

An examination of an iterative method for the solution of the phase problem in optics and electron optics: I. Test calculations

To cite this article: D L Misell 1973 *J. Phys. D: Appl. Phys.* **6** 2200

View the [article online](#) for updates and enhancements.

You may also like

- [Four-step phase-shifting digital holography simultaneously sensing dual-wavelength information using a monochromatic image sensor](#)
Tatsuki Tahara, Ryota Mori, Yasuhiko Arai et al.
- [In-focus quantitative intensity and phase imaging with the numerical focusing transport of intensity equation method](#)
Xiaolin Tian, Xin Meng, Wei Yu et al.
- [An examination of an iterative method for the solution of the phase problem in optics and electron optics: II. Sources of error](#)
D L Misell

ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

An examination of an iterative method for the solution of the phase problem in optics and electron optics: I. Test calculations

D L Misell†

Department of Physics, Queen Elizabeth College, Campden Hill Road, London, W8 7AH

Received 18 May 1973, in final form 19 July 1973

Abstract. An evaluation is made of an iterative method for determining the amplitude and phase from the image intensity recorded in optical systems. The method, which requires two images recorded at different lens defocus values, is tested with simulated data subject to error arising from the photographic recording of the image. In the case of error-free data, the solution for the phase distribution appears to be indeterminate to within a constant. The results for photographic noise levels of up to 20% of the maximum image intensity reflect the small effect of error on the calculated phase distribution. The calculation of phase distributions for both symmetric and asymmetric amplitude-phase distributions shows that the use of two images, taken at defocus values differing by about 100 nm in electron optics and about 1 mm in optics (depending on the numerical aperture of the objective lens), may be used to determine the complex object wavefunction in both dark-field and bright-field optics.

1. Introduction

In both optics and electron optics the image intensity $j_1(r_i)$ is measured as a function of image coordinates r_i , and an attempt is made to infer from this intensity distribution the structure of the object (specimen) through its light or electron scattering properties. The transmission function of the specimen is a complex function $\psi_0(r_0) = \epsilon(r_0) + i\eta(r_0)$, for object coordinates r_0 , reflecting the wave nature of the interaction between the incident radiation and the specimen; in electron optics this interaction occurs through the potential distribution in the object and ψ_0 should reflect the atomic positions. In electron optics there is the particular problem of inelastic electron scattering which contributes to the final image as a relatively unstructured background (Misell 1973a). In optics a lens system modifies $\psi_0(r_0)$ to $\psi_1(r_i)$ causing substantial distortions to ψ_0 as a result of lens defects, such as spherical aberration, axial astigmatism and defocus of the objective lens. There is also the problem resulting from the removal from the image of the higher spatial frequencies of the object caused by using an aperture in the back focal plane of the objective lens in an attempt to reduce the lens aberration effects in the image plane. For coherent illumination of the specimen, we can write the linear relation between ψ_0 and ψ_1 (for unit magnification and using the isoplanatic approximation—Born and Wolf 1959) as

$$\psi_1(r_i) = \int \psi_0(r_0) G_1(r_i - r_0) dr_0 \quad (1)$$

† Now at: National Institute for Medical Research, The Ridgeway, Mill Hill, London, NW7 1AA.

neglecting the contribution of inelastic electron scattering. $G_1(\mathbf{r})$ is the resolution function of the lens system, including the effects of lens apertures.

If ψ_1 could be measured then the inversion of equation (1) to give ψ_0 would be a well-defined problem—image deconvolution. However, as only $j_1(\mathbf{r}_i) = |\psi_1(\mathbf{r}_i)|^2$ can be measured, the equation to be solved for $\psi_0(\mathbf{r}_0)$ is nonlinear; that is,

$$j_1(\mathbf{r}_i) = |\psi_1(\mathbf{r}_i)|^2 = \left| \int \psi_0(\mathbf{r}_0) G_1(\mathbf{r}_i - \mathbf{r}_0) d\mathbf{r}_0 \right|^2. \quad (2)$$

The solution of equation (2) for ψ_0 , where only the modulus of the complex function of equation (1) is determined, is a problem as yet to be satisfactorily solved in optics or electron optics; even the problem of the uniqueness of the solution of equation (2) has not yet been evaluated. In bright-field electron microscopy, allowing the unscattered component of the transmitted electron beam to contribute to the image, an approximate linear relation between $j_1(\mathbf{r}_i)$ and $\psi_0(\mathbf{r}_0)$ can be developed (eg Hoppe 1970, 1971, Erickson and Klug 1971, Lenz 1971, Thon 1971, Frank 1972, Hoenders 1972). Writing $\psi_0(\mathbf{r}_0) \simeq 1 + \epsilon(\mathbf{r}_0) + i\eta(\mathbf{r}_0)$, where ϵ and η are much less than unity (weak phase/weak amplitude object) gives

$$j_1(\mathbf{r}_i) = 1 + 2 \operatorname{Re} [\psi(\mathbf{r}_i)] + \operatorname{Re}^2 [\psi(\mathbf{r}_i)] + \operatorname{Im}^2 [\psi(\mathbf{r}_i)] \quad (3)$$

with

$$\operatorname{Re} [\psi(\mathbf{r}_i)] = \int \eta(\mathbf{r}_0) q'(\mathbf{r}_i - \mathbf{r}_0) d\mathbf{r}_0 + \int \epsilon(\mathbf{r}_0) q(\mathbf{r}_i - \mathbf{r}_0) d\mathbf{r}_0$$

and the last two squared terms in equation (3) are neglected in comparison with the second term. $-q'(\mathbf{r})$ and $q(\mathbf{r})$ are respectively the imaginary and real parts of the resolution function $G(\mathbf{r})$. Thus equation (3) gives a linear relationship between $j_1(\mathbf{r}_i)$ and $\epsilon(\mathbf{r}_0)$, $\eta(\mathbf{r}_0)$, the two unknown parameters characterizing the object wavefunction, and at least two electron micrographs, taken at different defocus values (thus varying $G(\mathbf{r})$ and $j(\mathbf{r})$) are required to determine both ϵ and η (Erickson and Klug 1971, Frank 1972); in this linear approximation the contribution of inelastic electron scattering to the image intensity is neglected. The linear approximation is valid only in bright-field microscopy, where the interaction terms ϵ , η in the equation for ψ_0 are small compared with unity (representing the unscattered amplitude), as it is explicitly assumed that η , representing the electron phase shift on transmission through the specimen, is less than unity (~ 0.1). The validity of this approximation for heavily stained materials is doubtful (Grinton and Cowley 1971). Similar approximations may be made in optics, particularly in phase contrast optical microscopy. In dark-field microscopy $\psi_0(\mathbf{r}_0) = \epsilon(\mathbf{r}_0) + i\eta(\mathbf{r}_0)$ and the second-order terms are the sole contribution to $j_1(\mathbf{r}_i)$. Thus dark-field microscopy is avoided because of the difficulties of image analysis and, in electron microscopy, dark-field optics is avoided because of the radiation damage in the specimen caused by the increased photographic recording times compared with bright-field optics. However, dark-field microscopy achieves a clear advantage over the bright-field configuration: namely, that all spatial frequencies of the object transmitted by the objective aperture are present in the image with unit weighting; in contrast, using the weak phase/weak amplitude approximation in bright-field optics, certain spatial frequencies are absent from the image because of zeros in the appropriate transfer function of the objective lens. One way to avoid the nonlinear relation (2) is by the use of incoherent illumination, giving a linear relation between $j_1(\mathbf{r}_i)$ and $|\psi_0(\mathbf{r}_0)|^2$; that is,

$$j_1(\mathbf{r}_i) = \int |\psi_0(\mathbf{r}_0)|^2 |G_1(\mathbf{r}_i - \mathbf{r}_0)|^2 d\mathbf{r}_0. \quad (4)$$

