Physics of Functional Materials and Devices
Prof. Amressh Chandra
Department of Physics
Indian Institute of Technology Kharagpur

Lecture – 15, Week 4
Nearly free electron model

Welcome to the second lecture of week 4. In the previous lecture, I had introduced to you the concept of free electron gas and the reasons why you move from the classical theory and then use the Sommerfeld theory to explain the phenomena's which were being explained using the free electron gas model. Let us move further and start discussing the case of nearly free electron model. What is the requirement? How do you explain the physics associated with nearly free electron model? The Bloch theorem, the periodicity and the requirement of periodic potential which is explained using the chronic Penney model and from there we will move and talk to you about the extended and reduced zone schemes associated with metals, insulators and semiconductors. As we had seen that the free electron model could not explain certain phenomena's. The major drawbacks associated with free electron model are the following.

It could not account for the differences observed in conductors and insulators. For examples, why the conductors have large number of free electrons, but the insulators do not have any. It could not explain the variation of resistivity with temperature. For example, you are not able to explain why resistivity was decreasing as we were increasing the temperature in insulators.

The various properties of semiconductors could not also be explained using the basic free electron gas model. To overcome this limitation which was being observed using the free electron gas model, a new theory which was slightly modified theory was proposed and this is now known as nearly free electron model. Nearly free electron model, what do I mean? What is nearly free? Either it is free or it is bound, but we are talking about nearly free electron model. In this model, it was assumed that the iron cores are at rest and the potential experience by the electrons in a crystal is periodic with a period equal to the lattice constant. That means you have periodicity and the nature of potential was following this periodicity.

So if you plot the v_x versus distance curve, then you had the periodicity which was similar to the lattice constant. The assumption of the model is based on the fact that the iron cores of crystals are distributed periodically on the lattice side. That is the way you are actually defining the crystal. You have periodic arrangements of the atoms or the building blocks and you have rotational and translational symmetries. The potential contribution due to the other free electrons are taken as constant.

This type of periodic potential extends up to infinity in all directions. Infinity does not mean that you are talking literally that there is no limit to the dimension. It is meaning that all along the length in let us say x direction from 0 to the end of the crystal, you have the similar periodicity. And this periodic potential was extending in all the directions except at the surface of the crystal. Why? Because beyond the surface you do not have any crystal.

You are outside the crystal. So, you cannot have the similar periodicity. But we had seen in the previous case itself that you were saying the electrons would remain within the crystal. Why? Because of the potential which is restricting the motion of these electrons within the crystals. So effective attractive forces coming in because of the positive ion cores would prevent the electrons to move out of the crystals.

This is what we have seen in the previous case. In the free electron gas model, there was no upper limit to the energy. So, you could have any energy possible. So, you could take energy possible. According to the free electron model, the one-dimensional Schrodinger equation for an electron in a constant potential.

So, it is not like you have potential which is an infinite potential. So, at x is equal to 0 and x is equal to L, you are not talking about a case where the potentials have a value which are much much larger than the energy available to the electron. So that is what is the meaning of infinite potential. The energy is associated with the particles which are inside a box are much much less than that of the potential wall. So, for a constant potential you write the Schrodinger equation as given by equation 1.

Now you have the solution of this equation and you will get the wave function as e^{ikx} where $E - V_0 = \frac{P^2}{2m}$ or it is equal to $\frac{\hbar^2}{2m}k$. Now when we consider a periodic potential, the Schrodinger equation is written as $\frac{d^2\varphi}{dx^2} + \frac{2m}{\hbar^2}(E - V(x))\varphi = 0$. And how are we getting this periodic potential? Let us say you have the atoms which are there and these atoms are arranged in a lattice. That is what this is indicating. So, you have periodic arrangement of atoms. There is a potential energy associated with each atom. So, you have the potential energy associated with each atom. The atom would be at the position of lowest energy level. That is why this is at the mean position. You have seen that you cannot have the overlapping potential energies.

