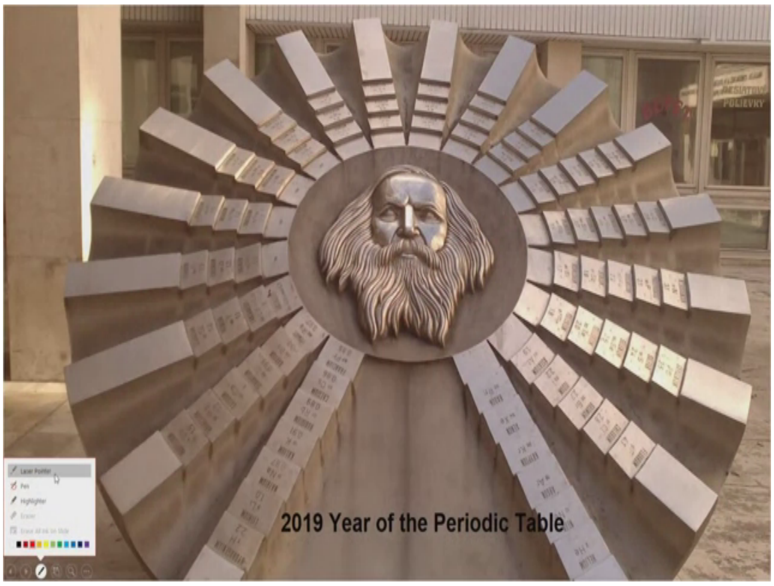
# Symmetry and Structure in the Solid State
Prof. T. N. Guru Row
Solid State and Structural Chemistry Unit
Indian Institute of Science, Bangalore

## Lecture – 51
## Data Reduction

So, as I was mentioning in the last class set of classes, maybe a couple of classes ago that this is the Year of Periodic table.

(Refer Slide Time: 00:37)



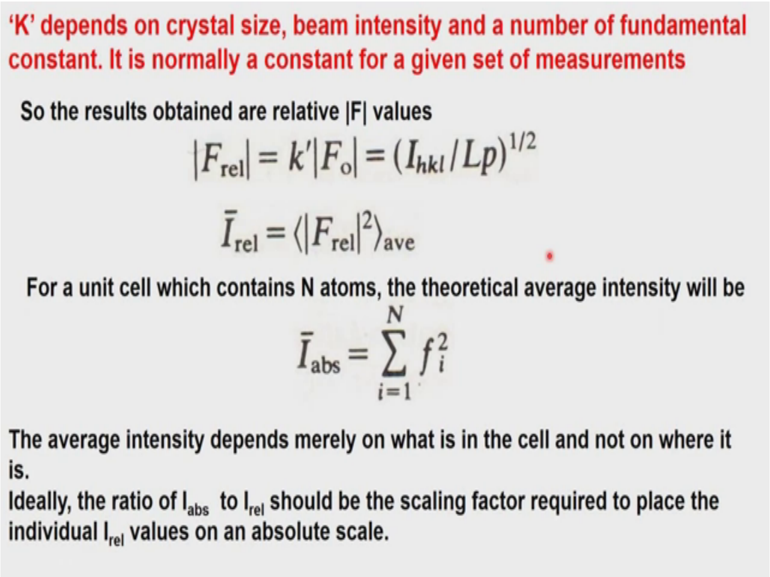
So, the United Nations has declared this as the Year of the Periodic Table and what is very crucial from the point of view of the periodic table is the way in which it is arranged, you see that Mandelif who is shown up here. This is the plot which is in front of the University of Bratislava in Slovenia.

And this is the one where we have the display of all the elements and you see that he also loved symmetry. So, all the elements have been arranged in a symmetric manner. You can see that all the rare earths are up here. A helium, neon, argon, krypton, xenon, radon and you see that the this is the completion of the sequence. Fluorine is up here and you can start from this side hydrogen lithium, sodium potassium, rubidium and so on. So, in fact, this is the way he preferred to probably have the arrangement. We do not know, but essentially now, periodic table appears in the current acceptable form I mean universal basis.