Equation (4) may be inverted without approximation to give information on

$|\psi_0(\mathbf{r}_0)|^2$ only; that is, the magnitude of $\epsilon^2 + \eta^2$. Evidently for a phase object $\psi_0(\mathbf{r}_0) \simeq \exp(i\eta(\mathbf{r}_0))$, $|\psi_0(\mathbf{r}_0)|^2 \simeq 1$ and the image gives no useful information on $\psi_0(\mathbf{r}_0)$. The assumption of incoherent illumination for the evaluation of $|\psi_0(\mathbf{r}_0)|^2$ was made by Stroke and Halioua (1972) for an image taken in the scanning transmission electron microscope; however, electron-optical instruments usually use partially coherent illumination, and the validity of equation (4) in image analysis is doubtful. The important question then arises whether, in optics, information on the amplitude $|\psi_0|$ is sufficient to define the structure of a specimen. In optics this question has a definite answer: the amplitude $|\psi_0|$ carries very little information on the specimen structure—the information is restricted to its absorption properties but excludes its phase shifting properties (O'Neill and Walther 1963, Pefina 1963, Walther 1963). In coherence theory the information on $|\psi_0|$ which is available is of little value in the determination of the complex degree of coherence (eg Wolf 1962, Mehta *et al* 1966, Nussenzveig 1966, Pefina 1972). In optics, using coherent illumination of the object, the paper by Kermisch (1970) shows clearly that the amplitude information can be discarded and the phase information alone used to reconstruct the object—the reverse is not true; a simple example is an object where the transmitted amplitude is unity but the phase shifts, due to optical path differences, are nonzero. Since most of the electron-optical theory has been derived from optics, there seems to be a strong case for attempting a determination of ψ_0 rather than $|\psi_0|$. The fundamental problem of relating ψ_0 to the object structure, even neglecting radiation damage, in electron microscopy is not easy to solve; only for a weak interaction of the incident beam with the specimen can $\psi_0(\mathbf{r}_0)$ be directly related to the object potential distribution $V(\mathbf{r}_0)$ (eg Grinton and Cowley 1971)—in this case there is a strong case for using bright-field optics and the associated linear equation (3). In the practical case of a specimen stained with heavy metal, plural interactions invalidate any simple relation between $\psi_0(\mathbf{r}_0)$ and $V(\mathbf{r}_0)$.

The purpose of this paper is to examine an iterative method for the retrieving from $j_1(\mathbf{r}_i)$ information on the complex form of the object wavefunction $\psi_0(\mathbf{r}_0)$ under the general conditions that neither $\epsilon(\mathbf{r}_0)$ nor $\eta(\mathbf{r}_0)$ need be small, and assuming coherent illumination of the specimen. The method, which has been outlined (Misell 1973b), requires at least two images taken at different defocus values of the objective lens; the difference in defocus between two images Δf may be measured in both optics and electron optics. In this paper an examination is made for test distributions of the viability of the method for iteration between the two image amplitudes $|\psi_1(\mathbf{r}_i)|$ and $|\psi_2(\mathbf{r}_i)|$, both images of the same object $\psi_0(\mathbf{r}_0)$. In §3 we examine image profiles where the object wavefunction $\psi_0(\mathbf{r}_0) = |\psi_0(\mathbf{r}_0)| \exp[i\phi_0(\mathbf{r}_0)]$ is represented by gaussian profiles of varying structure, for both the amplitude $|\psi_0|$ and phase angle ϕ_0 . We also investigate the effect of asymmetry on the validity of the solutions for the phase distribution. Photographic noise is superimposed on the two image intensity distributions $j_1(\mathbf{r}_i)$ and $j_2(\mathbf{r}_i)$ to simulate the experimental situation. The iterative method presented in this paper is shown to work in both dark-field and bright-field optics, the latter situation having already been considered extensively in the literature. In §2 we examine the nonlinear equation that relates $|\psi_1(\mathbf{r}_i)|$ to $|\psi_2(\mathbf{r}_i)|$, the two image amplitudes, and the phase distribution $\phi_1(\mathbf{r}_i)$ of one of the images; the non-uniqueness of the solution for $\phi_1(\mathbf{r}_i)$ is discussed.

The present method to determine the phase distribution from two image amplitudes using different focal planes for the image recording is closely related to the method of Gerchberg and Saxton (1972), which uses the diffraction plane and image plane information as a basis for an iterative scheme. This method has been applied by Gerchberg (1972) to the bright-field image of stained catalase crystals, where the phase distribution

appears not to depend too critically on the image contrast, which may be taken to be zero; that is, $j_1(r_i) = \text{constant}$. The present method, in contrast to that given by Gerchberg and Saxton, uses the contrast differences between two images to obtain a solution for the phase distribution. The method given by Gerchberg and Saxton, in common with the present method, has not yet examined the uniqueness problem in detail. In optics the non-uniqueness of various methods for the solution of the phase problem has been detailed (eg Wolf 1962, O'Neill and Walther 1963, Roman and Marathay 1963, Walther 1963, Nussenzveig 1967, Peřina 1972).

2. Mathematical basis of the method

The basic idea of the iterative scheme is to use the objective lens as a nonlinear phase shifting device, to create two images of the same object by a defocus change. The two image wavefunctions $\psi_1(r_i)$ and $\psi_2(r_i)$, corresponding to defocus values Δf_1 and Δf_2 , may be related to the integral equation (Misell 1973b)

$$\psi_2(r) = \int \psi_1(r') G(r-r') dr' \quad (5)$$

where $G(r)$ is the Fourier transform of the objective lens transfer function, $T(\nu)$, for a defocus difference $\Delta f = \Delta f_2 - \Delta f_1$; that is,

$$G(r) = q(r) - iq'(r) = F^{-1} [\exp(-iK_0\Delta f\nu^2\lambda_0^2/2) B(\nu)]. \quad (6)$$

$K_0 = 2\pi/\lambda_0$ for radiation of wavelength λ_0 ; ν is the spatial frequency transmitted by the aperture function $B(\nu)$. Since the lens aberration terms—eg spherical aberration and axial astigmatism—are common to both ψ_1 and ψ_2 , only the difference term involving Δf is required to give a relation between ψ_1 and ψ_2 . Equation (5) gives a unique ψ_2 from a given ψ_1 , a different Δf giving a different ψ_2 . The spatial frequency ν is related to a spacing r in the object by $\nu r \simeq 1$; and the maximum value of ν , ν_{max} , giving a minimum spacing $r_{\text{min}} \simeq 1/\nu_{\text{max}}$, is related to the semi-angle of the objective aperture α by $\nu\lambda_0 = \alpha$. In electron microscopy a typical value for ν_{max} is 4 nm^{-1} ($\alpha = 0.015 \text{ rad}$) for 100 keV incident electrons ($\lambda_0 = 3.7 \text{ pm}$) giving $r_{\text{min}} \simeq 0.25 \text{ nm}$, and in optics with $\lambda_0 = 500 \text{ nm}$, $\nu_{\text{max}} \simeq 100 \text{ mm}^{-1}$ ($\alpha \simeq 0.05 \text{ rad}$) giving $r_{\text{min}} \simeq 10 \text{ }\mu\text{m}$.

The iterative scheme uses the two image intensities $j_1(r)$ and $j_2(r)$ using equation (5) as the basis for iteration (Misell 1973b). Starting with $j_1 \rightarrow |\psi_1|$, we assign a set of phase angles to $|\psi_1|$ to give $|\psi_1| \exp(i\phi_1)$, which is then convoluted with G and a first approximation to ψ_2 , ψ_2' , is determined. The modulus of ψ_2' is compared with the actual amplitude distribution $|\psi_2|$ (determined from j_2) and the wavefunction is reset to $|\psi_2| \exp(i\phi_2)$. A convolution of this modified ψ_2 with the function G' (corresponding to the resolution function for a defocus difference $-\Delta f = \Delta f_1 - \Delta f_2$) gives a result for ψ_1' which may be compared with the actual amplitude distribution $|\psi_1|$. Resetting ψ_1' to $|\psi_1| \exp(i\phi_1)$ the iteration is continued until a comparison of $|\psi_1'|$ with $|\psi_1|$ gives only a specified small difference. This method is used to obtain the results in §3. However, the iterative method is unsuited to an evaluation of the uniqueness of the phase distribution, and we examine below an explicit nonlinear relationship between $|\psi_1|$, $|\psi_2|$ and the phase angle ϕ_1 of the first image. A method for the solution of this nonlinear equation is indicated but so far a viable numerical method has not yet been determined.