So, you have this going up to infinite, going to 0, but you do not see the whole closing parabolic potential here. So you have certain regions which are not allowed. This is physically explained or believed to explain the stability of atoms in a periodic potential. Similar concept is then used to explain the origin of band gaps in these periodic potentials. So what is the periodic potential in a lattice with a constant having a lattice constant a is the value of V(x).

For example, this is the value of V(x) and the potential after x plus a would be same. Why it should be same? It is same because you are not able to distinguish between one atom or the other. They are similar. So if you move from one point in the lattice to the next, you should get the same potential. You can choose origin at this point and I can choose origin at the other point.

It makes no difference. The crystal structure remains the same. It is just a change of origin because you cannot distinguish between this atom and the other atom. Hence, what should happen? As you move the distance which is given by the lattice parameter, the potential should be same. That is the origin of this postulate at V (x+A) is equal to V(x).

The solution of this equation which is the Schrodinger equation for the periodic potential V(x) is governed by the famous theorem known as the Bloch's theorem. What is Bloch's theorem? It states the Schrodinger equation in a periodic potential possesses the form $\varphi\,(x) = u_k(x)e^{ikx}$ where $u_k(x)$ has the same periodicity as that of the crystal lattice. That means $u_k(x + a)$ is equal to the value of $u_k(x)$. The solution of the equation 4 takes the form of equation 5 according to the Bloch's theorem. Similarly, you can then write from moving from one dimension to three dimension and you can write the form in terms of r vector.

Now you have periodicity. This was explained using the Kroneck-Penning model. What was the model? So you have atoms. You have periodicity. Now the potential energy of an electron in a linear array of positive nuclei is assumed to have the form of a periodic array of square wells with a period of A plus B.

So it was considered as a periodic array of square wells so that your mathematics could become easy. The model illustrates the behavior of electrons in a periodic potential by assuming simple one dimensional model of a periodic potential. So you try to explain the whole phenomena using the one dimensional periodic potential. What is that? This means for x greater than 0 less than a at the bottom of the well V is equal to 0. So you will have the value of V equal to 0 at the bottom of the well.

And the electron is assumed to be in the vicinity of the nucleus. Whereas outside the well that is b is more than x but it is less than a. So you have this region that is region 2. The potential energy is assumed to be V_0. So you have a barrier in this region.

So you have a barrier of width 0 and this is 0 and this is minus b. So this is what you actually assume. If you take this region 2 and region 1 which we saw in the previous slide you will get the corresponding Schrodinger equations. In region 2 it would be $\frac{d^2\varphi}{dx^2} + \frac{2m}{\hbar^2}$ (E – V_0)φ = 0 whereas in region 1 you do not have any potential and you are assuming that v is taking the form of 0 and hence you will get $\frac{d^2\varphi}{dx^2} + \frac{2m}{\hbar^2}$ E φ = 0. Assuming that the energy of the electron is less than the value of V_0.

What are we believing? That electrons are unable to jump over the potential wall. We can define two real quantities capital K and capital Q. $K^2 = \frac{2mE}{\hbar^2}$, and $Q^2 = \frac{2m(V_0 - E)}{\hbar^2}$. So using the definitions of K and Q what do you get? If you place them in equation 6 and 7 respectively what would you get? You would get $\frac{d^2\varphi}{dx^2} + K^2 \text{ P } \varphi = 0$, $\frac{d^2\varphi}{dx^2} - Q^2 \text{ P } \varphi = 0$. In region 2 and region 1 respectively.

We have just taken the definition of K and capital Q respectively. Now what are we believing? We are believing that, these Schrodinger equations are going to follow the Bloch's theorem. So what would be the form of the wave function? They would take the form of psi which is a function of x would be equal to u_k which is a function of x, e^{ikx}. And what is u_k? It is a periodic function in x with periodicity a + b. Hence $u_k(x)$ is equal to u_k (ax + a + b).

