(****)

International School of Electron Microscopy, Erice, Italy 9..

'historic' detailed + rigorous

ELECTRON MICROSCOPY IN MATERIAL SCIENCE

"Ettore Majorana" International Centre for Scientific Culture
1970 International School of Electron Microscopy
a NATO Advanced Study Institute
Sponsored by CNR-MPI
Erice, April 4-18

EDITOR
U. VALDRÈ

1971

ACADEMIC PRESS NEW YORK AND LONDON

ACADEMIC PRESS INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS INC. (LONDON) LTD.
Berkeley Square House, London W.1

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 70-157136

PRINTED IN ITALY

Transfer of Image Information in the Electron Microscope

FRIEDRICH ALFRED LENZ

Transfer of Image Information in the Electron Microscope

F. A. LENZ

Institut für Angewandte Physik, Universität Tübingen - Tübingen, Germany

1. General theory.

The purpose of the electron microscope as that of any other optical o electron optical imaging device is to transmit information about propertie of an object to an image. Therefore, we may consider it as an informatio channel and use some of the concepts and methods of information theory t describe the imaging properties of an electron microscope. In order t illustrate some of the basic concepts, let us start with the transfer o a signal which is a function of one variable only. An example is the transmis sion of an electrical signal along a telephone line. In this case, the signa may be a voltage or current, and the variable on which it depends is time The input signal $S_0(t)$ which is entered into the transmission line on the inpu end gives rise to an output signal $S_1(t)$ at the output end of the line. If th transmission line is any good, the receiver at the output end should be abl to conclude from the output signal $S_1(t)$ he is receiving on at least some o the information contained in the input signal $S_0(t)$. In the case of an imaging device the input and output signals depend on at least two variables x and y if an object surface is imaged to an image surface. x and y may stand fo co-ordinates in these surfaces. If three-dimensional information on the objec is to be transmitted, the input and output signals are functions of thre variables. Using vector denotation, an input signal $S_0(r_0)$ is fed into the trans-mission system at the object (input), and an output signal $S_1(r_1)$ is received at the image (output). If the imaging device is any good, the receiver at the output end should be able to conclude from the output signal $S_1(r_1)$ (the image) on at least some of the information contained in the input signal $S_0(r_0)$. If the transmission system is free of noise, the output signal S_1 will depend only on the input signal S_0 and on nothing else. Noise does not have to be audible: In the case of the electron microscope it means the source of any part of the output signal S_1 which is not due to the input signal S_0 but to such causes as mechanical vibrations of the microscope column, granularity of the photographic emulsion or fingerprints of a technical assistant on the micrograph.

Let us first neglect noise, not because there is not any but because it makes the theory simpler. Then there is a unique relation between input and output signal. In other words: Two or more different shots of the same object taken under exactly equal conditions should give two or more exactly identical micrographs. If there is some noise on the transmission line, and one has a reproducible input signal, one better records the output signal repeatedly in order to be able to distinguish which part of the output signal is real information and which part is due to noise.

We have seen that, neglecting noise, there is some unique relation between input and output signal. We call a system linear if this relation is linear. In other words, if the response of the system to one input signal S_0 is S_1 and the response to another input signal S_0' is S_1', then an input signal $\alpha S_0+\beta S_0'$ would, in a linear system, produce an output signal $\alpha S_1+\beta S_1'$ for arbitrary α and β. It is easier to treat linear than nonlinear systems. We shall therefore take care to define our input and output signals S_0 and S_1 so that they are related to each other by a linear relation at least to a good approximation. The transfer of electrical signals in electrical transmission lines can be made well enough linear. If, in an electron microscope, we define input and output signals as the amplitudes of an electron wave in the object and the image, they are also linearly related. This follows directly from the linearity of Schrödinger's or Dirac's wave equation. If we declare the mass thickness of the object as the input signal and the optical density of the developed photographic plate as the output signal, the linearity between input and output are no longer self-evident but at best a tolerable approximation. Most transfer theories are restricted to the case of linear transfer.

One function which can be used to describe the relation between input and output signals in a linear system is its impulsive response $G(t, t')$ or $G(r_1, r_0)$, respectively. It describes the response of the system to a short pulse $S_0(t)=\delta(t-t_0)$ in a one-dimensional transfer system or to an object consisting of one point only in an image transfer system, *i.e.* $S_0(r_0)=\delta(r_0-r_0')$. The delta function describing the short pulse or the object point, respectively,

is defined so that $\delta(t-t_0)$ equals zero for all times $t\neq t_0$ but is so large for $t=t_0$ that

$$\int \delta(t-t_0)\,\mathrm{d}t=1 \tag{1.1}$$

if the interval of integration contains the time $t=t_0$. If the interval of integration does not contain $t=t_0$, the value of the integral equals zero. Correspondingly, the delta function in two-dimensional space is defined so that

$$\iint \delta(r_0-r_0')\,\mathrm{d}r_0=\iint \delta(x_0-x_0')\,\delta(y_0-y_0')\,\mathrm{d}x\,\mathrm{d}y=1 \tag{1.2}$$

if the two-dimensional interval of integration contains the point r_0' with the co-ordinates x_0', y_0'. Otherwise, the value of the integral equals zero. This definition of the delta function can be extended to more than two dimensions. The definition of the delta function implies that

$$\int_{-\infty}^{+\infty} A(t')\,\delta(t-t')\,\mathrm{d}t'=A(t)\,;\qquad \iint A(r_0')\,\delta(r_0-r_0')\,\mathrm{d}r_0'=A(r_0)\,. \tag{1.3}$$

In other words: Any arbitrary function $A(t)$ can be written as a linear superposition of delta functions $\delta(t'-t)$ with a weight function $A(t')$. Since we have assumed that $G(t, t')$ is the response of the linear system to the input signal $\delta(t-t')$, the output signal $S_1(t)$ of an arbitrary input signal

$$S_0(t)=\int_{-\infty}^{+\infty} S_0(t')\delta(t-t')\,\mathrm{d}t' \tag{1.4}$$

can be written as

$$S_1(t)=\int_{-\infty}^{+\infty} S_0(t')G(t, t')\,\mathrm{d}t'\,. \tag{1.5}$$

The transfer properties of a linear transfer system are therefore completely described by its impulsive response $G(t, t')$. Mathematicians and theoretical physicists refer to the impulsive response as « Green's function ». For signals with more than one dimension, we have correspondingly

$$S_1(r_1) = \iint S_0(r_0) G(r_1, r_0)\, \mathrm{d}r_0\,, \tag{1.6}$$

where $G(r_1, r_0')$ is the impulsive response of the linear imaging system to a delta function $\delta(r_0 - r_0')$. The integration (1.6) is extended over the object surface.

The transfer properties of a good electric transmission line should not depend on time. In other words: If the same message $S_0(t)$ is transmitted at two different times t_1 and t_2, say today and tomorrow, then the two input signals $S_0(t-t_1)$ and $S_0(t-t_2)$ should produce the same output signals $S_1(t-t_1)$ and $S_1(t-t_2)$, apart from a shift $t_2 - t_1$ in time. This independence of the transfer properties on time can be expressed by saying that the impulsive response is a function not of the two separate variables t and t' but only a function of one variable, *viz.* the difference $t-t'$:

$$G(t, t') = G(t-t')\,. \tag{1.7}$$

The response to a short pulse at time $t = t'$ will be the same as to a pulse at $t = t''$, only with a time delay of $t'' - t'$ between both. If the signal has more than one dimension such as in imaging devices, the corresponding property of the system would be that the image disk of an object point at position $r_0 = r_0'$ is the same as the image disk of an object point at $r_0 = r_0''$, only displaced to another position in the image. The shift in the image may be different from $r_0'' - r_0'$ because the image may be magnified with respect to the object. This desirable property of an imaging system that all object points at $r_0 = r_0'$ would produce an image disk of equal shape around the point Mr_0' (M is the magnification) in the image plane or in the image space is called isoplanacy. It can be expressed by saying that the impulsive response is a function not of two separate vectors r_1 and r_0 but only of the difference $r_1 - Mr_0$.

$$G(r_1, r_0) = G\left(\frac{r_1}{M} - r_0\right). \tag{1.8}$$

The condition of isoplanacy is not precisely satisfied in optical and electron optical imaging systems. If the system has aberrations depending on r_0 such as distortion, third-order astigmatism, or coma, the isoplanacy condition (1.8) is violated, *i.e.* the image disk of an off-axis point looks different from that

of an axis point. Aberrations depending only on the initial *direction* of an electron trajectory such as spherical aberration, defocusing, axial astigmatism and axial coma do not affect isoplanacy. If the field of view is sufficiently small, the condition of isoplanacy can always be considered to be approximately satisfied.

