

PC2232 Physics for Electrical Engineers

AY2014/15 Semester II

Part II: Quantum Mechanics

Mankei Tsang

E-mail address: mankei@nus.edu.sg

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING, NATIONAL UNIVERSITY OF SINGAPORE, 4 ENGINEERING DRIVE 3, SINGAPORE 117583

DEPARTMENT OF PHYSICS, NATIONAL UNIVERSITY OF SINGAPORE, 2 SCIENCE DRIVE 3, SINGAPORE 117551

PLEASE NOTE.

- Important equations are **boxed**. If you understand what they mean and where they come from you will probably do well. Simply memorizing them will not help much.
- * denotes advanced topics. They won't be the focus of exams but it is good to read them to gain a deeper understanding.
- There are **Exercises** throughout these lecture notes. They are probably more difficult than the tutorials and the exams but can be used to test and enhance your understanding of the discussions around them. **Answers** are given if they are relevant to future discussions. **Questions** are similar but more conceptual problems.
- Please **do not share** any lecture material, including these notes, **outside the class**; they may contain copyrighted material not suitable for circulation.
- Notations:
 - We follow the physics convention and use i for the imaginary number. Note that engineering literature often uses j instead. They mean the same thing.
 - Vectors are denoted by ***bold italic*** face.
 - Unit vectors are denoted by a hat \hat{e} on a vector.
 - Some symbols are recycled and may have different meanings in different contexts. For example, V may mean volume in one context and voltage in another, θ is used many times to denote an angle, ϕ is commonly used to denote an arbitrary phase, n usually means refractive index but may also be used to mean an integer index in a sum, while \hat{n} is usually a normal unit vector.
 - Beware of similar-looking symbols such as k (wavenumber) and κ (greek letter "kappa," used to denote the spring constant).

CHAPTER 1

From Classical Physics to Quantum Physics

1.1. Review of classical physics

1.1.1. Force. We have learned previously that a charged particle, such as an electron, experiences a Lorentz force in electromagnetic fields. A force \mathbf{F} is a vector quantity, and according to Newton's second law, the rate of change of momentum \mathbf{p} is equal to the force,

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}. \quad (1.1)$$

This equation is also valid for a massive particle in a gravitational field, which obeys general relativity. Gravitational fields obey Einstein field equation (http://en.wikipedia.org/wiki/Einstein_field_equations), an analog of Maxwell's equations. \mathbf{F} is the sum of all forces experienced by the particle.

In the Newtonian (nonrelativistic) limit, the momentum is the mass times the velocity \mathbf{v} :

$$\mathbf{p} \approx m\mathbf{v}, \quad (1.2)$$

and the velocity is the rate of change of the position:

$$\frac{d\mathbf{r}}{dt} = \mathbf{v}. \quad (1.3)$$

Given a Lorentz force and/or a gravitational force, we know how particles move through the equations of motion. Given what we know about the masses and charges of the particles and their positions and velocities, we can solve Maxwell's equations and Einstein field equation given such sources to calculate the electromagnetic fields and the gravitational fields. Thus the particles and the fields are coupled. The equations may be difficult to solve by hand or even by computers, but in principle they can be done.

1.1.2. Energy and momentum. A consequence of these equations of motion is that certain combinations of these physical quantities do not change in time and are thus **conserved**. We find that, for example, the total **energy** of a system of particles and fields is the same, even if they may exchange energy through interactions. We have talked about the energy of electromagnetic fields in Part I, and let us now consider the energy of a particle. In the Newtonian limit and assuming that the force \mathbf{F} depends just on the position \mathbf{r} and not time, the energy of a particle is the sum of **kinetic energy** and **potential energy** U :

$$E \approx \frac{p^2}{2m} + U, \quad (1.4)$$

where p is the magnitude of the momentum \mathbf{p} and U is the potential energy for a given force field $\mathbf{F}(\mathbf{r})$:

$$U(\mathbf{r}) = U_0 - \int_{\mathbf{r}_0}^{\mathbf{r}} \mathbf{F}(\mathbf{r}') \cdot d\mathbf{r}', \quad (1.5)$$

which is a line integral of the force. U_0 is a reference value for the potential energy and can be defined as any constant; the physics turns out to be the same if we add any constant to the energy, as long as we keep it as part of the total energy and never change it throughout the calculation. The energy is given by a more complicated expression when relativity and the full Lorentz force are taken into account. There is also **momentum conservation**, in which case the total momentum of a system as a vector quantity should remain the same. These conservation laws result from the equations of motion and are convenient because they simplify calculations by imposing constraints on the possible solutions.

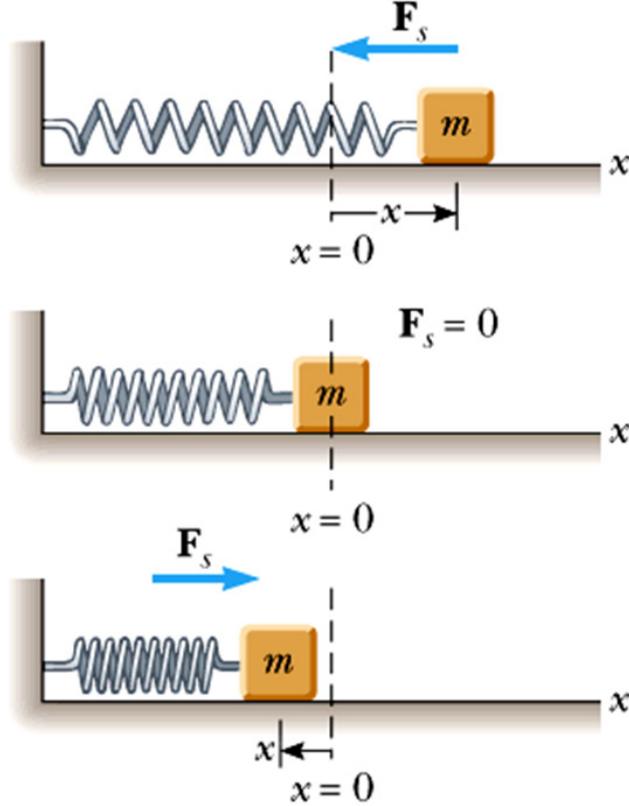


FIGURE 1.1. A mass-spring system as an example of a harmonic oscillator. The force opposes the displacement and is proportional to it.

1.1.3. Harmonic oscillator. For a simple example, consider a mass attached to a spring in one dimension, as shown in Fig. 1.1. This is an example of a **harmonic oscillator**. The force by the spring on the mass is a restoring force proportional to the position of the mass x :

$$F = -\kappa x, \quad (1.6)$$

where κ is the spring constant. The potential energy of the mass with this force is then

$$U(x) = U_0 + \int_0^x \kappa x' dx' = U_0 + \frac{1}{2} \kappa x^2. \quad (1.7)$$

If we are given this $U(x)$, the force is

$$F = -\frac{dU(x)}{dx} = -\kappa x, \quad (1.8)$$

and we see why the constant U_0 does not matter: the force is the same regardless. To visualize this problem, we can imagine a ball rolling near the bottom of a potential valley. Without friction, the ball would roll around indefinitely, as shown in Fig. 1.3.

1.1.4. Electrostatic potential. A more complicated example is the electrostatic potential. Consider a positively charged nucleus with charge e of a hydrogen atom and the electrostatic attractive force experienced by an electron around it, as shown on the left of Fig. 1.4. The nucleus is much heavier than the electron, so we can assume that the nucleus does not move much and for simplicity assume it to be at the origin of our coordinate system. The force is

$$\mathbf{F} = -e\mathbf{E} = -\frac{e^2}{4\pi\epsilon_0 r^2} \hat{\mathbf{r}}, \quad (1.9)$$

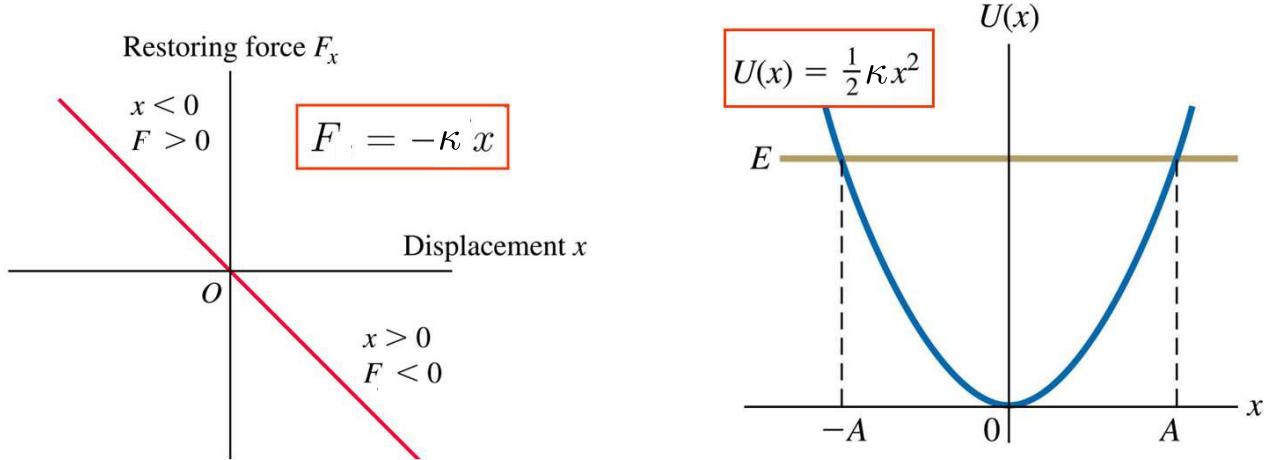


FIGURE 1.2. Potential energy of a particle in a harmonic oscillator is a parabola with a minimum. One can add any constant to the potential energy without changing the force and thus the physics.

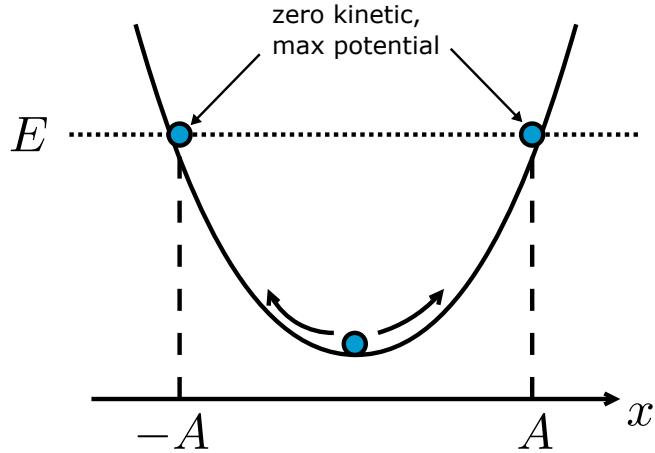


FIGURE 1.3. The solution for the particle position in a harmonic oscillator is $x(t) = A \cos(\omega_0 t + \phi)$ where $\omega_0 = \sqrt{\kappa/m}$ and A and ϕ depend on the initial conditions. Depending on its position, the particle may have different portions of kinetic energy and potential energy, but the total energy is conserved.

where r is the radial direction and \hat{r} is the radial unit vector. The potential is

$$U(r) = U_0 + \int_0^r \left(\frac{e^2}{4\pi\epsilon_0 r^2} \hat{r} \right) \cdot dr \quad (1.10)$$

$$= U_0 + \int_0^r \frac{e^2}{4\pi\epsilon_0 r^2} dr \quad (1.11)$$

$$= U_0 - \frac{e^2}{4\pi\epsilon_0 r}. \quad (1.12)$$

This Coulomb potential determines the potential energy of an electron as a function of its distance away from the nucleus. If we set $U_0 = 0$, the potential is always negative, it is $-\infty$ at $r = 0$, and approaches zero when infinitely far from the nucleus ($r \rightarrow \infty$), as plotted on the right of Fig. 1.4.

1.1.5. Determinism. Classical physics assumes that the physical quantities position r , velocity v , momentum p , mass m , charge q , electromagnetic fields (E, H), and gravitational fields are intrinsic properties of the world. They all have precise values, and they change in space and time according to the equations of motion given those values. A philosophical implication of this thinking is **determinism**: given initial conditions of these physical quantities, they evolve through precise equations of motion and their values are completely

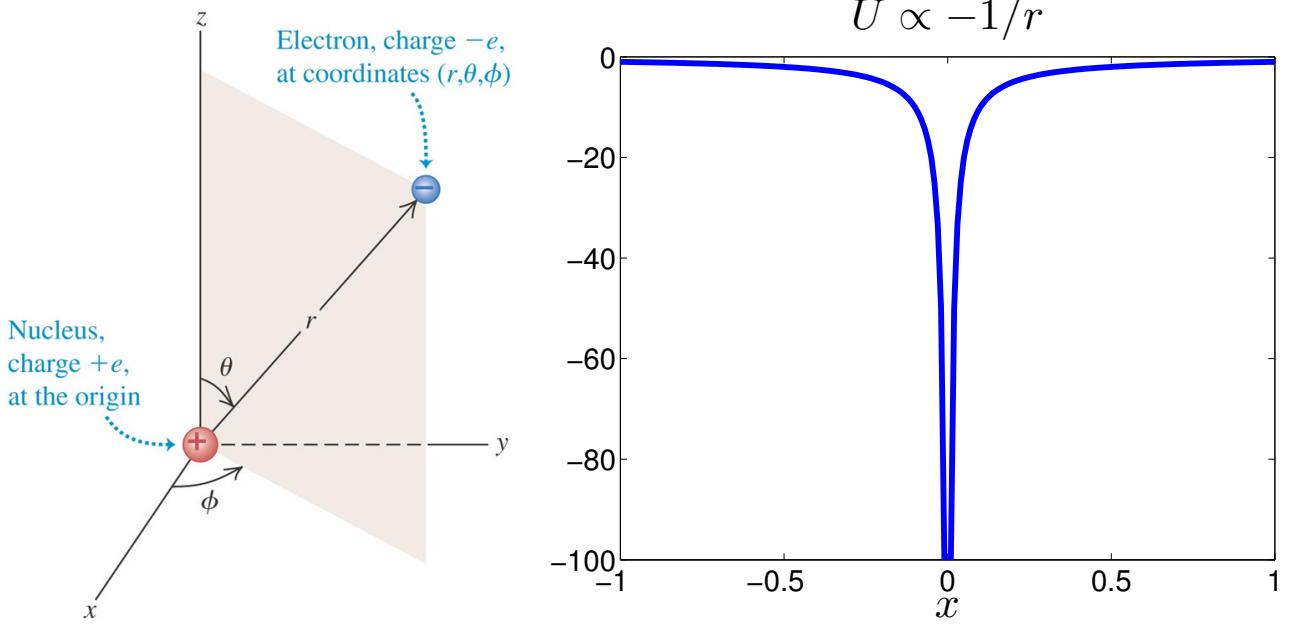


FIGURE 1.4. Right: A cross section along x of the Coulomb potential experienced by the electron $U(r) \propto -1/r = -1/\sqrt{x^2 + y^2 + z^2}$. When $y = 0$ and $z = 0$, $-1/r = -1/|x|$.

determined from the initial conditions. There is no room for them to wiggle or deviate from their designated paths. A mighty being that has precise information about all these variables in the universe at a given time and infinite computing capacities to solve the equations of motion will be able to perfectly predict the future and also infer the past. This is called a Laplace demon (http://en.wikipedia.org/wiki/Laplace%27s_demon).

Although we use this kind of procedure to some extent in practice for our predictions, such as fluid dynamics calculations in weather forecast, a full Laplace's demon for the universe is impractical because it is not possible for one to know about everything in the universe exactly, from the motion of astronomical objects down to individual atoms. Even if we have that information, the equations of motion will be way too difficult to solve by any means. To predict the full behavior of a room of gas, for example, we will need to know the positions and velocities of all $\sim 10^{24}$ gas molecules, not to mention the electromagnetic fields and gravitational fields in all space that the particles produce or experience.

1.1.6. Statistical mechanics. This difficulty of deterministic physics leads us to another branch of classical physics: **statistical mechanics**, which was pioneered by Ludwig Boltzmann (http://en.wikipedia.org/wiki/Ludwig_Boltzmann, Fig. 1.5). To deal with a room of gas molecules, for example, we throw our hands up and say that it is not possible to know everyone of them and solve their equations of motion exactly. What we do instead is to assign **probabilities** to each possible situation, make some simplifying assumptions about the probabilities, and calculate how the gas molecules behave on average. From such considerations rise the concepts of **entropy** and **temperature**.

1.2. Basic ideas of probability theory

1.2.1. Shell game. I shall illustrate the basic concepts of probabilities through a simple example: consider a ball, and it can have just two possible positions. We assign $x = 0$ if the ball is in the left cup and $x = 1$ if the ball is in the right cup, as shown in Fig. 1.6.

Consider an initial time t_0 and let x_0 be the initial ball position. If I prepare the setup completely **randomly**, there would be 50% chance it's in one cup, and 50% chance for the other.

$$P(x_0 = 0, t_0) = \frac{1}{2}, \quad P(x_0 = 1, t_0) = \frac{1}{2}. \quad (1.13)$$



FIGURE 1.5. Tombstone of Ludwig Boltzmann at the Vienna Central Cemetery, Austria.

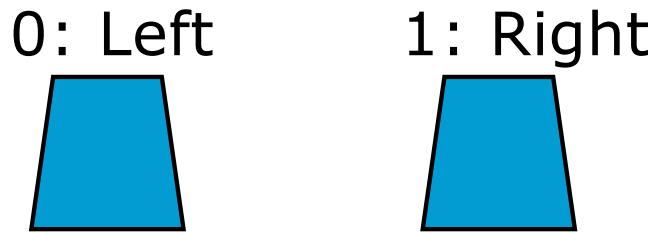


FIGURE 1.6. A ball can be in either of the cups. Assign $x = 0$ if the ball is in the left cup and $x = 1$ if it is in the right cup.

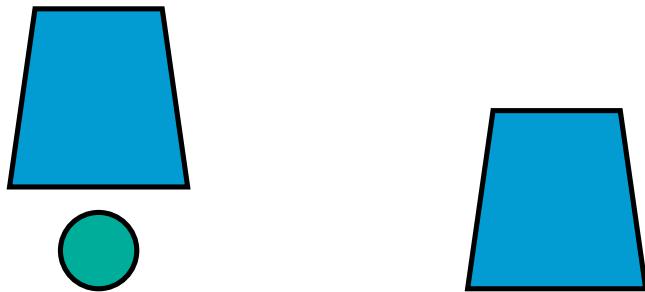


FIGURE 1.7. If I know for sure that the ball is in the left cup, the probability $P(x_0 = 0, t_0) = 1$, and the other probability is zero.

If we have seen that the ball is on the left, as shown in Fig. 1.7, that the probability $P(x_0 = 0, t_0)$ is one, meaning that we are certain it is on the left, while $P(x_0 = 1, t_0) = 0$, meaning that we are certain it won't be on the right.

In general, the probability of the ball in any cup is 1, so says that

$$\sum_{x_0} P(x_0, t_0) = 1, \quad (1.14)$$

which is a **normalization condition** for the probability distribution.

1.2.2. Deterministic transition. Suppose that the cups switch places, as in Fig. 1.8. Let t be the time after switching and x be the new position of the ball. The switch is represented by the equation

$$x = 1 - x_0, \quad (1.15)$$

such that, if $x_0 = 0$ initially, $x = 1$, and vice versa. This is an extremely simplified example of an equation of motion. In terms of probabilities, the probability distribution of x at time t , given that the ball was at x_0 at t_0 , becomes

$$P(x = 1, t|x_0 = 0, t_0) = 1, \quad P(x = 0, t|x_0 = 1, t_0) = 1, \quad (1.16)$$

and, by the normalization of probabilities,

$$P(x = 0, t|x_0 = 0, t_0) = 1 - P(x = 1, t|x_0 = 0, t_0) = 0, \quad (1.17)$$

$$P(x = 1, t|x_0 = 1, t_0) = 1 - P(x = 0, t|x_0 = 1, t_0) = 0, \quad (1.18)$$

that is, the probability that the ball stays at the same place is zero. This is an example of **deterministic transition**.

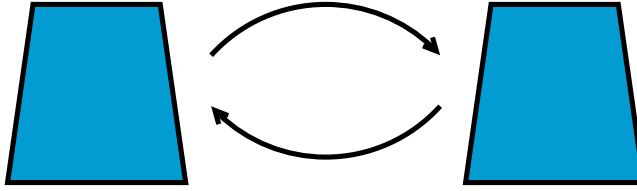


FIGURE 1.8. The cups have switched places.

In general, given the initial probabilities $P(x, t)$ and transition probabilities $P(x, t|x_0, t_0)$, the new probabilities are determined by

$$P(x, t) = \sum_{x_0} P(x, t|x_0, t_0)P(x_0, t_0), \quad (1.19)$$

which is nothing but a statement about the marginal probability distribution. If the ball is on the left initially, $P(x_0 = 0, t_0) = 1$, after switching, $P(x = 1, t) = 1$ and $P(x = 0, t) = 0$. Everything is deterministic; you know that the ball will for sure be on the right after the switch. This is a special case of a Laplace demon: you know the initial condition, the transition is deterministic, and you know the final outcome. An example of such a Laplace demon is found here: https://www.youtube.com/watch?v=4jRtK4_OgIE.

If $P(x_0 = 0, t_0) = P(x_0 = 1, t_0) = 1/2$, we are completely unsure which cup the ball is in. After switching, $P(x = 1, t) = 1/2$ and $P(x = 0, t) = 1/2$. Even though the transition is deterministic, the uninformative initial probabilities mean that the position remains completely unknown afterwards.

1.2.3. Mixing. If a random number of shuffles have taken place (and you don't know the number, or the player cheated like <https://www.youtube.com/watch?v=czCwt6fAaIA> and you have no idea what's going on),

$$P(x = 0, t|x_0 = 0, t_0) = P(x = 1, t|x_0 = 0, t_0) = \frac{1}{2}, \quad (1.20)$$

$$P(x = 0, t|x_0 = 1, t_0) = P(x = 1, t|x_0 = 1, t_0) = \frac{1}{2}. \quad (1.21)$$

If you know $x_0 = 0$, initially $P(x_0 = 0, t_0) = 1$, but after the shuffles, the new probabilities are $P(x = 0, t) = P(x = 1, t) = 1/2$. You no longer know where the ball is. This mixing transition completely removes any information you have about the ball position.

In general, the transition properties can be somewhere in between deterministic transitions and complete mixing. The general rule for the evolution of probabilities is, again, Eq. (1.19).

1.3. Disagreements with experiments

Classical physics is extremely successful on the macroscopic scale. It predicts the motions of astronomical objects accurately, it explains many things in our daily life, such as rainbows, the colors of the sky, and engines in our cars, and it has a self-consistent mathematical structure. It becomes **less accurate** when we study experiments on a small scale however, and there are **experimental observations** that seem to contradict predictions from classical physics, such as

- **Blackbody radiation:** An optical cavity with perfectly absorbing walls at a high temperature contains light with a certain spectrum. This spectrum is at odds with classical electrodynamics and statistical mechanics.
- **Photoelectric effect:** When light shines on a piece of metal, it can excite electrons to come off the metal into free space. Electrons are excited only if the light is above a certain frequency and **the energy of each electron then scales with the optical frequency**. This does not agree with any classical model of light and matter.

Historically, the experimental observations of these effects and their disagreement with classical physics led to doubts about classical physics and eventually a revolution that overhauls the classical framework. A thorough explanation of these effects require advanced concepts in quantum mechanics, and we shall defer it to later chapters.

1.4. Basic ideas of quantum mechanics

Physicists eventually came to realize a new framework that agrees with all experimental results and supersedes classical physics: **quantum mechanics**. Instead of regarding quantities such as position, velocity, and fields to be the fundamental quantities, quantum mechanics postulates a **complex wavefunction** as the fundamental quantity in physics. The wavefunction is a function of classical physics quantities such as position and momentum, and our observations of the world are linked to the wavefunction only indirectly through probabilities: whenever we measure a physical quantity, such as the position of an electron or a photon, the outcome is uncertain but it obeys a **probability distribution that is the magnitude squared of the wavefunction**. The wavefunction evolves in time through **Schrödinger's equation**, which is analogous to Maxwell's equations but actually a bit simpler mathematically if we study its nonrelativistic limit for one particle.



FIGURE 1.9. Consider a ball in either of the two cups. The position x has two possibilities: $x = 0$ (left cup) or $x = 1$ (right). A basic idea of quantum mechanics is to assign a complex number, called a wavefunction, to each possibility.

1.4.1. Wavefunction. To understand the basic concepts of quantum mechanics, here we shall strip quantum mechanics to its bare bones by considering again one particle in only two possible positions: left ($x = 0$) or right ($x = 1$), as depicted in Fig. 1.9. x is called a quantum **observable**. In general, there are many other types of observables that may have more possibilities or are continuous variables, but we shall focus on the binary case to illustrate the basic concepts. The first basic idea of quantum mechanics is to assign a set of **complex** numbers $\psi(x = 0)$ and $\psi(x = 1)$ to the possibilities. $\psi(x)$ is called a **wavefunction** of x . This assignment is like the assignment of probabilities to x , except that a probability distribution is **nonnegative**, while a wavefunction is **complex**.

1.4.2. Evolution in time. As the system evolves with time, the wavefunction $\psi(x, t)$ also changes in time. In general, the evolution is **linear** and the wavefunction $\psi(x, t)$ at a later time is determined by the wavefunction at an earlier time $\psi(x, t_0)$ as follows:

$$\psi(x, t) = \sum_{x_0} \psi(x, t|x_0, t_0) \psi(x_0, t_0), \quad (1.22)$$

where $\psi(x, t)$ is the wavefunction of x at time t , $\psi(x_0, t_0)$ is the wavefunction of $x = x_0$ at time t_0 , and $\psi(x, t|x_0, t_0)$ is a transition matrix that determines the evolution. We can write it in the matrix notation if we

want:

$$\begin{pmatrix} \psi(0, t) \\ \psi(1, t) \end{pmatrix} = \begin{pmatrix} \psi(0, t|0, t_0) & \psi(0, t|1, t_0) \\ \psi(1, t|0, t_0) & \psi(1, t|1, t_0) \end{pmatrix} \begin{pmatrix} \psi(0, t_0) \\ \psi(1, t_0) \end{pmatrix}, \quad (1.23)$$

and $\begin{pmatrix} \psi(0, t) \\ \psi(1, t) \end{pmatrix}$ is called the **quantum state** of x at time t . Eq. (1.22) looks like the transitions of probabilities in Eq. (1.19), except that now the transition amplitudes $\psi(x, t|x_0, t_0)$ are **complex**, and the notation $\psi(x, t|x_0, t_0)$ is written to make it resemble the notation for classical transition probabilities. We shall not worry too much about the specific form of the transition amplitudes for now and will just assume that they are given.

1.4.3. Measurement. Now consider a measurement of the particle position at time t . Measurement is done by a **macroscopic object**, for example, your eyes or a camera. The outcome is **random** and sometimes I will see the particle is on the left, sometimes I will see it's on the right. The probability distribution of this random outcome is equal to the magnitude squared of the wavefunction:

$$P(x, t) = |\psi(x, t)|^2. \quad (1.24)$$

This is called **Born's rule**. The wavefunction is normalized like the probability:

$$\sum_x |\psi(x, t)|^2 = \sum_x P(x, t) = 1, \quad (1.25)$$

which becomes also a normalization condition for the wavefunction at any time. This means that

$$|\psi(0, t)|^2 + |\psi(1, t)|^2 = |\psi(0, t_0)|^2 + |\psi(1, t_0)|^2, \quad (1.26)$$

and $\psi(x, t|x_0, t_0)$ must be a **unitary matrix**.

- **Exercise:** From the normalization condition for wavefunctions, prove that the transition amplitude function $\psi(x, t|x_0, t_0)$ must be a unitary matrix.

1.4.4. Wavefunction collapse. Although the measurement outcome can be random depending on the probability distribution $|\psi(x, t)|^2$, after the act of measurement and the outcome is known, the wavefunction **collapses** to match the outcome, such that, if I get outcome $x = 0$, the wavefunction becomes $\psi(0) = 1$, $\psi(1) = 0$ after the measurement, and if I get outcome $x = 1$, the wavefunction becomes $\psi(0) = 0$, $\psi(1) = 1$ instead, as depicted in Fig. 1.10. A subsequent measurement immediately after the first will produce the same outcome as the first.

1.4.5. Incompatible observables. To be concrete, consider now the **spin** of an electron. Classically, we can think of an electron spin heuristically as a self-rotation of the electron, like the rotation of the earth around the north-pole-south-pole axis. This spin can be represented by a vector in real three-dimensional space; for earth it is a vector pointing in the north pole direction, as shown in Fig. 1.11. The magnitude of the spin vector is determined by the **angular momentum** of the rotation.

The spin vector of an electron has three components:

$$\mathbf{s} = s_x \hat{\mathbf{x}} + s_y \hat{\mathbf{y}} + s_z \hat{\mathbf{z}}. \quad (1.27)$$

Let us focus on the component s_z first. It turns out that, for an electron, s_z can only have two possible values.¹ We shall not worry about its magnitude for now and just say it is either $s_z = 1/2$ or $s_z = -1/2$. According to quantum mechanics, there is a wavefunction associated with this binary observable given by $\psi_z(s_z)$. It evolves in time according to Eq. (1.22), and also obeys the rules for quantum measurement and wavefunction collapse like what we discussed earlier.

The new feature to be introduced here is the wavefunctions for the other two spin components $\psi_x(s_x)$ and $\psi_y(s_y)$. For some very deep reason [1], it turns out that **these wavefunctions depend on one another**. Suppose

¹There is a very deep reason in relativistic quantum field theory behind why a spin component of an electron has two possible values [1].

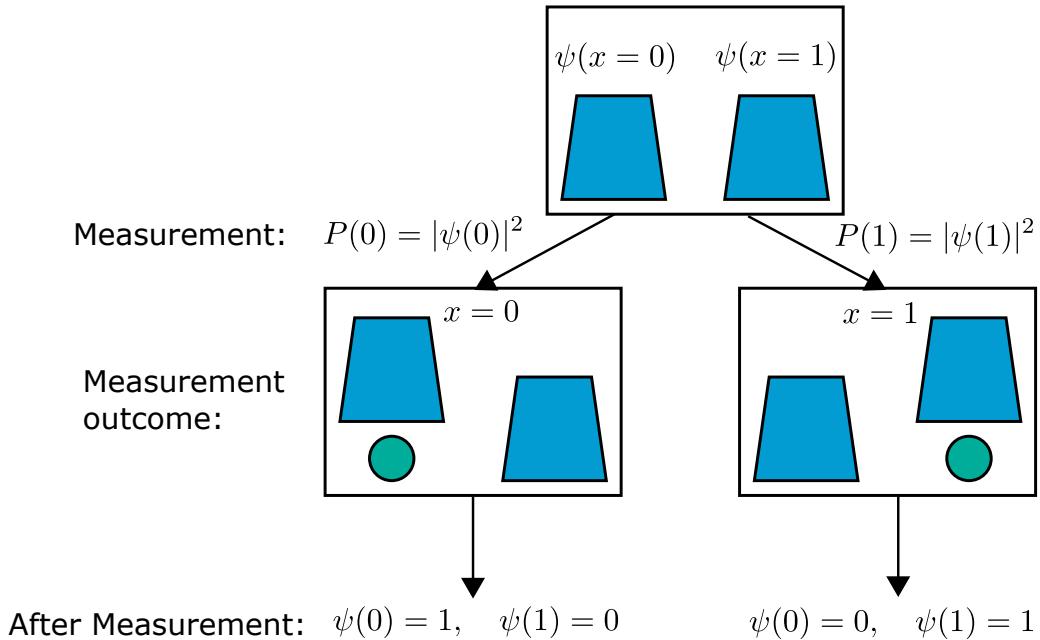


FIGURE 1.10. Measurement and wavefunction collapse. Before the measurement, the outcome is uncertain, and the probability distribution is given by $|\psi(x)|^2$. After the measurement, the wavefunction collapses such that a subsequent measurement immediately after the first will produce the same outcome as the first.

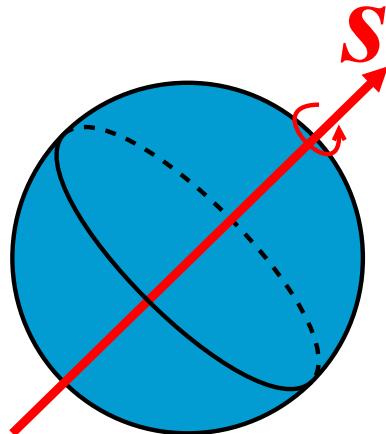


FIGURE 1.11. The spin of the earth and the associated spin vector s , which points in the direction of the north pole and has the magnitude equal to the angular momentum associated with the rotation. The electron spin can be visualized this way.

that $\psi_z(s_z)$ is given. Then

$$\begin{pmatrix} \psi_x \left(s_x = +\frac{1}{2} \right) \\ \psi_x \left(s_x = -\frac{1}{2} \right) \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \psi_z \left(s_z = +\frac{1}{2} \right) \\ \psi_z \left(s_z = -\frac{1}{2} \right) \end{pmatrix}, \quad (1.28)$$

such that the probability distribution when one measures the observable s_x is $|\psi_x(s_x)|^2$, and

$$\begin{pmatrix} \psi_y\left(s_y = +\frac{1}{2}\right) \\ \psi_y\left(s_y = -\frac{1}{2}\right) \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -i/\sqrt{2} \\ 1/\sqrt{2} & i/\sqrt{2} \end{pmatrix} \begin{pmatrix} \psi_z\left(s_z = +\frac{1}{2}\right) \\ \psi_z\left(s_z = -\frac{1}{2}\right) \end{pmatrix}, \quad (1.29)$$

such that the probability distribution for the measurement of s_y is $P_y(s_y) = |\psi_y(s_y)|^2$. Each wavefunction is related to another through a **unitary matrix**. Given $\psi_x(s_x)$, for example, one can invert the unitary matrix above to get back $\psi_z(s_z)$ and then get $\psi_y(s_y)$. s_x , s_y , and s_z are called **incompatible observables** because their wavefunctions and therefore their probability distributions are interrelated, one can measure one of the observables at each time but never simultaneously. Quantum mechanics specifically prohibits the simultaneous measurements and simultaneous precise knowledge about any pair of incompatible observables.

- **Exercise:** Given the wavefunction $\psi_x(s_x)$, express $\psi_y(s_y)$ and $\psi_z(s_z)$ in terms of it. Similarly, express $\psi_x(s_x)$ and $\psi_z(s_z)$ in terms of $\psi_y(s_y)$.

For example, suppose

$$\begin{pmatrix} \psi_z\left(s_z = +\frac{1}{2}\right) \\ \psi_z\left(s_z = -\frac{1}{2}\right) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (1.30)$$

which means that, if I measure s_z I will for sure get $s_z = +1/2$. What if I instead measure s_x of this state?

$$\begin{pmatrix} \psi_x\left(s_x = +\frac{1}{2}\right) \\ \psi_x\left(s_x = -\frac{1}{2}\right) \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}. \quad (1.31)$$

Now I will get a $|1/\sqrt{2}|^2 = 50\%$ chance of obtaining $s_x = +1/2$, and a 50% chance of obtaining $s_x = -1/2$. The precise knowledge of one observable thus leads to a reduction of knowledge about the other incompatible observables.

1.5. Continuous variables

So far we have been studying binary observables to spell out the basic ideas of quantum mechanics. In the next few chapters we shall focus on observables with continuous possible values instead, such as the three dimensional position r of an electron. Despite all the nontrivial concepts we have introduced earlier, the wavefunction in three space dimensions turns out to behave in a remarkably similar way to complex solutions for electromagnetic waves, as we shall see later in the next chapters.

1.5.1. One-dimensional position. Considering a one-dimensional continuous position x for now, the wavefunction is still written as $\psi(x, t)$, but the probability distribution for a continuous variable is defined slightly differently in terms of a **probability density**:

$$P(x, t) = |\psi(x, t)|^2. \quad (1.32)$$

The concept of a probability density is somewhat like charge density or energy density, except that now the infinitesimal value

$$P_x(x, t)dx = \text{probability of observing } x \text{ in } [x, x + dx] \quad (1.33)$$

is the probability of observing x within an infinitesimal length, as depicted in Fig. 1.12. To find the probability of observing x within an interval, one simply sums up all the infinitesimal probabilities in the form of an integral:

$$\int_a^b P_x(x, t)dx = \text{probability of finding } x \text{ in } [a, b], \quad (1.34)$$

as demonstrated in Fig. 1.13. The normalization condition is hence

$$\int_{-\infty}^{\infty} P_x(x, t)dx = 1. \quad (1.35)$$

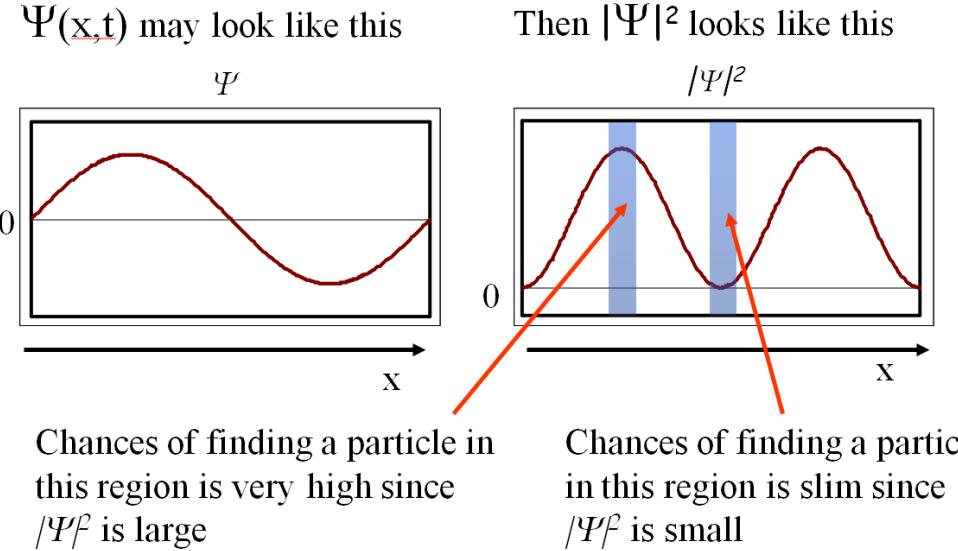


FIGURE 1.12. $|\psi(x,t)|^2$ is the probability density, meaning that $|\psi(x,t)|^2 dx$ is the probability of observing x in an infinitesimal interval.

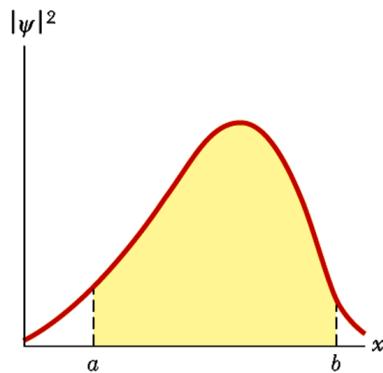


FIGURE 1.13. The probability of observing x within an interval can be obtained by integrating the probability density over that interval.

- **Exercise:** Consider a wavefunction in terms of the three-dimensional position r . Write down a reasonable relation between the probability density in three dimensions and the wavefunction and explain its meaning and the normalization condition in terms of the probability of observing r in a volume.

1.5.2. Momentum. It turns out that position and momentum are incompatible observables, and the momentum wavefunction is the **Fourier transform** of the position wavefunction:

$$\phi(k, t) = \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp(-ikx) \psi(x, t). \quad (1.36)$$

This kind of transform is a continuous generalization of a unitary matrix multiplication, like how Fourier transform is a continuous generalization of Fourier series. The probability density for the measurement of the momentum observable is then

$$P_k(k, t) = |\phi(k, t)|^2, \quad \int_{-\infty}^{\infty} dk P_k(k, t) = \int_{-\infty}^{\infty} dk |\phi(k, t)|^2 = 1. \quad (1.37)$$

Since the position and momentum are incompatible, the laws of quantum mechanics prohibit their simultaneous measurements. They also obey an **uncertainty relation**: if the position is known precisely, the momentum has

a large uncertainty, and vice versa. k here has the unit of m^{-1} . In mks units, the momentum is

$$p = \hbar k, \quad (1.38)$$

where \hbar (“h-bar”) is the **Planck’s constant**. Planck’s constant is usually defined as

$$h \approx 6.626 \times 10^{-34} \text{ J/s}, \quad (1.39)$$

but more often than not we will use \hbar instead, which is

$$\hbar = \frac{h}{2\pi}. \quad (1.40)$$

- **Exercise:** Given a momentum wavefunction $\phi(k)$, how do you determine the position wavefunction $\psi(x)$?
- **Exercise:** Given a wavefunction $\psi(r)$ with respect to a three-dimensional position r , guess the form of the momentum wavefunction $\phi(k)$ and determine its normalization condition.

1.5.3. *Position-momentum uncertainty relations. To illustrate the uncertainty relations between position and momentum, suppose that

$$\psi(x) \propto \delta(x - x_0), \quad (1.41)$$

where δ is the Dirac delta function centered at x_0 . Never mind that the probability density is unnormalizable, let’s focus on the qualitative picture and the spread of the wavefunctions here. Since $\delta(x - x_0) = \infty$ when $x = x_0$ and is zero when $x \neq x_0$, the probability density is sharply peaked at x_0 ; we have precise knowledge about the position.

The momentum wavefunction is

$$\phi(k) \propto \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp(-ikx) \delta(x - x_0) = \frac{1}{\sqrt{2\pi}} \exp(-ikx_0), \quad (1.42)$$

leading to a constant momentum distribution:

$$|\phi(k)|^2 \propto 1. \quad (1.43)$$

This says that the momentum distribution has infinite width, or in other words, all momenta have equal probabilities, and the momentum is thus completely random when the position is precise.

Suppose instead that

$$\psi(x) \propto \exp(ik_0x), \quad (1.44)$$

which looks like the spatial part of a plane wave solution in electromagnetics. $|\psi(x)|^2 \propto 1$, all positions have equal probabilities and the position is thus completely random, but now

$$\phi(k) \propto \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} \exp(-ikx) \exp(ik_0x) = \sqrt{2\pi} \delta(k - k_0), \quad (1.45)$$

the momentum distribution is peaked at $k = k_0$, and the momentum thus has a definite value at k_0 .

More generally, define the standard deviation of $|\psi(x)|^2$ and the standard deviation of $|\phi(k)|^2$ as

$$\Delta x \equiv \left[\int_{-\infty}^{\infty} dx (x - \langle x \rangle)^2 |\psi(x)|^2 \right]^{\frac{1}{2}}, \quad \Delta k \equiv \left[\int_{-\infty}^{\infty} dk (k - \langle k \rangle)^2 |\phi(k)|^2 \right]^{\frac{1}{2}}, \quad (1.46)$$

$$\langle x \rangle \equiv \int_{-\infty}^{\infty} dx x |\psi(x)|^2, \quad \langle k \rangle \equiv \int_{-\infty}^{\infty} dk k |\phi(k)|^2. \quad (1.47)$$

It can be proved (http://en.wikipedia.org/wiki/Uncertainty_principle#Robertson_E2.80.93Schr.C3.B6dinger_uncertainty_relations) that, for any $\psi(x)$,

$$\Delta x \Delta k \geq \frac{1}{2}. \quad (1.48)$$

The standard deviation is a measure of uncertainty, and the product of these position and momentum uncertainties cannot go below a certain value.

- **Exercise:** What is the position-momentum uncertainty relation in MKS units?

Answer:

$$\boxed{\Delta x \Delta p \geq \frac{\hbar}{2}.} \quad (1.49)$$

- **Exercise:** Express the uncertainty relations in terms of percentage errors with respect to the typical size and momentum of macroscopic objects, such as soccer balls and cars. Note how the right-hand side is an extremely small number. Now do the same for an electron with respect to the size of a modern-day transistor and typical electron momentum inside a semiconductor. Use the saturation velocity in silicon (http://en.wikipedia.org/wiki/Velocity_saturation) to compute the typical momentum. Using Moore's law (http://en.wikipedia.org/wiki/Moore%27s_law), estimate roughly when you expect to see the uncertainty relation to become important in transistors. What effects do you think the uncertainty relation will have on the operation of transistors? For example, think about the requirement on Δx such that the electron doesn't leak out of the transistor, and think about the relation between the electric current and the electron momentum and how the momentum uncertainty affects the measurement of current.

CHAPTER 2

Schrödinger Equation

2.1. Free particle in three space dimensions

We will now focus on the wavefunction $\psi(\mathbf{r}, t)$ of a nonrelativistic particle in three space dimensions. \mathbf{r} is the three-dimensional position vector, as shown in Fig. 2.1.

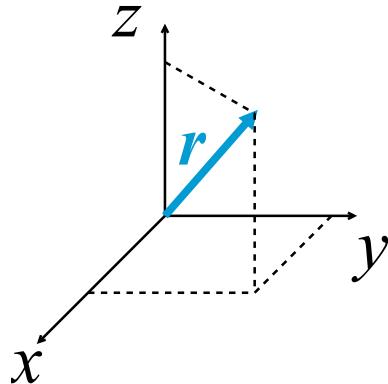


FIGURE 2.1. Position vector \mathbf{r} in three dimensions.

In free space, the wavefunction obeys the following **Schrödinger equation**:

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi, \quad (2.1)$$

where $\hbar = h/(2\pi)$ is the Planck's constant we just introduced in the previous chapter, m is the mass of the particle, and ∇^2 is the Laplacian operator that we encountered in electromagnetics:

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (2.2)$$

Like the electromagnetic wave equation, Schrödinger equation is also written as a partial differential equation, except that the latter involves the first-order partial time derivative and ψ is a complex scalar, not a real vector. $|\psi(\mathbf{r}, t)|^2$ is the **probability density** of finding the particle at position \mathbf{r} . This means that

$$|\psi(\mathbf{r}, t)|^2 dx dy dz = \text{probability of finding the particle inside a box with volume } dx dy dz \text{ at position } \mathbf{r}, \quad (2.3)$$

where the box is pictured in Fig. 2.2. Like the charge density or energy density, the total probability of finding the particle inside a volume is given by a volume integral:

$$\text{Probability of finding the particle inside a volume} = \int_{\text{volume}} |\psi(\mathbf{r}, t)|^2 dx dy dz, \quad (2.4)$$

as depicted in Fig. 2.3.

2.2. Single-frequency plane wave

A useful solution of Eq. (2.1) is the **single-frequency plane wave** solution:¹

$$\psi = \tilde{\psi} \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t). \quad (2.5)$$

¹I now avoid the term monochromatic because the frequency of a quantum wavefunction is not associated with color.

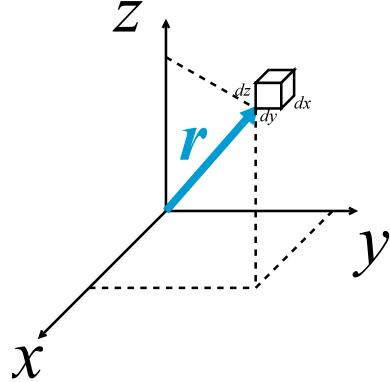
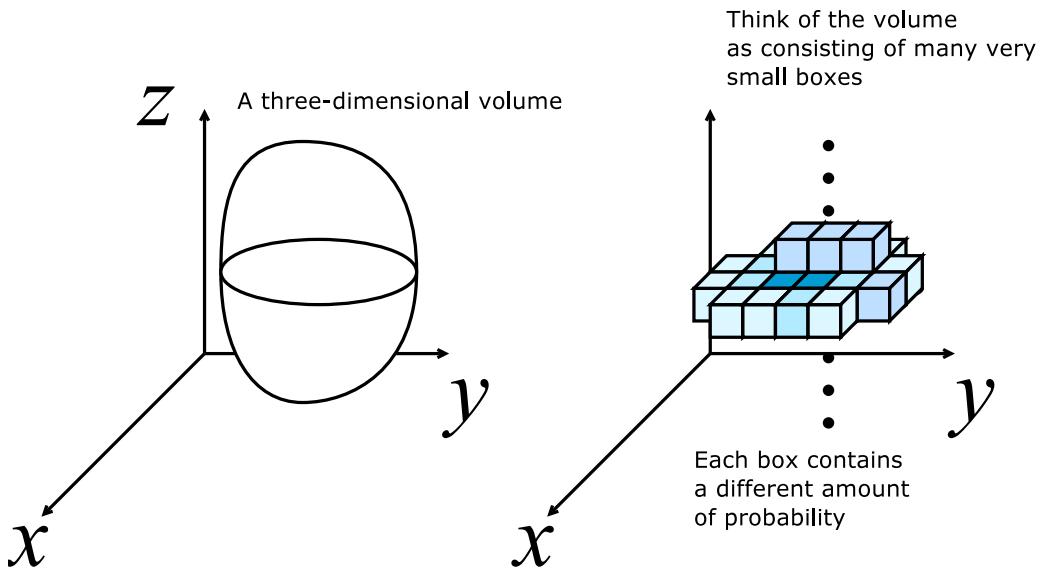
FIGURE 2.2. A box with infinitesimal volume at position \mathbf{r} .

