Chapter3: Quantum Control and Solid State Physics

1 Free Electrons

The free electron wavefunction is $\psi = \exp(ikx)$. The momentum is $\frac{\hbar}{i}\frac{\partial}{\partial x}$. This gives the kinetic energy $\epsilon = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m}$, which for $\epsilon = \hbar\omega$ gives,

$$\omega(k) = \frac{\hbar k^2}{2m}. \tag{1}$$

The dispersion is a parabola as shown below in figure 1 A.

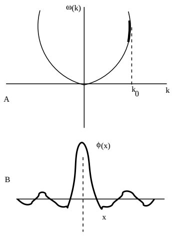

Figure 1: Figure A shows the dispersion $\omega(k)$ vs k for a free electron. Figure B shows a wavepacket centered at k_0.

Now, consider a wavepacket centered at k_0, shown the figure 1 A, B. The packet takes the form

$$\phi(x) = \frac{1}{\sqrt{N}}\sum_j \exp(ik_j x), \quad \phi(x,t) = \frac{1}{\sqrt{N}}\sum_j \exp(-i\omega(k_j)t)\exp(ik_j x), \tag{2}$$

where $\omega(k_j) = \omega(k_0) + \omega'(k_0)\Delta k_j$ where $\Delta k_j = k_j - k_0$. Denote $v_g = \omega'(k_0) = \frac{\hbar k_0}{m}$, as the group velocity. Then

$$\phi(x,t) = \frac{1}{\sqrt{N}}\exp(i(k_0 x - \omega(k_0)t))\sum_j \exp(i\Delta k_j(x - v_g t)). \tag{3}$$

The function $f(x) = \frac{1}{\sqrt{N}}\sum_j \exp(i\Delta k_j x) = \frac{2}{\sqrt{N}}\sum_j \cos(\Delta k_j x)$, is centered at origin with width $\propto (\Delta k)^{-1}$ as shown in figure 1 B. Then

$$|\phi(x,t)| = |f(x - v_g t)|, \tag{4}$$

the free electron wavepacket moves with a group velocity v_g.

Now lets apply an electrical field E in the x direction at $t = 0$. Then the Schröedinger equation is

$$i\hbar\frac{\partial \psi}{\partial t} = \frac{1}{2m}(-i\hbar\frac{\partial}{\partial x})^2 + eEx\ \psi. \tag{5}$$

The equation is satisfied by time varying wavevectors $\exp(ik(t)x)$, where $k(t) = k - \frac{eEt}{\hbar}$, with energy (dispersion) $\omega(k(t)) = \frac{\hbar(k(t))^2}{2m} = \frac{\hbar(k-\frac{eEt}{\hbar})^2}{2m}$, so that the wavefunction becomes

$$\exp(-i\int_0^t \omega(k(\tau))\ d\tau)\exp(ik(t)x).$$

The initial wavepacket $\phi(x)$ evolves to $\phi(x,t)$, where,

$$\phi(x) = \frac{1}{\sqrt{N}}\sum_j \exp(ik_j x), \quad \phi(x,t) = \frac{1}{\sqrt{N}}\sum_j \exp(-i\int_0^t \omega(k_j(\tau))d\tau)\exp(ik_j(t)x). \tag{6}$$

The group velocity

$$v_g(t) = \frac{\hbar k(t)}{m} = \frac{\hbar(k - \frac{eEt}{\hbar})}{m}; \quad \frac{dv_g(t)}{dt} = -\frac{eE}{m}. \tag{7}$$

The electron wavepacket simply accelerates the way we know from classical mechanics. Being more pedagogical, we have

$$\begin{aligned} \phi(x,t) &= \frac{1}{\sqrt{N}}\sum_j \exp(-i\int_0^t \omega(k_j(t))\ )\exp(ik_j(t)x) \\ &= \frac{1}{\sqrt{N}}\exp(-i\int_0^t \omega(k_0(t))\ )\exp(ik_0(t)x)\sum_j \exp(i\Delta k_j(x - \int_0^t v_g(\sigma)d\sigma)). \end{aligned} \tag{8}$$

The wavepacket evolves with instantaneous velocity $v_g(t)$.

The above method can be generalized to arbitrary potential. Consider the Schröedinger equation

$$i\hbar\frac{\partial \psi}{\partial t} = (-\frac{\hbar^2}{2m}\frac{\partial}{\partial x^2} - eV(x))\psi. \tag{9}$$

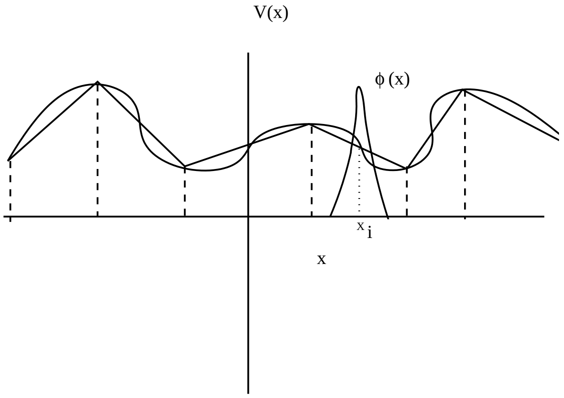

Figure 2: Figure shows linear approximation of potential $V(x)$. The wavepacket $\phi(x)$ is confined to a cell.

We approximate the potential V by piecewise linear potential such that $V(x) = V(x_i) + V'(x_i)\delta x$, where $\delta x = x - x_i$, as shown in figure 2. We call these regions of linearized potential, cells. We can rewrite the potential in a cell as $V(x) = U(x_i) + V'(x_i)x$

We assume that the wavepacket has large k_0 such that $\Delta k \sim \sqrt{k_0}$ is large and therefore for the wavepacket, $\Delta x \sim (\Delta k)^{-1}$ is small so that it fits well within one cell. Then in this cell, the Schröedinger equation takes the form

$$i\hbar\frac{\partial\psi}{\partial t} = (-\frac{\hbar^2}{2m}\frac{\partial}{\partial x^2} - eV'(x_i)x - eU(x_i))\psi. \tag{10}$$

Since the wavepacket is confined to a cell, it evolution would be same if the potential we have was not only true in the cell but globally true. This is because the wavepacket doesn't know what the potential is outside the cell, its confined to the cell. Then lets solve the Schröedinger equation with this potential assumed globally true and see how wavepacket evolves.

Then as before for the Schröedinger equation is solved by wavevector $\psi = \exp(ik(t)x)$. Let $x(t)$ denote coordinates of center of wavepacket, then

$$k(t) = k + \frac{e\int_0^t V'(x(\tau))d\tau}{\hbar}, \quad \omega(k(t)) = \frac{\hbar(k + \frac{e\int_0^t V'(x(\tau))d\tau}{\hbar})^2}{2m} - \frac{e\int_0^t U(x(\tau))d\tau}{\hbar}. \tag{11}$$

The group velocity

$$v_g(t) = \frac{\hbar(k + \frac{e\int_0^t V'(x(\tau))d\tau}{\hbar})}{m}; \quad \frac{dv_g(t)}{dt} = \frac{eV'(x(t))}{m}. \tag{12}$$

This is classical mechanics. Therefore at high energies where k_0 is large and wavepacket is well confined, i.e., over the packet width, the second order change of potential is small,

$V''(x)\Delta x \ll V'(x)$. A linearized potential is a good approximation and evolution in quantum mechanics mimics classical mechanics.

2 Electrons in Periodic Potential

We now turn to electrons in a periodic potential created by periodic arrangement of ions in crystal/solid. Once again to ease exposition, we first analyze everything in in one dimension.

Consider a periodic potential with period a.

$$U(x) = \sum_{l=1}^{n} V(x - a_l) = \sum V(x - la).$$

where

$$\begin{aligned} V(x) &= 2V_0 \cos^2(\frac{\pi x}{a}), \quad \frac{-a}{2} \leq x \leq \frac{a}{2} && (13) \\ &= 0 \quad |x| \geq \frac{a}{2}. && (14) \end{aligned}$$

The potential is shown below in Fig. 3.

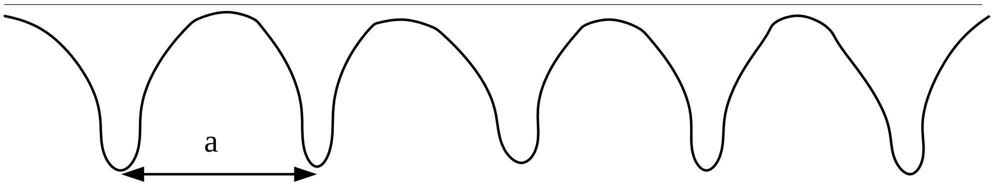

Figure 3: Figure depicts the periodic potential in Eq. (13).

How does the wavepacket evolve in this potential. This potential is rapidly varying on length sacels of a. To treat the particle classically, the packet width should be confined to $\Delta x \sim a$. This gives $\Delta k = a^{-1}$. This means that packet should have $k_0 \gg a^{-1}$. However if we think of lattice of length $L = na$ lattice points, we have wavevectors $k_j = \frac{2\pi j}{L} = \frac{2\pi j}{na}$, for $j = 0, \cdot, n$. In each lattice site we have one electron, so if we start filling our k_j, we fill all k_j between $-\left[\frac{\pi}{a}, \frac{\pi}{a}\right]$. We have filled energy levels according to minimum energy. Infact we can fill two electrons in every k_j, so that we fill all k_j between $-\left[\frac{\pi}{2a}, \frac{\pi}{2a}\right]$. Then our $k_0 \in -\left[\frac{\pi}{2a}, \frac{\pi}{2a}\right]$. Then its not true $k_0 \gg a^{-1}$. Therefore when we treat electrons in lattice potential of ions the potential varies quite rapidly and it is not correct to treat the problem classically. We have to treat the problem quantum mechanically. Let see how we do this.

Let a be the lattice constant such that periodic potential $V(x) = V(x + a)$. We can Fourier decompose

$$V(x) = \sum_k V_k \exp(ikx). \quad V_{-k} = V_k^*.$$

For example when $V(x)$ is as in Eq. (13), we have

$$V(x) = V_0(1 + \exp(i\frac{2\pi}{a}) + \exp(-i\frac{2\pi}{a}))$$

what does this potential do. It couples wavevectors $k \pm \frac{2\pi m}{a}$. Then suppose we have a wavefunction of the form

$$\phi(x) = \sum_m b_m \exp(i(k + \frac{2\pi m}{a})x)$$

then we can write the Schröedinger equaion as

$$\frac{d}{dt}\begin{bmatrix} \vdots \\ b_{-1} \\ b_0 \\ b_1 \\ \vdots \end{bmatrix} = -iH \begin{bmatrix} \vdots \\ b_{-1} \\ b_0 \\ b_1 \\ \vdots \end{bmatrix},$$

where,

$$H = \begin{bmatrix} \ddots & \dots & \dots & \dots & \dots \\ 0 & \frac{\hbar^2(k+\frac{2\pi}{a})^2}{2m} + V_0 & V_0 & \dots & 0 \\ 0 & V_0 & \frac{\hbar^2 k^2}{2m} + V_0 & V_0 & 0 \\ 0 & 0 & V_0 & \frac{\hbar^2(k-\frac{2\pi}{a})^2}{2m} + V_0 & V_0 \\ 0 & \dots & \dots & \dots & \ddots \end{bmatrix}. \tag{15}$$

The V_0 on the diagonal is just a phase factor, we can remove it to get

$$H = \begin{bmatrix} \ddots & \dots & \dots & \dots & \dots \\ 0 & \frac{\hbar^2(k+\frac{2\pi}{a})^2}{2m} & V_0 & \dots & 0 \\ 0 & V_0 & \frac{\hbar^2 k^2}{2m} & V_0 & 0 \\ 0 & 0 & V_0 & \frac{\hbar^2(k-\frac{2\pi}{a})^2}{2m} & V_0 \\ 0 & \dots & \dots & \dots & \ddots \end{bmatrix}. \tag{16}$$

The eigenfunctions of the Bloch electrons take the general form

$$\psi_k(x) = \exp(ikx)u_{k,n}(x),$$

where k is the crystal momentum lying in range $\frac{-\pi}{a} \leq k \leq \frac{\pi}{a}$, $u_{k,n}(x+a) = u_{k,n}(x)$ is the periodic part of the wavefunction and n is the band index. We have

$$u_{k,n}(x) = \sum_k b_k \exp(-i\frac{2\pi kx}{a})$$

The corresponding energies $\omega_{k,n}$ are eigenvalues of the system Hamiltonian

$$H = \begin{bmatrix} \ddots & \dots & \dots & \dots & \dots \\ 0 & \frac{\hbar^2(k+\frac{2\pi}{a})^2}{2m} + V_0 & V_0 & \dots & 0 \\ 0 & V_0 & \frac{\hbar^2k^2}{2m} + V_0 & V_0 & 0 \\ 0 & 0 & V_0 & \frac{\hbar^2(k-\frac{2\pi}{a})^2}{2m} + V_0 & V_0 \\ 0 & \dots & \dots & \dots & \ddots \end{bmatrix}. \tag{17}$$

which for different l, couples free electron states $\exp(i(k+\frac{2\pi l}{a}))$, with the periodic potential.

If $(\dots, b_1, b_0, b_{-1}, \dots)'$ is the eigenvector of H in Eq. (17) corresponding to energy $\epsilon_{k,n} = \hbar\omega_{k,n}$, then

$$u_{k,n}(x) = \sum_l b_l \exp(i\frac{2\pi l}{a}x); \quad \psi_{k,n}(x) = \sum_l b_l \exp(i(k+\frac{2\pi l}{a})x).$$

$$H = \begin{bmatrix} \ddots & \dots & \dots & \dots & \dots \\ 0 & \frac{\hbar^2(k+\frac{2\pi}{a})^2}{2m} & V_0 & \dots & 0 \\ 0 & V_0 & \frac{\hbar^2k^2}{2m} & V_0 & 0 \\ 0 & 0 & V_0 & \frac{\hbar^2(k-\frac{2\pi}{a})^2}{2m} & V_0 \\ 0 & \dots & \dots & \dots & \ddots \end{bmatrix}. \tag{18}$$

Let us focus on a two by two block as shown below. For $k \in [0, \frac{\pi}{a}]$, the remaining matrix in Eq. (17) can be taken to be diagonal as offdiagonal entries are truncated.

$$H = \begin{bmatrix} \frac{\hbar^2k^2}{2m} & V_0 \\ V_0 & \frac{\hbar^2(k-\frac{2\pi}{a})^2}{2m} \end{bmatrix}. \tag{19}$$

In fig. (4), we sketch $\omega(k)$ as function of k (right half).

