

Image Formation

Mauro Gemmi

Earth Science Department University of Milan, via Botticelli 23, I-20133 Milano Italy
e-mail: mauro.gemmi@unimi.it

1. Amplitude and phase objects.

If we "illuminate" a crystal with a parallel (not convergent) monochromatic electron beam at the exit surface of our sample the electron wave function can be described as

$$\Psi(x, y) = A(x, y) \exp[-i\phi(x, y)] \quad (1)$$

where x and y are the coordinates on the surface. The exit wave function is not a plane wave but instead it has an amplitude and a phase that are different in different points of the surface. The fact that the amplitude of the wave has changed and it is not constant over the surface, it is due to a non homogeneous absorption inside the crystal. From quantum mechanics it is known that the wave function is not an observable quantity, but it is its square modulus that gives the probability density of finding the electron in the correspondent position. Therefore the intensity of an ideal¹ image is given by

$$I(x, y) = |\Psi(x, y)|^2 = A(x, y)^2 \quad (2)$$

and the contrast is dominated by the absorption effects. In this case the crystal can be considered mainly an "amplitude object" since the contrast in the image is due to the amplitude variation of the exit wave. When the crystal is very thin instead (i.e. near the edge), the absorption can be neglected. The effect of the interaction of the electron beam with the matter results only in a phase change dependent on the position

$$\Psi(x, y) = \exp[-i\phi(x, y)] \quad (3)$$

This approximation is called "phase object" approximation. It is worth to note that if our imaging system is ideal a phase object doesn't give contrast in the image

$$I(x, y) = |\Psi(x, y)|^2 = 1 \quad (4)$$

A phase object is invisible if the optical system is perfectly in focus! We will see in the following how in a real optical system, like the electron microscope, we can exploit the aberrations (that are defects of our imaging systems!) to obtain contrast from a phase object.

2. Phase object approximation for electrons.

The phase object approximation for electrons is very useful since there is a simple relation between the phase shift $\phi(x, y)$ and the electrostatic potential inside the crystal. From quantum mechanics it is known that a particle having momentum p behaves² like a wave with wave length given by the famous De Broglie equation

$$\lambda = \frac{h}{p} \quad (5)$$

where h is the Plank constant. An electron accelerated in the gun of an electron microscope by an accelerating voltage $-V_0$ has a total energy of $E=eV_0$ if e is the absolute value of the electron charge. Before reaching the crystal this energy is completely kinetic. Inside the crystal the energy splits into kinetic T and potential energy $-eV(\mathbf{r})$ since the nuclei and the electronic clouds of the atoms produce an electrical field that scatters the incoming electron beam. Using the conservation of the energy we can derive the momentum p before and inside the crystal.

¹ With ideal optical system we mean a system without aberration that reproduces, in its image plane, a perfect copy of the exit wave function. In an electron microscope the exit wave function is the analogous of the object in a conventional optical system.

² The wave behavior is described by a probability wave having this wave length.

Before

$$p = \sqrt{2mT} = \sqrt{2meV_0}$$

only kinetic
also electric potential

Inside

$$T = E + eV(\vec{r}) = eV_0 + eV(\vec{r})$$

$$p = \sqrt{2mT} = \sqrt{2m(eV_0 + eV(\vec{r}))}$$

In terms of wave length using the De Broglie formula we have

$$\lambda_{\text{before}} = \frac{h}{\sqrt{2meV_0}}$$

$$\lambda_{\text{inside}} = \frac{h}{\sqrt{2m(eV_0 + eV(\vec{r}))}}$$
(6)

Since the wave length is different inside the crystal the electron wave function undergoes a phase shift that for a crystal of infinitesimal thickness dz is given by

$$d\phi = \left(\frac{2\pi}{\lambda_{\text{inside}}} - \frac{2\pi}{\lambda_{\text{before}}} \right) dz = \frac{2\pi}{\lambda_{\text{before}}} \left(\sqrt{1 + \frac{V(\vec{r})}{V_0}} - 1 \right) dz = \frac{\pi}{\lambda_{\text{before}} V_0} V(\vec{r}) dz + \dots$$

The approximation of considering $\frac{V(\vec{r})}{V_0}$ small is reasonable since the crystal potential involves energy of several eV whereas the accelerating voltage is around 10^5 Volts.

In order to obtain the total phase shift for a thickness t we have to integrate the last equation

$$\phi(x, y) = \sigma \int_0^t V(x, y, z) dz = \sigma N_{\text{cell}} \int_{\text{unit cell}} V(x, y, z) dz = \sigma N_{\text{cell}} V_{\text{proj}}(x, y)$$
(7)

where $\sigma = \frac{2\pi me\lambda}{h^2}$ is called interaction constant and N_{cell} is the number of unit cells crossed by the beam.

Consequently, if the crystal is thin enough that can be considered a phase object the wave function of a monochromatic plane wave at the exit surface becomes

$$\Psi(x, y) = \exp[-i\sigma N_{\text{cell}} V_{\text{proj}}(x, y)]$$
(8)

As can be seen the phase of the exit wave function carries a lot of information on the crystal structure since it is proportional to the projected electrostatic potential. It is a two-dimensional map of the structure projected along the electron beam. Unfortunately, as pointed out in the previous section, a perfect image of this wave function doesn't show any contrast and the phase of the wave is lost!

It must be noticed that whereas in crystallography the word phase is always referred to the phase of the structure factor,³ in electron microscopy it can have different meanings. The reader must always check if the phase means the structure factor phase or the phase of the exit wave function, like in this case. In particular the reconstruction of the phase of the exit wave function is not the phase problem in crystallography that is the recovering of the structure factor phases. Even if these two tasks give both information about the crystal structure they are separate problems in different spaces, the first one in direct space the second in reciprocal space.

3. Weak phase object approximation and phase contrast.

The expression (8) can be further simplified. If the crystal is so thin that the phase shift is small compared to 2π , we can consider only the first two terms of the Taylor series of the exponential

$$\Psi(x, y) = 1 - i\sigma N_{\text{cell}} V_{\text{proj}}(x, y)$$
(9)

This formula is known as the weak phase object approximation for obvious reasons. If we calculate in this approximation the contrast of a perfect image, we still obtain a uniform contrast since terms of second order must be neglected. In Fourier space the same equation becomes (see mathematical appendix)

³ The structure factor is the Fourier transform of the quantity responsible for the scattering. In x-ray diffraction is the Fourier transform of the electronic density, in electron diffraction is the Fourier transform of the electrostatic potential.

$$\tilde{\psi}(\mathbf{u}) = \delta(\mathbf{u}) - i\sigma N_{cell} \tilde{V}(\mathbf{u}) = \delta(\mathbf{u}) - i\sigma \frac{tS}{\Omega} \frac{\lambda}{\sigma S} F(\mathbf{u}) = \delta(\mathbf{u}) - i \frac{\lambda t}{\Omega} F(\mathbf{u}) \quad (10)$$

where the Dirac delta function $\delta(\mathbf{u})$ represents the intense direct beam, $F(\mathbf{u})$ is the structure factor, t is the thickness and S is the area of one unit cell projected along the incident beam direction. Since the structure factor amplitudes are proportional to the scattered beam amplitudes⁴, from (9) and (10) we can conclude that we don't see the effects of the scattering in the image because they give a pure imaginary contribution to the wave function. If we are able to rotate in the complex plane (see complex number in the appendix) the phase of the scattered beams (i.e. the phase of the structure factors) ± 90 degrees with respect to the direct beam we can see contrast arising from the scattered electrons. In fact neglecting the second order terms we have

$$\left| 1 - i\sigma N_{cell} V_{proj}(x, y) \exp(\pm i \frac{\pi}{2}) \right|^2 = 1 \pm 2\sigma N_{cell} V_{proj}(x, y) \quad (11)$$