FIGURE 2.3. The probability of finding the particle in a volume is the volume integral of the probability density.

This should remind you about the monochromatic plane-wave solution in electromagnetics. Recall that \mathbf{k} is the wavevector, and its direction denotes the propagation direction, as shown in Fig. 2.4. $\tilde{\psi}$ is the complex amplitude of the plane wave. In quantum mechanics we no longer need to take the real part, to link to experiments we just transform it to a wavefunction with respect to the observable under measurement and take the magnitude squared for the probability density.

To verify that Eq. (2.5) is the solution of the Schrödinger equation, let's plug it into Eq. (2.1). The left-hand side is

$$i\hbar \frac{\partial \psi}{\partial t} = i\hbar(-i\omega)\tilde{\psi} \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t), \quad (2.6)$$

and the right-hand side is

$$-\frac{\hbar^2}{2m} \nabla^2 \psi = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \tilde{\psi} \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t) \quad (2.7)$$

$$= \frac{\hbar^2}{2m} (k_x^2 + k_y^2 + k_z^2) \tilde{\psi} \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t). \quad (2.8)$$

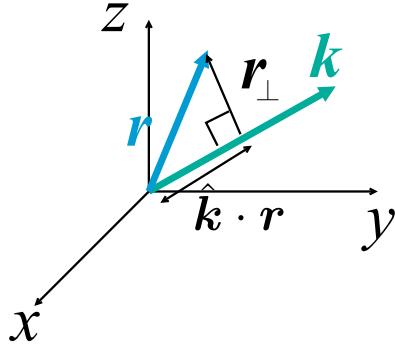


FIGURE 2.4. The plane wave propagates in the direction of \mathbf{k} . The period of the wave along \mathbf{k} is the wavelength and given by $\lambda = 2\pi/k$. The wave is constant along any direction perpendicular to the wavevector.

By now I hope you are familiar with the steps involved. Equating the left-hand side and the right-hand side of the Schrödinger equation,

$$\hbar\omega\tilde{\psi}\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega t)=\frac{\hbar^2}{2m}(k_x^2+k_y^2+k_z^2)\tilde{\psi}\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega t), \quad (2.9)$$

and dividing by $\tilde{\psi}\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega t)$,

$$\boxed{\hbar\omega=\frac{\hbar^2}{2m}(k_x^2+k_y^2+k_z^2)}. \quad (2.10)$$

This is a **dispersion relation** between the wave oscillation in time, as specified by ω , and the wave oscillation in space, as specified by the magnitude of \mathbf{k} .

2.3. *Probability current

Analogous to the current density \mathbf{J} for charged particles or the Poynting vector \mathbf{S} for electromagnetic waves, we can also define a probability current density as follows:

$$\mathbf{J}=\frac{\hbar}{2mi}(\psi^*\nabla\psi-\psi\nabla\psi^*). \quad (2.11)$$

It has the unit of probability per area per second. If you integrate this vector field using a surface integral $\int \mathbf{J} \cdot d\mathbf{A}$, you get the probability current (probability per second) through the surface.

- **Exercise:** Prove a continuity equation for probability density $\rho=|\psi|^2$ and \mathbf{J} using the free-particle Schrödinger equation. The answer is

$$\frac{\partial\rho}{\partial t}+\nabla\cdot\mathbf{J}=0. \quad (2.12)$$

This is a statement about probability conservation and confirms that \mathbf{J} is indeed the current density.

- **Exercise:** For a plane wave $\tilde{\psi}\exp(i\mathbf{k}\cdot\mathbf{r}-i\omega t)$, show that the current density is

$$\frac{\hbar\mathbf{k}}{m}|\tilde{\psi}|^2. \quad (2.13)$$

This is similar to the charge current density we studied in Part I: the current density is the velocity (group velocity $\hbar\mathbf{k}/m$) times a “stuff” density $|\tilde{\psi}|^2$

2.4. Correspondence with classical physics

A wavefunction that oscillates in time $\propto \exp(-i\omega t)$ is associated with a definite energy given by

$$\boxed{E=\hbar\omega.} \quad (2.14)$$

A wavefunction that oscillates in space $\propto \exp(i\mathbf{k}\cdot\mathbf{r})$ is associated with a definite momentum given by

$$\boxed{\mathbf{p}=\hbar\mathbf{k}.} \quad (2.15)$$

The wavefunction is nonzero everywhere in space and time however, and the particle is delocalized in space and time. Strictly speaking, this plane-wave solution is not normalizable, meaning that the volume integral over all space is always infinite unless $\tilde{\psi} = 0$, but it often serves as a good approximation as well as building block of more complicated solutions in quantum mechanics, just like the role of plane waves in electromagnetics.

The dispersion relation in Eq. (2.10) that links ω and \mathbf{k} then also links the energy and momentum as follows:

$$\boxed{E = \hbar\omega = \frac{\hbar^2 k^2}{2m} = \frac{p^2}{2m}}, \quad (2.16)$$

where k is the magnitude of \mathbf{k} and p is the magnitude of \mathbf{p} . It agrees with the kinetic energy of a nonrelativistic particle.

The wavelength λ of a single-frequency plane wave is its spatial period. It is given by

$$\lambda = \frac{2\pi}{k}. \quad (2.17)$$

Since k is related to the momentum, it follows that a quantum particle with a certain momentum will have a wavelength λ for its wavefunction. In terms of the momentum magnitude p ,

$$\boxed{\lambda = \frac{2\pi\hbar}{p} = \frac{h}{p}}. \quad (2.18)$$

This is known as the **de Broglie wavelength**. It associates a classical particle property (momentum) to a wave property (wavelength). Note the inverse relation: the higher the momentum, the smaller the period.

- *Exercise: Consider a plane-wave solution

$$\psi(\mathbf{r}, t) \propto \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t). \quad (2.19)$$

The energy is $E = \hbar\omega$, and momentum is $\hbar\mathbf{k}$. Given the relativistic energy-momentum relation $E^2 = p^2c^2 + m^2c^4$ and assuming that $E \geq 0$, write down the relation between ω and the magnitude of \mathbf{k} . Assume the Newtonian limit $p^2c^2 \ll m^2c^4$ and express E as a quadratic function with respect to p . Show that E is a sum of rest energy and kinetic energy. Write ω as a function of k in this limit. Show that, if we consider only the excess energy above the rest energy, we get back the dispersion relation from the free-particle Schrödinger equation.

Answer: For nonnegative E , we can write

$$E = \sqrt{p^2c^2 + m^2c^4}, \quad \hbar\omega = \sqrt{\hbar^2k^2c^2 + m^2c^4}, \quad \omega = \sqrt{k^2c^2 + \frac{m^2c^4}{\hbar^2}}. \quad (2.20)$$

(Note that we get back the dispersion relation for electromagnetic waves when $m = 0$). Let us rearranging the right-hand side as follows:

$$E = \sqrt{m^2c^4 \left(1 + \frac{p^2c^2}{m^2c^4} \right)} = mc^2 \sqrt{1 + \frac{p^2}{m^2c^2}}. \quad (2.21)$$

If $p \ll mc$, we can use the first two terms in the Taylor series ($\sqrt{1 + \epsilon} \approx 1 + \epsilon/2$ for $\epsilon \ll 1$) to obtain

$$E \approx mc^2 \left(1 + \frac{p^2}{2m^2c^2} \right) = mc^2 + \frac{p^2}{2m}, \quad (2.22)$$

which consists of the rest energy mc^2 and kinetic energy $p^2/(2m)$. Substituting $E = \hbar\omega$ and $p = \hbar k$ gets us the dispersion relation for a nonrelativistic massive particle.

2.5. Electron diffraction

Spherical and cylindrical wave solutions also exist for an electron wavefunction that obeys the free-particle Schrödinger equation. To compute the diffraction pattern, we can use the same procedures we have studied in terms of Fresnel and Fraunhofer diffraction and sum up all the cylindrical or spherical waves that come from point sources or slits on a screen. Once we have computed $\psi(\mathbf{r}, t)$ at some detector plane, we can just take the magnitude squared to get the probability density, like how we got the intensity under the paraxial approximation. An example of a double-slit interference is depicted in Fig. 2.5. Details about the experiment are in Fig. 2.6.

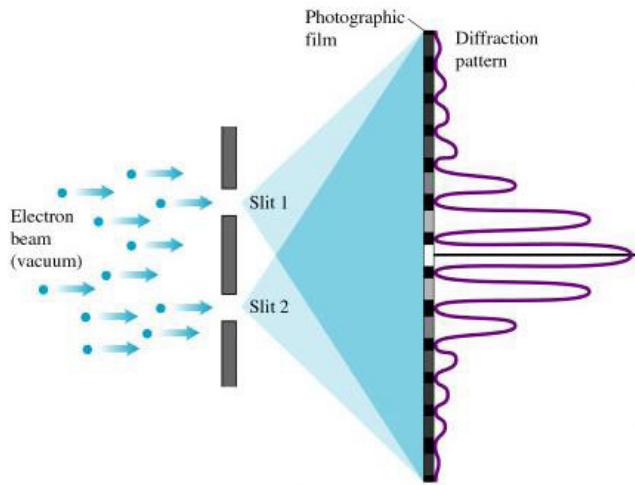


FIGURE 2.5. Electron double-slit interference. The resulting probability density is just like the optical intensity in an optical double-slit interference experiment, as they are both the magnitude squared of a complex wave solution and the resulting wave comes from the superposition of many cylindrical waves from the slits. Image from [2].

The buildup of interference pattern for electrons is known as **wave-particle** duality: each electron will diffract and interfere like waves when you are not looking, but if you measure its position, there is a wavefunction collapse and the outcome reveals that the electron is at a single position. You can obtain the interference pattern for the electron only after many measurements.

2.6. Electron microscopes

Recall that the focus spot size of an optical lens must always be somewhat larger than the wavelength λ . The resolution of microscopes using electrons as waves can be a lot higher than optical microscopes because electron wavelength can be much smaller and can focus a lot better.

- **Exercise:** Suppose that an electron is accelerated by a voltage $V = 1 \text{ kV}$. What is its wavelength?

Answer: The potential energy of an electron at the beginning of acceleration is $e \approx 1.6 \times 10^{-19} \text{ C}$ times the voltage, and after the acceleration is done it is converted to the kinetic energy, so the kinetic energy will be

$$E = eV \approx (1.6 \times 10^{-19} \text{ C}) (1 \text{ kV}). \quad (2.24)$$

Given this energy, we can use the relation given by Eq. (2.16) to calculate the momentum magnitude:

$$p = \sqrt{2mE}, \quad (2.25)$$

where the electron mass is $m = 9.11 \times 10^{-31} \text{ kg}$. From Eq. (2.18), the de Broglie wavelength becomes

$$\lambda = \frac{h}{p} \sim 4 \times 10^{-11} \text{ m}, \quad (2.26)$$

which is much smaller than the typical optical wavelength $\sim 500 \text{ nm}$ in an optical microscope.

Two examples are the scanning electron microscope (SEM) and the transmission electron microscope (TEM), as shown in Fig. 2.7.

Some nice images from an electron microscope can be found at <http://www.iflscience.com/technology/some-spectacular-sem-images-microscopic-world>. A movie that discusses the principles of SEM can be found at <http://legacy.mos.org/sln/sem/sem.mov>.

2.7. *General solution of the free-particle Schrödinger equation

In the previous chapter, we mentioned that a wavefunction evolves in time through a unitary matrix, but so far we have just written down a differential equation given by Eq. (2.1) and haven't shown how the unitarity

- (1) Assume that each electron passes through the slits one by one, and before the slits each electron has the plane wave solution.
- (2) The screen blocks the electron and only allows it to go through the slits. For simplicity, let's consider two slits much narrower than the wavelength. From each slit, a cylindrical wave comes out, and the electron wavefunction is a superposition of the two cylindrical waves from the two slits.
- (3) **Note that the electron is interfering with itself.** Each cylindrical wave is simply a different component of the wavefunction for **one electron**.
- (4) At the detector screen, the probability density of the electron is the magnitude squared of the wavefunction. The probability density in theory is

$$|\psi(x, z, t)|^2 \propto \cos^2\left(\frac{\pi x d}{\lambda z}\right), \quad (2.23)$$

where d is the distance between the two slits, λ is the (de Broglie) wavelength, and z is the propagation distance from the slits to the film. This result should not surprise you, as it comes from the same considerations as optical diffraction.

- (5) The detector screen is effectively measuring and indicating the position of the electron by a white dot. According to the probability density, there will be places where the electron is more likely to arrive (constructive interference) and places where the electron is not going to arrive (destructive interference).
- (6) At the beginning, a few white dots show up, indicating the random outcomes of such measurements. But as more and more electrons are recorded, more electrons should arrive at places where each is more likely to arrive and less electrons at places where each is less likely to arrive.
- (7) The final distribution reproduces the probability density predicted from the wave theory and quantum measurement.

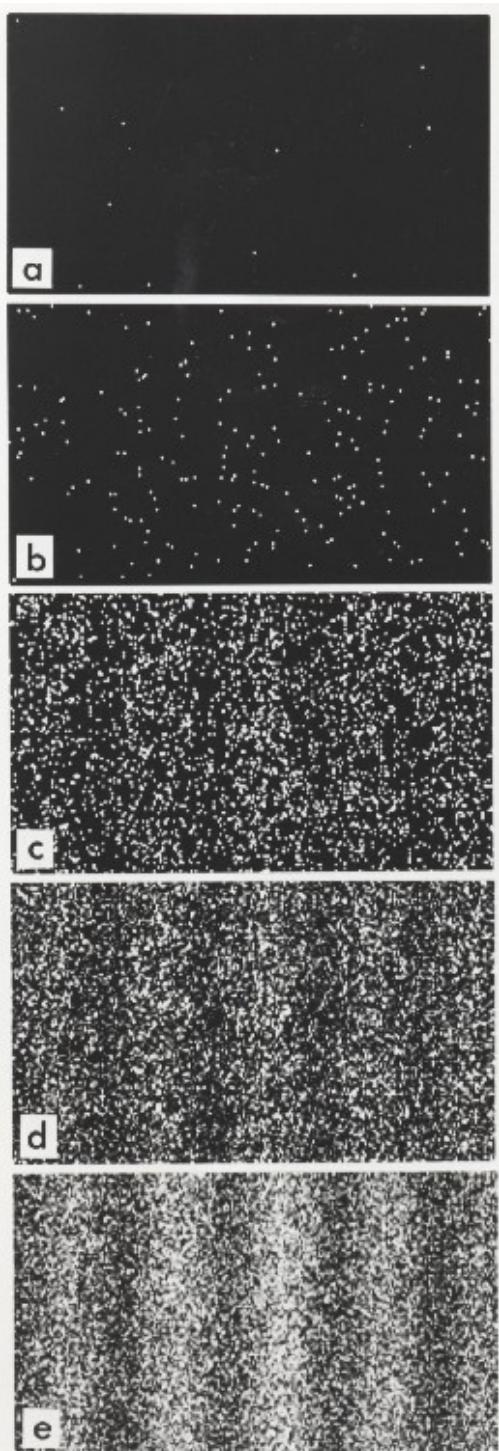


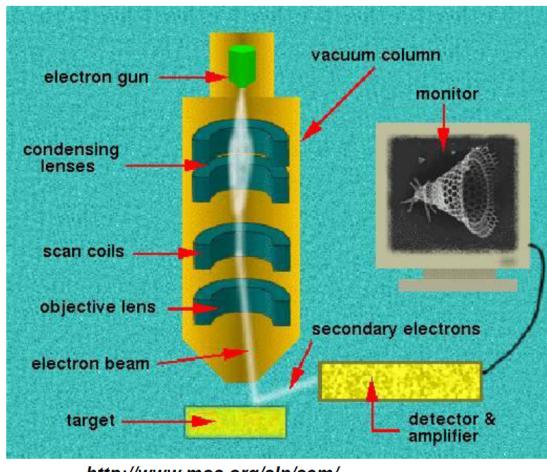
FIGURE 2.6. An electron double-slit experiment, as reported in [3] with results shown in the figure on the right.

comes into play. The general solution of Eq. (2.1) turns out to be

$$\psi(\mathbf{r}, t) = \int_{\text{all space}} d^3 r_0 \left(\frac{m}{2\pi i \hbar t} \right)^{3/2} \exp\left(\frac{im}{2\hbar t} |\mathbf{r} - \mathbf{r}_0|^2 \right) \psi(\mathbf{r}_0, 0), \quad (2.27)$$

which can be obtained using Fourier transform techniques. This transform is unitary and a continuous-variable generalization of a unitary matrix.

It can be shown that, in general, the differential form of the Schrödinger equation will always lead to a unitary transform that conserves probability, but in general the unitary transform is very difficult to deal with



See the animation on the working principle of a Scanning Electron Microscope



- 1986 Physics Nobel Prize
- Ernst Ruska

“for his fundamental work in electron optics, and for the design of first electron microscope”

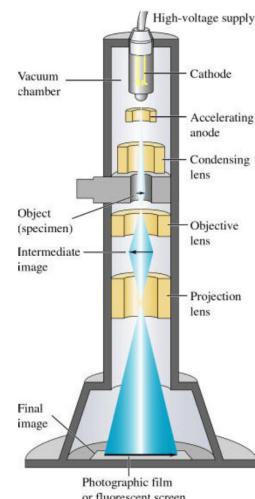


FIGURE 2.7. Scanning and transmission electron microscopes. From a PC2232 AY13/14 slide.

beyond the free-particle case (unless you use the Feynman path integral) and we usually prefer the differential form.

CHAPTER 3

Schrödinger Equation with Potential

The free-particle Schrödinger equation is important if we want to study electron microscopes, but our main interest in this module is to study the behavior of electrons **confined** by the attractive forces from the positively charged nuclei in a material, such as atoms, molecules, and in particular solids and semiconductors. In quantum mechanics, we have already seen how the time evolution of the wavefunction is related to the energy via $E = \hbar\omega$, and it suggests to us that we should focus on the **potential energy** due to the force in quantum mechanics.

Suppose that the force on a particle is given. We integrate it using a line integral in Eq. (1.5) to obtain the potential energy $U(\mathbf{r})$. Then the quantum particle in the presence of this potential will obey the following Schrödinger equation:

$$\boxed{i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + U\psi.} \quad (3.1)$$

This is similar to the free-particle version given by Eq. (2.1), but with one additional term that is simply the multiplication of the potential $U(\mathbf{r})$ with the wavefunction $\psi(\mathbf{r}, t)$.

- **Exercise:** Suppose that $\psi(\mathbf{r}, t)$ is a solution of the Schrödinger equation. What about $e^{i\theta}\psi(\mathbf{r}, t)$, where θ is a constant that doesn't depend on \mathbf{r} and t ? Does θ affect the probability distribution of any observable?

Answer: Since

$$\frac{\partial}{\partial t} (e^{i\theta}\psi) = e^{i\theta} \frac{\partial \psi}{\partial t}, \quad \nabla^2 (e^{i\theta}\psi) = e^{i\theta} \nabla^2 \psi, \quad U(e^{i\theta}\psi) = e^{i\theta} U\psi, \quad (3.2)$$

$e^{i\theta}\psi$ also satisfies Schrödinger equation if ψ does. To calculate the probability distribution for any observable, it involves a unitary transform \hat{u} on ψ (think of \hat{u} as a unitary matrix multiplying a vector ψ). But since $\hat{u}(e^{i\theta}\psi) = e^{i\theta}\hat{u}\psi$ (think of matrix multiplication: a constant $e^{i\theta}$ multiplying a vector ψ can be taken out of the product), the resulting probability distribution $\psi^\dagger \hat{u}^\dagger \hat{u}\psi$ doesn't depend on $e^{i\theta}$.

We have already mentioned that, if the wavefunction oscillates in time like $\psi \propto \exp(-i\omega t)$, the energy of the particle is equal to $\hbar\omega$. This association holds in general, and we can look for single-energy/single-frequency solutions of Eq. (3.1) by writing the wavefunction as a spatially varying part $\tilde{\psi}(\mathbf{r})$ and a single-frequency time dependence as follows:

$$\psi(\mathbf{r}, t) = \tilde{\psi}(\mathbf{r}) \exp(-i\omega t). \quad (3.3)$$

Substituting this into Eq. (3.1), we obtain

$$i\hbar \frac{\partial \psi}{\partial t} = i\hbar(-i\omega)\tilde{\psi}(\mathbf{r}) \exp(-i\omega t) \quad (3.4)$$

$$= \exp(-i\omega t) \left(-\frac{\hbar^2}{2m} \nabla^2 + U \right) \tilde{\psi}(\mathbf{r}). \quad (3.5)$$

The $\exp(-i\omega t)$ term appears on both sides and goes away, leaving an equation governing the spatially varying part of the wavefunction:

$$\boxed{\hbar\omega\tilde{\psi}(\mathbf{r}) = E\tilde{\psi}(\mathbf{r}) = \left[-\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}) \right] \tilde{\psi}(\mathbf{r}),} \quad (3.6)$$

which specifies the wavefunction for a given total energy E . Eq. (3.6) is known as the **time-independent** Schrödinger equation, as it depends only on the position \mathbf{r} while the time dependence has been separated from this equation.

- **Exercise:** Given $\psi_0(\mathbf{r}, t)$, which is a solution of

$$i\hbar \frac{\partial \psi_0}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi_0 + U(\mathbf{r}) \psi_0, \quad (3.7)$$

solve

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + U(\mathbf{r}) \psi + U_0 \psi, \quad (3.8)$$

where U_0 is a constant. If $\psi_0 \propto e^{-i\omega_0 t}$ and $\psi \propto e^{-i\omega t}$, how are their frequencies and energies related?

Answer: We rewrite Eq. (3.8) as

$$i\hbar \left(\frac{\partial}{\partial t} + i \frac{U_0}{\hbar} \right) \psi = -\frac{\hbar^2}{2m} \nabla^2 \psi + U(\mathbf{r}) \psi \quad (3.9)$$

and multiply $\exp(iU_0 t/\hbar)$ on both sides:

$$\exp \left(\frac{iU_0 t}{\hbar} \right) i\hbar \left(\frac{\partial}{\partial t} + i \frac{U_0}{\hbar} \right) \psi = \left[-\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}) \right] \psi \exp \left(\frac{iU_0 t}{\hbar} \right). \quad (3.10)$$

The left-hand side can be rewritten as

$$i\hbar \frac{\partial}{\partial t} \left[\exp \left(\frac{iU_0 t}{\hbar} \right) \psi \right] \quad (3.11)$$

as can be seen using the product rule of differentiation, so

$$i\hbar \frac{\partial}{\partial t} \left[\exp \left(\frac{iU_0 t}{\hbar} \right) \psi \right] = \left[-\frac{\hbar^2}{2m} \nabla^2 + U(\mathbf{r}) \right] \left[\psi \exp \left(\frac{iU_0 t}{\hbar} \right) \right]. \quad (3.12)$$

Lo and behold, this is the same as Eq. (3.7) if we equate

$$\psi_0 = \psi \exp \left(\frac{iU_0 t}{\hbar} \right). \quad (3.13)$$

In other words, if we know how to solve for ψ_0 , ψ is automatically given by $\psi_0 \exp(-iU_0 t/\hbar)$. This means that adding a constant U_0 to the Schrödinger equation will not change the solution, apart from a new factor of $\exp(-iU_0 t/\hbar)$. This factor would not appear in the probability distribution for any observable, so we are free to introduce any U_0 without changing the physics.

If $\psi_0 \propto e^{-i\omega_0 t}$, then

$$\psi \propto \exp(-i\omega_0 t) \exp \left(-\frac{iU_0 t}{\hbar} \right), \quad (3.14)$$

meaning that ψ would have a frequency at

$$\omega = \omega_0 + \frac{U_0}{\hbar}, \quad (3.15)$$

and energy of

$$E = \hbar\omega_0 + U_0, \quad (3.16)$$

which consists of the energy for ψ_0 plus a constant U_0 .

We will simplify Eq. (3.6) even further and assume $U(\mathbf{r}) = U(x)$ is a function of one coordinate x only, and the wavefunction doesn't change along y and z such that $\frac{\partial \psi}{\partial y} = 0$ and $\frac{\partial \psi}{\partial z} = 0$. Then $\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2}$, and we obtain a one-dimensional time-independent Schrödinger equation:

$$E \tilde{\psi}(x) = \left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + U(x) \right] \tilde{\psi}(x).$$

(3.17)

For brevity, we shall omit the tilde from now on.

3.1. Examples of potentials

3.1.1. Flat potential. Our very first example is a **flat potential** $U(r) = U_0$, where U_0 is a constant. The Schrödinger equation is

$$E\psi(x) = \left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U_0\right)\psi(x). \quad (3.18)$$

The plane-wave solution also works here:

$$\psi(x) = \tilde{\psi} \exp(ikx). \quad (3.19)$$

Substituting this into Eq. (3.18), we obtain a relation between the energy, the wavenumber k , and the potential U_0 :

$$E = \frac{\hbar^2 k^2}{2m} + U_0. \quad (3.20)$$

We find that the total energy E is the sum of the kinetic energy $\hbar^2 k^2 / (2m)$ and the potential energy U_0 , as one would expect from classical physics. Given the **linearity** of the Schrödinger equation, any superposition of solutions is also a solution, and we can write in general

$$\psi(x) = A \exp(ikx) + B \exp(-ikx), \quad (3.21)$$

$$k = \sqrt{\frac{2m}{\hbar^2}(E - U_0)}, \quad (3.22)$$

for a given energy E .

An interesting situation occurs if $E < U_0$, such that the total energy is smaller than the potential energy. In classical physics, this is not allowed, because the kinetic energy has to be nonnegative. In quantum mechanics, however, we can make k imaginary if $E < U_0$:

$$k = i\sqrt{\frac{2m}{\hbar^2}(U_0 - E)}. \quad (3.23)$$

The plane-wave solution becomes an **evanescent** solution:

$$\psi(x) = A \exp\left[-\sqrt{\frac{2m}{\hbar^2}(U_0 - E)}x\right] + B \exp\left[\sqrt{\frac{2m}{\hbar^2}(U_0 - E)}x\right], \quad (3.24)$$

which decays or grows exponentially.

3.1.2. Infinite square well. Our next example is the **infinite square well** plotted in Fig. 3.1:

$$U(x) = \begin{cases} 0, & 0 \leq x \leq L, \\ \infty, & x < 0 \text{ or } x > L, \end{cases} \quad (3.25)$$

which has infinite potential on the sides but flat zero potential within the well. To understand what it means, consider the force that a particle will experience:

$$F(x) = -\frac{\partial U}{\partial x} = \begin{cases} +\infty, & x = 0, \\ 0, & 0 < x < L, \\ -\infty, & x = L, \end{cases} \quad (3.26)$$

which means that the particle experiences an infinite force pushing it to the right at $x = 0$, an infinite force pushing it to the left at $x = L$, but no force inside the well. This is the situation for a particle between two **walls**: Right at each wall, there is a huge force pushing the particle away from the walls such that the particle cannot penetrate the wall, but otherwise the particle experiences no force between the two walls.

3.1.3. Finite square well. A more realistic situation is the **finite square well**, as plotted in Fig. 3.2:

$$U(x) = \begin{cases} 0, & 0 \leq x \leq L, \\ U_0, & x < 0 \text{ or } x > L. \end{cases} \quad (3.27)$$

There is a force pushing the particles at each wall, but now the potential beyond the walls is finite, meaning that a classical particle with a total energy higher than U_0 can have nonzero kinetic energy and exists outside the walls.

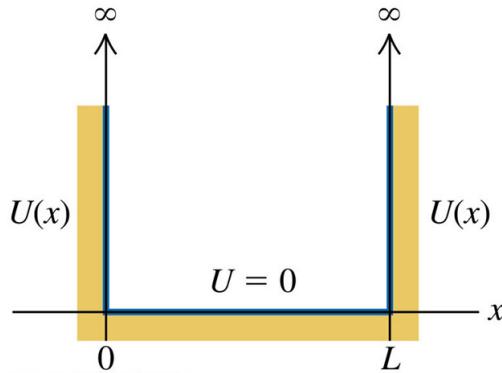


FIGURE 3.1. Infinite square well.

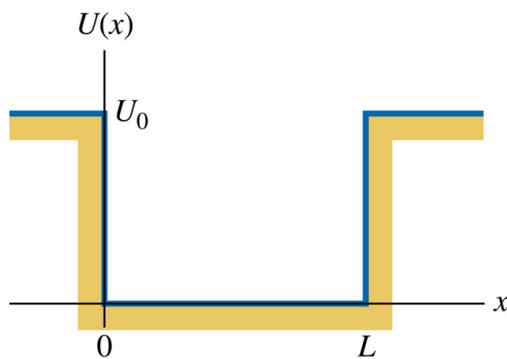


FIGURE 3.2. Finite square well.

3.1.4. Harmonic oscillator. Another important example of potential well is the **harmonic well**:

$$U(x) = \frac{1}{2}\kappa x^2, \quad (3.28)$$

which we encountered in Sec. 1.1.3. A particle under this potential is called a **harmonic oscillator**, and classically experiences a force that opposes the displacement and is proportional to it.

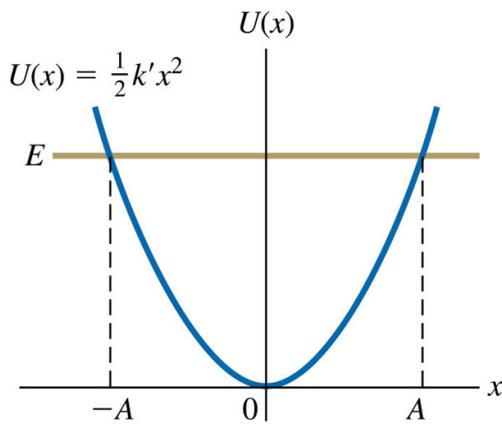


FIGURE 3.3.

3.1.5. Potential barrier. Our final example is the potential barrier, as plotted in Fig. 3.4:

$$U(x) = \begin{cases} U_0, & 0 \leq x \leq L, \\ 0, & x < 0 \text{ or } x > L. \end{cases} \quad (3.29)$$

where the forces at the two walls push the particle away from the barrier.

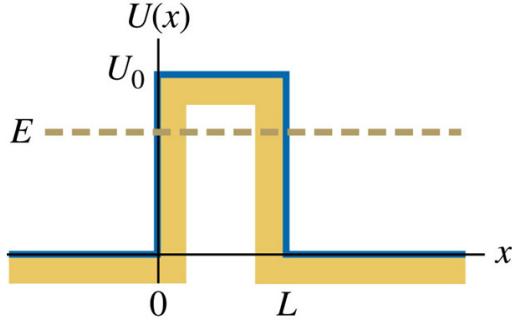


FIGURE 3.4. A square potential barrier.

3.2. Infinite square well

We shall now solve for the wave solution with the infinite square well plotted in Fig. 3.1. The potential function is

$$U(x) = \begin{cases} 0, & 0 \leq x \leq L, \\ \infty, & x < 0 \text{ or } x > L, \end{cases} \quad (3.30)$$

which is infinite outside the well, and that means that the wavefunction $\psi(x)$ must be **zero** there:

$$\psi(x) = 0, \quad x < 0, \quad x > L. \quad (3.31)$$

This fact comes from the evanescent wave solution in Eq. (3.24), which says that, for any finite energy E , $\psi(x)$ is either zero or infinity for $U_0 \rightarrow \infty$, and we have to choose the zero solution for the wavefunction to be normalizable.

The zero wavefunction outside the walls also sets **boundary conditions** for the wave inside the well. We know from Sec. 3.1.1 that a general solution inside the well, if $E > 0$, should be

$$\psi(x) = A \exp(ikx) + B \exp(-ikx), \quad 0 \leq x \leq L, \quad (3.32)$$

$$E = \frac{\hbar^2 k^2}{2m}, \quad (3.33)$$

for arbitrary complex constants A and B . This solution represents left-propagating and right-propagating waves inside the well. Since $\psi(x=0) = 0$, it sets the following boundary condition:

$$\psi(x=0) = A + B = 0. \quad (3.34)$$

We can visualize this boundary condition as a relation between the input amplitude B for the left-propagating wave and output amplitude A for the right-propagating wave, as shown on the left of Fig. 3.5. Eq. (3.34) then suggests that the **reflection coefficient** s_{22} , defined as the ratio between the output and input amplitudes, is

$$s_{22} \equiv \frac{A}{B} = -1. \quad (3.35)$$

Since the wavefunction is zero beyond the wall, the transmission coefficient is 0 naturally.

We can now do the same for the wall on the right-hand side, as shown on the right of Fig. 3.5. At $x = L$, the boundary condition is

$$\psi(x=L) = A \exp(ikL) + B \exp(-ikL) = 0. \quad (3.36)$$

The input is now $A \exp(ikL)$ and the output is $B \exp(-ikL)$, and the input and output amplitudes have acquired phase factors due to propagation. The reflection coefficient is also -1 there:

$$\frac{B e^{-ikL}}{A e^{ikL}} = s_{22} = -1. \quad (3.37)$$

Combining Eq. (3.34) and (3.37), we obtain a condition on k :

$$s_{22}^2 \exp(2ikL) = 1. \quad (3.38)$$

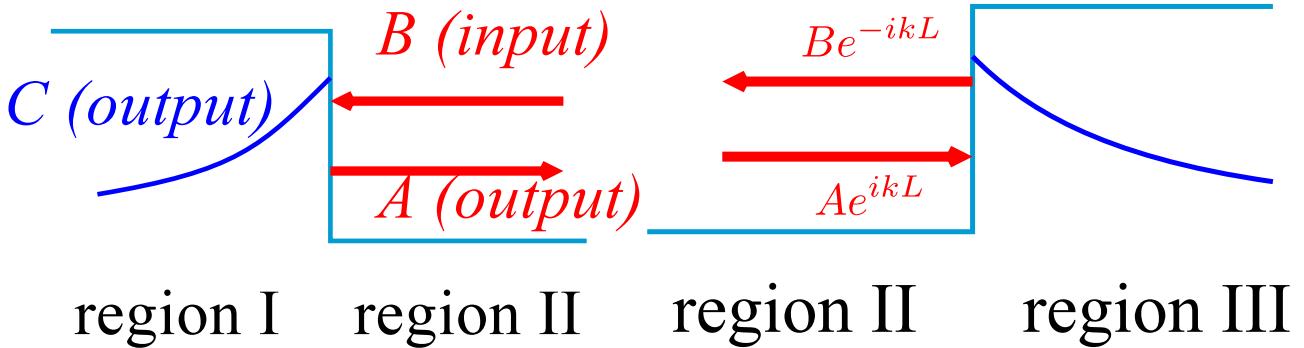


FIGURE 3.5. Left: Input-output analysis for the left wall. There is zero output beyond the wall for an infinite square well. Right: Input-output analysis for the right wall.

Just like the closed optical cavity with perfectly conducting mirrors we studied in Part I, this condition is a statement about the **round-trip factor** $s_{22}^2 \exp(2ikL)$ that the wave picks up by completing a round trip inside the well. This round-trip factor has to be 1 for our assumed solution to hold. Since $s_{22} = -1$, it means that

$$\exp(2ikL) = 1, \quad (3.39)$$

and the phase of the complex exponential has to be integer-multiples of 2π :

$$2kL = 2\pi n, \quad n \text{ integer.} \quad (3.40)$$

Given L , there are discrete wavenumbers that satisfy this relation:

$$k = \frac{\pi n}{L}, \quad n = 1, 2, \dots \quad (3.41)$$

We assume n to be positive because $\psi = 0$ if $n = 0$ and negative n turns out to give the same wavefunction as a positive n . The wavefunction becomes

$$\psi(x) = A \exp(ikx) + B \exp(-ikx) = A [\exp(ikx) - \exp(-ikx)] = 2iA \sin(kx), \quad 0 \leq x \leq L. \quad (3.42)$$

A can then be determined by the normalization condition:

$$\int_0^L dx |\psi(x)|^2 = 4|A|^2 \int_0^L dx \sin^2(kx) = 1. \quad (3.43)$$

• **Exercise:** Find $|A|$.

The phase of A would depend on the initial condition in the time-dependent case and not very important. The wavefunction for the first few n are plotted on the left of Fig. 3.6. They are **standing waves** that satisfy the boundary conditions, just like standing waves in a closed optical cavity.

Given the restriction on the wavenumber k in Eq. (3.41), there is also a restriction on E :

$$E = \frac{\hbar^2 k^2}{2m} = \frac{\hbar^2 \pi^2 n^2}{2m L^2}, \quad n = 1, 2, \dots \quad (3.44)$$

Now the energy is **quantized** and depends **quadratically** on the integer. This is the major difference between a quantum particle inside a well and an optical cavity. Whereas the frequency of light depends linearly on the wavenumber ($\omega \propto k$), the frequency of the wavefunction here depends quadratically on the wavenumber ($\omega \propto k^2$). The **discrete energy levels** are plotted on the right of Fig. 3.6, and we see the first sign why quantum mechanics is called quantum: it has quantized energy levels.

Let's look at the solution for the lowest energy E_1 :

$$E_1 = \frac{\hbar^2 \pi^2}{2m L^2}. \quad (3.45)$$

It is above the zero potential inside the well, meaning that there is a nonzero kinetic energy even if the particle is already at the lowest possible energy. This is called the **zero-point** or **ground-state** energy, and is another difference between quantum mechanics and classical mechanics: The lowest energy for a classical particle is simply the potential energy without any kinetic energy, as the particle is at rest. For a quantum particle, it is

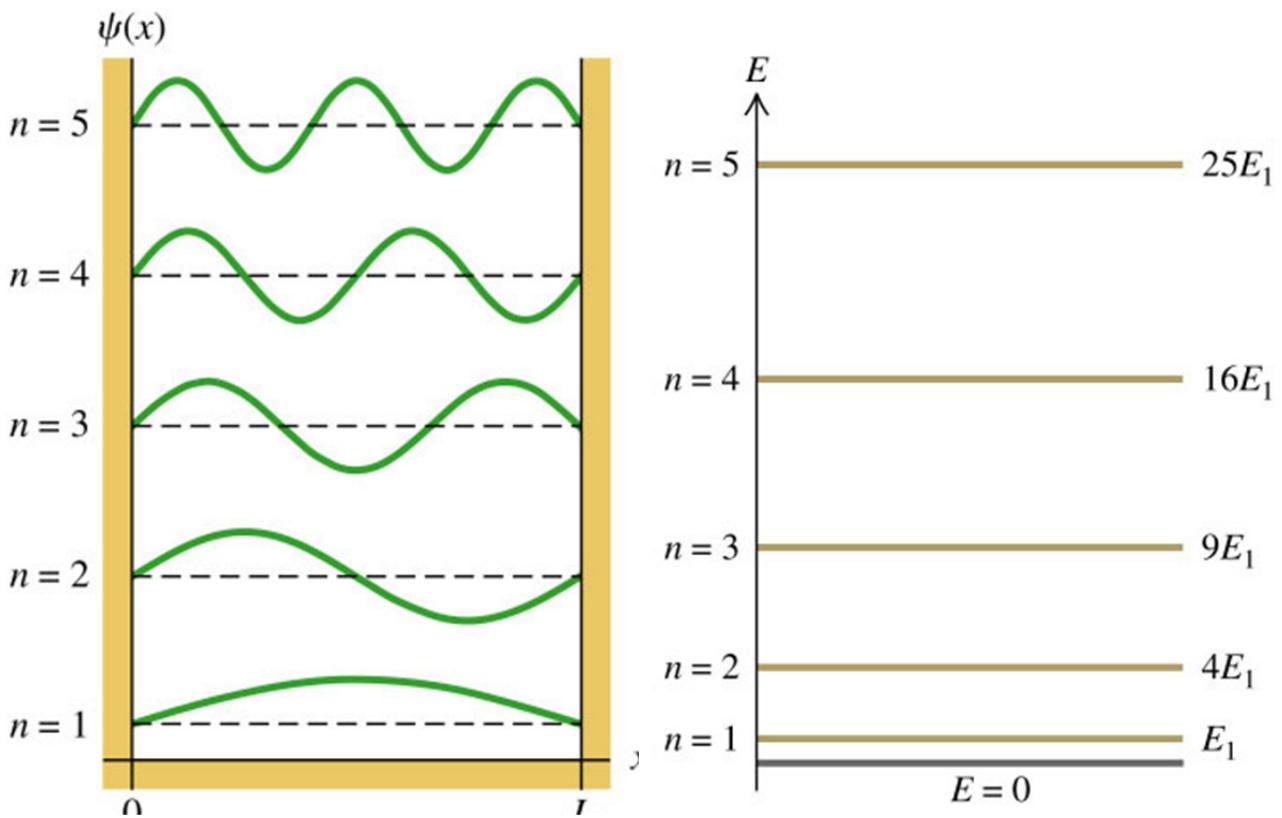


FIGURE 3.6. Left: standing wave solutions for the first few quantum numbers n . Right: the discrete energy levels. As we go up in n , the separation between adjacent energy levels becomes more and more sparse.

always traveling around the well, k is nonzero, and therefore E is somewhat above the potential energy. A particle that has a wavefunction with the lowest possible energy is said to be in the **ground state**.

3.3. *Quantum well

Many lasers and light-emitting diodes (LEDs) are made from quantum well structures in semiconductors, as shown in Fig. 3.7. In such structures, the middle layer has a lower potential for electrons, which are confined in that layer. The wavenumber along that direction is then quantized, although the electrons are still free to move in the two other directions. The electrons in such a layer is called a two-dimensional electron gas.

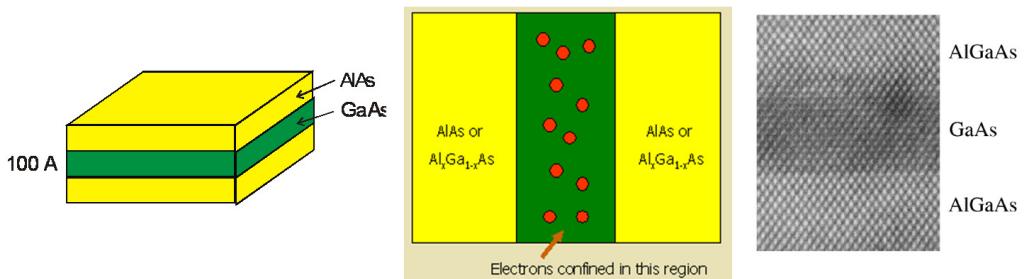


FIGURE 3.7. A quantum well structure. The sandwich of different semiconductor materials leads to a lower potential in the middle gallium arsenide layer and the electrons are confined within the layer.

The 2014 Nobel Prize in Physics was awarded to Isamu Akasaki, Hiroshi Amano and Shuji Nakamura “for the invention of efficient blue light-emitting diodes which has enabled bright and energy-saving white light sources.” (http://www.nobelprize.org/nobel_prizes/physics/laureates/2014/

[popular.html](#)), as shown in Fig. 3.8. The efficiency of their invention is in no small part due to the thin quantum well structure, which traps electrons (and holes) within a very thin region for the efficient generation of light. To compute the emission spectrum of a quantum-well LED, it is necessary to account for the discreteness of the wavenumber along the confinement direction [4].

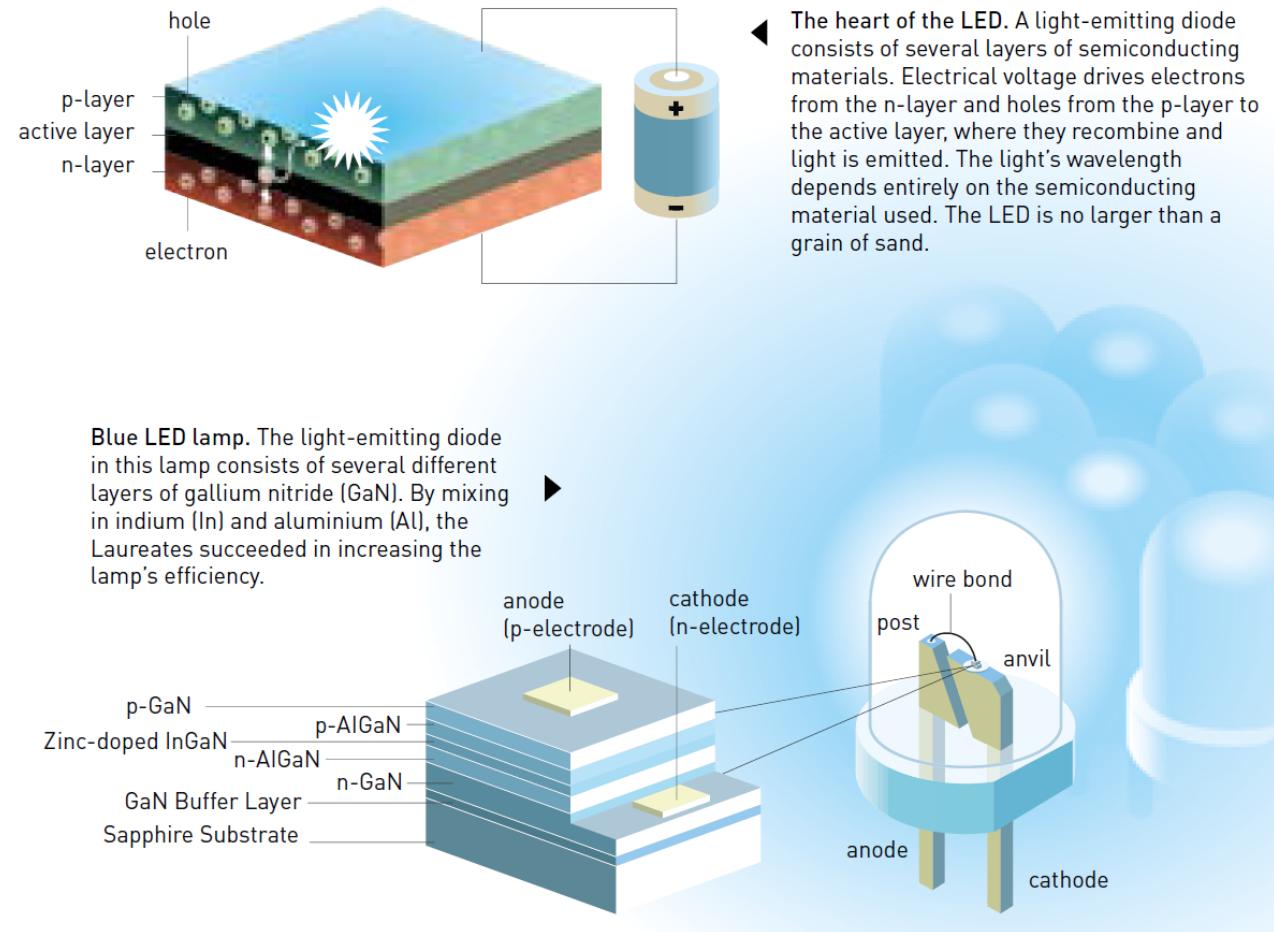


FIGURE 3.8. From www.nobelprize.org

3.4. Finite square well

3.4.1. Solutions in each region. We now consider the more realistic model of a square well with finite potential:

$$U(x) = \begin{cases} 0, & 0 \leq x \leq L, \\ U_0, & x < 0 \text{ or } x > L, \end{cases} \quad (3.46)$$

as plotted in Fig. 3.2.

