Phil. Trans. Roy. Soc. Lond. B. **261**, 105–118 (1971) [105]
Printed in Great Britain

Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs

By H. P. Erickson and A. Klug, F.R.S.
Medical Research Council Laboratory of Molecular Biology, Cambridge

[Plates 17 and 18]

The effects of defocusing and spherical aberration in the electron microscope image are most simply and directly displayed in the Fourier transform of the image. We have investigated the process of image formation by determining the changes in the transform of the image of a thin crystal of catalase, which has discrete diffraction maxima in the resolution range of 10 to 2.5 nm, as a function of defocusing. The changes in amplitude and phase of these diffraction maxima have been measured and compared with the predictions of a first-order theory of image formation. The theory is generally confirmed, and the transfer function of the microscope is completely determined by finding the relative contributions from phase and amplitude contrast. A 'true' maximum contrast image of the catalase crystal, compensated for the effects of defocusing, is reconstructed from the set of micrographs in the focal series. The relation of this compensated image to individual underfocused micrographs, and the use of underfocus contrast enhancement in conventional electron microscopy, are discussed.

This approach and the experimental methods can be extended to high resolution in order to compensate for spherical aberration as well as defocusing. In as much as spherical aberration is the factor presently limiting the resolution of electron lenses, this could provide a considerable extension of the resolution of the electron microscope.

Introduction

In the analysis of structure from electron micrographs it is important to know how contrast enhancement and artefacts from defocusing and aberrations affect the image in the electron microscope. These effects are displayed much more simply and directly in the Fourier transform of the image than in the image itself, and are best analysed in terms of the transform. The analysis of these effects through the image transform is of particular interest because of the involvement of the transform in systems for the analysis of periodic structures in electron micrographs, and in the system used for three-dimensional reconstruction by electron microscopy (De Rosier & Klug 1968).

Conventional microscopy today is largely concerned with the imaging of details from 10 or 20 nm down to 2 nm resolution. Especially with the microscopy of biological specimens there is generally little preservation of meaningful structural details beyond 2 nm resolution. As modern electron lenses are essentially perfect to this resolution, the only electron optical factor affecting the image is defocusing. This is important in practical microscopy since micrographs are normally taken somewhat under focus, both because it is technically more difficult to obtain a perfectly in-focus image, and because the defocusing produces a useful enhancement of image contrast. We have analysed the effects of defocusing theoretically and experimentally, and in terms of the results of the investigation can specify conditions for the proper use of underfocus contrast enhancement and the nature of artefacts that will occur with excessive defocusing.

At higher resolution the spherical aberration of the electron lens affects the image in a manner very similar to defocusing. Under optimum conditions modern microscopes can record image details at a point to point resolution of about 0.2 nm. As discussed below, however, the

spherical aberration causes significant alterations in the Fourier components of the image corresponding to resolutions greater than 0.45 nm. In this sense it is understood to limit the valid resolution of the image to 0.45 nm. The approach we have used, of processing the image transorm to measure and compensate for the effects of defocusing, may be directly extended to high resolution to compensate for the effects of spherical aberration. The compensated image would be valid and free from artefacts to much higher resolution than the 0.45 nm presently attainable.

The extension of the resolution to atomic and subatomic levels is of considerable interest for the future of electron microscopy, but is somewhat of an academic exercise at present, with specimen preparation techniques generally limiting the meaningful resolution to 2 nm. In addition, experimental work at resolutions of 0.5 nm and better is at present very difficult and unreliable because of a variety of secondary factors, such as instrumental instabilities and specimen damage, affecting the imaging. We have therefore confined our experimental study, which includes a demonstration of the compensation system, to the medium resolution range immediately accessible with available specimens and practical routine microscopy. The extension of the results and the systems to high resolution, when improvements in specimen preparation and instrumental techniques make it experimentally practicable, should be relatively straightforward.

Definitions and assumptions in the theory

Phase contrast is produced by the interference of the scattered electron wave that passes through the objective aperture with the unscattered or background electron wave. Amplitude contrast is produced by the loss of electrons which are scattered outside the objective aperture. In the case where the scattered electron wave is only a small fraction of the unscattered wave, which is more or less generally true in the microscopy of thin specimens, a particularly simple first order or linear approximation can be invoked. The derivation of this first-order theory, and the nature of the errors introduced by second-order effects, will be treated in detail elsewhere (Erickson 1971). The definitions and results of the derivation will be given here.

Under normal conditions of microscopy the thickness of the specimen is much less than the defocusing increment needed to produce changes in the image, and all layers of the specimen can thus be considered to be at the same level of focus. The specimen can then be treated as a two-dimensional object and described by a function, $\sigma(x, y)$, giving the projected density in atoms per unit area at object plane coordinates x, y. The object transform is defined as

$$T^0(\alpha, \phi) = \iint \sigma(x, y) \exp\{-(2\pi i/\lambda)\,[x\alpha \cos\phi + y\alpha \sin\phi]\}\,dx\,dy, \tag{1}$$

where the transform is given in circular coordinates, α/λ and ϕ. α can be identified physically as the angle of scattering in the microscope, λ is the electron wavelength, and α/λ is the reciprocal radial coordinate, which designates the spatial frequency, or less strictly the reciprocal resolution, of object details contributing to the transform. ϕ is the azimuthal coordinate.

In the wave theory of image formation the effects of spherical aberration and defocusing are attributed to a phase shift, $\chi(a)$, which the scattered electron wave undergoes at the diffraction plane of the microscope. This is a function of scattering angle α:

$$\chi(\alpha) = 2\pi/\lambda\,[-\tfrac{1}{4}C_s\alpha^4 + \tfrac{1}{2}\Delta f\alpha^2], \tag{2}$$

where C_s is the coefficient of spherical aberration and Δf is the defocusing (positive for a weak or underfocused lens).

The effect of an objective aperture can be accounted for by defining an aperture function, $A(\alpha)$, which is zero for those scattering angles obstructed by the aperture and unity otherwise. The atomic scattering factor, $f(\alpha)$, for elastic scattering of electrons by a single atom, also enters the derivation.

Transform of the phase-contrast image

Phase contrast is the predominant factor in the medium and high resolution image. The transform of the phase-contrast image, T^{i}_{ph}, is related to the transform of the object (equation (1)) by the expression:

$$T^{i}_{ph}(\alpha, \phi) = -T^{0}(\alpha, \phi)\, A(\alpha) f(\alpha) \sin \chi(\alpha). \tag{3}$$

A zero-order term, equivalent to the transform of the average background intensity over the area of the image, has been left out of the right-hand side of equation (3).

Equation (3) shows that the relation of the object to the image formed in the presence of aberrations is very direct and simple when expressed in terms of the Fourier transforms. The image transform is directly proportional to the object transform, modulated by the three factors, the most important of which is $\sin \chi(\alpha)$. The result is especially simple because each of these factors is a real number and therefore modulates only the amplitudes of the complex transform, leaving the phases unchanged. Object details of resolution corresponding to a particular spatial frequency, α/λ, are thus imaged with a contrast proportional to $\sin \chi(\alpha)$.