In the isoplanatic approximation we can write eqs (1.5) and (1.6) as

$$S_1(t) = \int_{-\infty}^{+\infty} S_0(t') G(t-t')\, \mathrm{d}t' \tag{1.9}$$

and

$$S_1(r_1) = \iint S_0(r_0)\, G\left(\frac{r_1}{M} - r_0\right) \mathrm{d}r_0\,. \tag{1.10}$$

Integrals of this type are called convolution integrals. To understand the physical meaning of the linear relation between the input signal S_0 and the output signal S_1 the following consideration may be useful.

The input signal which we have considered above as a linear superposition of delta functions, can, according to Fourier's theorem, also be considered as a linear superposition of sinusoidal functions:

$$S_0(t) = \int_{-\infty}^{+\infty} s_0(f) \exp\left[-2\pi i f t\right] \mathrm{d}f\,. \tag{1.11}$$

Because of the linearity of the transfer system, each Fourier component $s_0(f)\exp[-2\pi i f t]$ of the input signal corresponding to a frequency f can be transformed to the corresponding Fourier component of the output signal and then summed up (or rather integrated up). In other words: In the expression

$$S_1(t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} s_0(f) \exp\left[-2\pi i f t'\right] \mathrm{d}f\, G(t-t')\, \mathrm{d}t' \tag{1.12}$$

we can *first* integrate over t' and *then* over f. The integration over t' is nothing else but a Fourier transform of G:

$$\int_{-\infty}^{+\infty} \exp\left[-2\pi i f t'\right] G(t-t')\, \mathrm{d}t' = \exp\left[-2\pi i f t\right] \int_{-\infty}^{+\infty} \exp\left[2\pi i f t'\right] G(t')\, \mathrm{d}t'\,. \tag{1.13}$$

Equation (1.12) and (1.13) can be interpreted as follows: The calculation of the output signal $S_1(t)$ from a given input signal $S_0(t)$ can be performed in the following steps: First, the Fourier transform $s_0(f)$ of $S_0(t)$ is formed. Then $s_0(f)$ is multiplied by the Fourier transform of the impulsive response G to obtain the Fourier transform of $S_1(t)$. The Fourier transform $T(f)$ of the impulsive response G is called the *transfer function* of the system:

$$T(f)=\int_{-\infty}^{+\infty}\exp[2\pi ift]\,G(t)\,\mathrm{d}t\,. \tag{1.14}$$

According to eqs. (1.12) and (1.13), the product of s_0 with the transfer function $T(f)$ yields the output signal by another Fourier transform:

$$S_1(t)=\int_{-\infty}^{+\infty}s_0(f)T(f)\exp[-2\pi ift]\,\mathrm{d}f\,. \tag{1.15}$$

Let us for a while assume that the input signal is a sine or cosine function:

$$S_0(t)=A\exp[-2\pi if_0t]\,. \tag{1.16}$$

A comparison with eq. (1.11) shows that this is equivalent with a Fourier transform $s_0(f)$ of S_0 (an « input spectrum »)

$$s_0(f)=A\delta(f-f_0)\,. \tag{1.17}$$

As we have seen, the output spectrum $s_1(f)$ is obtained by multiplying the input spectrum $s_0(f)$ by the transfer function

$$s_1(f)=AT(f)\,\delta(f-f_0)\,. \tag{1.18}$$

The output signal $S_1(t)$ is, according to eq. (1.15), the Fourier transform of the output spectrum $s_1(f)$:

$$S_1(t)=\int_{-\infty}^{+\infty}s_1(f)\exp[-2\pi ift]\,\mathrm{d}f=AT(f_0)\exp[-2\pi if_0t]\,. \tag{1.19}$$

In other words: A sinusoidal input function with frequency f_0 and amplitude A is received at the output as a function of the same frequency but with an amplitude $AT(f_0)$. A transmission system for which $T(f_0)$ equals one for all frequencies f_0 would be an ideal system because the output signal would always be identical with the input signal. According to Fourier's theorem any arbitrary input function can be written as a linear superposition of sinusoidal functions with different frequencies f and amplitudes $s_0(f)$. Each of them is transmitted and forms a Fourier component $s_1(f)=T(f)s_0(f)$ at the output. They only have to be linearly superimposed to form the output signal $S_1(t)$.

If the signals are two-dimensional as in image transfer systems the same reasoning can be applied. We have, however, to use different variables $\boldsymbol{r}_0$ and $\boldsymbol{r}_1$ in the input and output signals, respectively, because the co-ordinates in the object and in the image plane do not have the same meaning.

Let us define the Fourier transforms of input and output signal and of the impulsive response G

$$S_0(\boldsymbol{r}_0)=\iint s_0(\boldsymbol{f})\exp[-2\pi i\boldsymbol{f}\boldsymbol{r}_0]\,\mathrm{d}\boldsymbol{f}\,;\quad s_0(\boldsymbol{f})=\iint S_0(\boldsymbol{r}_0)\exp[2\pi i\boldsymbol{f}\boldsymbol{r}_0]\,\mathrm{d}\boldsymbol{r}_0\,; \tag{1.20}$$

$$S_1(\boldsymbol{r}_1)=\iint s_1(\boldsymbol{f})\exp\left[-\frac{2\pi i}{M}\boldsymbol{f}\boldsymbol{r}_1\right]\mathrm{d}\boldsymbol{f},\quad s_1(\boldsymbol{f})=\iint S_1(\boldsymbol{r}_1)\exp\left[\frac{2\pi i}{M}\boldsymbol{f}\boldsymbol{r}_1\right]\frac{\mathrm{d}\boldsymbol{r}_1}{M^2}, \tag{1.21}$$

$$T(\boldsymbol{f})=\iint G(\boldsymbol{t})\exp[2\pi i\boldsymbol{f}\boldsymbol{t}]\,\mathrm{d}\boldsymbol{t}\,. \tag{1.22}$$

In eq. (1.22), $\boldsymbol{t}$ stands as a substitution for

$$\boldsymbol{t}=\frac{\boldsymbol{r}_1}{M}-\boldsymbol{r}_0\,. \tag{1.23}$$

As in the case of one-dimensional signal functions it can again be shown that it follows from eq. (1.10) that

$$s_1(\boldsymbol{f})=T(\boldsymbol{f})s_0(\boldsymbol{f})\,. \tag{1.24}$$

The « space frequency » $\boldsymbol{f}$ is now a vector with two components f_x and f_y. The area elements $\mathrm{d}\boldsymbol{r}_0$ and $\mathrm{d}\boldsymbol{r}_1$ in eqs (1.20) and (1.21) stand for