First consider the region I outside the left wall ($x < 0$), as shown in Fig. 3.9:

$$E\psi(x) = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} + U_0\psi(x), \quad x < 0. \quad (3.47)$$

If we would like a normalizable solution, the propagating-wave solutions $\exp(\pm ikx)$ won't work. The other option is the evanescent wave solutions, and if we want it to be normalizable, we should choose one that decays exponentially away from the well as $x \rightarrow -\infty$. Thus we assume an energy that is below the potential outside the well:

$$E < U_0 \quad (3.48)$$

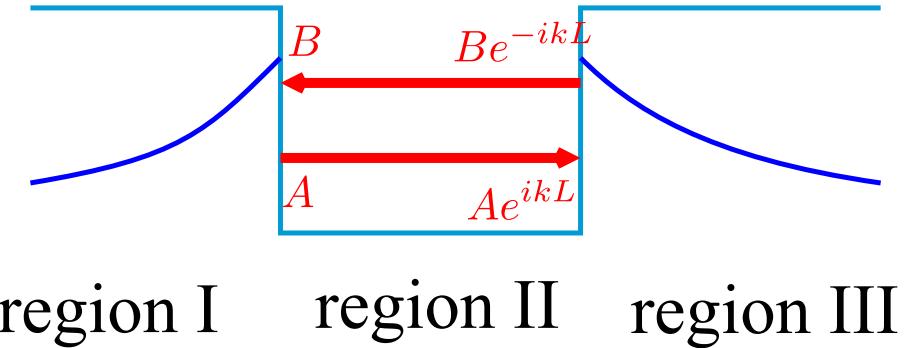


FIGURE 3.9. Input-output analysis for a finite square well.

and let

$$\psi_I(x) = C \exp(\alpha x), \quad (3.49)$$

where $\psi_I(x)$ is the part of $\psi(x)$ in region I ($x < 0$), C is a constant to be determined later, and α is a positive **decay coefficient** that depends on E according to Eq. (3.24):

$$\alpha = \sqrt{\frac{2m}{\hbar^2}(U_0 - E)}. \quad (3.50)$$

Similarly, for the region III outside the right wall ($x > L$),

$$E\psi(x) = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} + U_0 \psi(x), \quad x > L, \quad (3.51)$$

we should choose an evanescent solution that decays exponentially as $x \rightarrow +\infty$:

$$\psi_{III}(x) = G \exp(-\alpha x), \quad (3.52)$$

where $\psi_{III}(x)$ denotes the wavefunction to the right of the wall ($x > L$), G is another amplitude, and α is the same as before.

Within the well ($0 < x < L$), let's assume propagating solutions:

$$\psi_{II}(x) = A \exp(ikx) + B \exp(-ikx), \quad (3.53)$$

with the wavenumber depending on the energy as follows:

$$k = \sqrt{\frac{2mE}{\hbar^2}}. \quad (3.54)$$

Given Eqs. (3.50) and (3.54), α and k are related:

$$\alpha^2 + k^2 = \frac{2mU_0}{\hbar^2}, \quad (3.55)$$

which suggests that we can plot α versus k on a circle, as shown in Fig. 3.10. The larger the k (momentum inside the well), the smaller the α (decay coefficient outside the well), meaning that there is more of the particle protruding outside the well when the inside momentum is large.

We now have all these undetermined constants C, G, A, B , not to mention α and k as a function of E . To proceed further, we need **boundary conditions**.

3.4.2. Boundary conditions. Let's come back to the Schrödinger equation:

$$E\psi(x) = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi(x)}{\partial x^2} + U\psi(x). \quad (3.56)$$

Observe that, if U, E and ψ are all finite (not infinite), we also would like the second-order derivative to be finite. For a finite second-order derivative, we require $\psi(x)$ and the first-order derivative $\frac{\partial \psi(x)}{\partial x}$ to be continuous everywhere. This means that, at $x = 0$, where the wavefunctions in the two regions meet,

$$\psi_I(x=0) = \psi_{II}(x=0), \quad \left. \frac{\partial \psi_I(x)}{\partial x} \right|_{x=0} = \left. \frac{\partial \psi_{II}(x)}{\partial x} \right|_{x=0}. \quad (3.57)$$

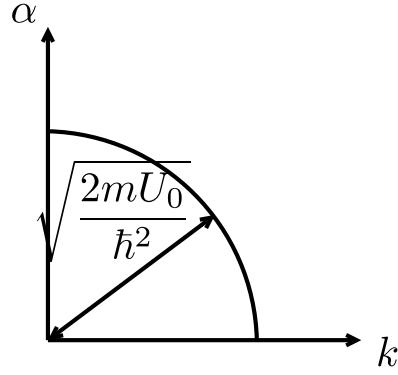


FIGURE 3.10. A relation between the decay coefficient α for the evanescent waves outside the well and the wavenumber k for the propagating waves inside the well. This comes from the fact that all the waves correspond to a particle with a given energy. The radius of the circle is $\sqrt{2mU_0/\hbar^2}$, where U_0 is the height of the potential outside the well.

Similarly,

$$\psi_{II}(x = L) = \psi_{III}(x = L), \quad \frac{\partial\psi_{II}(x)}{\partial x}\Big|_{x=L} = \frac{\partial\psi_{III}(x)}{\partial x}\Big|_{x=L}. \quad (3.58)$$

For a more graphical illustration, see the caption of Fig. 3.11.

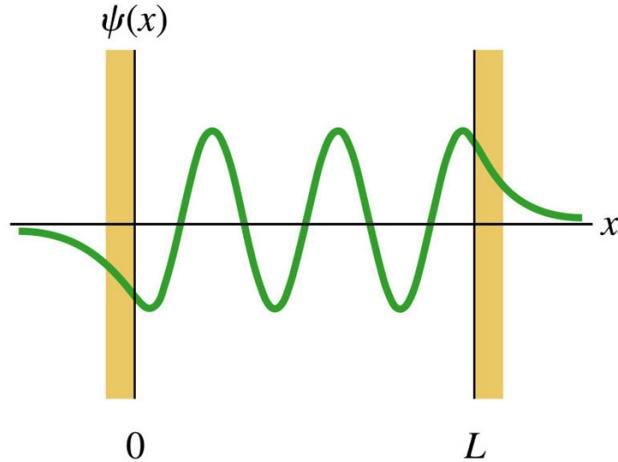


FIGURE 3.11. This is a possible solution for the wavefunction in the square well. Note that it is continuous at $x = 0$ ($\psi_I(x = 0) = \psi_{II}(x = 0)$) and also at $x = L$ (no jump). Also note that the slope of the wavefunction is the same coming from the left or from the right of a point (no kink). These boundary conditions come from the requirement of a finite second-order derivative.

3.4.3. Input-output analysis. We now study Fig. 3.9 and spell out the transmission and reflection coefficients at each wall. At the left wall, ψ_I given by Eq. (3.49) has to match ψ_{II} given by Eq. (3.53) at $x = 0$ according to Eq. (3.57). This means that

$$\psi_I(0) = C = \psi_{II}(0) = A + B, \quad \psi'_I(0) = C\alpha = \psi'_{II}(0) = ikA - ikB. \quad (3.59)$$

We can combine these two equations to give

$$A + B = C = \frac{ik}{\alpha}(A - B), \quad A = \frac{ik + \alpha}{ik - \alpha}B. \quad (3.60)$$

Just as in Sec. 3.2, we can consider B to be an input amplitude to the left wall and A to be an output amplitude arising from the reflection by the left wall. A reflection coefficient can then be defined as

$$s_{22} = \frac{A}{B} = \frac{ik + \alpha}{ik - \alpha}. \quad (3.61)$$

Note that both k and α are real here, resulting in $|s_{22}|^2 = |A|^2/|B|^2 = 1$. Since the input wave has a probability density $\propto |B|^2$ and the output wave has a probability density $\propto |A|^2$, the unity $|s_{22}|^2$ implies that the input and output probabilities are equal, and the entire amount of the particle that comes to the wall is reflected. This is precisely the phenomenon of **total internal reflection** that we encountered in optics: the reflection is total, and the transmitted wave is evanescent. We can define a transmission coefficient as

$$s_{12} = \frac{C}{B} = \frac{2ik}{ik - \alpha}, \quad (3.62)$$

which relates the amplitude C in region I to B . We can visualize this reflection roughly as a particle that runs a little bit into the classically forbidden region but is ultimately reflected completely, as shown in Fig. 3.12.

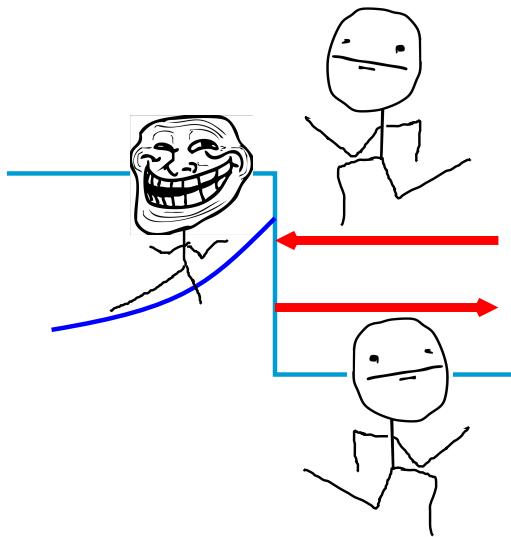


FIGURE 3.12. For a finite square well, a quantum particle can enter a classically forbidden region in the form of an evanescent wave, but it must be totally reflected eventually.

On the right wall, we can regard Ae^{ikL} as an input amplitude, Be^{-ikL} as a reflected amplitude, and $\psi_{III}(x = L) = Ge^{-\alpha L}$ as the transmitted amplitude. The situation is the same as that on the left wall:

$$\frac{Be^{-ikL}}{Ae^{ikL}} = s_{22} = \frac{ik + \alpha}{ik - \alpha}, \quad (3.63)$$

$$\frac{Ge^{-\alpha L}}{Ae^{ikL}} = s_{12} = \frac{2ik}{ik - \alpha}. \quad (3.64)$$

The key is to combine Eq. (3.61) and (3.63) to obtain

$$s_{22}^2 e^{2ikL} = 1. \quad (3.65)$$

This is the same equation for the round-trip factor as that for the infinite square well in Eq. (3.38), except that the reflection coefficient s_{22} now depends on k and α according to Eq. (3.61):

$$\left(\frac{ik + \alpha}{ik - \alpha} \right)^2 e^{2ikL} = 1. \quad (3.66)$$

This equation is yet another relation between α and k . Now we need to combine Eq. (3.66) and the earlier Eq. (3.55) to find the permissible k for a finite square well.

- **Exercise:** If the potential beyond the right wall ($x > L$) is a constant U_1 not equal to U_0 , how would Eq. (3.65) be modified? You should start from the beginning of this section (Sec. 3.4) and modify all the steps accordingly.

3.4.4. *Finding the wavenumber. The procedure is as follows: First we simplify Eq. (3.66). There are two solutions to Eq. (3.66):

$$\left(\frac{ik + \alpha}{ik - \alpha} \right) e^{ikL} = 1 \quad \text{or} \quad \left(\frac{ik + \alpha}{ik - \alpha} \right) e^{ikL} = -1. \quad (3.67)$$

For each solution, put the denominator on the right-hand side and multiply both sides by $e^{-ikL/2}$:

$$(ik + \alpha)e^{ikL/2} = (ik - \alpha)e^{-ikL/2} \quad \text{or} \quad (ik + \alpha)e^{ikL/2} = -(ik - \alpha)e^{-ikL/2}. \quad (3.68)$$

Grouping the terms,

$$2\alpha \cos \frac{kL}{2} = 2k \sin \frac{kL}{2} \quad \text{or} \quad 2i\alpha \sin \frac{kL}{2} = -2ik \cos \frac{kL}{2}. \quad (3.69)$$

These result in

$$\alpha = k \tan \left(\frac{kL}{2} \right) \quad \text{or} \quad -k \cot \left(\frac{kL}{2} \right). \quad (3.70)$$

Multiplying both sides by L , we obtain an equation in terms of dimensionless numbers αL and kL :

$$\alpha L = kL \tan \left(\frac{kL}{2} \right) \quad \text{or} \quad -kL \cot \left(\frac{kL}{2} \right). \quad (3.71)$$

In terms of these dimensionless numbers, Eq. (3.55) can also be rewritten as

$$(\alpha L)^2 + (kL)^2 = \frac{2mU_0L^2}{\hbar^2}. \quad (3.72)$$

We then plot Eqs. (3.71) and (3.72) as $Y = \alpha L$ along the vertical axis versus $X = kL$ along the horizontal axis, as shown in Fig. 3.13.

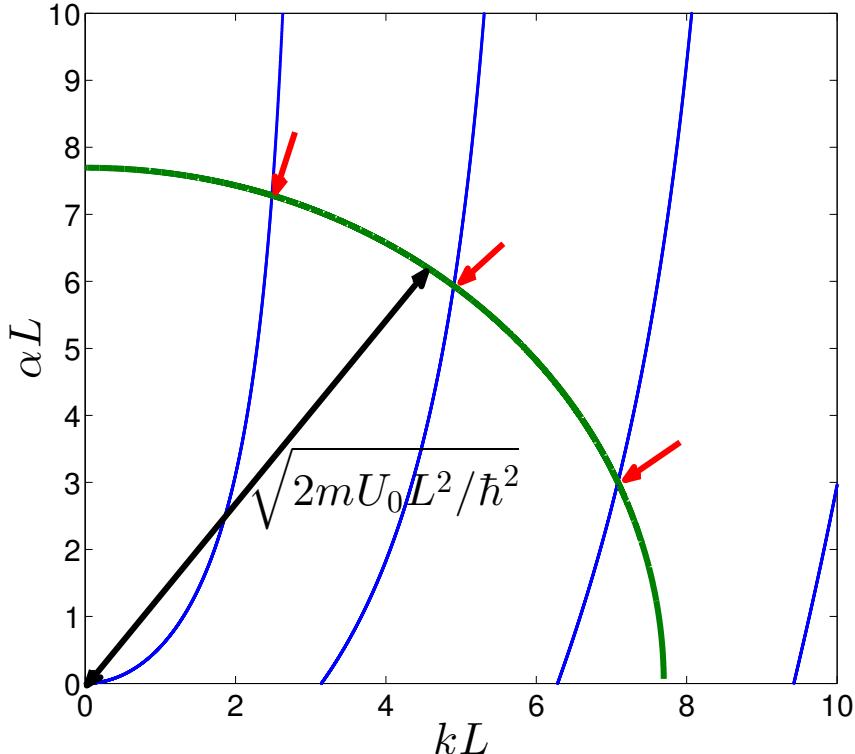


FIGURE 3.13. The two relations between evanescent-wave decay coefficient α and wavenumber k that must be satisfied for a finite square well. Both α and k are positive by assumption. The set of blue curves are a plot of Eq. (3.71) and the circle with radius $\sqrt{2mU_0L^2/\hbar^2}$ is a plot of Eq. (3.72). Only the discrete points on the intersections between the two sets of curves, denoted by the red arrows, can satisfy both relations.

Only the intersections between the two sets of curves can satisfy both relations and therefore represent the final restrictions on k . Only some discrete values of k will do, and note that, given the dimensionless number $\sqrt{2mU_0L^2/\hbar^2}$, there are only a finite number of intersections, leading to a finite number of k , whereas there are infinitely many k levels for an infinite square well.

If we change the potential well depth U_0 or the width L , the relation between the dimensionless numbers $Y = \alpha L$ and $X = kL$ governed by Eq. (3.71) remains the same:

$$Y = X \tan\left(\frac{X}{2}\right) \text{ or } -X \cot\left(\frac{X}{2}\right), \quad (3.73)$$

but the radius of the curve governed by Eq. (3.72) increases as a function of $\sqrt{2mU_0L^2/\hbar^2}$:

$$X^2 + Y^2 = \frac{2mU_0L^2}{\hbar^2}. \quad (3.74)$$

This means that, as we increase this radius by increasing U_0 or L , there will be more and more intersections, leading to more and more permissible k values. A nice demonstration of this can be found at <http://phet.colorado.edu> (bound-states).

3.4.5. Example. The circle in Fig. 3.13 is actually plotted for a radius of

$$\sqrt{\frac{2mU_0L^2}{\hbar^2}} = \sqrt{6}\pi, \quad (3.75)$$

following an example in Ref. [2]. In the figure we see that there are three intersections, corresponding to the three wavefunctions in the square well. The three values of k have to be found numerically by combining Eqs. (3.71) and (3.72), but once they are found the energy E is determined from Eq. (3.54), or $E = \hbar^2k^2/(2m)$, α is determined from either Eq. (3.71) or (3.72), and s_{22} and s_{12} both depend on k and α . The resulting wavefunctions are obtained by noting that all the amplitudes end up being proportional to B :

$$A = s_{22}B \text{ from Eq. (3.61)}, \quad (3.76)$$

$$C = s_{12}B \text{ from Eq. (3.62)}, \quad (3.77)$$

$$G = s_{12}Ae^{ikL}e^{\alpha L} = s_{12}s_{22}Be^{ikL}e^{\alpha L} \text{ from Eqs. (3.64) and (3.76)}. \quad (3.78)$$

The wavefunction is then proportional to B everywhere, and B , the remaining unknown, can be determined using the normalization condition. The three resulting wavefunctions are plotted in Fig. 3.14. Each wavefunction consists of a spatially oscillating component inside the well due to the interference between plane waves and evanescent waves on the sides. These normalizable solutions are also called **bound states** in general. The energy is always below U_0 , otherwise α would not be real, the waves outside the well would not be evanescent, and we would not get a normalizable solution. Note that there is a nonzero probability for the particle to be found outside the well, where the potential energy there is above the total energy E . This would have been forbidden in classical physics.

3.4.6. Scattering states. For $E > U_0$, the wave solutions outside the well are also propagating waves. They are not normalizable but can be used to approximate or build traveling solutions. They are useful for studying the scattering of waves (transmission and reflection) by the potential well.

For $E < 0$, we are forced to assume evanescent waves inside the well, but it turns out that it is impossible to satisfy the boundary conditions using evanescent waves everywhere. No wave solution thus exists for $E < 0$.

3.4.7. Infinite-square-well limit. If $U_0 \rightarrow \infty$, the solutions for a finite square well should approach those for an infinite square well in Sec. 3.2. The easiest way to see this is to observe that $\alpha \rightarrow \infty$ for $U_0 \rightarrow \infty$ according to Eq. (3.50), leading to

$$\lim_{\alpha \rightarrow \infty} s_{22} = \lim_{\alpha \rightarrow \infty} \frac{ik + \alpha}{ik - \alpha} = -1, \quad (3.79)$$

which coincides with the reflection coefficient for a wall of the infinite square well. The simpler analysis in Sec. 3.2 then follows.

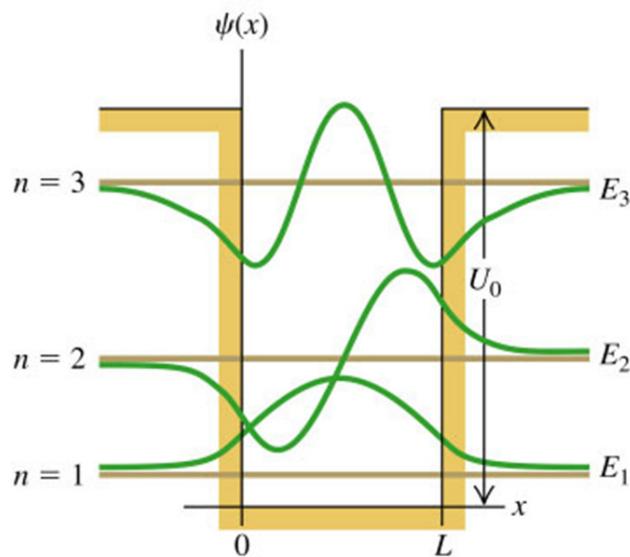


FIGURE 3.14. Wavefunctions of the three bound states in a finite square well with $\sqrt{2mU_0L^2/\hbar^2} = \sqrt{6}\pi$. From [2].

CHAPTER 4

Harmonic Oscillator

The harmonic oscillator, with a potential given by

$$U(x) = \frac{1}{2}\kappa x^2, \quad (4.1)$$

was discussed in Sec. 1.1.3 in the context of classical physics and again mentioned in Sec. 3.1.4. It is important for two reasons:

- (1) A more realistic potential function may look something like Fig. 4.1, but the particle can be trapped near a local minimum, in which case we can approximate $U(x)$ with a parabola and the problem becomes easier to solve.

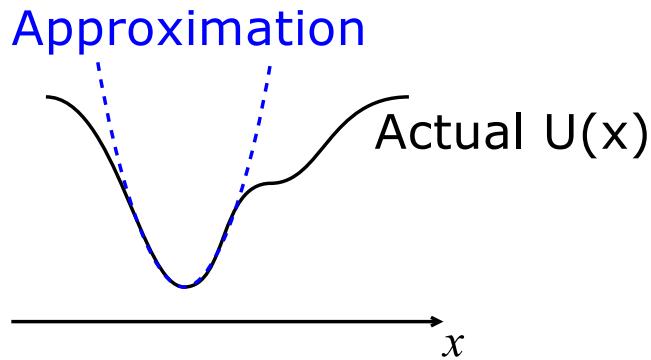


FIGURE 4.1. The parabolic approximation of a potential function $U(x)$ near a local minimum.

- (2) The quantum harmonic oscillator is a basic building block of quantum field theory. A full theory of quantum electromagnetics, for example, regards each plane wave with a certain wavevector as a quantum harmonic oscillator.

We will not worry about these two applications in this module, but they are so important for more advanced quantum mechanics that we will mention this briefly in this chapter.

4.1. Hermite-Gaussian wavefunctions

The Schrödinger equation for a harmonic oscillator reads

$$E\psi = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + \frac{1}{2}\kappa x^2 \psi. \quad (4.2)$$

The solution is

$$\psi_n(x) = A_n H_n(\beta x) \exp\left(-\frac{1}{2}\beta^2 x^2\right), \quad n = 0, 1, 2, \dots, \quad (4.3)$$

where β is a constant defined as

$$\beta \equiv \left(\frac{m\kappa}{\hbar^2}\right)^{1/4}, \quad (4.4)$$

$H_n(\beta x)$ is a Hermite polynomial (http://en.wikipedia.org/wiki/Hermite_polynomials), and A_n is a normalization constant. The first few wavefunctions are plotted in Fig. 4.2. These wavefunctions are called Hermite-Gaussian functions.

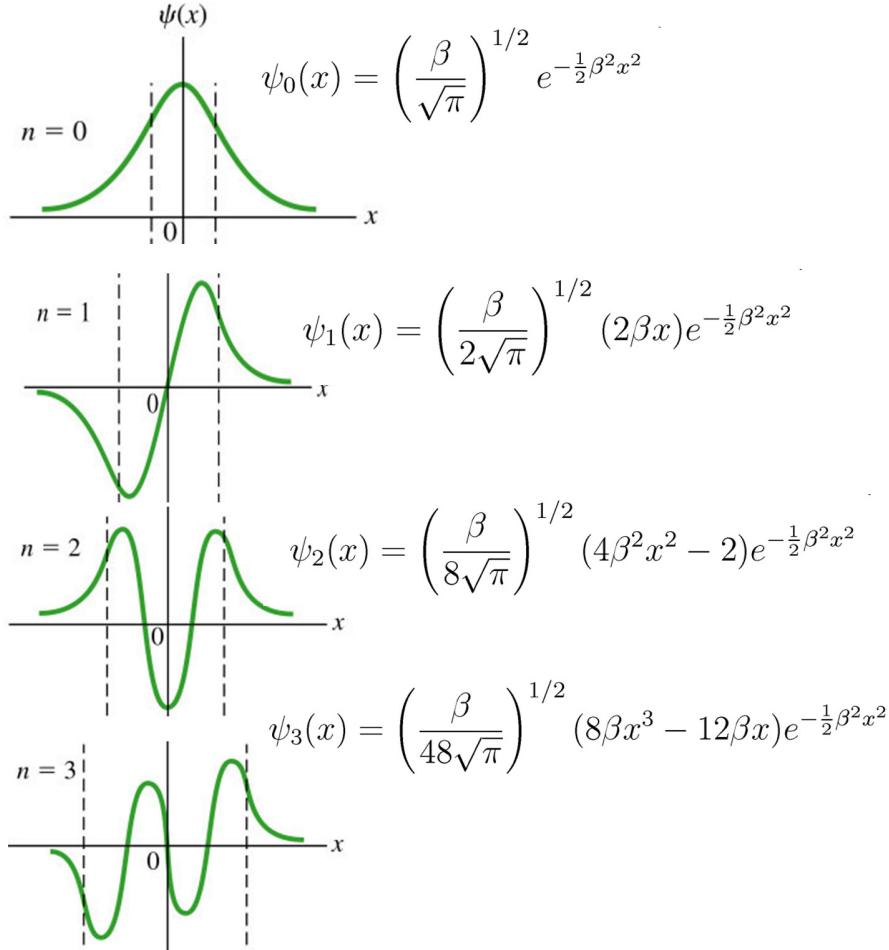


FIGURE 4.2. The wavefunctions for the first few energy levels for a quantum harmonic oscillator.

In practice, almost no one cares about the specific forms of the wavefunctions for the harmonic oscillator; rather, it is the energy for each of these wavefunctions n that is the most important:

$$E_n = \left(n + \frac{1}{2} \right) \hbar\omega_0, \quad n = 0, 1, 2, \dots \quad (4.5)$$

Note that

$$\omega_0 \equiv \sqrt{\frac{\kappa}{m}} = \text{classical frequency} \quad (4.6)$$

is the **classical frequency** for the harmonic oscillator. It is not the frequency for the quantum wavefunction in the time-dependent case, which oscillates at frequency

$$\text{quantum frequency} = \frac{E_n}{\hbar} = \left(n + \frac{1}{2} \right) \omega_0. \quad (4.7)$$

Also note that the energy levels, as shown on the left of Fig. 4.3, are **equally spaced**, unlike, for example, the energy levels for an infinite square well that scale quadratically with an integer.

The difference between the adjacent energy levels is $\hbar\omega_0$. It is an extremely small energy for oscillators in everyday life, where the discrete energy levels are so closely spaced that one would not have noticed the discreteness. On the other hand, for optical frequencies $\omega_0 \sim 10^{15}$ rad/s, $\hbar\omega_0 \sim 10^{-19}$ J, which is comparable to the typical electron energy which would be the electron charge $e \approx 1.6 \times 10^{-19}$ C times a typical voltage, say, 1 V, or $E = 1$ eV (electron-Volt). The separation between the discrete energy levels then becomes a lot more noticeable.

Fig. 4.4 compares the classical physics and the quantum physics of a harmonic oscillator side-by-side.

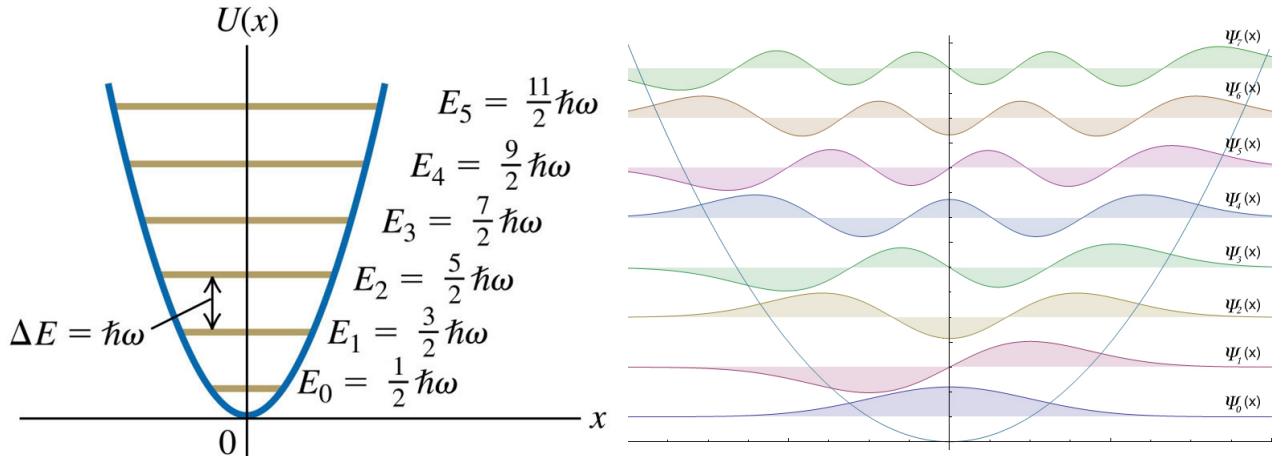
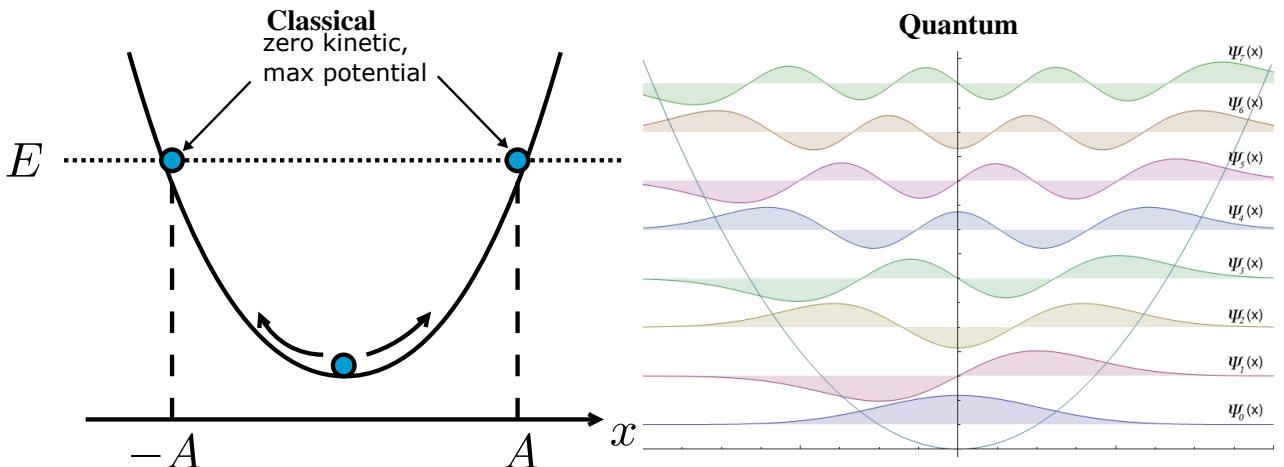


FIGURE 4.3. Left: the discrete energy levels for a quantum harmonic oscillator, plotted together with the quadratic potential. Right: the Hermite-Gaussian wavefunctions together with the energy levels. From <http://en.wikipedia.org/wiki/File:HarmOsziFunktionen.png>.



- At each time, the particle can exist at only one point in space.
- Energy can be arbitrary as long as it is above the minimum potential value $U(0)$. The lowest possible energy is $U(0)$, when the particle is at rest. $E - U(x)$ is the kinetic energy, and as the particle moves around, it has different portions of kinetic energy and potential energy depending on its position.
- A particle cannot be in regions where its energy E is below the potential energy $U(x)$
- The general solution is $x(t) = A \cos(\omega_0 t + \phi)$, where A and ϕ depend on the initial conditions for position and velocity. The classical physics turns out to be an approximation of the time-dependent quantum mechanics if we define the initial condition of the wavefunction properly.

- For a given energy, the wavefunction is spread everywhere except at nodes where $\psi_n(x) = 0$.
- Only discrete energy levels are allowed. The lowest possible energy is slightly above the minimum potential energy.
- In regions where $E < U(x)$, $\psi(x)$ is nonzero, although it decays rapidly away from the center. It is possible to find the particle in such classically forbidden regions.
- To obtain the evolution in the time dependent case, one should write the initial wavefunction $\psi(x, 0)$ as a superposition of the Hermite-Gaussian wavefunctions. Then each of these Hermite-Gaussian wavefunctions will evolve in time with a factor $\exp(-iE_n t/\hbar)$. Summing them back up gives us the wavefunction at a later time.

FIGURE 4.4. Comparison of the classical physics and the quantum physics of a harmonic oscillator.

4.2. Generic features of quantum bound states

We have now studied a few examples of potential wells: the **infinite square well**, the **finite square well**, and the **harmonic oscillator**. We have focused on **normalizable** solutions, which are concentrated near each potential minimum and are also called **bound states**, as they are confined by attractive forces to stay near the well and don't go outside the well much. We can now summarize a few generic features of these bound states for any potential well:

- (1) **Energy exists at discrete levels**, as standing-wave solutions exist only for such energies.
- (2) **Evanescence waves**: As long as $U(x)$ is not infinite, it is possible for the wavefunction to be nonzero where $U(x) > E$, although it decays rapidly in those regions.
- (3) **Energy must be below the potential energy $U(x)$ for $x \rightarrow \pm\infty$** . If the energy is above the potential energy far away from the well, there is a positive kinetic energy that can carry the wave away to infinity and one would not have a bound state anymore. We have traveling wave solutions (scattering states) instead, which are not normalizable.
- (4) **Zero-point energy**: The lowest bound-state energy is always strictly above the minimum potential energy. *This is because any wavefunction in space is a superposition of waves with different momenta, and unless the wavefunction is a constant everywhere it always has some components that are traveling and have some positive kinetic energy.
- (5) Note, however, that the following features depend on the specific potential function:
 - The specific energy levels and the spacings between them. For example, $E_n \propto n^2$ for an infinite square well and $E_n \propto n + 1/2$ for a harmonic oscillator.
 - The evanescent waves may decay in different ways depending on the potential (exponential $\exp(\pm\alpha x)$ for a flat potential outside the well, and Gaussian decay for a harmonic oscillator).

CHAPTER 5

Tunneling

We would now like to study traveling waves (scattering states) in quantum mechanics. They behave very similarly to optical waves in the Mach-Zehnder or Fabry-Pérot interferometers, but we shall take this opportunity to study something different: the phenomenon of tunneling, as introduced in Fig. 5.1.

We have already seen that a quantum wave can exist in regions where its energy is below the potential there. It must be **evanescent** and cannot carry any real momentum, but the wavefunction is nonzero and leads to a nonzero probability that the particle will be found there.

The evanescent wave decays rapidly in a forbidden region, but when it meets another region that has a lower potential energy than the energy of the particle, it can be **transmitted** as a traveling wave again.

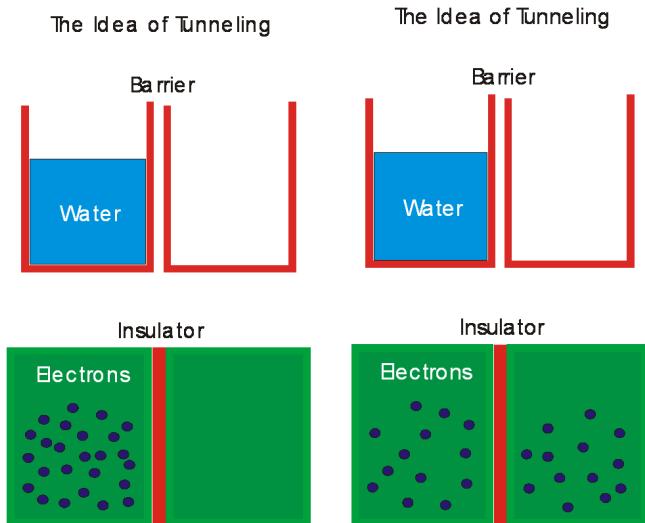


FIGURE 5.1. Top: classically, water in a bucket cannot leak into another bucket as long as the bucket has no hole, because the bucket wall strongly repels the water molecules as they get near it and the molecules do not have enough energy to overcome the repulsion. Bottom: in quantum mechanics, each electron has a nonzero probability to tunnel to the other side even if it has a lower energy than the potential barrier.

Let us study a simple example of tunneling by assuming the potential barrier shown in Fig. 5.2.

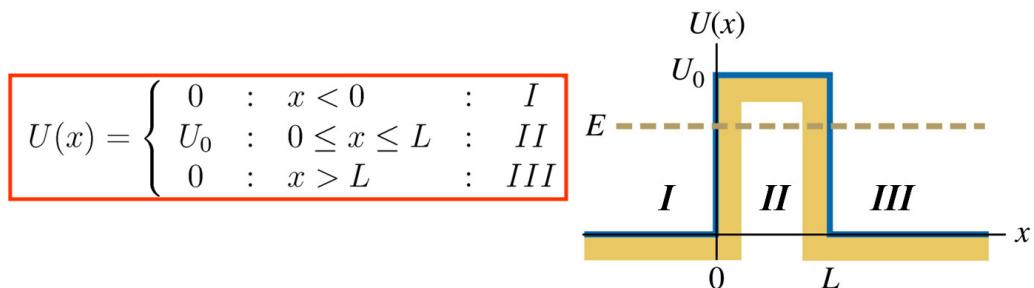


FIGURE 5.2.

The solutions we assume for each region are given in Fig. 5.3. An incident wave from the left with wavefunction $A \exp(ikx)$ impinges on the potential barrier. There will be a reflected wave given by $B \exp(-ikx)$,

and a final transmitted wave towards the right given by $C \exp(ikx)$. Our goal is to solve for the reflection coefficient B/A and the transmission coefficient C/A and find out the probability of the particle tunneling to the other side, when it has a lower energy than the potential barrier.

$$\psi(x) = \begin{cases} Ae^{ikx} + Be^{-ikx} & : I \\ Fe^{\alpha x} + Ge^{-\alpha x} & : II \\ Ce^{ikx} & : III \end{cases}$$

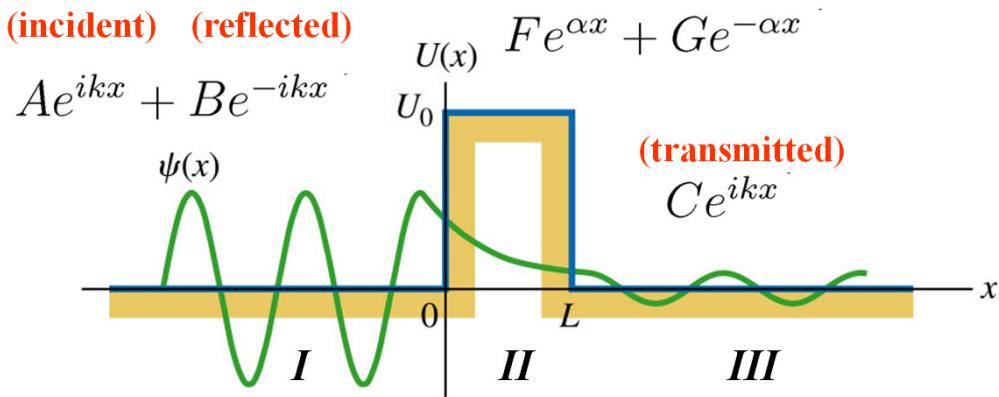


FIGURE 5.3.

The wave inside the barrier is evanescent, and it consists of not only the wave transmitted from the left wall ($\exp(-\alpha x)$) but also the reflection of the $\exp(-\alpha x)$ wave at the right wall, resulting in another evanescent wave that is $\propto \exp(\alpha x)$. To proceed further, we shall undertake an input-output analysis not unlike what we did for a Fabry-Pérot interferometer.

5.1. Input-output analysis

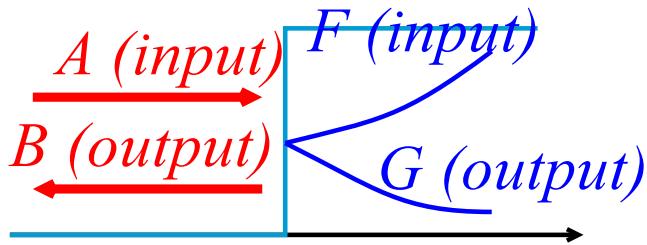


FIGURE 5.4. The setup of input and output waves on the left wall of the potential barrier. The waves inside the barrier are evanescent, but otherwise the input-output analysis is the same as before.

Let's focus on the left wall at $x = 0$ between region I and region II, as depicted in Fig. 5.4. We now have two inputs, one coming from the left ($A \exp(ikx)$), and one coming from the right ($F \exp(\alpha x)$). The only new feature here is that the waves in region II are evanescent, but otherwise the math proceeds in the same fashion as what we did for a Fabry-Pérot interferometer.

To get the scattering matrix for this wall, we use the boundary conditions in Sec. 3.4.2.

$$\psi_I(0) = A + B = \psi_{II}(0) = F + G, \quad (5.1)$$

$$\psi'_I(0) = ikA - ikB = \psi'_{II}(0) = \alpha F - \alpha G. \quad (5.2)$$

Let's put the outputs (B and G) on the left-hand side and the inputs (A and F) on the right-hand side:

$$B - G = -A + F, \quad (5.3)$$

$$-ikB + \alpha G = -ikA + \alpha F. \quad (5.4)$$

This is a system of equations relating two inputs and two outputs. It can be solved by writing them in matrix form:

$$\begin{pmatrix} 1 & -1 \\ -ik & \alpha \end{pmatrix} \begin{pmatrix} B \\ G \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -ik & \alpha \end{pmatrix} \begin{pmatrix} A \\ F \end{pmatrix}. \quad (5.5)$$

We can multiply both sides by the inverse of the left matrix to obtain

$$\begin{pmatrix} 1 & -1 \\ -ik & \alpha \end{pmatrix}^{-1} \begin{pmatrix} 1 & -1 \\ -ik & \alpha \end{pmatrix} \begin{pmatrix} B \\ G \end{pmatrix} = \begin{pmatrix} B \\ G \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -ik & \alpha \end{pmatrix}^{-1} \begin{pmatrix} -1 & 1 \\ -ik & \alpha \end{pmatrix} \begin{pmatrix} A \\ F \end{pmatrix}, \quad (5.6)$$

$$\begin{pmatrix} B \\ G \end{pmatrix} = \begin{pmatrix} \frac{ik+\alpha}{ik-\alpha} & \frac{-2\alpha}{ik-\alpha} \\ \frac{2ik}{ik-\alpha} & -\frac{ik+\alpha}{ik-\alpha} \end{pmatrix} \begin{pmatrix} A \\ F \end{pmatrix}. \quad (5.7)$$

In other words, the scattering matrix for the left interface is

$$\begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} \frac{ik+\alpha}{ik-\alpha} & \frac{-2\alpha}{ik-\alpha} \\ \frac{2ik}{ik-\alpha} & -\frac{ik+\alpha}{ik-\alpha} \end{pmatrix}. \quad (5.8)$$

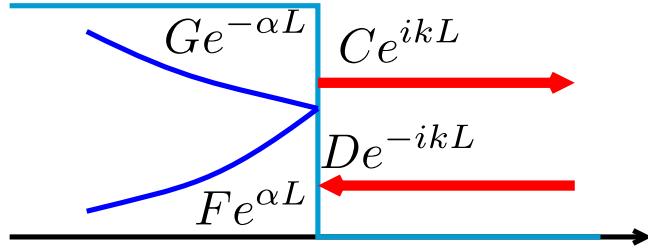


FIGURE 5.5. Input-output analysis for the right wall.

Now that we have obtained the scattering matrix for the left wall, the matrix for the right wall at $x = L$ becomes trivial. We just need to remember that the amplitudes there should be evaluated at $x = L$, and there can be another input wave coming from the right towards the barrier given by $D \exp(-ikx)$. The input-output relation is thus

$$\begin{pmatrix} Ce^{ikL} \\ Fe^{\alpha L} \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} De^{-ikL} \\ Ge^{-\alpha L} \end{pmatrix}. \quad (5.9)$$

At the right wall, G becomes an input and F becomes an output, as shown in Fig. 5.5, and we should remember that the waves have additional factors when evaluated at $x = L$.

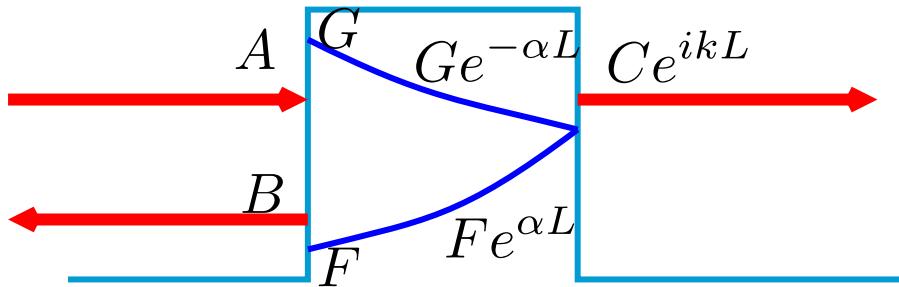


FIGURE 5.6. Input-output analysis for the potential barrier.

For the complete potential barrier shown in Fig. 5.6, there will be **multiple reflections** of these evanescent waves between the two walls, and we can choose any of the two methods in Part I, such as the partial-wave

analysis, to complete our task. The answer is

$$\frac{Ce^{ikL}}{A} = \frac{s_{12}s_{21}e^{-\alpha L}}{1 - s_{22}^2 e^{-2\alpha L}}, \quad \frac{B}{A} = s_{11} + \frac{s_{12}s_{22}s_{21}e^{-2\alpha L}}{1 - s_{22}^2 e^{-2\alpha L}}. \quad (5.10)$$

These answers should not surprise you; the only difference between them and the coefficients for a Fabry-Pérot is the propagation factors inside the barrier: For a Fabry-Pérot, there is a phase delay $\exp(jkL)$ introduced by propagation for length L inside a dielectric, whereas propagation of an evanescent wave introduces a factor of

$$\exp(-\alpha L) = \exp\left[-\sqrt{\frac{2m}{\hbar^2}(U_0 - E)L}\right]. \quad (5.11)$$

The decay coefficient α decreases monotonically with increasing E , and there is no resonance behavior to speak of. Nonetheless, the important point here is that the **transmission probability**,

$$T \equiv \frac{|Ce^{ikL}|^2}{|A|^2} = \frac{|s_{12}s_{21}e^{-\alpha L}|^2}{|1 - s_{22}^2 e^{-2\alpha L}|^2}, \quad (5.12)$$

is nonzero. This is the tunneling probability, because $|C|^2$ is proportional to the probability current of the transmitted wave, and the ratio $|C|^2/|A|^2$ tells us what fraction of the incoming current is transmitted.

If the evanescent wave decay is severe, the round-trip factor $s_{22}^2 \exp(-2\alpha L)$ would have a very small magnitude relative to 1. This means that multiple reflections will not be important, and we can just keep the lowest-order partial wave:

$$T \approx |s_{12}s_{21}e^{-\alpha L}|^2 = \frac{16\alpha^2 k^2}{(\alpha^2 + k^2)^2} \exp(-2\alpha L), \quad (5.13)$$

Recall that $\alpha^2 \propto U_0 - E$ as in Eq. (3.50), $k^2 \propto E$ as in Eq. (3.54), and $\alpha^2 + k^2 \propto U_0$ as in Eq. (3.55). As a function of the ratio between energy E and potential height U_0 , the tunneling probability becomes

$$T \approx 16 \left(1 - \frac{E}{U_0}\right) \frac{E}{U_0} \exp\left(-2\sqrt{\frac{2mU_0L^2}{\hbar^2}} \sqrt{1 - \frac{E}{U_0}}\right). \quad (5.14)$$

The important point here is that the tunneling probability is very low ($T \ll 1$) when $E \ll U_0$ and the energy is nowhere near the potential height, as one might expect.

A time-dependent simulation of tunneling can be found at <http://phet.colorado.edu> (quantum-tunneling).

5.2. *Applications: Flash memory, tunneling in a transistor, scanning-tunneling microscope (STM)

The flash memory (USB thumb drives, solid-state hard disks, etc.) store bits by trapping electrons in a special transistor. The writing and erasing of the bit are done by transferring electrons from an island through quantum tunneling. Check out http://ocw.mit.edu/courses/electrical-engineering-and-computer-science-6-007-electromagnetic-energy-from-motors-to-lasers-spring-2011/lecture-notes/MIT6_007S11_lec43.pdf.

- **Exercise:** Consider an electron initially trapped inside a leaky potential well, which consists of two identical potential barriers separated by distance L . The potential is 0 between the barriers, each barrier has a potential height U_0 and width d , and the potential outside the barriers is 0, as shown in Fig. 5.7. What is the probability that an electron with energy E is still there after some time t , assuming that the probability of tunneling T is very low?

Answer: We'll need time-dependent quantum mechanics to do this exactly, but we can also estimate it in a rough semiclassical way. We think of the electron as a classical particle with group velocity $v_g = \hbar k/m$. It bounces back and forth between the two barriers, and for each bounce there is a tunneling probability T that it will tunnel through the barrier and go away. In each bounce, the probability that the electron will be reflected and stay inside is thus $1 - T$. Now, the particle takes a time of L/v_g to go from one barrier to another, so in time t the number of bounces is around

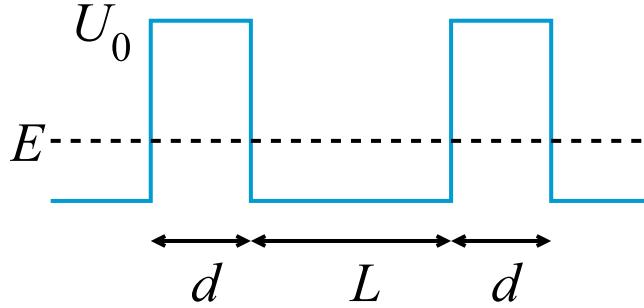


FIGURE 5.7. A leaky potential well, where the two barriers have finite width d .