This was first noticed for photons by Zernike in 1935, that also proposed a way of realizing such a phase contrast in an usual optical microscope. Zernike exploited the fact that in the back focal plane (BFP) of the

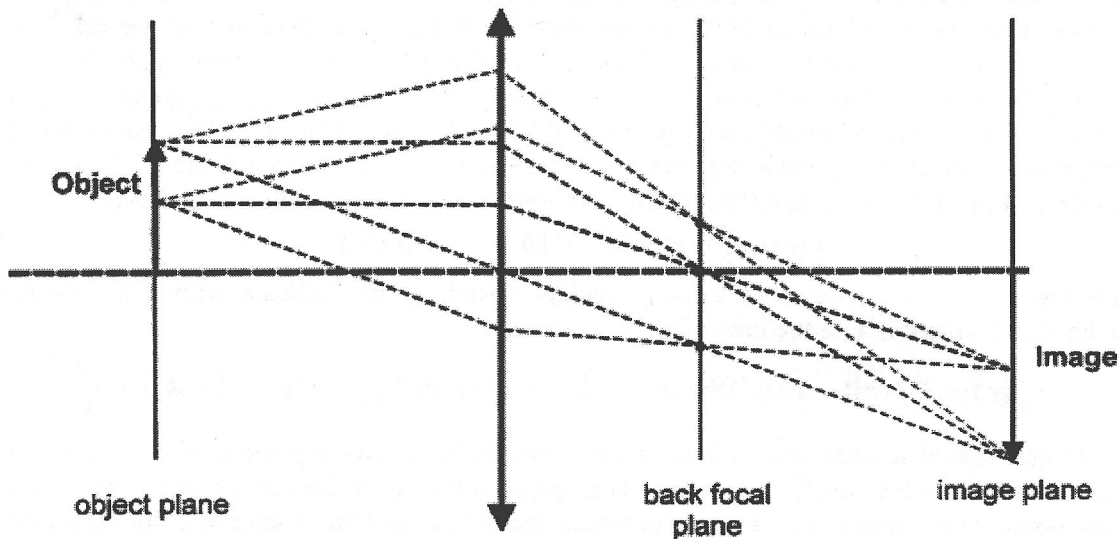


Fig. 1 Parallel rays coming from different points of the object are focused in the same points in the back focal plane.

objective lens we can find the Fourier transform of the exit wave function. In fact from geometrical optics we know that parallel rays arising from different points of the object are focused in the same point of the BFP and the position of this point depends only on the direction of the parallel rays. Consequently in the BFP there is a one to one correspondence between its points P and directions \mathbf{u} of the incoming rays: $P=P(\mathbf{u})$. In electron optics a set of parallel rays represents a set of plane waves. The Fourier transform $\tilde{\psi}(\mathbf{u})$ of the exit wave function is a decomposition in plane waves and it says that the plane wave $\exp[2\pi i(\mathbf{u} \cdot \mathbf{x})]$ is present with a weight $\tilde{\psi}(\mathbf{u})$ in the exit wave. This plane wave is focused in the BFP in a point $P(\mathbf{u})$ and since it has a weight equal to $\tilde{\psi}(\mathbf{u})$, a picture of the BFP gives in that point an intensity proportional to $|\tilde{\psi}(\mathbf{u})|^2$. Thus the lens forms in the BFP the Fourier transform of the exit wave function and its squared amplitude can be recorded in a photographic plate: this is what we do every time we record a diffraction pattern. Zernike proposed to use the fact that in the BFP of an optical microscope only the unscattered photons⁵ are focused on the optical axis. He suggested to put in the BFP a phase-changing plate made of a glass substrate on which a small transparent dielectric dot has been deposited. The dot is centered on the optical axis and has a thickness and an index of refraction such that it retards the phase of the focused light by either $\pi/2$ radians or $3\pi/2$ radians (90 or 270 degrees). The unscattered particles are intercepted by the shifting filter and their

⁴ This is true only if the phase object approximation is valid and it is false if dynamical effects are relevant.

⁵ The same argument can be applied both to photons in an optical microscope and to electrons in an electron microscope. In the last case, as it is explained later, the phase shift must be obtained using different methods.

phase is shifted. The result is a phase contrast image in the image plane, as already discussed for the electron case in the formula (11). The question is how to do this for electrons? How to produce this phase shift on the central beam or on the scattered beams? Surely we can't put something inside the column, electrons are not photons! To answer this question we have to describe how a general optical system transfers the information from the object plane to the image plane. What we will discover is that both the focus and the spherical aberration introduce phase shifts in the back focal plane that are a continuous function of the distance from perfect focus and of the scattering vector \mathbf{u} . There is a special focus called Shertzer for which a shifting close to 270 degrees ($-\pi/2$ in radians) is applied to all the scattered beams inside the resolution limit of the microscope.

4. Point spread function.

An optical system is a device that transfers information from an object plane to an image plane with a certain magnification. The better is the system the smaller are the deformations induced by the system during the transfer process. An ideal optical system would reproduce the object exactly as it is but on a different scale. A real optical system instead introduces errors and the final image is always slightly blurred and distorted with respect to the object. There is a very elegant and practical way to describe the imaging properties of an optical system. It is enough to know how the system transfers to the image plane a point object. If we know this we can calculate and predict the image of every object. A point object in \mathbf{r}_0 is represented mathematically by a Dirac delta function $\delta(\mathbf{r}-\mathbf{r}_0)$. Its image $h(\mathbf{r}-M\mathbf{r}_0)$ is the point spread function of the optical system and it is centered around the point $M\mathbf{r}_0$, M being the magnification of the system. Every object can be described by an object function $f(\mathbf{r})$ that for electrons is the exit wave function. The mathematical properties of the Dirac delta function tell us that we can write $f(\mathbf{r})$ as a convolution with $\delta(\mathbf{r}-\mathbf{r}_0)$

$$f(\mathbf{r}) = \int f(\mathbf{r}')\delta(\mathbf{r}-\mathbf{r}')d\mathbf{r}' = (f \otimes \delta)(\mathbf{r}) \quad (12)$$

Thus, since the electron microscope is a linear optical system, if we indicate with S the operator that transform the object function into the image function we have

$$f_{im}(\mathbf{r}) = S[f(\mathbf{r})] = \int f(\mathbf{r}')S[\delta(\mathbf{r}-\mathbf{r}')]d\mathbf{r}' = \int f(\mathbf{r}')h\left(\frac{\mathbf{r}}{M}-\mathbf{r}'\right)d\mathbf{r}' = (f \otimes h)\left(\frac{\mathbf{r}}{M}\right) \quad (13)$$

