$\hat{H} = \frac{\hat{p}^2}{2m} + \hat{V}. \tag{1.90}$$

Now,

$$\hat{E}|\psi\rangle = \hat{H}|\psi\rangle. \tag{1.91}$$

Or we could rewrite the expression as

$$i\hbar \frac{d}{dt}\psi(x,t) = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x,t) + V(x,t)\psi(x,t). \tag{1.92}$$

All these equations are statements of Schrödinger"s wave equation. We can employ whatever form is most convenient.

A Summary of Operators

Note there is no operator for time.

Position	$\hat{x}$	x	Energy	$\hat{E}$	$i\hbar\frac{d}{dt}$
Wavenumber	$\hat{k}$	$-i\frac{d}{dx}$	Potential Energy	$\hat{V}$	V
Angular Frequency	$\hat{\omega}$	$i\frac{d}{dt}$	Kinetic Energy	$\hat{T}$	$\frac{\hat{p}^2}{2m}$
Momentum	$\hat{p}$	$-i\hbar\frac{d}{dx}$	Hamiltonian	$\hat{H}$	$\frac{\hat{p}^2}{2m}+\hat{V}$

Table 1.2. All the operators used in this class.

The Time Independent Schrödinger Equation

When the potential energy is constant in time we can simplify the wave equation. We assume that the spatial and time dependencies of the solution can be separated, i.e.

$$\Psi(x,t) = \psi(x)\zeta(t) \tag{1.93}$$

Substituting this into Eq. (1.92) gives

$$i\hbar\psi(x)\frac{d}{dt}\zeta(t) = -\frac{\hbar^2}{2m}\zeta(t)\frac{d^2}{dx^2}\psi(x) + V(x)\psi(x)\zeta(t) \tag{1.94}$$

Dividing both sides by $\psi(x)\zeta(t)$ yields

$$i\hbar\frac{1}{\zeta(t)}\frac{d}{dt}\zeta(t)=-\frac{\hbar^2}{2m}\frac{1}{\psi(x)}\frac{d^2}{dx^2}\psi(x)+V(x). \tag{1.95}$$

Now the left side of the equation is a function only of time while the right side is a function only of position. These are equal for all values of time and position if each side equals a constant. That constant turns out to be the energy, E, and we get two coupled equations

$$E\zeta(t)=i\hbar\frac{d}{dt}\zeta(t) \tag{1.96}$$

and

$$E\psi(x)=-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x)+V(x)\psi(x). \tag{1.97}$$

The solution to Eq. (1.96) is

$$\zeta(t)=\zeta(0)\exp\left[-i\frac{E}{\hbar}t\right] \tag{1.98}$$

Thus, the complete solution is

$$\Psi(x,t)=\psi(x)\exp\left[-i\frac{E}{\hbar}t\right] \tag{1.99}$$

By separating the wavefunction into time and spatial functions, we need only solve the simplified Eq. (1.97).

There is much more to be said about this equation, but first let"s do some examples.

Free Particles

In free space, the potential, V, is constant everywhere. For simplicity we will set $V = 0$.

Next we solve Eq. (1.97) with $V = 0$.

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x)=E\psi(x) \tag{1.100}$$

Rearranging slightly gives the second order differential equation in slightly clearer form

$$\frac{d^2\psi}{dx^2}=-\frac{2mE}{\hbar^2}\psi \tag{1.101}$$

A general solution is

$$\psi(x)=\psi(0)\exp[ikx] \tag{1.102}$$

where

$$k=\sqrt{\frac{2mE}{\hbar^2}} \tag{1.103}$$

Inserting the time dependence (see Eq. (1.99)) gives

$$\psi(x,t)=\psi(0)\exp\left[i(kx-\omega t)\right] \tag{1.104}$$

where

$$\omega = \frac{E}{\hbar} = \frac{\hbar k^2}{2m} . \tag{1.105}$$

Thus, as expected the solution in free space is a plane wave.

The Square Well

Next we consider a single electron within a square potential well as shown in Fig. 1.20. As mentioned in the discussion of the photoelectric effect, electrons within solids are bound by attractive nuclear forces.

By modeling the binding energy within a solid as a square well, we entirely ignore fine scale structure within the solid. Hence the square well, or *particle in a box*, as it is often known, is one of the crudest approximations for an electron within a solid. The simplicity of the square well approximation, however, makes it one of the most useful problems in all of quantum mechanics.

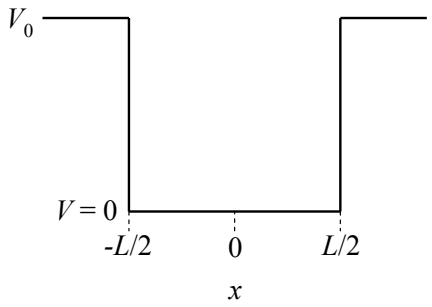


Fig. 1.20. The square well. Within the solid the potential is defined to be zero. In free space, outside the solid, the repulsive potential is V_0.

Since the potential changes abruptly, we treat each region of constant potential separately. Subsequently, we must connect up the solutions in the different regions. Tackling the problem this way is known as a *piecewise* solution.

Now, the Schrödinger Equation is a statement of the conservation of energy

$$\textit{total energy}\,(E) = \textit{kinetic energy}\,(KE) + \textit{potential energy}\,(V) \tag{1.106}$$

In classical mechanics, one can never have a negative kinetic energy. Thus, classical mechanics requires that $E > V$. This is known as the *classically allowed* region.

But in our quantum analysis, we will find solutions for $E < V$. This is known as the *classically forbidden* regime.

(i) The classically allowed region

Rearranging Eq. (1.97) gives the second order differential equation:

$$\frac{d^2\psi}{dx^2} = -\frac{2m(E-V)}{\hbar^2}\psi \tag{1.107}$$

In this region, $E > V$, solutions are of the form

$$\psi(x) = Ae^{ikx} + Be^{-ikx} \tag{1.108}$$

or

$$\psi(x) = A\sin kx + B\cos kx \tag{1.109}$$

where

$$k = \sqrt{\frac{2m(E-V)}{\hbar^2}} \tag{1.110}$$

In the classically allowed region we have oscillating solutions.

(ii) The classically forbidden region

In this region, $E < V$, and solutions are of the form

$$\psi(x) = Ae^{\alpha x} + Be^{-\alpha x} \tag{1.111}$$

i.e. they are either growing or decaying exponentials where

$$\alpha = \sqrt{\frac{2m(V-E)}{\hbar^2}}\,. \tag{1.112}$$

Matching piecewise solutions

The Schrödinger Equation is a second order differential equation. From Eq. (1.107) we observe the second derivative of the wavefunction is finite unless either E or V is infinite. Infinite energies are not physical, hence if the potential is finite we can conclude that $\frac{d\psi}{dx}$ and $\psi(x)$ are continuous everywhere.

That is, at the boundary ($x = x_0$) between piecewise solutions, we require that

$$\psi_-(x_0) = \psi_+(x_0) \tag{1.113}$$

and

$$\frac{d}{dx}\psi_-(x_0) = \frac{d}{dx}\psi_+(x_0) \tag{1.114}$$

Bound solutions

Electrons with energies within the well ($0 < E < V_0$) are *bound.* The wavefunctions of the bound electrons are localized within the well and so they must be normalizable. Thus, the wavefunction of a bound electron in the classically forbidden region (outside the well) must decay exponentially with distance from the well.

A possible solution for the bound electrons is then

$$\psi(x)=\begin{cases} Ce^{\alpha x} & \text{for } x \le -L/2 \\ A\cos(kx)+B\sin(kx) & \text{for } -L/2 \le x \le L/2 \\ De^{-\alpha x} & \text{for } x \ge L/2 \end{cases} \tag{1.115}$$

where

$$\alpha = \sqrt{\frac{2m(V_0 - E)}{\hbar^2}} \tag{1.116}$$

and

$$k = \sqrt{\frac{2mE}{\hbar^2}}, \tag{1.117}$$

and A, B, C and D are constants.

The limit that $V_0 \to \infty$ (the infinite square well)

At the walls where the potential is infinite, we see from Eq. (1.116) that the solutions decay immediately to zero since $\alpha \to \infty$. Thus, in the classically forbidden region of the infinite square well $\psi = 0$.

The wavefunction is continuous, so it must approach zero at the walls. A possible solution with zeros at the left boundary is

$$\psi(x) = A\sin\left(k\left(x + L/2\right)\right) \tag{1.118}$$

where k is chosen such that $\psi = 0$ at the right boundary also:

$$kL = n\pi \tag{1.119}$$

To normalize the wavefunction and determine A, we integrate:

$$\int_{-\infty}^{+\infty} |\psi(x)|^2 \, dx = \int_{-L/2}^{L/2} A^2 \cos^2\left(n\pi x/L + n\pi/2\right) dx \tag{1.120}$$

From which we determine that

$$A = \sqrt{2/L} \tag{1.121}$$

The energy is calculated from Eq. (1.117)

$$E = \frac{\hbar^2 k^2}{2m} = \frac{\hbar^2 \pi^2 n^2}{2mL^2}. \tag{1.122}$$

The first few energies and wavefunctions of electrons in the well are plotted in Fig. 1.21.

Two characteristics of the solutions deserve comment:

1. The bound states in the well are quantized – only certain energy levels are allowed.

2. The energy levels scale inversely with L^2. As the box gets smaller each energy level and the gaps between the allowed energy levels get larger.

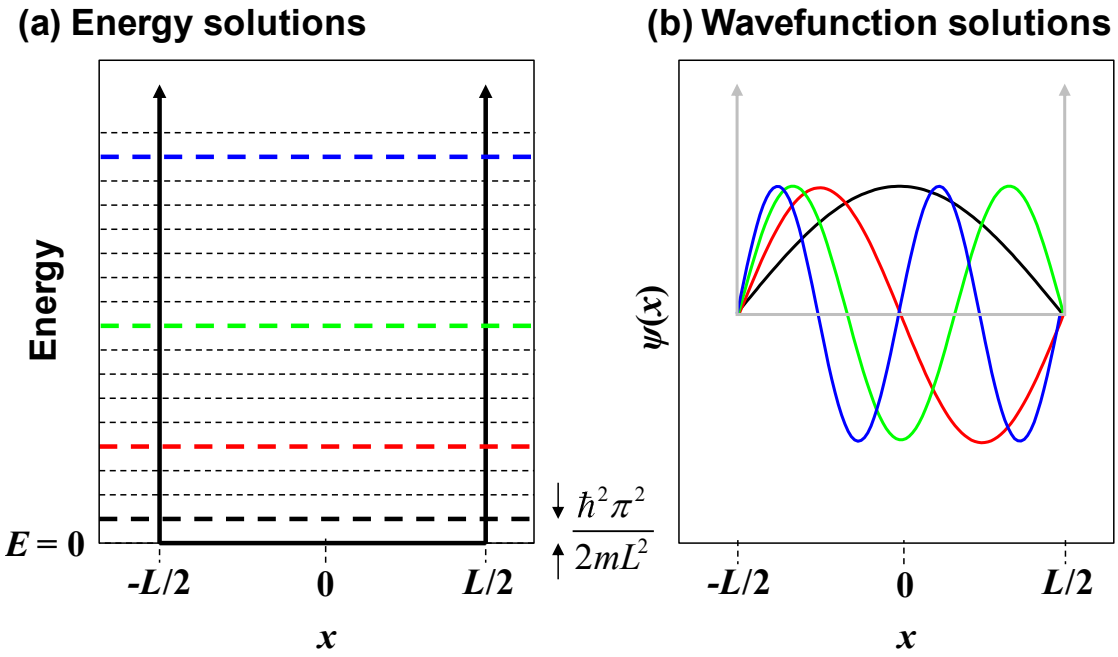


Fig. 1.21. The lowest four states for a single electron in an infinite square well. Note that we have plotted $\psi(x)$ not $|\psi(x)|^2$.

The Finite Square Well

When the confining potential is finite, we can no longer assume that the wavefunction is zero at the boundaries of the well. For a finite confining potential, the wavefunction penetrates into the barrier: the lower the confining potential, the greater the penetration.

From Eq. (1.115) the general solution for the wavefunction within the well was

$$\psi(x) = A\cos(kx) + B\sin(kx), \tag{1.123}$$

but from the solutions for the infinite well, we can see that, within the well, the wavefunction looks like

$$\psi(x) = A\cos(kx) \tag{1.124}$$

or

$$\psi(x) = A\sin(kx) \tag{1.125}$$

The simplification is possible because the well potential is symmetric around $x = 0$. Thus, the probability density $|\psi(x)|^2$ should also be symmetric,[†] and indeed both $\psi(x) = \sin(kx)$ and $\psi(x) = \cos(kx)$ have symmetric probability distributions even though $\psi(x) = \sin(kx)$ is an antisymmetric wavefunction.

We"ll consider the symmetric ($\cos(kx)$) and antisymmetric ($\sin(kx)$) wavefunction solutions separately.

[†] After all, if the potential is the same in both left and right halves of the well, there is no reason for the electron to be more probable in one side of the well.

(i) Symmetric wavefunction

We can assume a solution of the form:

$$\psi(x)=\begin{cases} e^{\alpha x} & \text{for } x \le -L/2 \\ A\cos(kx) & \text{for } -L/2 \le x \le L/2 \\ e^{-\alpha x} & \text{for } x \ge L/2 \end{cases} \tag{1.126}$$

where

$$k=\sqrt{\frac{2mE}{\hbar^2}} \tag{1.127}$$

and

$$\alpha=\sqrt{\frac{2m(V_0-E)}{\hbar^2}} \tag{1.128}$$

Note the solution as written is not normalized. We can normalize it later.

The next step is to evaluate the constant A by matching the piecewise solutions at the edge of the well. We need only consider one edge, because we have already fixed the symmetry of the solution.

At the right edge, equating the amplitude of the wavefunction gives

$$\psi(L/2)=A\cos(kL/2)=\exp[-\alpha L/2] \tag{1.129}$$

Equating the slope of the wavefunction gives

$$\psi'(L/2)=-kA\sin(kL/2)=-\alpha\exp[-\alpha L/2] \tag{1.130}$$

Dividing Eq. (1.130) by Eq. (1.129) to eliminate A gives

$$\tan(kL/2)=\alpha/k\,. \tag{1.131}$$

But α and k are both functions of energy

$$\tan\left(\frac{\pi}{2}\sqrt{\frac{E}{E_L}}\right)=\sqrt{\frac{V_0-E}{E}} \tag{1.132}$$

where we have defined the infinite square well ground state energy

$$E_L=\frac{\hbar^2\pi^2}{2mL^2}\,. \tag{1.133}$$

As in the infinite square well case, we find that only certain, discrete, values of energy give a solution. Once again, the energies of electron states in the well are quantized. To obtain the energies we need to solve Eq. (1.132). Unfortunately, this is a transcendental equation, and must be solved numerically or graphically. We plot the solutions in Fig. 1.22.

(ii) Antisymmetric wavefunction

Antisymmetric solutions are found in a similar manner to the symmetric solutions. We first assume an antisymmetric solution of the form:

$$\psi(x)=\begin{cases} e^{\alpha x} & \text{for } x \le -L/2 \\ A\sin(kx) & \text{for } -L/2 \le x \le L/2 \\ -e^{-\alpha x} & \text{for } x \ge L/2 \end{cases} \tag{1.134}$$

Then, we evaluate the constant A by matching the piecewise solutions at the edge of the well. Again, we need only consider one edge, because we have already fixed the symmetry of the solution.

At the right edge, equating the amplitude of the wavefunction gives

$$\psi(L/2)=A\sin(kL/2)=-\exp[-\alpha L/2] \tag{1.135}$$

Equating the slope of the wavefunction gives

$$\psi'(L/2)=kA\cos(kL/2)=\alpha\exp[-\alpha L/2] \tag{1.136}$$

Dividing Eq. (1.130) by Eq. (1.129) to eliminate A gives

$$\cot(kL/2)=-\alpha/k\,. \tag{1.137}$$

Expanding α and k in terms of energy

$$-\cot\left(\frac{\pi}{2}\sqrt{\frac{E}{E_L}}\right)=\sqrt{\frac{V_0-E}{E}}\,. \tag{1.138}$$

In Fig. 1.22, we solve for the energy. Note that there is always at least one bound solution no matter how shallow the well. In Fig. 1.23 we plot the solutions for a confining potential $V_0 = 5.E_L$.

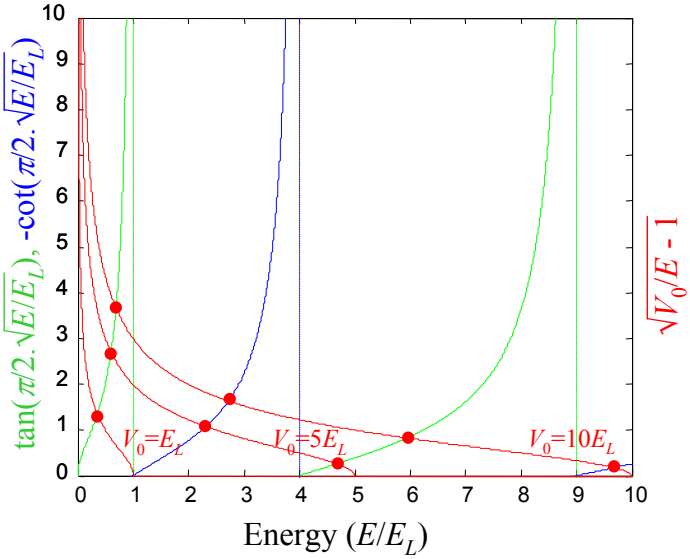


Fig. 1.22. A graphical solution for the energy in the finite quantum well. The green and blue curves are the LHS of Eqns. (1.132) and (1.138), respectively. The red curves are the RHS for different values of the confining potential V_0. Solutions correspond to the intersections between the red lines and the green or blue curves.

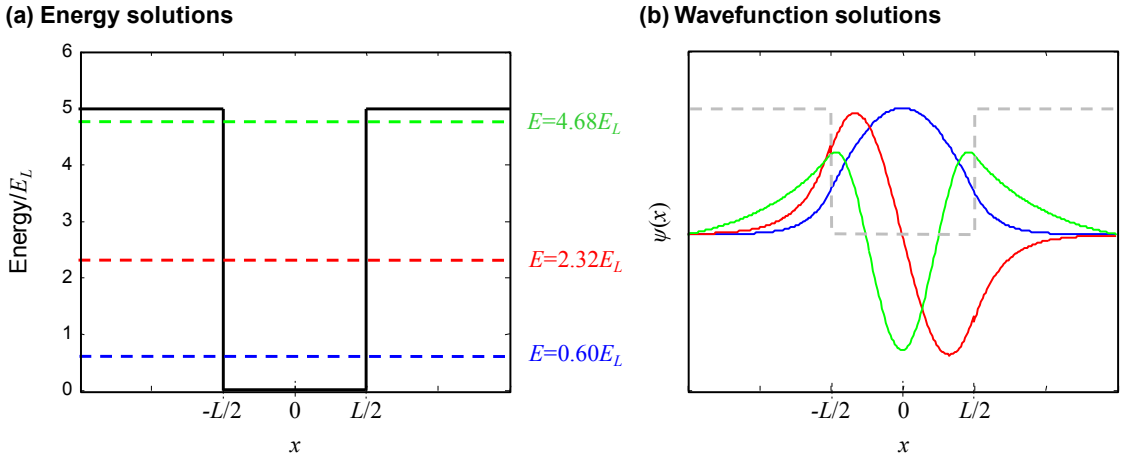


Fig. 1.23. The three bound states for electrons in a well with confining potential $V_0 = 5.E_L$. Note that the higher the energy, the lower the effective confining potential, and the greater the penetration into the barriers.

Potential barriers and Tunneling

Next we consider electrons incident on a potential barrier, as shown in Fig. 1.24.

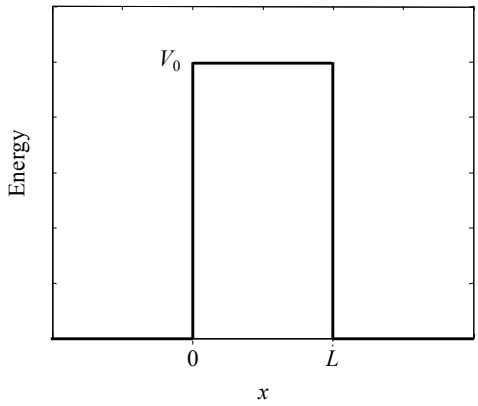


Fig. 1.24. A potential barrier.

We will assume that the particle is incident on the barrier from the left. It has some probability of being reflected by the barrier. But it also has some probability of being transmitted even though its energy may be less than the barrier height. Transmission through a barrier is known as *tunneling*. There is no equivalent process in classical physics – the electron would need sufficient energy to jump over the barrier.

Once again, we solve the time-independent Schrödinger Equation. To the left and right of the barrier, the electron is in a classically allowed region. We model the electron in these regions by a plane wave; see Eq. (1.108) and the associated discussion. On the other hand, within the barrier, if the energy, E, of the electron is below the barrier potential, V_0,

the barrier is a classically forbidden region. The solution in this region is described by expanding and decaying exponentials; see Eq. (1.111) and the associated discussion.

Analyzing the potential piece by piece, we assume a solution of the form

$$\psi(x)=\begin{cases} e^{ikx}+re^{-ikx} & \text{for } x\le 0 \\ ae^{\alpha x}+be^{-\alpha x} & \text{for } 0\le x\le L \\ te^{ikx} & \text{for } x\ge L \end{cases} \tag{1.139}$$

where once again

$$k=\sqrt{\frac{2mE}{\hbar^2}} \tag{1.140}$$

and

$$\alpha=\sqrt{\frac{2m(V_0-E)}{\hbar^2}} \tag{1.141}$$

The intensity of the incoming plane wave is unity. Hence the amplitude of the reflected wave, r, is the reflection coefficient and the amplitude of the transmitted wave, t, is the transmission coefficient. (The reflectivity and transmissivity is $|r|^2$ and $|t|^2$, respectively).

Next we match the piecewise solutions at the left edge of the barrier. Equating the amplitude of the wavefunction gives

$$\psi(0)=1+r=a+b \tag{1.142}$$

Equating the slope of the wavefunction gives

$$\psi'(0)=ik-ikr=\alpha a-\alpha b\,. \tag{1.143}$$

At the right edge of the barrier, we have

$$\psi(L)=ae^{\alpha L}+be^{-\alpha L}=te^{ikL} \tag{1.144}$$

and

$$\psi'(L)=a\alpha e^{\alpha L}-b\alpha e^{-\alpha L}=ikte^{ikL}\,. \tag{1.145}$$

Thus, we have four simultaneous equations. But these are a pain to solve analytically. In Fig. 1.25 we plot solutions for energy much less than the barrier, and energy close to the barrier. It is observed that the tunneling probability is greatly enhanced when the incident electron has energy close to the barrier height. Note that the wavefunction decay within the barrier is much shallower when the energy of the electron is large. Note also that the reflection from the barrier interferes with the incident electron creating an interference pattern.

When the electron energy is much less than the barrier height we can model the wavefunction within the barrier as simply a decaying exponential. The transmission probability is then approximately

$$T\approx\exp\left[-2\alpha L\right]. \tag{1.146}$$

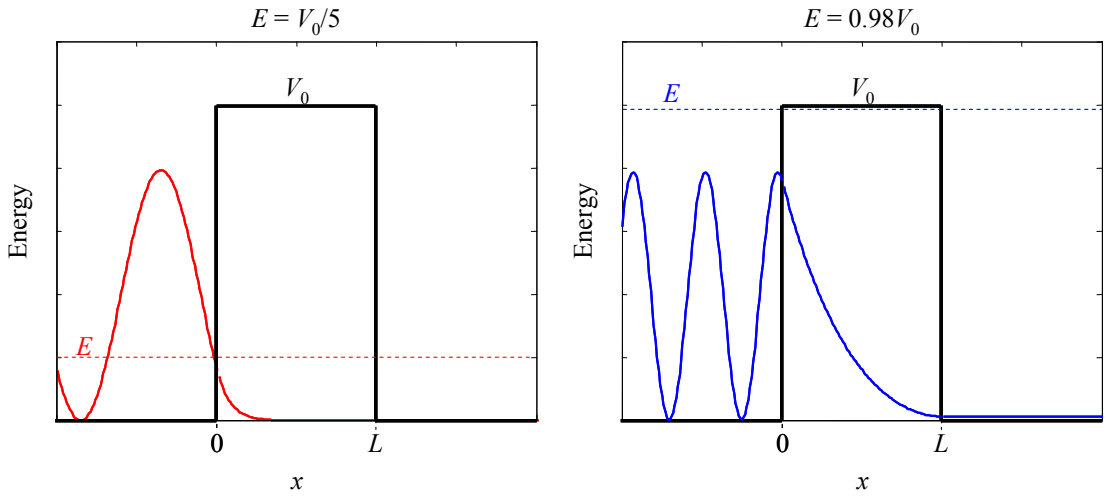


Fig. 1.25. Plots of the wavefunction for an electron incident from the left. **(a)** When the electron energy is substantially below the barrier height, tunneling is negligible. **(b)** For an electron energy 98% of the barrier height, however, note the non-zero transmission probability to the right of the barrier.

Problems

1. Suppose we fire electrons through a single slit with width *d*. At the viewing screen behind the aperture, the electrons will form a pattern. Derive and sketch the expression for the intensity at the viewing screen. State all necessary assumptions.

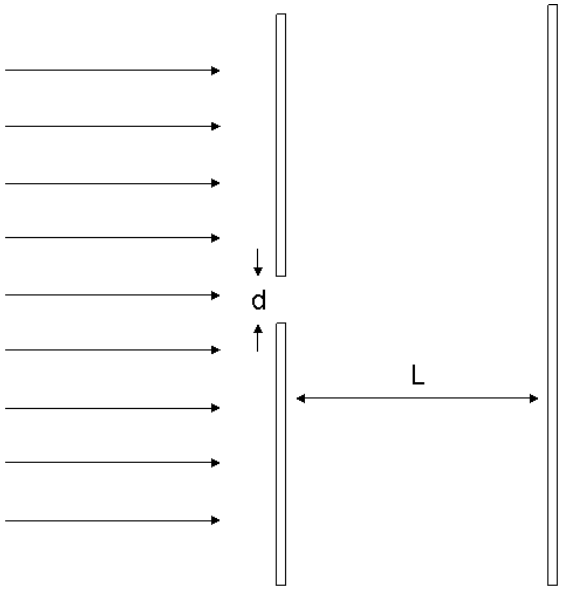


Fig. 1.26. The geometry of a single slit diffraction experiment.

How does this pattern differ from the pattern with two slits discussed in class?

2. Compare the different patterns at the viewing screen with d = 20Å and L = 100nm for fired electrons with wavelengths of 10Å, 100Å, and 1000Å. Explain.

3. Show that a shift in the position of a wavepacket by x_0 is equivalent to multiplying the k-space representation by $\exp[-ikx_0]$. Also, show that a shift in the k-space representation by k_0 is equivalent to multiplying the position of the wavepacket by $\exp[ik_0x]$. Show that similar relations hold for shifts in time and frequency.

4. Find $|F(\omega)|^2$ where $F(\omega)$ is the Fourier Transform of an exponential decay:

$$F(\omega) = \mathcal{F}[e^{-at}u(t)]$$

where $u(t)$ is the unit step function.

5. Show that if

$$(\Delta x)^2 = \langle x^2 \rangle - \langle x \rangle^2$$

then

$$\Delta x = \sigma$$

where

$$\psi(x) = \exp\left[-\frac{1}{4}\frac{\left(x - x_0\right)^2}{\sigma^2} \right]$$

6. Show the following:

(a) $\langle k \rangle = \dfrac{\langle A | k | A \rangle}{\langle A | A \rangle} = \dfrac{\langle \psi | -i\dfrac{d}{dx} | \psi \rangle}{\langle \psi | \psi \rangle}$

(b) $\langle \omega \rangle = \dfrac{\langle A | \omega | A \rangle}{\langle A | A \rangle} = \dfrac{\langle \psi | i\dfrac{d}{dt} | \psi \rangle}{\langle \psi | \psi \rangle}$

7. A free particle is confined to move along the x-axis. At time t=0 the wave function is given by

$$\psi(x, t=0) = \begin{cases} \dfrac{1}{\sqrt{L}} e^{ik_0 x} & \dfrac{-L}{2} \le x \le \dfrac{L}{2} \\ 0 & \textit{otherwise} \end{cases}$$

(a) What is the *most* probable value of momentum?

(b) What are the *least* probable *values* of momentum?

(c) Make a rough sketch of the wave function in *k*-space, A(*k, t*=0).

8. The commutator of two operators $\hat{A}$ and $\hat{B}$ is defined as

$$\left[\hat{A}, \hat{B} \right] = \hat{A}\hat{B} - \hat{B}\hat{A}$$

Evaluate the following commutators:

(a) $\left[\hat{x},\hat{x}^2\right]$ (b) $\left[\hat{p},\hat{p}^2\right]$ (c) $\left[\hat{x}^2,\hat{p}^2\right]$ (d) $[\hat{x}\hat{p},\hat{p}\hat{x}]$

9. Consider the wavefunction

$$\psi(x)=\exp\left[-\frac{1}{2}\frac{x^2}{\sigma^2}(1+iC)\right]$$

Show that

$$|\sigma_x||\sigma_k|\geq\frac{1}{2}\sqrt{1+C^2}$$

(Hint: Show that $\left|\sigma_k{}^2\right|=\dfrac{\langle\psi|-\dfrac{d^2}{dx^2}|\psi\rangle}{\langle\psi|\psi\rangle}$)

10. For the finite square well shown below, calculate the reflection and transmission coefficients ($E > 0$).

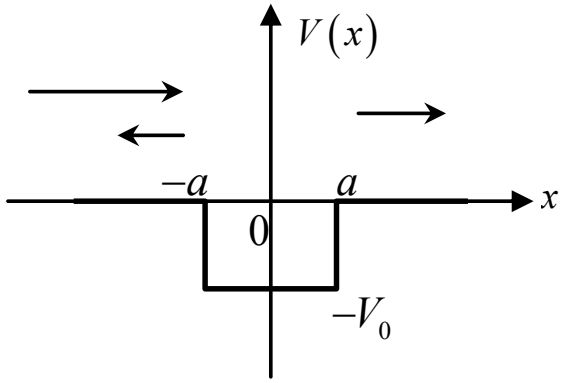


Fig. 1.27. A finite potential well.

11. Consider an electron in the ground state of an infinite square well of width L. What is the expectation value of its velocity? What is the expectation value of its kinetic energy? Is there a conflict between your results?

12. Derive the reflection and transmission coefficients for the potential $V(x) = A\delta(x)$, where $A>0$.

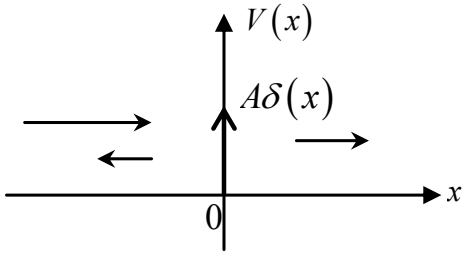


Fig. 1.28. A delta function potential.

One method to solve is by taking the Schrödinger equation with $V(x) = A\delta(x)$

$$-\frac{\hbar^2}{2m}\frac{\partial^2\psi(x)}{\partial x^2} + A\delta(x)\psi(x) = E\psi(x)$$

and integrating both sides from $-\varepsilon$ to ε for ε very small to get the constraint

$$-\frac{\hbar^2}{2m}\int_{-\varepsilon}^{+\varepsilon}\frac{\partial^2\psi(x)}{\partial x^2}dx + A\psi(0) \simeq 2\varepsilon E\psi(0) \simeq 0$$

$$\Rightarrow \left.\frac{d\psi}{dx}\right|_{x=\varepsilon} - \left.\frac{d\psi}{dx}\right|_{x=-\varepsilon} = \frac{2mA}{\hbar^2}\psi(0)$$

13. Consider two quantum wells each with width w separated by distance d.

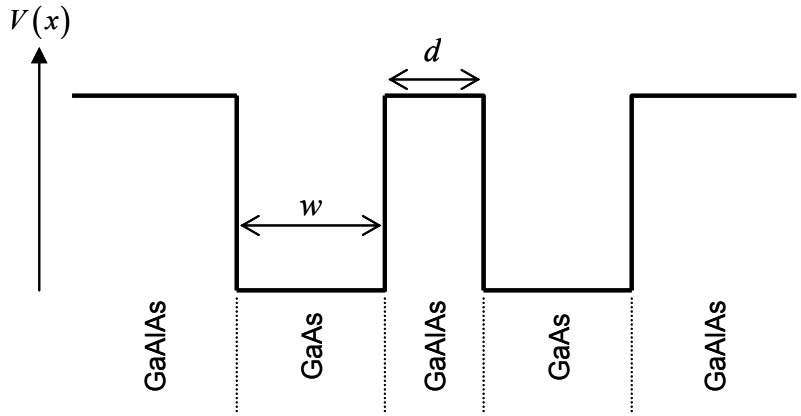


Fig. 1.29. The structure of the quantum wells.

(a) From your understanding of wavefunctions for a particle in a box, plot the approximate probability density for each of the two lowest energy modes for this system when the two quantum wells are isolated from each other.

(b) Plot the two lowest energy modes when the two quantum wells are brought close to each other such that $d<<w$?

(c) Next we wish to take a thin slice of another material and insert it in the structure above (where $d \ll w$) to kill one of the two modes of part **(b)** but leave the other unaffected. How would you choose the material (compared to the materials already present in the structure), and where should this new material be placed?

Part 2. The Quantum Particle in a Box

The goal of this class is to calculate the behavior of electronic materials and devices. In Part 1, we have developed techniques for calculating the energy levels of electron states. In this part, we will learn the principles of electron statistics and fill the states with electrons. Then in the following sections, we will apply a voltage to set the electrons in motion, and examine non-equilibrium properties like current flow, and the operation of electronic switches.

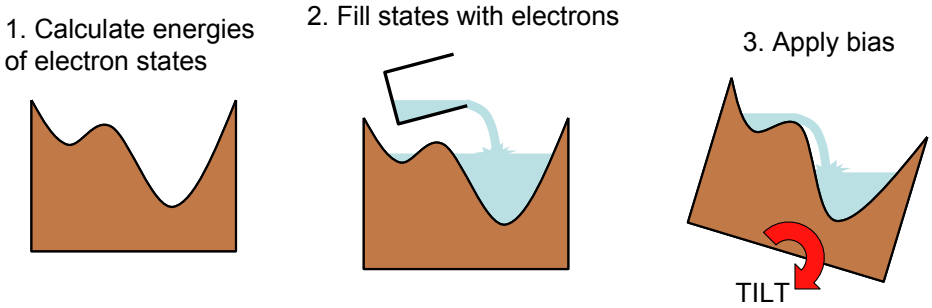


Fig. 2.1. The next two parts in diagram form.

How many electrons? Fermi-Dirac Statistics

In the previous section, we determined the allowed energy levels of a particle in a quantum well. Each energy level and its associated wavefunction is known as a „state". The **Pauli exclusion principle** forbids multiple identical electrons from occupying the same state simultaneously. Thus, one might expect that each state in the conductor can possess only a single electron. But electrons also possess *spin*, a purely quantum mechanical characteristic. For any given orientation, the spin of an electron may be measured to be +1/2 or -1/2. We refer to these electrons as spin up or spin down.

Spin up electrons are different to spin down electrons. Thus the exclusion principle allows *two* electrons per state: one spin up and one spin down.

Next, if we were to add electrons to an otherwise „empty" material, and then left the electrons alone, they would ultimately occupy their *equilibrium* distribution.

As you might imagine, at equilibrium, the lowest energy states are filled first, and then the next lowest, and so on. At T = 0K, state filling proceeds this way until there are no electrons left. Thus, at T = 0K, the distribution of electrons is given by

$$f(E,\mu) = u(\mu - E), \tag{2.1}$$

where u is the unit step function. Equation (2.1) shows that all states are filled below a characteristic energy, μ, known as the **chemical potential**. When used to describe electrons, the chemical potential is also often known as the Fermi Energy, $E_F = \mu$. Here, we will follow a convention that uses μ to symbolize the chemical potential of a contact, and E_F to describe the chemical potential of a conductor.

Equilibrium requires that the electrons have the same temperature as the material that holds them. At higher temperatures, additional thermal energy can excite some of the electrons above the chemical potential, blurring the distribution at the Fermi Energy; see Fig. 2.2.

For arbitrary temperature, the electrons are described by the Fermi Dirac distribution:

$$f(E,\mu)=\frac{1}{1+\exp\left[(E-\mu)/kT\right]} \tag{2.2}$$

Note that Eq. (2.2) reduces to Eq. (2.1) at the T = 0K limit. At the Fermi Energy, $E = E_F = \mu$, the states are half full.

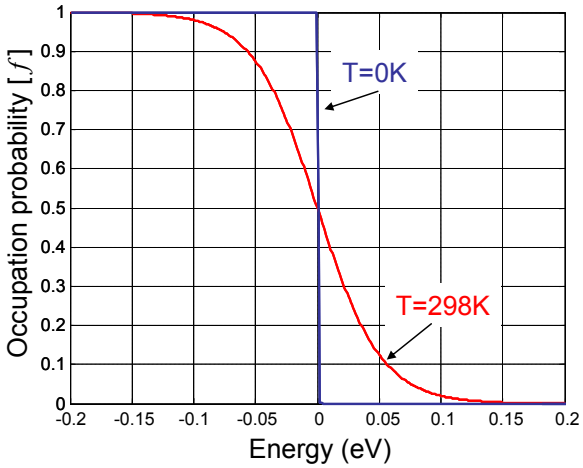


Fig. 2.2. The occupation probability for electrons at two different temperatures. The chemical potential in this example is μ = 0.

It is convenient to relate the Fermi Energy to the number of electrons. But to do so, we need to know the energy distribution of the allowed states. This is usually summarized by a function known as the density of states (DOS), which we represent by $g(E)$. There will be much more about the DOS later in this section. It is defined as the number of states in a conductor per unit energy. It is used to calculate the number of electrons in a material.

$$n=\int_{-\infty}^{+\infty} g(E)\,f(E,\mu)\,dE \tag{2.3}$$

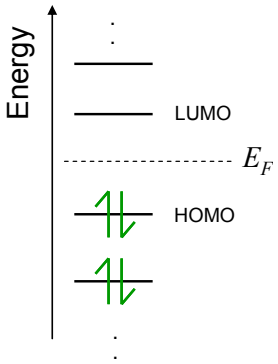


Fig. 2.3. At left, we show a possible energy structure for a molecule at T = 0K. You can think of a molecule as a little dot that confines electrons in every direction. Thus, like the 1 dimensional quantum well studied previously, the energy levels, or states, in a molecule are discrete. For a molecule each state is often described as a molecular orbital. Below the Fermi energy, each molecular orbital, contains two electrons, one spin up and one spin down. We represent these electrons with upward and downward pointing arrows. The highest occupied molecular orbital is frequently abbreviated as the HOMO. The lowest unoccupied molecular orbital is the LUMO.

Part 2. The Quantum Particle in a Box

Current

The electron distribution within a material determines its conductivity. As an example, let‟s consider some moving electrons in a Gaussian wavepacket. The wavepacket in turn can be described by the weighted superposition of plane waves. Now, we know from the previous section that, if the wavepacket is not centered on $k = 0$ in k-space, then it will move and current will flow.

There is another way to look at this. Note that there are plane wave components with both $+k_z$ and $-k_z$ wavenumbers. Thus, *even when the electron is stationary*, *components of the wavepacket are traveling in both directions*. But if each component moving in the $+k_z$ direction is balanced by an component moving in the $-k_z$ direction, there is no net current.

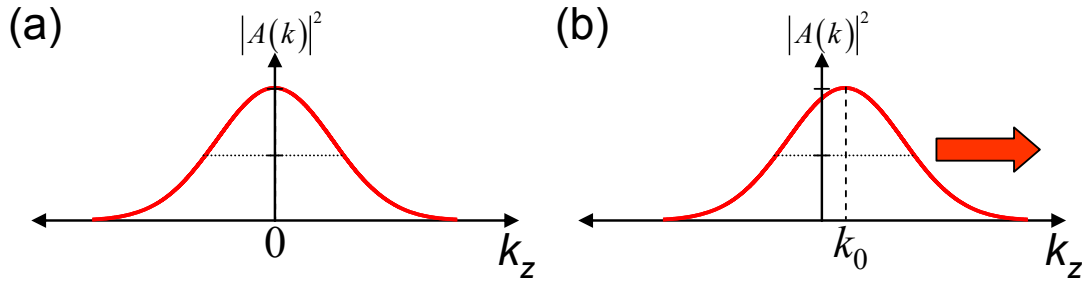


Fig. 2.4. (a) A wavepacket with no net velocity. Note that each plane wave component with a + k_z wavenumber is compensated by a plane wave component with a – k_z wavenumber. **(b)** A wavepacket with a net velocity in the positive z direction is asymmetric about $k_z = 0$.

We‟ll show in this section that we can apply a similar analysis to electrons within a conductor. For example, electrons in a wire occupy states with different wavenumbers, known as k-states. Each of these states can be modeled by a plane wave and there are states that propagate in both directions.

Recall that for a plane wave

$$E = \frac{\hbar^2 k_z^2}{2m} \tag{2.4}$$

This relation between energy and wavenumber is known as a dispersion relation. For plane waves it is a parabolic curve. Below the curve there are no electron states. Thus, the electrons reside within a certain *band* of energies.[†]

[†] Strictly a band needs an upper as well as a lower limit to the allowed energy, whereas the simple plane wave model yields only a lower limit. Later we‟ll also find upper limits in more accurate models of materials. Note also that 0-d materials such as molecules or quantum dots do not have bands because the electrons are confined in all directions and cannot be modeled by a plane wave in any direction.

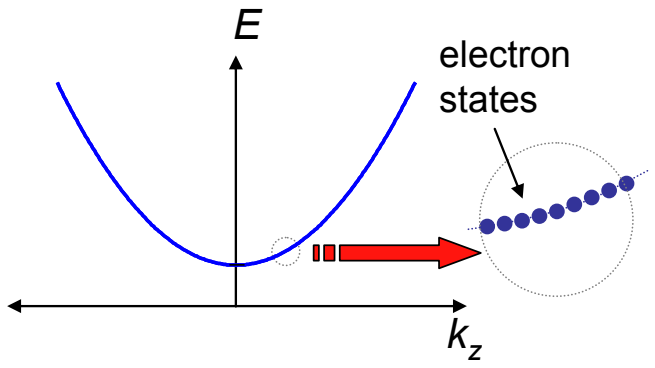


Fig. 2.5. Electrons in a wire occupy states with different energies and wavenumbers.

Under equilibrium conditions, the wire is filled with electrons up to the Fermi energy, E_F. The electrons fill both $+k_z$ and $-k_z$ states, and propagate *equally* in both directions. No current flows. We say that these electrons are compensated.

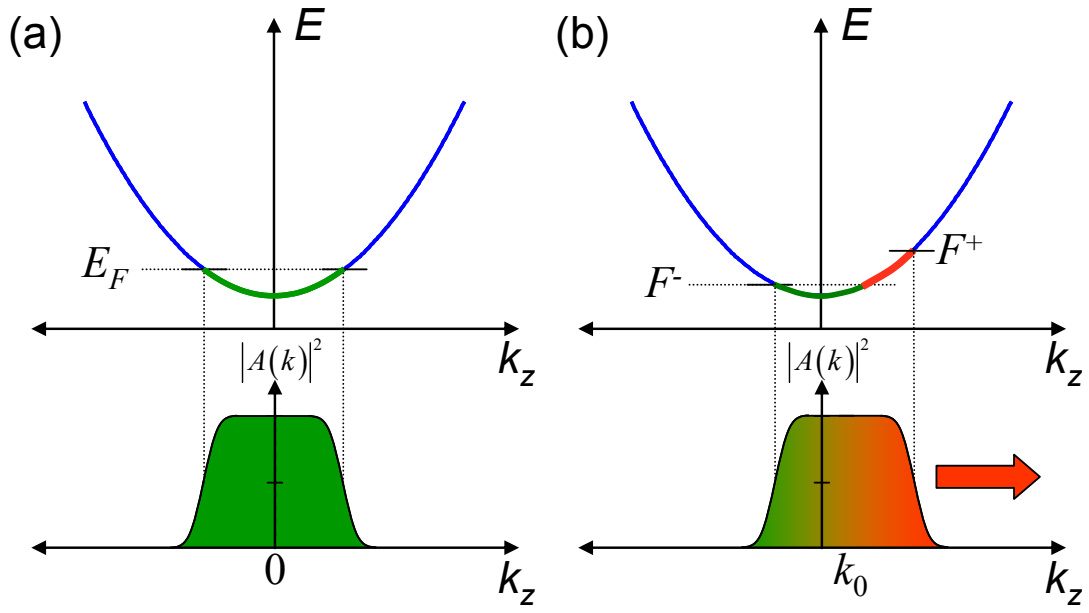
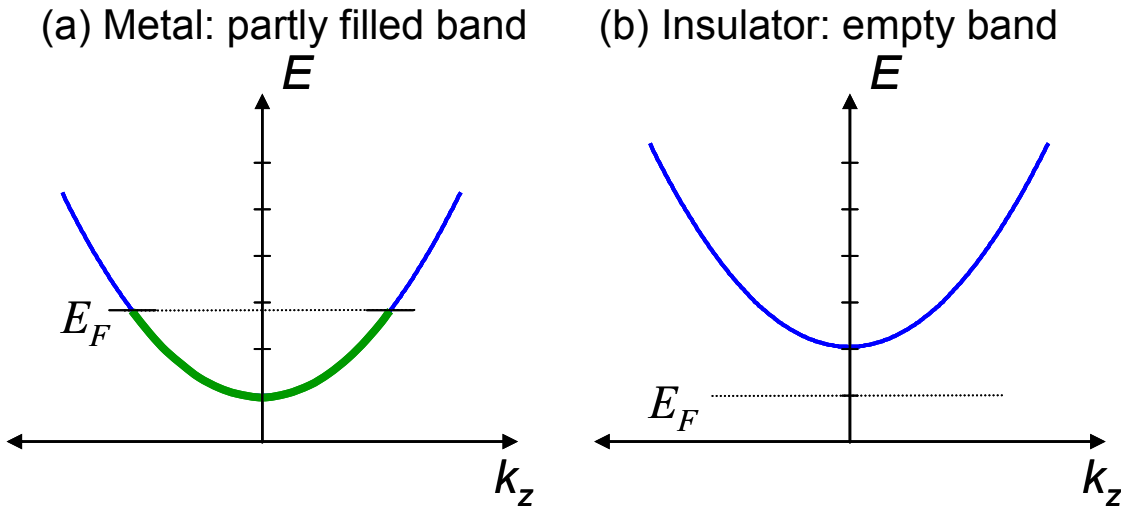


Fig. 2.6. **(a)** Under equilibrium conditions, electrons fill up the lowest energy k-states first. Since equal numbers of $+k_z$ and $-k_z$ states are filled there is no net current. **(b)** When $+k_z$ and $-k_z$ states are filled to different levels, there is a net current.

For a net current to flow there must be difference in the number of electrons moving in each direction. Thus, electrons traveling in one direction cannot be in equilibrium with electrons traveling in the other. We define two *quasi* Fermi levels: F^+ is the energy level when the states with $k_z > 0$ are half full, F^- is the corresponding energy level for states with $k_z < 0$. We can see that current flow is associated with a difference in the quasi

Fermi levels, *and* the presence of electron states between the quasi Fermi levels. If there are no electrons between F^+ and F^-, then the material is an insulator and cannot conduct



charge.

Fig. 2.7. Examples of a metal **(a)**, and an insulator **(b)**.

To summarize: current is carried by *uncompensated* electrons.

Metals and Insulators

Metals are good conductors; a small difference between F^+ and F^- yields a large difference between the number of electrons in $+k_z$ and $-k_z$ states. This is possible if the bands are partially filled at equilibrium.

If there are no electrons between F^+ and F^-, then the material is an insulator and cannot conduct charge. This occurs if the bands are completely empty or completely full at equilibrium. We have not yet encountered a band that can be completely filled. These will come later in the class.

Thus, to calculate the current in a material, we must determine the number of electrons in states that lie between the quasi Fermi levels. It is often convenient to approximate Eq. (2.2) when calculating the number of electrons in a material. There are two limiting cases:

(i) degenerate limit: $E_F - E_C \gg kT$.

As shown in Fig. 2.8 (a), here the bottom of the band, E_C, is much less than the Fermi energy, E_F, and the distribution function is modeled by a unit step:

$$f(E) = u(E_F - E) \tag{2.5}$$

(ii) non-degenerate limit: $E_C - E_F \gg kT$.

As shown in Fig. 2.8 (b), here $E_C \geq E_F$ and the distribution function reduces to the Boltzmann distribution:

$$f(E) = \exp\left[-(E - E_F)/kT\right] \tag{2.6}$$

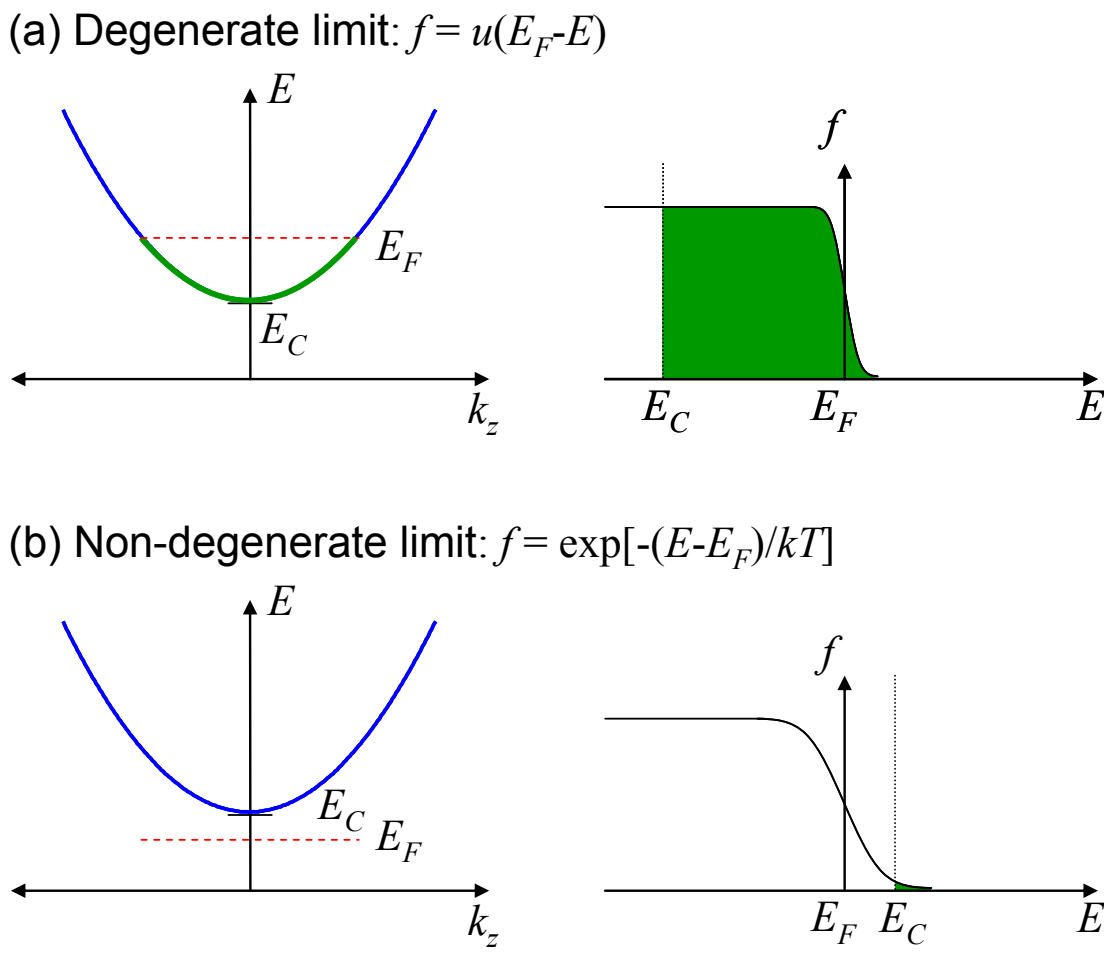


Fig. 2.8. Two limiting cases when calculating the number of electrons in a material. **(a)** If E_F is within a band, then thermal blurring of the electron distribution is not significant, and we can simply integrate up to E_F. This is the so-called degenerate case. **(b)** On the other hand, if the filled states are due solely to thermal excitation above E_F, the filling fraction falls off exponentially. This is the non-degenerate limit.

Part 2. The Quantum Particle in a Box

The Density of States

To determine whether a material is a metal or an insulator, and to calculate the magnitude of the current under applied bias, we need the density of states (DOS), which as you recall is a measure of the number of states in a conductor per unit energy. In this part, we will calculate the DOS for a variety of different conductors.

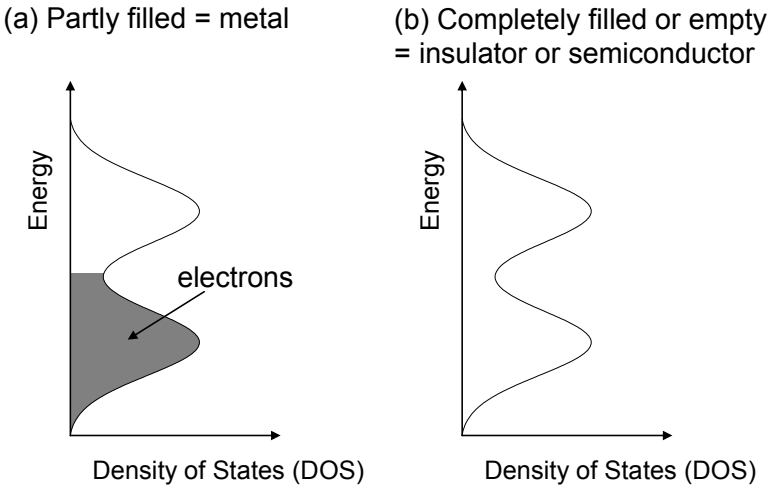


Fig. 2.9. Partly filled states give metallic behavior, while completely filled or empty states correspond to insulators or semiconductors.

To calculate the density of states we will employ two assumptions: (i) we will model the conductor as a homogeneous box, and (ii) we will assume periodic boundary conditions.

The particle in a box

Modeling our electronic material as a box allows us to ignore atoms and assume that the material is perfectly homogeneous. We will consider boxes in different dimensions: either three dimensions (typical bulk materials), 2-d (quantum wells), 1-d (quantum wires), or 0-dimensions (this is a quantum dot). The label „quantum" here refers to the confinement of electrons. When we say that an electron is „confined" in a low dimensional material we mean that critical dimensions of the material are on the order of the wavelength of an electron. We've seen that when particles are confined, their energy levels become discrete.

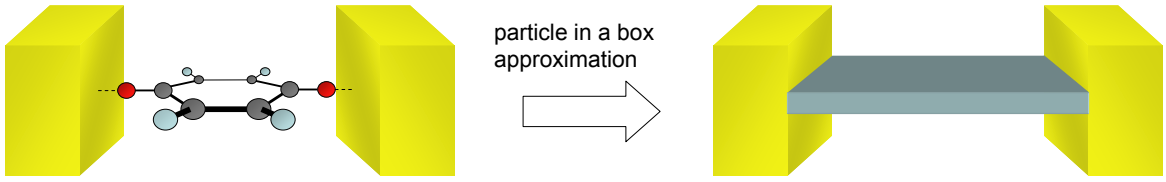


Fig. 2.10. The „particle in a box" takes a complex structure like a molecule and approximates it by a homogeneous box. All details, such as atoms, are ignored.

In quantum dots, electrons are confined in all three dimensions, in quantum wires, electrons are confined in only two dimensions and so on. So when we say that a given structure is 2-d, we mean that the electron is *unconfined* in 2 dimensions. In the unconfined directions, we will assume that the electron described by a plane wave.

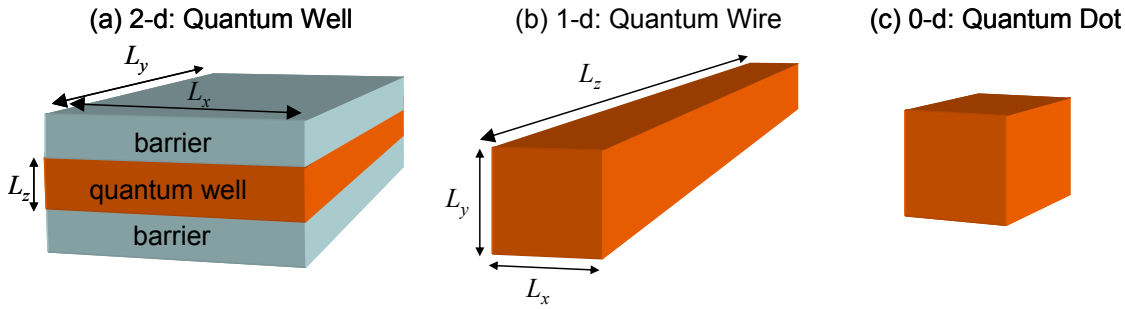


Fig. 2.11. (a) In quantum wells, electrons are confined only in one dimension. Quantum wells are usually implemented by burying the confining material within a barrier material. **(b)** Quantum wires confine electrons in two dimensions. The electron is not confined along the wire. **(c)** In a quantum dot, an electron is confined in three dimensions.

The Schrödinger Equation in Higher Dimensions

Analyzing quantum wells and bulk materials requires that we solve the Schrödinger Equation in 2-d and 3-d. The equation in 1-d

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}+V(x)\right]\psi(x)=E\psi(x) \tag{2.7}$$

is extended to higher dimensions as follows:

(i) The Kinetic Energy operator

In 1-d

$$\hat{T}=\frac{\hat{p}_x^{\,2}}{2m} \tag{2.8}$$

Now, the magnitude of the momentum in 3-d can be written

$$|\mathbf{p}|^2=p_x^2+p_y^2+p_z^2 \tag{2.9}$$

Where p_x, p_y and p_z are the components of momentum on the x, y and z axes, respectively. It follows that in 3-d

$$\hat{T}=\frac{\hat{p}_x^{\,2}}{2m}+\frac{\hat{p}_y^{\,2}}{2m}+\frac{\hat{p}_z^{\,2}}{2m}=-\frac{\hbar^2}{2m}\left(\frac{d^2}{dx^2}+\frac{d^2}{dy^2}+\frac{d^2}{dz^2}\right) \tag{2.10}$$

(ii) Separable Potential – Quantum Well

A quantum well is shown in Fig. 2.11 (a). We will assume that the potential can be separated into x, y, and z dependent terms

$$V(x,y,z)=V_x(x)+V_y(y)+V_z(z) \tag{2.11}$$

For example, a quantum well potential is given by

Part 2. The Quantum Particle in a Box

$$\begin{aligned} V_x(x) &= 0 \\ V_y(y) &= 0 \\ V_z(z) &= V_0 u(z-L) + V_0 u(-z) \end{aligned} \tag{2.12}$$

where in the infinite square well approximation $V_0 \to \infty$, and u is the unit step function.

For potentials of this form the Schrödinger Equation can be separated:

$$\begin{aligned} &\left[-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + V_x(x)\right]\psi(x,y,z) + \left[-\frac{\hbar^2}{2m}\frac{d^2}{dy^2} + V_y(y)\right]\psi(x,y,z) \\ &+\left[-\frac{\hbar^2}{2m}\frac{d^2}{dz^2} + V_z(z)\right]\psi(x,y,z) = \left(E_x + E_y + E_z\right)\psi(x,y,z) \end{aligned} \tag{2.13}$$

The wavefunction can also be separated

$$\psi(x,y,z) = \psi_x(x)\psi_y(y)\psi_z(z) \tag{2.14}$$

From Eqs. (1.118) and (1.119), the solutions to the infinite quantum well potential are

$$\psi(x,y,z) = \psi_x(x)\psi_y(y)\psi_z(z) = \sqrt{\frac{2}{L}}\sin\left(n\frac{\pi z}{L}\right).\exp\left[ik_x x\right].\exp\left[ik_y y\right] \tag{2.15}$$

with

$$E = E_x + E_y + E_z = \frac{\hbar^2 k_x^{\,2}}{2m} + \frac{\hbar^2 k_y^{\,2}}{2m} + \frac{n^2\hbar^2\pi^2}{2mL^2} \tag{2.16}$$

This dispersion relation is shown in Fig. 2.12 for the lowest three modes of the quantum well.

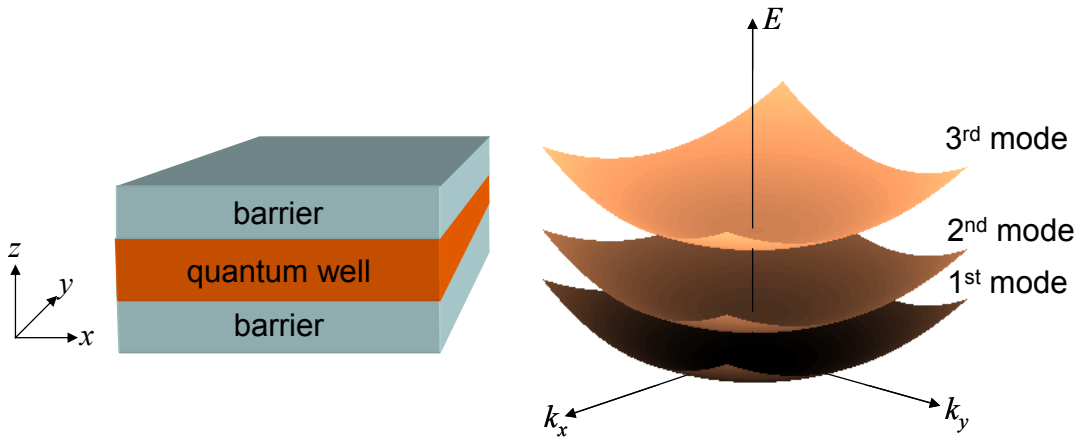


Fig. 2.12. A quantum well confines electrons in 1 dimension.

(ii) Separable Potential – Quantum Wire

A quantum wire with rectangular cross-section is shown in Fig. 2.11(b). Again, we will assume that the potential is infinite at the boundaries of the wire:

$$V(x,y,z) = V_0 u(-x) + V_0 u(x - L_x) + V_0 u(-y) + V_0 u(y - L_y), \tag{2.17}$$

where $V_0 \to \infty$. The associated wavefunction is confined in the x-y plane and composed of plane waves in the z direction, thus we chose the trial wavefunction

$$\psi(x,y,z) = \psi(x,y) e^{ik_z z} \tag{2.18}$$

Inserting Eq. (2.18) into the Schrödinger equation gives (for $0 \le x \le L_x$ and $0 \le y \le L_y$):

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x,y) - \frac{\hbar^2}{2m}\frac{d^2}{dy^2}\psi(x,y) + \frac{\hbar^2 k_z^2}{2m}\psi(x,y) = E\psi(x,y) \tag{2.19}$$

Since the potential goes to infinity at the edges of the wire, $\psi(x=0) = \psi(x=L_x) = \psi(y=0) = \psi(y=L_y) = 0$. Thus, the solution is

$$\psi(x,y) = \psi_0 \sin(k_x x)\sin(k_y y), \tag{2.20}$$

where

$$k_x = \frac{n_x \pi}{L_x}, \quad n_x = 1,2,..., \quad k_y = \frac{n_y \pi}{L_y}, \quad n_y = 1,2,... \tag{2.21}$$

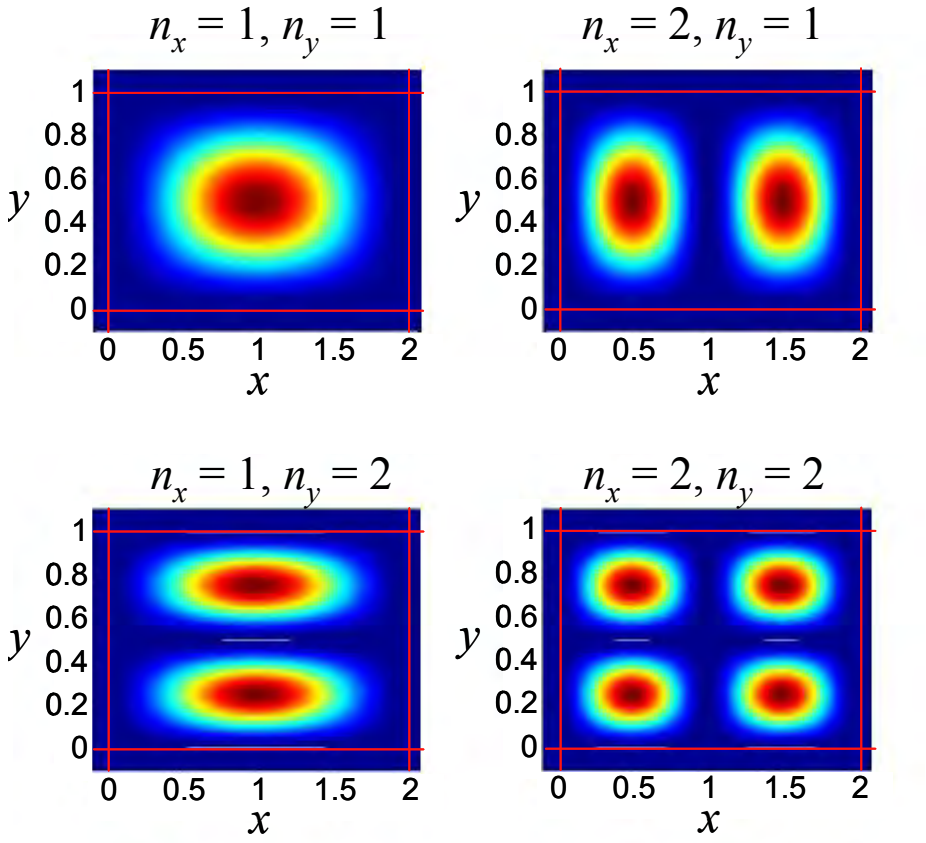


Fig. 2.13. The first four modes of the quantum wire. Since in this example, $L_x > L_y$ the $n_x = 2$, $n_y = 1$ mode has lower energy than the $n_x = 1$, $n_y = 2$ mode.

Thus, the constraint in the *x*- and *y*-directions defines the discrete energy levels

$$E_{n_x,n_y} = \frac{\hbar^2\pi^2}{2m}\left(\frac{n_x^2}{L_x^2}+\frac{n_y^2}{L_y^2}\right), \quad n_x, n_y = 1,2,... \tag{2.22}$$

The total energy is

$$E_{n_x,n_y} = \frac{\hbar^2\pi^2}{2m}\left(\frac{n_x^2}{L_x^2}+\frac{n_y^2}{L_y^2}\right)+\frac{\hbar^2 k_z^2}{2m}, \quad n_x, n_y = 1,2,... \tag{2.23}$$

This dispersion relation is plotted in Fig. 2.14 for the lowest three modes.

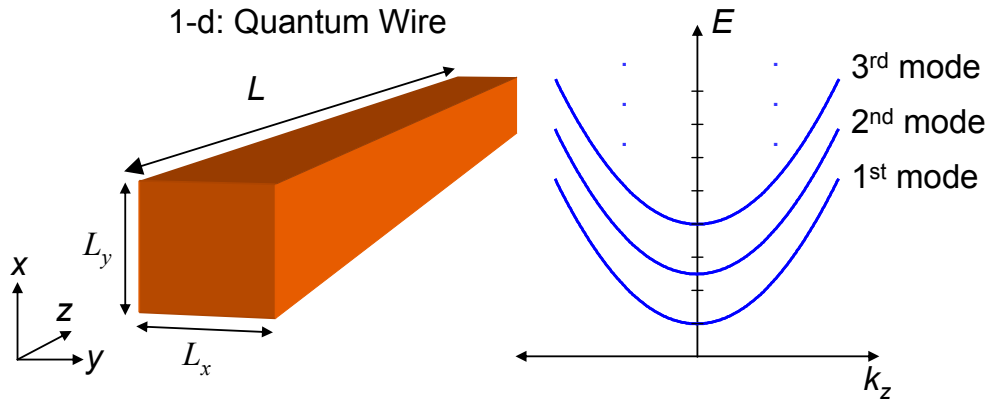


Fig. 2.14. A quantum wire confines electrons in 2 dimensions.

The 0-d DOS: single molecules and quantum dots confined in 3-d

The 0-d DOS is a special case because the particle is confined in all directions.

Like a particle in a well with discrete energy levels, we might assume that the density of states in a 0-d might be a series of delta functions at the allowed energy levels. This is indeed true for an *isolated* 0-d particle.

The lifetime of a charge in an orbital of an isolated particle is infinite. From the uncertainty principle, infinite lifetimes are associated with perfectly discrete energy states in the isolated molecule, i.e. if $\Delta t \to \infty$, $\Delta E \to 0$.

But when, for example, a molecule is brought in contact with a metal electrode, the electron may eventually escape into the metal. Now, the electron"s lifetime on the molecule is finite, and hence the molecule"s energy levels should also exhibit a finite width. Thus, molecular energy levels are broadened in a coupled metal-molecule system – the greater the coupling, the greater the broadening of the molecular energy levels.

Let"s assume that there are two electrons in the molecular orbital. Let"s assume that the lifetime of these electrons on the molecule is τ. Furthermore, let"s assume that the decay of the electron probability on the molecule is exponential. Then the time dependence of the electron probability in the molecular orbital is:

$$|\psi(t)|^2 = \frac{2}{\tau}\exp\left[-\frac{t}{\tau}\right]u(t) \tag{2.24}$$

One choice for the corresponding wavefunction is

$$\psi(t) = \sqrt{\frac{2}{\tau}}\exp\left[-i\frac{E_0 t}{\hbar} - \frac{t}{2\tau}\right]u(t) \tag{2.25}$$

We have included a complex term in the exponent to account for the phase rotation of the electron. The characteristic angular frequency of the phase rotation is $E_0/\hbar$, where E_0 is the energy of the molecular orbital in the isolated molecule.

Next we use a Fourier transform to convert from the time to the energy domain. The Fourier transform of $\psi(t)$ gives

$$A(\omega) = \frac{\sqrt{2/\tau}}{1/2\tau + i(E_0/\hbar - \omega)} \tag{2.26}$$

Now from the definition of normalization in angular frequency we require that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|A(\omega)|^2\,d\omega = \int_{-\infty}^{\infty} g(E)\,dE \tag{2.27}$$

where $g(E)$ is the density of states. Thus,

$$g(E) = \frac{1}{2\pi}|A(\omega)|^2\frac{d\omega}{dE} = \frac{2}{\pi}\frac{\hbar/2\tau}{(E-E_0)^2 + (\hbar/2\tau)^2} \tag{2.28}$$

This function is known as a Lorentzian – it is characteristic of an exponential decay in the time domain; see Fig. 2.15. The width of the Lorentzian at half its peak value (full width at half maximum, or FWHM) is $\hbar/\tau$. The coupling between the molecule and the contact can be described either by the lifetime of the electron on the molecule τ, or a coupling energy, Γ, equal to the FWHM of the Lorentzian density of states and given by

$$\Gamma = \hbar/\tau \tag{2.29}$$

Thus, as expected, a stronger interaction between a molecule and a contact is correlated with a shorter electron lifetime on the molecule and a broader molecular density of states. In well coupled systems Γ may approach 1 eV, corresponding to an electron lifetime of $\tau \sim 4$fs.

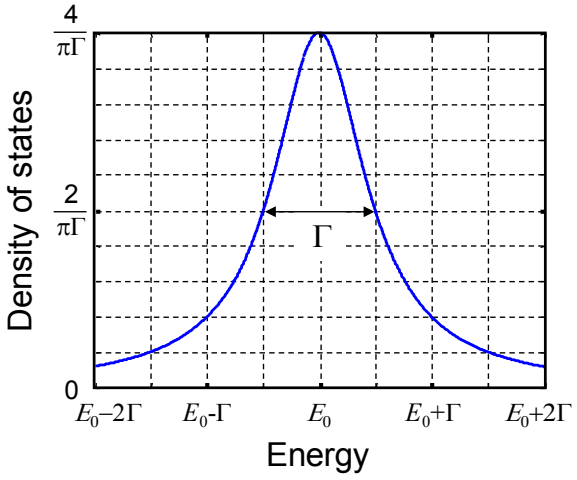


Fig. 2.15. The density of states in a molecule is broadened when the molecule is brought in contact with a metal. The effect is significant if the molecule is small, and the lifetime of electrons on the molecule changes dramatically in the presence of the metal.

Part 2. The Quantum Particle in a Box

Periodic boundary conditions

Usually, our material does not exist in isolation, but rather it may be connected to contacts, for example. So we have a problem: When determining its energy structure, how do we treat the boundaries between our material and the rest of the physical world?

First of all, if the material is big enough (for example, if a quantum wire is long enough), the boundaries will not significantly affect the electron states in the majority of the material. If this is true, we can choose any *boundary conditions* that are convenient. Indeed, we shall continue for the moment assuming that we can choose convenient boundary conditions. But note that in nanoscale devices, boundary conditions can be problematic; see the previous discussion on 0-d materials coupled to contacts.

When boundaries do not dominate the properties of the material, the usual choice is *periodic boundary conditions*. As shown in Fig. 2.16, to apply periodic boundary conditions, we take the wavefunction of the material, and make infinite copies in the unconfined directions. This gives us the ability to analyze electrons traveling in those directions in the material. After all, if just a single, isolated copy of the material was studied, the material could not support traveling electrons, only standing waves.

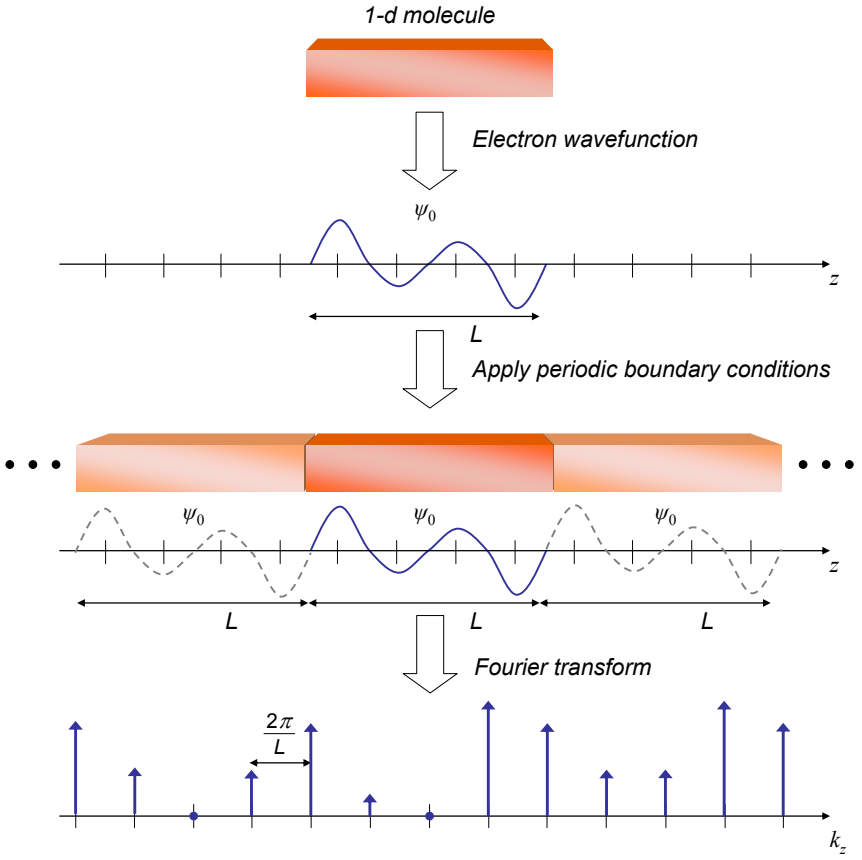


Fig. 2.16. Given periodic boundary conditions, only certain *k* values are allowed in the „unconfined" direction of a quantum wire.

But forcing periodicity in real space affects the Fourier transform of the wavefunction. In *k*-space, the periodic wavefunction is discrete. For example, on the long axis of a quantum wire of length *L*, the allowed *k*-values are spaced by

$$\Delta k = \frac{2\pi}{L} \tag{2.30}$$

Each allowed *k*-value corresponds to a plane wave, and each allowed *k*-value corresponds to a discrete electron wavefunction with a characteristic energy. As we shall see, knowing the separation of *k*-states in *k*-space allows us to easily *count* the number of electron states in the material.

The other way to think about the limitation to certain discrete *k* values in a periodic material is to recall that any periodic structure supports modes. Consequently, there are only certain allowed wavevectors for delocalized electrons in a periodic molecule. We characterize these modes by their wavevector, *k* given by $k = 2\pi n / Na_0 = 2\pi n / L$, where *L* is the length of the molecule. The allowed *k* states are therefore:

$$k = 0, \pm\frac{2\pi}{L}, \pm\frac{4\pi}{L}, \pm\frac{6\pi}{L}, \ldots \tag{2.31}$$

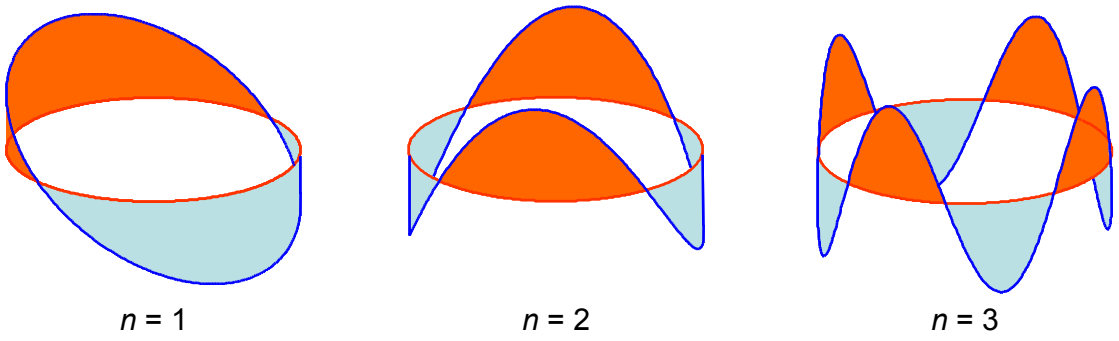


Fig. 2.17. Several modes on a ring. Because the ring is periodic, only the wavevectors, $k = 2\pi n/L$ give stable states, where *L* is the perimeter of the ring and n is an integer.

Thus, in a conductor where we have applied periodic boundary conditions, the spacing of the allowed *k* states is determined by the length of conductor.

The 1-d DOS: quantum wires confined in 2-d

The density of states, $g(E)$ is defined as the number of allowed states within energy range *dE*, i.e. the total number of states within the energy range $-\infty < E < E_F$ is

$$n_s(E_F) = \int_{-\infty}^{E_F} g(E)\,dE \tag{2.32}$$

To determine $g(E)$ we will count *k* states and then use the relation between *E* and *k* (known as the dispersion relation) to change variables from *k* to *E*.

We showed above that the energy of electrons in a quantum wire is

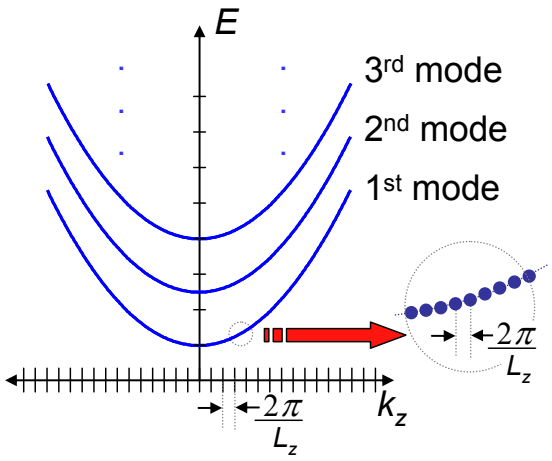


Fig. 2.18. The dispersion relation of a quantum wire after application of periodic boundary conditions.

$$E_{n_x,n_y} = \frac{\hbar^2\pi^2}{2m}\left(\frac{n_x^2}{L_x^2}+\frac{n_y^2}{L_y^2}\right)+\frac{\hbar^2 k_z^2}{2m}, \quad n_x, n_y = 1,2,... \tag{2.33}$$

Thus, counting the *x*-*y* modes is straight forward, since in the confined potential they are discrete. But to count the modes in the *z*-direction *we impose periodic boundary conditions*.

This should be OK if the wire is sufficiently long since the boundaries of the wire are then less significant. Periodic boundary conditions cause k_z to be quantized, and each allowed value of *k*-space occupies a length $2\pi/L_z$.

For convenience, we will integrate with respect to the magnitude of k_z. Since we are integrating $|k_z|$ from 0 to ∞, not $-\infty < k_z < \infty$, there is an extra factor of two to account for modes with negative k_z, and an additional factor of two to account for the two possible electron spins per *k* state.

$$n_s\left(|k_z|\right) = 2\times 2\times \int_0^{|k_z|} \frac{1}{2\pi/L_z} dk \,. \tag{2.34}$$

Next we need to change variables in Eq. (2.34), i.e. we need *g*(*E*) where

$$n_s\left(E_F\right) = \int_{-\infty}^{E_F} g\left(E\right) dE \tag{2.35}$$

Now $|k_z|$ is related to the energy by

$$E - E_{n_x,n_y} = \frac{\hbar^2 |k_z|^2}{2m}, \quad E \geq E_{n_x,n_y} \tag{2.36}$$

Using the dispersion relation of Eq. (2.36) in Eq. (2.34) gives,

$$g\left(E\right) dE = \frac{2L}{\pi}\sqrt{\frac{m}{2\hbar^2}} \sum_{n_x,n_y} \frac{u\left(E - E_{n_x,n_y}\right)}{\sqrt{E - E_{n_x,n_y}}} dE \,, \tag{2.37}$$

where u is the unit step function. The DOS is plotted in Fig. 2.19. Note that the flat region in the dispersion relation as $k \rightarrow 0$ yields infinite peaks in the DOS at the bottom of each band. The peaks have finite area, however, since the wire contains a finite number of states.