$N \approx t/(L/v_g) = v_g t/L$. The probability that the electron will still be inside after N bounces is therefore

$$(1 - T)^N \approx \exp(-TN) \approx \exp\left(-\frac{T v_g}{L} t\right), \quad (5.15)$$

where I have used the approximation $\exp(-T) \approx 1 - T$ for $T \ll 1$ (from Taylor series) in reverse to write the probability as an exponential decay. We can use this to estimate how long a flash memory cell stores the electrons before the total charge decays significantly. The decay is quicker for electrons with higher group velocity $\hbar k/m = \sqrt{2E/m}$, a higher tunneling probability T , or a smaller distance L between the two barriers.

Also check out these articles on the effect of tunneling for transistors: http://en.wikipedia.org/wiki/MOSFET#Increased_gate-oxide_leakage, <http://spectrum.ieee.org/semiconductors/devices/the-tunneling-transistor>.

For a description of STM, see Figs. 5.8, 5.9, and 5.10. The vacuum between the tip and the sample being measured is a potential barrier. The measured tunneling current is proportional to the tunneling probability T . Since T is extremely sensitive to the width of the potential barrier (exponential relation), the current is very sensitive to the distance between the tip and the sample.

5.3. *Frustrated total internal reflection

The phenomenon of tunneling can also be seen in optics. When total internal reflection occurs, the evanescent wave in the lower refractive index region can be transmitted as a propagating wave again in another dielectric, as shown in Fig. 5.11.

5.4. *Resonant tunneling

Let us consider the leaky potential well again, but now there is an input wave from the left, as shown in Fig. 5.12. Because of tunneling, each potential barrier acts as a partial mirror. There would be multiple reflections between the two barriers, and the final transmission on the right-hand side has the same behavior as that of a Fabry-Pérot cavity. This means that, if the energy of the input wave matches the resonance condition, it can be transmitted 100% in principle.

While 100% transmission with a Fabry-Pérot cavity is not so surprising in the context of optics, in quantum mechanics this seems a lot more counterintuitive: It's like saying that it is difficult to run through one wall, but it is easy to run through two walls provided you have the right energy.

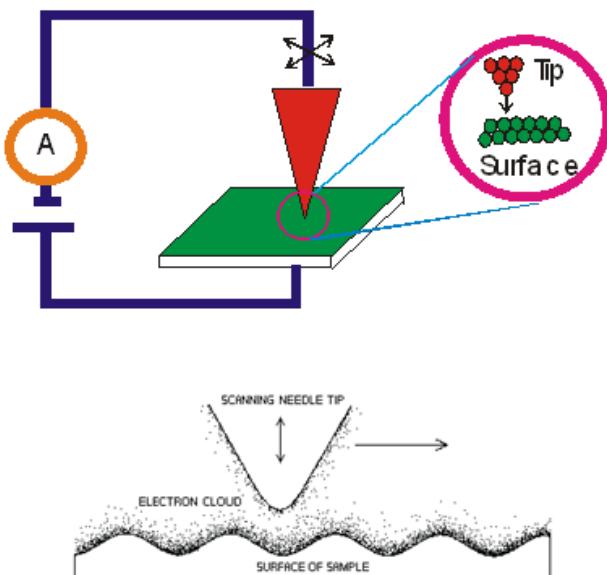
- **Question:** If the resonant tunneling probability can be 100%, why don't we observe it more often in daily life? Why don't we see people tunneling through a room?

Answer: Remember that our analysis uses the **time-independent** Schrödinger equation. In the time-dependent case, our analysis is equivalent to a **steady-state** analysis, assuming the wavefunction to oscillate in time forever like $\exp(-i\omega t)$. Resonant tunneling occurs only if the frequency ω matches the resonant frequency of the cavity very well, especially if the potential barriers are high. An input wave with a precise frequency is possible only if it lasts for a very long time, according to the uncertainty relation for time and frequency. This means that resonant tunneling in reality takes an extremely long time of oscillation with a precise frequency for the transmission to be completed, and we don't

Feel around in the Nanoworld



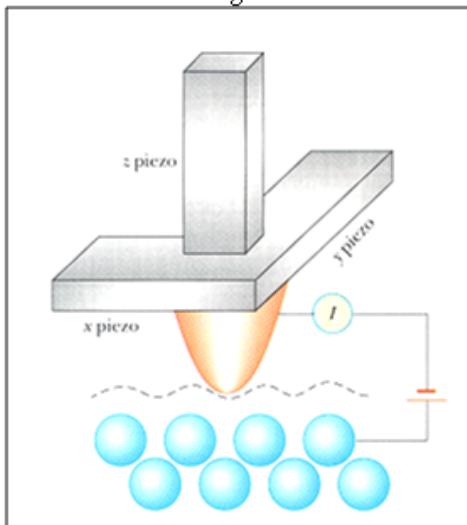
Scanning Tunneling Microscope (STM) works by sensing the tunneling current between a sharp tip and a conducting surface when the tip is brought close to the surface.



G. Binnig, H. Rohrer, Scientific American, Vol 253, Aug. 1985, page 40-46

FIGURE 5.8. From PC2232 AY13/14

STM schematic diagram.



- A conducting **very sharp tip** ($< 1 \text{ nm}$) is positioned very close to the surface of a specimen (about a nanometre). Electrons will tunnel through the barrier with an exponential probability with surface/tip distance.
- If the XYZ piezoelectric drivers receive a feedback signal to keep the tip current constant, then the distance will also be constant.
- If the tip is raster scanned across the surface, **keeping the tip current constant**, then tip will follow the surface atom profile.
- **The tip can be accurately positioned because of the exponential current change with the tip-surface distance.**

• **The STM can measure the height of surface features to within 0.001 nm, approximately 1/100 of an atomic diameter.**

FIGURE 5.9. From PC2232 AY13/14

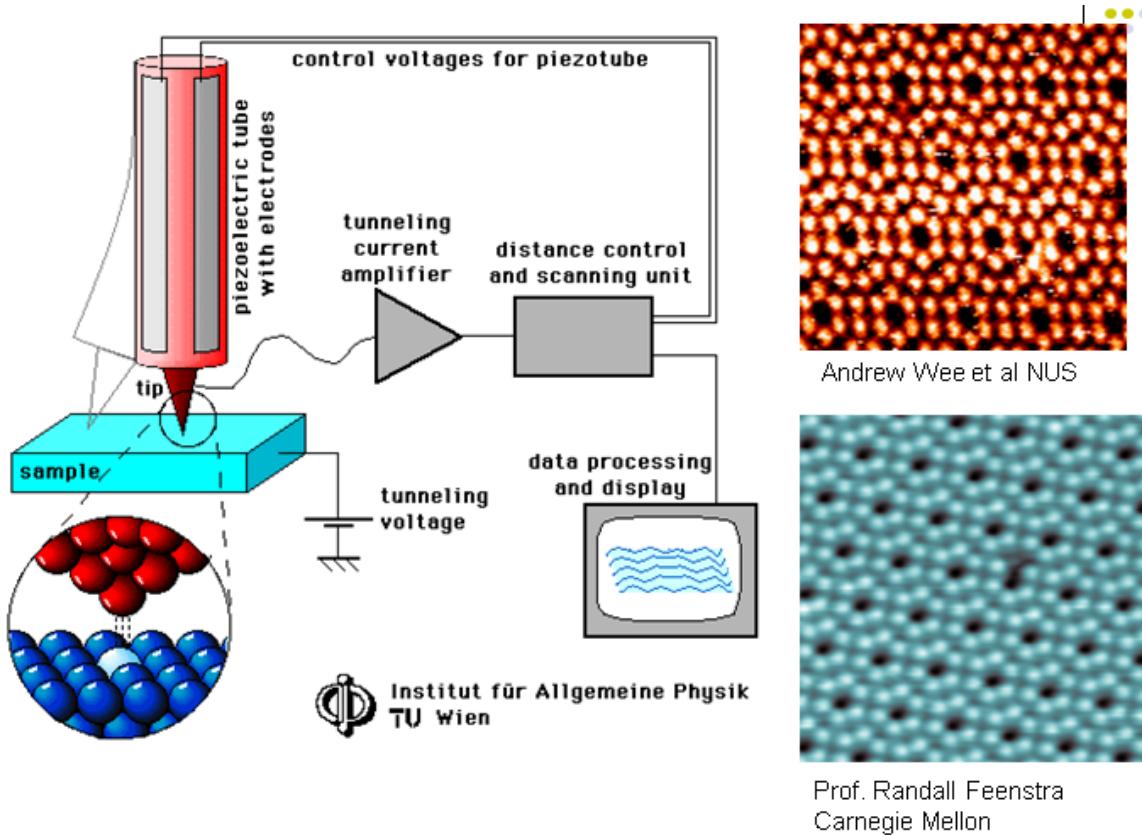


FIGURE 5.10. From PC2232 AY13/14

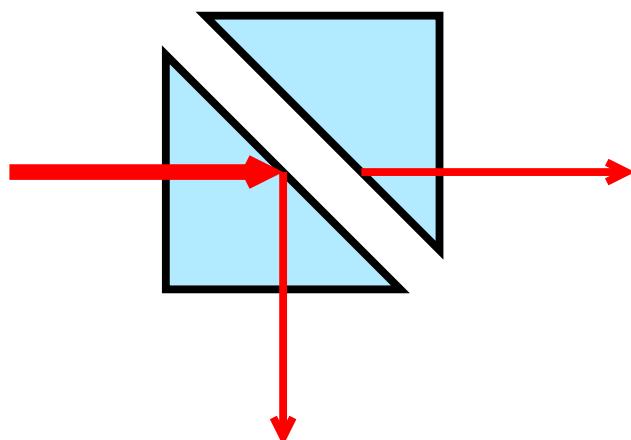


FIGURE 5.11. Frustrated total internal reflection is an optical tunneling phenomenon. An optical evanescent wave exists in the gap when the incident angle is beyond the critical angle, and this evanescent wave can be transmitted as a traveling wave if it meets another dielectric.

observe it in everyday life if we don't wait for that long or our wavefunctions don't oscillate precisely enough.

Resonant tunneling can actually be observed and exploited in nano-electronic devices; see, for example, http://en.wikipedia.org/wiki/Resonant-tunneling_diode. A similar resonant tunneling phenomenon can occur in optics if the input wave is resonant with a Fabry-Perot slab inside the forbidden region, as shown in Fig. 5.13.

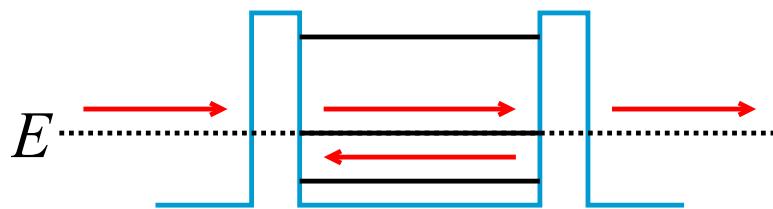


FIGURE 5.12. Resonant tunneling through two barriers. The two barriers act like a cavity and if the energy of the input wave is resonant with the waves inside the cavity, it is possible to obtain 100% transmission.

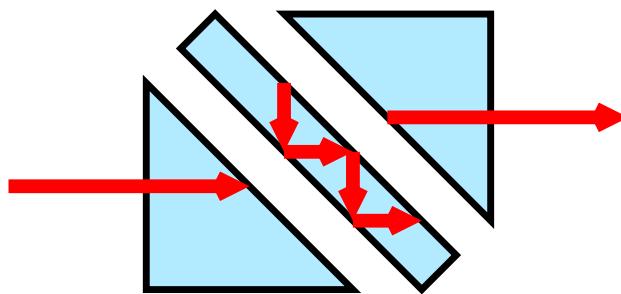


FIGURE 5.13. Optical resonant tunneling. An input wave can be coupled to a Fabry-Pérot slab through evanescent coupling. If the input is resonant with the waves inside the slab, the transmission on the other side can be 100% in the lossless case.

CHAPTER 6

Coupled Potential Wells

6.1. Energy-level splitting

Consider two identical potential wells separated by a certain distance, as shown in Fig. 6.1. For simplicity, suppose that each well has only one bound state with one energy level E . The standing-wave solution for the time-independent Schrödinger equation that we solved for one well in Sec. 3.4 is not valid here anymore, because the wave in one well can tunnel to the other well in time.

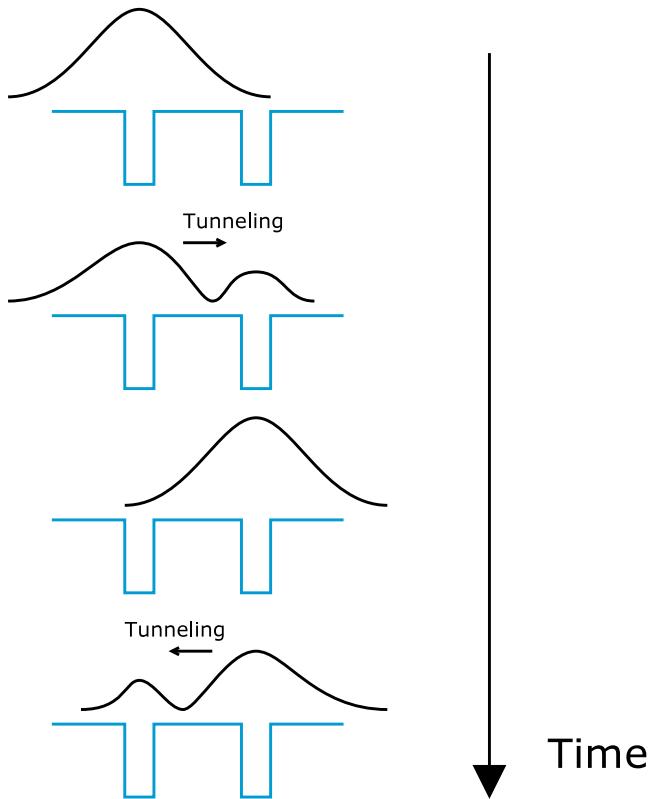


FIGURE 6.1. Oscillations of the wave between the two wells in the time domain via tunneling. The probability densities are plotted.

Let's consider the time-dependent case according to the **time-dependent Schrödinger equation** given by Eq. (3.1). Suppose that we start with the standing-wave solution $\tilde{\psi}_1(x)$ that is valid for the left well as the initial wavefunction. If there were just this one well, the standing wave will oscillate in time like

$$\psi(x, t) = \tilde{\psi}_1(x) \exp(-iEt/\hbar). \quad (6.1)$$

With two wells, however, the wave will begin to tunnel to the right well. If we think about the probability density, as the tunneling goes on, there will be more and more probability of the particle appearing in the right well, as shown in Fig. 6.1.

What happens next requires us to solve Eq. (3.1), but I will just tell you what will happen: eventually most of the particle will tunnel to the right and the wavefunction at some time will look like the standing-wave solution of the right well only (let's say it is $\tilde{\psi}_2(x)$), and then it starts to tunnel back to the left well again. What this means is that the particle will experience **real oscillations** between the two wells, as the particle tunnels

back and forth between the two wells. The wavefunction in time will look something like

$$\psi(x, t) \approx [\tilde{\psi}_1(x) \cos(\Delta\omega t) + i\tilde{\psi}_2(x) \sin(\Delta\omega t)] \exp(-iEt/\hbar), \quad (6.2)$$

where $\tilde{\psi}_1(x)$ is the wave solution for the left well only (an example is sketched in top-left of Fig. 6.2), $\tilde{\psi}_2(x)$ is the wave solution for the right well only (top-right of Fig. 6.2), and E is the original energy for $\tilde{\psi}_1$ in the case of one well. $\Delta\omega$ is some sort of frequency that depends on how close the wells are and the height of the potential barrier; it determines the speed of the tunneling oscillation.

At times $t = 0, \pi/\Delta\omega, 2\pi/\Delta\omega, \dots$, $\sin(\Delta\omega t) = 0$, $\psi(x, t)$ is proportional to $\tilde{\psi}_1(x)$ only, the particle is mainly in the left well, while at $t = \pi/(2\Delta\omega), 3\pi/(2\Delta\omega), 5\pi/(2\Delta\omega), \dots$, $\cos(\Delta\omega t) = 0$, $\psi(x, t)$ is proportional to $\tilde{\psi}_2(x)$ only, and the particle is mainly in the right well. Any time in-between, the time-dependent solution $\psi(x, t)$ is a varying superposition of $\tilde{\psi}_1(x)$ and $\tilde{\psi}_2(x)$. $\cos(\Delta\omega t)$ controls the amplitude of $\tilde{\psi}_1(x)$ in the overall wavefunction, while $\sin(\Delta\omega t)$ controls the amplitude of $\tilde{\psi}_2(x)$.

What is the time-independent solution for the two wells then? We can rewrite the cosine and sine in Eq. (6.2) as complex exponentials:

$$\psi(x, t) \approx \left[\tilde{\psi}_1(x) \frac{e^{i\Delta\omega t} + e^{-i\Delta\omega t}}{2} + i\tilde{\psi}_2(x) \frac{e^{i\Delta\omega t} - e^{-i\Delta\omega t}}{2i} \right] \exp(-iEt/\hbar) \quad (6.3)$$

$$= \frac{1}{2} [\tilde{\psi}_1(x) + \tilde{\psi}_2(x)] e^{i\Delta\omega t - iEt/\hbar} + \frac{1}{2} [\tilde{\psi}_1(x) - \tilde{\psi}_2(x)] e^{-i\Delta\omega t - iEt/\hbar} \quad (6.4)$$

$$= \frac{1}{2} [\tilde{\psi}_1(x) + \tilde{\psi}_2(x)] e^{-i(E/\hbar - \Delta\omega)t} + \frac{1}{2} [\tilde{\psi}_1(x) - \tilde{\psi}_2(x)] e^{-i(E/\hbar + \Delta\omega)t} \quad (6.5)$$

We end up having a superposition of two solutions that oscillate like $\exp(-j\omega t)$ with two different frequencies. The first component $\propto \tilde{\psi}_1 + \tilde{\psi}_2$ oscillate at a frequency of

$$\frac{E}{\hbar} - \Delta\omega \quad (6.6)$$

and the second component $\propto \tilde{\psi}_1 - \tilde{\psi}_2$ oscillate at a frequency of

$$\frac{E}{\hbar} + \Delta\omega. \quad (6.7)$$

Given these single-frequency solutions and these frequencies in the time-dependent case, we know that $\tilde{\psi}_1 + \tilde{\psi}_2$ will satisfy the time-**independent** Schrödinger equation with an energy

$$E - \hbar\Delta\omega, \quad (6.8)$$

and $\tilde{\psi}_1 - \tilde{\psi}_2$ will satisfy the time-independent equation with an energy

$$E + \hbar\Delta\omega. \quad (6.9)$$

Thus, when we put two potential wells next to each other, we get two energy levels that are slightly different from the original energy for each well, as shown in Fig. 6.2. This is known as **energy level splitting**, as the presence of a potential well closeby perturbs an energy level and splits it into two.

6.2. *Coupled-wave analysis

I would now like to explain in detail how to get Eq. (6.2) using something called coupled-wave or coupled-mode analysis. In the context of solid-state physics it is similar to something called a tight-binding analysis. This method is also very useful in optics to analyze coupled waveguides and coupled resonators, so let me introduce it here.

We start with the time-dependent Schrödinger equation for a wavefunction $\psi(x, t)$ in one space dimension:

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + U(x)\psi. \quad (6.10)$$

The potential is assume to be the sum of two wells, one on the left, on the right, as shown in Fig. 6.3:

$$U(x) = U_1(x) + U_2(x). \quad (6.11)$$

The analysis will work as long as for any form of $U_1(x)$ and $U_2(x)$, but you can think of it as two square wells in Fig. 6.3 if you want.

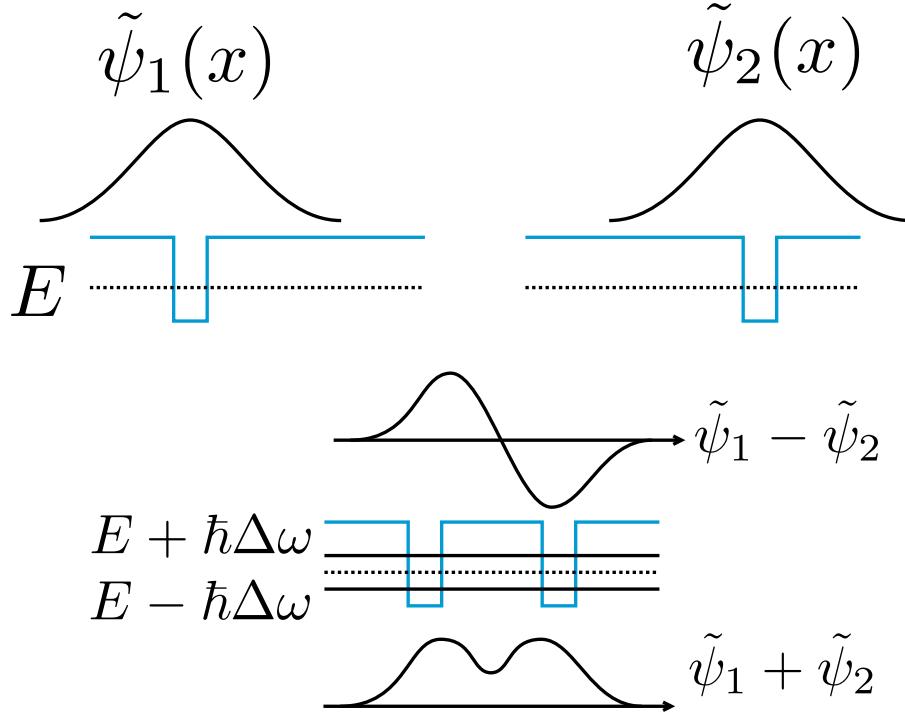


FIGURE 6.2. Top: $\tilde{\psi}_1(x)$ is the wavefunction for one well with energy E , and $\tilde{\psi}_2(x)$ is the wavefunction for the other well with the same energy. When the two wells are put close to each other, the wavefunction for the time-independent Schrödinger equation is a superposition of the two, and there are two energy levels instead: one slightly below the original, one slightly above.

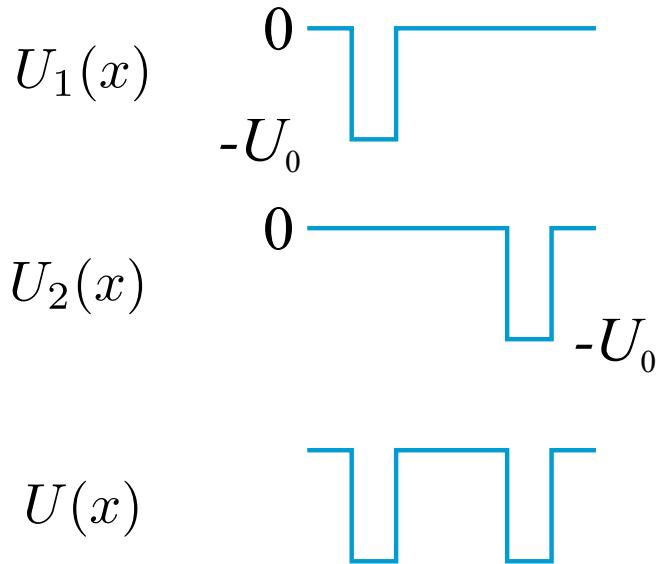


FIGURE 6.3. The potential for two wells can be expressed as a sum of two potential wells.

To find an approximate solution of Eq. (6.10), we start by assuming that we have found the standing-wave solutions if there were just one well. For $U_1(x)$ alone, we assume that we have found a $\tilde{\psi}_1$ with energy E that satisfies the time-independent Schrödinger equation:

$$E\tilde{\psi}_1 = -\frac{\hbar^2}{2m} \frac{\partial^2 \tilde{\psi}_1}{\partial x^2} + U_1 \tilde{\psi}_1. \quad (6.12)$$

For simplicity, we will just assume that $U_2(x)$ is a shifted version of $U_1(x)$, so by itself $U_2(x)$ will lead to a solution $\tilde{\psi}_2(x)$ that is simply a shifted version of $\tilde{\psi}_1(x)$ with the same energy E . In other words, we assume that we have found a $\tilde{\psi}_2$ solution that satisfies the time-independent equation with $U_2(x)$ alone in the following way:

$$E\tilde{\psi}_2 = -\frac{\hbar^2}{2m} \frac{\partial^2 \tilde{\psi}_2}{\partial x^2} + U_2 \tilde{\psi}_2. \quad (6.13)$$

When $U(x) = U_1(x) + U_2(x)$, we might be inclined to write

$$\psi(x, t) = A_1 \tilde{\psi}_1(x) \exp(-i\omega t) + A_2 \tilde{\psi}_2(x) \exp(-i\omega t)? \quad (6.14)$$

where $\omega = E/\hbar$. You can check that this is not a solution of Eq. (6.10) however. This solution simply does not take into account the dynamics of tunneling. Instead, we will assume the following form:

$$\psi(x, t) = A_1(t) \tilde{\psi}_1(x) \exp(-i\omega t) + A_2(t) \tilde{\psi}_2(x) \exp(-i\omega t). \quad (6.15)$$

$A_1(t)$ and $A_2(t)$ are now time-dependent coefficients that we'd like to solve. Plugging this into Eq. (6.10) and using the product rule of differentiation, the left-hand side becomes

$$i\hbar \frac{\partial \psi}{\partial t} = i\hbar \left(\frac{dA_1}{dt} \tilde{\psi}_1 e^{-i\omega t} - i\omega A_1 \tilde{\psi}_1 e^{-i\omega t} + \frac{dA_2}{dt} \tilde{\psi}_2 e^{-i\omega t} - i\omega A_2 \tilde{\psi}_2 e^{-i\omega t} \right) \quad (6.16)$$

$$= \left(i\hbar \frac{dA_1}{dt} \tilde{\psi}_1 + EA_1 \tilde{\psi}_1 + i\hbar \frac{dA_2}{dt} \tilde{\psi}_2 + EA_2 \tilde{\psi}_2 \right) e^{-i\omega t}, \quad (6.17)$$

while the right-hand side is

$$\begin{aligned} -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial x^2} + (U_1 + U_2) \psi &= \left[A_1 \left(-\frac{\hbar^2}{2m} \frac{\partial^2 \tilde{\psi}_1}{\partial x^2} + U_1 \tilde{\psi}_1 + U_2 \tilde{\psi}_1 \right) \right. \\ &\quad \left. + A_2 \left(-\frac{\hbar^2}{2m} \frac{\partial^2 \tilde{\psi}_2}{\partial x^2} + U_1 \tilde{\psi}_2 + U_2 \tilde{\psi}_2 \right) \right] e^{-i\omega t}. \end{aligned} \quad (6.18)$$

Equating both sides and divide both sides by $e^{-i\omega t}$,

$$\begin{aligned} i\hbar \frac{dA_1}{dt} \tilde{\psi}_1 + EA_1 \tilde{\psi}_1 + i\hbar \frac{dA_2}{dt} \tilde{\psi}_2 + EA_2 \tilde{\psi}_2 &= A_1 \left(-\frac{\hbar^2}{2m} \frac{\partial^2 \tilde{\psi}_1}{\partial x^2} + U_1 \tilde{\psi}_1 + U_2 \tilde{\psi}_1 \right) \\ &\quad + A_2 \left(-\frac{\hbar^2}{2m} \frac{\partial^2 \tilde{\psi}_2}{\partial x^2} + U_1 \tilde{\psi}_2 + U_2 \tilde{\psi}_2 \right). \end{aligned} \quad (6.19)$$

This looks horrendous, but we can actually use Eqs. (6.12) and (6.13) to simplify the right-hand side, as $\tilde{\psi}_1$ and $\tilde{\psi}_2$ are solutions of the one-well time-independent Schrödinger equations. We get

$$i\hbar \frac{dA_1}{dt} \tilde{\psi}_1 + EA_1 \tilde{\psi}_1 + i\hbar \frac{dA_2}{dt} \tilde{\psi}_2 + EA_2 \tilde{\psi}_2 = A_1 (E \tilde{\psi}_1 + U_2 \tilde{\psi}_1) + A_2 (E \tilde{\psi}_2 + U_1 \tilde{\psi}_2). \quad (6.20)$$

There are terms appearing on both sides ($EA_1 \tilde{\psi}_1$ and $EA_2 \tilde{\psi}_2$) that cancel. If we also divide both sides by $i\hbar$, we get

$$\frac{dA_1}{dt} \tilde{\psi}_1 + \frac{dA_2}{dt} \tilde{\psi}_2 = -\frac{i}{\hbar} (A_1 U_2 \tilde{\psi}_1 + A_2 U_1 \tilde{\psi}_2). \quad (6.21)$$

We really want to know $A_1(t)$ and $A_2(t)$ as a function of t , but here they are coupled to functions $\tilde{\psi}_1(x)$, $\tilde{\psi}_2(x)$, and $U_1(x)$ and $U_2(x)$ that depend on x . To eliminate the x -dependence in Eq. (6.21), we can do two things:

- (1) Multiply both sides by $\tilde{\psi}_1^*(x)$ and integrate over all x :

$$\frac{dA_1}{dt} \int_{-\infty}^{\infty} dx |\tilde{\psi}_1|^2 + \frac{dA_2}{dt} \int_{-\infty}^{\infty} dx \tilde{\psi}_1^* \tilde{\psi}_2 = -\frac{i}{\hbar} \left(A_1 \int_{-\infty}^{\infty} dx U_2 |\tilde{\psi}_1|^2 + A_2 \int_{-\infty}^{\infty} dx \tilde{\psi}_1^* U_1 \tilde{\psi}_2 \right). \quad (6.22)$$

To simplify this, we make the following assumptions:

(a) If $\tilde{\psi}_1$ is already normalized, the first integral on the left-hand side is 1:

$$\int_{-\infty}^{\infty} dx |\tilde{\psi}_1|^2 = 1. \quad (6.23)$$

(b) For the next integral, we assume that $\tilde{\psi}_1$ and $\tilde{\psi}_2$ do not have much overlap in space, which would mean

$$\left| \int_{-\infty}^{\infty} dx \tilde{\psi}_1^* \tilde{\psi}_2 \right| \ll \int_{-\infty}^{\infty} dx |\tilde{\psi}_1|^2. \quad (6.24)$$

For example, if you look at Fig. 6.2, $\tilde{\psi}_1^*(x)\tilde{\psi}_2(x)$ should be a small number overall because their peaks don't overlap, whereas $|\tilde{\psi}_1(x)|^2$ can be larger because the peaks of $\tilde{\psi}_1^*(x)$ and $\tilde{\psi}_1(x)$ coincide.

(c) For the integrals on the right-hand side of Eq. (6.22), we also assume

$$\left| \int_{-\infty}^{\infty} dx U_2 |\tilde{\psi}_1|^2 \right| \ll \left| \int_{-\infty}^{\infty} dx \tilde{\psi}_1^* U_1 \tilde{\psi}_2 \right|. \quad (6.25)$$

This is because $|\tilde{\psi}_1|^2$ has a suppressed tail near the right well so $|\tilde{\psi}_1|^2 U_2$ has a small magnitude, whereas $\tilde{\psi}_1^* \tilde{\psi}_2$ is relatively larger near the left well as $\tilde{\psi}_1$ has its peak there.

The assumptions above basically say that there is not much overlap between the two single-well solutions. Eq. (6.22) becomes

$$\frac{dA_1}{dt} \approx -\frac{i}{\hbar} A_2 \int_{-\infty}^{\infty} dx \tilde{\psi}_1^* U_1 \tilde{\psi}_2. \quad (6.26)$$

Let us define a coupling coefficient κ_{12} as

$$\kappa_{12} \equiv \frac{1}{\hbar} \int_{-\infty}^{\infty} dx \tilde{\psi}_1^* U_1 \tilde{\psi}_2, \quad (6.27)$$

so that we can write a simple ordinary differential equation in terms of A_1 and A_2 :

$$\frac{dA_1}{dt} \approx -i\kappa_{12} A_2. \quad (6.28)$$

For simplicity, we will assume that κ_{12} is real.

(2) We can also multiply Eq. (6.21) with $\tilde{\psi}_2^*(x)$ and integrate over x ,

$$\frac{dA_1}{dt} \int_{-\infty}^{\infty} dx \tilde{\psi}_2^* \tilde{\psi}_1 + \frac{dA_2}{dt} \int_{-\infty}^{\infty} dx |\tilde{\psi}_2|^2 = -\frac{i}{\hbar} \left(A_1 \int_{-\infty}^{\infty} dx \tilde{\psi}_2^* U_2 \tilde{\psi}_1 + A_2 \int_{-\infty}^{\infty} dx \tilde{\psi}_2^* U_1 \tilde{\psi}_2 \right), \quad (6.29)$$

and make approximations similar to above. The result is

$$\frac{dA_2}{dt} \approx -i\kappa_{21} A_1, \quad \kappa_{21} \equiv \frac{1}{\hbar} \int_{-\infty}^{\infty} dx \tilde{\psi}_2^* U_2 \tilde{\psi}_1. \quad (6.30)$$

With the two wells being shifted versions of each other and otherwise identical, it turns out that

$$\kappa_{21} = \kappa_{12}^*, \quad (6.31)$$

and since κ_{12} is already assumed to be real, its conjugate is the same as itself, and $\kappa_{21} = \kappa_{12}$.

We can now write Eq. (6.28) and (6.30) side-by-side:

$$\frac{dA_1}{dt} \approx -i\kappa_{12} A_2, \quad \frac{dA_2}{dt} \approx -i\kappa_{12} A_1. \quad (6.32)$$

You probably know how to solve this system of differential equations. The simplest way is to do another derivative:

$$\frac{d^2 A_1}{dt^2} \approx -i\kappa_{12} \frac{dA_2}{dt} \approx -\kappa_{12}^2 A_1. \quad (6.33)$$

The solution is sines and cosines. If we assume that $A_1(0) = 1$ and $A_2(0) = 0$,

$$A_1(t) \approx A_1(0) \cos(\kappa_{12}t) \quad A_2(t) \approx -iA_1(0) \sin(\kappa_{12}t). \quad (6.34)$$

Alternatively, we can write it in terms of complex exponentials:

$$A_1(t) \approx \frac{1}{2} A_1(0) e^{i\kappa_{12} t} + \frac{1}{2} A_1(0) e^{-i\kappa_{12} t}, \quad (6.35)$$

$$A_2(t) \approx -\frac{1}{2} A_1(0) e^{i\kappa_{12} t} - \frac{1}{2} A_1(0) e^{-i\kappa_{12} t}. \quad (6.36)$$

The key point here is that $A_1(t)$ and $A_2(t)$ oscillate at two frequencies: κ_{12} and $-\kappa_{12}$. The final solution for $\psi(x, t)$, going back to Eq. (6.15), is

$$\psi = A_1(t) \tilde{\psi}_1(x) e^{-i\omega t} + A_2(t) \tilde{\psi}_2(x) e^{-i\omega t} \quad (6.37)$$

$$\approx \frac{1}{2} [\tilde{\psi}_1(x) - \tilde{\psi}_2(x)] e^{-i(\omega - \kappa_{12})t} + [\tilde{\psi}_1(x) + \tilde{\psi}_2(x)] e^{-i(\omega + \kappa_{12})t}. \quad (6.38)$$

This suggests that the time-dependent $\psi(x, t)$ oscillates at two frequencies $\omega \pm \kappa_{12}$, one slightly below the original ω , one slightly above. These two frequencies are then associated with two energy levels split from the original one.

We can get back the solutions suggested in the previous section if κ_{12} is real and negative and we associate the frequency shift $\Delta\omega$ with the magnitude of the coupling coefficient $|\kappa_{12}|$:

$$\Delta\omega \equiv |\kappa_{12}| = \left| \int_{-\infty}^{\infty} dx \tilde{\psi}_1^* U_2 \tilde{\psi}_2 \right|. \quad (6.39)$$

The magnitude of the frequency shifts depends on the overlap between $\tilde{\psi}_1^*$, U_2 , and $\tilde{\psi}_2$.

6.3. *Multiple potential wells

What happens when we have multiple potential wells, such as that shown in Fig. 6.4?



FIGURE 6.4. Multiple potential wells with the same depth and same spacing between adjacent wells.

Suppose that there are N equally spaced potential wells with the same depth. It is not difficult to generalize the coupled-mode analysis in the previous section to multiple wells, if we assume

$$\psi(x, t) = \sum_{l=1}^N A_l(t) \tilde{\psi}_l(x) \exp(-i\omega t), \quad (6.40)$$

where $\tilde{\psi}_l(x)$ is a standing-wave solution for the n th well and $A_l(t)$ is a time-varying amplitude. Think of $A_l(t)$ as an **envelope** over each $\tilde{\psi}_l(x)$.

Making approximations similar to those in the previous section, it is possible to obtain the following coupled-mode equations:

$$\frac{d}{dt} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{pmatrix} \approx -i \begin{pmatrix} 0 & \kappa_{12} & 0 & 0 & \dots \\ \kappa_{12} & 0 & \kappa_{12} & 0 & \dots \\ 0 & \kappa_{12} & 0 & \kappa_{12} & \dots \\ \vdots & & & & \\ 0 & \dots & 0 & \kappa_{12} & 0 \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_N \end{pmatrix}, \quad (6.41)$$

where κ_{12} is the coupling coefficient that characterizes the overlap between two nearby single-well wavefunctions; we have assumed that the overlap between single-well wavefunctions farther apart are negligible. The

key point about the solution for this system of differential equations is that each $A_l(t)$ will be a superposition of N frequency components. For large N , the solution can be approximated as¹

$$A_l(t) \approx \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} B_n \exp \left[\frac{i2\pi n(l-1)}{N} \right] \exp(-i\Delta\omega_n t), \quad (6.42)$$

where B_n are the coefficients determined by discrete Fourier transform of the initial conditions for $A_l(0)$ at $t = 0$:

$$B_n = \frac{1}{\sqrt{N}} \sum_{l=1}^N A_l(0) \exp \left[-\frac{i2\pi n(l-1)}{N} \right]. \quad (6.43)$$

These frequencies $\Delta\omega_n$ are the **eigenvalues** of the matrix in Eq. (6.41) and correspond to N frequency shifts to the original frequency ω for one well. For large N , the eigenvalues can be shown to be

$$\Delta\omega_n \approx 2\kappa_{12} \cos \left(\frac{2\pi n}{N} \right), \quad n = 0, 1, \dots, N-1, \quad (6.44)$$

which means that all the frequency shifts are always bounded **in a band**:

$$-2\kappa_{12} \leq \Delta\omega_n \leq 2\kappa_{12}. \quad (6.45)$$

Some examples are plotted in Fig. 6.5. The integer n plays a similar role to the n in infinite square well; just like how n determines the momentum k in the square well, here it determines the effective wavenumber, also called “quasi-momentum,” for the envelope.

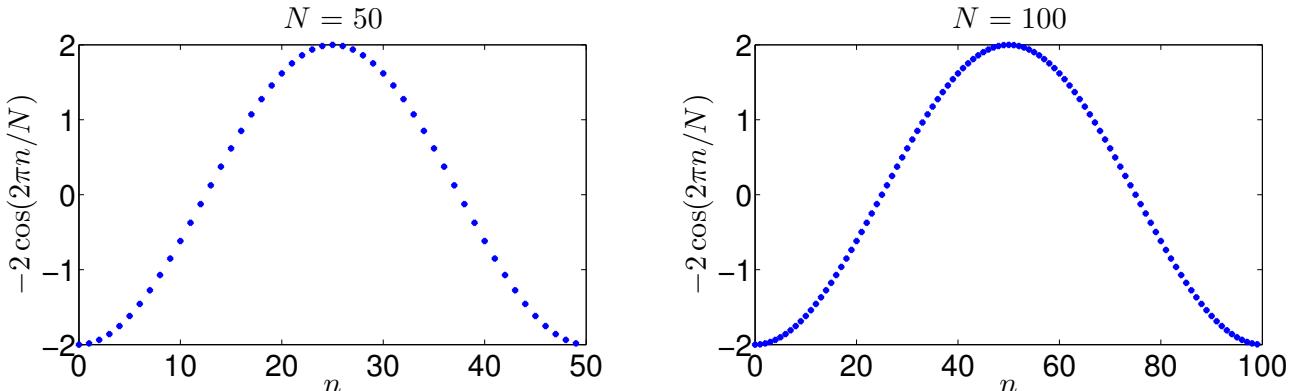


FIGURE 6.5. Plot of frequency shifts $\Delta\omega_n = 2\kappa_{12} \cos(2\pi n/N)$ versus n , assuming $\kappa_{12} = -1$. For larger N , there are simply more closely spaced eigenvalues within the same band $-2 \leq \Delta\omega_n \leq 2$.

If we increase the number of wells N , while keeping the spacing between adjacent wells and thus κ_{12} constant, the increasing number of frequency shifts $\Delta\omega_n$ become more and more closely spaced within the band. In terms of energy levels, it just means that N coupled potential wells lead to each single-well energy split into N levels, and **as the number of wells increases, there are more and more energy levels within an energy band**.

Meanwhile, the wavefunction that corresponds to each new energy level is a **superposition** of all the waves from all wells. Mathematically, they are given by $\sum_l A_l \psi_l(x)$ when A_l is an **eigenvector** of the matrix in Eq. (6.41). Some examples are shown in Fig. 6.6. If you look at these plots, they look kinda like sinusoidal waves on a larger scale. This is because A_l acts like an envelope that modulates each $\psi_l(x)$.

¹The solution comes from assuming that the matrix in Eq. (6.41) can be approximated as a circulant matrix (http://en.wikipedia.org/wiki/Circulant_matrix) for large N , in which case its eigenvalues and eigenvectors are easy to solve.

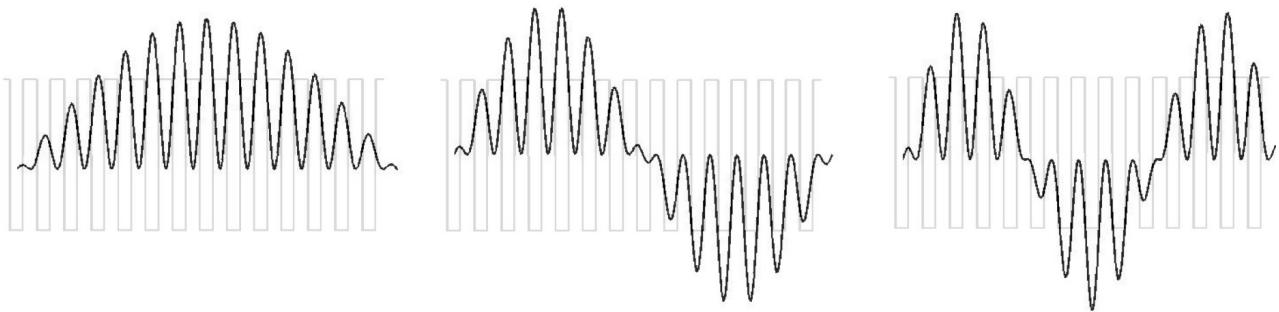


FIGURE 6.6. Some time-independent solutions for the wavefunctions of the coupled potential wells. They are global superpositions of waves from the individual wells. Figures from http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-007-electromagnetic-energy-from-motors-to-lasers-spring-2011/lecture-notes/MIT6_007S11_lec45.pdf.

6.4. Energy bands and bandgaps

Let us summarize what we have learned from the coupled-wave analysis:

- Consider a certain wavefunction in one potential well with a certain energy E .
- When two potential wells are brought near each other, the single-well wavefunction is no longer the solution for two wells. Instead, two new solutions appear with energies close to E . The energy level E is split into two, one slightly lower than E , one slightly higher.
- The amount of energy splitting is a function of how close the two wells are.
- If we bring N potential wells together, each single-well energy will be split into N energy levels.
- If the separation between adjacent wells is kept constant, the N energy levels are always bounded within an **energy band**.
- As N increases, the increasing number of energy levels become more and more closely spaced within the band.

While we've done the coupled-wave analysis using just one space dimension, we could have done it for three space dimensions and the features above in fact still hold. Consider an atom, which consists of a positively charged nucleus and some negatively charged electrons attracted to the nucleus. For the electrons, the nucleus acts as a 3D potential well, and the electron energy will also be quantized. The following will happen when we bring atoms together:

- When atoms are extremely far away from one another, we can consider them to be isolated and each atom will have discrete energy levels for their electrons. Two of these discrete energy levels, labeled $1s$ and $2s$, are shown in Fig. 6.7 for $r \rightarrow \infty$.
- Now let us bring two atoms together, as the two potential wells from the two positively-charged nuclei get closer to each other, *each* single-atom energy level is **split into two**, and the amount of splitting depends on the distance between the two, as shown in the left figure of 6.7 as r gets smaller.
- Now imagine N atoms brought together. **Each single-atom energy level splits into N levels**, and for a given separation r , these N energy levels have to be within an **energy band**. The center figure of 6.7 shows this for $N = 5$.
- As the number of atoms N increases, there are more and more energy levels. These energy levels become more and more closely spaced within a band for a given r .
- A typical solid has $N \sim 10^{23}$ atoms in ~ 1 gram. When there are so many energy levels within a band, it is as if almost any energy within the band is permissible. This gives rise to an **effectively continuous** spectrum of energies within a band, as shown in the right figure of 6.7.
- In a solid, the actual separation among atoms is the result of an equilibrium due to repulsions among the positive nuclei and attractions to the electrons that hold the nuclei together like glue. This separation determines the actual energy bands of a solid.

- In-between two energy bands, there can still be a **gap** of energy values that the electrons cannot have. Such gaps are called **bandgaps**, as shown in Fig. 6.8.

ether.

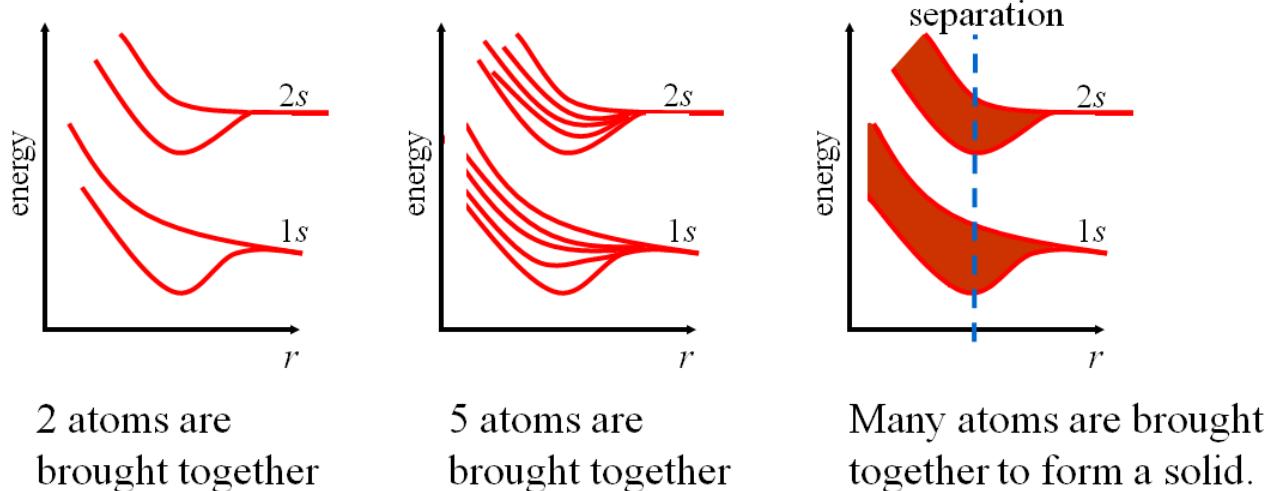


FIGURE 6.7. Many atoms brought together will form new energy levels split from each original single-atom energy levels. For a large number of atoms, these split energy levels are so close that they form energy bands that are effectively continuous.