The physical significance is best discussed with reference to the graphs of $-\sin \chi(\alpha)$ in figure 1. At $\Delta f = 0$, (figure 1*a*), $\chi(\alpha)$ is determined solely by spherical aberration, and $\sin \chi(\alpha)$ is essentially zero except for higher resolutions, $\alpha/\lambda > 0.7$ nm^{-1}. Lower resolution components of the object transform do not contribute to the in-focus image through phase contrast. The optimum defocusing for high resolution phase-contrast microscopy is around 90 nm under focus (figure 1*b*), because $-\sin \chi(\alpha)$ is then close to -1.0 over a large range of resolutions, for values of α/λ between 1 and 2.3 nm^{-1}. The negative sign which appears in equation (3), and which is included in the graphs in figure 1, indicates a reduction in electron intensity over areas of high mass density, the condition of normal contrast. The image transform is thus identical to the object transform over this region (ignoring now the factors $A(\alpha)$ and $f(\alpha)$) and we can say that the object transform is contributing to the image with maximum, or 'true', phase contrast. If $\sin \chi(\alpha) = 1$ for all α the microscope would be equivalent to a perfect phase contrast microscope. Actually for values of α/λ less than 1 nm^{-1} and greater than 2.3 nm^{-1}, $\sin \chi(\alpha)$ is less than unity. Corresponding components of the object transform will contribute to the image with reduced contrast, but otherwise without distortion or artefact.

Artefacts are introduced at higher resolution when $\sin \chi(\alpha)$ goes through zero and changes sign. Corresponding parts of the object transform will contribute to the image, but because of the sign change the contrast will be reversed, i.e. image details of corresponding reciprocal spacing will be black where they should be white. It is in this sense that the optical resolution of the microscope is limited by spherical aberration to about 0.45 nm ($\alpha/\lambda = 2.3$ nm^{-1}) since at higher resolution details are being imaged with alternating reversed and normal contrast.

At larger Δf the peak of normal phase contrast becomes sharper and moves to lower values of α/λ. At 500 nm under focus (figure 1*c*) details between 5 and 2 nm resolution, $\alpha/\lambda = 0.2$ to 0.5 nm^{-1}, will be imaged with normal phase contrast. Artefacts will occur for $\alpha/\lambda > 0.7$ nm^{-1} so the resolution of this image is limited to 1.5 nm. This will be a particularly useful defocusing for biological microscopy when one wants to obtain a high contrast image of details in this medium resolution range, and is not concerned about artefacts at higher resolution.

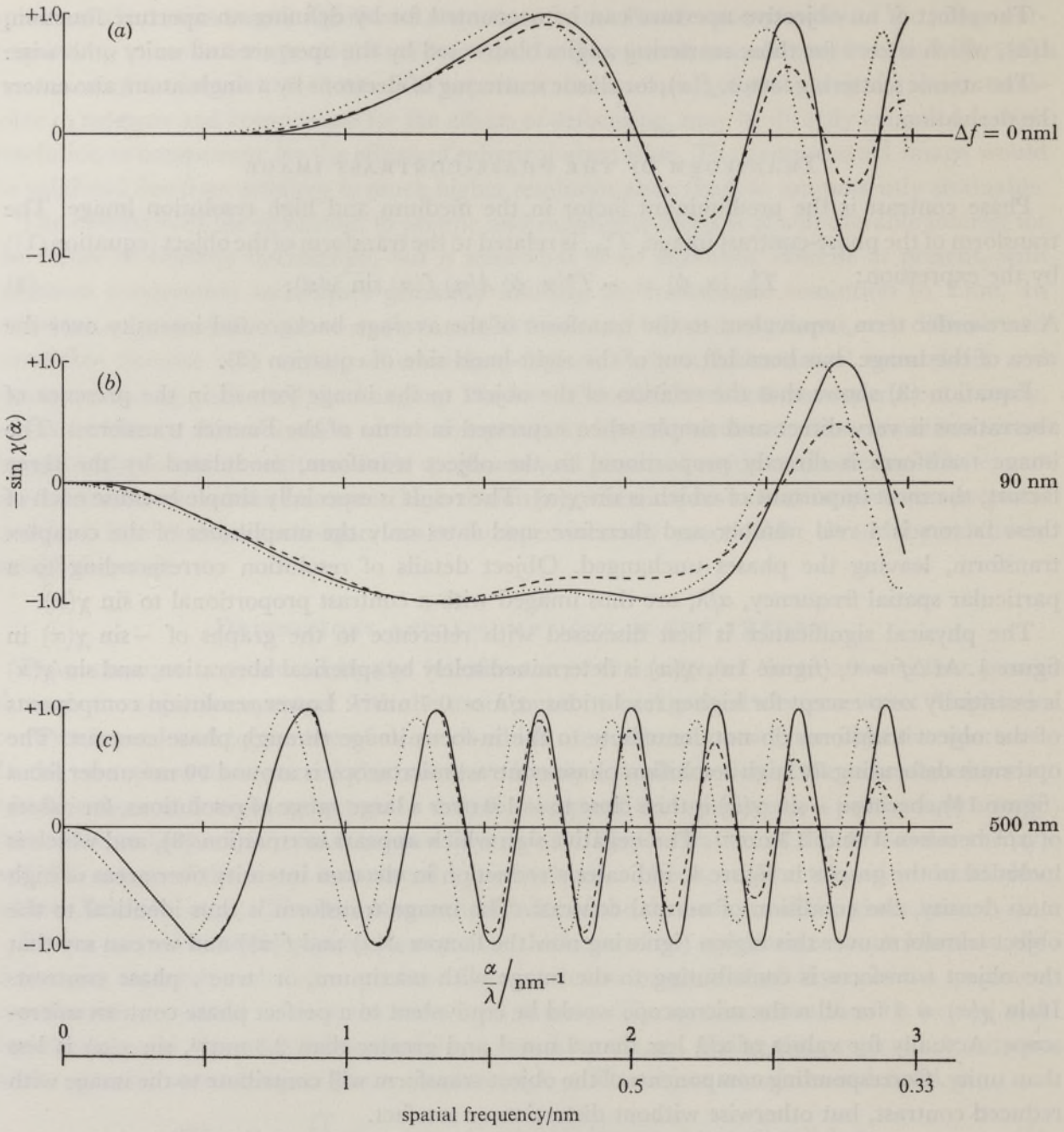

FIGURE 1. The phase-contrast transfer function, $-\sin\chi(\alpha)$, plotted as a function of α/λ, in nm^{-1}, for $\lambda = 0.0042$ nm, $C_s = 1.3$ mm, and for the indicated values of Δf. A negative value of this function implies that the corresponding region of the object transform is contributing to the image with normal contrast, i.e. a subtraction from the background electron intensity over regions of high mass density. The solid curves are for pure phase contrast. The dashed curves are corrected for the effects of chromatic aberration with normal electrica instabilities, by averaging over a range of Δf of ± 20 nm. The dotted curves are corrected for the effects of the partial coherence of the electron source, assuming a 100 μm diameter condenser aperture. These corrections are discussed by Erickson (1971).

THE TRANSFORM OF THE AMPLITUDE CONTRAST AND OF THE TOTAL IMAGE

The transform of the image produced by amplitude contrast is given by an expression similar to equation (3)

$$T^{i}_{amp}(\alpha, \phi) = -T^{0}(\alpha, \phi)\, A(\alpha) f(\alpha)\, Q(\alpha) \cos\chi(\alpha). \qquad (4)$$

The transfer function here is proportional to cos $\chi(\alpha)$ because the amplitude contrast is defined in the theoretical derivation as that due to the real part of the scattered wave. Phase contrast, for which the transfer function is proportional to sin $\chi(\alpha)$, is identified with the imaginary part. The factor $Q(\alpha)$ gives the magnitude of the real part of the scattered wave relative to that from the imaginary part, or, in other words, the ratio of amplitude contrast to phase contrast.

