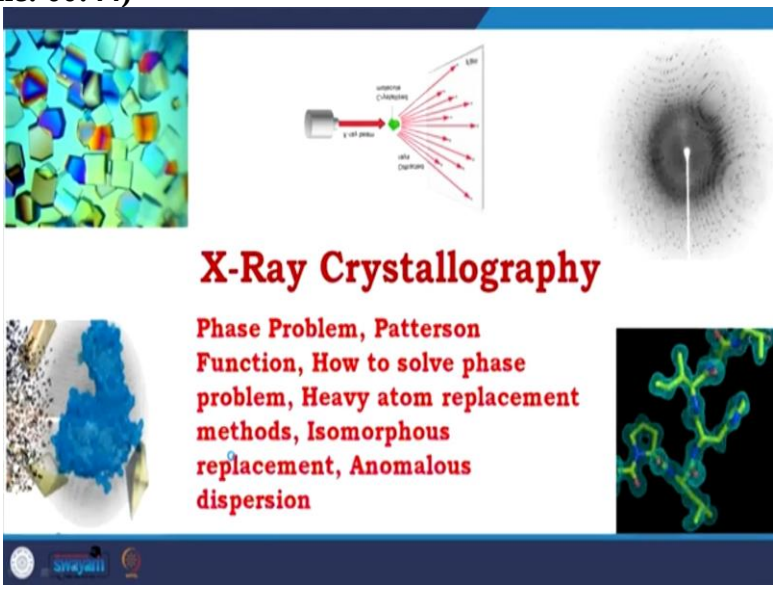
Structural Biology
Prof. Saugata Hazra
Department of Biotechnology
Indian Institute of Technology, Roorkee

Lecture - 23
X-Ray Crystallography - Phase Problem - Part 1

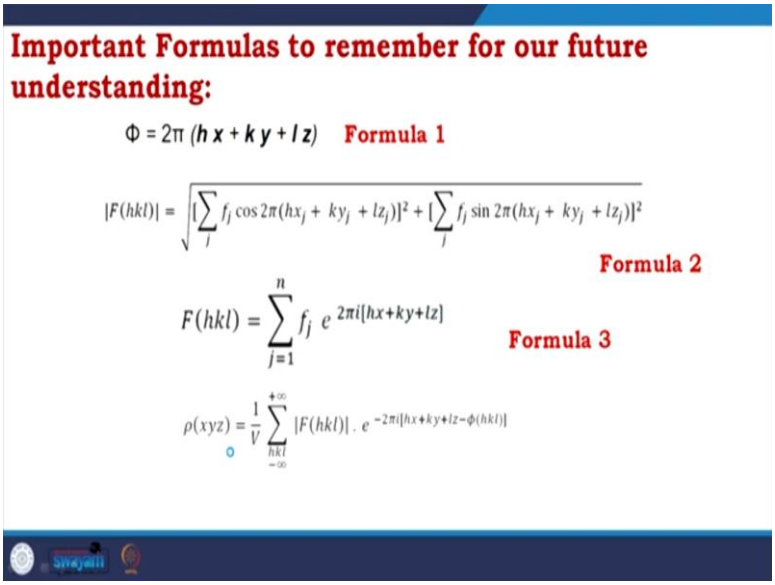
Hi everyone, welcome again to the course on structural biology. We are continuing with the structural biology techniques. We are talking about X-ray crystallography. And today, we are going to discuss the phase problem, the Patterson function.

(Refer Slide Time: 00:44)



How to solve the phase problem? Heavy atom replacement methods to solve phase problems and some techniques relating to heavy atom replacements are isomorphous replacement, anomalous dispersions, etc.

(Refer Slide Time: 01:00)



I will start where we end in the previous class, a few formulas we have picked up are important. One is the

$\varphi = 2\pi (h x + k y + l z)$

$F(hkl) = \sqrt{[\sum f_j \cos 2\pi (hx_j + ky_j + lz_j)]^2 + [\sum f_j \sin 2\pi (hx_j + ky_j + lz_j)]^2}$

These we call formula 2

$F(hkl) = \sum f_j e^{2\pi [hx + ky + lz]}$

$\sum j = 1$ to n for the n atom present

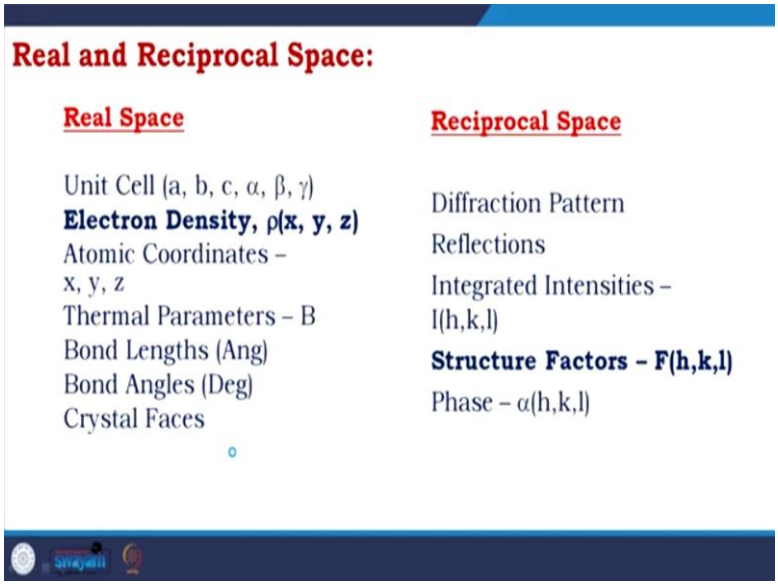
that is the formula 3 which is about structure factor.

The fourth formula is about electron density

$\rho(xyz) = 1 / V \sum | F(hkl) * e^{2\pi [hx + ky + lz - \varphi(hkl)]}$

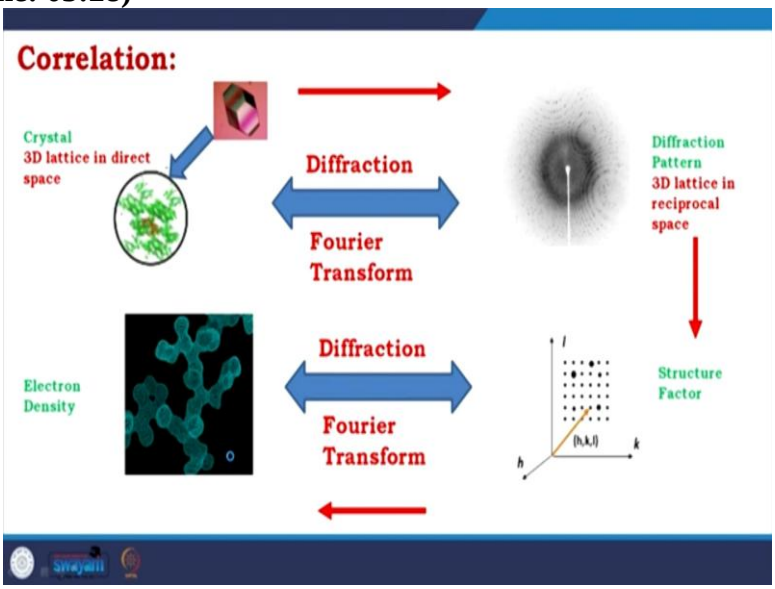
summation hkl - infinity to + infinity

(Refer Slide Time: 02:23)



I would also talk a little bit about the real and reciprocal spaces, which we discussed in detail while discussing crystal lattice. In the real space, we get information about unit cells a, b, c and alpha, beta, gamma, distance and angles, and electron density which I just talked about rho x, y, z. Atomic coordinates x, y, z. Thermal parameters B, which we will talk about later in detail. Bond lengths in angstrom, bond angles in degrees, crystal faces. From the reciprocal space, we will learn about diffraction patterns, reflections, and integrated intensities: I (h, k, l.) Structure factors F (h, k, l.) Phase alpha (h, k, l).

(Refer Slide Time: 03:28)

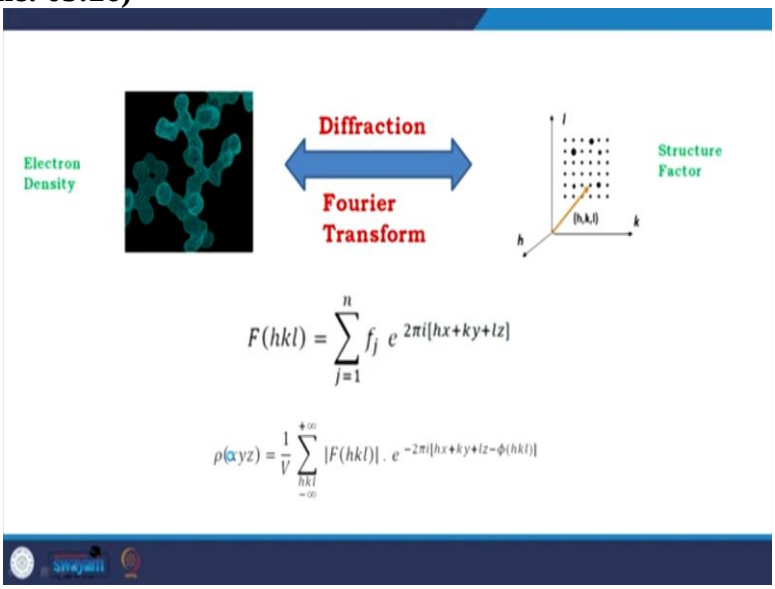


I would also like to talk about some correlations. So, throughout the data collection and processing story, we talked about crystal, unit cell, lattice. From a crystal, we get the information of 3D lattice in direct space. Whereas you know that this is a diffraction pattern, this represents a 3D lattice from a

diffraction pattern. So, the crystal represents a 3D lattice, and the diffraction pattern also represents a 3D lattice. Still, this is a 3D lattice in direct space, this is the 3D lattice in the reciprocal space, and the conversion from crystal to diffraction pattern is a diffraction experiment. Whereas when you want to do the theoretical calculation, this is the Fourier transform. Similarly, this is the electron density, and this is the structure factor.

These are also interconvertible in the experimental term, diffraction, whereas Fourier transforms in the theoretical term. If we now consider the entire thing, it starts by, definitely start by getting a pure protein, but then the process of crystallography starts after getting a crystal. So, from crystal to obtaining a diffraction pattern by diffraction, by applying X-ray then from diffraction pattern the intensity of spots would give us the structure factor.

(Refer Slide Time: 05:26)



And then from structure factor by inverse Fourier, we get electron density also the relation of electron density and structure factor, now we know the formula. In contrast, the structure factor

F (hkl) = $\sum f_j e^{2\pi [hx + ky + lz]}$

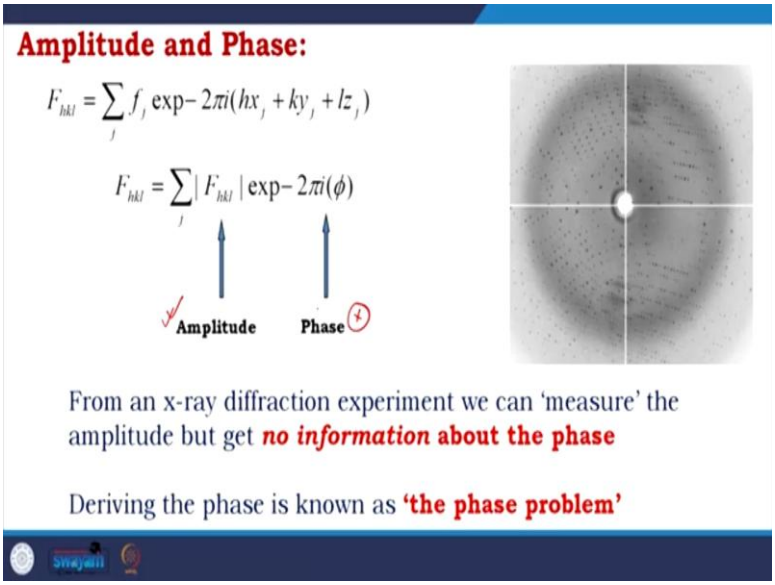
$\sum$ j = 1 to n for the n atom present

whereas the electron density

ρ (xyz) = 1 / V $\sum$ | F (hkl) * $e^{2\pi [hx + ky + lz - \varphi (hkl)]}$

summation hkl - infinity to + infinity

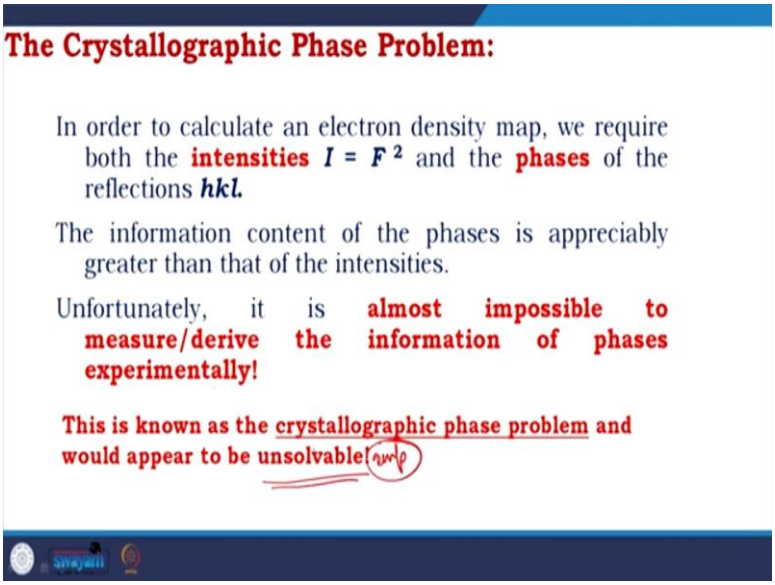
(Refer Slide Time: 06:17)



We know that F (hkl) has two terms, f_j and exponential value of $e^{2\pi [hx + ky + lz]}$. So, if you see that if j summation gives us the absolute value of the structure factor, which is amplitude which comes directly from the diffraction pattern, the other term is the phase. We can measure the amplitude from an X-ray diffraction experiment discussed in detail, but we get no information about the phase. So, we get this, but we do not get information.