This is a very nice result! In order to obtain the wave function in the image plane it suffices to calculate the convolution of the exit wave function with the point spread function of the microscope. The equation (13) says that the point spread function is the way in which the optical system is sampling the object and every point \mathbf{r} of the image receives a contribution from all the points \mathbf{r}' of the object for which $h(\mathbf{r}/M-\mathbf{r}')$ is different from zero. As we have seen in the previous chapter if we pass in Fourier space sometimes the equations become simpler and they can be easily interpreted. Furthermore in an optical system the Fourier space coincides with the BFP and it can be observed directly. The convolution that appears in equation (13) is not an intuitive operation but if we calculate the Fourier transform it becomes a product⁶. We can then summarize the effect of our microscope on the image as a product of the Fourier transform of the point spread function $H(\mathbf{u})$ with the Fourier transform of the exit wave $\tilde{\Psi}(\mathbf{u})$. If we call $\tilde{\Psi}_{im}(\mathbf{u})$ the Fourier transform of the wave function in the image plane we have

$$\tilde{\Psi}_{im}(\mathbf{u}) = \tilde{\Psi}(\mathbf{u})H(\mathbf{u}) \quad (14)$$

The function $H(\mathbf{u})$ is usually called amplitude transfer function

5. The amplitude transfer function: the effect of spherical aberration and defocus

The amplitude transfer function of an electron microscope is made by the product of three terms

$$H(\mathbf{u}) = A(\mathbf{u})E(\mathbf{u})B(\mathbf{u}) \quad (15)$$

- $A(\mathbf{u})$: the aperture function. It is a cut off function due to the objective diaphragm that we put in the back focal plane in order to eliminate the contributions of high spatial frequencies. It is a step function equal to 1 for those values of the scattering vector \mathbf{u} that fall inside the aperture and to 0 elsewhere.
- $E(\mathbf{u})$: the envelope function. It is a damping function that attenuates the effect of the higher spatial frequencies. It is due mainly to the fact that the incoming electrons don't have the same energy

⁶ This is a beautiful property of Fourier transform. By passing in Fourier space convolution becomes product and product becomes convolution $FT[f \otimes g] = FT[f]FT[g]$ and $FT[fg] = FT[f] \otimes FT[g]$

(chromatic aberration) and the beam is not perfectly parallel but it is slightly convergent. Therefore the sample it is not "illuminated" by a monochromatic plane wave and the resulting image is moderately blurred.

- $B(\mathbf{u})$: the aberration function. It is a complex exponential

$$B(\mathbf{u}) = \exp[i\chi(\mathbf{u})] \quad (16)$$

that produces a phase shift $\chi(\mathbf{u})$ in the Fourier space. It is due to the spherical aberration of the system and to the focus conditions.

The spherical aberration is a defect of an optical system caused by an inhomogeneous deflection of the off axis rays. The further off axis the electron is, the more strongly is bent back towards the axis. To take into account this effect we have to consider that in the equations describing the motion of the electron in the magnetic field of the lenses the inclination β of the trajectories with respect to the optical axis always appears as $\sin(\beta)$ which can be expanded in Taylor series as

$$\sin(\beta) = \beta - \frac{\beta^3}{3!} + \frac{\beta^5}{5!} + \dots$$

In the paraxial theory where $\sin(\beta) \approx \beta$ it can be shown that it exists a plane (image plane) in which a faithful image can be obtained. However if the aperture of our optical system is quite large, the cubic term cannot be ignored and third order aberrations must be considered. The result of the cubic term is that an area surrounding a certain object point can contribute to the same point in the image giving raise to a disk of confusion in the object plane of radius

$$r_c = C_s \beta^3 \quad (17)$$

The maximum path difference between rays coming from this confusion disk (see fig. 2) is

$$\Delta = r_c \sin(\beta) \approx r_c \beta$$

that in differential form gives

$$d\Delta = r_c d\beta = C_s \beta^2 d\beta$$

The corresponding phase shift is

$$d\chi(\beta) = \frac{2\pi}{\lambda} d\Delta = \frac{2\pi}{\lambda} C_s \beta^2 d\beta \quad (18)$$

Taking into account the Bragg law

$$2\sin(\theta_b) \approx 2\theta_b = \beta = \lambda u$$

and integrating (18) between 0 and β we obtain the total phase shift

$$\chi(u) = \frac{\pi}{2} C_s \lambda^3 u^4 \quad (19)$$

This is not a rigorous derivation of $\chi(u)$. If the reader is not satisfied he can consult the Spence's book quoted in the references.

The effect of the focus can be calculated in an analogous way. In fact if we underfocus or over focus the objective lens we obtain a disk of confusion in the object plane having radius

$$r_f = \varepsilon \beta$$

where ε is the distance of the object plane from the plane that is in focus. ε is positive for overfocus (strengthening of the lens) and negative for underfocus (weakening of the lens). ε is usually known as defocus since high resolution

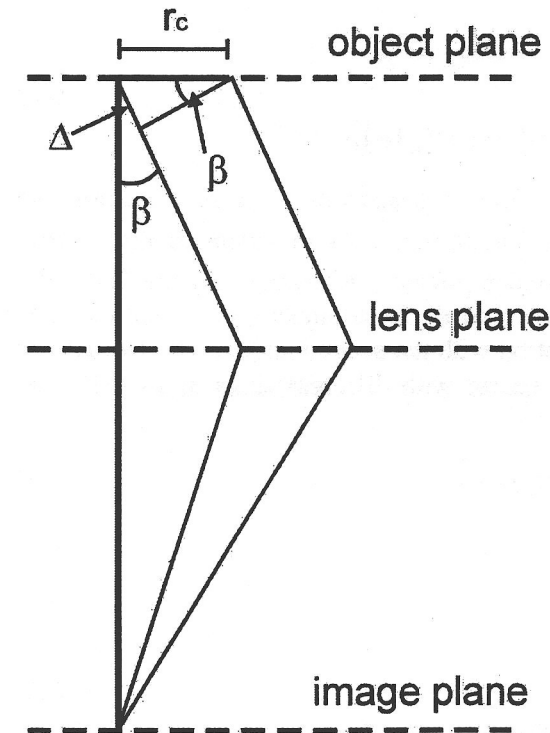


Fig. 2 Ray geometry due to spherical aberration

images are taken underfocused. Following the same procedure adopted for spherical aberration with r_c replaced by r_f we obtain the following phase shift

$$\chi(u) = \pi \varepsilon \lambda u^2$$

The total phase shift due to $B(\mathbf{u})$ is then

$$\chi(u) = \pi \varepsilon \lambda u^2 + \frac{\pi}{2} C_s \lambda^3 u^4 \quad (20)$$

This phase shift $\chi(\mathbf{u})$ can play the role of the phase-changing plate of the Zernike's microscope. Furthermore $\chi(\mathbf{u})$ can be tuned by changing the defocus ε . Unfortunately the optimal condition of a phase shift constant over the reciprocal space and equal to $\pi/2$ radians or $3\pi/2$ radians cannot be achieved exactly since $\chi(\mathbf{u})$ is a varying function of \mathbf{u} . However we have optimal defocus values for which we are close to this condition.

Summarising we have found a way to obtain a phase contrast image with electrons using the imaging properties of an optical system. In order to obtain the same phase shift obtained by Zernike in an optical microscope with the phase-changing plate we have to underfocus the electron microscope. It is astonishing that to see with electrons we exploit the spherical aberration which is a defect of the lens and we record images not in focus!