Amplitude contrast is traditionally attributed to the loss of electrons scattered outside the objective aperture. In the simplest approximation $Q(\alpha)$ is the ratio of the integrated intensity of the electrons scattered outside the objective aperture of the amplitude of the electron wave inside the aperture. Numerical calculations (Erickson 1971) indicate that it should vary from less than 0.1 for carbon to about 0.4 for a heavy element.

The atomic scattering factor $f(\alpha)$ which is used in this derivation, and which appears in equations (3) and (4), is the real scattering factor derived from the Born approximation. An exact derivation of electron scattering gives a complex $f(\alpha)$, in which the imaginary part has about the same relative magnitude, 0.1 to 0.4, as the $Q(\alpha)$ given above (Reimer 1969). The complex $f(\alpha)$ must be used in an exact derivation of image formation, but in the first-order theory used here the Born approximation is appropriate. The amplitude contrast term of equation (4) is equivalent to the contribution from the imaginary part of $f(\alpha)$ that would be obtained from an exact derivation. A more rigorous description of amplitude contrast is given by the second-order theory of image formation (Erickson 1971). In view of these theoretical complexities, and the approximations involved in deriving the numerical estimates, we should emphasize that in the present work we have been able to measure the factor $Q(\alpha)$ experimentally. The empirical values that are obtained, and which are used in the discussion, are consistent with the above estimates but do not rely on a particular theoretical derivation.

In general amplitude and phase contrast will both contribute, independently, to the image, and the total image transform can be written

$$T^{i}_{tot}(\alpha, \phi) = -T^{0}(\alpha, \phi) f(\alpha) A(\alpha) [\sin \chi(\alpha) + Q(\alpha) \cos \chi(\alpha)]. \quad (5)$$

The transfer function, $[\sin \chi(\alpha) + \cos \chi(\alpha)]$, can be written alternatively as $\sin [\chi(\alpha) + \Phi(\alpha)]$. This is a useful formulation for showing that the effect of a small fraction of amplitude contrast can be treated as an additional phase shift, $\Phi(\alpha)$, in the phase contrast transfer function. The zeros and peaks of the transfer function are accordingly shifted to the left (cf. dotted curve in figure 4). One should note particularly that the transfer function does not then go to zero at low values of α/λ. Thus low resolution Fourier components, which would be weak or missing in the pure phase image, are not lost when amplitude contrast is included (figures 4 and 5).

Compensation of spherical aberration and defocusing

Obviously the conventional microscope, when operated with a small degree of underfocusing as in figure 1*b*, is almost a perfect phase-contrast microscope out to a resolution of 0.45 nm. It is actually better than the curve for pure phase contrast indicates because the reduced phase contrast at the lower resolutions is largely compensated by amplitude contrast, as will be shown below in the discussion of the experimental measurements. In the range of validity of equation (3) or equation (5) it should be possible to extend the resolution well beyond the 0.45 nm limit set by spherical aberration by correcting the transform for the contrast reduction and contrast reversal of the higher resolution components. For this one would

obtain the transform of the image and determine the true object transform from it, essentially by dividing the image transform by sin $\chi(\alpha)$. The compensated image is then reconstructed directly by inverse Fourier transformation. The value of sin $\chi(\alpha)$ can be readily calculated from the electron optical parameters, or determined directly from the micrograph (Thon 1966). Of course where sin $\chi(\alpha)$ and the amplitude of the image transform go to zero this process becomes indeterminate and corresponding components of the object transform cannot be determined from this single micrograph. These may be only a small part of the whole transform and may be negligible for the reconstruction of the compensated image. If not they would have to be determined from other micrographs at slightly different values of defocusing where the zeros of sin $\chi(\alpha)$ occur at different values of α. A system like this has been proposed and outlined by Schiske (1968).

Experimental methods

Thin crystals of catalase, negatively stained with uranyl acetate, were used as a test specimen for experimental analysis because their transform is particularly suitable. These crystals are essentially two-dimensional (the ones actually used probably consist of two layers, about 20 nm thick) periodic structures, and the transform thus consists of a set of discrete, well-defined spots on the corresponding reciprocal lattice. The amplitude and phase of these discrete spots are readily determined by computer processing, giving a quantitative measure of the variations in the image transform with defocusing at the spatial frequency corresponding to each spot.

Micrographs were taken with a Philips EM-300, operating at 80 kV, using a 100 μm second condenser aperture, and no objective aperture. A focal series of twelve micrographs, ranging from in focus to 2000 nm underfocus, was obtained (figure 2). The amount of defocusing was determined for each micrograph by measuring the radius of the rings of maximum and minimum noise intensity in the optical transform, and comparing these with the peaks and zeros of sin $\chi(\alpha)$ calculated for small increments of Δf. The modulation of the diffraction pattern of the noise image by the transfer function, and its relation to defocusing, has been previously demonstrated by Thon (1966). The small fraction of amplitude contrast in the high resolution range where the rings were measured, $\alpha/\lambda > 0.4\ \text{nm}^{-1}$, was negligible for this determination. The theoretical curves used for this comparison were corrected for the partial coherence from the 100 μm condenser aperture (Erickson 1971) as this has a significant effect on the determination of Δf. The values of Δf obtained in this way are accurate to about 20 nm.

Optical diffraction–qualitative analysis

Figure 2, plate 17, shows the image used for the analysis, and the corresponding optical diffraction pattern, at three different defocusings. Figure 2*b* shows the diffraction pattern most clearly, and this is diagrammed and the indexing indicated in figure 3. The spatial frequency of the spots varies from $1/8.9\ \text{nm}^{-1}$ for the (0, 2) spot, to $1/2.4\ \text{nm}^{-1}$ for the (1, 7) spot. The expected change in the amplitude of the spots is shown qualitatively in these optical diffraction patterns. At $\Delta f = 0$, figure 2*a*, phase contrast is zero for this resolution range (cf. figure 1*a*) and the image is due entirely to amplitude contrast. The transform is generally weak and many of the higher resolution components are missing. At 540 nm underfocus (figure 2*b*), where we