Now, most of the e m structures are made up of elements and it is the symmetry in these elements which actually carries through to the symmetry of diffraction. As we have already realised that we assume that the atoms are spherical and scattering that is coming from the atom is essentially from the electron density which surrounds these atoms and therefore all these atoms scatter and we know that the scattering factor falls off with respect to sin, theta by lambda. It also depends upon the value of z which is indicated in this periodic table.

So, there is a very close resemblance and very close association of what we get from an X-ray diffraction experiment; what we get as symmetry information with the arrangement of atoms in the periodic table. Towards the end of the course, we will see whether we can see some details of these individual elements and their properties from the point of view of doing what is known as quantum crystallography; that will be the last part of the discussion in this course.

(Refer Slide Time: 02:56)



'K' depends on crystal size, beam intensity and a number of fundamental constant. It is normally a constant for a given set of measurements

So the results obtained are relative |F| values

$$|F_{\mathrm{rel}}| = k'|F_{\mathrm{o}}| = (I_{hkl}/Lp)^{1/2}$$

$$\bar{I}_{\mathrm{rel}} = \langle |F_{\mathrm{rel}}|^2 \rangle_{\mathrm{ave}}$$

For a unit cell which contains N atoms, the theoretical average intensity will be

$$\bar{I}_{\mathrm{abs}} = \sum_{i=1}^{N} f_i^2$$

The average intensity depends merely on what is in the cell and not on where it is.

Ideally, the ratio of $I_{abs}$ to $I_{rel}$ should be the scaling factor required to place the individual $I_{rel}$ values on an absolute scale.

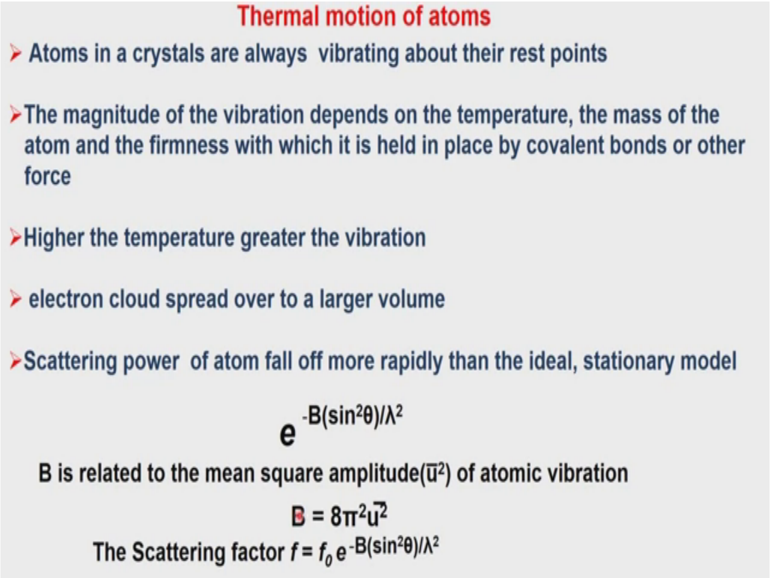
So, what we will do now is to go back to our the scale factor issue and the one once we have done the Lp corrections as is shown here the Lorentz polarization correction. We can therefore; get the relative values of the structure factors in terms of k prime and F naught. The value of k prime is the one which we have to determine and it depends upon crystal size, beam intensity and the number of fundamental constants. It is normally a constant for a given set of measurements. So, we can actually calculate the value of k.

So, we can write the what much we have the F relative we can write the I relative as in this form. The F of relative square averaged over different ranges of sin theta by lambda and for a unit cell can we when it contains n atoms, the theoretical average intensity is I abs is sum over i to n of f i square.

What it essentially tells us is that the average intensity depends merely on what is in the cell in terms of the electron density and not where it is located. So, the absolute value depends essentially on the sum of i equals 1 to n of f i of square. For all the atoms that are present inside the unit cell and so this gives us an idea of how we can go from I relative and get to the actual I abs through the use of determining. This is so called Structural scale factor. So, the ratio of I abs to I relative therefore should be the scaling factor as I mentioned and this is allowed.

Because of the fact that we have now all the electrons in the scattering material represented in terms of the absolute value of the intensity. So, as I mentioned in the previous class, wherever we measure the intensities whichever instrument we use, whichever radiation we use, we can actually put the observations on an absolute scale by calculating this. I absolute from the I relative values, but working out the right kind of scale factor. This process in which we do this analysis is due to Wilson and it is known as the Wilson Plot.