6. Contrast transfer function

In order to see how the transfer function directly influences the image we calculate the Fourier transform of the square modulus of the wave function in the image plane

$$I_{im}(\mathbf{u}) = FT\left(|\Psi_{im}(x, y)|^2\right) = FT\left(\Psi(x, y) \otimes h(x, y)\right)^2 = \tilde{\Psi}_{im}(\mathbf{u}) \otimes \left(FT\left(\Psi_{im}^*(x, y)\right)\right) = \tilde{\Psi}_{im}(\mathbf{u}) \otimes \tilde{\Psi}_{im}^*(-\mathbf{u}) \quad (21)$$

This is an experimental quantity. It can be easily obtained by most of the image processing software directly from the recorded image. From equations (14), (15) and (16) we obtain, for $\tilde{\Psi}_{im}(\mathbf{u})$, the expression

$$\tilde{\Psi}_{im}(\mathbf{u}) = \tilde{\Psi}(\mathbf{u}) A(\mathbf{u}) E(\mathbf{u}) (\cos[\chi(\mathbf{u})] + i \sin[\chi(\mathbf{u})]) \quad (22)$$

If we substitute in (22) the weak phase object approximation (10), since both $A(0)$ and $B(0)$ are equal to 1 and $\chi(0)=0$ we obtain

$$\begin{aligned} \tilde{\Psi}_{im}(\mathbf{u}) &= \delta(\mathbf{u}) + \frac{\lambda t}{\Omega} F(\mathbf{u}) A(\mathbf{u}) E(\mathbf{u}) (\sin[\chi(\mathbf{u})] - i \cos[\chi(\mathbf{u})]) \\ \tilde{\Psi}_{im}^*(-\mathbf{u}) &= \delta(\mathbf{u}) + \frac{\lambda t}{\Omega} F(\mathbf{u}) A(\mathbf{u}) E(\mathbf{u}) (\sin[\chi(\mathbf{u})] + i \cos[\chi(\mathbf{u})]) \end{aligned} \quad (23)$$

We have used the property that the structure factor is the Fourier transform of a real quantity so $F^*(\mathbf{u}) = F(-\mathbf{u})$. These expressions must be inserted in the convolution (21) to calculate the Fourier transform of the image and apparently the result seems to be quite complicated. Although only the first order terms in $F(\mathbf{u})$ must be taken into account since we are considering a weak phase object approximation. The first order terms arise only from the convolution of the delta function with the second addendum of equations (23). Furthermore the $\cos[\chi(\mathbf{u})]$ terms cancel out because they appear with different signs in eq (23). The final result is

$$I_{im}(\mathbf{u}) = \delta(\mathbf{u}) + 2 \frac{\lambda t}{\Omega} F(\mathbf{u}) A(\mathbf{u}) E(\mathbf{u}) \sin[\chi(\mathbf{u})] \quad (24)$$

This equation is usually written as

$$I_{im}(\mathbf{u}) = \delta(\mathbf{u}) + \frac{\lambda t}{\Omega} T(\mathbf{u}) F(\mathbf{u}) \quad (25)$$

and

$$T(\mathbf{u}) = 2 A(\mathbf{u}) E(\mathbf{u}) \sin[\chi(\mathbf{u})] \quad (26)$$

is called the contrast transfer function (CTF). Examples of this function for a modern field emission microscope of 200 kV are shown in fig. 3. The envelop part $E(\mathbf{u})$ has a damping effect and it limits the maximum frequency transferred by the microscope. The cut off value u_0 imposed by this function is called the information limit, since $1/u_0$ gives the size of the smallest detail detectable in the image (see fig. 3b). It is misleading to relate u_0 directly to the resolution since the information transferred by $T(\mathbf{u})$ beyond its first zero is distorted by the change of sign. For this reason it is more common to consider the resolution of the microscope (at a certain defocus) fixed by the first zero u_r of $T(\mathbf{u})$ (see fig. 3b). By looking at equation (23) and (26) we see that $\sin[\chi(\mathbf{u})]$ is transferring the information to our image, whereas the $\cos[\chi(\mathbf{u})]$ transfers information to the imaginary part of $\Psi_{im}(x, y)$ that gives contribution only to the second order in the

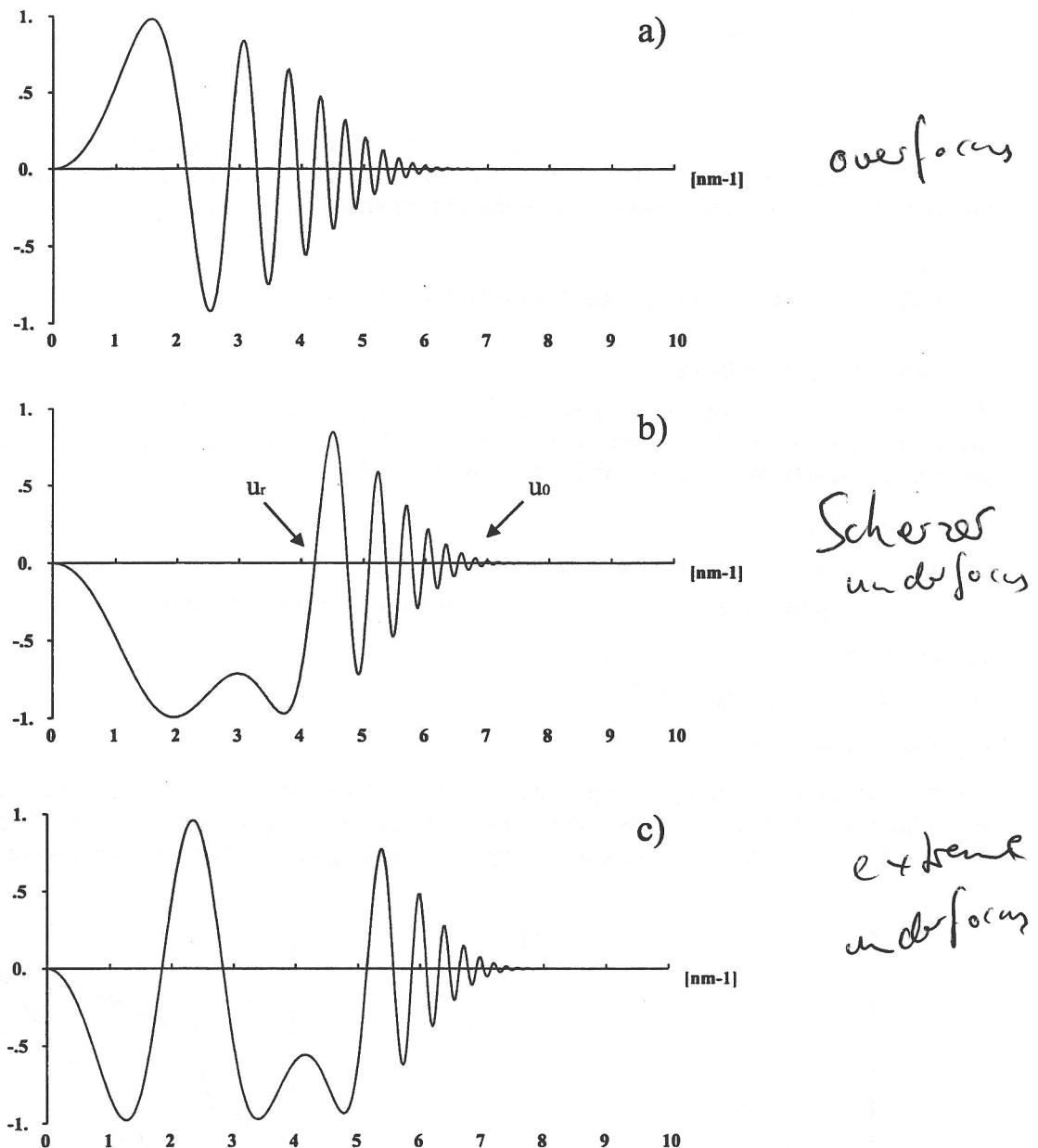


Fig. 3 Contrast transfer functions calculated for a 200kV Philips Tecnai F20 field emission gun microscope at different defocus. The C_s of the lens is 1.2mm. a) Positive defocus $\epsilon=70\text{nm}$. b) Scherzer defocus $\epsilon=-70\text{nm}$. c) $\epsilon=-130\text{nm}$