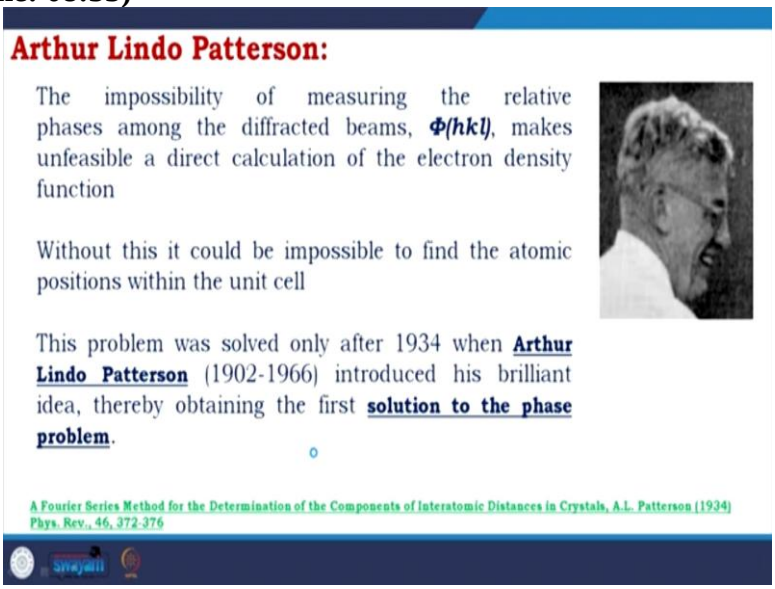
Is it we do not get it what else today? We will discuss that. But more importantly, as we are getting the amplitude and do not get the phase, we are unable to calculate electron density. This is called the phase problem. As you all know now, the phase problem is critical and famous in protein crystallography.

(Refer Slide Time: 07:52)



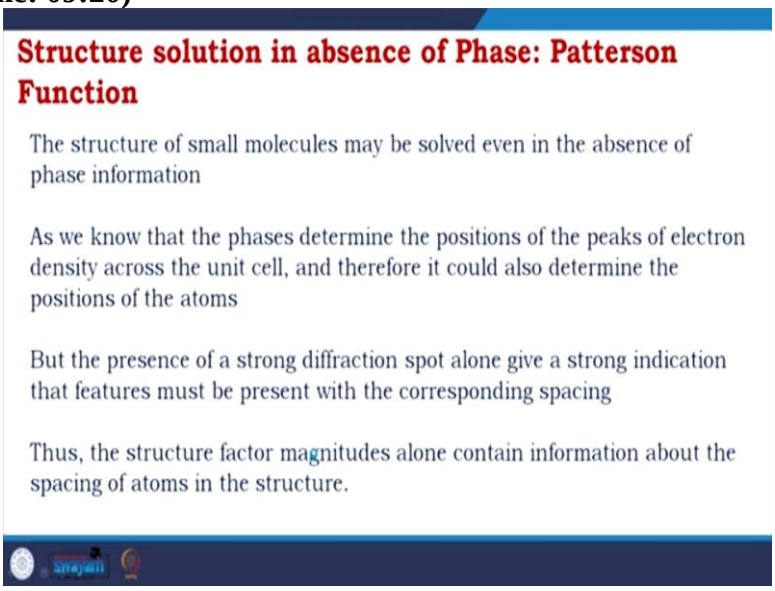
So, to calculate an electron density map, we require both the intensities and the phases of the reflection. The information content of the phases is appreciably greater than that of the intensities. Unfortunately, it is almost impossible to measure the information of the phases experimentally, and this is known as the crystallographic phase problem and appears to be unsolvable, very important.

(Refer Slide Time: 08:33)



And when we are talking about that, we have to introduce Arthur Lindo Patterson. So, the impossibility of measuring the relative phases among the diffracted beams φ (hkl) makes a direct calculation of the electron density function unfeasible. Without this, it could be impossible to find the atomic position within the unit cell. The problem was solved only after 1934 when Patterson introduced his brilliant idea, thereby obtaining the first solution to the phase problem. What is his idea? We are going to discuss this in detail.

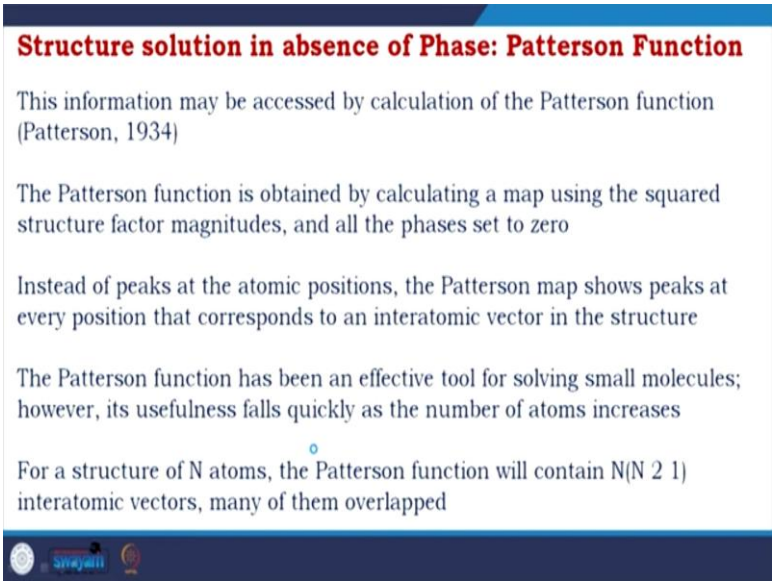
(Refer Slide Time: 09:20)



Structure solution in the absence of phase the Patterson function. The structure of small molecules may be solved even in the absence of phase information. As we know that the phases determine the positions of the peaks of electron density across the unit cell, it could also determine the atom's position. But the presence of a strong diffraction spot alone indicates that features must be present with the corresponding spacing.

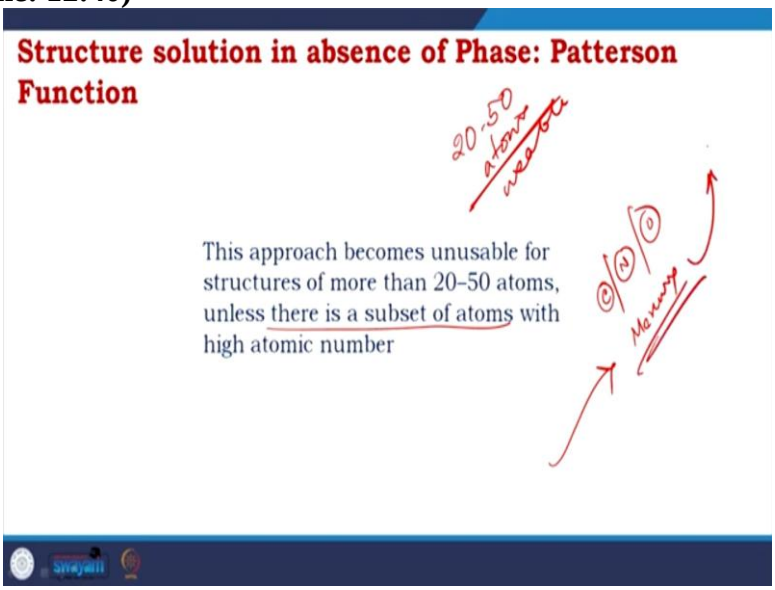
Thus the structure factor magnitudes alone contain information about the spacing of atoms in the structure. So, we have discussed but just to your memory refreshed, if you see the amplitude? The amplitude is if you look at the diffraction pattern, you see different spots dot. Some of the dots are very intense, whereas some are not so intense. So, you get the intensity i^2 which is proportional to the structure factor. But also, if you see, there is a pattern of the space difference, the difference between the spaces of the dots. When I was talking about the phase, the phase comes because of these differences. So, from the diffraction pattern, it could also, there is a chance that we could have gained some information about the phase.

(Refer Slide Time: 11:22)



This information may be accessed by calculation of the Patterson function. The Patterson function is obtained by calculating a map using the squared structure factor magnitude and zero phases. Instead of peaks at the atomic position, the Patterson map shows peaks at every position corresponding to an interatomic vector in the structure. The Patterson function has been an effective tool for solving small molecules. However, its usefulness falls quickly as the number of atoms increases. So, you may get to calculate it without getting the information about the phase, but more the atoms coming more complexity the method gets to be less effective. For a structure of N atoms, the Patterson function will contain N (N 2 1) interatomic vectors, many of which overlapped.

(Refer Slide Time: 12:40)

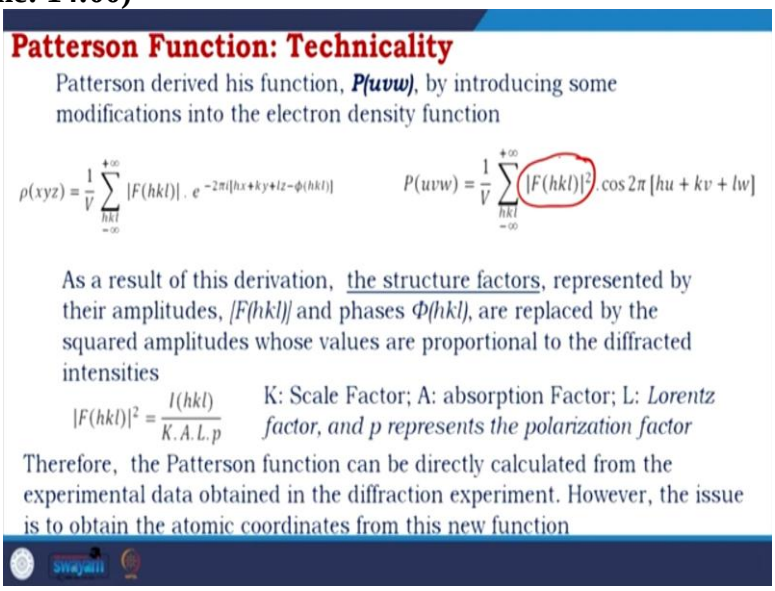


This approach becomes unusable for the structure of more than 20 to 50 atoms unless there is a subset of atoms with a high atomic number. So, two things, one for small molecule this is

good, you understand it, a second thing is coming which is great, and it is now for what in protein crystallography Patterson is used. If you see again, this approach becomes unusable for a structure of more than 20 to 50 atoms. So, for structures of 20 to 50 atoms usable, this is one part. The second part is unless there is a subset of atoms with a high atomic number.

So, in protein, if you remember and I will talk about this, there is carbon, nitrogen, and oxygen. Now, if you put mercury, the intensity is much higher, so you have a chance of using it; you get it.