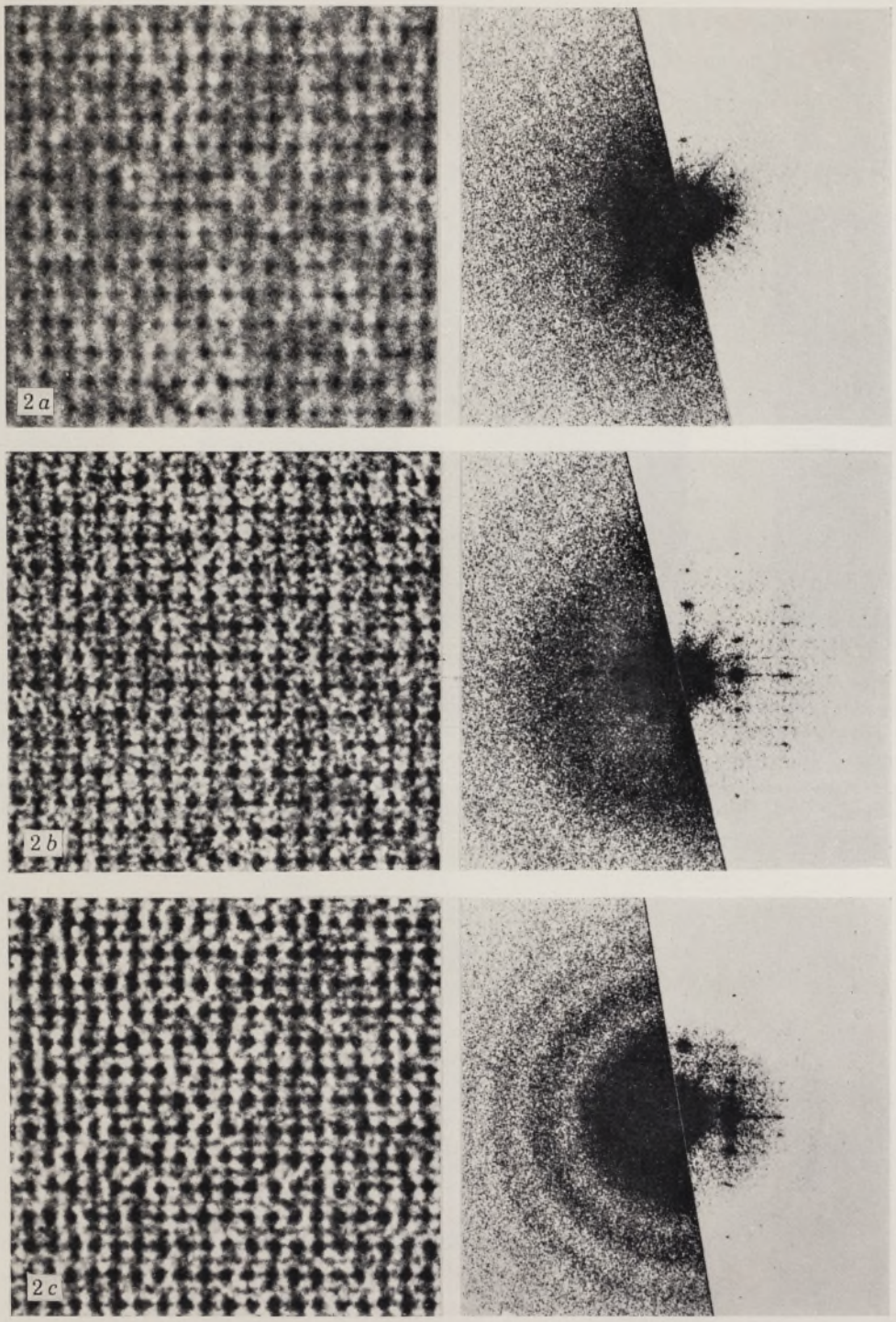

FIGURE 2. The image of a negatively stained catalase crystal and the corresponding optical diffraction pattern at different values of defocusing: (*a*) in focus, (*b*) 540 nm under focus, (*c*) 1450 nm under focus. The apparent lattice spacings in the micrographs are 6.55 nm in the horizontal and 8.9 nm in the vertical direction. The reciprocal lattice spacings in the diffraction pattern are $1/6.55\ \mathrm{nm}^{-1}$ and $1/17.8\ \mathrm{nm}^{-1}$ (corresponding to the true vertical repeat of 2×8.9 nm). The left half of the optical diffraction pattern has been overexposed to show the low intensity spots and the effect of the transfer function on the noise pattern. The highest resolution diffraction spot, indices (1, 7), corresponds to a reciprocal spacing of $1/2.4\ \mathrm{nm}^{-1}$. The indexing is indicated in figure 3.

(*Facing p.* 110)

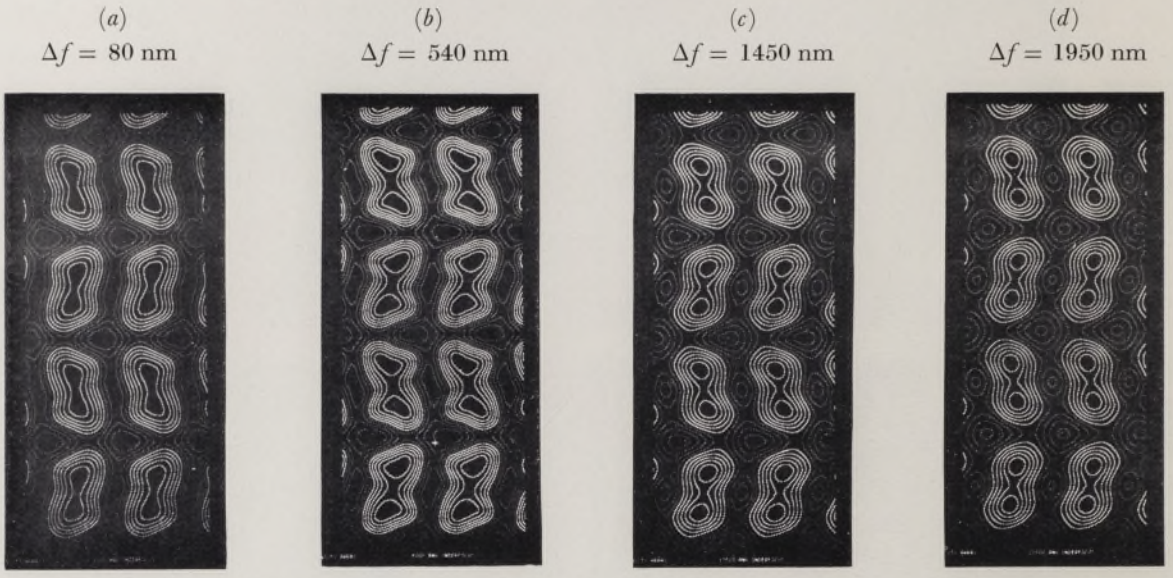

FIGURE 6. Noise filtered, averaged images reconstructed from the discrete catalase transforms. (*a*) is reconstructed from the compensated 'true' object transform in figure 3. As discussed in the text this is essentially the same as that reconstructed from a single image 80 nm under focus. (*b*), (*c*) and (*d*) are reconstructed from the transforms of individual underfocused micrographs. The images are presented here as contour plots with the same number of contours between maximum and minimum density. The absolute level of contrast is thus the same in each of these images. (*b*) is the best underfocus for 2 nm resolution. (*c*) and (*d*) are excessively underfocused.

expect phase contrast to be enhanced over the resolution range $\alpha/\lambda = 0.2$ to $0.5\ \mathrm{nm}^{-1}$ (cf. figure 1*c*, especially the dotted curve corrected for partial coherence), the intensity of the discrete spots, and of the image noise, is increased considerably. Many of the higher resolution components, which were weak or missing in the in focus transform are quite strong here. The effect of excessive defocusing is displayed in figure 2*c*, where the first zero of the transfer function, seen as the ring of low intensity noise in the overexposed left half, passes through the region of the catalase transform. The (0, 6) spot, which was reasonably intense in figure 2*b*, is at a spatial frequency, α/λ, near this zero and, as we expect, it is missing in this image transform. In addition, the (1, 7) spot now falls in the next ring of maximum noise intensity, or under the second peak of the transfer function. It should therefore be contributing to the image with reversed contrast although the sign change is not indicated in the optical diffraction pattern. Since at this defocusing the higher resolution components of the catalase transform are missing or are contributing to the image with reversed contrast we can say that this image is excessively underfocused.

Computer-generated transforms—quantitative comparison with theory

To investigate these effects more quantitatively we have used the programs available in this laboratory, developed largely by DeRosier & Moore (1970), for obtaining the complex Fourier transform of an electron micrograph by digital computer. This gives numerically the amplitude and also the phase angle, of each of the diffraction spots and makes possible a quantitative measure of the changes seen qualitatively in the optical diffraction patterns.