(Refer Slide Time: 05:07)



Thermal motion of atoms

- Atoms in a crystals are always vibrating about their rest points
- The magnitude of the vibration depends on the temperature, the mass of the atom and the firmness with which it is held in place by covalent bonds or other force
- Higher the temperature greater the vibration
- electron cloud spread over to a larger volume
- Scattering power of atom fall off more rapidly than the ideal, stationary model

$$e^{-B(\sin^2\theta)/\lambda^2}$$

B is related to the mean square amplitude($\overline{u}^2$) of atomic vibration

$$B = 8\pi^2\overline{u^2}$$

The Scattering factor $f = f_0 e^{-B(\sin^2\theta)/\lambda^2}$

Of course, these scattering factors are also affected by the thermal motion of the atoms; if the atoms are not at rest, they are vibrating in their rest positions and the magnitude of the vibration therefore, depends on the temperature, the mass of the atom, the firmness with which it is held in place by covalent bonds and other forces. So, higher the temperature greater is the vibration, electron cloud spreads over a larger volume and scattering power of the atom falls off rapidly with the ideal stationary model with the formula e minus B sin squared theta by lambda square representing the fall off.

So, B is called the B by (Refer Time: 05:51) factor and it is actually representing how far the atom is moving away from its mean position. The mean square amplitude of such a situation is u average square. Then, B is equal to 8 pi square u average square and this is known as the B by 1 factor. So, the scattering factor gets modified as a consequence; F becomes F naught; e to the minus B sin square theta by lambda square. So, the modifications that get introduced in F can be represented in the form of a behaviour of sin theta pi by lambda and the variations with respect to that.

(Refer Slide Time: 06:33)

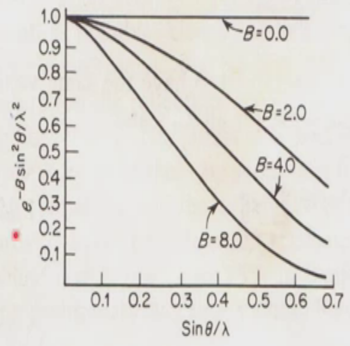


**Figure 7.4.** Temperature factor $e^{-B(\sin^2\theta)/\lambda^2}$ as a function of $(\sin\theta)/\lambda$.

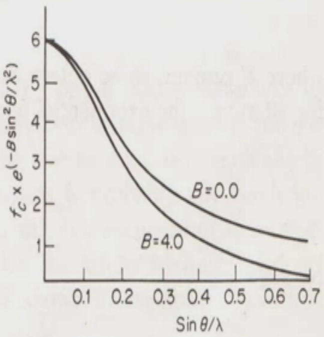


**Figure 7.5.** The product $f_0 e^{-B(\sin^2\theta)/\lambda^2}$ as a function of $(\sin\theta)/\lambda$.

For example, if we put now e minus B sin squared theta by lambda square on the y axis, plot it against sin theta by lambda. As the value of B increases, of course B equals 0 will be very constant. There is no thermal parameter associated with it, but that is not very ideal case which is impossible case. In fact, because the atoms are always moving they are not at rest any time. So, as the B value increases you notice that this value which is e minus B sin square theta by lambda, the so called thermal power factor. It influences the curve in such a way that we get this fall off. So, essentially what therefore, happens is if we now consider the scattering factor and multiply it by this thermal parameter, we see depending upon the revival of factor the scattering factor fall off f versus sin theta by lambda takes different asymptotic paths.