(Refer Slide Time: 14:00)



Patterson derived his function P (u v w). So, now, this XYZ is changing to u v w by introducing some modifications into the electron density function. How? Let us take a look. So, this is

$\rho (xyz) = 1 / V \sum | F (hkl) * e^{2\pi i [hx + ky + lz - \varphi (hkl)]}$

summation hkl - infinity to + infinity

Patterson function

$\rho (uvw) = 1 / V \sum | F (hkl) |^2 * \cos 2\pi (hu + kv + lw)$

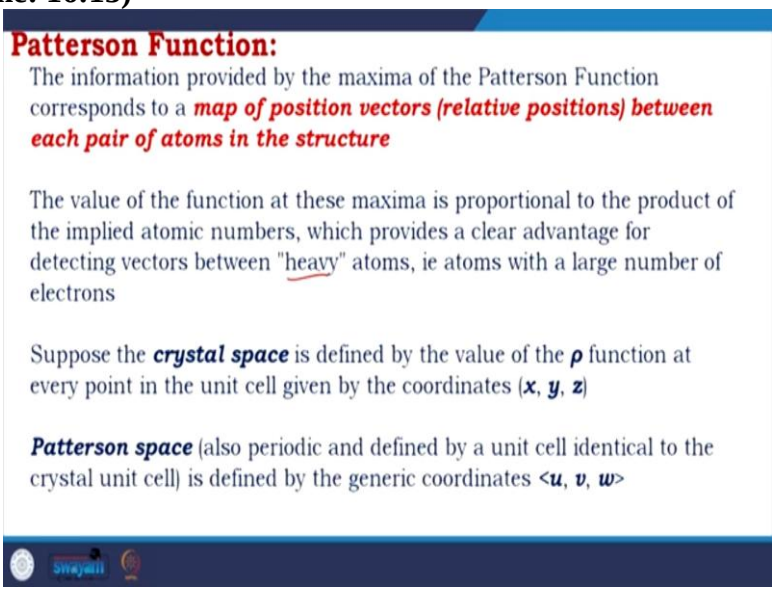
As a result of this derivation, the structure factors represented by their amplitudes and phases are replaced by the squared amplitudes whose values are proportional to their diffracted intensities. So, there is no phase and the amplitude is squared. And this comes to ultimately to the formula

$| F (hkl) |^2 = I (hkl) / K.A.L. p$

K is the scale factor, A is the absorption factor, L is the Lorentz factor and p represents the polarization factor.

So, the Patterson function can be directly calculated from the experimental data obtained in the diffraction experiment because now we do not see the involvement of a phase. However, the issue is to obtain the atomic coordinates from this new function.

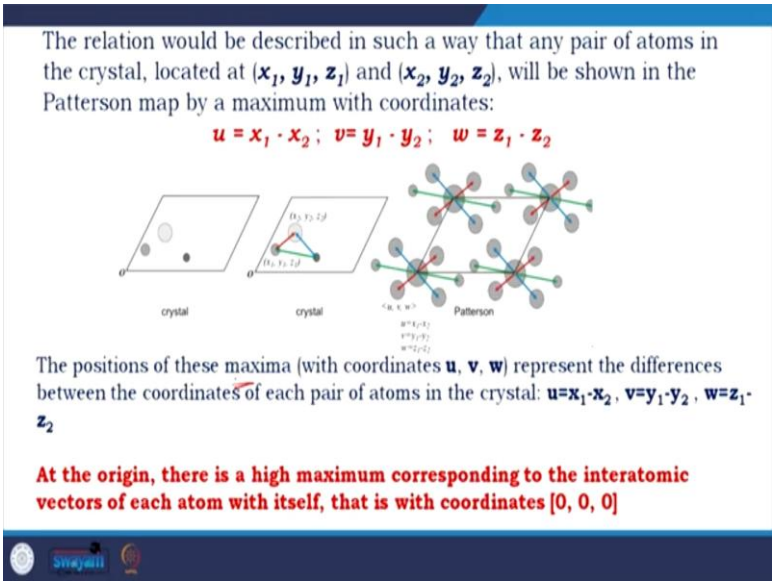
(Refer Slide Time: 16:13)



The information provided by the maxima of the Patterson function corresponds to a map of position vectors which are relative positions between each pair of atoms in the structure. The value of the function at these maxima is proportional to the product of the implied atomic numbers, which provides a clear advantage for detecting vectors between heavy atoms that is atomic with a large number of electrons.

Again, we will discuss this, but you could easily now understand that carbon, nitrogen, and oxygen. Now, suppose you include transition metals mercury, gold, platinum. In that case, you have a lot of electrons, which gives a lot of intensity because, in this method, intensity plays a huge role in the scattering, where you have the importance of the number of electrons. Suppose the value of rho (ρ) function defines the crystal space at every point in the unit cell given by the coordinates x, y, z. Patterson space which is also periodic and defined by a unit cell that is identical to the crystal unit cell, is defined by the generic coordinates (u, v, w)

(Refer Slide Time: 17:34)

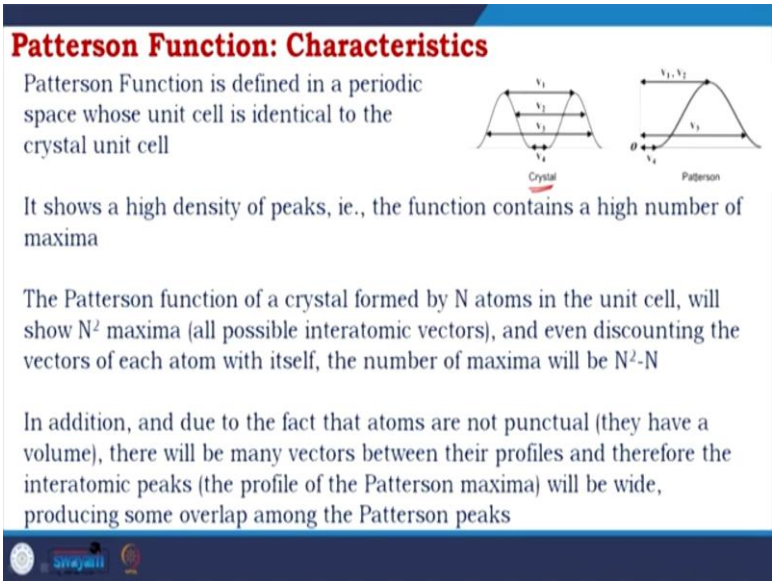


The relation would be described in such a way that any pair of atoms in the crystal, located at x_1, y_1, z_1, and x_2, y_2, z_2 will be shown in the Patterson map by a maximum with coordinates: $u = x_1 . x_2$; $v = y_1 . y_2$; $w = z1 - z_2$.

So, if you see the points x_1, y_1, z_1, and x_2, y_2, z_2 coordinates here, you will see how u v w is coming as we already talked about, $u = x_1 . x_2$; $v = y_1 . y_2$; $w = z1 - z_2$.

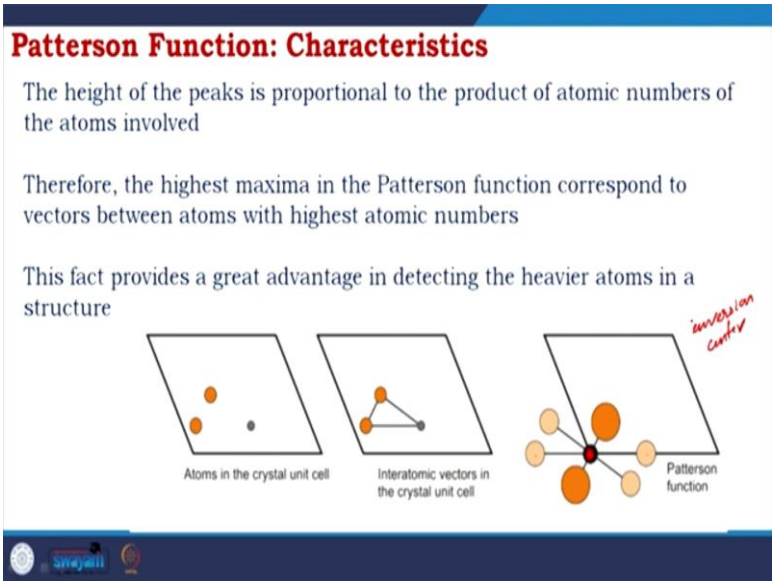
The positions of these maxima represent the differences between the coordinates of each pair of atoms in the crystal. At the origin, which is 0, 0, 0, there is a high maximum corresponding to the interatomic vector of each atom with itself.

(Refer Slide Time: 18:51)



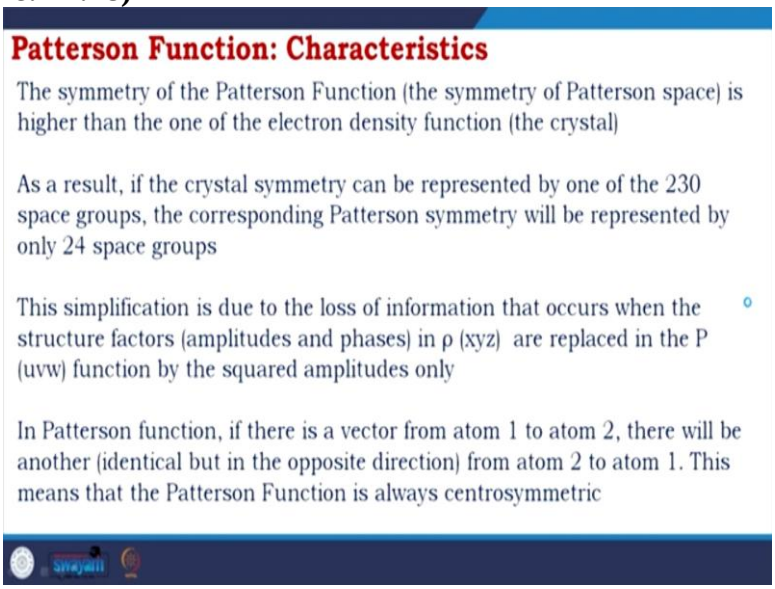
So, we now understand mathematically how u, v, w change up this equation and how it put a lot of importance on the diffraction amplitude, neglecting the phase. Patterson function is defined in a periodic space whose unit cell is identical to the crystal unit cell. It shows a high density of peaks because of the maxima. That is, the function contains a high number of maxima. The Patterson function of a crystal formed by an atom in the unit cell will show N square maxima all possible interatomic vectors and even discounting the vectors if you do that, of each atom with itself, the number would be N^2. N. In addition, because atoms are not punctual and have a volume, there will be many vectors between their profiles. Therefore, the interatomic peaks will be wide, producing some overlap among the Patterson peaks. You could see that it would be wide here, and there are overlaps if you compare the crystals, the peaks, and the Pattersons.

(Refer Slide Time: 20:14)



The height of the peaks is proportional to the product of atomic numbers of the atoms involved; therefore, the highest maxima in the Patterson function correspond to vectors between atoms with the highest atomic numbers. Again giving an advantage to the heavy atoms, this fact provides a great advantage in detecting the heavier atoms in a structure. So, if you see, this is the atom in the crystal unit cell. These are the interatomic vectors in the crystal unit cell, the Patterson function. Also, if you see, I talked about these in the symmetry topic. If you look at this point, everyone is the same and equal distance. So, this is an inversion center. You get an inversion center or center of the inversion I.