2.1. A nonlinear equation

The image wavefunction $\psi_1(\mathbf{r})$ is written as (omitting all subscripts i)

$$\psi_1(\mathbf{r}) = \epsilon_1(\mathbf{r}) + i\eta_1(\mathbf{r}) = |\psi_1(\mathbf{r})| \exp(i\phi_1(\mathbf{r})) \quad (7)$$

where the phase angle $\phi_1(\mathbf{r}) = \tan^{-1}(\eta_1(\mathbf{r})/\epsilon_1(\mathbf{r}))$. From equation (5), $\psi_2(\mathbf{r})$ can be written in terms of $\psi_1(\mathbf{r})$, equation (7), as

$$\psi_2(\mathbf{r}) = (|\psi_1(\mathbf{r})| \exp(i\phi_1(\mathbf{r}))^*(q(\mathbf{r}) - iq'(\mathbf{r}))) \quad (8)$$

where the $*$ indicates the convolution of two functions, ψ_1 and G in this case. The expression for $|\psi_2|^2$, the second image intensity, is

$$\begin{aligned} |\psi_2(\mathbf{r})|^2 = & [(|\psi_1(\mathbf{r})| \cos \phi_1(\mathbf{r}))^* q(\mathbf{r}) + (|\psi_1(\mathbf{r})| \sin \phi_1(\mathbf{r}))^* q'(\mathbf{r})]^2 \\ & + [(|\psi_1(\mathbf{r})| \sin \phi_1(\mathbf{r}))^* q(\mathbf{r}) - (|\psi_1(\mathbf{r})| \cos \phi_1(\mathbf{r}))^* q'(\mathbf{r})]^2 \\ = & (F_1(\mathbf{r}) + F_2(\mathbf{r}))^2 + (F_3(\mathbf{r}) - F_4(\mathbf{r}))^2. \end{aligned} \quad (9)$$

This nonlinear integral equation relates the measured distributions $|\psi_1(\mathbf{r})|$ and $|\psi_2(\mathbf{r})|$ and the known functions $q(\mathbf{r})$, $q'(\mathbf{r})$ to the required phase angle $\phi_1(\mathbf{r})$. We can first show from equation (9) that not only is $\phi_1(\mathbf{r})$ indeterminate to within a factor $2n\pi$, but it is also indeterminate to within a constant ϕ_c ; that is, $\phi_1(\mathbf{r}) = \phi_1(\mathbf{r}) + \phi_c$ is also a solution of equation (9). Writing $\phi_1(\mathbf{r}) = \phi_1(\mathbf{r}) + \phi_c$ in equation (9) gives the following result for $|\psi_2(\mathbf{r})|^2$:

$$\begin{aligned} |\psi_2(\mathbf{r})|^2 = & \cos^2 \phi_c (F_1(\mathbf{r}) + F_2(\mathbf{r}))^2 + \sin^2 \phi_c (F_4(\mathbf{r}) - F_3(\mathbf{r}))^2 + 2 \sin \phi_c \cos \phi_c (F_1(\mathbf{r}) + F_2(\mathbf{r})) \\ & \times (F_4(\mathbf{r}) - F_3(\mathbf{r})) + \cos^2 \phi_c (F_3(\mathbf{r}) - F_4(\mathbf{r}))^2 + \sin^2 \phi_c (F_1(\mathbf{r}) + F_2(\mathbf{r}))^2 \\ & + 2 \sin \phi_c \cos \phi_c (F_3(\mathbf{r}) - F_4(\mathbf{r})) (F_1(\mathbf{r}) + F_2(\mathbf{r})). \end{aligned} \quad (10)$$

Equation (10) reduces exactly to the form of equation (9). This type of non-uniqueness for $\phi_1(\mathbf{r})$ is unimportant, since we are interested in the variation of ϕ_1 about a mean or constant background level and the absolute level of ϕ_1 is unimportant. For example, in the electron microscopy of a thin object, the phase angle is related to the projection of the object potential distribution $V(\mathbf{r}, z)$ onto the (x, y) plane:

$$\phi_0(\mathbf{r}) \simeq -\text{const.} \int_0^t V(\mathbf{r}, z) dz \quad (11)$$

for specimen thickness t . Thus the variation in $\phi_0(\mathbf{r})$ gives a direct measure of the variation in the potential distribution across the object, and the constant background level of $\phi_0(\mathbf{r})$ gives information only on the absolute level of $V(\mathbf{r})$. Thus in the present method for evaluating the phase distribution, we determine solutions which differ from the actual test values by an arbitrary constant ϕ_c . However, the present work does not exclude the possibility of the solution for the phase being non-unique in other ways (eg Gerchberg and Saxton 1972).

2.2. A method for the solution of the nonlinear equation

We give here an outline of a method to linearize the integral equation (9), but at present we do not have a viable numerical method to solve the resulting linear integral equation. We rewrite equation (9) as

$$\begin{aligned} B(\mathbf{r}) = & \left[\int f(\mathbf{r}, \mathbf{r}') \cos [\phi_1(\mathbf{r}') + \beta(\mathbf{r}')] d\mathbf{r}' \right]^2 + \left[\int f(\mathbf{r}, \mathbf{r}') \sin [\phi_1(\mathbf{r}') + \beta(\mathbf{r}')] d\mathbf{r}' \right]^2 \\ = & I_1^2(\mathbf{r}) + I_2^2(\mathbf{r}) \end{aligned} \quad (12)$$

where

$$\left. \begin{aligned} B(\mathbf{r}) &= |\psi_2(\mathbf{r})|^2 \\ f(\mathbf{r}, \mathbf{r}') \cos \beta(\mathbf{r}') &= |\psi_1(\mathbf{r}')| q(\mathbf{r} - \mathbf{r}') \\ f(\mathbf{r}, \mathbf{r}') \sin \beta(\mathbf{r}') &= -|\psi_1(\mathbf{r}')| q'(\mathbf{r} - \mathbf{r}'). \end{aligned} \right\} (13)$$

If $\phi_1(\mathbf{r})$ is an initial approximation to the phase distribution, then an improvement to $\phi_1(\mathbf{r})$ can be obtained by substituting $\phi_1(\mathbf{r}) + \Delta\phi_1(\mathbf{r})$ in equation (12) to give a linear equation in $\Delta\phi_1(\mathbf{r})$, provided the corrections to $\phi_1(\mathbf{r})$, $\Delta\phi_1(\mathbf{r})$ are first-order. Equation (12) becomes

$$B(\mathbf{r}) = I_1^2(\mathbf{r}) - 2I_1(\mathbf{r}) \int f(\mathbf{r}, \mathbf{r}') \Delta\phi_1(\mathbf{r}') \sin(\phi_1(\mathbf{r}') + \beta(\mathbf{r}')) d\mathbf{r}' + I_2^2(\mathbf{r}) + 2I_2(\mathbf{r}) \int f(\mathbf{r}, \mathbf{r}') \Delta\phi_1(\mathbf{r}') \cos(\phi_1(\mathbf{r}') + \beta(\mathbf{r}')) d\mathbf{r}'. \quad (14)$$

If we write $I_1(\mathbf{r}) = g(\mathbf{r}) \sin \gamma(\mathbf{r})$ and $I_2(\mathbf{r}) = g(\mathbf{r}) \cos \gamma(\mathbf{r})$, equation (14) becomes

$$\frac{1}{2} \left(\frac{B(\mathbf{r}) - g(\mathbf{r})^2}{g(\mathbf{r})} \right) = \int f(\mathbf{r}, \mathbf{r}') \Delta\phi_1(\mathbf{r}') \cos(\phi_1(\mathbf{r}') + \beta(\mathbf{r}') + \gamma(\mathbf{r})) d\mathbf{r}' \quad (15)$$

which is a linear integral equation in $\Delta\phi_1(\mathbf{r})$.