The eigenevalues of (19) can be written in closed form. For

$$H = \begin{bmatrix} \mu_1 & V_0 \\ V_0 & \mu_2 \end{bmatrix}, \tag{20}$$

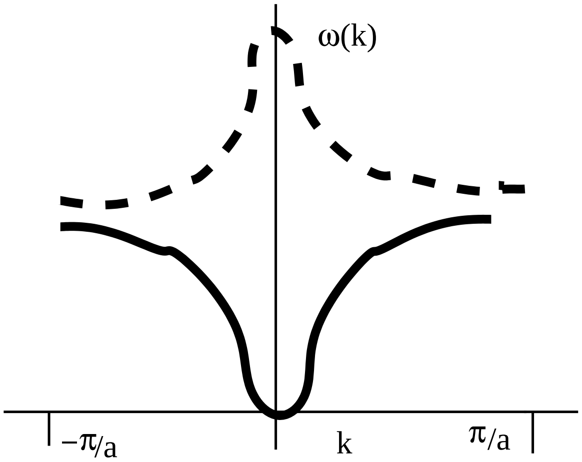

Figure 4: Figure depicts the ebergy ω_k as function of k for periodic potential in Eq. (**??**).

we have the eigenevalues

$$\frac{\mu_1+\mu_2}{2} \pm \sqrt{(\frac{\mu_1-\mu_2}{2})^2 + V_0^2}.$$

At $k=\frac{\pi}{a}$, we have $\mu_1=\mu_2$ and the difference of eigenvalues is called the band gap.

For $k \in [-\frac{\pi}{a}, 0]$, we focus on the following subblock of Eq. (17).

$$H = \left[\begin{array}{cc} \frac{\hbar^2 k^2}{2m} & V_0 \\ V_0 & \frac{\hbar^2 (k+\frac{2\pi}{a})^2}{2m} \end{array} \right]. \tag{21}$$

This helps us sketch $\omega(k)$ as function of k (left half).

We now consider when only $n=0$ is populated, we suppress band index from now with $n=0$ implied. Now consider a wavepacket around $k=k_0$ written as

$$\phi(x) = \frac{1}{\sqrt{N}} \sum_j \psi_{k_j}(x) = \frac{1}{\sqrt{N}} u(x) \sum_j \exp(ik_j x),$$

where $u_k(x)$ for k around k_0 are approximated to be same $u(x)$. Then this wavepacket evolves as

$$\begin{aligned} \phi(x,t) &= \frac{u(x)}{\sqrt{N}} \sum_j \exp(-i\omega(k_j)t) \exp(ik_j x) \\ &= \frac{u(x)}{\sqrt{N}} \exp(-i\omega(k_0)t) \exp(ik_0 x) \sum_j \exp(i\Delta k_j (x - v_g t)). \end{aligned}$$

where $v_g = \frac{d\omega}{dk}|_{k_0}$. The wavepacket evolves with velocity v_g.

Now lets analyze the evolution of the wavepacket in the presence of an electric field E, say in x direction. We again introduce time varying $k(t)$ with

$$k(t) = k + \frac{eA(t)}{\hbar}, \quad A(t) = -Et.$$

then

$$H(t) = \begin{bmatrix} \ddots & \dots & \dots & \dots & \dots \\ 0 & \frac{\hbar^2(k+\frac{2\pi}{a}+\frac{eA(t)}{\hbar})^2}{2m} + V_0 & V_1 & \dots & 0 \\ 0 & V_{-1} & \frac{\hbar^2(k+\frac{eA(t)}{\hbar})^2}{2m} + V_0 & V_1 & 0 \\ 0 & 0 & V_{-1} & \frac{\hbar^2(k-\frac{2\pi}{a}+\frac{eA(t)}{\hbar})^2}{2m} + V_0 & V_1 \\ 0 & \dots & \dots & \dots & \ddots \end{bmatrix}. \tag{22}$$

Then observe

$$\omega_{k,n}(t) = \omega_{k+\frac{eA(t)}{\hbar},n} = \omega_{k-\frac{eEt}{\hbar},n}. \tag{23}$$

This is how energies of $H(t)$ change when we apply electric field. To solve for time varying Schröedinger equation we have to realize that for moderate E, the change of $H(t)$ is adiabatic and hence we really just follow the eigenvectors of $H(t)$. To see how this adiabatic following works, say at $t = 0$, we are in the ground state say $X(0)$ (note $X(0) = (\dots, a_1, a_0, a_{-1}, \dots)'$, where $\psi_k(x) = \sum_l a_l \exp(i(k + \frac{2\pi l}{a})x)$.) then $X(t)$ satisfies the Schröedinger equation

$$\dot{X} = \frac{-i}{\hbar} H(t) X.$$

Lets diagnolize $H(t)$ as

$$H(t) = \Theta(t)\Lambda(t)\Theta'(t),$$

where $\Theta(t)$ is matrix of eigenvectors and $\Lambda(t)$ eigenvalues. Let us assume that $V_{-l} = V_l$ in Eq. (22), that is to say periodic potential is symmetric around origin. Then $H(t)$ is symmetric and $\Theta(t)$ a real rotation matrix. As we will see, this will ensure we don't have any geometric phases in our adiabatic evolution. Then we have $\dot{\Theta}(t) = \Omega(t)\Theta$, where $\Omega(t)$ is a skew symmetric matrix and we get for $Y(t) = \Theta'(t)X(t)$,

$$\dot{Y} = (\frac{-i}{\hbar} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ \vdots & \dots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix} + \underbrace{\Theta'\Omega(t)\Theta}_{\bar{\Omega}})Y. \tag{24}$$

where λ_1 is the smallest eigenvalue and so on. Note $Y(0) = (1, 0, \dots, 0)'$.

Now there is a gap between the lowest eigenvalue and higher eigenvalues which is atleast the band gap in range of eV. $\bar{\Omega}_{1j}(t)$ are comparatively small as we will show in following, and therefore they average out and $Y(t)$ simply evolves at $Y(t) = \exp(-\frac{i}{\hbar}\int \lambda_1(\sigma))Y(0) = \exp(-i\int \omega_k(\sigma))Y(0)$ and we get that we just adiabatically follow the eigenvector of H to get

$$\psi(t) = \exp(-i\int \omega_k(\sigma)d\sigma)\psi_{k+\frac{eA(t)}{\hbar}} = \exp(-i\int \omega_{k-\frac{eE\sigma}{\hbar}}d\sigma)u_{k-\frac{eEt}{\hbar}}(x)\exp(ik(t)x).$$

To estimate how big are elements $\bar{\Omega}_{1j}(t)$, observe

$$\begin{aligned} \dot{H} &= \Theta(t)\dot{\Lambda}(t)\Theta'(t) + [\Omega, H] \\ \Theta'(t)\dot{H}\Theta(t) &= \dot{\Lambda}(t) + [\bar{\Omega}, \Lambda] \\ \|[\bar{\Omega}, \Lambda]\| &\leq \|\Theta'(t)\dot{H}\Theta(t)\| = \|\dot{H}\|. \end{aligned}$$

Let $\bar{\Omega}_*(t) = \max\{\bar{\Omega}_{1j}(t)\}$, then we have

$$(\lambda_1 - \lambda_2)\bar{\Omega}_*(t) \leq \|[\bar{\Omega}, \Lambda]\| \leq \|\dot{H}\|. \tag{25}$$

Note we truncate H in Eq. (17) to finite size, because when l on diagonal terms of H become large the corresponding offdiagonals are rapidly truncated and in computing the ground state the submatrix with few small l suffices. Usually $n = 10$ is appropriate when V_l are of order eV.

When we compute $\dot{H}$ we only have terms on diagonal and they are of the form

$$\frac{eE\hbar}{m}(k + \frac{eA(t)}{\hbar}),$$

then

$$\|\dot{H}\| = \frac{eE}{\sqrt{m}}\sqrt{\sum_l \frac{\hbar^2(k + \frac{2\pi l}{a} + \frac{eA(t)}{\hbar})^2}{m}} = \frac{eE}{\sqrt{m}}\sqrt{2tr(\bar{H})}. \tag{26}$$

where $\bar{H}$ is just H without V_0 on the diagonal. Then $\bar{\Omega}_*(t)$ from Eq. (25) satisfies ,

$$\hbar\bar{\Omega}_*(t) \leq \frac{\hbar eE}{\sqrt{m}(\lambda_1 - \lambda_2)}\sqrt{2tr(\bar{H})}. \tag{27}$$

Observe $(\lambda_1 - \lambda_2)$ is few eV and $tr(\bar{H})$ is also tens of eV, but for field say $E = 10V/m$, the term

$$\frac{\hbar e E}{\sqrt{m}} \sim \frac{10^{-33}\times 10^{-18}}{10^{-15}} = 10^{-36}.$$

$$\frac{\hbar e E}{\sqrt{m}(\lambda_1-\lambda_2)}\sqrt{tr(\bar{H})} = 10^{-36}/\sqrt{10^{-18}} = 10^{-27}J = 10^{-8}eV.$$

Now we observe that $(\lambda_1 - \lambda_2)$ is of order eV while $\hbar\bar{\Omega}_*(t)$ is of order $10^{-8}eV$, then in Eq. (24), we have elements $\bar{\Omega}_{1l}$ average out, giving us a pure adiabatic evolution.

It should be noted that in the equation

$$\omega_k(t) = \omega_{k-\frac{eEt}{\hbar}}.$$

we treat $k - \frac{eEt}{\hbar}$ in reduced Brillouin zone i.e, for $b > 0$, $k - \frac{eEt}{\hbar} = \frac{\pi}{a} + b$ is reduced to $-\frac{\pi}{a} + b \in [-\frac{\pi}{a}, \frac{\pi}{a}]$.

Now lets envisage a situation where we turn on a electric field in a conductor for some time τ and switch it off. Then $A(t) = -\int_0^t E(\tau)d\tau$ rises from 0 to a steady value A. This can be achieved as shown below in figure 5, where a conducting loop is pierced by a solenoid of area a_r. When current in the solenoid is turned on from 0 to a steady state value, that creates a magnetic field $B(t)$ which goes from 0 to a steady state value B and hence establishes a transient electric field in the conducting loop and finally results in a steady A in the loop such that

$$A = \frac{Ba_r}{2\pi r}.$$

Then the electron wavefunction $\psi_k = \exp(ikx)u_k(x)$ is adiabatically transformed to $\exp(i(k + \frac{eA}{\hbar})x)u_{k+\frac{eA}{\hbar}}(x)$. The initial wavepacket

$\phi(x) = \frac{1}{\sqrt{N}}u_{k_0}(x)\sum_j \exp(k_j x)$ is adiabatically transformed to

$$\phi(x) = \frac{1}{\sqrt{N}}u_{k_0+\frac{eA}{\hbar}}(x)\sum_j \exp((k_j + \frac{eA}{\hbar})x).$$

If initially the wavepacket was moving with a group velocity $v_g = \frac{d\omega(k)}{dk}|_{k_0}$, now it moves with a group velocity $v_g = \frac{d\omega(k)}{dk}|_{k_0+\frac{eA}{\hbar}}$. This is shown in Fig. 6, where $E(t)$ is transiently turned on and off and $A(t)$ reaches a steady state value A. The figure shows how A in the conducting loop shifts the energy of a wavepacket and thereby changes the group velocity and hence accelerates the wavepacket. Lets estimate the shift $\frac{eA}{\hbar}$ for the example we have. Suppose solenoid has $N = 10$ turns per c.m. and carries a final current of $I = 1$ milli Ampere. Then it establishes a B field of $B = \mu_0 NI$. Let r be .1 m and $a_r = 1$ cm^2. Then we have $\frac{eA}{\hbar} = 10^4\text{m}^{-1}$. At half filling $k_0 = \frac{\pi}{2a}$, with $a = 1A^\circ$, $k_0 \sim 10^{10}\text{m}^{-1}$ and is displaced .0001% by electric field.

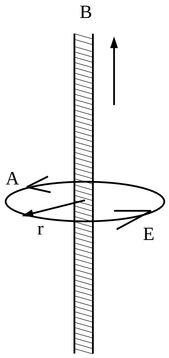

Figure 5: Fig. shows a solenoid inside a loop conductor. As current in solenoid is turned on a magnetic field B is established inside the solenoid, which creates a E in the loop that rises and decays and A that goes from 0 to a steady state value.

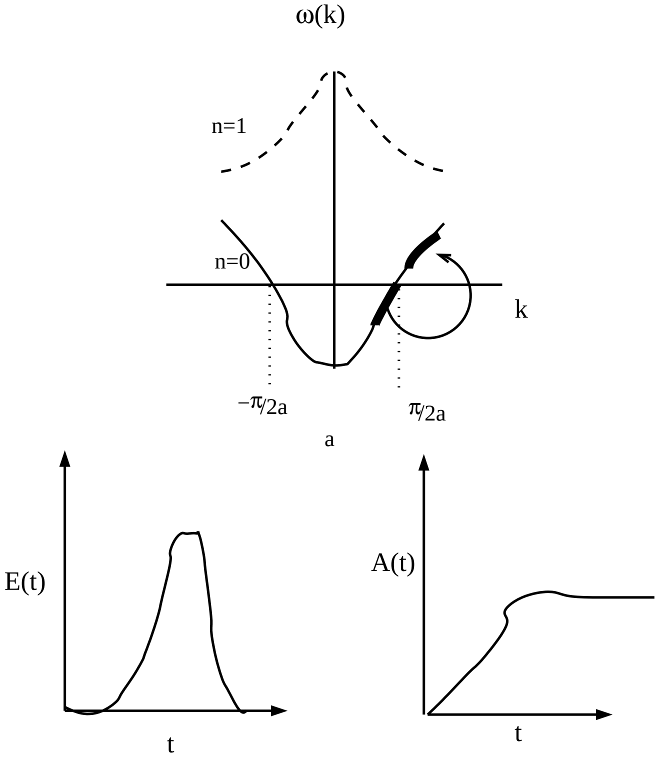

Figure 6: Bottom figure shows how application of current in solenoid in Fig. 5 establishes a transient E and A in the conducting loop which shifts the energy of a wavepacket and thereby changes the group velocity and hence accelerates the wavepacket as in top figure.

