ANALYSIS OF OBJECT-IMAGE RELATIONSHIPS IN ELECTRON MICROSCOPY BY IMAGE PROCESSING TECHNIQUES

Thesis by

Henri M. Horgen

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1975

(Submitted July 1, 1974)

A mes parents

ACKNOWLEDGMENTS

I wish to express my deepest gratitude for the stimulating advice and for the continual encouragement that were extended to me during the course of this research by my adviser, Professor R. E. Villagrana. Special thanks are due to Dr. D. M. Maher for the provision of lattice fringe and weak-beam micrographs, for many fruitful discussions, and particularly for giving the opportunity to perform computations at the Bell Telephone Laboratories in Murray Hill. Grateful acknowledgment is made to Dr. R. Nathan for enabling me to carry out this work at the Image Processing Applications Group of the Jet Propulsion Laboratory in Pasadena, I would like to thank Dr. J. Frank for encouraging me to perform an analysis of the practical limitations of the Z discrimination technique and for providing most of the computer programs used in this restoration scheme. Thanks are also due to Dr. N. George of the Institute faculty for the use of the facilities of his laboratory.

The partial support of this research under the Caltech President's Fund CIT 31174 is gratefully acknowledged.

I am thankful for the assistance of Mr. R. Hammond during the implementation of the non-column approximation computer program. My sincere appreciation is offered to Mr. H. Arnal and the technicians of JEOL for their experimental prowess in the specimen preparation and the data collection of the focus series. Special thanks are also due to Mr. F. Youngkin for his outstanding work in photography, and to Mrs. E. Gyrak and Mrs. Ruth Stratton for their unfailing patience during the typing of the manuscript.

Finally, I would like to thank my relatives and all my friends for their understanding and moral support, which enabled me to reach the end of this long quest.

ABSTRACT

By considering an electron microscope to be an information channel, it is shown that the correspondence between the object and the image can be linked to the electron optical characteristics of the instrument and to the statistical properties of the noise. A discussion of the image formation of a cluster of atoms is introduced in order to demonstrate the main contrast mechanisms that operate at the atomic level. A direct extension of this analysis to more complex specimens gives rise to a wave optical theory of image formation, which is used to present the concept of the amplitude transfer function. This formalism greatly simplifies in the case of weakly scattering objects, where the total object wave is linearly related to both the projected potential distribution of the specimen and the amplitude attenuation of the incident beam. For this category of specimens, there exists a linear relationship between the image intensity in bright-field and the total object wave. Phase and amplitude contrast transfer functions describe the perturbing influence of the objective lens aberrations on the phase and the amplitude of the object wave. If spatial and chromatic incoherence effects are included in this formalism, it is shown that the linearity between image intensity and object is preserved.

The validity of the approximations of the wave optical theory is first checked by studying the effect of defocusing on the transfer conditions of the phase and amplitude contrast mechanisms. A medium resolution experiment is conducted on a bright-field image of bovine liver catalase. The results demonstrate a qualitative agreement

between experiment and theory.

A reconstruction scheme is next implemented on a through-focus series of a specimen of gold on carbon. This scheme is analyzed critically prior to a description of the experimental results. It is demonstrated that under certain conditions this technique is capable of restoring the total object wave and simultaneously achieving a selective contrast enhancement at heavy atom locations. This potential Z discrimination is tested experimentally, and the difficulties encountered during the processing are discussed. A qualitative estimate of all contrast mechanisms that contribute to a high-resolution bright-field micrograph can be inferred from this analysis. The problems which one faces in a quantitative interpretation of micrographs at the atomic level are also discussed, and possible ways to circumvent these problems are mentioned.

Next, image processing schemes for improving the signal-to-noise ratio of an image are applied to micrographs of crystalline specimens. The enhancement of lattice fringe images is demonstrated for both silicon and gold specimens. Periodic images can be processed in either real space or Fourier space and an analysis of these processing modes is presented.

By enhancing a weak-beam image of a dissociated-dislocation dipole in germanium, a quantitative comparison to a simulated image is rendered possible. The resolving power of the weak-beam technique is analyzed for this particular example. It is found that by selecting a diffraction geometry so that the systematic reflections are dynamically

interacting, four dislocation peaks are individually resolved. Problems associated with the contrast interpretation are discussed in conjunction with a calculation of the image contrast. Finally, suggestions for further study of the atomic structure of crystalline defects by the weak-beam method are given.

PREFACE

High-resolution electron microscopy is becoming a widely used tool for investigating structures at the atomic level. Of primary consideration to the microscopist is the interpretation of image intensities and their relationship to the structure which is being examined. The cumulative adverse effects of the lens aberrations on image resolution, and of the system noise on image contrast, render a visual interpretation of details of atomic dimension impractical. Noise is introduced by the limited number of electrons that are collected during the image recovery, so that its magnitude is a function of the resolution range at which one operates. Therefore the possibility of interpreting image details in terms of the structure of the object depends strongly on the information content of the specimen. If the conditions of observation are defined with sufficient accuracy, and if a manipulation of the recorded intensities by digital computing methods is undertaken, it is possible to infer the structural properties of the object from its electron microscope image.

This thesis describes the problems to which one is exposed when a one to one correspondence between image and object is desired, given the resolution capability of available instruments. We consider first the problem of obtaining information about the atomic arrangement of a weakly scattering specimen using micrographs taken in bright-field. Then we investigate the potential use of an electron microscope for studying the local arrangement of atomic planes in an imperfect crystal. The methods that will be employed combine high-resolution electron

microscopy in conjunction with subsequent image processing algorithms, and theoretical image contrast analysis.

In order to illustrate how the microscope distorts the phase relationship of the electrons as they propagate from the specimen plane to the image plane, we describe in Chapter One the correspondence between object and image electron wave for a specimen that is a finite assemblage of atoms. For a general specimen, we show that this relationship is characterized by the amplitude transfer function of the instrument.

The next chapter presents a description of the processing algorithms which are implemented in the remainder of this work. The information loss suffered by the data during their conversion from a continuous to a discrete distribution is analyzed; criteria for minimizing distortions associated to the conversion step are established.

In Chapter Three, we limit our considerations to cases where a relatively simple relationship between image intensity and the projected potential of the object can be defined. A detailed analysis of the underlying assumptions of the wave optical theory of image formation is presented, and the domain of applicability of this theory is checked by an experimental study performed at medium resolution. An image reconstruction scheme is next implemented, which restores the phase and amplitude components of the object wave by combining bright-field micrographs taken at different defocus. It is shown that this resolution extending scheme also allows a selective enhancement of the contrast from heavy atoms, if certain experimental conditions are

fulfilled. The complicated procedures for the calculation of correlation functions between the micrographs and for the accurate determination of the pupil function are examined. The problems which one encounters in extending the present instrumental resolution to the atomic range are discussed and suggestions are given in view of the existing technology.

Electron statistics are often the limiting influence when a quantitative interpretation of image contrast is desired. In order to circumvent this problem, one performs a linear averaging of the data, which attenuates the contribution of statistically random noise to the signal. Chapter Four describes two examples of noise reduction algorithms. In the first, the noise that is superimposed onto lattice fringe images is eliminated by extracting the periodic information content of the image. In the second example, the background noise in a weak-beam micrograph of an imperfect crystal is partially removed by multiple averaging of micrographs that have been collected sequentially. The result of this averaging is then compared to a theoretical image simulation. This comparison serves as an illustration of the difficulties which one faces when attempting to obtain a quantitative agreement between experiment and theory. From these investigations, it appears that the weak-beam method could prove to be a powerful means for investigating the strain fields of isolated defects when the data necessary for image contrast analysis can be measured.