calculation of the square modulus $I_{im}(x, y) = |\Psi_{im}(x, y)|^2$. Consequently, as already worked out by following Zenike's considerations, the best conditions are reached when the $\sin[\chi(\mathbf{u})]$ remains close to -1 or 1 for the widest possible interval in Fourier space. The form of $\chi(\mathbf{u})$ written in (20) implies that if the spherical aberration term is not compensated by a negative defocus ϵ , the $\sin[\chi(\mathbf{u})]$ oscillates too fast and the resolution is compromised (see fig 3a). Therefore the $\sin[\chi(\mathbf{u})]$ must be negative for small \mathbf{u} and the optimal shape is obtained for a negative defocus. This defocus is called Scherzer defocus and it is a function only of the C_s and the accelerating voltage of the microscope

$$\epsilon_{sch} = -\sqrt{\frac{4}{3}} C_s \lambda \quad (27)$$

For this value the $\sin[\chi(\mathbf{u})]$ is negative and close to -1 for a wide range of \mathbf{u} (see fig. 3b). The first zero is

$$u_{sch} = 1.51 C_s^{-\frac{1}{4}} \lambda^{\frac{3}{4}} \quad (28)$$

Inverting the (28) we obtain the resolution

$$r_{sch} = \frac{1}{1.51} C_s^{1/4} \lambda^{3/4} \quad (29)$$

From this last formula it is quite obvious in which way we can increase the performance of the microscope as a high resolution instrument. We have two possibilities either we decrease the electron wave length by increasing the accelerating voltage or we reduce the spherical aberration.

A. Mathematical appendix on Fourier transform

A.1 Complex numbers

Before starting to discuss the Fourier transform and its properties we must become familiar with several mathematical concepts. First of all we must know what complex numbers are, because the Fourier transforms are complex functions of real variables. A complex number is defined as

$$z = x + i y \quad (1)$$

where x and y are real numbers and i is the imaginary unit. We can treat these numbers as ordinary numbers with the only caution that $i^2 = -1$. For example if we want to sum z_1 with z_2 we have:

$$z_1 + z_2 = x_1 + x_2 + i (y_1 + y_2) \quad (2)$$

whereas if we multiply them we obtain

$$z_1 z_2 = x_1 x_2 - y_1 y_2 + i (x_1 y_2 + x_2 y_1)$$

Another way of representing a complex number is as a 2D-vector having components (x, y) . The x and y axes are called real and imaginary axis respectively. In fact if $y=0$, z is real while, if $x=0$, z is purely imaginary. In this case the sum (2) can be represented using the parallelogram rule for summing two vectors (see fig. 1b).

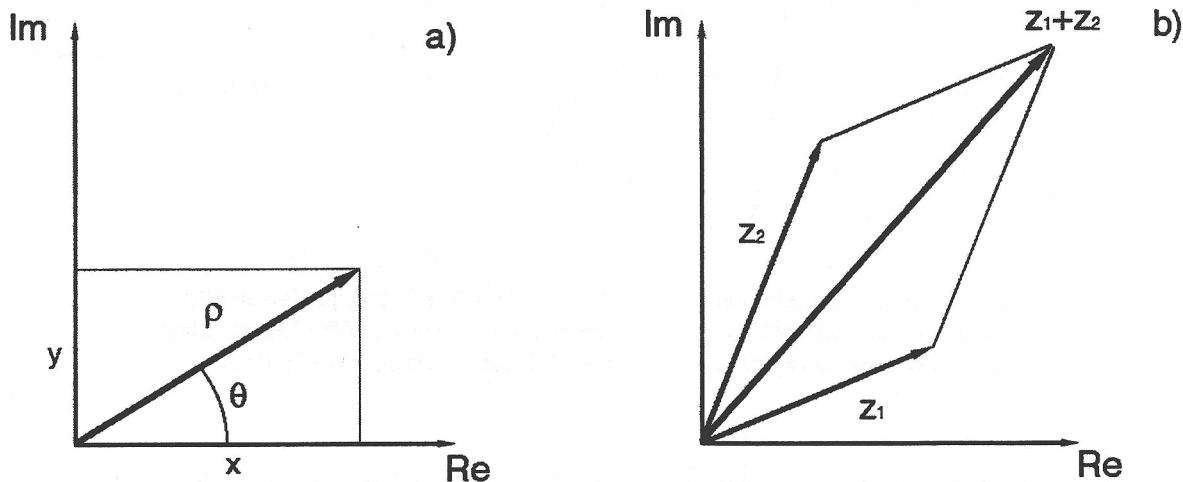


Fig1.a) Representation of a complex number as a vector. b) Sum of complex numbers using the parallelogram law.

We can identify this vector also using its length ρ (called modulus or amplitude) and the angle θ (called phase) formed with the positive direction of the x axis (see fig. 1a). In this case we can write

$$z = \rho \cos(\theta) + i \rho \sin(\theta) \quad (3)$$

where

$$\rho = \sqrt{x^2 + y^2}, \quad \theta = \arctan\left(\frac{y}{x}\right)$$

The eq. (3) can be rewritten in a compact form as:

$$z = \rho \exp(i\theta) \quad (4)$$

This is the most compact and useful form of writing a complex number and consequently the most common. This is also the form in which structure factors are usually written and every crystallographer should be familiar with it. We can think of the amplitude as the magnitude of the complex number and the phase as its sign. This is 0 for positive real numbers, 90 degrees for positive imaginary numbers, 180 degrees for negative real numbers, 270 (or -90) degrees for negative imaginary numbers and something in between for ordinary complex numbers. A final remark: if the exponential of $i\theta$ sounds you strange I suggest you to consider it just as an algebraic expression useful for calculation, keeping always in mind that θ is an angle consequently it is periodic with period 360° (or 2π if you measure it in radians).

A.2 Periodic functions in 2 and 3 dimensions.

In electron crystallography we are dealing with periodic functions of two and three variables like for example the contrast in an HREM image or the crystal potential, then we must have clear how to describe the periodicity in more than one dimension. As reference function we chose the cosine wave. A periodic function in one dimension is a function that repeats itself after a certain interval called period. To define it you have to specify the period, the shape of the function inside the period and the value of the function in the origin.