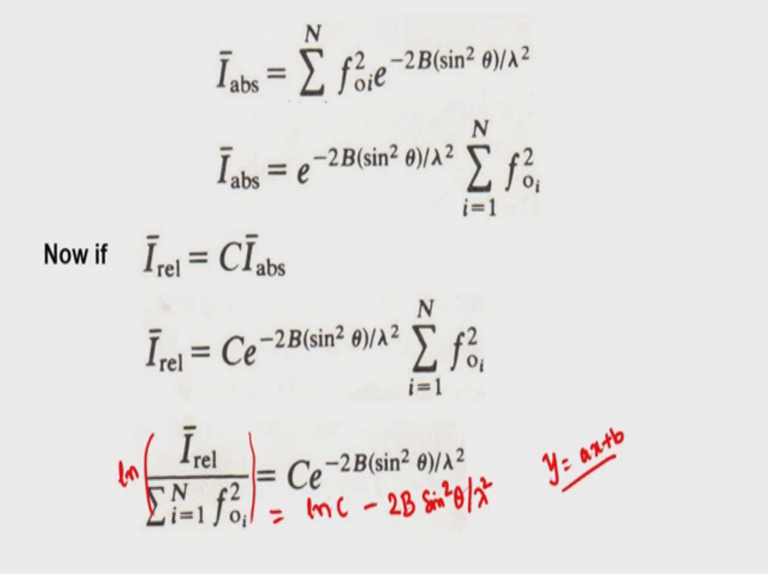
So, the extent to which the scattering occurs at somewhat an angle like 0.3 sin theta by lambda, you see that it gets reduced considerably, if the thermal parameter is large. So, the contribution from the atom now essentially has to be read out from this curve to various scattering angles. So, thus the scattering angle increases, we have reduction in the F value and the further reduction is caused by the presence of this so called thermal parameter. So, having kept this in mind and looking at these two diagrams, we see that if the temperature is a very crucial factor in our all our X-ray diffraction measurements.

So, unless otherwise there is a phase transition, it is suggested that the material under investigation, the data is better collected at low temperatures. The lower the temperature, the better is the approximation of F sin theta by lambda to ideality and as a consequence we will get the entire contribution to the diffraction condition. So, the diffraction; the reciprocal lattice point intensities will improve when we reduce the temperature and at the same time of course, there is a little bit of contraction of the lattice and this contraction of the lattice can be anticipated and evaluated. It is a straightforward proportionality factor and therefore, it is not a serious issue. The serious issue will be when there is a phase transition which we are not going to consider in this particular course, we are not discussing phase transitions. So, as we go down to lower temperatures.

So, it is generally suggested that the ideal temperature would be to collect the data at helium temperatures; something like 10 degree Kelvin or something. However, you know getting the 10 degree Kelvin is a very serious issue, we will have to spend a lot of money and as a consequence most of the data is suggested that it is collected at liquid nitrogen temperatures. Liquid nitrogen temperature, the boiling point is 77 so Kelvin. So, we collect the data around 100 K.

So, most of the data, it is suggested that one collects it at 100K to get better resolutions number one and secondly, more reflections as a consequence because there is a contraction of the lattice all right. But the reciprocal lattice even though it contracts more and more observable diffraction spots will appear because the reciprocal lattice volume depends upon the main unit cell. So, as a result we get the larger values associated with the scattering coefficient and therefore, the overall scattering improves considerably.

(Refer Slide Time: 10:05)



$$\bar{I}_{\text{abs}} = \sum^{N} f_{\text{o}i}^{2} e^{-2B(\sin^2\theta)/\lambda^2}$$

$$\bar{I}_{\text{abs}} = e^{-2B(\sin^2\theta)/\lambda^2} \sum_{i=1}^{N} f_{\text{o}_i}^{2}$$

Now if $\bar{I}_{\text{rel}} = C\bar{I}_{\text{abs}}$

$$\bar{I}_{\text{rel}} = Ce^{-2B(\sin^2\theta)/\lambda^2} \sum_{i=1}^{N} f_{\text{o}_i}^{2}$$

$$\ln\left(\frac{\bar{I}_{\text{rel}}}{\sum_{i=1}^{N} f_{\text{o}_i}^{2}}\right) = Ce^{-2B(\sin^2\theta)/\lambda^2}$$

$$= \ln C - 2B \sin^2\theta/\lambda^2$$

$\underline{y = ax+b}$

So, having noticed that aspect we will now see what happens when we want to put it on a proper scaling. So, the I abs which we are talking about the absolute intensity therefore, now not just depends upon the f i square, but it depends upon f which is at rest which we call it as 0 i square that is the atomic scattering factor with respect to the sin theta by lambda. The measure of the z value, but that gets modified by the exponential minus 2 B sin square theta by lambda square.

