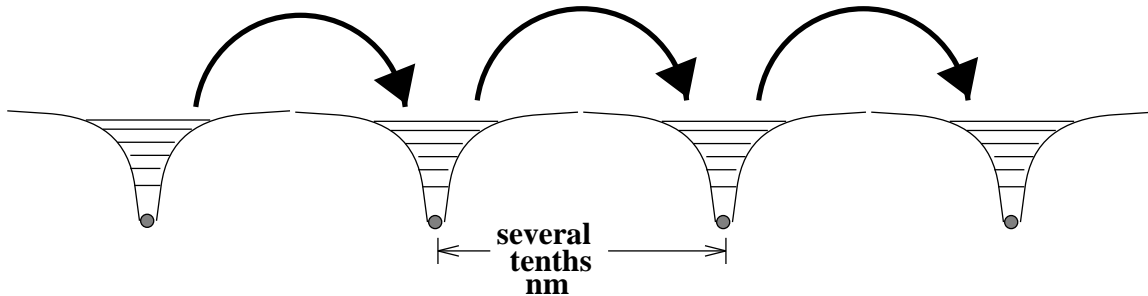


Energy Bands in Crystals

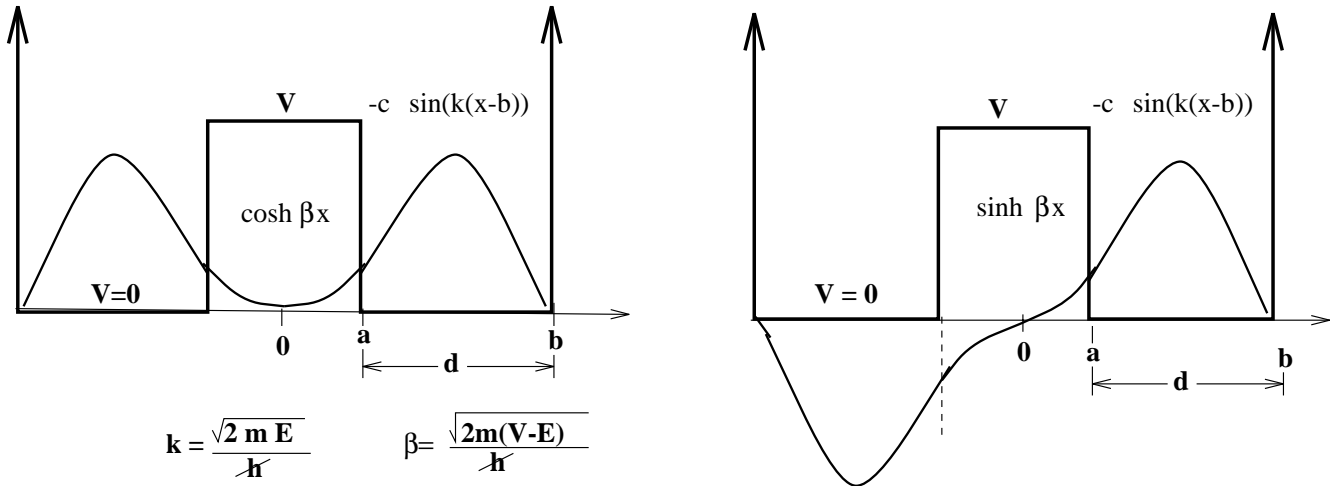
This chapter will apply quantum mechanics to a one dimensional, periodic lattice of potential wells which serves as an analogy to electrons interacting with the atoms of a crystal. We will show that as the number of wells becomes large, the allowed energy levels for the electron form nearly continuous energy bands separated by band gaps where no electron can be found. We thus have an interesting quantum system which exhibits many dual features of the quantum continuum and discrete spectrum.



The energy band structure plays a crucial role in the theory of electron conductivity in the solid state and explains why materials can be classified as insulators, conductors and semiconductors. The energy band structure present in a semiconductor is a crucial ingredient in understanding how semiconductor devices work.

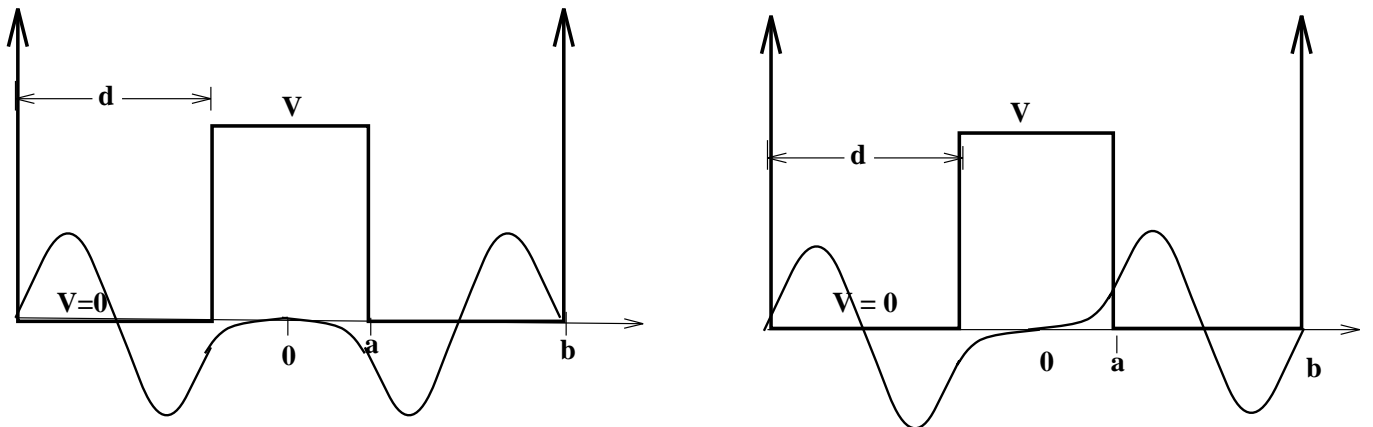
Energy levels of “Molecules”

By a “molecule” a quantum system consisting of a few periodic potential wells. We have already considered a two well molecular analogy in our discussions of the ammonia clock.



Recall for reasonably large hump potentials V , the two lowest lying states (a even ground state and odd first excited state) are very close in energy. As $V \rightarrow \infty$, both the ground and first excited state wave functions become completely isolated within a well with little wave function penetration into the classically forbidden central hump. The electron in either well has a wavelength which approaches $\lambda \rightarrow 2d$ and the both the ground and first excited state energies become degenerate at $E_2 \rightarrow E_1 \rightarrow \hbar^2 \pi^2 / (2md^2)$. The incredibly small $\Delta E = E_2 - E_1$ in ammonia makes this molecule useful in technological applications such as radiofrequency clocks and MASERS.

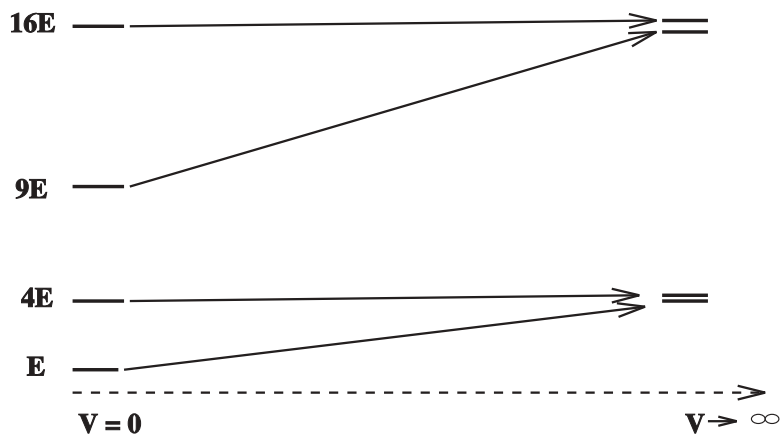
Here is a sketch of the second and third excited states in the large V limit.



As $V \rightarrow \infty$, $\lambda_4 \rightarrow \lambda_3 \rightarrow d$ and both states become nearly degenerate at

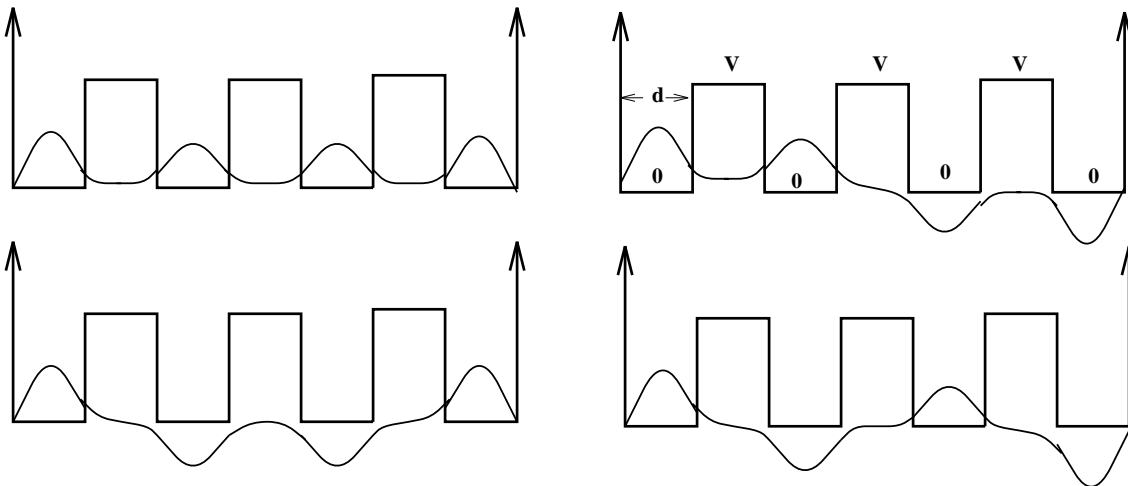
$E_4 \rightarrow E_3 \rightarrow 4 \hbar^2 \pi^2 / (2md^2)$. We could call the $V \rightarrow \infty$ limit the tightly bound limit since the electrons are tightly bound in their quantum wells owing to the large well depth.

At the other extreme we have $V \rightarrow 0$ or the “loosely” bound limit. In this limit, the pattern of energy levels becomes completely different. As $V \rightarrow 0$, we have a single, infinite square well of width $= 2(a+d)$ and an energy spectrum given by the familiar form $E_n = \hbar^2 \pi^2 n^2 / (2m(a+d)^2)$. The below sketch summarizes how one might expect the energy level pattern for the 4 lowest energy states to change for a well with $a \ll d$ as one varies goes from the weak to strong binding limit.

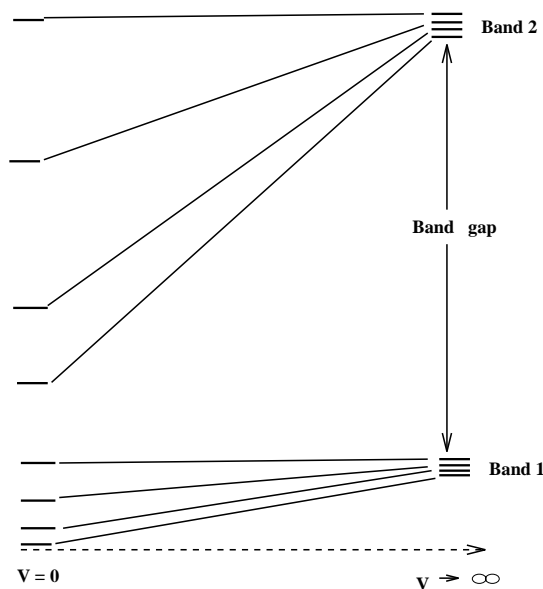


The reader can easily confirm the energy levels at the two extremes. For convenience, I drew straight lines between the $V = 0$ and $V \rightarrow \infty$ energy levels. In reality they will be non-crossing curves that one would have to obtain by the solving the transcendental equations for various values of V . This type of energy diagram which shows the evolution of the energy levels as the atoms become increasingly more isolated is often called a *correlation diagram*.