3. Numerical tests of the method

We consider the particular application of the iterative method to high-resolution electron microscopy with a potential image resolution 0.1–0.3 nm; the scaling of various parameters to the resolution achieved in the electron microscopy of biological specimens ~1 nm and in optics will be discussed below. The model taken for $\psi_0(\mathbf{r}_0)$ is of gaussian form and for simplicity we assume that $\psi_1(\mathbf{r}_i)$ is the convolution of $\psi_0(\mathbf{r}_0)$ with the aperture function to give only a diffraction limited image ψ_1 ; in practice ψ_0 will be convoluted with a resolution function, including the effects of lens aberrations, but the results are more conveniently presented using this simplified model. With an objective aperture of $\alpha = 0.015$ rad, corresponding to $\nu_{\max} = 4 \text{ nm}^{-1}$ for 100 keV electrons ($\lambda_0 = 3.7 \text{ pm}$), the resolution in the diffraction limited image is about 0.25 nm, and the gaussian chosen for $|\psi_0|$ has an amplitude–phase distribution of radial half-width about 0.5 nm and an intensity distribution of half-width about 0.3 nm. Initially we have taken the forms of the amplitude and phase distributions to be identical, with various composite gaussian profiles (§3.1), followed by examples where $|\psi_0(\mathbf{r}_0)|$ and $\phi_0(\mathbf{r}_0)$ have different gaussian profiles (§3.2), and finally we examine the effects of asymmetry in $|\psi_0|$ or ϕ_0 on the solution for the phase distribution (§§3.3 and 3.4). All these examples are taken in dark-field optics, where there is no background contribution from the unscattered component of the transmitted electron beam. In §3.5 we demonstrate the validity of the present iterative scheme in bright-field optics, where the weak phase/weak amplitude approximation may be valid.

Using the gaussian forms for $|\psi_0(\mathbf{r}_0)|$ and $\phi_0(\mathbf{r}_0)$ we calculate

$$\psi_1(\mathbf{r}) = \psi_0(\mathbf{r}) \star (\text{aperture function}); \quad (16)$$

and, choosing a defocus difference of $\Delta f = 100 \text{ nm}$, we calculate the wavefunction $\psi_2(\mathbf{r})$ using equation (5). In theory a defocus difference as small as 10 nm could be used in the test calculations that follow, but when photographic noise is added to the image intensities, in addition to other systematic errors (eg defocus error) the differences between

the two images must be significant, and a defocus difference of 50–100 nm is the smallest practical difference for Δf . The two image intensities $j_1(r_i)$ and $j_2(r_i)$ are calculated from $|\psi_1(r_i)|$ and $|\psi_2(r_i)|$ respectively, choosing the maximum values of $|\psi_1|$ to give a maximum optical density of almost unity, and photographic noise is superimposed on the intensity distributions j_1 and j_2 , as would occur in an experimental situation. The object in the following test calculations is then to determine the phase angle $\phi_1(r_i)$, which should have a similar form to the object phase distribution $\phi_0(r_0)$.

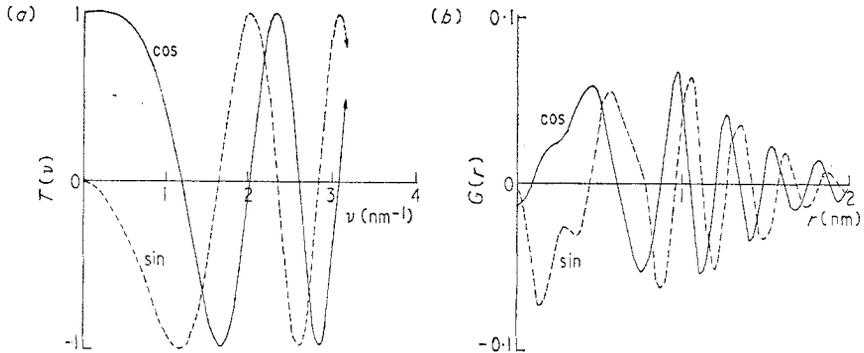


Figure 1. The transfer function $T(v)$ and the resolution function $G(r)$ corresponding to a defocus value $\Delta f=100$ nm; the full curve is the real part (cos) and the broken curve is the imaginary part (sin) of $T(v)$ or $G(r)$.

In figure 1 we show the optical transfer function $T(v)$ (figure 1(a)—real (cosine) and imaginary (sine) parts) giving the resolution function $G(r)$ (figure 1(b)) for a defocus $\Delta f=100$ nm. Note that since $|T(v)|=1$ in dark-field optics all spatial frequencies are given unit weighting in the image. Because of the profile of $G(r)$, we expect $|\psi_2(r_i)|$ to be significantly distorted from the form of $|\psi_1(r_i)|$, this distortion depending on how much structure $|\psi_1|$ displays.

3.1. Amplitude and phase distributions of identical form

Figure 2(a) shows the case of a single gaussian form for both ψ_0 and ϕ_0 as affected by the aperture function to give $\psi_1(r_i)=|\psi_1(r_i)| \exp(i\phi_1(r_i))$ (full curves) with the maximum value of $\phi_1 \simeq \pi$ in order to give a significant phase variation across the specimen. The broken curve $|\psi_2|$ of figure 2(a) shows the second image amplitude, which is used in the iteration with $|\psi_1|$. The iteration scheme outlined in §2 is terminated when the sum squared difference between the actual value of $|\psi_1|$ and the calculated estimate $|\psi_1'|$

$$\sum_{\text{all points}} (|\psi_1(r_i)| - |\psi_1'(r_i)|)^2$$

attains a prescribed value, say 0.001, relative to the total energy density

$$\sum_{\text{all points}} |\psi_1(r_i)|^2$$

in the case of error-free data. In this case the difference between $|\psi_1|$ and $|\psi_1'|$ is not significant enough to show graphically in figure 2(a). The calculated phase distribution ϕ_1' , which differs from ϕ_1 by a constant ϕ_c , is used to scale ϕ_1' to match ϕ_1 (see §2.1)—this scaling constant ϕ_c is chosen arbitrarily only for the case of perfect data. In the

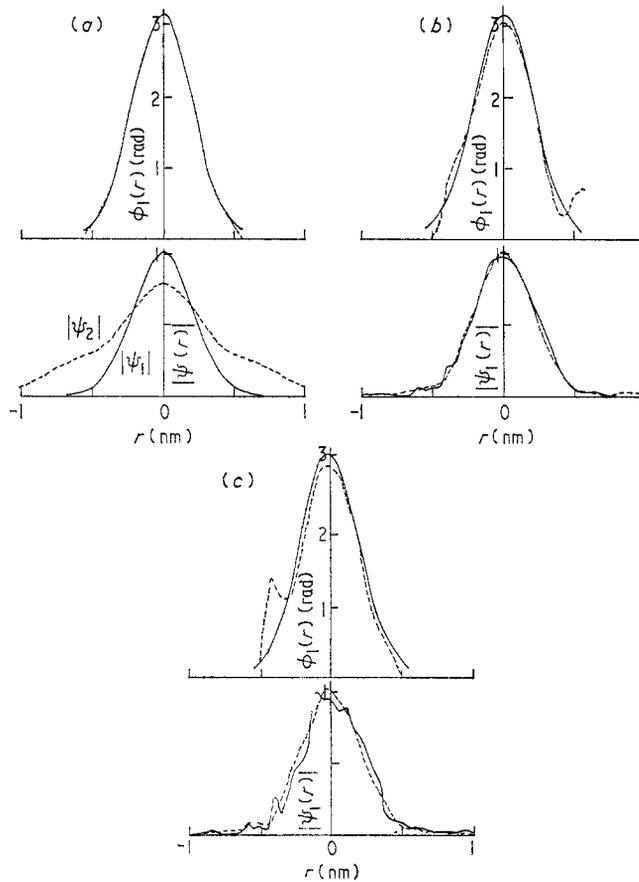


Figure 2. The image amplitude $|\psi_1|$ and the phase distribution $\phi_1(r)$ for a gaussian function (full curves). The broken curve in the amplitude diagram is: (a) the second image amplitude $|\psi_2|$ for a defocus difference $\Delta f=100$ nm; (b) the fitted image amplitude $|\psi_1|$ when the image intensity is subject to error of 10% (full curve); (c) as in (b) but for 20% error. The broken curve in the phase diagram is the fitted phase distribution.

results that are presented for the noise affected data, the solution for ϕ_1' is not renormalized but the same scaling constant ϕ_c taken from the result for perfect data is used; thus the effect of noise on the phase solution can be clearly seen in absolute terms. The phase solution for perfect data is shown as the broken curve in figure 2(a); relatively large deviations from the actual ϕ_1 occur at radial distances larger than 0.5 nm because of numerical errors arising from the small values of $|\psi_1|$ above 0.5 nm.