From this equation you can write what? You can write $\frac{d\varphi}{dx}$ and $\frac{d^2\varphi}{dx^2}$ can be determined. Once you have these two values you can substitute them in equations 9 and what will you obtain? You will obtain $\frac{d^2u_1}{dx^2} + 2ik\frac{du_1}{dx} + (K^2 - k^2)u_1 = 0$ and similarly for Q^2 substitution you will get the equation. Using the Bloch's theorem your general solution for $u_1(x)$ and $u_2(x)$ would then become equal to $u_1(x)$ is equal to a $e^{i(K-k)x}$ + b $e^{-i(K+k)^2}u_1(x)$ and similarly you would get the values for $u_2(x)$ and you would define two additional constants c and d. We know the boundary conditions. Using those boundary conditions you will then get what? You will get a + b = c + d and you will get a $e^{ik(a-b)} = q\ (c - d)$ and the remaining two equations would be obtained by using the boundary conditions.

Using the root of solution which is the determinant root of solving these equations we would then get the values of $\frac{q^2 + k^2}{2qk} \cdot \frac{(Sinhb)^3}{Sinka + (Coshb)^3}\ (Coska) = CosK(a + b)$. Using the assumptions as potential barrier V_0 is infinite and the barrier width b is trailing to 0. You will get b value which is nearly 0, will then give you, sinh(Qb) is equal to Qb because we are taking the b tending to 0 and $(Coshb)^3$ equal to 1. Replace these values in the previous equation and you will get $\frac{q^2 + k^2}{2qk}\ (Sin(Ka) + Cos(Ka)) = Coska$. We know that $\frac{q^2 + k^2}{2qk} = \frac{mv^2}{qk\hbar^2}$. Replace this into equation 14 and you will get $\frac{m\,V_0 b}{\hbar^2\ k\ (sinka + coska)} = Cos(Ka)$. if you define this value as p what will you get you will have $\frac{pSinka}{ka + Coska} = Coska$ and p is defined here. Now, if you have this equation this must be satisfied for the solutions to wave equation to exist. Now, to have these values which would satisfy this equation you will find that only those values of ka are allowed for which the left hand side of the equation lies between +1 and -1. Can you have cos values which are more than 1 or less than -1? No, you have values ranging from ± 1 for cos. The other values of ka are not allowed and hence they will lie within a region that is called as the forbidden energy levels.

That is why you start seeing some forbidden energy bands because you cannot have coska having values more than +1 or less than -1. The energy spectrum of the electron hence would consist of alternate regions of allowed energy bands and forbidden bands. So, you will have some allowed energy bands and then you will have the forbidden energy bands. The width of the allowed energy band increases with k a or the energy. So, you will have a larger number of energy bands which are allowed.

The width of a particular energy band decreases with the increase in the value of p that is with increasing binding energy of the electron. So, if you have higher binding energy obviously the width of the band would be much higher. What is the way of changing the width of allowed and disallowed that is the forbidden energy bands? Obviously go on changing the value of p and you will have different forms of energies associated with lattice that is forming the crystal. If p is infinite for example what will happen? Sine ka is equal to 0 that means ka is equal to nΠ hence k would be nΠ by a and then you can plot the values and you will get the band structure. For example in this case for p is equal to 6Π.

If p is equal to 0 then what will happen? CosKa is equal to coska that means $K^2 = k^2$. What is the resultant? It means E = $\frac{P^2}{2m}$ the conventional model from where you started building the whole model. So, you have range from 0 to infinity and you get the two extremes. The relation infers that coska takes a specific value of each of the allowed energy value of E. Coska is an even periodic function with period 2Πn by a where n is any integer.

Hence, E is also an even periodic function of k. E is a periodic function of k with a period of $\frac{2\pi}{a}$. There exists a general relation and you will find that it is conventionally followed and you have a relation between E and k and it is commonly known as E-k schemes or E-k diagrams and that is what you may be hearing by different set of people. The general and convenient way of expressing these periodic functions, periodic lattices, the arrangement of atoms are in terms of the extended zone scheme or the reduced zone scheme. So, the notable points of the extended zone schemes are the solid lines in this figure.