We next sketch lowest lying state wave functions for a one dimensional well model of a 4 atom molecule.



These are four tightly bound states, which have either even or odd parity. In the tight binding limit, we expect all 4 low lying states to be nearly degenerate with a wavelength of $\lambda = 2d$ and an energy of $E = \hbar^2\pi^2/(2md^2)$. The energy level diagram as one goes from the loosely bound to tightly bound limit might crudely resemble the following:



We note as $V \rightarrow \infty$ the lowest 8 states naturally organize themselves into 2 energy “bands” with four electron orbitals (neglecting spin) per band. The bands shrink as the electron becomes more tightly bound to a given well and

the band gap increases. Indeed the band clustering of electronic energy levels is observed in molecules although in a more complicated way owing to the effects of spin and exchange symmetry discussed in the chapter **Assorted Aspects of Atomic Physics**.

The forgoing introduction has all been very qualitative. Can one actually solve for the bound states of such finite multiwell structures? The answer is fortunately yes but the task at first appears daunting. The four well problem illustrated above has 7 quantum regions 4 of which are classically allowed and 3 of which are classically forbidden. The wave function for each of the 4 classically allowed regions is of the form $\psi_i = \alpha_i \sin(kx) + \gamma_i \cos(kx)$; while the wave function in the 3 classically forbidden regions is of the form $\psi_i = \delta_i \exp(\beta x) + \epsilon_i \exp(-\beta x)$. Off hand there appear to be 14 unknown amplitude coefficients and one unknown energy. One of the amplitudes can be eliminated by normalization, two amplitudes can be eliminated since there are nodes on either end of the molecule which leaves a total 12 unknowns including the bound state energy. There are 12 remaining boundary conditions, continuity and derivative continuity across each of the 6 boundaries, which enable one to write a hellacious transcendental equation for the bound state energies involving 12 equations and 12 unknowns!

Of course, the problem becomes twice as easy to solve once one exploits the parity symmetry of the potential well by placing our origin in the center of the central potential hump. We could then begin searching for even or odd wave functions. For the case of the even wave functions, in the central hump region, we know the wave function must be of the form of $\psi = \cosh(\beta x)$. We need only solve for the boundary conditions on the right hand side of the well because of the symmetry. We can then write the transcendental equation down by matching 3 sets of boundary conditions which apply to the right hand side of the well and solve for the 5 unknown amplitude coefficients and energy. This results in a less hellacious transcendental equation involving “only” 6 equations in 6 unknowns.

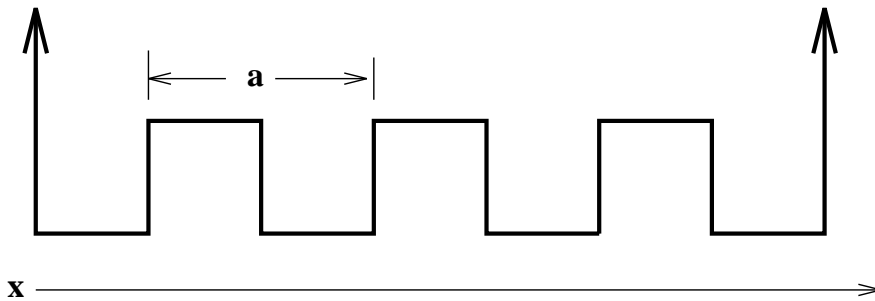
Perhaps this is within the analytic range of a really persistent and careful student, but what if we “upped the ante”, and asked you to solve a more realistic example consisting of the roughly 10^{20} quantum wells representing electrons interacting with a periodic, crystalline lattice. Using the methodology described above, this problem would likely be too much for the fastest supercomputer. Interestingly enough, for the case of a really huge number of atoms in a periodic lattice, a new symmetry appears which makes the problem of writing an exact transcendental equation to compute the band structure fairly simple.

Exploiting Translational Symmetry in Periodic Structures

The new symmetry which we wish to exploit in modeling the electronic structure of one dimensional crystals is called *translational symmetry*. By translational symmetry we mean the potential energy has the property that:

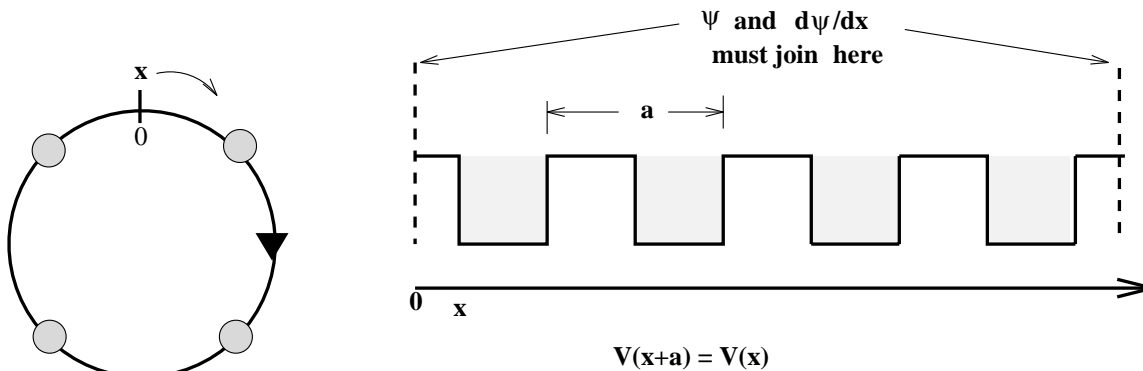
$$V(x + a) = V(x) \tag{1}$$

where a is called a “lattice” spacing which gives a repeat distance scale for the crystal. Although the four potential wells in the system illustrated below are fairly similar, the potential does not have true translation symmetry.



The reason is that the end sections which require a node break the symmetry. The well is not truly the same even if one translates the potential by a distance a since ultimately one will hit the end of the well. If one considered such a structure repeated thousands of times, one would expect such a lattice be very nearly periodic with nearly exact translational symmetry since only the two end “atoms” , a very small fraction of the total, would break the symmetry.

One can have true translational symmetry if instead one considers a ring lattice of say four atoms.



As we go around the ring lattice through one complete cycle, we are back to where we started and hence the wave function and its derivatives must be continuous at the origin as indicated above. The true translation symmetry of the ring lattice offers us the promise of setting up a transcendental equation for the energy levels of an electron in the lattice by considering only one of the many periodic regions of the well, such as the region from $0 < x < a$.

Since the potential in a ring lattice is periodic with lattice spacing a , we might be tempted to think that the wave function itself is periodic in the sense:

$$\psi(x + a) = \psi(x) \quad \text{not true!} \quad (2)$$

It turns out that this is not so. The underlying symmetry of the situation applies to the PDF of the electron or the modulus of the wave function, not its wave function itself.

$$|\psi(x + a)| = |\psi(x)| \quad (3)$$

Recall that a very similar thing happens when we applied parity considerations to one dimensional bound states in symmetric wells. The actual condition relating

the left and right wave functions was $\psi(-x) = \pm\psi(x)$. For the 1st, 3rd , 5th ... excited states of a symmetric well, the wave function on the left differ from the wave function on the right by a -1 phase factor.

Applying this lesson to the case of a translationally invariant potential we anticipate:

$$\text{If } V(x+a) = V(x) \text{ then } \psi(x) = e^{i\chi(x)} u(x) \text{ where } u(x+a) = u(x) \quad (4)$$

In words, Eq. (4) tells us that if a potential is periodic with a lattice spacing a , the wave function must be of the form of a possibly x dependent phase (described by a phase factor function $\chi(x)$) times a translationally symmetric function called $u(x)$. There is no implied translational invariance to the phase $\exp(i\chi(x))$ part of $\psi(x)$. Recall from the chapter on **Quantum Mechanics in Multidimensions** the inclusion of such a position dependent phase means that even a stationary state wave function (representing a spinning electron cloud) can carry a probability current. In this context the phase factor can allow for the steady-state flow of electrons through the lattice.

What could be the functional form of such a position dependent phase factor? To answer this, lets look at the wave function for the case of constant potential. The constant potential represents the most trivial case for a periodic potential since $V(x+a) = V(x)$ for all a . We know the solutions of the time independent Schrödinger Equation for this case are of the traveling wave form:

$$\text{For } V(x) = \text{constant} \quad : \quad \psi(x) = e^{ikx} = e^{i\chi(x)} u(x) \quad (5)$$

In this case of the constant, periodic potential, the position dependent phase was of the form $\exp(ikx)$ and the periodic function became $u(x) = 1$. I will show shortly that the position dependent phase always has the form $\exp(ikx)$ and the wave function for the general problem of the periodic potential can always be

represented by:

$$\psi_k(x) = e^{ikx} u_k(x) \quad \text{where } u_k(x+a) = u_k(x) \quad (6)$$

You will note that I put the k subscript on both the wave function ψ_k as well as the periodic function $u_k(x)$. One can think of the variable k as labeling the eigenstates in much the same way as $+$ or $-$ label even or odd eigenstates for a simple bound state well. Wave functions of the form of Eq. (6) are known as Bloch waves. In the loose binding limit the $u_k(x)$ will serve as a fairly gentle modulation of the quantum mechanical traveling wave $\psi(x) = \exp(ikx - i\omega t)$. The Bloch wave electron thus moves more or less freely through the crystal with a slight modulation near the lattice sites.

The validity of such solutions to periodic wave functions in one dimensional systems is known as Floquet's theorem and in three dimensions is known as Bloch's theorem. Below is a simple, one dimensional proof.

Why do Bloch Waves work?

Of course any function can be written in the form of a Bloch wave: $\psi_k(x) = e^{ikx} u_k(x)$ where u_k is an arbitrary function. Our demonstration of the validity of the Bloch wave function rests on proving that $u_k(x)$ is periodic in the sense $u_k(x+a) = u_k(x)$. As we will show, this proof rests on demonstrating that ψ_k is an eigenstate of the translational operator.