TABLE OF CONTENTS

CHAPTER ONE

OBJECT-IMAGE RELATIONSHIPS

CHAPTER ONE

OBJECT-IMAGE RELATIONSHIPS

1.1 Introduction

An electron microscope can be viewed as an optical channel that transmits information about the structural and scattering properties of an object. The incident coherent electron beam is first coded by the specimen, then undergoes various distortions from lens aberrations in the transmission, and is finally recorded by image sensors. The information flow undergoes a sequence of transformations involving information degradation, which prevents a quantitative interpretation of the magnified image[1,2]. In general, these changes in the data stream are caused by the interaction of the electron beam with the specimen, processing of the electron beam by the microscope, and the receiver that is used for image recovery.

In electron microscopy, concepts from the field of Information Theory can be used for image evaluation, prediction, restoration, and enhancement. These concepts also provide useful tools for determining the structural relationships that exist in the object. Viewing the microscope as a transmission channel, we can specify for each input the average mutual information of the channel outputs[3]. For instance, analytical estimates of the channel capacity of a continuous coherent channel have been derived for a weak phase object with additive white Gaussian noise. These estimates yield criteria for assessing the potential image quality. However, these criteria neglect incoherent links in the imaging chain (e.g., photographic recording) that spread

the information and introduce random noise[2,4].

In this chapter we shall use the wave optical theory of image formation to describe how an electron microscope transmits information about an object which consists of a finite number of randomly arranged atoms[5]. We shall then introduce the concept of the amplitude transfer function (ATF), in the isoplanatic approximation, for a continuous object distribution.[6] Finally, we shall analyze the optical noise that is introduced into the image by the recording sensors and by the quantum nature of the electron wave[4].

1.2 Imaging of Atoms with the Electron Microscope

1.2.1 Theoretical Foundations

An electron microscope can be operated in either an imaging mode or a diffraction mode. The optics of these two modes of operation are illustrated in the ray diagrams shown in Fig. 1-1. A schematic diagram showing the coordinate system of the object plane, the back focal plane, and the image plane is given in Fig. 1-2 [7].

The scattering of electrons by a weakly scattered object can be described using the kinematical theory of electron diffraction. According to this theory, the scattered amplitude, $\Psi(\xi,\eta)$ at the aperture plane is proportional to the Fourier transform of the scattering potential of the object. If we assume that a plane-wave beam of monochromatic electrons is incident on an object consisting of N_a atoms located at (x_o^j, y_o^j, z_o^j), then the scattered wave is the sum of the waves scattered by the individual atoms with attention to phase:

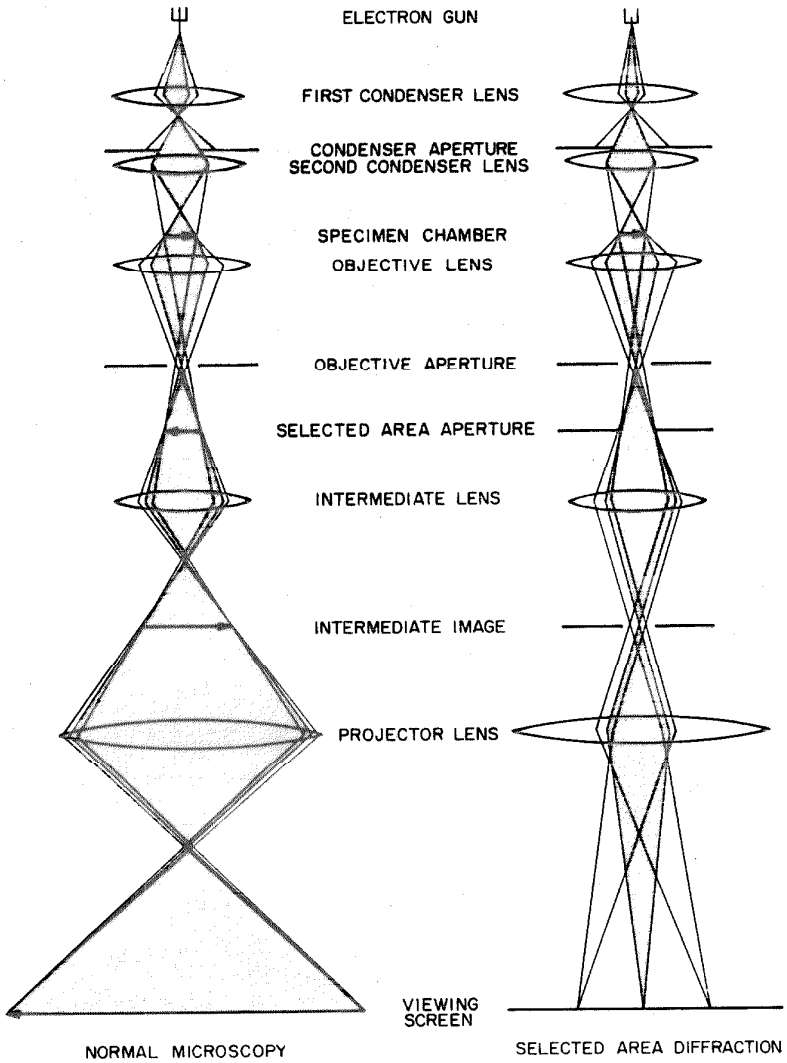


Fig. 1-1. Ray diagram of the two basic modes of operation of an electron microscope. Selected area diffraction is obtained by decreasing the focal length of the intermediate lens.

$$\Psi(\xi,\eta) = \frac{iA_o}{L_o} \sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) - 2\pi iK(\frac{\xi x_o^j + \eta y_o^j}{L_o})] \qquad (1.1)$$

where A_o is the amplitude of the incident electron beam, $K = 1/\lambda$ is the electron wave vector, and $\eta_j(\theta)$ is the anomalous phase shift of the wave which is scattered by the j^{th} atom. Since atoms act as phase objects in the first Born approximation, the scattered electron wave undergoes a phase shift of $\pi/2$. According to the optical theorem, an additional phase term, $\eta_j(\theta)$, must be introduced into the scattering amplitudes in order to conserve the number of particles[8]. The magnitude of $\eta_j(\theta)$ increases with atomic number, Z , and decreases with increasing electron energy.