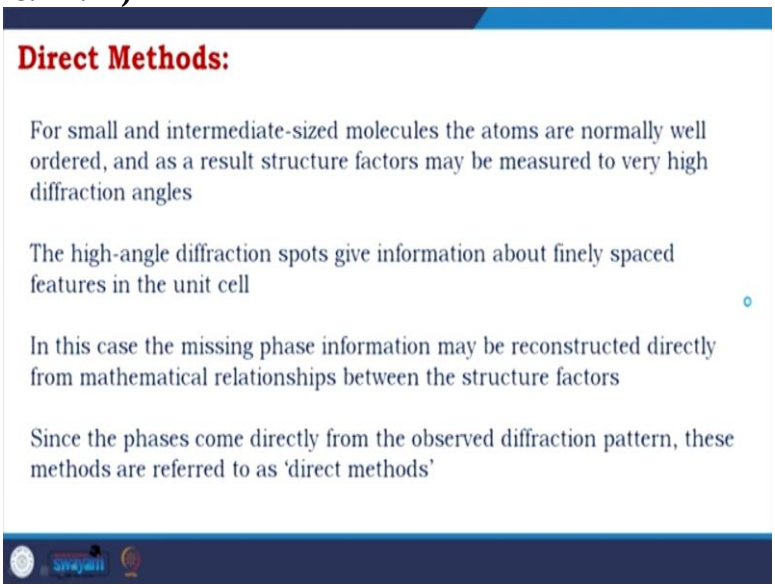
(Refer Slide Time: 21:18)



And it would be very important when we are talking about symmetry. The symmetry of the Paterson function or the symmetry of the Patterson space is higher than one of the electron density functions, the crystal. As a result, if the crystal symmetry can be represented by one

of the 230 space groups you already know, only 24 will represent the corresponding Patterson symmetry. This simplification is due to the loss of information when the structure factors in ρ (xyz) are replaced in the P (u v w) function by the squared amplitudes only. Also, in the Paterson function, if there is a vector from atom 1 to atom 2, there will be another vector identical but in the opposite direction from atom 2 to atom 1. This means that the Patterson function is always centrosymmetric.

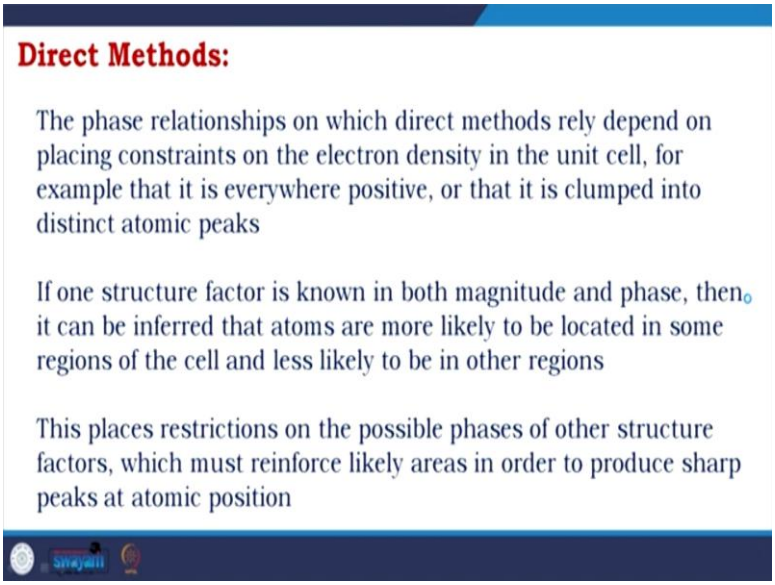
(Refer Slide Time: 22:22)



Talking about the direct methods to solve the phase problem for small and intermediate-sized molecules, the atoms are normally well ordered. As a result, structure factors may be measured to very high diffraction angles. The high angle diffraction spots give information about finely spaced features in the unit cell. The missing phase information may be reconstructed directly from mathematical relationships between the structure factors.

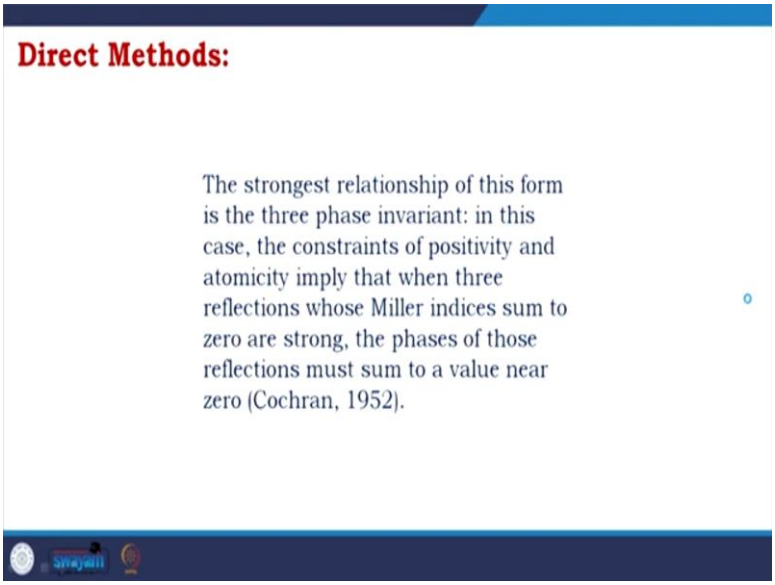
Let us think about NaCl crystal or something here. You know that you have only two atoms, sodium, and chlorine. And you know their nature and also know the space and all. So, you could reconstruct. They are less complex. Since the phases come directly from the observed diffraction pattern, these methods are direct methods.

(Refer Slide Time: 23:26)



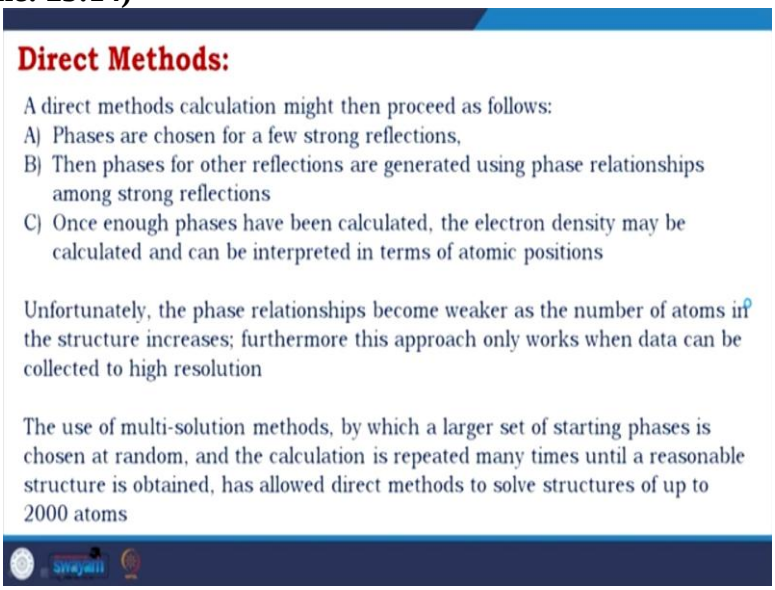
The phase relationships on which direct methods rely depend on placing constraints on the electron density in the unit cell. For example, it is everywhere positive or clumped into distinct atomic peaks. So, as you know, when you have a very complex protein, it is difficult to find them out. But when you have a small molecule, you know the information also, it is easier. If one structure factor is known in both magnitude and phase then, it can be inferred that atoms are more likely to be located in some regions of the cell and less likely to be in other regions. I was talking about the small molecules where the information is simpler, and it is more regular, it is more definitive and all. This places restrictions on other structure factors' possible phases, which must reinforce likely areas to produce sharp peaks at the atomic position. So, in one way, it is simpler, and they are as if you imagine, if there is NaCl as we are talking about, there would be more redundancy so that peak intensity would be more and all those things.

(Refer Slide Time: 24:50)



The strongest relationship of this form is the three-phase invariant. In this case, the constraints of positivity and atomicity imply that when three reflections whose mirror indices sum to zero are strong, the phases of those reflections must sum to a value near zero.

(Refer Slide Time: 25:14)

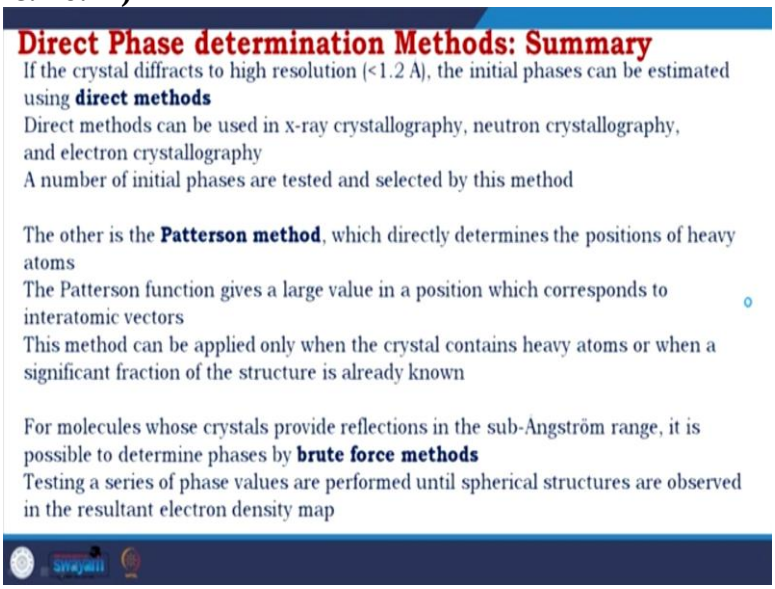


A direct method calculation might then proceed as follows. One, phases are chosen for a few strong reflections, then phases for other reflections are generated using phase relationships amongst strong reflections. Once enough phases have been calculated, the electron density may be calculated and interpreted in terms of atomic positions. So, first, the phases are chosen for strong reflections because they are easy to get. Phases for other reflections are generated using phase relationship the mathematical distance among strong reflections.

Once enough phases have been calculated, the electron density may be calculated and interpreted in terms of atomic position. Unfortunately, the phase relationship becomes weaker as the number of atoms in the structure increases. Furthermore, this approach only works when data can be collected to high resolution. So, the requirements are simpler compounds with fewer atoms and high resolution.

Using the multi-solution method, a larger set of starting phases is chosen at random. The calculation is repeated many times until a reasonable structure is obtained allowed direct methods to solve the structure of up to 2000 atoms.

(Refer Slide Time: 26:47)



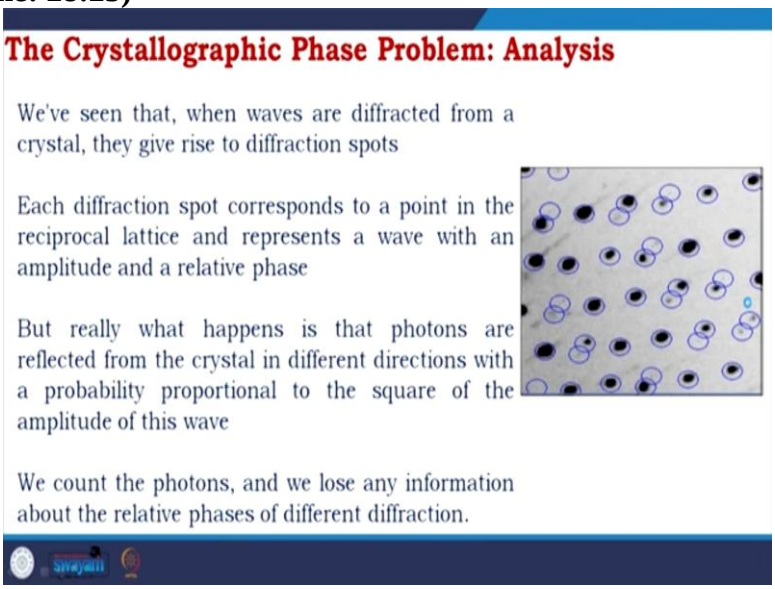
Let us get a summary of this. So, there are three methods: direct method, one is called Patterson function-based method, and another is called brute force method. If the crystal diffracts to a high resolution of less than 1.2 angstroms, the initial phases can be estimated using direct methods. Direct methods can be used in X-ray crystallography, neutron crystallography, and electron crystallography. Several initial phases are tested and selected by this method.

The other one is the Patterson method which directory determines the position of heavy atoms. The Patterson function gives a large value corresponding to interatomic vectors. This method can be applied only when the crystal contents, heavy atoms, or a significant structure fraction are already known.

For molecules whose crystals reflect the sub angstrom range, it is possible to determine phases by the brute force method.