Both the $+$ or $-$ labeling for parity eigenstates and the k labeling for Bloch waves have special significance. Recall in the chapter on **Quantum Mechanics in Multidimensions** we discussed the concept of a symmetry operator and applied it to rotational symmetry. We can think of both parity and translational symmetry as quantum mechanical operators as well. For example we can define the parity operator \check{P} according to:

$$\check{P}\psi(x) = \psi(-x) \quad (7)$$

For the case of a symmetric potential, $V(x) = V(-x)$, the parity operator com-

commutes with the Hamiltonian since $\partial^2/\partial(-x)^2 = \partial^2/\partial x^2$:

$$\begin{aligned} \check{P} \check{H} \psi(x) &= \check{P} \left(\frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right) \psi(x) \\ &= \left(\frac{-\hbar^2}{2m} \frac{\partial^2}{\partial(-x)^2} + V(-x) \right) \psi(-x) = \left(\frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right) \psi(-x) = \check{H} \check{P} \psi(x) \\ \text{therefore } \check{P} \check{H} \psi(x) &= \check{H} \check{P} \psi(x) \text{ or } [\check{P}, \check{H}] = 0 \end{aligned} \quad (8)$$

As discussed in the chapter on **Quantum Measurements**, the eigenstates of the Hamiltonian can be chosen to be simultaneous eigenfunctions of a commuting operator such as parity. We can label the parity eigenstates with + or - and note that the odd and even states have a parity eigenvalue of ± 1 as indicated below:

$$\check{P} \psi_{\pm}(x) = \psi_{\pm}(-x) = \pm 1 \psi_{\pm}(x) \quad (9)$$

Essentially the same things are true of translational invariance which we can define by an operator \check{T} (not to be confused with kinetic energy!) where

$$\check{T}\psi(x) = \psi(x+a) \quad (10)$$

For the case of a translationally symmetric potential $V(x) = V(x+a)$ and the translation operator commutes with the Hamiltonian since $\partial^2/\partial(x+a)^2 = \partial^2/\partial x^2$:

$$\begin{aligned} \check{T} \check{H} \psi(x) &= \check{T} \left(\frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right) \psi(x) \\ &= \left(\frac{-\hbar^2}{2m} \frac{\partial^2}{\partial(x+a)^2} + V(x+a) \right) \psi(x+a) = \left(\frac{-\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right) \psi(x+a) = \check{H} \check{T} \psi(x) \\ \text{therefore } \check{T} \check{H} \psi(x) &= \check{H} \check{T} \psi(x) \text{ or } [\check{T}, \check{H}] = 0 \end{aligned} \quad (11)$$

We thus should be able to choose eigenstates of the Hamiltonian which are also

eigenstates of the translational operator or:

$$\tilde{T} \psi_k(x) = \psi_k(x + a) = \lambda_k \psi_k(x) \quad (12)$$

Since we have a translationally symmetric PDF, we note that the eigenvalue λ_k can only be a phase factor which we can arbitrarily write in terms of the lattice spacing as $\lambda_k = \exp(ika)$. We thus require:

$$\psi(x + a) = \tilde{T}\psi_k(x) = \lambda_k \psi_k(x) = e^{ika} \psi(x) \quad (13)$$

We now can substitute $\psi_k = \exp(ikx) u_k(x)$ into Eq. (13) and try to show that $u_k(x)$ is periodic.

$$\begin{aligned} e^{ik(x+a)} u_k(x + a) &= \psi_k(x + a) = \tilde{T} \psi(x) = e^{ika} \psi_k(x) = e^{ika} e^{ikx} u_k(x) \\ &= e^{ik(x+a)} u_k(x) \quad \text{Therefore } u_k(x + a) = u_k(x) \end{aligned} \quad (14)$$

The translational symmetry of the Hamiltonian implies that the Bloch wave functions are eigenvectors of the translational operator. The subscript k , labels the state ψ_k by its translational eigenvalue $\exp(ika)$.

What values of k are possible?

For the case of the constant potential, k is a continuous real variable which is related to the momentum via $p = \hbar k$. For the case of a finite ring lattice, we will show that only discrete values of k are possible. Let us say the ring lattice consists of N atoms with lattice spacing a . Since the lattice is on a ring we must insist that the wave function has a unique value once goes completely about a complete circumference (Na) of the ring. Hence:

$$\psi(x + Na) = \psi(x) \quad (15)$$

One can easily build up a complete translation through Na by making N repeated

translations of a using Eq. (13):

$$\psi(x + a) = e^{ika} \psi(x) \quad , \quad \psi(x + 2a) = e^{ika} \psi(x + a) = e^{i2ka} \psi(x)$$

$$\text{Therefore : } \psi(x + Na) = \left(e^{ika} \right)^N \psi(x) = e^{iNka} \psi(x) \quad (16)$$

The Eq. (16) condition implies:

$$\psi(x + Na) = e^{iNka} \psi(x) = \psi(x) \quad \text{or} \quad e^{iNka} = 1 \quad (17)$$

The latter condition implies Nka must be an integral multiple of 2π

$$Nka = 2\pi n \quad \text{or} \quad k = n \frac{2\pi}{Na} \quad \text{where } n \in \{0, \pm 1, \pm 2, \dots\}$$

Hence k itself lies on a discrete lattice

$$k \in \left\{ 0, \pm \frac{2\pi}{Na}, \pm 2 \frac{2\pi}{Na}, \dots \right\} \quad (18)$$

In the next section, we will show that the first energy band exists in the k range $|k| < \pi/a$. The second energy band exists from $\pi/a < |k| < 2\pi/a$ and so on. The discrete lattice of k described by Eq. (18) tells us that there are N energy levels in each energy band where N is the number of atoms in the crystal.² Generally N is huge, and the k approaches a continuous variable and the energy levels approach a continuous band as well.

What does k mean physically?

² We note that there were 2 energy levels per “band” in our two atom molecule and 4 energy levels per band in our four atom molecule which seems to confirm this point. As the exercises will point out, this is actually an “accident” rather than a confirmation since our molecules are not truly periodic and thus the wave functions are not actual Bloch waves.

One is very tempted when viewing the Bloch form $\psi_k = \exp(ikx) u_k(x)$ to equate $\hbar k$ with the electron momentum. A little reflection will show that this cannot be true for an electron experiencing a periodic potential. The k quantum number which labels the Bloch wave function is a constant of motion. Unless $V(x)$ is constant, the actual electron momentum changes as the electron feels the force due to the crystal lattice and momentum cannot be used to label the wavefunction for an electron in a crystal. Hence k is not the true momentum but is actually more useful than the true momentum since it is a constant of motion.

The object of quantum game when applied to an electron in a crystal is to use translational invariance in the form of Eq. (13) to compute E_k (or in the continuum limit $E(k)$). Of course it is always worthwhile to find the allowed energies of a quantum system. The problem of computing the “dispersion” relation $E(k)$, which we will undertake in the next section, provides much richer fruits than just this.

The k quantum number plays a critical role in describing how an electron absorbs energy. As an example consider applying an external (i.e. non-lattice) force, F_{ext} , on an electron traveling in a one-dimensional crystal for a small time interval, Δt . We will say while this force is being applied the electron is traveling with a velocity \bar{v} . Recall $F_{\text{ext}}\bar{v}$ is the instantaneous power acting on the electron and hence $\Delta E = F_{\text{ext}}\bar{v} \Delta t$ must be the energy absorbed by the electron during this small time interval. Given that the electron can only be in a quantum state described by k , the change in energy must be accompanied by a change in k or:

$$\Delta E = F_{\text{ext}}\bar{v} \Delta t = \frac{\partial E}{\partial k} \times \Delta k \quad (19)$$

We can write Eq. (19) in the suggestive form:

$$\frac{1}{\hbar} \frac{\partial E}{\partial k} \times \hbar \frac{\Delta k}{\Delta t} = \frac{1}{\hbar} \frac{\partial E}{\partial k} \times \hbar \frac{dk}{dt} = \bar{v} \times F_{\text{ext}} \quad (20)$$

Although we haven’t proved it, the correct assignment of the \bar{v} and F_{ext} terms in

Eq. (20) is:

$$\bar{v} = \frac{1}{\hbar} \frac{\partial E}{\partial k} \quad \text{and} \quad F_{\text{ext}} = \hbar \frac{dk}{dt} \quad (21)$$

The expression $F_{\text{ext}} = \hbar dk/dt$ is very reminiscent of Newton's Second Law or the force acting on a particle is equal to rate of change of the particle's momentum. But to apply Newton's law we would need to consider both the external forces plus all internal forces due to interactions with nuclei of the lattice and all other electrons. In most cases these internal forces would be much larger than any external forces one could apply through external electric or magnetic fields. Hence k is not the physical momentum but rather is a conserved quantum number that stays constant if no external forces are applied. It is often called the *crystal* momentum. Similarly

$$\bar{v} = \frac{1}{\hbar} \frac{\partial E}{\partial k} \quad (22)$$

is extremely reminiscent of the formula for group velocity of a free particle wave packet which we first met in the chapter on **Quantum Measurement**:

$$\nu = \frac{d\omega}{dk} = \frac{1}{\hbar} \frac{dE}{dk} \quad (23)$$

However the Eq. (23) only makes sense for a free particle while Eq. (22) applies to an electron traveling within a varying, periodic potential. It is relatively easy³ to derive Eq. (22) by first order perturbation theory, where in this derivation \bar{v} becomes the average velocity ($\bar{v} = d\langle x \rangle/dt$) of the electron in the crystal.

Eq. (22) says that an electron in a crystal which is an eigenstate of k and E_k where $dE/dk \neq 0$, actually travels freely through the crystal with a constant average velocity. The electron flows effortlessly through the crystal quantum tunneling over every potential well within the lattice.⁴

³ A good treatment is in Appendix E of **Solid State Physics** by Ashcroft & Mermin

⁴ In fact this is not true because vibrational energy transfer (phonon's) with the lattice except in the case of superconductors.

Interestingly enough, when an external force is applied to the electron (by applying an electric field to the crystal), it generally accelerates with the “wrong” acceleration. The electron appears to have a different “mass” in the crystal than it does in a vacuum.

In order to see this, we revisit Eq. (21) with the goal of computing the average acceleration or $\bar{a} \equiv d\bar{v}/dt$. Since $\bar{v} = \hbar^{-1}\partial E/\partial k$ the only way one can change \bar{v} is to change k and then the change in $\partial E/\partial k$ is proportional to the double derivative: $\partial E^2/\partial k^2$.

$$\Delta\bar{v} = \frac{1}{\hbar}\Delta\left(\frac{\partial E}{\partial k}\right) = \frac{1}{\hbar}\left(\frac{\partial^2 E}{\partial k^2} \times \Delta k\right) \rightarrow \bar{a} = \frac{\Delta\bar{v}}{\Delta t} = \frac{1}{\hbar}\frac{\partial^2 E}{\partial k^2} \times \frac{\Delta k}{\Delta t} \quad (24)$$

We now use $F_{\text{ext}} = \hbar\bar{a}$ to write Eq. (24) in the form:

$$\bar{a} = \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2} F_{\text{ext}} = \frac{F_{\text{ext}}}{m_*} \quad (25)$$

where in Eq. (25) we defined the “effective” mass as $F_{\text{ext}} = m_*\bar{a}$. Hence:

$$\frac{1}{m_*} = \frac{1}{\hbar^2}\frac{d^2 E}{dk^2} \quad (27)$$

We have written an *effective mass*, m_* , rather than the free particle mass m of the electron. There is no guarantee in a crystal that $\hbar^{-2} d^2 E/dk^2$ will equal the free particle mass ($mc^2 = 0.511 \times 10^6 \text{ eV}$) of an electron. As we will show in the next section all things are possible: sometimes $\hbar^{-2} d^2 E/dk^2 > m$, sometimes $\hbar^{-2} d^2 E/dk^2 < m$ and sometimes $\hbar^{-2} d^2 E/dk^2 < 0$. Sometimes an electron in a crystal can actually accelerate in a direction **opposite** to the applied force!

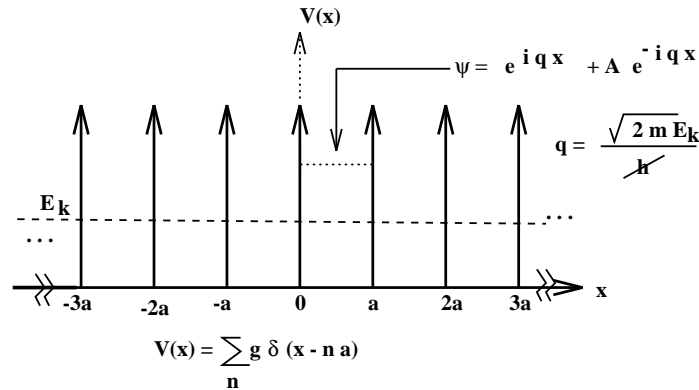
So what does crystal momentum k mean physically? We have seen that semiclassically k is sort of an “external” electron momentum which is effectively blind to the internal forces created by the lattice potential. In the absence of any forces external to the crystal itself, k is a constant of motion which serves

to label the Bloch wave function. In very close correspondence with the concept of group velocity, $\hbar^{-1} d^2 E/dk^2$ represents the average velocity of an electron traveling through the crystal. When an *external* force is applied the electron changes crystal momentum in a way very similar to Newton's second law $F = \hbar dk/dt = m_* d\bar{v}/dt$. However the effective mass, m_* which controls the average acceleration of the crystal bound electron is generally not the free particle mass, but rather is related to $E(k)$ through the relation: $m_*^{-1} = \hbar^{-2} d^2 E/dk^2$. To compute the "inertia" of the electron one must use the Bloch condition to solve the Schrödinger Equation and find $E(k)$. Lets do exactly that.