Considering first the phase angle, which is not recorded in the optical diffraction pattern, we recall from equation (3) or equation (5) that the phase of the image transform should not

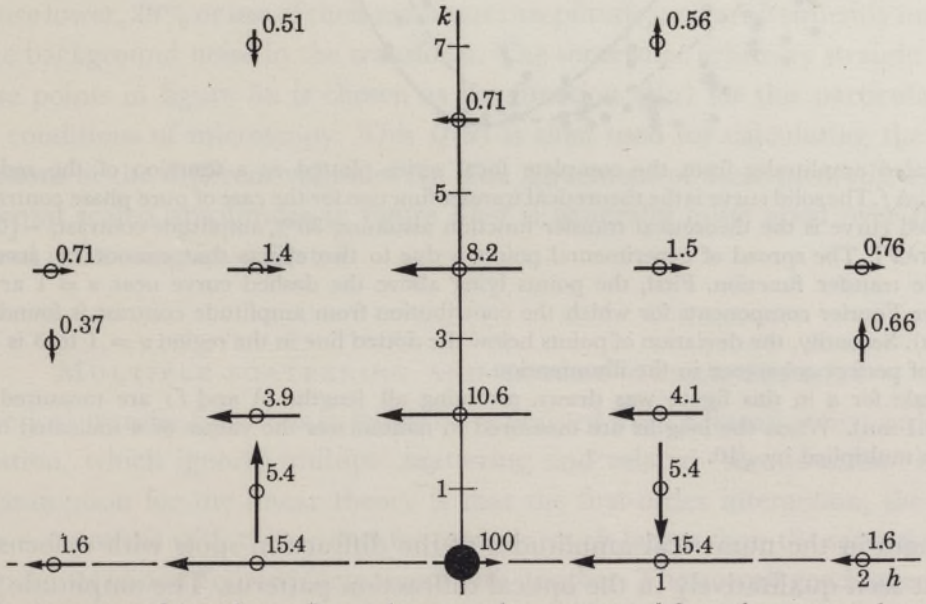

Figure 3. The compensated transform of a catalase crystal reconstructed from the computed transforms of the set of micrographs in the focal series. The direction of the arrows indicates the phase angle of the diffraction spot and the numbers give the maximum amplitude for that spot, relative to the zero-order component of amplitude 100. The phase angles depend on the origin used in calculating the transform. A point at the bottom left-hand corner of the images in figure 6 was used here, which lies at the intersection of the glide axis and mirror line in the plane group, *pgm*, of the image. The fact that the phase angles are all multiples of 90° confirms the plane group symmetry.

be affected by the defocusing. This prediction was confirmed in the numerical transforms as the phases of the diffraction spots, shown in figure 3, varied only 10 or 15° with different degrees of moderate underfocusing. With excessive defocusing the phase of the higher resolution components, falling under the second peak of the transfer function (e.g. the (1, 7) spot in figure 2*c*), did change phase by exactly 180°. This is equivalent to the change of sign of the transfer function from minus to plus, and confirms that these components are contributing to these images with reversed contrast.

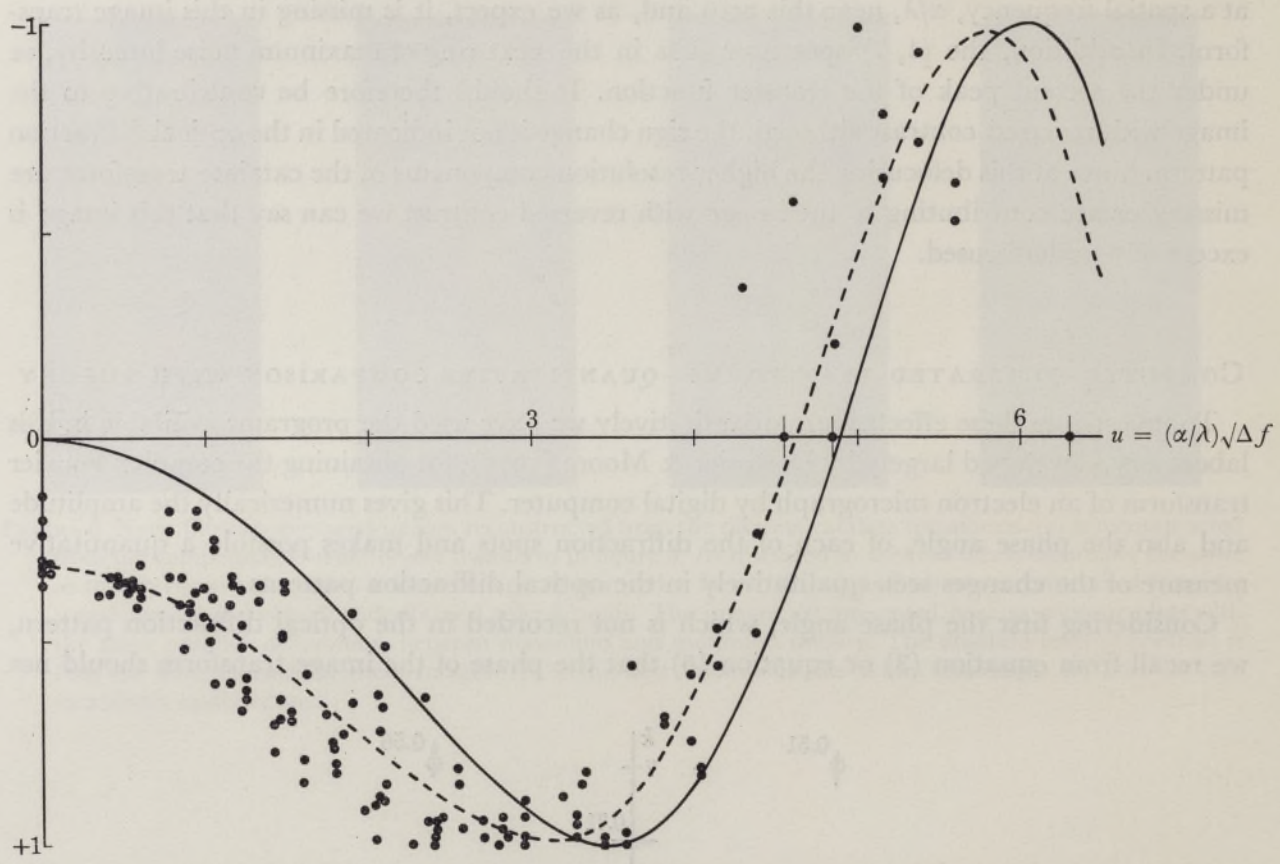

FIGURE 4. Scaled amplitudes from the complete focal series plotted as a function of the reduced variable $u = (\alpha/\lambda)\sqrt{\Delta f}$. The solid curve is the theoretical transfer function for the case of pure phase contrast, $-\sin \pi\lambda u^2$. The dashed curve is the theoretical transfer function assuming 35% amplitude contrast, $-[0.93 \sin \pi\lambda u^2 + 0.35 \cos \pi\lambda u^2]$. The spread of experimental points is due to two effects that cannot be accounted for by the simple transfer function. First, the points lying above the dashed curve near $u = 1$ are mostly from high-order Fourier components for which the contribution from amplitude contrast is found to be reduced (figure 5*a*). Secondly, the deviation of points below the dotted line in the region $u = 1$ to 3 is due largely to the lack of perfect coherence in the illumination.

The scale for u in this figure was drawn assuming all lengths (λ and f) are measured in ångströms (1 Å = 0.1 nm). When the lengths are measured in nanometres the values of u indicated on the abscissa should be multiplied by $\sqrt{10}$.