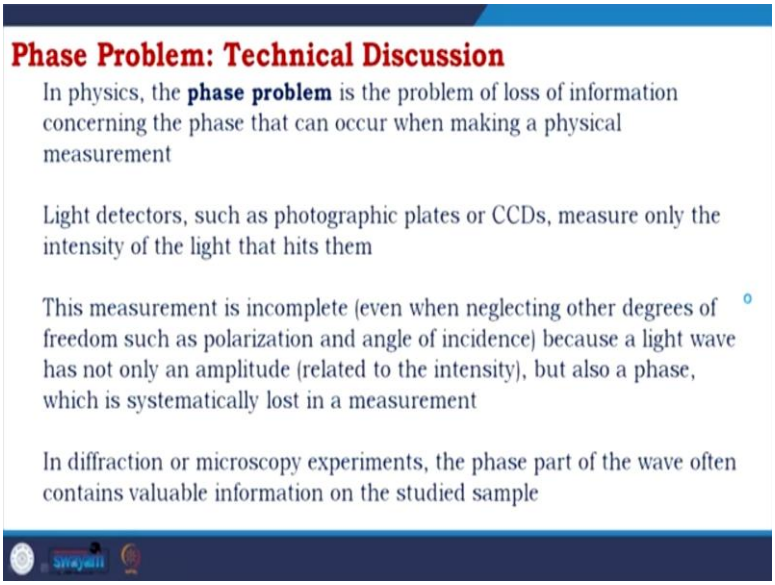
In this method, we have to test a series of phase values which will be performed until spherical structures are observed in the resultant electron density map. So, this is about the direct method and its possibilities, advantages, disadvantages.

(Refer Slide Time: 28:25)



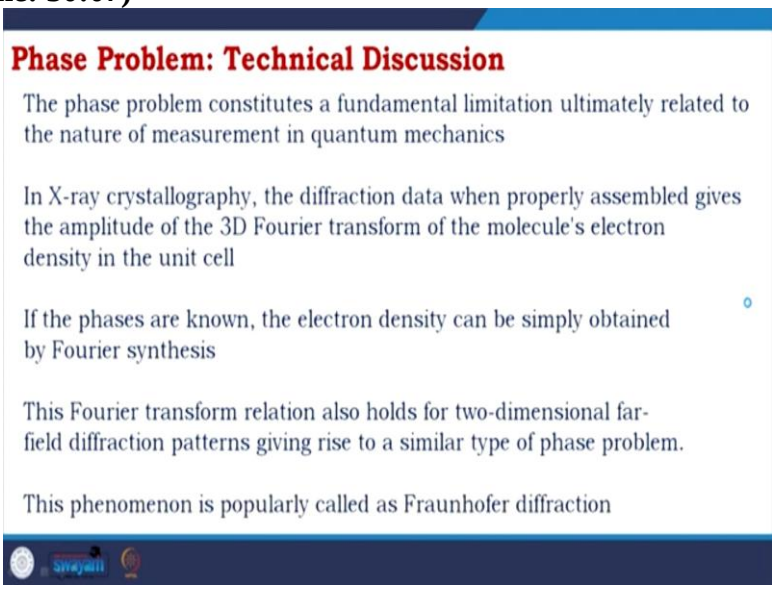
Again, come back to the fifth problem. Why did the first problem technically happen? We have seen that when waves are diffracted from a crystal, they give rise to diffraction spots. Each diffraction spot corresponds to a point in the reciprocal lattice and represents a wave with an amplitude and a relative phase. But in reality, what happens is that photons are reflected from the crystal in different directions, with a probability proportional to the square of the amplitude of this wave. We count the photons, and we lose any information about the relative phases of different diffractions. So, we count the photons, but we lose any information about the relative phases.

(Refer Slide Time: 29:18)



In physics, the phase problem is the loss of information concerning the phase when making a physical measurement. Light detectors, such as photographic plates or CCDs, measure only the intensity of the light that hits them. This measurement is incomplete, and when I say incomplete, we are neglecting other degrees of freedom, such as polarization and angle of incidence, because a light wave has an amplitude and a phase that is systematically lost in the measurement. In diffraction or microscopy experiments, the phase part of the wave often contains valuable information on the studied sample.

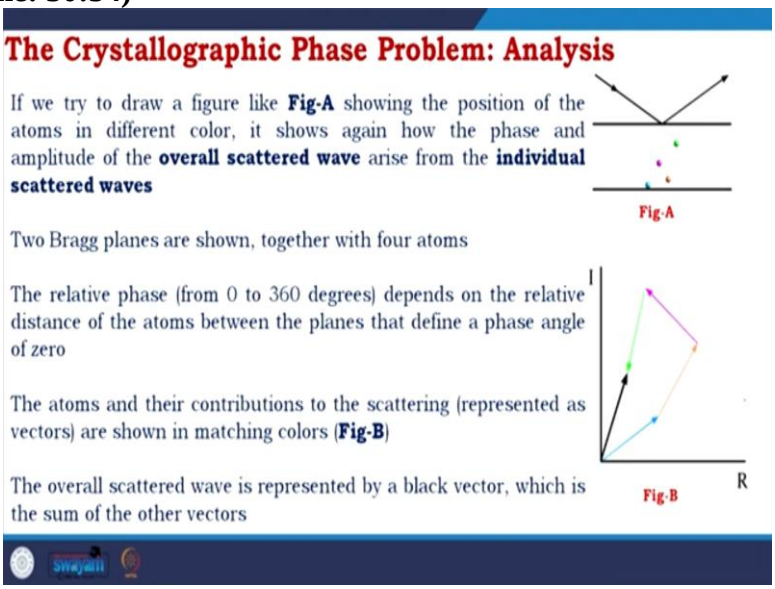
(Refer Slide Time: 30:07)



The phase problem constitutes a fundamental limitation ultimately related to the nature of the measurement in quantum mechanics. In X-ray crystallography, when properly assembled, the diffraction data gives the amplitude of that 3D Fourier transform of the molecule's electron density in the unit cell. If the phases are known, the electron density can be obtained by

Fourier synthesis. This Fourier transform relation also holds for 2-dimensional far-field diffraction patterns that lead to a similar phase problem. This 2-dimensional phenomenon is popularly called Fraunhofer diffraction.

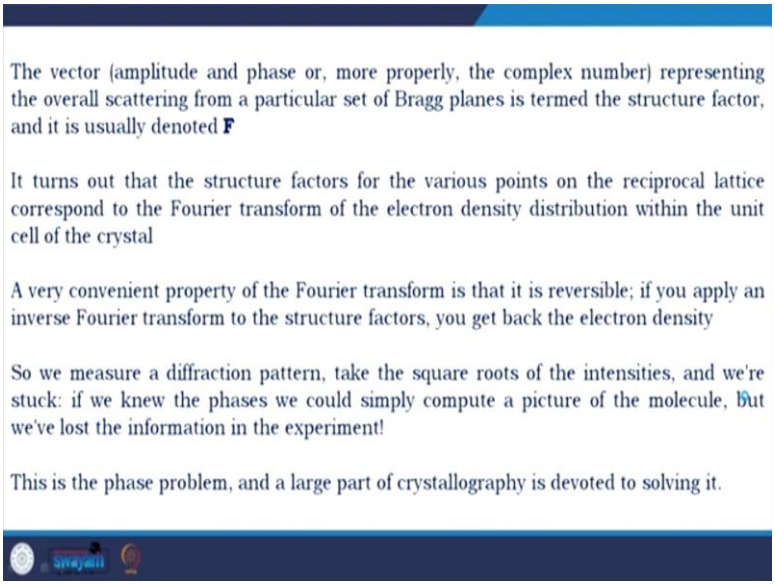
(Refer Slide Time: 30:54)



So, if you see figure A, there are four atoms and 2 Bragg planes. Now, if we try to draw a figure like a figure is swinging the position of the atoms in different colors, so four colors are showing four different atoms, it shows again how the phase and amplitude of the overall scattered wave arise from the individual scattered waves. So, every atom has its scattering. There are four atoms and 2 Bragg planes. The relative phase from 0 to 360 degrees depends on the relative distance of the atoms between the planes that define a phase angle of zero.

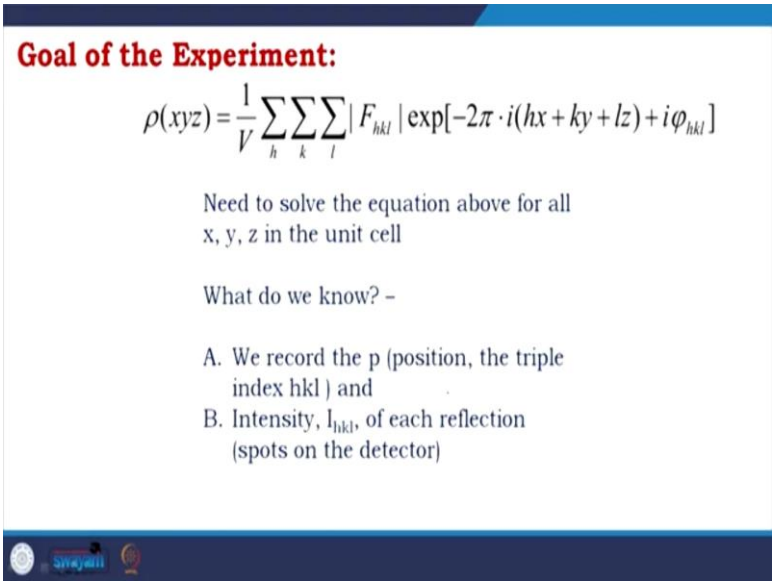
So, if you look at the plot up imaginary and real in figure B, the atoms and their contributions to the scattering are shown in matching colors. The overall scattered wave is represented by a black vector, the sum of the other factors.

(Refer Slide Time: 32:24)



The vector, amplitude, and phase or, more properly, the complex number representing the overall scattering from a particular set of Bragg planes is termed the structure factor, and it is usually denoted as F. It turns out that the structure factors for the various points on the reciprocal lattice correspond to the Fourier transform of the electron density distribution within the unit cell of the crystal. A very convenient property of the Fourier transform is that it is reversible if you apply an inverse Fourier transform to the structure factor, you get back the electron density which we have discussed earlier. So, we measure a diffraction pattern by taking the intensities' square root, and we are stuck. If we knew the phases, we could compute a picture of the molecule, but we have lost the information through our experiment. So, we talked about what the problem is now. Through theory, we try to understand that this is the phase problem, and a large part of the development of crystallography is devoted to solving it.

(Refer Slide Time: 33:47)



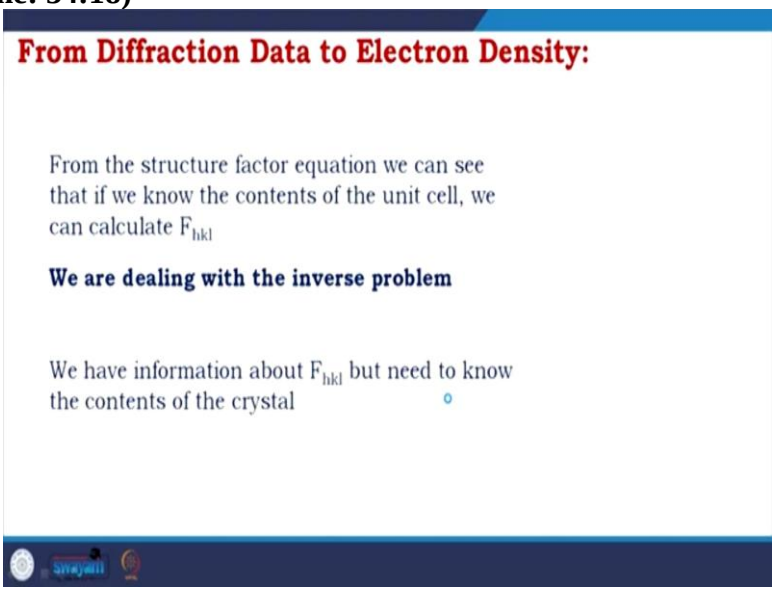
So, this electron density equation needs to solve the equation for all xyz in the unit cell.

ρ (x y z) = 1 / V $\sum\sum\sum$ F (hkl) $e^{-2\pi i (hx + ky + lz + \varphi_{hkl})}$

What do we know?

We record the ρ (the position, the triple index hkl), and intensity is I_{hkl}, proportional to each reflection's structure factor square. We get from the spots in the detector the diffraction pattern.