Kronig-Penney Model

The Kronig-Penney model is a solveable system which consists of a periodic rectangular lattice potential. By "solving" the system, we are looking for the energy as a function of the crystal momentum, $E(k)$ ⁴ or E_k . Perhaps the simplest to solve version would be an infinite, periodic lattice of δ functions or:

$$V(x) = \sum_{n=-\infty}^{n=+\infty} g \delta(x - na) \quad (28)$$



⁴ In this business k which labels the translational eigenvalue for the state is an input quantity rather than an unknown. Each k should be selected according to the allowed values given by Eq. (18). For the case of an infinite Kronig-Penney lattice, k is a continuous variable.

Since we are in 0 potential region except right at the each δ function we write a general traveling wave solution in the region from $0 < x < a$ of the form:

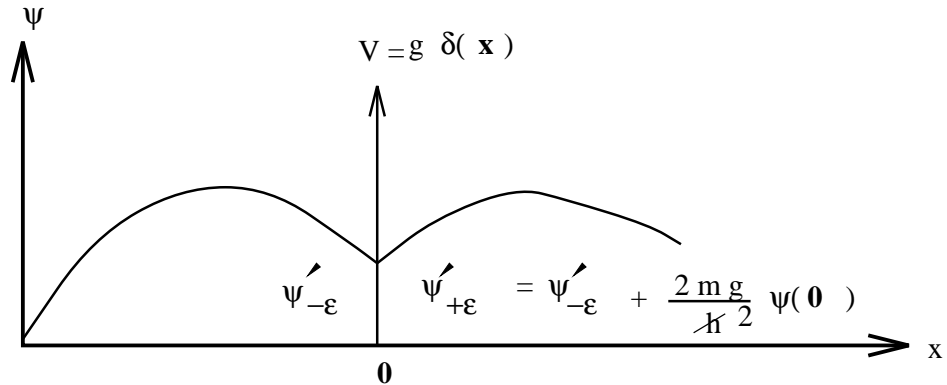
$$\psi(x) = e^{iqx} + A e^{-iqx} \quad \text{for } 0 < x < a \quad (29)$$

I have arbitrarily set the amplitude of the $\exp(iqx)$ piece to unity and I use the symbol q to stand for $2\pi/\lambda$ to distinguish it from the “crystal momentum” Bloch variable, k . There are two unknowns: E_k and the unknown amplitude A . We clearly need two boundary conditions.

The first of these boundary conditions is continuity of the wave function which we will take in the vicinity of the origin:

$$\psi(+\epsilon) = \psi(-\epsilon) = \psi(0) \quad (30)$$

where ϵ stands for an infinitesimal number. The extreme singularity of a Dirac δ -function creates a slope discontinuity which replaces the usual boundary condition of a continuous derivative. Recall in the chapter on **Quantum Tunneling** we showed that a Dirac δ -function potential creates a slope discontinuity in the wave function (at the point $x = 0$) as illustrated below



$$\frac{\partial \psi}{\partial x} \Big|_{+\epsilon} - \frac{\partial \psi}{\partial x} \Big|_{-\epsilon} = \frac{2mg}{\hbar^2} \psi(0) \quad \text{for } V(x) = g\delta(x) \quad (31)$$

Now we appear to have a problem in applying the boundary conditions given by Eq. (30) and Eq. (31) since our wave function $\psi(x) = \exp(iqx) + A \exp(-iqx)$

only applies in the open interval $+\epsilon < x < a - \epsilon$, and yet we need to get to the negative side of the δ function at 0.

This is where the crystal momentum comes in! We invoke the Bloch condition through Eq. (13)

$$\psi(x + a) = e^{ika} \psi(x)$$

Specifically we apply it to both the ψ and its derivative for a translation a in the vicinity of $x = -\epsilon$:

$$e^{ika} \psi(-\epsilon) = \psi(a - \epsilon) \quad , \quad e^{ika} \frac{\partial \psi}{\partial x} \Big|_{-\epsilon} = \frac{\partial \psi}{\partial x} \Big|_{a-\epsilon}$$

$$\text{or} \quad \psi(-\epsilon) = e^{-ika} \psi(a - \epsilon) \quad , \quad \frac{\partial \psi}{\partial x} \Big|_{-\epsilon} = e^{-ika} \frac{\partial \psi}{\partial x} \Big|_{a-\epsilon} \quad (32)$$

We can use Eq. (32) to evaluate the wave function and its derivative just left of the origin, since we know that $\psi(x) = \exp(iqx) + A \exp(-iqx)$ is valid at $x = a - \epsilon$:

$$\psi_{-\epsilon} = e^{-ika} \psi(a - \epsilon) = e^{-ika} (e^{iqa} + A e^{-iqa})$$

$$\frac{\partial \psi}{\partial x} \Big|_{-\epsilon} = e^{-ika} \frac{\partial \psi}{\partial x} \Big|_{a-\epsilon} = e^{-ika} \frac{\partial}{\partial x} (e^{iqx} + A e^{-iqx}) = e^{-ika} (iqe^{iqa} - iqAe^{-iqa}) \quad (33)$$

Using the $-\epsilon$ forms, the continuity condition becomes:

$$1 + A = e^{-ika} (e^{iqa} + A e^{-iqa}) = e^{i(q-k)a} + A e^{-i(q+k)a} \quad (34)$$

The derivative discontinuity condition , Eq. (31), becomes:

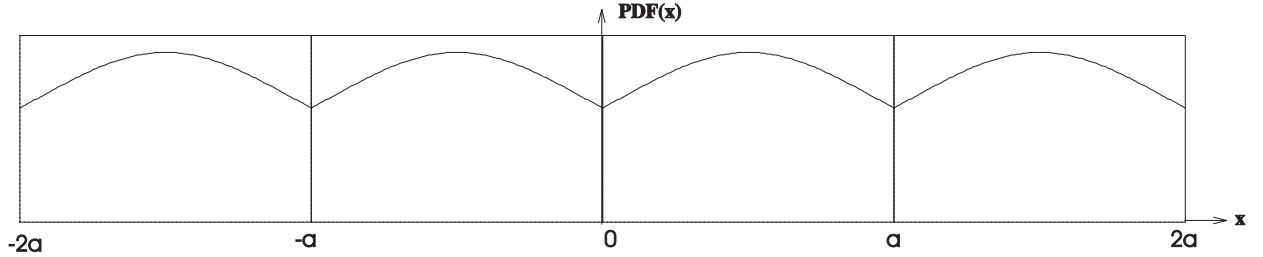
$$(iq - iqA) - e^{-ika} (iqe^{iqa} - iqAe^{-iqa}) = \frac{2mg}{\hbar^2} (1 + A) \quad (35)$$

where we use $\partial \psi / \partial x \Big|_{+\epsilon} = (iq - iaA) \exp(iq\epsilon) \rightarrow (iq - iaA)$. We can solve both Eq. (34) and (35) for the unknown A and equate the two expressions. After

some algebra, we obtain the transcendental equation for q (or E):

$$\cos(qa) + \frac{mg}{\hbar^2 q} \sin(qa) = \cos(ka) \quad \text{where } q = \sqrt{2mE}/\hbar \quad (36)$$

The left hand side (LHS) of Eq. (36) is a function of q or indirectly E only; while the right hand side (RHS) is a function of k only. Hence Eq. (36) serves as an elegant mapping for assigning a translational eigenvalue (or crystal momentum) , k , for each electron energy or vice versa. For the case of the Kronig-Penny potential, for a given k , one can find q using Eq. (36). One can then solve for A using Eq. (34) and construct the PDF in the region ($0 < x < a$) as $PDF = |u_k(x)|^2 \propto |e^{iqx} + A e^{-iqx}|^2$. We can then use translational invariance $u_k(x) = u_k(x + a)$ to generate the PDF in all other regions. An example might look like the below figure:



In this example we are in the middle of the first band. You can see the discontinuity in the PDF slope that occurs at each δ -function. If g were to decrease to zero, $u_k(x)$ would approach a constant and $\psi_k \rightarrow \exp(ikx)$ which is the solution in a field free region. For finite g , the modulations of the PDF become very large near the band edges at $k \in \{\pm\pi/a, \pm2\pi/a, \pm3\pi/a...\}$ since $A \rightarrow \pm 1$ as we will show shortly.

We next discuss why there are energy gaps based on the transcendental equation. As k is varied continuously, $\cos(ka)$ sweeps through the range $1 \geq \cos(ka) \geq -1$. The LHS of Eq. (36) is pseudo-periodic in q with an amplitude⁵

of roughly⁶

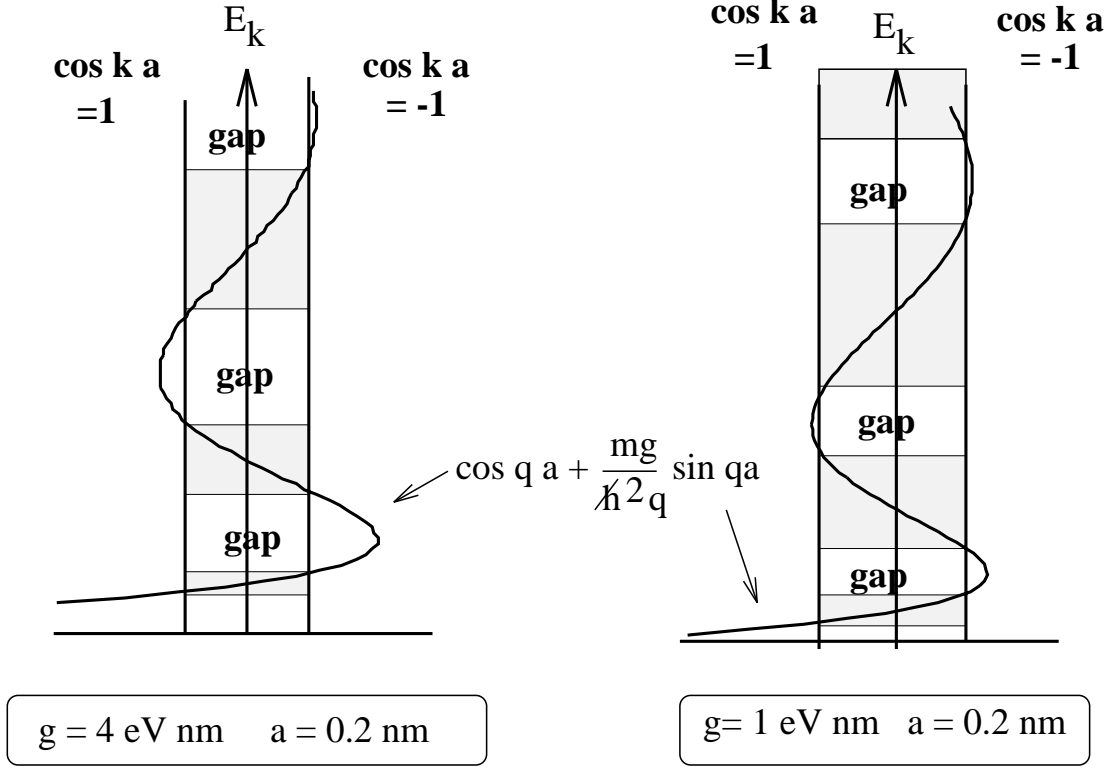
$$\text{amplitude} \approx \sqrt{1 + \frac{m^2 g^2}{\hbar^4 q^2}} \geq 1 \quad (37)$$

Since the LHS swings past ± 1 which are the limits of the RHS, there will be forbidden q (or E values of the electron which can never be matched by the RHS) and therefore cannot be assigned a valid translational eigenvalue k . These electron energies are forbidden, and lie in an energy “gap”.