The changes in the numerical amplitudes of the diffraction spots with defocusing was the same as that seen qualitatively in the optical diffraction patterns. The amplitude of each spot increased with underfocusing, reached a maximum at a particular value of Δf, and fell toward zero with greater defocusing. Higher resolution components reached the maximum for moderate amounts of underfocusing, lower resolution components only for much greater defocusing. By examining the transforms from the complete focal series the maximum amplitudes of each of the spots was determined and the maximum contrast transform shown in figure 3 was reconstructed.

This is essentially the true object transform, compensated for the effects of defocusing, and will be discussed below in relation to the reconstruction of the compensated image.

For further analysis and comparison of the experimental results with the theoretical expression, equation (5), it is convenient to scale the amplitudes of the spots in each image transform by dividing by the maximum amplitude. This is tantamount to dividing each image transform by the object transform, which we have determined above (figure 3). The scaled amplitudes will vary between plus and minus one, and should be equal to the value of the transfer function, $[P(\alpha) \sin \chi(\alpha) + Q(\alpha) \cos \chi(\alpha)]$. ($P(\alpha)$ is a normalization factor, related to $Q(\alpha)$, chosen to make the maximum value of the transfer function unity.)

For the initial comparison with the theory the scaled amplitudes from the complete focal series, for all combinations of Δf and α/λ, are plotted in figure 4 as a function of the variable $u = (\alpha/\lambda)\Delta f$. For pure phase contrast all the points should follow the common transfer function, $-\sin \pi\lambda u^2$. Assuming 35 % amplitude contrast, constant for all values of α, the transfer function should be $-[0.93 \sin \pi\lambda u^2 + 0.35 \cos \pi\lambda u^2]$. Most of the experimental points lie between these two curves in figure 4, and the points from lower resolution Fourier components generally lie closer to the curve with amplitude contrast. There are some consistent deviations due to the effects of partial coherence. The variation of amplitude contrast with α and the effects of partial coherence cannot be included in a single transfer function plotted as a function of the reduced variable u. For this it is necessary to plot the scaled amplitudes from each individual image as a function of α/λ and to compare these with individual transfer functions calculated for the corresponding Δf. This is presented in figure 5.

To calculate the theoretical transfer functions it is first necessary to know $Q(\alpha)$. This is given directly in the plot for $\Delta f = 0$ (figure 5*a*), since $\sin \chi(\alpha) = 0$ and $\cos \chi(\alpha) = 1$ over this entire resolution range. In this case, where the image is due completely to amplitude contrast, the lower resolution spots are seen to be about 40 % of their maximum amplitude. Higher resolution components are lower, 20% or less of their maximum amplitude, and are frequently indistinguishable from the background noise in the transform. The somewhat arbitrary straight line drawn through these points in figure 5*a* is chosen as the function $Q(\alpha)$ for this particular specimen under these conditions of microscopy. This $Q(\alpha)$ is then used for calculating the theoretical transfer functions at the different values of Δf. The agreement of these theoretical curves with the experimental scaled amplitudes in figure 5*b–f* is generally quite good over the range of resolutions and defocusing investigated.

Multiple scattering and second-order effects

The theoretical transfer functions in figure 5 are derived from a linear or first-order theory of image formation, which ignores multiple scattering and related 'second-order' effects. The important assumption for the linear theory is that the first-order interaction, the product of each Fourier component with the zero-order term, is much larger than the second-order interactions, the product of one Fourier component with another. The overall good agreement of the experimental data with the first-order theoretical curves indicates that the second-order effects are mostly unimportant. A notable exception is the consistent deviation of the (1, 4) spot from the theoretical curve. This deviation can be attributed to a significant second-order interaction of the (1, 0) spot with the (0, 4) spot. This second-order effect will appear in the transform at the (1, 4) position, the sum of the indices of the two interacting components. It will be

significant because the two interacting components are quite strong, while the observed amplitude at the (1, 4) position is weak. The product of the amplitudes of the two components, $15.4 \times 8.2 = 126$, is in fact comparable to the product of the observed (1, 4) term with the zero

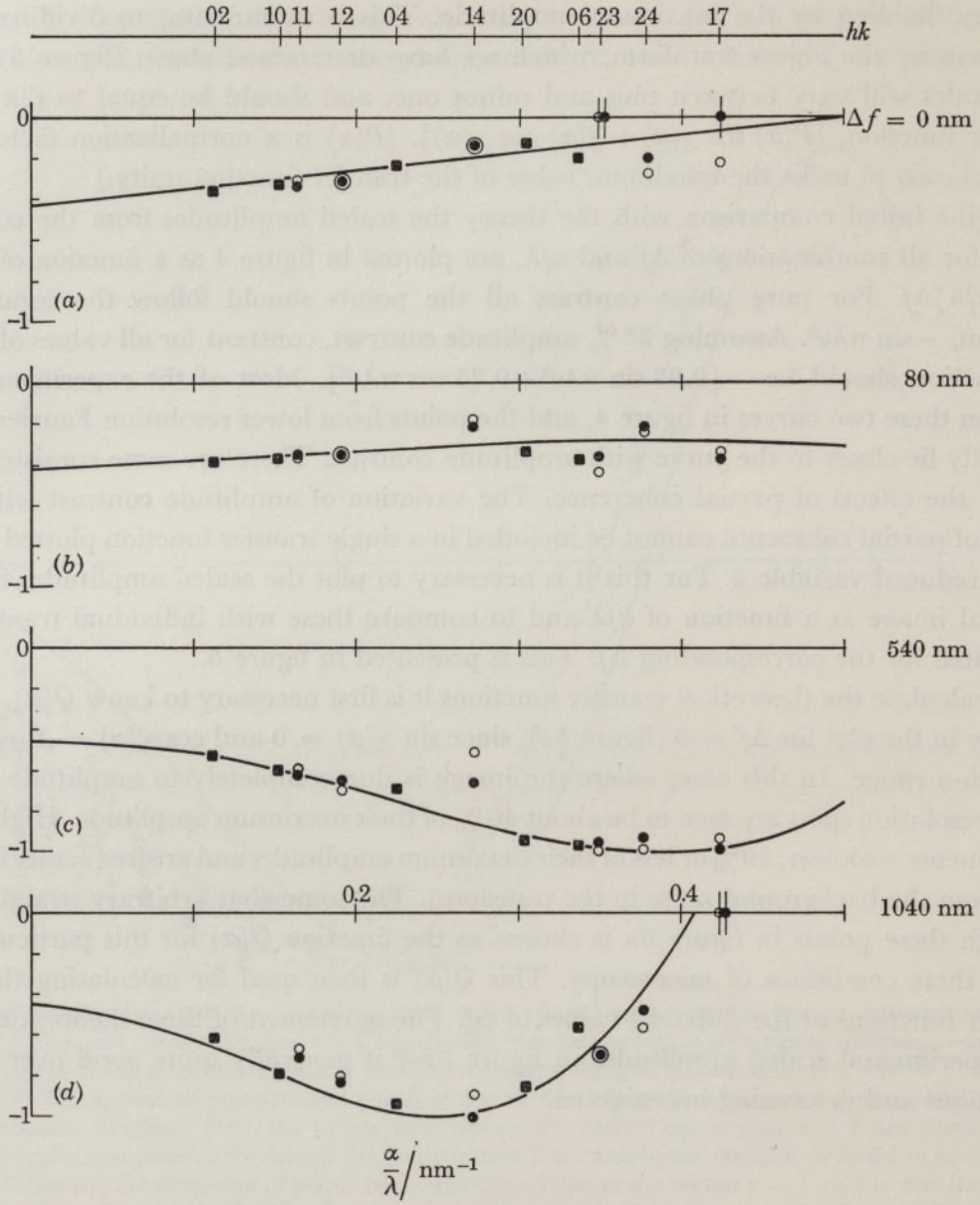

FIGURE 5. Scaled amplitudes of diffraction spots from individual micrographs at the indicated degree of underfocus. The indices (h, k) are given across the top of the diagram. ●, h positive; ○, h negative; ■, $(0, k)$ and $(h, 0)$ spots. For $\Delta f = 0$ nm the straight line drawn through the points is chosen as the experimentally determined estimate of the function $Q(\alpha)$. The curves for the other cases are the theoretical transfer function $-[P(\alpha) \sin \chi(\alpha) + Q(\alpha) \cos \chi(\alpha)]$. The curves are also corrected for the partial coherence of the electron source as mentioned in figure 1.