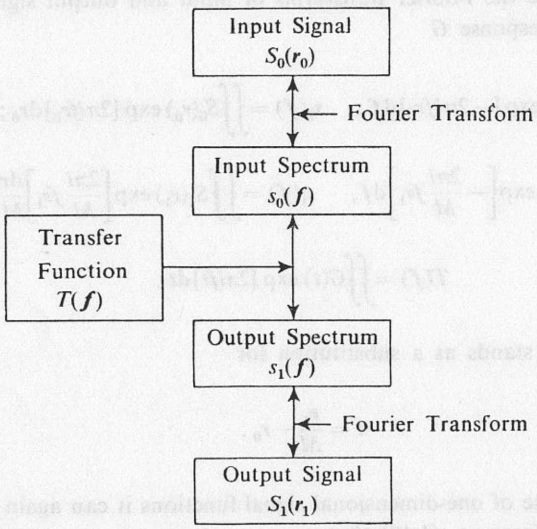
$$\mathrm{d}\boldsymbol{r}_0=\mathrm{d}x_0\mathrm{d}y_0\,,\quad \mathrm{d}\boldsymbol{r}_1=\mathrm{d}x_1\mathrm{d}y_1\,, \tag{1.25}$$

such as the element df stands for

$$df = df_x \, df_y \,. \qquad (1.26)$$

The Fourier transforms (1.20) correspond to the expansion of the input (= object) signal in a series of sinusoidal components each of which is denoted by its space frequency f and its amplitude $s_0(f)$. The vector f with the components f_x and f_y denotes the direction of the sinusoidal component (plane wave) associated with each Fourier component. The vector f is perpendicular to the wave fronts of this plane wave, and its length $|f|$ is the inverse of the repeat of the sinusoidal component.

Using the concept of the transfer function, the linear relation between the input and output signal can be described by the following diagram.



If the transfer function or the impulsive response of a system is known, the relation between S_0 and S_1 is uniquely defined, and one can conclude on S_1 if S_0 is known and *vice versa*. If, on the other hand, the relation between S_0 and S_1 were known empirically by taking a great number of micrographs of different objects with known properties, one would be able to determine the transfer function $T(f)$.

2. Amplitude transfer and contrast transfer function.

In the general theory we have derived relations between input and output signals in linear transmission systems but we have not specified what the physical nature of these signals is in electron microscopy. Since the image is transferred by means of electrons and since the propagation of these electrons in space can be described by a linear wave equation such as Schrödinger's, an obvious definition would be to identify the input signal with the wave amplitude in the object plane and the output signal with the wave amplitude in the image plane. The condition of linearity is exactly fulfilled in this case. The transfer function can be derived from a study of the propagation of the electron wave through the lenses and apertures of the electron optical imaging system.

The input signal depends on the conditions of illumination and on the interaction of the illuminating beam with the object. Let us first assume the illumination to be coherent in direction of the optical axis which we identify with the z axis of a Cartesian or cylindrical system of co-ordinates The wave amplitude of the incoming primary wave from the condenser, before it enters the object, would be a plane wave $\exp[2\pi ikz]$ with a wave number k depending on the acceleration voltage U

$$k = \frac{1}{\lambda} = \frac{1}{h}\sqrt{2em_0U\left(1+\frac{eU}{2m_0c^2}\right)} = \frac{1}{h}\sqrt{2em_0U^*} \,. \qquad (2.1)$$

In eq. (2.1), h denotes Planck's constant. In high-resolution transmission microscopy the object can be considered as nonabsorbing. Practically all electrons entering the object from the condenser side leave it again on the image side because the probability for all interactions removing electrons from the beam, such as backscattering or bremsstrahlung production close to the short wavelength limit, is very small. The interaction between the primary electron beam and a thin object can therefore be understood as a local distortion of the electron wavefronts due to the local variations of the electrostatic potential within and between the atoms. Within an atom the potential is more positive than in the surrounding vacuum, and consequently the local wavelength is shorter than the vacuum wavelength. The resulting distortion of the wavefronts may be referred to as phase-shifting, diffraction or scattering, three different names for the same physical process.

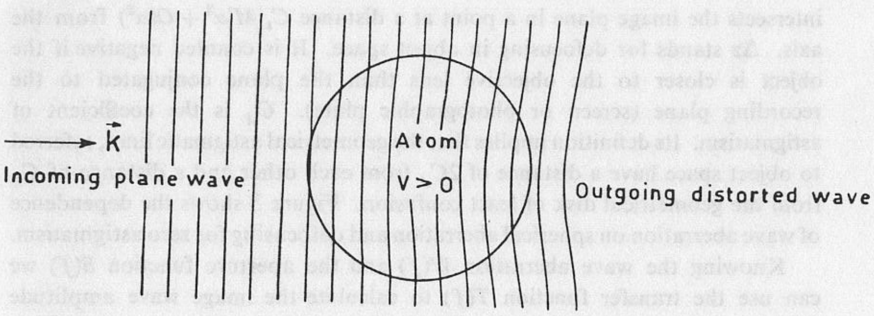


Fig. 1. – Interaction of atom and electron wave.

The amplitude of the distorted wave after passing through the object, which, without this interaction, would be a constant, is now

$$S_0(x_0, y_0) = \exp[i\eta(x_0, y_0)], \quad S_0(\boldsymbol{r}_0) = \exp[i\eta(\boldsymbol{r}_0)] \tag{2.2}$$

a complex function of the co-ordinates x_0, y_0 in the object plane. S_0 is the input signal, and $\eta(\boldsymbol{r}_0)$ is the phase shift. The fact that the object is treated as nonabsorbing is expressed by the constance of object current density

$$j(\boldsymbol{r}_0) \sim |S_0(\boldsymbol{r}_0)|^2 = 1 \tag{2.3}$$

immediately behind the object.

For weak phase objects, *i.e.* if $\eta(\boldsymbol{r}_0)$ is small compared with 2π for all $\boldsymbol{r}_0$, we have

$$\eta(\boldsymbol{r}_0) = 2\pi \int_0^t \left(\frac{1}{\lambda(x_0, y_0, z)} - \frac{1}{\lambda_0}\right) dz = \\ = \frac{2\pi e m \lambda}{h^2} \int_0^t \varphi(x_0, y_0, z)\, dz = \frac{2\pi e}{h v} \int_0^t \varphi(x_0, y_0, z)\, dz . \tag{2.4}$$

The integration in eq. (2.4) is extended over the thickness t of the object. Equation (2.4) is relativistically correct if the relativistic expressions for m, λ and v are used.

All information about the object which the electron wave is carrying is contained in $S_0(\boldsymbol{r}_0)$ or $\eta(\boldsymbol{r}_0)$, respectively. In order to calculate the corresponding

output signal $S_1(\boldsymbol{r}_1)$, *i.e.* the wave amplitude in the image plane, we have to know the transfer function T or its Fourier transform, the impulsive response. The impulsive response G is the response of the imaging system to a point source in the object, *i.e.* the image wave amplitude in the diffraction disk which forms the image of an object point. The classical method of determining the wave amplitude in the diffraction disk is the application of Kirchhoff's integral formula. If the surface of integration in Kirchhoff's integral is the back focal plane of the objective lens, the evaluation of the integral is equivalent to the Fourier transform leading from the output spectrum s_1 to the output signal S_1. Kirchhoff's integration is extended only over the transparent part of the objective aperture. It has further to take into account the phase shift due to aberrations and defocusing. This phase shift is closely related to the wave aberration which is defined as the local distance between the real wave front and an ideal wave front, *i.e.* a sphere around the geometrical image point.