So, you see the solid lines. represent the E-k relationship in the extended zone scheme. The corresponding dotted lines in the parabolic curve is associated with the free electron in the constant potential, but the solid lines are the cases where you get for nearly free electron. The E-k curve of the extended zone scheme is not continuous. So, you have cases where you see discontinuity and this is where this is happening at points where k is equal to plus minus $\frac{n\pi}{a}$. The allowed values of k defines the boundaries of the Brillouin zone.

The first Brillouin zone extends from $\frac{-\pi}{a}$ to $\frac{+\pi}{a}$, the second one from $\frac{\pi}{a}$ to 2 $\frac{\pi}{a}$ and $\frac{-\pi}{a}$ to $\frac{-2\pi}{a}$, the third one is from $\frac{-2\pi}{a}$ to minus $\frac{-3\pi}{a}$ on the left side and then $\frac{2\pi}{a}$ to $\frac{3\pi}{a}$ on the right side. So, for example, if you had taken a lattice, so I had taken and then I had drawn the potential.

So, I had drawn the potentials associated with a particular atom and chose a origin at this point. So, this is the origin.

Now, you had the E-k diagram which you saw. For free electron model what was the assumption that the electron can take any value. So, it can go and take any values and you have a potential which is like a parabola. But in the free electron model you said no, near the boundaries you will have discontinuities and you will have a gap in the middle. This is what you saw in the previous curve. Near the boundaries you will have forbidden values and these energies would not be allowed.

Now, suppose a second student takes the same example but chooses the origin here and then does all the calculation, plots everything and you have certain relations. Would those two values be different? Obviously, those values would not be different. Why? Because you can choose origin anywhere in the lattice. Now, what is the first Brillouin zone? The first Brillouin zone is the region which is defining the area which is belonging to me and only me, that means to a given atom and only to that atom.

So, the region around a given atom. If you find the relation between the second Brillouin zone, the third Brillouin zone, what is the value? You will find that the summation of the energies in the second Brillouin zone is equal to the energy of first Brillouin zone. So, you can then reduce the whole extended zone scheme to one curve and that is where the origin of reduced zone scheme comes from. So, if you move origin from one place to the other it does not matter, you will have the similar wave like variation in the potential and the first Brillouin zone carries all the information about the material and if you just take the information in the first Brillouin zone, you have the complete information about the material and that is why it is called as the reduced zone scheme. So, you can just reduce the complete information about a material in this first Brillouin zone and you will be able to understand about the complete material and that is why what you see in books, what you see in numericals, what you see in common figures published in papers, literatures is the reduced zone scheme and people do not report the extended zone scheme and it is limited to $\frac{-\pi}{a}$ to $\frac{+\pi}{a}$. The wave vector k belonging to this region is called the reduced wave vector and this is the energy band gap.

It is the energy band gap. Sometimes you would also read certain books and and you will find you will get a term which is called as flat band diagram. What is that? If you plot this curve and just I plot it in terms of certain bands. This continuous lines means what? This means that all the energy levels are allowed. Then you have a band gap and then you have another set of bands where the energy levels are allowed. This is the band gap, but if you look into what I have plotting, it is just like plates over plates.

So, it is a flat band. So, you can just deconvolute the complete information in terms of flat band diagrams. Flat band diagram, the reduced zone scheme or the extended zone scheme

give you the same information. What are those informations? Those informations are what are the available energy levels? If there is any gap between the two levels that is a level lower and the level higher and if there is a gap, what is the magnitude of this band gap? Based on the magnitude of this band gap, then you define the types of materials either as insulator or semiconductor. If these two bands are overlapping, then you have metals and if these two bands have certain gap, you can have conductors or insulators. So, you have seen that the nearly free electron model considers that electrons are not totally free, but it encounters the potential due to the presence of positive ion cores.

The nearly free electron model was able to explain the origin of band gap in solids which are semiconductors and insulators and why there were no band gaps in metals. And you have seen extended and reduced zone schemes carry the same information which were being mathematically obtained by solving the periodic function and the corresponding Schrodinger equations. These are the books which were followed to prepare the information that were presented in today's lecture and I thank you for attending lecture 2 of week 4.