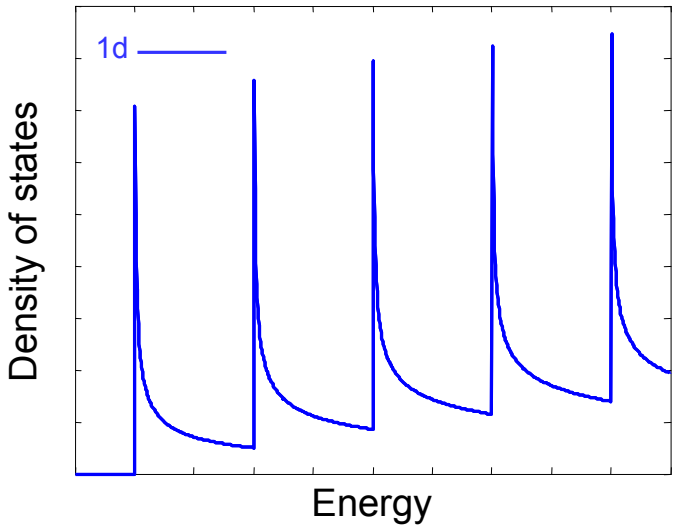


Fig. 2.19. The density of states for a quantum wire (1d).

Periodic Boundary Conditions in 2-d

Applying periodic boundary conditions to 2d materials follows the same principles as in 1d.

Let"s assume that the long axes of the quantum well are aligned with the x and y axes, and that the dimensions of the quantum well are $L_x \times L_y$. When we apply periodic boundary conditions, the infinite system is periodic on both the x-axis (period L_x) and the y-axis (period L_y).

First, let"s consider periodicity on the x-axis; see Fig. 2.20.

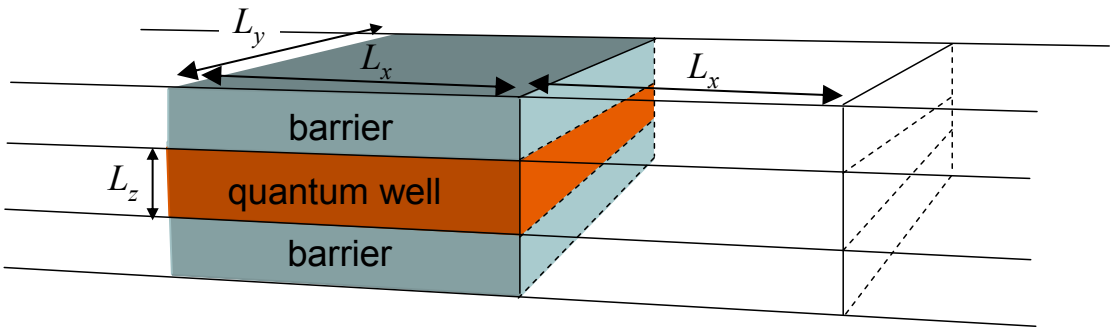


Fig. 2.20. The quantum well, assuming periodic boundary conditions on the x-axis.

Part 2. The Quantum Particle in a Box

On the x-axis the wavefunction is a plane wave.

$$\psi_x(x) = \exp[ik_x x] \tag{2.38}$$

Under periodic boundary conditions, only discrete k_x values are allowed

$$k_x = n_x \frac{2\pi}{L_x} \tag{2.39}$$

where n_x is an integer.

Similarly, for periodicity on the y-axis:

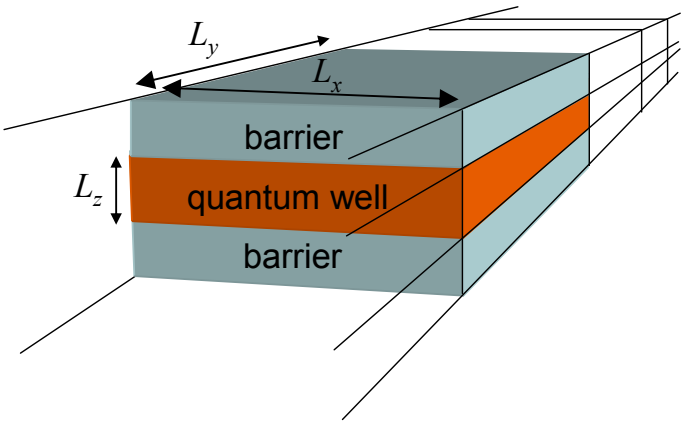


Fig. 2.21. The quantum well, assuming periodic boundary conditions on the y-axis.

the wavefunction on the y-axis

$$\psi_y(y) = \exp[ik_y y] \tag{2.40}$$

is restricted to discrete k_y values:

$$k_y = n_y \frac{2\pi}{L_y} \tag{2.41}$$

where n_y is an integer.

Thus, in k-space the allowed k-states are spaced regularly, with:

$$\Delta k_x = \frac{2\pi}{L_x}, \quad \Delta k_y = \frac{2\pi}{L_y} \tag{2.42}$$

Overall, the area occupied in k-space per k-state is:

$$\Delta k^2 = \Delta k_x \Delta k_y = \frac{2\pi}{L_x}\frac{2\pi}{L_y} = \frac{4\pi^2}{A} \tag{2.43}$$

where A is the area of the quantum well.

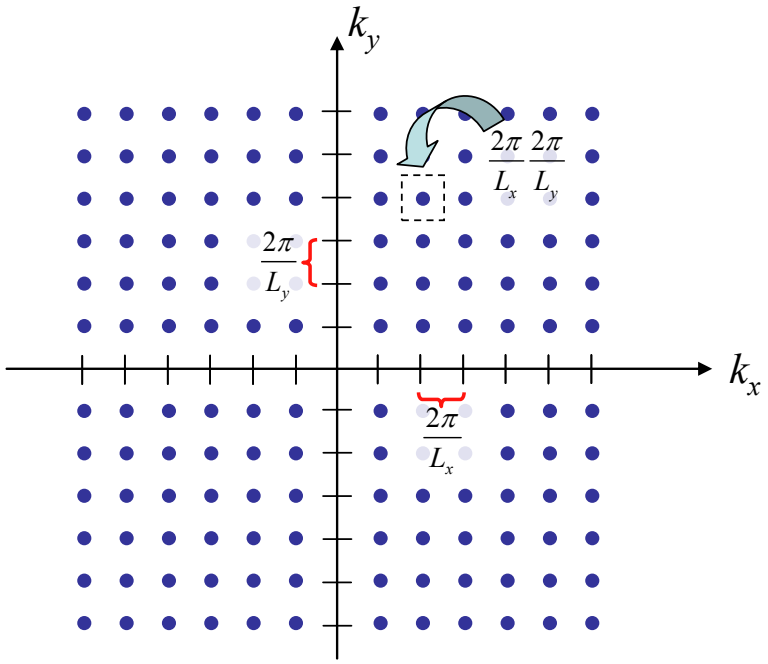


Fig. 2.22. In *k*-space only certain discrete values are allowed. Each state occupies an area of $4\pi^2/L_xL_y$.

The 2-d DOS: quantum wells confined in 1-d

We showed above that the energy of electrons in a quantum well is

$$E = \frac{\hbar^2\pi^2}{2mL_z^2}n^2 + \frac{\hbar^2\left(k_x^2 + k_y^2\right)}{2m}, \quad n = 1, 2, \ldots \tag{2.44}$$

For the DOS calculation, the specifics of the confining potential are irrelevant; we note only that the electron is unconfined in two dimensions. If the quantum well has area $L_x \times L_y$ then each allowed value of *k*-space occupies an area of $2\pi/L_x \times 2\pi/L_y$.

It is convenient to convert to cylindrical coordinates (k,ϕ,z) where *k* is the magnitude of the *k*-vector in the *x*-*y* plane. The number of states within a ring of thickness *dk* is then

$$n_s(k)dk = 2 \times \frac{1}{4\pi^2/A} \times 2\pi k dk \tag{2.45}$$

where $A = L_x \times L_y$, and again we have multiplied by two to account for the electron spin.

Now *k* is related to the energy by

$$E - E_n = \frac{\hbar^2 k^2}{2m}, \quad E \geq E_n \tag{2.46}$$

Thus, from Eq. (2.46),

$$g(E)\,dE = \frac{Am}{\pi\hbar^2}\sum_n u(E - E_n)\,dE\,, \tag{2.47}$$

where u is the unit step function. The DOS is plotted in Fig. 2.24.

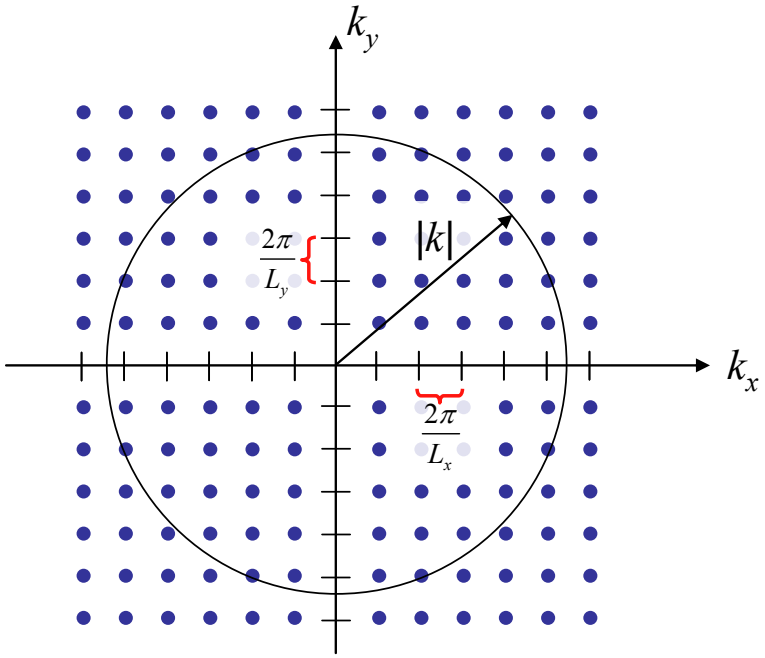


Fig. 2.23. We calculate the number of k-states within a circle of radius $|k|$.

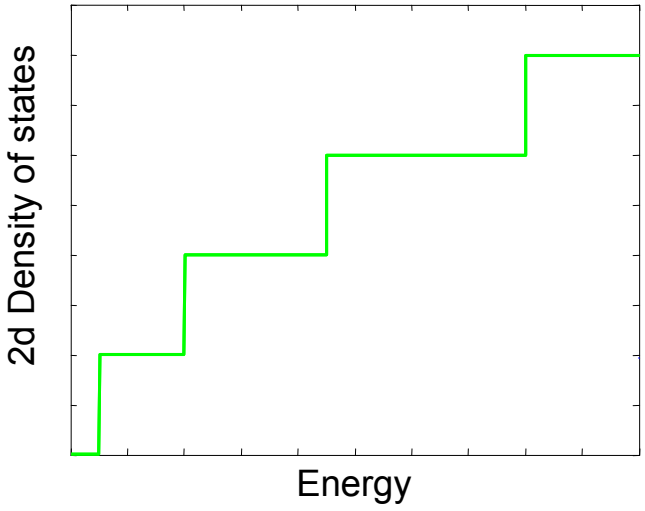


Fig. 2.24. The density of states for a quantum well (2d).

Periodic boundary conditions in 3-d

In three dimensions, again only discrete values of k are allowed. This time the *volume* of k-space per allowed state is

$$\Delta k^3 = \Delta k_x \Delta k_y \Delta k_z = \frac{2\pi}{L_x}\frac{2\pi}{L_y}\frac{2\pi}{L_z} = \frac{8\pi^3}{V} \tag{2.48}$$

where V is the volume of the material.

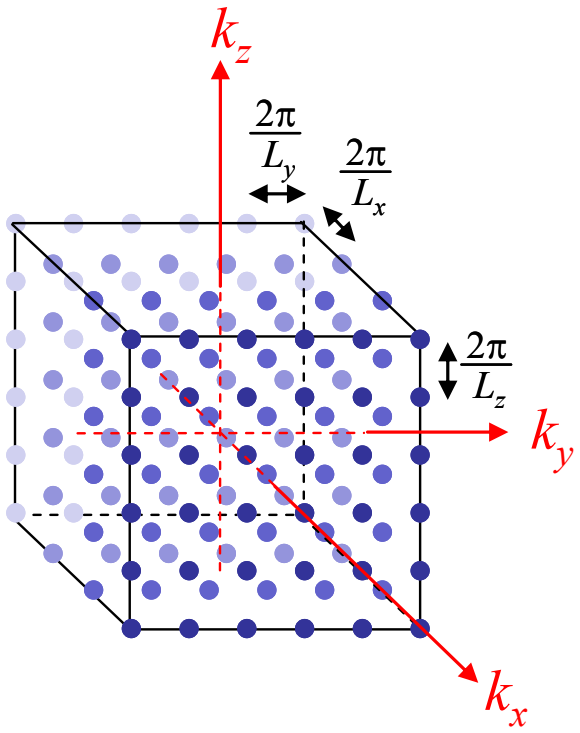


Fig. 2.25. In k-space only certain discrete values are allowed. Each state occupies a volume of $8\pi^3/L_xL_yL_z$.

In summary, the k-space occupied per state is

1-d	$\Delta k = \frac{2\pi}{L}$
2-d	$\Delta k^2 = \frac{4\pi^2}{A}$
3-d	$\Delta k^3 = \frac{8\pi^3}{V}$

Table 2.1. The k-space occupied per state in 1, 2 and 3 dimensions.

Part 2. The Quantum Particle in a Box

The 3-d DOS: bulk materials with no confinement

In 3-d, there is no electron confinement. The only constraint on k_x, k_y, or k_z are the periodic boundary conditions. We have just shown that if the system has volume $L_x \times L_y \times L_z$ then each allowed value of k-space occupies a volume of $2\pi/L_x \times 2\pi/L_y \times 2\pi/L_z = 8\pi^3/V$.

To determine the number of allowed states, we will integrate over all k-space. It is convenient to do this in spherical coordinates. If k is the magnitude of the $\boldsymbol{k}$ vector, the number of modes within a spherical shell of thickness dk is then

$$n_s(k)dk = 2 \times \frac{1}{8\pi^3/V} \times 4\pi k^2 dk . \tag{2.49}$$

where $V = L_x \times L_y \times L_z$ $V = L_x \times L_y \times L_z$, and the factor of two accounts for electron spin. The unconfined wavefunctions within our 3-d box are plane waves in all directions, i.e. the wavefunction could be described by

$$\psi(x,y,z) = \psi_0 e^{ik_x x} e^{ik_y y} e^{ik_z z} \tag{2.50}$$

Substituting into the Schrödinger Equation gives

$$-\frac{\hbar^2}{2m}\left(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}\right)\psi = E\psi \tag{2.51}$$

Which gives

$$\frac{\hbar^2}{2m}\left(k_x^2 + k_y^2 + k_z^2\right) = E \tag{2.52}$$

Rearranging:

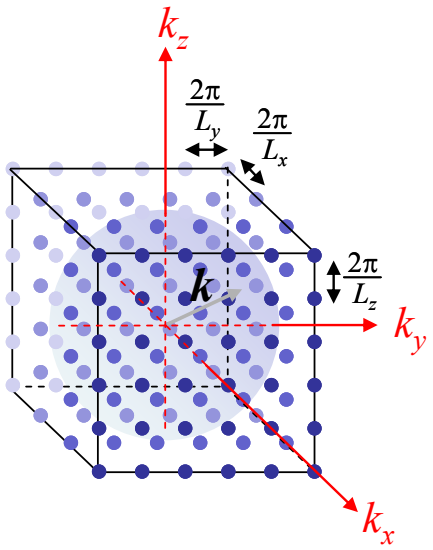


Fig. 2.26. Construction used for calculating the DOS for a 3d system.

$$E = \frac{\hbar^2 k^2}{2m} \tag{2.53}$$

Using Eq. (2.53) to relate E to k gives:

$$g(E)dE = \frac{V}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{\frac{3}{2}}\sqrt{E}dE \tag{2.54}$$

where $g(E)$ is the density of states per unit energy.

A comparison of the density of states in 1-d, 2-d and 3-d materials is shown in Fig. 2.27.

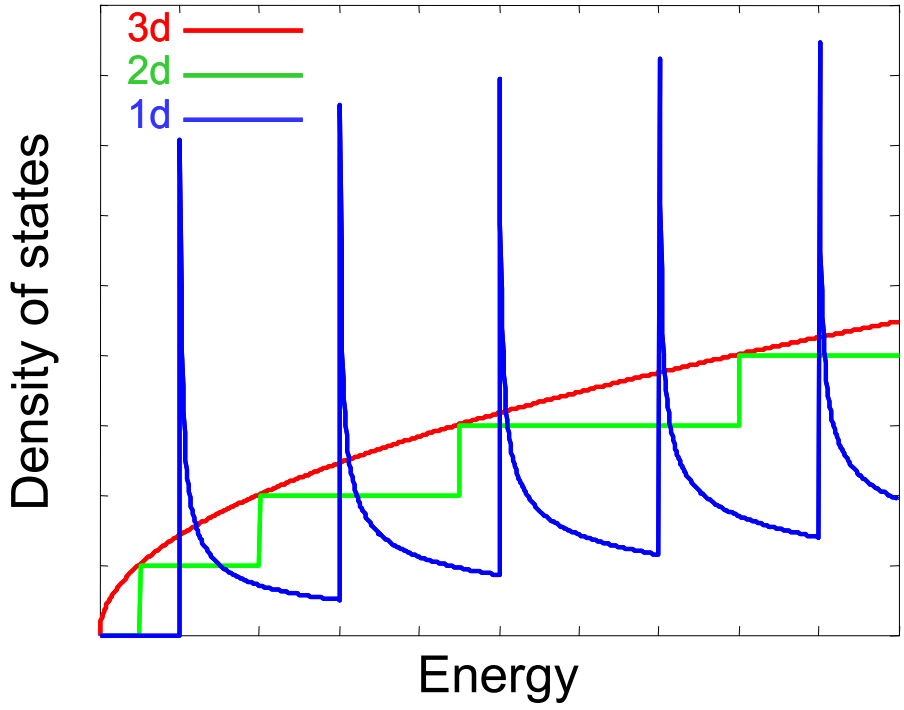


Fig. 2.27. Normalized densities of states for bulk materials (3d), quantum wells (2d), and molecular wires (1d).

Part 2. The Quantum Particle in a Box

Problems

1. i) Using MATLAB, generate Fig. 2.13.

ii) If $L_y = L_x \cdot \sqrt{\frac{3}{5}}$ which mode has a lower energy $\{n_x = 3, n_y = 1\}$ or $\{n_x = 2, n_y = 2\}$

2. The transistor illustrated below has an oxide ($\varepsilon = 4\varepsilon_0$) thickness d = 1.2nm and channel depth L_z = 2.5nm. Considering the channel as a quantum well, how many modes are filled when a voltage V=1V is applied to the gate?

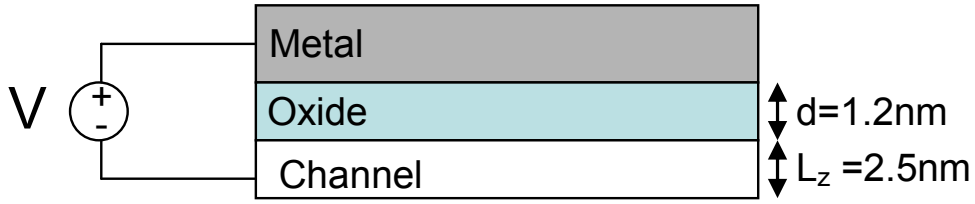


Fig. 2.28. A simple model of a modern transistor.

Hint: calculate the charge in the channel using the expression for a parallel plate capacitor.

3. A particular conductor of length L has the dispersion relation:

$$E_1(k) = 5 + 2V\cos(ka)$$
$$E_2(k) = 10 - 2V\cos(ka) \quad , \quad |k| < \frac{\pi}{a}$$

where V and a are positive constants.

i) Sketch the dispersion relation.

ii) Calculate the density of states in terms of E, V, and a.

4. In general, the degenerate approximation for the electron distribution function $f(E,E_F)$ works when the density of states is large and slowly varying above and below the Fermi level. The non degenerate approximation works best when the density of states at the Fermi level is much smaller than the density of states at higher energies.

Here, we consider these approximations in a Gaussian density of states. The Gaussian varies fairly slowly near its center, but it decreases extremely rapidly in its tails. The electron population in a Gaussian density of states is given by

$$n = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(E/\sigma)^2} f(E, E_F)\, dE \,,$$

where $f(E,E_F)$ is the Fermi function. For a particular range of the Fermi level, E_F, the electron population may be approximated as:

$$n \approx \int_{-\infty}^{E_F} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(E/\sigma)^2} dE$$

This is the degenerate limit.

As the Fermi level decreases, the electron population is better calculated in the non-degenerate limit. Derive the minimum Fermi level E_F for the degenerate limit to hold as a function of temperature T and standard deviation ζ.

Hint: estimate the minimum Fermi level by examining the energetic distribution of electrons in the non-degenerate limit.

Part 3. Two Terminal Quantum Dot Devices

In this part of the class we are going to study electronic devices. We will examine devices consisting of a quantum dot or a quantum wire conductor between two contacts. We will calculate the current in these „two terminal" devices as a function of voltage. Then we will add a third terminal, the gate, which is used to independently control the potential of the conductor. Then we can create transistors, the building-block of modern electronics. We will consider both nanotransistors and conventional transistors.

We will begin with the simplest case, a quantum dot between two contacts.

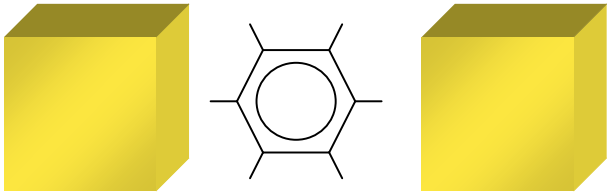


Fig. 3.1. A molecule between two contacts. We will model the molecule as a quantum dot.

Quantum Dot / Single Molecule Conductors

As we saw in Part 2, a quantum dot is a 0-d conductor; its electrons are confined in all dimensions. A good example of a quantum dot is a single molecule that is isolated in space. We can approximate our quantum dot or molecule by a square well that confines electrons in all dimensions. One consequence of this confinement is that the energy levels in the isolated quantum dot or molecule are discrete. Typically, however, the simple particle-in-a-box model does not generate sufficiently accurate estimates of the discrete energy levels in the dot. Rather, the material in the quantum dot or the structure of the molecule defines the actual energy levels.

Fig. **3.2** shows a typical square well with its energy levels. We will assume that these energy levels have already been accurately determined. Each energy level corresponds to a different molecular orbital. Energy levels of bound states within the well are measured with respect to the **Vacuum Energy**, typically defined as the potential energy of a free electron in a vacuum. Note that if an electric field is present the vacuum energy will vary with position.

Next we add electrons to the molecule. Each energy level takes two electrons, one of each spin. The **highest occupied molecular orbital** (HOMO) and the **lowest unoccupied molecular orbital** (LUMO) are particularly important. In most chemically stable materials, the HOMO is completely filled; partly filled HOMOs usually enhance the reactivity since they tend to readily accept or donate electrons.

Earlier we stated that charge transport occurs only in partly filled states. This is best achieved by adding electrons to the LUMO, or subtracting electrons from the HOMO. Modifying the electron population in all other states requires much more energy. Hence we will ignore all molecular orbitals except for the HOMO and LUMO.

Fig. **3.2** also defines the **Ionization Potential** (I_P) of a molecule as the binding energy of an electron in the HOMO. The binding energy of electrons in the LUMO is defined as the **Electron Affinity** (E_A) of the molecule.

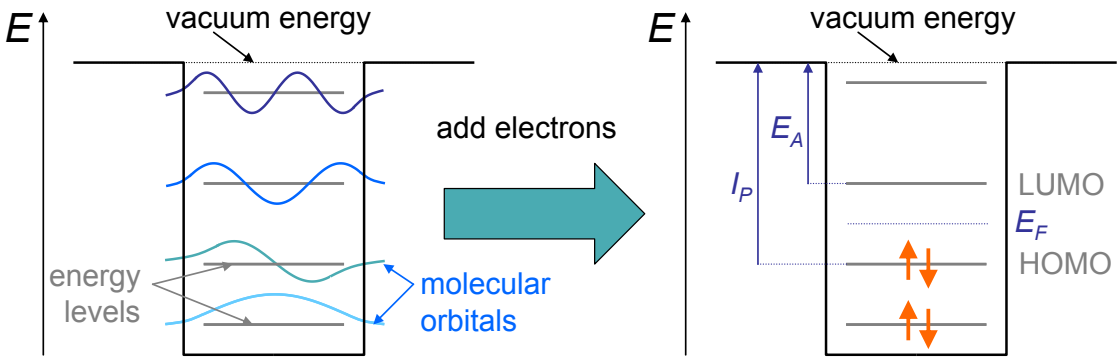


Fig. 3.2. A square well approximation of a molecule. Energy levels within the molecule are defined relative to the vacuum energy – the energy of a free electron at rest in a vacuum.

Contacts

There are three essential elements in a current-carrying device: a conductor, and at least two contacts to apply a potential across the conductor. By definition the contacts are large: each contact contains many more electrons and many more electron states than the conductor. For this reason a contact is often called a reservoir. We will assume that all electrons in a contact are in equilibrium. The energy required to promote an electron from the Fermi level in the contact to the vacuum energy is defined as the **work function** (Φ).

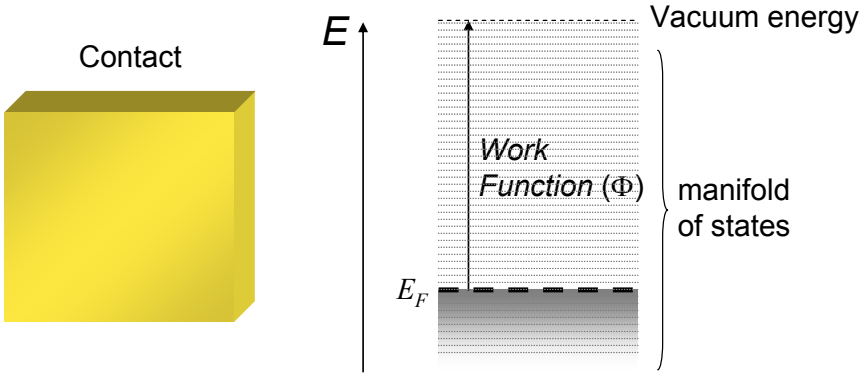


Fig. 3.3. An energy level model of a metallic contact. There are many states filled up with electrons to the Fermi energy. The minimum energy required to remove an electron from a metal is known as the work function.

Part 3. Two Terminal Quantum Dot Devices

Metals are often employed as contacts, since metals generally possess very large numbers of both filled and unfilled states, enabling good conduction properties. Although the assumption of equilibrium within the contact cannot be exactly correct if a current flows through it, the large population of mobile electrons in the contact ensures that any deviations from equilibrium are small and the potential in the contact is approximately uniform. For example, consider a large metal contact. Its resistance is very small, and consequently any voltage drop in the contact must be relatively small.

Equilibrium between contacts and the conductor

In this section we will consider the combination of a molecule and a single contact.

In the absence of a voltage source, the isolated contact and molecule are at the same potential. Thus, their vacuum energies (the potential energy of a free electron) are identical in isolation.

When the contact is connected with the molecule, equilibrium must be established in the *combined* system. To prevent current flow, there must be a uniform Fermi energy in both the contact and the molecule.

But if the Fermi energies are different in the isolated contact and molecules, how is equilibrium obtained?

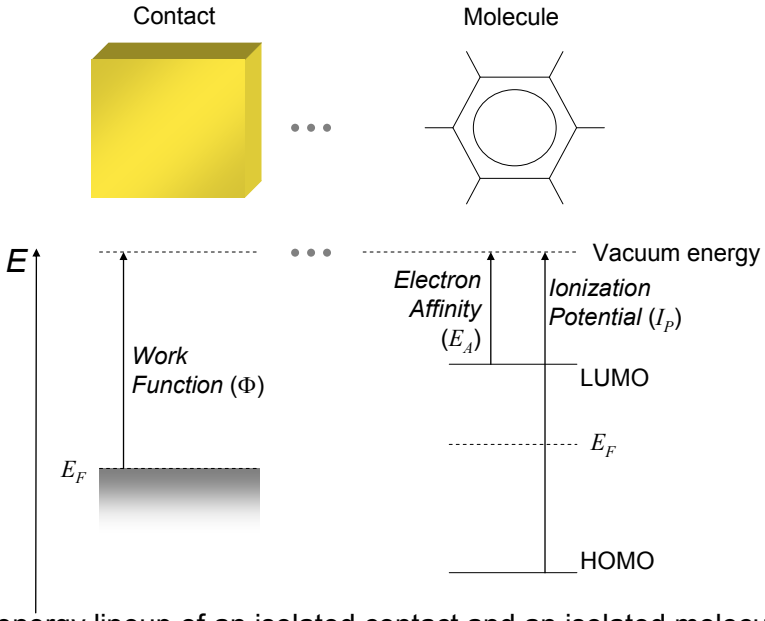


Fig. 3.4. The energy lineup of an isolated contact and an isolated molecule. If there is no voltage source in the system, the energy of a free electron is identical at the contact and molecule locations. Thus, the vacuum energies align. The Fermi energies may not, however. But at equilibrium, the Fermi energies are forced into alignment by charge transfer.

Since Fermi levels change with the addition or subtraction of charge, equilibrium is obtained by charge transfer between the contact and the molecule. Charge transfer changes the potential of the contact relative to the molecule, shifting the relative vacuum energies. This is known as „charging". Charge transfer also affects the Fermi levels as electrons fill some states and empty out of others. Both charging and state filling effects can be modeled by capacitors. We"ll consider electron state filling first.

(i) The Quantum Capacitance

Under equilibrium conditions, the Fermi energy must be constant in the metal and the molecule. We can draw an analogy to flow between water tanks. The metal is like a very large tank. The molecule, with its much smaller density of states, behaves as a narrow column. When the metal and molecule are connected, water flows to align the filling levels.

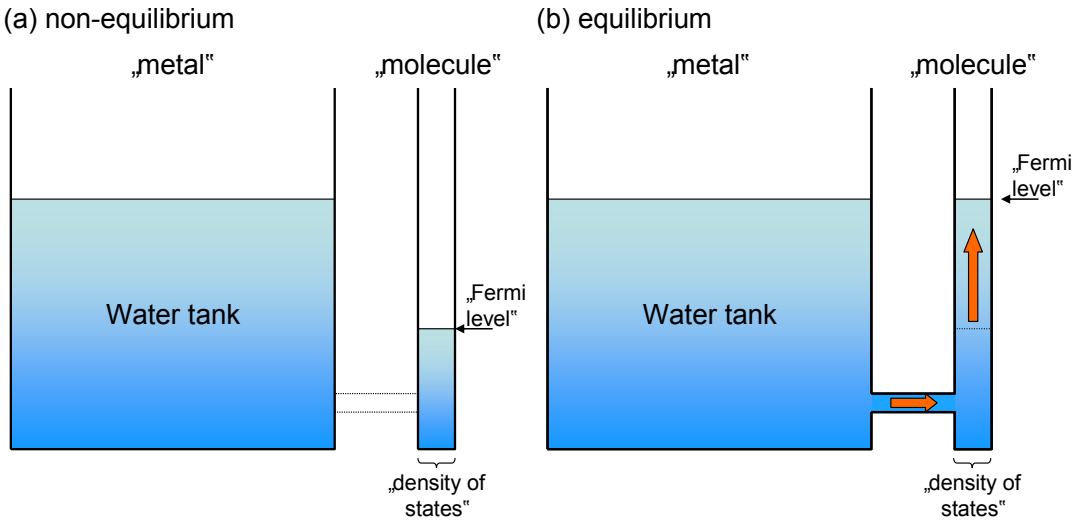


Fig. 3.5. An analogy for electron transfer at the interface between a metal and a molecule. The size of the water tank is equivalent to the density of states. The Fermi level is equivalent to the water level. If the „metal" has a sufficiently large density of states, then the change in its water level is imperceptible.

But a molecule will not necessarily have a uniform density of states as shown in Fig. 3.5. It is also possible that only a fractional amount of charge will be transferred. For example, imagine that some fractional quantity δn electrons are transferred from the contact to the molecule. It is possible for the wavefunction of the transferred electron to include both the contact and the molecule. Since part of the shared wavefunction resides on the molecule, this is equivalent to a fractional charge transfer.

But if δn were equal to +1, the LUMO would be half full and hence the Fermi energy would lie on the LUMO, while if δn were -1, the HOMO would be half full and hence the Fermi energy would lie on the HOMO. In general, the number of charges on the molecule is given by

$$n = \int_{-\infty}^{+\infty} g(E) f(E, E_F) dE \tag{3.1}$$

where $g(E)$ is the density of molecular states per unit energy. For small shifts in the Fermi energy, we can linearize Eq. (3.1) to determine the effect of charge transfer on E_F. We are interested in the quantity dE_F/dn. For degenerate systems we can simplify Eq. (3.1):

$$n = \int_{-\infty}^{E_F} g(E) dE \tag{3.2}$$

taking the derivative with respect to the Fermi energy gives:

$$\frac{dn}{dE_F} = g(E_F) \tag{3.3}$$

We can re-arrange this to get:

$$\delta E_F = \frac{\delta n}{g(E_F)} \tag{3.4}$$

Thus after charge transfer the Fermi energy within the molecule changes by $\delta n/g$, where g is the density of states per unit energy.

Sometimes it is convenient to model the effect of filling the density of states by the „quantum capacitance" which we will define as:

$$C_Q = q^2 g(E_F) \tag{3.5}$$

i.e.

$$\delta E_F = \frac{q^2}{C_Q} \delta n \tag{3.6}$$

If the molecule has a large density of states at the Fermi level, its quantum capacitance is large, and more charge must be transferred to shift the Fermi level.

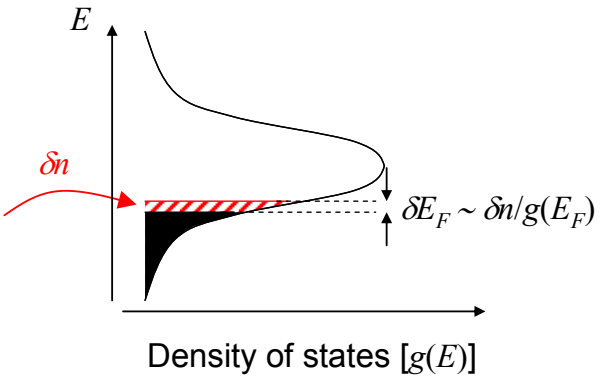


Fig. 3.6. Transferring charge changes the Fermi level in a conductor. The magnitude of the change is determined by the density of states at the Fermi level, and often expressed in terms of a „quantum capacitance".

We can also calculate the quantum capacitance of the contact. Metallic contacts contain a large density of states at the Fermi level, meaning that a very large number of electrons must be transferred to shift its Fermi level. Thus, we say that the Fermi energy of the

contact is „*pinned*" by the density of states. Another way to express this is that the quantum capacitance of the contact is approximately infinite.

The quantum capacitance can be employed in an equivalent circuit for the metal-molecule junction. But we have generalized the circuit such that each node potential is the Fermi level, not just the electrostatic potential as in a conventional electrical circuit.

In the circuit below, the metal is modeled by a voltage source equal to the chemical potential μ_1 of the metal. Prior to contact, the Fermi level of the molecule is E_F^0. The contact itself is modeled by a resistor that allows current to flow when the Fermi levels on either side of the contact are misaligned. Charge flowing from the metal to the molecule develops a potential across the quantum capacitance. But note that this is a change in the Fermi level, not an electrostatic potential. It is also important to note that the quantum capacitance usually depends on the Fermi level in the molecule. The only exception is if the density of states is constant as a function of energy. Thus, a constant value of C_Q can only be employed for small deviations between μ_1 and E_F^0.

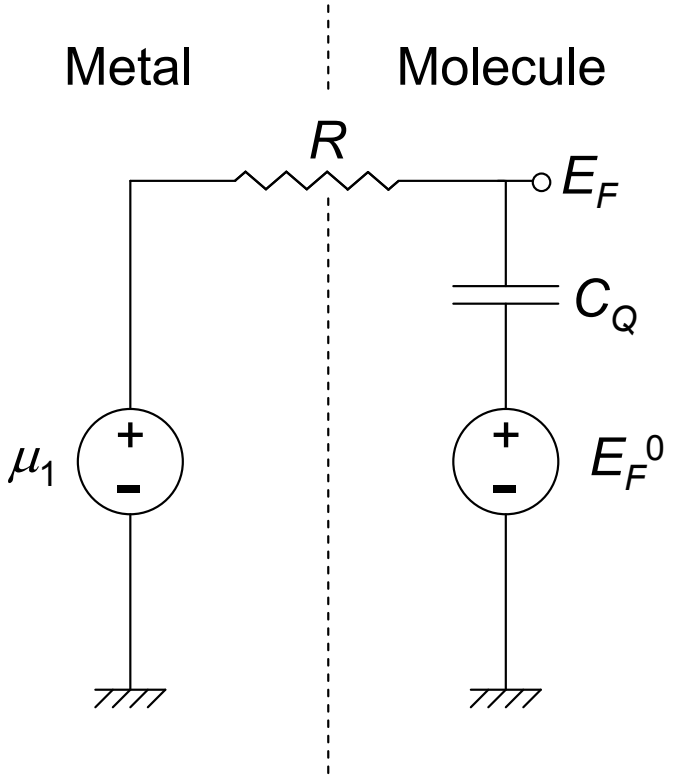


Fig. 3.7. A small signal model for the metal-molecule junction. The effects of charging are not included. The resistor will be characterized further in later sections.

(ii) Electrostatic Capacitance

Unfortunately, the establishment of equilibrium between a contact and the molecule is not as simply as water flow between two tanks. Electrons, unlike water, are charged. Thus, the transfer of electrons from the contact to a molecule leaves a net positive charge on the contact and a net negative charge on the molecule.

Charging at the interface changes the potential of the molecule relative to the metal and is equivalent to shifting the entire water tanks up and down. Charging assists the establishment of equilibrium and it reduces the number of electrons that are transferred after contact is made.

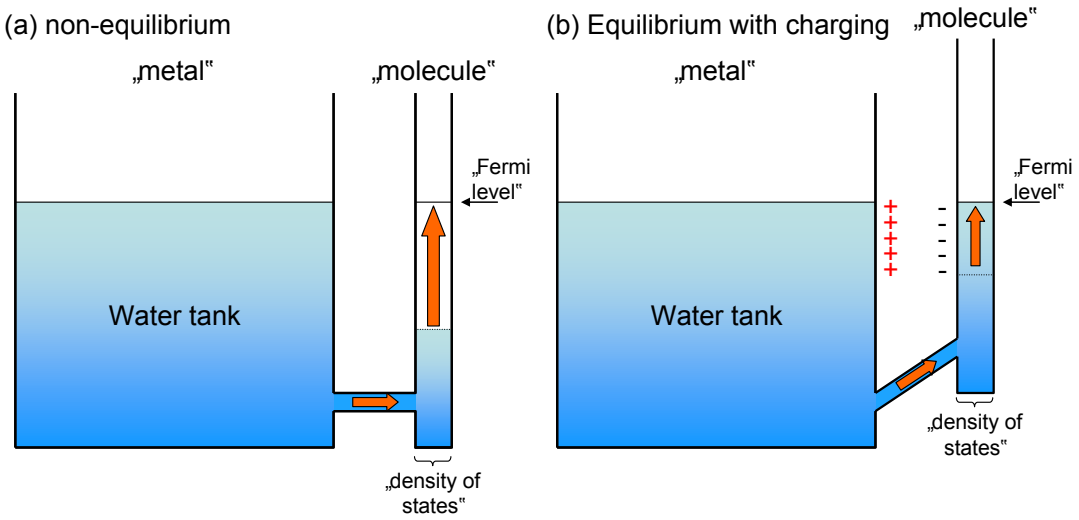


Fig. 3.8. Electrons carry charge and shift the potential when they are transferred between a metal and a molecule. The resulting change in potential is equivalent to lifting up the „molecule" column of water. The water levels must ultimately match, but now less water is required to be transferred.

The contact and the molecule can be considered as the two plates of a capacitor. In Fig. 3.9 we label this capacitor, C_{ES} - the electrostatic capacitance, to distinguish it from the quantum capacitance discussed in the previous section.

When charge is transferred at the interface, the capacitor is charged, a voltage is established and the molecule changes potential. The change in the molecule's potential per electron transferred is known as the **charging energy** and is reflected in a shift in the vacuum energy. From the fundamental relation for a capacitor:

$$C_{ES} = \frac{Q}{V} \tag{3.7}$$

where V is the voltage across the capacitor. We can calculate the change in potential due to charging:

$$U_C = qV = \frac{q^2}{C_{ES}} \delta n \,. \tag{3.8}$$

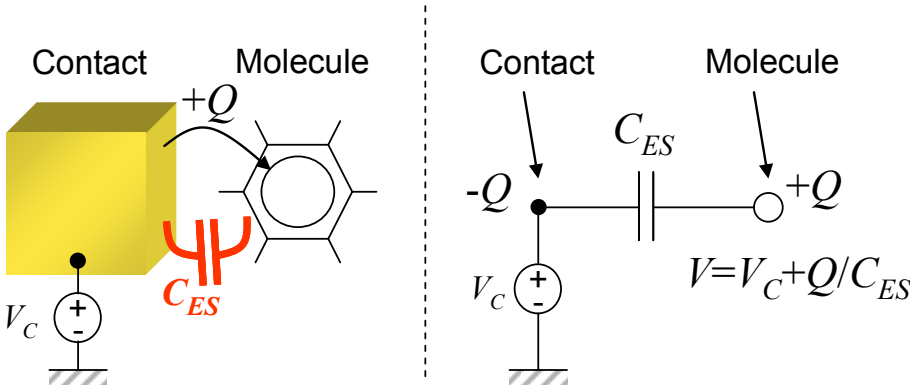


Fig. 3.9. A contact and a molecule can be modeled as two plates of a parallel capacitor. When charge is transferred, this electrostatic capacitance determines the change in electrostatic potential, and hence the shift in the vacuum energy.

We will find that δn is a dynamic quantity – it changes with current flow. It can be very important in nanodevices *because the electrostatic capacitance is so small*. For the small spacings between contact and conductor typical of nanoelectronics (e.g. 1 nm), the charging energy can be on the order of 1V per electron.

Summarizing these effects, we find that the Fermi energy of the neutral molecule, E_F^0, is related to the Fermi energy of the metal-molecule combination, E_F, by

$$E_F = \delta n/g + \frac{q^2}{C_{ES}}\delta n + E_F^0 \tag{3.9}$$

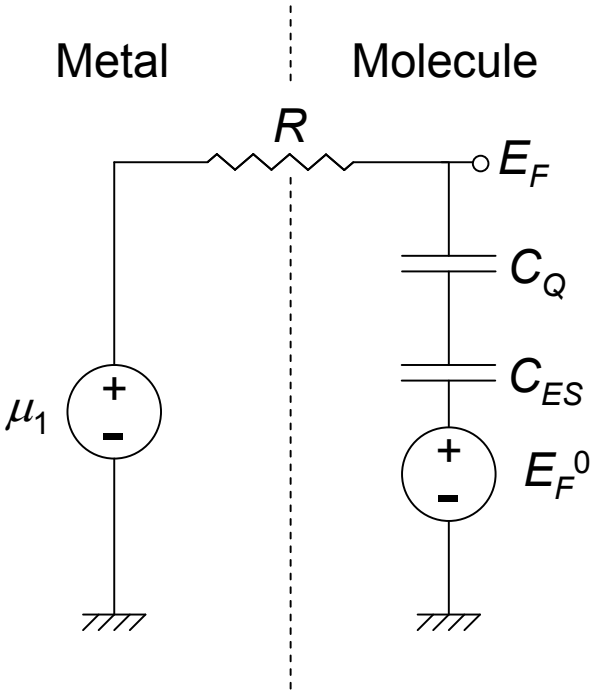


Fig. 3.10. A small signal model for the metal-molecule junction, including the effects of charging. The resistor will be characterized further in later sections.

Or, in terms of the quantum capacitance:

$$E_F = \frac{q^2}{C_Q}\delta n + \frac{q^2}{C_{ES}}\delta n + E_F^0 \tag{3.10}$$

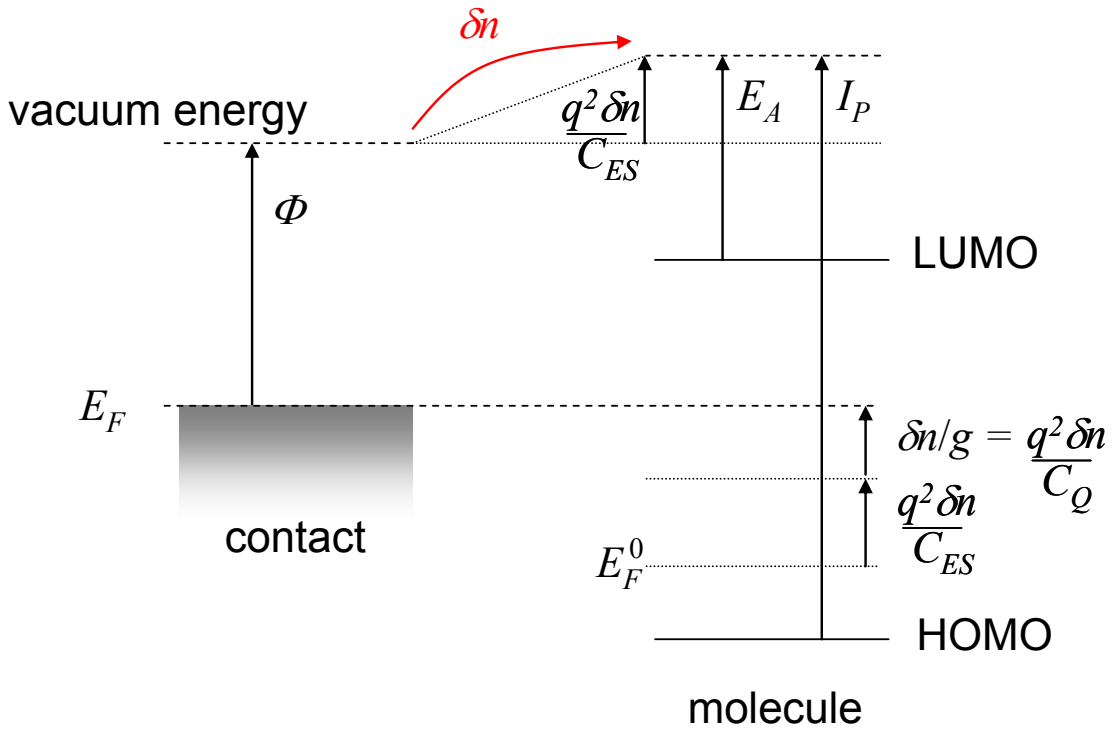


Fig. 3.11. Changes in energy level alignment when charge is transferred from the metal to a molecule. Charging of the molecule corresponds to applying a voltage across an interfacial capacitor, thereby changing the potential of the molecule. Consequently, the vacuum level shifts at the molecule"s location, shifting all the molecular states along with it. In addition, the transferred charge fills some previous empty states in the molecule. Both effects change the Fermi energy in the molecule.

Calculation of the electrostatic capacitance

(i) Isolated point conductors

For small conductors like single molecules or quantum dots, it is sometimes convenient to calculate C_{ES} by assuming that the conductor is a sphere of radius R. From Gauss"s law, the potential at a point with radius r from the center of the sphere is:

$$V = \frac{Q}{4\pi\varepsilon r} \tag{3.11}$$

where $r > R$, ε is the dielectric constant and Q is the net charge on the sphere.
If we take the potential at infinity to be zero, then the potential of the sphere is $V = Q/4\pi\varepsilon R$ and the capacitance is

$$C_{ES} = \frac{Q}{V} = 4\pi\varepsilon R \tag{3.12}$$

The notable aspect of Eq. (3.12) is that the electrostatic capacitance scales with the size of the conductor. Consequently, the charging energy of a small conductor can be very large. For example, Eq. (3.12) predicts that the capacitance of a sphere with a radius of $R = 1\text{nm}$ is approximately $C_{ES} = 10^{-19}$ F. The charging energy is then $U_C = 1.6\text{eV}$ per charge.

(ii) Conductors positioned between source and drain electrodes

In general, the potential profile for an arbitrary distribution of charges must be calculated using Gauss"s law. But we can often make some approximations. The source and drain contacts can sometimes be modeled as a parallel plate capacitor with

$$C = \frac{\varepsilon A}{d} \tag{3.13}$$

where A is the area of each contact and d is their separation. This approximation is equivalent to assuming a uniform electric field between the source and drain electrodes. This is valid if $A >> d$ and there is no net charge between the contacts. For source and drain electrodes separated by a distance l, the source and drain capacitances at a distance z from the source are:

$$C_S(z) = \frac{\varepsilon A}{z}, \quad C_D(z) = \frac{\varepsilon A}{l - z}. \tag{3.14}$$

The potential varies linearly as expected for a uniform electric field.

$$U(z) = -qV_{DS}\frac{1/C_S(z)}{1/C_D(z) + 1/C_S(z)} = -qV_{DS}\frac{z}{l}. \tag{3.15}$$

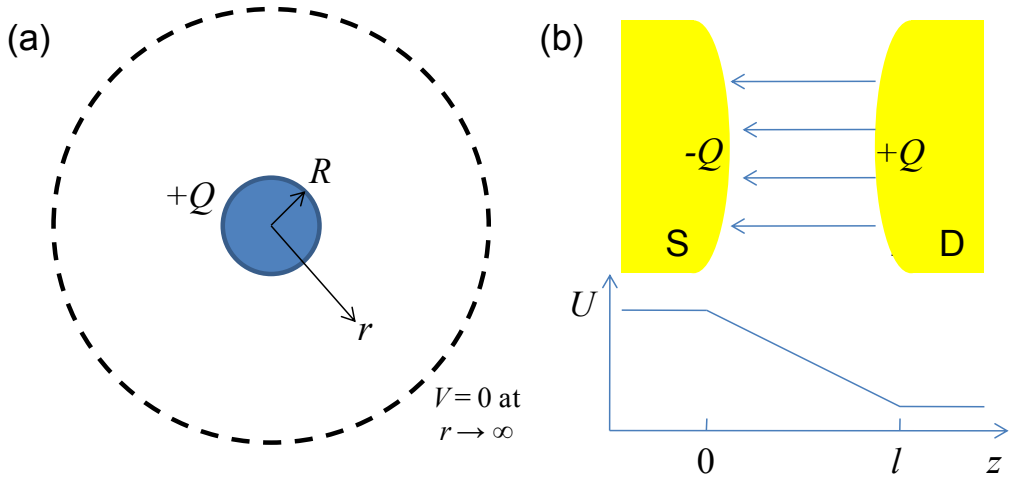


Fig. 3.12. (a) The capacitance of an isolated 0-d conductor is calculated by assuming the potential at infinity is zero. **(b)** A uniform electric field between the source and drain yields a linearly varying potential. The source and drain capacitors can be modeled by parallel plates.

Part 3. Two Terminal Quantum Dot Devices

Current Flow in Two Terminal Quantum Dot/Single Molecule Devices

In this section we present a simplified model for conduction through a molecule. It is based on the „toy model" of Datta, *et al.*[†] which despite its relative simplicity describes many of the essential features of single molecule current-voltage characteristics.

The contact/molecule/contact system at equilibrium is shown in Fig. 3.13. At equilibrium, $\mu_1 = E_F = \mu_2$. Since there are two contacts, this is an example of a two terminal device. In keeping with convention, we will label the electron injecting contact, the source, and the electron accepting contact, the drain. We will model the molecule by a quantum dot. This is accurate if the center of the molecule is much more conductive than its connections to the contacts.

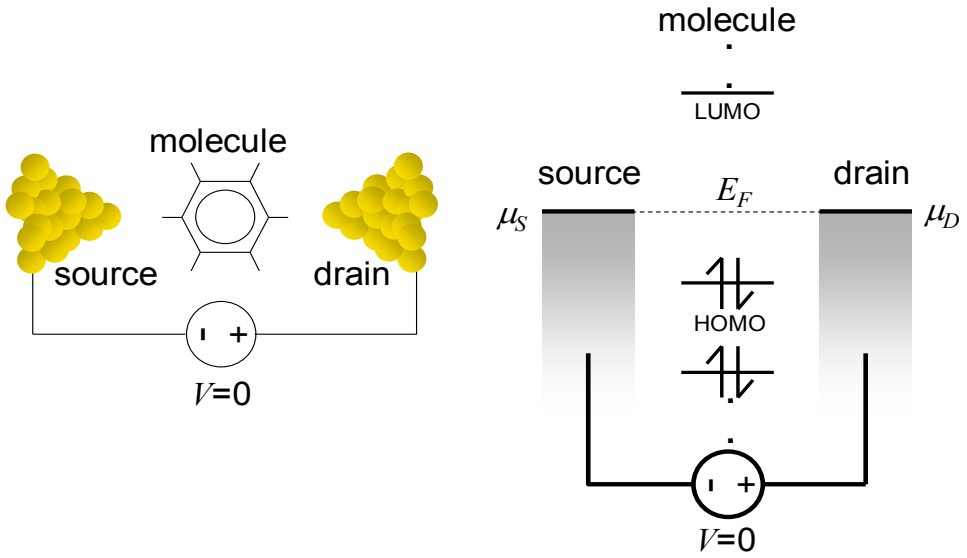


Fig. 3.13. A two terminal device with a molecular/quantum dot conductor. At equilibrium no current flows and the Fermi levels are aligned.

Now, when we apply a potential between the source and drain contacts we shift Fermi level of one contact with respect to the other, i.e.

$$\mu_D - \mu_S = -qV_{DS} \tag{3.16}$$

There are two effects on the molecule:
(i) *The electrostatic effect*: the potential at the molecule is changed by the electric field established between the contacts. The energy levels within the molecule move rigidly up or down relative to the contacts.
(ii) *The charging effect*: Out of equilibrium, a current will flow and the amount of charge on the molecule changes. It may increase if current flows through the LUMO, or decrease if current flows through the HOMO.

[†] S. Datta, „Quantum transport: atom to transistor" Cambridge University Press (2005).
F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

Unfortunately, these effects are linked: moving the molecular energy levels with respect to the contact energy levels changes the amount of charge supplied to the molecule by the contacts. But the charging energy associated with charge transfer in turn changes the potential of the molecule.

We will first consider static and charging effects independently.

(i) Electrostatics: The Capacitive Divider Model of Potential

Our two terminal device can be modeled by a quantum dot linked to the source and drain contacts by two capacitors, C_S and C_D, respectively. The values of these capacitors depend on the geometry of the device. If the molecule is equi-spaced between the contacts we might expect that $C_S \sim C_D$. On the other hand, if the molecule is closely attached to the source but far from the drain, we might expect $C_S >> C_D$. (Recall that the capacitance of a simple parallel plate capacitor is inversely proportional to the spacing between the plates.)

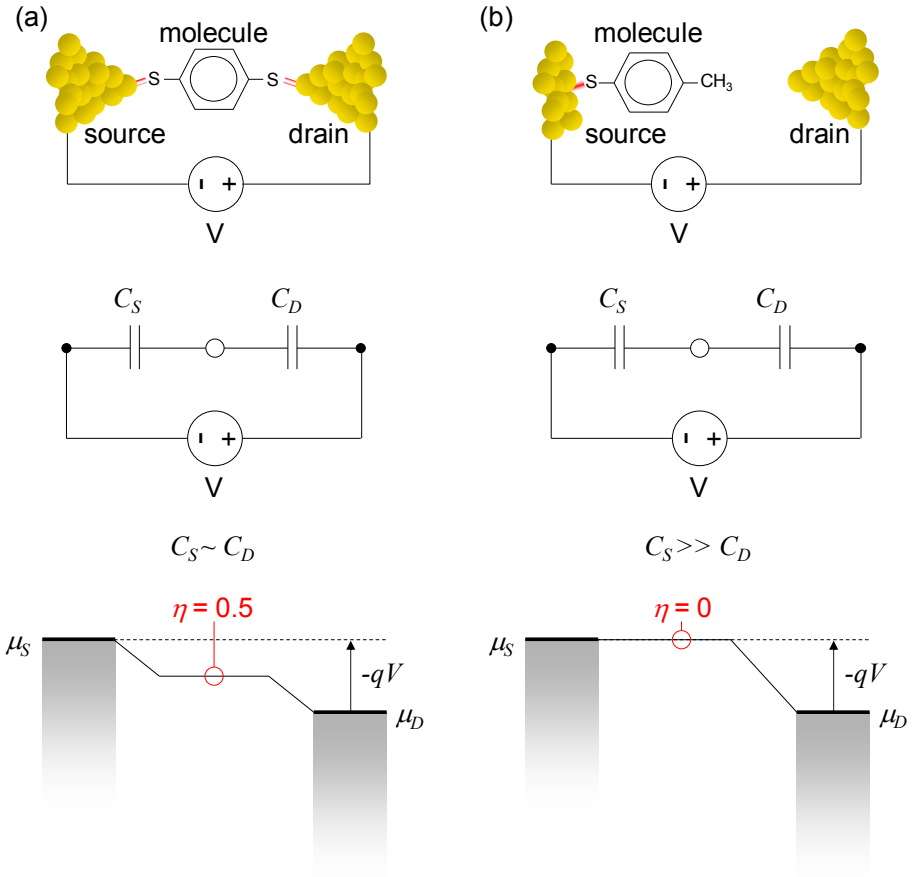


Fig. 3.14. Two single molecule two terminal devices accompanied by possible potential profiles in the molecular conductor. **(a)** symmetric contacts, **(b)** asymmetric contacts. We are concerned with the voltage in the center of the molecule. This is determined by the voltage division factor, η. It can be obtained by from a voltage divider constructed from capacitors. Adapted from F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

These two potential profiles are shown in Fig. 3.14. The voltage is calculated from the capacitive divider. Thus, an applied voltage, V, shifts the chemical potentials of both the source and drain contacts:[†]

$$E_F = -\frac{1/C_S}{1/C_D + 1/C_S} qV_{DS} + \mu_S \tag{3.17}$$

It is convenient to use the Fermi energy of the molecule at equilibrium as a reference, i.e. if we set $E_F = 0$:

$$\begin{aligned} \mu_S &= +\frac{C_D}{C_S + C_D} qV_{DS} \\ \mu_D &= -\frac{C_S}{C_S + C_D} qV_{DS} \end{aligned} . \tag{3.18}$$

We can define a voltage division factor, η.[†] It gives the fraction of the applied bias that is dropped between the molecule and the source contact, *i.e.*

$$\eta = \frac{C_D}{C_S + C_D} \tag{3.19}$$

As shown in Fig. 3.15, the voltage division factor determines in part whether conduction occurs through the HOMO or the LUMO. If $\eta = 0$, then the molecular energy levels are fixed with respect to the source contact. As the potential of the drain is increased, conduction eventually occurs through the HOMO. But if the potential of the drain is decreased, conduction can occur through the LUMO. The current-voltage characteristic of this device will exhibit a gap around zero bias that corresponds to the HOMO-LUMO gap. [†]

If $\eta = 0.5$, however, then irrespective of whether the bias is positive or negative, current always flows through the molecular energy level closest to the Fermi energy. In this situation, which is believed to correspond to most single molecule measurements,[1] the gap around zero bias is *not* the HOMO-LUMO gap, but, in this example, four times the Fermi energy – HOMO separation.[§]

The voltage division factor is a crude model of the potential profile, which more generally could be obtained from Poisson‟s equation. η is also likely to vary with bias. At high biases, there may be significant charge redistribution within the molecule, leading to a change in η.[†]

[†] F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

[§] It is possible to experimentally distinguish between $\eta = 0.5$ and $\eta = 0$ by choosing contact metals with different work functions. If the conductance gap is observed to change then it cannot be determined by the HOMO-LUMO gap, and hence $\eta \neq 0$

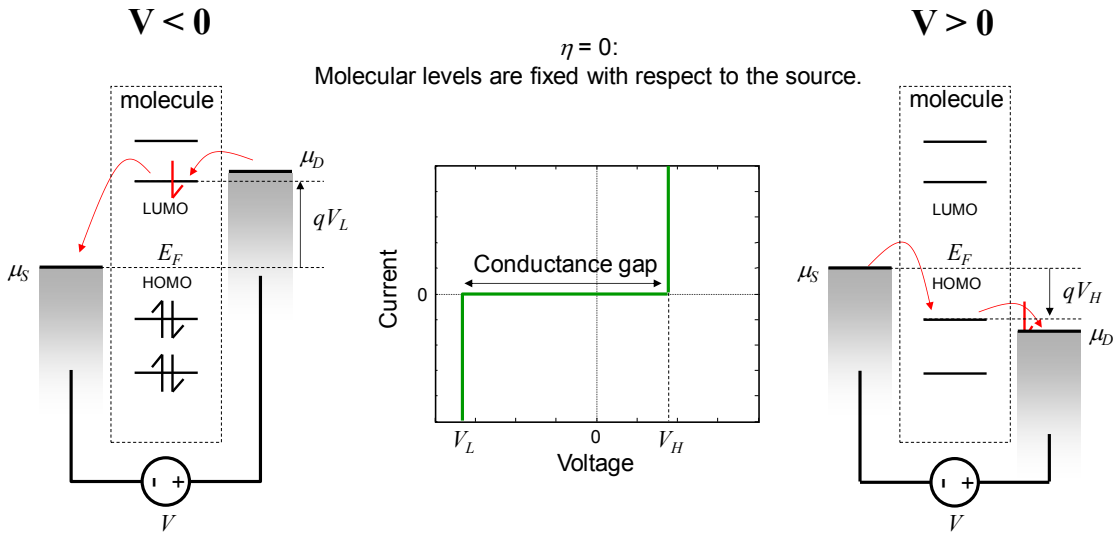


Fig. 3.15. The voltage division factor is crucial in determining the conduction level in a single molecule device. In this example, when $\eta = 0$, conduction always occurs through the HOMO when the applied bias is positive, and through the LUMO when the applied bias is negative. The conductance gap is determined by the HOMO-LUMO separation. Adapted from F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

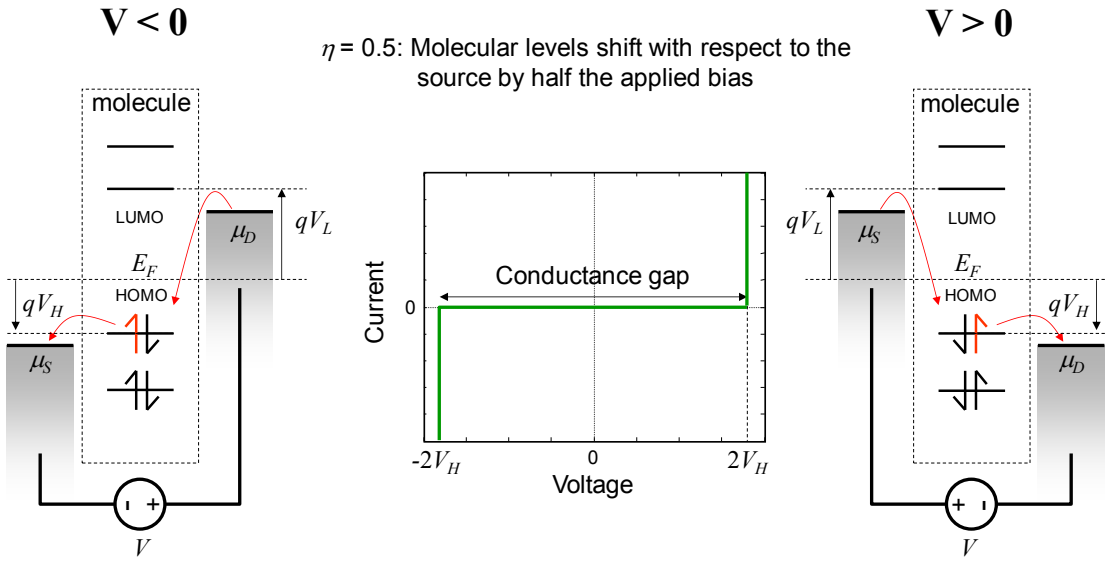


Fig. 3.16. When $\eta = 0.5$, conduction always occurs through the molecular orbital closet to the Fermi Energy. In this example that is the HOMO, irrespective of the polarity of the applied bias. Adapted from F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

(ii) Charging

Previously, we defined the charging energy as the change in the molecule‟s potential per additional electron. To calculate the net effect of charging we need the number of electrons transferred.

At equilibrium, the number of electrons on the molecule is determined by its Fermi energy.

$$N_0 = \int_{-\infty}^{\infty} g(E) f(E, E_F) dE \tag{3.20}$$

Under bias, the electron distribution on the molecule is no longer in equilibrium. We will define the number of electrons under bias as *N*.

Thus, the change in potential at the molecule due to charging is

$$U_C = \frac{q^2}{C_{ES}}(N - N_0) \tag{3.21}$$

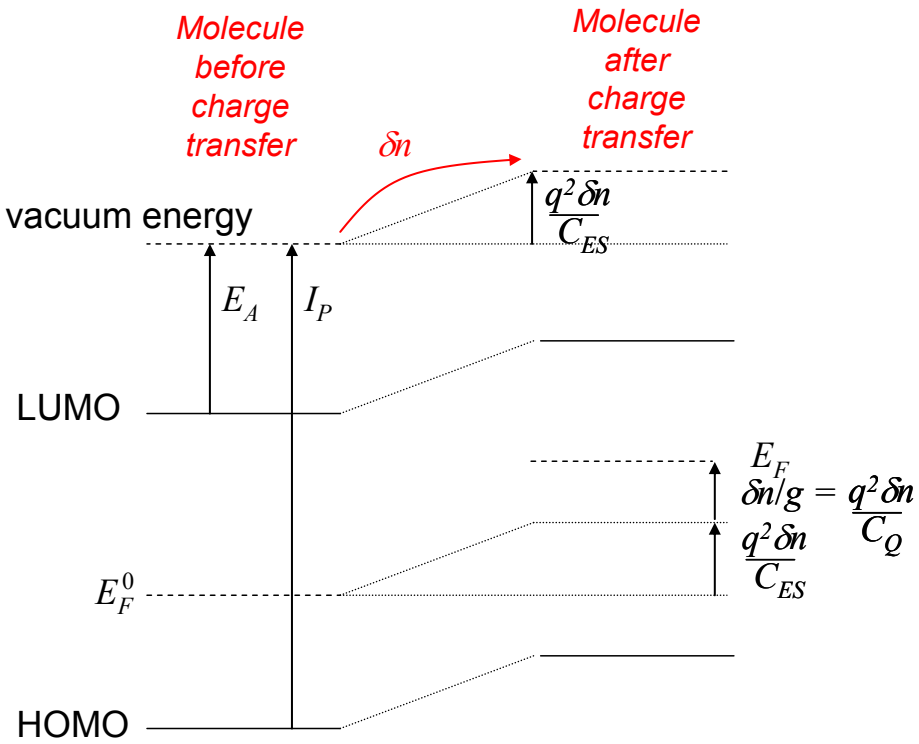


Fig. 3.17. The effect of charging on a molecule. The addition of electrons shifts the molecular potential (and hence all orbitals within the molecule) in order to repel the addition of more electrons. Note that although we have shown the expected change in the Fermi level, this is only meaningful if the molecule remains in equilibrium. Adapted from F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

Summary

The net change in potential at the molecule, U, is the sum of electrostatic and charging effects:

$$U = U_{ES} + U_C \tag{3.22}$$

By applying the source-drain voltage relative to a ground at the molecule we have forced $U_{ES} = 0$ in Fig. 3.16. But it will not always be possible to ignore electrostatic effects on U if the ground is positioned elsewhere. Analyses of transistors, for example, typically define the source to be ground.

We model the effect of the change in potential by rigidly shifting all the energy levels within the molecule, i.e.

$$g \to g(E - U) \tag{3.23}$$

Calculation of Current

Let‟s model the net current at each contact/molecule interface as the sum of two components: the contact current, which is the current that flows into the molecule, and the molecule current, which is the current that flows out of the molecule.

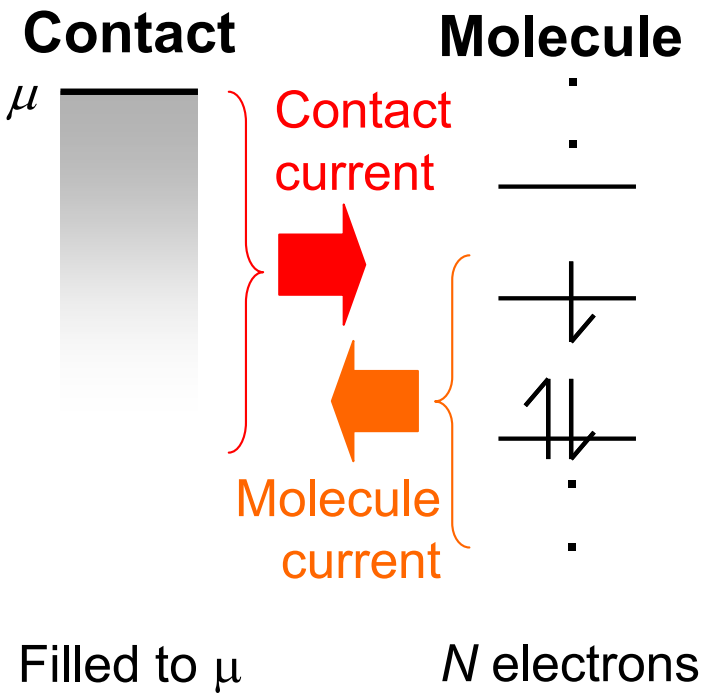


Fig. 3.18. The net current at a contact/molecule interface can be broken into a contact current – the current that flows out of the contact - and a molecule current – the current that flows out of the molecule. At equilibrium, these currents must balance.

(i) The contact current

This current is the number of available states in the molecule filled per second. Electrons in the contact are filled to its chemical potential. They cannot jump into higher energy states in the molecule. The total number of electrons that can be transferred is simply equal to the number of states.

At the source contact, we get

$$N_S = \int_{-\infty}^{\infty} g(E-U) f(E,\mu_S)\, dE \tag{3.24}$$

where $g(E\text{-}U)$ is the molecular density of states shifted by the net potential change. Similarly, if at the drain contact then the number of electrons, N_D, that could be transferred level is

$$N_D = \int_{-\infty}^{\infty} g(E-U) f(E,\mu_D)\, dE \tag{3.25}$$

Let"s define the transfer rate at the source and drain contacts as $1/\tau_S$ and $1/\tau_D$, respectively. Then the contact currents are

$$I_S^C = q\frac{N_S}{\tau_S}, \quad I_D^C = -q\frac{N_D}{\tau_D} \tag{3.26}$$

Note that we have defined electron flow out of the source and into the drain as positive.

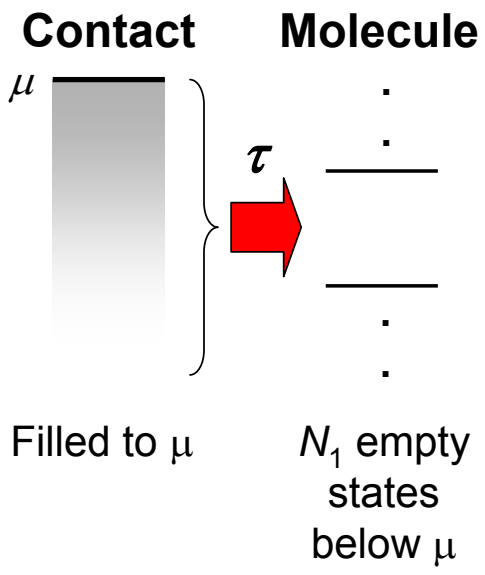


Fig. 3.19. The contact current is the rate of charge transfer from the contact to the molecule. Only states in the molecule with energies below the chemical potential of the contact may be filled. The transfer rate of a single electron from the contact is $1/\tau_S$.

(ii) The molecule current

Now, if we add electrons to the molecule, these electrons can flow back into the contact, creating a current opposing the contact current. The molecule current is the number of electrons transferred from the molecule to the contact per second.

Thus, the molecule currents into the source and drain contacts are

$$I_S^M = -q\frac{N}{\tau_S}, \quad I_D^M = q\frac{N}{\tau_D} \tag{3.27}$$

where we have again defined electron flow out of the source and into the drain as positive.

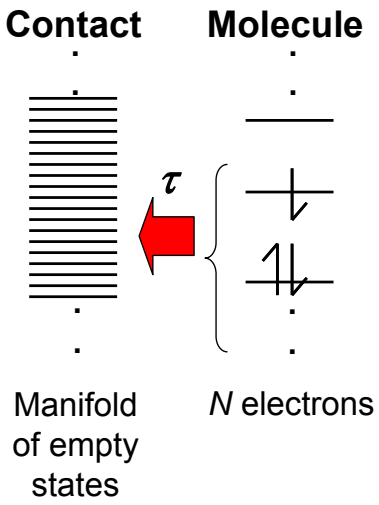


Fig. 3.20. The molecule current is the rate of charge transfer from the molecule to the contact. Note that the electrons on the molecule are not necessarily in equilibrium. The lifetime of a single electron on the molecule is τ.

From Eqns. (3.26) and (3.27) the net current at the source contact is

$$I_S = \frac{q}{\tau_S}\left(N_S - N\right) \tag{3.28}$$

and the net current at the drain contact is

$$I_D = \frac{q}{\tau_D}\left(N - N_D\right) \tag{3.29}$$

Note that we have assumed that the transfer rates in and out of each contact are identical. For example, let"s define τ_S^{M} as the lifetime of an electron in the molecule and $1/\tau_S^{\mathrm{C}}$ as the rate of electron transfer from the source contact. It is perhaps not obvious that $\tau_S^{\mathrm{M}} = \tau_S^{\mathrm{C}}$, but examination of the inflow and outflow currents at equilibrium confirms that it must be so. When the source-molecule junction is at equilibrium, no current flows. From Eqns. (3.20), (3.21) and (3.24), we have $N_S = N$. Thus, for $I_S = 0$ we must have $\tau_1^{\mathrm{M}} = \tau_1^{\mathrm{C}}$.

Equating the currents in Eqns. (3.28) and (3.29) gives†

$$I = q\int_{-\infty}^{\infty} g(E-U)\frac{1}{\tau_S+\tau_D}\left(f(E,\mu_S)-f(E,\mu_D)\right)dE \tag{3.30}$$

and

$$N = \int_{-\infty}^{\infty} g(E-U)\frac{\tau_D f(E,\mu_S)+\tau_S f(E,\mu_D)}{\tau_S+\tau_D}dE \tag{3.31}$$

The difficulty in evaluating the current is that it depends on U and hence N. But Eq. (3.31) is not a closed form solution for N, since the right hand side also contains a N dependence via U. Except in simple cases, this means we must iteratively solve for N, and then use the solution to get I. This will be discussed in greater detail in the problems accompanying this Part.

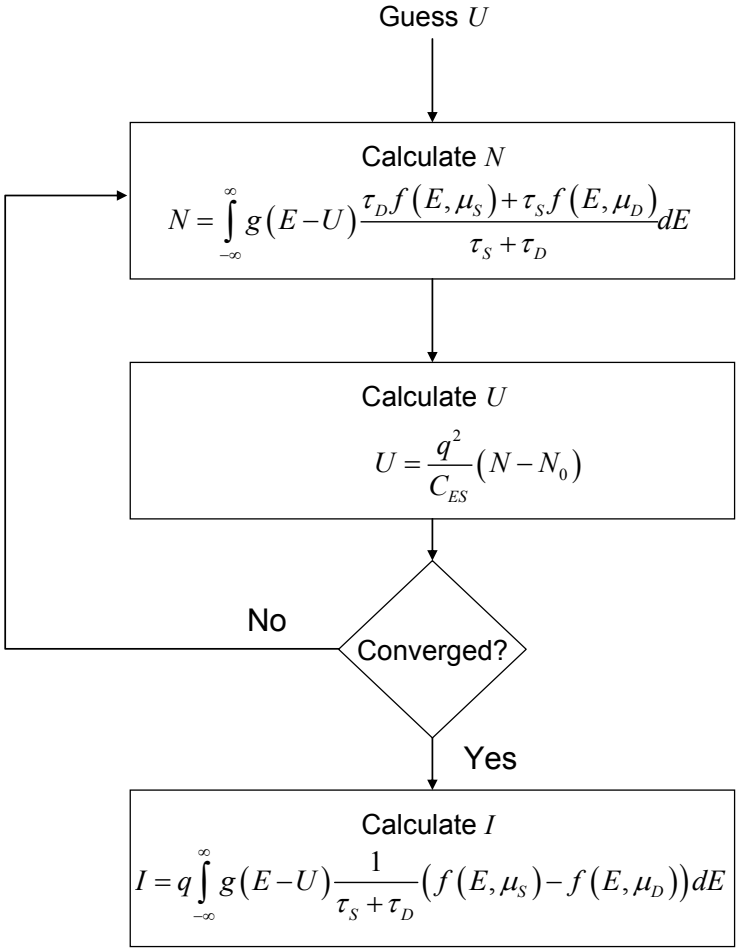


Fig. 3.21. A flow diagram describing an iterative solution to the IV characteristics of two terminal molecular devices. Adapted from F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

† F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

Analytic calculations of the effects of charging

The most accurate method to determine the *IV* characteristics of a quantum dot device is to solve for the potential and the charge density following the scheme of Fig. 3.21. This is often known as the self consistent approach since the calculation concludes when the initial guess for the potential U has been modified such that it is consistent with the value of U calculated from the charge density.

Unfortunately, numerical approaches can obscure the physics. In this section we will make some approximations to allow an analytic calculation of charging. We will assume operation at $T = 0$K, and discrete molecular energy levels, *i.e.* weak coupling between the molecule and the contacts such that $(\Gamma = \hbar/\tau) \to 0$.

Let"s consider a LUMO state with energy E_{LUMO} that is above the equilibrium Fermi level. Under bias, the energy of the LUMO is altered by electrostatic and charging-induced changes in potential. When we apply the drain source potential, it is convenient to assume that the molecule is ground. Under this convention, the only change in the molecule"s potential is due to charging. Graphically, the physics can be represented by plotting the energy level of the molecule in the presence and absence of charging. In Fig. 3.22, below, we shade the region between the charged and uncharged LUMOs. Now

$$U_C = \frac{q^2}{C_{ES}}(N - N_0), \tag{3.32}$$

and at $T = 0$K, $N_0 = 0$ for the LUMO in Fig. 3.22. Thus, the area of the shaded region is proportional to the charge on the molecule.

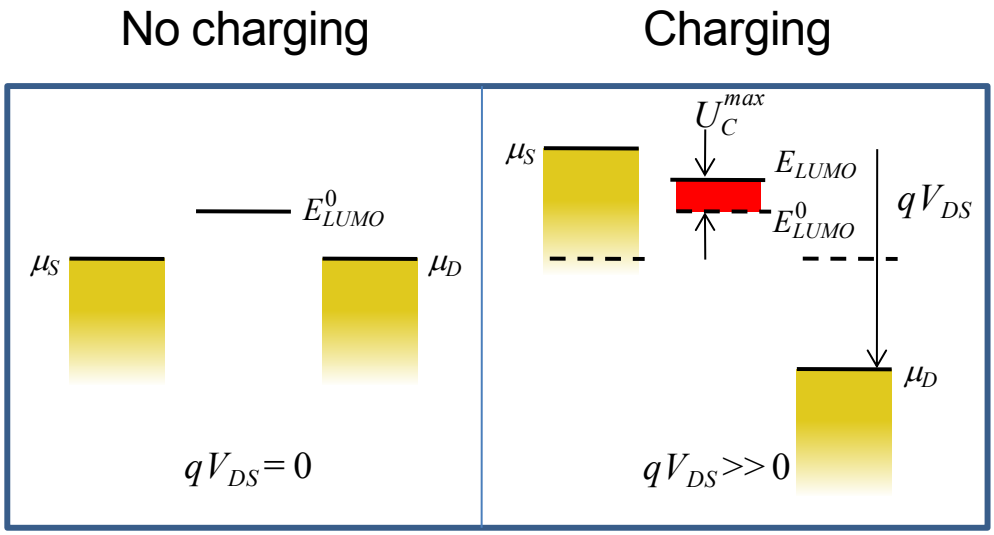


Fig. 3.22. Under bias, the energy levels of the molecule can be shifted by electrostatic and charging-induced changes in the potential. If we assume the molecule is ground, then the electrostatic changes in potential alter the source and the drain. Only charging then alters the molecular energy level. The difference between the charged and uncharged molecular energy levels is proportional to the charge on the molecule and is shaded red.

The graphical approach is a useful guide to the behavior of the device. There are three regions of operation, each shown below.