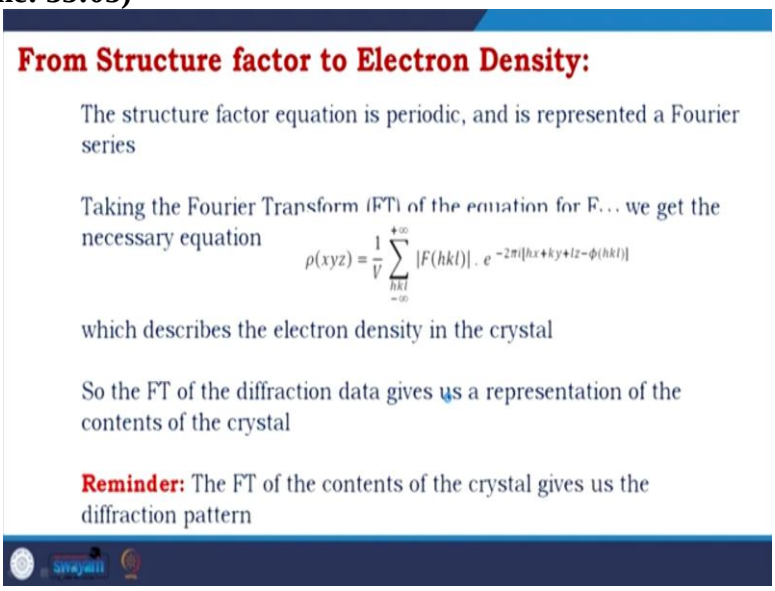
(Refer Slide Time: 34:18)



As we talked about, the measured intensities are proportional to the coefficient of the electron density equation. So, from the structure factor equation, we can see that if we know the contents of the unit cell, we can calculate F (hkl). But actually, we are dealing with the inverse problem. We have information about F (hkl). We got it from the intensity from the

spots but need to know the content of the crystal. We have F (hkl), and we want to get the electron density which will help us spot the atoms to understand what is there in the crystal.

(Refer Slide Time: 35:05)



The structure factor equation is periodic and is represented as a Fourier series. Taking the Fourier transform of the equation for F (hkl), we get the necessary equation.

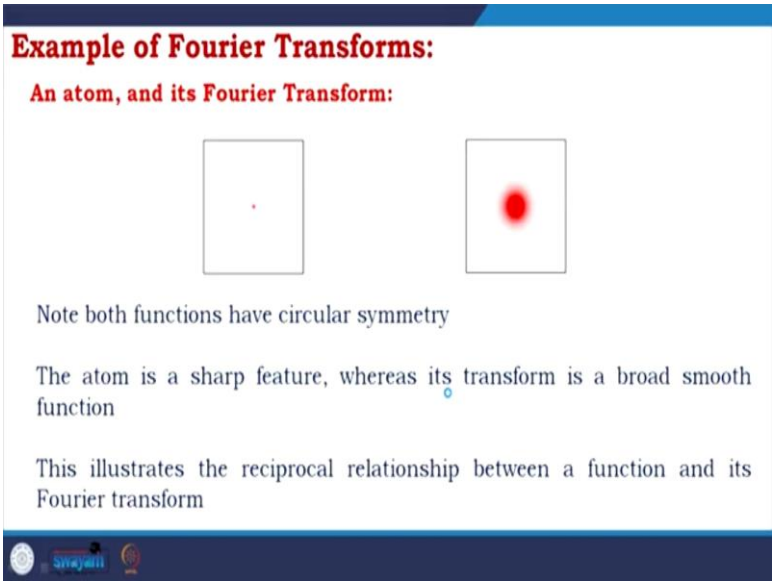
ρ (xyz) = 1 / V $\sum$ | F (hkl) * e $^{2\pi i [hx + ky + lz - \varphi (hkl)]}$

summation hkl - infinity to + infinity

This is the electron density equation that describes the electron density of the crystal.

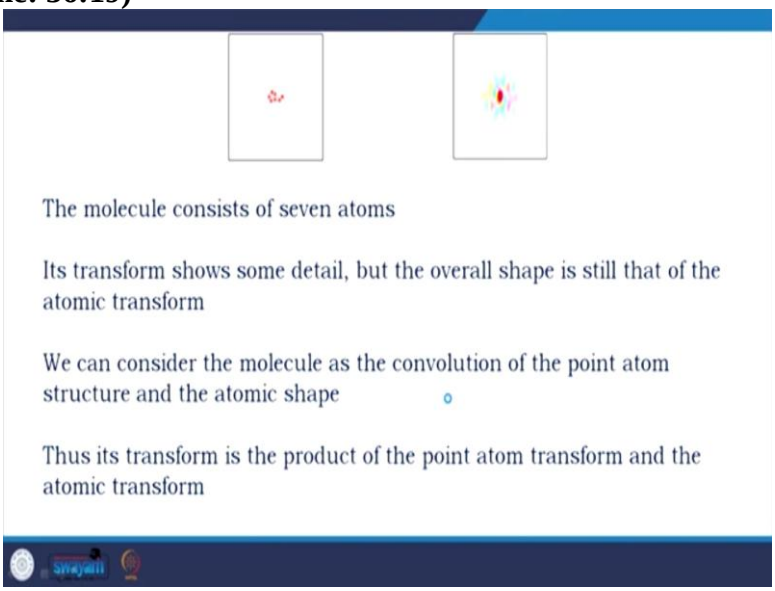
So, the Fourier transform of the diffraction data gives us a representation of the crystal's contents. And note that the Fourier transform of the content of the crystal gives us the diffraction pattern.

(Refer Slide Time: 35:48)



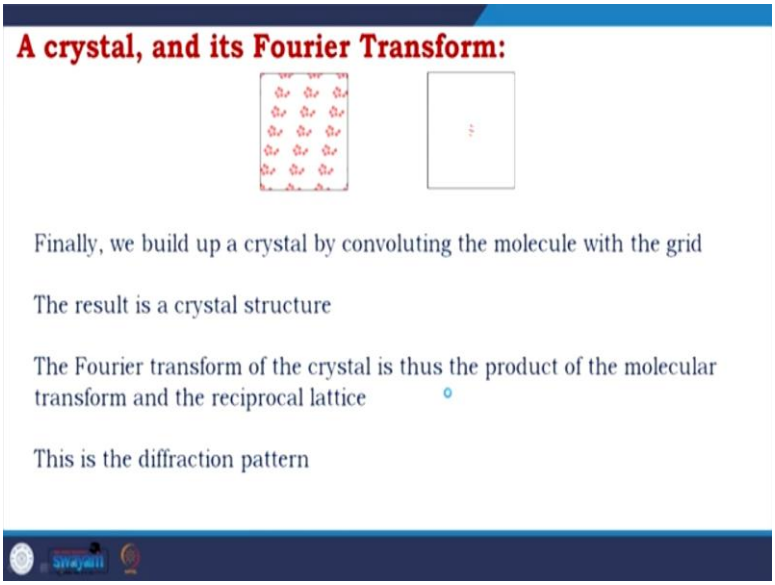
So, let us take an example of an atom and its Fourier transform. So, if this is an atom, the Fourier transform is like that. So, both functions have circular symmetry. If you see, the atom is a sharp feature, whereas its transform is a broad, smooth function. This illustrates the reciprocal relationship between a function and its Fourier transform.

(Refer Slide Time: 36:19)



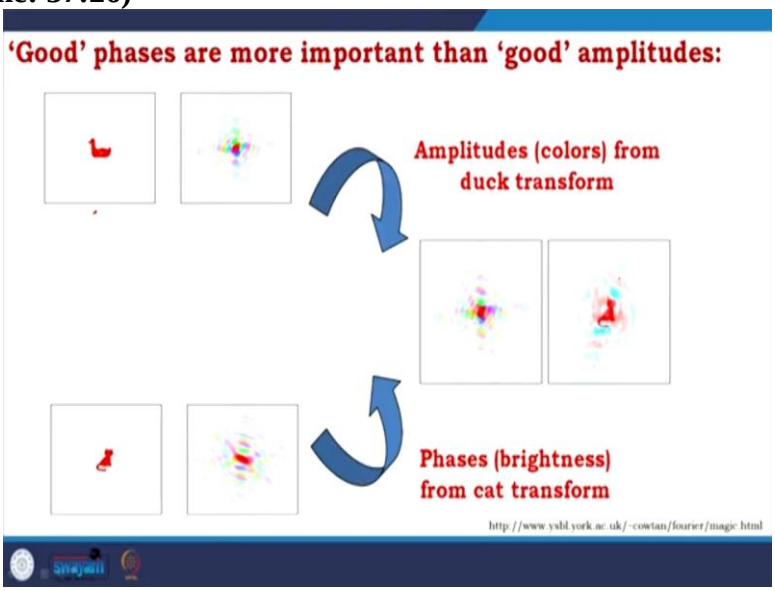
Now we take a bigger molecule, consisting of 7 atoms, and it is Fourier transform shows some detail, but the overall shape is still that of the atomic transform. We can consider the molecule as the convolution of the point atom structure and the atomic shape. Thus its transform is the product of the point atom transform and the atomic transform.

(Refer Slide Time: 36:56)



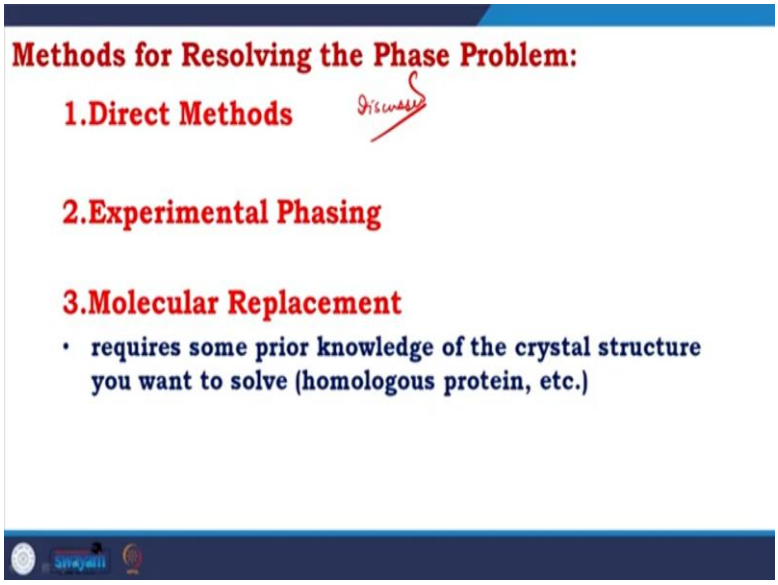
Now we are going to crystal, and we start getting a pattern. Finally, we build up a crystal by convoluting the molecule with the grid. The result is a crystal structure. The Fourier transform of the crystal is thus the product of the molecular transform and the reciprocal lattice. We also talked about this and this is the diffraction pattern.

(Refer Slide Time: 37:26)



So, now, if you have pictures with phase and amplitude, if we take amplitude or color of a picture from the duck transform, we get phases brightness from the cat transform. You see that you could see the cat, which means that good phases are more important than good amplitude.

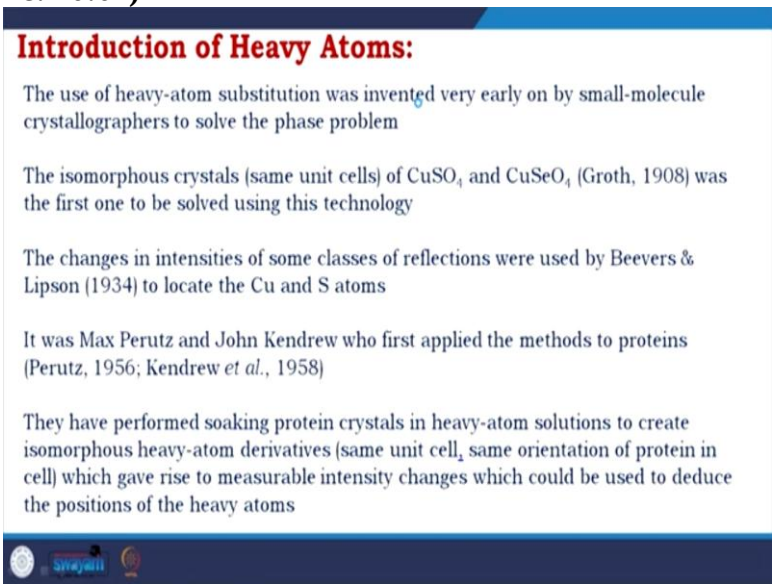
(Refer Slide Time: 38:08)



So, what are the methods to solve the phase problem? The direct method we have discussed already. Another two categories are experimental Phasing. We will use heavy atoms. The third one is the molecular replacement, which requires prior knowledge of the crystal structure you want to solve.

Molecular replacement is a hybrid technique taking the concept of theory and experiment together. It is a great contribution from Michael Rossman, and we will discuss these and their effect and role in changing protein crystallography in detail.