An important point to note is the form used for the initial phase distribution ϕ_1 in the iterative scheme. Evidently if one has some idea of the functional form of ϕ_1 convergence is extremely rapid (~ 5 iterations). However, in dark-field optics this is most unlikely and it would seem that the initial estimates for ϕ_1 should be based on a slowly varying function of r_i . In practice, random number initial values generated in the range 0 to 2π achieve the best convergence (~ 20 iterations). The discussion of convergence of this type of iterative scheme has been given by Gerchberg and Saxton (1972) and they find that random number initial values are most suited to the iteration scheme. We have verified that different random number sequences for ϕ_1' give similar profiles for

ϕ_1' , differing only in the constant ϕ_c . The author, however, is not satisfied with such a random approach to the phase solution, and a more systematic iterative scheme, such as that suggested in §2.2, is likely to be more predictable in its convergence properties.

In figures 2(b) and 2(c) we show the results for the phase solution where photographic noise is superimposed on both image intensity distributions j_1 and j_2 . The photographic grain noise γ , which is proportional to the square root of the optical density $j^{1/2}$, is characterized by the variation of the optical density about a mean value j_m with a standard deviation σ in the normal gaussian distribution; that is, the probability of an optical density j is determined by

$$\frac{1}{(2\pi)^{1/2}\sigma} \exp\left(-\frac{1}{2\sigma^2}(j-j_m)^2\right) \quad (17)$$

for a region with mean density j_m about an image point r_i (see eg Valentine 1966). Assuming an electron-optical magnification of 500 000 for recording j on a photographic emulsion, σ varies from about 0.02 (fine grain) to 0.2 (coarse grain). In the present test calculations we chose the larger values for σ of 0.1 and 0.2 to demonstrate the adverse effects of the noise on the solution for the phase distribution. At low optical densities, ~ 0.1 , the gaussian relation (17) is invalid and Poisson statistics are used to describe the effects of grain noise. The image intensities subject to photographic noise are calculated using the expression

$$j+\gamma$$

in accordance with the above analysis. The square roots of j_1 and j_2 give the image amplitudes that are used in the iterative scheme. Figures 2(b) and 2(c) show $|\psi_1|$ for $\sigma=0.1$ and $\sigma=0.2$ respectively (full curves) and the fitted results $|\psi_1'|$ (broken curves) which represent a least-squares fit to $|\psi_1|$. There is little point in continuing the iterations beyond a convergence of ~ 0.1 measured relative to the total energy density of the image (depending on the noise level), because photographic noise has strictly invalidated the linear relation (5) between ψ_1 and ψ_2 . The phase distributions ϕ_1' calculated (broken curves) using the same ϕ_c derived in the error-free result of figure 2(a) are shown in comparison with the actual gaussian ϕ_1 (full curves). The solution determined for ϕ_1' is a fair representation of the original gaussian form, the large deviations occurring in the region where the amplitudes $|\psi_1|$ and $|\psi_2|$ are significantly affected by noise, namely, at optical densities below 0.2 (corresponding to an image amplitude of about 0.4).

Similar results to those shown in figure 2 are presented in figure 3 for overlapping gaussian distributions separated by 0.4 and 0.6 nm respectively. In figure 3 (Ia) the broader gaussian form for $|\psi_1|$ gives an excellent solution for ϕ_1 (full and broken curves coincide) and the adverse effect of photographic noise—3(Ib, Ic)—on the phase solutions out to radial distances of 0.5 nm is reduced because of the broader intensity distributions. In contrast, the results presented in figure 3(II) for a more structured $|\psi_1|$ and $|\psi_2|$ show the serious effect on the solution for ϕ_1 when the noise level is about 20% of the maximum intensity. However, photographic emulsions with such a large grain noise (corresponding to x-ray emulsions) are seldom used in electron microscopy, unless extremely fast recording is required. The results of figures 2 and 3 give some indication of the noise levels that can be tolerated for simple profiles and show that the present method of phase solution should be restricted to images with noise levels below 10%.

In order to define the effect of photographic noise on $|\psi_1|$ and $|\psi_2|$, and to ensure that the noise we have added to the images of figures 2 and 3 does not correspond only

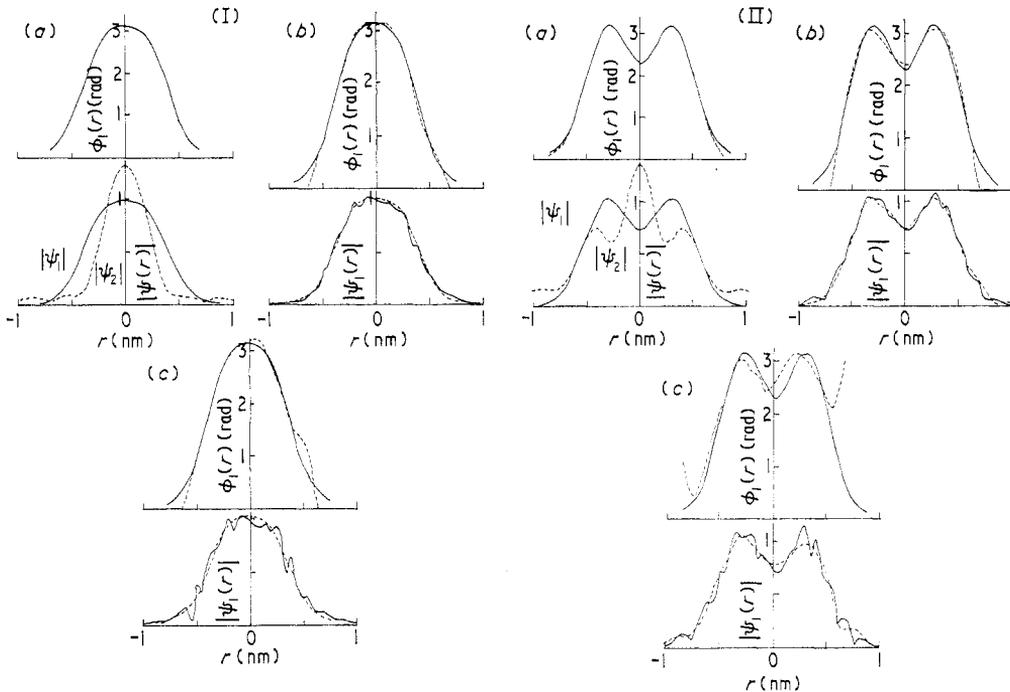


Figure 3. (I) The image amplitude $|\psi_1|$ and the phase distribution $\phi_1(r)$ for overlapping gaussian functions (full curves). The broken curve in the amplitude diagram is: (a) the second image amplitude $|\psi_2|$ for a defocus difference $\Delta f=100$ nm; (b) the fitted image amplitude $|\psi_1|$ when the image intensity is subject to error of 10% (full curve); (c) as in (b) but for 20% error. The broken curve in the phase diagram is the fitted phase distribution. (II) As for (I) but with a different spatial separation of the overlapping gaussian distributions.

to high frequencies (which may be excluded by the aperture function used in equation (5)), we examine the noise spectrum of the $|\psi_1|$ in figure 3(IIc), corresponding to a noise level of about 20%. The power spectrum of $|\psi_1|$, that is

$$|S(\mathbf{v})|^2 = [F(|\psi_1(r_i)|)]^2$$

is presented in figure 4 for the error-free $|\psi_1|$ (full curve) and the noise affected $|\psi_1|$ (broken curve). Within the frequency range (4 nm^{-1}) transmitted by the objective aperture function using equation (5) there is a significant alteration in the power spectrum of $|\psi_1|$. However, a substantial proportion of the noise content of $|\psi_1|$ occurs at frequencies greater than 4 nm^{-1} , and the total energy density eliminated by the aperture corresponds to about 40% of the noise spectrum.