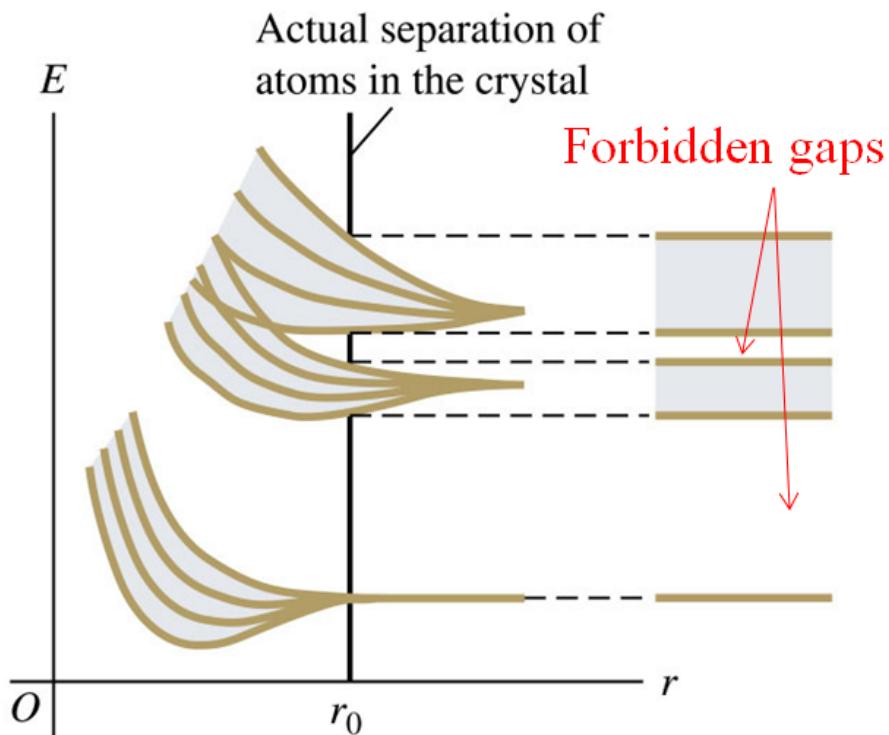


FIGURE 6.8. Many atoms brought together will form new energy levels split from each original single-atom energy levels. For a large number of atoms, these split energy levels are so close that they form energy bands that are effectively continuous. There are still gaps between the energy bands, and electrons cannot have energy within those bandgaps.

We will have more to say about energy bands and bandgaps, once we've learned more about atoms.

*For a large number of potential wells, there is a more accurate way of analyzing the energy bands of a crystal. It is called Bloch theory, which assumes that $U(x)$ is periodic. For an introduction, see, for example, Chap. 5.3 in Ref. [5].

CHAPTER 7

Three-Dimensional Square Well

The real world has three space dimensions, and after a lot of warm-up exercises with the one-dimensional time-independent Schrödinger equation we should finally tackle the 3D version. I will rewrite the time-independent Schrödinger equation introduced in Eq. (3.6) here:

$$E\psi(\mathbf{r}) = -\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + U(\mathbf{r})\psi(\mathbf{r}), \quad (7.1)$$

where

$$\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z} \quad (7.2)$$

is the three-dimensional position vector and

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (7.3)$$

is the Laplacian. The normalization condition is

$$\int_{\text{all space}} dV |\psi(\mathbf{r})|^2 = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \int_{-\infty}^{\infty} dz |\psi(x, y, z)|^2 = 1. \quad (7.4)$$

There are a few cases of the potential function $U(\mathbf{r})$ for which the Schrödinger equation can be solved exactly. We have already studied one example, the free-particle case with $U(\mathbf{r}) = 0$ everywhere, in Chap. 2. For another example, consider a 3D infinite square well, as shown in Fig. 7.1, which serves as a good model for **conduction electrons** in metals or semiconductors.

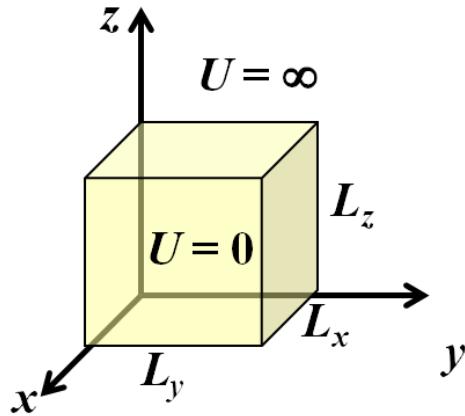


FIGURE 7.1. A three-dimensional infinite square well.

Consider a box with lengths L_x , L_y , and L_z , as shown in Fig. 7.1. Inside the box, the potential is 0; outside the box, the potential is ∞ . The infinite potential on the sides of the box means that the wavefunction must be zero there to maintain finite energy, just like the 1D infinite square well. The **boundary conditions** become

$$\psi(x = 0) = 0, \quad \psi(x = L_x) = 0, \quad \psi(y = 0) = 0, \quad \psi(y = L_y) = 0, \quad \psi(z = 0) = 0, \quad \psi(z = L_z) = 0. \quad (7.5)$$

7.1. Separable solutions

A common technique of solving a multi-variable partial differential equation is to assume the wavefunction is given by a product of single-variable functions. Here we assume

$$\psi(x, y, z) = X(x)Y(y)Z(z). \quad (7.6)$$

This kind of technique doesn't always work, but it turns out to work for the infinite square well. With this assumption, a second-order partial derivative looks like

$$\frac{\partial^2}{\partial x^2}\psi(x, y, z) = Y(y)Z(z)\frac{d^2X(x)}{dx^2}, \quad (7.7)$$

because Y and Z are assumed to be functions of y and z respectively and do not depend on x . Similarly,

$$\frac{\partial^2}{\partial y^2}\psi(x, y, z) = X(x)Z(z)\frac{d^2Y(y)}{dy^2}, \quad \frac{\partial^2}{\partial z^2}\psi(x, y, z) = X(x)Y(y)\frac{d^2Z(z)}{dz^2}. \quad (7.8)$$

Inside the box, $U = 0$, and the Schrödinger equation given by Eq. (7.1) becomes

$$EXYZ = -\frac{\hbar^2}{2m} \left(YZ\frac{d^2X}{dx^2} + XZ\frac{d^2Y}{dy^2} + XY\frac{d^2Z}{dz^2} \right). \quad (7.9)$$

Let's divide both sides by XYZ :

$$E = -\frac{\hbar^2}{2m} \left(\frac{1}{X}\frac{d^2X}{dx^2} + \frac{1}{Y}\frac{d^2Y}{dy^2} + \frac{1}{Z}\frac{d^2Z}{dz^2} \right). \quad (7.10)$$

If I re-arrange the equation in this way:

$$\frac{\hbar^2}{2m} \frac{1}{X(x)} \frac{d^2X(x)}{dx^2} = -E - \frac{\hbar^2}{2m} \left[\frac{1}{Y(y)} \frac{d^2Y(y)}{dy^2} + \frac{1}{Z(z)} \frac{d^2Z(z)}{dz^2} \right], \quad (7.11)$$

the left-hand side is a function of x only, and right-hand side is a function of y and z only. The only way for them to be equal is if **they are both equal to a constant**. With some hindsight, I will define this constant as

$$\frac{1}{X(x)} \frac{d^2X(x)}{dx^2} = -k_x^2. \quad (7.12)$$

Similar to Eq. (7.11), I can re-arrange Eq. (7.10) such that $(1/Y)d^2Y/dy^2$ is on the left-hand side and the rest of the terms are on the right. Again, the left-hand side would be a function of y only, while the right-hand side would be a function of x and z . This means that I can write

$$\frac{1}{Y(y)} \frac{d^2Y(y)}{dy^2} = -k_y^2. \quad (7.13)$$

Doing the same thing for $(1/Z)d^2Z/dz^2$, I get

$$\frac{1}{Z(z)} \frac{d^2Z(z)}{dz^2} = -k_z^2. \quad (7.14)$$

If we put these constants back in Eq. (7.10), we find that k_x , k_y , and k_z must obey the following relation with respect to the energy E :

$$E = \frac{\hbar^2(k_x^2 + k_y^2 + k_z^2)}{2m}. \quad (7.15)$$

Now, it should not surprise you that the solutions for Eqs. (7.12), (7.13) and (7.14) are

$$X = A_x \exp(ik_xx) + B_x \exp(-ik_xx), \quad (7.16)$$

$$Y = A_y \exp(ik_yy) + B_y \exp(-ik_yy), \quad (7.17)$$

$$Z = A_z \exp(ik_zz) + B_z \exp(-ik_zz), \quad (7.18)$$

for some complex constants A_x , B_x , A_y , B_y , A_z , and B_z and k_x , k_y , k_z assumed to be positive here. If we multiply these back together as $\psi = XYZ$, we get plane-wave solutions that we took for granted in Chap. 2. We did not have to go through the separable solution method if we accepted plane waves as solutions already, but the method will be indispensable when we study the hydrogen atom later.

7.2. Boundary conditions

We still need to match the boundary conditions given by Eqs. (7.5). We will just check the boundary conditions one-by-one. The first one is $\psi(x = 0) = 0$:

$$\begin{aligned}\psi(x = 0) &= X(x = 0)YZ \\ &= (A_x + B_x)[A_y \exp(ik_y y) + B_y \exp(-ik_y y)][A_z \exp(ik_z z) + B_z \exp(-ik_z z)] = 0.\end{aligned}\quad (7.19)$$

For this to hold at any y and z , we should have

$$B_x = -A_x. \quad (7.20)$$

By the same token, with $\psi(y = 0) = 0$ and $\psi(z = 0) = 0$,

$$B_y = -A_y, \quad B_z = -A_z, \quad (7.21)$$

and we get

$$\psi(x, y, z) = (2i)^3 A_x A_y A_z \sin(k_x x) \sin(k_y y) \sin(k_z z). \quad (7.22)$$

Now, we also have $\psi(x = L_x) = 0$ on the other side. This means that

$$\psi(x = L_x) = (2i)^3 A_x A_y A_z \sin(k_x L_x) \sin(k_y y) \sin(k_z z) = 0, \quad (7.23)$$

and we need $\sin(k_x L_x)$ to vanish there also, leading to

$$k_x L_x = \pi n_x, \quad n_x = 1, 2, \dots \quad (7.24)$$

Similarly,

$$k_y L_y = \pi n_y, \quad n_y = 1, 2, \dots, \quad k_z L_z = \pi n_z, \quad n_z = 1, 2, \dots \quad (7.25)$$

The integers n_x , n_y , and n_z denote the number of **half-waves** that the wavefunction has inside the box along each dimension. Some isosurface plots of the resulting probability densities can be found at http://www.nextnano.de/nextnano3/tutorial/3Dtutorial_QD_cubic.htm; each plot indicates the surfaces where the probability density is constant for a certain wavefunction.

7.3. Final solutions

In terms of the integers n_x, n_y, n_z , we get the following solution for the wavefunction:

$$\boxed{\psi(x, y, z) = A \sin\left(\frac{\pi n_x x}{L_x}\right) \sin\left(\frac{\pi n_y y}{L_y}\right) \sin\left(\frac{\pi n_z z}{L_z}\right)}, \quad (7.26)$$

where A is some constant that we can compute via the normalization condition. The energy according to Eq. (7.15) becomes

$$\boxed{E_{n_x, n_y, n_z} = \frac{\hbar^2 \pi^2}{2m} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2} \right)}. \quad (7.27)$$

Energy is again discrete, but we now have **three quantum numbers** n_x, n_y, n_z to denote each solution. This is a new feature for three space dimensions.

7.4. Energy degeneracy

The ground-state energy is

$$E_{1,1,1} = \frac{\hbar^2 \pi^2}{2m} \left(\frac{1}{L_x^2} + \frac{1}{L_y^2} + \frac{1}{L_z^2} \right). \quad (7.28)$$

If the box has equal lengths $L_x = L_y = L_z = L$, we have three possible combinations of the quantum numbers for the next lowest energy:

$$E_{1,1,2} = E_{1,2,1} = E_{2,1,1} = \frac{\hbar^2 \pi^2}{2m L^2} (2^2 + 1^2 + 1^2). \quad (7.29)$$

These energy levels are called **degenerate**, as shown in Fig. 7.2. Each combination of quantum numbers denotes different number of half-waves along the three dimensions, but the different wavefunctions end up

having the same energy. This phenomenon will be important when we study density of states and Fermi-Dirac statistics, where we care more about the energy and not so much about the quantum numbers.

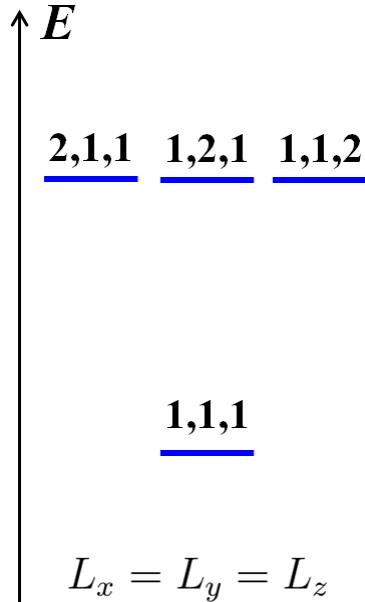


FIGURE 7.2. The lowest energy occurs for $n_x = n_y = n_z = 1$. For the next energy levels, three combinations of quantum numbers share the same energy, even though they correspond to different wavefunctions. These three wavefunctions are called degenerate states.

- **Exercise:** How many degenerate levels are there if

$$E = \frac{\hbar^2\pi^2}{2mL^2} (2^2 + 2^2 + 1^2) \quad (7.30)$$

7.5. Momentum-space diagram

We can denote the states in a 3D well using the **3D momentum space**. We know that the three momentum components are restricted to have magnitudes given by

$$k_x = \frac{\pi}{L_x}, \frac{2\pi}{L_x}, \frac{3\pi}{L_x}, \dots \quad k_y = \frac{\pi}{L_y}, \frac{2\pi}{L_y}, \frac{3\pi}{L_y}, \dots \quad k_z = \frac{\pi}{L_z}, \frac{2\pi}{L_z}, \frac{3\pi}{L_z}, \dots \quad (7.31)$$

We can then use a dot to denote each state in the three-dimensional momentum space, as shown in Fig. 7.3. Since we are simply looking at the magnitude of the momentum components, we can just focus on the octant where they are all positive.

For a given energy E , we know from Eq. (7.15) that the states must lie on the sphere that obeys the equation

$$k_x^2 + k_y^2 + k_z^2 = \frac{2mE}{\hbar^2}. \quad (7.32)$$

For a given E , we can then count the number of dots on the sphere with radius $\sqrt{2mE/\hbar^2}$ to compute the number of degenerate levels. This picture will be important when we study **density of states**.

- **Exercise:** Suppose that L_x, L_y, L_z are very large. How many states, approximately, are there in a k -space box with lengths $\Delta k_x, \Delta k_y$, and Δk_z ? Does it matter where the box is?
- **Exercise:** Again suppose that L_x, L_y, L_z are very large. How many states are there, approximately, that have energies E within the range $0 < E \leq E_{\max}$? This sort of calculation is useful later on when we discuss density of states.

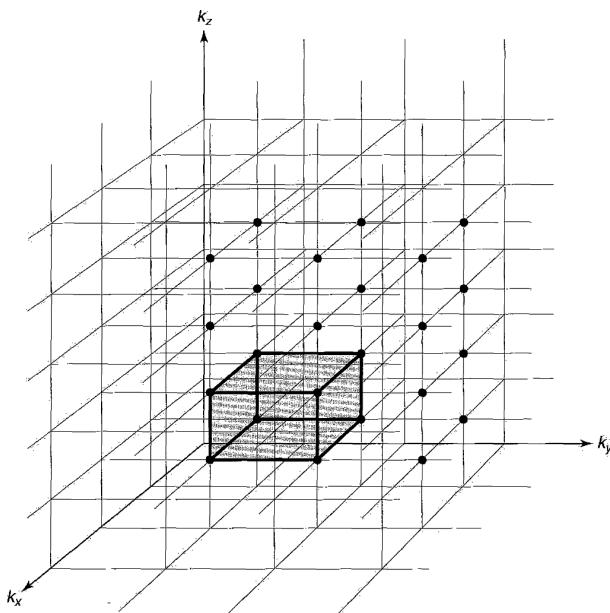


FIGURE 7.3. The discrete states in a 3D well can be denoted by dots in the momentum space. Figure from [5]. Note that each dot is separated from adjacent ones by a distance of π/L_x , π/L_y , or π/L_z in the \mathbf{k} space.

CHAPTER 8

Hydrogen Atom

The first triumph of the Schrödinger equation came when it was able to explain the energy levels of an electron in a hydrogen atom, which consists of an electron and a positively charged nucleus with charge $+e$. Since the nucleus is much heavier than the electron, we can assume that the nucleus is not moving in time. Assume that the nucleus is a point charge at $\mathbf{r} = 0$. The electric field from the nucleus attracts the electron towards the nucleus, and the potential for the electron is a Coulomb potential:

$$U(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0} \frac{e^2}{r}, \quad (8.1)$$

where r is the distance between the nucleus and the electron position \mathbf{r} , as shown in Fig. 8.1 and ϵ_0 is the free-space permittivity. If I plot the Coulomb potential along one direction, say, x , it would look like the right figure of 8.1. Note that $U(r)$ is always negative, $-\infty$ at the single point of $r = 0$, and $U(r) \rightarrow 0$ as $r \rightarrow \infty$. The resulting bound states must have negative energy E as well.

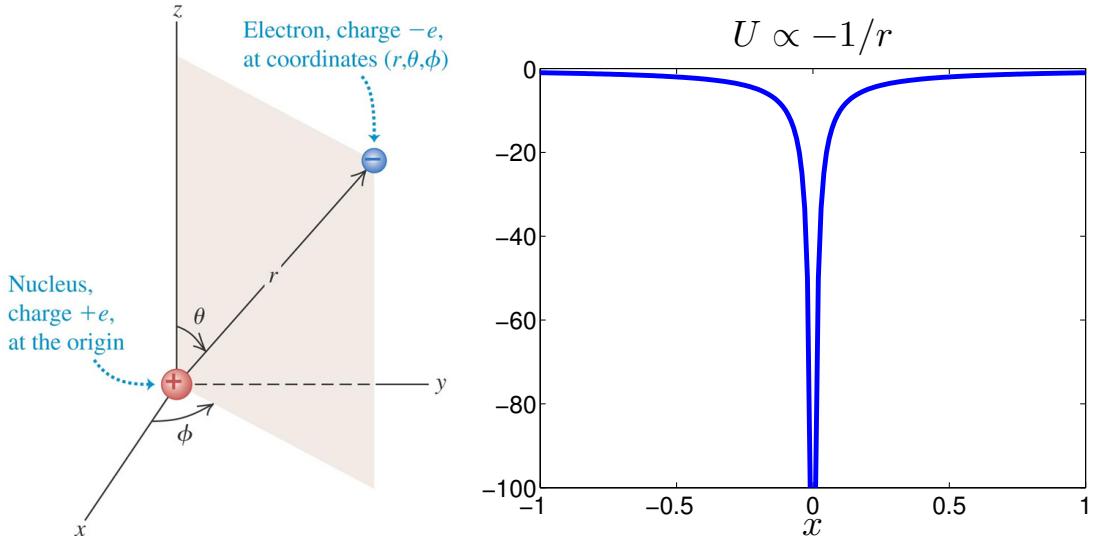


FIGURE 8.1. Left (from [2]): A model of hydrogen atom. The electron is trapped by a Coulomb potential that depends on the charges and the distance between them. Right: the Coulomb potential along one direction.

With $U(\mathbf{r})$ being a function of r only, it is begging us to use **spherical coordinates**, as shown in Fig. 8.2. Calculus in spherical coordinates is summarized in Fig. 8.3.

8.1. Separable solutions

With the Laplacian in spherical coordinates, the time-independent Schrödinger equation given by Eq. (7.1) can be rewritten as

$$\frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} + \frac{2m}{\hbar^2} [E - U(r)] \psi = 0. \quad (8.2)$$

This turns out to be easier to solve because $U(r)$ is a function of r only.

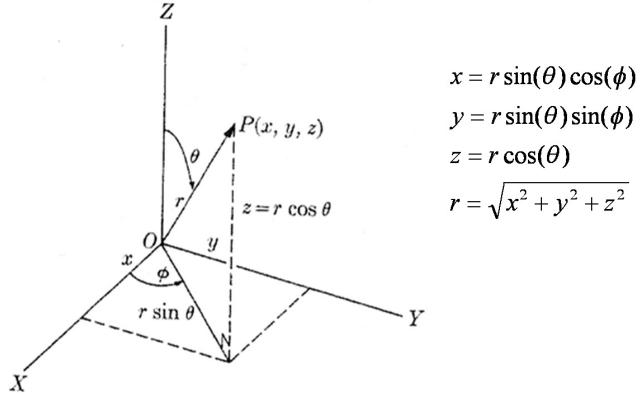


FIGURE 8.2. Spherical coordinates. From PC2232 AY13/14

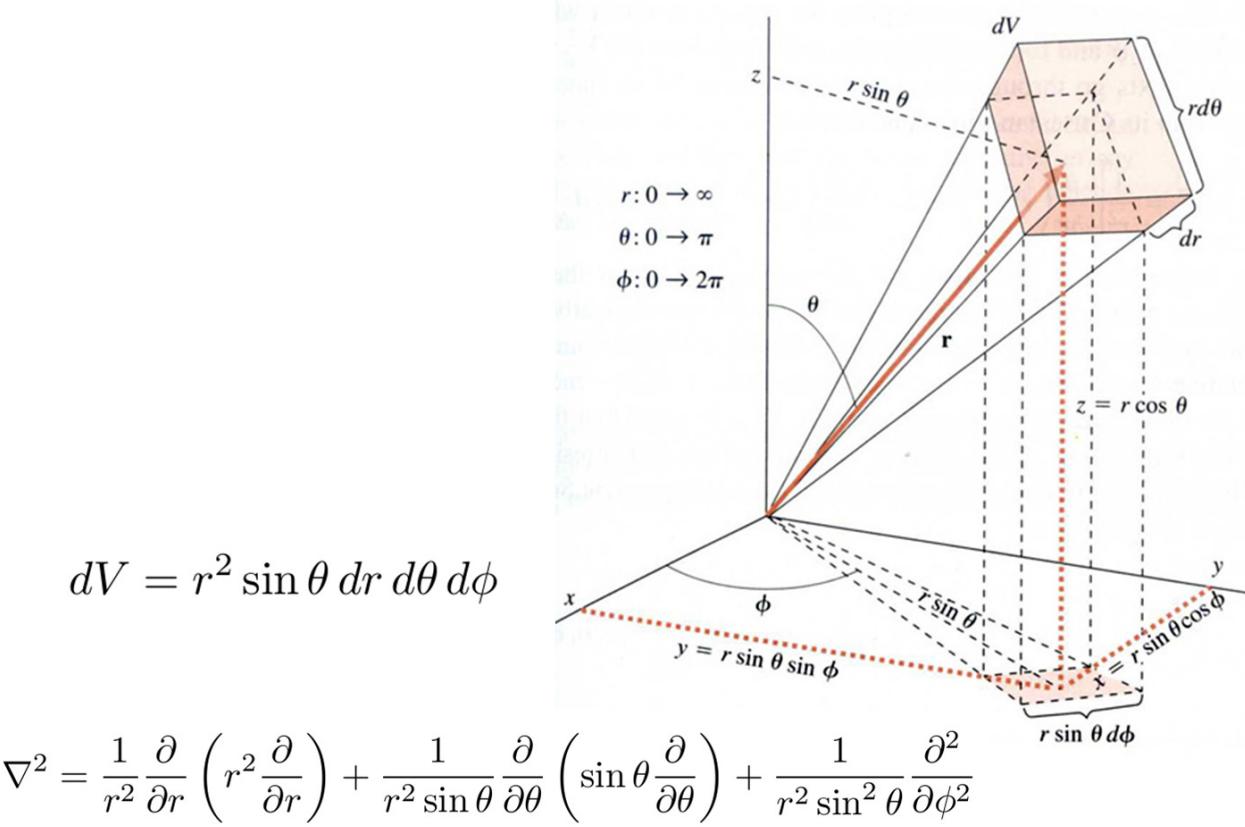


FIGURE 8.3. Calculus in spherical coordinates. From PC2232 AY13/14

Let's multiply both sides by $r^2 \sin^2 \theta$:

$$\sin^2 \theta \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) + \sin \theta \frac{\partial}{\partial \theta} \left(\sin \theta \frac{\partial \psi}{\partial \theta} \right) \frac{\partial^2 \psi}{\partial \phi^2} + \frac{2mr^2 \sin^2 \theta}{\hbar^2} [E - U(r)] \psi = 0. \quad (8.3)$$

To solve this, we will again turn to the assumption of separable functions: assume $\psi(\mathbf{r})$ to be a **product** of functions of r, θ, ϕ :

$$\psi(\mathbf{r}) = R(r)\Theta(\theta)\Phi(\phi). \quad (8.4)$$

With this assumption, we see that the partial derivatives can be written in terms of ordinary derivatives, such as

$$\frac{\partial \psi}{\partial r} = \Theta\Phi \frac{dR}{dr}, \quad \frac{\partial \psi}{\partial \theta} = R\Phi \frac{d\Theta}{d\theta}, \quad \frac{\partial \psi}{\partial \phi} = R\Theta \frac{d\Phi}{d\phi}, \quad \text{etc.} \quad (8.5)$$

Substituting $\psi = R\Theta\Phi$ into the Schrödinger equation,

$$\Theta\Phi \sin^2 \theta \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + R\Phi \sin \theta \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + R\Theta \frac{d^2\Phi}{d\phi^2} + \frac{2mr^2 \sin^2 \theta}{\hbar^2} [E - U(r)] R\Theta\Phi = 0, \quad (8.6)$$

and dividing both sides by $R\Theta\Phi$, we get

$$\frac{\sin^2 \theta}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \frac{1}{\Phi} \frac{d^2\Phi}{d\phi^2} + \frac{2mr^2 \sin^2 \theta}{\hbar^2} [E - U(r)] = 0. \quad (8.7)$$

Let's put $\frac{1}{\Phi} \frac{d^2\Phi}{d\phi^2}$ to the right-hand side:

$$\frac{\sin^2 \theta}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \frac{2mr^2 \sin^2 \theta}{\hbar^2} [E - U(r)] = -\frac{1}{\Phi} \frac{d^2\Phi}{d\phi^2}. \quad (8.8)$$

The left-hand side is a function of r and θ only, and the right-hand side is a function of ϕ only. This can happen only if both sides are equal to a constant. I will define the constant by

$$-\frac{1}{\Phi} \frac{d^2\Phi}{d\phi^2} = m_l^2. \quad (8.9)$$

The left-hand side of Eq. (8.8) is also equal to m_l^2 . Rearranging terms,

$$\frac{\sin^2 \theta}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{2mr^2 \sin^2 \theta}{\hbar^2} (E - U(r)) = m_l^2 - \frac{\sin \theta}{\Theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right), \quad (8.10)$$

and dividing both sides by $\sin^2 \theta$, we get

$$\frac{1}{R} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \frac{2mr^2}{\hbar^2} [E - U(r)] = \frac{m_l^2}{\sin^2 \theta} - \frac{1}{\Theta \sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right). \quad (8.11)$$

The left-hand side is a function of r only, and the right-hand side is a function of θ only. They must be equal to another constant; I'll assume that the constant is $l(l+1)$.

To summarize, assuming separable functions for $\psi(r, \theta, \phi) = R(r)\Theta(\theta)\Phi(\phi)$, we are able to separate the Schrödinger equation into three ordinary differential equations:

$$\frac{d^2\Phi}{d\phi^2} + m_l^2 \Phi = 0, \quad (8.12)$$

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left(\sin \theta \frac{d\Theta}{d\theta} \right) + \left[l(l+1) - \frac{m_l^2}{\sin^2 \theta} \right] \Theta = 0, \quad (8.13)$$

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \left\{ \frac{2m}{\hbar^2} [E - U(r)] - \frac{l(l+1)}{r^2} \right\} R = 0, \quad (8.14)$$

with two constants m_l and l .

8.2. Magnetic quantum number

Consider Eq. (8.12) first. The solution is

$$\boxed{\Phi(\phi) = \Phi(0) \exp(im_l\phi)}. \quad (8.15)$$

Consider a periodic function in this **azimuthal** direction, as shown in Fig. 8.4. Since $\phi + 2\pi$ is the same azimuthal coordinate as ϕ and indicate the same angle in space, the wavefunction really has just one value whether it is $\Phi(\phi)$ or $\Phi(\phi + 2\pi)$. This means that we must assume

$$\Phi(\phi) = \Phi(\phi + 2\pi), \quad (8.16)$$

resulting in

$$\Phi(0) \exp(im_l\phi) = \Phi(0) \exp[i(m_l\phi + 2\pi m_l)], \quad (8.17)$$

or

$$\exp(i2\pi m_l) = 1. \quad (8.18)$$

This implies that m_l must be an integer:

$$m_l = 0, \pm 1, \pm 2, \dots \quad (8.19)$$

m_l is called the magnetic quantum number. It is the number of periods for the wavefunction in the azimuthal ϕ direction.

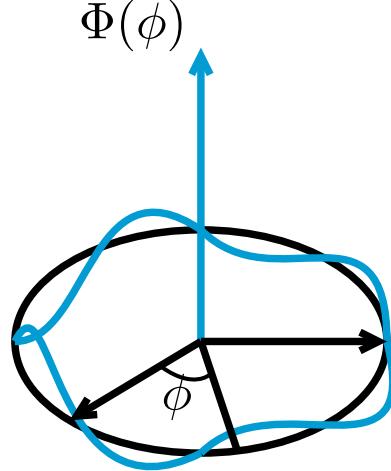


FIGURE 8.4. A wavefunction must be periodic with a period 2π in the azimuthal direction ϕ .

8.3. Orbital quantum number

Consider Eq. (8.13) next. The solution for $\Theta(\theta)$ depends on m_l and l . It can be shown that there is a solution to Eq. (8.13) provided that

$$|m_l| \leq l, \quad (8.20)$$

and l is a nonnegative integer. We can start counting by listing l :

$$l = 0, 1, 2, \dots, \quad (8.21)$$

and for each l , we list all the m_l such that

$$m_l = 0, \pm 1, \pm 2, \dots, \pm l. \quad (8.22)$$

For example, for $l = 3$, we can have $m_l = m_3 = -3, -2, -1, 0, 1, 2, 3$. The solutions are known as the associated Legendre functions, which depend on l and m_l . We will write them as $\Theta_{lm_l}(\theta)$; the first few are shown in Fig. 8.5.

$$\begin{aligned} l = 0 & \left[\Theta_{00}(\theta) = \frac{1}{\sqrt{2}} \right] \\ l = 1 & \left[\begin{array}{l} \Theta_{10}(\theta) = \frac{\sqrt{6}}{2} \cos \theta \\ \Theta_{1\pm 1}(\theta) = \frac{\sqrt{3}}{2} \sin \theta \end{array} \right] \\ l = 2 & \left[\begin{array}{l} \Theta_{20}(\theta) = \frac{\sqrt{10}}{4} (3 \cos^2 \theta - 1) \\ \Theta_{2\pm 1}(\theta) = \frac{\sqrt{15}}{2} \sin \theta \cos \theta \\ \Theta_{2\pm 2}(\theta) = \frac{\sqrt{15}}{4} \sin^2 \theta \end{array} \right] \\ m_l = 0, \pm 1, \pm 2, \dots, \pm l & \end{aligned}$$

FIGURE 8.5. Associated Legendre functions.

8.4. *Spherical harmonics

So far we have solved for $\Phi(\phi)$ and $\Theta(\theta)$ without specifying $U(r)$ at all. These solutions do not depend on r and is true for any potential that is a function of r only. If we combine them as follows:

$$Y_{l m_l}(\theta, \phi) = \Phi_{m_l}(\phi)\Theta_{l m_l}(\theta) \quad (8.23)$$

$Y_{l m_l}(\theta, \phi)$ are known as spherical harmonics. They often occur when one solves a wave equation with spherical symmetry. It is a function of the angle coordinates θ and ϕ and also two quantum numbers l and m_l . Spherical plots of these functions are shown in Fig. 8.6.

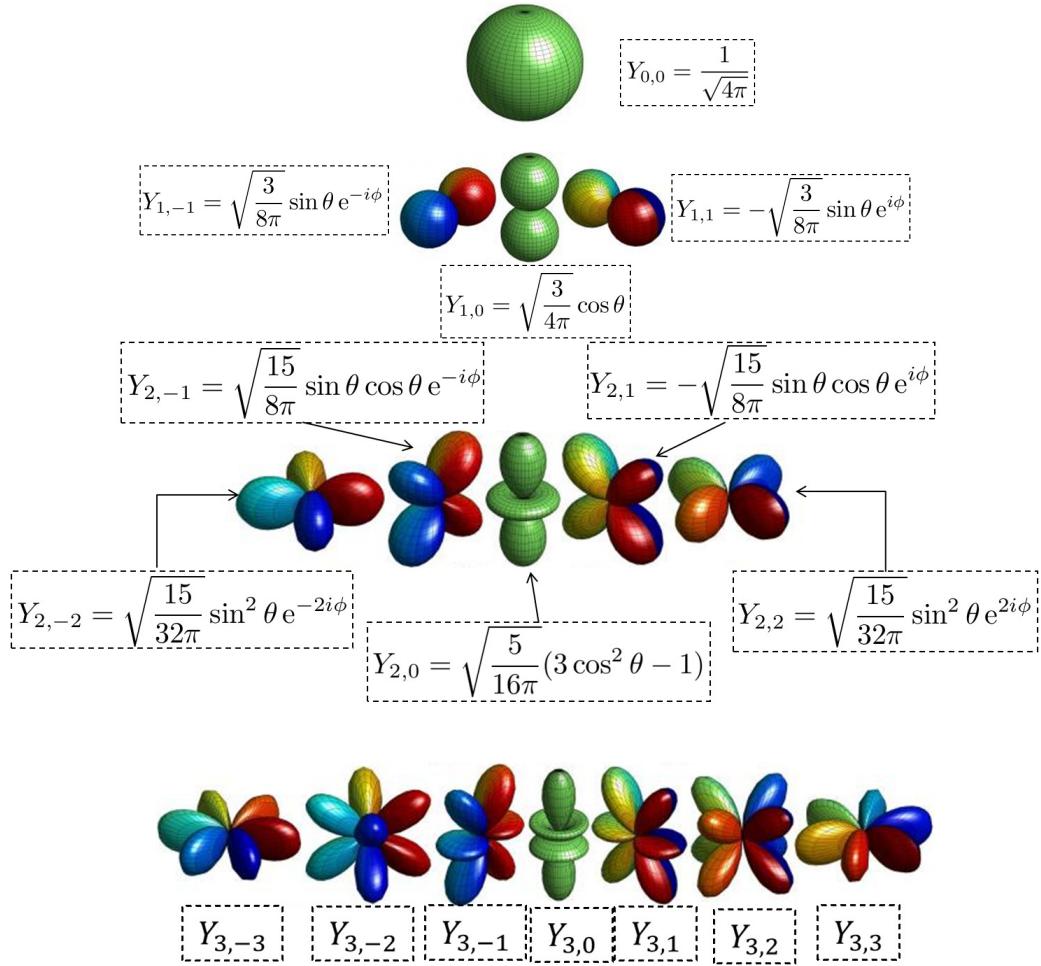


FIGURE 8.6. Some spherical plots of spherical harmonics $Y_{l m_l}(\theta, \phi)$. A spherical plot uses the radial coordinate to represent the function value; each point on the surface uses the distance from the origin to indicate the value $Y_{l m_l}(\theta, \phi)$ as a function of the angles given by θ and ϕ (<http://mathworld.wolfram.com/SphericalPlot.html>). The colors denote the varying phase of the complex functions. Pictures from http://zoi.utia.cas.cz/Spherical_harmonics_image.

The spherical harmonics are important for chemistry, electromagnetic scattering, image processing, etc., but actually not so important for us; the important point to note here is the quantum numbers l and m_l ; for each orbital quantum number l , there are $2l + 1$ magnetic quantum numbers m_l . They are like the quantum numbers n_x, n_y, n_z for a 3D infinite square well, except that l and m_l are for the angle coordinates and there is a restriction on the possible m_l given an l .

8.5. Principal quantum number

There is a final quantum number that we need to figure out. Consider Eq. (8.14):

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR}{dr} \right) + \left[\frac{2m}{\hbar^2} \left(E + \frac{e^2}{4\pi\epsilon_0 r} \right) - \frac{l(l+1)}{r^2} \right] R = 0. \quad (8.24)$$

Note that this equation depends on the orbital quantum number l . For bound states, it can be shown that the energy levels must be discrete and given by

$$E_n = -\frac{1}{(4\pi\epsilon_0)^2} \frac{me^4}{2\hbar^2} \left(\frac{1}{n^2} \right) = -\frac{13.60 \text{ eV}}{n^2}, \quad n = 1, 2, 3, \dots \quad (8.25)$$

where eV is the electron-Volt unit for energy. n is called the **principal quantum number**. Note how the energy spacings are different from those for square wells or harmonic oscillators. The ground state is $n = 1$, in which case $E_1 = -13.60 \text{ eV}$. The next energy level is $E_2 = -13.6/2^2 \text{ eV}$, etc., as shown in Fig. 8.7.

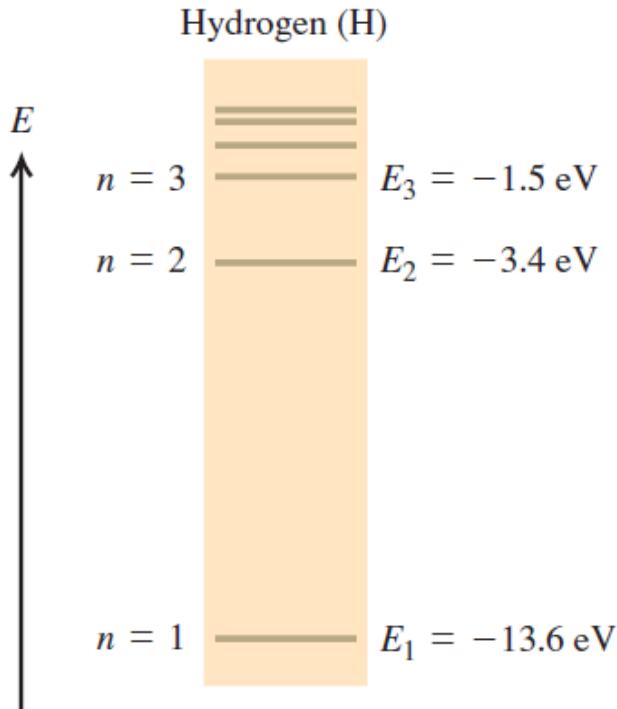


FIGURE 8.7. The three lowest energy levels of a hydrogen atom. From [2].

- **Exercise:** Suppose that the nucleus now has charge $+Ze$ instead of $+e$, where Z is an integer. How do the energy levels for an electron change?

For a given n , a bound-state solution also requires l to be smaller than n , or

$$l = 0, 1, \dots, n-1. \quad (8.26)$$

The radial solutions $R_{nl}(r)$ are associated Laguerre functions. The first few are shown in Fig. 8.8. Note that they all decay exponentially away from the center. The constant a_0 in the functions is

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} \approx 0.0529 \text{ nm}, \quad (8.27)$$

which tells you that the electron probability density in the radial direction has a width on the order of 1 Angstrom.

$$\begin{aligned}
R_{10}(r) &= \frac{1}{a_0^{3/2}} 2e^{-r/a_0} \\
R_{20}(r) &= \frac{1}{(2a_0)^{3/2}} 2 \left(1 - \frac{r}{2a_0}\right) e^{-r/2a_0} \\
R_{21}(r) &= \frac{1}{(2a_0)^{3/2}} \frac{r}{\sqrt{3}a_0} e^{-r/2a_0} \\
R_{30}(r) &= \frac{1}{(3a_0)^{3/2}} \left(2 - \frac{4r}{3a_0} + \frac{4r^2}{27a_0^2}\right) e^{-r/3a_0} \\
R_{31}(r) &= \frac{1}{(3a_0)^{3/2}} \frac{4\sqrt{2}r}{9a_0} \left(1 - \frac{r}{6a_0}\right) e^{-r/3a_0} \\
R_{32}(r) &= \frac{1}{(3a_0)^{3/2}} \frac{2\sqrt{2}r^2}{27\sqrt{5}a_0^2} e^{-r/3a_0}
\end{aligned}$$

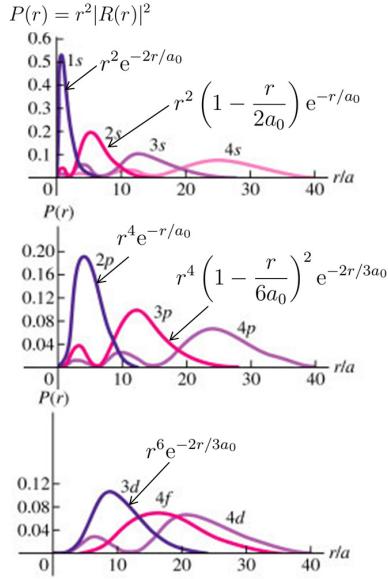


FIGURE 8.8. Left: first few associated Laguerre functions. Right: some plots of the radial probability densities. From PC2232 AY13/14

8.6. Quantum numbers

We see that 3 quantum numbers are required to specify the state of an electron in hydrogen atom. We start with n :

$$n = 1, 2, \dots \quad (\textbf{principal quantum number}). \quad (8.28)$$

Once n is given, we have a set of l :

$$l = 0, 1, 2, \dots, n - 1 \quad (\textbf{orbital quantum number}). \quad (8.29)$$

For each l , there is a set of m_l :

$$m_l = 0, \pm 1, \pm 2, \dots, \pm l \quad (\textbf{magnetic quantum number}). \quad (8.30)$$

For historic reasons, a letter instead of number is often used to refer to the orbital number l in the so-called spectroscopic notation, as shown in Table 8.1. For example, $(n = 3, l = 0)$ is called the $3s$ orbital and $(n = 3, l = 1)$ is called the $3p$ orbital. In chemistry and atomic physics, each principal quantum number n is also called a **shell**, and each l is also called a **subshell**.

value of l	0	1	2	3	4	5
Letter	s	p	d	f	g	h

TABLE 8.1. First few letters of the spectroscopic notation that correspond to the orbital quantum numbers l . The rest can be found on, e.g., http://en.wikipedia.org/wiki/Spectroscopic_notation.

- **Exercise:** For $n = 1, 2, 3$, list all the possible quantum numbers.

8.7. Probability densities

Let us plot the probability densities to obtain a mental picture of the wavefunctions. For the s orbitals, $l = 0, m_l = 0$, the wavefunctions are radially symmetric, as shown in Fig. 8.9.

For more complicated 3D plots, see <http://www.webelements.com/shop/shopimages/products/extras/POS0007-A2-orbitron-2010-800.jpg>.

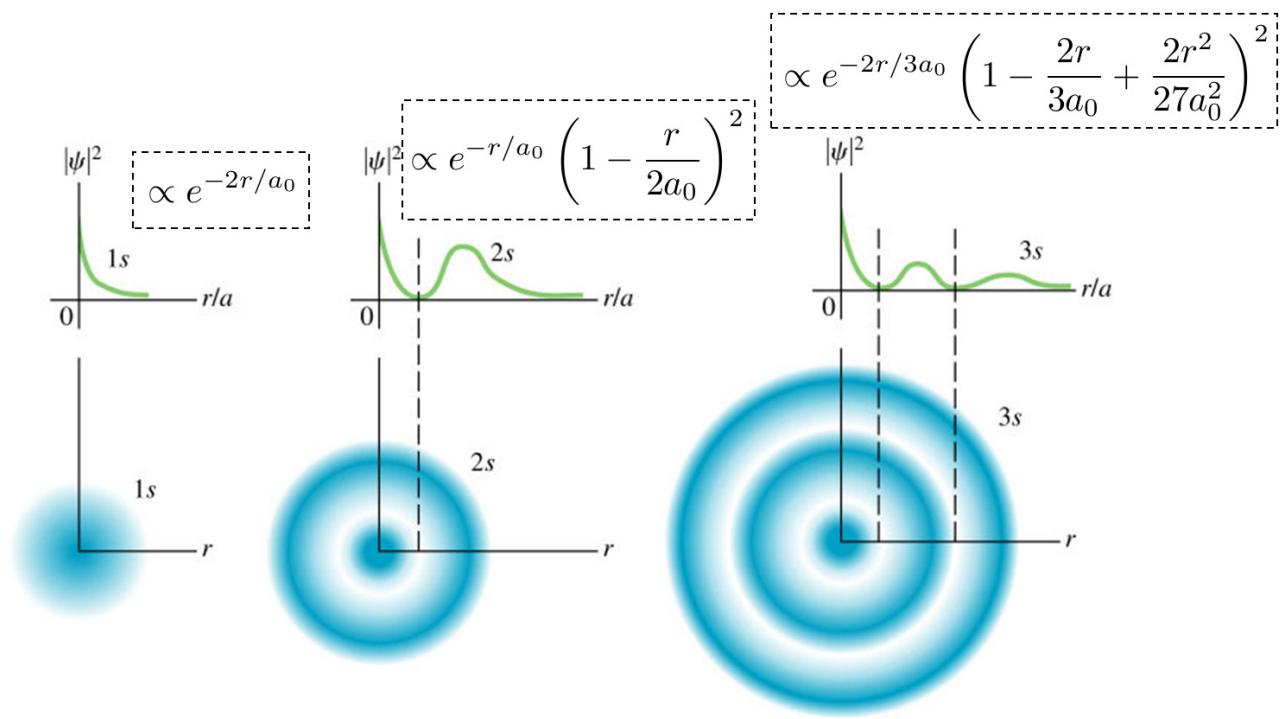


FIGURE 8.9. Cross sections of the probability densities of a few radially symmetric *s* orbitals.
From PC2232 AY13/14

CHAPTER 9

Angular Momentum

9.1. Orbital angular momentum

In classical physics, the angular momentum for an object with position \mathbf{r} and momentum \mathbf{p} with respect to a reference point at the origin is defined as

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \quad (9.1)$$

as depicted in Fig. 9.1. It is a useful quantity because it is conserved when a potential depends only on r , such as the gravitational attraction from the sun to the earth. For example, if both \mathbf{r} and \mathbf{p} are in the x - y plane, as in Fig. 9.1, the angular momentum points in the \hat{z} direction.

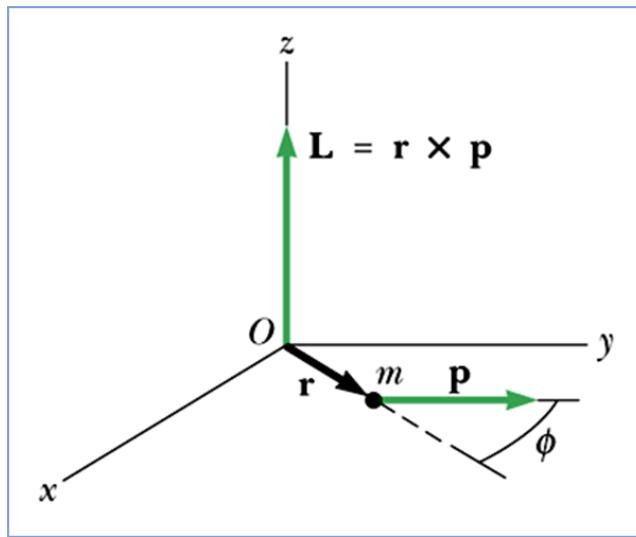


FIGURE 9.1. Angular momentum vector. From PC2232 AY13/14

To explain angular momentum in quantum mechanics, recall that the plane-wave wavefunction $\exp(i\mathbf{k} \cdot \mathbf{r})$ propagating in the direction \mathbf{k} has a definite momentum given by $\hbar\mathbf{k}$. Analogous to a plane wave, $\exp(im_l\phi)$ represents a wave moving in the azimuthal ϕ direction around the z axis. It turns out that $\exp(im_l\phi)$ has a definite angular momentum component along the z direction:

$$L_z = m_l \hbar, \quad m_l = -l, -(l-1), \dots, l-1, l. \quad (9.2)$$

For a given l , for some complicated reason which has to do with operators, it is associated with a definite magnitude of the angular momentum vector given by

$$L = \sqrt{l(l+1)}\hbar. \quad (9.3)$$

We still have this relation among the components of the angular momentum:

$$L_x^2 + L_y^2 + L_z^2 = L^2, \quad (9.4)$$

which means that $L_x^2 + L_y^2$ is

$$L_x^2 + L_y^2 = [l(l+1) - m_l^2] \hbar^2. \quad (9.5)$$

It turns out that L_x and L_y are **incompatible observables** with respect to L_z , such that, if L_z has a definite value, L_x and L_y are random. For a given l , we can plot the angular momentum vector in a sphere like Fig. 9.2,

where $l = 2$. There we have $m_2 = -2, -1, 0, 1, 2$, but L_x and L_y are random, meaning that the angular momentum vector can be anywhere on each circle.