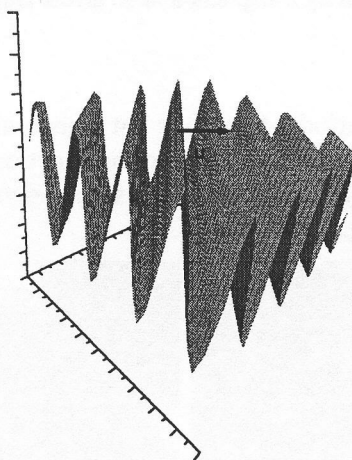


Fig 2 . Graph of the function (5) where $\mathbf{u} = (1,1)$

For the cosine the period is 2π , the shape is the cosine wave and the value in the origin is 1⁷. It is worth to recall that if we consider the cosine function as the real part of a complex number (see eq. 3) a shift of the origin corresponds to a shift of the phase θ ⁸. In two or more dimensions things are more complex because we have more degrees of freedom. We can understand easily that we need something more by looking at the waves on the surface of a lake. We can describe them in an approximate way as two dimensional cosine waves, but in this case beside the period, the shape and the origin we need to specify also the direction of the wind! Then, when we are in more than one dimension, a periodic wave is characterised also by a direction of propagation identified by a vector \mathbf{u} .⁹ In the case of our cosine wave we have:

$$f(\mathbf{r}) = \cos(2\pi \mathbf{u} \cdot \mathbf{r}) \quad (5)$$

In fig. 2 to clarify this concept we display the graph of a two-dimensional cosine wave having direction $\mathbf{u}=(1,1)$. The length of \mathbf{u} is the frequency of the wave (the inverse of the period). Along lines perpendicular

⁷ There is no difference between cosine and sine function except for the choice of the origin.

⁸ This will be very important when one has to fix the structure factor phases on the basis of symmetry. (See Sven Hovmöller lecture on "Symmetry in real and reciprocal space.")

⁹ In this lecture notes we are going to write vectors with bold letters.

to \mathbf{u} the cosine has the same value, therefore \mathbf{u} identifies lines where the phase of the wave is the same¹⁰ (see fig. 3). In three dimensions the situation is exactly the same except that \mathbf{u} has three components instead of 2 and consequently it identifies equal phase planes instead of lines.

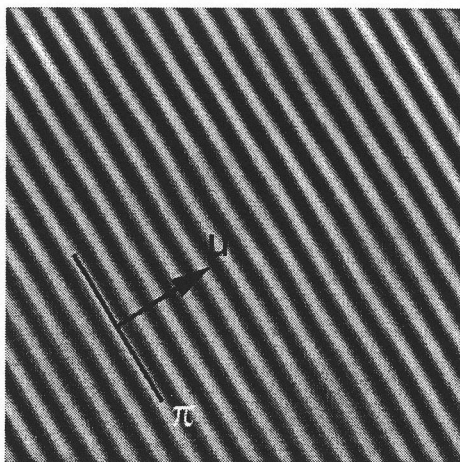


Fig. 3 Gray scale map of the cosine function (5). The equal phase lines π and the vector \mathbf{u} are indicated by a segment and an arrow respectively.

A. 3 Fourier transform

The cosine waves are the fundamental periodic functions but a general periodic function in two dimensions¹¹ is more complicated. Instead of having one definite period the function is repeated regularly in the same way inside an area. This area is called unit cell and it tiles the space identifying with its sides a lattice Λ ¹². An example of such a function is given in fig. 4.

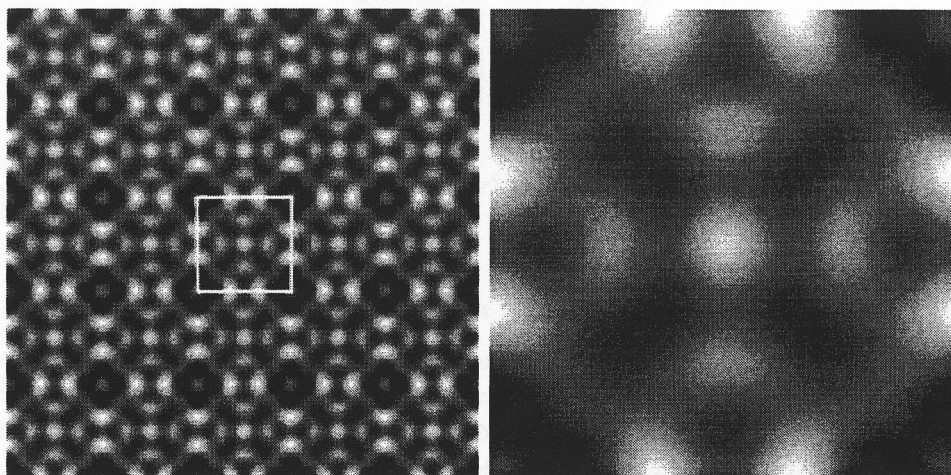


Fig. 4 Example of two dimensional periodic function. On the left a grey scale map is displayed and the unit cell is drawn. On the right an enlargement of the motif inside the unit cell is shown.

Suppose we want to know which fundamental periodicity of the form (5) are hidden in a function like that one displayed. The Fourier transform gives us exactly this information. It is defined in the following way:

¹⁰ In this case I call phase the argument of the cosine function, but I suggest to the reader to think always about our cosine wave as the real part of a complex number. In this way the phase we are speaking about is the phase of the complex number.

¹¹ We are going to analyse the 2D case for simplicity but in principle there is no difference between 2D and 3D Fourier transform, you have just to add one variable. What is really different is the crystallographic symmetry that in 3D is much more rich.

¹² In 3D this is the Bravais lattice of the crystal.

$$\tilde{f}(\mathbf{u}) = \text{FT}[f(\mathbf{r})] \equiv \frac{1}{S} \int_S f(\mathbf{r}) \exp[2\pi i(\mathbf{u} \cdot \mathbf{r})] d\mathbf{r} \quad (6)$$

where S is the area of the unit cell (in 3D is the volume of the unit cell), and the integral is calculated inside the unit cell. To understand the meaning of the $\tilde{f}(\mathbf{u})$ we have to recall that (Fourier inversion formula) the periodic function $f(\mathbf{r})$ can be written as

$$f(\mathbf{r}) = \int \tilde{f}(\mathbf{u}) \exp[-2\pi i(\mathbf{u} \cdot \mathbf{r})] d\mathbf{u} \quad (7)$$

Since the function $f(\mathbf{r})$ is real $\tilde{f}(-\mathbf{u}) = \tilde{f}^*(\mathbf{u})$. If we call $\phi(\mathbf{u})$ the phase of $\tilde{f}(\mathbf{u})$ we have

$$f(\mathbf{r}) = \int |\tilde{f}(\mathbf{u})| \cos[2\pi i(\mathbf{u} \cdot \mathbf{r}) - \phi(\mathbf{u})] d\mathbf{u} \quad (8)$$

where the integration is carried out over half of the frequency space. This formula is very interesting and tells us that the function $f(\mathbf{r})$ can be viewed as a superposition of cosine waves each one weighted by the amplitude and shifted by the phase of the Fourier transform calculated in the corresponding frequency vector \mathbf{u} .