3.2. Amplitude and phase distributions of different radial half-widths

Maintaining the symmetrical gaussian forms for $|\psi_0|$ and ϕ_0 , we consider the phase solutions where ϕ_0 is either of smaller or larger radial half-width than $|\psi_0|$. For the phase distribution in figure 5(I) we have taken for ϕ_0 a double gaussian which has a radial half-width of about 0.5 of that of $|\psi_0|$, $|\psi_0|$ being of similar form to the $|\psi_1|$ given in figure 3(IIa). However, the $|\psi_1|$ given in figure 5(Ia) arises from a convolution of ψ_0

with the aperture function to give a diffraction limited $|\psi_1|$, which will depend on the form of ϕ_0 . The difference between the $|\psi_1|$ in figure 3(IIa) and that in figure 5(Ia) would not depend on the functional form of ϕ_0 if we had used an infinite aperture (with no lens aberrations in the system). Thus the use of an aperture in the image forming system seems to be a requirement for a unique $|\psi_1|$ for a given ϕ_0 (see O'Neill and Walther 1963, Walther 1963). Then the effect of the aperture on the image for the simple gaussian form for $|\psi_0|$ depends on the form of ϕ_0 ; this clearly seen by the comparison of the $|\psi_1|$, $|\psi_2|$ curves in figures 3 (IIa) and 5(Ia). For the form of ϕ_1 chosen for figure 5(I), the fitted phase distributions ϕ_1' are very good, even for the noise affected results 5(Ib, c).

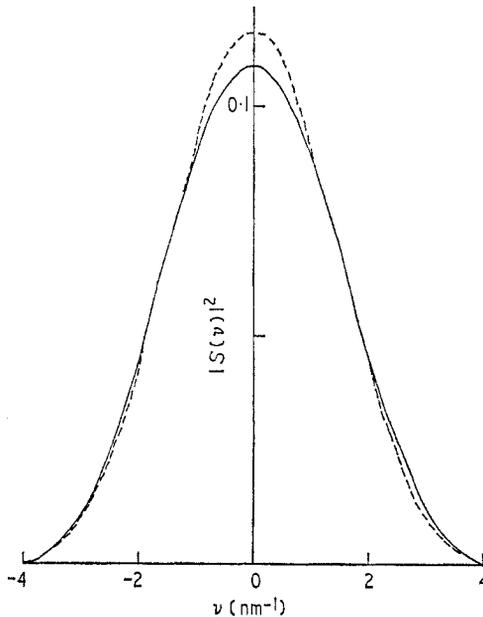


Figure 4. The power spectrum $|S(v)|^2$ for the image amplitude $|\psi_1|$ shown in figure 3(II). The full curve is $|S(v)|^2$ for the error-free data (figure 3(IIa)) and the broken curve represents $|S(v)|^2$ for the noise affected data (figure 3(IIc)—noise level 20% of the maximum image intensity).

The effect of the noise is less marked than in the corresponding results in figures 3(IIb, c), because the phase distribution in figure 5(I) has fallen to very small values at radial distances where $|\psi_1|$ and $|\psi_2|$ are not too seriously affected by noise.

If we now take a broader gaussian form for ϕ_0 approximately twice the radial half-width of $|\psi_0|$ (corresponding to figure 3(I)), $|\psi_1|$ is hardly affected by the aperture and $|\psi_2|$ is correspondingly similar to $|\psi_1|$ (contrast figure 3(Ia)). In fact, in the limit of $\phi_0 \rightarrow$ constant, $|\psi_1|$ and $|\psi_2|$ are almost identical and the phase shift introduced by the defocus Δf integrates out of equation (5). Thus in high-resolution microscopy it could be difficult to determine a phase variation ϕ_0 which is relatively unstructured (broad) compared to the amplitude distribution. However, it is unlikely that $|\psi_0|$ will reflect more structural information than ϕ_0 ; ϕ_0 is most likely to be more sensitive than $|\psi_0|$ to structural variations as in figure 5(I). The fitted results for ϕ_1 in figure 5(II) appear to be very sensitive to error, but this is because in the region of image space where ϕ_1

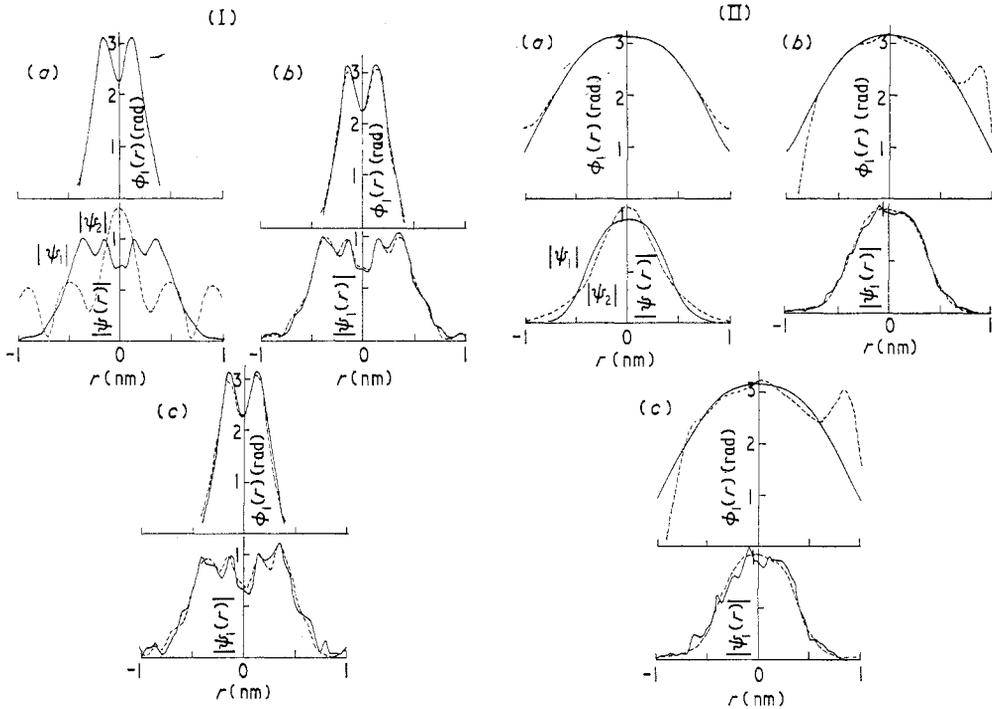


Figure 5. (I) The image amplitude $|\psi_1|$ and the phase distribution $\phi_1(r)$ for overlapping gaussian functions (full curves)— $\phi_1(r)$ is sharper than $|\psi_1|$. The broken curve in the amplitude diagram is: (a) the second image amplitude $|\psi_2|$ for a defocus difference $\Delta f=100$ nm; (b) the fitted image amplitude $|\psi_1|$ when the image intensity is subject to error of 10% (full curve); (c), as in (b) but for 20% error. The broken curve in the phase diagram is the fitted phase distribution. (II) As for (I), but now the phase distribution $\phi_1(r)$ is broader than the amplitude distribution $|\psi_1|$.

is still significant, the amplitudes $|\psi_1|$ and $|\psi_2|$ are dominated by noise effects. In the case of perfect data (figure 5(IIa)) the poor result for ϕ_1 above 0.5 nm, where $\phi_1 \sim 0.5$ (ϕ_1)_{max}, arises from the small values of $|\psi_1|$ and $|\psi_2|$, 0.1 of their maximum values, resulting in numerical integration errors.