Each point in the back focal plane of the objective lens corresponds to one space frequency $\boldsymbol{f}$. If the object were a periodic structure whose object signal contained only one or a small number of space frequencies, then the wave function in the back focal plane would be zero except for steep local intensity maxima, one for each space frequency. In other words, the wave function in the back focal plane is the diffraction pattern of the object with a diffraction length equal to the focal length of the objective lens. Each space frequency $\boldsymbol{f}$ in the object (input signal) corresponds to one Bragg angle, *i.e.* one direction of a diffracted wave in object space. Each direction in object space corresponds to one point in the back focal plane. These two statements can be combined into one, saying that each space frequency $\boldsymbol{f}$ corresponds to one point $\boldsymbol{r}_B$ in the back focal plane

$$\boldsymbol{r}_B = l\lambda \boldsymbol{f} . \tag{2.5}$$

In eq. (2.5) l denotes the focal length of the objective lens (the letter f being reserved for space frequencies). Equation (2.5) should look familiar to people who have worked with electron diffraction of crystals where the position vector $\boldsymbol{r}$ of an intensity maximum in the diffraction diagram is equal to the product of the diffraction length l, the wavelength λ and the reciprocal lattice vector $\boldsymbol{f}$ denoting a space frequency (inverse of the spacing of lattice planes) in the periodic structure of the crystal. The effect of the aberrations is to shift the phase of the wave function in the back focal plane where the phase shift depends on $\boldsymbol{r}_B$ which, according to eq. (2.5) can be interpreted as a phase shift depending on space frequency.

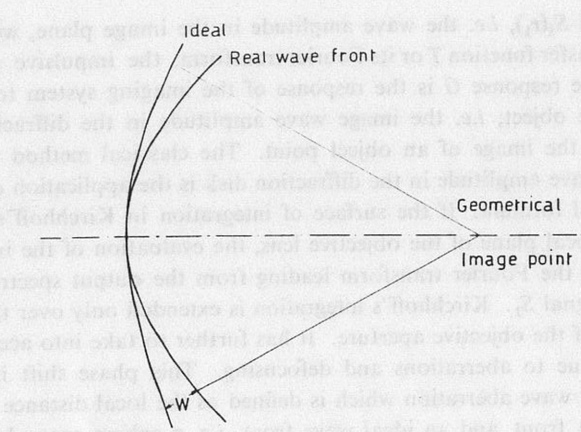


Fig. 2.

The first Fourier transformation transforming the input signal $S_0(r_0)$ into the input spectrum $s_0(f)$ corresponds to the formation of the diffraction pattern of the object neglecting lens aberrations and apertures. These are taken into account by the transfer function

$$T(f) = \frac{1}{M}\exp\left[\frac{2\pi i}{\lambda}W(f)\right]B(f), \tag{2.6}$$

which describes the phase shift $2\pi W/\lambda$, and the effect of an aperture by the aperture function $B(f)$. W is the wave aberration introduced by aberrations and defocusing. $B(f)$ is assumed to equal 1 in the transparent parts of the aperture, and to vanish for the opaque parts. If the lens suffers from spherical aberration, axial astigmatism and defocusing the wave aberration can be written as

$$\left.\begin{aligned} W(r_B) &= \frac{C_s}{4l^4}(x_B^2+y_B^2)^2 + \frac{\Delta z}{2l^2}(x_B^2+y_B^2) - \frac{C_A}{2l^2}(x_B^2-y_B^2),\\ W(f) &= \frac{C_s}{4}\lambda^3 f^4 + \frac{\Delta z}{2}\lambda^2 f^2 - \frac{C_A}{2}(f_x^2-f_y^2)\lambda^2. \end{aligned}\right\} \tag{2.7}$$

In eq. (2.7), C_s is the third-order spherical aberration coefficient. Its definition is the usual one, *i.e.* it implies that a geometrical electron trajectory leaving the axis point of the object plane under an angle α against the axis

intersects the image plane in a point at a distance $C_s|M|\alpha^3 + O(\alpha^5)$ from the axis. Δz stands for defocusing in object space. It is counted negative if the object is closer to the objective lens than the plane conjugated to the recording plane (screen or photographic plate). C_A is the coefficient of astigmatism. Its definition implies that the geometrical astigmatic lines, referred to object space have a distance of $2C_A$ from each other and a distance of C_A from the geometrical disk of least confusion. Figure 3 shows the dependence of wave aberration on spherical aberration and defocusing for zero astigmatism.

Knowing the wave aberration $W(f)$ and the aperture function $B(f)$ we can use the transfer function $T(f)$ to calculate the image wave amplitude $S_1(r_1)$ if we know the object wave amplitude $S_0(r_0)$, *i.e.* we can conclude from a given object on the corresponding image and *vice versa*. But unfortunately, wave amplitudes are not observable quantities. What we can observe in the image are such quantities as current density, contrast, optical density, etc., and they are not linearly related to any property of the object. It can, how-

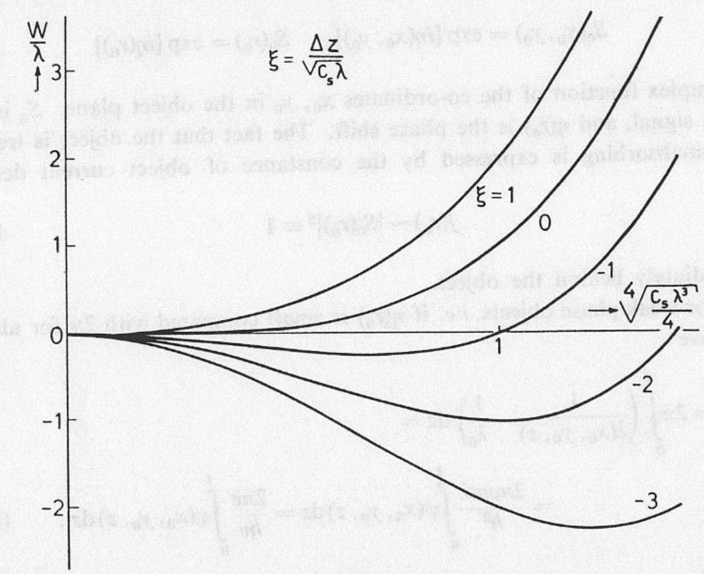


Fig. 3.

ever, be shown that the contrast in the image of a weak phase object is at least approximately a linear function of the phase shift η. To show this, let us treat the amplitude transfer of a weak phase object. For $\eta \ll 2\pi$, eq. (2.2) can be written as

$$S_0(r_0) = \exp[i\eta(r_0)] = 1 + i\eta(r_0) + O(\eta^2). \tag{2.8}$$

The object spectrum follows by Fourier transformation, neglecting second and higher-order terms in η,

$$s_0(f) = \int S_0(r_0) \exp[2\pi i f r_0] dr_0 = \delta(f) + i \int \eta(r_0) \exp[2\pi i f r_0] dr_0. \tag{2.9}$$

$s_0(f)$ describes the angular distribution of the wave behind the object. The delta function stands for the undiffracted primary beam in axial direction. The second term on the right-hand side is the complex scattering amplitude of the object. If $\eta(r_0)$ in eq. (2.9) is replaced by the expression in eq. (2.4) one obtains

$$s_0(f) = \delta(f) + \frac{ie}{\hbar v} \iiint \varphi(x_0, y_0, z) \exp[2\pi i f r_0] dx_0 dy_0 dz. \tag{2.10}$$

We see that the second term on the right-hand side is a three-dimensional Fourier transform of the potential distribution within the scatterer. The integral on the right-hand side is known as the scattering amplitude of the scatterer. In the special case that the scatterer is an atom, it is called the atom form amplitude. Its absolute square is the differential scattering cross-section. Let us introduce an abbreviation $A(f)$ for this quantity:

$$A(f) = \int \eta(r_0) \exp[2\pi i f r_0] dr_0, \tag{2.11}$$

so that eq. (2.9) can be written as

$$s_0(f) = \delta(f) + iA(f). \tag{2.12}$$

According to eq. (1.24), the image (output) spectrum $s_1(f)$ follows from the object (input) spectrum by multiplication with the amplitude transfer function

$$s_1(f) = T(f) s_0(f) = T(0)\,\delta(f) + iA(f)T(f). \tag{2.13}$$

Performing the inverse Fourier transformation we obtain the output signal

(the image wave amplitude)

$$S_1(r_1) = T(0) + i \int A(f) T(f) \exp\left[-2\pi i f \frac{r_1}{M}\right] df. \tag{2.14}$$

In bright field microscopy, $B(0) = 1$, and it follows from eq. (2.6) that $T(0) = 1/M$. The first term on the right-hand side of eq. (2.14) describes the bright background of the bright field image whose current density is M^{-2} times the primary current density in the object. The second term describes a small modulation of this background. It is small because we have assumed the phase shift η is small and because A is defined as the Fourier transform of this phase shift. In dark field microscopy, $B(0) = T(0) = 0$, and the background is dark. It is evident from eq. (2.14) that the contrast in a dark field image of a weak phase object exceeds that of the bright field image.