So, this modification of interaction of minus 2 B sin squared theta by lambda square will take into account, the thermal parameter. So, therefore, we get the I absolutely as equal to e to the minus B sin square theta by lambda square. We take it out; we will assume that all the atoms of the same thermal parameter. So, we take it out and minus 2 B sin square theta by lambda square comes up and then, we have the summation over i to n of f o i square in which now varies with respect to the nature of the atom. So, essentially what it tells us is that we now assume that all the atoms are vibrating at the same temperature.

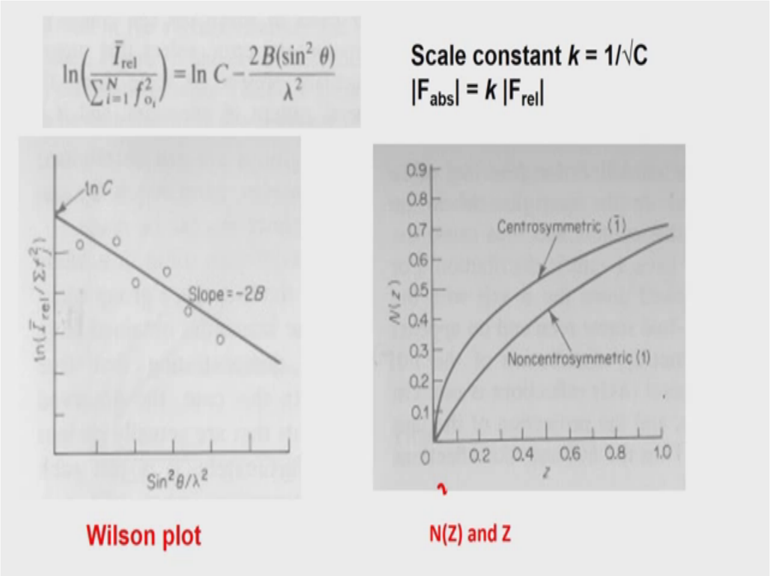
And therefore, the temperature effect is identical on all these scattering factors of all the atoms which are in this structure maybe there are 30 atoms, oxygen, nitrogen, chlorine, bromine, whatever. So, the fall off of f versus sin theta by lambda is fairly well defined in terms of the behaviour of f and then, we modify it with a thermal parameter which is now modifying every one of these fall off of f sin theta by lambda for every atom.

So, when we do this, we get a value of the absolute value of the intensity. Now, what we have measured from the experiment is the relative intensity. So, we can write relative intensity is equal to some constant times the absolute intensity. So, we can rewrite this expression, substitute for the absolute intensity as e to the minus 2 B sin squared theta by lambda square summation over i to n of f o i square.

So, this now, represents the relation or the value that we should get for the I relative which we have measured. So, this is the measured value, this is the absolute value we are looking for. So, if you take the ratio of I relative to the sum over i e to n of f o i square, this ratio is essentially equal to a constant C times an exponential factor. So, this, now exponential factor if you take logarithms on both sides. So, you take the logarithm of this quantity, logarithm of this and logarithm of that you can write this in the form of a straight line equation.

The reason is the following, we can write it like we can write the equation like y equals a x plus b. So, this will be like a straight line that is because we take the logarithms on both sides. So, this then will become log C minus two B sin squared theta by lambda square. So, this now, looks like y equals a x plus b right and therefore, we get a straight line and we plot that.

(Refer Slide Time: 13:41)



$$\ln\left(\frac{\bar{I}_{rel}}{\sum_{i=1}^{N} f_{o_i}^2}\right) = \ln C - \frac{2B(\sin^2\theta)}{\lambda^2}$$

Scale constant $k = 1/\sqrt{C}$

$|F_{abs}| = k\,|F_{rel}|$

**Wilson plot**

**N(Z) and Z**

Here is the expression given logarithm of I relative by sigma, i going from 1 to n and so, we get this expression and this is like y equals n x plus C equation. So, when we plot this logarithm value against sin squared theta by lambda squared, we will be ending up getting a straight line; the intercept is log C and the slope will be corresponding to 2 B.