The propagation of the electrons from the aperture plane to the image plane can be described by Huygens' principle, according to which each point in the diffraction pattern emits wavelets. Those wavelets that pass through the objective aperture recombine at the image plane, after being phase shifted by the lens aberrations. In the general case, the phase shift introduced by these aberrations depends not only on the initial direction of the electron trajectory but also on the object plane coordinates[9]. If we assume that the imaging system is isoplanatic (i.e., it produces an image disk of equal shape for whatever the object point location), the resulting spatial invariance of the aberration function greatly simplifies the analysis. Isoplanacy is violated if such aberrations as distortion, third-order astigmatism, or coma are significant. In practice the only resolution-limiting aberrations originate from the objective lens, whose aberration function describes the deviation of the wave front from the Gaussian reference sphere and

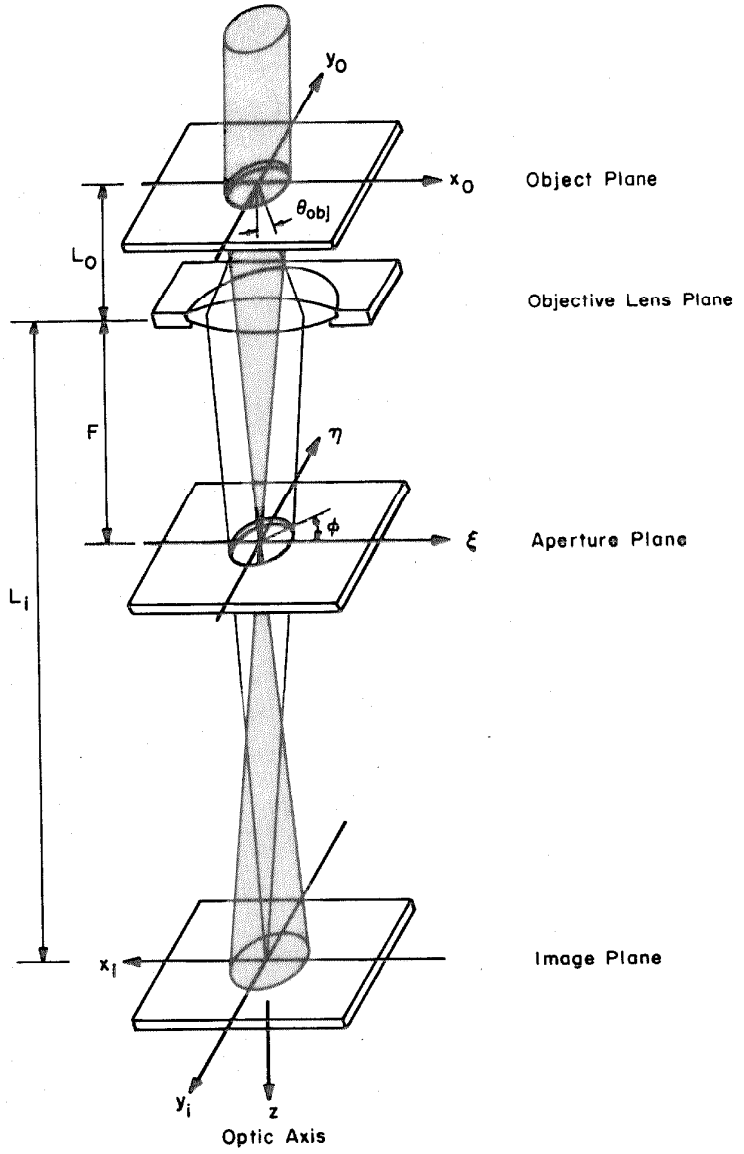


Fig. 1-2. Perspective drawing of the coordinate systems used in this work

is expressed as[10]:

$$\gamma(\theta,\phi) = \frac{\pi}{2\lambda} C_s \theta^4 - \frac{\pi}{\lambda} [\Delta Z + \frac{\Delta Z_a}{2} \sin 2(\phi-\phi_o)]\theta^2 \qquad (1.2)$$

with the following list of definitions:

θ = scattering angle

ϕ = azimuthal angle at the back focal plane

C_s = spherical aberration constant of the objective lens (in mm)

ΔZ = defocus value between the object and the conjugate object plane (in Å)

ΔZ_a = focus difference of the axial astigmatism (in Å)

ϕ_o = reference azimuthal angle of the axial astigmatism

By inspecting Fig. 1-2, we see that the coordinates (ξ,η) are linked to the polar coordinates (θ,ϕ) by the relationships

$$\begin{aligned} \xi &= L_o \tan\theta \cos\phi \simeq L_o\theta \cos\phi \\ \eta &= L_o \tan\theta \sin\phi \simeq L_o\theta \sin\phi \end{aligned} \qquad (1.3)$$

We may thus write the kinematically scattered wave in the Fraunhofer approximation as

$$\psi_s(x_i,y_i) = \frac{1}{L_i\lambda} \iint_{B_{obj}} \Psi(\xi,\eta) \exp[-i\gamma(\theta,\phi)-2\pi i K(\xi x_i+\eta y_i)/L_i] d\xi\, d\eta \qquad (1.4)$$

where B_{obj} is the open area of the objective aperture.

In order to simplify the notation, we will assume that x_i and y_i have been scaled to object-plane coordinates by dividing them through by the magnification M, and inverting their sign in order to

account for the image reversal by the lens system. If we replace $\Psi(\xi,\eta)$ in Eq. (1.4) by the analytical expression given in Eq. (1.1), we obtain for the image wave amplitude[11,12,13]

$$\psi_s(x_i,y_i) = \frac{iA_o}{ML_o^2\lambda} \int_{B_{obj}} \int \{ \sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) - i\gamma_j(\theta,\phi) + 2\pi iK(\xi(x_i - x_o^j) + \eta(y_i - y_o^j))/L_o] \, d\xi \, d\eta \qquad (1.5)$$

where L_i has been replaced by ML_o, and $\gamma_j(\theta,\phi)$ denotes the aberration phase shift of the wave scattered by the j^{th} atom. Since the displacement of the j^{th} atom from the conjugate object plane is $(\Delta Z + z_o^j)$, the phase term $\gamma_j(\theta,\phi)$ contains the z coordinate of this atom and may be expressed as

$$\gamma_j(\theta,\phi) = \gamma(\theta,\phi) \exp[-i\pi z_o^j \theta^2/\lambda] \qquad (1.6)$$

Let us now introduce the reciprocal coordinate system defined by

$$k_x = \frac{\theta}{\lambda} \cos\phi \simeq \xi/\lambda L_o$$
$$k_y = \frac{\theta}{\lambda} \sin\phi \simeq \eta/\lambda L_o \qquad (1.7)$$

where $\underset{\sim}{k} = (k_x,k_y)$ is a Fourier transform variable. By changing the variables (ξ,η) into (k_x,k_y) in Eq. (1.5), the elastically scattered wave amplitude at the image plane is given by

$$\psi_s(\underset{\sim}{r}_i) = \frac{iA_o\lambda}{M} \int_{B_{obj}} \{ \sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) + i\pi z_o^j \theta^2/\lambda - 2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j] \} \cdot \exp[-i\gamma(\underset{\sim}{k}) + 2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_i] \, d\underset{\sim}{k} \qquad (1.8)$$

where $\underset{\sim}{r}_i = (x_i, y_i)$, $\underset{\sim}{r}_o^j = (x_o^j, y_o^j)$, and $\gamma(\underset{\sim}{k})$ stands for $\gamma(\theta,\phi)$.