Let us define contrast C in the bright field image by

$$C(r_1) = \frac{|S_1(r_1)|^2 - 1/M^2}{1/M^2} = M^2 |S_1(r_1)|^2 - 1. \tag{2.15}$$

Replacing S_1 in eq. (2.15) from eq. (2.14) and neglecting second order terms we obtain

$$C(r_1) = iM \int A(f) [(T(f) - T^*(-f)] \exp\left[-2\pi i f \frac{r_1}{M}\right] df. \tag{2.16}$$

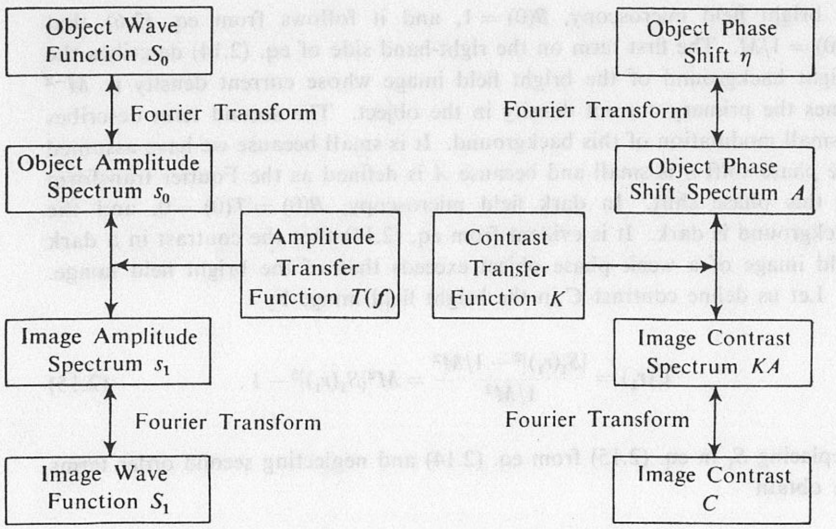
If the aperture function $B(f) = B(-f)$, *i.e.* if $B(f)$ has two-fold symmetry around the optical axis and if further $W(f) = W(-f)$ then we have

$$C(r_1) = 2 \int A(f) B(f) \sin\left(\frac{2\pi}{\lambda} W(f)\right) \exp\left[-2\pi i f \frac{r_1}{M}\right] df. \tag{2.17}$$

Equations (2.17) and (2.11) define a linear relation between the real contrast $C(r_1)$ and the real phase shift $\eta(r_0)$. If we define a contrast transfer function

$$K(f) = 2B(f) \sin\left(\frac{2\pi}{\lambda} W(f)\right), \tag{2.18}$$

this relation can be interpreted as follows: The input signal for contrast transfer is now $\eta(r_0)$. Its input spectrum $A(f)$ is multiplied by the contrast transfer function $K(f)$ to obtain the output spectrum. The inverse Fourier transform (2.17) then generates the output signal, *i.e.* the image contrast. This is explained in the following diagrams.



While phase contrast can be understood and explained only using wave optical aspects, another type of contrast has been discussed since the early days of electron microscopy, the so-called scattering absorption or amplitude contrast. It can be explained without using wave-optical concepts by saying that the atoms in the object scatter a fraction of the incoming electron current by scattering angles large enough to be intercepted by the objective aperture. The characteristic features of this type of contrast can also be explained in terms of the amplitude transfer theory. Let us suppose that the phase shift $\eta(r_0)$ is so large that it makes sense to continue the expansion (2.8) by an additional second-order term:

$$S_0(r_0) = \exp[i\eta(r_0)] = 1 + i\eta(r_0) - \frac{\eta^2(r_0)}{2} + O(\eta^3)\,. \tag{2.19}$$

Let us, for the sake of simplicity, consider a sinusoidal variation of phase

shift $\eta(r_0)$ with the space frequency $f_x = f_0$; $f_y = 0$:

$$\eta(r_0) = \eta_0 \cos(2\pi f_0 x_0)\,. \tag{2.20}$$

Then the object wave function is

$$S_0(r_0) = 1 + \frac{i}{2}\eta_0[\exp[2\pi i f_0 x_0] + \exp[-2\pi i f_0 x_0]] - \\ -\frac{1}{8}\eta_0^2[\exp[4\pi i f_0 x_0] + 2 + \exp[-4\pi i f_0 x_0]] + O(\eta_0^3)\,. \tag{2.21}$$

The corresponding object spectrum is

$$s_0(f) = \delta(f) + \frac{i}{2}\eta_0\delta(f_y)[\delta(f_x + f_0) + \delta(f_x - f_0)] - \\ -\frac{\eta_0^2}{8}\delta(f_y)[\delta(f_x + 2f_0) + 2\delta(f_x) + \delta(f_x - 2f_0)] + O(\eta_0^3)\,. \tag{2.22}$$

Let us now assume that the space frequency is so high, and the objective aperture is so narrow that $T(f_0, 0)$ and $T(2f_0, 0)$ both vanish. In this case we have an image spectrum

$$s_1(f) = T(f)s_0(f) = \frac{1}{M}\delta(f)\left(1 - \frac{1}{4}\eta_0^2\right) \tag{2.23}$$

and

$$S_1(r_1) = \frac{1}{M}\left(1 - \frac{1}{4}\eta_0^2\right). \tag{2.24}$$

The effect is a uniform reduced background intensity, and the space frequency f_0 is not resolved. An example is a thin foil of some amorphous material imaged under conditions at which the atoms or other local variations of potential are not resolved. Then regions containing many such atoms or potential variations appear darker in the image than regions containing less scatterers. This type of « area » contrast is compared with phase contrast in the following table.

	Space frequencies f_0 and $2f_0$ intercepted by aperture $T(f_0)=T(2f_0)=0$	$2f_0$ intercepted, f_0 not intercepted $T(f_0)\neq 0;\ T(2f_0)=0$	f_0 and $2f_0$ not intercepted $T(f_0)\neq 0;\ T(2f_0)\neq 0$
η small $\sin\eta \ll 1$ $\eta^2 \ll \eta$	No contrast	Phase contrast linear in η, f_0 resolved	Phase contrast linear in η, f_0 resolved
η larger	Amplitude contrast proportional with η^2, f_0 not resolved	Nonlinear phase contrast containing higher harmonics. Loss in background intensity, f_0 resolved	

Example: Image of a phase edge.

Let us assume that an object consists of two half-planes each of which is homogeneous, but because of a difference in thickness or in mean inner potential they produce different phase shifts:

$$S_0(r_0)=\begin{cases}\exp[-i\varphi/2]\,, & \text{for } x_0\leqslant 0\,,\\ \exp[i\varphi/2]\,, & \text{for } x_0>0\,.\end{cases} \tag{2.25}$$

In order to simplify the problem let us assume that

$$T(f)=\begin{cases}1\,, & \text{for } |f|\leqslant f_0\,,\\ 0\,, & \text{for } |f|>f_0\,.\end{cases} \tag{2.26}$$

This corresponds to a circular objective aperture within which the wave aberration is negligible.