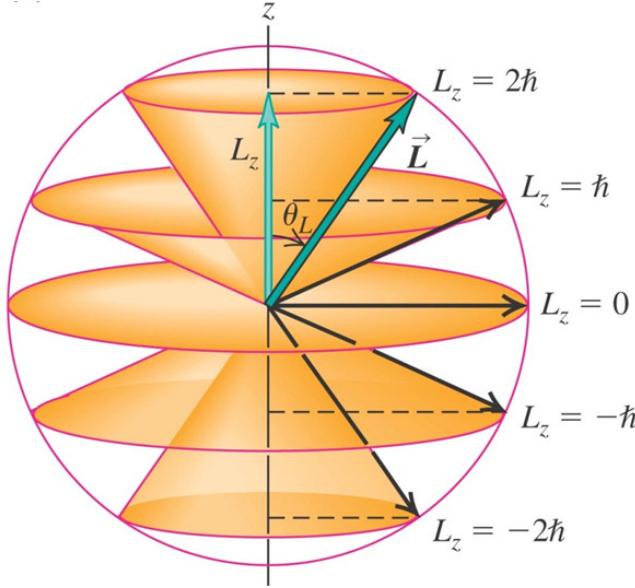


FIGURE 9.2. The angular momentum vector for $l = 2$. There are 5 possible values for L_z . For each l and m_l , $L_x^2 + L_y^2 = L^2 - L_z^2$ is known, such that the vector must be on a circle, but L_x and L_y are incompatible observables with respect to L_z , and their specific values are unknown. From Ref. [2]

9.2. Magnetic moment

In classical electromagnetism, a current loop acts like a magnet with a magnetic moment. For a nonrelativistic electron rotating in a circle, the magnetic moment vector μ is given by

$$\mu = -\frac{e}{2m} \mathbf{L}. \quad (9.6)$$

We are interested in the magnetic moment because a new potential energy appears for a magnet in an external magnetic field \mathbf{B} :

$$U = -\mu \cdot \mathbf{B}, \quad (9.7)$$

such that, to minimize this energy, a magnet will try to align with the direction of the magnetic field such that the dot product is maximized. A magnet also produces its own magnetic field pointing from its north to its south (http://en.wikipedia.org/wiki/Magnetic_field), meaning that if we put two magnets side-by-side, they would tend to be anti-aligned to minimize the energy.

- **Question:** What is direction of the magnetic moment of the earth?

Answer: It is actually pointing towards the south pole, such that a magnetic compass will be anti-aligned with the earth magnet and points to the north.

9.3. Spin angular momentum

It turns out that each electron has its own intrinsic angular momentum called spin, specified by a spin quantum number s_z ¹, as discussed earlier in Sec. 1.4.5. This is regardless of any potential that the electron experiences. To visualize spin, we can think of the electron like the earth, the orbital angular momentum corresponds to the orbit of earth around the sun, and the spin as the self-rotation around its north-pole-south-pole axis, as shown in Fig. 1.11. This analogy should not be taken literally, but is helpful to visualize how it behaves.

¹*There is a very deep reason behind this spin in relativistic quantum field theory [1], but we will just take it as given here.

As it turns out, the spin quantum number can have two possible values only:

$$s_z = -\frac{1}{2}, \frac{1}{2}, \quad (9.8)$$

regardless of the other quantum numbers. For a given set of the other quantum numbers n, l, m_l in the hydrogen atom, the wavefunction of the electron can be written as

$$\psi_{n,l,m_l}(\mathbf{r}, s_z) = \psi_{n,l,m_l}(\mathbf{r})\psi_z(s_z). \quad (9.9)$$

If $\psi_z(s_z = 1/2) = 1$ and $\psi_z(s_z = -1/2) = 0$, the spin has a definite value $s_z = 1/2$; for $\psi_z(s_z = 1/2) = 0$ and $\psi_z(s_z = -1/2) = 1$, the spin has a definite value $s_z = -1/2$.

Given a s_z , the z component of the spin angular momentum is

$$S_z = s_z \hbar. \quad (9.10)$$

The wavefunctions for the other components are determined from $\psi_z(s_z)$ and were discussed in Sec. 1.4.5, but we won't worry about those here. We just need to remember that the other spin components are also quantized and each can have two possible values, it doesn't matter how we define our axes, we will always get just two possible values when we measure a component of the spin vector.

9.4. Spin magnetic moment

The magnetic moment due to spin turns out to be

$$\boldsymbol{\mu} = -g_e \frac{e}{2m} \mathbf{S}, \quad g_e \approx 2.0023. \quad (9.11)$$

g_e is called the gyromagnetic ratio. Its value has been measured experimentally and agrees with relativistic quantum electrodynamics calculations to about 15 significant figures.

A magnet is made of ferromagnetic materials, such as iron, in which the spins of electrons can be induced to align in one direction to produce a large net magnetic field ². The magnetic moment due to orbital angular momentum doesn't really contribute to magnetism, and it is the spin that does the job.

An essential application of magnetic materials is **magnetic hard disks**, which consist of ferromagnetic material with magnetic domains, the direction of magnetic moment in each domain represents bits (http://en.wikipedia.org/wiki/Magnetic_storage).

- **Exercise:** Zeeman effect: In a magnetic field, there is an additional energy $-\boldsymbol{\mu} \cdot \mathbf{B}$. Suppose that $\mathbf{B} = B_0 \hat{z}$. Describe what happens to the energy of a bound state with quantum numbers (n, l, m_l, s_z) . For a given n , how many energy levels are there in the presence of the magnetic field?

9.5. Stern-Gerlach experiment

A silver atom has a net magnetic moment due to the spin of one of its electrons. In a magnetic field \mathbf{B} , a new potential energy appears as in Eq. (9.7). Suppose that the magnetic field is inhomogeneous:

$$\mathbf{B} = B(z) \hat{z}, \quad (9.12)$$

such that the magnetic field is pointing in the \hat{z} direction and also varies along the z direction. Then the force on the silver atom is

$$\mathbf{F} = -\nabla U = \nabla(\boldsymbol{\mu} \cdot \mathbf{B}) = \hat{z} \frac{dB(z)}{dz} (\boldsymbol{\mu} \cdot \hat{z}). \quad (9.13)$$

The force is in the z direction and proportional only to $\boldsymbol{\mu} \cdot \hat{z}$, the z component of the magnetic moment vector $\boldsymbol{\mu}$. Since $\boldsymbol{\mu}$ is proportional to the spin vector \mathbf{S} , the force is proportional to S_z . In the setup shown in Fig. 9.3, a beam of silver atoms pass through such a magnetic field. Each atom experiences a force depending on S_z , and because S_z can have only two values, each atom can experience a force either pointing up or pointing down with a specific magnitude. This leads to the atoms deflected along the z direction at two possible displacements. This experiment was the first demonstration of the quantization of angular momentum.

²*In classical EM, magnets prefer to be anti-aligned. The spins in a ferromagnetic material are aligned due to a very complicated quantum effect called exchange interaction: <http://en.wikipedia.org/wiki/Ferromagnetism>

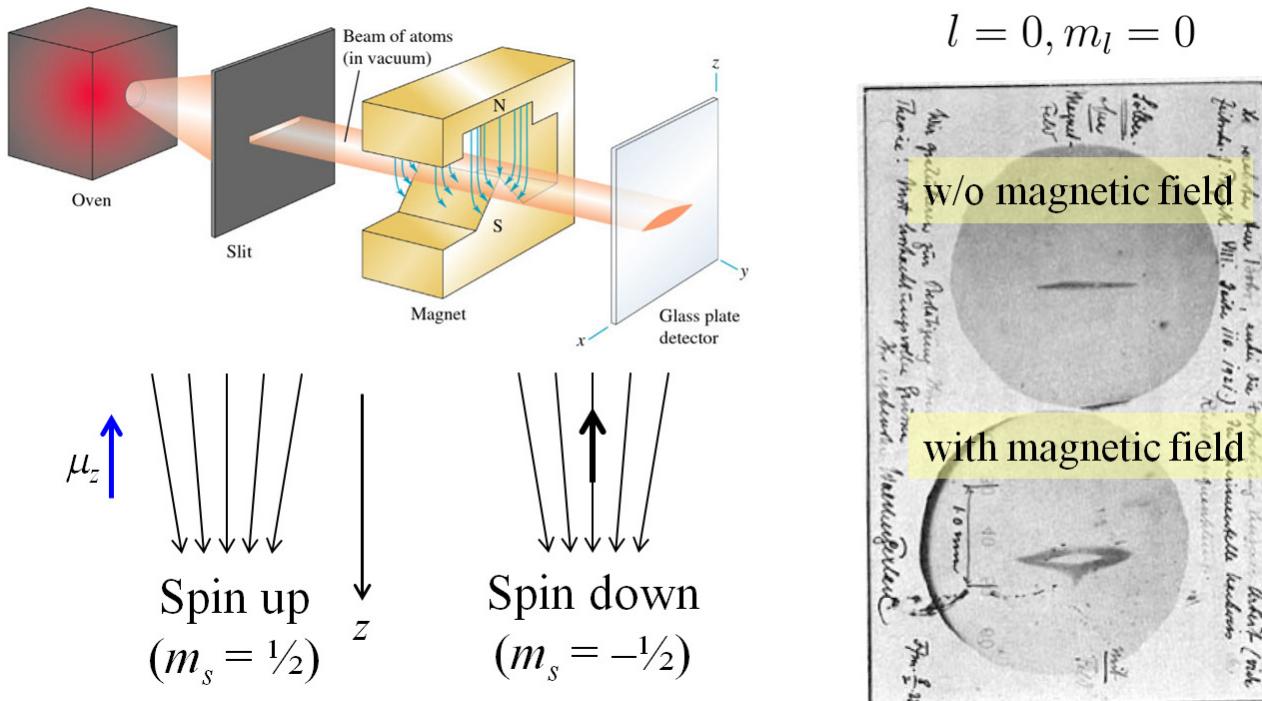


FIGURE 9.3. Stern-Gerlach experiment. An inhomogeneous magnetic field leads to a force acting on each silver atom depending on its magnetic moment. The two peaks in the distribution demonstrate that the magnetic moment is quantized.

9.6. Quantum numbers

For an electron in a hydrogen atom, we have **four quantum numbers**, including spin:

$$n = 1, 2, \dots \quad \text{principal quantum number} \quad (9.14)$$

$$l = 0, 1, 2, \dots, n-1 \quad \text{orbital quantum number} \quad (9.15)$$

$$m_l = 0, \pm 1, \pm 2, \dots, \pm l \quad \text{magnetic quantum number} \quad (9.16)$$

$$s_z = \frac{1}{2}, -\frac{1}{2} \quad \text{spin quantum number} \quad (9.17)$$

- **Exercise:** Given an energy with a principal quantum number n , how many possible states are there?

Answer:

- n values of l
- For each l , $2l + 1$ values of m_l
- 2 spins for each l, m_l

$$\sum_{l=0}^{n-1} 2(2l+1) = 2[1 + 3 + \dots + (2n-1)] = 2n^2. \quad (9.18)$$

There are more and more **degenerate** states with the same energy for increasing n .

CHAPTER 10

Atoms beyond Hydrogen

An atom in general can have Z number of electrons and a nucleus with positive charge Ze , as shown in Fig. 10.1. Z is called the atomic number. The Z electrons can have different positions $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_Z$, each has a different kinetic energy, and the potential energy for each electron depends on its position and is also different. Not only would each electron experience attraction towards the nucleus, it would also be **repeled by the other negatively charged electrons**.

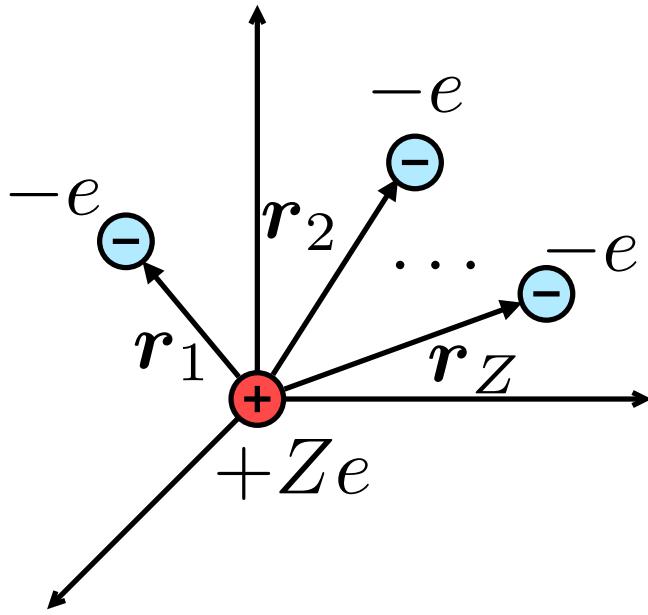


FIGURE 10.1.

The right way of dealing with this in quantum mechanics is to assume a **multi-electron** wavefunction given by $\psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_Z)$. Note that this wavefunction is now very difficult to visualize. If you had trouble picturing three dimensions, the wavefunction here now has $3Z$ dimensions! For a helium atom, $Z = 2$, there are 6 dimensions to deal with, for a lithium atom, $Z = 3$, there are 9 dimensions, etc.

The multi-electron wavefunction obeys the Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} (\nabla_1^2 + \nabla_2^2 + \dots + \nabla_Z^2) \psi + [U(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_Z)] \psi, \quad (10.1)$$

where each Laplacian is with respect to the position of an electron; for example, in Cartesian coordinates, if the position vector for the a th electron has the following Cartesian coordinates:

$$\mathbf{r}_a = x_a \hat{\mathbf{x}} + y_a \hat{\mathbf{y}} + z_a \hat{\mathbf{z}}, \quad (10.2)$$

then the Laplacian for this position is

$$\nabla_a^2 = \frac{\partial^2}{\partial x_a^2} + \frac{\partial^2}{\partial y_a^2} + \frac{\partial^2}{\partial z_a^2}. \quad (10.3)$$

The potential is even more complicated. For each electron at \mathbf{r}_a , it should include the attractive potential from the nucleus at the origin:

$$\left(-\frac{Ze^2}{4\pi\epsilon_0} \frac{1}{r_a} \right), \quad (10.4)$$

and also the repulsive potential from other electrons:

$$\sum_{b \neq a} \frac{e^2}{4\pi\epsilon_0} \frac{1}{|\mathbf{r}_a - \mathbf{r}_b|}. \quad (10.5)$$

See <http://9gag.com/gag/5624331> for the full equation. It is no longer possible to solve the resulting Schrödinger equation without approximations.

10.1. *Fermions

There is yet another restriction on the multi-electron wavefunction coming from the fact that electrons are **fermions**. Assume that all electrons have the same spin (say, $s_{z1} = 1/2, s_{z2} = 1/2, \dots$) for the moment. To understand what fermions mean, consider the multi-electron wavefunction

$$\psi(\mathbf{r}_1, \dots, \underbrace{\mathbf{r}_a}_{\text{position of } a\text{th electron}}, \dots, \underbrace{\mathbf{r}_b}_{\text{position of } b\text{th electron}}, \dots, \mathbf{r}_Z). \quad (10.6)$$

Suppose that we **switch the positions of the a th electron and the b th electron**, such that the a th electron is at position \mathbf{r}_b and the b th electron is at position \mathbf{r}_a .

$$\psi(\mathbf{r}_1, \dots, \underbrace{\mathbf{r}_b}_{\text{position of } a\text{th electron}}, \dots, \underbrace{\mathbf{r}_a}_{\text{position of } b\text{th electron}}, \dots, \mathbf{r}_Z). \quad (10.7)$$

For fermions, there is a restriction on the multi-particle wavefunction such that, after such a switch in position, the wavefunction must be equal to the original one with a **minus sign**:

$$\psi(\mathbf{r}_1, \dots, \mathbf{r}_b, \dots, \mathbf{r}_a, \dots, \mathbf{r}_Z) = -\psi(\mathbf{r}_1, \dots, \mathbf{r}_a, \dots, \mathbf{r}_b, \dots, \mathbf{r}_Z), \quad (\text{for fermions}) \quad (10.8)$$

for any a and b . We say that the wavefunction is **anti-symmetric**.

For **bosons**, such as photons, the wavefunction is **symmetric** instead, and there is no minus sign in Eq. (10.8) for a multi-bosonic wavefunction.

To appreciate what the anti-symmetry means, let's consider just two electrons. The anti-symmetric requirement on the wavefunction becomes

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = -\psi(\mathbf{r}_2, \mathbf{r}_1). \quad (10.9)$$

For example, if we assume that the wavefunction is a product of a single-electron function $\chi(\mathbf{r}_1)$ of \mathbf{r}_1 and the same function $\chi(\mathbf{r}_2)$ of \mathbf{r}_2 , it will not be a legal two-electron wavefunction:

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \chi(\mathbf{r}_1)\chi(\mathbf{r}_2) = \psi(\mathbf{r}_2, \mathbf{r}_1) \neq -\psi(\mathbf{r}_2, \mathbf{r}_1). \quad (10.10)$$

This is because $\chi(\mathbf{r}_1)\chi(\mathbf{r}_2)$ is equal to $\psi(\mathbf{r}_2, \mathbf{r}_1)$ without the minus sign.

One way of getting a legal two-electron wavefunction is to assume the following:

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\sqrt{2}} [\chi_1(\mathbf{r}_1)\chi_2(\mathbf{r}_2) - \chi_1(\mathbf{r}_2)\chi_2(\mathbf{r}_1)], \quad (10.11)$$

for **two different single-electron functions** χ_1 and χ_2 . You can check that it is always equal to $-\psi(\mathbf{r}_2, \mathbf{r}_1)$ as long as χ_1 and χ_2 are different. The factor of $1/\sqrt{2}$ is just a normalization factor. In general, $\psi(\mathbf{r}_1, \mathbf{r}_2)$ doesn't have to be in the form given by Eq. (10.11), but it is a very useful assumption in practice, as functions with smaller number of variables are often easier to deal with.

For multiple electrons, we can get an anti-symmetric wavefunction $\psi(\mathbf{r}_1, \dots, \mathbf{r}_Z)$ from a collection of single-variable functions $\chi_1(\mathbf{r}), \chi_2(\mathbf{r}), \dots, \chi_Z(\mathbf{r})$ using the Slater determinant:

$$\psi(\mathbf{r}_1, \dots, \mathbf{r}_Z) = \frac{1}{\sqrt{Z!}} \begin{vmatrix} \chi_1(\mathbf{r}_1) & \chi_2(\mathbf{r}_1) & \dots & \chi_Z(\mathbf{r}_1) \\ \chi_1(\mathbf{r}_2) & \chi_2(\mathbf{r}_2) & \dots & \chi_Z(\mathbf{r}_2) \\ \vdots & \vdots & \dots & \vdots \\ \chi_1(\mathbf{r}_Z) & \chi_2(\mathbf{r}_Z) & \dots & \chi_Z(\mathbf{r}_Z) \end{vmatrix}. \quad (10.12)$$

You can check that for $Z = 2$ it reduces to Eq. (10.11). **All the functions χ_1, χ_2, \dots are different from one another.**

There is a connection between spin and whether a particle is a fermion or a boson. If the spin quantum number s_z of a particle takes on half-integer values, it must be a fermion, and if s_z is an integer, the particle

must be a boson. The reason behind this can be found in relativistic quantum field theory, which we obviously won't go into detail.

We have not yet taken spin into account in our wavefunction. Including the spin degree of freedom, each electron not only has a position \mathbf{r}_a but also a spin value s_{za} :

$$\psi(\mathbf{r}_1, s_{z1}, \mathbf{r}_2, s_{z2}, \dots, \mathbf{r}_Z, s_{zZ}). \quad (10.13)$$

The single-electron wavefunction that we can use in a Slater determinant can then be written as

$$\chi_1(\mathbf{r}, s_z), \chi_2(\mathbf{r}, s_z), \dots, \chi_Z(\mathbf{r}, s_z). \quad (10.14)$$

Both fermions and bosons are called **indistinguishable particles**. For fermions, the minus sign only affects how you construct the multi-particle wavefunction, but once you are given a wavefunction of fermions, you can switch the positions and spins of any two fermions without affecting the resulting probability density or physical properties such as energy and momentum.

10.2. Pauli exclusion principle

We will now assume that the multi-electron wavefunction consists of each electron occupying a single-electron state (given by one of the $\chi_1(\mathbf{r}, s_z)$, $\chi_2(\mathbf{r}, s_z)$, ... discussed in the previous section). **Because of the anti-symmetric requirement, once an electron is in one of the single-electron states, no other electron can occupy the same state.** This is called the Pauli exclusion principle.

For a multi-electron atom, it is an okay approximation to assume that each single-electron state is essentially the same as the quantum state for a hydrogen atom, except that the energy corresponding to each single-electron state will be different, owing to the increased nucleus charge Ze as well as repulsions from the other electrons. For the spin, we will assume that it is either spin up:

$$\psi_z^\uparrow(s_z) = \begin{cases} 1, & s_z = 1/2, \\ 0, & s_z = -1/2, \end{cases} \quad (10.15)$$

or spin down:

$$\psi_z^\downarrow(s_z) = \begin{cases} 0, & s_z = 1/2, \\ 1, & s_z = -1/2. \end{cases} \quad (10.16)$$

10.3. Periodic table

With all these assumptions and approximations, we are now ready to discuss the **qualitative** features of the quantum states of the electrons in an atom. Let's start with the **hydrogen atom**. At zero temperature (absolute zero in Kelvin), the electron will want to be in the lowest energy possible, so it should be in an $n = 1$ shell. For $n = 1$, there is just one $l = 0$ and one $m_l = 0$, so it's just a $1s$ orbital. For simplicity, we can assume that its spin is given by $s_z = 1/2$, such that its state in terms of the hydrogen orbital wavefunction $\psi_{n,l,m_l}(\mathbf{r})$ would be

$$\chi_1(\mathbf{r}, s_z) = \psi_{1,0,0}(\mathbf{r})\psi_z^\uparrow(s_z). \quad (10.17)$$

Let's continue with the **helium atom** ($Z = 2$). The nucleus now has a $+2e$ charge, and there are two electrons. The potential due to the nucleus is now multiplied by 2, and we also need to worry about the repulsion between the two electrons. In terms of wavefunctions, we can assume that each electron occupies a single-electron state that looks like a hydrogen orbital, but modified suitably to account for increased nucleus potential and electron-electron repulsion. Again, the two electrons want to occupy the states with lowest energies. What happens at the end is that one electron will be in the $1s$ orbital with spin up, and the other electron can't be in the same state but can still be in the $1s$ orbital if it is spin down:

$$\chi_1(\mathbf{r}, s_z) = \psi_{1,0,0}^{(2)}(\mathbf{r})\psi_z^\uparrow(s_z), \quad (10.18)$$

$$\chi_2(\mathbf{r}, s_z) = \psi_{1,0,0}^{(2)}(\mathbf{r})\psi_z^\downarrow(s_z). \quad (10.19)$$

Note that I have used the superscript (2) to denote that $\psi_{1,0,0}^{(2)}(\mathbf{r})$ is a **hydrogen-like** orbital with the same quantum numbers n, l, m_l but with details suitably modified for the $Z = 2$ helium atom.

Let's do another one: the **lithium atom** ($Z = 3$). There are three electrons. At zero temperature, two will occupy the lowest $1s$ orbitals, and the third electron will go into the $n = 2$ orbitals. For $n = 2$, l can be 0 ($2s$

orbitals, $m_l = 0$) or 1 ($2p$ orbitals, $m_l = -1, 0, 1$). Unlike the hydrogen atom, the $2s$ orbital in a lithium atom turns out to have a lower energy than the $2p$ orbital, so the electron will fill a $2s$ state first:

$$\chi_1(\mathbf{r}, s_z) = \psi_{1,0,0}^{(3)}(\mathbf{r})\psi_z^{\uparrow}(s_z), \quad (10.20)$$

$$\chi_2(\mathbf{r}, s_z) = \psi_{1,0,0}^{(3)}(\mathbf{r})\psi_z^{\downarrow}(s_z). \quad (10.21)$$

$$\chi_3(\mathbf{r}, s_z) = \psi_{2,0,0}^{(3)}(\mathbf{r})\psi_z^{\uparrow}(s_z). \quad (10.22)$$

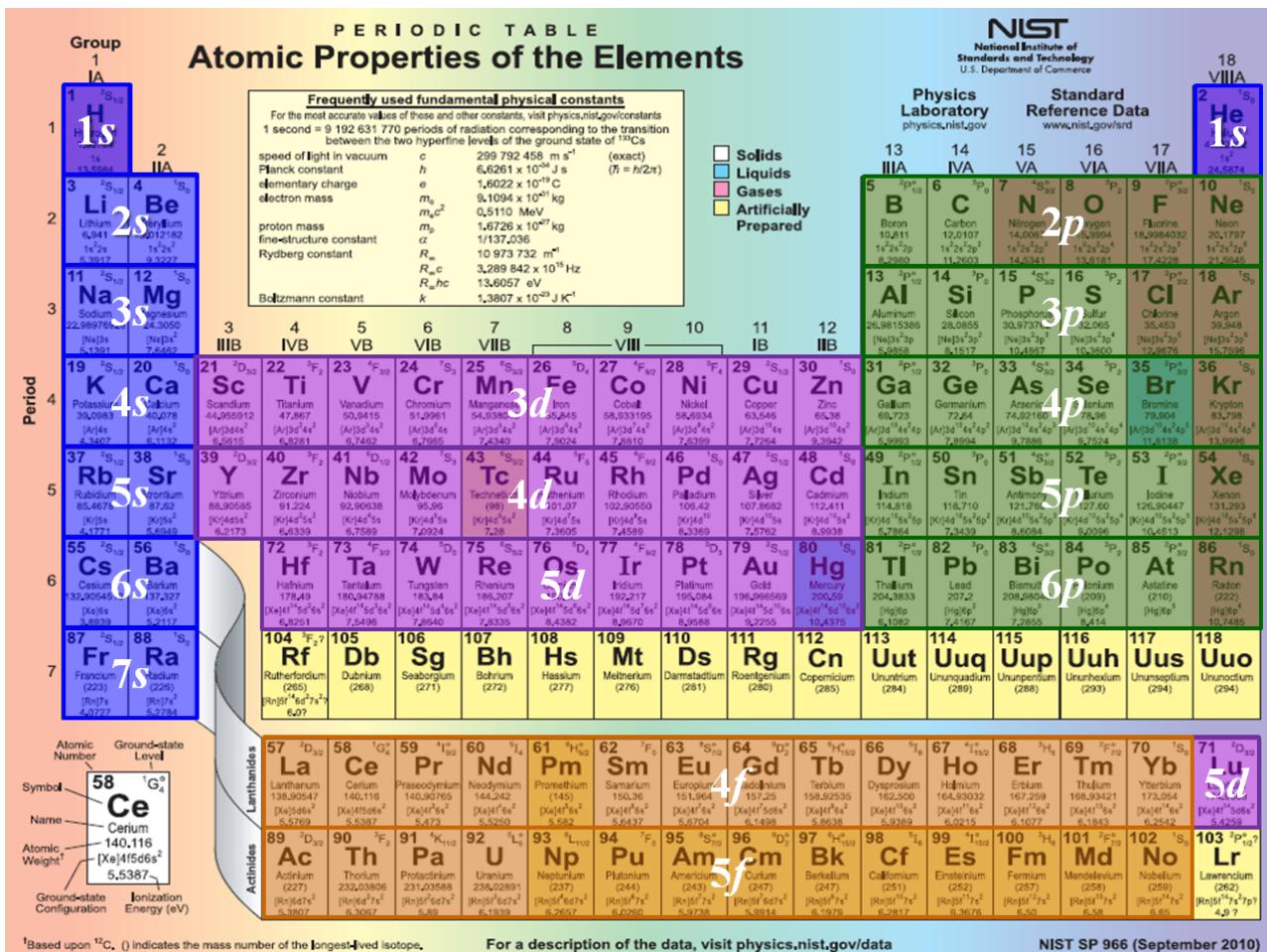
The superscript (3) now denotes that the hydrogen-like orbital is modified for the $Z = 3$ lithium atom.

We can keep going up in Z , and we would end up with the table in Fig. 10.2. Because of electron repulsions, some orbitals end up having quite different energies than the hydrogen-atom calculation would indicate. We can also list the electron configurations in the **periodic table**, as shown in Fig. 10.3.

Table 41.3 Ground-State Electron Configurations

Element	Symbol	Atomic Number (Z)	Electron Configuration
Hydrogen	H	1	$1s$
Helium	He	2	$1s^2$
Lithium	Li	3	$1s^2 2s$
Beryllium	Be	4	$1s^2 2s^2$
Boron	B	5	$1s^2 2s^2 2p$
Carbon	C	6	$1s^2 2s^2 2p^2$
Nitrogen	N	7	$1s^2 2s^2 2p^3$
Oxygen	O	8	$1s^2 2s^2 2p^4$
Fluorine	F	9	$1s^2 2s^2 2p^5$
Neon	Ne	10	$1s^2 2s^2 2p^6$
Sodium	Na	11	$1s^2 2s^2 2p^6 3s$
Magnesium	Mg	12	$1s^2 2s^2 2p^6 3s^2$
Aluminum	Al	13	$1s^2 2s^2 2p^6 3s^2 3p$
Silicon	Si	14	$1s^2 2s^2 2p^6 3s^2 3p^2$
Phosphorus	P	15	$1s^2 2s^2 2p^6 3s^2 3p^3$
Sulfur	S	16	$1s^2 2s^2 2p^6 3s^2 3p^4$
Chlorine	Cl	17	$1s^2 2s^2 2p^6 3s^2 3p^5$
Argon	Ar	18	$1s^2 2s^2 2p^6 3s^2 3p^6$
Potassium	K	19	$1s^2 2s^2 2p^6 3s^2 3p^6 4s$
Calcium	Ca	20	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2$
Scandium	Sc	21	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d$
Titanium	Ti	22	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^2$
Vanadium	V	23	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^3$
Chromium	Cr	24	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^3 3d^5$
Manganese	Mn	25	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^5$
Iron	Fe	26	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^6$
Cobalt	Co	27	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^7$
Nickel	Ni	28	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^8$
Copper	Cu	29	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^3 d^{10}$
Zinc	Zn	30	$1s^2 2s^2 2p^6 3s^2 3p^6 4s^2 3d^{10}$

FIGURE 10.2. Electron configuration for some atoms at zero temperature. At zero temperature, the electrons occupy the states with the lowest energies, subject to Pauli exclusion principle. For example, for silicon, $1s^2 2s^2 2p^6 3s^2 3p^2$ means 2 electrons in the $1s$ orbital, 2 in $2s$, 6 in $2p$, 2 in $3s$, and 2 in $3p$. The hydrogen-like orbitals have different energies than those for a hydrogen atom because of increased nucleus charge and electron repulsion. For example, starting from potassium, the $4s$ orbital turns out to have a lower energy than the $3d$ orbitals, so the $4s$ orbital is filled first before the $3d$ orbitals. From [2].



CHAPTER 11

Solids

If we were interested in chemistry, next we would study chemical bonds, such as ionic bounds and covalent bonds, and crystal structures; see Ref. [2] for an introduction to those topics. Here I will skip all those and go straight into semiconductors, especially silicon, which are of more interest to electrical engineers.

11.1. Silicon atom

Consider first a single silicon atom. It has an atomic number $Z = 14$, the nucleus has charge $+14e$, and the electron configuration is $1s^2 2s^2 2p^6 3s^2 3p^2$, which means there are 2 electrons in the $1s$ orbital, 2 in $2s$, 6 in $2p$, 2 in $3s$, and 2 in $3p$, as depicted in Fig. 11.1. Due to electron repulsions (an effect called screening) the $3p$ orbitals turn out to have higher energy than the $3s$ orbitals, and electrons fill the $3s$ orbitals first before the remaining two electrons partially fill the $3p$ orbitals.

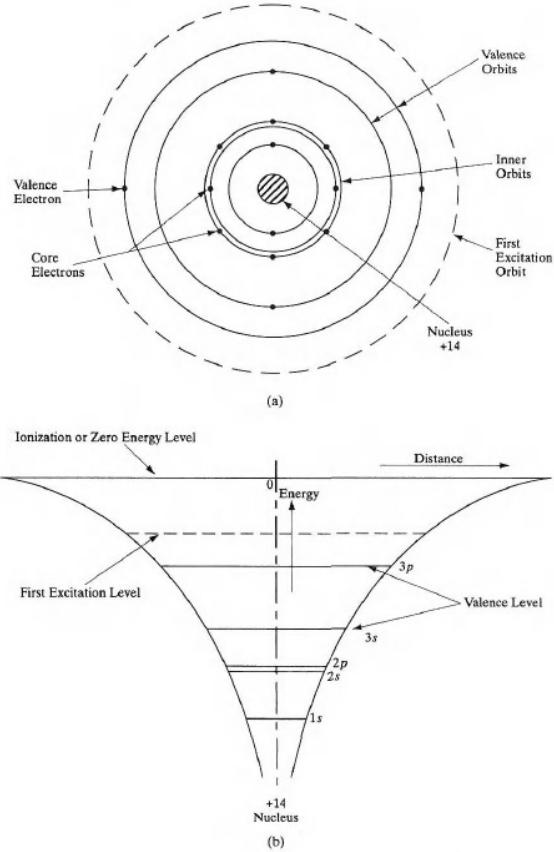


Figure 2-8
Electronic structure and energy levels in a Si atom: (a) The orbital model of a Si atom showing the 10 core electrons ($n = 1$ and 2), and the 4 valence electrons ($n = 3$); (b) energy levels in the coulombic potential of the nucleus are also shown schematically.

FIGURE 11.1. Electrons and their energy levels in a silicon atom. From [6]

11.2. Silicon crystal

If we bring many silicon atoms together, they can form a silicon crystal. It is helpful to recall the discussion in Chap. 6. When we bring N silicon atoms together, each energy level of an atom starts to split in N levels. The splitting is very small for the $n = 1, 2$ states because their wavefunctions don't overlap much, but for the $3s$ and $3p$ orbitals in the $n = 3$ shell the wavefunctions are more diffuse and lead to a larger energy splitting, as shown in Fig. 11.2. When the atoms are far away, there are a total of $2N$ states in the $3s$ orbital (2 spins times N atoms) occupied by $2N$ electrons, and in the $3p$ orbital there are a total of $6N$ states (2 spins times 3 states in $m_l = -1, 0, 1$ times N atoms) partially filled by the other $2N$ electrons.

When the atoms get closer, the resulting wavefunction for each energy level is no longer the individual orbitals of each atom, but a superposition of these orbitals across all the atoms. Picture Fig. 6.6, except that we now have atoms in a 3D crystal and the orbitals look more complicated.

As the atoms get closer and closer, the splitting of the two energy bands becomes so severe that the bands start to overlap. What this means is that some states made from the $3s$ orbitals and some states made from the $3p$ orbitals have identical energies. For each energy, we can now have degenerate quantum states made from a superposition of $3s$ orbitals and $3p$ orbitals. The result is known as orbital hybridization (http://en.wikipedia.org/wiki/Orbital_hybridisation), as shown in Fig. 11.3. We end up having four sp^3 orbitals for each silicon atom, and as the atoms get closer, electrons in these orbitals form **covalent bonds** with the other electrons in neighboring atoms, as shown in Fig. 11.4. Each bond consists of two sp^3 orbitals, one from each atom, and the two electrons in each bond have opposite spins. The bonds form because the electrons turn out to have lower energies when they are in this configuration.

As the atoms get even closer to the actual equilibrium spacing of atoms in a silicon crystal, the wavefunctions made of the sp^3 orbitals form a energy band called the **valence band**. There are $4N$ states in the valence band, as there are 4 sp^3 orbitals for each atom.

All the discussions so far are quite heuristic and belong more to the domain of chemistry. For our purpose, we just need to know that, at zero temperature, the valence band is completely filled with electrons, while above a certain energy **bandgap** ($E_g \approx 1.1$ eV), there is another band called the **conduction band**, which remains unoccupied.

11.3. Insulators, semiconductors, conductors

For a material to conduct current under a voltage, the electrons need to move. For them to move, they need to be accelerated and gain more kinetic energy. Since electrons are waves, we need to study how the waves behave under an external voltage. For example, multiple potential wells in one dimension under a voltage would look like Fig. 11.5.

To be rigorous, we would need to solve the Schrödinger equation again for this new potential. In practice, however, the potential drop is usually very gentle across many atoms, and we can think of the potential wells as pretty much level in each region of space, the band structures we have derived still hold locally, and the voltage drop only makes the overall energy of each band structure in each region to drop with respect to the band structure of the other regions, as shown in Fig. 11.6. When this approximation is valid, we can simply draw a **simplified band diagram**, as shown in the bottom of Fig. 11.6, that plots the band structure in each region with respect to the position inside the material. You will find these diagrams a lot in the study of semiconductor devices.

We can now have a heuristic picture of how electrons move in a material. Consider first a material with valance band all filled with electrons and no electron in the conduction band at zero temperature, such as silicon. For an electron to move in another region, there must be a vacant energy level for it to move into. For the electrons in the valence band, as shown in Fig. 11.7, the energy levels that the electron can move into are all occupied, and it also cannot move into the forbidden bandgap. This means that the electrons simply cannot move, and we have an **insulator**, that is, no current under an applied voltage.¹ Silicon at zero temperature is in fact an excellent insulator.

What makes silicon and other materials such as germanium a **semiconductor** at higher temperatures is that their bandgaps are relatively small. At nonzero temperatures, more energy is available to the electrons, and some electrons will be in the conduction band, leaving behind some vacancies in the valance band, as shown in

¹*Under very high voltage across a small region, it is actually possible for a valence-band electron to tunnel to the conduction band. See <http://spectrum.ieee.org/semiconductors/devices/the-tunneling-transistor>.

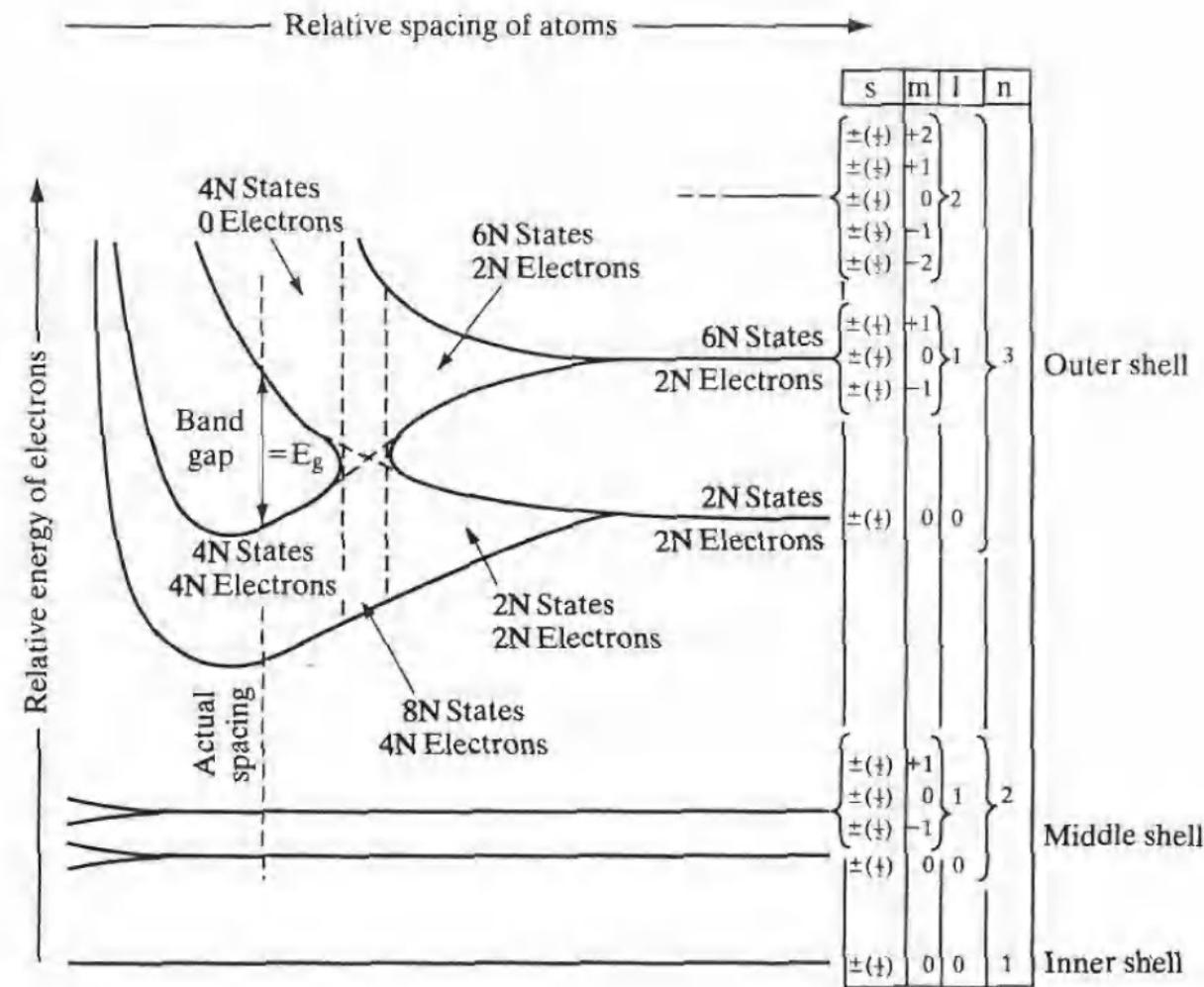


FIGURE 11.2. Energy of the electrons in N silicon atoms, as the atoms are brought together. The focus is the $2N$ electrons in the $3s$ orbitals and the $2N$ electrons in the $3p$ orbitals. As the atoms are brought closer together, each energy level splits into a large number of closely spaced energy levels with wavefunctions consisting of a superposition of the single-atom wavefunctions over all atoms. Level crossing leads to a new valence band that consists of $4N$ sp^3 hybridized orbitals, fully occupied, and an unoccupied conduction band at zero temperature. From [6]

Fig. 11.8. Now there are some vacant levels for valence-band electrons to move into. Also, there are also a lot of vacant energy levels for the conduction-band electrons. This leads to a higher conductivity.²

To model the behavior of electrons in the valence band, it is often more convenient to keep track of the relatively few **vacancies** near the band edge, rather than the vast number of valence-band electrons. We will model each vacancy as a **positively charged particle** called a **hole**, and instead of thinking of a valence-band electron moving into a vacancy on the right, we will think about a **hole moving to the left**, as shown in Fig. 11.9. The conduction-band electrons and the valence-band holes are both called **charge carriers**.

A hole, being positively charged, experiences the negative of the potential experienced by an electron. We don't usually plot the hole potential separately, and it is understood implicitly that a hole will want to go up in an electron potential. In Fig. 11.9, for example, the hole will want to move to the left. For a hole to move to

²*Note that this picture of an electron keeping its energy as it moves is idealistic. In practice, an electron can collide with the nuclei and other defects in the crystal and lose its energy to them as it moves. Due to these collisions, the electron cannot accelerate in a material under a voltage, but rather moves in a constant **drift velocity** on average.

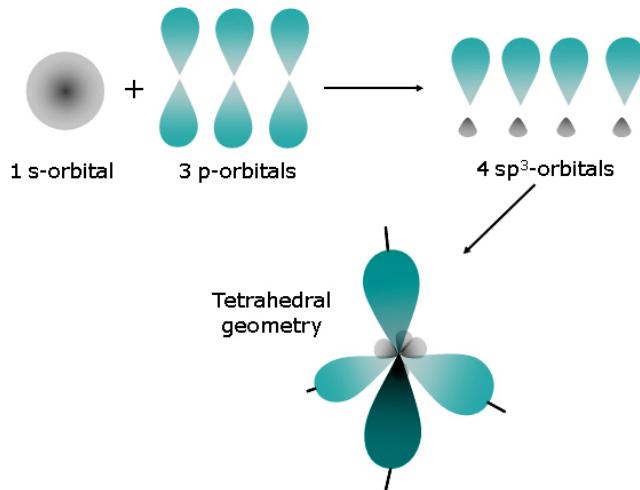


FIGURE 11.3. When the atoms are very close together, some wavefunctions made of the $3s$ orbitals and some wavefunctions made of the $3p$ orbitals can have identical energies. This means that they can be made of a superposition of these orbitals. In silicon, they superimpose to form sp^3 orbitals that result in covalent bonds. From [http://www.ntu.ac.uk/cels/molecular_geometry/hybridization/Sp₃_hybridization/](http://www.ntu.ac.uk/cels/molecular_geometry/hybridization/Sp3_hybridization/)

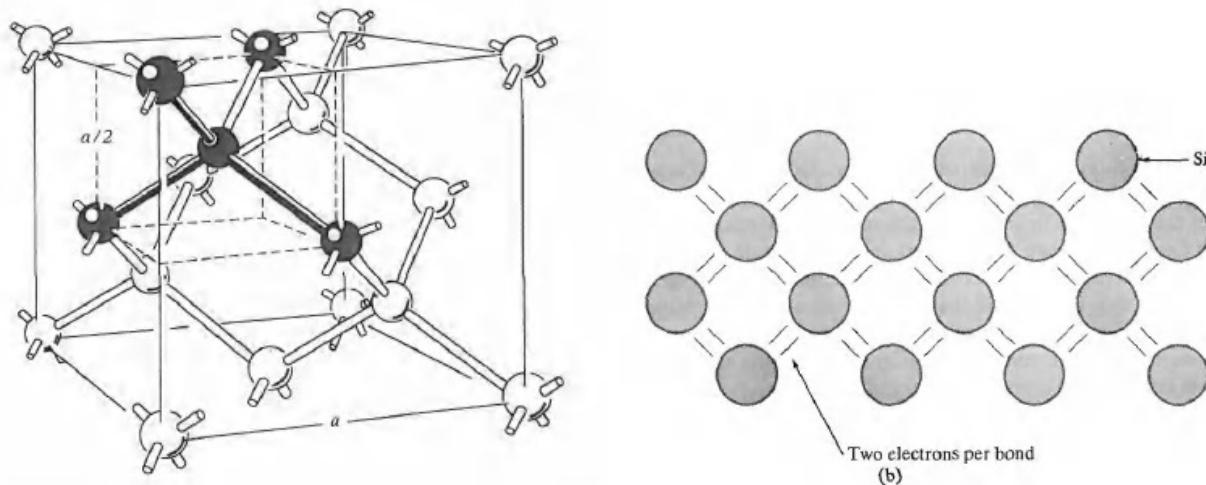


FIGURE 11.4. Left: The diamond cubic crystal structure of silicon. Right: each covalent bond consists of two sp^3 orbitals from two neighboring atoms and two electrons with opposite spins. [6]

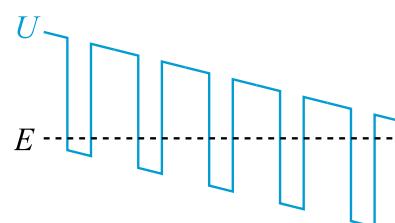


FIGURE 11.5. The potential of multiple 1D potential wells with an applied voltage. If we think of the electrons as classical particles, they would experience a force pushing them to the right. We can think of the kinetic energy as $E - U$ in each well, which increases as the potential drops.