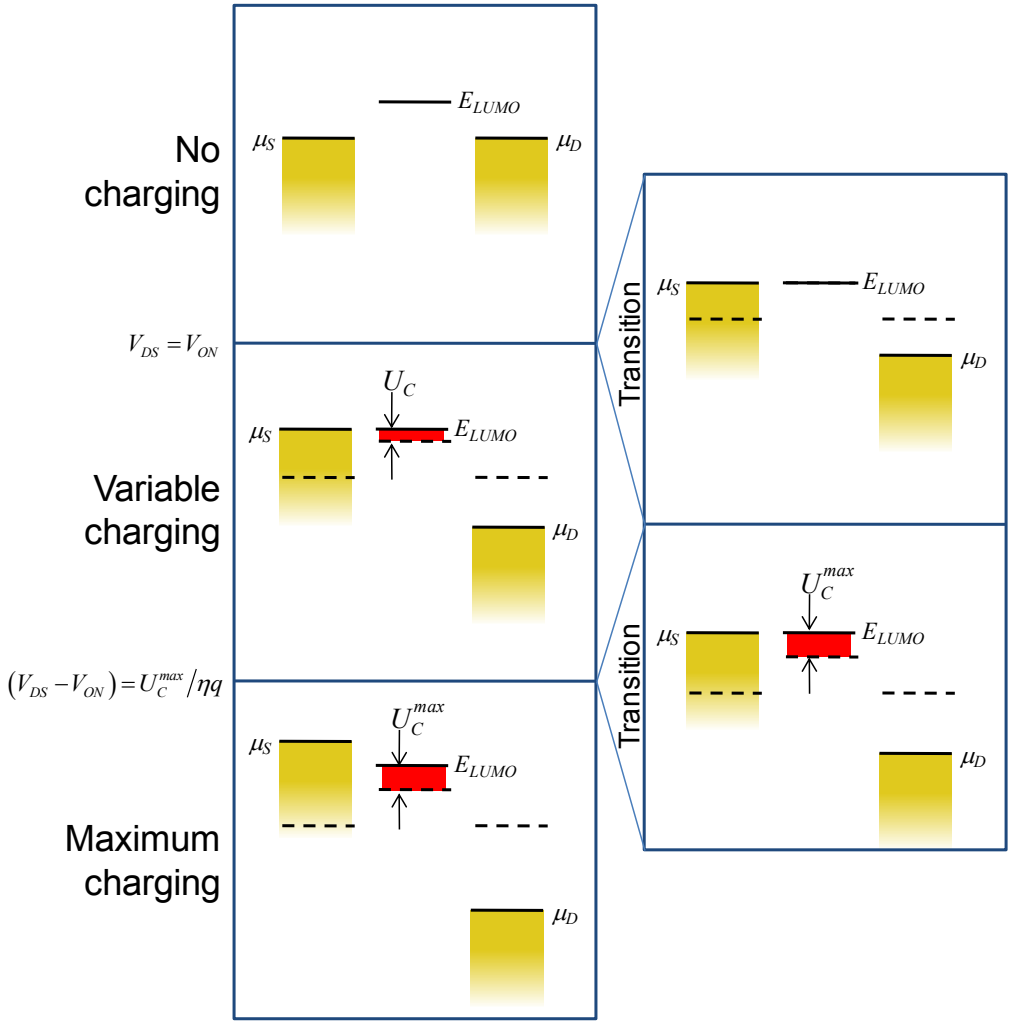


Fig. 3.23. The three regions of operation for a two terminal quantum dot device with discrete energy levels at T = 0K. Current flow requires the source to inject carriers into the molecular energy levels. The onset of current flow occurs when the chemical potential of the source is resonant with the LUMO. Additional drain-source bias charges the molecule, and the current increases linearly with the molecular charge. Finally, a maximum charge density is reached. Further increases in applied bias do not increase the charging energy or the current flow.

(i) No charging

At $T = 0$K and $V_{DS} = 0$ there is no charge in the LUMO. Charging cannot occur unless electrons can be injected from the source into the LUMO. So as V_{DS} increases, charging remains negligible until the LUMO energy is aligned with the chemical potential of the source. Thus, for $\mu_S < E_{LUMO}$ the charging energy, $U_C = 0$.

In this region, $I_{DS} = 0$. We define the bias at which current begins to flow as $V_{DS} = V_{ON}$. V_{ON} is given by

$$V_{ON} = \frac{E^0_{LUMO} - \mu_S}{q\eta} \tag{3.33}$$

where E^0_{LUMO} is the LUMO energy level at equilibrium.

(ii) Maximum charging

For $\mu_S > E_{LUMO}$, the charge on the LUMO is independent of further increases in V_{DS}. It is a maximum. From Eq. (3.31), the LUMO"s maximum charge is

$$N^{max} = \frac{2\tau_D}{\tau_S + \tau_D} \tag{3.34}$$

Consequently, the charging energy is

$$U_C^{max} = \frac{2q^2}{C_{ES}} \frac{\tau_D}{\tau_S + \tau_D} \tag{3.35}$$

For all operation in forward bias, it is convenient to calculate the current from Eq. (3.29). At $T = 0$K, the charges injected by the drain $N_D = 0$. Consequently,

$$I_{DS} = \frac{qN^{max}}{\tau_D} = \frac{2q}{\tau_S + \tau_D} \tag{3.36}$$

Maximum charging occurs for voltages $\mu_S > E_{LUMO}$. We can rewrite this condition as

$$\left(V_{DS} - V_{ON}\right) > U_C^{max} / \eta q \tag{3.37}$$

(iii) Variable charging

Charging energies between $0 < U_C < U_C^{max}$ require that $\mu_S = E_{LUMO}$. Assuming that the molecule is taken as the electrostatic ground, then $\mu_S = E_{LUMO} = U_C$ for this region of operation. Calculating the current from Eq. (3.29) gives

$$I_{DS} = \frac{qN}{\tau_D} \tag{3.38}$$

Then, given that $U_C = q^2 N / C_{ES}$, we can rearrange Eq. (3.38) to get

$$I_{DS} = \frac{C_{ES} U_C}{q\tau_D} \tag{3.39}$$

Then, from Eqs. (3.18) and (3.19) and noting that $\mu_S = E_{LUMO} = U_C$,

$$I_{DS} = \frac{C_{ES}}{\tau_D} \eta \left(V_{DS} - V_{ON}\right). \tag{3.40}$$

This region is valid for voltages

$$0 < \left(V_{DS} - V_{ON}\right) < U_C^{max} / \eta q\,. \tag{3.41}$$

The full *IV* characteristic is shown below. Under our assumptions the transitions between the three regions of operation are sharp. For $T > 0$K and $\left(\Gamma = \hbar/\tau\right) > 0$ these transitions are blurred and are best calculated numerically; see the Problem Set.

Part 3. Two Terminal Quantum Dot Devices

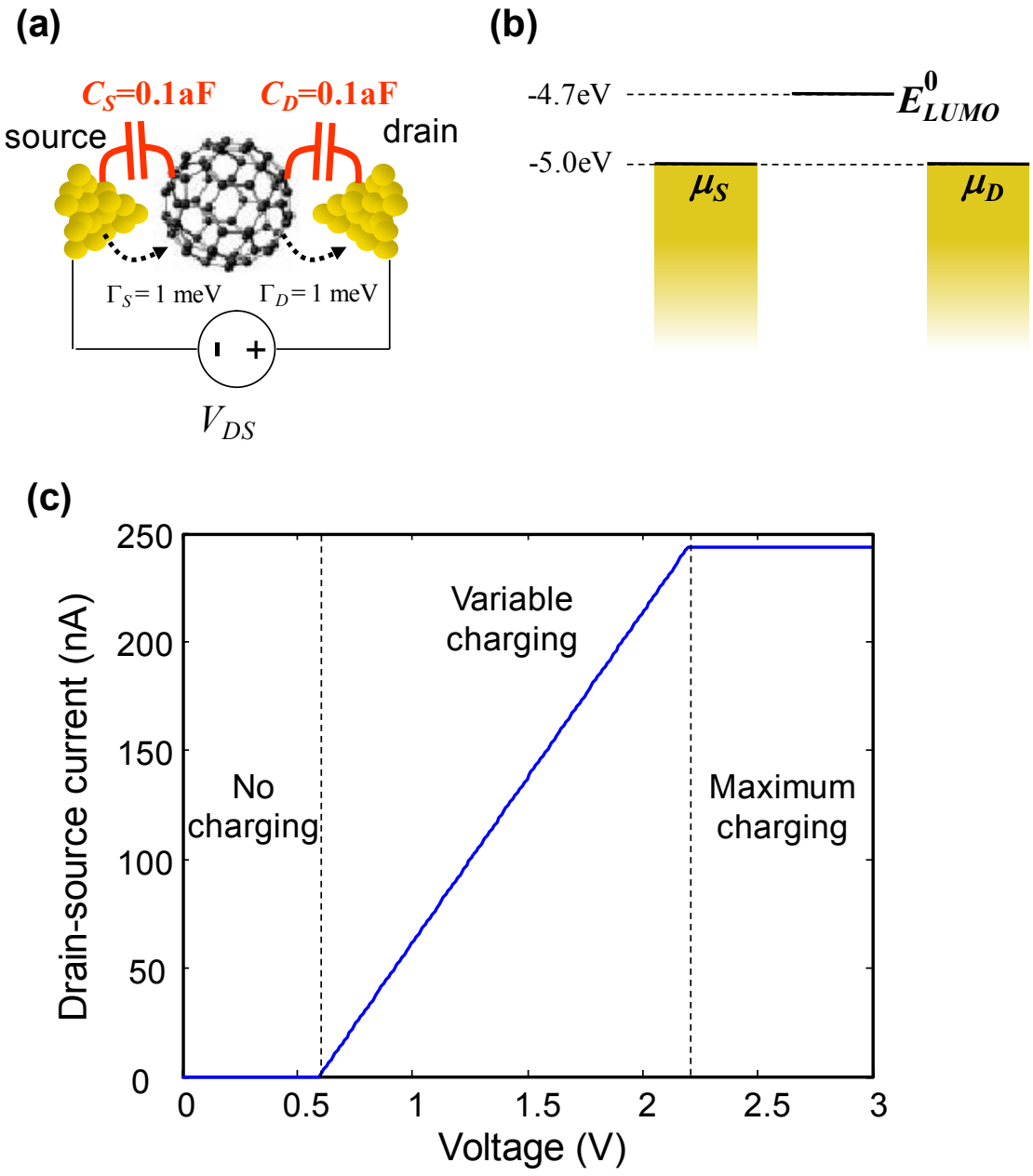


Fig. 3.24. (a) An example of a single molecule device subject to strong charging effects. The coupling between the molecule and the contacts is small relative to the applied voltage, *i.e.* Γ_S/q = Γ_D/q = 1mV. The source and drain capacitances yield a large charging potential per electron q^2/C_{ES} = 0.8 eV. The voltage division factor is η = 0.5. **(b)** The offset between the LUMO and the contact work functions is 0.3 eV, consequently, V_{ON} = 0.3/η = 0.6V **(c)** The current-voltage characteristic.

A small signal circuit model

In the discussion of the establishment of equilibrium between a contact and a molecule we introduced a generalized circuit model where each node potential is the Fermi level, not the electrostatic potential as in a conventional electrical circuit.

We can extend the model to two terminal, and even three terminal devices. It must be emphasized, however, that the model is only valid for small signals. In particular, the model is constrained to small V_{DS}. We assume that the density of states is constant and the modulation in V_{DS} must be smaller than kT/q so that we can ignore the tails of the Fermi distribution.

Let"s consider current injected by the source

$$I_S = \frac{q}{\tau_S}\left(N_S - N\right) \tag{3.42}$$

This can be rewritten as

$$I_S = \frac{q}{\tau_S}\int_{-\infty}^{+\infty} g\left(E-U\right)\left(f\left(E,\mu_S\right)-f\left(E,E_F\right)\right)dE \tag{3.43}$$

For small differences between the source and drain potentials, and at $T = 0$K, we get

$$I_S = \frac{C_Q}{\tau_S}\frac{\left(\mu_S - E_F\right)}{q} \tag{3.44}$$

Thus, each contact/molecule interface is Ohmic in the small signal limit. Defining $R_S = \tau_S/C_Q$, and $R_D = \tau_D/C_Q$.We can model the contact/molecule/contact as shown in Fig. 3.25.

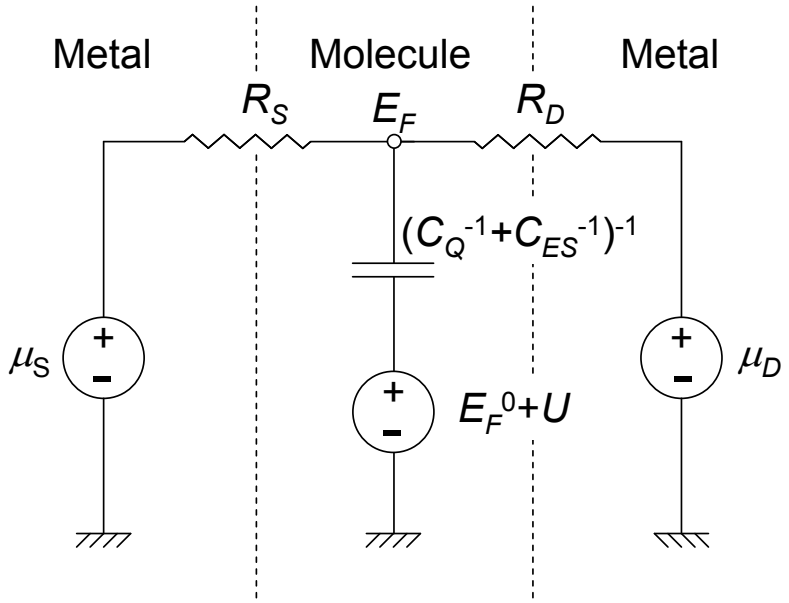


Fig. 3.25. A small signal model for two terminal metal/molecule/metal circuits. Note that the potential U must be determined separately (e.g. by using a capacitive divider circuit).

Part 3. Two Terminal Quantum Dot Devices

The Ideal Contact Limit[†]

Interfaces between molecules and contacts vary widely in quality. Much depends on how close we can bring the molecule to the contact surface. Here, we have modeled the source and drain interfaces with the parameters τ_S and τ_D. If electron injection is unencumbered by barriers or defects then these lifetimes will be very short. We might expect that the current should increase indefinitely as the injection rates decrease. But in fact we find a limit – known as the quantum limit of conductance. We will examine this limit rigorously in the next section but for the moment, we will demonstrate that it also holds in this system.

We model ideal contacts by considering the current under the limit that $\tau_S = \tau_D \to 0$. Note that the uncertainty principle requires that the uncertainty in energy must increase if the lifetime of an electron on the molecule decreases. Thus, the density of states must change as the lifetime of an electron on a molecule changes.

Let"s assume that the energy level in the isolated molecule is discrete. In Part 2, we found a Lorentzian density of states for a single molecular orbital with net decay rate $\tau_S^{-1} + \tau_D^{-1}$:

$$g(E-U)dE = \frac{2}{\pi}\frac{(\hbar/2\tau_S + \hbar/2\tau_D)}{(E-U-E_0)^2 + (\hbar/2\tau_S + \hbar/2\tau_D)^2}dE \tag{3.45}$$

If we take the limit, we find that the molecular density of states is uniform in energy:

$$\lim \tau_S = \tau_D \to 0 \quad g(E-U)dE = \frac{8}{h}\frac{1}{1/\tau_S + 1/\tau_D}dE \tag{3.46}$$

Substituting into Eq. (3.30) for $\tau_S = \tau_D$ gives

$$I = \frac{2q}{h}\int_{-\infty}^{+\infty} f(E,\mu_S) - f(E,\mu_D)dE \tag{3.47}$$

At $T = 0$K,

$$f(E,\mu) = u(\mu - E) \tag{3.48}$$

where u is the unit step function, and the integral in Eq. (3.47) gives

$$I = \frac{2q}{h}(\mu_S - \mu_D) \tag{3.49}$$

Note $-qV_{DS} = (\mu_D - \mu_S)$, thus the conductance through a single molecular orbital is

$$G = \frac{2q^2}{h} \tag{3.50}$$

The equivalent resistance is G^{-1}= 12.9 kΩ. Thus, even for ideal contacts, this structure is resistive. We will see this expression again in the next section. It is the famous quantum limited conductance.

[†] This derivation of the quantum limit of conductance is due to S. Datta, „Quantum transport: atom to transistor" Cambridge University Press (2005).

Problems

1. (a) A 1nm × 1nm molecule is 30Å away from a metal contact. Calculate the electrostatic capacitance using the parallel plate model of the capacitor. Find the change in potential per charge added to the molecule, $U_{ES}/\delta n$.

(b) A 50nm × 50nm molecule is 30Å away from a metal contact. Calculate the electrostatic capacitance using the parallel plate model of the capacitor. Find the change in potential per charge added to the molecule, $U_{ES}/\delta n$.

(c) A 100nm × 100nm molecule is 30Å away from a metal contact. Calculate the electrostatic capacitance using the parallel plate model of the capacitor. Find the change in potential per charge added to the molecule, $U_{ES}/\delta n$.

2.(a) Assume the molecule in problem 1(a) has a uniform density of states of $g(E) = 3\times10^{20}$/eV and Fermi level at E_F^0 = -5.7eV in isolated space. The metal has a work function of 5eV. Sketch all of the energy levels in equilibrium after the metal contact and molecule are brought into contact. Find the number of charges, δ_n, transferred from the molecule to the metal (or vice versa.).

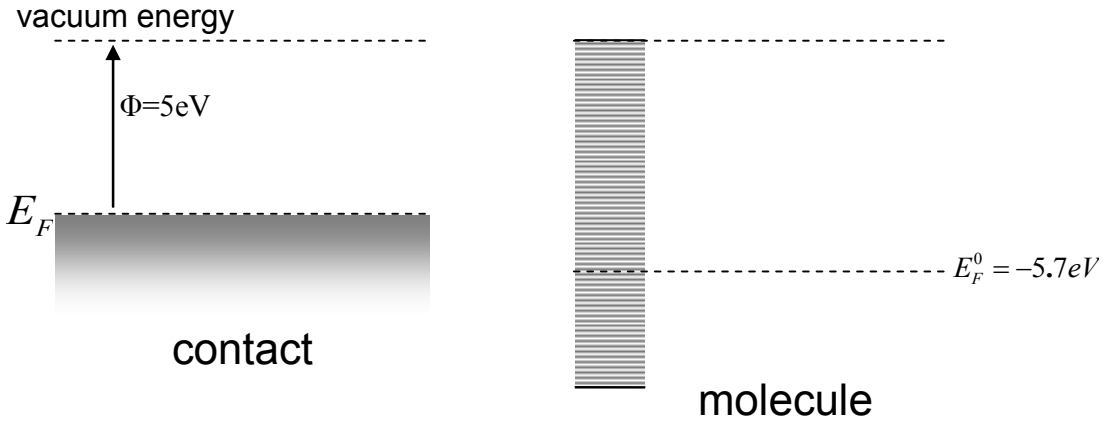


Fig. 3.26. A separated contact and molecule. The molecule has a uniform density of states.

(b) Repeat part (a) for the molecule in problem 1(b) using the same density of states and Fermi levels.

(c) Repeat part (a) for the molecule in problem 1(c) using the same density of states and Fermi levels.

Part 3. Two Terminal Quantum Dot Devices

3. Consider the molecule illustrated below with E_A = 2eV, I_P = 5eV, and E_F^0 = 3.5eV.

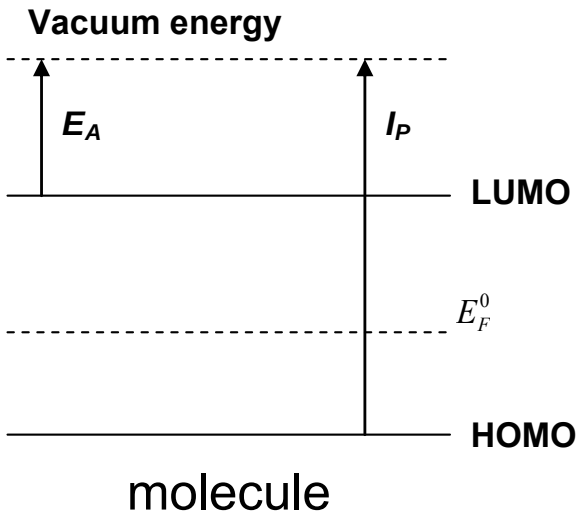


Fig. 3.27. Energy levels within a molecule.

(a) What is C_Q when the electron lifetime is $\tau = \infty, 1ps, \ 1fs$?

(b) For each of the lifetimes in part (a), what is the equilibrium Fermi level when $\delta n = 0.1$ charge is added to the molecule.

(c) What is the equilibrium Fermi level when $\tau_{HOMO} = .1ps \ and \ \tau_{LUMO} = 1ps$?

4. A quantum well is brought into contact with a metal electrode as shown in the figure below.

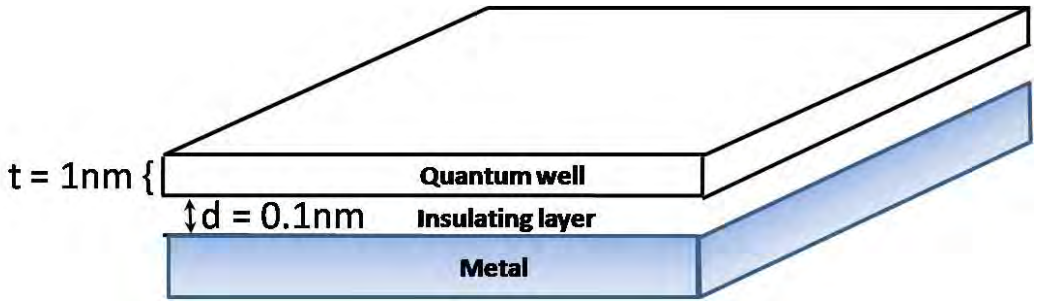


Fig. 3.28. The quantum well in contact with a metal surface.

(a) Calculate the DOS in the well between 0 and 2eV assuming an infinite confining potential and that the potential inside the well is zero.

Continued….

(b) On contact charge can flow between the metal and the quantum well. But assume that on contact the well is still separated from the metal by a 0.1nm thick layer with dielectric constant $5\times8.84\times10^{-12}$ F/m. Calculate the surface charge density at equilibrium, for an initial separation of **(i)** 0.6eV and **(ii)** -0.6eV, between the quantum well E_F and the chemical potential of the metal. Assume T=0.

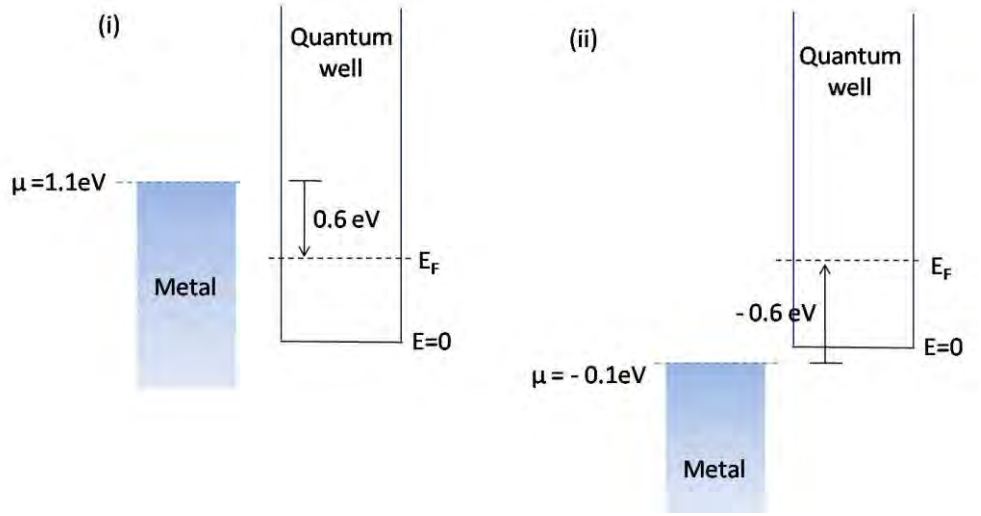


Fig. 3.29. Two different energetic alignments between the quantum well and the metal.

(c) Plot the vacuum energy shift at each interface at equilibrium, for part **b (i)** and **(ii)**, above.

The next question is adapted from an example in „Introductory Applied Quantum and Statistical Mechanics" by Hagelstein, Senturia and Orlando, Wiley Interscience 2004.

5. In this problem we consider charge injection from a *discrete energy level* rather than a metal. Consider charge transport through a quantum dot buried within an insulator. The materials are GaAs|GaAlAs|GaAs|GaAlAs|GaAs with thicknesses 1000Å|40Å|100Å|40Å|1000Å. The potential landscape of this device is modeled as below with V_0=0.3 eV, b=50Å, and a=90Å. Let the effective mass of an electron in GaAs and GaAlAs be $0.07m_e$, where m_e is the mass of an electron.

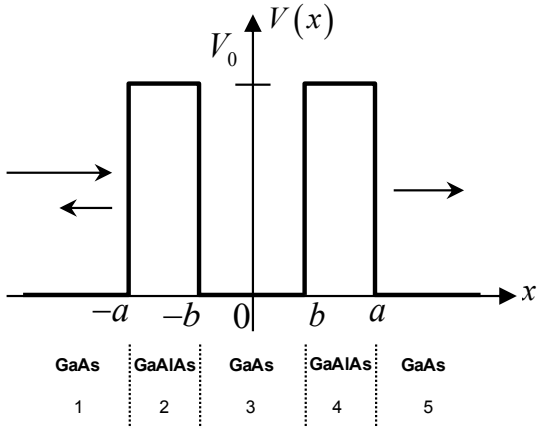


Fig. 3.30. The energy levels for a quantum dot buried within a tunnel barrier.

Part 3. Two Terminal Quantum Dot Devices

Considering only energies below V_0, the wavefunction is piecewise continuous with

$$\begin{cases} \psi_1 = e^{ikx} + Be^{-ikx} \\ \psi_2 = Ce^{-\alpha x} + De^{\alpha x} \\ \psi_3 = Fe^{ikx} + Ge^{-ikx} \\ \psi_4 = He^{-\alpha x} + Ie^{\alpha x} \\ \psi_5 = Je^{ikx} \end{cases}, \text{ where } k = \frac{\sqrt{2mE}}{\hbar} \text{ and } \alpha = \frac{\sqrt{2m(V_0 - E)}}{\hbar}$$

(a) Match the boundary conditions and find $\overline{\overline{M}}$ such that $\overline{\overline{M}}\,\overline{C} = \overline{A}$, where

$$\overline{C} = \begin{pmatrix} B \\ C \\ D \\ F \\ G \\ H \\ I \\ J \end{pmatrix} \text{ and } \overline{A} = \begin{pmatrix} e^{-ika} \\ ike^{-ika} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

(b) The transmission coefficient is $T = |J|^2 = \left|\overline{C}(8)\right|^2$, where $\overline{C} = \overline{\overline{M}}^{-1}\overline{A}$. Determine T numerically or otherwise. It is plotted below as a function of electron energy, E. Verify that the resonant energies where $T = 1$ are within a factor of two of the eigenenergies of an infinite square well with width $L=2b$. Note that this approximation for the resonant energies works better for deeper square wells (V_0 big).

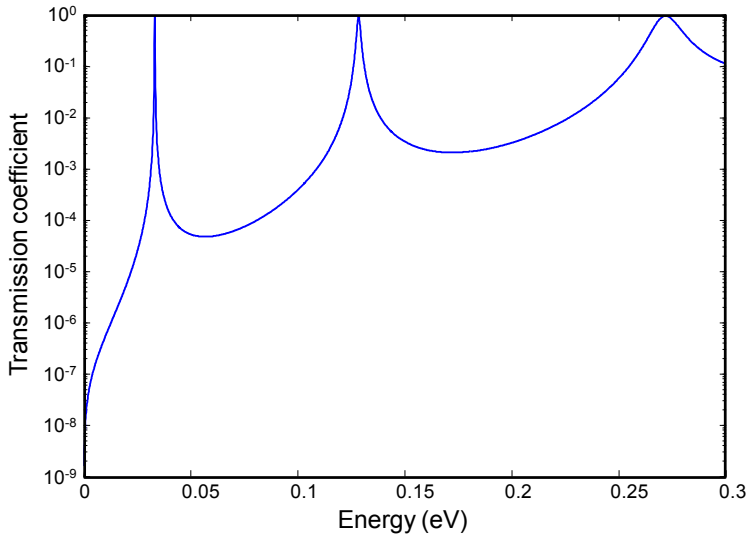


Fig. 3.31. Your solution to part (b) should look like this.

(c) Why do the widths of the resonances in the transmission coefficient increase at higher electron energy?
(d) Now suppose we apply a voltage across this device. Electrons at the bottom of the conduction band E_C in the GaAs on the left side will give net current flow which is proportional to the transmission coefficient T.

Let‟s first consider only one discrete energy level E_0 in the dot as shown in part (a) of the figure below. Assume the potential drop is linear (electric field F is constant) across the whole device, as shown in part (b) of the figure below.

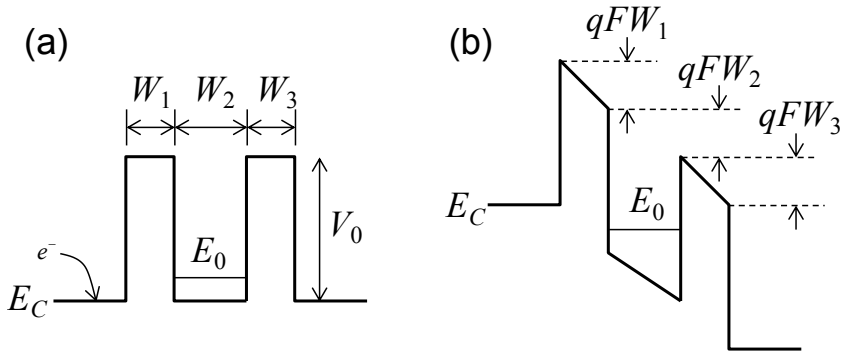


Fig. 3.32. (a) The Quantum Dot structure at zero bias, and **(b)** under an applied bias.

Sketch out qualitatively what you think the current-vs-voltage curve will look like. It should look quite different to the *IV* characteristic of a quantum dot with metal contacts. Explain the difference.

(e) Approximating E_0 as the ground state of an infinite square well, what is the expression for the resonant voltage in terms of W_2, assuming $W_1=W_3$?

(f) Without solving for the wavefunction, sketch qualitatively what the probability density of the lowest eigenstate of an infinite well will look like when distorted by an electric field. Where in the well has the highest probability of finding an electron?

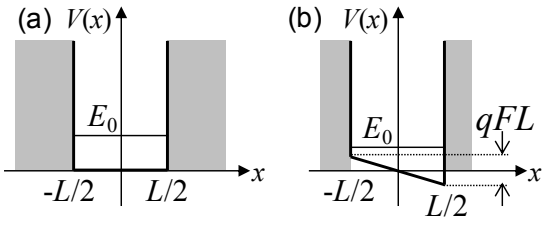


Fig. 3.33. The quantum dot under bias.

(g) Now, if we consider the multiple discrete energy levels in the dot, what will the current-voltage curve look like?

(h) Analytically determine the current through the dot at the 0.27eV resonance. Hint: consider the width of the resonance.

Part 3. Two Terminal Quantum Dot Devices

6. Consider the two terminal molecular device shown below. Note that this calculation differs from the previous calculation of charging by considering transport through the HOMO rather than the LUMO.

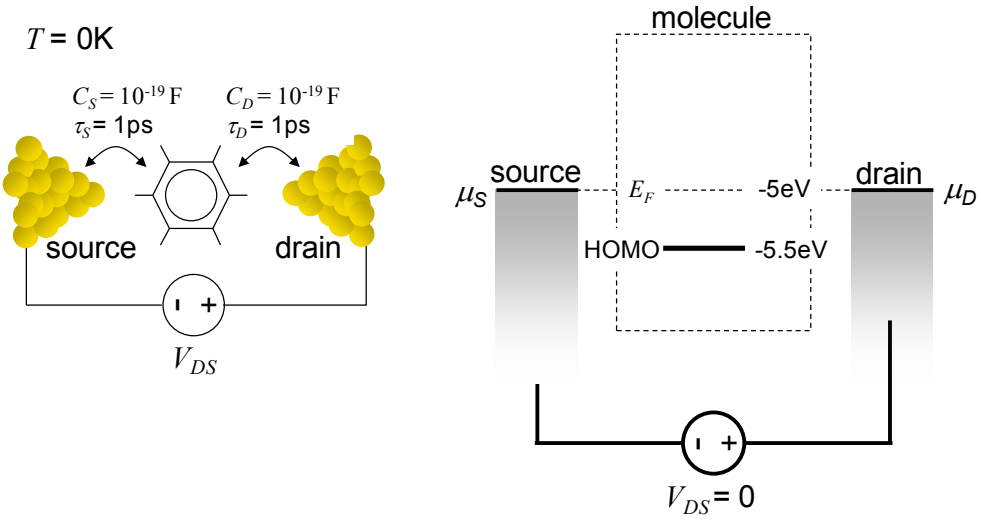


Fig. 3.34. The device structure for an analytical calculation of charging effects.

(a) Estimate the width of the HOMO from τ_S and τ_D.

(b) Assuming that the molecule can be modeled as a point source conductor of radius 2nm, calculate the charging energy per electron. Compare to the charging energy determined from the capacitance values shown in Fig. 3.34.

(c) Assuming that the charging energy is negligible, calculate the I_{DS}-V_{DS} characteristic and plot it.

(d) Now consider charging with q^2/C_{ES} = 1eV. How does charging alter the maximum current and turn on voltage (the lowest value of V_{DS} when current flows)?

(e) Show that the number of electrons on the HOMO is at *least*

$$N = 2\frac{\tau_D}{\tau_S + \tau_D}$$

(f) What is the maximum charging energy when q^2/C_{ES} = 1eV?

(g) Assuming that the charging energy is negligible, plot the energy level of the HOMO together with the source and drain workfunctions for V_{DS} = 2V. On the same plot, indicate the energy level of the HOMO when q^2/C_{ES} = 1eV and V_{DS} = 2V. What is the charging energy at this bias?

(h) Calculate the I_{DS}-V_{DS} characteristic when q^2/C_{ES} = 1eV. Plot it on top of the I_{DS}-V_{DS} characteristic calculated for negligible charging.

7. Consider the two terminal molecular device shown in the figure below. This question considers conduction through both the HOMO and LUMO as well as the effect of mismatched source and drain injection rates and capacitances.

Note that: $T = 0\text{K}$; $C_S = 2C_D$; $\eta_S = 1\text{ps}$; $\eta_D = 9\text{ps}$

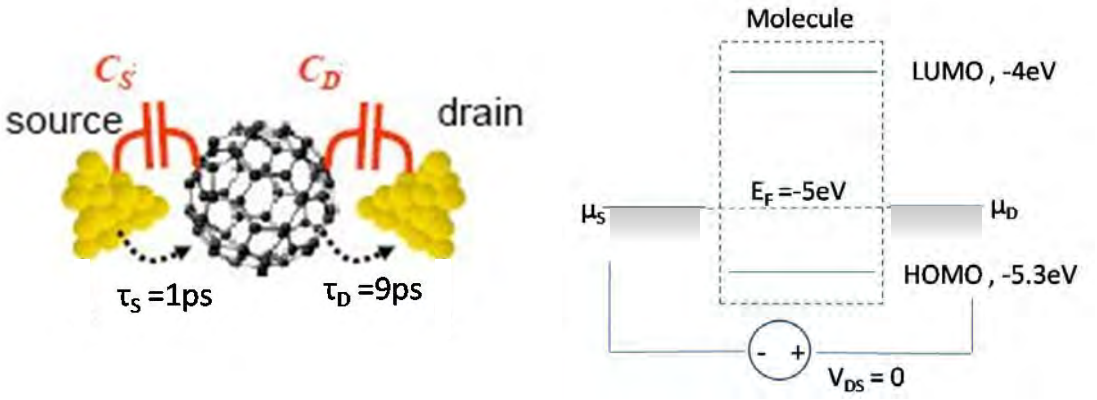


Fig. 3.35. A two terminal molecular device.

(a) Estimate the width of the HOMO and LUMO from η_S and η_D.

(b) If C_D = 1pF calculate the charging energy per electron.

(c) Plot the current-voltage characteristic (IV) from V_{DS} = -10V to V_{DS} = 10V assuming that the charging energy equals zero.

(d) Assume the charging energy is now 1eV per electron. What is C_D?

(e) Plot the IV from V_{DS} = 0V to V_{DS} = 10V assuming that the charging energy is 1eV per electron.

(f) Plot the IV from V_{DS} = -10V to V_{DS} = 0V assuming that the charging energy is 1eV per electron. **Hint:** you should find a region of this IV characteristic in which the HOMO and LUMO are charging together, increasing the current but leaving the net charge on the molecule unchanged. Consequently, your IV characteristic should exhibit a step change in current at a particular voltage.

Part 3. Two Terminal Quantum Dot Devices

8. A quantum wire with square cross-section with a 1-nm thickness side is bent and fused into a circular ring with radius R= 2nm as shown below.

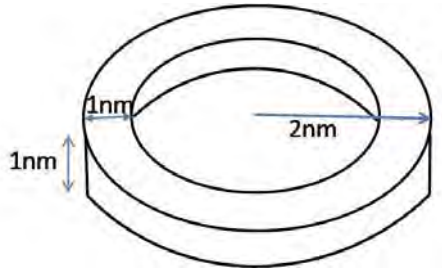


Fig. 3.36. A quantum wire bent into a ring.

(a) Plot the DOS in the wire from E = 0 to E = 0.8eV. Assume an infinite confining potential and that the potential in the ring is V = 0.

(b) Next the ring is placed between contacts as shown. What is the charging energy?

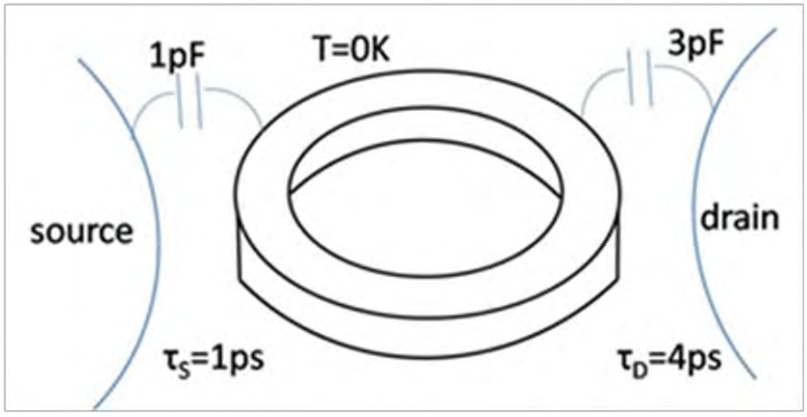


Fig. 3.37. The ring between contacts.

(c) Plot the current versus voltage for V from 0 to 1.1V.

(d) A magnetic field is applied perpendicular to the ring. Sketch the changes in the IV.

The next three questions have been adapted from F. Zahid, M. Paulsson, and S. Datta, „Electrical conduction in molecules". In *Advanced Semiconductors and Organic Nanotechniques*, ed. H. Korkoc. Academic Press (2003).

9. (a) Numerically calculate the current-voltage and conductance-voltage characteristics for the system shown in Fig. 3.38 with the following parameters:

$$\eta = 0.5$$
$$E_F = \text{-}5.0 \text{ eV}$$
$$\text{HOMO} = \text{-}5.5 \text{ eV}$$
$$\Gamma_S = \Gamma_D = 0.1 \text{ eV}$$
$$T = 298\text{K}$$

Set the charging energy to zero, i.e. take $C_{ES} \to \infty$.

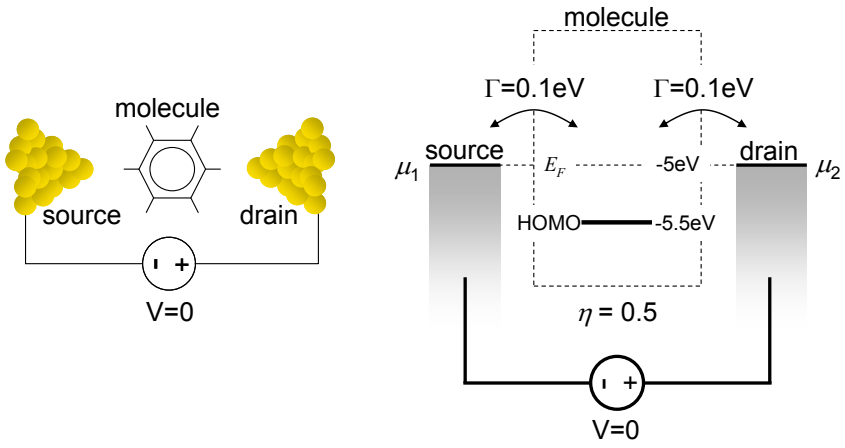


Fig. 3.38. The two terminal molecular device for this problem.

Hints

(1) Despite the statement that $\Gamma_S = \Gamma_D = 0.1$ eV, the HOMO in this problem is assumed to be infinitely sharp. Simplify Eqns. (3.30) and (3.31) for $g(E\text{-}U) = 2\delta(E\text{-}U\text{-}\varepsilon)$, where ε is the energy of the HOMO.

(2) You will need to implement the flow chart shown in Fig. 3.21. If your solution for U oscillates and does not converge, try setting

$$U = U_{old} + \alpha\left(U_{calc} - U_{old}\right),$$

where U_{calc} is the solution to Eq. (3.21), U_{old} is the previous iteration's estimate of U and α is a small number that may be reduced to obtain convergence.

(b) Repeat the numerical calculation of part (a) with $q^2/C_{ES} = 1$ eV.

(c) Explain the origin of the conductance gap. What determines its magnitude?

(d) Write an analytic expression for the maximum current when $q^2/C_{ES} = 0$.

(e) Explain why the conductance is much lower when the charging energy is non-zero.

Part 3. Two Terminal Quantum Dot Devices

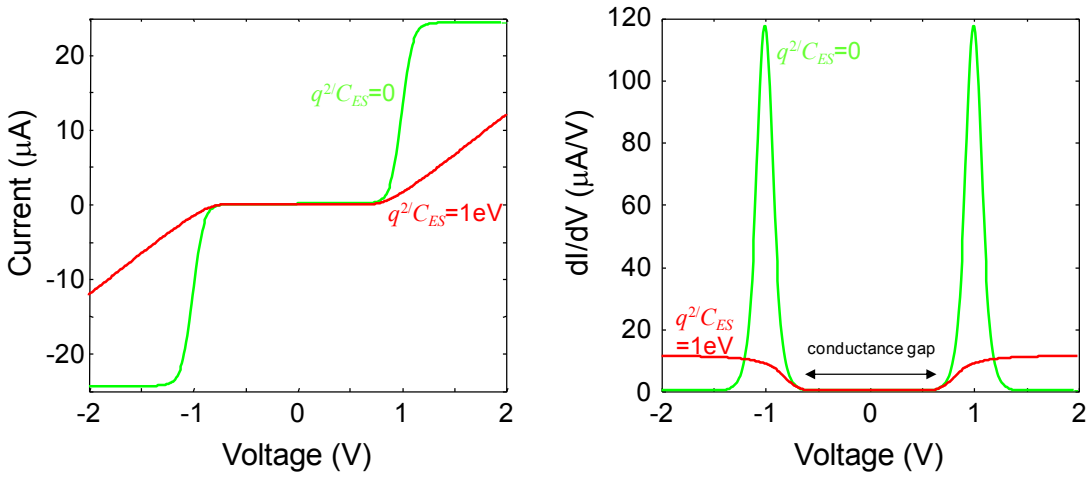


Fig. 3.39. Your solution should look like this.

10. Next, we add a LUMO level at -1.5 eV.

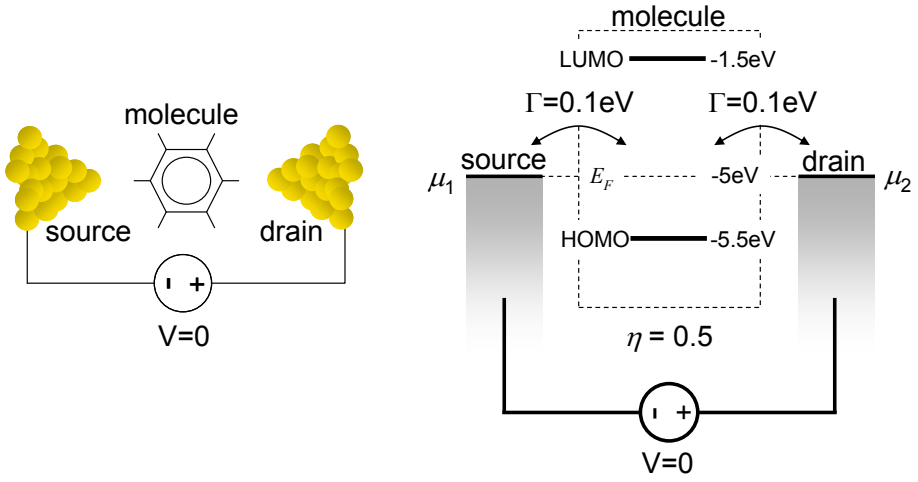


Fig. 3.40. The two terminal molecular device for this problem.

(a) Numerically calculate the current-voltage and conductance-voltage characteristics for $q^2/C_{ES} = 1$ eV and $E_F = -2.5$ eV.

(b) Repeat the numerical calculation for $q^2/C_{ES} = 1$ eV and $E_F = -3.5$ eV.

(c) Repeat the numerical calculation for q^2/C_{ES} = 1 eV and E_F = -5.0 eV (same as Q8.b).

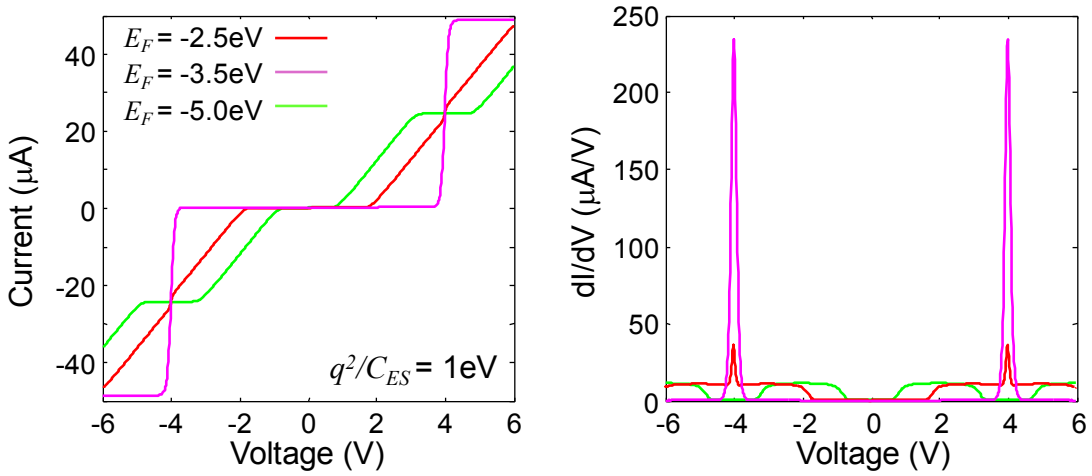


Fig. 3.41. Your solution should look like this.

(d) Why are the current-voltage characteristics uniform? Hint: what would happen if $\eta \neq 0.5$? Identify the origin of the transitions in the IV curve.

(e) Why is the effect of charging absent when E_F = -3.5 eV?

11. Next, we consider a Lorentzian density of states rather than simply a discrete level.

$$g(E)dE = \frac{1}{\pi}\frac{\Gamma}{(E-\varepsilon)^2 + (\Gamma/2)^2}dE \tag{51}$$

where $\Gamma = \Gamma_S + \Gamma_D$ and ε is the center of the HOMO. Ignore the LUMO, *i.e.* consider the system from problem 9.

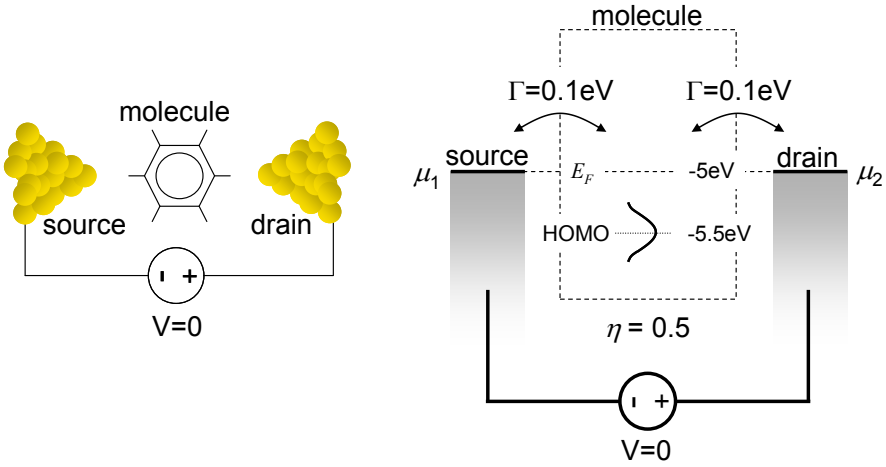


Fig. 3.42. The two terminal molecular device for this problem.

Part 3. Two Terminal Quantum Dot Devices

(a) Numerically compare the current-voltage and conductance-voltage characteristics for a discrete and broadened HOMO with $q^2/C_{ES} = 1$ eV.

(b) How would you expect the IV to change if $\Gamma_S > \Gamma_D$? Explain.

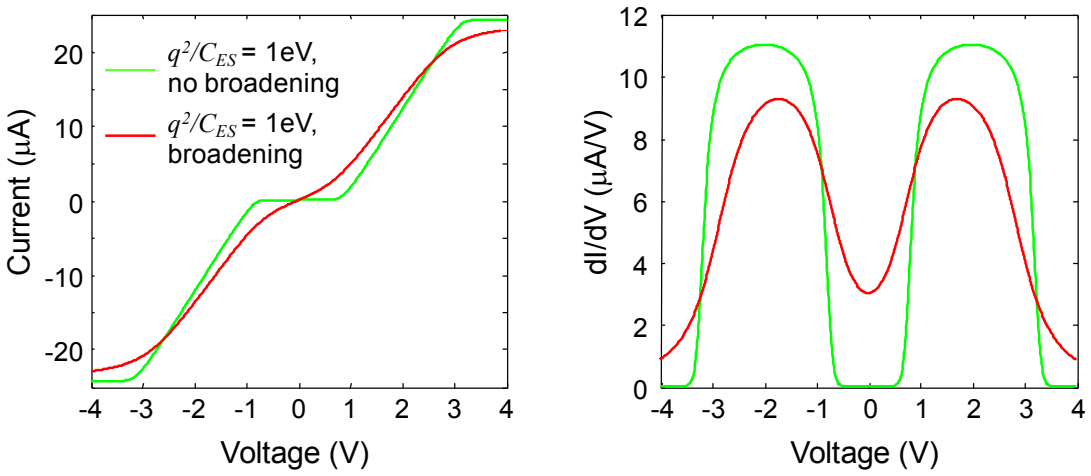


Fig. 3.43. Your solution should look like this.

12. The following problem considers a 2-terminal conductor under illumination. The light produces an electron transfer rate of αN_H from the HOMO to the LUMO. The light also causes an electron transfer rate of αN_L from the LUMO to the HOMO, where α is proportional to the intensity of the illumination, and the electron populations in the LUMO and HOMO are N_L and N_H, respectively.

Assume $C_S = C_D$, that the LUMO and HOMO are delta functions, and $T = 300$K. Also, assume that under equilibrium in the dark, the Fermi Energy is midway between the HOMO and LUMO.

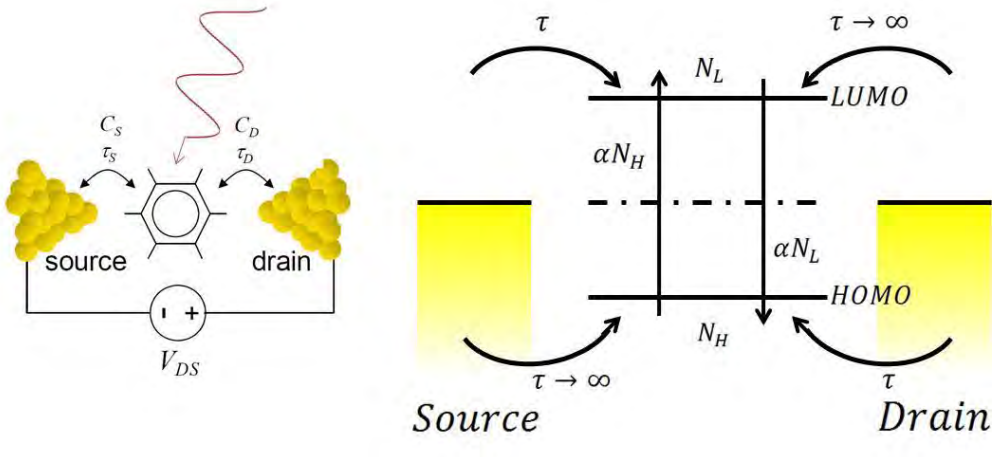


Fig. 3.44. A model of a single molecule solar cell.

Next, imagine that the contacts are engineered to have the following characteristics:
Transfer rate between Drain and HOMO = $1/\tau$.
Transfer rate between Drain and LUMO = 0.
Transfer rate between Source and HOMO = 0.
Transfer rate between Source and LUMO = $1/\tau$.

(a) Determine the short circuit current for this system. (*i.e.* let $V_{DS} = 0$, what is the current that flows through the external short circuit?)

(b) Determine the open circuit voltage for this system. (*i.e.* Disconnect the voltage source, what is the voltage that appears across the terminals of the molecule?)

(c) Repeat **(a)** and **(b)** with the addition of an additional electron transfer rate βN_L from the LUMO to the HOMO, where β is independent of the intensity of the illumination.

13. This problem refers to the 2 terminal molecular conductor below.

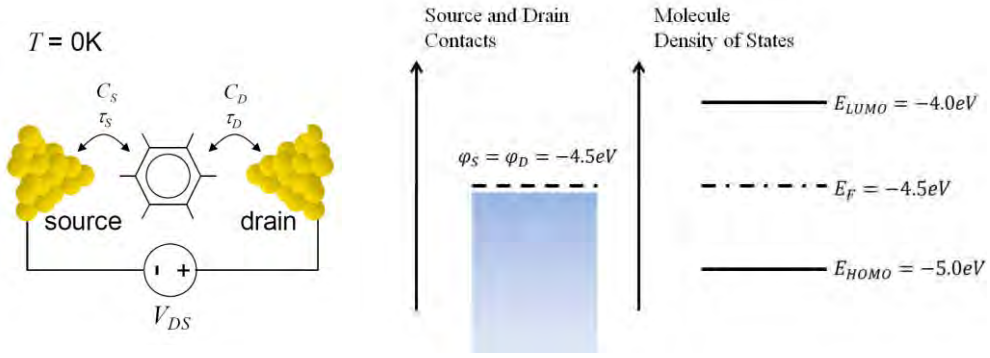


Fig. 3.45 Equilibrium between a molecule and a contact requires charge transfer.

(a) When $\tau_S = 10$ fs, $\tau_D = 5$ fs, calculate the actual molecular density of states versus energy. Determine the full width half maximum of HOMO and LUMO.

(b) Based on the actual density of states calculated in part c), find the number of electrons and the charging energy when the molecule is brought into contact with the metal electrode and reached equilibrium (no applied voltage). Also sketch the energy diagram at equilibrium. Assume that the charging energy per electron is 1eV and $\tau_S = 10$ fs, $\tau_D = 5$ fs.

Hint: You will need your calculator to solve this. You might use $\int \frac{1}{1+x^2} dx = \tan^{-1}(x)$.

Part 4. Two Terminal Quantum Wire Devices

Let"s consider a quantum wire between two contacts. As we saw in Part 2, a quantum wire is a one-dimensional conductor. Here, we will assume that the wire has the same geometry as studied in Part 2: a rectangular cross section with area $L_x.L_y$. Electrons are confined by an infinite potential outside the wire, and can only flow along its length; arbitrarily chosen as the z-axis in Fig. 4.1.

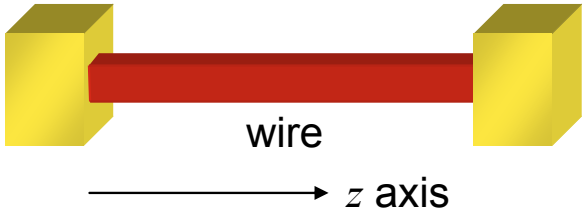


Fig. 4.1. A quantum wire between two contacts.

Under these assumptions, if we model the electrons by plane waves in the z direction we get

$$E = \frac{\pi^2\hbar^2}{2m}\left(\frac{n_x^2}{L_x^2}+\frac{n_y^2}{L_y^2}\right)+\frac{\hbar^2 k_z^{\,2}}{2m}, \quad n_x, n_y = 1,2,... \tag{4.1}$$

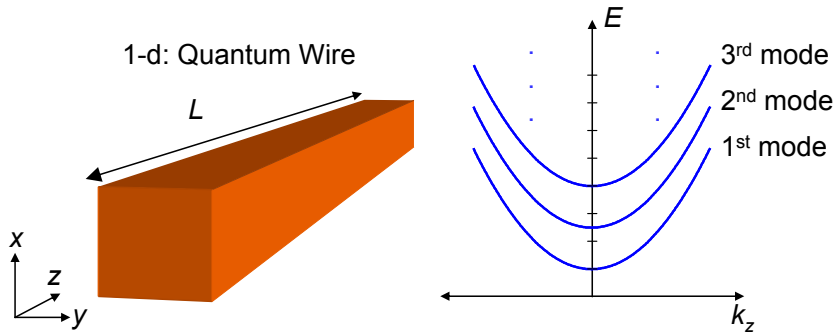


Fig. 4.2. Plane waves in a quantum wire have parabolic energy bands.

Recall that for current to flow there must be difference in the number of electrons in $+k_z$ and $-k_z$ states. As in Part 2, we define two *quasi* Fermi levels: F^+ for states with $k_z > 0$, F^- for states with $k_z < 0$. Thus, current flows when electrons traveling in the $+z$ direction are in equilibrium with each other, but not with electrons traveling in the $-z$ direction. For example, in Fig. 4.3, current is carried by the uncompensated electrons in the $+k_z$ states.

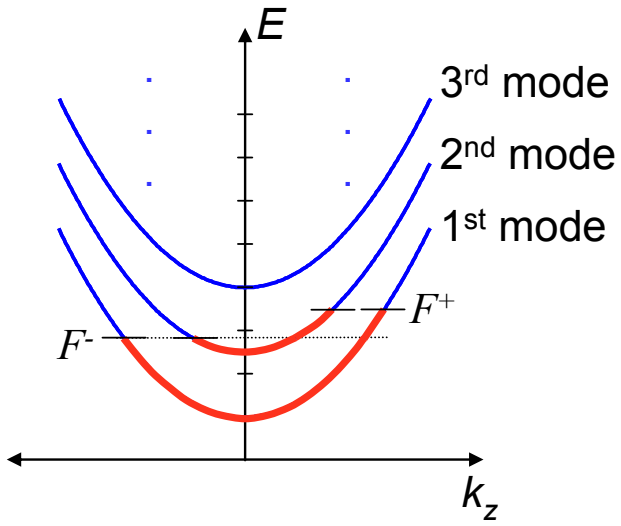


Fig. 4.3. Current flows when the quasi Fermi levels differ for $+k_z$ and $-k_z$. states.

Scattering and Ballistic Transport

Next, let‟s assume that electrons travel in the wire without scattering, *i.e.* the electrons do not collide with anything in the wire that changes their energy or momentum. This is known as „ballistic‟ transport - the electron behaves like a projectile traveling through the conductor.

Electron scattering is usually caused by interactions between electrons and the nuclei. The probability of an electron collision is enhanced by defects and temperature (since the vibration of nuclei increases with temperature). Thus, the scattering rate can be decreased by lowering the temperature, and working with very pure materials. But all materials have some scattering probability. So, the smaller the conductor, the greater the probability that charge transport will be ballistic. Thus, ballistic transport is a nanoscale phenomenon and can be engineered in nanodevices.

For ballistic transport the electron has no interaction with the conductor. Thus, *the electron is not necessarily in equilibrium with the conductor*, i.e. the electron is not restricted to the lowest unoccupied energy states within the conductor.

But scattering can bump electrons from high energy states down to lower energies. There are two categories of scattering: elastic, where the scattering event may change the momentum of the electron but its energy remains constant; and inelastic, where the energy of the electron is not conserved. Equilibrium may be established by inelastic scattering.

Electron scattering is the mechanism underlying classical resistance. We will spend a lot more time on this topic later in this part.

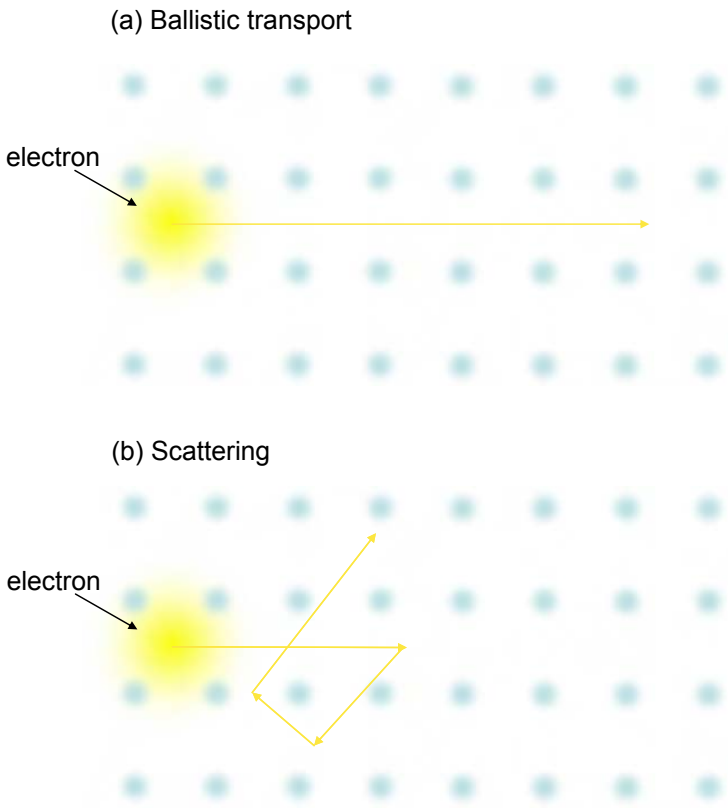


Fig. 4.4. Here, we represent an electron traveling through a regular lattice of nuclei. If the electron travels ballistically it has no interaction with the lattice. It travels with a constant energy and momentum and will not necessarily be in equilibrium with the material. If, however, the electron is scattered by the lattice, then both its energy and momentum will change. Scattering assists the establishment of equilibrium within the material.

Equilibrium between contacts and the conductor

When the contact is connected with the wire, equilibrium must be established. For example, if F is higher than the chemical potential of the contact, μ, then electrons will diffuse from the wire into empty states in the contact. This is known as depletion. If the electrons are not replenished from another source, the loss of electrons lowers the Fermi level within the wire. In addition, since the wire has lost negative charge, it becomes positively charged, establishing an electric field that counteracts the diffusion of electrons out of the wire. Ultimately equilibrium is established when the Fermi level in the wire equals the chemical potential of the contact. If all the electrons diffuse out of the wire then the wire is said to be fully depleted.

If the chemical potential of the contact is higher than the Fermi level in the wire. Electrons diffuse into the wire from the contact, raising the Fermi level. This is known as accumulation. The addition of negative charge also establishes an electric field that counteracts the diffusion of electrons from the contact to the wire.

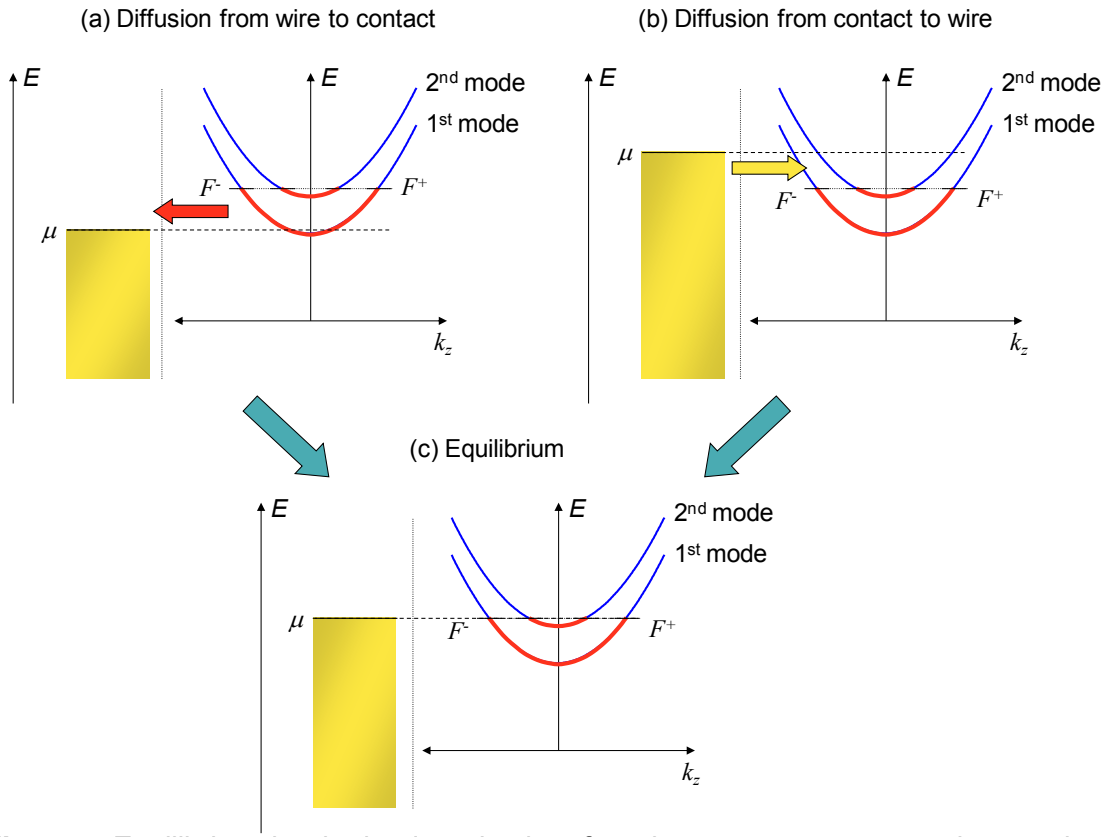


Fig. 4.5. Equilibrium is obtained at the interface between a contact and a conductor when the diffusion currents in and out of the conductor match.

Bias

Now, what happens when a voltage is applied between the contacts? Recall that applying voltage shifts the relative potential energies of each contact, i.e. μ_D-μ_S=-qV_{DS}, where the chemical potential of the source is μ_S and the chemical potential of the drain is μ_D.

As in the equilibrium case, charges flow from each contact, ballistically through the conductor and into the other contact. Thus, all states with $k_z > 0$ are injected by the source and have no relation with the drain. Similarly, electrons with $k_z < 0$ are injected by drain.

But now the injected currents do not balance. Conductor states in the energy range between μ_S and μ_D are uncompensated and only be filled by the source, yielding an electron current flowing from source to drain in Fig. 4.6.

The quasi Fermi level for electrons with $k_z > 0$, F^+ must equal the electrochemical potential of the left contact, *i.e.*

$$F^+ = \mu_S.$$

Similarly,

$$F^- = \mu_D.$$

Thus, current can only flow when there is a difference between the chemical potentials of the contacts. This shouldn"t be surprising, since the difference between chemical potentials is simply related to the voltage by μ_D-μ_S=-qV_{DS}.

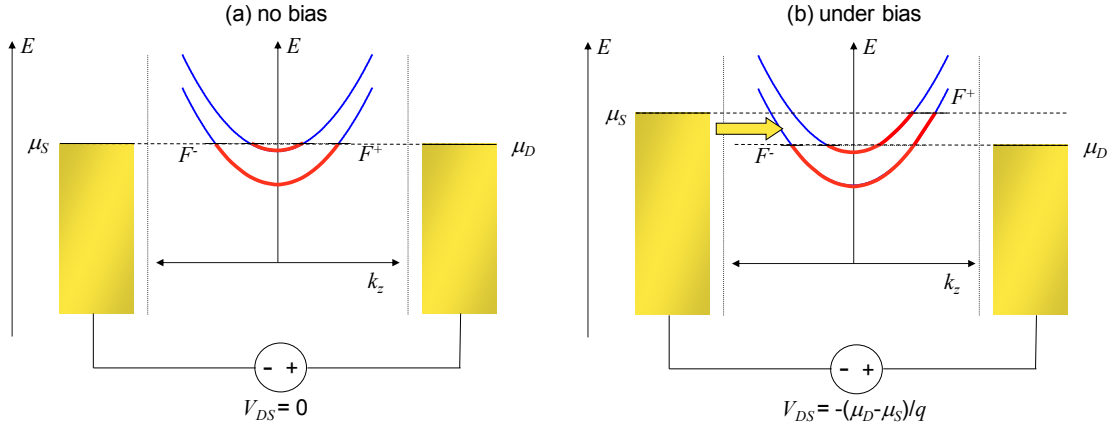


Fig. 4.6. Under bias the Fermi Levels of each contact shift. Diffusion from states in the contact with the higher potential causes a current.

The Spatial Profile of the Potential

As shown in Fig. 4.7, below, the wire may be described by its dispersion relation or by its density of states (DOS). The dispersion relation describes a band of conducting states. The bottom of the band is known as the conduction band edge. It corresponds to the lowest energy for a plane wave state in the wire. The conduction band edge is particularly important because its position controls the current flow in the wire. If it is below the source work function then electrons are readily injected into the wire. In contrast, if the conduction band edge is above the source workfunction, then current flow requires electrons with additional thermal energy.

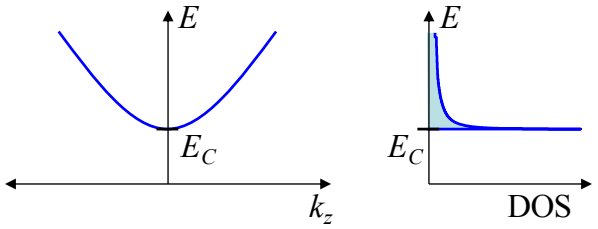


Fig. 4.7. Two representations of a 1-d quantum wire. The dispersion relation at left shows a band of energies available for conduction. The density of states at right drops to zero below the conduction band edge (E_C).

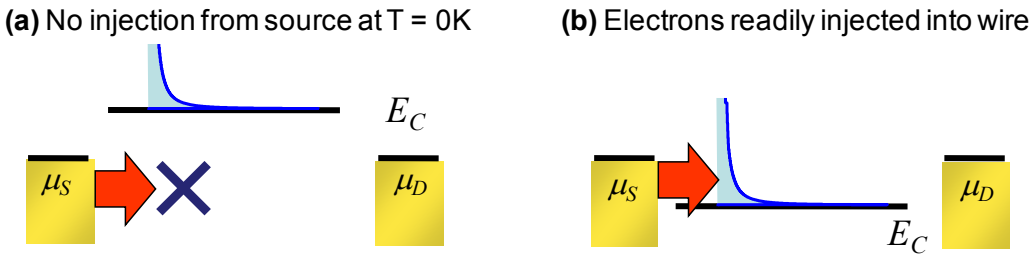


Fig. 4.8. The position of the conduction band edge, E_C, determines whether charge can be injected from the source into the wire.

The application of a bias may later the position of the conduction band edge by changing the local electrostatic potential, U. Two examples are shown in Fig. 4.9, below.

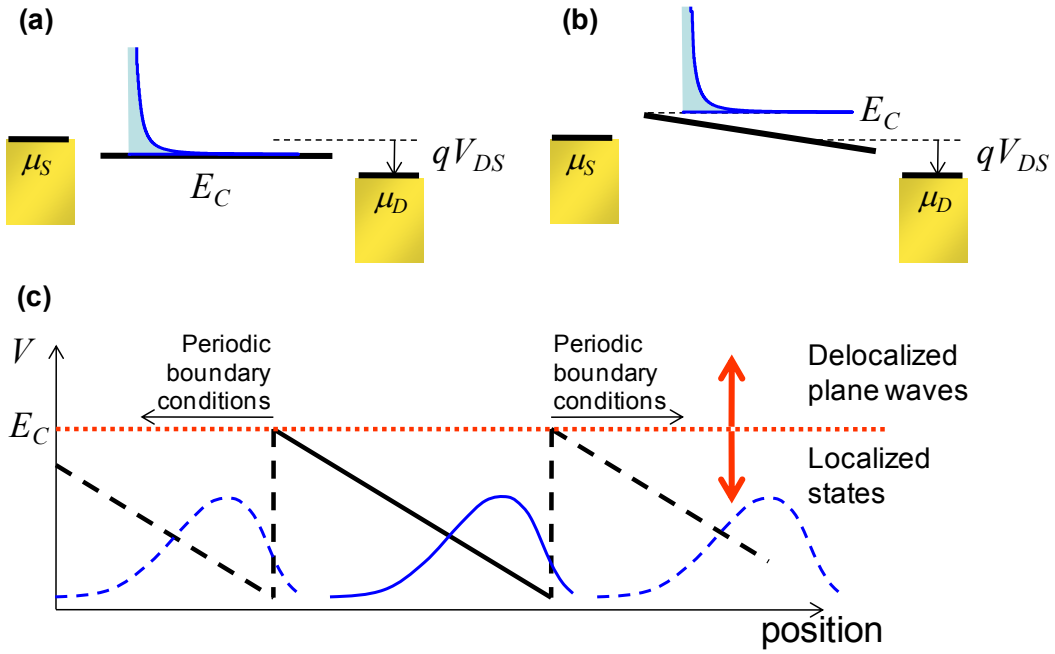


Fig. 4.9. (a) A metallic wire under bias. **(b)** An insulating or nanoscale wire under bias. Note that the conduction band edge corresponds to the maximum potential in the wire. This is explained in **(c)**, where we consider the electronic wavefunctions in the wire under bias. We have applied periodic boundary conditions to help demonstrate that electronic states with energies below the maximum potential are localized. These states may only be accessed by tunneling from the source. Electronic states above the maximum potential are delocalized plane waves. Consequently, the conduction band edge is positioned at the point of the maximum repulsive potential in the wire.

The first example in Fig. 4.9 demonstrates a metallic wire under bias. In this limit there is no potential variation along the wire. In (b) of Fig. 4.9 we present a wire with varying potential along its length. The conduction band edge, E_C, occurs at the point of maximum repulsive potential. This is explained in (c) of Fig. 4.9. Electronic wavefunctions in the

wire with energies below E_C are localized and may only be accessed from the source by tunneling through a repulsive potential. Due to the relatively low rate of injection into these states we will ignore current through these modes in this class. Above E_C, however, the electron wavefunctions are delocalized plane waves which readily transport electrons between the contacts.

Determining the potential profile of the wire can be Consider a point, z, on the wire. The electrostatic potential at point z is given by

$$U(z) = -qV_{DS}\frac{C_D(z)}{C_{ES}(z)} \tag{4.2}$$

where C_D is the capacitance linking the point at z to the drain, and C_S is the capacitance linking the point at z to the source, and $C_{ES}(z)$ is the total electrostatic capacitance at z; $C_{ES}(z) = C_S(z) + C_D(z)$.

If we assume that source and drain capacitances can be modeled by parallel plate capacitors, we found in Eq. (3.15) that the potential varies linearly between the contacts. However, charging of the conductor can change the potential profile by opposing changes induced by the drain source voltage. Adding the effect of charging to Eq. (4.2) gives:

$$U = -qV_{DS}\frac{C_D}{C_{ES}} + \frac{q^2(N-N_0)}{C_{ES}} \tag{4.3}$$

Remember that all these electrostatic capacitances vary with position along the wire. Next, let‟s assume that applied bias is small and consequently the change in charge is small i.e. $\delta N = N - N_0$. We can relate δN to the density of states at the Fermi level in the wire, $g(E_F)$, and the change in potential, δU.

$$\delta N = -g(E_F)\delta U \tag{4.4}$$

Combining Eq. (4.4) with a small signal Eq. (4.3) gives

$$\delta U = -q\delta V_{DS}\frac{C_D}{C_{ES}} - \frac{q^2 g(E_F)}{C_{ES}}\delta U \tag{4.5}$$

Substituting the quantum capacitance $C_Q = q^2 g(E_F)$ and collecting terms gives

$$\delta U = -q\delta V_{DS}\frac{C_D}{C_{ES} + C_Q} \tag{4.6}$$

The equivalent circuit is shown in Fig. 4.10. Note that the quantum capacitance depends on the position of the Fermi level within the DOS. Because the potential shifts the DOS relative to the Fermi level, the quantum capacitance also depends on the potential.

In this class we‟ll consider two extreme cases: $C_Q >> C_{ES}$ and $C_Q << C_{ES}$. The former case corresponds to a perfectly metallic wire. The later case can correspond to either a perfect insulator or a nanoscale conductor, which due to its size, has very few electronic states.

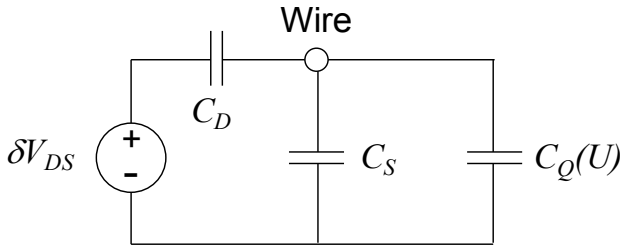


Fig. 4.10. The equivalent circuit to determine the potential in a nanowire under bias. When the quantum capacitance, C_Q, is large the potential in the wire varies little with applied bias. This corresponds to the behavior of a metallic wire. In insulators or smaller wires with fewer electronic states, the potential varies with position in the wire.

(i) Perfectly metallic wires

In the limit that $C_Q >> C_{ES}$, Eq. (4.6) reduces to $\delta U = 0$, meaning that the potential is fixed along the length of the wire by the large density of states at the Fermi Level. The potential of the wire relative to the contacts is then determined by the contact properties, in particular the coupling coefficients τ_S and τ_D. The potential is determined from the analysis of Fig. 3.25.

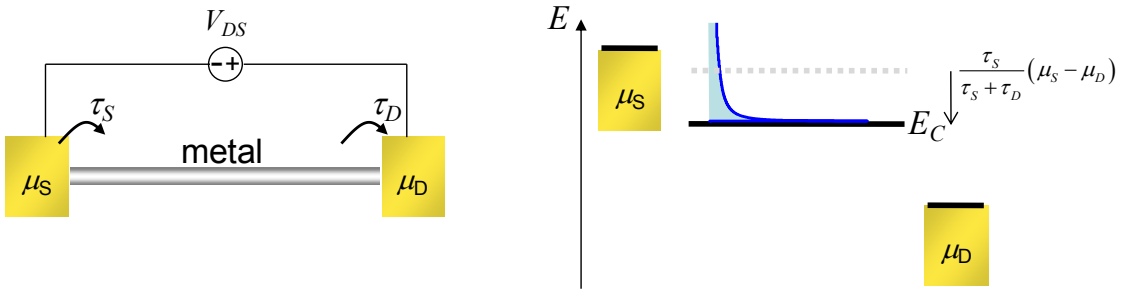


Fig. 4.11. The potential of a metallic wire is constant along its length. The position of the conduction band edge can be modeled by a voltage divider where the source and drain coupling coefficients τ_S and τ_D, respectively, represent the source and drain contact resistors.

(ii) The Insulator/Nanoscale Limit

The spacing between k states in a 1-d conductor is simply $\Delta k = 2\pi/L$, where L is the length of the wire. When L is small there are few states available for electrons. Consequently, insulators and many very small conductors have relatively few states for electrons at the Fermi level. We can ignore charging effects in these conductors. The potential is then simply described by Eq. (4.2).

Part 4. Two Terminal Quantum Wire Devices

The quantum limit of conductance

We've seen that in a quantum wire, current flow requires a difference in the quasi Fermi levels for electrons moving with and against the current. Furthermore, only electrons between the quasi Fermi levels, i.e. $F^- < E < F^+$ carry current.

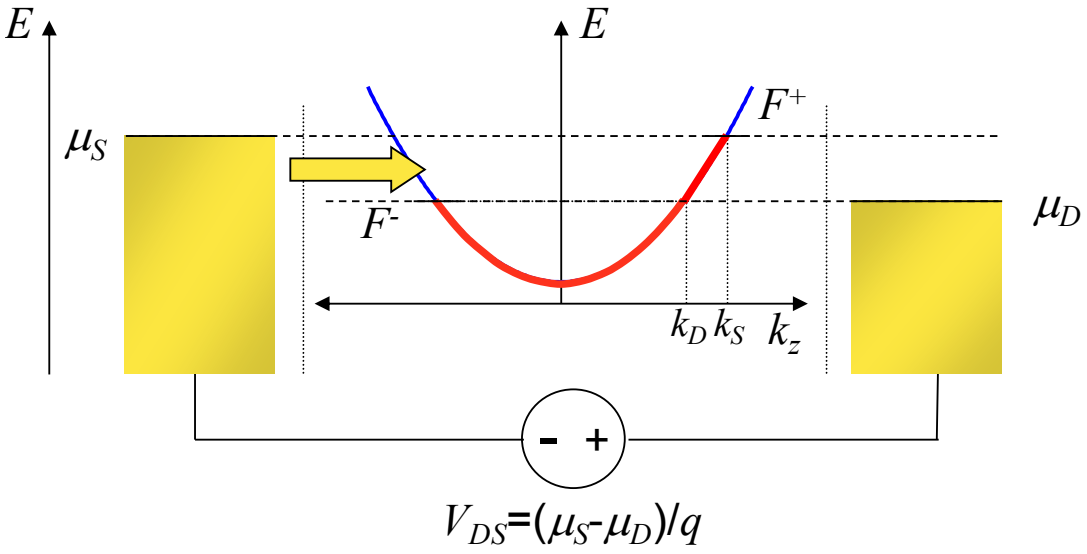


Fig. 4.12. In a single mode wire under bias, k states between k_D and k_S contain uncompensated electrons.