We can now understand conduction in solids. Consider a half filled band in a conductor (say the conducting loop in above example) as shown in the figure 7a. There is no net velocity of electrons. There are as many wavepackets moving in the positive direction as negative direction. While the wavepacket centered at k_0 has group velocity $v_g = \frac{d\omega}{dk}|_{k_0}$, there is a wavepacket at $-k_0$ with group velocity $-v_g$. There is no net current, though electrons are itenerant. Now say we apply a transient electric field as in solenoid example above which shifts the energy $\omega_k \to \omega_{k+\frac{eA}{\hbar}}$. Then we reach a configuration as shown in Fig. 7b. Then we see we have excess of wavepackets with positive v_g. In the solenoid example we calculated the excess to be around .0001%. This gives net conduction and current. Therefore half filled bands conduct. Now imagine a insulator where band is completely filled, then all $k \in [-\frac{\pi}{a}, \frac{\pi}{a}]$ are occupied. After application of E we have $\omega_k \to \omega_{k+\frac{eA}{\hbar}}$, but because we are in the reduced Brillouin zone, nothing happens, as all energies are simply shifted, the energy of one wavepacket takes the value of another and so on. Although individual wavepackets are accelerated and deaccelerated the sum total of velocities is still zero. It is a common understanding that in a insulator all $k's$ are filled so electric field cannot accelerate a wavepacket and everything is stuck. In our picture, wavepackets are accelerated/deaccelerated but sum total of v_g of all the wavepackets remains zero.

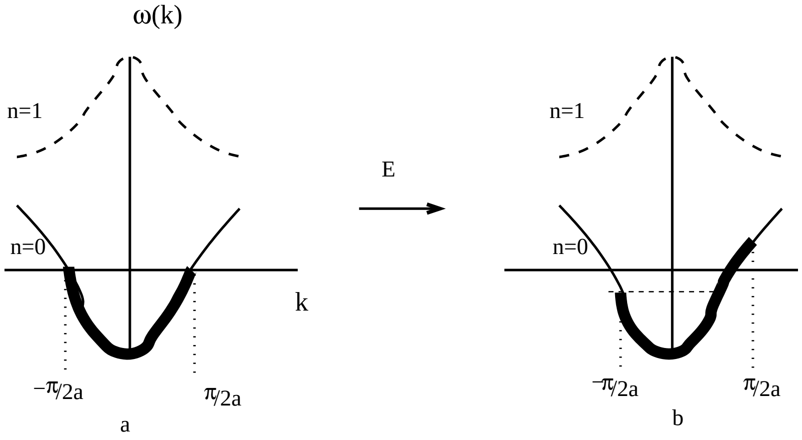

Figure 7: Fig. a shows a half filled band. Fig. b shows how by application of electric field, the energies are shifted.

In an insulator as the electron wavefunction $\psi_k = \exp(ikx)u_k(x)$ is adiabatically transformed to $\exp(i(k + \frac{eA}{\hbar})x)u_{k+\frac{eA}{\hbar}}(x)$, when $k + \frac{eA}{\hbar} > \frac{\pi}{a}$, then putting it back in reduced Brillouin zone let $k + \frac{eA}{\hbar} = \frac{2\pi}{a} - k'$, with $k' \in [\frac{-\pi}{a}, \frac{\pi}{a}]$, then observe $u_{k+\frac{eA}{\hbar}}(x)$ really means $\exp(-i\frac{2\pi}{a}x)u_{-k'}(x)$.

Now in presence of an electric field we have,

$$\omega_k(t) = \omega_{k-\frac{eEt}{\hbar}}. \tag{28}$$

Then the group velocity satisfies,

$$\begin{aligned} v_g(t) &= \frac{d\omega_{k-\frac{eEt}{\hbar}}}{dk}, \qquad (29) \\ \frac{dv_g(t)}{dt} &= \frac{d}{dk}\frac{d\omega_{k-\frac{eEt}{\hbar}}}{dt} = \frac{d^2\omega_{k-\frac{eEt}{\hbar}}}{dk^2}\frac{eE}{\hbar} = \hbar^{-2}\frac{d^2\epsilon_{k-\frac{eEt}{\hbar}}}{dk^2}eE = -\frac{1}{m^*}eE, \qquad (30) \end{aligned}$$

where $m^* = \hbar^2(\frac{d^2\epsilon_{k+\frac{eEt}{\hbar}}}{dk^2})^{-1}$ is the effective mass.

For pedagogical reasons we have restricted to 1D , we can easily generalize the above to 3D. Let ∇' be gradient written as a column vector and ∇ be gradient written as a row vector. Writing v_g as column 3 vector,

$$\begin{aligned} v_g(t) &= \nabla'_k\omega_{k+\frac{eEt}{\hbar}}, \qquad (31) \\ \frac{dv_g(t)}{dt} &= \nabla'_k\frac{d\omega_{k-\frac{eEt}{\hbar}}}{dt} = \nabla'_k\nabla_k\omega_{k-\frac{eEt}{\hbar}}\frac{eE}{\hbar} = \hbar^{-2}\nabla'_k\nabla_k\epsilon_{k-\frac{eEt}{\hbar}}eE = -\frac{1}{m^*}eE, \qquad (32) \end{aligned}$$

where $m^* = \hbar^2(\nabla'_k\nabla_k\epsilon_{k+\frac{eEt}{\hbar}})^{-1}$ is the effective mass matrix.

We have studied the acceleration of the wavepacket in presence of electric field. We may easily generalize it to a magnetic field $B\hat{z}$ in the $\hat{z}$ direction. We just note that

$$\begin{aligned} v_g(t) &= \nabla'_k\omega_{k(t)}. \qquad (33) \\ \frac{dv_g(t)}{dt} &= \nabla'_k\frac{d\omega_{k(t)}}{dt} = \nabla'_k\nabla_k\omega_{k(t)}\dot{k}(t) = \hbar^{-1}\nabla'_k\nabla_k\epsilon_{k(t)}\dot{k} = \frac{\hbar}{m^*}\dot{k}. \qquad (34) \end{aligned}$$

Then we get,

$$\frac{dv_g}{dt} = -\frac{e}{m}v_g \times B. \tag{35}$$

Thus we derived wavepacket dynamics of an electron in periodic potential. Of course, the whole treatment in this paper is in the absence of a resistance. The wavepacket in reality scatters of phonons and impurites. The present paper only details the dynamics of electron wavepacket between collisions with lattice.

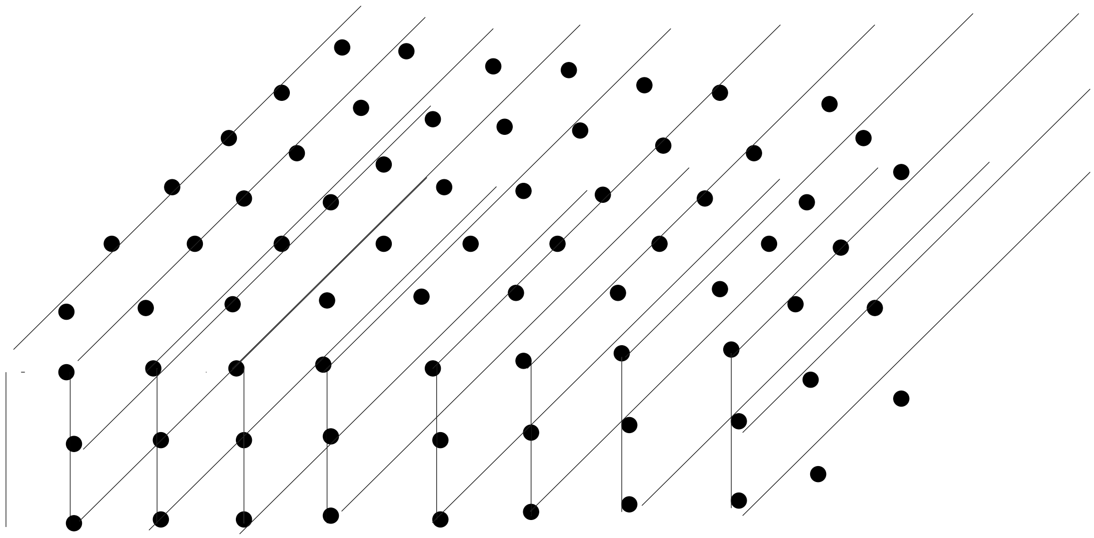

Figure 8: Fig. shows a 3D lattice made from periodic arrangement of positively charged ions.

3 Periodic potential in three dimensions: real solids

In last section, we studied one dimensional periodic lattices, and showed how electron energies get organized as bands, where each band is made up of block vectors with their wavevector taking on values between $[\frac{-\pi}{a}, \frac{-\pi}{a}]$. All this can be generalized to three dimensions. We can talk about plain wave states $\exp(-i(k_x x + k_y y + k_z z))$, which we abbreviate (k_x, k_y, k_z), for the electrons and how periodic potential couples them. For example periodic potential can be simplified to

$$V(x, y, z) = V_0 \cos^2(\frac{\pi x}{a}) \cos^2(\frac{\pi y}{a}) \cos^2(\frac{\pi z}{a}) \tag{36}$$

which is a periodic potential in three dimensions, with period a. The potential will couple plain wave states $(k_x \pm \frac{2\pi l}{a}, k_y \pm \frac{2\pi m}{a}, k_z \pm \frac{2\pi n}{a})$. Let

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

Then for each (k_x, k_y, k_z), we can form the matrix

$$H = \begin{bmatrix} \ddots & \dots & \dots & \dots & \dots \\ 0 & E(k_x + \frac{2\pi}{a}, k_y, k_z) & V_0 & \dots & 0 \\ 0 & V_0 & E(k_x, k_y, k_z) & V_0 & 0 \\ 0 & 0 & V_0 & E(k_x - \frac{2\pi}{a}, k_y, k_z) & V_0 \\ 0 & \dots & \dots & \dots & \ddots \end{bmatrix}. \tag{37}$$

when we diagonalize this matrix, we get the energies $\epsilon_n(k_x, k_y, k_z)$, where n indices the various bands. The corresponding eigenvectors are

$$u_n(k_x, k_y, k_z) = \sum_{lmn} b_{lmn}(k_x \pm \frac{2\pi l}{a}, k_y \pm \frac{2\pi m}{a}, k_z \pm \frac{2\pi n}{a}).$$

we do it for all $-\frac{\pi}{a} \leq k_x, k_y, k_z leq \frac{\pi}{a}$ and obtain energy bands. Fig. 10 shows energybands for periodic lattice in two dimensions.

In this approach, we have outlined, we treat the lattice as periodic potential and solve for electron energies and wavefunctions. There is nothing wrong with this approach, and it is most natural yet there is a shortcoming to it. We have neglected completely the screening of ion charge due to electrons. Therefore the energies we get are much lower than what we will obtain if we accounted for screening/repulsion due to electronic charge. Therefore our eigenvectors and energies do not reflect the true ground states. How do we get true electron energies and wavevectors, taking into acount the effect of electron screening. There is approximate mathod for doing this it is called the tight binding approximation.

4 Tight Binding Approximation

The basic idea is as follows. Consider a periodic array of say hydrogen atoms. If we take an isolated hydrogen atom then the electron around its nucleus orgainizes as orbitals with energies ϵ_k. Now suppose we bring hydrogen atoms together, as long as they are far nothing changes, as each hydrogen atom has its own electron and they are isolated because electron of one hydrogen atom sees no potential due to another becacuse the electron in the other atom screens its charge. Then if we have a array of isolated hydrogen atoms, well separated then the enrgies and eigenevectors are as for individual ones. But now suppose we bring the atoms further close so that electron of atom A enters the electron clound of electron B. Then the electron of atom A will begin to also see some potential due to ion B, as its charge is not fully screened and this will modify the wavefunction and energies of electron A.

Let us call the potentials of ions V_A and V_B and the wavefunctions ϕ_A and ϕ_B each with energy ϵ_0. There is a transition from ϕ_A to ϕ_B given by

$$\langle \phi_B V_B \phi_A \rangle = -t.$$

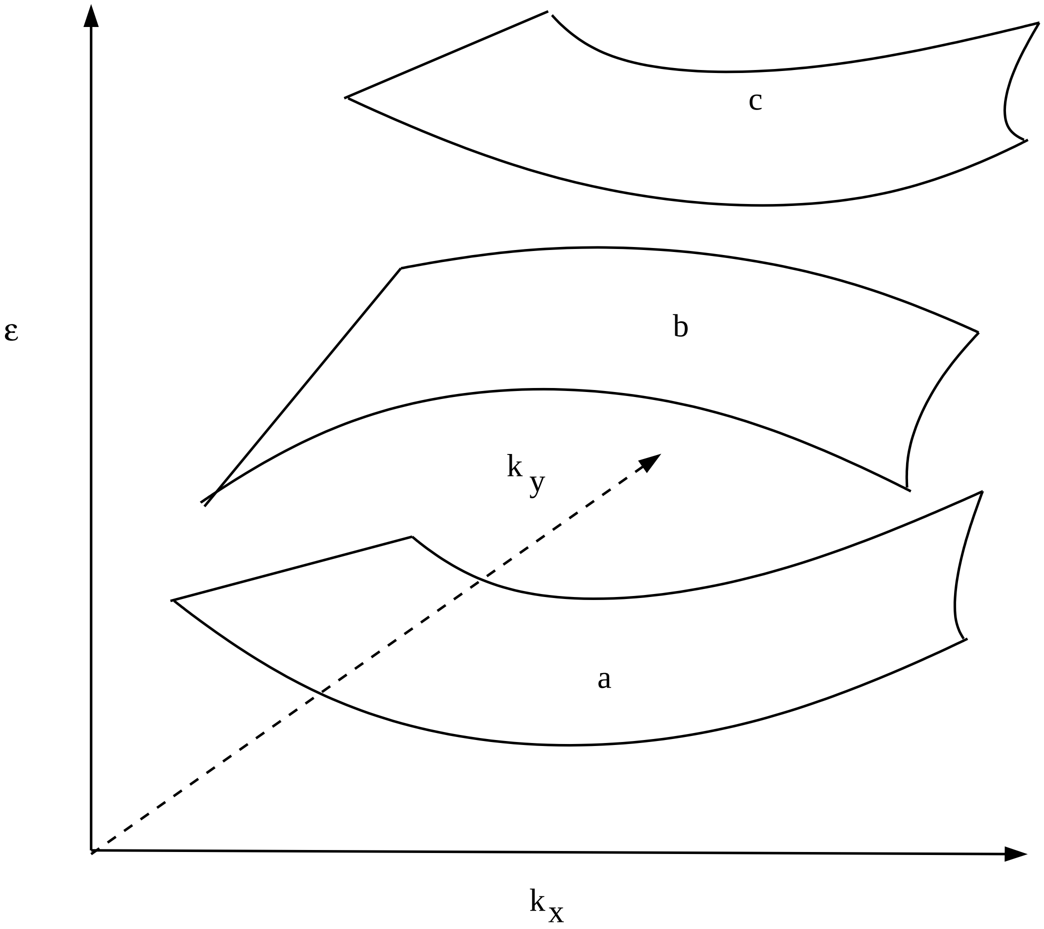

Figure 9: Fig. shows energy bands for periodic two dimensional lattice.

and similarly by symmetry there is transition from ϕ_B to ϕ_A given by

$$\langle \phi_A V_A \phi_B \rangle = -t.$$

Then in the basis ϕ_A and ϕ_B, we can write the Hamiltonian as

$$H = \begin{bmatrix} \epsilon_0 & -t \\ -t & \epsilon_0 \end{bmatrix} \tag{38}$$

The eigenvalues now are $\epsilon_0 \mp t$ and eigenvectors $\frac{1}{2}(\phi_A \pm \phi_B)$. The two electrons will now accupy the lower energy orbital $\frac{1}{2}(\phi_A + \phi_B)$. The new orbitals are linear combination of atomic orbital (LCAO).