order term, $1.5 \times 100 = 150$ (see figure 3). The behaviour of this term as a function of defocusing, and the deviation from the linear theory, can be explained by more detailed consideration of the second-order effects (Erickson 1971). The important point for the present discussion is that the experimental results are almost completely accounted for by the linear theory of image formation. The occasional small deviations occur where second-order effects are expected to be

important, and are accounted for when second-order terms are included in the theoretical derivation.

Second-order or multiple-scattering effects have been noted previously in a study of negatively stained cell walls (Finch, Klug & Nermut 1967). Here the second-order interactions were between two superposed lattices which were rotated relative to each other. These second-order interactions were consequently separated spatially from the first-order components in the image transform, and could be found and identified in a single optical diffraction pattern. Here too the second-order effects represented only a small correction to the image predicted on the first-order theory.

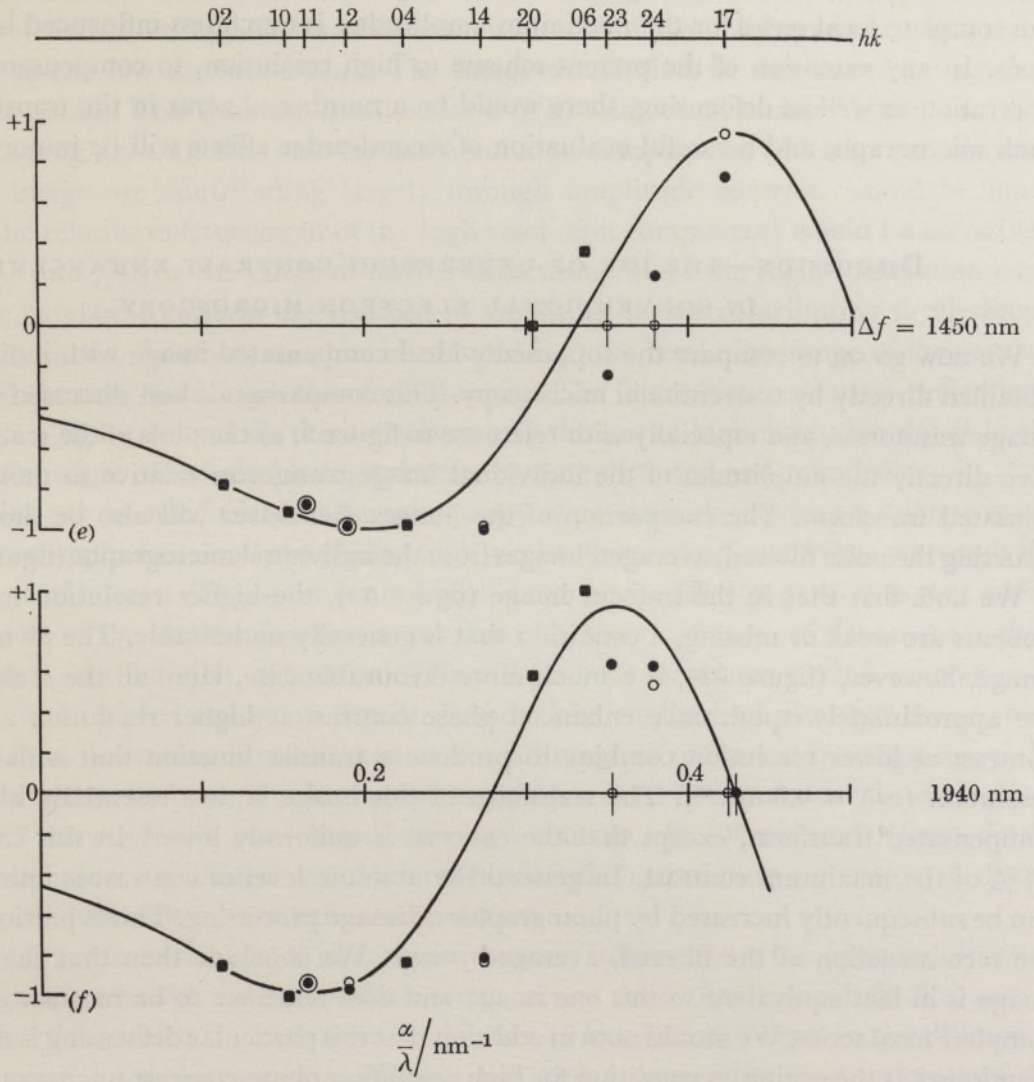

FIGURE 5*e* and *f*. For legend see previous page.

RECONSTRUCTION OF THE COMPENSATED IMAGE OF CATALASE

The transform shown in figure 3, in which each diffraction spot has its maximum amplitude and correct phase, represents the 'true' object transform of the catalase crystal. The 'true' image, compensated for the effects of defocusing, can be reconstructed directly from it by inverse Fourier transformation. Only the discrete diffraction spots are included in this transformation so the non-periodic noise in the image is eliminated. The reconstructed image is thus noise filtered and is also averaged over all the repeating units in the image. The

compensated reconstructed image is shown in figure 6*a*. The morphological units of the crystal seen here are consistent with the structure of four subunits with 222 symmetry proposed from the analysis of helical tubes of catalase (Kiselev, DeRosier & Klug 1968).

The compensated image could be reconstructed from a single micrograph once we know the factor $Q(\alpha)$, assuming none of the discrete Fourier components falls exactly on a zero of the transfer function. The second-order deviations from the linear theory noted above could, however, present complications or lead to errors, especially in compensating an excessively defocused image, for which the first-order transfer function has several zeros where second-order effects dominate. The compensated transform (figure 3), which we have derived by scanning the complete focal series for the maximum amplitudes, is much less influenced by these deviations. In any extension of the present scheme to high resolution, to compensate for spherical aberration as well as defocusing, there would be a number of zeros in the transfer function of each micrograph, and a careful evaluation of second-order effects will be important.

Discussion—the use of underfocus contrast enhancement in conventional electron microscopy

We now go on to compare the supposedly ideal compensated image with individual images obtained directly by conventional microscopy. This comparison is best discussed in terms of the image transforms, and especially with reference to figure 5, as the plots of the scaled amplitudes give directly the amplitudes of the individual image transforms relative to those of the compensated transform. The comparison of the images themselves will also be shown by reconstructing the noise filtered, averaged images from the individual micrographs (figure 6, plate 18).