An equivalent expression for the scattered wave, $\psi_s(\underset{\sim}{r}_i)$, is found by using the Fourier transform operator, $\mathcal{F}$, and by defining an objective lens pupil function, $P(\underset{\sim}{k})$, as

$$P(\underset{\sim}{k}) = b(\underset{\sim}{k})\, e^{-i\gamma(\underset{\sim}{k})} \tag{1.9}$$

where $b(\underset{\sim}{k})$ is the objective aperture function with

$$b(\underset{\sim}{k}) = \begin{cases} 1 & \text{for } \underset{\sim}{k} \in B_{obj} \\ 0 & \text{elsewhere} \end{cases}$$

The scattered wave can then be written as

$$\psi_s(\underset{\sim}{r}_i) = \frac{iA_o\lambda}{M}\, \mathcal{F}^{-1}\, \{F_f(k) \cdot P(\underset{\sim}{k})\} \tag{1.10}$$

where $F_f(\underset{\sim}{k})$ is called the object Fresnel transform and is defined by[12]

$$F_f(\underset{\sim}{k}) = \sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) + i\pi z_o^j \theta^2/\lambda - 2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j] \tag{1.11}$$

We shall now investigate under what conditions an object may be approximated by a planar arrangement of atoms, since this permits an estimate of $\psi_s(\underset{\sim}{r}_i)$ without a detailed knowledge of the atomic arrangement of the object. Let us introduce $\tilde{F}(k_x,k_y,k_z)$ as being the Fourier transform, or "structure factor", of an aperiodic array of atoms. Then, by analogy with its counterpart in periodic structures, we can define this structure factor as[14]

$$\tilde{F}(k_x,k_y,k_z) = \sum_{j=1}^{N_a} |f_j(\theta)|\, e^{i\eta_j(\theta)} \exp[-2\pi i(k_x x_o^j + k_y y_o^j + k_z z_o^j)] \qquad (1.12)$$

Now we know that the wave vector of an elastically scattered wave lies on the Ewald sphere. In Fig. 1-3 we illustrate the generation of this dispersion surface by a wave vector which satisfies the conditions (cf. Eq. 1.7):

$$k_z = \frac{1}{\lambda}(\cos\theta - 1) \simeq -\frac{\theta^2}{2\lambda} = -\frac{\lambda}{2}(k_x^2 + k_y^2) \qquad (1.13)$$

where the paraxial approximation has been employed. Then, by replacing k_z in Eq. (1.12) by its value given by Eq. (1.13), we see that we retrieve $F_f(\underset{\sim}{k})$ from $\tilde{F}(k_x,k_y,k_z)$. Therefore, for small angles θ, the three-dimensional object transform, when sampled by the Ewald sphere yields $F_f(\underset{\sim}{k})$. Thus, the Fresnel transform can be thought of as the projection of an array of atoms onto a curved surface.

For thin specimens we may neglect the curvature of the Ewald sphere; hence the Fresnel transform becomes the "structure factor", $F(\underset{\sim}{k})$ of a planar arrangement of atoms:

$$F(\underset{\sim}{k}) = \sum_{j=1}^{N_a} |f_j(\theta)|\, e^{i\eta_j(\theta)} \exp(-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j) \qquad (1.14)$$

where $F(\underset{\sim}{k})$ is derived from the three-dimensional array of atoms projected along the k_z axis. The above approximation is valid as long as the thickness dependence phase term, $\exp(i\pi z_o^j \theta^2/\lambda)$, remains negligible[15]. If we call t_o the specimen thickness, this approximation requires that

$$\theta^2 t_o/\lambda \ll 1 \qquad (1.15)$$

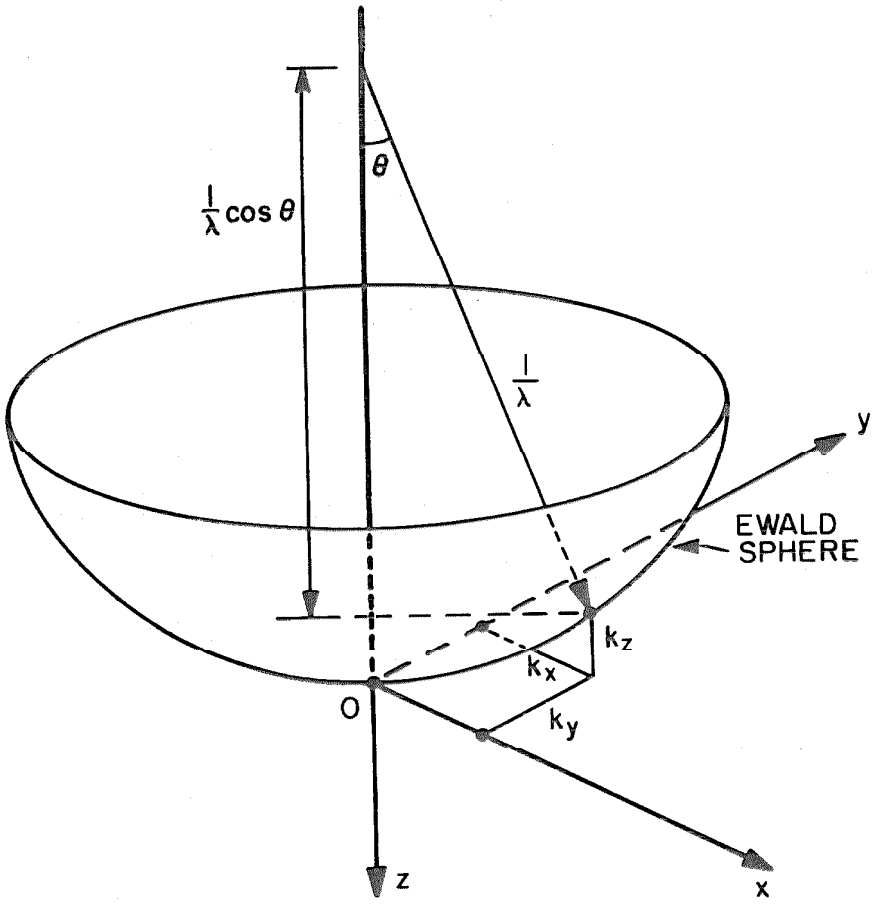


Fig. 1-3. Diagram of the Ewald sphere and its associated coordinate system

Let us introduce the diffraction-limited instrument resolution, ρ_s, given by

$$\rho_s = \lambda/\theta_{obj} \tag{1.16}$$

Then inequality (1.15) may be written as

$$t_o << \rho_s^2/\lambda = t_M \tag{1.17}$$

where t_M is a threshold value for the specimen thickness, beyond which one cannot ignore the z dependence in the wave function. For example, at a resolution of $\rho_s = 3\mathring{A}$, and for 100 keV electrons, a planar representation of the object breaks down for thickness values greater than: $t_M = 240\mathring{A}$.

We shall now derive an analytical expression for the wave scattered by a single atom, in order to study the various contrast mechanisms that govern the formation of an atomic image. For a circularly symmetric array of atoms, and when the azimuthal dependence of $\gamma(k)$ due to astigmatism can be ignored, the amplitude of the scattered wave given by Eq. (1.8) simplifies to

$$\psi_s(\rho_i) = \frac{i2\pi A_o}{M\lambda} \int_0^{\theta_{obj}} \{ \sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) + i\pi z_o^j \theta^2/\lambda - i\gamma(\theta)]\} J_o(2\pi\theta\rho_i^j/\lambda)\theta \, d\theta \tag{1.18}$$

where $\rho_i = |\underset{\sim}{r}_i|$, $\rho_i^j = |\underset{\sim}{r}_i - \underset{\sim}{r}_o^j|$, and J_o is the zero-order Bessel function. In particular, the amplitude distribution of the wave scattered by a single atom is given by

$$\psi_{s1}(\rho_i) = \frac{i2\pi A_o}{M\lambda} \int_0^{\theta_{obj}} |f(\theta)| \exp[i\eta(\theta) - i\gamma(\theta)] J_o(2\pi\theta\rho_i/\lambda)\theta \, d\theta \tag{1.19}$$

From Eq. (1.19), we see that the image intensity, $|\psi_{s1}|^2$, is just an Airy disc modulated by the scattering factor term. Consequently, the image of an assemblage of atoms consists of overlapping Airy discs.