Thus we have seen how atomic orbitals are modified when we bring atoms closer. We

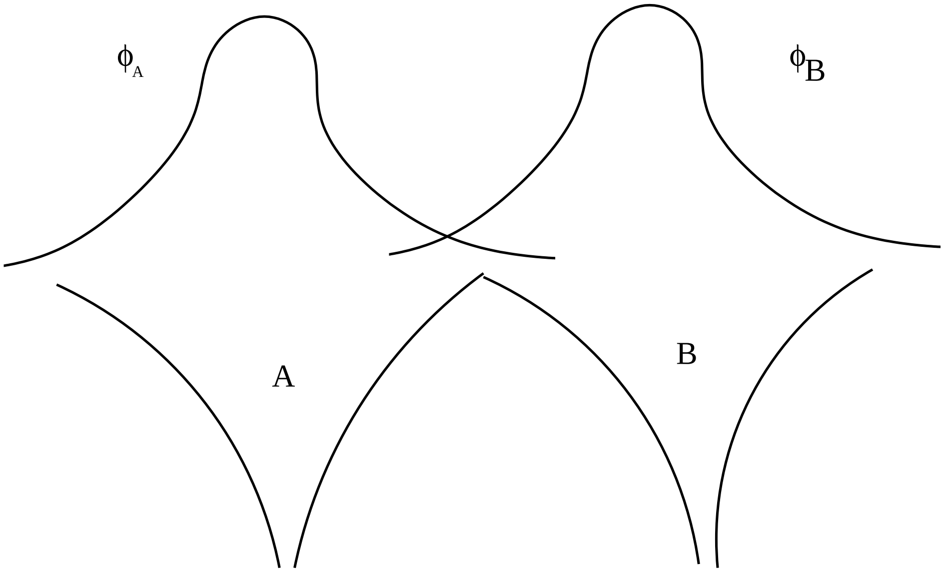

Figure 10: Fig. shows two potential wells A and B and their corresponding orbitals with overlap between them.

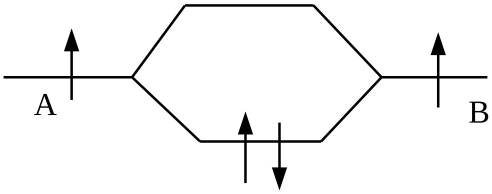

Figure 11: Fig. shows two atomic orbitals ϕ_A and ϕ_B and their linear combinations with energy $\epsilon_0 - t$ and $\epsilon_0 + t$.

form molecular orbitals. Now we extend this to a periodic array say first in one dimension, also called a 1 d chain.

$$H = \begin{bmatrix} \epsilon_0 & -t & 0 & \dots & \dots & -t \\ -t & \epsilon_0 & -t & 0 & \ddots & \vdots \\ 0 & -t & \epsilon_0 & -t & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & -t & \epsilon_0 & -t \\ -t & 0 & \dots & \dots & -t & \epsilon_0 \end{bmatrix} \tag{39}$$

where we have introduced a transition between 1 and n to close the chain. H above is a

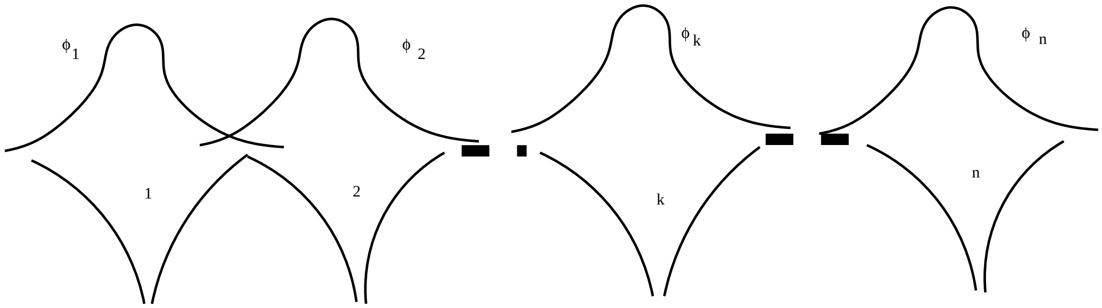

Figure 12: Fig. shows atomic orbitals ϕ_1 to ϕ_n and their potential wells.

circulant matrix of the form

$$\begin{bmatrix} r_0 & r_1 & r_2 & \dots & \dots & r_n \\ r_n & r_0 & r_1 & r_2 & \ddots & \vdots \\ r_{n-1} & r_n & r_0 & r_1 & r_2 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ r_2 & \dots & \dots & r_n & r_0 & r_1 \\ r_1 & r_2 & \dots & \dots & r_n & r_0 \end{bmatrix} \tag{40}$$

A circulant marix is always diaganolized by a DFT matrix whose j^{th} column is $\frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ \omega^{j-1} \\ \omega^{2(j-1)} \\ \dots \\ \omega^{(n-1)(j-1)} \end{pmatrix}$ with eigenvalue $\sum_k r_k (\omega^{j-1})^{k-1}$ where ω in n^{th} root of unity $\omega = \exp(i\frac{2\pi}{n})$. Then our eigenvalues are

$$\epsilon_0 - 2t\cos\frac{2\pi j}{n}, \quad 0 < j < n-1.$$

which can be written as

$$\epsilon(k) = \epsilon_0 - 2t\cos ka, \quad \frac{\pi}{a} < k < \frac{\pi}{a}, \tag{41}$$

This is called a dispersion relation. The eigen vectors are of the form

$$\Psi(x) = \sum_l \exp(ikla)\phi(x - la) \tag{42}$$

Recall the eigenfunction for the Block electrons in a periodic potential. It has the form $\exp(ikx)p(x)$ where $p(x)$ is periodic with period a. The orbital in 42 is similar, except its exponential part is sampled at lattice points.

Thus starting from isolated atomic orbitals, with one energy ϵ_0, we branch into many energies as we decrease the spacing between atoms. The energy gets broadened into a band. This is shown in figure

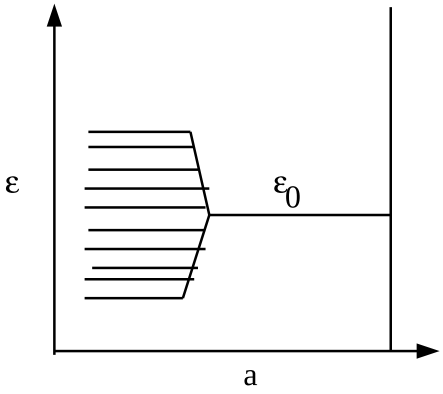

Figure 13: Fig. shows how atomic orbital energy ϵ_0 gets branced into many energies as we decrease the separation a between orbitals.

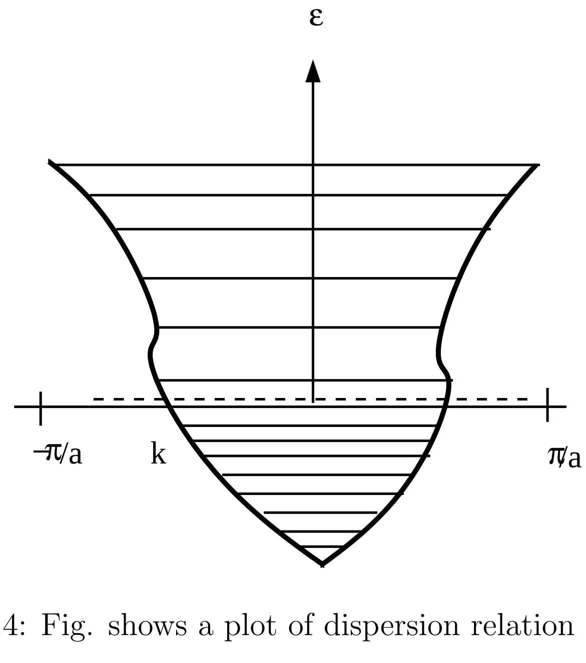

Figure 14: Fig. shows a plot of dispersion relation in Eq. (41).

Now coming back to 1 d chain of hydrogen like atoms, we found that given N atoms, we get N wavevectors ranging from $\frac{-\pi}{a}$ to $\frac{\pi}{a}$, we have N orbitals with sepration of $\frac{2\pi}{N}$ between them. These are shown in figure 14. By Pauli exclusion principle, each orbital can carry only 2 electrons with opposite spins. Since we have N electrons, if we fill the energy levels starting from minimum energy, we get only $\frac{N}{2}$ filled orbitals as each can hold two electrons. The remaining $\frac{N}{2}$ orbitals are empty. When we discussed chain of periodic potentials, we showed how half filled bands conduct. Same is true now. After we do tight binding approximation, we still have the same principle, half filled bands conduct. We can form electron wavepackets

$$\frac{1}{\sqrt{n}}\sum_j \Psi_j = \sum_l (\frac{1}{\sqrt{n}}) \sum_j \exp(ikla))\phi(x - la) \tag{43}$$

centered at wavevector k_0. Then this will evolve with group velocity $\hbar^{-1}\frac{d\epsilon(k)}{dk}|_{k_0}$.

Now imagine an atom with many orbitals with energies $\epsilon_1, \ldots, \epsilon_N$. Then as we bring these atoms together, these energies will get broadened as shown in fig (15).

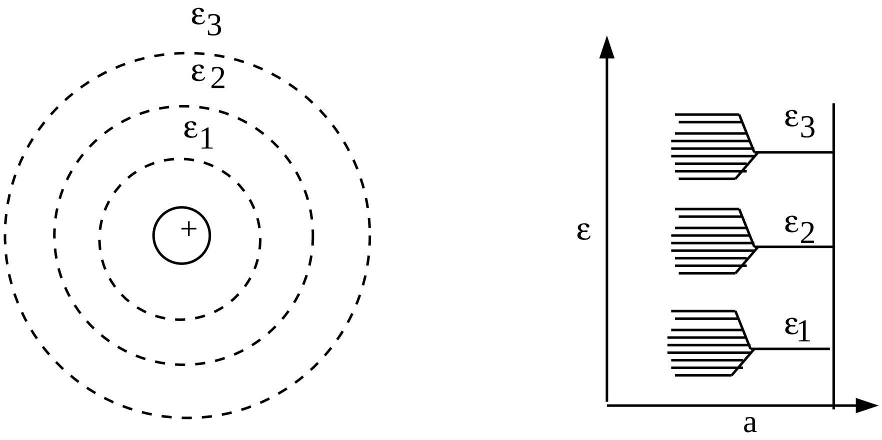

Figure 15: Fig. shows how many orbitals with energies $\epsilon_1, \ldots, \epsilon_N$ get broadened as we decrease the separation a between orbitals.

The tight binding approximation has been described using a 1-d chain. In real solids, we have periodic arrangement of ions in three dimensions. Then we have to solve tight binding approximation for three dimensions. To do this let recall 1-dimension in a different format. There we solve a eigenvalue problem of the form

$$-t(x_{l-1} + x_{l+1}) = \epsilon x_l \tag{44}$$

This is solved by using a wave-ansatz $x_l = \exp(ikla) = \omega^l$ with $\epsilon = -2t\cos(ka)$, with $\frac{-\pi}{a} \leq k \leq \frac{\pi}{a}$.

How does this problem look in two dimensions with periodic arrangement of ions with spacing a. We find we have equation,

$$-t(x_{l-1,m} + x_{l+1,m} + x_{l,m-1} + x_{l,m+1}) = \epsilon x_{lm} \tag{45}$$

This is solved by using a wave-ansatz $x_{lm} = \exp(ik_x la)\exp(ik_y ma)$ with $\epsilon = -2t(\cos(k_x a) + \cos(k_y b))$ with $\frac{-\pi}{a} \leq k_x, k_y \leq \frac{\pi}{a}$. We plot below the function $\epsilon(k_x, k_y)$.

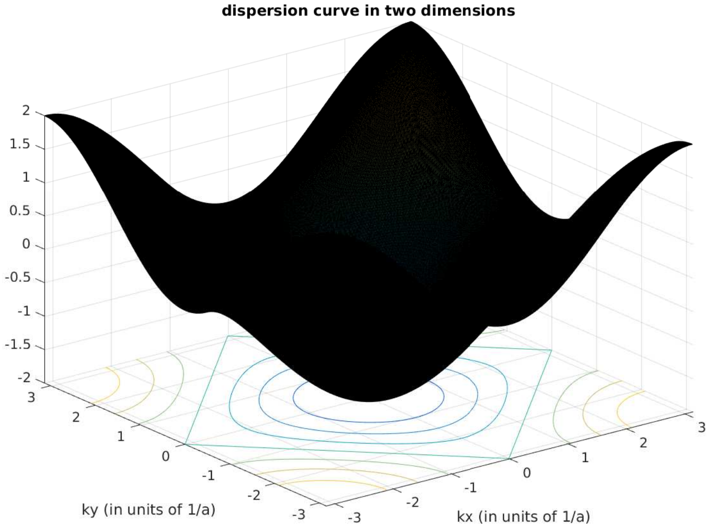

Figure 16: Fig. shows plot of $\epsilon = -2t(\cos(k_x a) + \cos(k_y b))$ as function of (k_x, k_y).

How does this problem look in three dimensions with periodic arrangement of ions with spacing a. We find we have equation,

$$-t(x_{l-1,m,n} + x_{l+1,m,n} + x_{l,m-1,n} + x_{l,m+1,n} + x_{l,m,n-1} + x_{l,m,n+1}) = \epsilon x_{l,m,n} \tag{46}$$

This is solved by using a wave-ansatz $x_{lmn} = \exp(ik_x la)\exp(ik_y ma)\exp(ik_z na)$ with $\epsilon = -2t(\cos(k_x a) + \cos(k_y a) + \cos(k_z a))$ with $\frac{-\pi}{a} \leq k_x, k_y, k_z \leq \frac{\pi}{a}$.