(Refer Slide Time: 40:02)

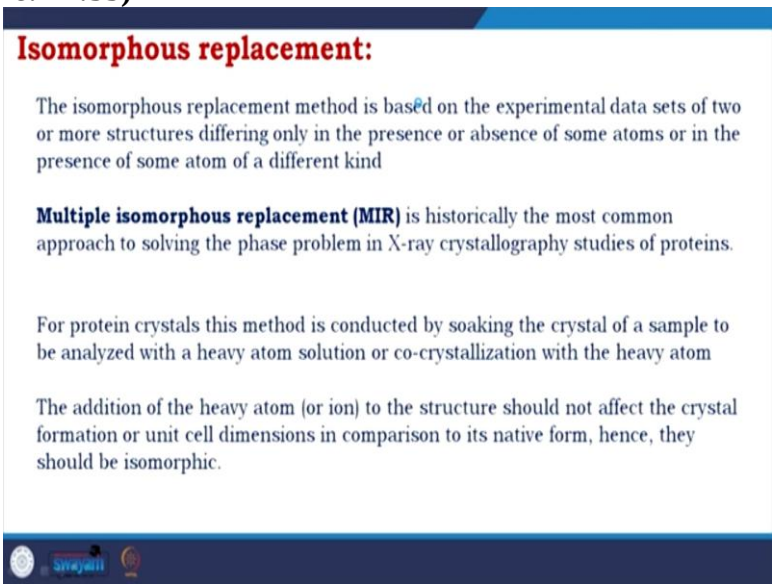


Small-molecule crystallographers invented heavy atoms substitution very early on to solve the phase problem. The isomorphous crystals, which are the same type as CuSo, and CuSeO4, were the first to be solved using that technology of isomorphous replacement.

Beevers and Lipson used the changes in intensities of some reflection classes to locate the copper and sulfur atoms.

Max Perutz and John Kendrew first applied the methods to proteins. They have performed soaking protein crystals in heavy atoms solutions to create isomorphous heavy atomic derivatives, same unit cell, the same orientation of protein in the cell, which gives rise to measurable intensity changes which could be used to deduce the position of heavy atoms.

(Refer Slide Time: 41:35)



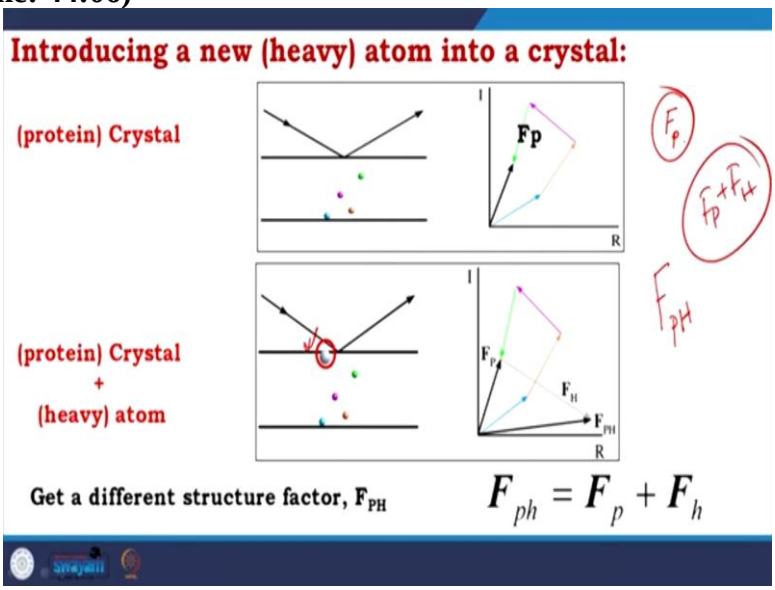
What is the isomorphous replacement? The isomorphous replacement method is based on the experimental data sets of 2 or more structures differing only in the presence or absence of some atoms or in some atom of a different kind. Multiple isomorphous replacement or MIR is historically the most common approach to solving the phase problem in X-ray crystallography studies of protein.

For protein crystals, this method is conducted by soaking the crystal of a sample to be analyzed with a heavy atom solution or co-crystallization with the heavy atom. So, I talked about it, but if you do not remember, we could soak a small molecule or a heavy metal into it when we get a crystal. Is this a challenging process? Yes, it is a very challenging process that I will discuss a little bit in the next class.

But this is probably the only way the other alternative is when you are growing a crystal you remember the drops we set up hanging drop, sitting drop, while we are doing the screening we add heavy atoms there. So, the first one where after the crystal is formed, you do the process is called soaking, you soak it. You have a heavy metal solution, and you put the crystal there. So, the crystal could soak a heavy metal.

Another one is co-crystallization. You add the heavy metal inside the protein. So they could crystallize together. The addition of the heavy atom of iron to the structure should not affect the crystal formation or unit cell dimension compared to its native form. Hence they should be isomorphic. So, if the effect you cannot continue with this. How to know? Again we will discuss this in the next class.

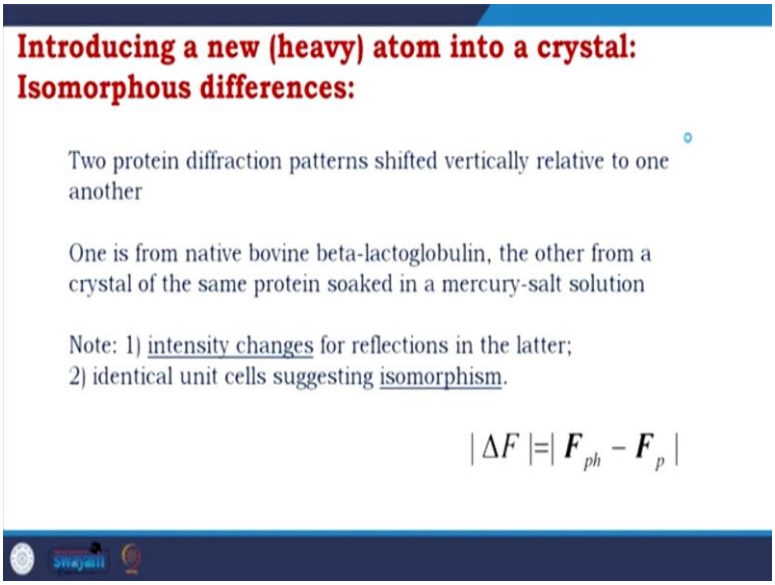
(Refer Slide Time: 44:06)



There are four atoms they contribute, and the black vector represents the sum. So, these represent the structure factor of the protein, Fp. When you add the heavy atom, here is the heavy atom. What do you get? You get Fp, and you get Fh. You get a different structure factor

$F_{ph} = F_p + F_h$

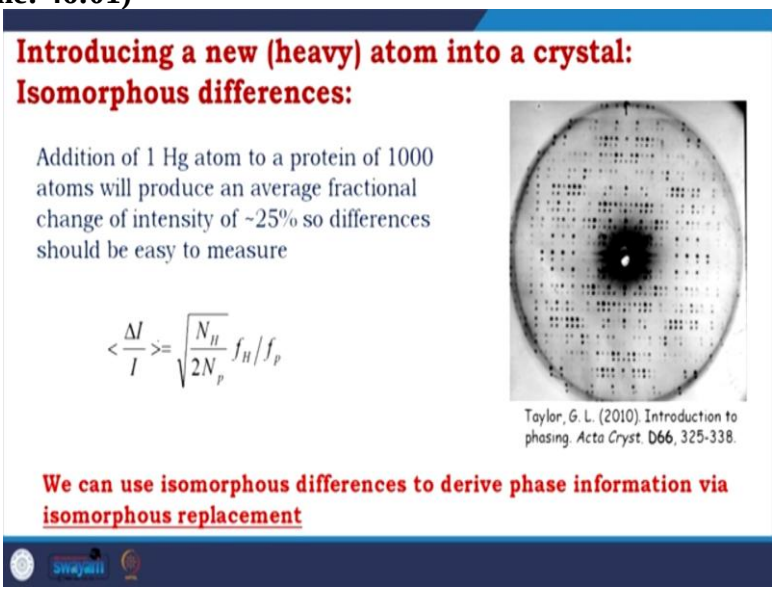
(Refer Slide Time: 45:16)



So, introducing a heavy atom into a crystal is isomorphous differences, so two protein diffraction patterns shifted vertically relative to one another. One is from native bovine beta-lactoglobulin and the other from a crystal of the same proteins soaked in a mercury salt solution. So, as heavy metal intensity changes for reflection in the latter, identical unit cells suggest isomorphism.

$|\Delta F| = |F_{ph} - F_p|$

(Refer Slide Time: 46:01)

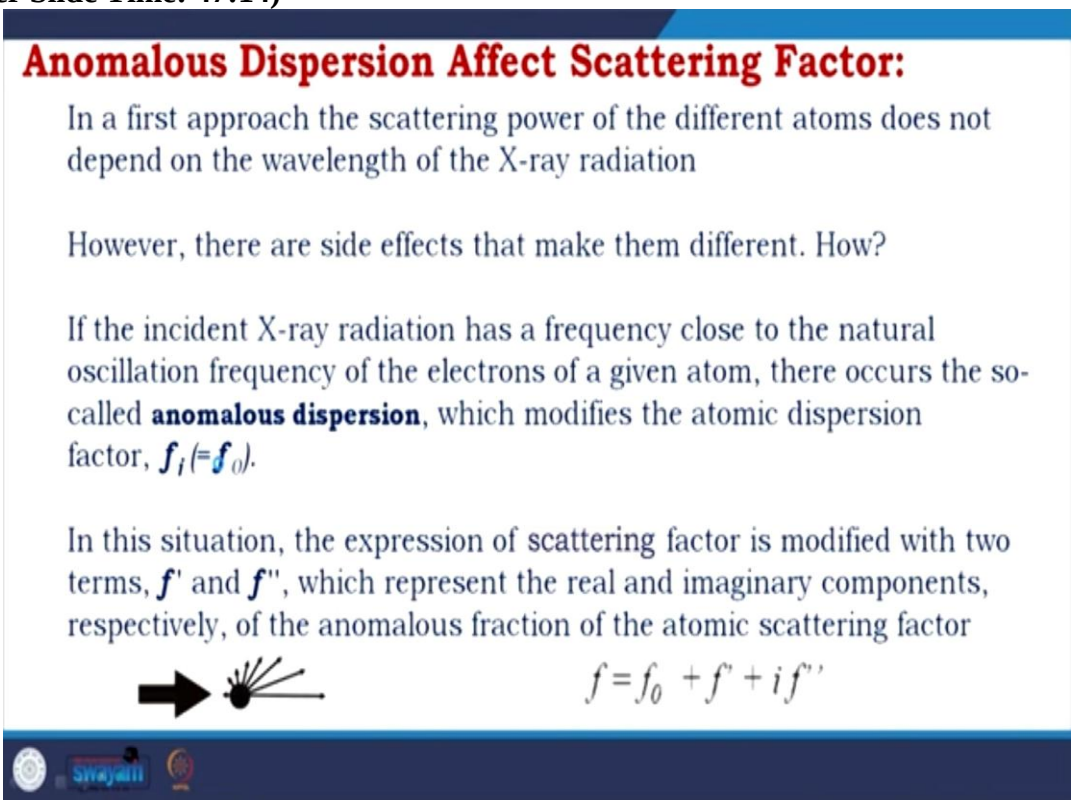


So, if you look at here, you see that there are sets, some of the sets, like if I take these sets are stronger and some are not so strong. So, what happened here? Adding one mercury atom to a protein of 1000 atoms will produce an average fractional change of intensity of 25%, so differences should be easy to measure. So, the intensity difference

$< \Delta I / I > = \sqrt{N_H} / 2 N_P f_H / f_P$

f_H is the contribution of scattering from the heavy atom, f_p is scattering from the protein. We can use isomorphous differences to derive phase information via isomorphous replacement.