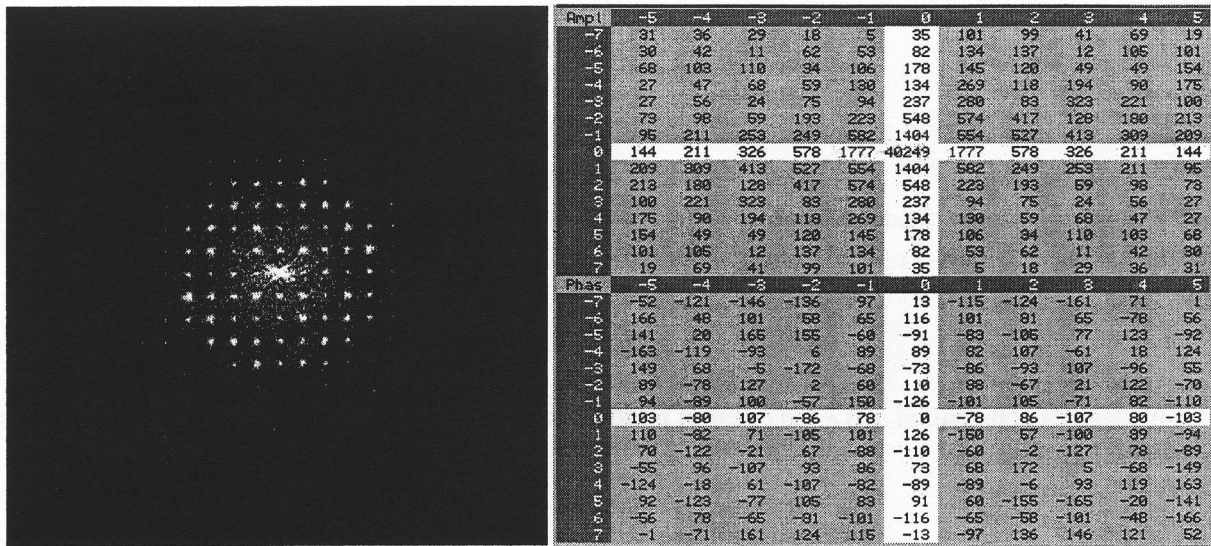


Fig.5: Fourier transform of the 2D periodic function displayed in fig. 4. On the right the digital map around the origin is displayed. Both the phase and the amplitude of the transform are printed.

As an example in fig. 5 the Fourier transform of fig.4 is displayed. As can be seen in the digital map shown on the right we must always taking into account that a Fourier transform is in general a complex function and what is usually represented on the screen is just its amplitude! Another important fact we can learn from fig. 5 is that, apart effects due to numerical problems, for a periodic function the amplitude of its Fourier transform is different from zero only on a discrete set of frequencies. These frequencies form a lattice called the reciprocal lattice¹³. This could be expected, in fact from one dimensional signal theory we know that a signal with a certain periodicity can be described as a superposition of cosine waves that are harmonics of this periodicity. Then the integrals (7) and (8) become series and if we indicate the reciprocal lattice with Λ^* we have:

$$f(\mathbf{r}) = \sum_{\mathbf{u} \in \Lambda^*} |\tilde{f}(\mathbf{u})| \cos[2\pi i(\mathbf{u} \cdot \mathbf{r}) - \phi(\mathbf{u})] \quad (9)$$

A.4 Crystal structure, crystal potential and structure factors

In an electron microscope, electrons are scattered by the electrostatic potential within the crystal. The potential is created by both the nuclei and electrons of the atoms in the crystal and it is a periodic 3D

¹³ This is another way of seeing the crystallographic reciprocal lattice.

function. The potential $V(\mathbf{r})$ at any position $\mathbf{r} = (x \ y \ z)$ within a unit cell depends on the types of atoms and their positions. $V(\mathbf{r})$, is the sum of contributions from all the atoms in the crystal:

$$V(\mathbf{r}) = \sum_{\mathbf{R} \in \Lambda} \sum_{j=1}^N V_j(\mathbf{r} - \mathbf{r}_j + \mathbf{R}) \quad (\text{in Volt}) \quad (10)$$

where V_j is the potential contributed by an atom j at position $\mathbf{r}_j = (x_j \ y_j \ z_j) = x_j \mathbf{a} + y_j \mathbf{b} + z_j \mathbf{c}$ (\mathbf{a} , \mathbf{b} and \mathbf{c} are the basic vectors of the unit cell) \mathbf{R} is a vector of the Bravais lattice Λ and N is the number of atoms in the unit cell. The Fourier components of the crystal potential are, by definition:

$$\tilde{V}(\mathbf{u}) = \text{FT}[V(\mathbf{r})] = \frac{1}{\Omega} \int_{\Omega} \sum_{j=1}^N V_j(\mathbf{r} - \mathbf{r}_j) \exp[2\pi i(\mathbf{u} \cdot \mathbf{r})] d\mathbf{r} \quad (11)$$

where FT is the Fourier transform operator, $\Omega = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ is the volume of the unit cell and the integration is over a whole unit cell. With the change of variable $\mathbf{w} = \mathbf{r} - \mathbf{r}_j$ this can be simplified as

$$\tilde{V}(\mathbf{u}) = \sum_{j=1}^N \exp[2\pi i(\mathbf{u} \cdot \mathbf{r}_j)] \frac{1}{\Omega} \int_{\Omega} V_j(\mathbf{w}) \exp[2\pi i(\mathbf{w} \cdot \mathbf{u})] d\mathbf{w} \quad (\text{in Volt})$$

that can be rewritten in a more compact and readable form as

$$\tilde{V}(\mathbf{u}) = \sum_{j=1}^N \tilde{V}_j(\mathbf{u}) \exp[2\pi i(\mathbf{u} \cdot \mathbf{r}_j)] \quad (12a)$$

Since the potential of a crystal is periodic, its Fourier components $\tilde{V}(\mathbf{u})$ are discrete. $\tilde{V}(\mathbf{u})$ is different from zero if $\mathbf{u} = (h \ k \ l) = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ is a discrete reciprocal-space vector belonging to the reciprocal lattice Λ^* (\mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* are the basic vectors of Λ^*). The structure factor $F(\mathbf{u})$ differ from $\tilde{V}(\mathbf{u})$ only by a multiplying constant which is related to the unit cell volume Ω (Spence and Zuo, 1992)

$$\tilde{V}(\mathbf{u}) = \frac{h^2}{2\pi m e \Omega} F(\mathbf{u}) = \frac{\lambda}{\sigma \Omega} F(\mathbf{u}) \quad (13)$$

where h is Planck's constant, m and e the electron relativistic mass and charge, λ the electron wavelength and σ the interaction constant

$$\sigma = \frac{2\pi m e \lambda}{h^2} \quad (\text{in } \text{V}^{-1} \text{ \AA}^{-1})$$

Taking into account equation (12a) and the relation (13) we can understand why the structure factor is usually written as

$$F(\mathbf{u}) = \sum_{j=1}^N f_j(\mathbf{u}) \exp 2\pi i(\mathbf{u} \cdot \mathbf{r}_j) \quad (\text{in } \text{\AA}) \quad (12b)$$

$f_j(\mathbf{u})$, called atomic scattering factor, is just the Fourier transform of the crystal potential produced by the j^{th} atom and the exponential term is coming from the change of variable $\mathbf{w} = \mathbf{r} - \mathbf{r}_j$ we have done during the calculation of the Fourier integral. The f_j values for all atoms and some ions are listed in the International Tables for Crystallography, Vol. C (1995). The crystal structure factors $F(\mathbf{u})$ are measured in \AA while the Fourier components $\tilde{V}(\mathbf{u})$ of the crystal potential are measured in Volt.