In general, the total current in a quantum wire is

$$I = qN/\tau \tag{4.7}$$

where N is the number of uncompensated electrons, and τ is their transit time (the time they take to cross from one end of the wire to the other). Let's use Eq. (4.7) to calculate the current in a single mode quantum wire at T=0K.

The velocity of electrons in the wire is given by the group velocity (see Problem 3)

$$v = \frac{1}{\hbar}\frac{dE}{dk} \tag{4.8}$$

As an aside, we note that if F^+-F^- is small, the current carrying electrons all move at approximately the equilibrium Fermi velocity.

$$v_F = \frac{1}{\hbar}\frac{dE}{dk}\bigg|_{E_F} \tag{4.9}$$

The transit time in Eq. (4.7) is related to the length of the wire, L, and the velocity of the uncompensated electrons:

$$\tau = \frac{L}{v} = L\Big/\left(\frac{1}{\hbar}\frac{dE}{dk}\right) \tag{4.10}$$

The number of uncompensated electrons is equal to the number of electrons in the states $k_D < k < k_S$, equivalent to $\mu_D < E < \mu_S$ in Fig. 4.12. Each k state occupies $\Delta k = 2\pi/L$. Recall also that there are two electrons per k state (one of each spin). Thus,

$$N = 2\int_{k_D}^{k_S} \frac{dk}{2\pi/L} \tag{4.11}$$

Equation (4.7) is then,

$$I = 2q\int_{k_D}^{k_S} \frac{dk}{2\pi/L}\frac{1}{\hbar L}\frac{dE}{dk} \tag{4.12}$$

Simplifying gives

$$I = \frac{2q}{h}\int_{k_D}^{k_S} dk\,\frac{dE}{dk} \tag{4.13}$$

Changing the variable of integration to energy gives

$$\begin{aligned} I &= \frac{2q}{h}\int_{\mu_D}^{\mu_S} dE \\ &= \frac{2q}{h}\left(\mu_S - \mu_D\right) \end{aligned} \tag{4.14}$$

Note $V_{DS} = -\left(\mu_D - \mu_S\right)/q$, thus

$$I = \frac{2q^2}{h}V \tag{4.15}$$

This expression demonstrates that the resistance of an ideal single mode wire is

$$R = \frac{h}{2q^2} \approx 12.9\ \text{k}\Omega \tag{4.16}$$

A multiple mode wire with M modes can be thought of as M single mode wires in parallel. Since parallel conductances add, the quantum limit is usually written as a conductance. For a multimode ballistic wire the conductance is

$$G_C = \frac{2q^2}{h}M \tag{4.17}$$

This is the famous quantum limited conductance. It yields the surprising conclusion that even ballistic conductors have a resistance, although this resistance is independent of the length of the conductor.

But a resistance implies that power is dissipated when a current flows. Given that electron transport in the wire is ballistic, where do the resistive power losses occur?

If we look at Fig. 4.12 we find that carriers entering the wire from the source propagate without change in potential until they reach the drain where they must come to equilibrium at chemical potential μ_D. Thus, the power is dissipated in the drain.

The quantum limit in conductance arises as a consequence of the interface between the contact with its (ideally) infinite modes and infinite number of electrons all at

equilibrium, and a conductor with a small number of modes supporting non-equilibrium electrons. Thus, the quantum limit can be thought of as a contact resistance.

Of course, as the number of modes in the conductor increases, the contact resistance decreases. In the classical limit, it can be completely ignored.

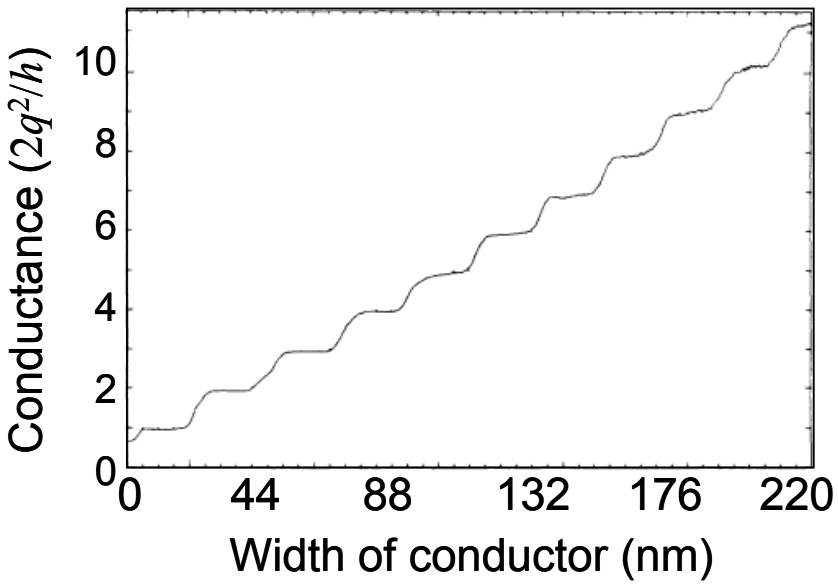


Fig. 4.13. Experimental data from van Wees, *et al* PRL **60,** 848 (1988) clearly showing conductance in a narrow conductor quantized in steps of $2q^2/h$.

The Landauer Formula[†]

We are now going to generalize the result of Eq. (4.15) by considering conduction at higher temperatures and in the presence of a scattering site.

Electrons flowing through the wire may be reflected by the scatterer. We define the transmission probability $\mathcal{T}$, of the scatterer, and assume that it acts equally on electrons flowing in either direction in the wire.

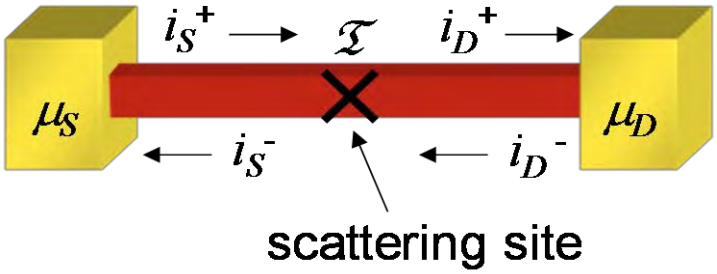


Fig. 4.14. A quantum wire containing a scatterer with transmission probability $\mathcal{T}$.

[†] This section is adapted from S. Datta, „Electronic Transport in Mesoscopic Systems" Cambridge (1995).

Let‟s define i_S^+ as the current carried by all electrons (compensated and uncompensated) in the $+k_z$ states in the wire adjacent to the source. Let i_S^- be the current carried by all electrons in the $-k_z$ states in the wire adjacent to the source. Similarly, we define i_D^+ and i_D^- as the currents entering and leaving the drain, respectively.

Generalizing Eq. (4.11) for wires with multiple modes and arbitrary temperatures, we calculate the number of electrons traveling in the $+k_z$ states adjacent to the source:

$$N_S^+ = 2\int_0^\infty \frac{dk}{2\pi/L} M\left(E(k)\right) f\left(E(k), \mu_S\right) \tag{4.18}$$

where the number of modes at energy E is $M(E)$, and as before $f(E,\mu)$ is the probability that a state of energy E is filled given the chemical potential μ. It follows that

$$\begin{aligned} i_S^+ &= \frac{2q}{h}\int_0^\infty M(E) f(E,\mu_S)\, dE \\ i_D^- &= \frac{2q}{h}\int_0^\infty M(E) f(E,\mu_D)\, dE \end{aligned} \tag{4.19}$$

and

$$\begin{aligned} i_D^+ &= \frac{2q}{h}\int_0^\infty \mathfrak{T} M(E) f(E,\mu_S) + (1-\mathfrak{T}) M(E) f(E,\mu_D)\, dE \\ i_S^- &= \frac{2q}{h}\int_0^\infty \mathfrak{T} M(E) f(E,\mu_D) + (1-\mathfrak{T}) M(E) f(E,\mu_S)\, dE \end{aligned} \tag{4.20}$$

The total current is $I = i_S^+ - i_S^- = i_D^+ - i_D$, this gives us the Landauer Formula

$$I = \frac{2q}{h}\int_0^\infty \mathfrak{T} M(E)\left(f(E,\mu_S) - f(E,\mu_D)\right) dE \tag{4.21}$$

Spatial variation of the electrochemical potential[†]

Next, we try to answer the question: *Where is the voltage dropped?*

Once again, let‟s consider a quantum wire at $T = 0$K. The wire has a single scatterer with transmission probability $\mathfrak{T}$. Uncompensated electrons emitted by the left contact are partly transmitted and partly reflected by the scatterer. Thus, to the right of the scatterer, only a fraction, $\mathfrak{T}$, of the $+k$ states in the energy range $\mu_D < E < \mu_S$ are filled. To the left of the scatterer, the fraction $(1-\mathfrak{T})$ of the $-k$ states in the energy range $\mu_D < E < \mu_S$ are filled; see Fig. 4.15(a).

After scattering the $+k$ states are no longer in equilibrium and the distribution of electrons in the $+k$ states can no longer be described by a quasi Fermi level. These electrons are said to be *hot*, and may travel some distance before they equilibrate. Similarly electrons in the $-k$ states are not in equilibrium to the left of the scatterer.

[†] This section is adapted from S. Datta, „Electronic Transport in Mesoscopic Systems“ Cambridge (1995).

In Fig. 4.15(b) we plot the *average* quasi Fermi level of both + k and - k states. The change in the average quasi Fermi levels can be interpreted as a potential change in the vicinity of the scatterer of $(1-\mathcal{T})(\mu_S-\mu_D)$.

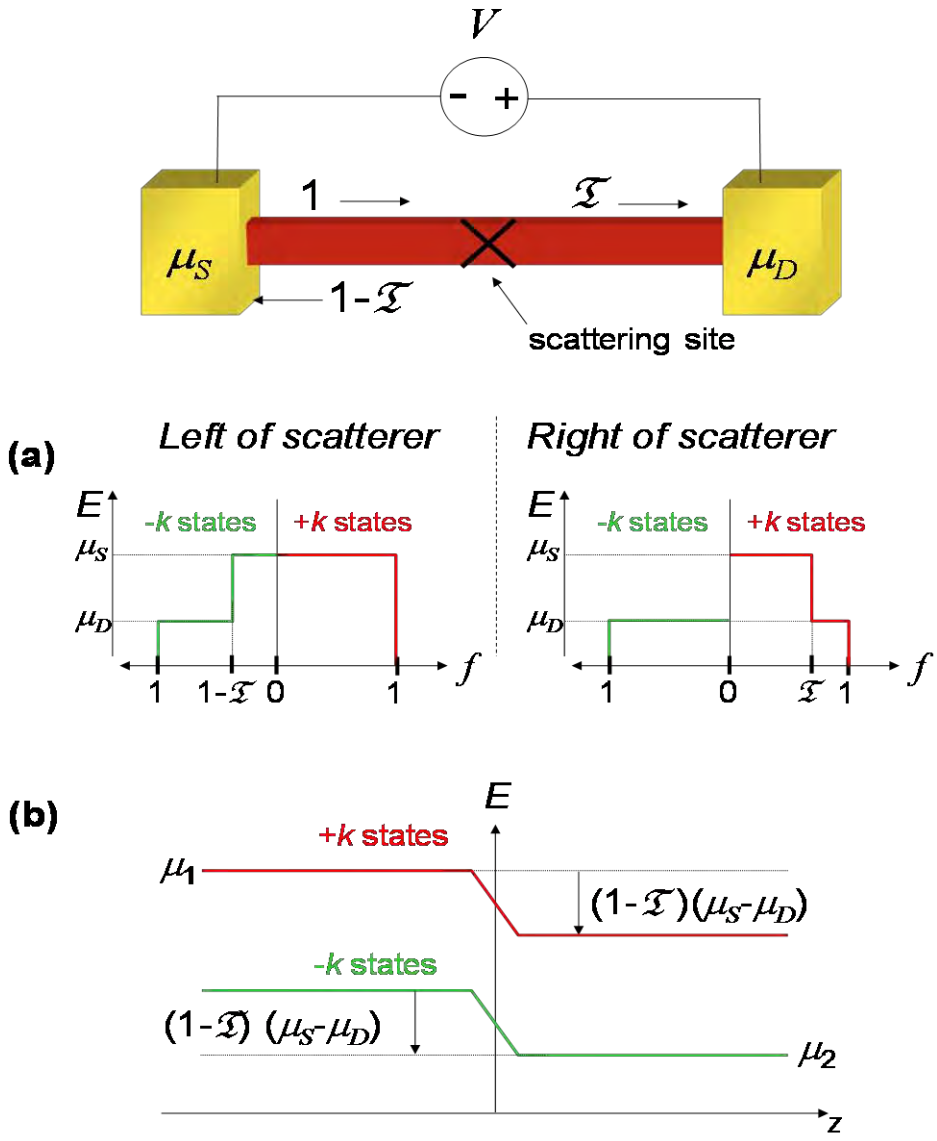


Fig. 4.15. (a) Distribution of electrons within a molecular wire that contains a scattering site. **(b)** The average quasi Fermi level of both $+k$ and $-k$ states changes at the scatterer. This can be interpreted as a change in potential at the scatterer. From S. Datta, „Electronic Transport in Mesoscopic Systems" Cambridge (1995).

But where is the heat dissipated?

It depends where the electrons relax into equilibrium. If the relaxation occurs within the contact, then once again all the heat is dissipated in the drain. Thus, although the average potential changes at the scatterer, heat is only dissipated where the electrons relax.

Ohm's law[†]

What happens when we increase the size of a conductor? Eventually, we should obtain Ohm"s law as the quantum phenomena transform into the familiar model of classical conduction:

$$V = IR \tag{4.22}$$

But a linear relationship between V and I is not particularly profound. Almost any system can be linearized over a sufficiently narrow range of voltage or current. It is more significant to evaluate the resistance, R, in terms of macroscopic quantities such as cross-sectional area, A, and length, L.

You might recall that resistance is classically defined as:

$$R = \frac{\rho L}{A}. \tag{4.23}$$

where ρ, the resistivity, is some material dependent quantity, usually determined by a measurement. Let"s see where this expression comes from – it will help illustrate differences between quantum and classical models of charge conduction.

At zero temperature, transmission formalism gives

$$G = \frac{2q^2}{h} M\mathfrak{T}, \tag{4.24}$$

where M is the number of modes, and $\mathfrak{T}$ is the net transmission coefficient. Rearranging this in terms of a resistance, we have

$$R = \frac{h}{2q^2}\frac{1}{M\mathfrak{T}} \tag{4.25}$$

To determine the net transmission coefficient, let"s break the conductor into a series of N elements labeled $i = 1\ldots N$, each containing a scattering site with transmission $\mathfrak{T}_i$; see Fig. 4.16.

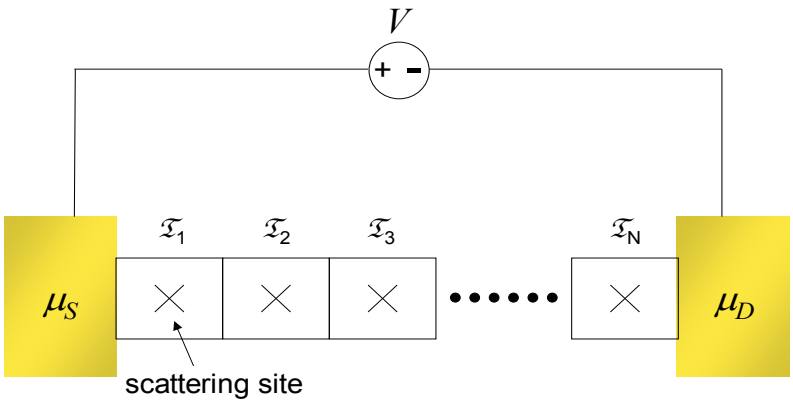


Fig. 4.16. The macroscopic conductor can be represented as a series of N scattering sites, each with transmission $\mathfrak{T}_i$.

[†] This section is adapted from S. Datta, „Electronic Transport in Mesoscopic Systems" Cambridge (1995).

Part 4. Two Terminal Quantum Wire Devices

For many scatterers there will be many reflections to consider. If the scattering mechanism preserves the phase information of the electrons, then multiple reflections can yield interference effects. Such scattering is said to be coherent. Here, we will consider only incoherent scattering that randomizes the electron phase.

Let"s begin with just two incoherent scatterers in series. The transmission for two incoherent scatterers in series is:

$$\mathcal{T}_{12} = \mathcal{T}_1\mathcal{T}_2 + \mathcal{T}_1\mathcal{T}_2\mathcal{R}_1\mathcal{R}_2 + \mathcal{T}_1\mathcal{T}_2\left(\mathcal{R}_1\mathcal{R}_2\right)^2 + \ldots \tag{4.26}$$

where $\mathcal{R}_i$ is the reflection from the ith scatterer, and $\mathcal{T}_i = (1\text{-}\mathcal{R}_i)$.

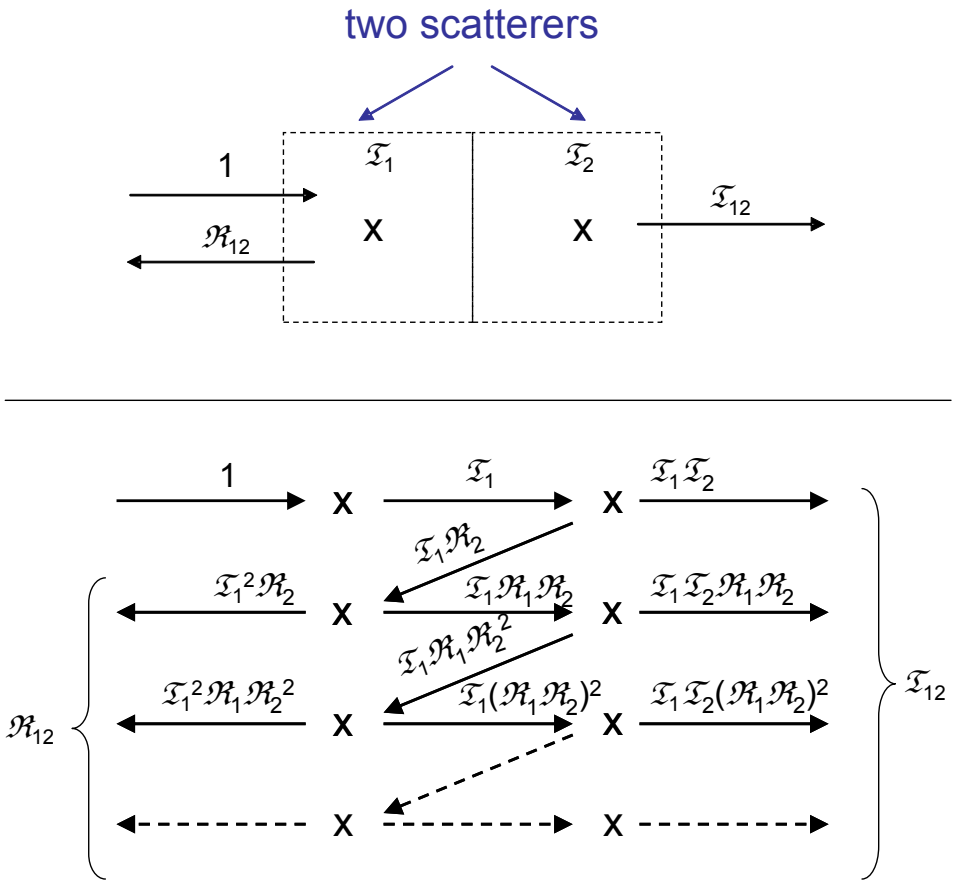


Fig. 4.17. Two scatterers generate an infinite number of reflections, but we can sum the geometric series. Adapted from S. Datta, „Electronic Transport in Mesoscopic Systems" Cambridge (1995).

This geometric series simplifies to

$$\mathcal{T}_{12} = \frac{\mathcal{T}_1\mathcal{T}_2}{1-\mathcal{R}_1\mathcal{R}_2}, \tag{4.27}$$

We can rearrange Eq. (4.27) to show

$$\frac{\mathcal{R}_{12}}{\mathcal{T}_{12}} = \frac{\mathcal{R}_1}{\mathcal{T}_1} + \frac{\mathcal{R}_2}{\mathcal{T}_2} \tag{4.28}$$

Thus, for N identical scatterers:

$$\frac{\mathfrak{R}}{\mathfrak{T}} = N\frac{\mathfrak{R}_i}{\mathfrak{T}_i} \tag{4.29}$$

Solving for the net transmission and using $\mathfrak{T} = (1-\mathfrak{R})$

$$\mathfrak{T} = \frac{\mathfrak{T}_i}{N\left(1-\mathfrak{T}_i\right)+\mathfrak{T}_i} \tag{4.30}$$

If we have λ scatterers per unit length, then

$$\mathfrak{T} = \frac{\mathfrak{T}_i}{\lambda L\left(1-\mathfrak{T}_i\right)+\mathfrak{T}_i} = \frac{L_0}{L+L_0} \tag{4.31}$$

Thus, Eq. (4.25) becomes

$$R = \frac{h}{2q^2}\frac{1}{M}\left(1+\frac{L}{L_0}\right) \tag{4.32}$$

where $L_0 = \mathfrak{T}/\lambda\left(1-\mathfrak{T}\right)$ is a characteristic length. The length dependence of resistance is clear from Eq. (4.32). The dependence on cross sectional area is due to the number of current-carrying modes in the conductor

$$M \approx \frac{{k_F}^2}{4\pi}A, \tag{4.33}$$

where k_F is the Fermi wavevector.

Thus, we find that resistance can be broken into two components, a resistance due to the contacts, and a resistance that scales with the length of the conductor.

$$R = R_C + R_B = \frac{\rho L_0}{A} + \frac{\rho L}{A} \tag{4.34}$$

The contact resistance is the quantum effect that is familiar to us from Landauer theory. But in large conductors, the contact resistance is overwhelmed and we get the familiar classical expression for resistance.

The Drude or Semi-Classical Model of Charge Transport

Quantum models of charge conduction are rarely applied outside nanoelectronics. For traditional applications, the semi-classical model of the German physicist Paul Drude is usually sufficient. Drude proposed that conductors contain immobile positive ions embedded in a sea of electrons. Unlike the quantum view, where those electrons occupy various states with different energies, Drude viewed electrons as indistinguishable.

In the quantum model of charge transport, current is carried by only that fraction of electrons close to the Fermi energy. The current carrying electrons move at approximately the Fermi velocity, $v_F = \hbar k_F/m$. The remaining electrons are compensated, *i.e.* equal numbers flow in each direction yielding no net current.

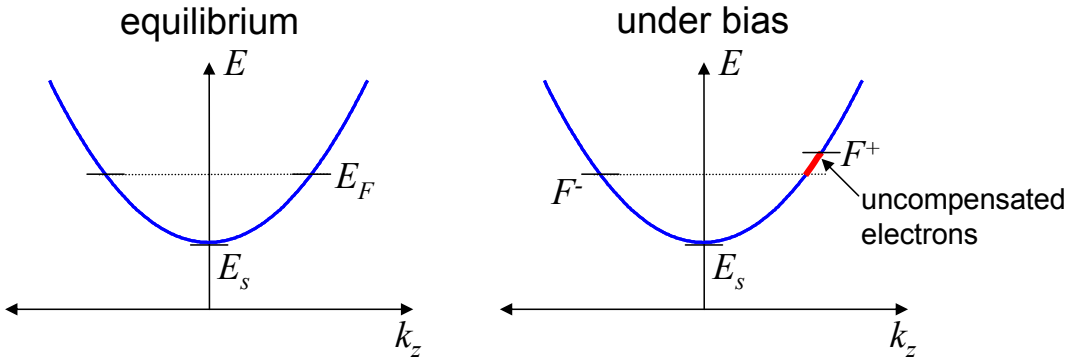


Fig. 4.18. Application of an electric field shifts the quasi Fermi levels for electrons moving with the field (F^+) and against the field (F^-).

But in the Drude model, current is carried by all electrons, moving at an average velocity known as the drift velocity, $\mathbf{v}_d$. Thus, the fundamental classical model for charge conduction is

$$\mathbf{J} = qn\mathbf{v}_d \tag{4.35}$$

where n is the density of electrons.

In the Drude model, all the electrons travel in the direction of the electric field, gathering energy from the field. Eventually each electron collides with something, a positive ion or another electron, at which point, the electron is stopped. It is then accelerated once more by the electric field, traveling in this stop-start manner through the conductor.

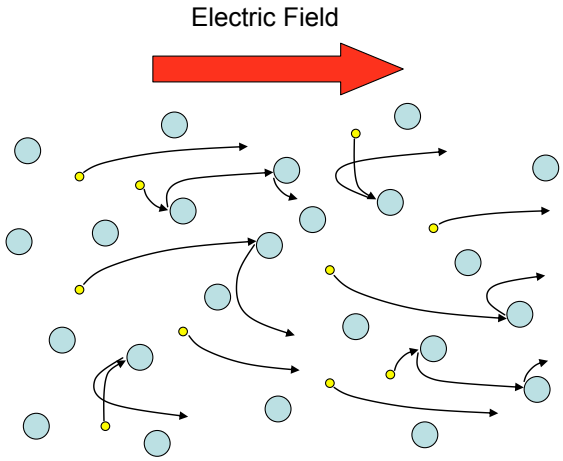


Fig. 4.19. Electron paths through scattering sites. The average time between collisions is the relaxation time, τ_m.

The conductivity of the material is characterized by τ_m, the relaxation time, the mean time between collisions.

The rate at which electrons gain momentum from the field, ε must be equal to the rate of losses due to scattering:[†]

[†] This derivation follows Ashcroft and Mermin, „Solid State Physics", Saunders College Publishing (1976).

$$\left.\frac{d\mathbf{p}}{dt}\right|_{scattering} = \left.\frac{d\mathbf{p}}{dt}\right|_{field}. \tag{4.36}$$

$$\frac{mv_d}{\tau_m} = q\varepsilon\,, \tag{4.37}$$

Rearranging Eq. (4.37), we can express the drift velocity and current density in terms of the relaxation time.

$$\mathbf{J} = \frac{nq^2\tau_m}{m}\boldsymbol{\varepsilon} \tag{4.38}$$

Comparing to Ohm‟s law (expressed in terms of the conductivity, $\sigma = 1/\rho$)

$$\mathbf{J} = \sigma\boldsymbol{\varepsilon} \tag{4.39}$$

where

$$\sigma = \frac{nq^2\tau_m}{m} \tag{4.40}$$

Mobility

One of the most important simplifications of the Drude model is *mobility*, defined as the ratio of the electric field to the drift velocity.

$$v_d = \mu\varepsilon \tag{4.41}$$

Using Eq. (4.37) we obtain

$$\mu = \frac{q\tau_m}{m} \tag{4.42}$$

Mobility is a very common metric for the quality of transistor materials. It typically peaks at several thousand cm^2/Vs in high quality transistor materials such as GaAs or InP. But as we have seen, the actual charge carrier velocity, v_F, has little relation to the electric field. So, why are Drude parameters such as mobility and conductivity useful quantities?

Effective Mass

So far, both the classical and quantum models of conduction have assumed that the current carrying electrons occupy pure planewave states. The dispersion relation of real materials, however, varies from the ideal parabola. We can approximate any dispersion relation by a plane wave if we allow the mass of the electron to vary. We call the modified mass the *effective mass*. Under this approximation, the electron is thought of as a classical particle and various complex phenmomena are wrapped up in the effective mass. For example, given dispersion relation $E(k)$, a Taylor expansion about $k = 0$ yields:

$$E(k) = E(0) + k\left.\frac{dE}{dk}\right|_{k=0} + \frac{1}{2}k^2\left.\frac{d^2E}{dk^2}\right|_{k=0} + \ldots \tag{4.43}$$

Approximating the dispersion relation by a plane wave gives

$$E(k) = E_0 + \frac{\hbar^2k^2}{2m^*} \tag{4.44}$$

Equating the quadratic terms in Eqns. (4.43) and (4.44) we get an expression for the effective mass

$$m^* = \hbar^2 \left(\frac{d^2 E}{dk^2} \right)^{-1} \tag{4.45}$$

The effective mass concept is commonly used in classical models of electron transport, especially models of mobility like Eq. (4.42).

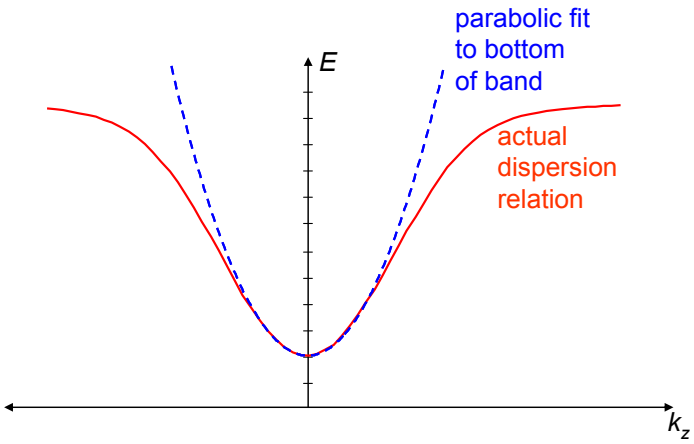


Fig. 4.20. We can model conduction in a material of arbitrary dispersion relation by assuming plane wave electron states with variable (effective) electron mass, m^*, obtained by fitting a parabola to the bottom of the band.

Comparing the quantum and Semi-Classical Drude models of conduction

(i) The mean free path

The Drude model gives a physically incorrect picture of charge conduction. Nevertheless it works quite well. The quantum model shows that rather than all the electrons moving at the drift velocity, as in the Drude model, only the uncompensated electrons carrying current, each moving at approximately the Fermi velocity:[3] Thus, the Drude model can be rearranged as

$$J = qn' v_F \tag{4.46}$$

where the uncompensated charge density is

$$n' = n \frac{v_d}{v_F} \tag{4.47}$$

We can also define the **mean free path**, L_m, as the average distance an electron travels between scattering events. The mean free path is related to the Fermi velocity by:

$$L_m = v_F \tau_m \tag{4.48}$$

Interestingly, the mean free path is approximately equal to the characteristic length L_0 in the derivation of Ohm's law.

(ii) Equilibrium and Non-equilibrium current flow

We can demonstrate the differences between the classical and ballistic limits using the analogy of water flow from one reservoir to another. The application of bias across a wire is equivalent to depressing the height of the drain reservoir relative to the source

reservoir. In the ballistic model water flowing from the source travels across the wire as a jet before relaxing to equilibrium in the drain. In the classical model the water minimizes its potential in channel.

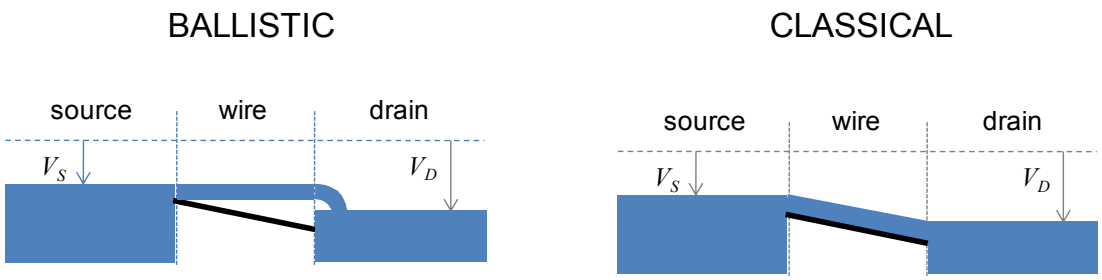


Fig. 4.21. A water flow analogy for ballistic and classical current flow. In the classical limit, the water is always in local equilibrium with the channel.

One way to think about classical transport is as the limit of a series of nanoscale ballistic wires interspersed by contacts. By definition, electrons in the contacts are in equilibrium. Thus contacts are different to the elastic scatterers we considered above, because electrons change energy in contacts. The limiting case of many closely spaced contacts is a continuously varying conduction band edge; see Fig. 4.22.

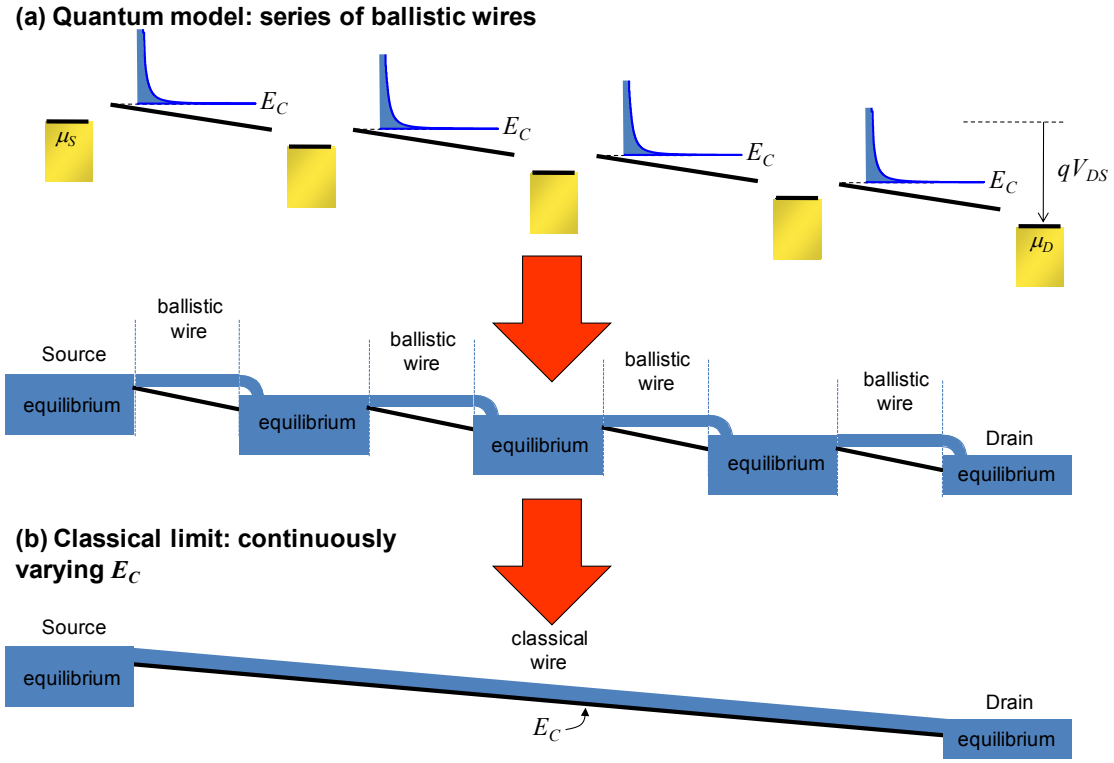


Fig. 4.22. In a classical wire, the conduction band edge varies continuously with position. The classical model can be imagined as the limiting case of many ballistic devices in series.

(iii) The length scale of ballistic conduction

To determine whether we should use the ballistic or semi-classical models of charge transport we need to know the likelihood of electron scattering in the channel. This depends on the channel length, and the quality of the semiconductor.

The number of scattering events in the channel is given by τ/τ_m where τ is the transit time of the electron, and τ_m is its average scattering time. Relating the transit time to the carrier velocity, and τ_m to the definition of mobility in Eq. (4.42) gives:

$$\frac{\tau}{\tau_m} = \frac{l/v}{m_{eff}\mu/q} = \frac{l^2/\mu V_{SD}}{m_{eff}\mu/q} = \frac{ql^2}{m_{eff}V_{SD}\mu^2} \tag{4.49}$$

This expression is plotted in Fig. 4.23 assuming a Si conductor with $V_{DS} = 1$V, $\mu = 300$ cm^2/Vs and $m_{eff} = 0.5 \times m_0$, where m_0 is the mass of the electron. It shows that silicon is expected to cross into the ballistic regime for lengths of approximately $l < 50$nm.

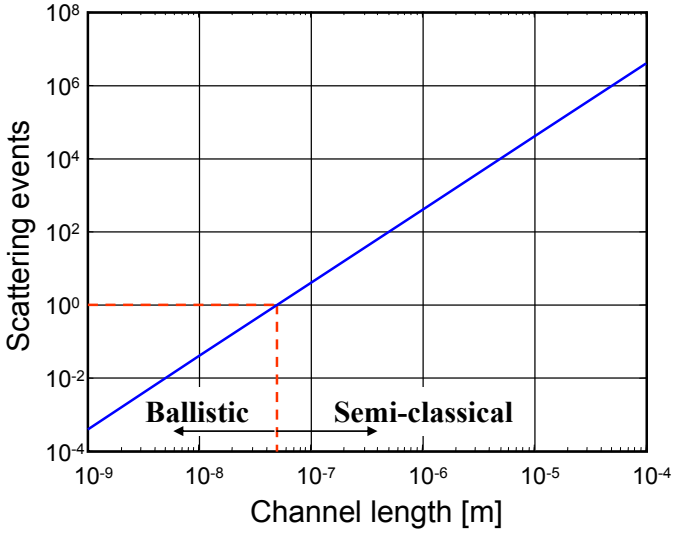


Fig. 4.23. The expected number of electron scattering events in Si as a function of the channel length. The threshold of ballistic operation occurs for channel lengths of approximately 50nm.

Problems

1. Consider the metal/nanowire/metal device shown below. Assume the nanowire is an ideal 1-dimensional conductor with no scattering.

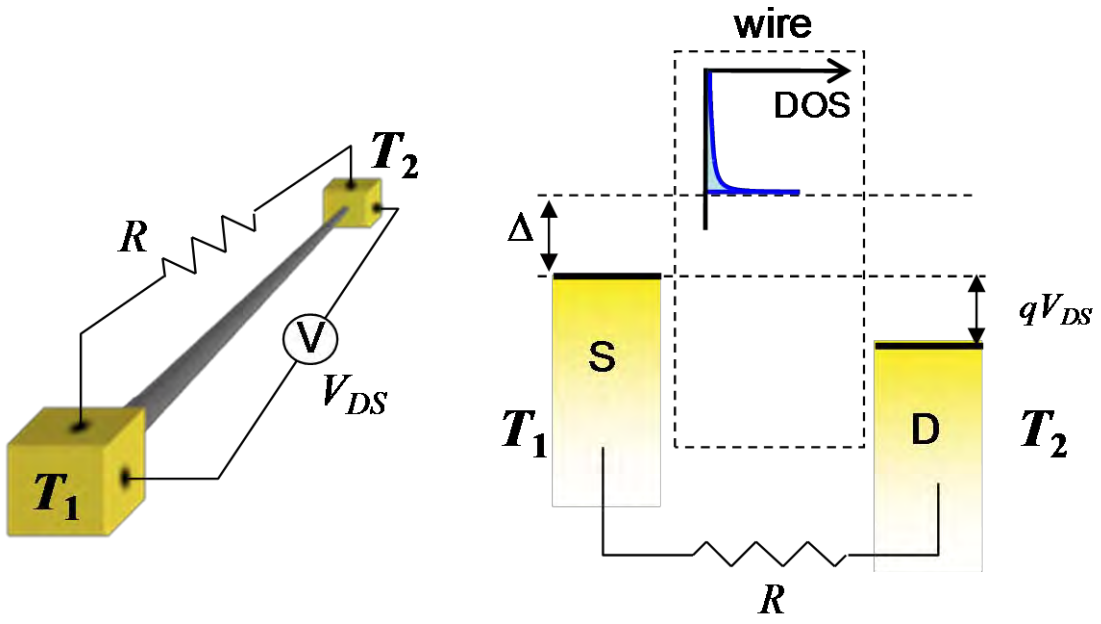


Fig. 4.24. A thermoelectric nanowire.

The source contact is heated to temperature T_1, while the drain contact remains at temperature T_2. Assume that the energy separation between the source and the bottom of the conduction band (Δ) is independent of bias. Assume also that $\Delta >> kT_1$.

(a) The contacts are now shorted together, i.e. $R \rightarrow 0$. What is the current that flows? (This is the „short circuit current").

(b) Next assume the contacts are returned to open circuit, i.e. $R \rightarrow \infty$. What is the voltage between the contacts? (This is the „open circuit voltage")

2. The dispersion relation for a relativistic particle is given by $E = \sqrt{p^2c^2 + (m_0c^2)^2}$ where $E = \hbar\omega$ and $p = \hbar k$. Find the group velocity of this particle.

3. The group velocity is given by $v_g = \frac{d}{dt}\langle x \rangle$. Show that $\frac{d}{dt}\langle x \rangle = \left\langle \frac{1}{\hbar}\frac{dE}{dk} \right\rangle$.

Part 4. Two Terminal Quantum Wire Devices

4. Find the effective mass for an electron in a conductor with the dispersion relation:

$$E(k) = 5 - 2V\cos(ka)\ ,\ |k| < \frac{\pi}{a}$$

where V and a are positive constants.

5. Graphene exhibits a photon-like dispersion relation. Assume the carrier velocity is independent of the carrier's energy and equal to the speed of light, c. Based on the high velocity of carriers in graphene it is often argued that graphene transistors will be faster than similar transistors constructed from other materials.

(a) Draw the dispersion relation of a graphene wire. Assume the wire has only one mode.

(b) Given a wire of length, l, constructed of ballistic graphene, assume we inject a carrier pulse as shown below. f is the distribution function, *i.e.* when $f = 1$ each state is completely full. What is the applied voltage?

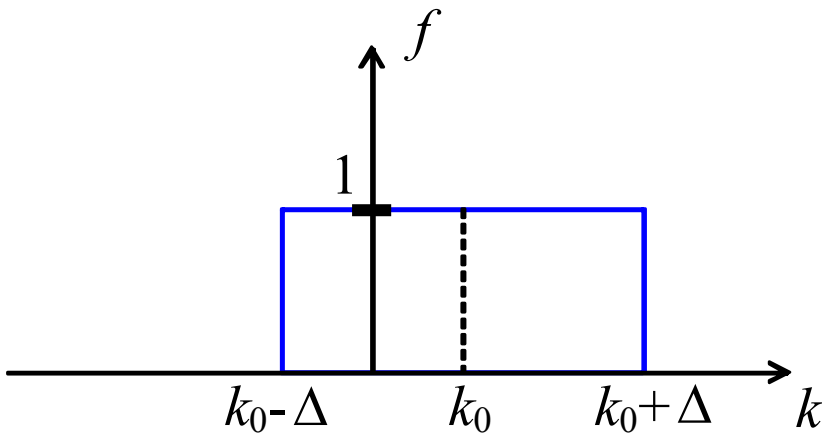


Fig. 4.25. The electron distribution in the graphene wire.

(c) How many carriers are contained in the pulse?

(d) Determine the current carried by the wire from the group velocity and number of carriers.

(e) What is the conductance of the wire?

(f) Now assume that the graphene is used to drive a load capacitance of value C. What is the time constant of the system? How does the graphene wire compare to other 1d wires?

6. This problem refers to the ballistic 1-D wire below. The X"s in the wire are representative of elastic scattering sites, each with transmission, T. Assume $E_C < \mu_S, \mu_D$ where

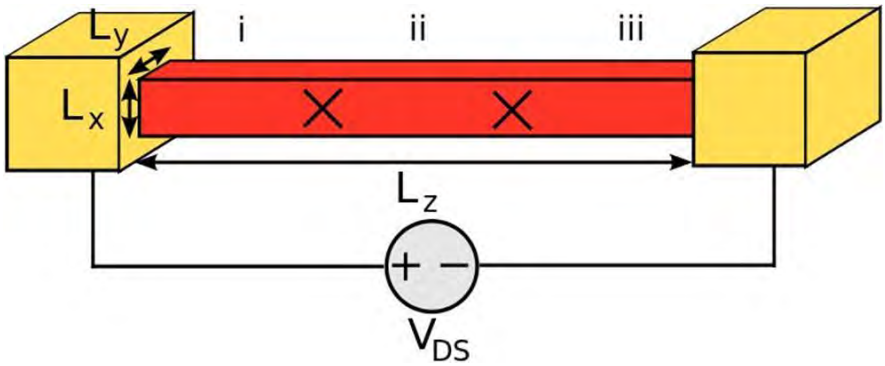
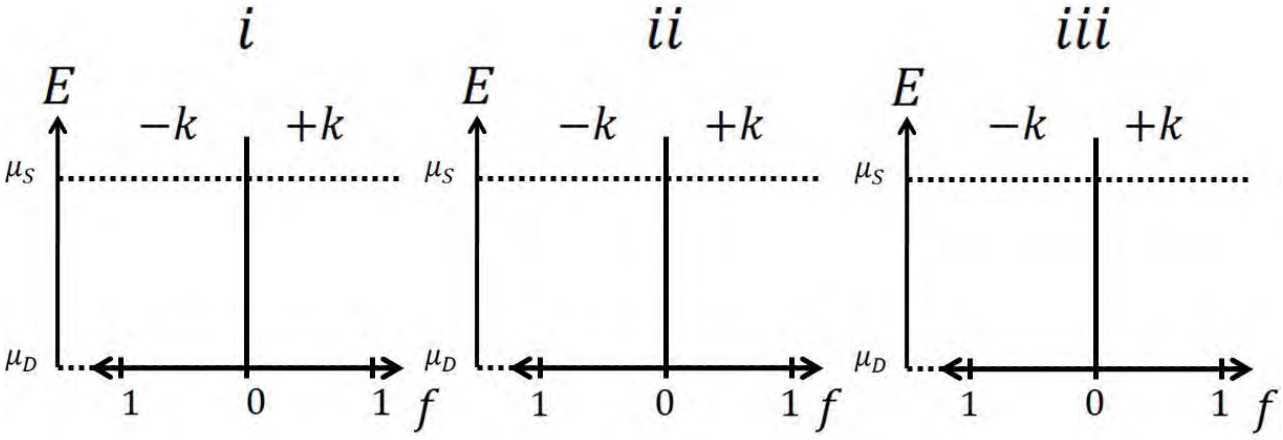
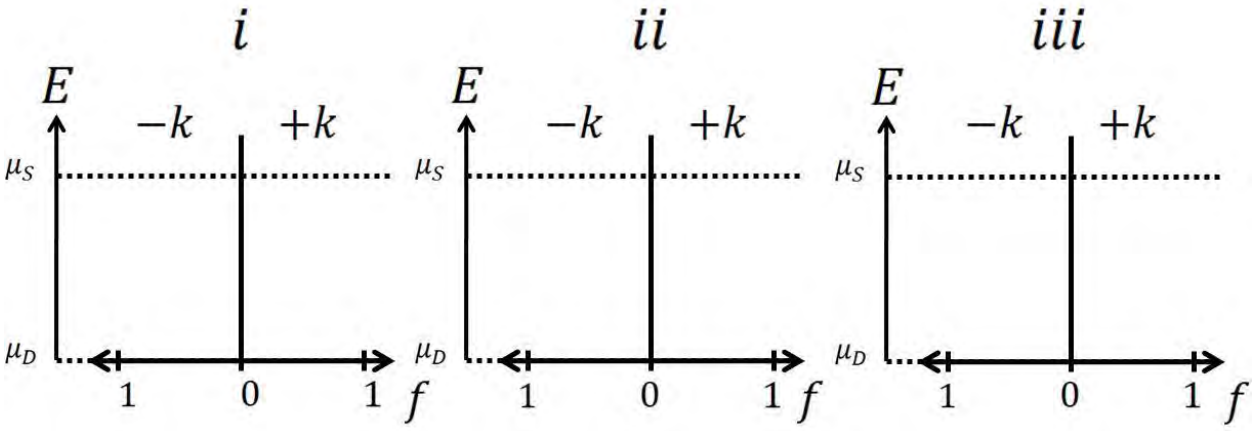


Fig. 4.26. A ballistic nanowire with two scattering sites.

(a) For T = 1.0, plot the filling function at positions (i), (ii), and (iii) along the z-axis.

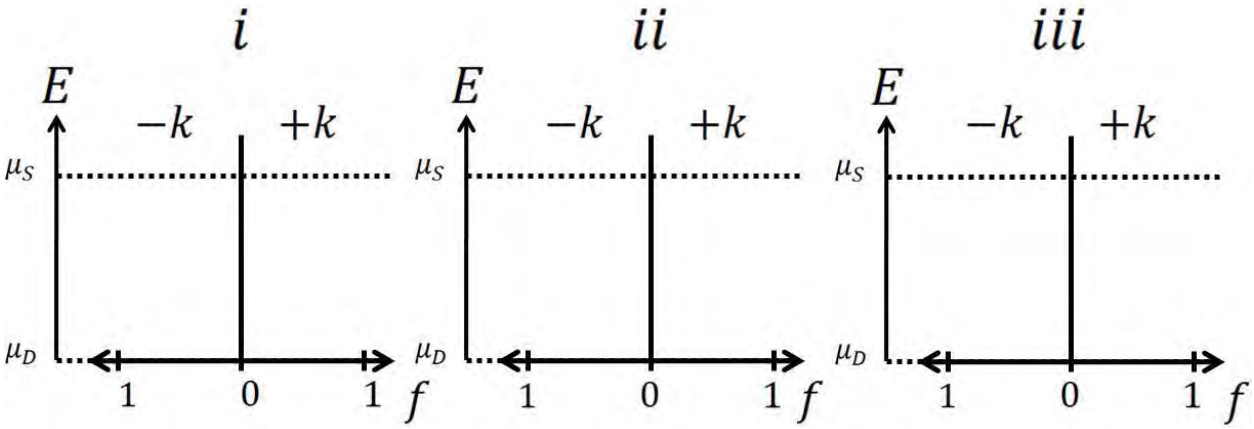


(b) For T = 0.5, plot the filling function at positions (i), (ii), and (iii) along the z-axis.



Part 4. Two Terminal Quantum Wire Devices

(c) Consider a very large number of scattering sites along the wire, each with T = 0.5. Plot the filling function at the source (i), at the midpoint of the wire (ii), and at the drain (iii).



Part 5. Field Effect Transistors

Field Effect transistors (FETs) are the backbone of the electronics industry. The remarkable progress of electronics over the last few decades is due in large part to advances in FET technology, especially their miniaturization, which has improved speed, decreased power consumption and enabled the fabrication of more complex circuits. Consequently, engineers have worked to roughly double the number of FETs in a complex chip such as an integrated circuit every 1.5-2 years; see Fig. 1 in the Introduction. This trend, known now as Moore‟s law, was first noted in 1965 by Gordon Moore, an Intel engineer. We will address Moore‟s law and its limits specifically at the end of the class. But for now, we simply note that FETs are already small and getting smaller. Intel‟s latest processors have a source-drain separation of approximately 65nm.

In this section we will first look at the simplest FETs: molecular field effect transistors. We will use these devices to explain field effect switching. Then, we will consider ballistic quantum wire FETs, ballistic quantum well FETs and ultimately non-ballistic macroscopic FETs.

(i) Molecular FETs

The architecture of a molecular field effect transistor is shown in Fig. 5.1. The molecule bridges the source and drain contact providing a *channel* for electrons to flow. There is also a third terminal positioned close to the conductor. This contact is known as the *gate*, as it is intended to control the flow of charge through the channel. The gate does not inject charge directly. Rather it is capacitively coupled to the channel; it forms one plate of a capacitor, and the channel is the other. In between the channel and the conductor, there is a thin insulating film, sometimes described as the „oxide‟ layer, since in silicon FETs the gate insulator is made from SiO_2. In the device of Fig. 5.1, the gate insulator is air.

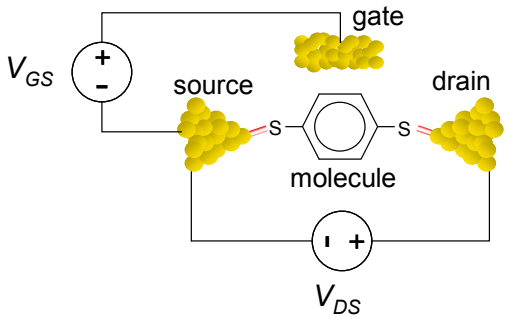


Fig. 5.1. A molecular FET. An insulator separates the gate from the molecule. The gate is not designed to inject charge. Rather it influences the molecule‟s potential.

Part 5. Field Effect Transistors

FET switching

In digital circuits, an ideal FET has two states, ON and OFF, selected by the potential applied to the gate. In the OFF state, the channel is closed to the flow of electrons even if a bias is applied between the source and drain electrodes. To close the channel, the gate must prevent the injection of electrons from the source. For example, consider the molecular FET in Fig. 5.2. Here, we follow FET convention and measure all potentials relative to a grounded source contact. For a gate bias of $V_{GS} < \sim 6.2$V little current flows through the molecule. But for $\sim 6.2\text{V} < V_{GS} < \sim 6.4$V the FET is ON and the channel is conductive.

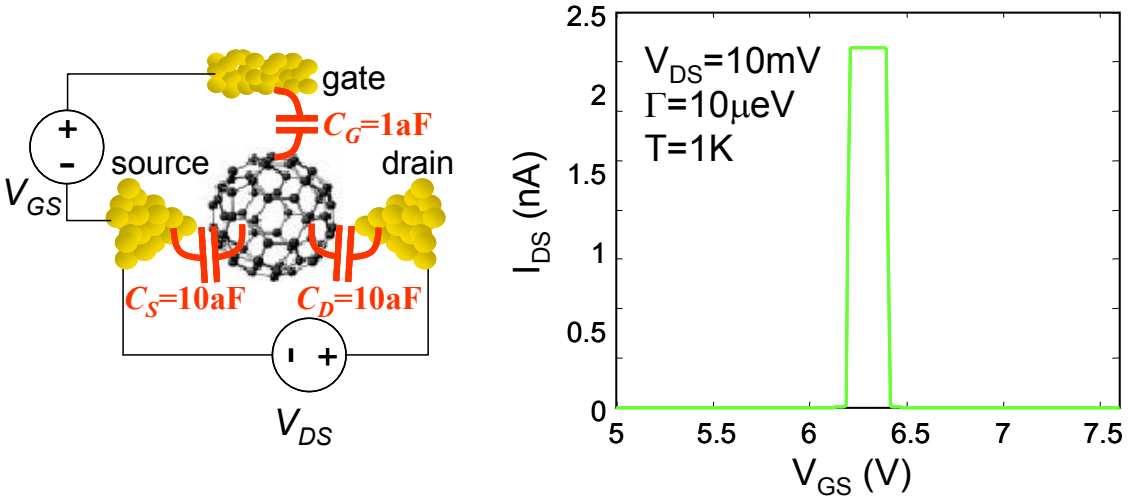


Fig. 5.2. The I_{DS}-V_{GS} characteristics of a FET employing the buckyball molecule C60. At equilibrium the source and drain chemical potentials are at -5eV, and the molecule's LUMO is at -4.7eV. The various electrostatic capacitances in the device are labeled. As the gate potential is increased, the LUMO is pushed lower. At approximately V_{GS} = 6.3V, it is pushed into resonance with the source and drain contacts and the current increases dramatically. The width of the LUMO determines the sharpness of the resonance. Note „aF" is the symbol for atto Farad (10^{-18} F).

As shown in Fig. 5.3, the transitions are much more gradual if the molecular energy level is broader. Similarly, increasing the temperature can also blur the switching characteristics.

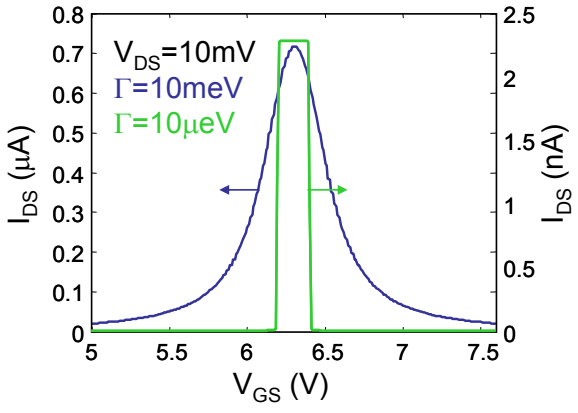


Fig. 5.3. A comparison between the switching characteristics of molecular FETs with broad and narrow energy levels. Note the different current scales.

The origin of FET switching is explained in Fig. 5.4. The gate potential acts to shift energy levels in the molecule relative to the contact chemical potentials. When an energy level is pushed between μ_1 and μ_2 electrons can be injected from the source. Correspondingly, the current is observed to increase. Further increases in gate potential push the energy level out of resonance and the current decreases again at ~ 6.4V.

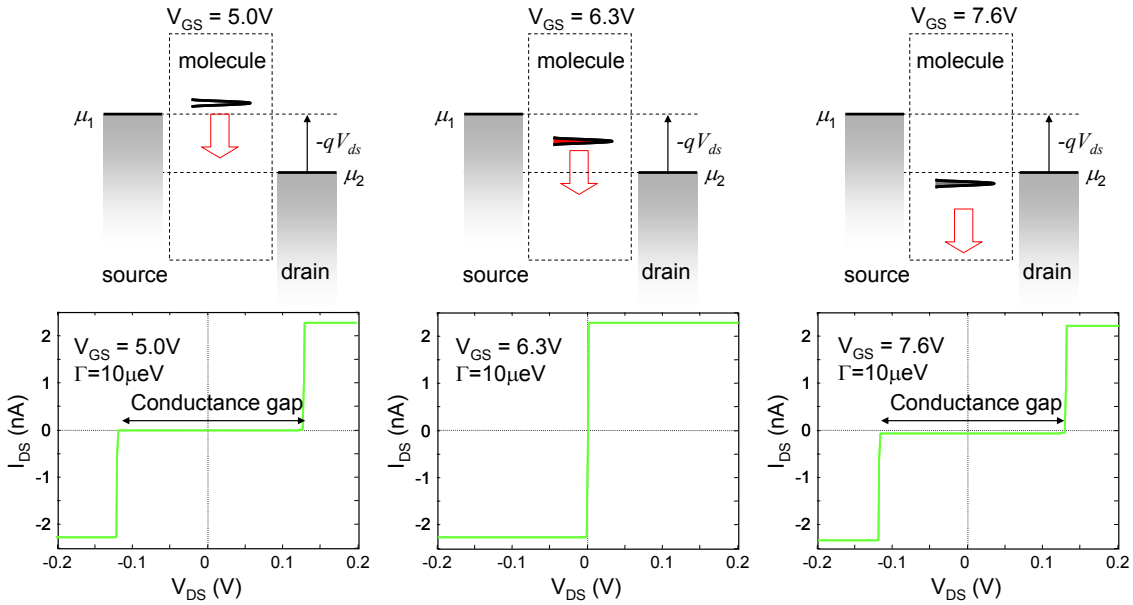


Fig. 5.4. The I_{DS}-V_{DS} characteristics of the FET from Fig. 5.2. Outside resonance a conductance gap opens because additional source-drain bias is required to pull the molecular level between the source and drain chemical potentials.

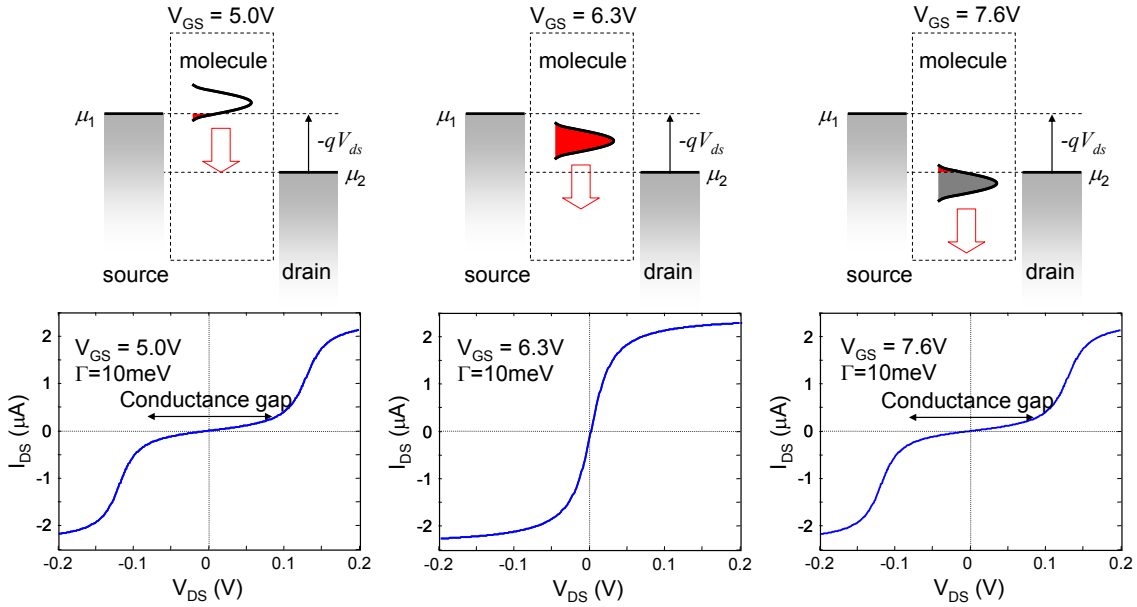


Fig. 5.5. The I_{DS}-V_{DS} characteristics of the FET from Fig. 5.2, except this time the width of the LUMO is 1000 x broader in energy. The corresponding IV shows more gradual transitions, a narrower conductance gap and much higher currents.

Part 5. Field Effect Transistors

FET Calculations

Unlike the two terminal case, where we arbitrarily set $E_F = 0$ and shifted the Source and Drain potentials under bias, the FET convention fixes the Source electrode at ground. There are two voltage sources: V_{GS}, the gate potential, and V_{DS}, the drain potential. We analyze the influence of V_{GS} and V_{DS} on the molecular potential using capacitive dividers and superposition:

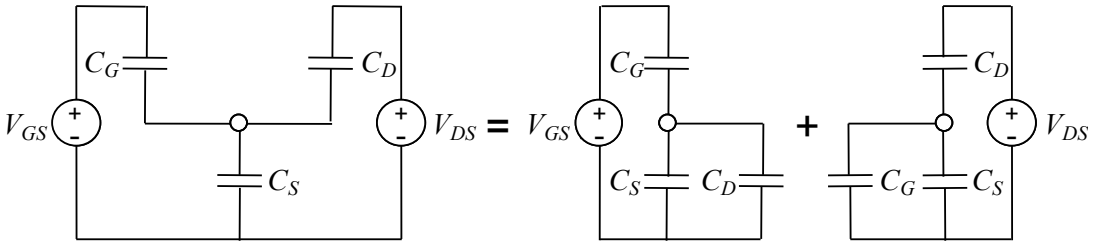


Fig. 5.6. Analyzing a molecular FET using a capacitive divider and superposition.

$$U_{ES} = -qV_{GS}\frac{1/(C_D + C_S)}{1/(C_D + C_S) + 1/C_G} - qV_{DS}\frac{1/(C_G + C_S)}{1/(C_G + C_S) + 1/C_D} \tag{5.1}$$

Simplifying, and noting that the total capacitance at the molecule is $C_{ES} = C_S + C_D + C_G$:

$$U_{ES} = -qV_{GS}\frac{C_G}{C_{ES}} - qV_{DS}\frac{C_D}{C_{ES}} \tag{5.2}$$

We also must consider charging. As before,

$$U_C = \frac{q^2}{C_{ES}}(N - N_0) \tag{5.3}$$

Recall that charging opposes shifts in the potential due to V_{GS} or V_{DS}. Thus, if charging is significant, the switching voltage increases; see Fig. 5.7.

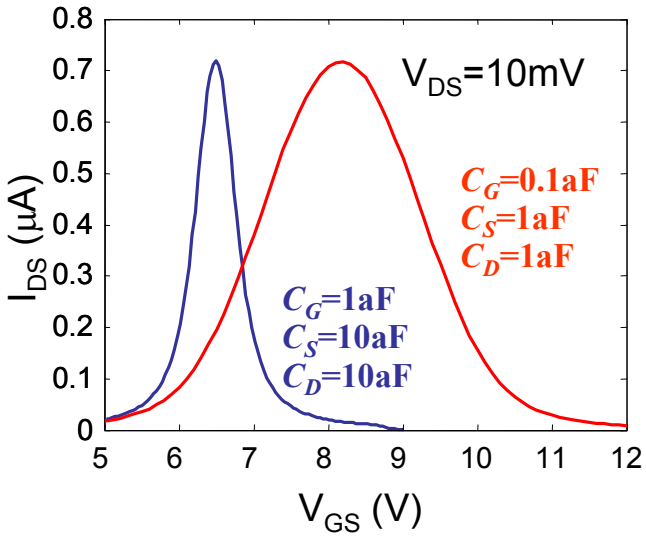


Fig. 5.7. I_{DS}-V_{GS} characteristics for the FET of Fig. 5.2 calculated under two different sets of capacitances. Charging is more important for the smaller capacitances. The switching voltage for this device is observed to increase to ~ 8.1V.

Adding the static potential due to V_{DS} and V_{GS} gives, the potential U in terms of the charge, N, and bias

$$U = -qV_{GS}\frac{C_G}{C_{ES}} - qV_{DS}\frac{C_D}{C_{ES}} + \frac{q^2}{C_{ES}}\left(N - N_0\right) \tag{5.4}$$

We also have an expression for potential for N in terms of U (see Eq. (3.31))

$$N = \int_{-\infty}^{\infty} g\left(E-U\right)\frac{\tau_D f\left(E,\mu_S\right)+\tau_S f\left(E,\mu_D\right)}{\tau_S+\tau_D}dE \tag{5.5}$$

As before, Eqns. (5.4) and (5.5) must typically be solved iteratively to obtain U. Then we can solve for the current using:

$$I = q\int_{-\infty}^{\infty} g\left(E-U\right)\frac{1}{\tau_S+\tau_D}\left(f\left(E,\mu_S\right)-f\left(E,\mu_D\right)\right)dE \tag{5.6}$$

Quantum Capacitance in FETs

Unfortunately, Eqns. (5.4) and (5.5) typically must be solved iteratively. But insight can be gained by studying a FET with a few approximations.

Another way to think about charging is to consider the effect on the channel potential of incremental changes in V_{GS} or V_{DS}. We can then apply simple capacitor models of channel charging to determine the channel potential in Eq. (5.6).

If the potential in the channel changes by δU then the number of charges in the channel changes by

$$\delta N = -g\left(E_F\right)\delta U \tag{5.7}$$

Note that we have assumed $T = 0$K, and note also the negative sign – making the channel potential more negative increases the number of charges.

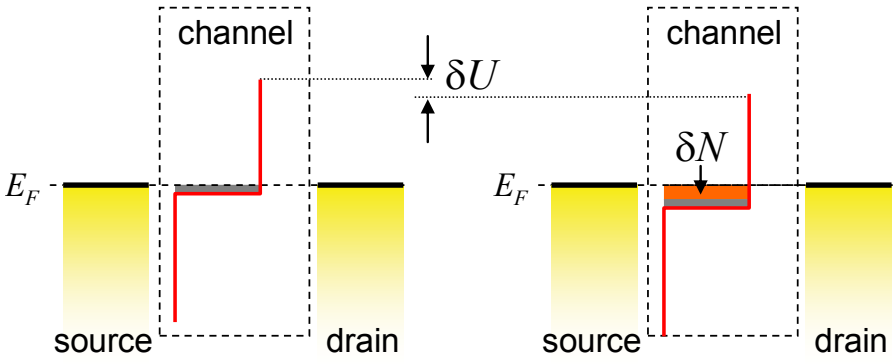


Fig. 5.8. A shift in the channel potential changes the number of charges in the channel.

Substituting back into Eq. (5.4) gives

$$\delta U = -q\delta V_{GS}\frac{C_G}{C_{ES}} - q\delta V_{DS}\frac{C_D}{C_{ES}} - \frac{q^2}{C_{ES}}g\left(E_F\right)\delta U \tag{5.8}$$

Collecting δU terms gives

$$\delta U = -q\delta V_{GS}\frac{C_G}{C_{ES}+C_Q} - q\delta V_{DS}\frac{C_D}{C_{ES}+C_Q} \tag{5.9}$$

Where we recall the **quantum capacitance** (C_Q):

$$C_Q = q^2 g\left(E_F\right) \tag{5.10}$$

Using the quantum capacitance, we can easily construct a small signal model for changes in V_{GS} or V_{DS}. See, for example the small signal V_{GS} model in Fig. 5.9. Note that the value of the quantum capacitance depends on the channel potential at the bias point.

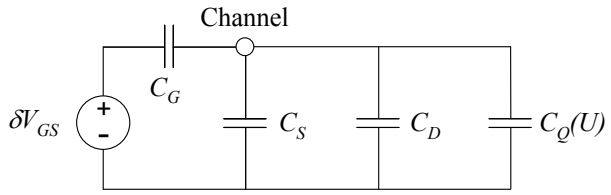


Fig. 5.9. A small signal model for the channel potential.

Using Eq. (5.7) we can also determine a small signal model for the charge in the channel.

$$q\delta N = \delta V_{GS}\frac{C_G C_Q}{C_{ES}+C_Q} \tag{5.11}$$

In the next section we will consider FET operation under two limiting cases: (i) when C_Q is large relative to C_{ES}, and (ii) when C_Q is small. The two cases typically correspond to the ON and OFF states of a FET, respectively.

Simplified models of FET switching

To further simplify the problem, we define two quantities, N_S and N_D, the charges injected into the channel from the source and drain contacts, respectively. Next, we assume that $\tau = \tau_S + \tau_D$, where $\tau_S = \tau_D$ and $C_G >> C_S$, C_D, Eqs (5.4), (5.5) and (5.6)become

$$U = -qV_{GS} + \frac{q^2}{C_G}\left(N - N_0\right) \tag{5.12}$$

$$N = \frac{N_S + N_D}{2} \tag{5.13}$$

$$I = \frac{q}{\tau}\left(N_S - N_D\right) \tag{5.14}$$

where

$$N_S = \int_{-\infty}^{\infty} g\left(E-U\right) f\left(E,\mu_S\right) dE \tag{5.15}$$

$$N_D = \int_{-\infty}^{+\infty} g\left(E-U\right) f\left(E,\mu_D\right) dE \tag{5.16}$$

Conduction in the FET is controlled by the number of electron states available to charges injected from the source. For switching applications, transistors must have an OFF state

where I_{DS} is ideally forced to zero. The OFF state is realized by minimizing the number of empty states in the channel accessible to electrons from the source. In the limit that there are no available states, the channel is a perfect insulator.

Switching between ON and OFF states is achieved by using the gate to push empty channel states towards the source chemical potential. The transition between ON and OFF states is known as the threshold. Although the transition is not sharp in every channel material, it is convenient to define a gate bias known as the *threshold voltage*, V_T, where the density of states at the source chemical potential $g(E_F)$ undergoes a transition.

The Zero Charging limit

As we saw in Part 3, charging-induced shifts in the energy levels of conductors can significantly complicate the calculation of IV characteristics. Equation (5.11) demonstrates that charging can be neglected if the quantum capacitance is much smaller than the electrostatic capacitance, *i.e.* $C_Q << C_{ES}$. For example, in Eq. (5.11), if $C_Q << C_{ES}$ then the charging, $\delta N \rightarrow 0$.

In the zero charging limit, Eq. (5.12) reduces to

$$U = -qV_{GS} \tag{5.17}$$

i.e. in this limit the channel potential simply tracks the gate bias.

Thus, in the zero charging limit, we can determine the current directly from Eq. (5.6), with the channel potential $U = -qV_{GS}$.

The zero charging limit almost always holds for insulators and transistors in the OFF state because the density of states at the Fermi level is small in both these examples. Determining whether a transistor remains in the zero charging limit in the ON state requires a comparison of C_Q and C_{ES}. Bulk devices very rarely operate within the zero charging limit in the ON state. But many small conductors contain relatively few states at the Fermi level even in the ON state, such that $C_Q << C_{ES}$ even when significant channel current is flowing.

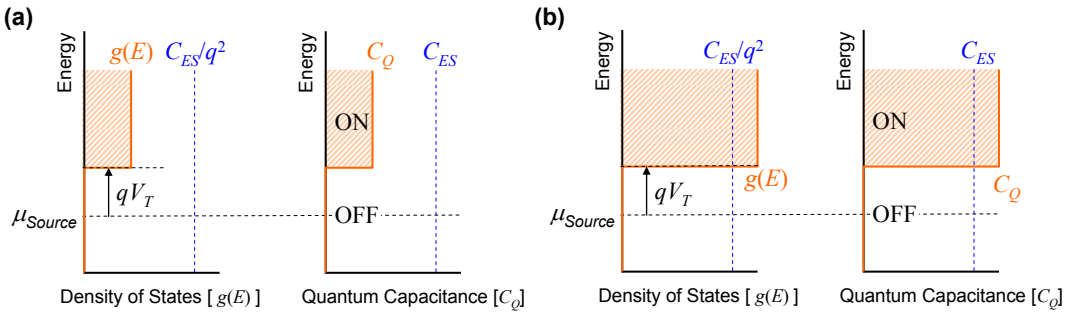


Fig. 5.10. We consider a channel material with a sharp transition in its density of states. In **(a)** we show a channel which remains in the insulator limit even in the ON regime. In **(b)** the channel states have sufficient density for the channel to be metallic in the ON regime.

The Strong Charging Limit

In metals and many large transistors in the ON state, the density of states at the Fermi level is sufficiently large that adding charges barely moves the channel potential. We say that the Fermi level is *pinned*. In the limit that $g(E_F) \to \infty$ then $\Delta U \to 0$.

In terms of the quantum capacitance, we find that if $C_Q >> C_{ES}$, Eq. (5.11) reduces to

$$q\delta N = \delta V_{GS} C_G \tag{5.18}$$

This limit is also known as **strong inversion** in conventional FET analysis. The channel transforms from an insulator to a metal. The transition occurs when the gate bias equals the threshold voltage, V_T, which is defined as the gate bias required to push the channel energy level down to the source workfunction.

In the strong charging/metallic limit, the gate and channel act as two plates of a capacitor. The charge in the channel then changes linearly with additional gate bias. In FETs, the channel potential relative to the source, $V(x)$, may also vary with position. Including the channel potential and the threshold voltage in Eq. (5.18) yields:

$$qN = C_G\left(V_{GS} - V_T - V(x)\right) \tag{5.19}$$

One way to interpret the metallic limit is to consider the difference between the actual position of the conduction band edge and its position in the absence of charging. The difference is proportional to the amount of charging; see Fig. 5.11. Note that the shaded region in the figure does not represent filled electron states below the conduction band. These electrons are in fact all at the bottom of the conduction band. Rather this the same graphical tool that we used in Part 3 to analyze charging within conductors.

The metallic or strong inversion limit is only maintained for $V_{GS} - V_T - V(x) > 0$. If $V(x)$ crosses the zero charging limit (V_{GS}-V_T) then charging decreases to zero. This is known as „pinch off".

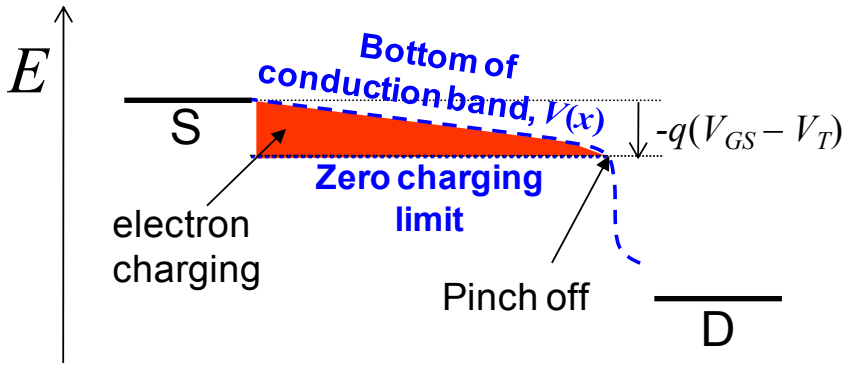


Fig. 5.11. Charging opposes gate-induced changes in the channel potential. Here we illustrate the effect of charging in a non-ballistic FET. In the absence of charging, increasing gate potential lowers the conduction band edge (the „zero charging limit"). Charging pushes the conduction band edge back up towards the source workfunction. The red shaded region is the difference between the actual channel potential and the zero charging limit. It represents the population of electrons in the conduction band. It does not represent filled states below the bottom of the conduction band.

The temperature dependence of current in the OFF state

Both nanoscale and larger transistors have a small quantum capacitance in the OFF state, which is also known as subthreshold since $V_{GS} < V_T$.

But even if the density of states is zero between $\mu_S > E > \mu_D$, at higher temperatures, some electrons may be excited into empty states well above the Fermi Energy. If the density of states is very low at the Fermi Energy, but higher far from the Fermi level, then we can model the Fermi distribution by an exponential tail. Recall that this is known as a non-degenerate distribution; see Fig. 5.12.

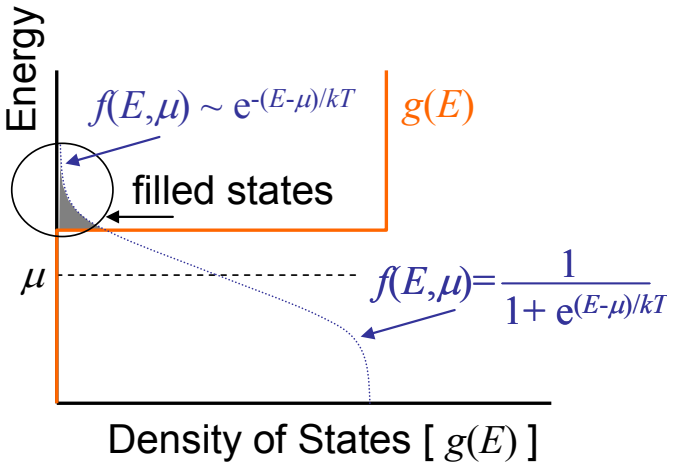


Fig. 5.12. If only the extreme tail states of the Fermi distribution are filled, then we can model the distribution by an exponential. This is common when the density of states at the Fermi Energy is small.

Equation (5.14) becomes

$$N_S = \int_{-\infty}^{\infty} g(E-U)\, e^{-(E-\mu_S)/kT}\, dE \tag{5.20}$$

Now changing the variable of integration to $E' = E - U$

$$N_S = \int_{-\infty}^{\infty} g(E')\, e^{-(E'+U-\mu_S)/kT}\, dE' \tag{5.21}$$

Simplifying

$$N_S = e^{-U/kT} \int_{-\infty}^{\infty} g(E')\, e^{-(E'-\mu_S)/kT}\, dE' \tag{5.22}$$

Similarly,

$$N_D = e^{-U/kT} \int_{-\infty}^{\infty} g(E')\, e^{-(E'-\mu_D)/kT}\, dE' \tag{5.23}$$

Thus, from Eq. (5.14) the current is

$$I = \frac{q}{\tau} \exp\left[\frac{qV_{GS}}{kT}\right] \cdot \int_{-\infty}^{\infty} g(E') \left(e^{-(E'-\mu_S)/kT} - e^{-(E'-\mu_D)/kT} \right) dE' \tag{5.24}$$

Equation (5.24) holds in the limit that $C_G >> C_S, C_D$. In general, we find that the current in the subthreshold region is

$$I = I_0 \exp\left[\frac{qV_{GS}}{kT}\frac{C_G}{C_{ES}}\right] \tag{5.25}$$

Taking logarithm of both sides we find,

$$\log_{10} I = \frac{q}{kT}\frac{C_G}{C_{ES}}(\log_{10} e)V_{GS} + \log_{10} I_0 \tag{5.26}$$

The slope, S, of the subthreshold regime is usually expressed gate volts per decade of drain current. At room temperature, the optimum, when $C_G >> C_S, C_D$, is

$$S = \frac{kT}{q}\frac{1}{\log_{10} e} \approx 60 \text{ mV/decade} \tag{5.27}$$

The slope becomes much sharper at low temperatures; see Fig. 5.13.

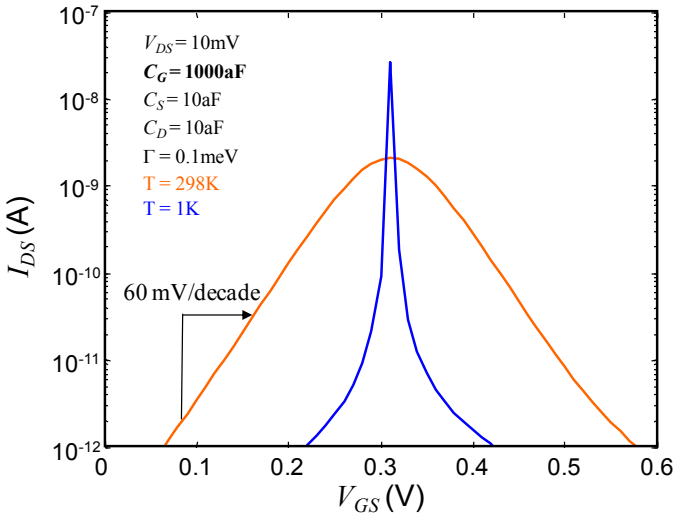


Fig. 5.13. A comparison of the switching characteristics of our C60 model FET at T = 1K and room temperature. In the OFF regime, the current varies exponentially with gate bias, i.e. a straight line on a log-linear plot. The slope at room temperature is 60 mV/decade of drain current.