We note first that in the in-focus image (figure 5*a*), the higher resolution transform components are weak or missing, a condition that is generally undesirable. The 80 nm underfocus image, however, (figure 5*b*), is a much more favourable case. Here all the scaled amplitudes are approximately equal, since enhanced phase contrast at higher resolution and amplitude contrast at lower resolution combine to produce a transfer function that is flat out to 2 nm resolution ($\alpha/\lambda = 0.5\ \mathrm{nm}^{-1}$). The transform of this image is thus essentially identical to the compensated transform, except that the contrast is uniformly lower. In this case it is about 35 % of the maximum contrast. In general the absolute level of contrast is unimportant as it can be subsequently increased by photographic or image processing. This is particularly true for the reconstruction of the filtered, averaged image. We conclude then that the compensated image is in fact equivalent to this one image and does not have to be reconstructed from the complete focal series. We should note in addition that this particular defocusing is that previously mentioned as the optimum condition for high resolution phase-contrast microscopy (figure 1*b*). Details of higher resolution than 2 nm will therefore be imaged with high contrast, and there will be no image artefacts from contrast reversal out to a resolution of 0.5 nm.

For normal microscopy of biological specimens, where one is not generally concerned with resolutions beyond 2 nm, further underfocusing, to the condition given in figure 5*c*, will frequently be advantageous. Here there is a two or threefold increase in contrast over the entire catalase transform, and a relative enhancement of the higher resolution components with respect to the lower. All parts of the object transform with α/λ from 0.3 to 0.45 nm^{-1} are contributing to the image with more than 90 % of maximum contrast, and the lower resolution components with 50 to 80 % of maximum. This relative enhancement of the higher resolution

details, combined with the overall high contrast, gives a sharp and pleasing image. The filtered image (figure 6*b*), reconstructed from this single underfocused micrograph, shows these effects clearly. The general shape of the morphological units is essentially identical to the 'true' image of figure 6*a* but the edges and the corners are sharpened. The true image is that for which all the scaled amplitudes are equal. Slight enhancement of the high resolution components, by underfocusing or by subsequent image processing, will not lead to serious artefacts, but excessive enhancement or, what is equivalent, a loss of low resolution components can be serious. In figure 5*c* the scaled amplitudes of the high resolution components are about twice as big as the low resolution ones, but the absolute amplitudes are still much lower. If the high resolution components were enhanced by a factor of five or ten instead of two, they would be comparable in magnitude to the low resolution ones. The image would then be dominated by spurious high resolution details. This situation could occur if $Q(\alpha)$ were much smaller than the value of about 0.3 obtaining here. In that case the low resolution components, which in a moderately underfocused image are contributing largely through amplitude contrast, would be much weaker, and the relative enhancement of the high resolution components would be excessive.

Figures 5*d*, *e* and *f* show the effect of further defocusing. Here the higher resolution components of the catalase transform are reduced in amplitude or are contributing to the image with reversed contrast. Artefacts are thus introduced into the resolution range of the catalase transform, and the underfocusing is therefore excessive. The filtered images reconstructed from these transforms (figure 6*c* and *d*), show artefacts and the loss of high resolution detail in the image as the flat edges of the morphological units become distorted and rounded.

These conclusions are not limited to the microscopy of periodic objects with a discrete transform, but can be applied generally to objects with a continuous transform. What matters is the resolution range of interest. The optimum defocusing can be specified in general as that which puts the first zero of the transfer function at a value of α/λ somewhat greater than the highest spatial frequency, or resolution, of interest. This will produce maximum overall contrast with enhancement of the higher resolution details. It should be emphasized also that the contrast enhancement from defocusing is not limited to a small range of reciprocal spacings, especially when the contribution from amplitude contrast is considered. At 500 nm under focus all object details out to 2 nm resolution are imaged with high contrast. At 100 nm under focus details out to 2 nm resolution are imaged with uniform but less than maximum contrast, and details out to 0.5 nm resolution now have enhanced contrast.

As a practical guide to the use of underfocus contrast enhancement the optical diffraction pattern will be most useful. Image artefacts occur only at resolutions beyond the first zero of the transfer function, which is clearly seen as the ring of low intensity noise in the diffraction patterns of underfocused micrographs. Without this tool a reasonably good estimate of the proper underfocusing can be obtained by judging the average size of the granularity in the image of the supporting or embedding film. Image details of about this same size and larger are being imaged with good contrast, and artefacts will occur only for image details smaller than this average background granularity. In general it can be recommended that, so long as excessive underfocusing is avoided, the best image is that which looks best in terms of sharpness of detail and contrast.

An important exception to the above conclusions and recommendations for using underfocus contrast enhancement is implied by expected dependence of $Q(\alpha)$ on the atomic number of the specimen. In many cases of practical microscopy, especially of biological specimens, one is

primarily interested in distinguishing the image of a small thickness of a stain, of high atomic number, from the superimposed image of the random granular structure or noise of the comparatively thick supporting or embedding film, of low atomic number. $Q(\alpha)$ should be significantly higher for the heavy elements than for the light ones so the image of the stain will have a larger contribution from amplitude contrast than will the noise from the supporting film. An amplitude contrast image, obtained by taking an in focus micrograph, would then give a better signal (stain):noise ratio than the underfocused image where the phase contrast is enhanced. This can, in fact, be understood as the basis for the concept that the in-focus, or 'para-focus' (Van Dorsten, Mellema & Premsela 1968) amplitude contrast image represents the 'true' image while the underfocus image with phase-contrast enhancement is 'noise' and artefact. In fact both the in-focus and the moderately underfocused images are valid or 'true', with artefacts only at resolutions beyond the first zero of the transfer function. The in-focus micrograph is of low overall contrast, but the contrast of the stain is enhanced relative to that of the low molecular mass substrate. The obvious extension of this idea would be to use the information in two or more micrographs at different defocusing to separate completely the images of heavy and light elements. This would involve processing and combining the image transforms in terms of the first-order theory discussed here. Quantitative experimental investigation of the scheme has not yet been attempted.

For the system of three-dimensional reconstruction (De Rosier & Klug 1968), the transforms of images viewing the object from different directions are combined to reconstruct the three-dimensional transform of the object. If all the different images are obtained with the same defocusing they will have the same transfer function. The three-dimensional transform reconstructed from these would then be modulated by this transfer function, where the reciprocal radial coordinate α/λ is interpreted to be the three-dimensional radial coordinate. The contrast properties and high resolution enhancement of the three-dimensional image reconstructed from this will be the same as those of the individual images.

Finally it is comforting to be able to confirm quantitatively that the moderately underfocused micrographs used in most biological microscopy are valid images, frequently the best possible in terms of resolution and contrast, with no artefacts in the low and medium resolution range of interest.

H. Erickson acknowledges support from the U.S. National Cancer Institute, post-doctoral fellowship number 5 F02 CA 23445—02.

References (Erickson & Klug)

Dorsten, A. C. van., Mellema, J. E. & Premsela, H. F. 1968 In *Electron microscopy 1968*, vol. 2 (ed. D. S. Bocciarelli). Rome: Tipografia Poliglotta Vaticana.
De Rosier, D. J. & Klug, A. 1968 *Nature, Lond.* **217**, 130.
De Rosier, D. J. & Moore, P. 1970 *J. molec. Biol.* **52**, 355.
Erickson, H. 1971 Manuscript in preparation.
Finch, J. T., Klug, A. & Nermut, M. V. 1967 *J. Cell Sci.* **2**, 587.
Kiselev, N. A., De Rosier, D. J. & Klug, A. 1968 *J. molec. Biol.* **35**, 561.
Reimer, L. 1969 *Z. Naturf.* **24***a*, 377.
Schiske, P. 1968 In *Electron microscopy 1968*, vol. 2 (ed. D. S. Bocciarelli). Rome: Tipografia Poliglotta Vaticana.
Thon, F. 1966 *Z. Naturf.* **21***a*, 476.