1.2.2 Imaging of a Single Atom

1.2.2.1 Contrast mechanisms

Two mechanisms product contrast during the formation of a single atom image. In the first, the contrast is due to the removal of elastically scattered electrons from the optical system by an objective aperture; this type of contrast is called diffraction or scattering contrast. This form of contrast can also be obtained by allowing one diffracted beam to pass down the optical axis of the microscope, and this is referred to as the dark field mode of imaging. In the second mechanism, several diffracted beams, including the transmitted beam, are allowed to reach the image plane; the mutual interference of these beams causes phase contrast[15]. We shall now evaluate the relative importance of both contrast mechanisms as a function of θ_{obj} by linking the various contrast factors to their corresponding scattering cross sections[16].

In the bright-field mode, the kinematically scattered wave recombines with the primary wave, $\psi_o = A_o/M$; hence, the bright-field intensity can be expressed as

$$I_B(\rho_i) = |\psi_o + \psi_{s1}(\rho_i)|^2 \tag{1.20}$$

where $\psi_{s1}(\rho_i)$ has been derived in Eq. (1.19). If we now expand $I_B(\rho_i)$ and collect the various factors of the expansion, we can write $I_B(\rho_i)$ as a sum of four terms

$$I_B(\rho_i) = I_o + I_D(\rho_i) + I_A(\rho_i) + I_P(\rho_i) \tag{1.21}$$

where $I_o = |\psi_o|^2$ is the transmitted beam intensity, $I_D(\rho_i) = |\psi_{sl}(\rho_i)|^2$ is the dark-field intensity, $I_A(\rho_i)$ is given by

$$I_A(\rho_i) = -\frac{4\pi A_o^2}{\lambda M^2} \int_0^{\theta_{obj}} |f(\theta)| \sin\eta(\theta)\, J_o(2\pi\theta\rho_i/\lambda)\cos\gamma(\theta)\theta\, d\theta \tag{1.22}$$

and $I_P(\rho_i)$ by

$$I_P(\rho_i) = \frac{4\pi A_o^2}{\lambda M^2} \int_0^{\theta_{obj}} |f(\theta)| \cos\eta(\theta)\, J_o(2\pi\theta\rho_i/\lambda)\sin\gamma(\theta)\theta\, d\theta \tag{1.23}$$

Let us now examine those factors that govern phase and amplitude contrast in the imaging process. If a very small objective aperture half angle is selected, we expect that the phase contrast contribution will vanish. Consequently, in the limit where θ_{obj} tends toward zero, we can make the approximations:

$$\gamma(\theta) \simeq \gamma(0) = 0$$

$$|f(\theta)| \sin\eta(\theta) \simeq |f(0)| \sin\eta(0) = \mathrm{Im}[f(0)] \tag{1.24}$$

where the validity of these approximations relies on the fact that the functions $\gamma(\theta)$, $|f(\theta)|$ and $\eta(\theta)$ have zero slopes at the origin[17]. Equations (1.22) and (1.23) then yield

$$I_A(\rho_i) \simeq -\frac{2\theta_{obj}}{\rho_i} I_o\, \mathrm{Im}[f(0)] J_1(2\pi\theta_{obj}\rho_i/\lambda) \tag{1.25}$$

and

$$I_P(\rho_i) \simeq 0 \tag{1.26}$$

Therefore, the term $I_P(\rho_i)$, which does not contribute to the image for small objective apertures, must be responsible for phase contrast, while the amplitude term $I_A(\rho_i)$ contributes to the scattering contrast. We conclude that both phase and scattering contrasts are accounted for, when we use complex atomic scattering factors[15].

Finally, it remains to be shown that the nonlinear dark-field term can usually be neglected in comparison to $I_A(\rho_i)$. If this were true, we could then establish a linear relationship between the recorded bright-field intensity and the scattering properties of an atom. By summing the dark-field term over the image plane we obtain the total elastic scattering cross-section for scattering inside the objective aperture:

$$2\pi \int_0^\infty I_D(\rho_i)\rho_i d\rho_i = 2\pi I_o \int_0^{\theta_{obj}} |f(\theta)|^2 \theta \, d\theta \simeq I_o \sigma_{obj} \tag{1.27}$$

In a similar fashion, if we integrate the scattering term $I_A(\rho_i)$, approximated for small θ_{obj} in Eq. (1.25), we obtain

$$2\pi \int_0^\infty I_A(\rho_i)\rho_i \, d\rho_i = -I_o \, 2\lambda \, \mathrm{Im}(f(0)) \tag{1.28}$$

By applying the optical theorem[18] to Eq. (1.28) we can express this result in terms of the total cross section for elastic scattering, σ_{el}:

$$-I_o \, \sigma_{el} = -I_o \, 2\lambda \, \mathrm{Im}(f(0)) \tag{1.29}$$

where

$$\sigma_{el} = 2\pi \int_0^\infty |f(\theta)|^2 \sin\theta \, d\theta$$

Scattering contrast is therefore brought about by two terms, $I_A(\rho_i)$ and $I_D(\rho_i)$, whose integral over the image plane

$$2\pi \int_0^\infty [I_A(\rho_i) + I_D(\rho_i)]\rho_i \, d\rho_i = -I_o(\sigma_{el} - \sigma_{obj}) \qquad (1.30)$$

yields the cross section for scattering outside the objective aperture.

1.2.2.2 Relative magnitude of phase and scattering contrast

In this section we shall consider the variations in phase and amplitude contrast that occur as we gradually open up the objective aperture, that is as we approach the high-resolution imaging range. We shall show that at the atomic resolution level the contrast of an atom reaches a maximum at specific electron microscope settings[16].

For small θ_{obj}, the phase contrast contribution to the bright-field image of an atom vanishes. At this medium resolution range, where an atom is not resolved, let us compare the contributions to scattering contrast from $I_A(\rho_i)$ and $I_D(\rho_i)$. Now since θ_{obj} is small, we can neglect the aberration term in the scattered wave function and make the approximation that $f(\theta) \simeq f(0)$. In these approximations Eq. (1.19) reduces to

$$\psi_{s1}(\rho_i) = \psi_{s1}(0) \, \frac{2J_1(2\pi \, \theta_{obj} \, \rho_i/\lambda)}{2\pi \, \theta_{obj} \, \rho_i/\lambda} \qquad (1.31)$$

We see from Eq. (1.31) that the image of an atom is an Airy pattern.

Equation (1.22) can now be written as

$$I_A(\rho_i) = - \pi I_o (\frac{\theta_{obj}}{\lambda})^2 \sigma_{el} (\frac{\psi_{sl}(\rho_i)}{\psi_{sl}(0)}) \tag{1.32}$$

and the dark-field term can similarly be represented as

$$I_D(\rho_i) = \pi I_o (\frac{\theta_{obj}}{\lambda})^2 <\sigma_{\theta_{obj}}> (\frac{\psi_{sl}(\rho_i)}{\psi_{sl}(0)}) \tag{1.33}$$

where $<\sigma_{\theta_{obj}}> = \pi\theta_{obj}^2 |f(0)|^2$ is an estimate for $\sigma_{\theta_{obj}}$ at small θ_{obj}. The ratio I_D/I_A has therefore the same order of magnitude as $\sigma_{\theta_{obj}}/\sigma_{el}$. Since the ratio of these cross-sections is much less than unity at medium resolution, the dark-field term, which is usually neglected in contrast calculations, does not alter the image significantly.