The below figure shows a rough sketch of the LHS and RHS of Eq. (36) as a function of E_k for two values of the δ -function strength g for an electron placed in a lattice with a spacing of $a = 0.2 \text{ nm}$. I have shaded those allowed regions where the LHS lies within ± 1 . These shaded regions, which form energy bands, represent possible electron energies since they can always be assigned a corresponding k value through Eq. (36). The first band extends from $k = \pm 0$ ($\cos ka = 1$) to $k = \pm \pi/a$ ($\cos ka = -1$). All subsequent band boundaries occur at integral multiples of $\pm \pi/a$.

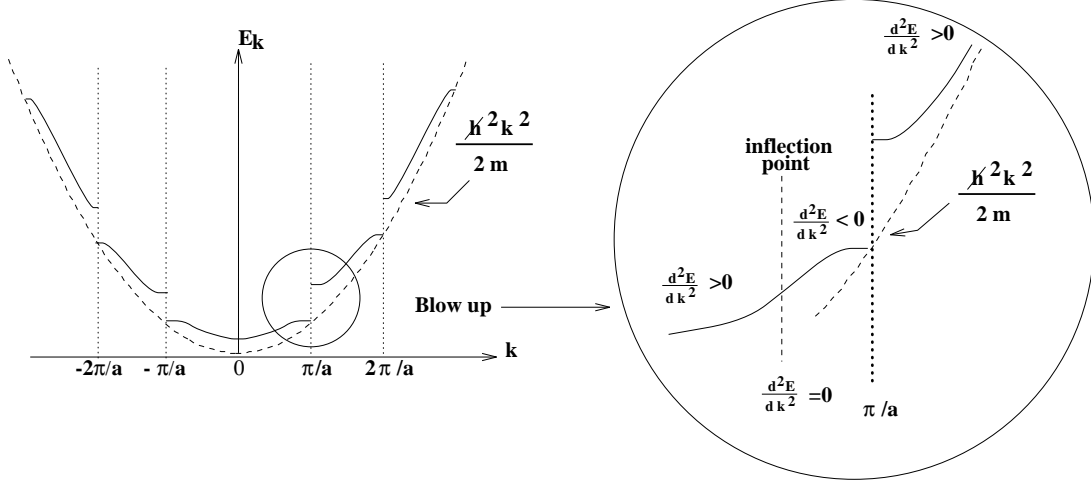
5 This is the amplitude in the limit that the variation in the LHS of Eq. (36) due to q is less important than the variation due to $\cos qa$ or $\sin qa$.

6 Recall from your knowledge of phasors that a wave $\psi = \alpha \sin kq + \beta \cos kq$ can be written as $\sqrt{\alpha^2 + \beta^2} \sin(ka + \phi_o)$ where $\tan \phi_o = \beta/\alpha$.



We note that as g is increased, the energy gap between allowed bands is increased, since according to Eq. (37), the LHS “amplitude” increases. Exactly the same effect was observed in our “molecule” model: as $V \rightarrow \infty$, the electron states became increasingly more isolated and degenerate causing the allowed “bands” to shrink and the gaps to grow. The limiting case is $g \rightarrow 0$, where the gaps disappear completely. This corresponds to the zero potential case where, of course, all kinetic energies are possible. As q increases in subsequent bands, the LHS “amplitude” approaches unity, and the allowed energy band region grows.

Another important way of viewing the band gap phenomenon is to solve the transcendental equation given in Eq. (36) for a variety of k 's and make a plot of E_k versus k . This is often called a *dispersion* curve. A crude sketch of such a curve for our δ -function lattice is shown below:



Since both the LHS and RHS of Eq. (36) involve sinusoidal functions the way of mapping q to k (or E_k to k) is somewhat ambiguous. We show the “extended zone” mapping which shows that E_k is monotonic in $|k|$.

The dashed parabola shows the dispersion relationship for a free electron which is:

$$E = \frac{\hbar^2 k^2}{2m} \quad \text{where} \quad mc^2 = 0.511 \times 10^6 \text{ eV} \quad (38)$$

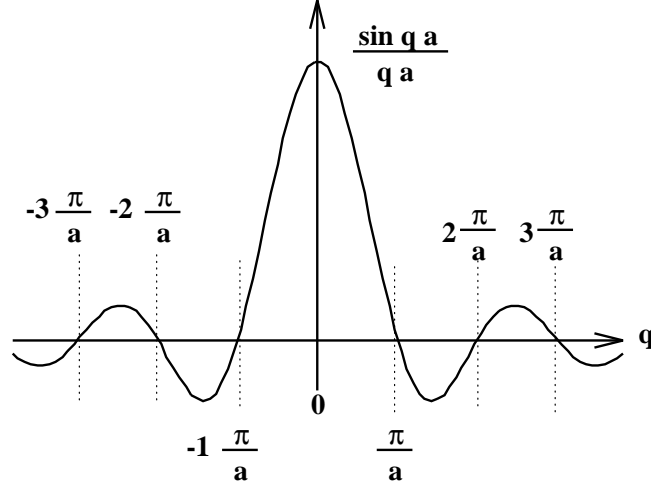
As you can see the dispersion relationship for the electron in the crystal is considerably different than that for a free particle. The most noticeable difference is the presence of band gap discontinuities which occur when k is an integral multiple of π/a . We further note the inverse effective mass, $m_*^{-1} = \hbar^{-2} d^2E/dk^2$, near $k = 0$ is significantly smaller than the inverse free particle mass. This means that the electron at the bottom of the lowest energy band ($k \approx 0$) in a Kronig-Penney crystal has much more inertia ($m_* > m$) than a free electron in vacuum.

We also note that the upper end (large $|k|$ end) of each band, touches the free electron parabola. It is easy to see why this is true by re-arranging the transcendental equation (36):

$$\frac{\sin(qa)}{qa} = \frac{\hbar^2}{mga} (\cos ka - \cos qa) \quad (39)$$

The left hand side of Eq. (39) goes to zero when q is a non-zero multiple of π/a

as illustrated below:



The vanishing of the left hand side of Eq. (39) implies that $k = q$. Since $E = q^2/(2m)$ we must also have $E = k^2/(2m)$ as well when k is a non-zero multiple of π/a .

We also note from the dispersion curve that $\partial E_k/\partial k = 0$ at each band boundary. This is also easy to see from Eq. (36). We can write

$$\frac{\partial E}{\partial k} = \frac{\partial}{\partial q} \left(\frac{\hbar^2 q^2}{2m} \right) \times \frac{dq}{dk} = \left(\frac{\hbar^2 q}{m} \right) \times \frac{dq}{dk} \quad (39b)$$

Eq. (36) is of the form $\text{LHS}(q) = \text{RHS}(k) = \cos ka$. Equating the change from the LHS due to a change in q with the change in the RHS due to a change in k we have:

$$\Delta q \frac{\partial \text{LHS}}{\partial q} = \Delta k \frac{\partial \cos ka}{\partial k} = -a \sin ka \times \Delta k = 0 \quad \text{if } k = n \frac{\pi}{a}$$

This means $\Delta q/\Delta k = 0 = dq/dk$ at each band gap and hence by Eq. (39b) we must have $\partial E/\partial k = 0$ as well.

We can use the fact that $\partial E/\partial k = 0$ at a band edge, and that E_k monotonically increases in k to argue that $\partial^2 E/\partial k^2 < 0$ just below the band edge and

$\partial^2 E/\partial k^2 > 0$ just above. For example, as one approaches $k \rightarrow n\pi/a$ from below, E_k increases up to its maximum value (where $\partial E/\partial k = 0$) at $k = n\pi/a$. This means E_k versus k is concave down when $k = n\pi/a - \epsilon$ and thus $\partial^2 E/\partial k^2 < 0$. The argument is reversed when $k = n\pi/a + \epsilon$. Now E_k increases monotonically above its minimum at the band edge and thus E_k is concave up.

We thus expect a dramatic change in m_* as one gets approaches each band gap. As you can see from the dispersion curve, the effective mass actually approaches infinity at the inflection point of $E(k)$. Beyond the inflection point, $d^2 E/dk^2$ is actually negative which means the electron accelerates in the direction **opposite** to the applied force! We see the mathematics of how this happens based on the Kronig-Penney model but is there a physical explanation? And is the result more general??

The concept of a negative effective mass for an electron in a crystal seems extremely non-intuitive. However there is a fairly simple and plausible explanation for this strange behavior near the top of the band gap which involves Bragg scattering. Recall from the chapter on **Wave Mechanics** that Bragg scattering occurs when wave scattering from successive crystal planes constructively interfere. In two dimensions, the Bragg scattering condition is:

$$2a \sin \theta = n \lambda \tag{40}$$

where θ is the angle that the electron or photon beam makes with the crystal plane. In a one dimensional crystal, there can only be backscattering, which means $\theta = 90^\circ$. The lowest order ($n = 1$) Bragg peak occurs when:

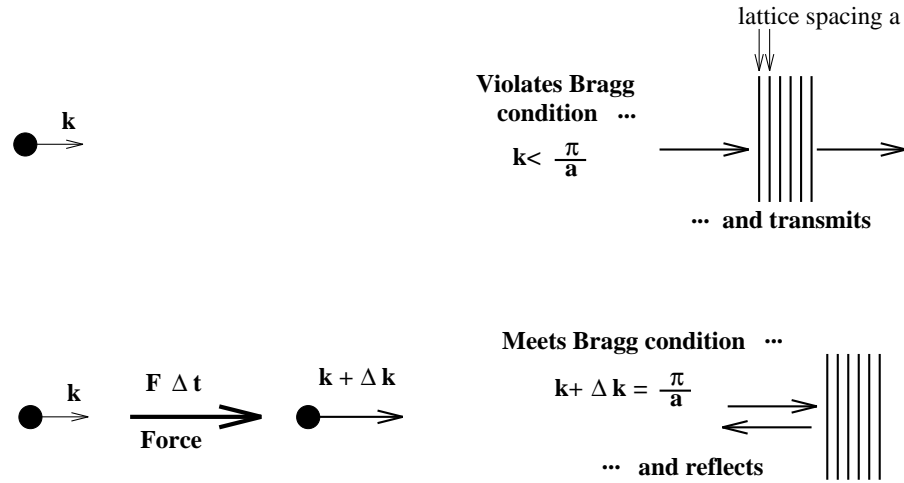
$$\lambda = \frac{h}{p} = \frac{h}{\hbar q} = \frac{2\pi}{q} = 2a \rightarrow q = \frac{\pi}{a} \tag{41}$$

Close to the top of the first energy band, $q \approx k$ and the condition for Bragg back scattering becomes

$$k = \frac{\pi}{a} \tag{42}$$

which you will notice occurs exactly at band gap boundary.

Let us imagine, that an electron is lying in a Bloch wave with k just slightly less than π/a . Recall that just below each band edge, $\partial E_k/\partial k = 0$ and the electron is at rest. We then turn on an electrical force for a period Δt . According to Eq. (23), the crystal momentum will be increased to: to $k' = k + \Delta k$ where $\Delta k = F\Delta t/\hbar$. This is very likely to result in a k' which matches the Bragg scattering condition $k' = \pi/a$ which will suddenly cause the electron waves to backscatter and accelerate against the direction of the applied force.