(Refer Slide Time: 47:14)

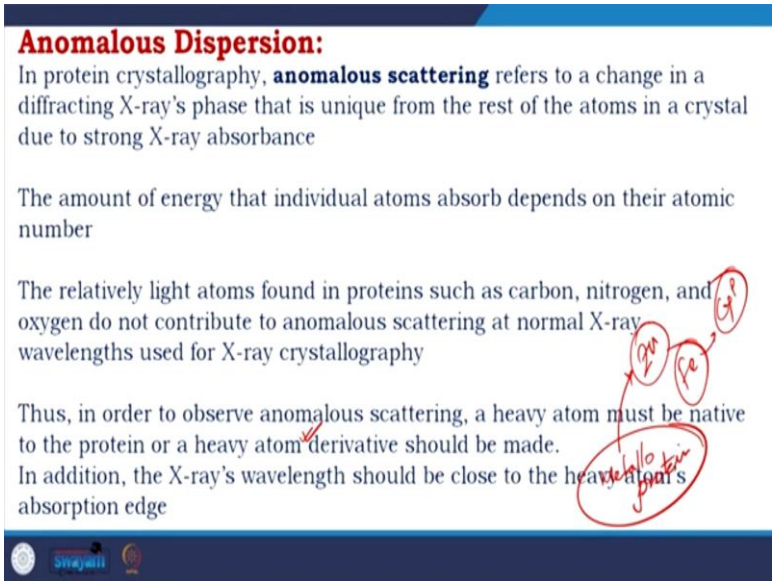


In a first approach, the scattering power of the different atoms does not depend on the wavelength of the X-ray radiation. However, there are side effects that make them different. If the incident X-ray radiation has a frequency close to the natural oscillation frequency of the electrons of a given atom.

There occurs the so-called anomalous dispersion, which modifies the atomic dispersion factor. In this situation, the expression of the scattering factor is modified with two terms

$f = f_0 + f' + if''$

(Refer Slide Time: 48:30)

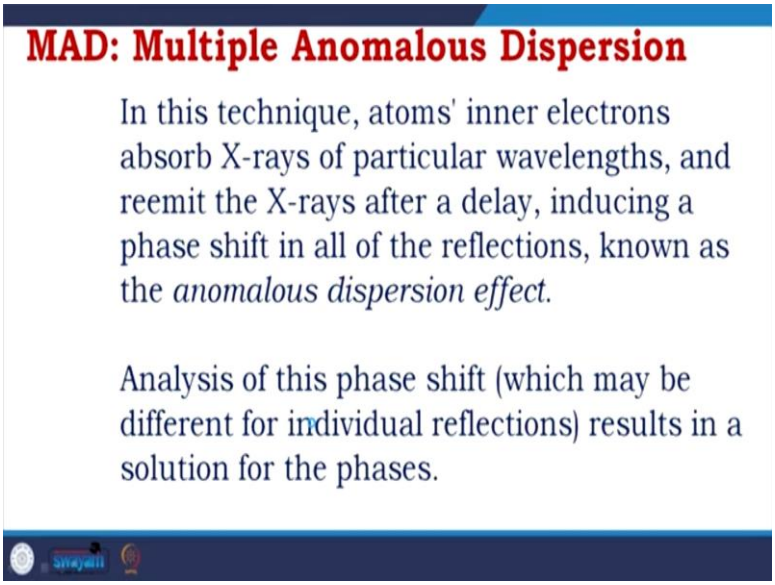


In protein crystallography, anomalous scattering refers to a change in a diffracting X rays phase that is unique from the other atoms in a crystal due to strong X-ray absorbance. The amount of energy that individual atoms absorb depends on their atomic number. The relatively light atoms found in proteins such as carbon, nitrogen, and oxygen do not contribute to anomalous scattering at normal X-ray wavelengths used for X-ray crystallography.

Thus to observe anomalous scattering, a heavy atom must be native to the protein, or a heavy atom derivative should be made. So, it would help if you had a heavy atom to do that a metal again and understand to observe anomalous scattering, a heavy atom must be native to the protein, or a heavy atom derivative should be made.

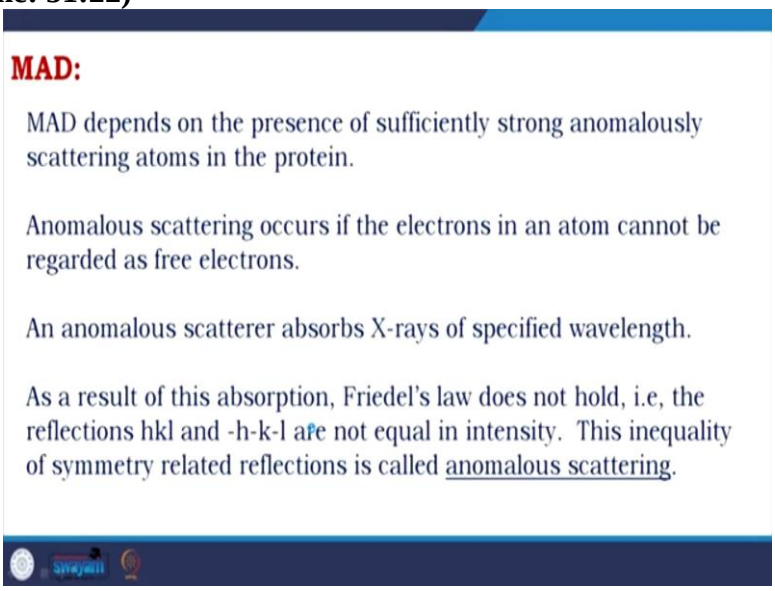
So, what is native? There are some proteins which are called Metalloproteins. In those proteins zinc, iron in their different oxidative states presents like, for iron, this is cytochrome p450. So, there are different DNA binding proteins with zinc and zinc fingers, so they are native, but if it is not native, you have to develop heavy-atom derivatives. In addition, the X-rays wavelength should be close to the heavy atoms absorption edge. So, you have to choose the heavy atom according to that.

(Refer Slide Time: 50:52)



MAD is called Multiple Anomalous Dispersion. In this technique, atoms' inner electrons absorb X rays of particular wavelengths and reemit the X rays after a delay inducing a phase shift in all reflections known as the anomalous dispersion effect. This phase shift (which may be different for individual reflections) results in a solution for the phases.

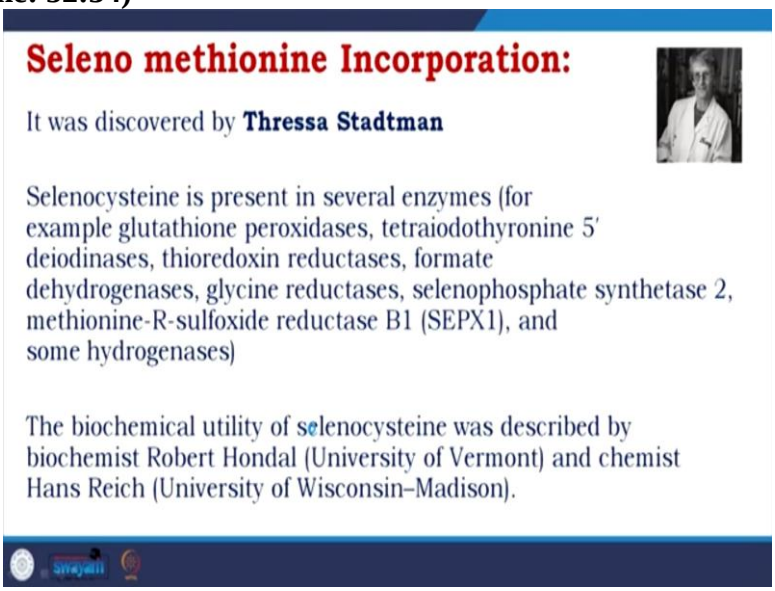
(Refer Slide Time: 51:22)



So, getting phase is a comparison of anything, any difference with the element with high electron content, with the element with anomalous anything the difference is made you are getting into your solution. MAD depends on sufficiently strong anomalously scattering atoms in the protein. As I talked about zinc, if you have one zinc and your protein is huge, then the presence of the zinc atom concerning the carbon, nitrogen, oxygen would be much low, you do not get good signal.

Anomalous scattering occurs if the electrons in an atom cannot be regarded as free electrons. An anomalous scattered absorbs X rays of a specified wavelength. As a result of this absorption, Friedel's law does not hold that the reflection hkl and - h - k - l are not equal in intensity. This inequality of symmetry-related reflections is called anomalous scattering.

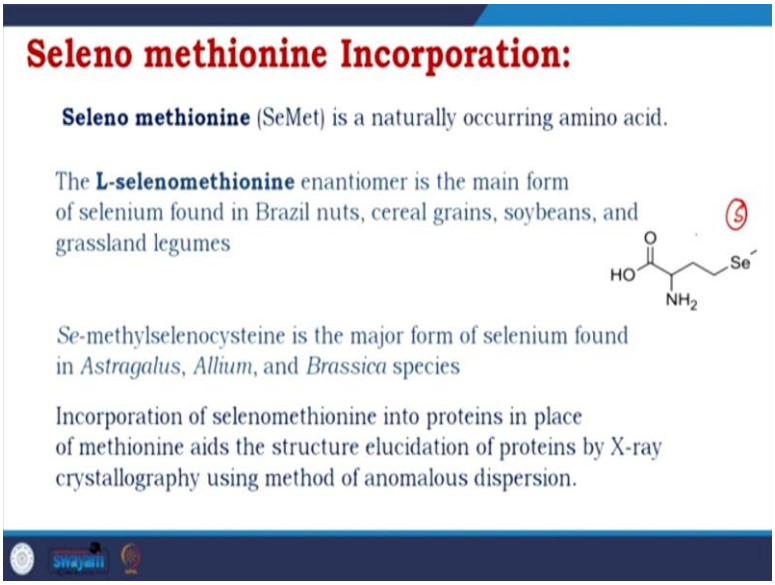
(Refer Slide Time: 52:34)



In the case of methionine, there is sulfur. The sulfur is replaced with selenium in some bacteria, which probably adapt to very high selenium conditions. That code is utilized in protein engineering to be introduced so that whatever methionines are there is changed to selenomethionine. It was discovered by Thressa Stadtman.

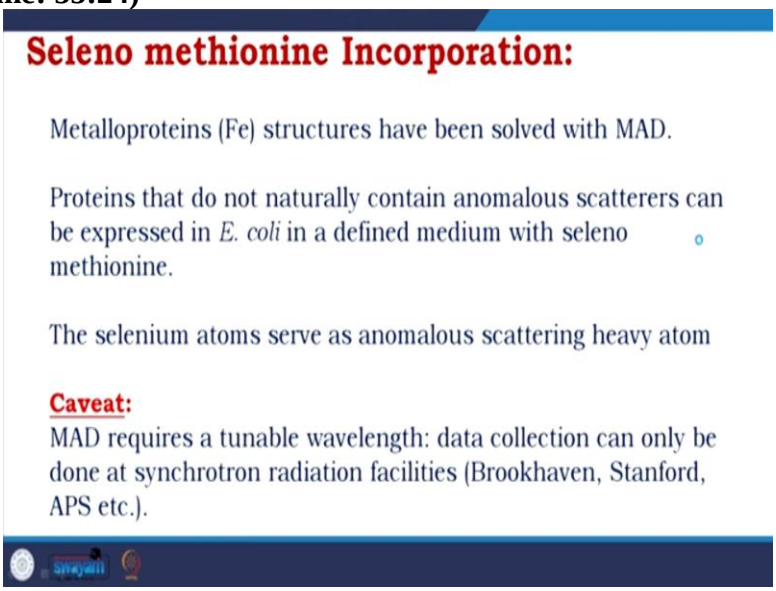
Selenocysteine is present in several enzymes glutathione peroxidases, tetraiodothyronine 5' prime deiodinases, thioredoxin reductases, formate dehydrogenases, glycine reductases, selenophosphate synthetase 2, methionine R sulfoxide reductase B1 or SEPX1, and some hydrogenases. Biochemist Robert Hondal described the biochemical utility of selenocysteine from University of Vermont and chemist Hans Reich, University of Wisconsin, Madison.

(Refer Slide Time: 54:15)



So, as I said, it is a naturally occurring amino acid. The L selenomethionine enantiomer is the main selenium found in Brazil nuts, cereal grains, soybeans, and grassland legumes. Seleno methylselenocysteine is the major selenium found in Astragalus, Allium, and Brassica species. Incorporation of selenomethionine into protein in place of methionine aids the structure elucidation of proteins by X-ray crystallography using the method of anomalous dispersion.

(Refer Slide Time: 55:24)



Metalloproteins structures have been solved with multiple anomalous dispersion (MAD). Proteins that do not naturally contain anomalous scatterers can be expressed in E coli in a defined medium with selenomethionine or selenocysteine. The Selenium atoms serve as anomalous scattering heavy atom. The caveat or MAD requires a tunable wavelength. The data collection can only be done at synchrotron radiation facilities.