At high resolutions phase contrast becomes the dominant contrast mechanism in bright-field imaging. In order to estimate the relative magnitude of the different contrast terms, we replace them by their upper bounds (e.g., $\sin \gamma(\theta) = 1$, $\cos \gamma(\theta) = 1$, and $f(\theta) = f(0)$). Then in these approximations, we can convert Eqs. (1.22), (1.23), and (1.33) into

$$I_A^{max} \simeq - \frac{2\pi}{\lambda} \theta_{obj}^2 \, Im(f(0)) \cdot I_o \tag{1.34a}$$

$$I_P^{max} \simeq \frac{2\pi}{\lambda} \theta_{obj}^2 \, Re(f(0)) \cdot I_o \tag{1.34b}$$

$$I_D^{max} \simeq (\frac{\pi\theta_{obj}^2}{\lambda})^2 \, |f(0)|^2 \cdot I_o \tag{1.34c}$$

where all contrast terms have been calculated at the origin $\rho_i = 0$, which corresponds to their peak value.

The ratio of scattering contrast versus phase contrast is then approximately equal to

$$|I_A^{max}|/I_P^{max} \simeq \mathrm{Im}(f(0))/\mathrm{Re}(f(0)) \simeq \eta(0) \qquad (1.35)$$

where $\eta(0)$ is the anomalous phase-shift angle at the origin. Calculations of $\eta(0)$ have led to a linear dependence of this quantity with Z, the atomic number[8]. For example, $\eta(0) = 0.02$ for a carbon atom, whereas $\eta(0) = 0.26$ for a gold atom. The anomalous phase shift is therefore significant for a specimen which is composed of heavy atoms. This drastic dependence of the scattering contrast upon Z can, in fact, be used to discriminate between atoms of different atomic numbers (see Chapter Three).

We shall now compare the magnitude of the dark-field term to the magnitude of the phase contrast term at high resolution. An estimate of their ratio can be found from Eq. (1.34), which yields

$$I_D^{max}/I_P^{max} \simeq \frac{\pi\theta_{obj}^2}{2\lambda}\,|f(0)| \qquad (1.36)$$

We notice that the contribution from $I_D(\rho_i)$ increases with Z ($\because |f(0)| \simeq Z^{1/3}$) [6], and with θ_{obj}. For a microscope operating at 100 keV with a resolution of $\rho_s = 2\text{Å}$, the ratio I_D^{max}/I_P^{max} takes values of 0.065 for a carbon atom and 0.15 for a gold atom.

Currently, the resolving power of an electron microscope does not exceed 2Å; so that phase contrast is the main contrast mechanism operating during the bright-field imaging of an atom. By combining Eqs. (1.21) and (1.23), we can approximate the overall atomic contrast,

$C_a(\rho_i)$, to

$$C_a(\rho_i) = \frac{I_B - I_o}{I_o} \simeq \frac{4\pi}{\lambda} \int_0^{\theta_{obj}} |f(\theta)| \cos \eta(\theta)\, J_o(2\pi\theta\rho_i/\lambda) \sin \gamma(\theta)\theta \, d\theta \tag{1.37}$$

The problem of finding the optimum microscope conditions for imaging a single atom reduces to one of maximizing $C_a(\rho_i)$. This optimization problem was first solved by Scherzer[10], who maximized the area under the curve $\sin \gamma(\theta)$ versus θ, which is called the phase-contrast transfer function (PCTF). The optimal choices for the objective aperture and for the focus are found to be

$$\theta_{obj}^{opt} = 1.41 \sqrt[4]{\lambda/C_s}$$

$$\Delta Z^{opt} = \sqrt{1.41\, \lambda C_s} \tag{1.38}$$

1.2.3 Imaging of Atom Clusters

The total scattered wave for an assembly of atoms is obtained by adding, with regard to phase, the individual amplitudes arising at that image point from each component atom[19] (cf. Eq. 1.8). Extensive contrast calculations have been carried out, at various electron optical settings, for simple atomic configurations such as linear chains and small clusters[11,15]. Such a calculation does not give an analytical expression for the image contrast in terms of the object structure and the optical parameters. It only yields a single numerical solution for a given set of experimental conditions. Therefore this procedure is impractical for studying the more complex atomic arrangements

usually found in real specimens. In the next section the influence of lens imperfections on the transfer of information from the object to the image plane is analyzed in a wave optical formalism.

1.3 Wave Theory of Image Formation

1.3.1 Object Wave Function

1.3.1.1 Introduction

For potential fields and crystal thicknesses for which the intensities of the diffracted electron beams are negligible compared with the incident beam, Cowley and Moody[5] have developed a wave optical approach, called the phase grating approximation, to describe the scattering of electrons. In this model the phase and amplitude changes of the wave function are determined for thin crystal slices perpendicular to the electron beam, with Fresnel diffraction taking place between these slices. If our specimen can be approximated by a single slice, this wave-optical formalism describes the scattering properties of the object by a transmission function. It can then be shown that the resulting image contrast can be explained by the same contrast mechanisms that were introduced in describing the image formation of an atom. In the wave optical approximation, phase contrast is produced by interference between the transmitted beam and waves that are elastically scattered within the objective aperture. Amplitude contrast arises from elastic and inelastic scattering out of the objective aperture.

1.3.1.2 Transmission function

In the small angle approximation, the progress of an electron wave through a thin slice may be represented by a convolution product

of the incident wave function with a propagation function, followed by a multiplication by the transmission function of the slice. Fast electrons propagate in spherical waves, which may be approximated by paraboidal waves, so that the phase change due to the propagation is given by $\exp[-i\pi\rho_o^2/\lambda t_s]$, where ρ_o is the lateral dimension of an object detail and t_s is the slice thickness.

We shall now consider when we can neglect Fresnel diffraction, as the electron wave traverses the slice. An equivalent way of approaching the problem consists in evaluating the spread of the wave across the slice. We may view the specimen as two-dimensional if this spread is less than the minimum resolvable detail, ρ_s, that is[20]

$$\theta_{obj}\, t_s << \rho_s \;\Rightarrow\; t_s << \frac{\rho_s^2}{\lambda} = t_M \qquad (1.39)$$

where ρ_s is the diffraction-limited resolution. By comparing Eq. (1.39) with Eq. (1.17) we see that we retrieve the same upper bound for the specimen thickness as we obtained for neglecting the curvature of the Ewald sphere.