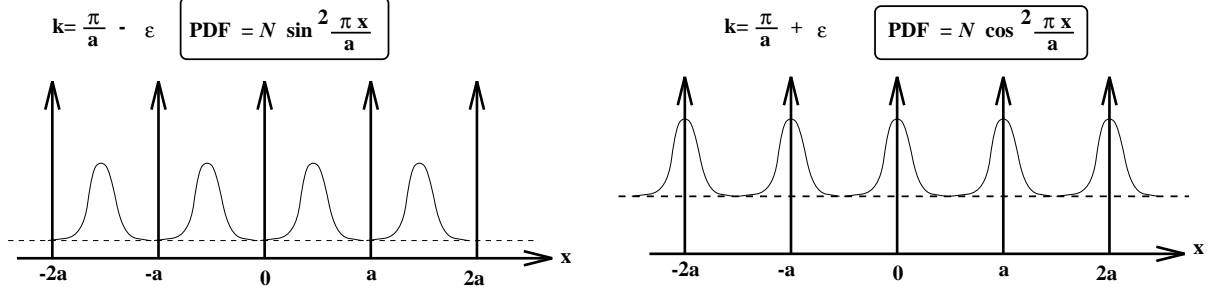
We see from the foregoing argument that the negative effective mass near a band boundary should be a general feature of all periodic potentials.

Why does the energy gap develop?

We can also describe the development of an energy gap as a consequence of Bragg scattering when k approaches a non-zero, integral multiple of π/a . When the Bragg scattering condition is approached, the electron wave is essentially completely reflected by the crystal lattice. The interference of an incident and equal amplitude reflected wave creates a standing wave pattern with the two forms:

$$\psi_{\pm} \propto e^{in\pi x/a} \pm e^{-in\pi x/a} \quad \text{where } n \in \{\pm 1, \pm 2, \pm 3 \dots\} \quad (43)$$

The below is a sketch of the PDF for the two possible standing wave patterns:



When one is just below the band gap $k < \pi/a$, one can use Equations (34) and (35) to show that the PDF is of the standing wave form $\mathcal{N} \sin^2(\pi x/a)$ form which tends to put the bulk of the electron probability away from the δ -function “perturbation” and diminishes $\langle \psi_- | \Delta V | \psi_- \rangle$. Just above the band gap $k > \pi/a$, the PDF is of the alternative standing wave form: $\mathcal{N} \cos^2(\pi x/a)$ form which tends to put the bulk of the electron probability at the δ -function “perturbations” and increases $\langle \psi_+ | \Delta V | \psi_+ \rangle$ relative to $\langle \psi_- | \Delta V | \psi_- \rangle$. Of course near the band edges, it is a very bad approximation to use the free particle wave function ($\psi_k = \mathcal{N} \exp(ikx)$) as the zero order wave function since when one comes close to satisfying the Bragg condition $k = \pm \text{integer } \pi/a$, the amplitude for reflected waves is very strong no matter what the lattice strength. Perturbation theory is only accurate when the change in the wave function is small.

The standing wave form that occurs just below and just above each band gap also explains why $\partial E_k / \partial k = 0$ at these k values. Eq. (43) tells us that just below and just above the band gap, the electron is in a state which is an equal mix of $p = \pm \hbar n \pi / a$ and thus $\langle p \rangle = 0$. Since $d\langle x \rangle / dt = \langle p \rangle / m$ we thus have $\bar{v} = 0$ for these standing wave solutions. Hence $\hbar^{-1} \partial E / \partial k = \bar{v} = 0$ whenever $k = n\pi/a \pm \epsilon$. We thus have an intuitive explanation of the main features of E_k for the Kronig-Penney model or any more complicated periodic potential near a band edge based on Bragg scattering.

- (1) $\partial E_k / \partial k = 0$ since ψ_k becomes a standing wave with no velocity since the electron Bragg reflects from either side of the lattice.
- (2) The standing wave pattern shifts from populating primarily at the lattice

sites to primarily away from the lattice sites as one passes through each band gap at $n\pi/a$, $n \in \pm 1, \pm 2, \dots$. This means one will have a discontinuity in energy on either side of the gap. In the case of the Kronig-Penney model with $g > 0$ the high energy is at $k = n\pi/a + \epsilon$ and the low energy is at $k = n\pi/a - \epsilon$.

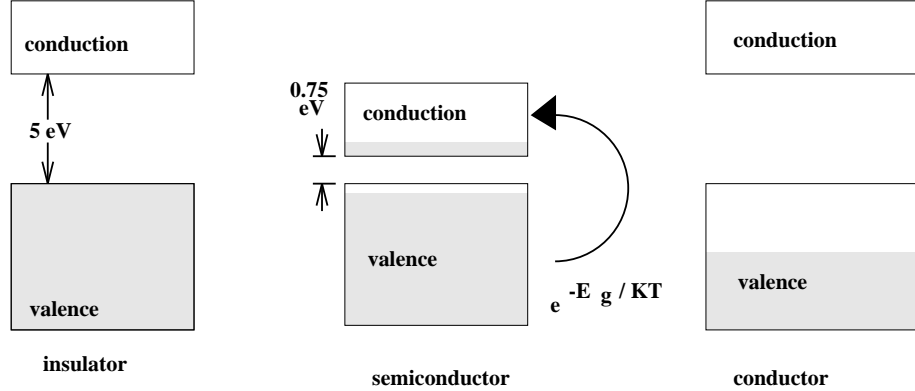
- (3) For the case of a δ -function lattice with $g > 0$, the $k = n\pi/a - \epsilon$ standing wave will have no PDF at the δ function and thus the electron in the lattice will have the same energy as a free electron or $E_k = (\hbar k)^2/(2m)$.

Insulators , Conductors , and Semiconductors

Band gap theory, combined with the Pauli exclusion principle, and some thermal physics, explains many of the basic conduction properties of crystals. Electrical conductivity involves those loosely bound, optically active electrons which are in the outer most incomplete shell of an atom. These are often referred to as the *valence* electrons. Our discussion of band structure and effective masses applies to these electrons. We also note that the number of electron orbitals (or different possible k values) within a band is $2N$ where N is the number of crystalline atoms in the sample.⁷ When the sample is cooled to absolute zero, the electrons will be stacked into their lowest available quantum state. The Pauli principle allows only one electron per orbital which are labeled by k and the electron spin state.

The below sketch is a somewhat oversimplified cartoon describing the difference between insulators, semiconductors, and insulators.

⁷ Ignoring spin we showed in a ring lattice consisting of N wells there were N allowed k values within each band. The factor of 2 is to allow for the \uparrow and \downarrow degree of spin freedom.



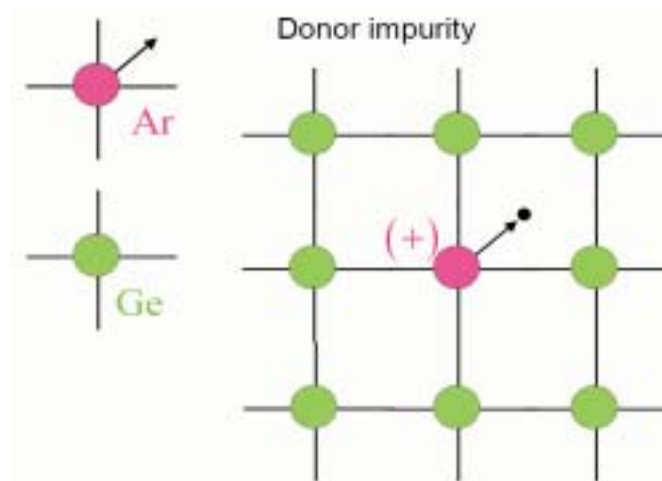
We begin with the insulator which is characterized by very high electrical resistivities $\rho \sim 10^{20} \Omega \text{ cm}$. An insulator will have just enough valence electrons to completely fill the $2N$ orbitals in a single band which is labeled as the valence band. An example would be a diamond crystal which consists of a periodic arrangement of 2 atoms per repeating (called *primitive*) cell with 4 valence electrons per atom. Hence the number of valence electrons per cell is just 2 times the number of wells which completely saturates the valence band with electrons. There is a rather large, 5.5 eV , energy gap between the valence and conduction band which means that it is very difficult to thermally excite electrons over this gap at room temperatures ($KT \sim 0.025 \text{ eV}$) since the Boltzman factor is only $\exp(-E_g/KT) = 3 \times 10^{-96}$. When an electrical field is applied to a diamond crystal it will create a force on the valence electrons. The electron will want to absorb the work created by this field and change its quantum state by changing its crystal momentum, k , according to Eq. (19). In the case of an insulator such as diamond, however, all of the available k levels within the band are filled. There is no k “slot” for the electron to absorb energy from the field and participate in a current flow. The only way around this would be to: (1) heat the crystal up significantly, (2) use electromagnetic fields associated with ultraviolet light so that electrons in the valence band can absorb enough energy to move to the conduction band, or (3) subject the electrons to the enormous fields required to provide the work necessary to place the electron in the conduction band and thus break down the insulator.

Germanium is in the same column of the periodic table as Carbon and is therefore also has 2 valence electrons per repeating cell and will completely fill up the valence band. However the Germanium band gap is only 0.75 eV and the population of thermally excited electrons at room temperatures is much larger. The Boltzman factor is for Germanium would be $\exp(-E_g/KT) = 1 \times 10^{-13}$ which is still a small number but much larger the factor for diamond. Germanium is known as a semiconductor. Semiconductors act as electrical insulators at absolute zero since there are no thermally activated vacancies in the valence band. The conductivity of a semiconductor will increase very rapidly with increasing temperature as the Boltzman factor “kicks in”. Typical semiconductors have room temperature resistivities which range from $\rho = 10^{-3} \rightarrow 10^9 \Omega cm$.

By way of contrast, an alkali metal such as sodium, will have only one valence electron per cell and will be have an exactly half filled valence band. There are thus plenty of empty k slots for an electron to occupy to enable it to absorb work from an applied field. As a result the alkali metals are classified as conductors. Typical conductor resistivities are $\rho \sim 10^{-6} \Omega cm$. In contrast to the case of a semiconductor, the electrical resistivity of a conductor *increases* with increasing temperature. In a conducting metal, the number of electrons which are free to travel is nearly unaffected by the the temperature. However once the electrons are set into motion by an external field, they can be “stopped” by transferring energy to a lattice vibration , called a “phonon”. Near room temperatures, the number of such phonons grows proportional to temperature resulting in a resistivity which is nearly proportional to temperature. The same dynamics will also affect semiconductors but is significantly less important than the exponential increase in the the number of valence and conduction band carriers due to the Boltzman factor. The difference between the temperature dependence of the resistivity of a conductor and semiconductor was known as early as 1840 and discussed in a publication of Michael Faraday.

Doped Semiconductors

Much of the technological importance of solid state physics rests in the ability to make designer semiconductors by incorporating random impurities within the semiconducting lattice. As we will argue shortly, the incorporation of such impurities can drastically affect the conductivity of the crystalline sample. We say that such a doped sample provides *extrinsic* conductivity where the pure sample exhibits intrinsic conductivity.



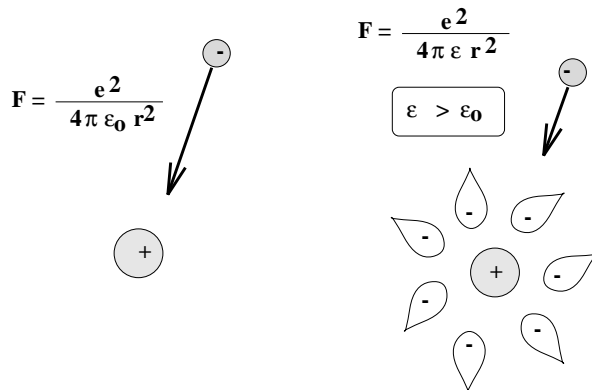
As an example, consider adding Arsenic impurities to a Germanium crystal. Arsenic has 5 valence electrons and lies adjacent in the periodic table to Germanium with 4 valence electrons. Four of the Arsenic valence electrons can participate in the covalent bonding and serve as sort of a “plug in replacement” for some of Germanium atoms in the crystalline lattice. The fifth Arsenic valence electron acts like a very loosely bound hydrogen-like atom which feels some attraction to the Arsenic ion because of its one net, unshielded charge. We often call each embedded Arsenic atom a *donor* since it donated an extra electron. What will be the energy level of the donated electron? In analogy with atomic energy levels, we anticipate that the donated electron energy levels will lie below the energy levels of the continuum. For the case of an isolated atom, the continuum is the lowest possible energy which an electron can have to completely escape the atom and travel to infinity with a kinetic energy just above zero.