In the following figure 17, we show how does energies of atomic orbitals in a 3D solid say (sodium or calcium) get broadened as we bring the atoms closer. For sodium (atomic number 11) we have $1s^2 2s^2 2p^6 3s^1$ only 1 electrons in $3s$ and energy level gets broadened we have two electrons in each orbital giving that the band is half filled as a result Sodium is a conductor. How about calcium (atomic number 12) it has $1s^2, 2s^2, 2p^6, 3s^2$ and has 2 electrons in $3s$

energy level gets broadened we have two electrons in each orbital giving that the s band is fully filled hence we should have calcium as insulator. However s band overlaps with p band and hence both of then get partially filled and we get a conductor. For Aluminium (atomic number 13) we have $1s^22s^22p^63s^23p^1$ and we have three electrons. Again we fill both s and p and the s band may get fully filled but p band is partially filled and we have conductor. Hence all these are metals, they are good conductors.

Now lets take next element silicon (atomic number 14) we have $1s^22s^22p^63s^23p^2$. Since silicon will occupy most of the remaining chapter, we spend some time studying its band structure. Its band structure is very interesting. There are atoms at corner of cube and at center of all faces of a cube. Such a structure is called fcc structure. Silicon lattice is made of such a fcc lattice and its dispaced version by an amount $\frac{a}{4}(1,1,1)$ as shown in fig. 18B. There is 1 3s orbital, and three 3p orbitals, which combine to form four sp^3 orbitals. Each silicon atom say at lattice position $\frac{a}{4}(1,1,,1)$ will form bonds with its four immediate nghs $(0,0,0)$, $a(\frac{1}{2},\frac{1}{2},0)$, $a(0,\frac{1}{2},\frac{1}{2})$ and $a(\frac{1}{2},0,\frac{1}{2})$ with one sp^3 orbital each, as shown in fig. 18B. The two sp^3 orbitals of the neighbouring silicon atoms strongly overlapp as shown in fig. 18B. Then as described in fig. 12, the two sp^3 atomic orbitals combine to form two molecular orbitals at energies $\epsilon + t$ and $\epsilon - t$. Now these molecular orbitals also overlap, and hence we have a picture like following fig. 19, where molecular orbitals are formed from atomic orbitals as in fig. 19a, which due to overlap between them get further broadened as in fig. 19b with Δ as the gap between the lower and higher manifold as shown in 19b.

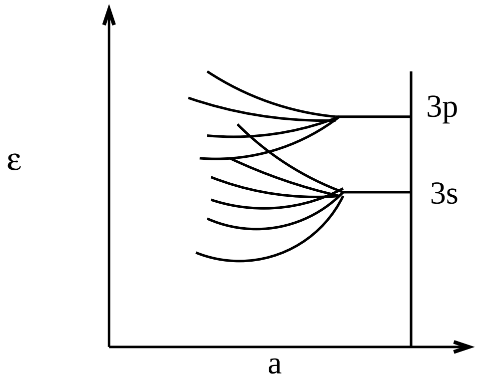

Figure 17: Fig. shows how 3s and 3p orbitals get broadened for an atom like sodium or calcium.

We develop a simple model to illustrate this process of first forming molecular orbitals from atomic orbitals and then broadening them. Consider a linear chain of atoms with strong overlap between atomic orbitals at site a_j and b_j with transfer integral $-t_1$ and weakoverlap between atomic orbitals at site a_{j+1} and b_j with transfer integral t_2. This is shown below in fig. 20.

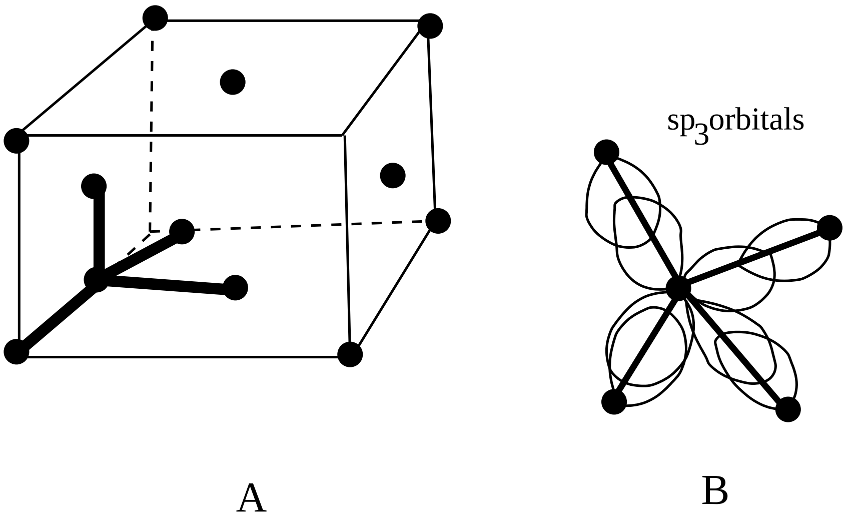

Figure 18: Fig. A shows two interspaced fcc lattices that make the silicon lattice.

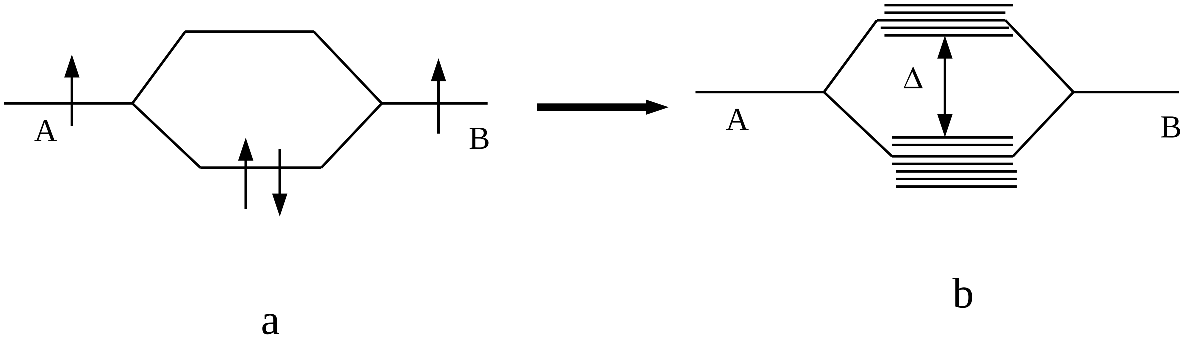

Figure 19: Fig. shows molecular orbitals are formed from atomic orbitals as in fig. a which due to overlap between them get further broadened as in fig. b.

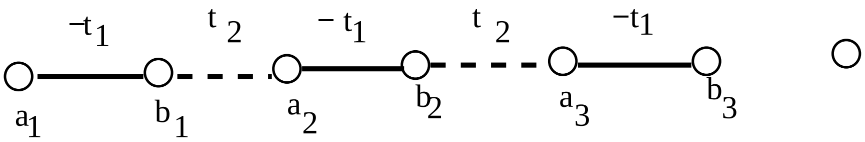

Figure 20: Fig. shows a linear chain of atoms with strong overlap between atomic orbitals at site a_j and b_j with transfer integral $-t_1$ and weakoverlap between atomic orbitals at site a_{j+1} and b_j with transfer integral t_2.

We can form the Hamiltonian

$$H = \begin{bmatrix} \epsilon_0 & -t_1 & 0 & \dots & \dots & t_2 \\ -t_1 & \epsilon_0 & t_2 & 0 & \ddots & \vdots \\ 0 & t_2 & \epsilon_0 & -t_1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & t_2 & \epsilon_0 & -t_1 \\ t_2 & 0 & \dots & \dots & -t_1 & \epsilon_0 \end{bmatrix} \tag{47}$$

to find orbitals, we solve the eigenvalue equation $Hx = \lambda x$, where $x = (x_1, x_2, \dots, x_{n-1}, x_n)$ Then expressing in terms of coordinates $y_1 = \frac{x_1+x_2}{\sqrt{2}}$ and $y_2 = \frac{x_1-x_2}{\sqrt{2}}$ etc, we get in terms of these new coordinates we get

$$\bar{H} = \begin{bmatrix} \epsilon_0 - t_1 & 0 & \frac{t_2}{2} & 0 & \vdots \\ 0 & \epsilon_0 + t_1 & 0 & \frac{-t_2}{2} & \vdots \\ \frac{-t_2}{2} & 0 & \epsilon_0 - t_1 & 0 & \vdots \\ 0 & \frac{t_2}{2} & 0 & \epsilon_0 + t_1 & \vdots \\ \dots & \dots & \dots & \dots & \vdots \end{bmatrix} \tag{48}$$

where coupling between $\epsilon_0 - t_1$ and $\epsilon_0 + t_1$ is truncated giving the above reduced Hamiltonian.

Thus we have two manifolds one with energies $\epsilon_0 - t_1$ which gets broadened to $\epsilon_0 - t_1 - \frac{t_2}{2}\cos(\frac{4\pi k}{n})$ with $0 \leq k \leq \frac{n}{2}$ and other gets broadened to $\epsilon_0 + t_1 + \frac{t_2}{2}\cos(\frac{4\pi k}{n})$. The corresponding eigenvectors are

$$(1, 0, \exp(i\frac{4\pi k}{n}), 0, \exp(i\frac{8\pi k}{n}), 0, \dots)$$

and

$$(0, 1, 0, \exp(i\frac{4\pi k}{n}), 0, \exp(i\frac{8\pi k}{n}), \dots),$$

which in terms of x_i coordinates take the form

$$(1, 1, \exp(i\frac{4\pi k}{n}), \exp(i\frac{4\pi k}{n}), \exp(i\frac{8\pi k}{n}), \exp(i\frac{8\pi k}{n}), \dots),$$

and

$$(1, -1, \exp(i\frac{4\pi k}{n}), -\exp(i\frac{4\pi k}{n}), \exp(i\frac{8\pi k}{n}), -\exp(i\frac{8\pi k}{n}), \dots).$$

which can be approximated to

$$(1, \exp(i\frac{2\pi k}{n}), \exp(i\frac{4\pi k}{n}), \exp(i\frac{6\pi k}{n}), \exp(i\frac{8\pi k}{n}), \exp(i\frac{10\pi k}{n}), \dots),$$

and

$$(1, -\exp(i\frac{2\pi k}{n}), \exp(i\frac{4\pi k}{n}), -\exp(i\frac{6\pi k}{n}), \exp(i\frac{8\pi k}{n}), -\exp(i\frac{10\pi k}{n}), \dots).$$

Rewriting these we get,

$$(1, \exp(i\frac{2\pi ka}{na}), \exp(i\frac{4\pi ka}{na}), \exp(i\frac{6\pi ka}{na}), \exp(i\frac{8\pi ka}{na}), \exp(i\frac{10\pi ka}{na}), \dots)$$

and

$$(1, \exp(i(\frac{\pi}{a}+\frac{2\pi k}{na})a), \exp(i2(\frac{\pi}{a}+\frac{2\pi k}{na})a), \exp(i3(\frac{\pi}{a}+\frac{2\pi k}{na})a), \exp(i4(\frac{\pi}{a}+\frac{2\pi k}{na})a), \exp(i5(\frac{\pi}{a}+\frac{2\pi k}{na})a), \dots)$$

Thus plotting energy vs the wavevector gives

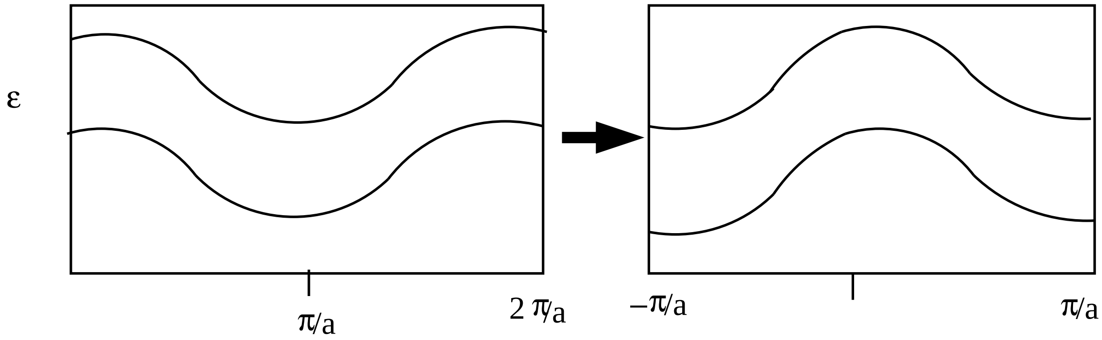

Figure 21: The figure shows indirect bandgap for the linear chain in Fig. 20.

Observe the maximum of lower energy manifold (called valence band) and minimum of the higher energy manifold (called conduction band) happens at different k. Hence this is a indirect band gap. This is basically what happens in a silicon crystal. The sp^3 orbital has two lobes a big positive lobe and a small negative lobe as shown in fig. 22A. Positive lobes from site 1 and 2 overlap to give a strong coupling like t_1 in fig. 20. The molecular orbital between sites 1 and 2 is then scattered by site 3 involving overlap of negative lobe of 1 and positive lobe of 3 as shown in fig. 22B. This gives a weak coupling like t_2 in fig. 20 of opposite sign to t_1. Hence the band structure of silicon also has a lower energy and higher energy manifold and it is indirect bandgap as in fig. 21, where now we have k values in three dimensions as opposed to one dimension as in fig. 21. In fig. 20 if we have one electron per atomic orbital then all the lower energy manifold will be filled with two electrons per orbital and we will end up in a insulator. However if the band gap is not too high (band gap in silicon is around 1.1 eV) some of the electrons from the valence band will be excited to conduction band at high temperatures. This will make the material a conductor. Such a material which is insulator at low temperatures and conductor at high is called a semiconductor. Silicon is a prime example.

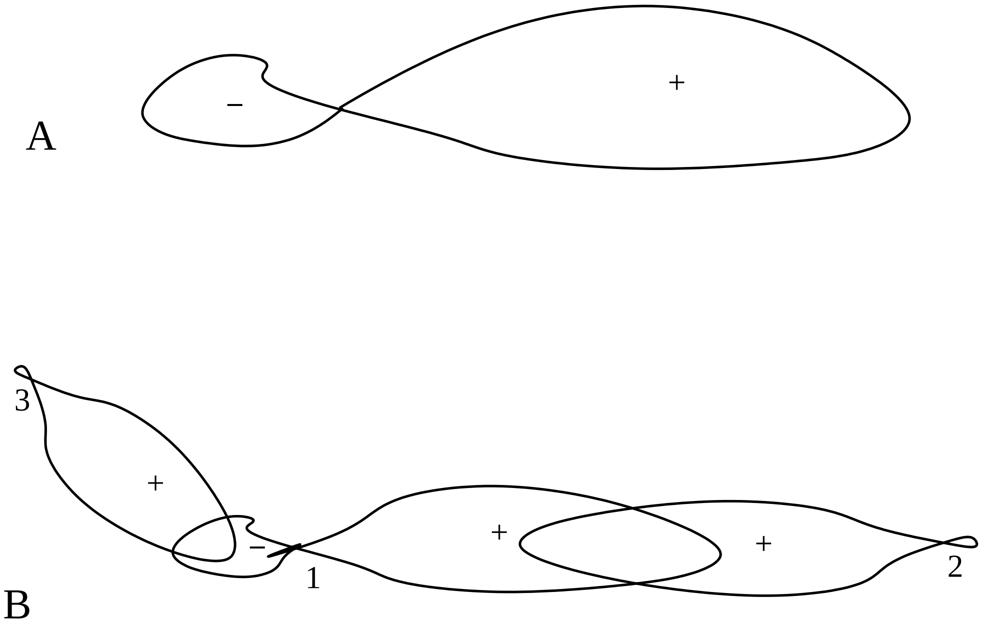

Figure 22: The figure A shows two lobes of a sp^3 orbital. Fig. B shows how these sp^3 orbitals overlap in a crystal.