If we neglect the effects of Fresnel diffraction, we can describe the diffraction from a thin specimen using a single slice. We shall now relate the scattering from this slice to the concept of complex transmittance; so that we may draw analogies with known results from the field of light optics. If we let U be the electron accelerating potential and $V(x_o,y_o,z_o)$ be the potential field describing the distribution of charged particles in the specimen, then the wavelength of the electron in the specimen is

$$\lambda(x_o,y_o,z_o) = h/\{2me[U + V(x_o,y_o,z_o)]\}^{1/2}$$
$$\simeq \lambda_o[1 + V(x_o,y_o,z_o)/U] \tag{1.40}$$

where $\lambda(x_o,y_o,z_o)$ is the electron wavelength at the point (x_o,y_o,z_o) in the object, and $\lambda_o = h/(2meU)^{1/2}$ is the vacuum electron wavelength. After traversing a specimen of thickness t_o, the phase of the electron wave will be changed relative to a wave in the absence of a field by an amount

$$\phi(x_o,y_o) = 2\pi \int_0^{t_o} \left(\frac{1}{\lambda(x_o,y_o,z_o)} - \frac{1}{\lambda_o}\right) dz_o \simeq -\frac{2\pi\lambda_o me}{h^2} \underbrace{\int_0^{t_o} V(x_o,y_o,z_o)dz_o}_{= V_{mean} \cdot t_o} \tag{1.41}$$

where $\phi(\underset{\sim}{r}_o)$ is the relative phase shift, and V_{mean} is the mean inner potential. Consequently, we can conclude that the phase delay is proportional to the projection in the beam direction of the potential distribution.

The preceding analysis allows us to describe the interaction of the incident beam with a thin specimen by the object transmission function[5] (Ob.TF)

$$\Lambda_t(\underset{\sim}{r}_o) = \exp[i\phi(\underset{\sim}{r}_o) - \mu(\underset{\sim}{r}_o)] \tag{1.42}$$

where $\mu(\underset{\sim}{r}_o)$ is an absorption function. This absorption term describes the decrease in the intensity of the transmitted beam, as a function of the number density of atoms in the specimen, projected onto the $\underset{\sim}{r}_o$ plane. The absorption of electrons by the object can be related to scattering contrast. In our discussion of the formation of the image

of a single atom, we showed that, in addition to the dark field term, anomalous phase shift generated scattering contrast. Absorption can then be described by the combined effects of the anomalous phase shift, the removal of electrons by the objective aperture, and the beam attenuation due to inelastic scattering events.

If we assume that only small phase and amplitude changes in the electron beam are induced by the specimen, then the transmission function given by Eq. (1.42) may be written as[21,22]

$$\Lambda_t(\underset{\sim}{r}_o) \simeq 1 + i\phi(\underset{\sim}{r}_o) - \mu(\underset{\sim}{r}_o) \qquad \text{for} \quad |\mu(\underset{\sim}{r}_o)| < |\phi(\underset{\sim}{r}_o)| \ll 1 \tag{1.43}$$

where only the first-order terms of the exponential expansion have been retained. For example, this approximation remains valid as long as $|\phi(\underset{\sim}{r}_o)| \leq 0.2$; if we replace $\phi(\underset{\sim}{r}_o)$ by its expression in Eq. (1.41), then this inequality becomes

$$t_o \leq 0.2 \frac{h^2}{2\lambda m e V_{mean}} = \tilde{t}_M \tag{1.44}$$

For 100 keV electrons, the upper bound is equal to $\tilde{t}_M^C \simeq 260\text{Å}$ for an amorphous specimen composed of carbon atoms ($V_{mean} \simeq 2$ eV), while for gold atoms ($V_{mean} \simeq 20$ eV) this bound decreases down to $\tilde{t}_M^{Au} \simeq 26\text{Å}$. The specimens which we have just described are considered to be weak phase and amplitude objects. In Chapter Three we will develop the formalism associated with the contrast analysis of such objects.

Equation (1.43) will not be valid, due to neglecting higher-order terms in both the phase and absorption functions, when the specimen

is thicker than $\sim 100\text{Å}$ [22,23]. When the specimen thickness is a fraction of the mean-free path for elastic scattering, the phase shifts are small and plural scattering events do not occur (cf. 3.2.3.2). Then the higher order terms of Eq. (1.43), which describe the effects of plural scattering, can be neglected.

1.3.1.3 Relationship between transmission function and scattering factor

We will now present an interpretation of the wave function in terms of the scattering properties of the constituent atoms of the specimen. For the sake of simplicity we shall assume that the specimen is composed of similar atoms, and that it obeys the weak phase-object approximation. By neglecting the absorption term, $\mu(\underset{\sim}{r}_o)$, the specimen transmission function reduces to

$$\Lambda_{ph}(\underset{\sim}{r}_o) \simeq 1 + i\phi(\underset{\sim}{r}_o) \tag{1.45}$$

The phase term can then be interpreted either as the phase shift suffered by the electron in passing through the specimen, or as the amplitude of the scattered wave.

Since we can view the specimen as a single slice, the total electron wave at its exit surface equals $A_o \Lambda_{ph}(\underset{\sim}{r}_o)$ where A_o is a normalization constant equal to the amplitude of the incident beam. The scattered electron wave $\Psi(\underset{\sim}{k})$ at the back focal plane can be expressed as

$$\Psi(\underset{\sim}{k}) = \frac{iA_o}{\lambda L_o} \Phi(\underset{\sim}{k}) \tag{1.46}$$

where $\Phi(k)$ is the transform of the object phase distribution. This

phase transform is calculated from $\phi(\underset{\sim}{r}_o)$ by

$$\Phi(\underset{\sim}{k}) \simeq \int_{A_{obt}} \phi(\underset{\sim}{r}_o) \exp(-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o)\, d\underset{\sim}{r}_o \tag{1.47}$$

where the integration is performed over the finite area of the object, A_{obt}.

According to Eqs. (1.1) and (1.14), the kinematically scattered wave at the diffraction plane, produced by an assembly of like atoms, can also be written as

$$\Psi(\underset{\sim}{k}) = \frac{iA_o}{L_o} f'(\theta) \sum_{j=1}^{N_a} \exp(-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j) = \frac{iA_o}{L_o} F_r(\underset{\sim}{k}) \tag{1.48}$$

where $f'(\theta)$ is the real atomic scattering amplitude, and $F_r(\underset{\sim}{k})$ is the corresponding structure factor. Then, by comparing the two equivalent relations for $\Psi(\underset{\sim}{k})$ in Eqs. (1.46) and (1.48), we are led to the identity:

$$\Phi(\underset{\sim}{k}) = \lambda F_r(\underset{\sim}{k}) = \lambda f'(\theta) \sum_{j=1}^{N_a} \exp(-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j) \tag{1.49}$$

Now, in order to relate the phase distribution $\phi(\underset{\sim}{r}_o)$ to the atomic arrangement within the specimen, we shall introduce the object density function $T(\underset{\sim}{r}_o)$, which is defined as the density of atoms projected along the incident beam direction. This density function is expressed in atoms per unit area and specified by the relationship:

$$\sum_{j=1}^{N_a} \exp(-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j) = \int_{A_{obt}} T(\underset{\sim}{r}_o) \exp(-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o) d\underset{\sim}{r}_o \simeq \mathcal{F}\{T(\underset{\sim}{r}_o)\} \tag{1.50}$$

Thus, by combining Eqs. (1.49) and (1.50), we obtain

$$\Phi(\underset{\sim}{k}) = \lambda f'(\theta)\ \mathcal{F}\{T(\underset{\sim}{r}_o)\} \tag{1.51}$$

or equivalently,

$$\phi(\underset{\sim}{r}_o) = \lambda\ T(\underset{\sim}{r}_o) * \mathcal{F}^{-1}\{f'(\theta)\} \tag{1.52}$$

The phase distribution is therefore proportional to a convolution product of the object density function with the transform of the atomic scattering amplitude. If the atoms in the object are considered to be point scatterers, then $f'(\theta)$ is a constant and $\phi(\underset{\sim}{r}_o)$ is directly proportional to the object density function.