The object spectrum is

$$s_0(f)=\int S_0(r_0)\exp[2\pi i f r_0]\,\mathrm{d}r_0=\delta(f_y)\left[\delta(f_x)\cos\frac{\varphi}{2}-\frac{1}{\pi f_x}\sin\frac{\varphi}{2}\right]. \tag{2.27}$$

Multiplication with the contrast transfer function (2.26) and Fourier transfor-

mation yields the image wave function

$$S_1(r_1)=\cos\left(\frac{\varphi}{2}\right)+\frac{2i}{\pi}\sin\left(\frac{\varphi}{2}\right)\mathrm{Si}\left(\frac{2\pi f_0 x_1}{M}\right), \tag{2.28}$$

where

$$\mathrm{Si}(u)=\int_0^u\frac{\sin v}{v}\,\mathrm{d}v \tag{2.29}$$

is the « sine integral ». For the image contrast we obtain

$$C(r_1)=\sin^2\left(\frac{\varphi}{2}\right)\left[\frac{4}{\pi^2}\mathrm{Si}^2\left(\frac{2\pi f_0 x_1}{M}\right)-1\right]. \tag{2.30}$$

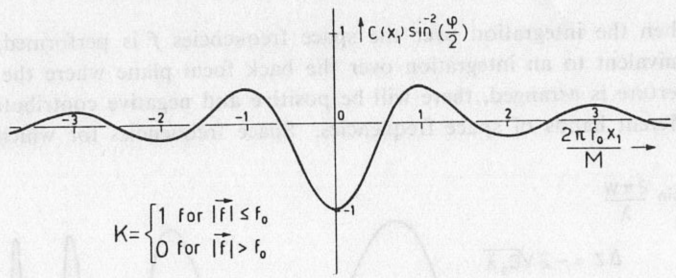


Fig. 4. – Contrast in the image of a phase edge.

3. Zonal plates and other interventions in the back focal plane of the objective.

Let us apply the contrast transfer theory to the case of a phase shifting point in the object, and let us ask the question how the aperture function $B(f)$ must be chosen if we want to achieve maximum bright field contrast in the image of this point. If the phase shifting interaction in the object is assumed to be localized in a point we have

$$\eta(r_0)=\eta_0\,\delta(r_0)\,. \tag{3.1}$$

The corresponding phase spectrum is

$$A(f)=\int\eta(r_0)\exp[2\pi i f r_0]\,\mathrm{d}r_0=\eta_0\,. \tag{3.2}$$

Multiplying by the transfer function we obtain the image contrast spectrum

$$A(f)K(f) = 2\eta_0 B(f) \sin\left(\frac{2\pi}{\lambda} W(f)\right). \tag{3.3}$$

By Fourier transform we obtain the image contrast

$$C(r_1) = 2\eta_0 \int B(f) \sin\left(\frac{2\pi}{\lambda} W(f)\right) \exp\left[-2\pi i f \frac{r_1}{M}\right] \mathrm{d}f. \tag{3.4}$$

The contrast $C(0)$ in the center of the image disk is

$$C(0) = 2\eta_0 \int B(f) \sin\left(\frac{2\pi}{\lambda} W(f)\right) \mathrm{d}f. \tag{3.5}$$

When the integration over the space frequencies f is performed, which is equivalent to an integration over the back focal plane where the objective aperture is arranged, there will be positive and negative contributions from different bands of space frequencies. Space frequencies for which the sine

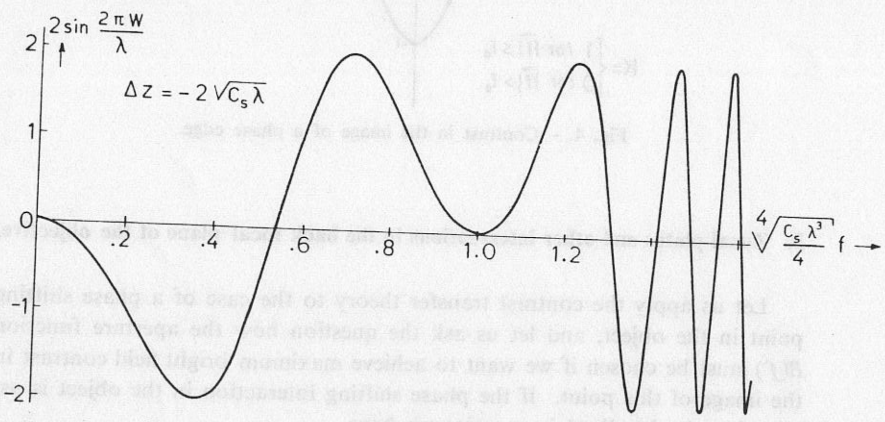


Fig. 5. – Contrast transfer function without aperture, $B(f) = 1$.

function in the integrand has a positive value, add to $C(0)$. For other space frequencies the sine function has a negative sign, and they will cancel at least part of the contrast. Hoppe's idea of using annular ring systems to

improve the electron microscopical image amounts to dimensioning th apertures so that either all negative contributions or all positive contribu tions to $C(0)$ are intercepted by the aperture stop. Figures 5 ÷ 7 show the con

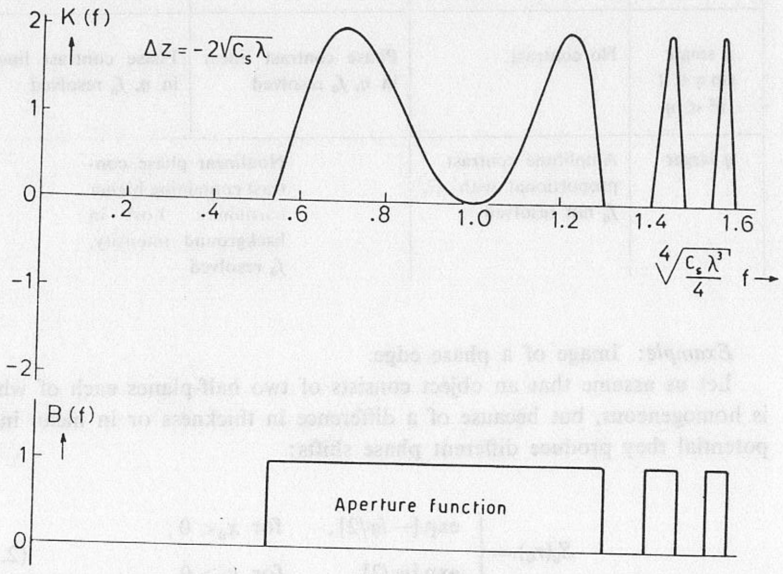


Fig. 6. – Contrast transfer function for maximum positive contrast, $B(0) = 1$.

trast transfer functions K and the aperture functions B for the case that the wave aberration is given by eq. (2.7) with $C_A = 0$ and $\Delta z = -2\sqrt{C_s \lambda}$. Figure 5 shows the contrast transfer function K if no aperture is used ($B \equiv 1$), Fig. 6 for an aperture which leaves through all positive contributions to contrast, and Fig. 7 the same for negative contributions. The aperture system which helps to image an object point with maximum contrast is not necessarily ideal for all other types of objects. Apertures consisting of a system of concentric rings leave through some bands of space frequencies and intercept others. If an observer is interested in properties of an object which are mainly in some fixed space frequency region, then it would be unwise to intercept a frequency band in this region. For example, if an observer is interested in atomic distances of the order of 1 Å, his objective aperture should be transparent in the region of space frequencies around 1 Å^{-1} which corresponds to an aperture radius of $r_B = l\lambda$ Å^{-1} according to eq. (2.5).