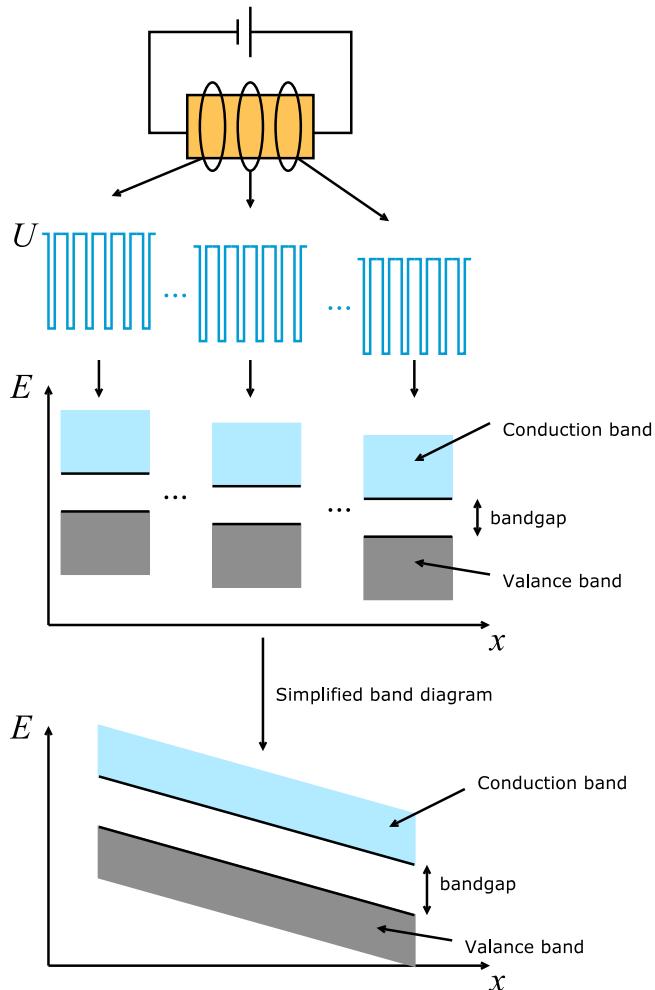


FIGURE 11.6. Top: when a solid is under an applied voltage, we assume that the potential wells in each region inside the solid are still pretty much level and the band structure is the same as before. The voltage leads to a drop in the overall potential in different regions, so the band structures are also shifted downwards as the potential drops. Bottom: When this approximation of gentle potential drop is valid, we can draw a simplified band diagram that plots the energy bands with respect to a real axis x .

another region, there must be an electron there for it to displace, and we think of an electron-occupied energy level as a vacancy for a hole.

Finally, consider materials such as aluminum, copper, etc. with electrons filling the conduction band partially at zero temperature, as shown in Fig. 11.10. There is no vacancy in the valence band, so electrons there cannot move, but there are a lot of vacancies in the conduction band for the electrons there to move. This means that such materials are excellent **conductors** even at zero temperature.

11.4. Doping

In addition to changing the temperature, it is possible to create carriers by introducing impurities to the crystal (**doping**).

First, suppose Group V elements (e.g., N, P, As, <http://en.wikipedia.org/wiki/Pnictogen>) are introduced into a crystal of group IV atoms such as silicon or germanium (http://en.wikipedia.org/wiki/Carbon_group), as shown in Fig. 11.11. Each of the Group V atoms has 5 electrons in its outermost shell, while each Group IV atom has 4 electrons in its outermost shell. Four of the electrons from each group-V atom would form covalent bonds with other group-IV atoms, and there is an additional donor electron that has energy very close to the conduction band edge. At usual temperatures, most of these electrons can be found in the conduction band and can conduct current. This is called an **n-type** semiconductor, where

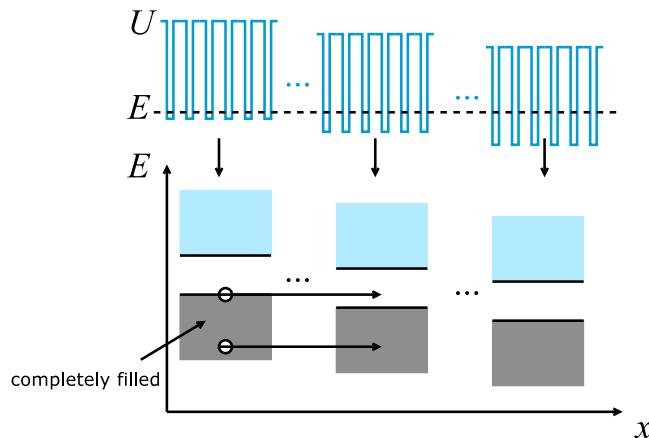


FIGURE 11.7. Top: For an electron with a given energy on the left to move, it must pick up kinetic energy $E - U$, and there must be a vacant energy level on the right for it to move into. Bottom: if the valence band is all filled with electrons, there is simply no available energy levels for the electrons on the left to move into. For example, the electron near the valence band edge cannot move because its energy is inside the bandgap in the next region. Even if there is an energy level on the right, it is occupied and the electron on the left cannot move into it.

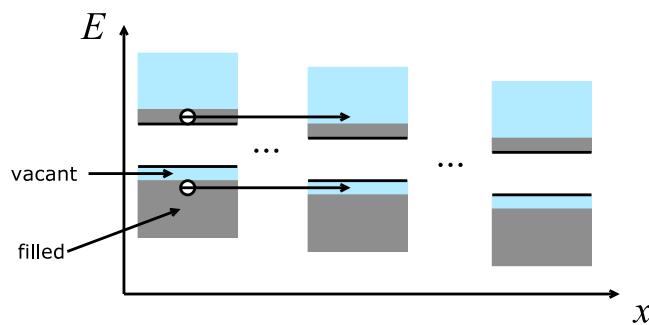


FIGURE 11.8. At higher temperatures, some electrons occupy the conduction band, leaving some vacant levels in the valence band. Now the valence-band electrons can move into those vacancies in the next regions, while the conduction-band electrons can also move into the vacant levels in the conduction band.

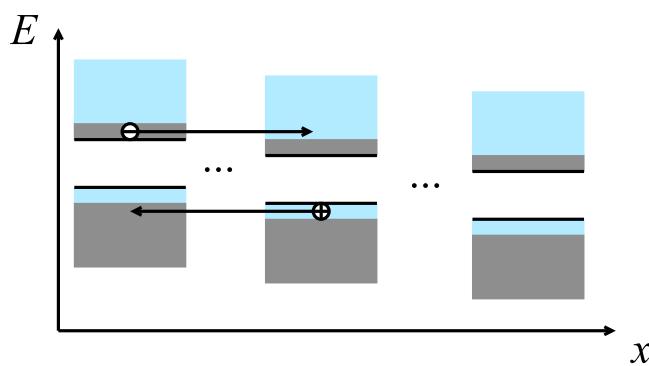


FIGURE 11.9. In the valence band, it is more convenient to think of each vacancy as a positively charged particle called a **hole**. Instead of thinking of valence-band electrons moving to the right here, we will think of a hole moving to the left.

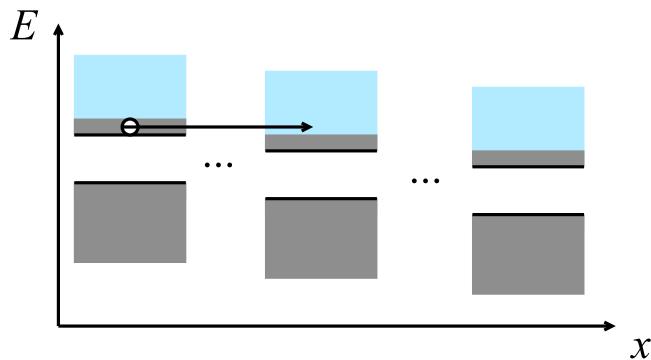


FIGURE 11.10. At zero temperature, if the electrons fill the conduction band partially, they can easily move into the vacancies in the next regions. The material is then a **conductor**.

there are a lot more conduction-band electrons than valence-band holes; the conduction-band electrons are the main charge carriers.

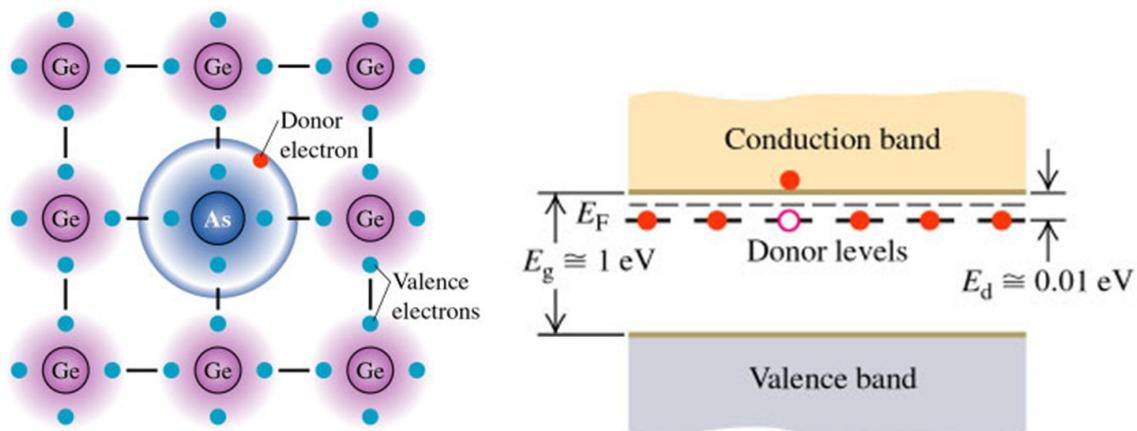


FIGURE 11.11. Left: A Group-V impurity atom arsenic (As) is in a Group-IV germanium (Ge) crystal. Four As electrons in its outermost shell form bounds with other germanium atoms, but there is an extra electron from the As atom that has an energy very close to the conduction band of the crystal. At usual nonzero temperatures most of these extra electrons in the “donor levels” can be found in the conduction band.

If we introduce instead Group III atoms (e.g., boron, aluminum, gallium, http://en.wikipedia.org/wiki/Boron_group) into a crystal of Group IV atoms, Each Group III atom will have a vacancy in the four covalent bonds, allowing valence-band electrons to occupy them and leaving behind a hole, as shown in Fig. 11.12. This is called a **p-type** semiconductor, where holes are the main charge carriers. Both n-type and p-type semiconductors are called **extrinsic semiconductors**, as their conductivities are mainly due to the impurity atoms.

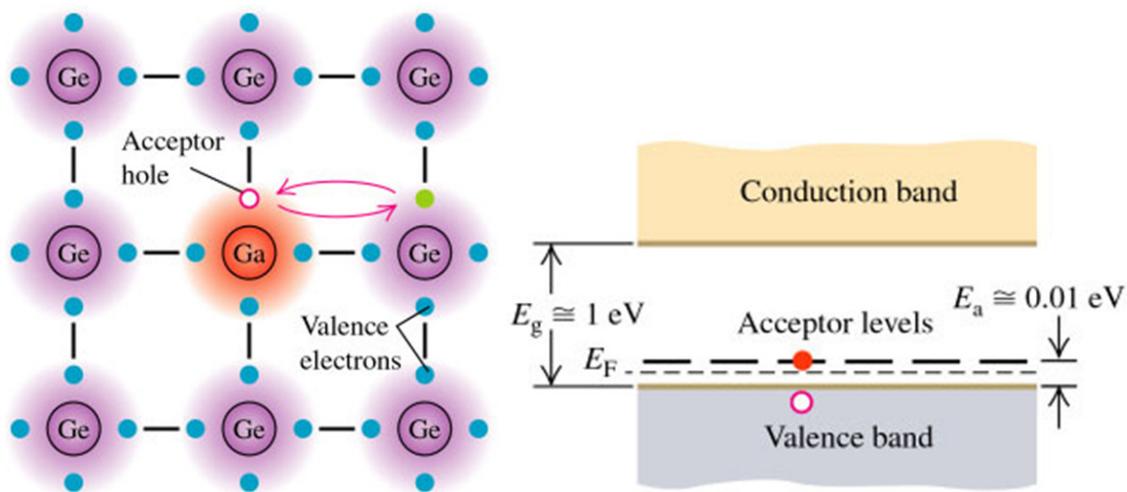


FIGURE 11.12. A Group-III atom gallium (Ga) in a Group-IV germanium crystal. Three electrons from the Ga atom form bonds with germanium atoms, but the covalent bonds in a group-IV crystal require four electrons from each atom, meaning that one of the bonds has a vacancy. Electrons in the valence band can easily move into such “acceptor” levels, leaving behind holes in the valence band.

CHAPTER 12

Temperature and Fermi Level

At a temperature of absolute zero (0 Kelvin), all electrons in a material are in the lowest energy states possible, but this is not a good model of devices at more usual temperatures, such as the room temperature (~ 300 K), when the electrons are hotter and can occupy higher energy levels. This is especially crucial for modeling the electrical properties of semiconductors, where the charge carrier concentrations are quite sensitive to the temperature.

To calculate the conductivity and the charge carrier concentrations, we need to know the number of electrons in the conduction band and the number of holes in the valence band at a certain temperature. This requires **statistical mechanics**, and we will introduce the basic concepts here so that you will understand where the concepts of **temperature** and **Fermi level** come from. The treatment below follows Refs. [5] and [7] closely.

12.1. Number of possibilities

Consider one energy level E_n , and d_n degenerate states having this same energy. Suppose that there are N_n electrons occupying them. How many different **possibilities** are there for the N_n electrons to occupy d_n states?¹ As an example, suppose $d_n = 4$, $N_n = 2$, as shown in Fig. 12.1. Electrons are fundamentally indistinguishable particles, and they are fermions, meaning that at most one electron can fill each state. If you think about it, there are a total of 6 possibilities, as shown in Fig. 12.2.

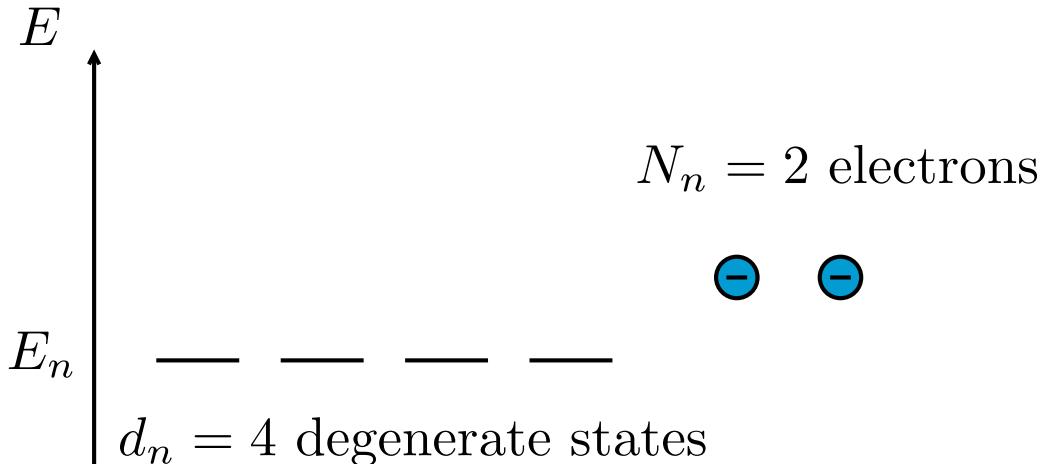


FIGURE 12.1. How many ways are there for $N_n = 2$ electrons to fill $d_n = 4$ states?

In general, the **number of different possibilities** for N_n electrons in d_n states is the **binomial coefficient** (also called “ d_n choose N_n ”, meaning the number of distinct ways choosing N_n states for electrons to occupy out of d_n available states):

$$\binom{d_n}{N_n} \equiv \frac{d_n!}{N_n!(d_n - N_n)!}, \quad (12.1)$$

where

$$d_n! \equiv d_n \times (d_n - 1) \times \dots \times 2 \times 1, \quad N_n! \equiv N_n \times (N_n - 1) \times \dots \times 2 \times 1, \text{ etc.} \quad (12.2)$$

¹*In statistical mechanics jargon, each possibility is also called a **microstate**.

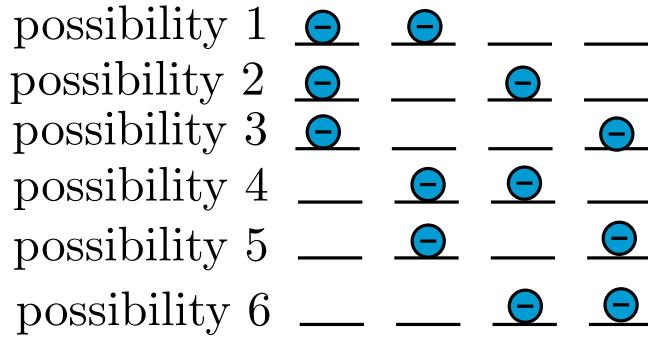
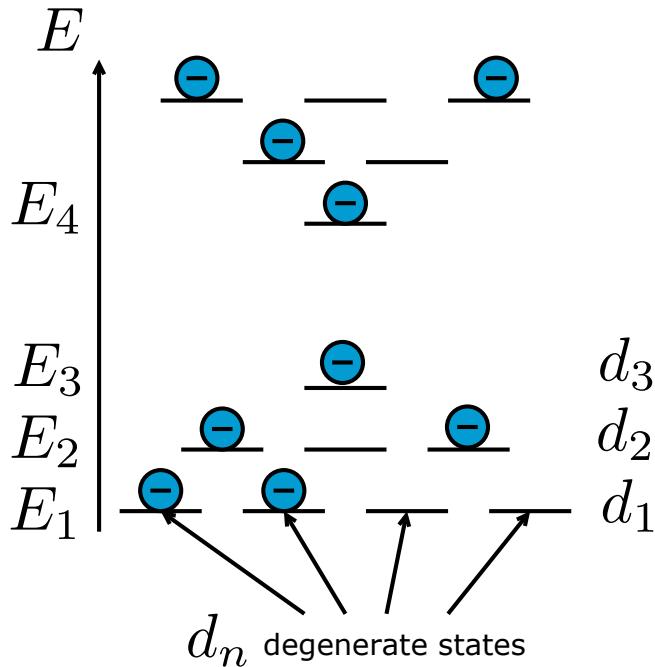


FIGURE 12.2. All 6 possibilities of 2 electrons filling 4 states.

is the factorial. For example, $\binom{4}{2} = 6$. This is a key concept in statistical mechanics: A given set of d_n and N_n corresponds to many different possibilities on a microscopic level. Here we don't really care about the specific possibilities, only the number given N_n and d_n .

12.2. Many energy levels

Now suppose that we have many energy levels given by E_1, E_2, \dots . Suppose that we have d_1 degenerate states for the energy level E_1 , d_2 degenerate states for E_2 , etc. Suppose also that N_1 electrons fill the d_1 degenerate states at energy level E_1 , N_2 electrons fill the d_2 degenerate states at energy level E_2 , etc. One example is shown in Fig. 12.3, where $d_1 = 4$, $d_2 = 3$, $d_3 = 1$, etc., and $N_1 = 2$, $N_2 = 2$, $N_3 = 1$, etc. How many possibilities are there for these N_1, N_2, N_3, \dots electrons to fill the d_1, d_2, d_3, \dots states?

FIGURE 12.3. One of the many possibilities for $N_1 = 2, N_2 = 2, N_3 = 1, N_4 = 1, N_5 = 1, N_6 = 2$ electrons to fill $d_1 = 4, d_2 = 3, d_3 = 1, d_4 = 1, d_5 = 2, d_6 = 3$ states.

There are $\binom{d_1}{N_1}$ different possibilities for energy level E_1 , $\binom{d_2}{N_2}$ possibilities for energy level E_2 , etc. To get the number of possibilities in total, we can just multiply all the numbers of possibilities for each energy level to get the answer:

$$W(N_1, N_2, \dots) = \prod_{n=1}^{\infty} \binom{d_n}{N_n}. \quad (12.3)$$

We call (N_1, N_2, \dots) a **configuration** that specifies the number of electrons at each energy level and W the **multiplicity** of the configuration, that is, the number of possibilities that the configuration possesses. Each configuration can contain many possible arrangements of electrons in the states. For example, Fig. 12.3 shows just one possibility for the $(N_1 = 2, N_2 = 2, N_3 = 1, N_4 = 1, N_5 = 1, N_6 = 2)$ configuration, but note that this configuration can possess other possibilities of arrangements. Some configurations may have higher multiplicities, some may have less. For example, if we have no electron at all ($N_n = 0$), then there is just one possibility for that configuration ($W = 1$). On the other hand, if we have the same number of electrons as the number of available states, ($N_n = d_n$), then there is also just one way for the electrons to fill all the available states and $W = 1$ also. We get more possibilities, and a higher multiplicity, only if the number of electrons at each energy level is a fraction of the available states.

The key assumption in statistical mechanics is to assume that the electrons continuously interact with one another through some physical interactions and exchange energy, such that they are free to jump around the energy levels, and after a while **all the possibilities end up having the same probability of actually happening**.²

If each possibility has the same probability, **the configuration that possess the most possibilities will have the highest probability of actually occurring**. In other words, the configuration that has the highest multiplicity will win. We would like to find this most probable configuration so that we will know the most probable number of electrons at each energy level.

There is one more thing to note: we will assume that the system is closed, and the total number of electrons N and the total energy E are **conserved**. This means that, all the configurations must be subject to the following **constraints**:

- **Number conservation:** The total number of electrons $N = \sum_n N_n$ is conserved.
- **Energy conservation:** The total energy $E = \sum_n N_n E_n$ is conserved.

12.3. Entropy

To find the most probable configuration, we need to find the combination of electron numbers (N_1, N_2, \dots) that maximizes the multiplicity $W(N_1, N_2, \dots)$. Mathematically, rather than maximizing W itself, we will maximize $\ln W(N_1, N_2, \dots)$ instead. It is analogous to how we express power in log scale in decibel units in engineering, except that here we are using the natural logarithm \ln and not $10 \log_{10}$. Since logarithm is a monotonically increasing function, the configuration that maximizes $\ln W(N_1, N_2, \dots)$ is also the one that maximizes $W(N_1, N_2, \dots)$.

We will refer to $\ln W$ using another name: the **entropy** of a configuration, which is defined as

$$S(N_1, N_2, \dots, N_N) \equiv \ln W(N_1, N_2, \dots, N_N). \quad (12.4)$$

This formula is the crowning achievement of Boltzmann; see Fig. 12.4. **Finding the most probable configuration is hence equivalent to finding the configuration that maximizes the entropy.**

12.4. Constrained optimization

We now have the task of finding the configuration (N_1, N_2, \dots) that maximizes the entropy $\ln W(N_1, N_2, \dots, N_N)$, subject to constraints $N = \sum_n N_n$ (electron number conservation) and $E = \sum_n N_n E_n$ (energy conservation). The mathematical tool we use is the **Lagrange multipliers** method (http://en.wikipedia.org/wiki/Lagrange_multiplier), which you should have learned in MA1505. Define the Lagrange function as

$$G(N_1, N_2, \dots, \alpha, \beta) = \ln W + \alpha \left(N - \sum_n N_n \right) + \beta \left(E - \sum_n N_n E_n \right). \quad (12.5)$$

²*In statistical mechanics jargon, this is known as the **ergodic hypothesis**. Note that we have not actually invoked classical or quantum mechanics to prove that it is true, it is more like a hypothesis that ends up being quite successful in explaining most experiments. It can be proved only in special cases.



FIGURE 12.4. Tombstone of Ludwig Boltzmann at the Vienna Central Cemetery, Austria. The formula above is the relation between entropy S , Boltzmann constant $k_B \approx 1.38 \times 10^{-23} \text{ J/K}$ (abbreviated as k there), and multiplicity W of a configuration. The Boltzmann constant in front of $\ln W$ is there for historic reasons.

where α and β are called **Lagrange multipliers**. If we differentiate the Lagrange function with respect to the multipliers and set the partial derivatives to zero, we get back the constraints:

$$\frac{\partial G}{\partial \alpha} = N - \sum_n N_n = 0, \quad N = \sum_n N_n, \quad (12.6)$$

$$\frac{\partial G}{\partial \beta} = E - \sum_n N_n E_n = 0, \quad E = \sum_n N_n E_n. \quad (12.7)$$

We also need to differentiate G with respect to the variables (N_1, N_2, \dots) and set those partial derivatives to zero to find the maximizing configuration. For each N_n , we get

$$\frac{\partial G}{\partial N_n} = \frac{\partial}{\partial N_n} (\ln W) - \alpha \frac{\partial}{\partial N_n} \left(N - \sum_m N_m \right) - \beta \frac{\partial}{\partial N_n} \left(E - \sum_m N_m E_m \right) \quad (12.8)$$

$$= \frac{\partial}{\partial N_n} (\ln W) - \alpha - \beta E_n = 0. \quad (12.9)$$

Note that, in the sum $\sum_m N_m$ and $\sum_m N_m E_m$, we use a new dummy index m so that it is not confused with the index n of the variable under partial derivative $\frac{\partial}{\partial N_n}$. The tricky part is

$$\frac{\partial}{\partial N_n} (\ln W) = \frac{\partial}{\partial N_n} \ln \left[\prod_m \frac{d_m!}{N_m!(d_m - N_m)!} \right], \quad (12.10)$$

which we will tackle in the next section.

12.4.1. *Stirling's approximation. The log function comes in handy here because it turns a product into a sum:

$$\ln \left[\prod_m \frac{d_m!}{N_m!(d_m - N_m)!} \right] = \sum_m \ln \frac{d_m!}{N_m!(d_m - N_m)!} = \sum_m \{ \ln(d_m!) - \ln(N_m!) - \ln[(d_m - N_m)!] \} \quad (12.11)$$

When we differentiate this with respect to N_n , only one term in the summation depends on N_n :

$$\begin{aligned} \frac{\partial}{\partial N_n} \sum_m \{\ln(d_m!) - \ln(N_m!) - \ln[(d_m - N_m)!]\} &= \frac{\partial}{\partial N_n} \{\ln(d_n!) - \ln(N_n!) - \ln[(d_n - N_n)!]\} \\ &= -\frac{\partial}{\partial N_n} \{\ln(N_n!) + \ln[(d_n - N_n)!]\} \end{aligned} \quad (12.12)$$

To simplify this, we invoke Stirling's approximation:

$$\ln(z!) \approx z \ln z - z \text{ for } z \gg 1. \quad (12.13)$$

We use this approximation for both the number of electrons N_n , assuming $N_n \gg 1$:

$$\ln(N_n!) \approx N_n \ln N_n - N_n, \quad (12.14)$$

and the number of vacancies $d_n - N_n$, assuming $d_n - N_n \gg 1$:

$$\ln[(d_n - N_n)!] \approx (d_n - N_n) \ln(d_n - N_n) - (d_n - N_n). \quad (12.15)$$

The final result is thus

$$\frac{\partial}{\partial N_n} \ln W \approx -\ln N_n + \ln(d_n - N_n). \quad (12.16)$$

12.5. Fermi-Dirac distribution

From $\frac{\partial G}{\partial N_n} = 0$ and Stirling's approximation, we finally get

$$\frac{\partial G}{\partial N_n} \approx -\ln N_n + \ln(d_n - N_n) - \alpha - \beta E_n = 0. \quad (12.17)$$

Rearranging the terms, we find that the number of electrons N_n as a fraction of the number of degeneracies d_n is

$$\boxed{\frac{N_n}{d_n} = \frac{1}{\exp(\alpha + \beta E_n) + 1}}. \quad (12.18)$$

This is called the **Fermi-Dirac distribution**. N_n , the number of electrons at an energy level, is a function of d_n , the number of degeneracies, and E_n , the energy of the level, as well as the Lagrange multipliers that we still need to determine.

Recall that α is a Lagrange multiplier arising from **number conservation**, and β is another Lagrange multiplier arising from **energy conservation**. To calculate them, we have to look at the constraints again:

$$\sum_n N_n = \sum_n \frac{d_n}{\exp(\alpha + \beta E_n) + 1} = N, \quad \sum_n N_n E_n = \sum_n \frac{d_n E_n}{\exp(\alpha + \beta E_n) + 1} = E. \quad (12.19)$$

Given all the other parameters (d_n, E_n, E, N), α and β are determined implicitly by these two equations. Simply put, α and β are convenient **fitting parameters** for the distribution given the total number and energy.

You may not have heard of α and β , but they are actually related to the **temperature** T (in Kelvin) and **Fermi level** E_F . In fact, the temperature and the Fermi level are **defined** in terms of these Lagrange multipliers by

$$\beta = \frac{1}{k_B T}, \quad \alpha = -\frac{E_F}{k_B T}. \quad (12.20)$$

In terms of T and E_F , the Fermi-Dirac distribution is

$$\boxed{f(E_n) = \frac{N_n}{d_n} = \frac{1}{\exp[(E_n - E_F)/(k_B T)] + 1}}. \quad (12.21)$$

It is actually not trivial to determine the Fermi level and the temperature for a given system. We need to know the energy levels E_n , the numbers of degeneracies d_n , the total number of electrons N , and the total energy E , and work backwards from Eqs. (12.19) to obtain them.

If this is the first time you've heard about these definitions, it might come as a shock that a physical quantity as basic and intimate as temperature arises from the mathematical concept of Lagrange multipliers. We usually think of temperature as some sort of energy, but this is not strictly correct. It is true that the temperature usually rises as the energy increases, e.g., when we boil water, but when the water is boiling, it stays at 100°C even as

we supply more energy to it. The relation between temperature and energy is quite nontrivial in general and depends on other details of the system.

Similarly, the Fermi level usually rises as more electrons are available, such that an n-type semiconductor (more electrons come from donors) has a higher Fermi level than a p-type semiconductor (electrons are taken by the acceptors), but again the Fermi level is usually calculated by working backwards from other facts of the system in solid-state physics [6].

- **Question:** What is the Fermi-Dirac distribution for holes?

- ***Exercise:** Bose-Einstein distribution: bosons are also indistinguishable particles, but unlike fermions, any number of bosons can occupy the same state. Convince yourself that the number of possibilities for N_n bosons in d_n states is

$$\binom{d_n + N_n - 1}{N_n}. \quad (12.22)$$

Using this and Stirling's approximation, confirm that the Bose-Einstein distribution for bosons at thermodynamic equilibrium is

$$f(E_n) = \frac{N_n}{d_n} \approx \frac{N_n}{d_n - 1} \approx \frac{1}{\exp(\alpha + \beta E_n) - 1}. \quad (12.23)$$

For photons, the photon number is usually not conserved. Show that its distribution at thermal equilibrium is

$$f(E_n) \approx \frac{1}{\exp(\beta E_n) - 1}. \quad (12.24)$$

12.6. Properties of Fermi-Dirac distribution

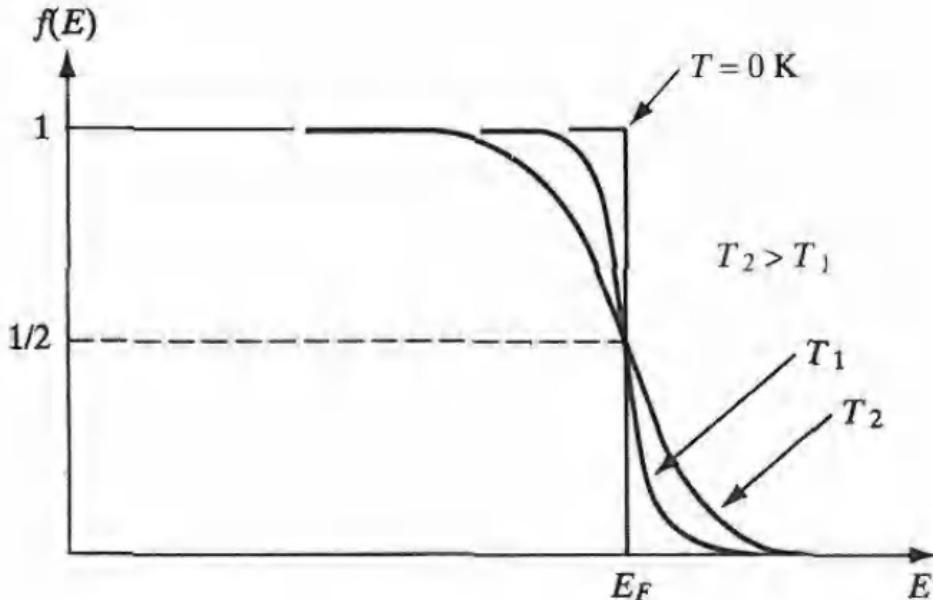


FIGURE 12.5. Fermi-Dirac distribution versus energy E at various temperatures. Note the simple shape for $T = 0$ K, and also the exponential tail for $E \gg E_F$ for higher temperatures.

If we are somehow given E_F and T , the Fermi-Dirac distribution can be plotted against energy, as in Fig. 12.5. At absolute zero temperature ($T = 0$ K), $N_n = d_n$ for all energy levels below E_F , meaning that all the energy levels below the Fermi level are all occupied, and $N_n = 0$ for all energy levels above E_F , meaning that the energy levels above the Fermi level are all vacant. This agrees with our earlier intuition that at zero temperature the electrons will want to occupy the lowest energy levels. If the temperature is somewhat higher, the Fermi-Dirac function "softens," more vacancies can be found below E_F and more occupations can be found above E_F . $f(E_F)$ right at the Fermi level is always 1/2.

For an energy much higher than the Fermi level ($E \gg E_F$), the Fermi-Dirac distribution can be approximated as

$$f(E) \approx \exp\left(-\frac{E - E_F}{k_B T}\right). \quad (12.25)$$

This means that the fraction of occupancies rapidly decays exponentially as the energy is further away from the Fermi level. It is often useful to remember that for $T = 300 \text{ K}$ (room temperature), $k_B T \approx 0.026 \text{ eV}$, so we are always thinking of $E - E_F$ as a ratio with $k_B T \approx 0.026 \text{ eV}$ when studying the Fermi-Dirac distribution.

- **Exercise:** Compute $f(E)$ for $T = 300 \text{ K}$ and $E - E_F = 0.1 \text{ eV}$, 1 eV , 5 eV . Do the same for $T = 1000 \text{ K}$. Note how the distribution rapidly changes for energy away from the Fermi level.

12.7. *Thermal equilibrium

Temperature in everyday life, such as that from the weather report or your thermometer, is a statement about the temperature of an **environment**. We use this number to determine whether it is hot or cold. In other words, we use the temperature to gauge how much or how fast we will lose or gain energy from the environment. To understand this concept of temperature in statistical mechanics, we need to consider not just one closed system but **two systems** which are free to **exchange energy with each other**.

Let the multiplicity of one system be $W(N_1, N_2, \dots)$ and that of another system be $W'(N'_1, N'_2, \dots)$. Multiplying the two gives the multiplicity of the two systems WW' . The entropy becomes $\ln(WW') = \ln W + \ln W'$. The energy levels (E_1, E_2, \dots) and (E'_1, E'_2, \dots) and degeneracies (d_1, d_2, \dots) and (d'_1, d'_2, \dots) of the two systems can be different. In fact, we don't even require the particles to be fermions, they can be bosons or distinguishable particles.

Suppose that the two systems can **exchange energy but not particles**. If we assign the same probability to all possibilities, the situation is known as **thermal equilibrium**. The constraints are then

$$\sum_n N_n = N, \quad \sum_m N'_m = N', \quad \sum_n N_n E_n + \sum_m N'_m E'_m = E. \quad (12.26)$$

The important point here is that the two systems can exchange energy and it is their total energy that is conserved. Again we can follow the Lagrange multipliers method and define

$$G = \ln W + \ln W' + \alpha \left(N - \sum_n N_n \right) + \alpha' \left(N' - \sum_m N'_m \right) + \beta \left[E - \left(\sum_n N_n E_n + \sum_m N'_m E'_m \right) \right]. \quad (12.27)$$

For the first system, if we compute $\frac{\partial G}{\partial N_n} = 0$, we find that $\frac{\partial}{\partial N_n} \ln W - \alpha - \beta E_n = 0$, which is exactly the same equation as Eq. (12.9) for one system, and we would still obtain the Fermi-Dirac distribution if we assume that the first system consists of electrons/fermions.

We could do the same for $\frac{\partial G}{\partial N'_m} = 0$, and we would get the distribution for the other system. The key point is this: the two systems end up sharing **the same Lagrange multiplier** β . With the temperature defined by $\beta = 1/(k_B T)$, **the two systems will have the same temperature at thermal equilibrium**.

For example, consider a pack of ice left in the environment. For simplicity let's assume that there's no evaporation or condensation. The ice and the environment are not at thermal equilibrium initially, but they keep exchanging energy, and given long enough time, they will reach a thermal equilibrium when they are at the same temperature. Since the environment contains much more energy than the ice, the total energy is essentially the same as the energy of the environment, and the final temperature will be that of the environment. For another example, consider a thermometer in a glass of hot water. Let's say that the thermometer initially is at room temperature, and the thermometer temperature will rise once it is in the hot water. The final temperature of the two systems will also be the same given long enough time to reach the thermal equilibrium.

12.8. *Thermodynamic equilibrium

What if the two systems can **exchange energy as well as particles**? We will encounter this a lot in semiconductor physics, such as pn junctions and metal contacts, where there are two different materials in

contact and they are free to exchange electrons and also energy. The constraints become

$$\sum_n N_n + \sum_m N'_m = N, \quad \sum_n N_n E_n + \sum_m N'_m E'_m = E. \quad (12.28)$$

Now the Lagrange function is

$$G = \ln W + \ln W' + \alpha \left(N - \sum_n N_n - \sum_m N'_m \right) + \beta \left[E - \left(\sum_n N_n E_n + \sum_m N'_m E'_m \right) \right]. \quad (12.29)$$

The two systems share **the same Lagrange multipliers** α and β . For electrons, each system will again obey the Fermi-Dirac distribution, and the two systems must have the **same temperature** and the **same Fermi level** at thermodynamic equilibrium.

12.9. Semiconductors and pn junction

The Fermi level depends nontrivially on the number of electrons, the total energy, and the energy levels. For example, in an intrinsic semiconductor, the Fermi level is somewhere inside the bandgap, as shown in Fig. 12.6. At zero temperature, the valence band is below the Fermi level and completely filled, while the conduction band is above the Fermi level and completely vacant. At higher temperatures, however, the valence band can have more vacancies, leading to the presence of thermally excited holes, and there can be more thermally excited electrons in the conduction band.

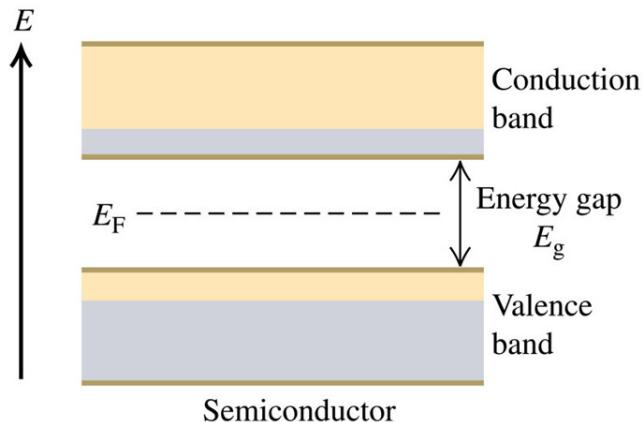


FIGURE 12.6. The Fermi level in an intrinsic semiconductor is somewhere inside the bandgap.

An n-type semiconductor has more electrons and more energy levels from the donors close to the conduction band. The Fermi level ends up being moved very close to the conduction band, and the donor electrons can be thermally excited into the conduction band, as shown in Fig. 12.7.

Similarly, a p-type semiconductor has vacant acceptor levels close to the valence band, leading to a lower Fermi level and electrons can be thermally excited to the acceptor levels, leaving behind holes in the valence band, as shown in Fig. 12.8.

When a p-type semiconductor and an n-type are in contact, they are free to exchange electrons and energy. At thermodynamic equilibrium, they must have the same Lagrange multipliers and thus the same temperature and the same Fermi level. What actually happens is that electrons migrate from the n-side to the p-side to create a potential between the two sides and maintain the same Fermi level across the pn junction, as shown in Fig. 12.9. More details should be covered in **EE2021 Devices and Circuits**.

- **Question:** Is a pn junction in thermodynamic equilibrium under an external applied voltage or current?

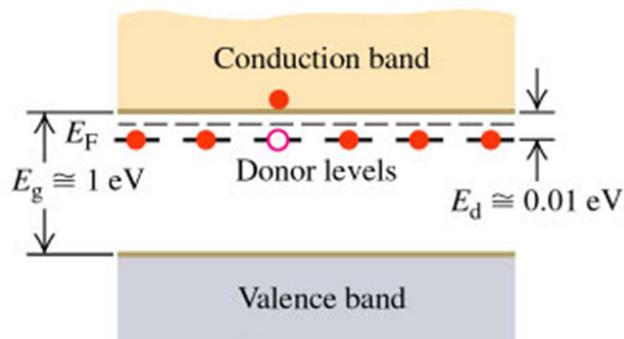


FIGURE 12.7. In an n-type semiconductor, the Fermi level is higher and closer to the conduction band due to the extra electrons and energy levels from the donors. The donor electrons can be thermally excited to the conduction band at nonzero temperatures.

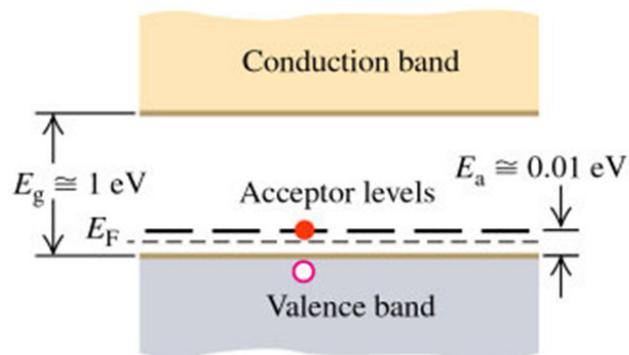


FIGURE 12.8. In a p-type semiconductor, the Fermi level is lower and closer to the valence band due to the vacant acceptor levels. Holes can be thermally excited at nonzero temperatures by electrons moving into the acceptor levels.

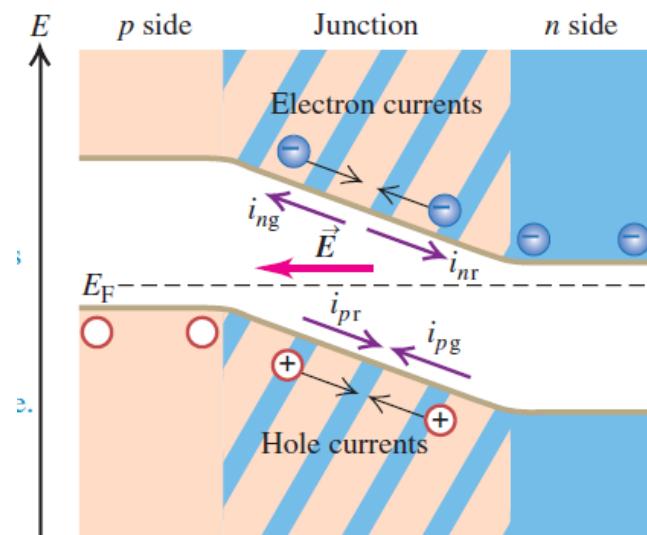


FIGURE 12.9. An n-type semiconductor and a p-type semiconductor can exchange electrons and energy when in contact. At thermodynamic equilibrium, they share the same temperature and Fermi level.

CHAPTER 13

Density of States

The Fermi-Dirac distribution determines the fraction of occupied states N_n/d_n at a given energy E_n , but to calculate the number of electrons N_n at each energy and determine the carrier densities in a semiconductor we still need to know d_n , the number of degenerate states. As we studied in Chap. 6, for a solid, there are a large number of energy levels within each band, and counting and keeping track of the number of states for all the energy levels become difficult. To tackle the problem, we will introduce a concept called density of states, which keeps track of the number of states as a function of energy.

13.1. Free electrons in a box

To study the density of states, we will focus on the model of free electrons in a box, the same problem we studied in Chap. 7. In that chapter, we learned that the three momentum components of the standing wave inside the box are all quantized:

$$k_x = \frac{\pi n_x}{L_x}, \quad k_y = \frac{\pi n_y}{L_y}, \quad k_z = \frac{\pi n_z}{L_z}, \quad (13.1)$$

with positive integers n_x, n_y, n_z ; these momenta can be indicated as dots in the momentum-space diagram shown in Fig. 13.1. Each point denotes two possible states for the two spins. It turns out that this is also a good model of electrons near the bottom of the conduction band for metals or semiconductors.¹ To model these conduction-band electrons, we have to make a couple of modifications to the energy:

$$E = E_c + \frac{\hbar^2}{2m_e} (k_x^2 + k_y^2 + k_z^2), \quad (13.2)$$

where E_c is the conduction band edge and m_e is the effective electron mass. In the momentum space, we can compute the energy for each state by noting the distance from the origin to each state, which is

$$k = \sqrt{k_x^2 + k_y^2 + k_z^2}, \quad (13.3)$$

and the energy is then $E_c + \hbar^2 k^2 / (2m_e)$.

Let us ask the reverse question: Given an energy E , how many degenerate states are there? In the momentum space, any state that has the following distance k from the origin will have energy E :

$$k = \sqrt{\frac{2m_e}{\hbar^2} (E - E_c)}. \quad (13.4)$$

This means that we can draw a spherical surface with radius $\sqrt{\frac{2m_e}{\hbar^2} (E - E_c)}$ and count the number of dots on the surface, then multiply by two for the two spins. This counting is tedious especially for large momenta, and we will make some approximations and use calculus to help.

13.2. Momentum-space calculus

Let us introduce the concept of **momentum-space density of states** $C(k_x, k_y, k_z)$:

$$\int_{\mathcal{V}} C(k_x, k_y, k_z) dk_x dk_y dk_z = \textbf{number of states within the } \mathbf{k}\text{-space volume } \mathcal{V}. \quad (13.5)$$

¹*The conduction-band electrons can be analyzed using a coupled-wave analysis, a 1D version of which was introduced in Sec. 6.3. If you look at the resulting wavefunctions shown in Fig. 6.6, they look somewhat like sinusoidal standing waves and obey similar quantization conditions. The coupled-mode equations are actually mathematically similar to the free-particle Schrödinger equation and will give similar energy levels near the bottom of an energy band, where the quasi-momentum is small.

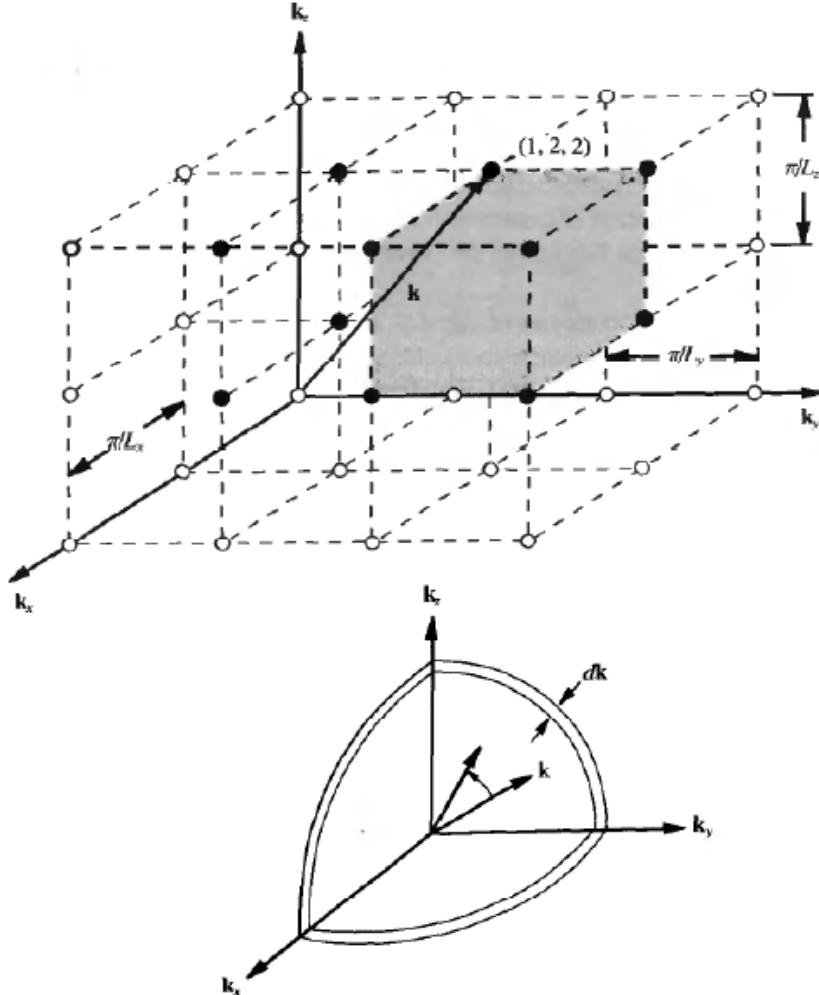


FIGURE 11.3. Mode diagram for a “large” semiconductor. The top sketch shows each individual allowed value of \mathbf{k} whereas the insert implies that the points are very close together so that they could be considered as a continuum.

FIGURE 13.1. From [8]

This is analogous to the concepts of charge density, energy density, or probability density, except that those concepts are in real space and $C(k_x, k_y, k_z)$ is in **momentum space**.

Furthermore, given the **regular lattice structure** of the states in momentum space shown in Fig. 13.1, we will **approximate** $C(k_x, k_y, k_z)$ as a **constant momentum-space density**. This is like approximating the density of discrete atoms in a solid in real space to be constant. For the momentum space, since each dot is separated from adjacent ones by distances $\pi/L_x, \pi/L_y, \pi/L_z$, our approximation will be okay if L_x, L_y, L_z are very large, so that the dots are very close to one another. We call this the **bulk approximation**.