We now look at another semiconductor, Gallium Arsenide (GaAs). Its structure is same as silicon as shown in fig. 18A, except now all a sites are Ga and b sites As. Outer electronic configuration of Ga is $4s^2 4p^1$ and As is $4s^2 4p^3$. Like silicon they also form sp^3 orbitals except now sp^3 orbitals of As are more electronegative and have lower energy compared to Ga. Then again we can write a simplified model as a linear chain in 20 where a sites have energy $\epsilon_1 = \epsilon_0 + \Delta\epsilon$ and b sites have energy $\epsilon_2 = \epsilon_0 - \Delta\epsilon$. Then

$$H = \begin{bmatrix} \Delta\epsilon & -t_1 & 0 & \dots & \dots & t_2 \\ -t_1 & -\Delta\epsilon & t_2 & 0 & \ddots & \vdots \\ 0 & t_2 & \Delta\epsilon & -t_1 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & t_2 & \Delta\epsilon & -t_1 \\ t_2 & 0 & \dots & \dots & -t_1 & \end{bmatrix} \tag{49}$$

to find orbitals, we solve the eigenvalue equation $Hx = \lambda x$, where $x = (x_1, x_2, \dots, x_{n-1}, x_n)$ Then using $d = \frac{t_1}{\Delta} \ll 1$ and expressing in terms of coordinates $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & d \\ -d & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ etc, we get in terms of these new coordinates we get for $\Delta\epsilon' = \sqrt{(\Delta\epsilon)^2 + t_1^2}$ and $t_2' = d\ t_2$

$$\bar{H} = \begin{bmatrix} \Delta\epsilon' & 0 & t_2' & 0 & \vdots \\ 0 & -\Delta\epsilon' & 0 & -t_2' & \vdots \\ -t_2' & 0 & \Delta\epsilon' & 0 & \vdots \\ 0 & t_2' & 0 & -\Delta\epsilon' & \vdots \\ \dots & \dots & \dots & \dots & \vdots \end{bmatrix} \tag{50}$$

Thus the energies are again in two bands the gallium band $\Delta\epsilon' - 2t_2' \cos(ka)$ and the Arsenic band $-\Delta\epsilon' + 2t_2' \cos(ka)$, where $k = \frac{4\pi m}{na}$ with $m = 0, \dots, \frac{n}{2}$. The minima of gallium band and maxima of arsenic band occur at same k as shown below in fig. 23. Therefore GaAs is a direct band gap material.

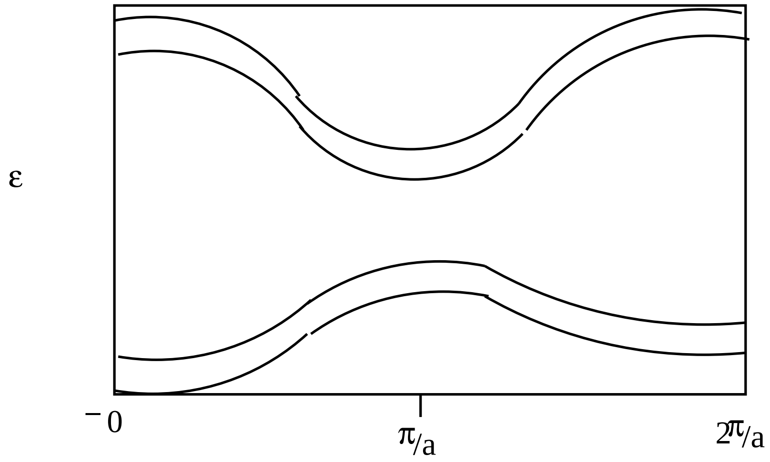

Figure 23: The figure shows direct bandgap for the linear chain in Fig. 20 modelling GaAs.

As remarked before in a semiconductor at low temperatures valence band is full and conduction band is empty, as we raise the temperature electrons from top of valence band are excited to conduction band. These excited electrons may fall back and the difference of energy of the bands may be emitted as light, given be relation $\hbar\omega = \Delta E$. The wavenumber of this light $k = \frac{\omega}{c}$. For ΔE in couple of eV, we have ω as 10^{14} and corresponds to $k = 10^6$. Compared to wavenumber $\frac{\pi}{a}$, this is negligible. Hence conservation of momentum says that wavenumebr of electron in conduction and valence band should be same, which is true in direct band-gap materials. hence direct bandgap materials can emit light. For instance GaAs emits red light. If we take GaP (Gallium Phosphide) then the bandgap is even bigger, as phoshide is more electronegative than As and hence ΔE is bigger. Giving light of higher frequence i.e. green light. It is even more in GaN (Gallium Nitride), giving blue light. With these three colors, we can make white light which is how our LED bulb works. We will say more about this when we discuss Light emitting diodes (LED).

We are now ready to discuss some devices where we will see quantum control.

4.1 Doping

Lets consider silicon again, and now suppose we replace some of them with phosphorus atoms. Silicon outer shell is $3s^2 3p^2$, while in phosphorus we have $3s^2 3p^3$. Thus phosphorus will give one more electron. We may just say that in silicon valence band is all filled, so the extra electrons given by phosphorus atom will just go into conduction band making silicon a conductor. We say silicon is n-doped. We may worry that sp^3 orbital of phosphorus will be more electronegative than silicon and therefore if we again model all this as a 1d chain, we have a situtation where we have a 1d chain where some sites (randomly distributed) have slightly less energy.

$$H = \begin{bmatrix} \epsilon_0 & -t & 0 & \dots & \dots & -t \\ -t & \epsilon_0 - \Delta & -t & 0 & \ddots & \vdots \\ 0 & -t & \epsilon_0 & -t & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & -t & \epsilon_0 & -t \\ -t & 0 & \dots & \dots & -t & \epsilon_0 \end{bmatrix} \tag{51}$$

We can treat ΔE energy increments as certain sites as perturbations. Then Hamiltonian can be written as a pure Hamiltonian H_0 without the energy increments, and a perturbation H_1 with increments .

$$H = \underbrace{\begin{bmatrix} \epsilon_0 & -t & 0 & \dots & \dots & -t \\ -t & \epsilon_0 & -t & 0 & \ddots & \vdots \\ 0 & -t & \epsilon_0 & -t & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & -t & \epsilon_0 & -t \\ -t & 0 & \dots & \dots & -t & \epsilon_0 \end{bmatrix}}_{H_0} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & \dots & \dots & 0 \\ 0 & -\Delta & 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & 0 & -t \\ 0 & 0 & \dots & \dots & 0 & 0 \end{bmatrix}}_{H_1} \tag{52}$$

Since H_0 is diagonalized with a DFT matrix U, the effect of the perturbation can be evaluated by looking at diagonal entries of $U^\dagger H_1 U$, which are just $-\Delta \frac{n_1}{n}$, where n_1 are number of sites at which we have phosphorus. Thus all energies are uniformly shifted down, because we have a more electronegative sample. For $n_1 \ll n$, the perturbation is small, and first order correction is good.

We talked about n doped silicon. We can replace silicon with Aluminium Al, whose outer shell is $3s^2 3p^1$. Not all valence band will be filled now and we will again get a conductor. The energy shifts now are positive. This is called a p doped silicon.

Most of modern electronics is playing with n and p doped silicon and making useful devices out of them. Next few sections are devoted to description of these devices. We start with most basic of these called a p-n junction.

5 p-n junction

Fig. 24a shows a pn junction with p-doped silicon on right and n doped silicon on the left. When we apply a positive voltage to p junction compared to n junction (forward bias), the following happens. The valence electron in the p region moves toward positve terminal, creating a positive p region, which pulls the conduction electrons of the n region into the p region and current flows, see fig 24b. Some of these electrons arriving from n region will recombine with holes in the valence band. Nonetheless, there is current flow in the forward bias.

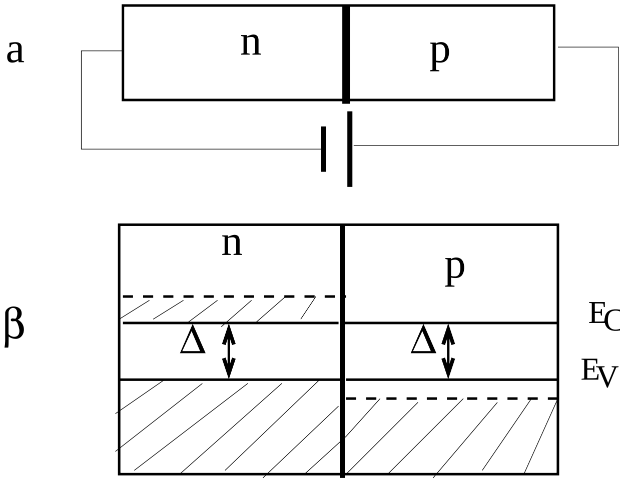

Figure 24: The figure shows a pn junction.

Now apply a negative voltage to the p junction compared to n junction (reverse bias), the following happens. The valence electrons in p region begin to move towards n junction, however all valence states in n region are filled and they donot move under an electric field, see fig 24b. Thus current cannot flow in p region in opposite direction, and we donot have current flow in reverse bias.

In practice, when we we bring a p and n junction in contact there is transfer of charge between them. The electrons in n region are in conduction band and at higher energy compared to electrons in the p region which are in valence band. Hence there is flow of electrons fron n region to p region. This leaves n region positively charged and p region negatively charged and an electric field develops at the interface as shown in fig. 25A, which prevents further flow of charges. Under this condition, if we make p terminal positive current won't flow. See the n region conduction electrons end up in gap region. This is shown in fig. 25B. To make the current flow we have to apply enough positive voltage (.7 V for silicon) on the p terminal so that it cancels the interface field. Then we have a condition like fig. 25C (forward bias) and current can flow. If we apply negative voltage to p terminal, we are in situtation 25D (reverse bias) and the electrons from p region end up in gap so no flow.

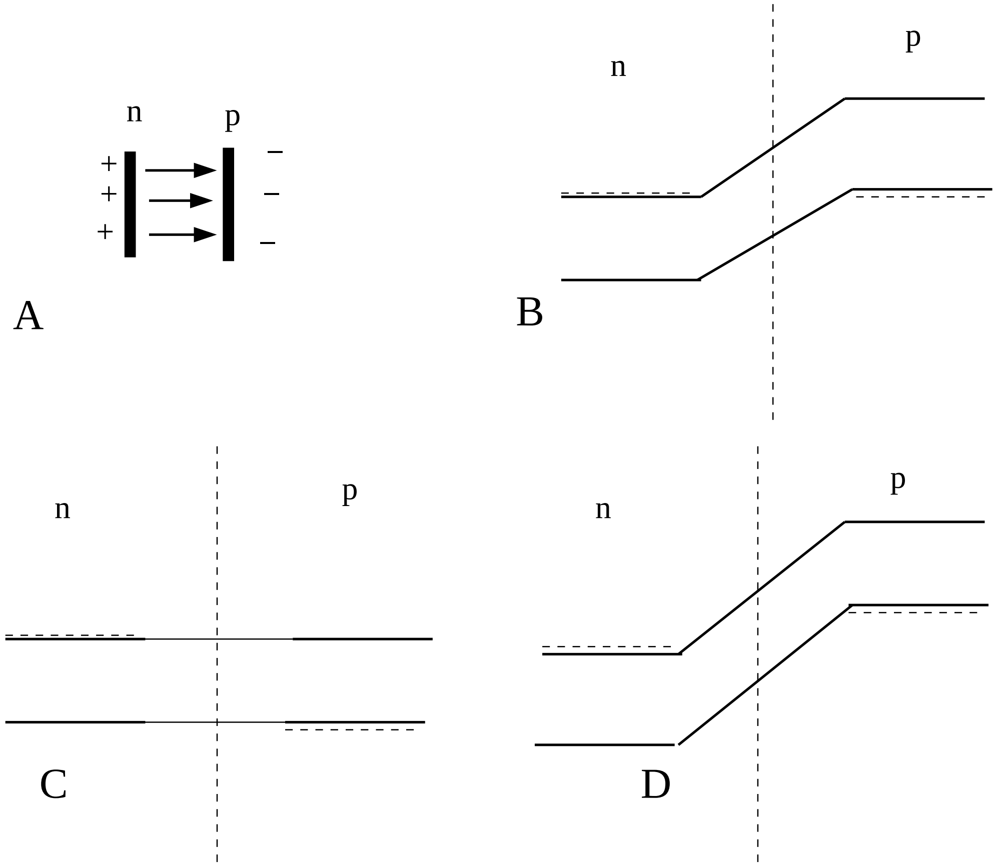

Figure 25: The fig A shows an electric field at the interface. Fig. B shows band energy diagram where fermi levels get alligned due to interface field. Fig. C shows band energy diagram under forward bias. Fig. D shows band energy diagram under reverse bias.

We saw p-n junction conducts when it is forward biased and doesnot conduct in the opposite direction. Therefore a p-n junction is used as an rectifier and is called a diode. It is shown in fig. 26B with arrow pointing towards n region. When diode is forward biased, there is .7 V drop across the diode and remaining voltage drop is across resistance. When diode is reverse biased, all drop across the diode and no voltage drop is across resistance. Volatge $V(t)$ across resistance is a rectified version of volatge $U(t)$ across source as in 26A. Rectifiers find use in all sorts of circuits including communication circuits.