Transconductance

A field effect transistor is a voltage controlled current source. Its input is the gate potential, and the output is the source-drain current. In applications, we usually desire that the FET amplifies small changes in V_{GS}. Thus, an important figure of merit for an FET is the transconductance, defined as

$$g_m = \frac{dI_{ds}}{dV_{gs}} \tag{5.28}$$

In the OFF state, the quantum capacitance is small and the gate‟s only influence on the FET is its electrostatic control of the channel potential. From Eq. (5.1), we see that this control is maximized when

$$C_G >> C_S, C_D \tag{5.29}$$

This is an important design goal for FETs. Under this limit the transconductance is commonly expressed as:

$$\frac{g_m}{I_{ds}} = \frac{1}{I_{ds}} \frac{dI_{ds}}{dV_{gs}} = \frac{q}{kT} \tag{5.30}$$

Good electrostatic control of the channel may be achieved either by increasing the dielectric constant of the gate insulator, or by reducing the thickness of the gate insulator.

A good rule of thumb is that the gate must be much closer to the channel than either the source or drain contacts. Fig. 5.14 shows the impact of varying C_G on our C60 molecular transistor. Increasing C_G shifts the switching gate voltage much lower.

Interestingly, to obtain this ideal characteristic, we increased C_G by three orders of magnitude relative to the more practical value used originally. This corresponds to increasing dielectric constant or reducing the gate-channel separation by three orders of magnitude.

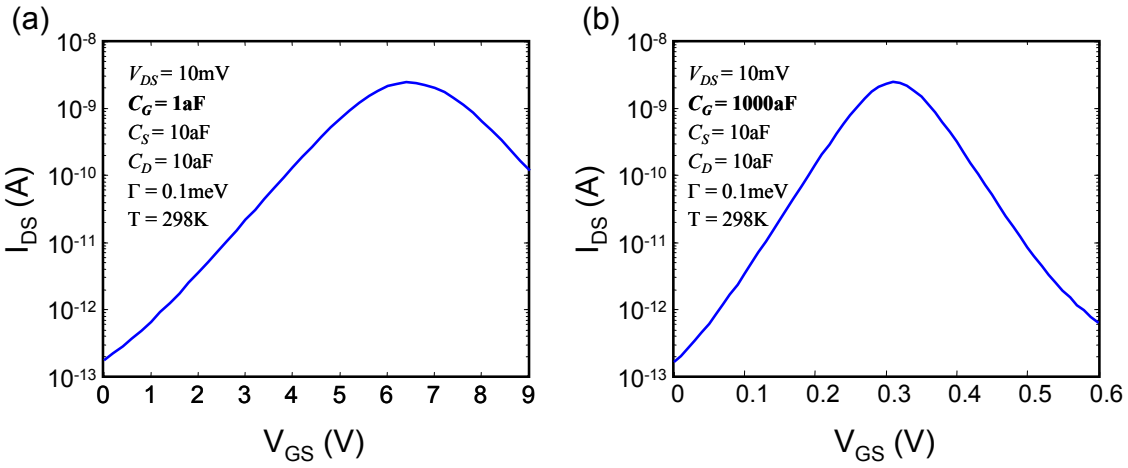


Fig. 5.14. A comparison of two C60 FETs. In **(a)** the gate has poor electrostatic control over the channel as evidenced by the small C_G. In **(b)** the control is better, and the switching voltage is much lower.

For a molecular transistor with source-drain separation of a few nanometers, the gate insulator should be only a few Ångstroms – too thin to sufficiently insulate the gate. This represents a possibly insurmountable obstacle to 0-d channel devices such as single molecule FETs.

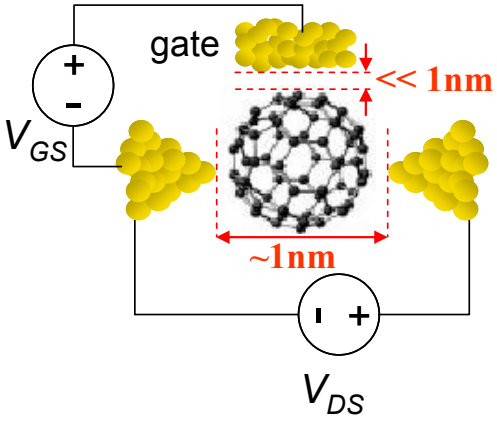


Fig. 5.15. For high transconductance, the gate capacitance must be much higher than the source or drain channel capacitances. This forces impractically small gate-channel separations in molecular transistors.

(ii) 1d and 2d FETs

The central equation of conduction is

$$I = \frac{q\left(N_S - N_D\right)}{\tau} \tag{5.31}$$

In 0-d the time constant, τ, was defined as the sum of the interfacial electron transfer time τ_S and τ_D, which in turn can be thought of as representations of the interaction energy between the 0-d conductor and the source and drain contacts: $\tau_S = \Gamma_S/\hbar$ and $\tau_D = \Gamma_D/\hbar$, respectively.

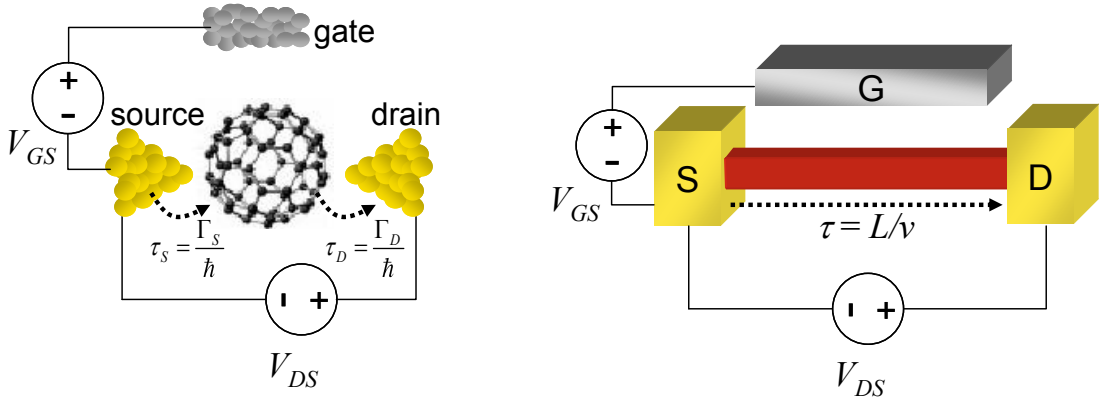


Fig. 5.16. (a) Electron transfer times in a 0-d conductor are related to the interaction energy Γ between the contact and the conductor. **(b)** In higher dimensions, we must determine the transit time from the electron velocity.

In higher dimensions, however, the electron transfer times at the contacts are less important. Rather, τ is the transit time for an electron in the conductor. It is given by

$$\tau = \frac{L_x}{v_x} \tag{5.32}$$

where L_x is the length of the channel, and v_x is the velocity component of the electron parallel to the source-drain current. It is important to note that in 1-d, 2-d and 3-d conductors the transit time is dependent on the energy of the electron since the electron velocity, v_x, is dependent on energy.

The other important change from the 0-d model concerns the density of states. In 0-d all states are accessible to electrons from both the source and drain contacts. But in higher dimensional ballistic devices, electrons injected from the source are only able to access states with momenta directed away from the source. We call these $+k$ states. Similarly, the drain only injects electrons into $-k$ states. Thus, we break the dispersion relation and density of states into two pieces, the density of $+k$ states is given by $g^{+}(E)dE$ and the density of $-k$ states is given by $g^{-}(E)dE$.

To summarize, in 1-d, 2-d and 3-d the fundamental equations for a transistor are:

$$U_{ES} = -qV_{GS}\frac{C_G}{C_{ES}} - qV_{DS}\frac{C_D}{C_{ES}} + \frac{q^2}{C_{ES}}\left(N - N_0\right) \tag{5.33}$$

$$N = N_S + N_D \tag{5.34}$$

$$I = \frac{q}{\tau}\left(N_S - N_D\right) \tag{5.35}$$

where

$$N_S = \int_{-\infty}^{\infty} g^{+}\left(E-U\right) f\left(E,\mu_S\right) dE \tag{5.36}$$

$$N_D = \int_{-\infty}^{+\infty} g^{-}\left(E-U\right) f\left(E,\mu_D\right) dE \tag{5.37}$$

The ballistic quantum wire FET.[†]

Consider the ballistic quantum wire FET shown in Fig. 5.17.

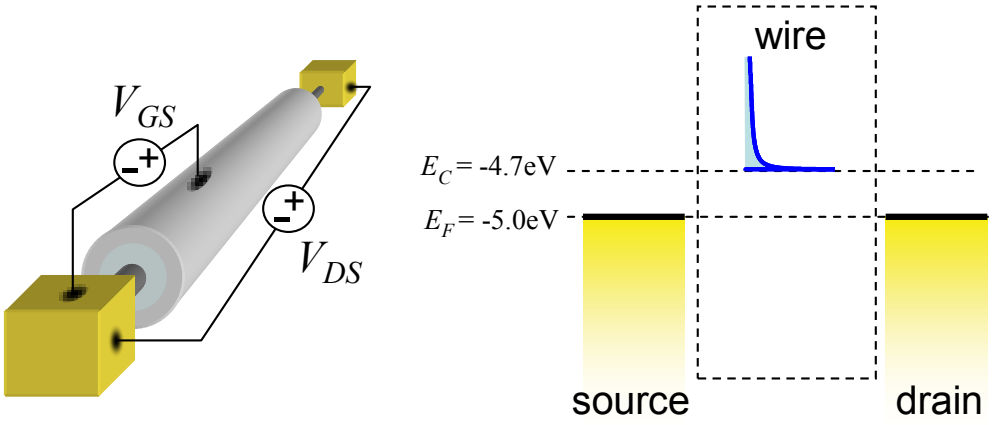


Fig. 5.17. A quantum wire FET. The gate is wrapped around the wire to maximize the capacitance between the channel and the gate. The length of the wire is L = 100nm, the gate capacitance is C_G = 50 aF per nanometer of wire length, and the electron mass, m, in the wire is $m = m_0$=9.1x10^{-31} kg.

We will assume that there is only one parabolic band in the wire.

From Eq. (2.37), the density of states in the wire is:

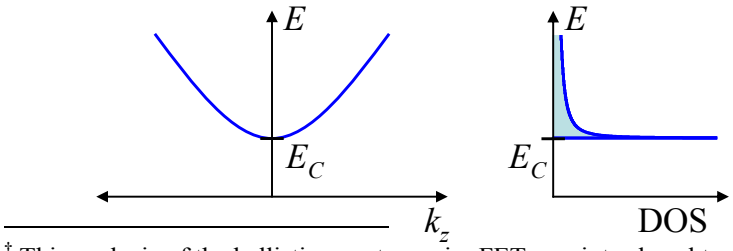


Fig. 5.18. The bandstructure and density of states in a single mode quantum wire.

[†] This analysis of the ballistic quantum wire FET was introduced to me by Mark Lundstrom at Purdue University. For a complete reference see Mark Lundstrom and Jing Guo, „Nanoscale Transistors: Physics, Modeling, and Simulation", Springer, New York, 2006.

$$g(E)dE = \frac{2L}{h}\sqrt{\frac{2m}{E-E_C}}u(E-E_C)dE, \tag{5.38}$$

where L is the length of the wire, and m is the electron mass in the wire. But only half of these states contain electrons traveling in the positive direction. Thus, we must divide Eq. (5.38) by two to yield:

$$g^+(E)dE = \frac{1}{2}\times\frac{2L}{h}\sqrt{\frac{2m}{E-E_C}}u(E-E_C)dE \tag{5.39}$$

Given the position of the Fermi Energy, this band is the conduction band. We will label the energy at the bottom of the conduction band, E_C. Since we model electrons moving along the wire as plane waves, within the parabolic band we have

$$E-E_C = \frac{\hbar^2k^2}{2m} = \frac{1}{2}mv^2 \tag{5.40}$$

We can rewrite Eq. (5.39) in terms of the velocity, v, of the electron:

$$g^+(E)dE = \frac{1}{2}\times\frac{4L}{hv(E)}u(E-E_C)dE \tag{5.41}$$

Now L/v is the transit time of an electron through the wire, thus

$$g^+(E)dE = \frac{1}{2}\times\frac{4\tau(E)}{h}u(E-E_C)dE\,. \tag{5.42}$$

We can substitute Eq. (5.42) into the expression for the current density (Eq. (5.31)) to obtain

$$I = \frac{2q}{h}\int_{-\infty}^{+\infty}u(E-E_C-U)\left(f(E,\mu_S)-f(E,\mu_D)\right)dE\,. \tag{5.43}$$

Quantum dot models of quantum wire transistor channels

Under bias we expect a spatial variation in the potential along a quantum wire. Current flow may also vary the charge density along the wire, which in turn affects the potential profile. Thus, the potential variation must be determined self consistently with the current flow.

We have seen in Part 4 that the conduction band edge in a ballistic conductor is determined by the point of maximum potential in the conductor. For electrostatic purposes, we will approximate this point on the quantum wire as a quantum dot, and then employ our discrete capacitive models of potential to calculate changes in the conduction band edge.

Usually the highest potential is located next to the source, because application of forward bias at the drain pulls the potential down along the channel.

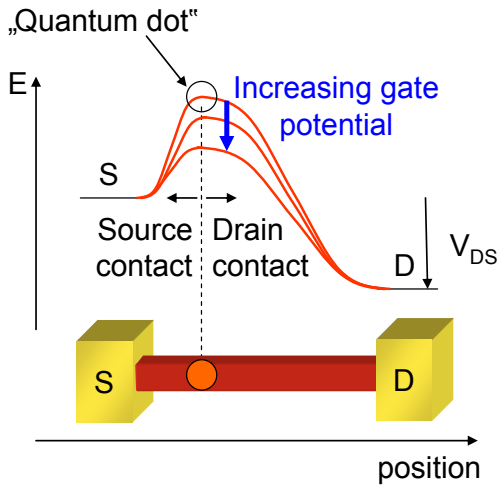


Fig. 5.19. An example of a typical potential profile along the length of a quantum wire as reflected by the bottom of the conduction band. The point of maximum potential acts as a barrier to the flow of current. As the gate bias increases, the barrier decreases, enhancing current flow. The point of maximum potential is modeled as a quantum dot with the same density of states as the quantum wire. The remainder of the wire is considered to be part of the contacts.

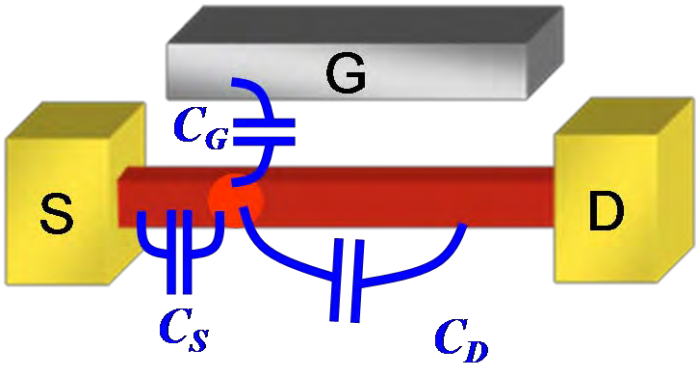


Fig. 5.20. Assuming the channel is modeled by a quantum dot we model the electrostatics of the transistor using capacitors.

Ballistic Quantum Wire FET Current-Voltage Characteristics at T = 0K.

The electrostatic capacitances are shown in Fig. 5.20 using the quantum dot model of the quantum wire. In this example we ignore source and drain capacitances. The gate capacitor was defined in Fig. 5.17 as C_G = 1 aF per nanometer of gate length.

We compare quantum and electrostatic capacitances in Fig. 5.21, we find that the single mode wire has relatively few states, hence its quantum capacitance is small, and above the band edge it operates in the zero charging/insulator regime; even in the ON state charging effects are negligible and we can take $U = -qV_{GS}$.

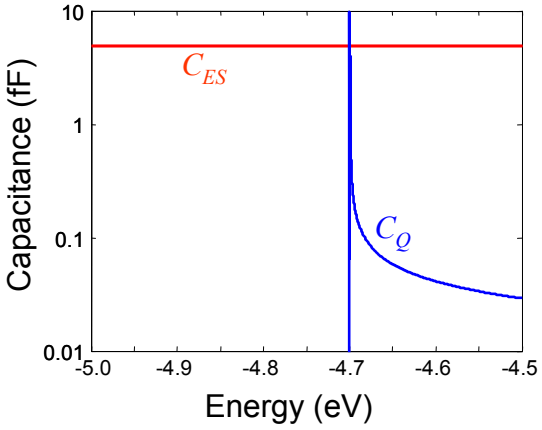


Fig. 5.21. A comparison between the electrostatic and quantum capacitances shows that that $C_Q >> C_{ES}$ only at the conduction band edge. But as the channel fills with charge, the wire returns to the insulator regime.

In forward bias (when the drain potential is lower than the source), there are three regimes of operation:

(a) OFF: $V_{GS} < V_T$

Let"s define the threshold voltage as the potential difference between the source and the conduction band minimum. Thus, in this example, $V_T = 0.3$ V. Recall that the gate potential is relative to the source potential. So when $V_{GS} < V_T$, electrons cannot be injected from the source. Hence no current can flow for positive drain voltages. This is the OFF state of the FET.

Note that source drain current can flow for $T > 0$K since the tail of the Fermi distribution for electrons in the source overlaps with states in the wire. The current follows Eq. (5.25).

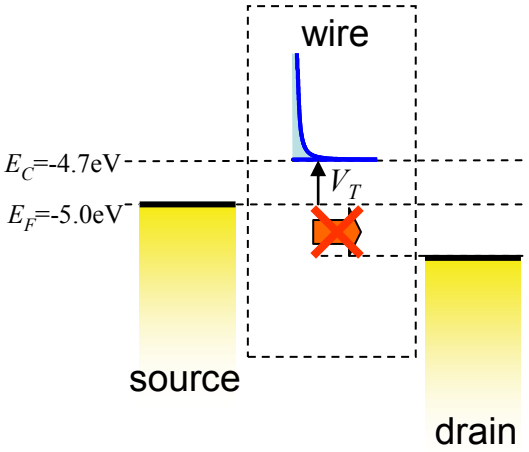


Fig. 5.22. Energy line up for FET in the OFF state. There are no channel states between the source and drain chemical potentials.

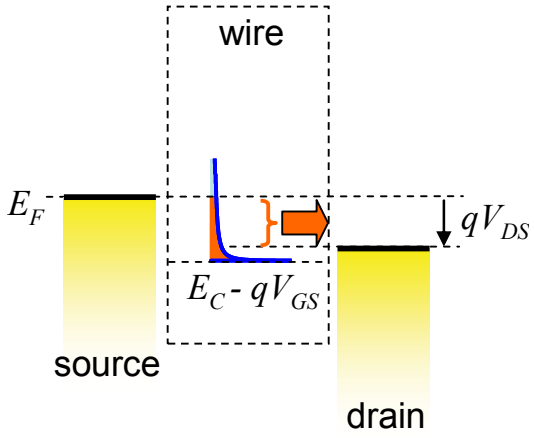
(b) The linear regime: $V_{GS} > V_T$, $V_{DS} < V_{GS}$-V_T

This is known as the linear regime because the current scales linearly with the drain source potential. Equation (5.43) reduces to

$$I_{DS} = \frac{2q^2}{h} V_{DS} \tag{5.44}$$

Note that the FET exhibits the quantum limit of conduction in this regime. Its transconductance, however, is zero.

Fig. 5.23. In the linear regime, the current is limited by the source drain potential.



(c) Saturation: $V_{GS} > V_T$, $V_{DS} > V_{GS}-V_T$

Once the drain potential exceeds V_{GS}-V_T, all the charge in the channel is uncompensated and injected into the drain. Thus, the current is limited by the gate potential. This is known as saturation.

$$I_{DS} = \frac{2q^2}{h}(V_{GS} - V_T) \tag{5.45}$$

The transconductance for a single mode wire in saturation is

$$g_m = \frac{2q^2}{h} \tag{5.46}$$

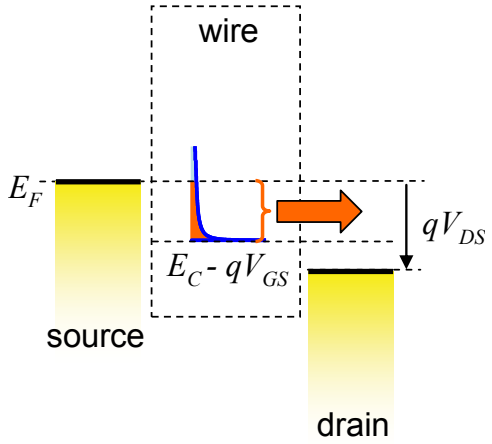


Fig. 5.24. In the saturation regime, the current is limited by the gate source potential.

Fig. 5.25 plots the forward bias characteristics of the FET both at T = 0K, and room temperature. At room temperature, the characteristics were determined numerically since the transition from linear to saturation regimes is blurred by thermal activation of electrons above the Fermi level.

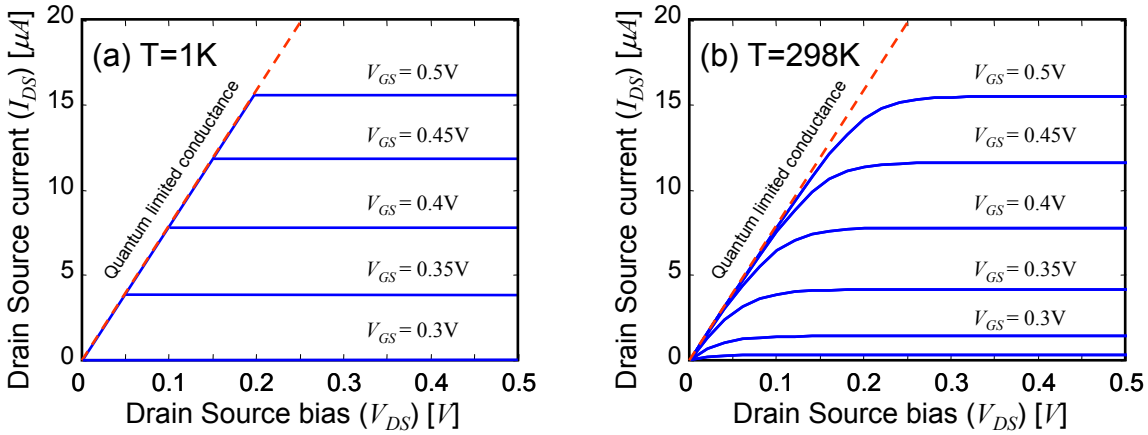


Fig. 5.25. Forward bias characteristics for a quantum wire FET at **(a)** T = 0K, and **(b)** room temperature.

Part 5. Field Effect Transistors

Ballistic Quantum Well FETs

To analyze the ballistic quantum well FET, let‟s begin with the master equation for current.

$$I = \frac{q\left(N_S - N_D\right)}{\tau} = \frac{q\left(N_S - N_D\right)v_x}{L} \tag{5.47}$$

We have defined conduction in the *x*-direction and the transit time is given by $\tau = L_x / v_x$.

Let‟s begin by considering the product $g.v_x$, which we will integrate to get $(N_S - N_D)v_x$. In *k*-space and circular coordinates, this is

$$g\left(k\right)v_x\left(k\right)kdkd\theta = 2\frac{1}{\left(2\pi\right)^2 / LW}\frac{\hbar k_x}{m}kdkd\theta \tag{5.48}$$

Simplifying further gives:

$$g\left(k\right)v_x\left(k\right)kdkd\theta = \frac{LW}{2\pi^2}\frac{\hbar}{m}k^2 dk\sin\theta d\theta \tag{5.49}$$

Converting the variable of integration back to energy using the dispersion relation $E = \hbar^2 k^2 / 2m + E_C + U$, and assuming conduction in just a single mode of the quantum well, yields

$$g\left(k\right)v_x\left(k\right)kdkd\theta = \frac{LW}{2\pi^2\hbar^2}\sqrt{2m\left(E - E_C - U\right)}u\left(E - E_C - U\right)dE\sin\theta d\theta \tag{5.50}$$

Substituting back into Eq. (5.47) and integrating over the +*k* hemisphere ($0 < \theta < \pi$) gives

$$I = \frac{qW}{\pi^2\hbar^2}\int_{-\infty}^{\infty}\sqrt{2m\left(E - E_C - U\right)}u\left(E - E_C - U\right)\left(f\left(E,\mu_S\right) - f\left(E,\mu_D\right)\right)dE \tag{5.51}$$

Below threshold the density of states is zero. Thus,

$$U = -q\eta^0 V_{GS} \tag{5.52}$$

where we neglect the effect of V_{DS}, and

$$\eta^0 = \frac{C_G}{C_S + C_D + C_G}. \tag{5.53}$$

The threshold voltage, V_T, is defined as the gate-source voltage required to turn the transistor ON, *i.e.* bring the bottom of the conduction band, E_C, down to the source workfunction. From Eq. (5.52) and requiring the $E_C + U = \mu_S$ at threshold, we get

$$V_T = \left(E_C - \mu_S\right)/\eta^0 q \tag{5.54}$$

Above threshold the density of states and hence the quantum capacitance is constant. Thus, the quantum well FET is the rare case where we can model charging phenomena *analytically*. Above threshold we have

$$U = -\eta q\left(V_{GS} - V_T\right) - \eta^0 qV_T \tag{5.55}$$

where again we neglect the effect of V_{DS}, and

$$\eta = \frac{C_G}{C_S + C_D + C_G + C_Q}. \tag{5.56}$$

Fom the 2-d DOS in Eq. (2.47), C_Q for a single mode quantum well is

$$C_Q = \frac{1}{2} q^2 \frac{mWL}{\pi\hbar^2}. \tag{5.57}$$

where we have only considered half the usual density of states (the $+k$ states). This is accurate in the saturation region because the drain cannot fill any states in the channel. The quantum capacitance increases in the linear region as the drain fills some $-k$ states leading to errors in the calculation of the current in the linear regime.

Noting that $V_T = E_C / \eta^0 q$ we can rewrite Eq. (5.55) above threshold as

$$E_C + U = \mu_S - \eta q\left(V_{GS} - V_T\right). \tag{5.58}$$

Now, we can simplify Eq. (5.51) to give us

$$I = \frac{qW}{\pi^2\hbar^2} \int_{\mu_S - \eta q(V_{GS}-V_T)}^{\infty} \sqrt{2m\left(E + \eta q\left(V_{GS} - V_T\right)\right)}\left(f\left(E,\mu_S\right) - f\left(E,\mu_D\right)\right)dE \tag{5.59}$$

At $T = 0$K, we can solve Eq. (5.59) in the linear regime ($V_{DS} < \eta(V_{GS} - V_T)$):

$$\begin{aligned} I &= \frac{qW}{\pi^2\hbar^2} \int_{\mu_S - qV_{DS}}^{\mu_S} \sqrt{2m\left(E - E_C + \eta q V_{GS}\right)}dE \\ &= \frac{qW}{\pi^2\hbar^2}\sqrt{\frac{8m}{9}}(\eta q)^{3/2}\left[\left(V_{GS} - V_T\right)^{3/2} - \left(V_{GS} - V_T - V_{DS}/\eta\right)^{3/2}\right] \end{aligned} \tag{5.60}$$

and in the saturation regime ($V_{DS} > \eta(V_{GS} - V_T)$):

$$\begin{aligned} I &= \frac{qW}{\pi^2\hbar^2} \int_{\mu_S - \eta q(V_{GS}-V_T)}^{\mu_S} \sqrt{2m\left(E + \eta q\left(V_{GS} - V_T\right)\right)}dE \\ &= \frac{qW}{\pi^2\hbar^2}\sqrt{\frac{8m}{9}}(\eta q)^{3/2}\left(V_{GS} - V_T\right)^{3/2} \end{aligned}. \tag{5.61}$$

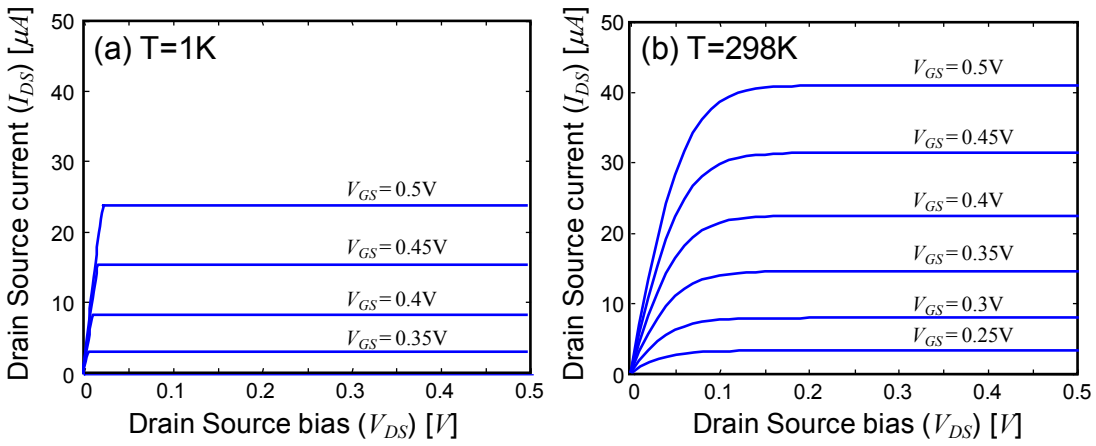


Fig. 5.26. Forward bias characteristics for a quantum well FET at **(a)** T = 1K, and **(b)** room temperature. The channel width is W = 120nm, and the electrostatic control over the channel is assumed to be ideal. Also, take m = 0.5×m_0.

Note that the saturation current goes as ~ $(V_{GS}\text{-}V_T)^{3/2}$ compared to the ballistic nanowire transistor, which goes as ~ $(V_{GS}\text{-}V_T)$. As we shall see, the conventional FET has a saturation current dependence of ~ $(V_{GS}\text{-}V_T)^2$.

(iii) Conventional MOSFETs

Finally, we turn our attention to the backbone of digital electronics, the non-ballistic metal oxide semiconductor field effect transistor (MOSFET).

The channel material is a bulk *semiconductor* – typically silicon. Here, we will consider a so-called n-channel MOSFET, meaning that the channel current is carried by electrons at the bottom of the conduction band of the semiconductor.

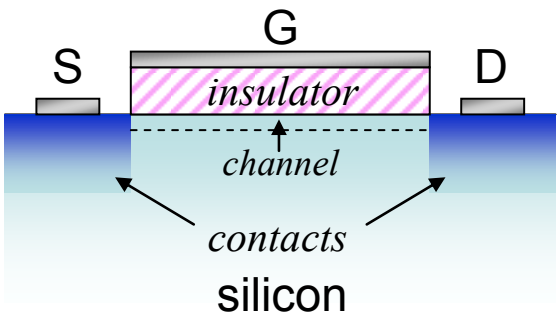


Fig. 5.27. An n-channel MOSET built on a silicon substrate. Phosphorous is diffused below the source and drain electrodes to form high conductivity contacts to the silicon channel beneath the insulator.

Now, let"s consider the various operating regimes of a conventional MOSFET.

(a) OFF: Subthreshold $V_{GS} < V_T$

Similar to the ballistic quantum wire FET, we can model channel current as injection over a barrier close to the source electrode.

Once again, let"s define the threshold voltage as the potential difference between the source Fermi energy and the conduction band minimum.[†]

As in the ballistic example, when $V_{GS} < V_T$, only the tail of the Fermi distribution for electrons in the source overlaps with empty states in the conduction band. The current follows Eq. (5.25).

$$I = I_0 \exp\left[\frac{qV_{GS}}{kT}\frac{C_G}{C_{ES}}\right] \tag{5.62}$$

Subthreshold characteristics determine the gate voltage required to switch the FET ON and OFF. From Eq. (5.27) the subthreshold slope is ideally 60mV/decade, meaning that a 60mV change in gate potential corresponds to a decade change in channel current.

[†] Actually, this is an overestimate of the threshold voltage because the density of states at the conduction band is so large that the transistor will often turn on when the Fermi level gets within a few *kT*. It also ignores the effect of charge trapped at the interface between the channel and the insulator.

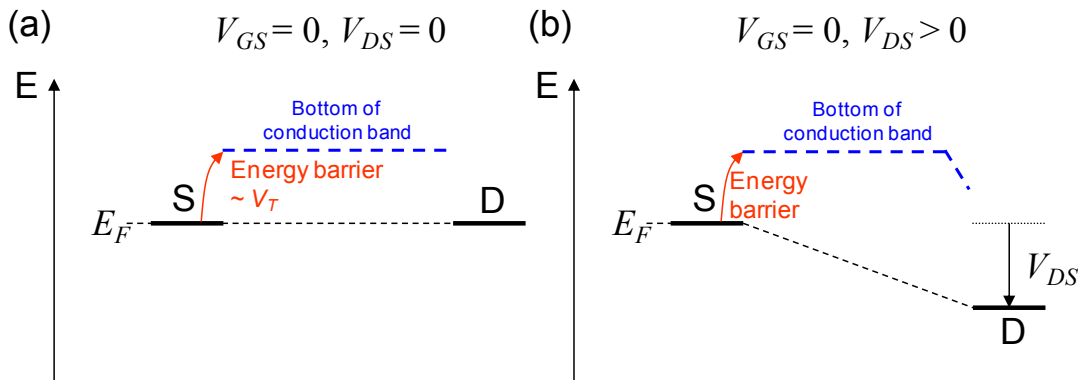


Fig. 5.28. Below threshold few electrons can be injected from the source into the conduction band, irrespective of the drain source potential.

(b) The linear regime: $V_{GS} > V_T$, $V_{DS} < V_{GS}-V_T$

As we shall see, this is known as the linear regime because the current scales linearly with the drain source potential. Consider a thin slice of the channel with width W, and length δx. For this analysis to hold, the length of this slice cannot be much shorter than the mean free path of the electron between scattering events. In a silicon transistor, we have shown that $\delta x >$ 50 nm (see the analysis associated with Fig. 4.23). Silicon transistors with channel lengths shorter than this should be analyzed in the ballistic regime.

Since the density of states above the conduction band is very large in a bulk semiconductor, a conventional MOSFET will enter the strong charging/metallic limit for $(V_{GS} - V) > V_T$, *i.e.* the number of charges, δN, in the slice is

$$q\delta N = \frac{C_G}{A} W \delta x \left(V_{GS} - V - V_T\right) \tag{5.63}$$

where $A = W.L$ is the surface area of the channel.

Now the current within the slice is given by

$$I = \frac{q\delta N}{\tau} \tag{5.64}$$

where τ is the lifetime of carriers within the channel slice.

Since scattering is important, we employ the classical model of charge transport to relate the charge carrier lifetime to velocity, v, and the length of the slice, δx.

$$I = q\delta N \frac{v}{\delta x} \tag{5.65}$$

Next we relate the charge carrier velocity to mobility

$$I = q\delta N \frac{\mu F}{\delta x} \tag{5.66}$$

Now, we must note that scattering causes the potential in the channel to vary with position. We define the channel potential $V(x)$ as a function of position in the channel. Thus, expressing the source-drain electric field in terms of the channel potential we have

$$I = q\frac{\mu}{\delta x}\delta N\frac{\delta V}{\delta x} \tag{5.67}$$

Next, we substitute Eq. (5.63) into Eq. (5.67), yielding

$$I = \mu\frac{C_G}{L}\left(V_{GS} - V - V_T\right)\frac{\delta V}{\delta x} \tag{5.68}$$

We solve this under the limit that $\delta x << L$, by integrating both sides with respect to x. Since the current is uniform throughout the channel, we obtain:

$$I.L = \int_0^L \mu\frac{C_G}{L}\left(V_{GS} - V - V_T\right)\frac{dV}{dx}dx \tag{5.69}$$

where L is the length of the channel. It is convenient to change the variable of integration on the righthand side to voltage. In the linear regime, the maximum channel potential is V_{DS}, hence:

$$I = \mu\frac{C_G}{L^2}\int_0^{V_{DS}}\left(V_{GS} - V - V_T\right)dV \tag{5.70}$$

The linear regime requires that the entire channel remains in the strong charging/metallic limit. This occurs if the gate to drain potential, V_{GD}, also exceeds V_T

$$V_{GD} > V_T \tag{5.71}$$

or we can re-write this as

$$V_{DS} < V_{GS} - V_T \tag{5.72}$$

Under this constraint, Eq. (5.70) yields

$$I = \mu\frac{C_G}{L^2}\left(\left(V_{GS} - V_T\right)V_{DS} - \frac{1}{2}{V_{DS}}^2\right) \tag{5.73}$$

It is standard to express this in terms of a gate capacitance *per unit channel area*, C_{OX}:

$$I = \mu\frac{W}{L}C_{OX}\left(\left(V_{GS} - V_T\right)V_{DS} - \frac{1}{2}{V_{DS}}^2\right) \tag{5.74}$$

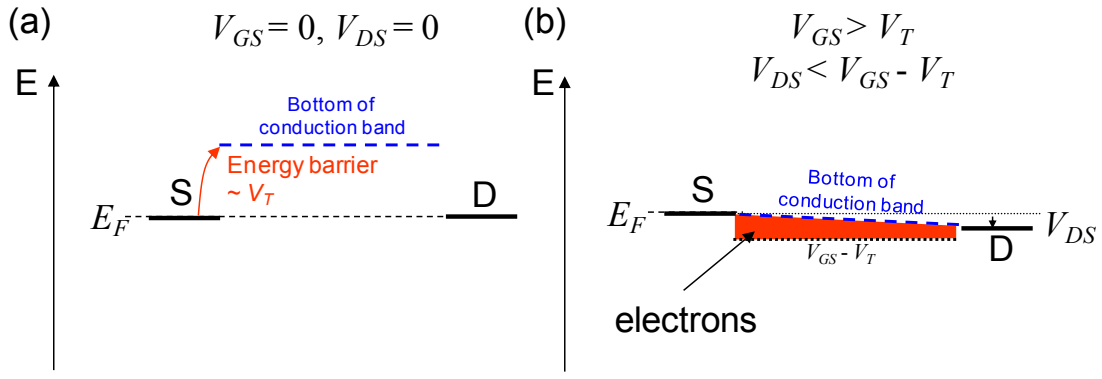


Fig. 5.29. Application of a gate source potential reduces the injection barrier between the source and the channel. The red shaded region represents the population of electrons in the conduction band. It does not represent filled states below the bottom of the conduction band.

(c) Saturation: $V_{GS} > V_T$, $V_{DS} > V_{GS}-V_T$

If the gate to drain potential exceeds threshold then the channel region close to the drain enters the zero charging regime, characterized by a high electric field and low density of mobile charges. The channel is said to pinch off and the current saturates because it is no longer dependent on V_{DS}. The strong charging/metallic region ends when the local channel potential $V = V_{GS} - V_T$

$$I = \int_0^{V_{GS}-V_T} \mu \frac{W}{L} C_{OX} \left(V_{GS} - V - V_T\right) dV \tag{5.75}$$

which gives

$$I = \mu \frac{W}{L} \frac{C_{OX}}{2} \left(V_{GS} - V_T\right)^2 \tag{5.76}$$

The *IV* characteristics of a non-ballistic MOSFET are shown in Fig. 5.31.

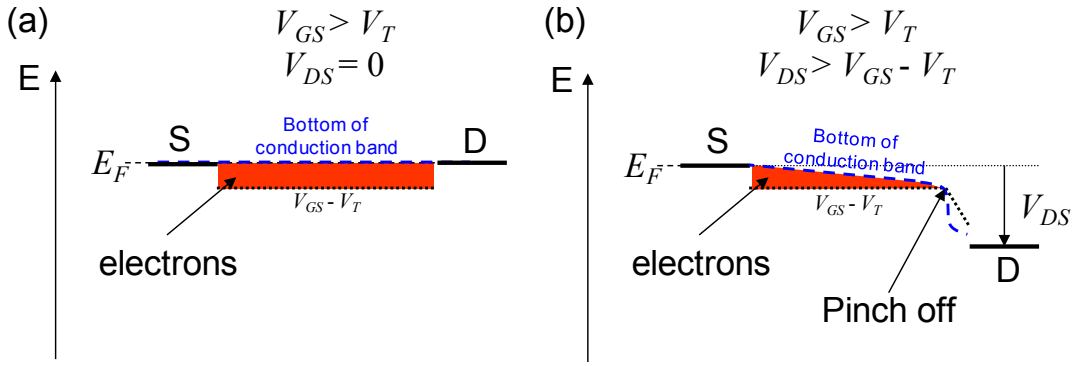


Fig. 5.30. As the drain-source potential increases, the channel near the drain enters the zero charging regime. The current is dependent only on the charge in the strong charging region, not on V_{DS}. The red shaded region represents the population of electrons in the conduction band. It does not represent filled states below the bottom of the conduction band.

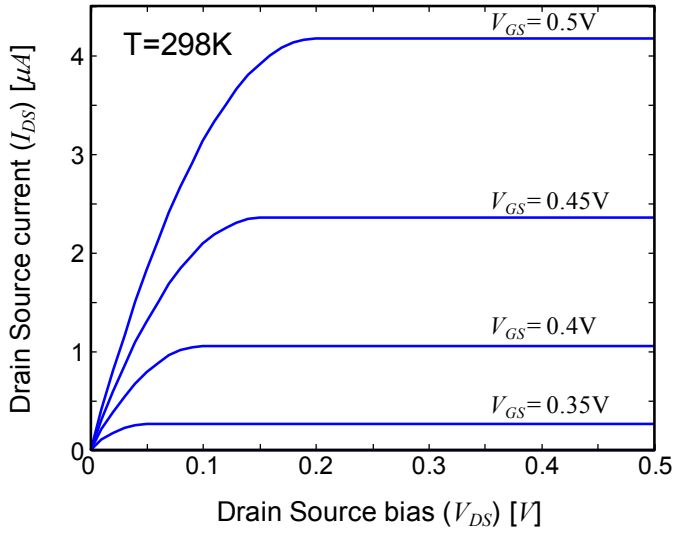


Fig. 5.31. The IV characteristics of a non-ballistic MOSFET with μ = 300 cm^2/Vs, L = 40nm, W = 3 × L, V_T = 0.3V, and C_G = 0.1 fF. Note that the use of the classical model for a transistor with such a short channel is inappropriate.

Comparison of ballistic and non-ballistic MOSFETs.

If we calculate the *IV* of a conventional MOSFET with a channel length in the ballistic regime, we obtain *IV* curves that are qualitatively similar to the ballistic result. For example, the classical model of a MOSFET with a channel length of 40 nm is shown in Fig. 5.31. It is qualitatively similar to Fig. 5.26. Both possess a linear and a saturation regime, and both exhibit identical subthreshold behavior. But the magnitude of the current differs quite substantially. The ballistic device exhibits larger channel currents due to the absence of scattering.

Another way to compare ballistic and non-ballistic MOSFETs is to return to the water flow analogy.† As before, the source and drain are modeled by reservoirs. The channel potential is modeled by a plunger. Gate-induced changes in the channel potential cause the plunger to move up and down in the channel. The most important difference between the ballistic and non-ballistic MOSFETs is the profile of the water in the channel. The height of the water changes in the non-ballistic device, whereas water in the ballistic channel does not relax to lower energies during its passage across the channel.

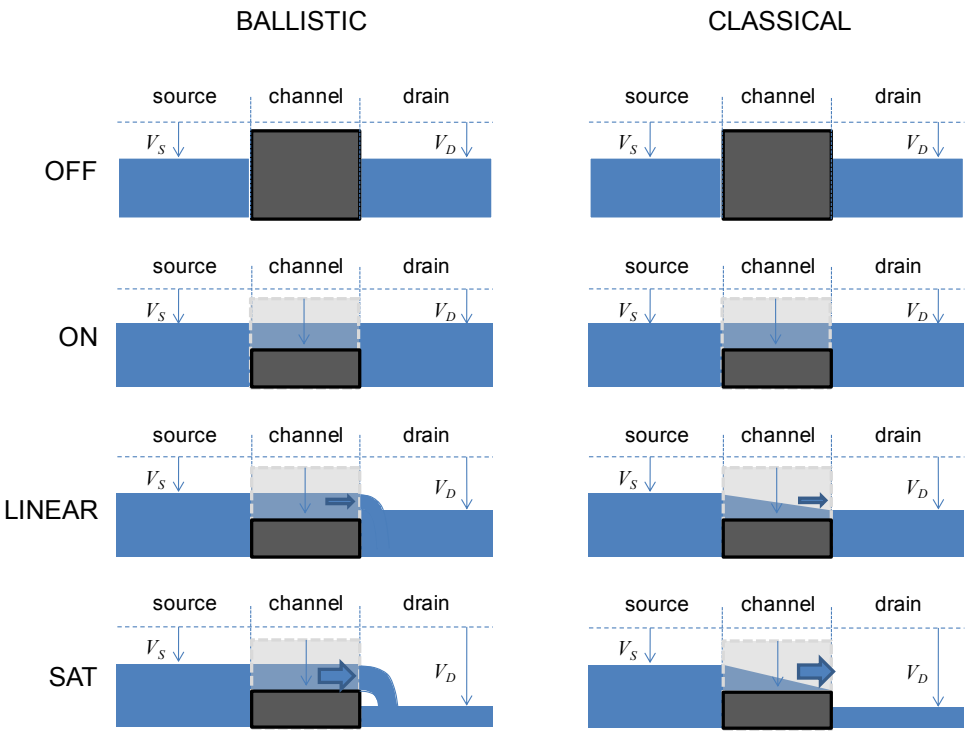


Fig. 5.32. The water flow analogy for the operation of ballistic and classical MOSFETs. Conduction in the channel is controlled by a plunger that models the channel potential. The transistors are turned ON by lowering the gate potential. Then, as the height of the drain reservoir decreases (corresponding to increased V_{DS}), the channel first enters the linear regime (where current flow is limited by V_{DS}) and then the saturation regime where the current is controlled only by the gate potential.

† For a more detailed treatment of the water analogy to conventional FETs see Tsividis, „Operation and Modeling of the MOS transistor", 2nd edition, Oxford University Press (1999).

Problems

1. Buckyball FETs

Park *et al.* have reported measurements of a buckyball (C60) FET. An approximate model of their device is shown in Fig. 5.33. The measured conductance as a function of V_{DS} and V_{GS} is shown in Fig. 5.34.

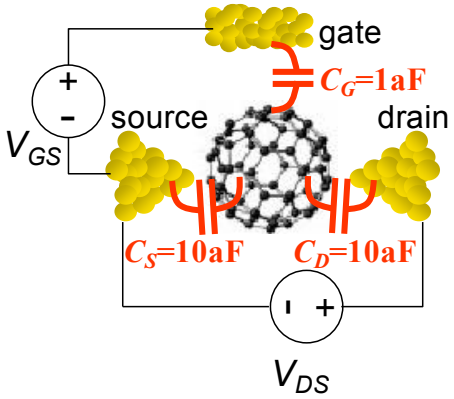


Fig. 5.33. The geometry of a C60 FET. In addition take the temperature to be T = 1K, and the molecular energy level broadening, Γ = 0.1eV. The LUMO is at -4.7 eV and the Fermi Energy at equilibrium is E_F = -5.0 eV

(a) Calculate the conductance (dI_{DS}/dV_{DS}) using the parameters in Fig. 5.33. Consider $5V < V_{GS} < 8V$ calculated at intervals of 0.2V and $-0.2V < V_{DS} < 0.2V$ calculated at intervals of 10mV.

(b) Explain the X-shape of the conductance plot.

(c) Note that Park, *et al.* measure a non-zero conductance in the upper and lower quadrants. Sketch their I_{DS}-V_{DS} characteristic at $V_{GS} \sim 5.9V$. Compare to your calculated I_{DS}-V_{DS} characteristic at $V_{GS} \sim 5.9V$. Propose an explanation for the non-zero conductances measured in the experiment in the upper and lower quadrants.

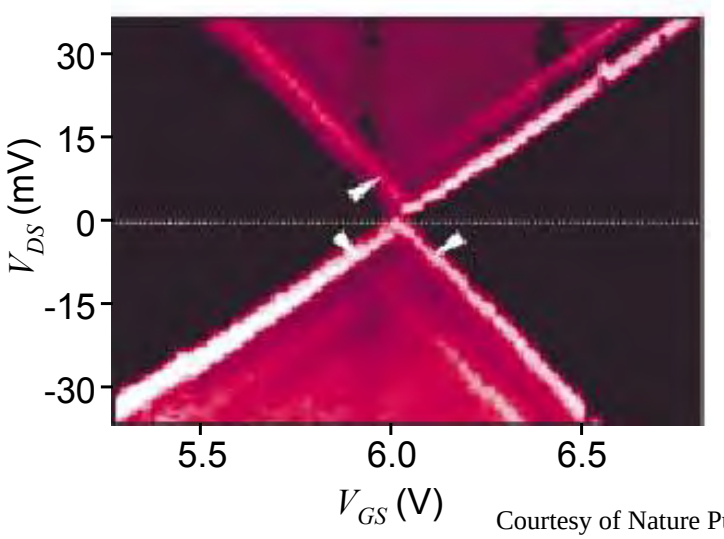


Fig. 5.34. The conductance (dI_{DS}/dV_{DS}) of a C60 FET as measured by Park *et al.* Ignore the three arrows on the plot. From Park, *et al.* "Nanomechanical oscillations in a single C60 transistor" Nature **407** 57 (2000).

(d) In Park *et al.*'s measurement the conductance gap vanishes at $V_{GS} = 6.0$V. Assuming that C_G is incorrect in Fig. 5.33, calculate the correct value.

Reference

Park, *et al*. "Nanomechanical oscillations in a single C60 transistor" Nature **407** 57 (2000)

2. Two mode Quantum Wire FET

Consider the Quantum Wire FET of Fig. 5.17 in the text.

Assume that the quantum wire has two modes at $E_{C1} = -4.7$eV, and $E_{C2} = -4.6$eV. Analytically determine the I_{DS}-V_{DS} characteristics for varying V_{GS} at $T = 0$K. Sketch your solution for $0 < V_{DS} < 0.5$, at $V_{GS} = 0.3$ V, 0.35 V, 0.4V, 0.45V and 0.5V.

Highlight the difference in the IV characteristics due to the additional mode.

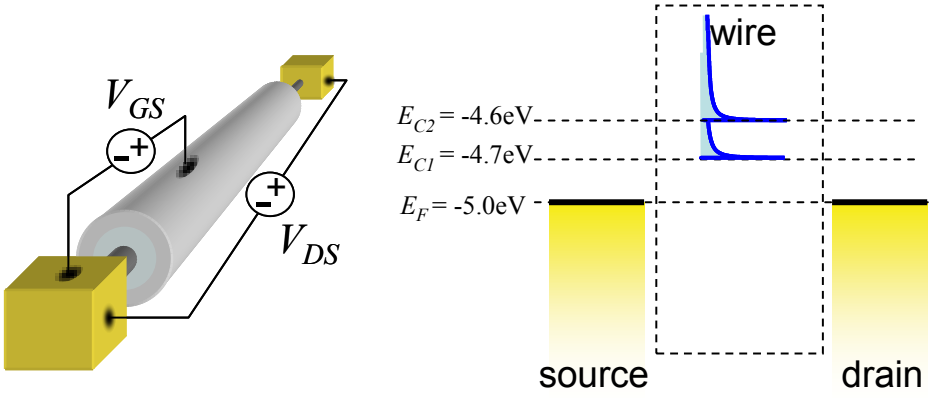


Fig. 5.35. A quantum wire FET with two modes. The length of the wire is L = 100nm, the gate capacitance is C_G = 50 aF per nanometer of wire length, and the electron mass, m, in the wire is $m = m_0$=9.1x10^{-31} kg. Assume C_S and C_D = 0.

3. 2-d ballistic FET

(a) Numerically calculate the current-voltage characteristics of a single mode 2-d ballistic FET using Eq. (5.51) and a self consistent solution for the potential, U. Plot your solution for $T = 1$K and $T = 298$K. In each plot, consider the voltage range $0 < V_{DS} < 0.5$, at $V_{GS} = 0.3$ V, 0.35 V, 0.4V, 0.45V and 0.5V. In your calculation take the bottom of the conduction band to be -4.7 eV, the Fermi Energy at equilibrium $E_F = -5.0$ eV, $L = 40$nm, $W = 3 \times L$, and $C_G = 0.1$ fF. Assume $C_D = C_S = 0$. Take the effective mass, m, as $m = 0.5 \times m_0$, where $m_0 = 9.1\text{x}10^{-31}$ kg.

You should obtain the IV characteristics shown in Fig. 5.26.

Continued on next page….

(b) Next, compare your numerical solutions to the analytic solution for the linear and saturation regions (Eqns (5.60) and (5.61)). Explain the discrepancies.

(c) Numerically determine I_{DS} vs V_{GS} at $V_{DS} = 0.5$V and $T = 298$K. Plot the current on a logarithmic scale to demonstrate that the transconductance is 60mV/decade in the subthreshold region.

(d) Using your plot in (c), choose a new V_T such that the analytic solution for the saturation region (Eq. (5.61)) provides a better fit at room temperature. Explain your choice.

4. An experiment is performed on the channel conductor in a three terminal device. Both the source and drain are grounded, while the gate potential is varied. Assume that $C_G >> C_S, C_D$.

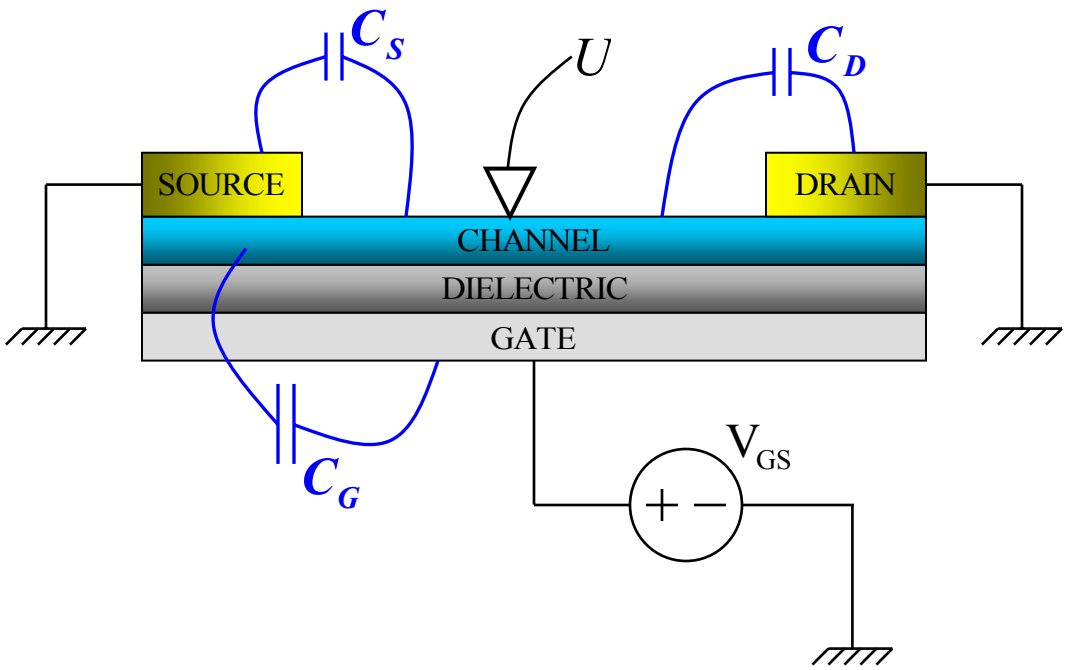


Fig. 5.36. Measuring the surface potential of a transistor channel.

The transistor is biased above threshold ($V_{GS} > V_T$). Measurement of the channel potential, U, shows a linear variation with increasing $V_{GS} > V_T$.

Under what conditions could the conductor be:

(i) a quantum dot (0 dimensions)?
(ii) a quantum wire (1 dimension)?
(iii) a quantum well (2 dimensions)?

Explain your answers.

Part 5. Field Effect Transistors

5. Consider a three terminal molecular transistor.

(a) Assume the molecule contains only a single, unfilled molecular orbital at energy Δ above the equilibrium Fermi level. Assume also that $C_G >> C_S$, C_D and $\Delta >> kT$. Calculate the transconductance for small V_{DS} as a function of T and V_{GS} for $V_{GS} << \Delta$.

Express your answer in terms of I_{DS}.

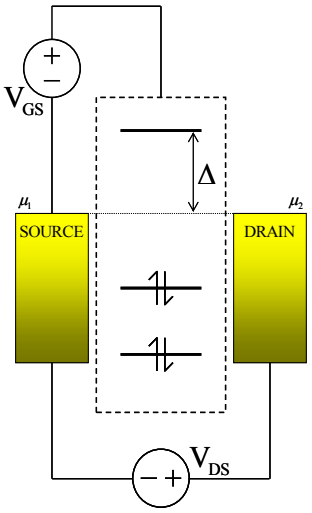


Fig. 5.37. A molecular transistor with discrete energy levels in the channel.

(b) Now, assume that the density of molecular states is

$$g(E) = \frac{1}{E_T} \exp\left[\frac{E-\Delta}{E_T}\right] u(\Delta - E)$$

where $E_T >> kT$. Calculate the transconductance for small V_{DS} as a function of T and V_{GS} for $V_{GS} << \Delta$. Assume $C_G \to \infty$.

Express your answer in terms of I_{DS}.

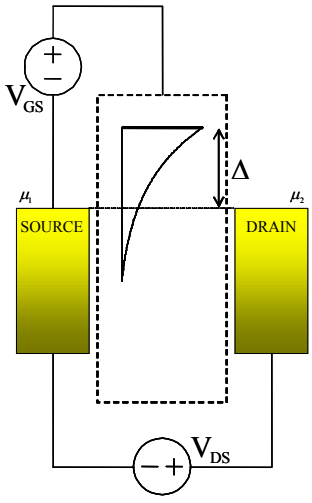


Fig. 5.38. A molecular transistor with an exponential DOS in the channel.

(c) Discuss the implications of your result for molecular transistors.

6. Consider the conventional n-channel MOSFET illustrated below:

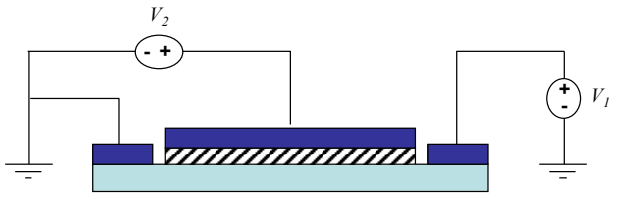


Fig. 5.39. The structure of a conventional MOSFET.

Assume V_T = 1V and V_2 = 0V. Sketch the expected IV characteristics (I vs. V_1) and explain, with reference to band diagrams, why the IV characteristics are not symmetric.

7. Consider a ballistic quantum well FET at T = 0K.

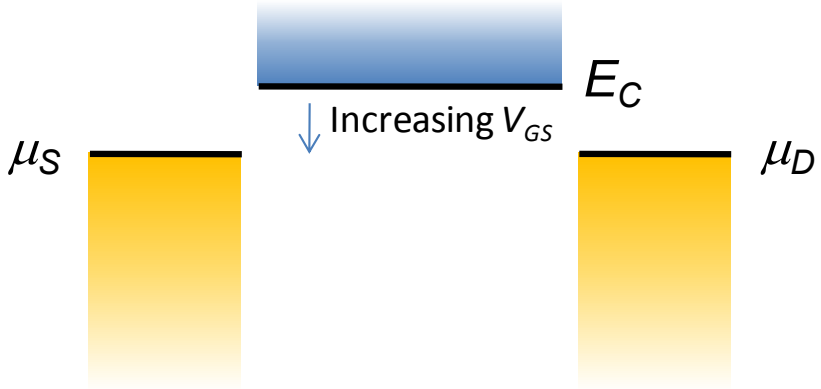


Fig. 5.40. A quantum well FET below threshold.

Recall that the general solution for a quantum well FET in saturation is:

$$I_{DS} = \frac{qW}{\pi^2\hbar^2}\sqrt{\frac{8m}{9}}\left(\eta q\left(V_{GS}-V_T\right)\right)^{3/2}$$

(a) In the limit that $C_Q >> C_{ES}$, the bottom of the conduction band, E_C, is „pinned" to μ_S at threshold. Show that under these conditions $\eta \to 0$.

(b) Why isn't the conductance of the channel zero at threshold in this limit?

(c) Given that $C_{OX} = C_G/(WL)$, where W and L are the width and length of the channel, respectively, show that the saturation current in this ballistic quantum well FET is given by

$$I_{DS} = \frac{8\hbar W}{3m\sqrt{\pi q}}\left(C_{OX}\left(V_{GS}-V_T\right)\right)^{3/2} \qquad (5.77)$$

Hint: Express the quantum capacitance in the general solution in terms of the device parameters m, W, and L.

Part 5. Field Effect Transistors

8. A quantum well is connected to source and drain contacts. Assume identical source and drain contacts.

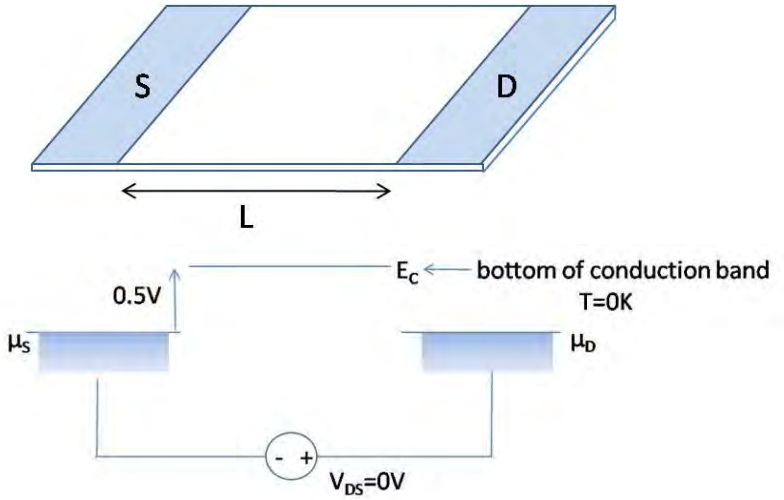


Fig. 5.41. A quantum well with source and drain contacts.

(a) Plot the potential profile along the well when $V_{DS} = +0.3$V.

Now a gate electrode is positioned above the well. Assume that $C_G \gg C_S, C_D$, except very close to the source and drain electrodes. At the gate electrode $\varepsilon = 4\times8.84\times10^{-12}$ F/m and $d = 10nm$. Assume the source and drain contacts are identical.

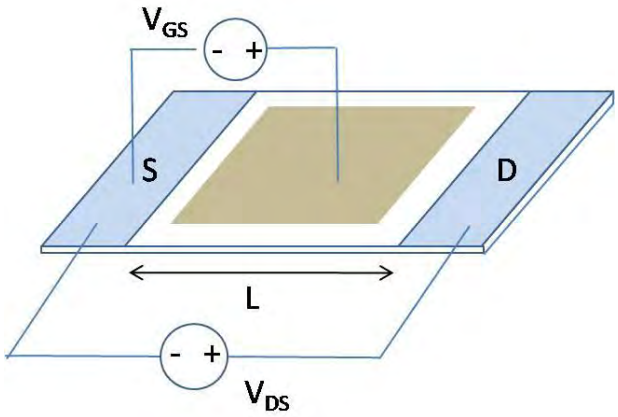


Fig. 5.42. The quantum well with a gate electrode also.

(b) What is the potential profile when $V_{DS} = 0.3$V and $V_{GS} = 0$V.

(c) Repeat (b) for $V_{DS} = 0$V and $V_{GS} = 0.7$V. Hint: Check the C_Q.

(d) Repeat (b) for $V_{DS} = 0.3$V and $V_{GS} = 0.7$V assuming ballistic transport.

(e) Repeat (b) for $V_{DS} = 0.3$V and $V_{GS} = 0.7$V assuming non-ballistic transport.

9. This problem considers a 2-d quantum well FET. Assume the following:

$T = 0\text{K}$, $L = 40\text{nm}$, $W = 120\text{nm}$, $C_G = 0.1\text{fF}$, $C_S = C_D = 0$

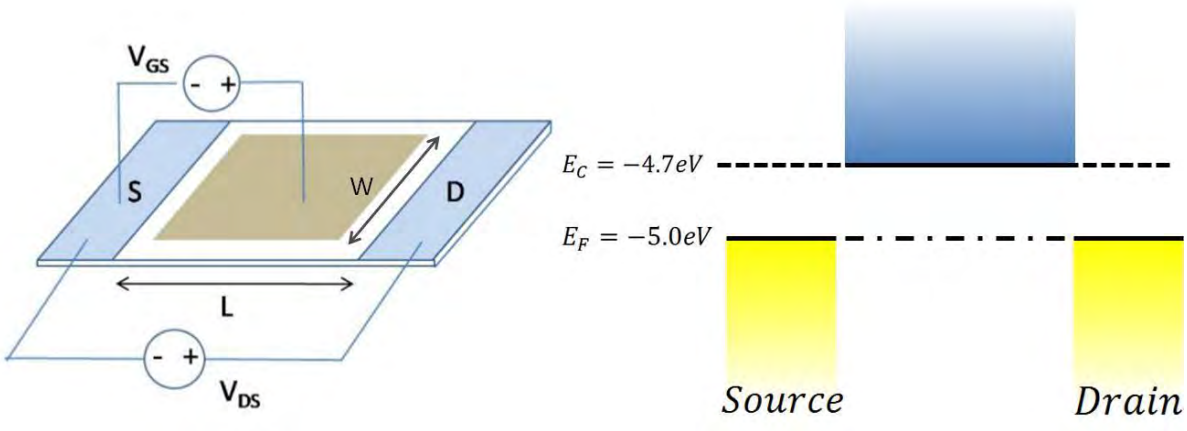


Fig. 5.43. A 2-d quantum well FET.

(a) Compare the operation of the 2-D well in the ballistic and semi-classical regimes. **Assume $C_Q \to \infty$ >> C_{ES} in both regimes.**
Take $\mu = 300\ \text{cm}^2/\text{Vs}$ in the semi-classical regime.

Plot I_{DS} vs V_{DS} for $V_{GS} = 0.5\text{V}$ and $V_{DS} = 0$ to 0.5V.

(**b)** Explain the difference in the IV curves. Is there a problem with the theory? If so, what?

Part 6. The Electronic Structure of Materials

Atomic orbitals and molecular bonds

The particle in the box approximation completely ignores the internal structure of conductors. For example, it treats an insulator such as diamond the same as a conductor such as gold. Despite this it can be surprisingly useful, as we have seen in the discussion of ballistic transistors.

We are concerned now with more accurate calculations of electronic structure. Unfortunately, exact solutions are not usually possible. Determining the energies and wavefunctions of multiple electrons in a solid is a classic „*many body problem*". For example, to solve for the electrons, we must know the exact position of each atom in the solid, and also calculate all interactions between multiple electrons.

Nevertheless, there is much to be learnt from a first principles analysis of electronic structure. We"ll begin at the bottom, with the hydrogen atom.

The hydrogen atom

Hydrogen is the simplest element. There are just two components: an electron and a positively charged nucleus comprised of a single proton.

The electron experiences the attractive potential of the nucleus. The nuclear potential is spherically symmetric and given by the Coulomb potential

$$V(r) = -\frac{Zq^2}{4\pi\varepsilon_0 r} \tag{6.1}$$

where r is the radial separation of the electron and the nucleus ε_0 is the dielectric constant and Z is the number of positive charges at the nucleus. For hydrogen there is one proton, and $Z = 1$.

Recall that a general expression for the kinetic energy operator in three dimensions is:

$$\hat{T} = -\frac{\hbar^2}{2m_e}\nabla^2 \tag{6.2}$$

where ∇^2 is the Laplacian operator.

In rectilinear coordinates (x,y,z)

$$\nabla^2 = \frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2} \tag{6.3}$$

In spherical coordinates (r,θ,ϕ)

$$\nabla^2 = \frac{1}{r}\frac{d^2}{dr^2}r + \frac{1}{r^2}\frac{1}{\sin^2\theta}\frac{d^2}{d\phi^2} + \frac{1}{r^2}\frac{1}{\sin\theta}\frac{d}{d\theta}\sin\theta\frac{d}{d\theta} \tag{6.4}$$

Thus, the Hamiltonian for the hydrogen atom is

$$\begin{aligned}\hat{H} &= -\frac{\hbar^2}{2m}\nabla^2 - \frac{Zq^2}{4\pi\varepsilon_0 r} \\ &= -\frac{\hbar^2}{2m}\left(\frac{1}{r}\frac{d^2}{dr^2}r + \frac{1}{r^2}\frac{1}{\sin^2\theta}\frac{d^2}{d\phi^2} + \frac{1}{r^2}\frac{1}{\sin\theta}\frac{d}{d\theta}\sin\theta\frac{d}{d\theta}\right) - \frac{Zq^2}{4\pi\varepsilon_0 r}\end{aligned} \tag{6.5}$$

This takes a bit of algebra to solve for the atomic orbitals and associated energies. An approximate solution (assuming a box potential rather than the correct Coulomb potential) is contained in Appendix 2.

The lowest energy solutions are plotted in Fig. 6.1, below.

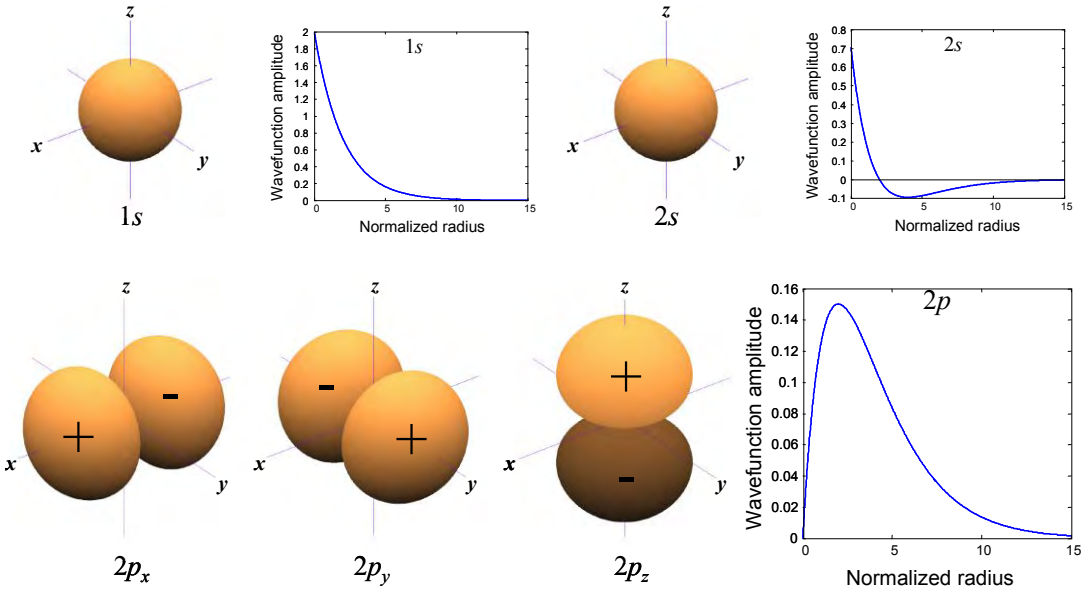


Fig. 6.1. The first five orbitals of the hydrogen atom together with their radial profiles.

Each of the solutions shown in Fig. 6.1 is labeled either *s* or *p*. These letters describe the angular symmetry of solution. They are the index for the **orbital angular momentum** of the electron. „*s*“ orbitals exhibit even symmetry about the origin in every dimension. Orbitals that exhibit odd symmetry about the origin in one dimension are labeled „*p*“. We show in Appendix 1 that the eigenfunctions of an electron restricted to the surface of a sphere are characterized by quantized angular momentum. We are only showing the *s* and *p* solutions but there are an infinite set of solutions, e.g. *s*, *p*, *d*, *f*… corresponding to orbital angular momenta of 0, 1, 2, 3…

The energy of each atomic orbital is also labeled by an integer known as the **principal quantum number**. Thus, the 1s orbital is the lowest energy s orbital, 2p and 2s orbitals are degenerate first excited states.

Knowledge of the exact atomic orbitals is not necessary for our purposes. Rather, we will use the orbitals as symbolic building blocks in the construction of molecular orbitals: electron wavefunctions in molecules.

Atoms to Molecules

We now seek to determine the electronic states of whole molecules – molecular orbitals. Although we will begin with relatively small molecules, the calculation techniques that we will introduce can be extended to larger materials that we don‟t usually think of as molecules: like Si crystals, for example.

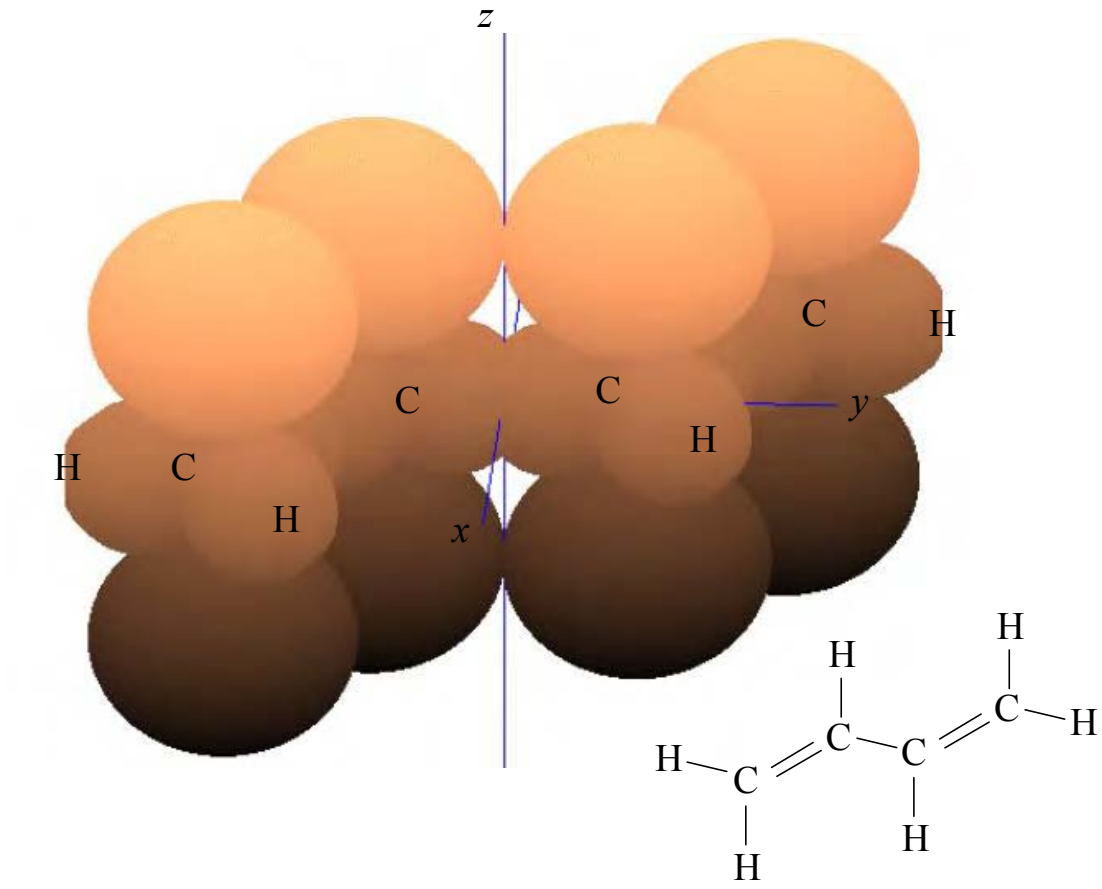


Fig. 6.2. The molecule 1,3-butadiene. Clouds of electron probability density are shown around each atom. They combine to form *molecular orbitals*.

In the previous discussion of atomic orbitals, we implicitly assumed that the nucleus is stationary. This is an example of the *Born-Oppenheimer approximation*, which notes that the mass of the electron, m_e, is much less than the mass of the nucleus, m_N. Consequently, electrons respond almost instantly to changes in nuclear coordinates.

In calculations of the electronic structure of molecules, we have to consider multiple electrons and multiple nuclei. We can simplify the calculation considerably by assuming

that the nuclear positions are fixed. The Schrödinger equation is then solved for the electrons in a static potential; see Appendix 3. Different arrangements of the nuclei are chosen and the solution is optimized.

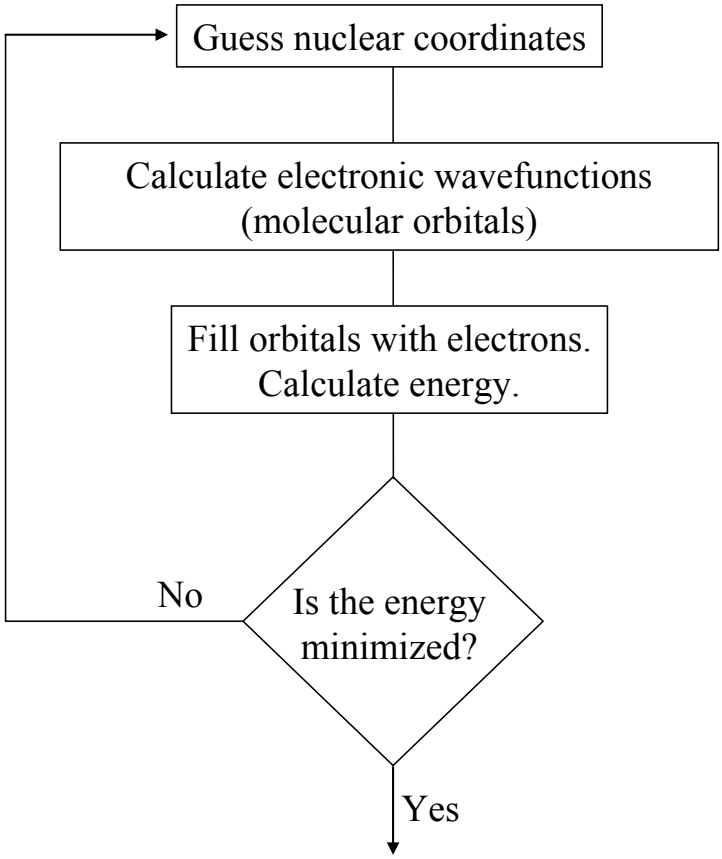


Fig. 6.3. Technique for calculating the electronic structure of materials.

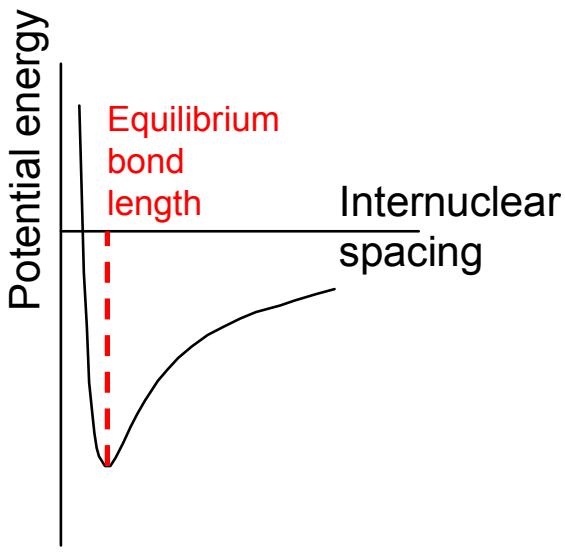


Fig. 6.4. The equilibrium internuclear spacing (bond length) in a molecule results from competition between a close-range repulsive force typically with exponential dependence on intermolecular spacing, and a longer-range attractive Coulomb force. Typically the molecular orbitals must be calculated for each internuclear spacing. The energy minima is the equilibrium bond length. Calculating the electronic states for fixed nuclear coordinates is an example of the Born-Oppenheimer approximation.

Molecular orbitals

Unfortunately, even when we apply the Born-Oppenheimer approximation and hold the nuclear coordinates fixed, the solution to the Schrödinger equation (Eq. (6.5)) is extremely complex in all but the simplest molecules. Usually numerical methods are preferred. But some conceptual insight may be gained by assuming that the molecular orbitals are linear combinations of atomic orbitals, *i.e.*, we write:

$$\varphi = \sum_r c_r \phi_r \tag{6.6}$$

where φ is the molecular orbital and ϕ is an atomic orbital. A filled molecular orbital with lower energy than the constituent atomic orbitals stabilizes the molecule and is known as a chemical bond.

We can define two types of molecular orbitals built from *s* and *p* atomic orbitals:

***σ molecular orbitals*:** These are localized between atoms and are invariant with respect to rotations about the internuclear axis. If we can take the *x*-axis as the internuclear axis, then both *s* and p_x atomic orbitals can participate in σ molecular orbitals. p_y and p_z atomic orbitals cannot contribute to s molecular orbitals because they each have zero probability density on the *x*-axis.

***π molecular orbitals*:** Electrons in π molecular orbitals are more easily shared between atoms. The probability density is not as localized as in a σ molecular orbital. A π molecular orbital is also not invariant with respect to rotations about the internuclear axis. linear combinations of p_y and p_z atomic orbitals form π molecular orbitals.