Now that we assume C is constant, we can get its value by imagining boxes in momentum space with volume

$$\frac{\pi}{L_x} \frac{\pi}{L_y} \frac{\pi}{L_z}. \quad (13.6)$$

There is one dot per box, implying two states for the two different spins per box. From Eq. (13.5), we get

$$\frac{\pi}{L_x} \frac{\pi}{L_y} \frac{\pi}{L_z} C = 2. \quad (13.7)$$

In other words, the momentum-space density of states for the free-electron model with the bulk approximation is

$$C(k_x, k_y, k_z) = \frac{2}{\pi^3} L_x L_y L_z. \quad (13.8)$$

13.3. Density of states

In solid-state physics we are more interested in the density of states **with respect to energy**:

$$\int_{E_a}^{E_b} D(E)dE = \text{number of states with energy between } E_a \text{ and } E_b. \quad (13.9)$$

In other words,

$$D(E)dE = \text{the number of states with energy between } E \text{ and } E + dE. \quad (13.10)$$

$D(E)dE$ is essentially the same as the d_n we studied in the previous chapter. It is more convenient to define this function $D(E)$ instead of the discrete d_n because the former allows us to use calculus to convert the momentum-space density $C(k_x, k_y, k_z)$ to $D(E)$.

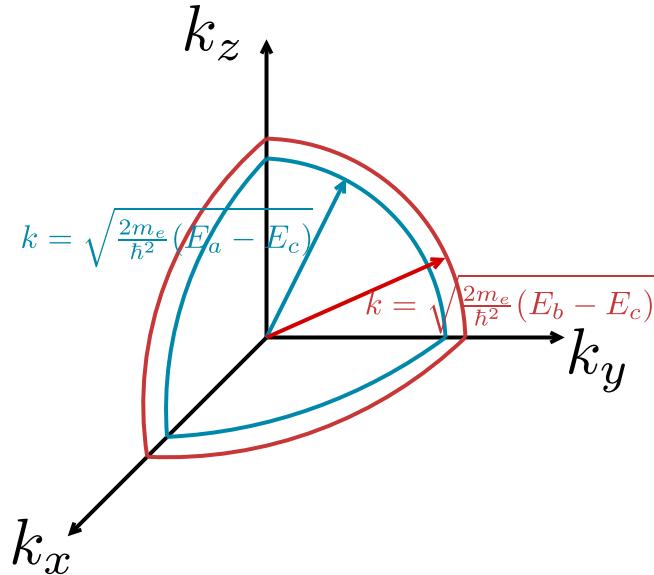


FIGURE 13.2. The number of states with energy between E_a and E_b is the number of states in the shell within two spherical surfaces with radii $k = \sqrt{\frac{2m_e}{\hbar^2}(E_a - E_c)}$ and $k = \sqrt{\frac{2m_e}{\hbar^2}(E_b - E_c)}$. If the momentum-space density $C(k_x, k_y, k_z)$ is constant, the number of states within the shell is C times the volume of the shell. If the difference between E_a and E_b is infinitesimal, the volume can be approximated by the surface area times the thickness dk of the shell.

Recall that a given energy E corresponds to a spherical surface in momentum space with radius $k = \sqrt{2m_e(E - E_c)/\hbar^2}$. Eq. (13.9) then implies that we should count the number of states in a shell between two spherical surfaces with radii $k = \sqrt{2m_e(E_a - E_c)/\hbar^2}$ and $k = \sqrt{2m_e(E_b - E_c)/\hbar^2}$ to get $\int_{E_a}^{E_b} D(E)dE$, as shown in Fig. 13.2. We have already assumed that the momentum-space density $C(k_x, k_y, k_z)$ is constant, meaning that the number of states in the shell will be the constant $C(k_x, k_y, k_z)$ times the volume of the shell. To get the differential $D(E)dE$, which is the number of states within the thin shell between $k = \sqrt{2m_e(E - E_c)/\hbar^2}$ and $k = \sqrt{2m_e(E + dE - E_c)/\hbar^2}$, we can approximate the volume by the surface area times the thickness of the shell:

$$D(E)dE = C \frac{4\pi k^2}{8} dk, \quad (13.11)$$

where C is the constant $C(k_x, k_y, k_z)$, $4\pi k^2/8$ is the surface area of a sphere $4\pi k^2$ divided by 8 for the area of the shell in one octant, and dk is the thickness of the shell given by

$$dk = \sqrt{\frac{2m_e}{\hbar^2}(E + dE - E_c)} - \sqrt{\frac{2m_e}{\hbar^2}(E - E_c)}. \quad (13.12)$$

It shouldn't surprise you that, for infinitesimal dE , we can use calculus to write

$$dk = \frac{dk}{dE} dE = \frac{d}{dE} \left[\sqrt{\frac{2m_e}{\hbar^2}(E - E_c)} \right] dE = \frac{m_e}{\hbar^2} \left[\frac{2m_e}{\hbar^2}(E - E_c) \right]^{-1/2} dE. \quad (13.13)$$

Now we can plug this back in Eq. (13.11), together with the constant C given by Eq. (13.8) and the fact $k^2 = 2m_e(E - E_c)/\hbar^2$, to obtain

$$D(E)dE = \underbrace{\frac{2}{\pi^3} L_x L_y L_z}_{\text{constant } C(k_x, k_y, k_z)} \underbrace{\frac{\pi}{2} \frac{2m_e}{\hbar^2}(E - E_c)}_{\text{surface area of shell } \pi k^2/2} \underbrace{\frac{m_e}{\hbar^2} \left[\frac{2m_e}{\hbar^2}(E - E_c) \right]^{-1/2}}_{\text{thickness of shell } dk} dE, \quad (13.14)$$

and the final result for conduction-band electrons in bulk is

$$D(E) = \frac{L_x L_y L_z}{\pi^2} \frac{m_e}{\hbar^2} \sqrt{\frac{2m_e}{\hbar^2}(E - E_c)}. \quad (13.15)$$

In solid-state physics, it is customary to define the density of states as $D(E)$ divided by the real-space volume:

$$\rho(E) = \frac{D(E)}{L_x L_y L_z}, \quad (13.16)$$

such that

$$\int_{E_a}^{E_b} \rho(E)dE = \text{number of states with energy between } E_a \text{ and } E_b \text{ per real-space volume.} \quad (13.17)$$

$$\rho(E)dE = \text{number of states with energy between } E \text{ and } E + dE \text{ per real-space volume.} \quad (13.18)$$

This quantity $\rho(E)$ is simply called the **density of states**. It has a more complicated physical meaning because it is also normalized by $L_x L_y L_z$, but it is more convenient if we want to compute carrier **densities** in real space.

The final result for conduction-band electrons in bulk is hence

$$\rho(E) = \frac{1}{\pi^2} \frac{m_e}{\hbar^2} \sqrt{\frac{2m_e}{\hbar^2}(E - E_c)}. \quad (13.19)$$

The most important point about this density of states here is its dependence on energy E : it is zero for $E < E_c$, and increases as $\sqrt{E - E_c}$ for $E \geq E_c$, as plotted in Fig. 13.3.

Note that the bulk approximation, which assumes a constant $C(k_x, k_y, k_z)$, is no longer valid for quantum wells, quantum wires, and quantum dots, where the sizes L_x , L_y , or L_z can be very small, the dots are more widely separated, and $\rho(E)$ will look quite different for those systems.

- **Exercise:** A certain quantum system has just one discrete energy level E_1 . What is its density of states $\rho(E)$?

13.4. Equilibrium carrier densities

As an example of how the density of states is used, suppose that we want to compute the conduction-electron density of a bulk semiconductor at thermodynamic equilibrium. This density, which is the number of electrons per unit real-space volume, will be given by the sum of the numbers of electrons with energy above the conduction band divided by the volume $L_x L_y L_z$:

$$n_0 = \frac{1}{L_x L_y L_z} \sum_{n; E \geq E_c} N_n, \quad (13.20)$$

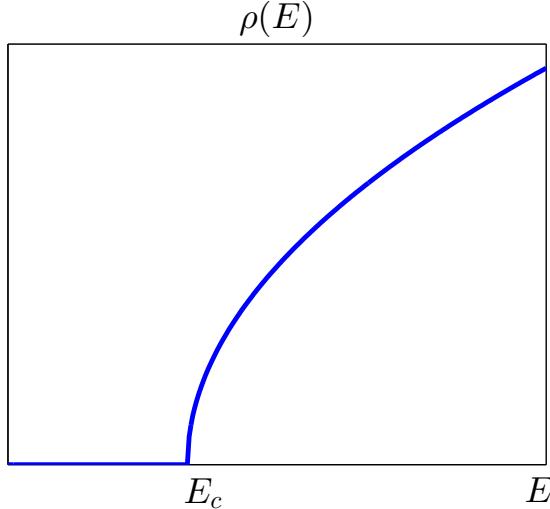


FIGURE 13.3. Density of states as a function of energy E . Below the conduction-band edge E_c , there is no state and the density is zero, and the density increases like $\sqrt{E - E_c}$ for higher energies.

where the statement $E \geq E_c$ denotes we consider only electrons with energy above the conduction-band edge. We can rewrite this in terms of d_n , the number of available states at each energy E_n , as

$$n_0 = \sum_{n; E \geq E_c} \frac{N_n}{d_n} \frac{d_n}{L_x L_y L_z}. \quad (13.21)$$

Now, N_n/d_n at thermodynamic equilibrium is given by the Fermi-Dirac distribution $f(E_n)$. We can also replace the sum by an integral if we replace $d_n/(L_x L_y L_z)$ by the density of states $\rho(E)dE$. The result is

$$n_0 = \int_{E_c}^{\infty} f(E) \rho(E) dE. \quad (13.22)$$

Similar calculations can be done for holes in the valence band. Some examples are shown in Fig. 13.4. These relations are often used backwards to compute the Fermi level E_F given n_0 and p_0 ; more details should be covered in **EE2021 Devices and Circuits**.

Besides carrier density calculations, the density of states and the Fermi-Dirac distribution are also useful for the computation of optical loss or gain in semiconductors for light-emitting diodes and lasers; these are covered in **EE4401 Optoelectronics** and **EE5517 Optical Engineering**.

- **Exercise:** Given the conduction-electron density n_0 , compute the Fermi level as a function of n_0 at zero temperature.
- **Exercise:** Assume $(E_c - E_F)/(k_B T) \gg 1$. Compute n_0 as a function of E_F , E_c , and temperature T . Express E_F as a function of n_0 . Use this to explain why the Fermi level of an n-type semiconductor is closer to the conduction band edge.
- ***Exercise:** Blackbody radiation: compute the density of states for a large box of photons if each photon has energy

$$E = \hbar\omega, \quad \omega = c\sqrt{k_x^2 + k_y^2 + k_z^2}, \quad (13.23)$$

where \mathbf{k} is quantized for standing waves in an optical cavity. Note that each photon has two possible spins, corresponding to two polarizations. Using the Bose-Einstein distribution with zero α (number of photons is not conserved), compute $\rho(E)f(E)$ for photons. Express it in terms of frequency $\omega = E/\hbar$ and compare the distribution with Planck's blackbody spectrum.

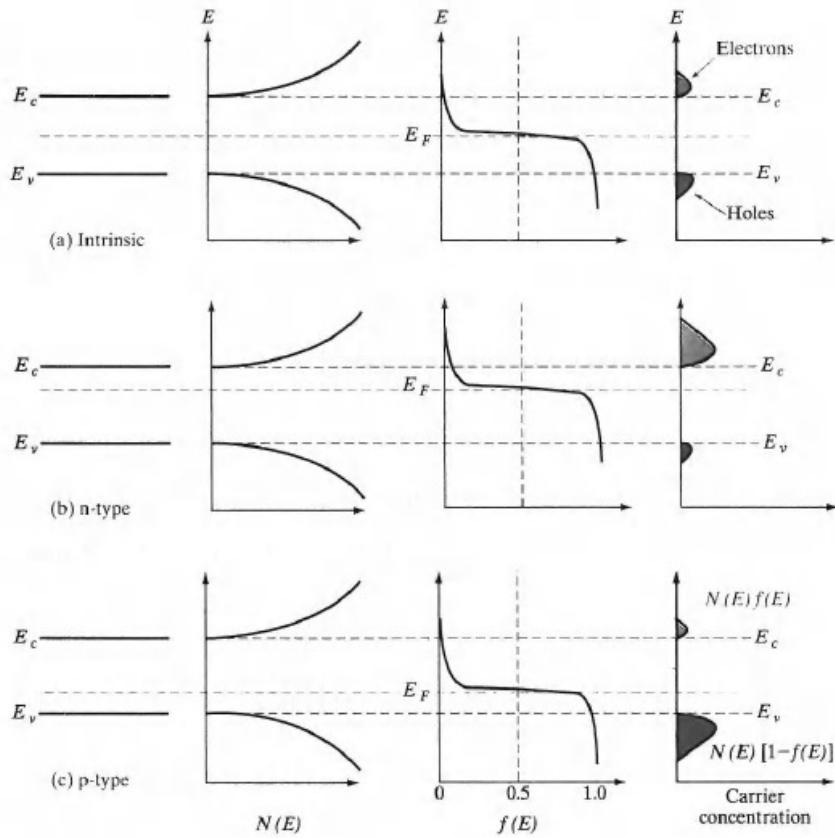


FIGURE 13.4. Multiplying the density of states (denoted by $N(E)$ here) by the Fermi-Dirac distribution gives the carrier density at each energy at thermodynamic equilibrium. Integrating over the energy gives the total carrier densities. The calculations for holes are similar. From [6]

CHAPTER 14

Photons

Consider the electron in the hydrogen atom. If it is at a certain energy level, the Schrödinger equation says that its wavefunction will oscillate forever in time like $\exp(-iEt/\hbar)$ and stays at that energy level. This is however not true in reality, as the electron can interact with electromagnetic fields to gain or lose energy. To describe their interactions we will need a quantum theory that includes both the charged particles and the electromagnetic fields. The full theory is formidable, however, and in this chapter we will rely on a more qualitative picture of **photons** to introduce the concepts.

14.1. Spontaneous emission

14.1.1. Hydrogen atom. To be concrete, we will first think about the electron energy levels in a hydrogen atom. There is just one electron, and the many energy levels are shown in Fig. 14.1. It turns out that, if the electron is initially at a higher energy level, it can **spontaneously** decay to a lower energy level, emitting an electromagnetic wave in the process, as depicted in Fig. 14.1. The process is called **spontaneous emission**. Due to **energy conservation**, the EM wave must carry the same energy as the energy lost by the electron. The quantum theory says that this wave is just like a particle, which we call a **photon**, with a certain wavefunction. The energy E of the EM wave with one photon is related to its frequency ω by

$$E = \hbar\omega. \quad (\text{one photon}) \quad (14.1)$$

For one photon, ω happens to coincide with the “quantum frequency” at which the one-photon wavefunction oscillates.

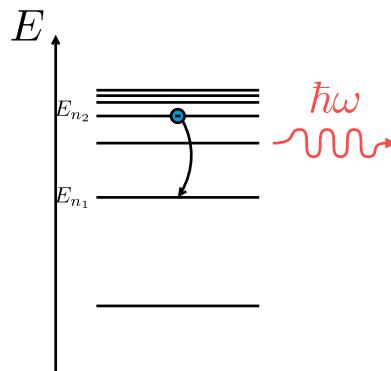


FIGURE 14.1. In a hydrogen atom, the electron can spontaneous decay to a lower energy level, emitting a photon, represented by the wiggly line, with energy $\hbar\omega$ equal to the energy lost by the electron. The EM wave that corresponds to the photon has a frequency ω .

For example, if we somehow excite a bunch of hydrogen atoms to some higher energy levels (by heating them up for example), they will eventually relax to lower energy levels and emit photons with various frequencies. Because the energy levels are discrete, the possible frequencies of these emitted photons are also discrete. Each frequency will depend on the initial energy and the final energy that the atom happens to have:

$$\omega = \frac{E_{n_2} - E_{n_1}}{\hbar}, \quad (14.2)$$

where E_{n_2} is the initial energy level for a hydrogen atom according to Eq. (8.25) and E_{n_1} is the final energy level. This means that we will observe optical emission with frequencies that obey the following relation:

$$\omega = \frac{1}{\hbar(4\pi\epsilon_0)^2} \frac{me^4}{2\hbar^2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right). \quad (14.3)$$

Recall that, in optics, we usually label each frequency by the free-space wavelength $\lambda_0 = 2\pi c/\omega$, so

$$\frac{1}{\lambda_0} = \frac{1}{2\pi c \hbar (4\pi\epsilon_0)^2} \frac{me^4}{2\hbar^2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right). \quad (14.4)$$

This is called the **Rydberg formula** (http://en.wikipedia.org/wiki/Rydberg_formula). Some of the possible transitions are shown in Fig. 14.2. In particular, the Balmer series are the discrete visible colors that correspond to transitions to the $n_1 = 2$ level, as shown in Fig. 14.3.

(b) Energy-level diagram for hydrogen, showing some transitions corresponding to the various series

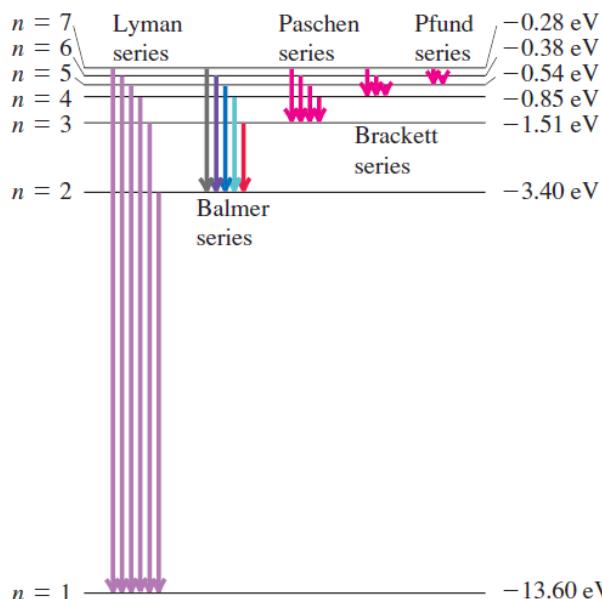


FIGURE 14.2. An electron in a hydrogen atom can spontaneously decay to a lower energy level, emitting a photon with frequency that corresponds to the energy lost. The various transitions shown in the figure, named after their discoverers, lead to the emission of photons with various colors. From [2].

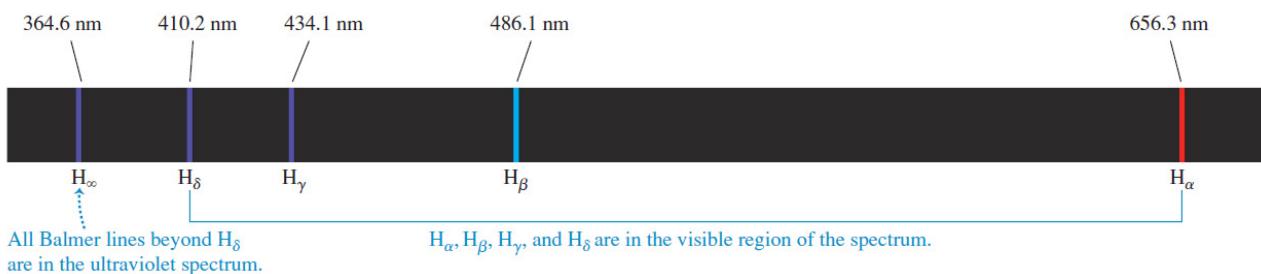


FIGURE 14.3. The Balmer series of spectral lines from hydrogen atoms. From [2].

- **Exercise:** For each spectral line in Fig. 14.3, work out the principal quantum numbers n_2 and n_1 that correspond to the initial and final energy levels of the hydrogen atom.

14.1.2. Media with multiple electrons. Spontaneous emission can happen with any atom or medium, not just hydrogen atom, as long as the medium is in a higher energy level initially and has a lower energy level for it to decay to. For example, consider an atom or a solid with multiple electrons at zero temperature, shown in Fig. 14.4. All the lowest energy levels are occupied, and it becomes impossible for any electron to decay to lower energy levels due to Pauli exclusion principle. Thus, **at zero temperature, there is no spontaneous emission.**

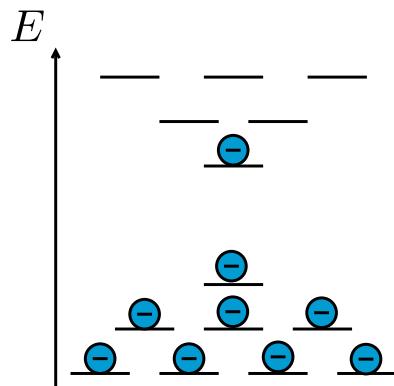


FIGURE 14.4. At zero temperature, the electrons occupy the lowest energy levels possible. It is not possible for any electron to decay to lower levels because they are all occupied and Pauli exclusion principle prevents more than 1 electron occupying the same state.

At higher temperatures, it is possible for some energy levels to be partially occupied. This means that it is possible for an electron at a higher energy level to decay to a lower vacant energy level spontaneously, emitting a photon in the process, as shown in Fig. 14.5. This is why a thermographic (infrared) camera (http://en.wikipedia.org/wiki/Thermographic_camera, <https://www.youtube.com/watch?v=pXAzZoWLzSc>) and even some snakes (http://en.wikipedia.org/wiki/Infrared_sensing_in_snakes) can measure the temperature by detecting spontaneously emitted infrared photons. A hot object can emit a variety of frequencies depending on its energy levels and the temperature, and infrared emission (wavelength around $10 \mu\text{m}$) happens to be the strongest at usual temperatures. The probability of spontaneous emission is higher if the temperature is higher, simply because there are more hot electrons and more vacant energy levels for them to relax to.

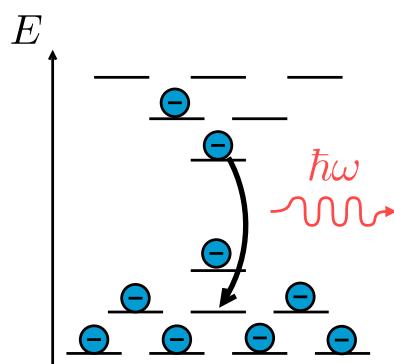


FIGURE 14.5. At nonzero temperatures, some electrons have higher energies, leaving some lower energy levels vacant. Spontaneous emission becomes possible if a higher-energy electron moves into the lower-energy vacancy. The higher the temperature, the more high-energy electrons and low-energy vacancies, leading to more spontaneous emission.

14.1.3. Semiconductors. For semiconductors, we think of vacant energy levels as **holes**. For spontaneous emission to happen, we need high-energy electrons and also lower-energy vacant energy levels. In the language of electrons and holes, this means that spontaneous emission occurs when an electron and a hole **recombine**, as shown in Fig. 14.6. **Light-emitting diodes (LEDs)** are essentially semiconductor devices that

inject conduction-band electrons and valence-band holes in the same region, so that they will recombine and emit photons. The electrons are usually close to the bottom of the conduction band, while the holes are close to the top of the valence band, meaning that the energy difference is essentially the bandgap E_g and the emitted optical frequency is essentially E_g/\hbar .

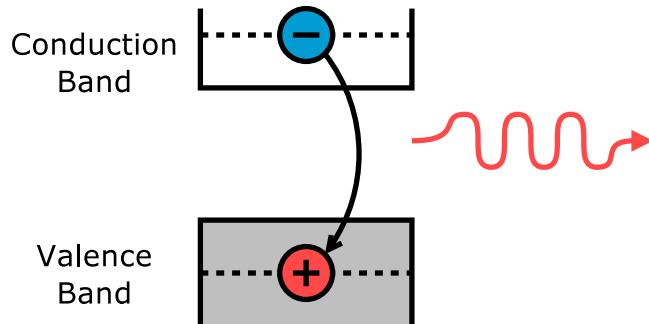


FIGURE 14.6. In a semiconductor, spontaneous emission occurs when an electron recombines with a hole. It takes an electron-hole pair for spontaneous emission to occur.

- **Question:** Is a light-emitting diode in thermodynamic equilibrium with its environment when it is turned on?
- **Exercise:** Find the bandgaps of the materials that make typical light-emitting diodes emitting in red, green, or blue and confirm that they correspond to the wavelengths emitted by such diodes.

14.1.4. Lasers. We mentioned in Part I that, according to the classical electromagnetic analysis, the electromagnetic energy of an optical cavity with gain can increase exponentially in time, but won't do so if the initial energy is zero. Spontaneous emission, which we didn't include in the classical analysis, is the mechanism that seeds the exponential energy increase: even if there is no photon initially, the optical medium can spontaneously emit some electromagnetic energy into the cavity to start the lasing process.

14.1.5. Spectroscopy. By measuring the optical frequencies emitted by a sample, we can infer its energy levels. This is called spectroscopy, which is a useful technique for identifying and detecting chemicals, e.g., in chemistry, biology, remote sensing, air pollutant monitoring, food safety, bomb detection, etc.

14.2. Absorption

If there is a photon with the right energy initially, it can also excite an electron to jump up to a higher energy level, provided that the higher energy level is vacant. This is called **absorption**.

A complication arises when there are multiple photons. Consider a classical electromagnetic plane wave for example. The quantum theory says that each electromagnetic plane wave acts like an energy level for photons. Multiple photons can be in the same EM wave, since they are bosons, and if this EM wave has a frequency ω , the total energy of n photons in the same EM wave is

$$E = n\hbar\omega. \quad (\text{n photons}) \quad (14.5)$$

This implies that the energy of an EM wave can exist only in **discrete levels (quanta)**. This quantization has nothing to do with the quantization of the wavevector for optical standing waves in an optical cavity, it is simply another condition that we have to accept here.¹

The multi-photon wavefunction would oscillate at a “quantum frequency” $E/\hbar = n\omega$, but we seldom care about this quantum frequency. **The frequency of the EM wave ω is the one that corresponds to the frequency we see in everyday life and optics experiments.**

Now imagine that we have $n + 1$ photons initially. Absorption only takes **one of these photons**, leaving n photons behind, while the energy level of the medium rises by an amount of $\hbar\omega$. This is depicted in Fig. 14.7.

For very large n , the optical energy can be approximated as continuous, and the continual absorption of photons in a medium can be modeled as an exponential decrease in electromagnetic energy. This would correspond to a classical model of lossy medium.

¹*You would have to learn the quantum theory of electromagnetic fields to find out why the EM energy is quantized; a good reference is [9].

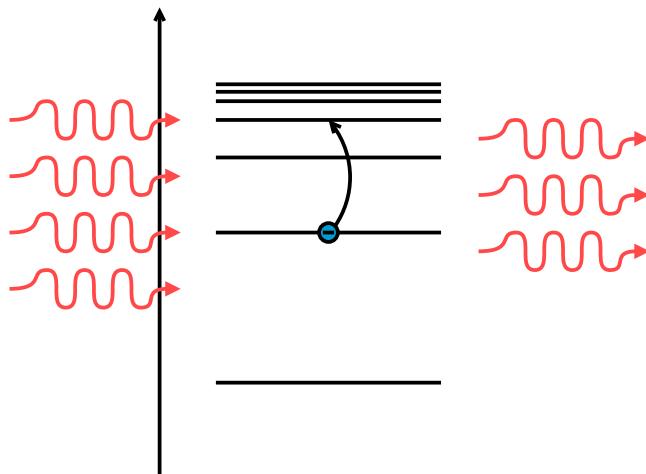


FIGURE 14.7. A hydrogen atom can absorb a photon if the EM frequency ω matches the energy difference $E = \hbar\omega$ for an upward transition.

14.2.1. Photodetectors. For absorption to occur, it requires an electron to be in a lower energy level initially and there is a vacant higher energy level. In the context of semiconductors, absorption creates a conduction-band electron and also a valence-band hole, thus generating an **electron-hole pair**, as shown in Fig. 14.8. This is the basis of **photodiodes** (<http://en.wikipedia.org/wiki/Photodiode>), which absorb photons to create charge carriers and produce “photocurrents.” Photodiodes are used in **image sensors** such as charge-coupled devices (CCD) and CMOS sensors (http://en.wikipedia.org/wiki/Active_pixel_sensor) in digital cameras.

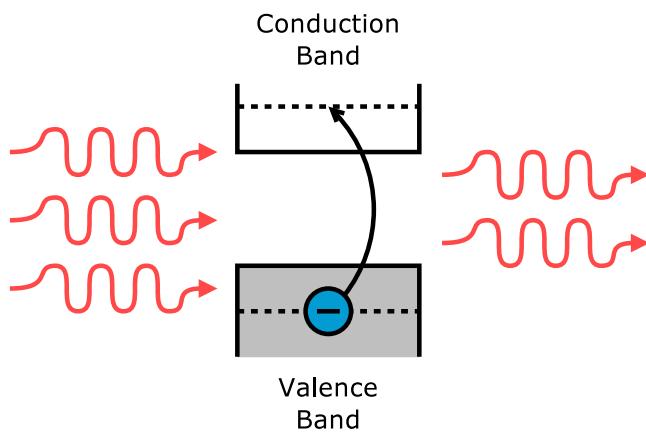


FIGURE 14.8. An electron-hole pair can be created by the absorption of one photon. The energy gained by the electron must be equal to the energy of the absorbed photon $E = \hbar\omega$.

14.2.2. Absorption spectroscopy. Absorption provides another way of doing spectroscopy: http://en.wikipedia.org/wiki/Absorption_spectroscopy, see also Fig. 14.9.

14.2.3. Photoelectric effect. If we shine light on a metal, some electrons can be knocked off the metal and become free (called photoelectrons), as depicted in Fig. 14.10. It is found that, if the frequency of the light ω is below a certain threshold, no photoelectron is produced no matter what the optical intensity is,, and photoelectrons appear only if the frequency ω is above a certain value. The maximum kinetic energy of each photoelectron is found to be

$$KE_{\max} = \hbar\omega - \varphi, \quad (14.6)$$

where φ is an energy called the work function of the metal. Increasing the optical intensity will increase the number of photoelectrons, but not the kinetic energy of each electron.

39.18 When a beam of white light with a continuous spectrum passes through a cool gas, the transmitted light has an absorption spectrum. The absorbed light energy excites the gas and causes it to emit light of its own, which has an emission spectrum.

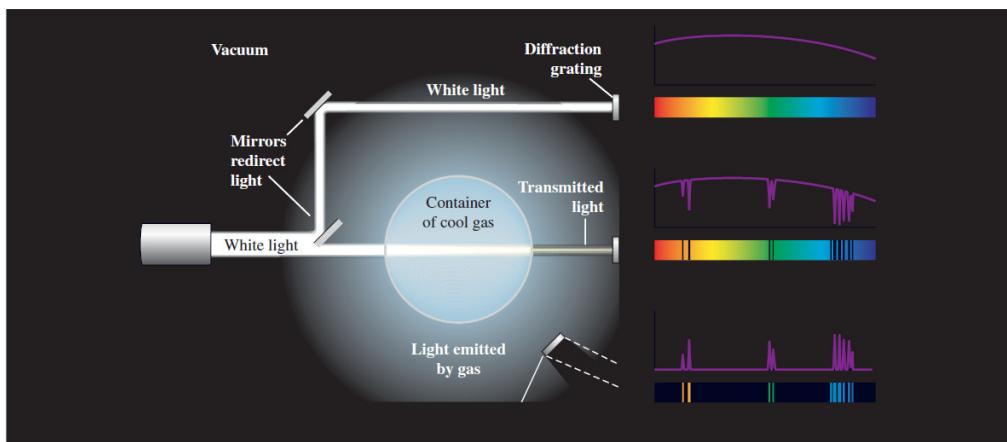


FIGURE 14.9. Spectroscopy can be done by studying the absorption or the spontaneous emission of a sample. From [2].

38.1 The photoelectric effect.

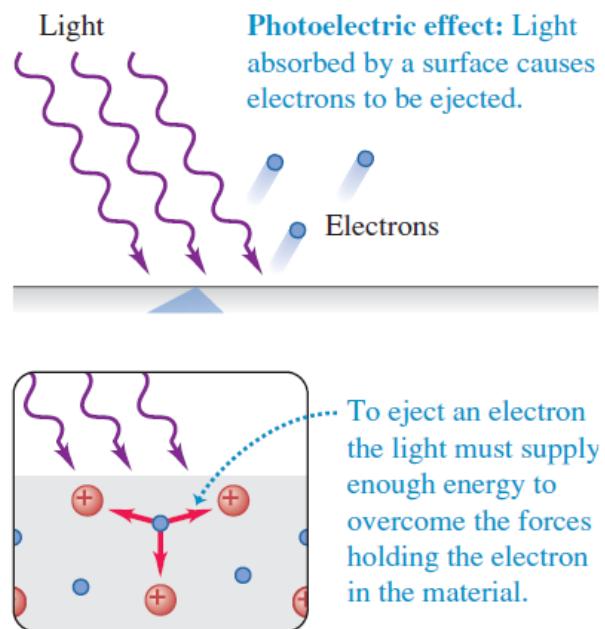


FIGURE 14.10. Photoelectric effect. From [2].

The quantum theory has a simple explanation: in a metal, there are electrons that fill up to the Fermi level at zero temperature, as shown in Fig. 14.11. An electron can jump from a bound state in the metal to a **scattering state** only if it absorbs a photon with energy $\hbar\omega$ that can overcome the energy difference between the bound-state energy and the potential energy outside the metal. An electron in a scattering state has a positive kinetic energy relative to the potential energy outside the metal and can move into free space. The minimum $\hbar\omega$ required to excite a photoelectron is then equal to the difference between the Fermi level and the outside potential energy. This difference is the work function φ . As the frequency ω is increased, it is possible for electrons below the Fermi level to be ejected also, but the maximum kinetic energy is attained only for electrons initially at the Fermi level.

Of all the brilliant discoveries by Einstein, his explanation of the photoelectric effect, using an early immature version of the quantum theory, was what got him the Nobel prize.

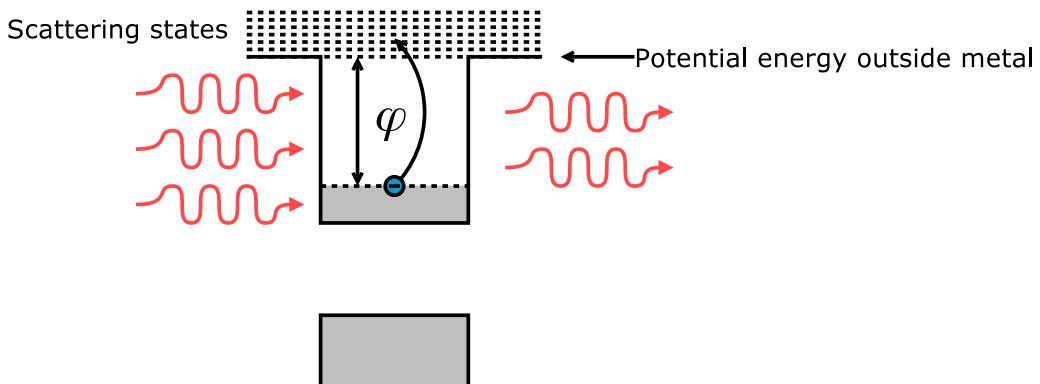


FIGURE 14.11. Inside a metal, electrons are filled up to the Fermi level in bound states. For an electron to escape the metal, it must pick up enough energy from an absorbed photon to move into a scattering state, which has a positive kinetic energy above the potential energy outside the metal. This can happen only if the energy of a photon exceeds the difference φ between the highest bound-electron energy (i.e., the Fermi level) and the outside potential energy.

14.3. Stimulated emission

On a microscopic level, physics is supposed to be time-symmetric: if it is possible for a ball to roll down a hill, converting potential energy to kinetic energy, it should also be possible for a ball with some initial kinetic energy to go up a hill and convert the kinetic energy to potential energy. By the same principle, we expect that the time-reversed version of absorption to be possible.

If initially there are n photons and a quantum system is at a higher energy level, it should be possible for the quantum system to relax to a lower energy, giving energy to the EM wave such that it has $n + 1$ photons afterwards. This is called **stimulated emission**. It is not so different from spontaneous emission, except that the initial photon number n is nonzero for stimulated emission.

For stimulated emission to happen, there must already be some nonzero initial electromagnetic energy, the quantum system must be in some higher energy state initially, and a lower energy state must be available for the system to decay to. This means that, for a quantum system with multiple electrons, there must be some vacant energy levels for a high-energy electrons to move into, just like Fig. 14.5. For a semiconductor, this means that there must be an electron-hole pair with the right energy difference to begin with, such that the electron can recombine with the hole and provide gain to the optical wave, as illustrated in Fig. 14.12. Stimulated emission is the quantum explanation of **optical gain** in a medium. Notable applications include **optical amplifiers** (which can compensate for optical losses in optical fiber communication systems) and **lasers** (gain medium in an optical cavity to produce a high optical output power).

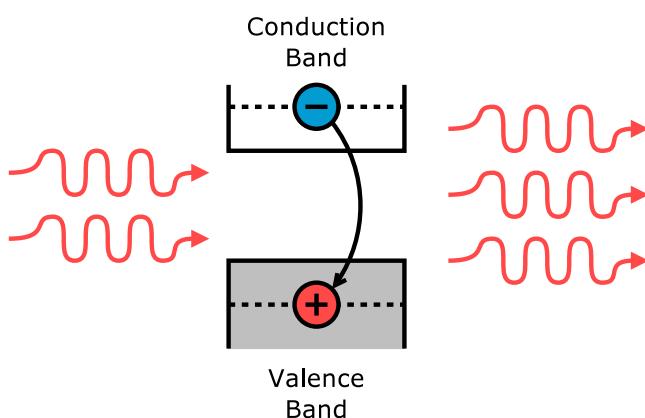


FIGURE 14.12. In a semiconductor, a conduction-band electron can recombine with a valence-band hole to provide one more photon to an EM wave in a stimulated emission process. Note that this is the time-reversed version of absorption shown in Fig. 14.8.

- **Exercise:** Find the wavelengths of laser diodes used in CD, DVD, and Blu-ray drives and confirm that they correspond to the bandgaps of the materials used to make the diodes.

14.4. Blackbody spectrum

39.30 A hollow box with a small aperture behaves like a blackbody. When the box is heated, the electromagnetic radiation that emerges from the aperture has a blackbody spectrum.

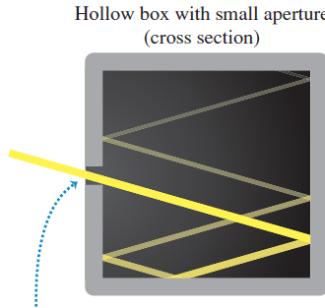


FIGURE 14.13. A box with walls at a certain temperature. The walls are assumed to be able to absorb all optical frequencies and therefore called a blackbody. Photons inside the box at thermal equilibrium with the walls will exhibit a blackbody spectrum. From [2].

Consider electromagnetic waves in a cavity with hot walls at a certain temperature T , as shown in Fig. 14.13. We assume that all kinds of energy levels are available at the walls, so that the photons are free to exchange energy with the walls at all possible frequencies. Photons can be absorbed or emitted, so the number of photons inside the cavity is not conserved, although the total energy of the walls and the photons must still be conserved. What is the optical spectrum at thermal equilibrium?

This spectrum is called the blackbody spectrum because the walls are assumed to be able to absorb all frequencies. In this context, a black body looks black only if it is at zero temperature and can only absorb; at higher temperatures a “black body” can also emit photons and won’t look black.

To compute the blackbody spectrum, we need two things: the **Bose-Einstein distribution** $f(E)$, which is the bosonic analog of the Fermi-Dirac distribution, and the **density of states** of EM waves $\rho(E)$ in a cavity. The energy E is the energy of each photon, so the EM-wave frequency will be $\omega = E/\hbar$.

Without photon-number conservation, the Bose-Einstein distribution is

$$f(E) = \frac{1}{\exp[E/(k_B T)] - 1}. \quad (14.7)$$

This can be derived using the kind of arguments we used in Chap. 12; see http://en.wikipedia.org/wiki/Bose-Einstein_statistics#Derivation_in_the_canonical_approach or Refs. [5, 7]. The only difference here is that photons are bosons, so multiple photons can occupy the same state, and the resulting $f(E)$ has different denominator.

The density of states for a large optical cavity, on the other hand, can be obtained by assuming a constant momentum-space density $C(k_x, k_y, k_z)$ like what we assumed in Chap. 13, assuming two spins (polarizations) for each single-photon state, and using the optical dispersion relation $E = \hbar\omega = \hbar ck$. The result is

$$\rho(E)dE = \frac{2}{\pi^3} \frac{\pi k^2}{2} dk, \quad \rho(E) = \frac{1}{\pi^2} \left(\frac{E}{\hbar c} \right)^2 \frac{1}{\hbar c}. \quad (14.8)$$

The photon-number density within energy interval E and $E + dE$ will then be

$$f(E)\rho(E)dE = \frac{1}{\pi^2(\hbar c)^3} \frac{E^2 dE}{\exp[E/(k_B T)] - 1}. \quad (14.9)$$

Let's convert this into frequency $\omega = E/\hbar$ to obtain the photon-number density within the frequency interval $\omega = E/\hbar$ to $\omega + d\omega = (E + dE)/\hbar$:

$$f(E)\rho(E)dE \Big|_{E=\hbar\omega} = \frac{1}{\pi^2 c^3} \frac{\omega^2 d\omega}{\exp[\hbar\omega/(k_B T)] - 1}, \quad (14.10)$$

and since each photon has energy $\hbar\omega$, the EM energy density in frequency interval between ω and $\omega + d\omega$ will become

$$Ef(E)\rho(E)dE \Big|_{E=\hbar\omega} = \frac{\hbar}{\pi^2 c^3} \frac{\omega^3 d\omega}{\exp[\hbar\omega/(k_B T)] - 1}. \quad (14.11)$$

This is Planck's law, which he derived by educated guessing from experimental data, before people knew anything about quantum mechanics. It was originally expressed in terms of the optical frequency in Hertz ν , such that $\hbar\omega = h\nu$, defining the Planck's constant as $h = 2\pi\hbar$. Here we have shown that the spectrum is a consequence of assuming EM waves to consist of bosons and using the maximum-entropy argument from statistical mechanics. If we just assume classical electromagnetism and statistical mechanics, the spectrum would blow up at small wavelengths and won't match the experimentally observed spectrum. Some plots of Planck's spectrum can be found in Fig. 14.14. Although the blackbody spectrum is derived using simple assumptions, it turns out to be a good approximation of the spontaneous emission spectra from many objects. For example, the solar spectrum obeys the blackbody spectrum approximately (<http://en.wikipedia.org/wiki/Sunlight>), even though the sun is the last thing one would call a black body in the usual sense.

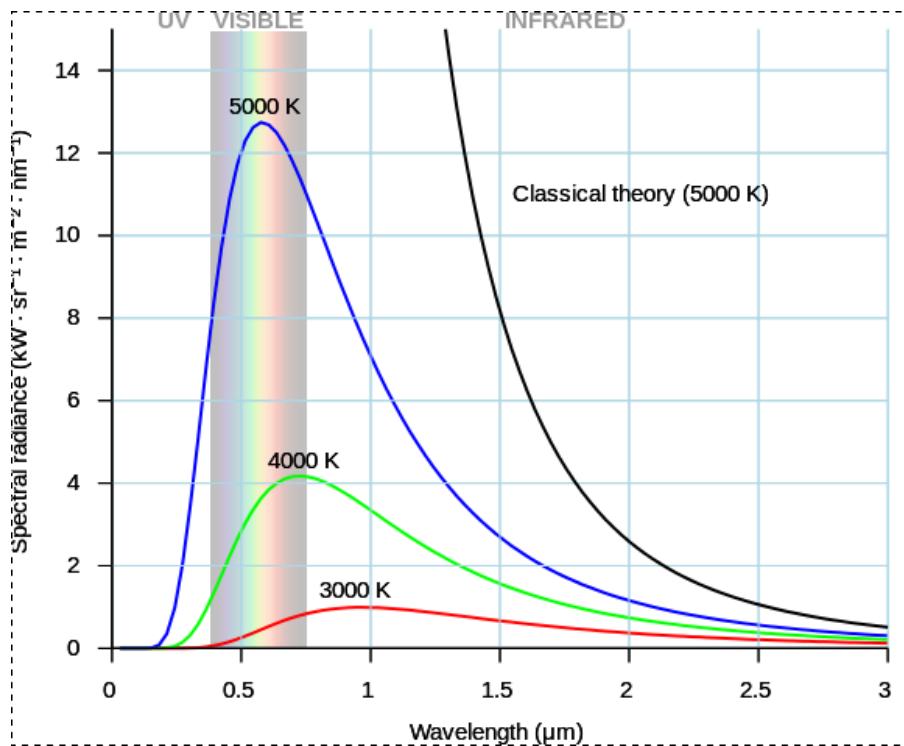


FIGURE 14.14. The blackbody spectrum for some temperatures. From http://en.wikipedia.org/wiki/File:Black_body.svg. As the temperature increases, the whole spectrum is increased and also the peak shifts to lower wavelengths.

- **Exercise:** From the blackbody spectrum, calculate the wavelength at which the spectrum is maximum. For human body temperature, what is this wavelength? Explain how an infrared thermal camera works by assuming that objects are blackbodies.

Spectrum of Solar Radiation (Earth)

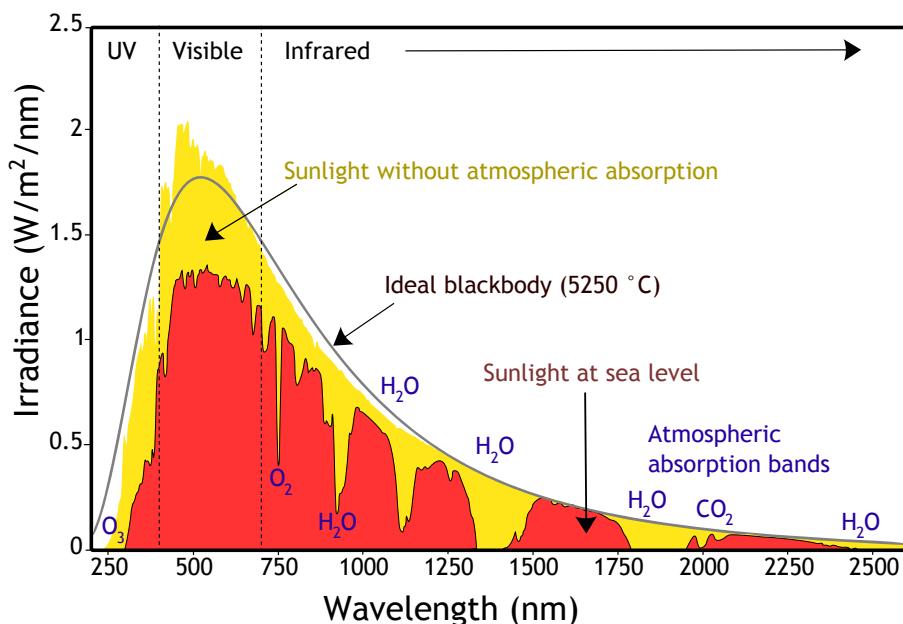


FIGURE 14.15. Comparison of the solar spectrum and the blackbody radiation. From http://en.wikipedia.org/wiki/File:Solar_spectrum_en.svg.

Bibliography

- [1] Steven Weinberg, *Quantum Theory of Fields* (University of Cambridge, Cambridge, 1995).
- [2] Hugh D. Young and Roger A. Freedman, *University Physics* (Pearson/Addison-Wesley, San Francisco, 2012).
- [3] A. Tonomura *et al.*, Am. J. Phys. **57**, 117 (1989) and image from <http://www.learningwithatlas-portal.eu/en/node/93624>,
- [4] Amnon Yariv and Pochi Yeh, *Photonics: Optical Electronics in Modern Communications* (Oxford University Press, New York, 2007).
- [5] David J. Griffiths, *Introduction to Quantum Mechanics* (Prentice-Hall, Upper Saddle River, 1995).
- [6] Ben G. Streetman and Sanjay K. Banerjee, *Solid State Electronic Devices* (Prentice-Hall, Upper Saddle River, 2006).
- [7] Amnon Yariv, *An Introduction to Theory and Applications of Quantum Mechanics* (Wiley, New York, 1982).
- [8] Joseph T. Verdeyen, *Laser Electronics* (Prentice Hall, New Jersey, 1995).
- [9] Amnon Yariv, *Quantum Electronics* (Wiley, New York, 1989).