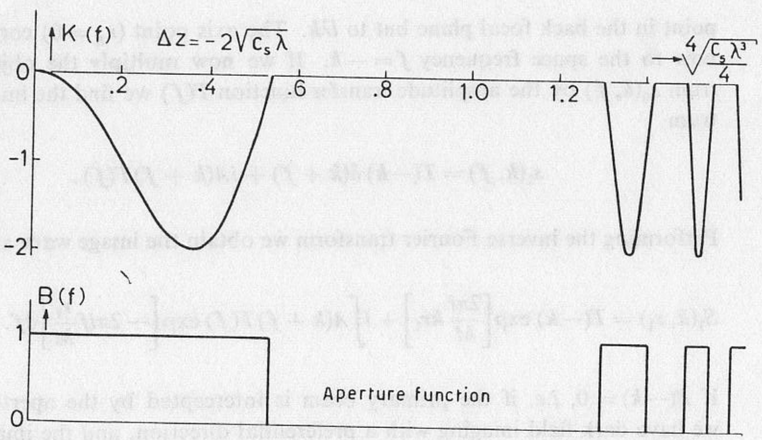


Fig. 7. – Contrast transfer function for maximum negative contrast, $B(0)=1$.

As we have seen, the Hoppe zonal plate is the aperture which optimizes the contrast of a point object in a bright field image. Let us now consider the effect of zonal plates on a dark field image. According to eq. (2.14) the wave amplitude S_1 in dark field is

$$S_1(r_1) = i\int A(f)T(f)\exp\left[-\frac{2\pi i}{M}fr_1\right]\mathrm{d}f. \tag{3.6}$$

According to eq. (3.2) we have for a point object $A=\eta_0$. In the geometrical image $r_1=0$ of this point we have

$$S_1(0) = i\eta_0\int T(f)\,\mathrm{d}f = \frac{i\eta_0}{M}\int \exp\left[-\frac{2\pi i}{\lambda}W(f)\right]B(f)\,\mathrm{d}f. \tag{3.7}$$

As in eq. (3.5) we have again an integrand whose real and imaginary parts are changing their signs. If different space frequency intervals are not to cancel each other's contributions to the absolute value of $S_1(0)$, $B(f)$ must again be chosen so that only ring-shaped areas of the objective aperture are transparent for which

$$n+c<\frac{W}{\lambda}<n+c+\frac{1}{2},\quad n \text{ integer},\ c \text{ arbitrary}. \tag{3.8}$$

For $c=0$, this is the same condition as for maximum positive bright field contrast. For $c=\frac{1}{4}$ it coincides with the condition for maximum negative bright field contrast. Since in dark field microscopy the phase relation of the diffracted electrons with respect to the primary electrons does not matter, any other value of c in the condition for the ring radii would also be acceptable. The most important conclusion is, however, that a Hoppe zone plate designed for maximum contrast in the bright field image of a point object will also maximize the intensity in the center of the dark field image of the same point object.

Most other interventions in the back focal plane such as a filament across the center intercepting the primary beam or narrow circular apertures surrounded by a phase shifting ring may have the effect of increasing contrast in the image of an object but not necessarily in the space frequency region in which the observer is interested. In order to design an optimum aperture one must know the space frequency region of main interest. Then one can design an aperture which produces a maximum of the amplitude or contrast transfer function around this space frequency of main interest. Having done this, one may expect to find this space frequency in all image areas corresponding to object areas in which this space frequency occurs, even if it does so only as a second or higher harmonic of a lower space frequency.

4. The effects of illumination on image transfer.

We have so far restricted ourselves to coherent illumination in the direction of the optical axis. Even when dark field images were discussed, it was assumed that the primary beam had axial direction, and the aperture was symmetric with respect to the axis. In practical dark field microscopy, however, conditions are often different: The primary beam is inclined with respect to the axis so that it does not intersect the back focal plane of the

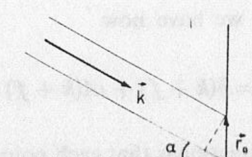


Fig. 8. – Oblique coherent illumination.

objective lens in its axis point but in another one. If the objective aperture is opaque in this point, the primary beam is intercepted, and a dark field image results. It is obvious that oblique illumination introduces a preferential direction in the image. In this case the transfer function is not a function of only the absolute value of the space frequency f but it also depends on its direction.

The oblique coherent illumination produces an additional phase shift $2\pi k r_0$ in excess of the one given in eq. (2.4) describing the interaction of primary beam and object. If we assume that the object is so thin that its thickness times the angle between k and the axis is smaller than the smallest details we want to observe we may treat the object as infinitely thin, and we have instead of eq. (2.4)

$$\eta(k, r_0) = 2\pi k r_0 + \eta(0, r_0)\,. \tag{4.1}$$

In eq. (4.1) and in the following we may treat k as a vector with two components k_x and k_y only, because r_0 lies in the object plane which we have assumed to be perpendicular to the axis. This is because $\eta(k, r_0)$ contains all information about the object which enters the transfer system, and because $\eta(k, r_0)$ does not depend on k_z. We can now apply the transfer theory to determine the image contrast, replacing $\eta(r_0)$ by $\eta(k, r_0)$. Equation (2.8) reads now, in the case of oblique coherent illumination

$$S_0(k, r_0) = \exp\,[i\eta(k, r_0)] = (1 + i\eta(0, r_0))\exp\,[2\pi i k r_0]\,. \tag{4.2}$$

The object spectrum becomes

$$s_0(k, f) = \int S_0(k, r_0)\exp\,[2\pi i f r_0]\,\mathrm{d}r_0 = \\ = \delta(k+f) + i\int \eta(0, r_0)\exp\,[2\pi i(k+f)r_0]\,\mathrm{d}r_0\,. \tag{4.3}$$

Using the abbreviation (2.11) we have now

$$s_0(k, f) = \delta(k+f) + iA(k+f)\,. \tag{4.4}$$

The physical meaning of this equation is that each point in the diffraction pattern of the object in the back focal plane has been shifted from $r_B = l\lambda k$ to $r_B = l\lambda(k+f)$. The primary beam ($f = 0$) no longer corresponds to the axis

point in the back focal plane but to $l\lambda k$. The axis point ($r_B = 0$) correspond now to the space frequency $f = -k$. If we now multiply the object spec trum $s_0(k, f)$ by the amplitude transfer function $T(f)$ we find the image spec trum

$$s_1(k, f) = T(-k)\,\delta(k+f) + iA(k+f)T(f)\,. \tag{4.5}$$

Performing the inverse Fourier transform we obtain the image wave amplitud

$$S_1(k, r_1) = T(-k)\exp\left[\frac{2\pi i}{M}kr_1\right] + i\int A(k+f)T(f)\exp\left[-2\pi i f\frac{r_1}{M}\right]\mathrm{d}f\,. \tag{4.6}$$

If $B(-k) = 0$, *i.e.* if the primary beam is intercepted by the aperture stop we have dark field imaging with a preferential direction, and the image wav amplitude becomes

$$S_1(k, r_1) = i\int A(k+f)T(f)\exp\left[-2\pi i f\frac{r_1}{M}\right]\mathrm{d}f\,. \tag{4.7}$$

If, on the other hand, $B(-k) = 1$ we have a bright field image with a wav amplitude

$$S_1(k, r_1) = \frac{1}{M}\exp\left[-\frac{2\pi i}{\lambda}W(-k)\right]\exp\left[\frac{2\pi i}{M}kr_1\right] + \\ + i\int A(k+f)T(f)\exp\left[-2\pi i f\frac{r_1}{M}\right]\mathrm{d}f\,. \tag{4.8}$$

It is not self-evident that coherent illumination always yields the best images It can be shown that even for an arbitrary incoherent illumination a con trast transfer function can be defined as long as weak phase objects are imaged and the isoplanatic approximation holds. The illumination is called in coherent if the condenser aperture α is large so that the beam can no longer be called parallel. If the variation in wave vector k within the primary beam is so large that the phase differences $2\pi k r_0$ (compare eq. (4.1)) vary by an amount comparable to or larger than 2π, then the phase relations between two points in a distance $|r_0|$ from each other are destroyed, and two such points are « incoherently illuminated ». If, on the other hand, the variation of k is so small that the phase differences $2\pi k r_0$ vary by less than $\pm\pi/2$ then two points at a distance $|r_0|$ from each other are « coherently illuminated »

Whether some illumination is coherent or not, depends not only on the condenser aperture but also on the size $|r_0|$ of the object details one wants to observe.