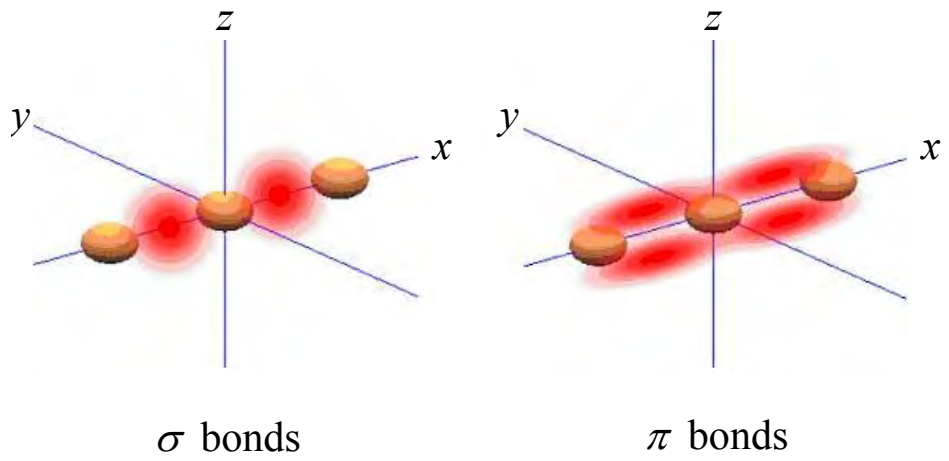


Fig. 6.5. Examples of σ and π bonds. σ bonds are localized between atoms whereas π bonds are delocalized above the internuclear axis.

Linear combination of atomic orbitals (LCAO)

The expansion of a molecular orbital in terms of atomic orbitals is an extremely important approximation, known as the Linear combination of atomic orbitals (LCAO). The atomic orbitals used in this expansion constitute the *basis set* for the calculation. Ideally, the number of atomic orbitals used should be infinite such that we could re-express any given wavefunction exactly in terms of a linear combination of atomic orbitals. In this case, we say that the basis set is also infinite. But computational limitations usually force the basis set to be finite in practice. Choice of the basis set is an especially important consideration in numerical simulations; for example we might consider *s*, *p* and *d* orbitals, but not *f* or higher orbitals.

In some cases, we can take good guesses at the weighting coefficients, c_r, based on the likely nuclear arrangement. However, depending on the nuclear arrangement, it often helps to define new atomic orbitals that are linear combinations of the familiar *s* and *p* atomic orbitals. These are known as symmetry adapted linear combinations (SALCs) because they are chosen based on the nuclear symmetry. They are also known as hybrid atomic orbitals. We discuss SALCs in Appendix 4.

The tight binding approximation

Each atom in a conductor typically possesses many electrons. We can simplify molecular orbital calculations significantly by neglecting all but a few of the electrons. The basis for discriminating between the electrons is energy. The electrons occupy different atomic orbitals: some electrons require a lot of energy to be pulled out the atom, and others are more weakly bound.

Our first assumption is that electrons in the deep atomic orbitals do not participate in charge transport. Recall that charge conduction only occurs though states close to the Fermi level. Thus, we are concerned with only the most weakly bound electrons occupying so-called *frontier* atomic orbitals.

In this class, we will exclusively consider carbon-based materials. Furthermore, we will only consider carbon in the triangular geometry that yields sp^2 hybridized atomic orbitals; see Appendix 4 for a full discussion. In these materials, each carbon atom has *one* electron in an unhybridized p_z orbital. The unhybridized p_z atomic orbital is the frontier orbital. It is the most weakly bound and also contributes to π molecular orbitals that provide a convenient conduction path for electrons along the molecule. We will assume that the molecular orbitals of the conductor relevant to charge transport are linear combinations of frontier atomic orbitals.

For example, let's consider the central carbon atom in Fig. 6.6. Assume that the atom is part of a triangular network and that consequently it contains one electron in a frontier p_z atomic orbital. Let's consider the effect of the neighboring carbon atom to the right of the central atom.

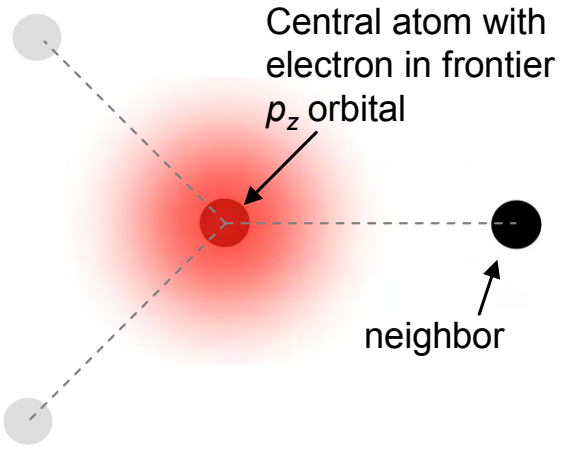


Fig. 6.6. One carbon atom with a single frontier electron and its neighboring nucleus. The Hamiltonian of the system contains potential terms for each of the two nuclei.

Assuming the positions of the atoms are fixed, the Hamiltonian of the system consists of a kinetic energy operator, and two Coulombic potential terms: one for the central atom and one for its neighbor:

$$H = T + V_1 + V_2 \tag{6.7}$$

Now, consider an integral of the form:

$$\langle \phi_r | H | \psi \rangle = \langle \phi_r | E | \psi \rangle \tag{6.8}$$

Following Eq. (6.6), the wavefunction in this two atom system can be written as

$$\psi = c_1 \phi_1 + c_2 \phi_2 \tag{6.9}$$

We can expand the LHS of Eq. (6.8) as follows:

$$\langle \phi_r | H | \psi \rangle = c_1 \langle \phi_r | T + V_1 | \phi_1 \rangle + c_1 \langle \phi_r | V_2 | \phi_1 \rangle + c_2 \langle \phi_r | T + V_2 | \phi_2 \rangle + c_2 \langle \phi_r | V_1 | \phi_2 \rangle \tag{6.10}$$

The RHS expands as

$$\langle \phi_r | E | \psi \rangle = c_1 E \langle \phi_r | \phi_1 \rangle + c_2 E \langle \phi_r | \phi_2 \rangle \tag{6.11}$$

The terms in these expansions are <u>not</u> equally important. We can considerably simplify the calculation by categorizing the various interactions and ignoring the least important.

(a) Overlap integrals

First of all, let's define the *overlap integral* between frontier orbitals on atomic sites s and r:

$$S_{sr} = \langle \phi_s | \phi_r \rangle . \tag{6.12}$$

These integrals yield the overlap between atomic orbitals at different sites in the solid. Spatial separation usually ensures that $S_{sr} << 1$ for $s \neq r$. Of course, for normalized atomic orbitals $S_{sr} = 1$ for $s = r$.

Fig. 6.7. The overlap between two adjacent atomic orbitals is shaded in yellow. In the tight binding approximation we will assume that the overlap between frontier atomic orbitals on different sites is zero.

(b) The self-energy

Next, let"s define the *self-energy*. At a particular atomic site, we have

$$T + V_r \left| \phi_r \right\rangle = \alpha_r \left| \phi_r \right\rangle \tag{6.13}$$

where α_r is the self energy, *i.e.*:

$$\alpha_r = \left\langle \phi_r \right| T + V_r \left| \phi_r \right\rangle . \tag{6.14}$$

The self energy, α, is defined to be negative for an electron in a positively charge nuclear potential. Note that if the interaction between the atoms is weak then the self energy is similar to the energy, E, of the combined system.

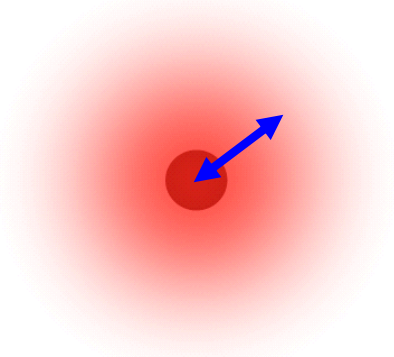


Fig. 6.8. The interaction between a nucleus and its frontier atomic orbital is known as the self energy.

(c) Hopping interactions

Let"s define the *hopping interaction* between different sites s and r:

$$\beta_{sr} = \left\langle \phi_s \right| V_s \left| \phi_r \right\rangle \tag{6.15}$$

The hopping interaction, β, is defined to be negative for an electron in a positively charge nuclear potential.

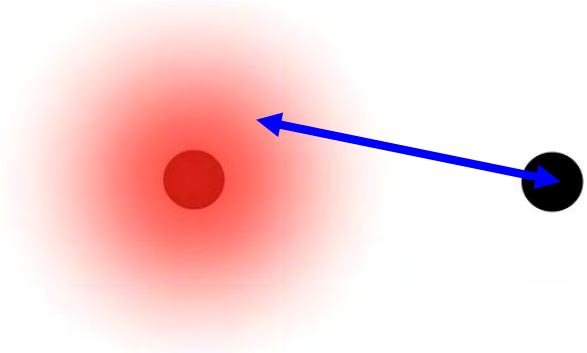


Fig. 6.9. The interaction between a nucleus and the neighboring frontier atomic orbital is known as the hopping interaction.

(d) The remaining interactions

The remaining interaction considers the interaction of a frontier orbital on one site with the potential on another site. It has the form

$$\langle \phi_r | V_s | \phi_r \rangle \tag{6.16}$$

where $s \neq r$. It may not be immediately evident that this interaction is usually much weaker than the hopping interaction of Eq. (6.15). But if the individual frontier orbitals decay exponentially with distance as exp[-*ka*] where *a* is the spacing between the atoms, then this terms behaves as exp[-2*ka*] whereas the hopping term and overlap integral S_{sr} for $s \neq r$ both follow exp[-*ka*].

Consequently, we will neglect this interaction.

Thus, Eq. (6.8) can be re-written for $r = 1$ and $r = 2$ as

$$\begin{aligned} c_1\alpha_1 + c_2\beta_{12} + c_2\alpha_2 S_{12} &= c_1 E + c_2 E S_{12} \\ c_1\beta_{21} + c_1\alpha_1 S_{21} + c_2\alpha_2 &= c_1 E S_{21} + c_2 E \end{aligned} \tag{6.17}$$

Terms containing only the self energy or energy, *E*, of the combined system are large. The small terms are highlighted below in red:

$$\begin{aligned} c_1\alpha_1 + c_2\beta_{12} + c_2\left(\alpha_2 - E\right)S_{12} &= c_1 E \\ c_1\beta_{21} + c_1\left(\alpha_1 - E\right)S_{21} + c_2\alpha_2 &= c_2 E \end{aligned} \tag{6.18}$$

Next, we note that the difference between the self energies, α_1 and α_2, and the energy, *E*, of the combined system may be small. Under this limit, we can reduce the equations further to

$$\begin{aligned} c_1\alpha_1 + c_2\beta_{12} &= c_1 E \\ c_1\beta_{21} + c_2\alpha_2 &= c_2 E \end{aligned} \tag{6.19}$$

Written as a matrix, we get

$$\begin{pmatrix} \alpha_1 & \beta_{12} \\ \beta_{21} & \alpha_2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = E \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \tag{6.20}$$

Thus, we can ignore the overlap integrals of separated atoms.

In summary, tight binding theory makes the following approximations:

1. Consider only frontier atomic orbitals
2. Consider only interactions between the frontier atomic orbitals of nearest neighbors. This is the tight binding approximation.
3. Ignore the overlap integrals of separated atoms, *i.e.* $S_{sr} = \delta_{sr}$. This is valid only when $\alpha_1 \approx \alpha_2 \approx E$. We will assume $S_{sr} = \delta_{sr}$ generally to simplify the mathematics.

The self energy, α, and the hopping interaction, β, could be calculated numerically given the potential and the frontier atomic orbital. But, in this class, we will not actually determine α and β. Rather we are interested in the form of the molecular wavefunctions and the dispersion relations for their energies. With this information we can determine whether the conductor is a metal or an insulator, and its density of states.

Solving for the energy

Considering the tight binding matrix of Eq. (6.20), non trivial solutions for the weighting factors, c_1 and c_2 are obtained from

$$\det\begin{vmatrix}\alpha_1 - E & \beta_{12} \\ \beta_{21} & \alpha_2 - E\end{vmatrix} = 0 \tag{6.21}$$

Let‟s assume that the hopping interactions are equal $\beta_{12} = \beta_{21} = \beta$. We‟ll consider two cases for equal and different self energies.

(a) Equal self energies $\alpha_1 = \alpha_2 = \alpha$

When $\alpha_1 = \alpha_2 = \alpha$, the energy is

$$E = \alpha \pm \beta \tag{6.22}$$

Substituting the energy back into Eq. (6.20) to obtain the coefficients c_1 and c_2 yields two normalized solutions:

$$\varphi = \frac{\phi_1 \pm \phi_2}{\sqrt{2}}\,. \tag{6.23}$$

These two orbitals can be defined by their **parity**: their symmetry if their position vectors are rotated. For example, we could exchange their coordinates. In this example the molecular orbital:

$$\varphi = \frac{\phi_1 + \phi_2}{\sqrt{2}} \tag{6.24}$$

does not change sign under exchange of electrons. It is classified as having *gerade* symmetry, denoted by *g*, where gerade is German for even. In contrast, the other orbital:

$$\varphi = \frac{\phi_1 - \phi_2}{\sqrt{2}} \tag{6.25}$$

does change sign under exchange of electrons. It is classified with *ungerade* symmetry, denoted by *u*, where ungerade is German for odd.

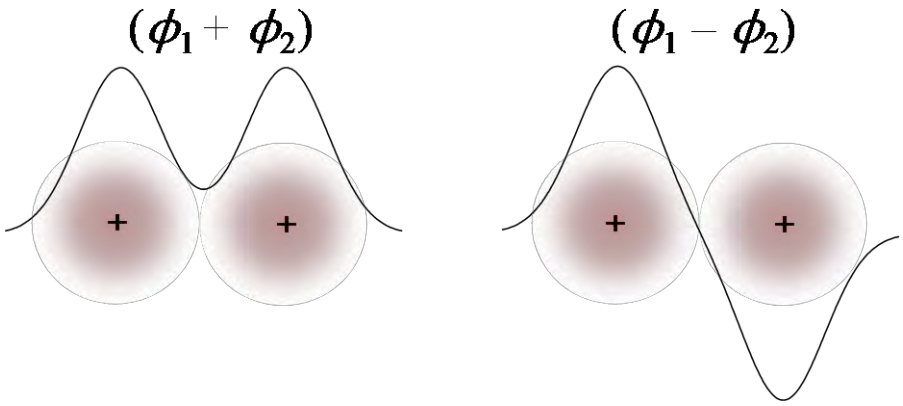


Fig. 6.10. The probability density plotted for the two linear combinations of two frontier orbitals. Due to the increased electron density between the nuclei, the $\Phi_1 + \Phi_2$ has lower energy.

Since the molecular orbital:

$$\varphi = \frac{\phi_1 + \phi_2}{\sqrt{2}} \tag{6.26}$$

has energy, $E = \alpha + \beta$, below that of the self energy, α, of each atomic orbital, the molecule is stabilized in this configuration. This is known as a bonding orbital because it describes a stable chemical bond. Recall that α and β are defined to be negative for an electron in a positively charge nuclear potential.

The other molecular orbital

$$\varphi = \frac{\phi_1 - \phi_2}{\sqrt{2}} \tag{6.27}$$

has energy, $E = \alpha - \beta$, greater than that of the self energy, α, of each atomic orbital. Thus, this configuration is not stable. It is known as an antibonding orbital.

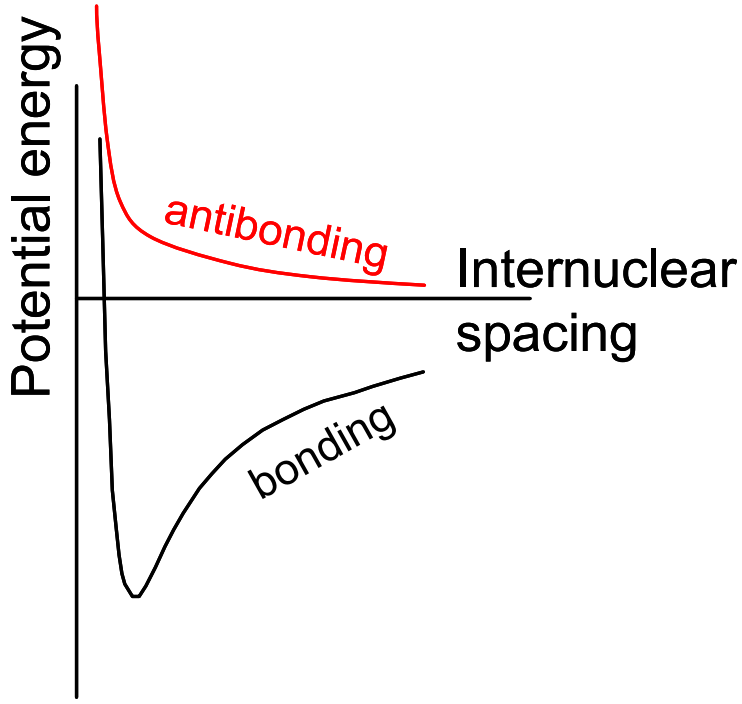


Fig. 6.11. Antibonding and bonding molecular potential energy curves. Note that the antibonding energy is typically substantially larger than shown.

(b) Different self energies

If $|\alpha_1 - \alpha_2| >> \beta$ and $\alpha_1 > \alpha_2$ then the solutions are:

$$E = \alpha_1 + \frac{\beta^2}{\alpha_1 - \alpha_2}, \quad E = \alpha_2 - \frac{\beta^2}{\alpha_1 - \alpha_2} \tag{6.28}$$

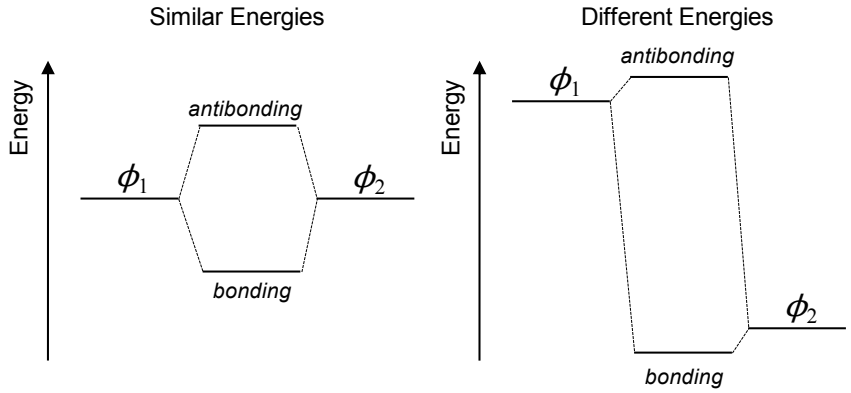


Fig. 6.12. The strongest bonds are formed from atomic orbitals with similar energies. In these diagrams the constituent atomic orbitals are shown at left and right. The molecular orbitals are in the center.

Thus, the splitting increases with the similarity in energy of the participating atomic orbitals, *i.e.* the bonding orbital becomes more stable. This is a general attribute of the interaction between two quantum states. The more similar their intial energies, the stronger the interaction.

Examples of tight binding calculations

Let"s consider a conductor consisting of four atoms, each of which provides a frontier atomic orbital containing a single electron. A molecular equivalent to this model conductor is 1,3-butadiene; see Fig. 6.13. Here each carbon atom contributes one electron in a frontier atomic orbital.

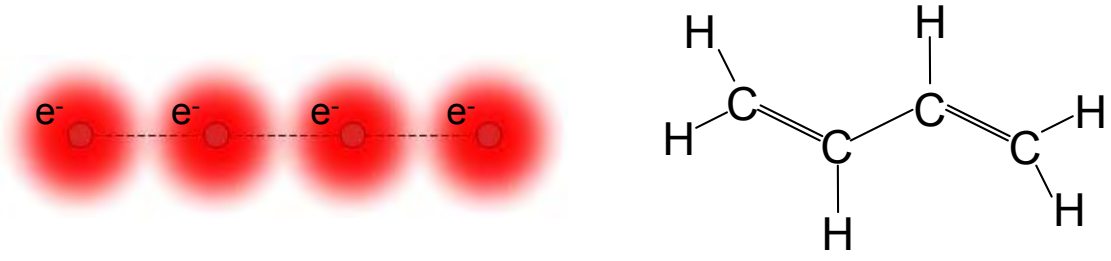


Fig. 6.13. (left) A model four atom conductor, where each atom contributes a single electron in a frontier atomic orbital. **(right)** An approximate chemical equivalent is 1,3-butadiene, where the carbon atoms provide the frontier orbitals.

We'll ignore the hydrogen atoms, since the frontier electrons are donated by the carbon atoms. Let's label the four carbon frontier atomic orbitals ϕ_1, ϕ_2, ϕ_3, and ϕ_4.

Following Eq. (6.6), we let the molecular orbitals be

$$\psi = c_1\phi_1 + c_2\phi_2 + c_3\phi_3 + c_4\phi_4 \tag{6.29}$$

where the c coefficients are yet to be determined.

Let's next consider integrals of the form:

$$\langle \phi_m | H | \psi \rangle = \langle \phi_m | E | \psi \rangle = E \langle \phi_m | \psi \rangle \tag{6.30}$$

Considering m = 1, 2, 3 and 4 in turn, we get four equations:

$$\begin{aligned}
\langle \phi_1 | H | \psi \rangle &= c_1\alpha_1 + c_2\beta_{12} = c_1 E \\
\langle \phi_2 | H | \psi \rangle &= c_2\alpha_2 + c_1\beta_{21} + c_3\beta_{23} = c_2 E \\
\langle \phi_3 | H | \psi \rangle &= c_3\alpha_3 + c_2\beta_{32} + c_4\beta_{34} = c_3 E \\
\langle \phi_4 | H | \psi \rangle &= c_4\alpha_4 + c_3\beta_{43} = c_4 E
\end{aligned} \tag{6.31}$$

You can think of each equation as describing the interactions between a particular carbon atom, and itself and its neighbors. Solving these equations gives the coefficients c_1, c_2, c_3, and c_4. To simplify, we will assume that the self energy at each carbon atom is the same, *i.e.*

$$\alpha = \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4.$$

In addition, we will assume that the hopping interactions between neighboring carbon atoms are the same, i.e.

$$\beta = \beta_{12} = \beta_{21} = \beta_{23} = \beta_{32} = \beta_{34} = \beta_{43}.$$

Perhaps the best way to solve the equations systematically is via a matrix. The equations can be re-written:

$$\begin{pmatrix} \alpha & \beta & 0 & 0 \\ \beta & \alpha & \beta & 0 \\ 0 & \beta & \alpha & \beta \\ 0 & 0 & \beta & \alpha \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = E \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} \tag{6.32}$$

This equation is of the familiar form

$$H|\psi\rangle = E|\psi\rangle \tag{6.33}$$

where the Hamiltonian is in the form of a matrix, and the wavefunction is a column vector containing the coefficients that weight the atomic orbitals:

$$H = \begin{pmatrix} \alpha & \beta & 0 & 0 \\ \beta & \alpha & \beta & 0 \\ 0 & \beta & \alpha & \beta \\ 0 & 0 & \beta & \alpha \end{pmatrix}, \quad \psi = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} \tag{6.34}$$

Expressing the Hamiltonian and wavefunction in this form is an example of *matrix mechanics*, a version of quantum mechanics formulated by Werner Heisenberg that is

convenient for many problems. Apart from this example, we won't pursue matrix mechanics in this class.

But it's worth taking a moment to examine the structure of the Hamiltonian matrix. Each row now describes the interactions between frontier orbitals on a carbon atom, and itself and its neighbors. The diagonal of the matrix contains the self-energies, and the off-diagonal elements are the hopping interactions. This particular example is *tridiagonal*, *i.e.* the matrix elements are zero, except for the diagonal, and its immediately adjacent matrix elements. Linear molecules with alternating single and double bonds always possess tridiagonal matrices.

After a bit of practice with tight binding calculations you should be able to skip directly to writing down the Hamiltonian matrix. For example, consider cyclobutadiene; shown below.

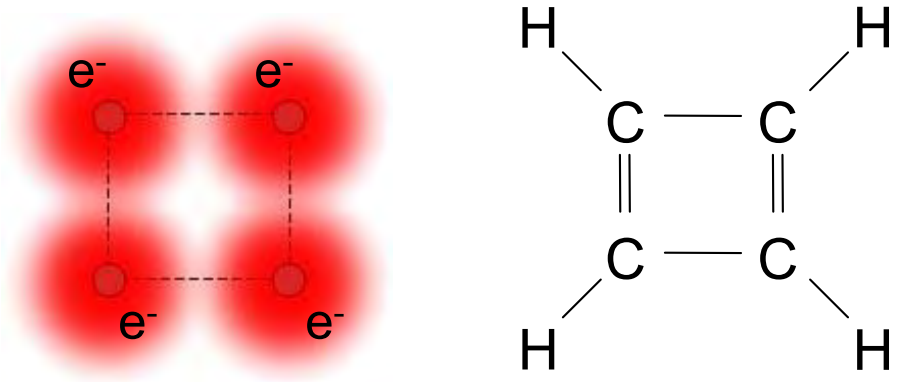


Fig. 6.14. A cyclic four atom molecule and its chemical equivalent, cyclo-butadiene.

Because of its ring structure, cyclobutadiene has additional hopping interactions between the carbons #1 and #4 that were on the ends of the chain in 1,3-butadiene. These interactions at the 1,4 and 4,1 positions are labeled in red in Eq. (6.35) below.

$$H = \begin{pmatrix} \alpha & \beta & 0 & \beta \\ \beta & \alpha & \beta & 0 \\ 0 & \beta & \alpha & \beta \\ \beta & 0 & \beta & \alpha \end{pmatrix} \tag{6.35}$$

As you can probably imagine, for all but the simplest molecules, these matrices can get extremely large and unwieldy. And solving them can be extremely computationally intensive. In fact, tight binding calculations are almost never done by hand. But some insight can be gained by analytically solving simple linear molecules.

Returning to 1,3-butadiene, rearranging Eq. (6.32), we get:

$$\begin{pmatrix} \alpha - E & \beta & 0 & 0 \\ \beta & \alpha - E & \beta & 0 \\ 0 & \beta & \alpha - E & \beta \\ 0 & 0 & \beta & \alpha - E \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix} = 0 \tag{6.36}$$

To find the non-trivial solution (i.e. solutions other than $c_1 = c_2 = c_3 = c_4 = 0$) we take the determinant:

$$\begin{vmatrix} \alpha - E & \beta & 0 & 0 \\ \beta & \alpha - E & \beta & 0 \\ 0 & \beta & \alpha - E & \beta \\ 0 & 0 & \beta & \alpha - E \end{vmatrix} = 0 \tag{6.37}$$

The solutions are:

$$E = \alpha \pm \beta \sqrt{\frac{3}{2} \pm \frac{\sqrt{5}}{2}} \tag{6.38}$$

And the eigenfunctions are:

$$c_j = \sin\left(jn \frac{\pi}{5} \right), \quad j, n = 1, 2, 3, 4. \tag{6.39}$$

These solutions are summarized in Fig. 6.15. The molecular orbitals are similar to the standing waves expected for a particle in a box.

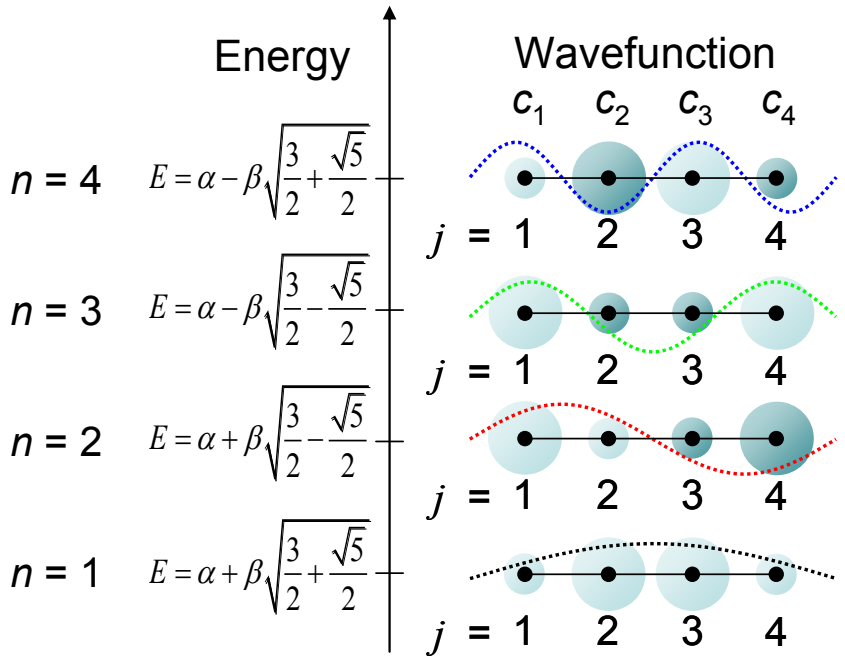


Fig. 6.15. The molecular orbitals and their energies for 1,3-butadiene. After „Molecular Quantum Mechanics", by Atkins and Friedman, 3rd edition, Cambridge University Press, 1997.

Polyacetylene

Next, let"s consider a longer chain of carbon atoms. Very long molecules are known as polymers, and a polymer equivalent of the idealized conductor in Fig. 6.16 is known as polyacetylene.

Specifically, let"s solve for a carbon chain of N atoms. Equation (6.37) is an example of a tridiagonal determinant. In general, an $N \times N$ tridiagonal determinant has eigenvalues:[†]

$$E_n = \alpha + 2\beta\cos\left(\frac{n\pi}{N+1}\right), \quad n = 1,2,...N\,. \tag{6.40}$$

and eigenvectors:

$$c_j = \sin\left(jn\frac{\pi}{N+1}\right), \quad j,n = 1,2,...N\,. \tag{6.41}$$

Note that Eqns. (6.40)-(6.41) reduce to Eqns. (6.38)-(6.39) by using the identity:
$\cos(2\pi/5) = 1/4\left(-1+\sqrt{5}\right)$

Thus, we have solved for the molecular orbitals in a molecule modeled by an arbitrarily long chain of frontier atomic orbitals, each containing a single electron.

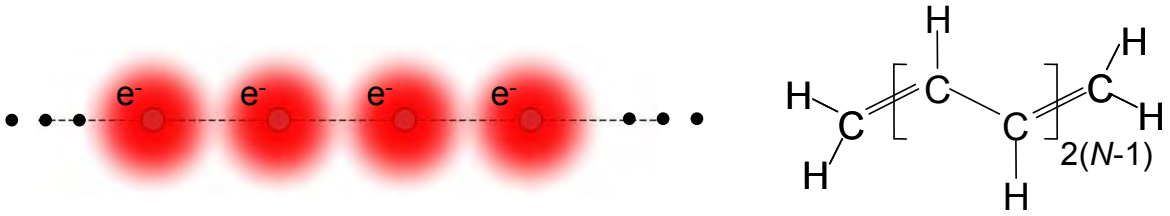


Fig. 6.16. (left) An infinite chain of atoms each contributing a single electron in a frontier orbital. **(right)** The equivalent polymer polyacetylene.

Next, let"s re-express our solutions for polyacetylene in terms of a wavevector, k. Note that because the atoms are discretely positioned in a chain, k is also discrete. There are only N allowed values of k.

Given $x = ja_0$, where a_0 is the spacing between carbon atoms, we get:

$$c(x) = \sin(kx) \tag{6.42}$$

and

$$E_n = \alpha + 2\beta\cos(ka_0) \tag{6.43}$$

where

$$k = \frac{\pi}{a_0}\frac{n}{N+1}, \quad n = 1,2,...N \tag{6.44}$$

[†] If you are interested and have a few spare hours you can try to prove this. After evaluating the first few determinants of simple triadiagonal matrices, N=1, N=2, N=3, etc.. find and solve a difference equation for the determinants as a function of the matrix dimension, N.

The dispersion relation of polyacetylene is plotted in Fig. 6.17. The energy states are restricted to energies $E = \alpha \pm 2\beta$, forming a *band*, of width 4β, centered at α. The bandwidth (4β) is directly related to the hopping interaction between neighboring carbon atoms. This is a general property: the stronger the interaction between an electron and the neighboring atoms, the larger the bandwidth. And as we shall, the broader the electronic bandwidth, the better the electron conduction within the material.

There are N states in the band, each separated by

$$\Delta k = \frac{\pi}{a_0 (N+1)}. \tag{6.45}$$

Note that the length of the chain is $L = (N\text{-}1)a_0$. Thus for long chains the separation between states in the band is approximately

$$\Delta k \approx \frac{\pi}{L} \tag{6.46}$$

Now each carbon atom contributes a single electron in the frontier atomic orbitals that comprise the molecular orbitals. Thus for a N-repeat polymer, there are N electrons. But each state holds two electrons, one of each spin. Filling the lowest energy states first, only the first $N/2$ k states are filled; see Fig. 6.17. Thus, the band is only half full, and so, if the polymer was connected to contacts we might expect polyacetylene to be a metal.

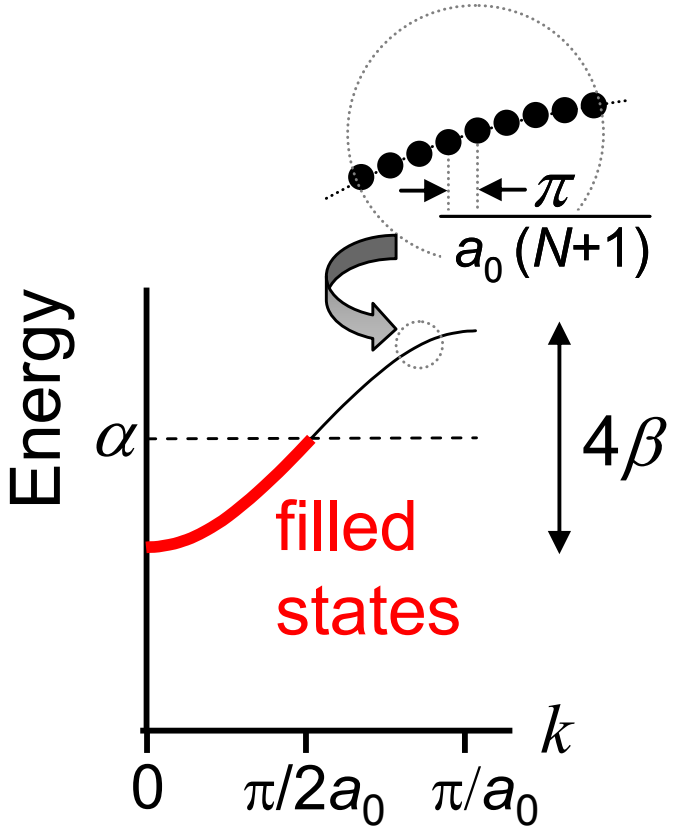


Fig. 6.17. The dispersion relation of polyacetylene as determined by a tight binding analysis. For N atoms, each donating a single electron in a frontier atomic orbital, there are N molecular orbitals with energies arranged in a band. Since the band of states is only half full this material might be expected to be a metal.

Crystals and periodic molecules

The particle in a box approximation is too crude for most problems. Tight binding, on the other hand, is often quite computationally intensive. But fortunately, we can make simplifications when the material is periodic. In this lecture, we are interested in first describing 1d, 2d and 3d periodic materials, and then calculating their wavefunctions. We"ll begin with some definitions.

The Primitive Unit Cell

A primitive unit cell of a periodic material is the smallest possible arrangement of atoms that can be copied to construct the entire material.

Primitive Lattice Vectors

Given a primitive unit cell, we can construct the periodic material by translating the unit cell by multiples of the primitive lattice vectors.

Some examples may help

Polyacetylene (average bond model)

The carbon backbone of polyacetylene consists of alternating single and double bonds. Thus, there are two possible configurations: single-double-single-double or double-single-double-single.

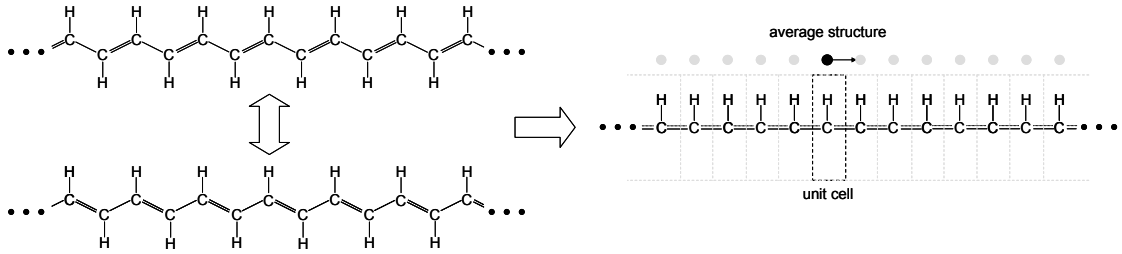


Fig. 6.18. The simplest model for polyacetylene is an average of the two possible alternating single-double bond configurations.

The simplest model assumes that every carbon atom in polyacetylene is identical. The carbon-carbon bonds are then an average of single and double, and the unit cell is a single carbon atom and its associated hydrogen atom. Under this model, to construct a polyacetylene chain, we should translate the primitive unit cell a distance a_0. Thus the primitive lattice vector is $\tilde{\mathbf{a}}_1 = a_0\tilde{\mathbf{x}}$, where we arbitrarily positioned the chain parallel to the x-axis.

Polyacetylene (alternating bond model)

A more accurate model of polyacetylene includes the effects of alternating single and double carbon-carbon bonds on the polymer backbone. Since double bonds contain a slightly higher electron density than single bonds, they are slightly shorter. Thus, the single-double bond alternation establishes a static deformation with twice the period, a_0, of the average bond model for polyacteylene. The periodic charge density established by the deformation is known as a *charge density wave*. The primitive lattice vector is $\tilde{\mathbf{a}}_1 = 2a_0\tilde{\mathbf{x}}$.

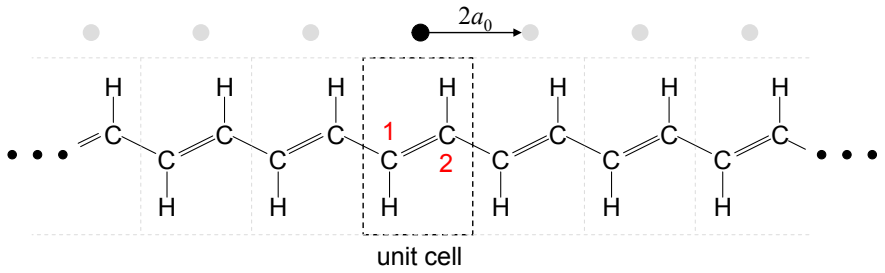


Fig. 6.19. A more accurate model for polyacetylene, including the alternating single and double bonds.

Graphene

Polyacetylene is a 1d chain of carbon atoms, each contributing one electron in a frontier atomic orbital. It is also possible to form 2d sheets of carbon atoms with a single electron in their frontier atomic orbitals. See for example graphene in Fig. 6.20.†

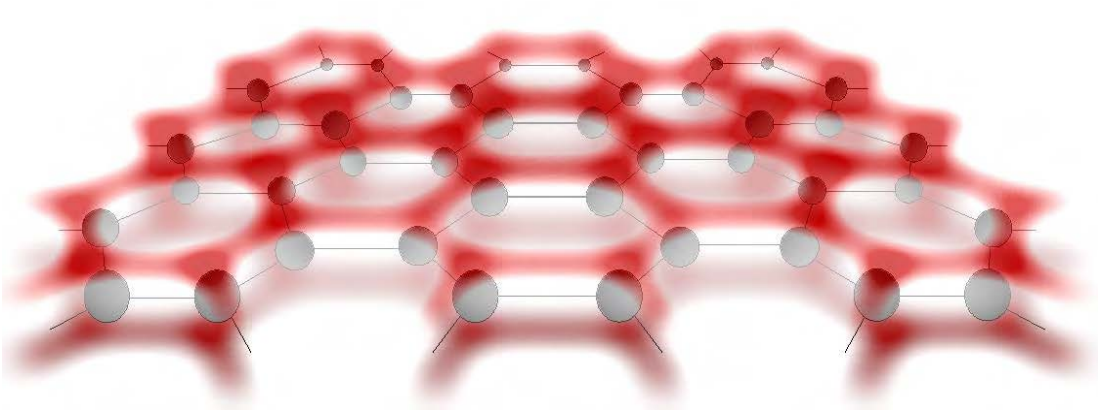


Fig. 6.20. Graphene is a 2d sheet of hexagonal carbon atoms. Electrons in frontier atomic orbitals are found above and below the plane.

† In graphene extended π orbitals are formed above and below the plane of a sheet of hexagonal carbon atoms, increasing the rigidity of the structure and enhancing charge transport.

Graphene may also be rolled up into cylinders to form *carbon nanotubes* – unique structures that we will consider in detail later on in the class.

The unit cell of graphene contains two carbon atoms labeled 1 and 2 in Fig. 6.21. The lattice is generated by shifting the unit cell with the primitive lattice vectors $\tilde{\mathbf{a}}_1 = a_0\left(-\sqrt{3}/2, 3/2\right)$ and $\tilde{\mathbf{a}}_2 = a_0\left(\sqrt{3}/2, 3/2\right)$, where a_0 is the carbon-carbon bond length.

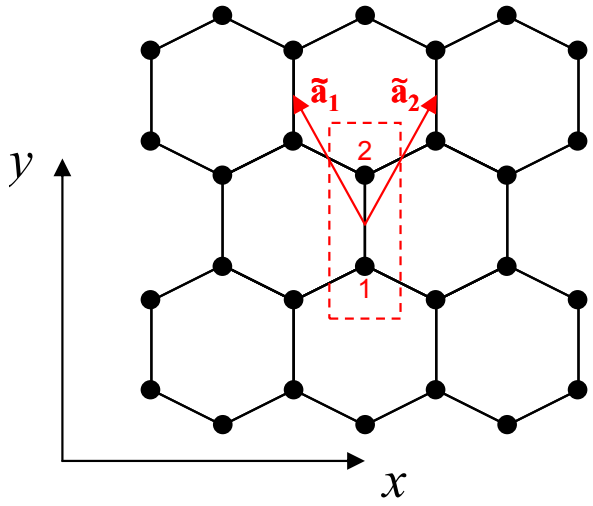


Fig. 6.21. A graphene lattice showing the unit cell and primitive lattice vectors.

Simple cubic, face centered cubic and diamond lattices

Fig. 6.22 shows the simplest 3-d crystal structure – the simple cubic lattice. The primitive lattice vectors are $\mathbf{a}_1 = a_0\hat{\mathbf{x}}$, $\mathbf{a}_2 = a_0\hat{\mathbf{y}}$, $\mathbf{a}_3 = a_0\hat{\mathbf{z}}$, where a_0 is the spacing between neighboring atoms. Very few materials, however, exhibit the simple cubic structure. The major semiconductors, including silicon and gallium arsenide, possess the same structure as diamond.

As also shown in Fig. 6.22, to describe the diamond structure, we first define the face centered cubic (FCC) lattice. Here the simple cubic structure is augmented by an atom in each of the faces of the cube. The primitive lattice vectors are:

$$\mathbf{a}_1 = \frac{a_0}{2}\left(\hat{\mathbf{x}} + \hat{\mathbf{z}}\right),\ \mathbf{a}_2 = \frac{a_0}{2}\left(\hat{\mathbf{y}} + \hat{\mathbf{z}}\right),\ \mathbf{a}_3 = \frac{a_0}{2}\left(\hat{\mathbf{x}} + \hat{\mathbf{y}}\right),$$

where a_0 is now the cube edge length.

In the diamond lattice, each atom is sp_3-hybridized. Thus, every atom is at the center of a tetrahedron. We can construct the diamond lattice from a face centered cubic lattice with a two atom unit cell. For example, in Fig. 6.22, our unit cell has one atom at (0,0,0), and another at a_0/4.(1,1,1).

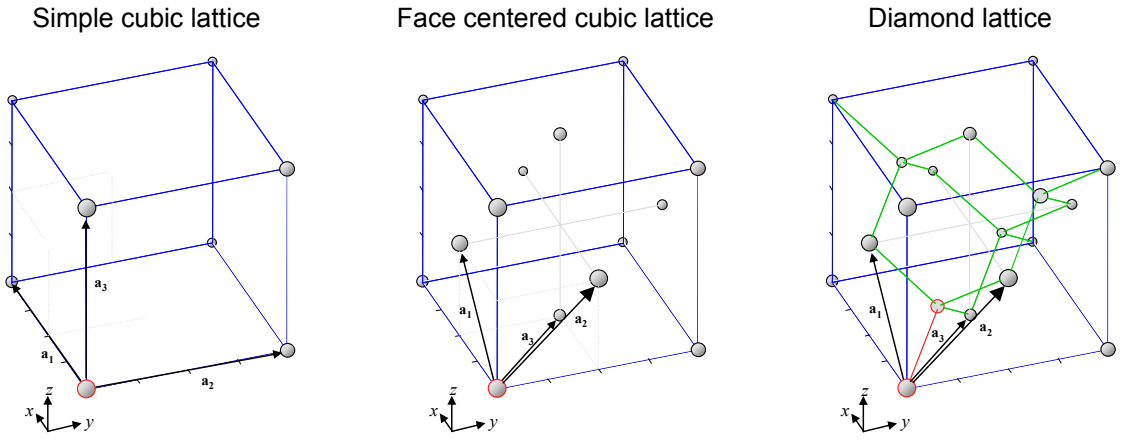


Fig. 6.22. A diamond lattice is simply a face-centered cubic lattice with a two atom unit cell (outlined in red).

Bloch functions: wavefunctions in periodic molecules

Wavefunctions in periodic materials are described by *Bloch functions*. To better understand their properties, it is instructive for us to derive Bloch functions.[†]

First, let"s consider a periodic molecule, comprised of unit cells translated by multiples of the primitive lattice vectors. Let the wavefunction of the unit cell be ϕ_0. Under the tight

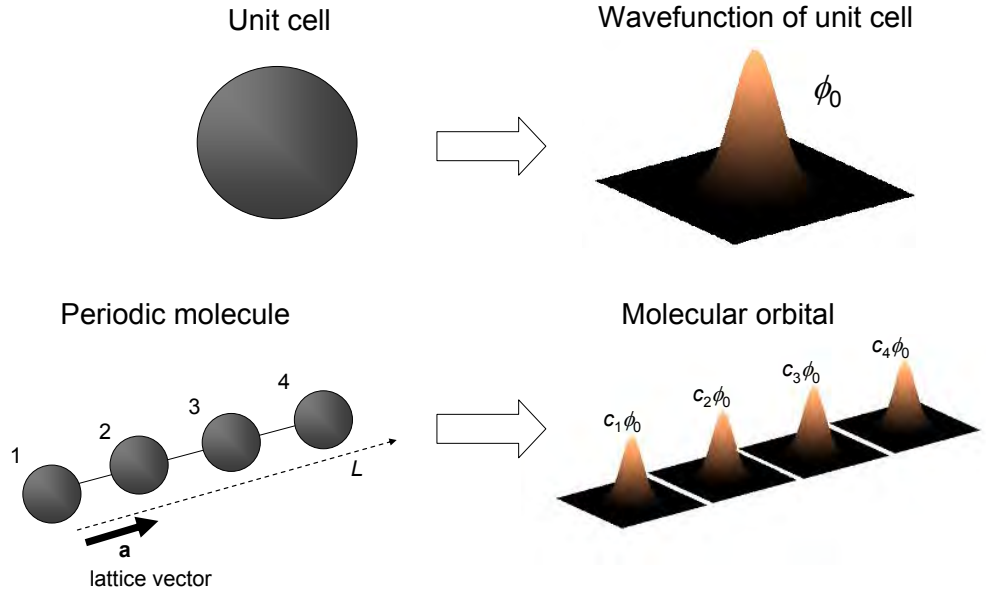


Fig. 6.23. The molecular orbitals of periodic molecules are linear combinations of the wavefunctions of the unit cells.

[†] Our method follows the derivation of Kittel in „Introduction to Solid State Physics", Wiley, 7th Edition, 1996.

binding approximation, the wavefunction of the unit cell is itself constructed from a linear combination of frontier atomic orbitals.

Now, the molecular orbitals will be composed of linear combinations of the wavefunction of the unit cell, i.e.

$$\psi = \sum_r c_r \phi_0 \tag{6.47}$$

where once again c_r is a set of coefficients. Note that, unlike approximations of molecular orbitals using linear combinations of frontier atomic orbitals, Eq. (6.47) is exact. We emphasize that ϕ_0 in Eq. (6.47) is the *exact* wavefunction of a unit cell of the complete molecule. In molecular orbital calculations ϕ_0 is typically calculated using tight binding, or another approximate technique. But for the moment we will assume that we know it exactly.

The aim of this derivation is to determine the coefficients c_r given that the material is periodic. Quite generally, we can relate the two coefficients c_1 and c_2 of the first two unit cells by

$$c_2 = \alpha c_1 \tag{6.48}$$

where α is some constant.

The symmetry of the material allows us to translate indistinguishably and consequently,

$$c_{r+1} = \alpha c_r \tag{6.49}$$

where $0 < r < N$, where N is the number of unit cells in the material.

Now if we *assume periodic boundary conditions*, we can compare the identical unit cells at r and $r + N$:

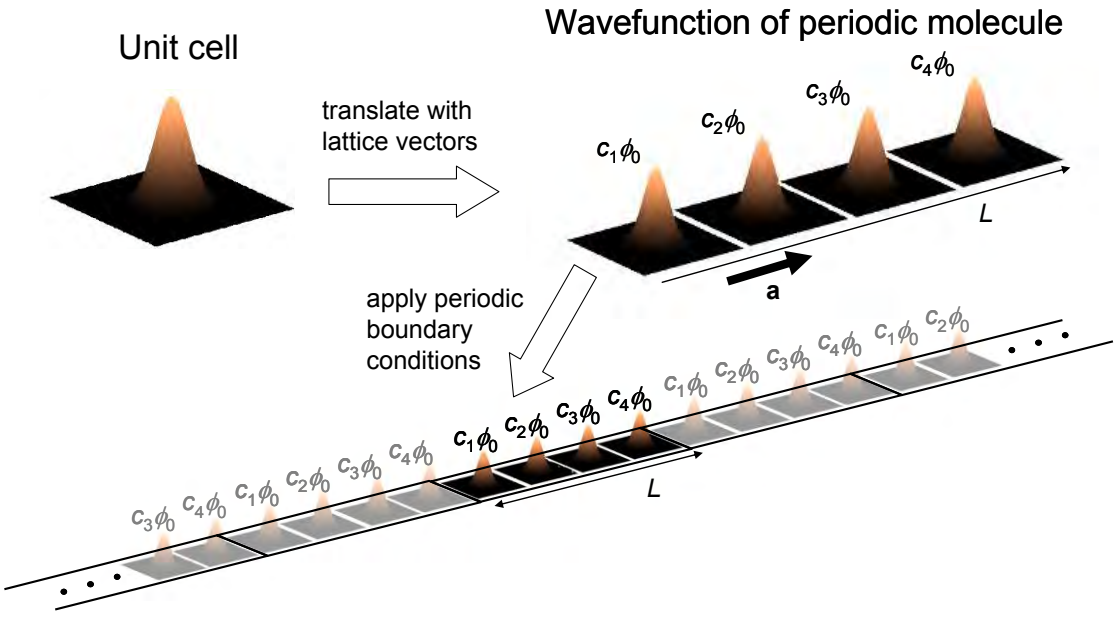


Fig. 6.24. The application of periodic boundary conditions to an already periodic molecule.

$$c_{N+1} = \alpha^N c_1 \tag{6.50}$$

But since $c_{N+1} = c_1$, α must be one of N roots of unity, i.e. $\alpha = \exp\left[i2\pi n/N\right]$, where n is an integer. Thus, the coefficients are phase factors; the wavefunction corresponding to each unit cell is modulated by a phase factor in a periodic molecule. Consequently, if we set $c_N = 1$ (which we can do since absolute phase is arbitrary):

$$c_r = \mathrm{e}^{i\frac{2\pi n}{N}r}. \tag{6.51}$$

Alternately, approximating the coefficients by a continuous function, we can write:

$$c(x) = \mathrm{e}^{ikx} \tag{6.52}$$

where

$$k = \frac{2\pi n}{L}. \tag{6.53}$$

Once again, only certain k values are allowed by the application of periodic boundary conditions. After all, standing waves in the molecule can possess only certain wavelengths. Recall also that Fourier transforms of periodic signals are discrete; see Fig. 6.25. Thus, it follows from a Fourier analysis of the coefficients that k must be discrete. In addition, the Fourier transform of the coefficients is itself periodic since the coefficients are discrete (recall discrete time Fourier series - DTFS).

The first Brillouin zone

Since there are only N distinct values of the coefficients (corresponding to one period of the Fourier transform), we typically restrict k to the N values in the range $-N/2 < n \le N/2$, i.e.

$$-\frac{\pi}{a_0} < k \le \frac{\pi}{a_0}. \tag{6.54}$$

This is known as the **first Brillouin zone**. Other values of k are either not permitted by periodic boundary conditions, or $c(x) = \exp[ikx]$ reduces to one of the N solutions. For example, consider $k = 2\pi(n+N)/L$:

$$c(x) = e^{i\frac{2\pi(n+N)}{L}x} = e^{i\frac{2\pi(n+N)}{Na_0}ra_0} = e^{i\frac{2\pi n}{N}r + i2\pi r} = e^{i\frac{2\pi n}{N}r} = e^{i\frac{2\pi n}{L}x} \tag{6.55}$$

where a_0 is the spacing between unit cells.

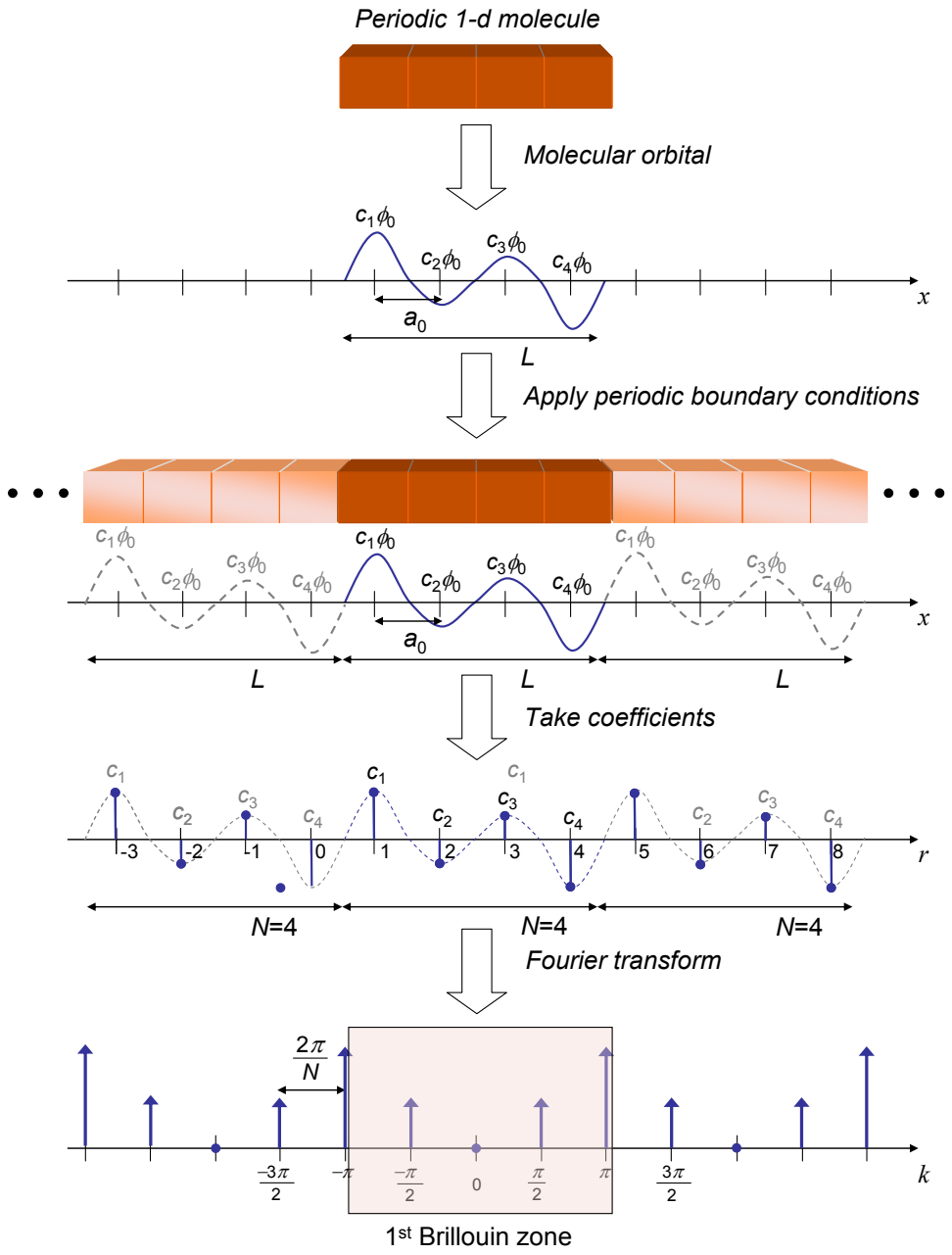


Fig. 6.25. A molecular orbital is described by linear combinations of the wavefunction of the unit cell. The coefficients, c_r, are phase factors. The phase coefficients are discrete – there are only N of them. Thus, the Fourier transform of the coefficients contains only N unique values (it is periodic). We can restrict the range of k values without losing information. Typically, we chose *k* values in the first Brillouin zone ($-\pi/a_0 < k \leq \pi/a_0$). Note also that the application of periodic boundary conditions fixes the spacing between k values at $2\pi/L$.

2-d and 3-d periodic materials

Applying Bloch functions to periodic 2d and 3d molecules follows the same principles as in 1d; see Fig. 6.26.

The molecular orbitals in 2-d and 3-d periodic materials are still composed of linear combinations of the wavefunction of the unit cell, ϕ_0, i.e.

$$\psi = \sum_r c_r \phi_0 \,. \tag{6.56}$$

Once again, when we apply periodic boundary conditions the area occupied in k-space per k-state is: (for 2-d and 3-d, respectively)

$$\Delta k^2 = \Delta k_x \Delta k_y = \frac{2\pi}{L_x}\frac{2\pi}{L_y} = \frac{4\pi^2}{A}, \quad \Delta k^3 = \Delta k_x \Delta k_y \Delta k_z = \frac{2\pi}{L_x}\frac{2\pi}{L_y}\frac{2\pi}{L_z} = \frac{8\pi^3}{V} \tag{6.57}$$

where A is the area of the molecule, and V is its volume.

3-d periodic materials are usually known as crystals. Si and the rest of the common semiconductor materials fall into the category of 3-d periodic materials.

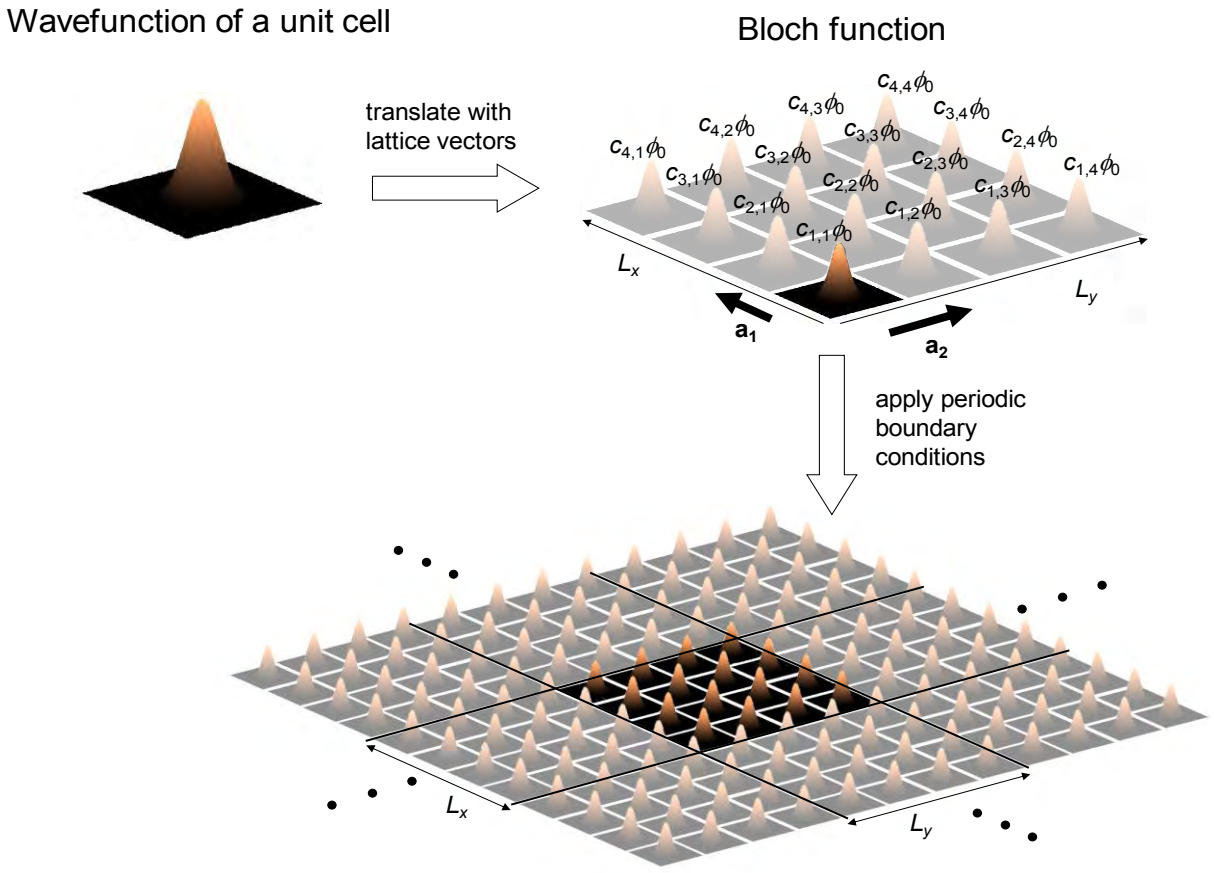


Fig. 6.26. An example of a 2-d periodic material.

Tight Binding Calculations in Periodic molecules and crystals

Polyacetylene (average bond model)

We now repeat the polyacetylene calculation, but this time we impose periodic boundary conditions and assume molecular wavefunctions of the Bloch form. The solutions are almost identical to the previous calculation in the absence of periodic boundary conditions, but there are some subtle yet important differences in the dispersion relation.

The unit cell of polyacetylene under the average bond model has only a single carbon atom. Let the wavefunction of the *j*th unit cell be $\phi(j)$, defined as the frontier atomic orbital of carbon. Since polyacetylene is periodic, we use a Bloch function to describe the molecular orbitals

$$\psi(x) = \sum_j e^{i\frac{2\pi n}{N}j} \phi(x - ja_0) \tag{6.58}$$

To derive the energy levels consider $\langle \phi_j | H | \psi(x) \rangle$. Now,

$$\langle \phi_j | H | \psi(x) \rangle = \varepsilon \langle \phi_j | \psi(x) \rangle \tag{6.59}$$

Under the tight binding approximation, this simplifies to

$$\beta \exp\left[i\frac{2\pi n}{N}(j-1) \right] + \alpha \exp\left[i\frac{2\pi n}{N} j \right] + \beta \exp\left[i\frac{2\pi n}{N}(j+1) \right] = \varepsilon \exp\left[i\frac{2\pi n}{N} j \right] \tag{6.60}$$

where $\beta = \langle \phi_j | H | \phi_{j+1} \rangle$ and $\alpha = \langle \phi_j | H | \phi_j \rangle$.

Simplifying gives (compare Eq. (6.43))

$$\varepsilon_n = \alpha + 2\beta \cos\frac{2\pi n}{N}. \tag{6.61}$$

Re-writing Eq. (6.61) gives

$$\varepsilon_k = \alpha + 2\beta \cos k a_0 \tag{6.62}$$

where

$$k = \frac{2\pi n}{N a_0} = \frac{2\pi n}{L} \tag{6.63}$$

Once again, we note that each carbon atom contributes a single electron to its frontier orbital, thus for a *N*-repeat polymer, there are *N* electrons.

We can determine whether polyacetylene is a metal or insulator by counting *k* states. The spacing between *k* states is $2\pi/L$. Thus in the first Brillouin zone, there must be $2\pi/a_0 \,/\, 2\pi/L = L/a_0 = N$ states. But each molecular orbital holds two electrons, one of each spin. Filling the lowest energy states first, only the first $N/2$ *k* states are filled; see Fig. 6.27. With only half its *k* states filled, polyacetylene might be expected to be a metal.

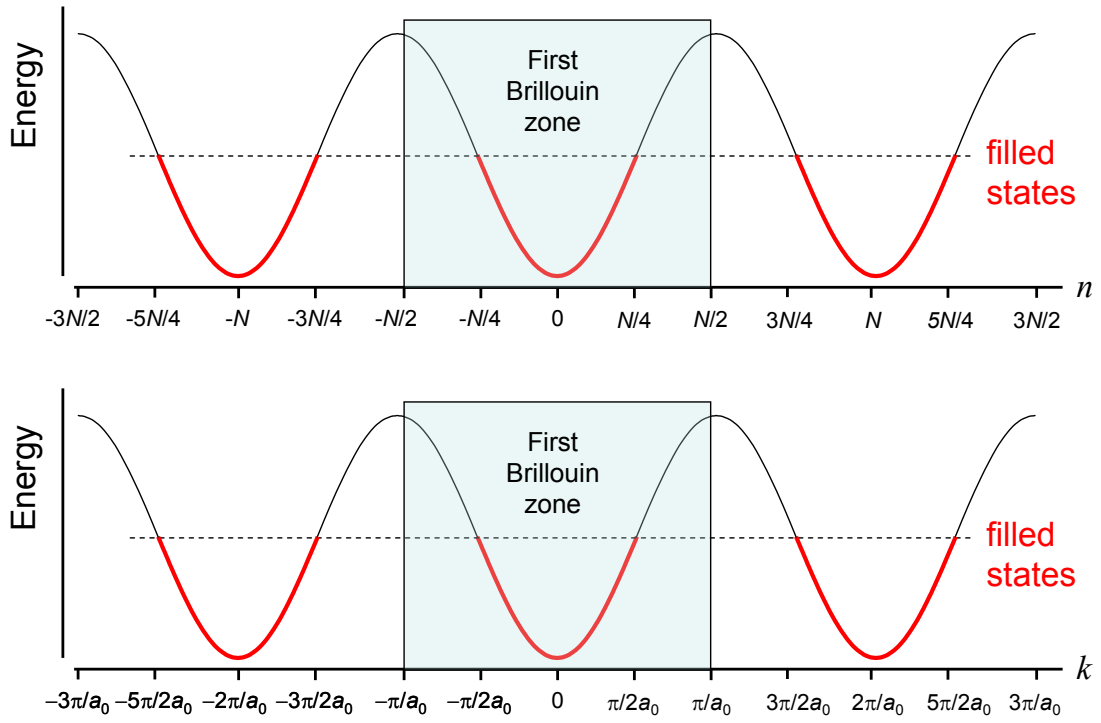


Fig. 6.27. Energy states in polyacetylene as determined by a tight binding analysis. Since the band of states is only half full this material might be expected to be a metal.

Question: *Why does the spacing between k states under periodic boundary conditions differ from that calculated for an isolated strand of polyacetylene – see Eq.*(6.46)*?*

Answer: In the previous section, we analyzed the allowed k states in a periodic linear molecule, polyacetylene. We did not employ periodic boundary conditions, thus we would expect that the solutions would only be comparable as the number, N, of unit cells increases, proportionately reducing the impact of the differing boundary conditions. But for $N\rightarrow\infty$ we still find Δk(*isolated polyacetylene*) = ½ Δk(*polyacetylene in periodic boundary conditions*).

The answer to this conundrum is that isolated polacetylene (i.e. actual polyacetylene – not polyacetylene with infinite copies to the left and right) can only support *standing* waves; there are no contacts that can inject charge, hence no solely left or right-propagating waves. Thus, considering both positive and negative values of k in isolated polyacetylene makes no sense. Rather, k ranges from 0 to π/a_0. There must be N states in this range, and we obtain $\Delta k=\pi/L$.

Given periodic boundary conditions, the polymer has infinite length. A wave could propagate to the left or right indefinitely. So we must consider both positive and negative values of k, i.e. k ranges from $-\pi/a_0$ to π/a_0. There must be N states in this range, and we obtain $\Delta k=2\pi/L$.

Interestingly, we are almost never interested in a completely isolated electronic material. For example, practical systems must have contacts to inject charge! Thus, periodic boundary conditions that allow for propagating waves often come closer to modeling practical systems.

Polacetylene (alternating bond model)

Next, let's see what happens to the dispersion relation under the alternating bond model. Now each unit cell has two carbon atoms; see Fig. 6.19. We'll model the unit cell with a *linear combination* of two frontier atomic orbitals because the contributions of each atomic orbital to the unit cell could vary.

Let the wavefunction of the *j*th unit cell be

$$\phi(j) = c_1\phi_1(j) + c_2\phi_2(j) \tag{6.64}$$

where $\phi_1(j)$ and $\phi_2(j)$ are the frontier atomic orbitals of the first and second carbon atom in the *j*th unit cell, respectively.

We must define two hopping integrals. For single bonds we have

$$\beta_S = \langle \phi_1(j-1) | H | \phi_2(j) \rangle \tag{6.65}$$

and for double bonds

$$\beta_D = \langle \phi_1(j) | H | \phi_2(j) \rangle . \tag{6.66}$$

As before, $\alpha = \langle \phi_1(j) | H | \phi_1(j) \rangle = \langle \phi_2(j) | H | \phi_2(j) \rangle$.

Assuming a wavefunction of the Bloch form (Eq. (6.58)) we take

$$\psi(x) = \sum_j e^{i\frac{2\pi n}{N}j} \phi(x - j2a_0), \tag{6.67}$$

where we note that the spacing between unit cells is now $2a_0$.

Let's now consider two overlap equations

$$\begin{aligned} \langle \phi_1(j) | H | \psi \rangle &= \varepsilon \langle \phi_1(j) | \psi \rangle \\ \langle \phi_2(j) | H | \psi \rangle &= \varepsilon \langle \phi_2(j) | \psi \rangle \end{aligned} . \tag{6.68}$$

Under the tight binding approximations, the LHS of Eq. (6.68) expands to

$$\begin{aligned} \langle \phi_1(j) | H | \psi \rangle &= \left(c_2\beta_S \exp\left[-i\frac{2\pi n}{N} \right] + c_1\alpha + c_2\beta_D \right) \exp\left[i\frac{2\pi n}{N} j \right] \\ \langle \phi_2(j) | H | \psi \rangle &= \left(c_2\alpha + c_1\beta_D + c_1\beta_S \exp\left[i\frac{2\pi n}{N} \right] \right) \exp\left[i\frac{2\pi n}{N} j \right] \end{aligned} \tag{6.69}$$

The RHS of Eq. (6.68) expands to

$$\varepsilon\langle\phi_1(j)|\psi\rangle = \varepsilon c_1 \exp\left[i\frac{2\pi n}{N}j\right]$$
$$\varepsilon\langle\phi_2(j)|\psi\rangle = \varepsilon c_2 \exp\left[i\frac{2\pi n}{N}j\right] \tag{6.70}$$

The solution for non-trivial c_1, c_2 is given by

$$\begin{vmatrix} \alpha-\varepsilon & \beta_S \exp\left[-i\frac{2\pi n}{N}\right]+\beta_D \\ \beta_S \exp\left[i\frac{2\pi n}{N}\right]+\beta_D & \alpha-\varepsilon \end{vmatrix} = 0 \tag{6.71}$$

i.e.

$$\varepsilon = \alpha \pm \sqrt{\beta_S{}^2 + \beta_D{}^2 + 2\beta_S\beta_D \cos 2ka_0} \tag{6.72}$$

In the alternating bond model, the period of polyacetylene is $2a_0$. Thus the number of distinct k values is $2\pi/2a_0$ / $2\pi/L = N/2$, where N is the number of carbon atoms in the polymer backbone.

But there are two solutions for the energy at each k value (i.e. there are two energy bands), so the total number of states is N. Since each state holds two electrons, we find that the bottom band is completely full and the top band is completely empty. Thus, the periodic potential formed by alternating single and double bonds opens a *band gap* at $k = \pi/2a_0$, completing transforming the material from a metal to an insulator/semiconductor! Obviously, the accuracy of the DOS calculation is critical.

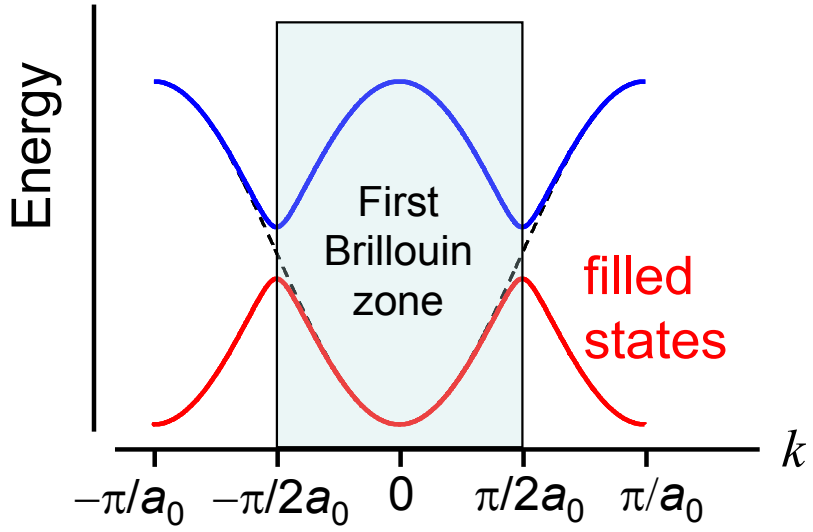


Fig. 6.28. A periodic perturbation with twice the interatomic spacing introduces a gap at the Fermi energy, transforming a metal into an insulator (wide bandgap semiconductor).

Graphene

Like polyacetylene in the alternating bond model, in graphene we have a unit cell with two carbon atoms. Let the wavefunction of the unit cell be

$$\phi = c_1\phi_1 + c_2\phi_2 \tag{6.73}$$

where ϕ_1 and ϕ_2 are the frontier atomic orbitals of the first and second carbon atom in the unit cell, respectively.

We assume a wavefunction of the Bloch form (Eq. (6.58)) but we re-write it in terms of a sum over all lattice vectors **R**:

$$\psi(\mathbf{x}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\bullet\mathbf{R}}\phi(\mathbf{x}-\mathbf{R}). \tag{6.74}$$

Next we define the hopping integral

$$\beta = \langle\phi_1|H|\phi_2\rangle \tag{6.75}$$

As before, $\alpha = \langle\phi_j|H|\phi_j\rangle$.

Let"s now consider two overlap equations

$$\begin{aligned}\langle\phi_1(\mathbf{R})|H|\psi(\mathbf{x})\rangle &= \varepsilon\langle\phi_1(\mathbf{R})|\psi(\mathbf{x})\rangle\\ \langle\phi_2(\mathbf{R})|H|\psi(\mathbf{x})\rangle &= \varepsilon\langle\phi_2(\mathbf{R})|\psi(\mathbf{x})\rangle\end{aligned}. \tag{6.76}$$

Under the Hückel/tight binding approximations, the LHS of Eq. (6.76)expands to

$$\begin{aligned}\langle\phi_1(\mathbf{R})|H|\psi(\mathbf{x})\rangle &= \left(c_1\alpha + c_2\beta\left(1+e^{-i\mathbf{k}\bullet\tilde{\mathbf{a}}_1}+e^{-i\mathbf{k}\bullet\tilde{\mathbf{a}}_2}\right)\right)e^{i\mathbf{k}\bullet\mathbf{R}}\\ \langle\phi_2(\mathbf{R})|H|\psi(\mathbf{x})\rangle &= \left(c_2\alpha + c_1\beta\left(1+e^{i\mathbf{k}\bullet\tilde{\mathbf{a}}_1}+e^{i\mathbf{k}\bullet\tilde{\mathbf{a}}_2}\right)\right)e^{i\mathbf{k}\bullet\mathbf{R}}\end{aligned}. \tag{6.77}$$

The RHS of Eq. (6.76) expands to

$$\begin{aligned}\varepsilon\langle\phi_1(\mathbf{R})|\psi(\mathbf{x})\rangle &= \varepsilon c_1 e^{i\mathbf{k}\bullet\mathbf{R}}\\ \varepsilon\langle\phi_2(\mathbf{R})|\psi(\mathbf{x})\rangle &= \varepsilon c_2 e^{i\mathbf{k}\bullet\mathbf{R}}\end{aligned} \tag{6.78}$$

The solution for non-trivial c_1, c_2 is given by

$$\begin{vmatrix}\alpha-\varepsilon & \beta\left(1+e^{-i\mathbf{k}\bullet\tilde{\mathbf{a}}_1}+e^{-i\mathbf{k}\bullet\tilde{\mathbf{a}}_2}\right)\\ \beta\left(1+e^{i\mathbf{k}\bullet\tilde{\mathbf{a}}_1}+e^{i\mathbf{k}\bullet\tilde{\mathbf{a}}_2}\right) & \alpha-\varepsilon\end{vmatrix} = 0 \tag{6.79}$$

i.e.

$$\varepsilon = \alpha \pm \beta\sqrt{3+2\cos\left(\mathbf{k}\bullet\tilde{\mathbf{a}}_1\right)+2\cos\left(\mathbf{k}\bullet\tilde{\mathbf{a}}_2\right)+2\cos\left(\mathbf{k}\bullet\left(\tilde{\mathbf{a}}_1-\tilde{\mathbf{a}}_2\right)\right)} \tag{6.80}$$

This is plotted in Fig. 6.29, where we have arbitrarily set $\alpha = 0$.

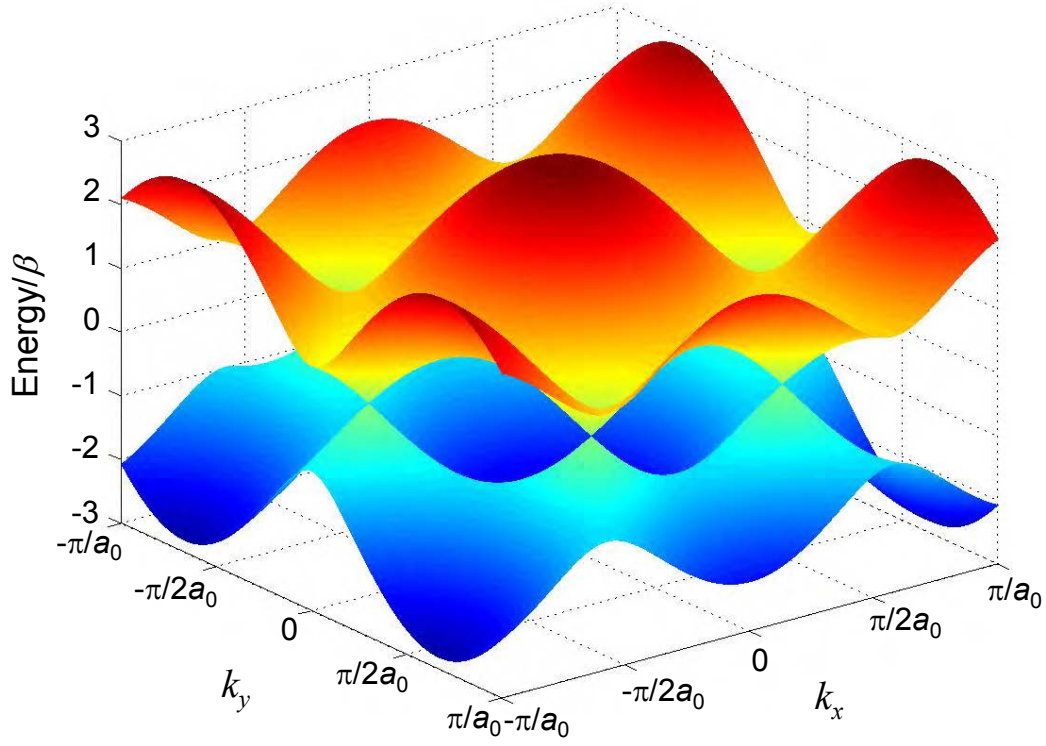


Fig. 6.29. The bandstructure of graphene.

Each unit cell contributes an orbital to a band; given N unit cells, each band has N states, or including spin, $2N$ states. Graphene, with two electrons per unit cell has two bands and $2N$ electrons. Thus, the lower band of graphene is completely filled.

We might therefore expect that graphene is an insulator, but the lower band touches the upper band at values of k known as the K points

$$\mathbf{K} = \left(\pm \frac{4\pi}{3\sqrt{3}a_0}, 0 \right), \left(\pm \frac{2\pi}{3\sqrt{3}a_0}, \pm \frac{2\pi}{3a_0} \right)$$

Thus, in these particular directions graphene conducts like a metal.

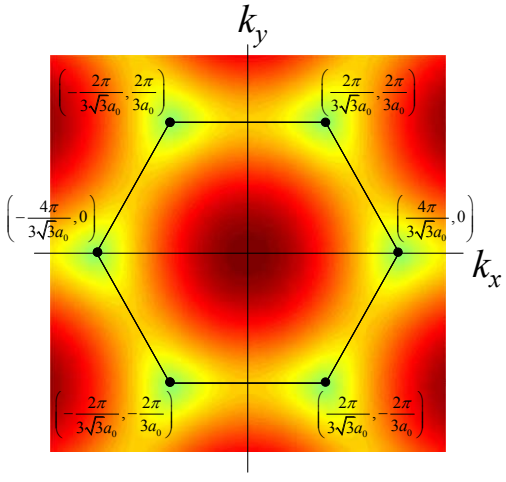


Fig. 6.30. The K points in graphene.

Carbon Nanotubes

Carbon nanotubes are remarkable materials. They are perhaps the most rigid materials known, and they have excellent charge transport properties.