So, we get therefore, the value of the scale constants which is k equals 1 over root of C. So, from log C, we can calculate this quantity and the thermal parameter, the overall thermal parameter associated with scattering can also be calculated by measuring the slope of this curve. This is the Wilson plot. I will spend a few minutes on this Wilson plot because this is something which is of crucial importance in structure analysis.

The nearness of the it is all not always ideally a straight line, it is nearly a straight line. It depends upon to some extent where the atoms are located inside the unit cell, even though we already told that the overall content of the scattering is what is there inside the unit cell rather than where it is, but it also matters that where it is; also matters. In the sense that suppose, we have several ring systems, then what happens is that we will not get a straight line. We will get some kind of a deviation from this straight line; somewhere around sin theta by lambda 0.18.

So, if we get a significant deviation at that point, we know that we have a ring system or a collection of 6-membered rings or 5-membered rings in the structure. The other thing is if the Wilson plot is really cranky, it is not looking like a straight line. It is never an ideal straight line, but if it is not mimicking a straight line, then we know there is something wrong in our data reduction. So, Wilson plot is not only used to find the scale and the overall thermal parameter, it is also in a way checks the quality of our diffraction measurements and also the data reduction procedure.

So, if we have done something wrong in the data reduction procedure, this will not be a straight line. So, Wilson plot therefore, is a very essential part and parcel it is a part and parcel of any structural determination. The advantage of doing this Wilson plot also is that we get a overall thermal parameter associated with the structure. So, we could associate this thermal parameter in the future improvement of the structure, when we do the so called refinement.

As I mentioned in the last class, there are two steps in which we determine the structure; the first step is to get the trial structure by trying to solve the phase problem, get enough information on the phases which though approximate are sufficient enough to give us the starting model and this starting model need not have the idealized geometries. So, what we do is we use the over determinacy in the data, the number of reflections which are available. Again, we put it back into the calculation.

So, we calculate the structure factor, we come take the observed structure factor and take the difference between the observed and the calculated and that quantity, the difference delta F is minimized and when we minimize that quantity, we get to a better and better structure. Geometrically as well as from the point of view of accounting for the entire electron density which is their inside the crystal structure.

So, the fit of the F naught and F c is very crucial and this is evaluated in terms of several figures of merit; one of them we discussed in the last class which is the r factor which is F naught minus F c modulus of F naught minus modulus of F t divided by modulus of F naught. So, these are some of the issues which will come out in when we do the Wilson plot. So, in it in a typical experiment, we collect the data, these data are measured the intensities are measured there on a relative scale and these relative intensities can be put on an absolute scale by this procedure which we just now discussed.

And based on the Wilson plot, we get the scale factor which converts the relative intensity to absolute intensities. So, that we can get this expression satisfied F abs equals k times F real and this is the rel is there real the relative intensity which we measured taking the square root of the intensity which we measure the relative structure factor.

So, the structure factor; absolute structure factor is what we need for our future phase solution and also for the Fourier transform which will then give us the electron density distribution. In this context, if we now plot this is known as an NZ plot. N is the fraction of the data which is represented in terms of the fraction Z. So, the number of data points with respect to the fraction Z of the collected data number; the number of data points. So, this is the fraction of data compared to the number of data points.

If we plot this; this gives us a curve which has the value of Z increases with respect to NZ. This is known as the NZ curve. This essentially tells us the nature of the distribution of intensities with respect to Non-centrosymmetric and with respect to Centrosymmetric systems. So, here is the situation where we can use this to a large extent, when we have no systematic absences which are depicted in the structure. And at the same time, we are looking for the presence of centric or a Non-centrosymmetric space group. The one of the easiest examples is from the triclinic system, when we have p 1 and p 1 bar these there are no systematic absences as we have discussed it thoroughly.

So, since there are no systematic absences, what we get is a situation where we analyze the intensity if this is the intensity distribution with respect to various sin theta by lambda ranges. So, the intensity distribution therefore, tells us that in the case of a non-centrosymmetric structure the curve which is shown below is generally followed and in the case of a central symmetric structure, the curve which is shown up is followed. So, one is an error function; the other is an exponential function.