3.3. Amplitude distribution asymmetric—phase distribution symmetric

It is useful to examine the validity of the phase solution, when the amplitude and phase forms of ψ_0 are of different functional form. In the first case we take $|\psi_0|$ to be asymmetric and ϕ_0 to be a symmetric gaussian (corresponding to the ϕ_0 of figure 3(Ia)). Here the second image amplitude (figure 6(Ia)) carries the information on the symmetric form of ϕ_0 . The phase solution for ϕ_1 for error-free data is excellent and even the errored results are acceptable. However, when the noise level becomes large the asymmetric information carried by $|\psi_1|$ causes a distortion of ϕ_1' in figure 6(Ic).

3.4. Amplitude distribution symmetric—phase distribution asymmetric

Considering the second type of asymmetry, namely an asymmetric ϕ_0 and a symmetric $|\psi_0|$ (corresponding to figure 3(IIa)), the result for $|\psi_1|$ does show some asymmetry

(figure 6(IIa)) but this useful information on the asymmetry of ϕ_0 is almost completely lost in the photographic noise: figures 6 (IIb, IIc). However, $|\psi_2|$ can be seen to carry substantial information on the asymmetry of ϕ_0 . The general conclusion is that asymmetric phase solutions are adversely affected by error superimposed on $|\psi_1|$ and $|\psi_2|$, the asymmetric information contained in these amplitudes being distorted by the noise.

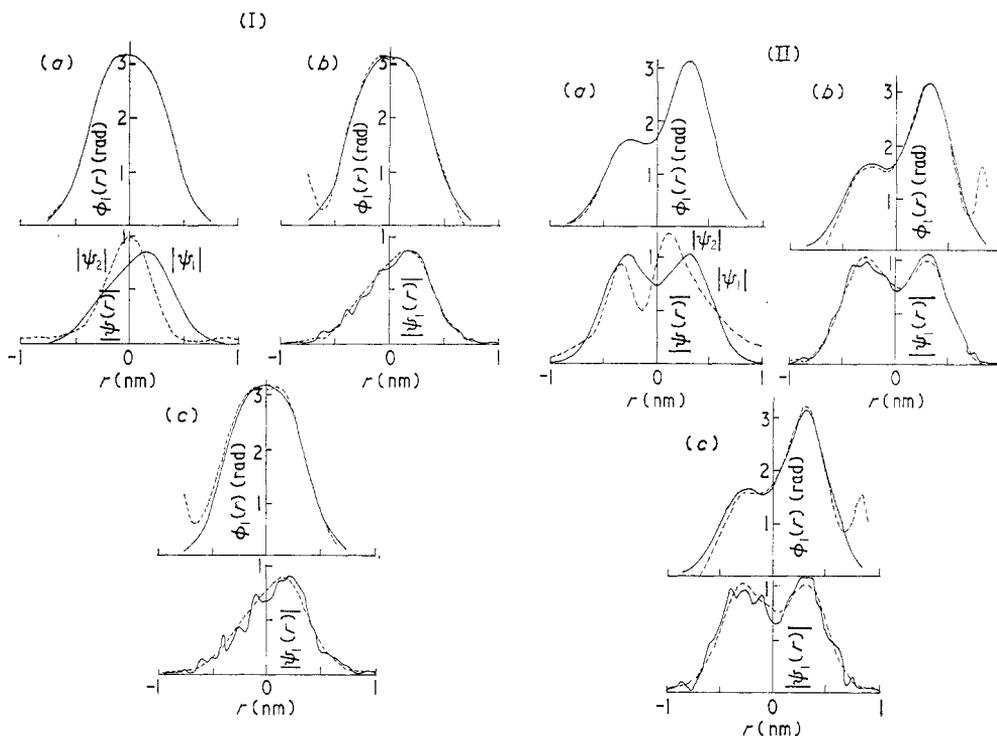


Figure 6. The image amplitude $|\psi_1|$ and the phase distribution $\phi_1(r)$ for overlapping gaussian functions (full curves)— $|\psi_1|$ is symmetrical and ϕ_1 is symmetrical. The broken curve in the amplitude diagram is: (a) the second image amplitude $|\psi_2|$ for a defocus difference $\Delta f=100$ nm; (b) the fitted image amplitude $|\psi_1|$ when the image intensity is subject to error of 10% (full curve); (c), as in (b) but for 20% error. The broken curve in the phase diagram is the fitted phase distribution. (II) As for (I), but now the amplitude distribution $|\psi_1|$ is symmetrical and the phase distribution ϕ_1 is asymmetrical.

3.5. Bright-field optics

In bright-field optics the object wavefunction $\psi_0(r_0)$ is allowed to combine with the unscattered wave, represented by unit amplitude (or some constant value), and the image wavefunction in bright-field optics consists of a modulation, which is not ψ_0 , on a constant background of unit amplitude. Thus the wavefunction forming the image is

$$\psi_b(r_0) = 1 + \psi_0(r_0) \quad (18)$$

and the phase angle is not just $\phi_0(r_0) = \tan^{-1}(\eta(r_0)/\epsilon(r_0))$ as in the dark-field case but $\tan^{-1}[\eta(r_0)/(1 + \epsilon(r_0))]$. The image wavefunctions $\psi_1(r_i)$ and $\psi_2(r_i)$ are calculated in the same way as in §2 for a $|\psi_0(r_0)|$ and a $\phi_0(r_0)$ of gaussian form, with the maximum value of $|\psi_0(r_0)| \sim 0.1$ to simulate the normal bright-field electron micrograph with an image

intensity contrast of about 0.2, the contrast being defined as $(j(r_i) - \text{background intensity}) / (\text{background intensity})$. The bright-field image amplitudes $|\psi_1|$ and $|\psi_2|$ corresponding to a defocus difference of $\Delta f = 100$ nm are shown in figure 7 for the three types of gaussian distribution for $|\psi_0|$ and ϕ_0 used in figures 2 and 3, namely (a) a single gaussian of radial half-width 0.5 nm, (b) two gaussians separated by 0.4 nm, (c) two gaussians separated by 0.6 nm. Note the dissimilar profiles for $|\psi_1|$ and $|\psi_2|$ in bright-field optics as compared with dark-field optics (figures 2(a), 3(Ia, IIa)), the difference arising from interference effects of the unscattered and object wavefunctions in the bright-field case. The most

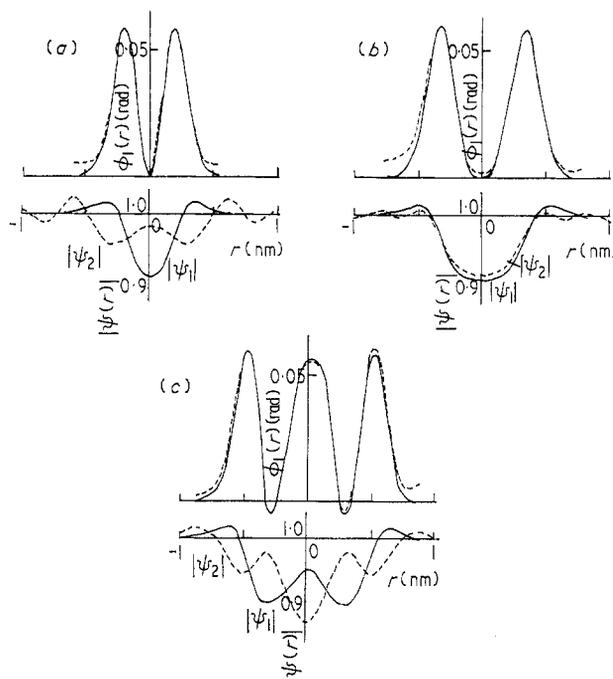


Figure 7. The bright-field image amplitudes $|\psi_1|$ and $|\psi_2|$ calculated for a defocus difference $\Delta f = 100$ nm, and for different gaussian profiles of the object wavefunction. The constant background level is 1.0. The phase distribution ϕ_1 (full curve) correspond to the image wavefunction ψ_1 and the broken curve is the fitted phase distribution.