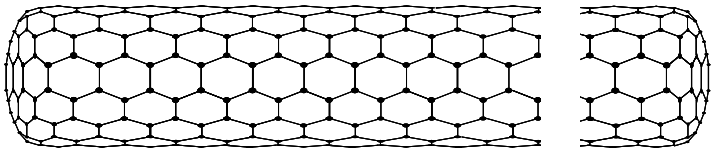


Fig. 6.31. An example of a carbon nanotube – composed of a rolled up graphene sheet. This particular tube is known as an armchair – you may be able to identify armchairs in the hexagonal lattice.

We will treat carbon nanotubes as rolled-up graphene sheets. The construction of a nanotube from a sheet of graphene can be imagined as in Fig. 6.32. We first draw the wrapping vector from one unit cell to another. When the tube is formed both ends of the wrapping vector will be connected. The wrapping vector is written

$$\mathbf{w} = n\tilde{\mathbf{a}}_1 + m\tilde{\mathbf{a}}_2 = (n, m) \tag{6.81}$$

The length of the wrapping vector determines the circumference of the tube, and as we shall see the vector (n,m) characterizes its electronic properties.

Next two parallel cuts are made perpendicular to the wrapping vector and the remaining piece is rolled up; see Fig. 6.33.[†]

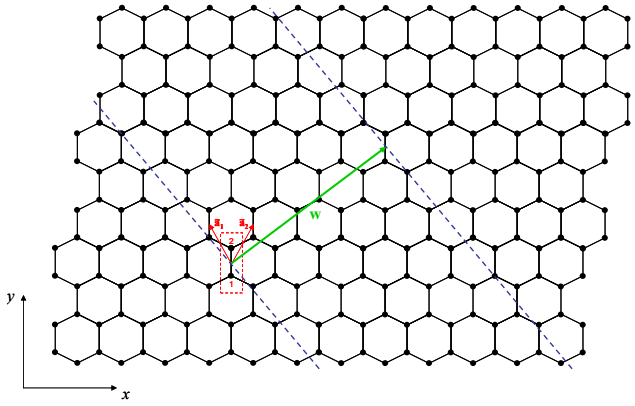


Fig. 6.32. Carbon nanotubes can be imagined to be constructed from rolled up pieces of graphene. In the example above, a wrapping vector, **w**, is drawn between two unit cells that will be connected when the tube is rolled up. The graphene sheet is cut perpendicular to **w**.

[†] Of course, carbon nanotubes are not actually made from graphene like this. There are many techniques including chemical vapor deposition using a catalyst particle that defines the width of the tube.

After the tube is rolled up, periodic boundary conditions are established on the circumference of the tube. Thus, only certain values of the wavevector, k, are allowed perpendicular to the tube axis, i.e.

$$\mathbf{k}\bullet\mathbf{w} = 2\pi l, \quad l \in \mathbb{Z} \tag{6.82}$$

If the allowed k-states include the K points of graphene then the carbon nanotube will be a metal, otherwise it is a semiconductor. For example, consider a (4,4) armchair nanotube as shown below. We also plot the K points for graphene in k space.

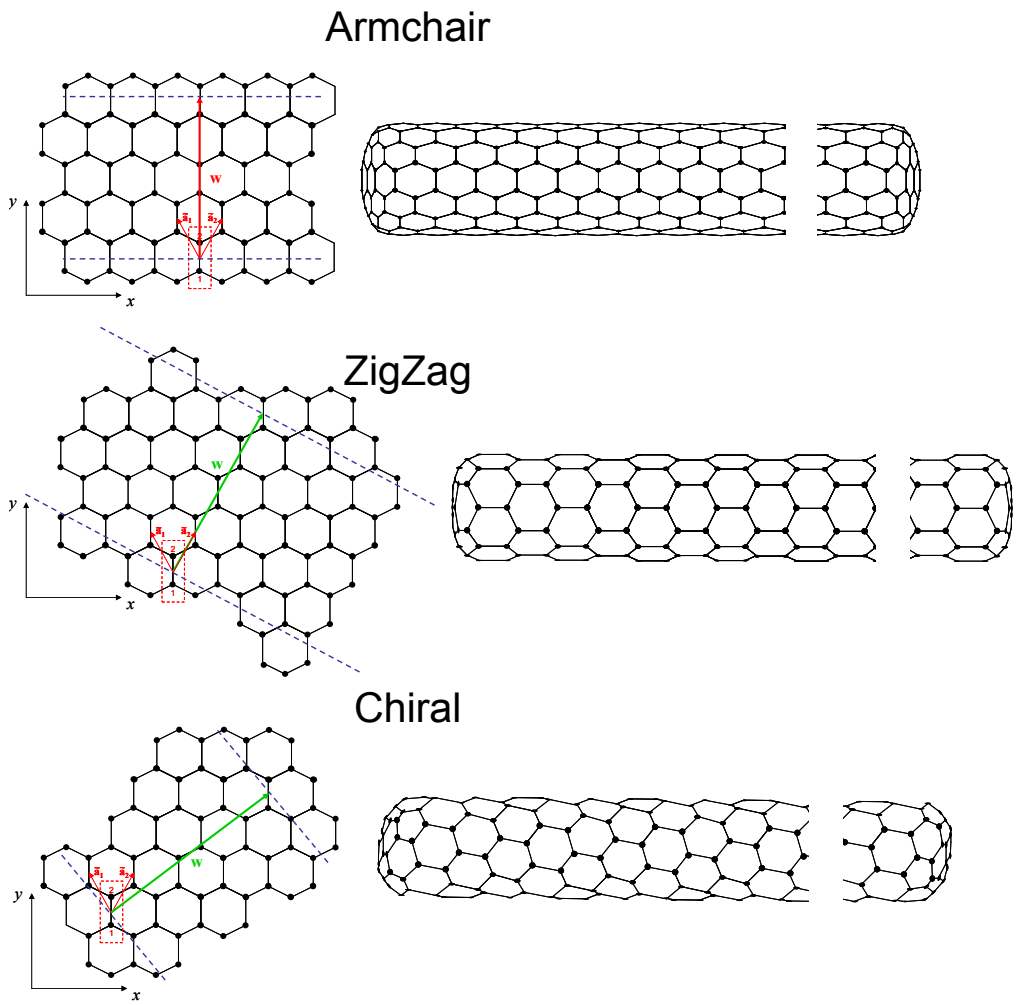


Fig. 6.33. Three types of nanotubes. The first two, armchair and zigzag are special cases with wrapping vectors (N,N) and (N,0) or (0,N), respectively. The third is the general case or chiral form with wrapping vector (n,m).

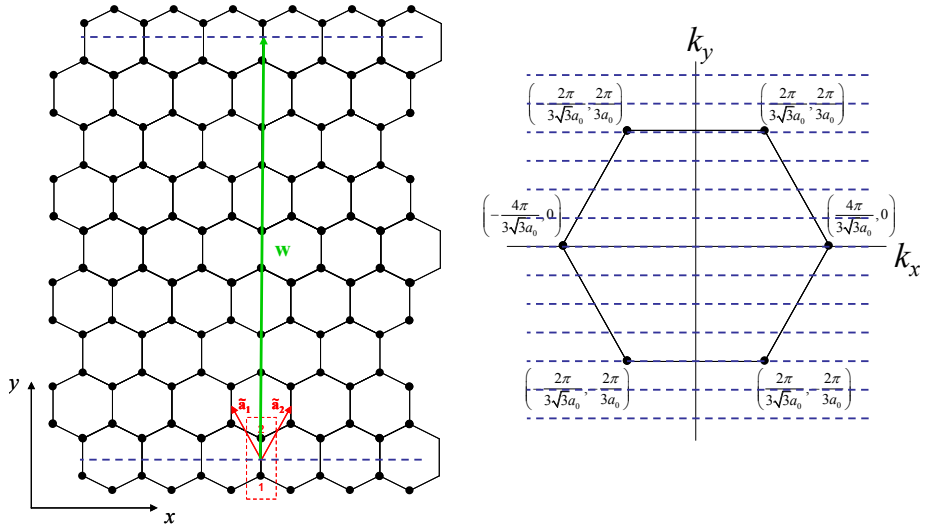


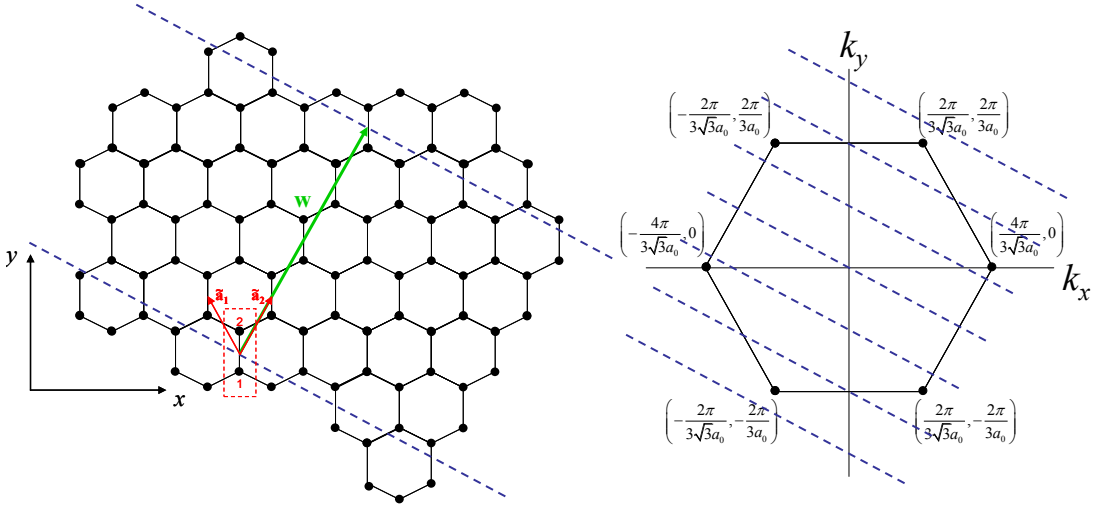
Fig. 6.34. At left, a (4,4) armchair nanotube. At right, the K points of graphene.

Let's begin by decomposing $\mathbf{k}$ into a component perpendicular to the tube axis, $\mathbf{k}_\perp$, and a component parallel to the tube axis, $\mathbf{k}_\parallel$. For a (4,4) tube $\mathbf{w} = 12a_0\hat{\mathbf{y}}$. Thus, the allowed values of $\mathbf{k}$ are given by

$$k_y = \frac{\pi l}{6a_0}. \tag{6.83}$$

As shown in Fig. 6.34, this set of allowed $\mathbf{k}_\perp$ values includes the K points. Thus (4,4) tubes are metallic.

Next, let's examine a (0,4) zigzag tube.



For a (0,4) tube $\mathbf{w} = 2\sqrt{3}a_0\hat{\mathbf{x}} + 6a_0\hat{\mathbf{y}}$. Thus, the allowed values of $\mathbf{k}_\perp$ are given by

Fig. 6.35. At left, a (0,4) zigzag nanotube. At right, the K points of graphene.

$$2\sqrt{3}k_x + 6k_y = \frac{2\pi l}{a_0}. \tag{6.84}$$

As shown in Fig. 6.35, this set of allowed **k** values does not include the K points. Thus (0,4) tubes are insulating/semiconducting.

Analytic approximations for the bandstructure of graphene and carbon nanotubes

Since the conduction properties of graphene are dominated by electrons occupying states at or near the K points, it is convenient to linearize the energy at $\boldsymbol{\kappa} = \mathbf{k} - \mathbf{K}$.

The exact tight binding solution from Eq. (6.80)is:

$$\varepsilon = \alpha \pm \beta\sqrt{3 + 2\cos\left(\mathbf{k}\bullet\tilde{\mathbf{a}}_1\right) + 2\cos\left(\mathbf{k}\bullet\tilde{\mathbf{a}}_2\right) + 2\cos\left(\mathbf{k}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right)} \tag{6.85}$$

We substitute $\mathbf{k} = \mathbf{K} + \boldsymbol{\kappa}$ and expand the cos(**K**+**κ**) terms as a Taylor series to second order in **κ**. This yields:

$$\varepsilon = \alpha \pm \beta\sqrt{\begin{array}{c} 3 + 2\cos\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_1\right) + 2\cos\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_2\right) + 2\cos\left(\mathbf{K}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right) \\ +2\boldsymbol{\kappa}\bullet\tilde{\mathbf{a}}_1\sin\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_1\right) + 2\boldsymbol{\kappa}\bullet\tilde{\mathbf{a}}_2\sin\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_2\right) + 2\boldsymbol{\kappa}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\sin\left(\mathbf{K}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right) \\ -\left(\boldsymbol{\kappa}\bullet\tilde{\mathbf{a}}_1\right)^2\cos\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_1\right) - \left(\boldsymbol{\kappa}\bullet\tilde{\mathbf{a}}_2\right)^2\cos\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_2\right) - \left(\boldsymbol{\kappa}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right)^2\cos\left(\mathbf{K}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right) \end{array}} \tag{6.86}$$

Next, we note some identities:

$$\cos\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_1\right) = \cos\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_2\right) = \cos\left(\mathbf{K}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right) = -\frac{1}{2} \tag{6.87}$$

$$\sin\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_1\right) = -\sin\left(\mathbf{K}\bullet\tilde{\mathbf{a}}_2\right) = -\sin\left(\mathbf{K}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right) \tag{6.88}$$

From these identities Eq. (6.86) reduces to

$$\varepsilon = \alpha \pm \beta\sqrt{\frac{1}{2}\left(\boldsymbol{\kappa}\bullet\tilde{\mathbf{a}}_1\right)^2 + \frac{1}{2}\left(\boldsymbol{\kappa}\bullet\tilde{\mathbf{a}}_2\right)^2 + \frac{1}{2}\left(\boldsymbol{\kappa}\bullet\left(\tilde{\mathbf{a}}_1 - \tilde{\mathbf{a}}_2\right)\right)^2} \tag{6.89}$$

Solving this (see the Problem Set) gives the approximate dispersion relation for graphene:

$$\varepsilon = \alpha \pm \frac{3}{2}\beta a_0 |\boldsymbol{\kappa}| \tag{6.90}$$

Since the speed of the charge carrier is given by the group velocity: $v = \hbar^{-1}\,\partial\varepsilon/\partial k$, we get

$$v = \frac{3}{2}\frac{\beta a_0}{\hbar} \tag{6.91}$$

For a_0 = 1.42Å and β = 2.5 eV, $v = 10^6$ m/s.

Now, for carbon nanotubes, the periodic boundary condition on the circumfrence is

$$(\boldsymbol{\kappa}+\mathbf{K})\bullet\mathbf{w}=2\pi l,\quad l\in\mathbb{Z} \tag{6.92}$$

Let"s consider each K point in turn:

For $\mathbf{K}=\left(\frac{4\pi}{3\sqrt{3}a_0},0\right)$

$$\begin{aligned}(\boldsymbol{\kappa}+\mathbf{K})\bullet\mathbf{w}&=\boldsymbol{\kappa}\bullet\mathbf{w}+\mathbf{K}\bullet(n\tilde{\mathbf{a}}_1+m\tilde{\mathbf{a}}_2)\\&=\boldsymbol{\kappa}\bullet\mathbf{w}+n\mathbf{K}\bullet\left(-\frac{\sqrt{3}}{2}a_0,\frac{3}{2}a_0\right)+m\mathbf{K}\bullet\left(\frac{\sqrt{3}}{2}a_0,\frac{3}{2}a_0\right)\\&=\boldsymbol{\kappa}\bullet\mathbf{w}+\frac{2\pi}{3}(m-n)\end{aligned} \tag{6.93}$$

Rearranging gives:

$$\boldsymbol{\kappa}\bullet\mathbf{w}=2\pi l+2\pi\frac{(n-m)}{3} \tag{6.94}$$

For $\mathbf{K}=\left(\frac{2\pi}{3\sqrt{3}a_0},\frac{2\pi}{3a_0}\right)$

$$\begin{aligned}\boldsymbol{\kappa}\bullet\mathbf{w}&=2\pi l+2\pi\frac{(2n+m)}{3}\\&=2\pi l+2\pi\frac{(3n-(n-m))}{3}\\&\equiv 2\pi l-2\pi\frac{(n-m)}{3}\end{aligned} \tag{6.95}$$

For $\mathbf{K}=\left(-\frac{2\pi}{3\sqrt{3}a_0},\frac{2\pi}{3a_0}\right)$

$$\begin{aligned}\boldsymbol{\kappa}\bullet\mathbf{w}&=2\pi l+2\pi\frac{(n+2m)}{3}\\&=2\pi l+2\pi\frac{((n-m)+3m)}{3}\\&\equiv 2\pi l+2\pi\frac{(n-m)}{3}\end{aligned} \tag{6.96}$$

The other K points follow by symmetry, and we can conclude that

$$\boldsymbol{\kappa}_\perp=\frac{2\pi}{|\mathbf{w}|}\left(l+\frac{(n-m)}{3}\right) \tag{6.97}$$

where we have separated $\boldsymbol{\kappa}$ into two components parallel, $\boldsymbol{\kappa}_\parallel$, and perpendicular, $\boldsymbol{\kappa}_\perp$ to the tube axis. From Eq. (6.90) we get

$$\varepsilon = \alpha \pm \frac{3\beta a_0}{d}\sqrt{\left(l+\left(\frac{n-m}{3}\right)\right)^2+\left(\frac{\boldsymbol{\kappa}_{\|} d}{2}\right)^2}\,. \tag{6.98}$$

where the tube circumference is $|\mathbf{w}| = \pi d$. Interestingly, Eq. (6.98) predicts that tubes are metallic when $\left[(n-m)/3\right] \in \mathbb{Z}$. Assuming that n and m are generated randomly, we expect that 1/3 of tubes should be metallic. Indeed, this seems to be the case in practice. Note also that for semiconducting tubes the band gap is inversely proportional to the tube diameter.

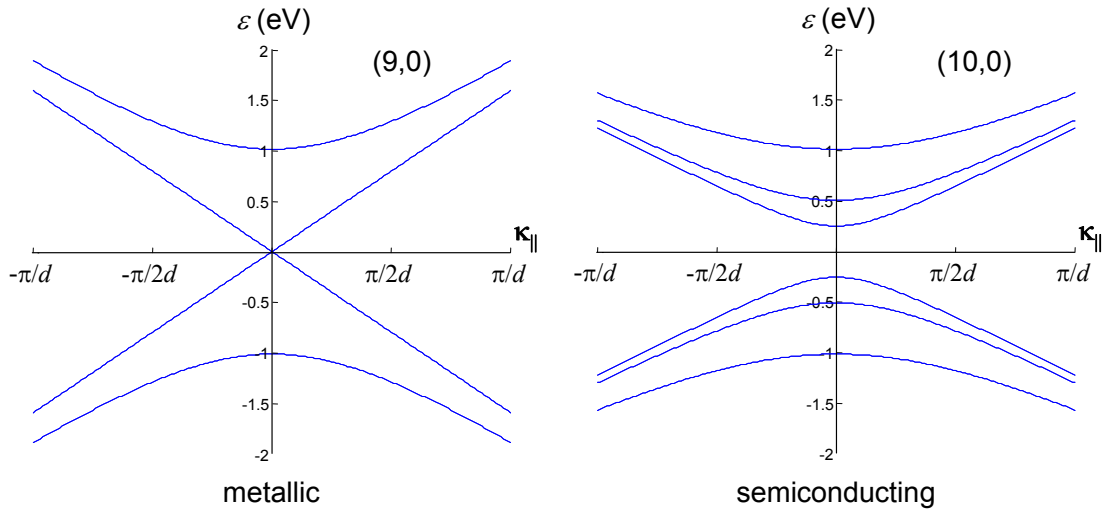


Fig. 6.36. Approximate band structures for metallic and semiconducting zigzag tubes.

Bandstructure of bulk semiconductors

As stated above, most of the common semiconductors are constructed from sp^3-hybridized atoms assembled in the diamond crystal structure.

Unfortunately, sp^3-hybridization makes the bandstructure calculation much harder. In our earlier sp^2-hybridized examples, we were able to ignore all of the atomic orbitals involved in σ bonds, and we considered only the π-bonding electrons from the unhybridized p atomic orbital. Since sp^3-hybridized materials only contain σ bonds, we can no longer employ this approximation and our calculation must include all four sp^3-hybridized atomic orbitals. In fact, to obtain reasonable accuracy, bandstructure calculations usually include the next highest unfilled orbital also.

For a diamond structure, with a two atom unit cell, this means that we must consider 10 atomic orbitals per unit cell (one set of five per atom in the unit cell).

The calculation itself follows the procedures we described for graphene. But now we must solve a 10x10 matrix. We will draw the line here in this class.

Band Gaps and Conduction and Valence bands

As shown in Fig. 6.37, if the Fermi energy separates two bands of allowed electron states, the upper band is empty and the lower band is full. The energy difference between the top of the filled band and the bottom of the empty band is known as the *band gap*. Empty bands that lie above the Fermi energy are known as *conduction bands*. Filled bands that lie below the Fermi energy are known as *valence bands*. In this class, we have mostly considered charge transport through the conduction band. Charge may also move through vacancies in the normally full valence band. These vacancies are known as holes. Holes are effectively positively charged because the semiconductor no longer has its full complement of electrons. Consequently, a MOSFET that conducts through its valence band is known as p-channel MOSFET. It requires a negative V_{GS} to turn on.

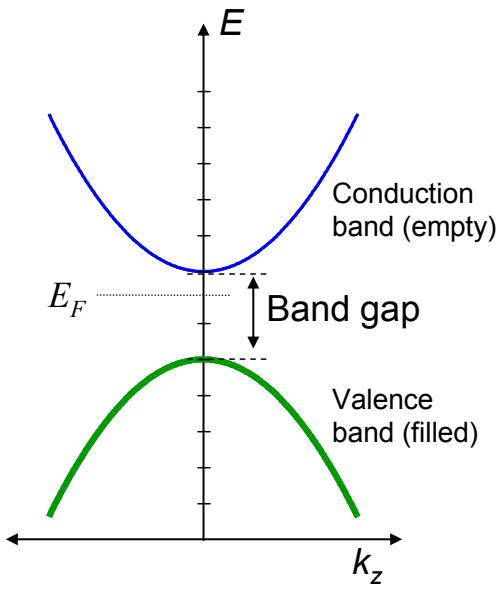


Fig. 6.37. The energy difference between the top of a filled (valence) band and the bottom of an empty (conduction) band is known as the band gap.

Band diagrams

Often the dispersion relation is simplified to show just the bottom of the conduction band, and the top of the valence band. This simplification is useful because usually the density of states is sufficiently large in a bulk semiconductor that the gate is prevented from pushing the Fermi level into the band. Plots showing the band edges as a function of position are known as a band diagrams.

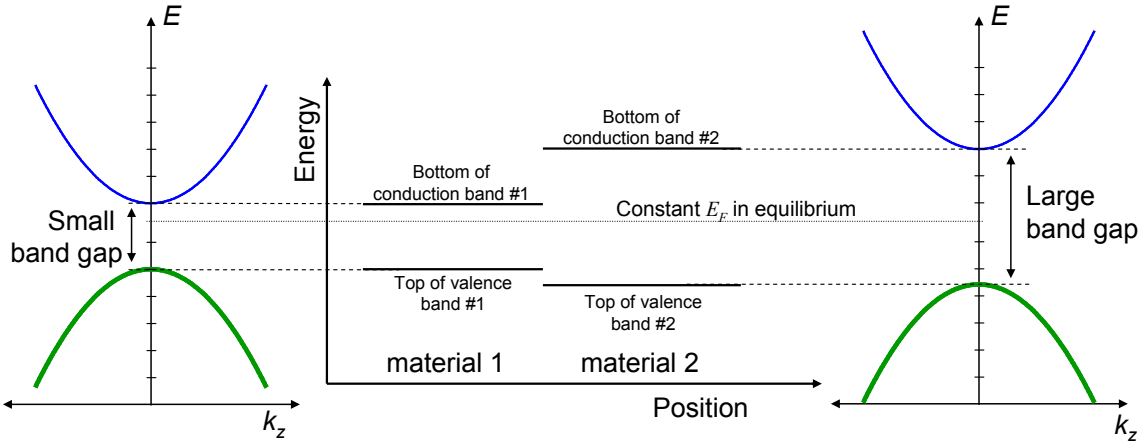


Fig. 6.38. An example of a band diagram. Here two insulators with different band gaps are connected. Note that the Fermi level must be constant in the two materials in equilibrium.

Semiconductors and Insulators

We have seen that insulators do not conduct because there are no uncompensated electrons. Considering both the conduction and valence band in the form shown in Fig. 6.38, it is evident that the material is an insulator when the Fermi energy lies in the bandgap. But if we introduce electrons to the conduction band, or remove electrons from the valence band, what was once an insulator can be transformed into a conductor. In fact if the Fermi energy is *close* to either band edge, then even a small movement in the Fermi energy can significantly modulate the conductivity. Such materials are known as *semiconductors* because it is easy to modulate them between the metallic and insulating regimes.

Sometimes, it can be difficult to distinguish between insulators and semiconductors. But insulators tend to have band gaps exceeding several electron-Volts and a Fermi energy close to the center of the band gap.

Problems

1. Consider the interaction of two carbon atoms each with one electron in a frontier $2p_z$ atomic orbital. Assuming the positions of the atoms are fixed, the Hamiltonian of the system consists of a kinetic energy operator, and two Coulombic potential terms: one for the central atom and one for its neighbor:

$$H = T + V_1 + V_2$$

Assume the wavefunction in this two atom system can be written as

$$\psi = c_1\phi_1 + c_2\phi_2$$

where ϕ_1 and ϕ_2 are the $2p_z$ atomic orbitals on the first and second carbon atoms, respectively, and c_1 and c_2 are constants.

The self energy is defined as

$$\alpha_r = \langle \phi_r | T + V_r | \phi_r \rangle$$

The hopping interactions are defined as

$$\beta_{sr} = \langle \phi_s | V_s | \phi_r \rangle$$

Earlier, we assumed that the *overlap integral* between frontier orbitals on atomic sites s and r could be approximated as

$$S_{sr} = \langle \phi_s | \phi_r \rangle = \delta_{sr}.$$

Do not make that assumption here and show that the electron energies of the system satisfy

$$\det(H - ES) = 0 \tag{6.99}$$

where H is a 2×2 Hamiltonian matrix and S is a 2×2 overlap matrix and E is a constant.

(a) Write each matrix in Eq. (6.99) in terms of the self energies, hopping integrals and overlap integrals.

(b) Under what conditions can you safely ignore the overlap integrals?

2. (a) Consider the potential $V(x) = -V_0\delta(x)$, sketched below.

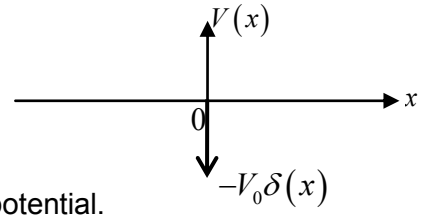


Fig. 6.39. A delta function potential.

(i) Show that the wavefunction is given by $\phi_1(x) = \sqrt{k}e^{-k|x|}$ where $k = \frac{mV_0}{\hbar^2}$

(ii) Show that the energy of the bound states ($E<0$) is $E_1 = -\frac{mV_0^2}{2\hbar^2}$.

(b) Now add a second delta function potential at x = a.

i.e. if the previous Hamiltonian was $H_1 = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2} - V_0\delta(x)$, the new Hamiltonian is $H = H_1 + V$ where $V = -V_0\delta(x-a)$

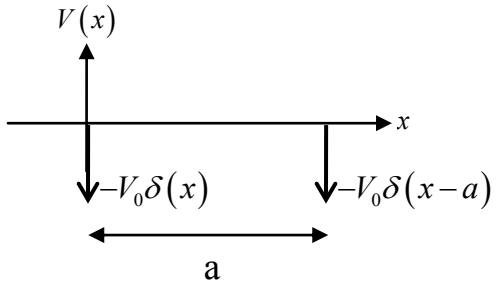


Fig. 6.40. Two delta function potentials.

Let the wavefunction of the new system be approximated by $\psi = c_1\phi_1 + c_2\phi_2$ where $\phi_2 = \phi_1(x-a)$ and c_1 and c_2 are constants.

The self energy is $\alpha = \langle\phi_1|H_1|\phi_1\rangle$

The hopping interaction is $\beta = \langle\phi_2|V|\phi_1\rangle$

In addition, define the overlap integral $S = \langle\phi_1|\phi_2\rangle$, and $\gamma = \langle\phi_1|V|\phi_1\rangle$

By evaluating the expressions

$$\langle\phi_1|H|\psi\rangle = E\langle\phi_1|\psi\rangle$$

and

$$\langle\phi_2|H|\psi\rangle = E\langle\phi_2|\psi\rangle$$

show that

$$\begin{pmatrix} \alpha+\gamma & \alpha S+\beta \\ \alpha S+\beta & \alpha+\gamma \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = E\begin{pmatrix} 1 & S \\ S & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Now show that $\alpha = \frac{-mV_0^2}{2\hbar^2}$, $\beta = \frac{-mV_0^2}{\hbar^2}e^{-ka}$, $S = (1+ka)e^{-ka}$, and $\gamma = \frac{-mV_0^2}{\hbar^2}e^{-2ka}$
Dropping terms containing e^{-2ka}, show that the matrix reduces to $E \approx \alpha \pm \beta$

3. For molecules where each carbon atom contributes at least one delocalized electron to a π orbital, we can use the perimeter free electron orbital theory approximation, which is described below.

Assume that the molecule in question is a circular ring of atoms and assume an infinite square well potential.

(a) Show that the energy levels of a molecule under this approximation are

$$E = \frac{h^2 m_l^2}{2m_e L^2},$$

where where m_l is an integer, and L is the perimeter of the molecule.

Hint: The Hamiltonian in polar coordinates is given by:

$$\hat{H} = \frac{\hbar^2}{2m_e}\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} + \frac{1}{r^2}\frac{d^2}{d\phi^2}\right)$$

(b) According to the perimeter free electron orbital theory approximation, the energy level structure of anthracene is shown in Fig. 6.41, below.

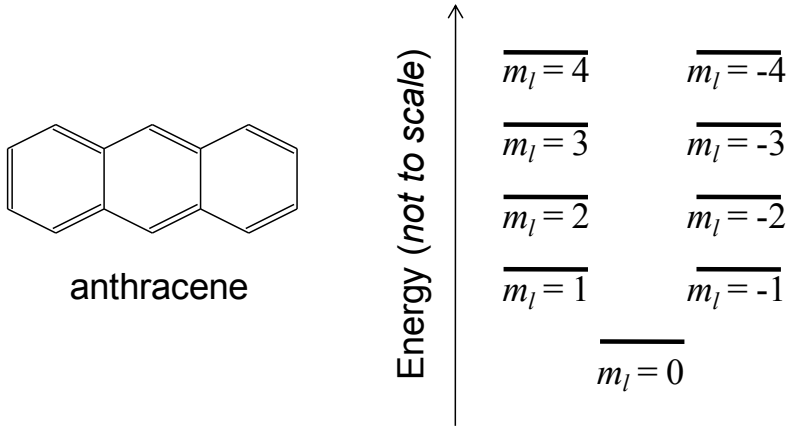


Fig. 6.41. The molecular and energetic structure of anthracene.

Continued on next page….

(i) Why is there a solution for $m_l = 0$ in anthracene but no solution for $n = 0$ in the infinite quantum well?
(ii) Why are there solutions for negative m_l in anthracene but no solutions for negative n in the infinite quantum well? *Hint:* consider the Pauli exclusion principle.

(c) Calculate the molecular orbitals and HOMO-LUMO gap of anthracene. Take $a = 1.38$Å as the C-C bond length. Assume each C atom donates 1 electron to the frontier orbitals.

4. Consider the periodic molecule consisting of two different alternating atom types illustrated below (frontier orbitals are shown).

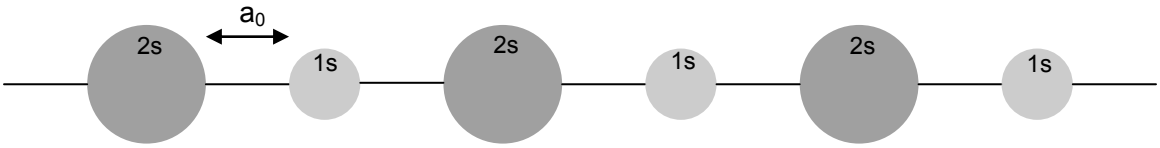


Fig. 6.42. A periodic array of two different atoms.

(a) How many atoms are in the unit cell in this molecule? Using periodic boundary conditions and assuming molecular wavefunctions of the Bloch form, find the energy levels.

(b) Find the density of states.

5. The band structure of molecular crystals

Let $\phi(\mathbf{r})$ be the HOMO of a typical molecule. As in most stable molecules, $\phi(\mathbf{r})$ is fully occupied and contains two electrons.

Unlike conventional crystalline semiconductors such as Si, the unit cells in a molecular crystal are held together by weak van der Waals forces. A typical value for the interaction between nearest neighbors in a van der Waals bonded solid is

$$\beta = \langle \phi(\mathbf{r}+\mathbf{R}) | H | \phi(\mathbf{r}) \rangle \approx -10 \text{ meV}$$

where H is the Hamiltonian for the interaction between nearest neighbors and **R** is the set of lattice vectors connecting the molecule at **r** to its nearest neighbors.

(a) Calculate the „valence" band structure of a cubic molecular crystal of this molecule. Let $\langle \phi(\mathbf{r}) | H | \phi(\mathbf{r}) \rangle = \alpha$. (See Fig. 6.43 below).

Continued on next page….

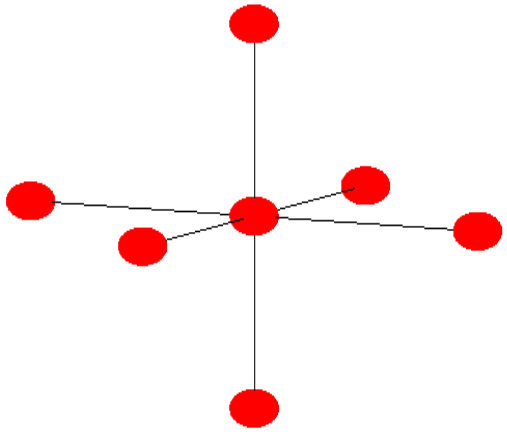


Fig. 6.43. The structure of a simple cubic molecular crystal

(b) Show that all molecular crystals with filled HOMOs are insulators.

6. Consider the following polymer:

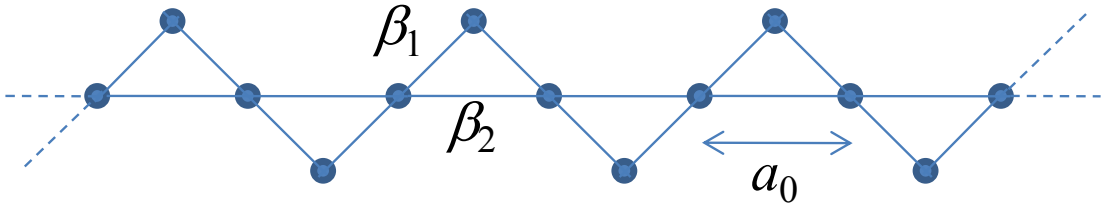


Fig. 6.44. A polymer.

Assume the spacing between atoms on the linear backbone is a_0, as shown. Also, assume all atoms are the same element, β_1 and β_2 are the hopping interactions between atoms as shown, the self energy at each atom is α, *and a*ssume each atom contributes one electron.

(a) What is the primitive unit cell and primitive lattice vector?

(b) Show that the dispersion relation is given by

$$E = \alpha + \beta_2 \cos(ka_0) \pm \sqrt{\beta_2^2 \cos^2(ka_0) + 2\beta_1^2 (1 + \cos(ka_0))}\,.$$

(c) Is the polymer metallic or insulating?

Part 6. The Electronic Structure of Materials

7. Graphene and carbon nanotube transistors

(a) With reference to the bandstructure of graphene shown below, explain why graphene when rolled up into nanotubes can be either metallic or semiconducting?

(b) Using the *k*-space plot shown below, determine whether the following *(n,m)* nanotubes are metallic or semiconducting. Recall that nanotubes are rolled-up graphene sheets with wrapping vector $\overline{w} = na_1 + ma_2 = (n,m)$.

i) (0,6)

ii) (N,N)

iii) (3,9)

iv) (3,5)

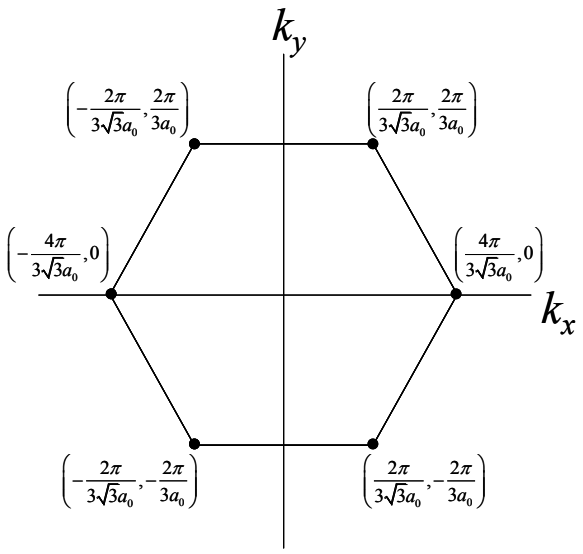


Fig. 6.45. The K points in graphene.

(c) At present, there is much interest in using graphene (as opposed to carbon nanotubes) as the channel material for field effect transistors. The idea is to fabricate entire chips on a single sheet of graphene.

First the graphene is deposited somehow (this is a technological challenge at present).
Next, the graphene is cut up.
Finally, contacts and gate insulators are deposited.

Why is the graphene cut up? Explain with reference to particle in a box models of conductors.

8. Carbon Nanotubes

(a) Prove the identities in Eq. (6.87) and Eq. (6.88).

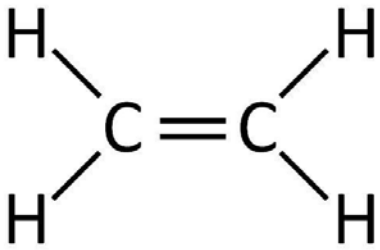
(b) Derive Eq. (6.90) from Eq. (6.89).

9. This question relates to the molecule shown below.

a) Write the Hamiltonian matrix for this molecule in terms of the tight binding parameters α, and β.

b) Write the energy for this molecular orbital in terms of α and β.

c) Compare the density of states of the HOMO and LUMO of the previous molecule to the one below.



Part 7. Fundamental Limits in Computation

This course has been concerned with the future of electronics, and especially digital electronics. At present digital electronics is dominated by a single architecture, Complementary Metal Oxide Semiconductor (CMOS), which is built on planar silicon field effect transistors. Steady improvements in the performance of CMOS circuits have been achieved by shrinking the feature sizes of the component transistors. This remarkable progress in electronics achieved over a period of > 30 years has come to underpin much of our economic life.

In this section, we address both practical and thermodynamic limits to silicon CMOS electronics. It is likely that these limits will dominate the future of the electronics industry.

Speed and power in CMOS circuits

As you should remember from 6.002, the archetype CMOS circuit is shown in Fig. 7.1. It is composed of two complementary FETs: the upper MOSFET is off for a high voltage input, and the lower MOSFET is off given a low input. The circuit is an inverter.

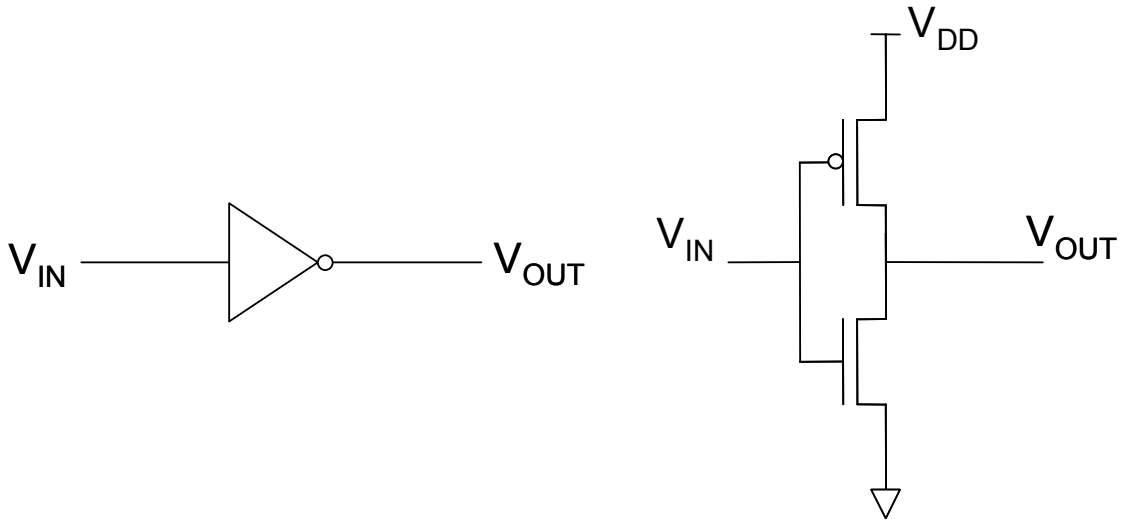


Fig. 7.1. A CMOS inverter consists of two complementary MOSFETs in series.

For constant voltage input, the circuit has two stable states, as shown in Fig. 7.2. Because one of the transistors is always off in steady state, the circuit ideally has no *static* power dissipation.

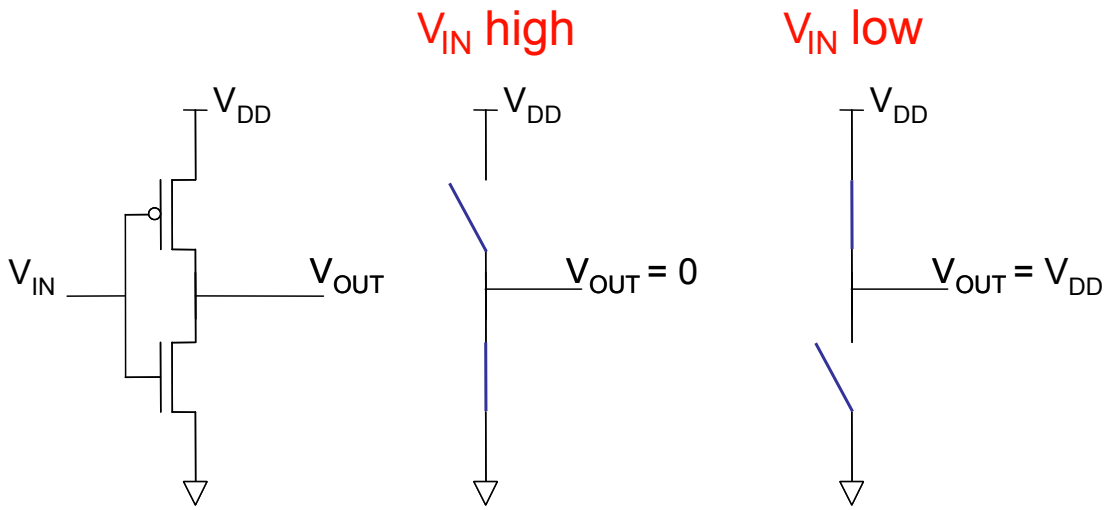


Fig. 7.2. The two steady state configurations of the inverter. No power is dissipated in either.

But when the input voltage switches the circuit briefly dissipates power. This is known as the *dynamic* power. We model the dynamics of a CMOS circuit as shown in Fig. 7.3. In this archetype CMOS circuit one inverter is used to drive more CMOS gates. To turn subsequent gates on an off the inverter must charge and discharge gate capacitors. Thus, we model the output load of the first inverter by a capacitor.

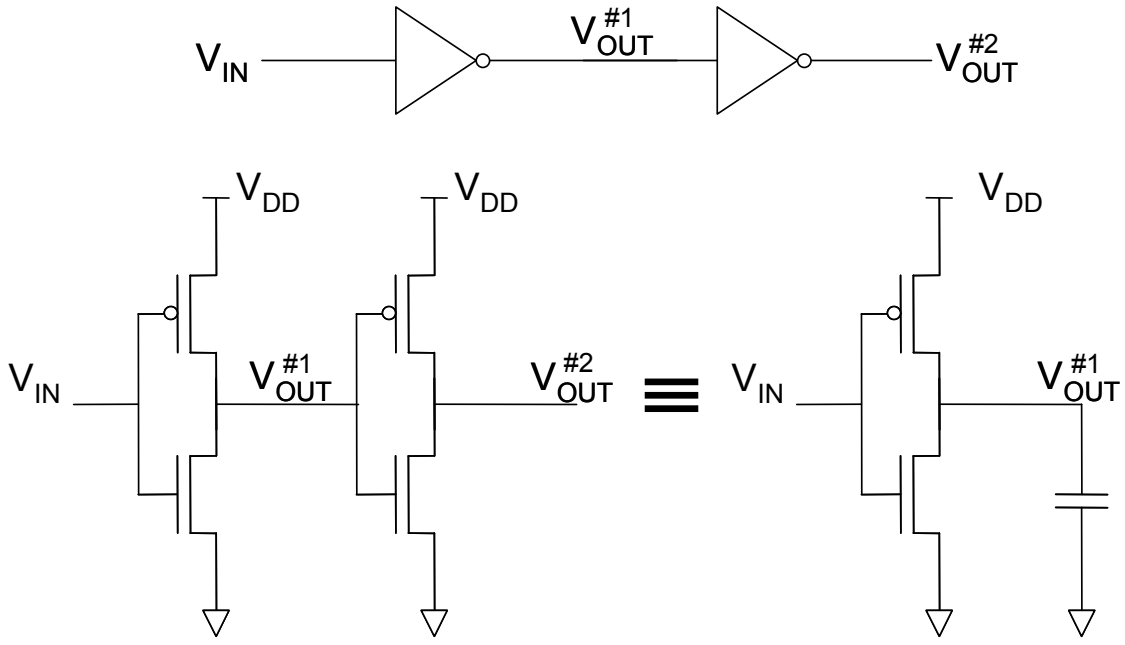


Fig. 7.3. Cascaded CMOS inverters. The first inverter drives the gate capacitors of the second inverter. To examine the switching dynamics of the first inverter, we model the second inverter by a capacitor.

Part 7. Fundamental Limits in Computation

We now consider the key performance characteristics of CMOS electronics.

The Power-Delay Product (PDP)

The power-delay product measures the energy dissipated in a CMOS circuit per switching operation. Since the energy per switching event is fixed, the PDP describes a fundamental tradeoff between speed and power dissipation – if we operate at high speeds, we will dissipate a lot of power.

Imagine an input transition from high to low to the inverter of Fig. 7.1.

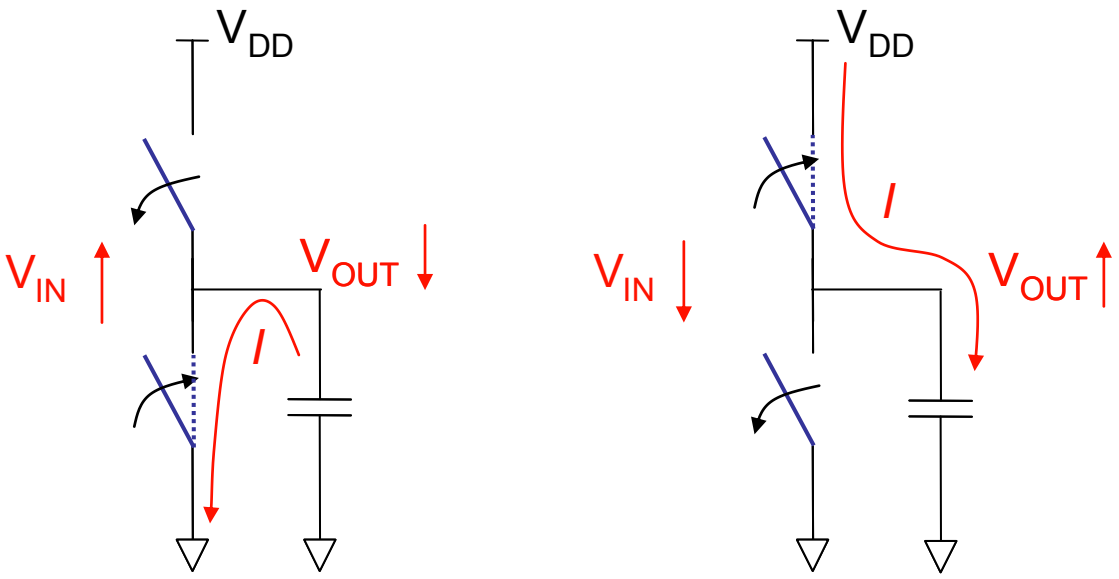


Fig. 7.4. Changes in the input voltage cause the output capacitor to charge or discharge dissipating power in the inverter.

If the output capacitor is initially uncharged, the energy dissipated in the PMOS FET is given by:

$$W = \int_0^{\tau/2} dt \left(V_{DD} - V_{OUT}\right) I \tag{7.1}$$

The current into the capacitor is given by:

$$I = C\frac{dV_{OUT}}{dt}, \tag{7.2}$$

Combining these expressions:

$$W = C\int_0^{\tau/2} dt \left(V_{DD} - V_{OUT}\right)\frac{dV_{OUT}}{dt} = C\int_0^{V_{DD}} dV_{OUT}\left(V_{DD} - V_{OUT}\right) = \frac{1}{2}CV_{DD}^{\,2}. \tag{7.3}$$

Similarly, in the second half of the cycle, when the capacitor is discharged through the NMOS FET, it is straightforward to show that again $W = \frac{1}{2}CV_{DD}^{\,2}$. Thus, the energy dissipated per cycle is:

$$PDP = CV_{DD}^{\,2}. \tag{7.4}$$

Switching Speed

The dynamic model of Fig. 7.4 relates the switching speed to the charging and discharging time of the gate capacitor.

$$f_{max} = \frac{I}{CV_{DD}} \tag{7.5}$$

Thus, switching speed can be improved by

(i) increasing the on current of the transistors
(ii) decreasing the gate capacitance by scaling to smaller sizes
(iii) decreasing the supply voltage (thereby decreasing the voltage swing during charge/discharge cycles

Scaling Limits in CMOS

Equation (7.4) demonstrates the importance of the gate capacitance. The capacitance is

$$C = \frac{\varepsilon A}{t_{ox}} \tag{7.6}$$

where A is the cross sectional area of the capacitor, t_{ox} is the thickness of the gate insulator and ε is its dielectric constant.

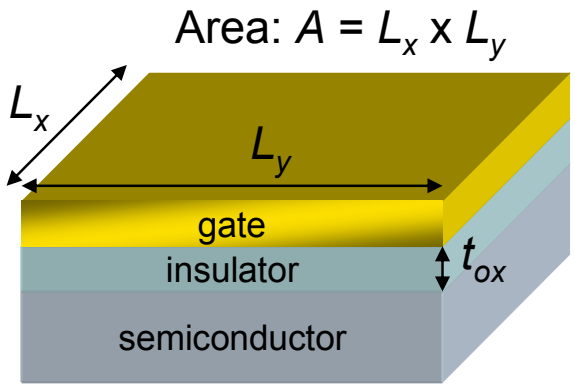


Fig. 7.5. The dimensions of a gate capacitor.

Now, if we scale all dimensions down by a factor s ($s < 1$), the capacitance decreases:

$$C(s) = \frac{\varepsilon s^2 A}{s t_{ox}} = sC_0 \tag{7.7}$$

From Eq. (7.4), reductions in C reduce the PDP, allowing circuits to run faster for a given power dissipation. Indeed, advances in the performance of electronics have come in large part through a continued effort of engineers to reduce the size of transistors, thereby reducing the capacitance and the PDP; see Fig. 7.6.

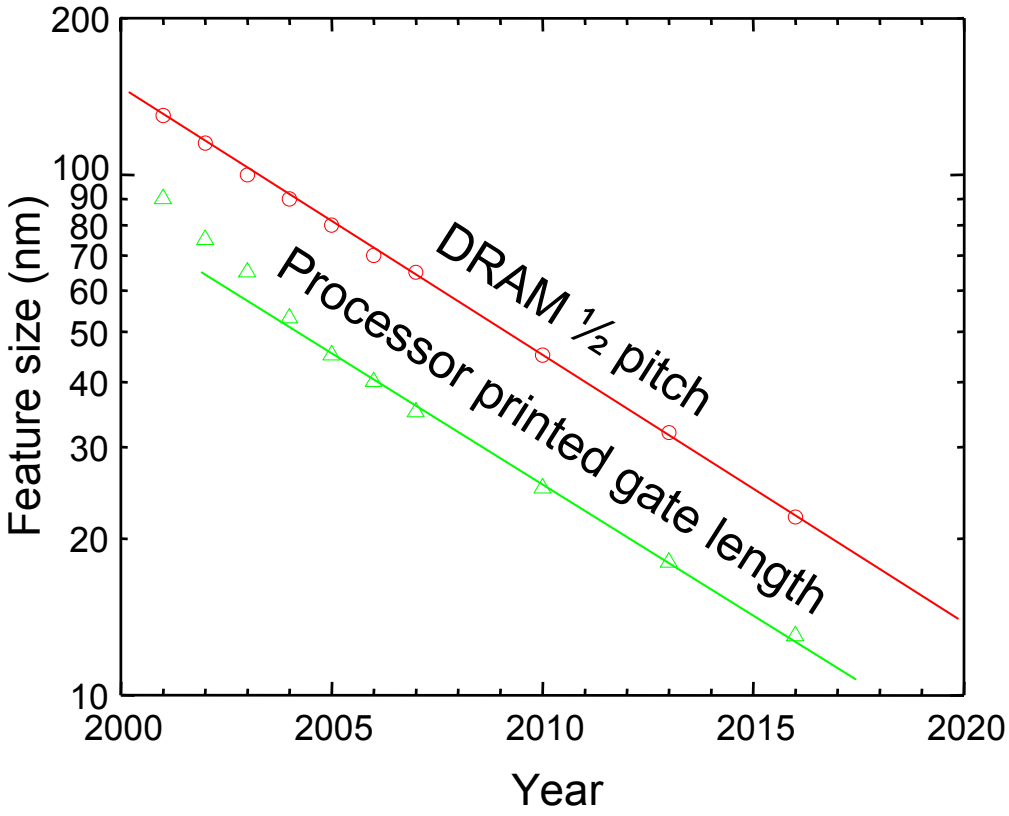


Fig. 7.6. The semiconductor roadmap predicts that feature sizes will approach 10 nm within 10 years. Data is taken from the 2002 International Technology Roadmap for Semiconductors update.

At present, however, there are increasing concerns that we are approaching the end of our ability to scale electronic components. There are at least two looming problems in electronics:

(i) *Poor electrostatic control*.

We saw in part 5 that gate control over charge in the channel requires $t_{ox} << L$, where L is the channel length. Now as the channel length, $L \rightarrow 10$ nm, $t_{ox} \rightarrow 1$ nm, i.e. the gate insulator is only several atoms thick! But the electric field across the gate must remain high to induce charge in the channel. Thus, reductions in feature sizes will eventually place severe demands on the gate insulator.

(ii) *Power density*

The electrostatic problem is fundamental, but it is possible that power concerns may obstruct the scaling of CMOS circuits prior to the onset of electrostatic issues. Power *density* is a particular concern since it does not benefit from continued reductions in component size. If the dimensions of a MOSFET are scaled down by a factor s ($s < 1$), $C \propto s$ (recall that capacitance is proportional to cross sectional area, and inversely

proportional to the spacing between the charges). But even if the PDP scales as s, the power *density* may increase because the number of devices per unit area increases as $1/s^2$.

The power densities of typical integrated circuits are approaching those of a light bulb filament (~ 100 W/cm^2). For comparison, the power density of the surface of the sun is ~ 6000 W/cm^2. Removal of the heat generated by an integrated circuit has become perhaps *the* crucial constraint on the performance of modern electronics. Indeed, the fundamental limit to power density appears to be approximately 1000 W/cm^2. In practice, using water cooling of a uniformly heated Si substrate with embedded micro channels, a power density of 790 W/cm^2 has been achieved with a substrate temperature near room temperature.

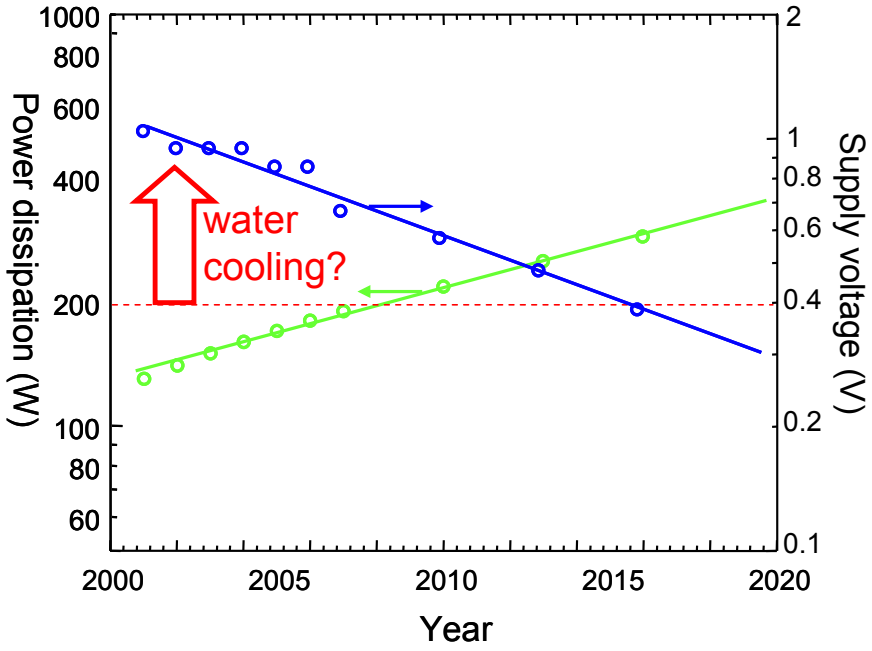


Fig. 7.7. The semiconductor roadmap predicts that supply voltages will drop to nearly 0.4V within 10 years. Power dissipation per chip is expected to increase to above 200W by 2008. It is expected that power dissipation in the shaded region will require significantly more expensive cooling systems. Data is taken from the 2002 International Technology Roadmap for Semi-conductors update.

As is evident from Eq. (7.4) above, the PDP also depends on the supply voltage V_{DD}. Ensuring that the total power dissipated per chip << 200 W has driven V_{DD} from 5V in early CMOS circuits to nearly 1V today. If the industry conforms to roadmap predictions, the supply voltage will eventually reach 0.4V by 2016.

But what is the ultimate limit to the PDP?

Part 7. Fundamental Limits in Computation

Brief notes on information theory and the thermodynamics of computation

We now examine the thermodynamics of computation.

(i) Minimum energy dissipated per bit

Assume we have a system, perhaps a computer, with a number of possible states. The uncertainty, or *entropy* of the computer is a measure of the number of states. Recall from thermodynamics that the Boltzmann-Gibbs entropy of a physical system is defined as

$$S = -k_B \sum_{i=1}^{N} p_i \ln p_i \,, \tag{7.8}$$

where the system has N possible states, each with probability p_i, and k_B is the Boltzmann constant.

The opposite of entropy and uncertainty is information. When the uncertainty of the system decreases, it gains information.

Now, the second law of thermodynamics can be restated as "all physical processes increase the total entropy of the universe". Let"s separate the universe into the computer, and everything else. The corresponding entropy of each system is given by

$$S_{universe} = S_{computer} + S_{everythingelse} \,. \tag{7.9}$$

Thus, thermodynamics requires

$$\Delta S_{universe} \geq 0 \,. \tag{7.10}$$

It follows that

$$\Delta S_{everythingelse} \geq -\Delta S_{computer} \,, \tag{7.11}$$

i.e. if the information within a computer increases during a computation, then the entropy decreases. This change in entropy within the computer must be at least balanced by an increase in the entropy of the remainder of the universe. The increase in entropy in the remainder of the universe is obtained by dissipating heat, ΔQ, from the computer.

According to thermodynamics the heat dissipated is

$$\Delta Q = T\Delta S_{everythingelse} \geq -T\Delta S_{computer} \tag{7.12}$$

Uncertainty and entropy can also be measured in bits. For example, how many bits are required to describe the computer with N states?

$$2^H = N \,. \tag{7.13}$$

Here, H is known as the **Shannon entropy**. If the states are equally probable, with probability $p = 1/N$, then the uncertainty reduces to:

$$H = \log_2 N = -\log_2 p \,. \tag{7.14}$$

Or more generally, if each state of the computer has probability p_i.

$$H = \left\langle -\log_2 p_i \right\rangle = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{7.15}$$

Comparing Eq. (7.8) with Eq. (7.15) and noting that $\ln p_i = \left(\ln 2\right) \log_2 p_i$ gives

$$\Delta Q = -k_B T \ln(2) \Delta H_{computer} \tag{7.16}$$

The heat must ultimately come from the power supply. Thus, the minimum energy required per generation of one bit of information is:

$$E_{min} = k_B T \ln(2) . \tag{7.17}$$

This minimum is known as the **Shannon-von Neumann-Landauer (SNL) limit**.

(ii) Energy required for signal transmission

Recall Shannon‟s theorem for the capacity, *c*, in bits per second, of a channel in the presence of noise.

$$c = b \log_2 \left(1 + \frac{s}{n} \right), \tag{7.18}$$

where *s* and *n* are the signal and noise power, respectively, and *b* is the bandwidth of the channel. The noise in the channel is at least $n = bk_B T$.

The energy required per bit transmitted is:

$$E_{min} = \lim_{s \to 0} \left\{ \frac{s}{c} \right\} = \lim_{s \to 0} \left\{ \frac{s}{b \log_2 (1 + s/n)} \right\} . \tag{7.19}$$

L‟Hôpital‟s rule gives

$$E_{min} = k_B T \ln(2) . \tag{7.20}$$

consistent with the previous calculation of E_{min}.

(iii) Consequences of E_{min}

It has been argued that since the uncertainty in energy, ΔE, within an individual logic element can be no greater than E_{min}, we can apply the Heisenberg uncertainty relations to a system operating at the SNL limit to determine the minimum switching time, *i.e.*[†]

$$\Delta E \Delta t \geq \hbar \tag{7.21}$$

Eq. (7.21) gives a minimum switching time of

$$\tau_{min} = \frac{\hbar}{\Delta E} = \frac{\hbar}{k_B T \ln(2)} = 0.04\,\text{ps} \tag{7.22}$$

Assuming that the maximum power density that we can cool is $P_{max} \sim 100\text{W/cm}^2$, the maximum integration density is

$$n_{max} = \frac{P_{max}}{E_{min} / \tau_{min}} = \frac{\hbar P_{max}}{{E_{min}}^2} \tag{7.23}$$

At room temperature, we get $n_{max} \sim< 10^{10}\,\text{cm}^{-2}$, equivalent to a switch size of 100 x 100 nm. This is very close to the roadmap value for 2016.

At lower temperatures, the power dissipation *on chip* is decreased, but the overall power dissipation actually increases due to the requirement for refrigeration.[4] Since the

[†] This argument, due to Zhirnov, *et al.* "Limits to Binary Logic Switch Scaling - A Gedanken Model", *Proceedings of the IEEE* **91**, 1934 (2003), has been used to argue that end of the roadmap Si CMOS is as good as charge based computing can get.

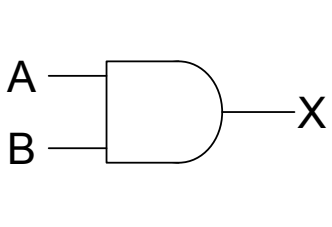
engineering constraint is likely to be on chip power dissipation – refrigeration may be one method for further increasing the density of electronic components.

Reversible computers

In the previous section, we defined computation as a process that increases information and decreases uncertainty. But if uncertainty (i.e. entropy) decreases within the computer, entropy must increase outside the computer. This is an application of the second law of thermodynamics, which states that all physical systems can only increase entropy over time.

Of all physical laws, the second law of thermodynamics is famous for defining the „arrow of time". The implication of the second law is that *computation is irreversible*, at least if the computation changes uncertainty.

For example, let's consider a two input AND gate. If one of the inputs to the AND gate is a zero, then the information in the other input is thrown away. Thus, the total number of states decreases when the inputs propagate to the output of an AND gate. Consequently, entropy decreases, heat is dissipated and AND gates are not reversible.



A	B	X
0	0	0
1	0	0
0	1	0
1	1	1

Fig. 7.8. AND gates are not reversible. If the output is zero, the inputs cannot be reconstructed.

The heat dissipated in the AND gate is calculated as follows. There are four possible input states. Assuming each is equi-probable the Shannon entropy is

$$H_{in} = -\log_2 1/4 = 2 \text{ bits} \tag{7.24}$$

There are two possible output states. The probability of the output X = 0 is ¾ and the probability of X = 1 is ¼.

$$H_{out} = -\frac{3}{4}\log_2 3/4 - \frac{1}{4}\log_2 1/4 \approx 0.811 \text{ bits} \tag{7.25}$$

Thus,

$$\Delta E = -k_B T \ln(2) \Delta H \approx 3.4 \times 10^{-21} \text{J} \tag{7.26}$$

But what if we designed a gate that did not throw away states during the computation? Such a system would be reversible, and more importantly it would not need to dissipate energy.

In fact, several reversible logic elements have been proposed. Perhaps the best known irreversible computer is the billiard ball computer pioneered by Fredkin.

An example of a billiard ball logic gate is shown in Fig. 7.9. Billiard balls are fired into the logic gate from positions A and B. If there is a collision, the balls are deflected to positions W and Z. If one ball is absent, however, an output at either X or Y is generated. We also need to assume that the balls obey the laws of classical mechanics; there is no friction and the collisions are perfectly elastic. Note that the number of states in a billiard ball logic elements does not change – the billiard balls are neither created nor destroyed.

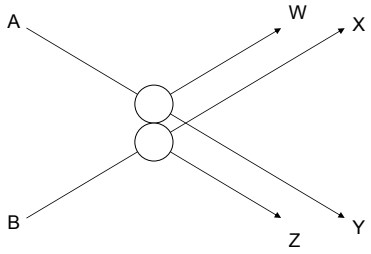


Fig. 7.9. A two ball collision gate. After Feynman, *Lectures on Computation.* Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

More complex devices are possible by adding „redirection gates" (walls). For example, Fig. 7.10 shows a switch made from collision and redirection gates.

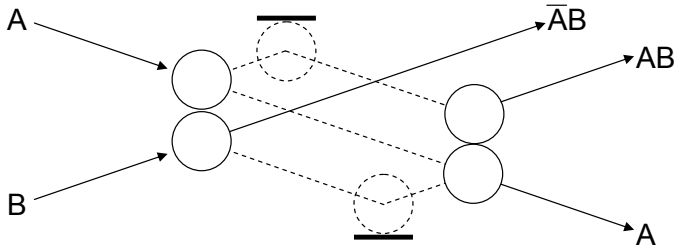
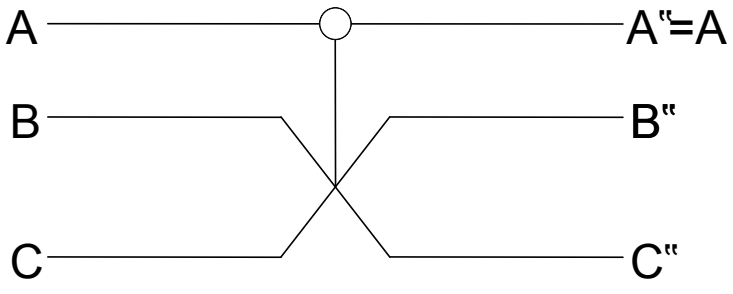


Fig. 7.10. A billiard ball switch. After Feynman, *Lectures on Computation.* Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

But given that many logic gates such as the AND gate are inherently non-reversible, the question arises: Can an arbitrary algorithm be implemented entirely from reversible elements? The answer is yes. Reversible computers can be constructed entirely of a fundamental reversible element known as the Fredkin gate, shown in Fig. 7.11.



A	B	C	A"	B"	C"
0	0	0	0	0	0
0	0	1	0	1	0
0	1	0	0	0	1
0	1	1	0	1	1
1	0	0	1	0	0
1	0	1	1	0	1
1	1	0	1	1	0
1	1	1	1	1	1

Fig. 7.11. The symbol for the Fredkin gate. A is unchanged. If A = 0 then B and C switch. If A = 1 then B and C remain unchanged. All logic elements may be formulated from reversible Fredkin gates. After Feynman, *Lectures on Computation.* Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

An implementation of a Fredkin gate with billiard balls is shown in Fig. 7.12.

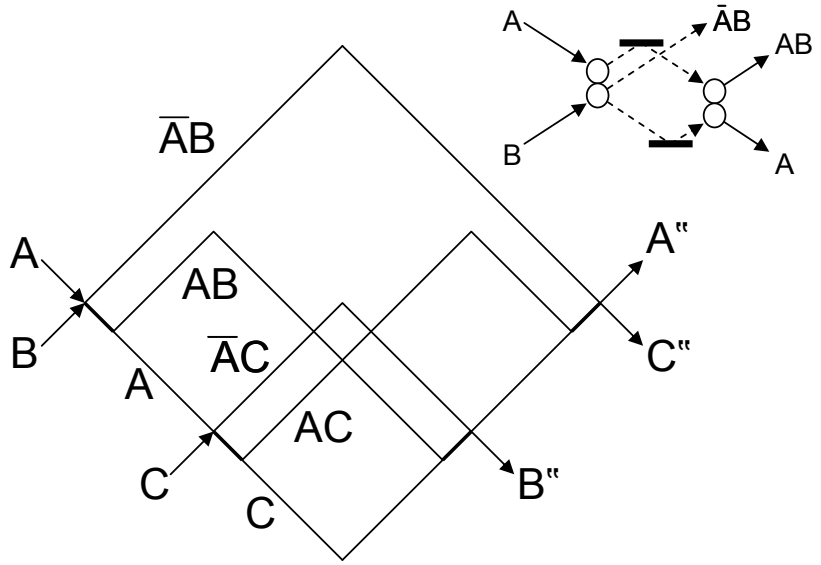


Fig. 7.12. A Fredkin gate constructed from four billiard ball switches. After Feynman, *Lectures on Computation.* Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

Reversible computers and noise

Reversible computers, however, remain extremely controversial in engineering circles. The catch is noise. Shannon‟s theorem, for example, requires $E_{min} = k_B T \ln(2)$ for the transmission of one bit of information in a noisy channel. This applies even in a reversible system such as the billiard ball collision gate. In fact, billiard ball gates are extremely sensitive to errors. Given a slight error in the trajectory or timing of one ball and a billiard ball computer would accrue a large number of errors.

A billiard ball computer could be made more robust and noise resistant by including trenches to guide the balls. But the trench guides the balls by dissipating that component of the ball‟s momentum that would otherwise drive it off its designed trajectory. Thus, the trenches inevitably lead to energy dissipation.

In contrast, let‟s briefly look at noise in CMOS circuits. The transfer function of a CMOS inverter is shown in Fig. 7.13. We see that close to the switching voltage, the inverter has very large gain, A_V:

$$A_V = \frac{dV_{out}}{dV_{in}} \gg 1 \tag{7.27}$$

The gain protects the inverter against noise. For example, consider two cascaded inverters. Assume some noise is added to the output of the first inverter. The *noise margin* tells us the minimum amount of noise required to cause an error at the output of the second inverter; see Fig. 7.14.

Thus, many device engineers argue that without gain no computation system is practical. And since reversible computers do not dissipate power it is not clear how they can amplify a signal, rendering them always subject to the adverse effects of noise.

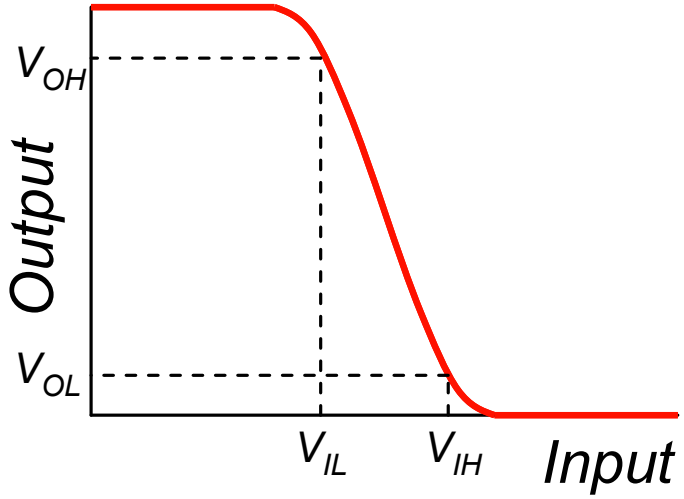


Fig. 7.13. Transfer characteristics of a CMOS inverter. V_{IL} and V_{IH} are defined as the threshold of low and high inputs, respectively. Note that the large gain means that $V_{OL} < V_{IL}$ and $V_{OH} > V_{IH}$, helping protect signal integrity against the effects of noise.

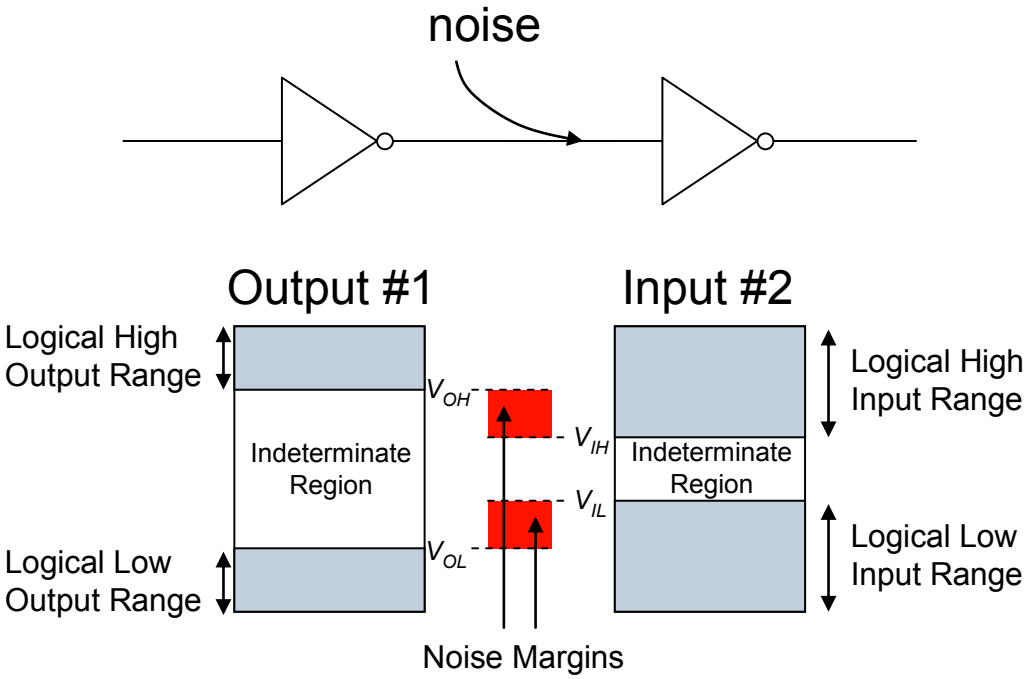


Fig. 7.14. The noise margin in a digital circuit is the minimum input noise voltage required to cause an error at the output of the next gate. The greater the gain, the greater the noise margin.

Part 7. Fundamental Limits in Computation

The future of electronics?

The immediate path is clear: we have not yet reached the limits of scaling, or the fundamental limits of field effect transistors. The electronics industry will push to smaller length scales to minimize the power delay product. It will also seek to exploit ballistic conduction in low dimensional materials, thereby increasing switching speeds.

It is realistic to expect that a future MOSFET might possess:

(i) ballistic transport and operation at the quantum limit of conductance
(ii) switching on and off at the optimum FET subthreshold slope of kT/q
(iii) scaling of all dimensions with a gate insulator thickness of ~ 1 nanometer

Traditionally, substantial materials development efforts have been devoted to improving the mobility of transistor channels. But because devices are already at the ballistic limit, the electrostatic design of nanotransistors will be a likely focus of materials development. We have seen that good electrostatic control of the channel can be achieved by maximizing the gate capacitance. For example, with a nanowire channel, the gate could be implemented as a concentric ring. Or a channel that consists of a single atomic layer (such as a grapheme sheet) might be preferable from the electrostatic viewpoint to a thicker layer of silicon, even though both will operate at the ballistic limit. Manufacturing such advanced structures may require a substantial amount of further development.

Beyond this, there appears to be only one major weakness of conventional FET technologies. There is a strong possibility that new technologies will demonstrate subthreshold slope far superior to kT/q. As we have seen, this will allow for dramatic reductions in operating voltage, and hence significantly lower power dissipation.

From a fundamental viewpoint, all transistors that operate in thermodynamic equilibrium, must exhibit an energy difference between their ON and OFF states. For example, the potential energy difference between the ON and OFF states of a FET is $\Delta E = \frac{1}{2}CV^2$, which can also be expressed as $\Delta E = \frac{1}{2}QV$, where Q is the total charge on the gate capacitor and V is the supply voltage. The fundamental limit in the OFF state current is the probability of thermal excitation from the OFF state to the ON state. That is:

$$I_{OFF} = I_{ON} \exp\left[-\tfrac{1}{2}QV/kT\right] \tag{7.28}$$

where I_{ON} is the maximum current associated with the ON state. But as we have seen, modern FETs do not operate at this limit because each electron in the channel is independent. In contrast to Eq. (7.28), the FET follows:

$$I_{OFF} = I_{ON} \exp\left[-qV/kT\right] \tag{7.29}$$

Except for a FET that operates with a single electron in the channel, the difference is substantial: a subthreshold slope of kT/Q versus kT/q. Indeed, at present transistor dimensions $Q >> 10^3\, q$.

So how can we approach the subthreshold limit?

It is thought that if all the charge in the channel behaves *collectively*, (*i.e.* all or none of the charge contributes to current) then it might be possible to switch closer to the limit. Perhaps the best examples of this principle are the voltage-dependent ion channels of biology, in which conformation changes may enable subthreshold slopes as sharp ≈ 10 mV/decade.[†]

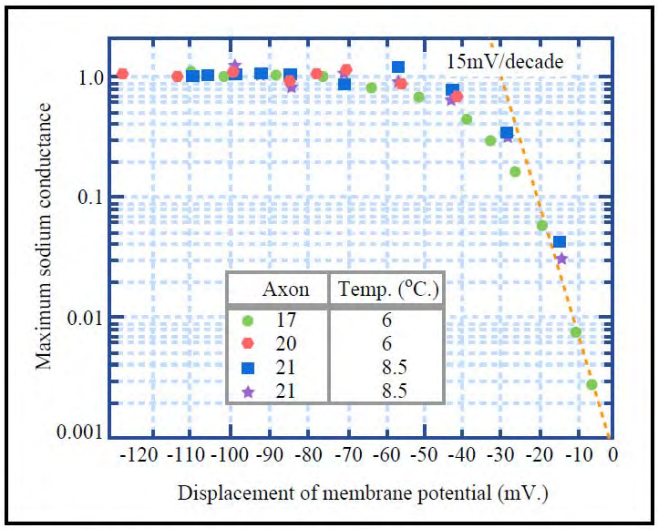


Figure by MIT OpenCourseWare.

Hodgkin and Huxley, J. Physiol. **116**, 449 (1952a)

Fig. 7.15. The switching characteristics of voltage gated sodium ion channels from the giant squid axon. Note the extremely sharp switching characteristics. Reproduced from Hodgkin and Huxley's classic 1952 series of papers.

Below, we show the structure and mechanism of the mechanical change in a voltage dependent K+ ion channel, as determined by MacKinnon, *et al*.[§] The channels sit in a membrane; when open they allow the diffusion of ions from one side of the membrane to the other.

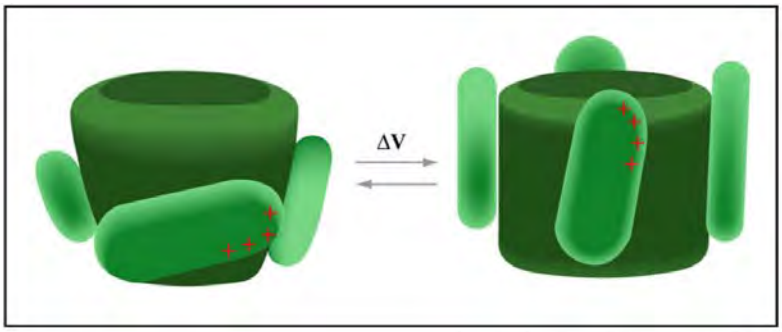


Figure by MIT OpenCourseWare.

Fig. 7.16. The voltage dependent K^+ ion channel has 4 charged paddles that rotate in an electric field, opening and closing a mechanical gate at the base of the channel. Reproduced from MacKinnon, *et al*.