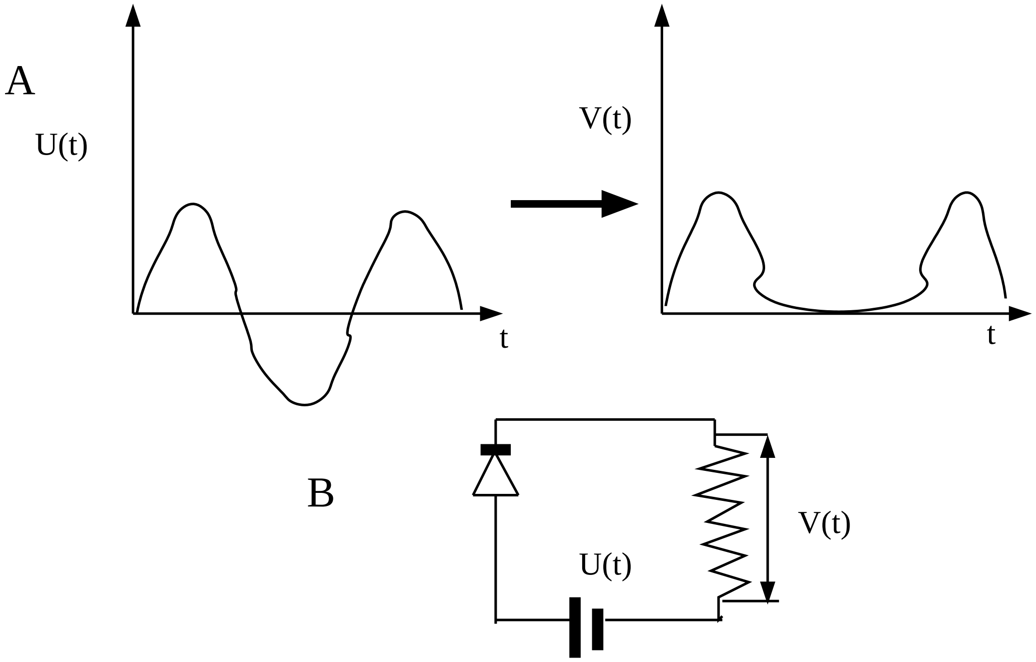

Figure 26: The fig A shows how an voltage is rectified. fig. B shows the circuit to do rectification.

We saw how to build a rectifier using semiconductor junctions. Its called a diode. Lets now see how to build an amplifier using semiconductor junctions.

6 Transistor

Transistor has two junctions as shown in 27A. A p region called base sandwitched between two n regions (called emitter and collector) as shown in fig 27A. One pn junction is forward biased (emitter and base) and other reverse biased (base and collector). Forward biased pn junctions suck valence elctrons from p region and **hurls** electrons from the n region in the p-region as shown in 27B. These electron are immediately sucked by the reverse biased junction and on their fraction of them α recombine and contribute to the p junction current the remaining $1 - \alpha$ fraction just fly by to other n region as shown in 27B. If I is the base

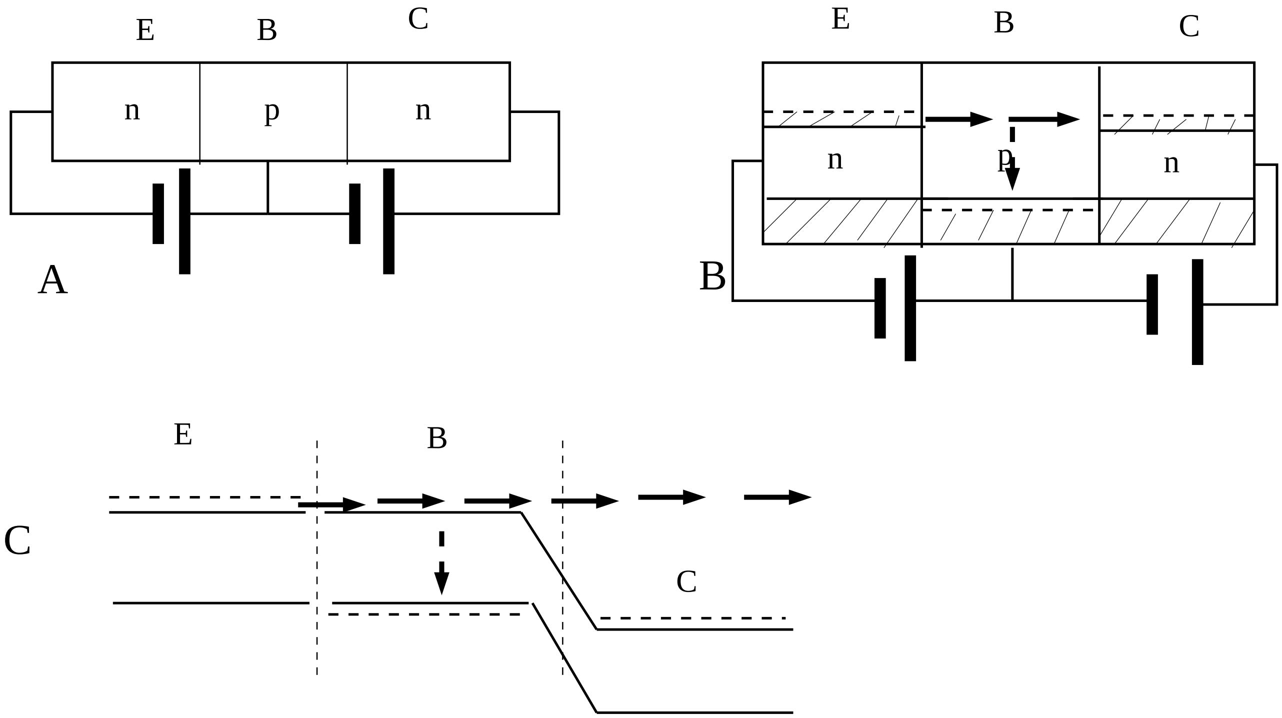

Figure 27: The fig A shows schematic ofa transistor. Fig. B shows a cartoon of how electrons and bands are in various regions without taking into account interface electric fields. Fig C shows the energy level diagram of a transistor.

current $\frac{1-\alpha}{\alpha} I = \beta I$ is the collector current and $(1+\beta)I$ the emitter current. α is small so β is large and is called the current gain of the transistor. Fig. 27C shows energy level diagram for the transistor.

6.1 Current Source and amplifier

Transistor acts a current source. Fig. 28A shows by use of a forward bias and resistance R_1, we establish a current in emitter, simply given by $I_e = \frac{V_1 - .7}{R_1}$. This gives a collector current $\frac{\beta}{\beta+1} I_e \sim I_e$ and the potential drop across resistance R_2 as $I_e R_2 = \frac{R_2(V_1 - .7)}{R_1}$. Thus the voltage V_1 is amplified to $\frac{R_2}{R_1} V_1$, where $\frac{R_2}{R_1}$ is the amplifier gain. V_1 can have oscillating parts as shown in figure 28B, which get amplified, the DC part of V_1 is used to bias the base emitter circuit as shown in figure 28B.

A good amplifier should have large input impedence as all voltage from a source appears across it then. Similarly a good amplifier should have low output impedence as all amplified

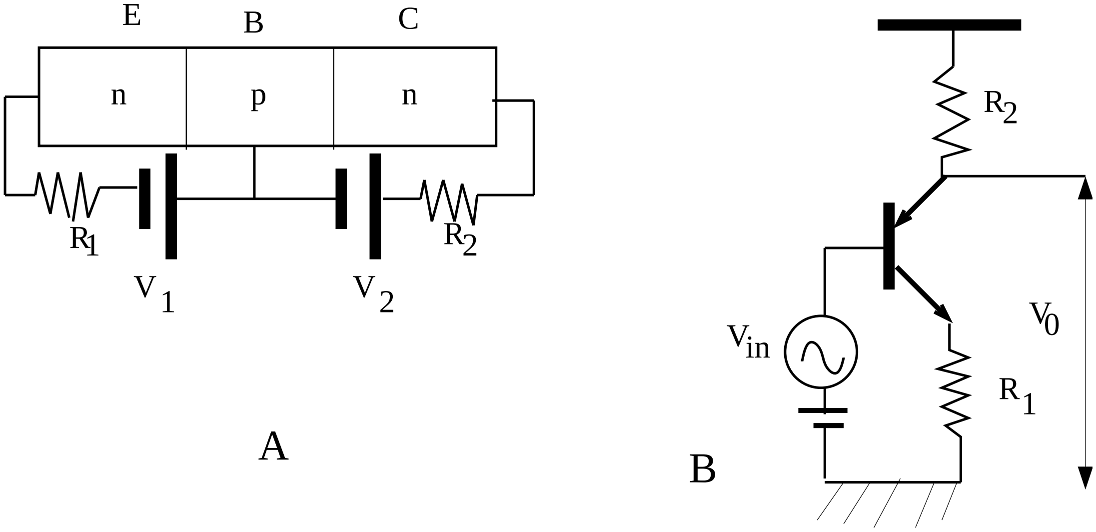

Figure 28: The fig A shows schematic of a transistor. Fig. B shows a cartoon of how electrons and bands are in various regions without taking into account interface electric fields. Fig C shows the energy level diagram of a transistor.

voltage appears across the amplifier. Lets compute the input and output impedence of the amplifier in 28B. Changing input voltage by ΔV gives $\Delta I_e = \frac{\Delta V}{R_1}$ and $\Delta I_b = \frac{\Delta V}{\beta R_1}$, then input impedence is $\frac{\Delta V}{\Delta I_b} = \beta R_1$. Similarly, if we measure the voltage across R_2 with a voltmeter with impdence R' then R_2 gets modified to $\frac{R_2 R'}{R_2+R'}$ and hence the voltage is $\frac{R'}{R_2+R'}\frac{R_2}{R_1}V_1$, hence the voltage is reduced by a factor $\frac{R'}{R_2+R'}$, this means output impdence is just R_2.

Until now we only amplified input voltage V. In many applications we want to amplify difference of voltage. Such an aplifier is called a differential amplifier or an opamp. It has two inputs V_1 and V_2 and output V_0 which is a $G(V_1 - V_2)$, where G is the amplification factor. The differential amplifier circuit is shown in 29A. The current through the emitter of transistor 1 is $\frac{V_1-V_2}{R_1}$ and hence voltage across output resistor R_0 is $\frac{(V_1-V_2)R_0}{R_1}$. The input impedence at lead 1 is simply βR_1 while at lead 2 we change the input by ΔV_2, then current in emitter 2 changes by $(\frac{1}{R_1}+\frac{1}{R_2})\Delta V_2$ and hence the input impedence is $\beta R'$ where $\frac{1}{R'} = \frac{1}{R_1}+\frac{1}{R_2}$, i.e. R_1, R_2 in parallel. The opamp is shown as a schematic in 29B, with two input leads and a output lead.

Now we show using opamp, how to build circuits with tunable gain. We do this by using concept of negative feedback. We assume that our transistors have very high gain β, so that input impedences are very high so that there is negligible current going in the leads. Further the gain is so high and any difference of the input voltages will lead to large current in the output transistor throwing it onto saturation. So we assume two leads have the same voltage. Now consider circuit 29C. The + and -1 lead have same voltage 0 and current $\frac{V}{R_1}$ goes directly to output instead of in the input lead and hence output voltage is $-V\frac{R_2}{R_1}$. We choose $R_2 \gg R_1$ to have good gain, this is called inverting amplifier. The gain can be tuned by changing R_1, R_2. There is alternate arrangement. See circuit 29D. Here $V = V_0\frac{R_1}{R_1+R_2}$ or $V_0 = (1+\frac{R_2}{R_1})V$, this is called noninverting amplifier. The gain can be tuned by changing R_1, R_2.

7 Field effect transistor

What we have studied is a Bipolar junction transistor. One of its main application is its use as an amplifier. Now we study another kind of transistor called Field effect transistor (FET). Its main application is it acts like a switch. FET is as shown in fig. 32A. There is p doped channel sandwiched between two n doped regions, called source and drain respectively. If we apply a voltage between source and drain no current will flow as shown in fig. 32B as p region has only carriers in valence band which cannot move to n regions as valence band is full. For conduction we need carriers in the p region in the conduction band. We arrange this by building a capacitor, a metal plate and a oxide dielectric and then we put positive voltage on the metal. The effect is to attract excess charges in the p region to the oxide-p interface. These excess charges will go to conduction band and we will be able to conduct, this is shown in 32C. We say we have formed an inversion layer at oxide-channel interface. We say we have turned on our switch, by applying a positive voltage to the metal plate also called gate. In 32D is shown a cartoon of a communication system that transmits when switch is on (transmit 1), and doesnot transmit when switch is off (transmit 0).

7.1 MESFET, GaAs, and High Mobility Transistors (HEMT)

We saw how we form an inversion layer by applying a positive voltage on the gate, pulling electrons. When we remove the voltage, we relax back and inversion layer disappers. How fast can be switch inversion layer on and off depends on how fast can we accelerate electrons, with electric field, which is inversely propotional to their effective mass or directly propotional to their mobility. In high speed communication, we need faster switching and hence we need to have high mobility electrons. Therefore instead of silicon, we use GaAs which has higher mobility. Recall in GaAs, the conduction band is the gallium band which have a transfer integral t that is a measure of how molecular orbitals scatter through As site. Since As is more electronegative than Si, we see larget t and larger mobility.