noticeable difference in the two types of optics is shown by the phase distribution $\phi_1(r_i)$ (full curves in figures 7(a), (b), (c)) which is now $\tan^{-1} [\eta_1 / (1 + \epsilon_1)]$ and not simply $\tan^{-1} (\eta_1 / \epsilon_1)$ as in dark-field optics. The bright-field forms for $\phi_1(r_i)$ are now very different from the form of $\phi_0(r_0)$, and only if the background subtraction can be made is it possible to gain information on the required form of $\phi_0(r_0)$. The main interpretative problem in bright-field optics is then the evaluation of the phase distribution, $\phi_1(r_i)$ in this case, requiring the subtraction of a constant amplitude contribution, which in general cannot be readily determined. The normal method used to eliminate a constant background is to take the Fourier transform of the data, when the constant term should transform in frequency space to a delta function $\delta(v)$, which may be filtered out. In figure 7 we present the phase solution ϕ_1' obtained by an iterative procedure using $|\psi_1|$ and $|\psi_2|$. Because of the rapid variations in ϕ_1 , the convergence to the measured image amplitudes $|\psi_1|$ and $|\psi_2|$ is not as rapid as in the dark-field case. The oscillatory

nature of the $|\psi_1|$ and $|\psi_2|$ shown in figure 7 also leads to numerical integration termination errors in the iterative procedure; the slow convergence of $|\psi_1|$ and $|\psi_2|$ to the constant background value of 1 should be contrasted with the rapid convergence of the dark-field image profiles of figures 2 and 3 to zero. As mentioned above, in order to extract from $\phi_1'(r_i)$ the useful information on the object wavefunction $\psi_0(r_0)$, we must use the background amplitude value (not necessarily unity as assumed here) to modify $\tan^{-1} [\eta/(1 + \epsilon)]$ to give values for $\tan^{-1} (\eta/\epsilon)$.

Since the modulation of the image intensity about the background of unity is only about 0.2 in bright-field electron microscopy (Grinton and Cowley 1971), the noise level, which is proportional to $j^{1/2}$, that is acceptable is considerably less than the 10–20% tolerated in dark-field optics. In fact, although a tolerable fit to $|\psi_1|$ and $|\psi_2|$ can be obtained for a noise level of 0.05 at a mean optical density of about unity, the phase distribution ϕ_1' does not reproduce any of the fine detail in ϕ_1 ; however, the values for ϕ_1' do correspond to a mean level of ϕ_1 . Thus bright-field optics do not seem suited to the present iterative scheme because of the drastic effects on the phase solution ϕ_1' for a normal noise level of 0.05, unless the modulation of the image intensity about the background is significantly greater than 0.2. Additionally, as mentioned in §1, the weighting of the object spatial frequencies in bright-field optics varies and some frequencies are attenuated substantially below a weighting of unity, and then the noise spectrum at these frequencies can completely mask the useful structural information.

The examples presented in this section for the phase solutions determined from two image amplitudes refer to high-resolution electron microscopy, achieving a potential image resolution of 0.25 nm (with $\nu_{\max} = 4 \text{ nm}^{-1}$). In the electron microscopy of biological specimens, the image resolution is probably limited by specimen preparation techniques (eg negative staining) and radiation damage to a resolution of about 1 nm ($\nu_{\max} = 1 \text{ nm}^{-1}$). The calculations presented here may be scaled by a factor of 4 along the r axis to simulate the lower-resolution figure, and the corresponding defocus difference between the two images would be 1600 nm (proportional to r_{\min}^2/λ_0) instead of 100 nm. In the case of such a large defocus value the effect of spherical aberration on the image intensity would be insignificant (Erickson and Klug 1971). The scaling of the results for electron microscopy explicitly assumes that wave optics describes the image contrast effects, rather than a scattering contrast model, which is effectively an intensity (incoherent) mechanism.

In optics with $\lambda_0 \simeq 500 \text{ nm}$, the potential image resolution is approximately 1–10 μm , and the results of figures 1–7 should be scaled appropriately, a factor of about $(4\text{--}40) \times 10^3$ for the r coordinate and the inverse factor for the ν coordinate. A typical defocus difference Δf , to produce the same image differences as presented in the electron-optical examples, would be 10 μm –1 mm for the resolution figures given above.

4. Conclusions

A viable method has been proposed for the evaluation of the amplitude–phase distribution in an image from two images taken at different objective lens defocus values. The method, which makes no assumptions regarding the form of the phase distribution and which is applicable to cases where the phase shifts are not small, has been illustrated by application to test distributions, simulating the situation in high-resolution electron microscopy. The effect of photographic noise on the images and the resulting errors in the phase distributions have been evaluated; in dark-field optics noise levels of about

10% of the maximum image intensity do not drastically affect the validity of the phase solution. The present iterative method works in bright-field optics but appears to be extremely sensitive to errors in the bright-field images, particularly when the modulation of the image intensity about the background (the contrast) is less than 0.2; in this case a noise level of about 0.02 for a mean optical density of unity, equivalent to about 10% on the actual modulated signal, is acceptable. Principally we propose that the present method using two images is a viable method in dark-field optics, where linear approximations are invalid. We have outlined the advantages of dark-field optics over the bright-field situation, but in electron microscopy the reduced electron signal in dark-field means increased photographic exposure times, which may result in increased specimen damage. In dark-field optics all spatial frequencies present in the object are present in the image with unit weighting, except those frequencies which are intercepted by the objective aperture, but in bright-field optics some spatial frequencies are attenuated or even removed from the image, and these spacings are difficult to reconstruct in the presence of noise. Whereas all the test calculations presented in §3 use only two images, in the experimental situation, where there is no indication of the correct form for the phase distribution, it would seem necessary to use a third image in the sequence to test the validity of the phase solutions obtained.

Other relevant questions, that arise in the discussion of the validity of the phase solutions, relate to the effects of various sources of error, namely, background (eg due to inelastic electron scattering), mismatching of the two images, and an error in determining the defocus difference between the two images. The effects of these errors on the phase solution are examined in the companion paper (part II).

Acknowledgments

The author is grateful to Dr K Pitman for advice on the solution of nonlinear integral equations and to Alan Greenaway for advice on photographic recording. The provision of excellent computing facilities by the University of London Computing Centre is acknowledged.

References

- Born M and Wolf E 1959 *Principles of Optics* (London: Pergamon)
- Erickson HP and Klug A 1971 *Phil. Trans. R. Soc., Lond.* **261B** 105–18
- Frank J 1972 *Biophys. J.* **12** 484–511
- Gerchberg RW 1972 *Nature, Lond.* **240** 404–6
- Gerchberg RW and Saxton WO 1972 *Optik* **35** 237–46
- Grinton G R and Cowley J M 1971 *Optik* **34** 221–33
- Hoenders B J 1972 *Optik* **35** 116–33
- Hoppe W 1970 *Acta Crystallogr.* **A26** 414–26
- 1971 *Phil. Trans. R. Soc., Lond.* **216B** 71–94
- Kermisch D 1970 *J. Opt. Soc. Am.* **60** 15–7
- Lenz F 1971 *Proc. 25th Anniv. Meeting of EMAG 1971: Inst. Phys. Conf. Ser. No. 10* pp 224–9
- Mehta CL, Wolf E and Balachandran AP 1966 *J. math. Phys.* **7** 133–8
- Misell DL 1973a *J. Phys. D: Appl. Phys.* **6** L1–5
- 1973b *J. Phys. D: Appl. Phys.* **6** L6–9
- Nussenzeig HM 1967 *J. math. Phys.* **8** 561–72
- O'Neill EL and Walther A 1963 *Optica Acta* **10** 33–40
- Pefina J 1963 *Optica Acta* **10** 337–40
- 1972 *Coherence of Light* (London: Van Nostrand Reinhold Company) pp 55–62

Roman P and Marathay A S 1963 *Nuovo Cim.* **30** 1452–64

Stroke G W and Halioua M 1972 *Optik* **35** 50–65

Thon F 1971 *Electron Microscopy in Material Science* ed U Valdré (New York: Academic Press)
pp 570–625

Valentine R C 1966 *Adv. Opt. Electron Microsc.* **1** 180–203 (New York: Academic Press)

Walther A 1963 *Optica Acta* **10** 41–9

Wolf E 1962 *Proc. Phys. Soc.* **80** 1269–72