For a crystal, the lowest energy for an electron to travel away from the Arsenic donor atom would be the lowest energy in the conduction band. Very naively we might expect that such that the electron *donated* by the Arsenic might lie with an energy relative to the conduction band given by the Bohr formula:

$$\Delta E = -\frac{e^4 m}{2(4\pi\epsilon_0)^2 \hbar^2} \frac{1}{n^2} = -\frac{13.6}{n^2} \quad (44)$$

However this is clearly too naive since we neglected the influence of the underlying periodic crystal which can drastically alter the semiclassical response of an electron to an external force by providing an effective mass which is significantly different than its real mass. For the case of an electron in the conduction band of a Germanium crystal the effective electron mass is about 1/5 of the actual electron mass ($m_* \approx 0.2 m$).

Another important effect to consider when considering an atom in a crystal as opposed to an atom in vacuum is the effect of the dielectric constant. In a dielectric, the Coulomb force will typically be lower than the force in vacuum due to the polarized distortion of electron clouds of the atoms comprising the material.



The electron clouds will tend to be pulled towards a positive charge which will create an internal (dipole) field which bucks the electrical field of the positive charge. As a result the ordinary Coulomb field or potential between the Arsenic ion and the donor electron will be reduced by a factor ϵ_0/ϵ .

One normally thinks of a dielectric constant, ϵ , as applying to an electrical insulator such as plastic or ceramic. In the case of a metal, the electrons have enormous mobility and can effectively screen out the electrical force between two charges. As you recall there is rarely an appreciable electrical field within a good conductor. Within a good conductor, $\epsilon \rightarrow \infty$. A semiconductor has an electron mobility somewhere in between an insulator and conductor and one might anticipate a fairly large ϵ . For a Germanium crystal $\epsilon \approx 16 \epsilon_o$. To the extent that the donor electron orbit is so large relative to the lattice spacing that we can safely average over many lattice sites, the lattice provides an “alternative vacuum” where $\epsilon_o \rightarrow \epsilon$ and $m \rightarrow m_*$ and

$$\Delta E = -\frac{e^4 m_*}{2(4\pi\epsilon)^2 \hbar^2} \frac{1}{n^2} = -\left(\frac{m_*}{m}\right) \left(\frac{\epsilon_o}{\epsilon}\right)^2 \frac{13.6}{n^2} \quad (45)$$

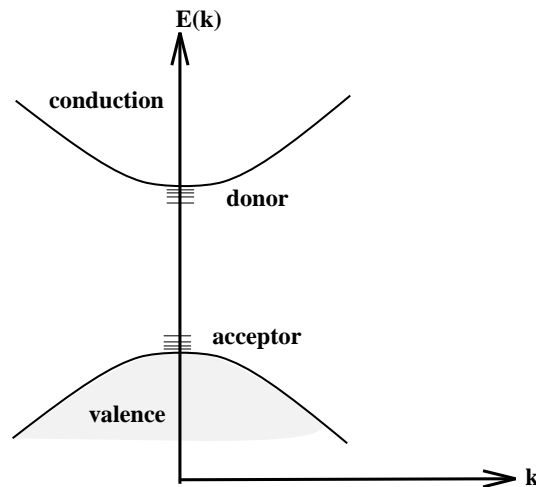
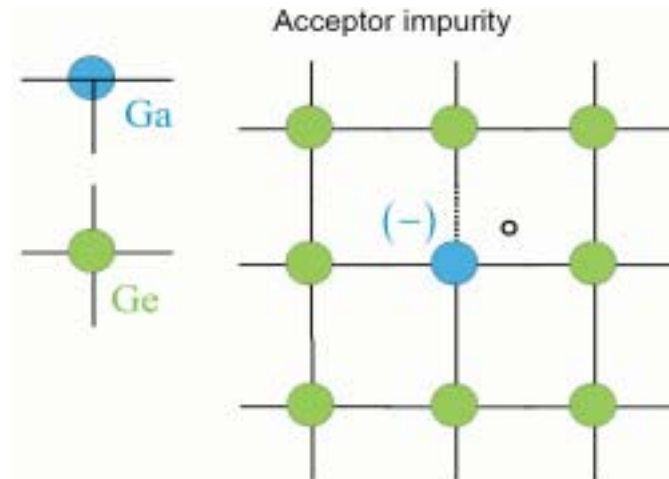
Because the effective mass is 1/5 the free mass, and the dielectric constant is $16 \times \epsilon_o$, we expect the lowest energy level of a donor electron from Arsenic to lie only 0.011 eV below its continuum or 0.011 eV below the conduction band. This is in good agreement with the actual value for Arsenic donors in Germanium of 0.013 eV. For the case of Arsenic donors in Silicon (which has a larger energy gap), the donors lie 0.05 eV below the conduction band.

The size of the electron cloud, given roughly by the Bohr radius, will also be affected by m_* and ϵ :

$$a_o = \frac{(4\pi\epsilon) \hbar^2}{e^2 m_*} = \left(\frac{\epsilon}{\epsilon_o}\right) \left(\frac{m}{m_*}\right) 0.0529 nm \quad (46)$$

The effective mass and dielectric constant of Germanium will both act to increase the Bohr radius by a factor of ~ 80 . This makes the electron cloud radius 4 nm or so which is over 25 times the lattice spacing. Hence it is not surprising that one can use average crystalline properties such as a dielectric constant and effective electron mass when computing the energy level for a donor atom.

One can also move the other way in the periodic table and dope a tetravalent Germanium crystal with Gallium which normally has three valence electrons. The effect of this will be to *accept* an electron from the valence band to form one of the required four covalent bonds to required to fix the Gallium in the lattice. This leaves missing electron or hole in the valence band and a set of very weakly bound hydrogen-like levels which exist above the valence band since the Gallium with the extra electron serves to repel valence electrons. Such impurities which create holes are called *acceptors*.

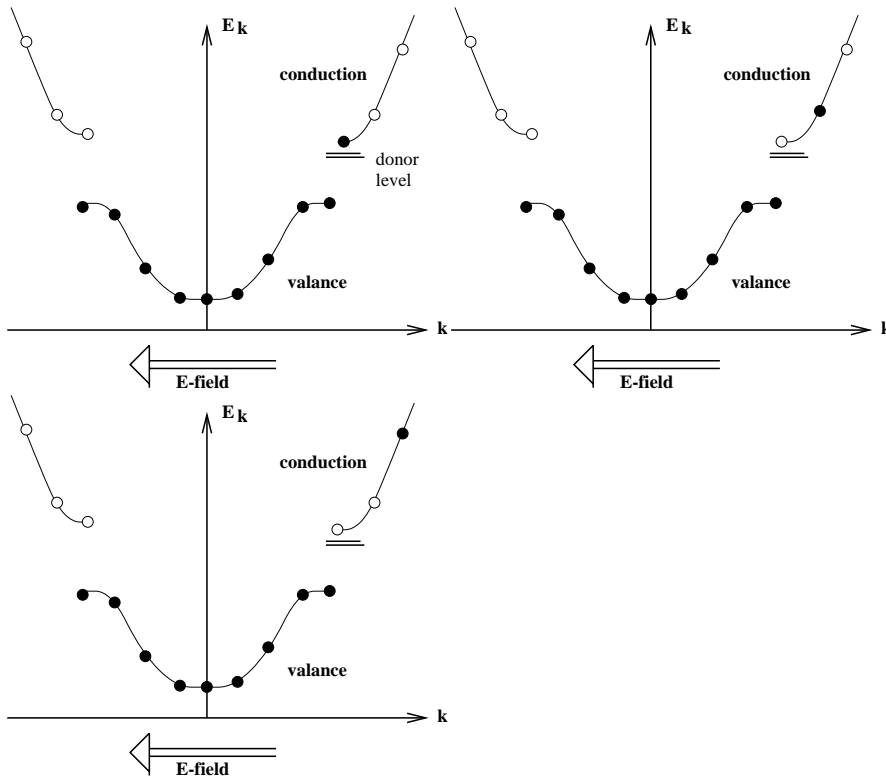


Because of the effective mass and dielectric effects, the donor and acceptor energy levels are within an energy which is comparable to $KT \approx 0.025 \text{ eV}$ from

the closest band. It is much easier to create conduction carriers by thermally exciting an electron from a donor level to the conduction band rather than thermally exciting all the way from the valence band.

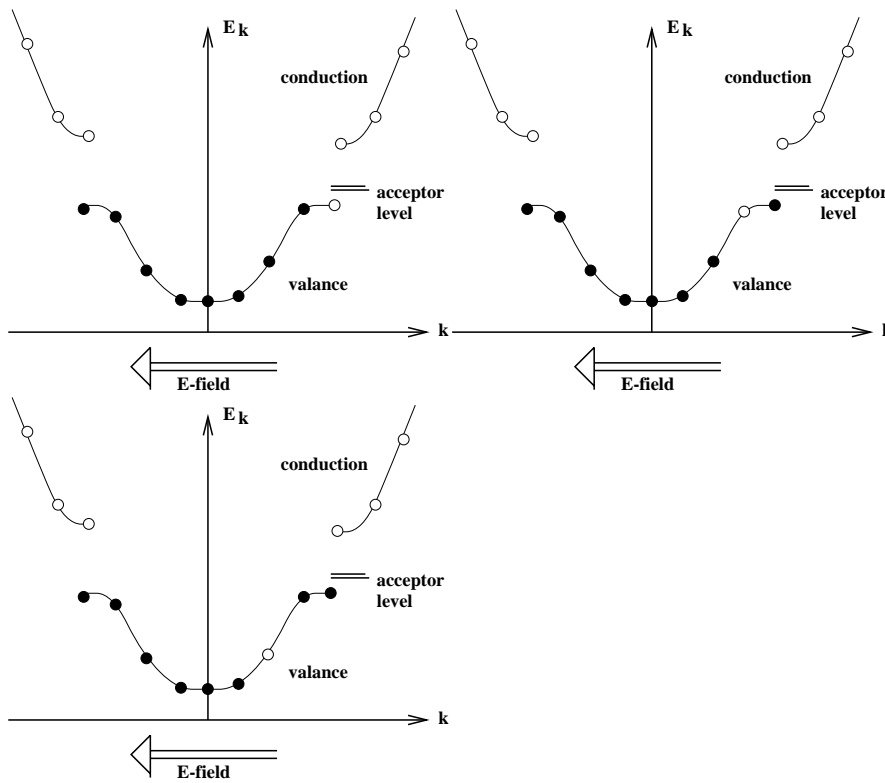
Under the influence of an external electric field, electrons will be able to move into the k slot of the hole.

Illustration of electron flow with donor impurities



Conversely an electron can be easily thermally excited from the valence band to an acceptor level and leave a hole in the valence band. The hole or k vacancy will appear to move opposite to the flow of the electrons and “flow” through the crystal as if it had a positive charge.