The structure factor $F(\mathbf{u})$ is the sum of all vectors (each with an amplitude $f_j(\mathbf{u})$ and a phase $2\pi \mathbf{u} \cdot \mathbf{r}_j$) for all the atoms in the unit cell, as shown in Fig. 6. $F(\mathbf{u})$ is complex and contains an amplitude part $|F(\mathbf{u})|$ and a phase part $\phi(\mathbf{u})$

$$F(\mathbf{u}) = |F(\mathbf{u})| \exp[i\phi(\mathbf{u})] \quad (14)$$

The phase $\phi(\mathbf{u})$ of the structure factor, unlike the amplitude, does not have a fixed value. It depends on where the origin is chosen in the unit cell. If the origin is shifted by a vector $\mathbf{r}_0 = (x_0 \ y_0 \ z_0)$, the new phase is related to the old phase by

$$\phi_{\text{new}}(\mathbf{u}) = \phi_{\text{old}}(\mathbf{u}) + 2\pi \mathbf{u} \cdot \mathbf{r}_0 \quad (15)$$

For centrosymmetric structures, the origin is always chosen at a centre of symmetry, and then $F(\mathbf{u})$ becomes a real number, i.e. the phase $\phi(\mathbf{u})$ can only be either 0° or 180° (see Fig. 6b).

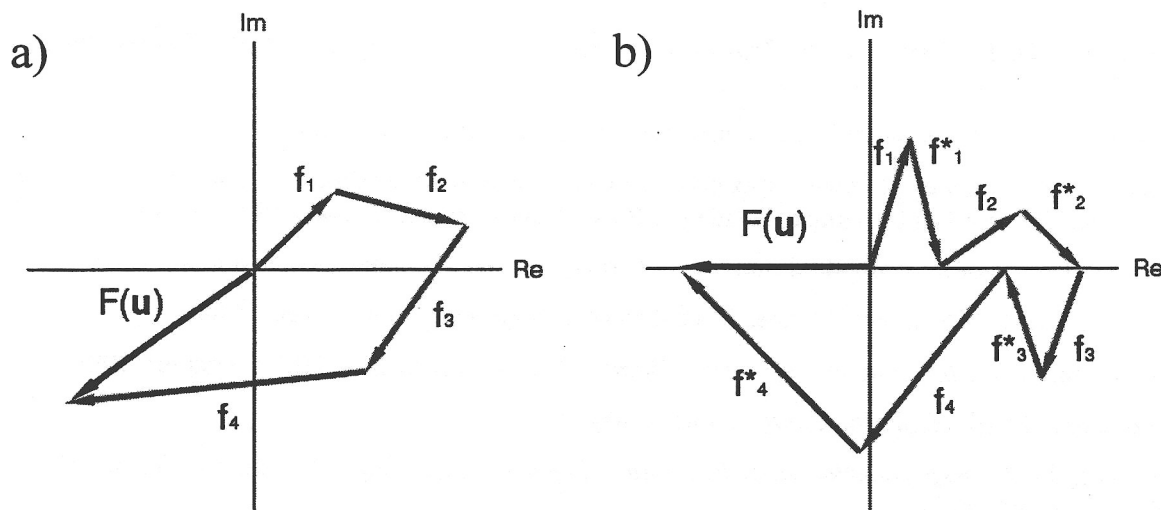


Fig. 6 Fourier synthesis of the crystal structure factor $F(\mathbf{u})$ presented in an amplitude-phase diagram. (a) A noncentrosymmetric crystal with 4 atoms in the unit cell. The atomic number increases with j and so does the scattering power f_j . Note that the structure factor phase $\phi(\mathbf{u})$ is close to the phase $2\pi\mathbf{u}\cdot\mathbf{r}_4$. This shows that heavier atoms play more important roles on the phase $\phi(\mathbf{u})$ than lighter atoms. (b) A centrosymmetric crystal with 4 unique atoms in the unit cell. For each atom pair, one at $\mathbf{r}_j = (x_j, y_j, z_j)$ and the other at $-\mathbf{r}_j = (-x_j, -y_j, -z_j)$, the amplitudes are the same ($= f_j$) while the phases are $2\pi\mathbf{u}\cdot\mathbf{r}_j$ and $-2\pi\mathbf{u}\cdot\mathbf{r}_j$, respectively. The sum of all vectors is always a real number, which means that the phase is either 0° or 180° for centrosymmetric crystals.

Normally, heavy atoms scatter electrons more strongly than light atoms. Thus, the vector sum of only the heavy atoms in a unit cell has a phase close to the phase of the total structure factor, coming from all the atoms in the unit cell, as shown in Fig. 6.

The potential $V(\mathbf{r})$ is a continuous three dimensional function in real space. On the other hand, the structure factors $F(\mathbf{u})$ are discrete points in reciprocal space. They can be represented by a list of amplitudes and phases columns $F_c(\mathbf{hkl})$ and $\phi_F(\mathbf{hkl})$. The potential $V(\mathbf{r})$ can be calculated from the structure factors $F(\mathbf{u})$ by inverse Fourier transformation:

$$V(\mathbf{r}) = \frac{\lambda}{\sigma \Omega} \sum_{\mathbf{h}} F(\mathbf{u}) \exp[-2\pi i(\mathbf{u} \cdot \mathbf{r})] \quad (16)$$

The potential at any point in the crystal can be calculated by adding the vectors $F(\mathbf{u})\exp[-2\pi i(\mathbf{u}\cdot\mathbf{r})]$ for all the structure factors $F(\mathbf{u})$, i.e. by Fourier synthesis. In fact, each vector of reflection \mathbf{u} , together with that of its Friedel pair $-\mathbf{u}$, generate a cosine wave (see Patterson, 1935):

$$F(\mathbf{u})\exp[-2\pi i(\mathbf{u}\cdot\mathbf{r})] + F(-\mathbf{u})\exp[-2\pi i(-\mathbf{u}\cdot\mathbf{r})] = 2|F(\mathbf{u})|\cos[\phi(\mathbf{u})-2\pi(\mathbf{u}\cdot\mathbf{r})] \quad (17)$$

If both the amplitudes $|F(\mathbf{u})|$ and the phases $\phi(\mathbf{u})$ of the structure factors for all reflections \mathbf{u} are known, the potential $V(\mathbf{r})$ can be obtained by adding a series of such cosine waves. Each cosine wave is a set of parallel planes (Bragg planes) of alternating high and low potential.

References

For a basic introduction to image formation and electron microscopy in general a very clear a complete book is

- Williams, D. B., Carter, C. B. *Transmission Electron Microscopy* in 4 Vol. (Plenum Press, New York 1996)

A more advanced but very clear and well written review about image formation is

- Lenz, F. A. Transfer of Image Information in the Electron Microscope in *Electron Microscopy in Material Science* Ed. U. Valdré (Academic Press, New York and London 1971) p.540-569.

Two very good books for studying the Fourier transforms and their relations with diffraction and imaging are

- Cowley, J.M. *Diffraction Physics*, 2nd edition (North-Holland, Amsterdam, 1984).
- Goodman, J.W., *Introduction to Fourier Optics*, 2nd edition (McGraw-Hill, Singapore 1996)

Another book about high resolution electron microscopy is

- Spence, J.C.H. *Experimental High Resolution Electron Microscopy* 2nd edition (Oxford University Press, Oxford 1988)

Other references in the text:

Spence, J.C.H. and J.M. Zuo, *Electron Microdiffraction*, (Plenum Press, New York and London, 1992) pxiv.