Therefore we use GaAs, which has larger mobility in the conduction band. The device is simply a n GaAs channel between n based source and drain. Instead of using oxide we simply

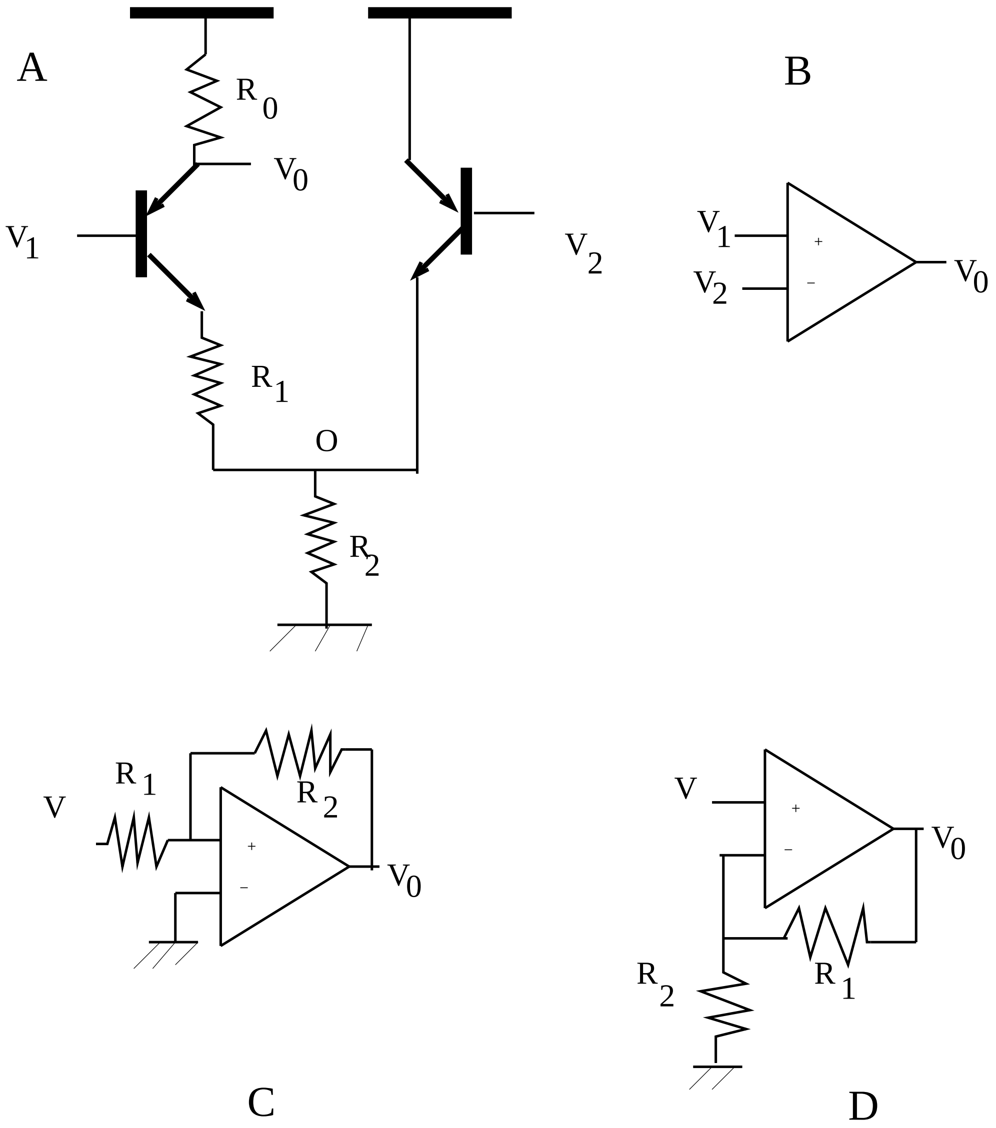

Figure 29: The fig A shows schematic of a differential amplifier. fig B shows symbolic representation of a differential amplifier. fig. C shows an inverting amplifier. fig. D shows a noninverting amplifier.

use a metal gate. If no voltage is applied to gate then channel conducts and on application of negative voltage will push electrons away from interface creating a depletion region and we stop conducting. Due to higher mobility it is faster to push/pull GaAs as compared to

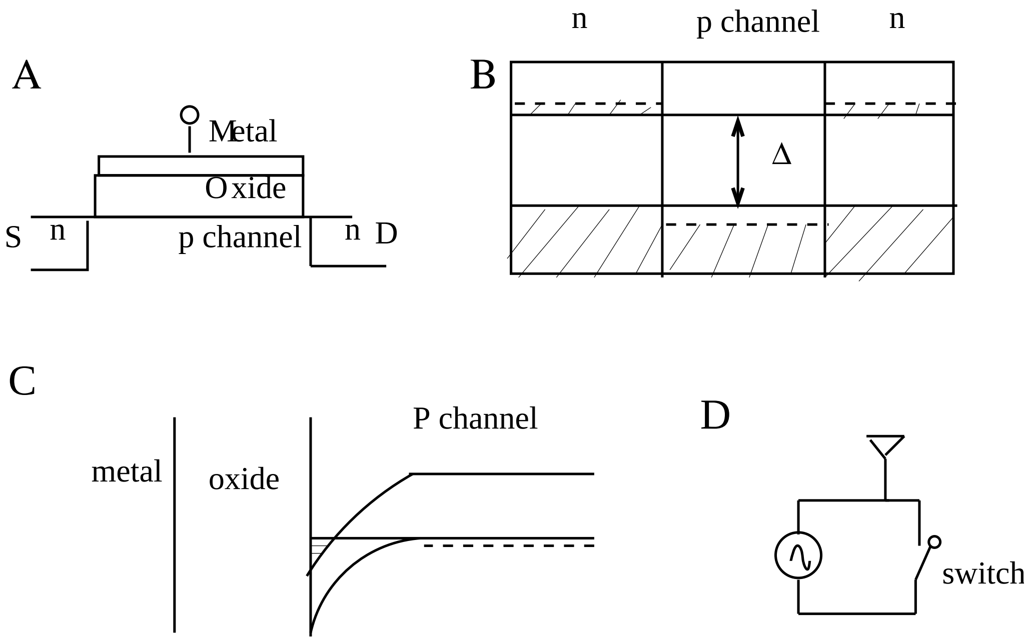

Figure 30: Fig. A shows the schematic of a MOSFET. Fig. B shows the bands for the source, channel and drain. Fig. C shows the formation of the inversion layer. Fig. D shows how FET can be used as a switch in a communication transmit circuit.

Si so it can be used for faster switching. We donot use oxide as in MOSFET because that will make n channel electrons go to oxide leading to depletion and it will already become non-conducting without a gate voltage. Why then we use oxide in MOSFET beacuse else electrons may flow directly from metal to channel and give conduction electrons without even applying any voltage. It all depends on what you want to do. InAs has even higher mobility beacuse Indium is more delocalized than Galium and has larger transfer integral.

The mobility is GaAs is still limited beacuse it is n doped (Si replaces Ga). We can think of Si as Ga with extra charge at doping sites which act as scattering impurities and limits mobility. To get better mobility we should use undoped GaAs. But now how will it work as there is no channel to begin with. We can get a channels, if we instead of oxide use n doped $\mathrm{Al}_{1-x}\mathrm{Ga}_x$ As. This material has even larger bandgap than GaAs as Al is more polar. Then electrons will flow from $\mathrm{Al}_{1-x}\mathrm{Ga}_x$ to GaAs creating a inversion layer called 2DEG (2 dimensional electron gas). This makes GaAs channel conducting. By putting a negative voltage on the gate, we repell these electrons and channel disappers. This arrangement has very high mobility and this FET is called High Mobility Transistors (HEMT). It is used for high speed communication applications.

8 Optoelectronics

8.1 Light emitting diodes and Semiconductor lasers

We have already discussed the basic idea of a light emitting diode. We have a direct band gap material like GaAs. We make a pn junction out of it. Then we put positive voltage on p junction and pull the valence electrons creating positively charged region which hurls conduction electrons from the n side. These conduction electrons then recombine with holes in the valence band emitting light. Depending on the band-gap ΔE we emit red light in GaAs, green light in GaP and blue light in GaN.

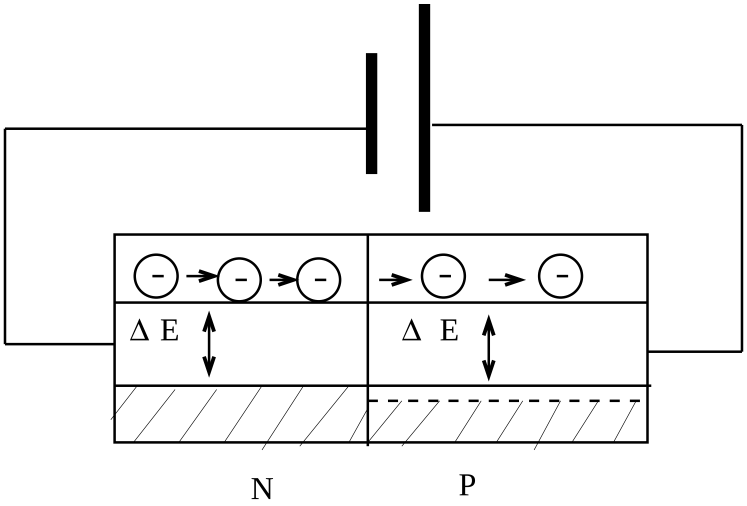

Figure 31: Fig. shows the schematic of a LED. A positive voltage on p junction pulls the valence electrons creating positively charged region which hurls conduction electrons from the n side. These conduction electrons then recombine with holes in the valence band emitting light.

The emission from LED is not coherent. As electrons fall from conduction to valence band they emit light of same frequency but the phase of light is not same i.e. it is incoherent. However if we build a cavity which can store this light by reflecting it from its ends. Then the emitted light will induce more emissions. This is called simulated emissions, very much like a two level system driven by electric field. if λ is the spontaneous emission rate and μ is the rate of simulated emission due to a single photon. Then as soon as a single photon is emitted it induces another emission and does it before another spontaneous emission takes place. Then we have two coherent photons which emit more and so on. If $\mu \gg \lambda$ we will emit many many photons before another spontaneous emission and they all will be coherent. Thus we have emitted coherent light, this is a laser as in your laser pointer.

8.2 solar cells

We talked about how in an LED when electrons are hurled in the p region they recombine with the holes and emit light. The opposite of this is that we shine light of right wavelength and promote electrons from valence to conduction band. Then we can collect these negative and positive carriers and make a battery out of them. For instance in the depletion region of a pn junction there is a inbuilt electric filed. The carriers that are generated in the region will travel to opposite side due to this field and form a battery. This is a solar cell. Its just a pn junction with light creating electron-hole pairs in the depletion region. Its schemaic is shown in figure

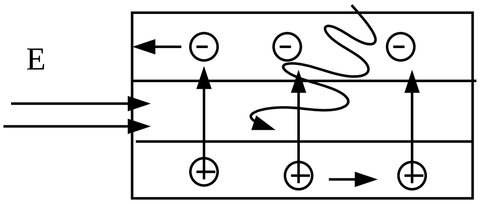

Figure 32: Fig. shows the schematic of a solar cell. Light creates electron hole pair. The electric field E in depletion region sorts them.

9 Fermi-Dirac Distribution

Consider a half filled band. Lower half is all filled and upper half empty. This is only true at 0 temperature. At finite temperatures we have a probability distribution that even upper orbitals have finite probability of occupation. This is given by Fermi-Dirac distribution. The probability of orbital with energy ϵ being occupied is given by

$$f(\epsilon) = \frac{1}{1 + \exp(\frac{\epsilon - \mu}{kT})},$$

where μ is a parameter called chemical potential. At $T = 0$, $\mu = \epsilon_F$ the fermi energy, else it is slightly above it. Then you can see at $T = 0$ nothing above ϵ_F is filled. Going by this formula even high unfilled bands have finite occupancy. If we consider a semiconductor like silicon, where valence band is full and conduction band is empty, at finite temperatures we find finite occupancy in conduction band which increases at high temperatures so that at temp like 1000 K, silicon is a conductor because conduction band has sizeable carriers and at low temperatures it is an insulator. This is shown in fig. 33. We can calculate

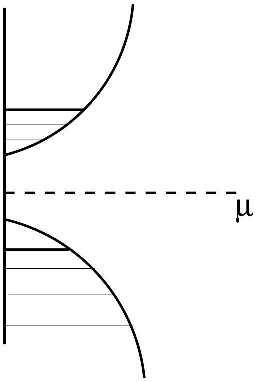

Figure 33: Fig. shows electrons and holes in conduction and valence band respectively at finite temperatures.

the concentration of carriers in conduction band n. if we take bottom of conduction band to be at 0 energy then this is simply

$$\int_0^\infty N(\epsilon) f(\epsilon) \tag{53}$$

where $N(\epsilon)$ is density of states at ϵ. Recall $\epsilon = \frac{\hbar^2 k^2}{2m}$, then $\frac{\Delta\epsilon}{\epsilon} = 2\frac{\Delta k}{k}$. number of states in band of width Δk is $4\pi k^2 \Delta k \frac{V}{(2\pi)^3} = V \frac{(2m)^{\frac{3}{2}}}{\hbar^3 (2\pi)^2} \sqrt{\epsilon} \Delta\epsilon$

Then

$$n = \frac{(2m)^{\frac{3}{2}}}{\hbar^3 (2\pi)^2} \int_{E_c}^\infty \sqrt{\epsilon} \exp(-\frac{(\epsilon - \mu)}{kT}) = \sqrt{2} \frac{m^{\frac{3}{2}}}{\hbar^3 \pi^2} \exp(-\frac{E_c - \mu}{kT}) \int_0^\infty x^2 \exp(-\frac{x^2}{kT}) = N_C \exp(-\frac{E_c - \mu}{kT}) \tag{54}$$

N_c is around $10^{19}/cm^3$ for Si at 300 K.

Similarly we can derive a formula for hole concentration p and is given by

$$p = N_V \exp(\frac{E_v - \mu}{kT}) \tag{55}$$

Then the product usig $n = p = n_i$

$$n_i^2 = np = N_C N_V \exp(-\frac{E_g}{kT}) \tag{56}$$

n_i is called intrinsic carrier concentration and depends on bandgap E_g. using $N_c = N_V \sim 10^{19}/cm^3$, and $E_g = 1.1eV$, $n_i = 10^{10}$ at room temperature.

if we dope say with donor concentration N_D, then $\mu \to \mu'$ and we can show we still have $np = n_i^2$. With $n = N_D$, we can find p. Furthermore

$$E_c - \mu' = kT\ ln(N_D/N_C). \tag{57}$$

If instead we dope with acceptors with concentration N_A.

$$\mu' - E_v = kT\ ln(N_A/N_V). \tag{58}$$

10 Problems

1. Using Eq. (19) and (21), sketch the curves in Fig. 4 , with $a = 3\ A^\circ$ and $V_0 = 100V$.
2. In above now sketch band 2 and 3 and calculate the gap.
3. A metallic bar of cross sction 1 mm^2 carries a current of 1 ampere. If free electron density is $10^{30}/m^3$, find the drift velocity of electrons.
4. In above suppose we subject electrons to an electric field of $1V/m$ for a femtosecond duration. What is the curent.
5. Consider metal calcium or magnesium with two electrons it the outer shell. This says that the conduction band will be full. Why is it a metal then.
6. In class we considered a molecular bond between two atomic orbitals each with energy ϵ and transfer element $-t$. This was a covalent bond. Generalize this to case when atomic orbitals have energy ϵ_A and ϵ_B respectively. what are the energies of two moelcular orbitals bonding and antibonding. This is called an ionic bond. Like table salt, NaCl.
7. A silicon ingot is doped with 10^{16} arsenic atoms/cm^3. Find the carrier concentrations and the Fermi level at room temperature (300 K).
8. Calculate the inbuilt potential for a silicon pn junction with $N_A = 10^{18}/cm^3$ and $N_D = 10^{15}/cm^3$ at 300K.
9. Consider a 2D lattice of atoms where atomic orbital at site (k, l) overlaps with $(k-1, l)$, $(k+1, l)$, $(k, l-1)$ and $(k, l+1)$ with strength (transfer integral) $-t$. The onsite energy is ϵ_0. Using tight-binding approximation, find the dispersion relation relating energy ϵ to wavevector.

10. Consider a 3D lattice of atoms where atomic orbital at site (k, l, m) overlaps with $(k-1, l, m)$, $(k+1, l, m)$, $(k, l-1, m)$ and $(k, l+1, m)$, $(k, l, m-1)$ amd $(k, l, m+1)$ with strength (transfer integral) $-t$. The onsite energy is ϵ_0. Using tight-binding approximation, find the dispersion relation relating energy ϵ to wavevector.