Illustration of hole flow with acceptor impurities



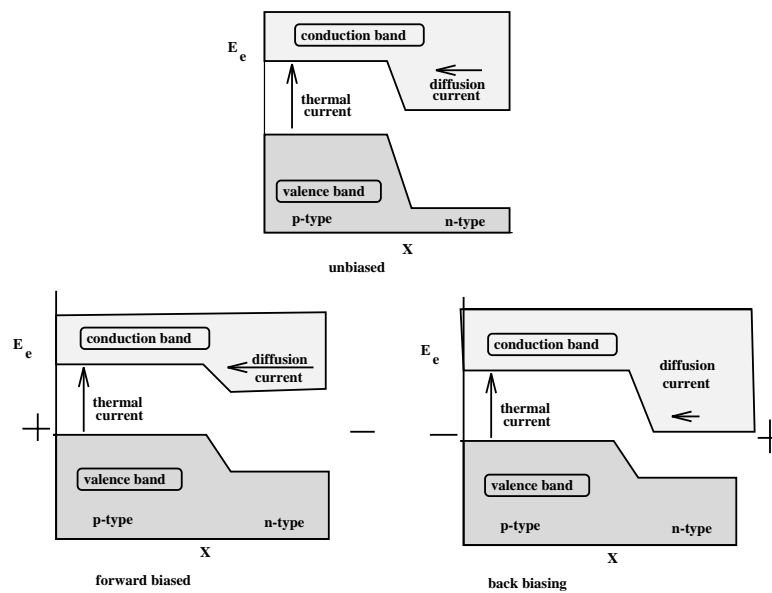
Even for modest impurity concentrations, the bulk of the electrical conductivity is due to extrinsic conduction created by thermal excitation involving the donor or acceptor levels.

Rectifiers

The ability to inject donor or acceptor impurities in a crystal allows one to create two distinctly different extrinsic semiconductors called n-type and p-type. The n-type semiconductor is doped with donor atoms. The electrical carriers are electrons in the conduction band which were thermally excited from the donor atoms. The p-type semiconductor doped with acceptor impurities. The electrical

carriers are holes which are left as the electrons are thermally excited from the valence band to the acceptor levels. Perhaps the simplest solid state device is a p-n junction where one engineers an abrupt change of impurity type within a relatively narrow junction region.

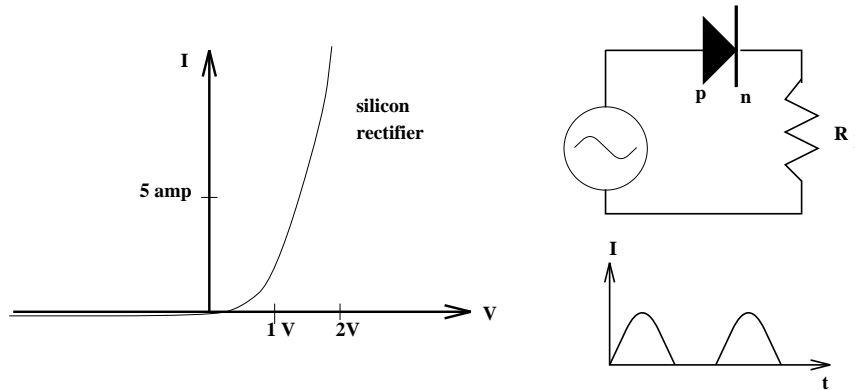
The diffusion of electrons from the n-type to p-type regions creates an electrostatic potential which warps the band structure. In equilibrium, with no external electric field, the diffusion current of electrons from the n to p-regions is balanced by an equal and opposite thermal current of electrons excited from the valence to conduction band.



As one forward biases the p-n junction, the band warping is reduced allowing n-region electrons to diffuse more readily to the p-region resulting in a much stronger diffusion current. The imbalance between the enhanced diffusion current over unchanged, thermal current allows one to have a net current through the diode. As one backward biases the p-n junction, the band warping is increased which significantly cuts down the diffusion current by significantly increasing potential gap for an electron to thermally diffuse into the n-type side of the junction.

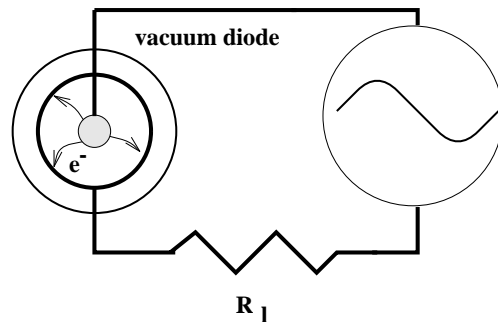
The result of this basic asymmetry in the carrier type is a much stronger

current flow in one direction than the other which allows the p-n junction to act as a rectifier which can be used to convert AC currents into DC currents as illustrated below:



A very crude version of a silicon junction rectifier was the cat whisker rectifier consisting of a metal silicon contact which was used to demodulate radiowaves as early as 1906.

The vacuum tube version of a diode rectifier relies on the thermionic emission of electrons from a hot filament. A current can flow only when the collection electrode is relatively positive with respect to the filament. Rectification in either the p-n junction or vacuum diode requires a position dependence in the carrier charge employed throughout the diode.



Important Points

1. We can model a one dimensional “molecule” by a semiperiodic array of rectangular quantum wells. In the tightly bound limit, there is little penetration of the wave function outside of each well and energy level for states with various allowed symmetries become nearly degenerate at a value given by energy levels of an individual well. This results in the formation of energy bands or energies which cluster very closely about the energy levels of a single well.
2. As the potential barrier between the “molecular” quantum well is lowered the size of the molecular band grows and the gap between bands shrinks.
3. We next considered a lattice of one dimensional wells arranged on a ring. This arrangement has true translational symmetry in the sense $V(x + a) = V(x)$ where a is the lattice spacing.
4. The fact that the potential is periodic, means that the PDF of an electron is periodic in the sense that $\text{PDF}(x + a) = \text{PDF}(x)$. Interestingly enough, the wave function itself is not periodic: $\psi(x + a) \neq \psi(x)$.
5. We can think of the lattice translation as an operator which commutes with the periodic Hamiltonian. This means that the wavefunctions can be selected to be simultaneous eigenfunctions of energy and translation. According to Bloch’s theorem, the translational eigenvalue equation is of the form:

$$T\psi_k(x) = \psi_k(x + a) = e^{ika} \psi_k(x)$$

and the wave function is of the form

$$\psi_k(x) = e^{ikx} u_k(x) \quad \text{where } u_k(x) \text{ is periodic : } u_k(x + a) = u_k(x)$$

6. The quantity k , which is related to the translational eigenvalue, is called the crystal momentum. This is the quantum number which is used to label the

eigenfunction in much the same way as principle quantum numbers labels the energy states of a hydrogen atom. In the case of a one dimensional, ring lattice with N wells, k can only take on values which are integral multiples of $2\pi/(Na)$. Quantum mechanics can be used to compute $E(k)$ or the energy of an electron orbital as a function of its crystal momentum k .

7. The crystal momentum, k , is not the true momentum of the electron since the true momentum varies with x along with the kinetic energy due to the changing lattice potential. However, the crystal momentum is a very useful concept in understanding the motion of the electron within the crystal. In particular the average velocity of an electron with crystal momentum k is $\bar{v} = \hbar^{-1} dE/dk$ and under the influence of an external force, a semiclassical argument shows that the crystal momentum varies in time just like regular mechanical momentum: $\hbar dk/dt = F$.
8. The electron within a crystal responds to an external force as if it had an effective mass (m_*) which can be computed from $E(k)$ according to:

$$\frac{1}{m_*} = \frac{1}{\hbar^2} \frac{d^2E}{dk^2}$$

The effective mass can be smaller than the free electron mass, larger than the free electron mass, and even negative.

9. We worked out $E(k)$ for the case of an infinite, one dimensional, periodic array of δ -functions as a special case of the Kronig-Penney model. The result was a transcendental equation of the form:

$$\cos(qa) + \frac{mg}{\hbar^2 q} \sin(qa) = \cos(ka) \quad \text{where } q = \sqrt{2mE}/\hbar \quad (36)$$

This transcendental equation has energy values where the LHS exceeds the maximum ± 1 range of $\cos(ka)$. These energy values cannot be mapped to a crystal momentum k and thus cannot be allowed energy levels. This specific model gives energy bands which shrink and energy gaps which grow in size as the strength of the δ -function is increased.

10. Discontinuity in $E(k)$ corresponding to the band gaps occur when k is a non-zero, integral multiple of $\pm\pi/a$. Before k reaches a band gap edge, the electron acquires a negative effective mass. Both the negative mass and band gap itself is a consequence of the electron approaching the condition for Bragg backward scattering from the crystalline planes which is $k = q = \pm n\pi/a$. Hence both effects are true in general and do not just apply to the Kronig-Penney model.
- (a) The application of a force when the electron has a crystal momentum just beneath $|k| = n \pi/a$ will cause it to match the condition, backscatter, and de-accelerate.
 - (b) When the Bragg condition is just met at the band edge, the Bloch waves degenerate into sine and cosine standing waves. The sine wave has a PDF which avoids the core of lattice potential; while the cosine wave has a PDF which concentrates near the core of the lattice. Hence one standing wave is significantly split away from the other to create an energy gap.
 - (c) The cosine and sine standing wave functions that occur just below or just above each band gap are states with equal parts $p = \pm\hbar n\pi/a$ and thus have $\bar{v} = \langle p \rangle / m = 0$. This means $\partial E / \partial k = 0$ whenever $k = n\pi/a \pm \epsilon$.
11. The number of possible k and spin orbitals within an energy band is just $2N$ where N is the number of electrons per lattice site or primitive cell. Often an element will have an even number of valence electrons per cell which will completely fill all available energy levels within a band when the crystal is at absolute zero temperature. In such a case, electrons cannot absorb energy from an external field and change their crystal momentum k without crowding into an already occupied orbital. This violates the Pauli Exclusion principle and no electrical current can flow under normal circumstances. When one has an odd number of valence electrons, such as

in an alkali metal, only half of the $2N$ available k slots in an energy band will be filled. Plenty of k slots will be available to allow the electrons to conduct a current in response to an applied crystal.

12. Semiconducting crystals have all their energy levels filled at very low temperatures and act as insulators. But semiconductors have relatively narrow band gaps of roughly 1 eV which allows for modest conductivity at room temperatures owing to electrons in the valence band being thermally excited to the conduction band with a relative population described by the Boltzmann factor $\exp(-\Delta E/(KT))$. The resistivity of semiconductor decreases very quickly with increasing temperature. The resistivity of a conventional conducting metal increases proportionately with temperature.
13. One can significantly increase the number of carriers through the insertion of donor or acceptor impurities which replace a small fraction of atoms in the underlying crystal. The valence electrons associated with the impurity lie just below the conduction band (donor) or just above the valence band (acceptor). The energy spectrum relative to the band edge is described by the Bohr formula but is very much reduced relative to $-13.6/n^2$ owing to effective mass ($m_* < m$) and dielectric ($\epsilon > \epsilon_o$) effects.
14. Since the impurity levels lie so close to the energy bands in a doped semiconductor, the extrinsic conductivity created by thermally activating either the bands or impurity levels easily dominates the intrinsic conductivity of a pure semiconductor due to thermal exciting electrons from valence to the conduction band. In an n-type semiconductor, which is doped with donor atoms, electrons are thermally (room temperature) excited from the donor levels and enter the conduction band. In a p-type semiconductor, which is doped with acceptor atoms, electrons are thermally excited out of the valence band and into the acceptor levels. In an n-type semiconductor, the majority current carriers are electrons in the conduction band. In a p-type semiconductor, the majority current carriers are the “holes” which

are left behind when thermally excited electrons leave the valence band for the close by acceptor level.

15. The ability to control which type of carrier exists in which area of a crystal allows one to make semiconductor junction devices which perform many of the same functions as older vacuum tubes. We discussed the operation of a p-n junction rectifier which functions as a diode vacuum tube. Other devices, not discussed, include a p-n-p junction which functions as a triode vacuum tube used in amplifiers, oscillators , and electronic switches.