So, as a consequence we can distinguish between p 1 and p 1 bar. So, it is also true that we can distinguish between p 2, p m and p 2 by m and by looking at certain statistics which are associated with projection reflections, it is possible to distinguish between p 2 and p m. Because p 2 and p m will give same system and there are no systematic absences so, also with p 2 by m. In other words, whenever we determine the structural features this way whether it is centric or non-centric, it is based upon the overall intensity distribution.

And all these approximations which we have used to find out the scale factor and thermal parameter evaluation is also incorporated into this. And therefore, it is not 100 percent reliable. So, the way we do that is either solve the structure as a central symmetric structure or preferably solve the structure in the non-centrosymmetric structure. Assume that it is a p 1 in fact any crystal structure take it from me, any crystal structure can be solved as p 1. Because if you solve the crystal structure as p 1 and then, determine where the molecules are, the molecules now show up the symmetry relationships between them.

So, if we can identify the symmetry relationships between the molecules, we can now downgrade these or upgrade the p 1 system to higher symmetry space group. So, it is possible therefore, to solve a structure of p 2 1 by C as a p 1, then we will find four molecules which will show the relationships, which will fit into the p 2 1 by C. So, in other words, the non-centrosymmetric structures in p 1 and p 1 bar, we can solve the structure in p 1, but then identify the 1 bar later on and put it across to p 1 bar. The indicated deviation from non-central symmetry to central symmetry, in this particular situation is only a guideline.

So, the intensity statistics is not a foolproof methodology to find out the whether it is centric or non-centric, it is a guideline to see whether probably one is centric the other is non-centric; we cannot conclude it. We can conclude only after the successful final refinement of the structure in the two space groups p 1 and p 1 bar. So, we just have to remember that when we come across such systems.

So, basically what we have done so far in the discussion now is to start from the fact that we have now measured the intensities. We have identified the coordinates h k l in reciprocal space so, we have the intensity measured. We collected it for the Lorenz and polarization effects which we discussed in the last discussion and having done that we now get the relative intensities. Now, we can also get the absolute intensity because the average intensity depends on what is there in the crystal; what is the total electron density that is there inside the cell, unit cell and not where it is and therefore, we have this expression.

Now, only thing now we have to worry about is the fact that they are all affected by the thermal motion. So, the thermal motion has to be introduced and that thermal motion which when we introduce introducing the concept of the (Refer Time: 23:35) factor, we find that the behavior of the F what is the sin theta product changes as this is shown here.

So, therefore, we modify the existing is sum over f of i square by the corresponding expression which will give us the I absolute.

Now, the I absolute therefore, can be written as equal to this quantity. We will take out the thermal parameter assuming that all our atoms in the structure are having the same similar thermal vibration factor. It is not the true case, but we can make an approximation because some of the light atoms may vibrate more, the heavy atoms may vibrate less; hydrogens may behave funnily we do not know how they are regions behave. So, all these things matter. But in principle we can have a overall thermal parameter associated with the diffraction condition.

So, then we now can calculate therefore, a quantity which is now the logarithm of I relative by sigma i f o i square which now represents the actual; I absolute value and then, we can get an expression which is something similar to a straight line which is y equals a x the a x plus b kind of a situation. So, we plot this quantity logarithm of this versus sin squared theta by lambda squared, we get a straight line. The intercept gives you the evaluation possible for the scale factor by means of which we put the relative structure factor onto an absolute scale. Only after putting it onto the absolute scale, we can attempt the structural determination protocol.

So, this is a must and I already listed out the importance of the other details which we can get from Wilson plot. The non centric distribution based on the intensity statistics is only an indicator, but that is a useful indicator to tell us which is which. So, in a practical situation when you are collecting data, the data is collected in such a way that we accumulate all the data of the h k i and the h k l; make a list of all that and then we do a statistical analysis. We calculate several parameters other than this essentially to see how the intensity statistics varies with respect to sin theta by lambda.

Now, the factor sin theta by lambda is very crucial because our atomic scattering factor falls off with respect to sin theta binomial. If data had not happened, then there was no problem because then we would have had a constant of the atomic scattering factor associated with one diffraction condition, of course modified by the thermal parameter.

But now it so happens that it falls off with respect to sin theta by lambda. So, what we do is, we will consider regions of equal value of F with respect to a range of sin theta by lambda. In other words, we assume we take the reciprocal lattice information, this limiting sphere and then divide the limiting sphere into a large number of concentric circles and these individual concentric circles now represent the reflections where the f versus sin theta by lambda is more or less slightly in like behaviour.