[†] Hodgkin and Huxley, J. Physiol. **116**, 449 (1952a)

[§] Y. Jiang, A. Lee, J. Chen, V. Ruta, M. Cadene, B.T. Chait and R. MacKinnon. Nature. **423**. 33-41 (2003)

Consider a membrane where there are N closed channels and N^* open channels. The ratio of open to closed channels is determined by the Boltzmann relation:

$$\frac{N^*}{N} = \exp\left[-\frac{U_{open} - U_{closed}}{kT}\right] \tag{7.30}$$

where U_{open} and U_{closed} are the energies of the open and closed conformations respectively. Under an electric field, we assume that Z charges move through a potential of ΔV, *i.e.*:

$$U_{open} = U_{closed} - Zq\Delta V\ . \tag{7.31}$$

The current through the ion channel is proportional to the number of open channels, N^*.

$$I \propto N^* \tag{7.32}$$

Since $N + N^*$ is a constant

$$I \propto \frac{N^*}{N+N^*} \approx \frac{N^*}{N} = \exp\left[\frac{Zq\Delta V}{kT}\right] \tag{7.33}$$

That is, the subthreshold slope is sharpened by a factor, Z, the effective[†] number of charges on the movable paddles.

$$\frac{\Delta V}{\log_{10} I} = \frac{kT}{Ze}\frac{1}{\log_{10} \mathrm{e}} \approx \frac{60}{Z}\ \text{mV/decade}\ . \tag{7.34}$$

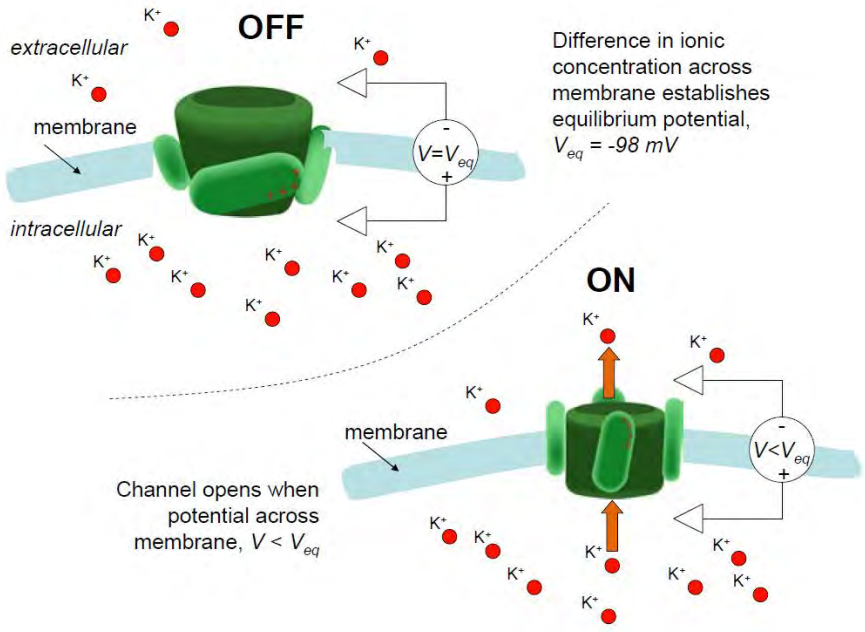


Fig. 7.17. Ion channels modulate the diffusion of ions through a membrane. The direction of ion current is determined by the concentration gradient. Typically, the ion channel preferentially passes ions of a particular size and charge. When it is open, the channel illustrated above selectively allows K+ ions to diffuse.

[†] Note that the effective number of charges is usually less than the actual number of charges on the movable structures in the ion channel because the charges are not usually free to move through the full potential ΔV across the membrane (the motion of the paddles is somewhat restricted).

The conclusion is that transistors are possible with subthreshold characteristics superior to those of conventional FETs. The ion channel shows that mechanically-coupling the charges together is one path to achieving the *collective* behavior that we desire. But the reliability of mechanical devices is questionable. Instead, it is possible that another collective phenomenon, like the switching of a magnetic domain in a ferromagnet, may be exploited to improve switching.

And beyond?

Researchers are currently pursuing a few ideas:

1. Reversible computing
The absence of power dissipation makes this a big prize, but concerns remain as to its noise immunity and fundamental practicality.

2. New information tokens
Transistors today use electrons to carry information. Instead, we might seek to use a different information token such as the spin of an electron or position in a mechanical switch. A change in information token could revolutionize electronics. But at present it is not clear what, for example, a spin-in spin-out transistor might look like, nor do we have a clear idea of the potential benefits of spin-based technology. For example, could it escape the Shannon-Von Neumann- Landauer limit?

3. Integration
More transistors per chip have traditionally meant more computing power. If we can't make transistors any smaller, perhaps we could shift to three dimensional circuits? A transition from two to three dimensional circuits could massively increase integration densities. But apart from the difficulty of fabricating such structures, we must also figure out how to cool them.

4. Architecture
The computing power of the brain clearly demonstrates the virtue of different approaches to certain problems such as pattern recognition. But it is not clear that our current model of electronics is suited to say, a shift to a neural network type architecture.

Whatever happens the stakes are high. As we approach the limits of CMOS, slow technological progress may reduce the need to update computers every few years. But the economic model of the electronics industry has come to rely on rapid technological change. Consequently, the rewards may be especially great for the next revolution in electronics technology.

Part 7. Fundamental Limits in Computation

Problems

Q1. Adiabatic Transistors

Consider the inverter shown below.

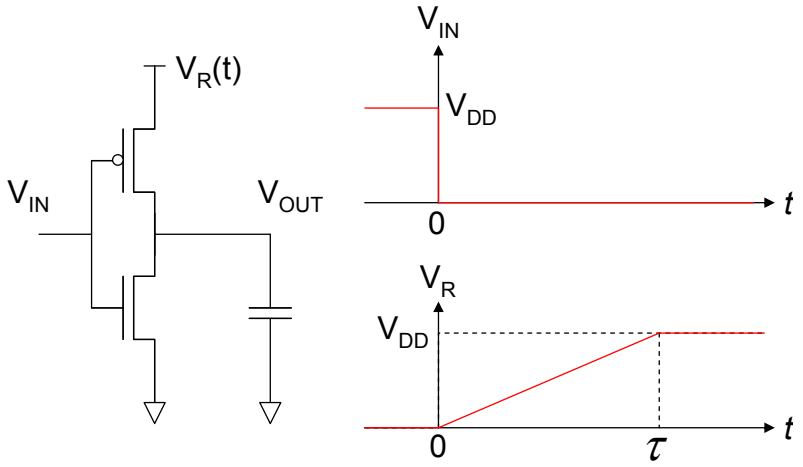


Fig. 7.18. An adiabatically-driven inverter.

Unlike in a conventional CMOS inverter, in this device, the supply voltage, V_R, adjusts during the switching operation. Initially the voltage on the output capacitor is zero, but at $t = 0$ the input voltage drops to zero. Also at $t = 0$, the supply voltage ramps from zero to the logic high voltage, V_{DD}.

Assume that the PMOS FET is modeled by a resistor, R.

(a) Show that the energy dissipated during the switching operation is

$$E = \frac{RC}{\tau} CV_{DD}^2 \text{ for } \tau >> RC.$$

This is known as an adiabatic switch, since switching occurs (in the limit) with no energy dissipation, i.e. we are adding charge to a capacitor using a vanishingly small excess voltage.
[Hint: You may assume V_{OUT} of the form $V_{OUT} = a + b\exp[-t/RC] + ct$ where a, b, and c are constants to be determined.]

(b) Show also that the energy dissipated reduces to the standard CMOS switching energy

$$E = \frac{CV_{DD}^2}{2} \text{ for } \tau << RC.$$

(c) The above example shows adiabatic switching when the capacitor voltage changes from low to high. Can it be implemented generally? i.e. consider the case when the capacitor voltage changes from high to low. And what happens when the capacitor does not change voltage during a cycle?

Q2. Cellular Automata

This question refers to a proposed architecture for molecular electronics: *Molecular Quantum-Dot Cellular Automata*. The figures are drawn from the reference.

In this architecture information is stored in bistable cells. An example cell is shown below:

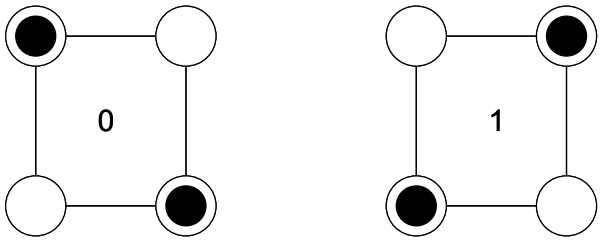


Fig. 7.19. A bistable cell for use in a cellular automata computer.

This cell consists of four electron traps positioned at the corners of a square. Only two of the traps are filled. From electrostatics, there are two stable states with the electrons at opposing corners of the square.

To transmit information, the cells are placed in a line. Information then propagates electrostatically, without current flow. It is argued that power dissipation is therefore eliminated and no interconnecting wires are required.

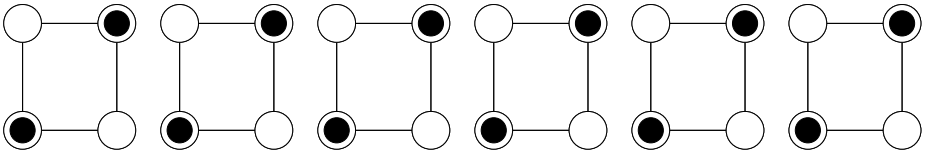


Fig. 7.20. A cellular automata wire.

By changing the topology, it is possible to make logic gates. For example, below we show an inverter.

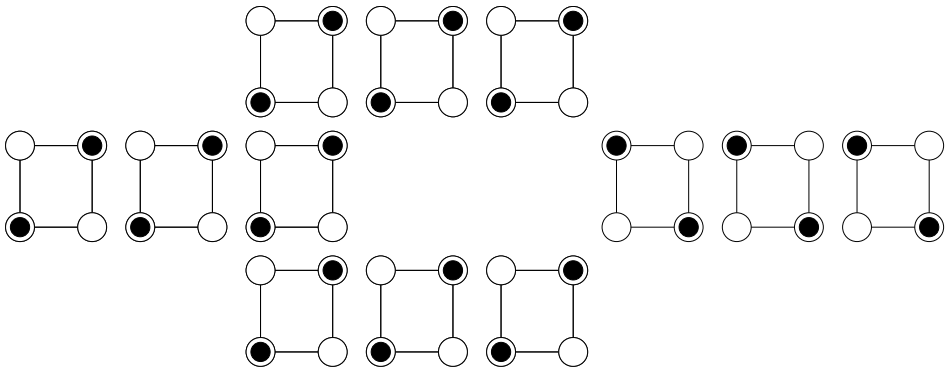


Fig. 7.21. A cellular automata inverter.

Question continued on next page….

(a) A proposed „majority gate" is shown below. The output Z is the majority of the inputs, A, B and C. *i.e.* if there are more 1 inputs than zero inputs then Z = 1, otherwise Z = 0. Use this gate to design a two input AND gate.

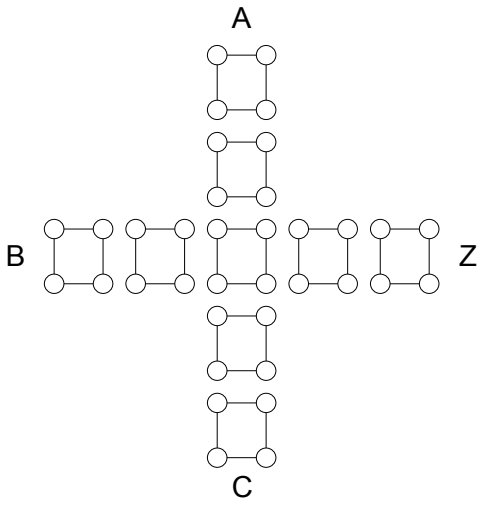


Fig. 7.22. A proposed majority gate.

(b) Is the majority gate truly dissipationless? Hint: calculate the entropy before and after a majority decision.

Reference: Lent, "Bypassing the transistor paradigm" Science **288** 1597 (2004)

Q3. Power delay products at the nanoscale

The power delay product is the minimum energy dissipated per bit of information processed. For a CMOS inverter the PDP is:

$$PDP = CV^2$$

where V is the supply voltage and C is the load capacitance as seen by the inverter. In this question, we will assume that the supply voltage is fixed.

(a) Determine the load capacitance as a function of the gate and quantum capacitances. Assume we can neglect all other capacitances.

(b) Consider a 2d field effect transistor (where $C_Q \to \infty$). If its dimensions are scaled by a factor s, how does the PDP scale?

(c) Now consider a quantum *wire* field effect transistor with $C_Q << C_G$. Its gate capacitance is given by

$$C_G = 2\pi\varepsilon \frac{l}{\log\left(r/a_0\right)}$$

where ε is the dielectic constant of the gate insulator, l is the gate length, r is the gate radius and a_0 is the 1d wire radius.

Assume that *l* and *r* are scaled by a factor *s*, how does the gate capacitance for a quantum *wire* field effect transistor scale?

(d) Now consider the impact of the quantum capacitance on the PDP on the quantum *wire* field effect transistor. How does the overall PDP scale? Is the scaling faster or slower than the equivalent PDP using large quantum *well* field effect transistors?

Q4. Mechanical transistors

Consider a mechanical switch.

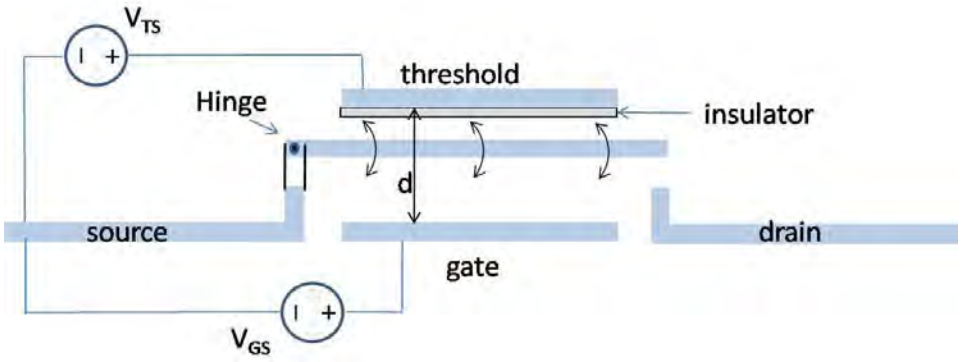


Fig. 7.23. A mechanical switch.

The conductor is pulled towards the gate electrode when $|V_{GS}| > |V_{TS}|$, switching the device on, and towards the threshold electrode when $|V_{GS}| < |V_{TS}|$ switching the device off. Assume two switches are wired together in a complementary logic circuit that drives a capacitive load as shown below.

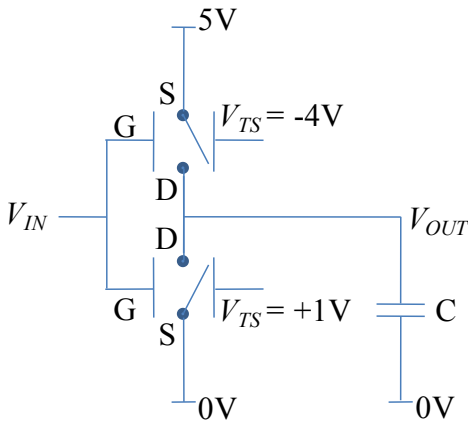


Fig. 7.24. A complementary logic circuit featuring mechanical switches.

(i) Plot steady state V_{OUT} versus V_{IN}, where V_{IN} ranges from 0 to 5V. Show that the circuit is complementary.

(ii) Assume V_{IN} is switched from $0V$ to $5V$ and then back to $0V$. How much energy is dissipated?

(iii) Consider one of the switches. Let C_T^{ON} and C_G^{ON} be the threshold-conductor capacitance, and the gate-conductor capacitance, respectively, in the ON state, and let C_T^{OFF} and C_G^{OFF} be the capacitances, respectively, in the OFF state. See the figure below.

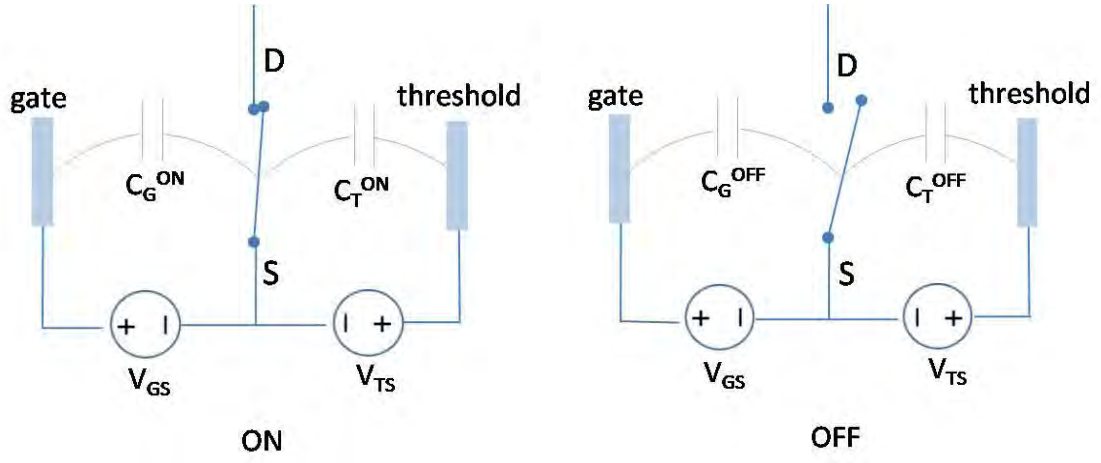


Fig. 7.25. Capacitive models of the switch in the ON and OFF configuration.

What is the energy stored in these capacitors in the (a) ON and in the (b) OFF positions as a function of V_{GS} and V_{TS}?

Now connect N switches all wired in parallel.

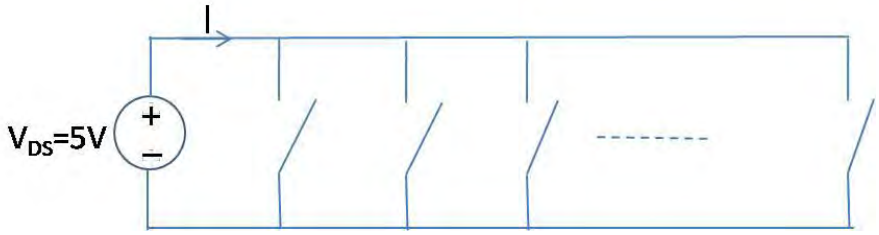


Fig. 7.26. *N* switches all wired in parallel.

Each switch has $V_{TS} = +1$V and resistance, $R = 100\Omega$. Assume all the gate electrodes are wired together at a potential V_{GS}. To simplify the analysis assume that $C_G^{ON} >> C_T^{ON}$ and also that $C_T^{OFF} >> C_G^{OFF}$. Furthermore, take $C_G^{ON} = C_T^{OFF} = C$.

(iv) Considering Boltzmann statistics, and the potential energy difference between the OFF and ON states, out of the N switches, what is the probable number of switches that are ON as a function of C, V_{GS} and V_{TS} when $|V_{GS}| < |V_{TS}|$?

Continued…

(v) Find I for the N switches as a function of V_{GS} and V_{TS} for $0 < V_{GS} < 5\text{V}$ (for $V_{TS} = 1\text{V}$).

(vi) Does the mechanical switch exhibit any benefit over conventional CMOS?

Q5. (a) Consider two identical balls each 1cm in diameter and of mass $m = 1\text{g}$. One is kept fixed, and the second is dropped directly on it from a height of $d = 10\text{cm}$. From the uncertainty principle alone, what is the expected number of times the moving ball bounces on the stationary ball before it misses the latter ball altogether? Assume the ball is dropped from an optimal initial state.

Hint: some parts of this problem can be solved classically.

(b) Discuss the implications of **(i)** for billiard ball computers.

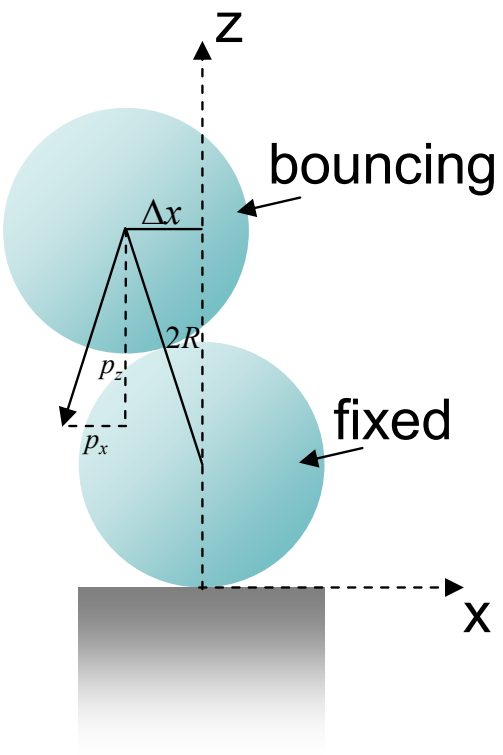


Fig. 7.27. An off-center collision between the fixed ball and the bouncing ball.

Q6. The following question refers to ion channel mechanical switches at $T = 300\text{K}$.

a) Assume that any given ion channel is either open with conductance $G = G_0$, or closed with conductance $G = 0$. Using Boltzmann statistics, write an expression for the conductance of a giant squid axon (with N ion channels in parallel) as a function of the applied membrane potential, V. Assume that the number of open channels at $V = 0$ is N_0.

Hint: Given Boltzmann statistics, the relative populations N_1 and N_2 of two states separated by energy dU are $N_1/N_2 = \exp(-dU/kT)$.

b) Where possible given the data in Fig. 7.15, evaluate your parameters.

c) Sketch a representative IV of a single ion channel.

Part 8. References

The following references were used to prepare aspects of these notes. They are recommended for those who wish to explore these topics in greater depth.

Quantum Transport

S. Datta, *Electronic Transport in Mesocopic Systems*.
Cambridge University Press 1995

S. Datta, *Quantum Transport: Atom to Transistor*.
Cambridge University Press 2005

Quantum Mechanics

P.L. Hagelstein, S.D. Senturia, and T.P. Orlando, *Introductory Applied Quantum and Statistical Mechanics*. Wiley 2004

P.W. Atkins and R.S. Friedman, *Molecular Quantum Mechanics.*
Oxford University Press, 3rd edition 1997

Ballistic Transistors

M. Lundstrom and J. Guo, *Nanoscale Transistors: Physics, Modeling, and Simulation*, Springer 2006.

Conventional MOS transistors

Y. Tsividis, *The MOS Transistor*.
Oxford University Press, 2nd edition 1999

Theory of computation

R. Feynman, *Lectures on Computation.*
Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996

Appendix 1. Electron Wavepacket Propagation

Stationary states and eigenfunctions

Until now, we have not considered the velocity of electrons because we have not considered the time dependence of solutions to the Schrödinger equation. In Part 1, we broke the full Schrödinger Equation into two coupled equations: an equation in time, and another in space. The separation is possible when the potential energy is constant in time. Then the spatial and time dependencies of the solution can be separated, i.e.

$$\Psi(x,t)=\psi(x)\zeta(t) \tag{A.1.1}$$

The time dependence is described by:

$$E\zeta(t)=i\hbar\frac{d}{dt}\zeta(t) \tag{A.1.2}$$

and the spatial dependence is given by

$$E\psi(x)=-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\psi(x)+V(x)\psi(x). \tag{A.1.3}$$

Solutions to these coupled equations are characterized by a time-independent probability density. The general solution to Eq. (A.1.2) is

$$\zeta(t)=\zeta(0)\exp\left[-i\frac{E}{\hbar}t\right] \tag{A.1.4}$$

and the probability density is:

$$\left|\Psi(x,t)\right|^2=\left|\psi(x)\zeta(t)\right|^2=\left|\psi(x)\right|^2\left|\zeta(0)\right|^2 \tag{A.1.5}$$

Because the solution does not evolve with time, it is said to be „stationary". These solutions are important and are known as „eigenfunctions". Eigenfunctions are extremely important in quantum mechanics. You can think of them as the natural functions for a particular system. Each eigenfunction is associated with a constant, known as the eigenvalue: in this case the constant energy, *E*.

An arbitrary wavefunction, however, will not necessarily be an eigenfunction or stationary. For example, consider a wavefunction constructed from two eigenfunctions:

$$\left|\psi(x)\right\rangle=a_1\left|\psi_1(x)\right\rangle+a_2\left|\psi_2(x)\right\rangle \tag{A.1.6}$$

where *a* and *b* are constants. This is known as a linear combination of eigenfunctions.
The full solution is

$$\Psi(x,t)=a_1\psi_1(x)e^{-i\frac{E_1}{\hbar}t}+a_2\psi_2(x)e^{-i\frac{E_2}{\hbar}t} \tag{A.1.7}$$

Substituting into the Schrödinger Equation gives

$$H\left|a_1\psi_1(x)e^{-i\frac{E_1}{\hbar}t}+a_2\psi_2(x)e^{-i\frac{E_2}{\hbar}t}\right\rangle=a_1E_1\left|\psi_1(x)e^{-i\frac{E_1}{\hbar}t}\right\rangle+a_2E_2\left|\psi_2(x)e^{-i\frac{E_2}{\hbar}t}\right\rangle \tag{A.1.8}$$

i.e., the linear combination is not necessarily itself an eigenfunction. It is not stationary: the phase of each eigenfunction component evolves at a different rate. The probability

density shows time dependent interference between each rotating phase term. For example, assuming that a_1, a_2, ψ_1 and ψ_2 are real:

$$\left|\Psi(x,t)\right|^2 = a_1^2\psi_1^{\,2}(x) + a_2^2\psi_2^{\,2}(x) + 2a_1a_2\psi_1(x)\psi_2(x)\cos\left(\left(E_2 - E_1\right)\frac{t}{\hbar}\right). \quad \text{(A.1.9)}$$

Completeness

We will not prove completeness in the class. Instead we merely state that the completeness property of eigenfunctions allows us to express any well-behaved function in terms of a linear combination of eigenfunctions. i.e. if $|\phi_n\rangle$ is an eigenfunction, then an arbitrary well-behaved wavefunction can be written

$$|\psi\rangle = \sum_n a_n |\phi_n\rangle \quad \text{(A.1.10)}$$

Completeness also requires that the potential be finite within the region of interest. For example, no combination of eigenfunctions of the infinite square well can ever describe a non-zero wavefunction amplitude at the walls.

Re-expressing the wavefunction in terms of a weighted sum of eigenfunctions is a little like doing a Fourier transform, except that instead of re-expressing the wavefunction in terms of a linear combination of exp[i*kx*] factors, we are using the eigenfunctions.[†]

The problem now is the determination of the weighting constants, a_n.

For this we need the next property of eigenfunctions:

Orthogonality

Eigenfunctions with different eigenvalues are orthogonal. *i.e.* the bracket of eigenfunctions corresponding to different eigenvalues is zero:

$$\langle\phi_j|\phi_i\rangle = 0, \quad for \quad i \neq j, \quad E_i \neq E_j \quad \text{(A.1.11)}$$

If different eigenfunctions have identical eigenvalues (*i.e.* same energy) they are known as *degenerate*.

Calculation of coefficients

Starting from Eq. (A.1.10) we have

$$|\psi\rangle = \sum_n a_n |\phi_n\rangle \quad \text{(A.1.12)}$$

Now, let"s take the bracket with an eigenfunction $\langle\phi_k|$

$$\langle\phi_k|\psi\rangle = \langle\phi_k|\sum_n a_n|\phi_n\rangle = \sum_n a_n\langle\phi_k|\phi_n\rangle \quad \text{(A.1.13)}$$

From the statement of orthogonality in Eq. (A.1.11) we have

[†] Note that exp[i*kx*] provides a continuous set of eigenfunctions for unbound states, i.e., the expansion in terms of eigenfunctions is the Fourier transform.

$$\langle \phi_k | \phi_n \rangle = \delta_{nk} \tag{A.1.14}$$

where δ_{nk} is the Kronecker delta function, i.e. $\delta_{nk} = 1$ only when $n = k$.

Thus,

$$a_k = \langle \phi_k | \psi \rangle . \tag{A.1.15}$$

These coefficients are very important.

Since the eigenfunctions are usually known, often the set of coefficients provides the interesting information in a particular problem. This may be clear from an example.

An example: the expanding square well

Consider an electron occupying the ground state of an infinite square well of length L. As shown in Fig. A 1.1, at time $t = 0$, the well suddenly triples in size. What happens to the electron?

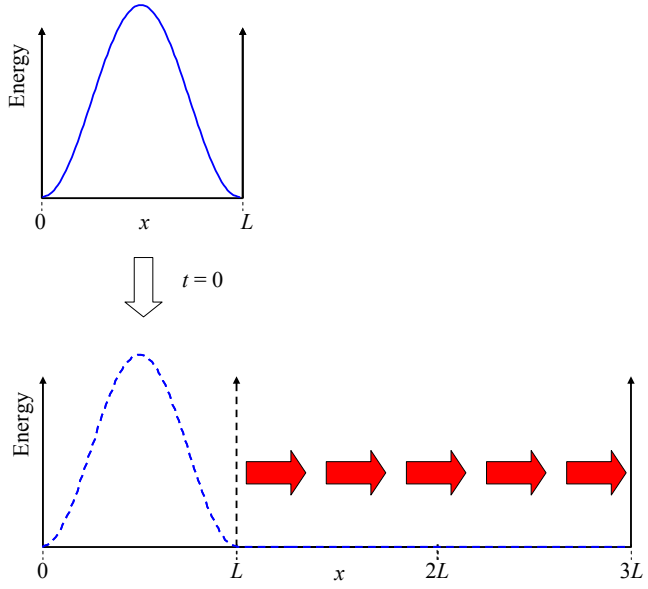


Fig. A 1.1. An electron is in the ground state of an infinite potential well. Suddenly the infinite well expands instantaneously to three times its previous size.

Let"s begin to answer this question by considering the wavefunction prior to the expansion of the well:[†]

$$\psi(x) = \sqrt{\frac{2}{L}} \sin(\pi x/L), \quad 0 \le x \le L \tag{A.1.16}$$

If we substitute this back into the Schrödinger Equation we can easily confirm that the effect of operating on this wavefunction with the Hamiltonian is the same as multiplying the wavefunction by a constant. i.e.

$$H \left| \sqrt{\frac{2}{L}} \sin(\pi x/L) \right\rangle = E \left| \sqrt{\frac{2}{L}} \sin(\pi x/L) \right\rangle \tag{A.1.17}$$

[†] Note that relative the previous square well analysis (see Fig. 1.21) we have shifted the x-axis here such that the left wall is at $x = 0$.

Thus, this wavefunction is an *eigenfunction* of the original square well, and the constant, *E*, the energy, is the *eigenvalue* corresponding to the eigenfunction.

Now, Eq. (A.1.16) is the lowest energy eigenfunction for the time independent infinite square well potential. The wavefunction evolves in time according to

$$E\zeta(t) = i\hbar\frac{d}{dt}\zeta(t) \tag{A.1.18}$$

Solving Eq. (A.1.2) gives:

$$\Psi(x,t) = \psi(x)\zeta(t) = \sqrt{\frac{2}{L}}\sin(\pi x/L)\exp\left[-i\frac{E}{\hbar}t\right] \tag{A.1.19}$$

where

$$E = E_L = \frac{\hbar^2\pi^2}{2mL^2}. \tag{A.1.20}$$

The probability density, however, is time independent:

$$\left|\Psi(x,t)\right|^2 = \frac{2}{L}\sin^2(\pi x/L), \quad 0 \le x \le L \tag{A.1.21}$$

We have verified that the eigenfunction is stationary, as it must be.

Now, when the well expands, the wavefunction cannot change instantaneously. To confirm this, consider a step change in the wavefunction in Eq. (A.1.2) – the energy would tend to ∞.

But the stationary states of the new well are

$$\psi(x) = \sqrt{\frac{2}{3L}}\sin(n\pi x/3L), \quad 0 \le x \le 3L \tag{A.1.22}$$

i.e. the wavefunction of the electron at $t = 0$ is not a stationary state in the expanded well. *To determine the evolution of the electron wavefunction in time, we must re-express the wavefunction in terms of the eigenfunctions of the expanded well.* We can then calculate the evolution of each eigenfunction from Eq. (A.1.2).

The wavefunction is now described as a *linear combination* of eigenfunctions:

$$\begin{aligned}\psi(x) &= \sqrt{\frac{2}{L}}\sin(\pi x/L), \quad 0 \le x \le L \\ &= \sum_n a_n\sqrt{\frac{2}{3L}}\sin(n\pi x/3L), \quad 0 \le x \le 3L\end{aligned} \tag{A.1.23}$$

where a_n is a set of constants, that weight the contributions from each eigenfunction.

We express the wavefunction as a linear combination of the eigenfunctions of the *expanded* well. From Eq. (A.1.15) we have

$$\psi(x) = \sum_n a_n\sqrt{\frac{2}{3L}}\sin(n\pi x/3L), \quad 0 \le x \le 3L \tag{A.1.24}$$

Thus, the coefficients are

$$a_n = \langle \phi_n | \psi \rangle = \int_0^L \sqrt{\frac{2}{L}} \sin(\pi x/L) \sqrt{\frac{2}{3L}} \sin(n\pi x/3L)\, dx \tag{A.1.25}$$

Solving gives

$$a_n = \begin{cases} \dfrac{1}{\sqrt{3}} & n = 3 \\ -\dfrac{6\sqrt{3}}{\pi} \dfrac{\sin(n\pi/3)}{n^2 - 9} & n \neq 3 \end{cases} \tag{A.1.26}$$

In Fig. A 1.2 we plot the cumulative effect of adding the weighted eigenfunctions. After about 10 eigenfunctions, the linear combination is a close approximation to the initial wavefunction.

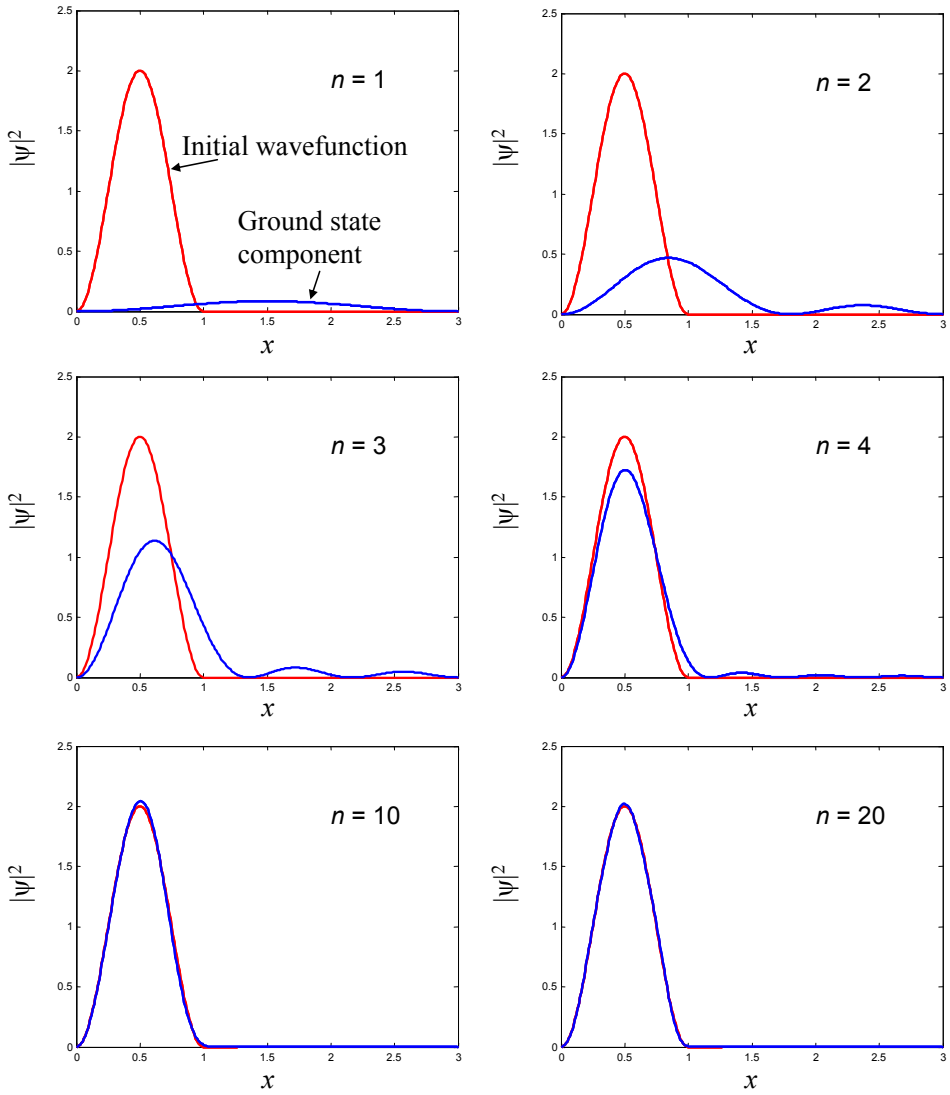


Fig. A 1.2. Here we show various approximations for the initial wavefunction. We need about 10 weighted eigenfunctions for a close match. The agreement gets better with addition eigenfunctions.

Next, we calculate the evolution of the wavefunction. From Eq. (A.1.24), we get

$$\Psi(x,t) = \sum_n a_n \sqrt{\frac{2}{3L}} \sin(n\pi x/3L) \exp\left[-i\frac{E_n}{\hbar}t\right], \quad 0 \le x \le 3L \tag{A.1.27}$$

where

$$E_n = \frac{n^2\pi^2\hbar^2}{2m(3L)^2} \tag{A.1.28}$$

The evolution of the wavefunction with time is shown in Fig. A 1.3, below.

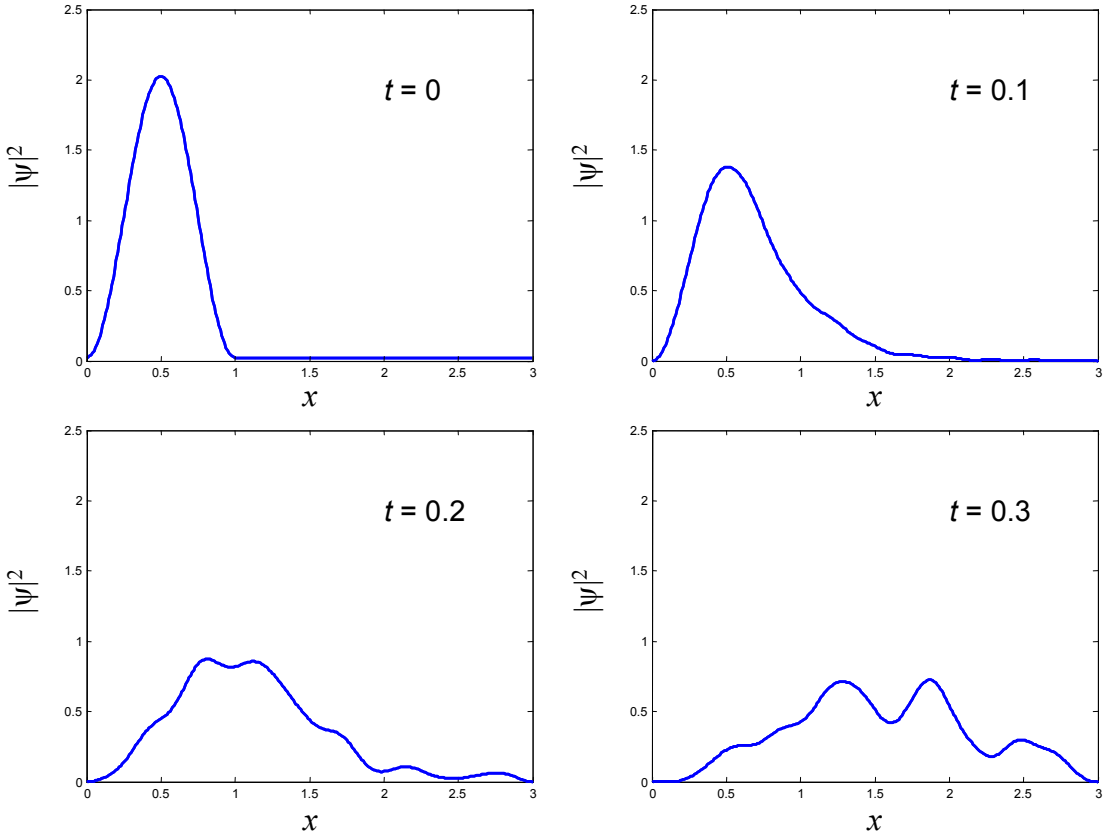


Fig. A 1.3. The evolution of the wavefunction after the expansion of the walls at t = 0.

Propagation of a Gaussian Wavepacket

Next we will examine the propagation of a Gaussian wavepacket in free space. Again, we will expand the wavefunction in terms of its eigenfunctions.

Consider an electron in free space. Let the initial wavepacket be a Gaussian.

$$\psi(x,0) = \left(\pi L^2\right)^{-1/4} \exp\left[-\frac{x^2}{2L^2}\right] \exp\left[ik_0 x\right] \tag{A.1.29}$$

Note that we have introduced a phase factor $\exp[ik_0x]$. Recall that multiplying the real space wavefunction by the phase factor $\exp[ik_0x]$, is equivalent to shifting the k-space wavefunction by k_0. Since the Fourier transform was centered at $k = 0$ prior to the shift, the phase factor shifts the expectation value of k to k_0. Hence the factor $\exp[ik_0x]$ gives the wavepacket has a non-zero average momentum.

Now, the eigenfunctions of the Schrödinger Equation in free space are the complex exponentials

$$\phi(k) = \exp\left[ikx\right] \tag{A.1.30}$$

where k is continuous.

Each eigenfunction evolves with time as

$$\phi(k,t) = \exp\left[ikx\right] \exp\left[-i\frac{E}{\hbar}t\right] \tag{A.1.31}$$

Expanding the wavefunction as a linear combination of these eigenfunctions we have

$$\psi(x,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} A(k)\phi(k,t)\,dk \tag{A.1.32}$$

$$\psi(x,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} A(k)\exp\left[ikx\right]\exp\left[-i\frac{E}{\hbar}t\right]dk \tag{A.1.33}$$

where $A(k)/2\pi$ describes the weighting of each complex exponential eigenfunction.

Since $\phi(k) = \exp\left[ikx\right]$, Eq. (A.1.32) evaluated at $t = 0$ is simply the inverse Fourier transform:

$$\psi(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} A(k)\exp\left[ikx\right]dk \tag{A.1.34}$$

Thus, $A(k)$ is determined from the Fourier transform of the wavefunction

$$A(k) = \left\langle \phi(k) \middle| \psi(x,0) \right\rangle = \int_{-\infty}^{+\infty} \psi(x,0)e^{-ikx}dx = \left(4\pi L^2\right)^{1/4} \exp\left[-\frac{L^2\left(k-k_0\right)^2}{2}\right] \tag{A.1.35}$$

Now, before we can substitute $A(k)$ back into Eq. (A.1.32) to get the time evolution of $\psi(x,t)$ we need to consider the possible k dependence of energy, E.

In general the relation between E and k is known as the **dispersion relation**.

Appendix 1. Electron Wavepacket Propagation

The dispersion relation is important because propagation of an electron in free space, or in a particular material is determined by the dispersion relation. Let‟s consider some examples.

(i) Linear dispersion relation

The dispersion relation determines how the wavepacket spreads in time. For example, if instead of an electron we were considering a photon with $E = \hbar\omega = \hbar ck$, the dispersion relation is linear and the photon does not spread as it propagates.

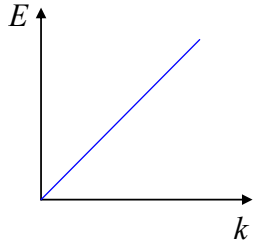


Fig. A 1.4. A linear dispersion relation.

The time dependent factor is

$$\exp\left[-i\frac{E}{\hbar}t\right] \equiv \exp\left[-i\omega t\right] = \exp\left[-ickt\right] \tag{A.1.36}$$

Solving Eq. (A.1.32) gives

$$\psi(x,t) = \frac{1}{\left(\pi L^2\right)^{1/4}} \exp\left[ik_0\left(x-ct\right)\right] \exp\left[-\frac{\left(x-ct\right)^2}{2L^2}\right]. \tag{A.1.37}$$

Thus, the probability density is simply the original function shifted linearly in time:

$$\left|\psi(x,t)\right|^2 = \frac{1}{\left(\pi L^2\right)^{1/2}} \exp\left[-\frac{\left(x-ct\right)^2}{L^2}\right] \tag{A.1.38}$$

(ii) Quadratic dispersion relation

For plane wave eigenfunctions, however, the dispersion relation is quadratic and we have

$$E = \frac{\hbar^2 k^2}{2m} \tag{1.39}$$

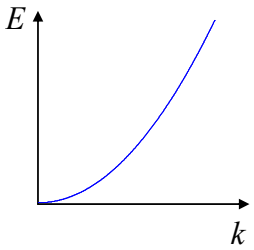


Fig. A 1.5. A quadratic dispersion relation.

As we have seen, particles in a box also have quadratic dispersion relations.

Thus the time dependent factor is

$$\exp\left[-i\frac{E}{\hbar}t\right] \equiv \exp\left[-i\omega t\right] = \exp\left[-i\frac{\hbar k^2}{2m}t\right] \tag{A.1.40}$$

Solving Eq. (A.1.32) gives

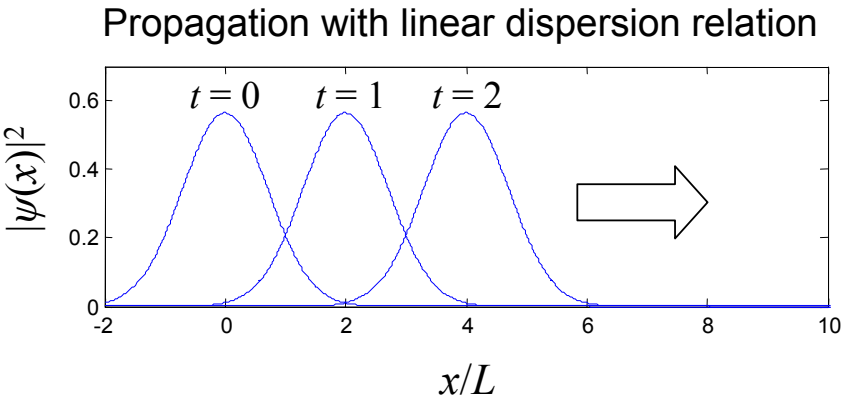
$$\psi(x,t) = \frac{1}{\left(\pi L^2\right)^{1/4}} \frac{1}{\sqrt{1+i\hbar t/mL^2}} \exp\left[i\left(k_0 x - \frac{\hbar {k_0}^2 t}{2m}\right)\right] \exp\left[-\frac{\left(x-\hbar k_0 t/m\right)^2}{2L^2\left(1+i\hbar t/mL^2\right)}\right] \tag{A.1.41}$$

and

$$\left|\psi(x,t)\right|^2 = \left(2\pi\left[\Delta x(t)\right]^2\right)^{-1/2} \exp\left[-\frac{\left(x-\hbar k_0 t/m\right)^2}{2\left[\Delta x(t)\right]^2}\right] \tag{A.1.42}$$

where

$$\left[\Delta x(t)\right]^2 = \frac{L^2}{2}\left(1+\left(\frac{\hbar t}{mL^2}\right)^2\right) \tag{A.1.43}$$



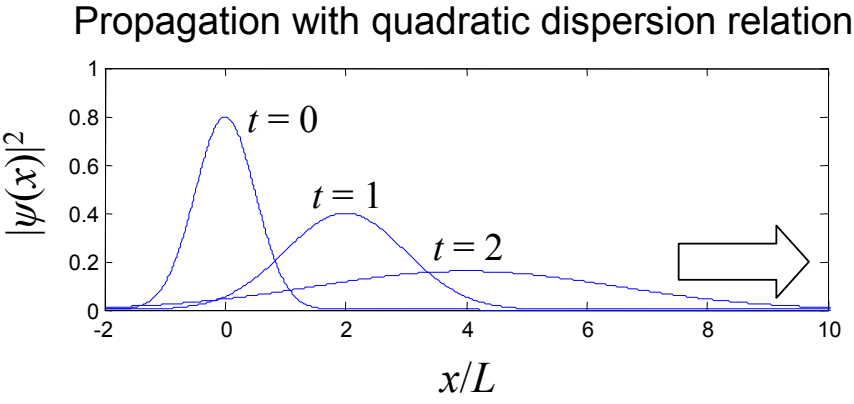


Fig. A 1.6. The evolution of a Gaussian wavepacket for linear and quadratic dispersion relations.

Group Velocity

As with a classical wave, the average velocity of the wavepacket is the **group velocity**, defined as the time derivative of the expectation value of position:

$$v_g = \frac{d}{dt}\langle x \rangle = \left\langle \frac{1}{\hbar}\frac{dE}{dk} \right\rangle = \left\langle \frac{d\omega}{dk} \right\rangle \tag{A.1.44}$$

If the wavepacket is highly peaked in k-space it is possible to simplify Eq. (A.1.44) by evaluating $d\omega/dk$ at the expectation value of k:

$$v_g = \left\langle \frac{d\omega}{dk} \right\rangle \approx \left. \frac{d\omega}{dk} \right|_{\langle k \rangle} \tag{A.1.45}$$

For the linear dispersion relation, $d\omega/dk$ is constant so we don't need the approximation:

$$v_g = \frac{d\omega}{dk} = c \tag{A.1.46}$$

i.e. the photon moves along at the speed of light, as expected.

For the quadratic dispersion relation, we have

$$v_g = \left\langle \frac{d\omega}{dk} \right\rangle = \left\langle \frac{\hbar k}{m} \right\rangle = \frac{\hbar k_0}{m} \tag{A.1.47}$$

Since $\hbar k_0$ is the expectation value of momentum, this is indeed the average velocity.

Problems

1. Consider the electron in the expanded well discussed in the notes. Numerically simulate the wave function as a function of time. Using the simulation or otherwise, determine the expectation value of energy before and after the well expands? Predict the behavior of the expectation values of position and momentum as a function of time.

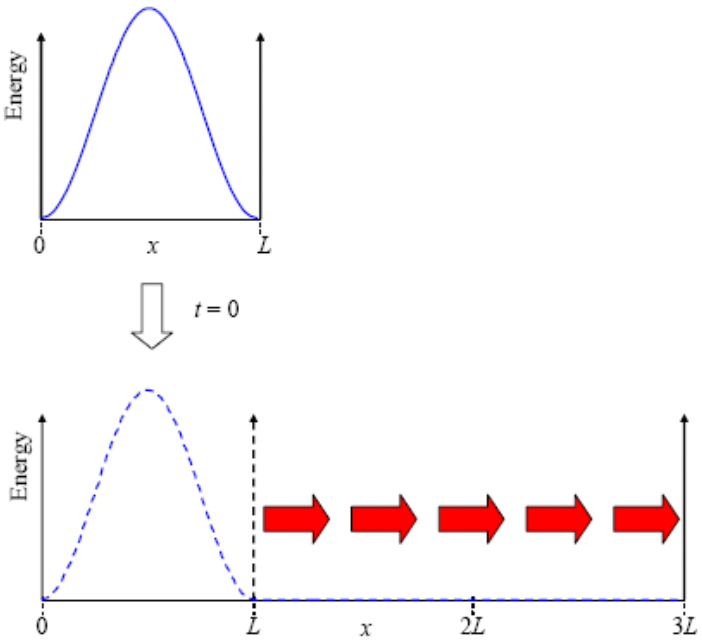


Fig. A 1.7. The expanding well.

2. Consider the wave function in an infinite square well of width L illustrated below at time $t = 0$. How will the wavefunction evolve for $t > 0$. Will the wave function ever return to its original position? If so, at what time $t = T$ will this occur?

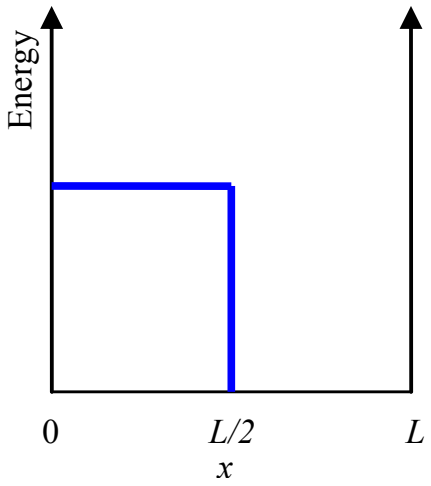


Fig. A 1.8. An initial state within a quantum well.

3. Derive the expression for the propagation of a Gaussian wavepacket with the linear dispersion relation $E = \hbar\omega = \hbar ck$. (Equation (A.1.37))

Appendix 2. The hydrogen atom

The box model of the hydrogen atom

Hydrogen is the simplest element. There are just two components: an electron, and a positively charged nucleus comprised of a single proton.

Perhaps the simplest model of the hydrogen atom employs our now familiar square potential wells. This approximation cannot be taken very far, but it does illustrate the origin of the shapes of some of the orbitals.

If we compare the smooth, spherically symmetric Coulomb potential to our box model of an atom, it is clear that the box approximation will give up a lot of accuracy in the calculation of atomic orbitals and energy. The box, however, does yield insights into the shape of the various possible atomic orbitals.

The box is a separable potential. Thus, the atomic orbitals can be described by a product:

$$\psi(x,y,z) = \psi_x(x)\psi_y(y)\psi_z(z) \tag{A.2.1}$$

If the wall have infinite potential, the possible energies of the electron are given by

$$E_{n_x,n_y,n_z} = \frac{\hbar^2\pi^2}{2m}\left(\frac{n_x^2}{L_x^2}+\frac{n_y^2}{L_y^2}+\frac{n_z^2}{L_z^2}\right) \tag{A.2.2}$$

where the dimensions of the box are $L_x \times L_y \times L_z$ and n_x, n_y and n_z are integers that correspond to the state of the electron within the box.

For example, consider the ground state of a box with infinite potential walls. $(n_x,n_y,n_z) = (1,1,1)$

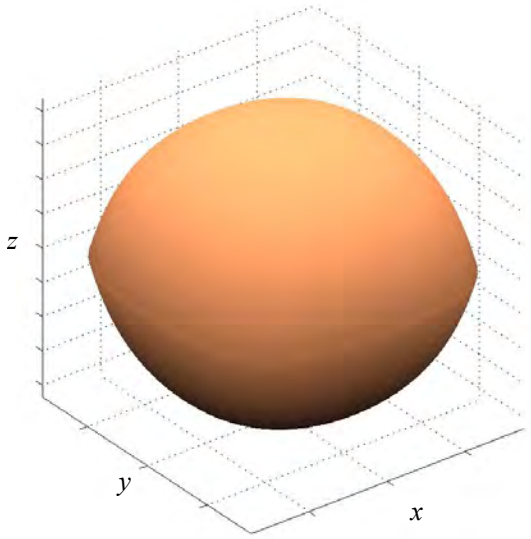


Fig. A 2.1. The ground state of a 3 dimensional box.
$(n_x,n_y,n_z) = (1,1,1)$

Appendix 2. The hydrogen atom

Now, consider the orbital"s shape if either $\psi_x(x)$ or $\psi_y(y)$ or $\psi_z(z)$ is in the first excited state: $(n_x,n_y,n_z) = (2,1,1)$, $(n_x,n_y,n_z) = (1,2,1)$ or $(n_x,n_y,n_z) = (1,1,2)$

The 1*s* orbital is similar to $(n_x,n_y,n_z) = (1,1,1)$. The *p* orbitals are similar to the first excited state of the box, i.e. $(n_x,n_y,n_z) = (2,1,1)$ is similar to a p_x orbital, $(n_x,n_y,n_z) = (1,2,1)$ is similar to a p_y orbital and $(n_x,n_y,n_z) = (1,1,2)$ is similar to a p_z orbital.

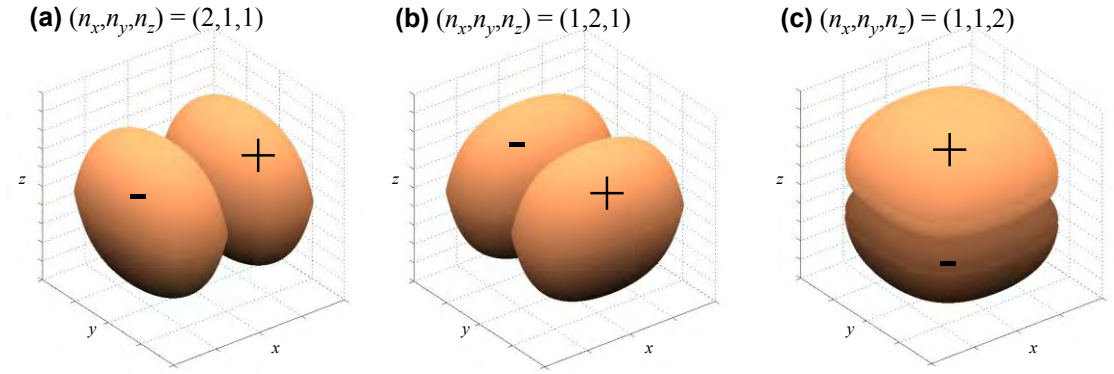


Fig. A 2.2. The first excited states of a 3 dimensional box. **(a)** $(n_x,n_y,n_z) = (2,1,1)$, **(b)** $(n_x,n_y,n_z) = (1,2,1)$, **(c)** $(n_x,n_y,n_z) = (1,1,2)$.

The approximation soon breaks down, however. The 2*s* orbital, which has the same energy as the 2*p* orbitals is most similar to the box orbital $(n_x,n_y,n_z) = (3,3,3)$, which has significantly higher energy. Nevertheless, the box does illustrate the alignment of the three *p* orbitals with the *x*, *y* and *z* axes.

Appendix 3. The Born-Oppenheimer approximation†

Consider the hydrogen atom Hamiltonian. Let the electron coordinate be x, and the nuclear coordinate be X. We will assume that the system is one dimensional for the purposes of explaining the approximation.

$$\hat{H} = -\frac{\hbar^2}{2m_e}\frac{d^2}{dx^2} - \frac{\hbar^2}{2m_N}\frac{d^2}{dX^2} - \frac{e^2}{4\pi\varepsilon_0|x-X|} \tag{A.3.1}$$

Now let‟s separate the solution, ψ, into an electron-only factor φ, and the nuclear-dependent factor χ:

$$\psi(x,X) = \varphi(x,X)\chi(X). \tag{A.3.2}$$

Substituting into Eq. (A.3.1) gives:

$$H\psi = -\frac{\hbar^2}{2m_e}\frac{d^2\varphi}{dx^2}\chi - \frac{\hbar^2}{2m_N}\left(\frac{d^2\varphi}{dX^2}\chi + 2\frac{d\varphi}{dX}\frac{d\chi}{dX} + \frac{d^2\chi}{dX^2}\varphi\right) + V(x,X)\varphi\chi \tag{A.3.3}$$

where we have replaced the Coulomb potential by V.

Now using the Born-Oppenheimer approximation, *i.e.* $m_e << m_N$, we approximate Eq. (A.3.3) by:

$$H\psi = -\frac{\hbar^2}{2m_e}\frac{d^2\varphi}{dx^2}\chi + V(x,X)\varphi\chi. \tag{A.3.4}$$

Next, canceling the nuclear-dependent factor χ:

$$H\varphi = -\frac{\hbar^2}{2m_e}\frac{d^2\varphi}{dx^2} + V(x,X)\varphi. \tag{A.3.5}$$

This equation is used to solve for the electron coordinates in a given nuclear configuration. The nuclear configuration is then optimized.

† This Appendix is adapted in part from *Molecular Quantum Mechanics* by Atkins and Friedman

Appendix 4. Hybrid Orbitals

Linear alignment with two neighbors (sp hybridization)

Consider three atoms in a line, as shown in Fig. A 4.1. Arbitrarily we align the atoms with the *x*-axis.

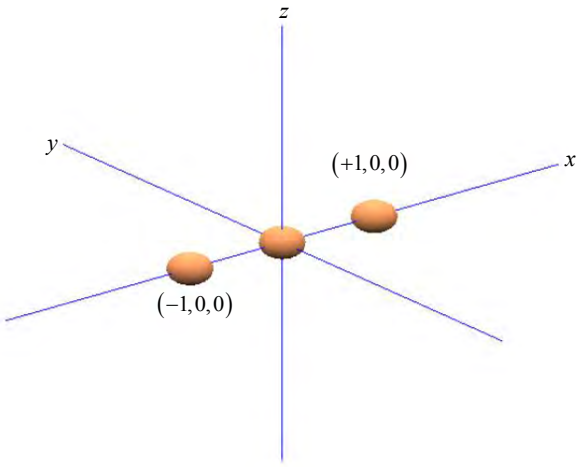


Fig. A 4.1. Three atoms in a line.

We wish to determine the contribution of the central atom"s orbitals to σ bonds. Recall that σ bonds have electron density on the axis between the atoms.

Now, if the basis set consists of *s* and *p* orbitals, only *s* and p_x orbitals can contribute to σ bonds on the *x*-axis. p_y and p_z orbitals have zero density on the *x*-axis and therefore cannot contribute to the σ bonds. They may contribute to π bonds however.

Let"s define the symmetry adapted atomic orbitals that contribute to σ bonds generally as:

$$\phi_\sigma = c_s\phi_s + c_{p_z}\phi_{p_z} \tag{A.4.1}$$

where c_s and c_{px} are the weighting coefficients for the s and px orbitals respectively. There are two σ bonds: one to the left, and one to the right. We"ll define the two symmetry adapted atomic orbitals that contribute to these σ bonds as ϕ_ζ^1 and ϕ_ζ^2 respectively.

The *s* orbital contributes equally to both symmetry adapted atomic orbitals. *i.e.*

$$|c_s|^2 = \frac{1}{2}, \quad c_s = \frac{1}{\sqrt{2}} \tag{A.4.2}$$

Since the p_x orbital is aligned with the *x*-axis, we can weight the p_x orbital components by the coordinates of the two neighboring atoms at $x = +1$ and $x = -1$,

$$\phi_\sigma^1 = c_s\phi_s + c_p\phi_{p_x}$$
$$\phi_\sigma^2 = c_s\phi_s - c_p\phi_{p_x} \qquad \text{(A.4.3)}$$

In the first orbital, we are adding the *s* and p_x orbitals in-phase. Consequently, we have maximum electron density in the positive *x*-direction. In the second, we are adding the *s* and p_x orbitals out of phase, yielding a maximum electron density in the negative *x*-direction.

Normalizing each orbital gives

$$|c_p|^2 = \frac{1}{2}, \quad c_p = \sqrt{\frac{1}{2}} \qquad \text{(A.4.4)}$$

Thus, the first symmetry adapted atomic orbital is

$$\phi_\sigma^1 = \tfrac{1}{\sqrt{2}}\left(\phi_s + \phi_{p_x}\right) \qquad \text{(A.4.5)}$$

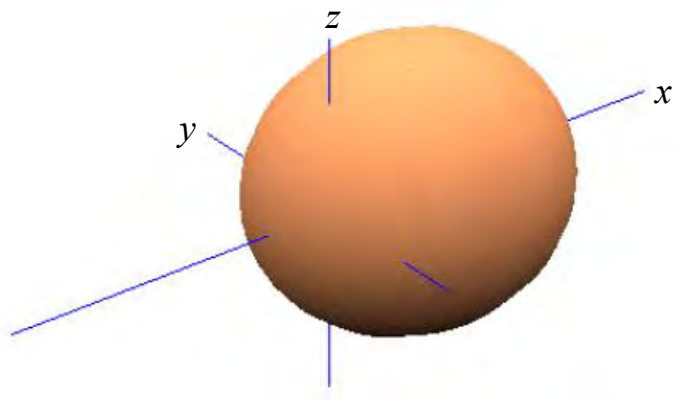


Fig. A 4.2. The *sp* hybrid-ized atomic orbital in the +*x* direction.

Similarly, the second symmetry adapted atomic orbital is

$$\phi_\sigma^2 = \tfrac{1}{\sqrt{2}}\left(\phi_s - \phi_{p_x}\right). \qquad \text{(A.4.6)}$$

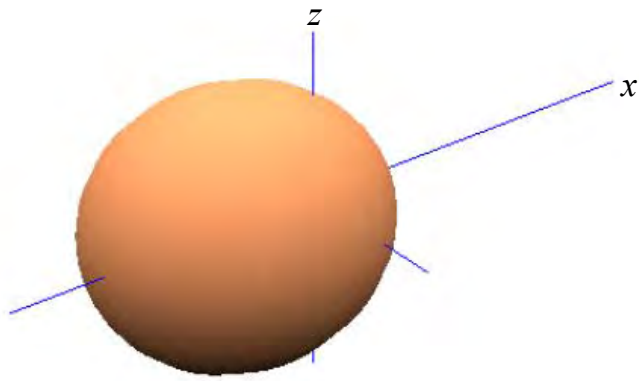


Fig. A 4.3. The *sp* hybrid-ized atomic orbital in the –*x* direction.

Thus, based purely on symmetry arguments, in a linear chain of atoms it is convenient to re-express the four atomic orbitals *s*, p_x, p_y and p_z, as

$$\phi_\sigma = \tfrac{1}{\sqrt{2}}\left(\phi_s \pm \phi_{p_x}\right), \qquad \text{(A.4.7)}$$

where p_y and p_z remain unaffected. This is known as *sp* hybridization since we have combined one *s* atomic orbital, and one *p* atomic orbital to create two atomic orbitals that contribute to σ bonds.

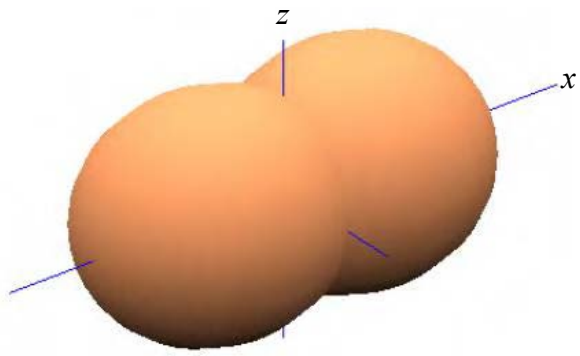


Fig. A 4.4. We plot both *sp* hybridized atomic orbitals. They point along the *x*-axis.

The remaining p_y and p_z atomic orbitals may combine in molecular orbitals with higher energy. The highest occupied molecular orbital (HOMO) is also known as the frontier molecular orbital. In an *sp*-hybridized material, the frontier molecular orbital will be a linear combination of p_y and p_z atomic orbitals. The frontier molecular orbital is relevant to us, because it more likely than deeper levels to be partially filled. Consequently, conduction is more likely to occur through the HOMO than deeper orbitals.

Now, σ bonds possess electron densities localized between atoms. But π bonds composed of linear combinations of *p* orbitals can be *delocalized* along a chain or sheet of atoms. Thus, if the HOMO is a π bond, it is much easier to push an electron through it; we"ll see some examples of this in the next section.

Planar alignment with three neighbors (sp^2 hybridization)

Consider a central atom with three equispaced neighbors at the points of a triangle; as shown in Fig. A 4.5. Arbitrarily we align the atoms on the *x*-*y* plane.

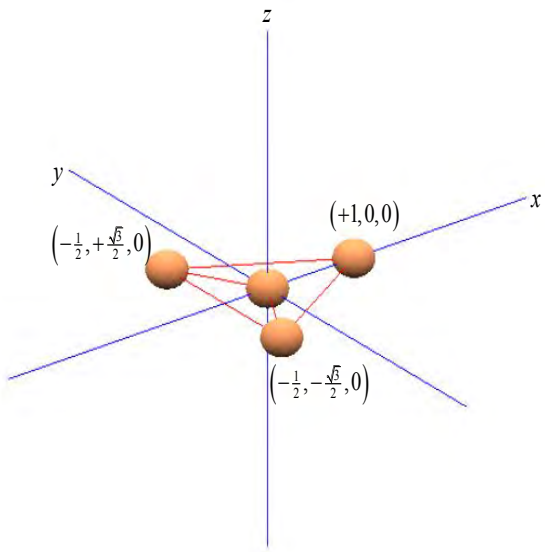


Fig. A 4.5. A central atom and its neighbors. Electrostatic repulsion can force the neighbors to the points of an equilateral triangle.

Once again, we wish to determine the contribution of the central atom"s orbitals to σ bonds. If the basis set consists of *s* and *p* orbitals, only *s*, p_x and p_y atomic orbitals can contribute to σ bonds in the *x*-*y* plane. p_z orbitals can only contribute to π bonds.

Let"s define the symmetry adapted atomic orbitals that contribute individually to σ bonds generally as:

$$\phi_\sigma = c_s\phi_s + c_{p_x}\phi_{p_x} + c_{p_y}\phi_{p_y} \tag{A.4.8}$$

The *s* orbital contributes equally to all three symmetry adapted atomic orbitals. i.e.

$$|c_s|^2 = \frac{1}{3}, \quad c_s = \frac{1}{\sqrt{3}} \tag{A.4.9}$$

Since the p_x orbital is aligned with the *x*-axis, and p_y with the *y*-axis, we can weight the *p* orbital components by the coordinates of the triangle of neighboring atoms

$$\begin{aligned}
\phi_\sigma^1 &= c_s\phi_s + c_p\left(+1\phi_{p_x} + 0\phi_{p_y}\right) \\
\phi_\sigma^2 &= c_s\phi_s + c_p\left(-\tfrac{1}{2}\phi_{p_x} + \tfrac{\sqrt{3}}{2}\phi_{p_y}\right) \\
\phi_\sigma^3 &= c_s\phi_s + c_p\left(-\tfrac{1}{2}\phi_{p_x} - \tfrac{\sqrt{3}}{2}\phi_{p_y}\right)
\end{aligned} \tag{A.4.10}$$

Normalizing each orbital gives

$$|c_p|^2 = \frac{2}{3}, \quad c_p = \sqrt{\frac{2}{3}} \tag{A.4.11}$$

Thus,

$$\begin{aligned}
\phi_\sigma^1 &= \frac{1}{\sqrt{3}}\phi_s + \sqrt{\frac{2}{3}}\phi_{p_x} + 0.\phi_{p_y} \\
\phi_\sigma^2 &= \frac{1}{\sqrt{3}}\phi_s - \frac{1}{\sqrt{6}}\phi_{p_x} + \frac{1}{\sqrt{2}}\phi_{p_y} \\
\phi_\sigma^3 &= \frac{1}{\sqrt{3}}\phi_s - \frac{1}{\sqrt{6}}\phi_{p_x} - \frac{1}{\sqrt{2}}\phi_{p_y}
\end{aligned} \tag{A.4.12}$$

This is known as sp^2 hybridization since we have combined one *s* atomic orbital, and two *p* atomic orbitals to create three atomic orbitals that contribute individually to σ bonds. The bond angle is 120º.

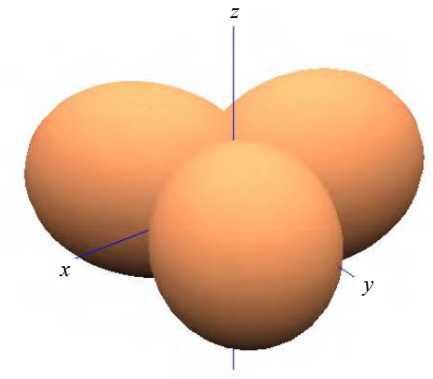


Fig. A 4.6. sp^2 hybridized molecular orbitals point to the vertices of a triangle.

The remaining p_z atomic orbitals will contribute to the frontier molecular orbitals of an sp^2 hybridized material; see for example ethene in Fig. A 4.7.

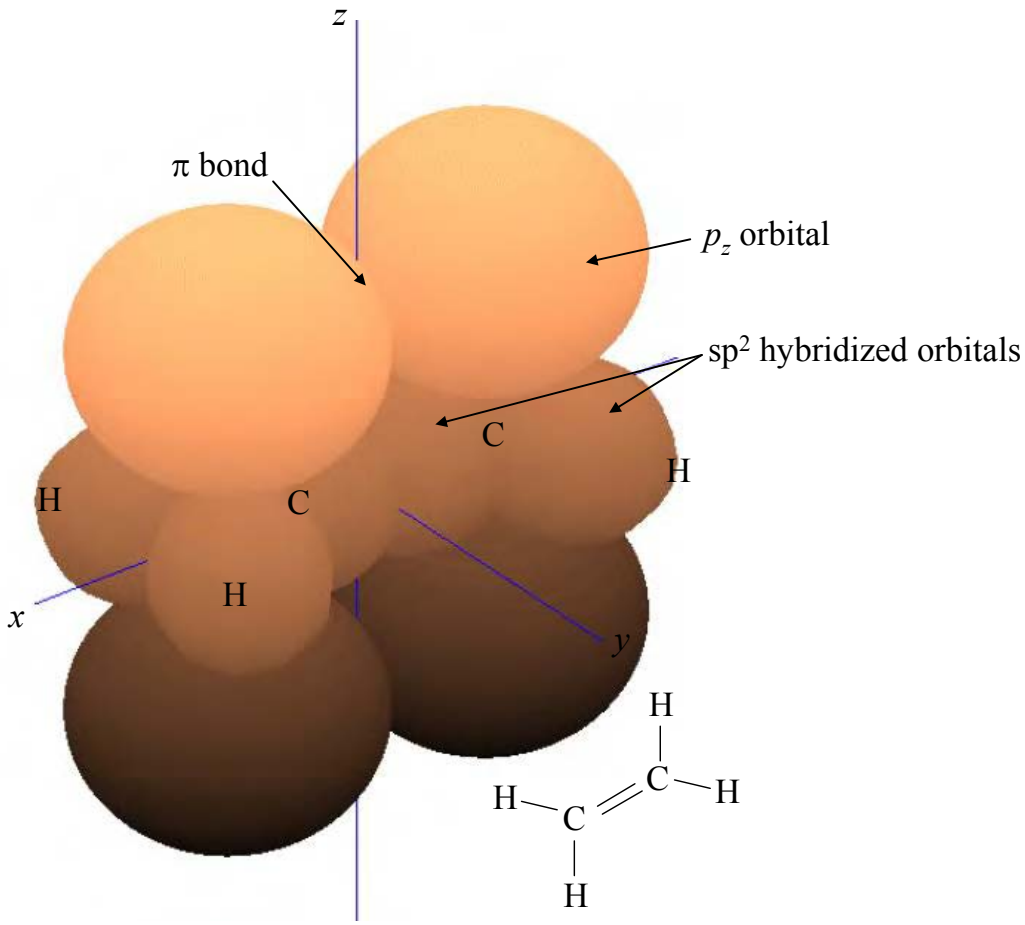


Fig. A 4.7. Ethene contains two sp^2 hybridized carbon atoms. The unhybridized p_z orbitals of carbon form π bonds.

As in the *sp* hybridized case, electrons in these π molecular orbitals may be delocalized. If electrons are delocalized over several neighboring atoms, then the molecule is said to be *conjugated*. Another sp^2 hybridized material was shown in Fig. 6.2. This is 1,3-butadiene, a chain of 4 × sp^2 hybridized carbon atoms. Note the extensive electron delocalization in the π bonds.

Some archetypal conjugated carbon-based molecules are shown in Fig. A 4.8. In each material the carbon atoms are sp^2 hybridized (surrounded by three neighbors at points of an equilateral triangle). Note that another typical characteristic of sp^2 hybridized materials is alternating single and double bonds.

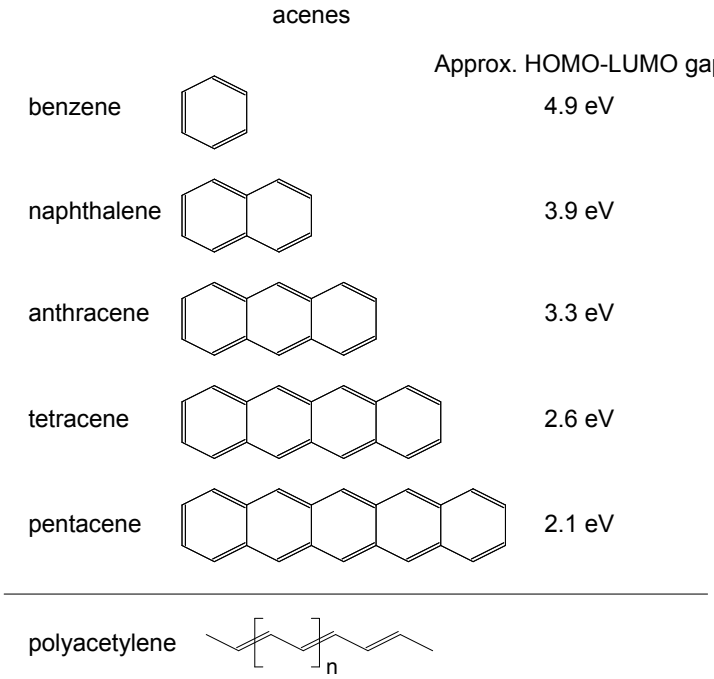


Fig. A 4.8. Examples of conjugated materials frequently employed in electronic devices. Note that the spacing between the HOMO and LUMO energy levels of electrons decrease as the molecules get bigger, consistent with particle in a box predictions of energy level spacing. Adapted from „Electronic Processes in Organic Crystals" by Pope and Swenberg, First Edition, Oxford University Press, 1982.

Tetrahedral alignment with four neighbors (sp^3 hybridization)

Consider a central atom with four equispaced neighbors. Repulsion between these atoms will push them to the points of a tetrahedron; see Fig. A 4.9.

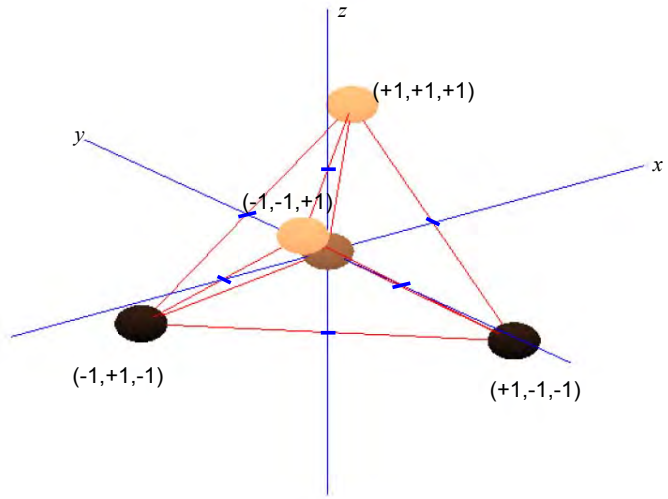


Fig. A 4.9. Electro-static repulsion forces four atoms surround-ing a central atom to the points of a tetrahedron.

Now, all atomic orbitals will contribute to σ bonds. There are no π bonds.

Appendix 4. Hybrid Orbitals

Let"s define the symmetry adapted atomic orbitals that contribute individually to σ bonds generally as:

$$\phi_\sigma = c_s\phi_s + c_{p_x}\phi_{p_x} + c_{p_y}\phi_{p_y} + c_{p_z}\phi_{p_z} \tag{A.4.13}$$

Once again, the *s* orbital contributes equally to all four symmetry adapted atomic orbitals. *i.e.*

$$\left|c_s\right|^2 = \frac{1}{4}, \quad c_s = \frac{1}{2} \tag{A.4.14}$$

Since the p_x orbital is aligned with the *x*-axis, p_y with the *y*-axis and p_z with the *z*-axis, we can weight the *p* orbital components by the coordinates of the triangle of neighboring atoms

$$\begin{aligned}
\phi_\sigma^1 &= c_s\phi_s + c_p\left(+1\phi_{p_x} + 1\phi_{p_y} + 1\phi_{p_z}\right)\\
\phi_\sigma^2 &= c_s\phi_s + c_p\left(-1\phi_{p_x} + 1\phi_{p_y} - 1\phi_{p_z}\right)\\
\phi_\sigma^3 &= c_s\phi_s + c_p\left(+1\phi_{p_x} - 1\phi_{p_y} - 1\phi_{p_z}\right)\\
\phi_\sigma^4 &= c_s\phi_s + c_p\left(-1\phi_{p_x} - 1\phi_{p_y} + 1\phi_{p_z}\right)
\end{aligned} \tag{A.4.15}$$

Normalizing each orbital gives

$$\left|c_p\right|^2 = \frac{1}{4}, \quad c_p = \frac{1}{2} \tag{A.4.16}$$

Thus,

$$\begin{aligned}
\phi_\sigma^1 &= \tfrac{1}{2}\left(\phi_s + \phi_{p_x} + \phi_{p_y} + \phi_{p_z}\right)\\
\phi_\sigma^2 &= \tfrac{1}{2}\left(\phi_s - \phi_{p_x} + \phi_{p_y} - \phi_{p_z}\right)\\
\phi_\sigma^3 &= \tfrac{1}{2}\left(\phi_s + \phi_{p_x} - \phi_{p_y} - \phi_{p_z}\right)\\
\phi_\sigma^4 &= \tfrac{1}{2}\left(\phi_s - \phi_{p_x} - \phi_{p_y} + \phi_{p_z}\right)
\end{aligned} \tag{A.4.17}$$

This is known as sp^3 hybridization since we have combined one *s* atomic orbital, and three *p* atomic orbitals to create four possible atomic orbitals that contribute individually to σ bonds. The bond angle is 109.5º.

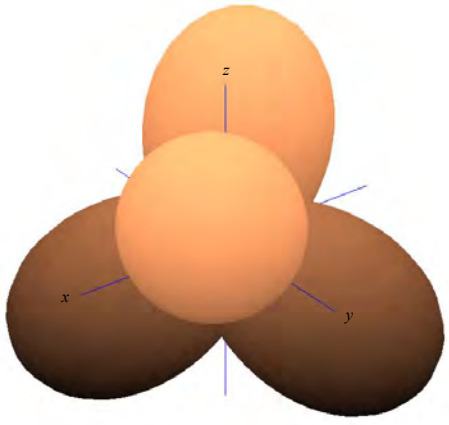


Fig. A 4.10. sp^3 hybridized orbitals point to the vertices of a tetraheron.

6.701 / 6.719 Introduction to Nanoelectronics

Spring 2010