According to eq. (4.2), the object wave function for an incoming electron with wave vector k can be written as

$$S_0(k, r_0) = S_0(0, r_0) \exp[2\pi i k r_0]. \tag{4.9}$$

According to eq. (1.10), the corresponding image wave function can be written as

$$S_1(k, r_1) = \int S_0(0, r_0) G\left(\frac{r_1}{M} - r_0\right) \exp[2\pi i k r_0] \, dr_0 . \tag{4.10}$$

The image current density is, apart from an irrelevant constant factor

$$|S_1(k, r_1)|^2 = \iint S_0(0, r_0) S_0^*(0, r_0') G\left(\frac{r_1}{M} - r_0\right) G^*\left(\frac{r_1}{M} - r_0'\right) \cdot \exp[2\pi i k(r_0 - r_0')] \, dr_0 \, dr_0' . \tag{4.11}$$

For incoherent illumination, all the current densities corresponding to different k vectors occurring in the primary beam are superimposed upon each other incoherently. The image current density becomes

$$j(r_1) = \int |S_1(k, r_1)|^2 F(k) \, dk = \iiint S_0(0, r_0) S_0^*(0, r_0') \cdot G\left(\frac{r_1}{M} - r_0\right) G^*\left(\frac{r_1}{M} - r_0'\right) F(k) \exp[2\pi i k(r_0 - r_0')] \, dr_0 \, dr_0' \, dk . \tag{4.12}$$

$F(k)$ is a distribution function describing the angular distribution of the primary beam from the condenser. It is defined so that $F(k)\,dk = F(k_x, k_y)\,dk_x\,dk_y$ is the probability that an incident electron has a direction such that the x and y components of its wave vector lie within the intervals $\{k_x, k_x + dk_x\}$ and $\{k_y, k_y + dk_y\}$. This distribution function is assumed to be normalized so that

$$\int F(k) \, dk = 1 . \tag{4.13}$$

The integration over k in eq. (4.12) can be performed if we introduce the

Fourier transform of the distribution function $F(k)$:

$$\Phi(r_0) = \int F(k) \exp[2\pi i k r_0] \, dk . \tag{4.14}$$

Then we have

$$j(r_1) = \iint S_0(0, r_0) S_0^*(0, r_0') G\left(\frac{r_1}{M} - r_0\right) G^*\left(\frac{r_1}{M} - r_0'\right) \Phi(r_0 - r_0') \, dr_0 \, dr_0' . \tag{4.15}$$

Let us now assume that we have a weak phase object, *i.e.* that $S_0(0, r_0)$ can be expressed by eq. (2.8)

$$S_0(0, r_0) = 1 + i\eta(r_0) . \tag{4.16}$$

Replacing S_0 from (4.16) in (4.15) and neglecting second-order terms in η we obtain for the image current density

$$j(r_1) = j_B + i \iint \eta(r_0) \Phi(r_0 - r_0') G\left(\frac{r_1}{M} - r_0\right) G^*\left(\frac{r_1}{M} - r_0'\right) dr_0 \, dr_0' - i \iint \eta(r_0') \Phi(r_0 - r_0') G\left(\frac{r_1}{M} - r_0\right) G^*\left(\frac{r_1}{M} - r_0'\right) dr_0 \, dr_0' \tag{4.17}$$

In eq. (4.17), j_B is an abbreviation for the background current density

$$j_B = \iint \Phi(r_0 - r_0') G\left(\frac{r_1}{M} - r_0\right) G^*\left(\frac{r_1}{M} - r_0'\right) dr_0 \, dr_0' . \tag{4.18}$$

If we again define contrast $C(r_1)$ by

$$C(r_1) = \frac{j(r_1) - j_B}{j_B} \tag{4.19}$$

we have

$$C(r_1) = \int \eta(r_0) \Gamma(r_1, r_0) \, dr_0 , \tag{4.20}$$

where the impulsive response Γ is given by

$$\Gamma(r_1, r_0) = \frac{i}{j_B} \int \left[\Phi(r_0 - r_0') G\left(\frac{r_1}{M} - r_0\right) G^*\left(\frac{r_1}{M} - r_0'\right) - \Phi(r_0' - r_0) \cdot G\left(\frac{r_1}{M} - r_0'\right) G^*\left(\frac{r_1}{M} - r_0\right) \right] dr_0' . \tag{4.21}$$

Substituting a new variable of integration

$$t = \frac{r_1}{M} - r_0', \tag{4.22}$$

it can be shown that the impulsive response Γ is a function not of r_1 and r_0 separately but only of the combination $r_1/M - r_0$. In other words: The isoplanacy condition is not destroyed by incoherent illumination. (4.20) can now be written as

$$C(r_1) = \int \eta(r_0)\, \Gamma\left(\frac{r_1}{M} - r_0\right) \mathrm{d}r_0 \,. \tag{4.23}$$

This is again a convolution integral so that we can define a transfer function for the Fourier transform of Γ. If the impulsive response $G(r_1/M - r_0)$ for coherent amplitude transfer is known, Γ can be calculated from eq. (4.21) for any arbitrary angular distribution $F(k)$ of the illuminating beam. The Fourier transform of Γ takes into account not only the electron optical properties of the imaging system behind the object but also the conditions of illumination. C, η and Γ are real functions. It should, however, be noted that, in the case of partial coherence, the linear terms in η may not be large compared to the second order terms.

REFERENCES

General Theory:

O. Scherzer: *Journ. Appl. Phys.*, **20**, 20 (1949).
K.-J. Hanszen, B. Morgenstern, *Zeits. angew. Phys.*, **19**, 215 (1965).
F. Lenz, *Laboratory Investigation*, **14**, 808 (1965); *Optik*, **22**, 270 (1965).

Zone Correction Plates:

W. Hoppe: *Naturwiss.*, **48**, 736 (1961).
W. Hoppe: *Optik*, **20**, 599 (1963).
G. Möllenstedt, R. Speidel, W. Hoppe, R. Langer, K. H. Katerbau and F. Thon: *Proc. 4th Eur. Conf. Electr. Micr. Rome 1968* (Rome, 1968), vol. **1**, p. 125.
F. Lenz, *Optik* **21**, 489 (1964).

Light Optical Theory of Phase Contrast:

F. Zernike: *Phys. Zeits.*, **36**, 848 (1935).

Optical Space Frequency Analysis by Fraunhofer Diffraction of Electron Micrographs:

A. Klug and J. E. Berger: *Journ. Mol. Biol.*, **10**, 565 (1964).
F. Thon: *Zeits. Naturfor.*, **21***a*, 476 (1966).

Image Reconstruction in Light Optical Microscopy:

A. Marechal and P. Croce: *Compt. Rend. Acad. Sci. Paris*, **237**, 607 (1953).

Image Reconstruction in Electron Microscopy:

K.-J. Hanszen: *Proc. 4th Eur. Conf. Electr. Micr., Rome 1968* (Rome, 1968), vol. **1**, p. 153,
P. Schiske: *Proc. 4th Eur. Conf. Electr. Micr., Rome 1968* (Rome, 1968), vol. **1**, p. 145.
D. J. De Rosier and A. Klug: *Nature*, **217**, 130 (1968).
R. Langer and W. Hoppe: *Optik*, **24**, 470 (1966-67); **25**, 413, 507 (1967).