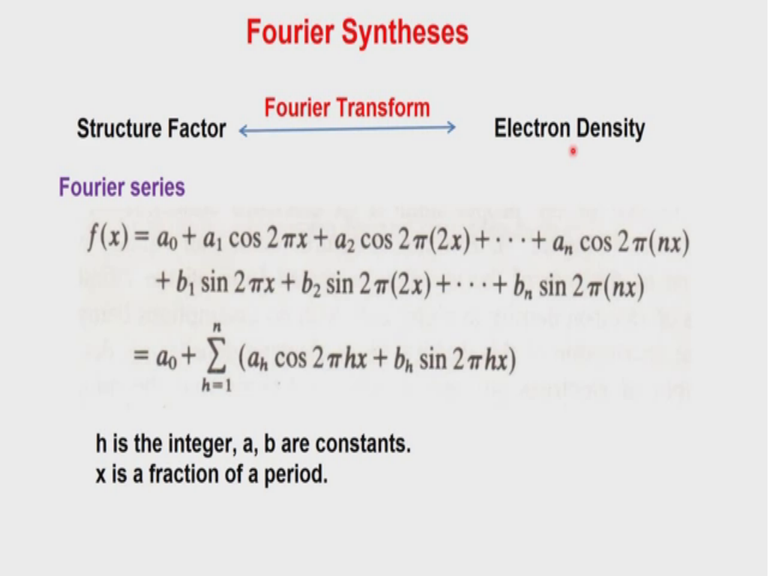
So, in other words if we go to the previous projection, you see that here if we take this region and go up to let us say 0.05, you see it is almost a straight line; then this part which is almost a straight line, that part which is almost a straight line. So, we actually take those regions where it is almost a straight line and that gives us the possibility of using this expression in a straightforward way. So, we therefore, take our limiting sphere, the reverse sphere related limiting sphere and divide this limiting sphere reflections which are appearing as h k l which are appearing as spots and these spots now form in different concentric circles, where the variation of F with respect to sin theta by lambda is minimal and that is very very important.

Otherwise the validity of the Wilson plot calculation here which is followed by the plot which we have done here becomes not very critical. So, this is now very important from the point of view of determining the scale factor and also the thermal parameter associated with it by looking at the slope which gives us minus 2 B, where B is the divisor effective.

So, at this stage, we can now go over to another aspect because now we what we are done is we have the experiment, we have now taken the; from that we have reduced the data after reducing the data, we have put the data on an absolute scale. So, what is our data? Our data is modulus of F abs absolute value of the modulus of the structure factor. So, what we therefore need obviously is, are the phases.

If we multiply these by the phases then we can do a Fourier transform to go to electron density. So, before we go to the Fourier transform, we should know something about what this Fourier series is all about and how the Fourier transform now brings in the issue of the presence of the variability associated with the relationship between F abs the absolute value of the structure factor multiplied by the phase and then the Fourier transformed taking us to the electron density. How does this come about? We should know a little bit of background information on what we call as Fourier series and at the same time, we should also know something about how we do this Fourier transforms. So, the next set of slides which we will consider will be the one which we will discuss now the concepts associated with the Fourier transforms.

(Refer Slide Time: 29:27)



## Fourier Syntheses

Structure Factor $\xleftrightarrow{\text{Fourier Transform}}$ Electron Density

Fourier series

$$
\begin{aligned}
f(x) &= a_0 + a_1 \cos 2\pi x + a_2 \cos 2\pi(2x) + \cdots + a_n \cos 2\pi(nx) \\
&\quad + b_1 \sin 2\pi x + b_2 \sin 2\pi(2x) + \cdots + b_n \sin 2\pi(nx) \\
&= a_0 + \sum_{h=1}^{n} (a_h \cos 2\pi hx + b_h \sin 2\pi hx)
\end{aligned}
$$

**h is the integer, a, b are constants.**
**x is a fraction of a period.**

So, we see here that we first look at the Fourier synthesis, with the point of view that the structure factor is a Fourier transform; it takes around two electron density.