1.3.2 Amplitude Transfer Function

1.3.2.1 Definition

In Sec. 1.2.1 we have described the propagation of the electrons from the object to the image plane by means of a pupil function, which modifies the object spectrum. This analysis relies upon the assumption that the illuminating beam is coherent, and that the distortions of the wave front by lens aberrations are consistent with the isoplanatic condition.

In the general case, the wave function in the image plane is related to the object wave function by a linear integral transform. The corresponding Green function of this transform describes the imaging properties of the system, and is therefore called the impulse response of the system. If $\underset{\sim}{K}$ denotes the wave vector of the incident electron beam, then the Green function, $G(\underset{\sim}{K},\underset{\sim}{r}_o,\underset{\sim}{r}_i)$, depends on the modulus of the wave vector if the illumination is not monochromatic, and on (K_x,K_y) if the electron wave vectors have an angular

distribution. Let us now limit our discussion to an isoplanatic system, for which the Green function is spatially invariant[6].

$$G(\underset{\sim}{K},\underset{\sim}{r}_o,\underset{\sim}{r}_i) = G(\underset{\sim}{K},\underset{\sim}{r}_i - \underset{\sim}{r}_o) \tag{1.53}$$

We shall now consider a perfectly coherent imaging system, in order to define the amplitude transmission function (ATF).

We showed in Sec. 1.2.1 that the imaging process could be visualized as a sequence of two successive Fourier transformations. The amplitude distribution of the wave at the diffraction plane is the Fourier transform of the wave function, $\phi_{ob}(\underset{\sim}{r}_o)$, at the exit surface of the specimen; hence

$$\Psi_{tot}(\underset{\sim}{k}) = \frac{1}{L_o\lambda}\,\mathcal{F}\{\psi_{ob}(\underset{\sim}{r}_o)\} = \frac{1}{L_o\lambda}\,S_o(\underset{\sim}{k}) \tag{1.54}$$

where $\Psi_{tot}(\underset{\sim}{k})$ is the wave at the back focal plane, and $S_o(\underset{\sim}{k})$ is the object spectrum. The image wave is then the inverse transform of $\Psi_{tot}(\underset{\sim}{k})$, which has been modified by the pupil function, so that we can write

$$\psi(\underset{\sim}{r}_i) = \frac{L_o\lambda}{M}\int \Psi_{tot}(\underset{\sim}{k})\,P(\underset{\sim}{k})\,\exp(2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)\,d\underset{\sim}{k} \tag{1.55}$$

where $\psi(\underset{\sim}{r}_i)$ is the total image wave function. If we insert Eq. (1.54) into (1.55), we then obtain

$$\psi(\underset{\sim}{r}_i) = \int_{A_{obt}} \psi_{ob}(\underset{\sim}{r}_o)d\underset{\sim}{r}_o\left\{\frac{1}{M}\int P(\underset{\sim}{k})\,\exp[2\pi i\underset{\sim}{k}\cdot(\underset{\sim}{r}_i-\underset{\sim}{r}_o)]d\underset{\sim}{k}\right\} \tag{1.56}$$

It then follows from Eq. (1.56) that the system impulse response can be described by

$$G(\underset{\sim}{r}_i - \underset{\sim}{r}_o) = \frac{1}{M} \int P(\underset{\sim}{k}) \exp[2\pi i \underset{\sim}{k} \cdot (\underset{\sim}{r}_i - \underset{\sim}{r}_o)] \, d\underset{\sim}{k} \tag{1.57}$$

It is now useful to introduce the concept of the ATF, $T(\underset{\sim}{k})$, which is defined as the Fourier transform of $G(\underset{\sim}{r}_i,0)$. The convolution theorem allows us to simplify Eq. (1.56) by expressing its transform as

$$S_i(\underset{\sim}{k}) = S_o(\underset{\sim}{k}) \, T(\underset{\sim}{k}) \tag{1.58}$$

where $S_i(\underset{\sim}{k})$ is the transform of $\psi(\underset{\sim}{r}_i)$, or equivalently the image spectrum and

$$T(\underset{\sim}{k}) = \mathcal{F}\{G(\underset{\sim}{r}_i,0)\} = \frac{1}{M} P(\underset{\sim}{k}) = \frac{1}{M} b(\underset{\sim}{k}) \, e^{-i\gamma(\underset{\sim}{k})} \tag{1.59}$$

We see that the imaging system acts as a linear spatially-invariant filter on the input data characterized by $T(\underset{\sim}{k})$.

Although the above analysis does not impose any conditions on the nature of the specimen, its practicality is limited for general objects. The reason for this limitation is that wave amplitudes are not observable quantities. However, we shall show in Chapter Three that a simple relationship exists between the object wave function and image intensity of weakly scattering objects. We shall also show that the coherence properties of the electron wave have a significant impact on the achievable resolution. Therefore, we shall give, in the next section, a brief summary of the effects of partial coherence on the imaging process.

1.3.2.2 Effect of partial coherence and chromatic aberration

A) Spatial incoherence

The angular spread of the electrons emitted from the gun can be

described as waves impinging upon the specimen at an angle to the optic axis. Since these electrons do not bear any phase relationships with each other, the image intensity is an incoherent superposition of elastically scattered waves with different wave vector components. In the first stage of the analysis, we shall assume that the incident beam is monochromatic, so that its incoherence can be described by an angular distribution. If we call $Q(\underset{\sim}{K}_o)$ the angular distribution of the beam on the specimen, where $\underset{\sim}{K}_o = (K_x, K_y)$, the image intensity can then be expressed as[24]

$$|\psi(\underset{\sim}{r}_i)|^2 = \int |\psi(\underset{\sim}{K}_o, \underset{\sim}{r}_i)|^2 \, Q(\underset{\sim}{K}_o) d\underset{\sim}{K}_o \tag{1.60}$$

where $\psi(\underset{\sim}{K}_o, \underset{\sim}{r}_i)$ is the image wave function corresponding to a given $\underset{\sim}{K}_o$.

Let us now examine what factors determine the angular spread of the electron beam in an electron microscope. Since electron microscopes use a double condenser lens system to focus the electron beam onto the specimen, the second-condenser aperture can be considered to be the exit pupil of the illumination system. We shall now show that this aperture is the effective source of illumination of the optical system.

We know, from the theory of partially coherent light, that an incoherent quasi-monochromatic,uniform, circular source gives rise in the exit pupil to a coherently illuminated area, whose radius is given by[9]

$$a_{coh} = 0.08\lambda/\theta_{ep} \tag{1.61}$$

where θ_{ep} is the half angle subtended by the pupil at the source. If

we apply the theorem to an electron microscope, with a second condenser aperture of 50 μm in radius that subtends a half angle of about 7.5×10^{-4} rad, we obtain $a_{coh} = 4\ \mu m \ll 50\ \mu m$. Since the coherently illuminated area is much smaller than the size of the aperture, the aperture acts as a distribution of independent sources. We can then consider the second condenser aperture to be incoherently illuminated.

Before proceeding with the analysis, we shall introduce a system of polar coordinates in the condenser aperture plane, similar to that used at the objective aperture plane. The polar coordinates in the condenser aperture plane are:

$$\theta_c = \lambda\sqrt{K_x^2 + K_y^2}$$

$$\phi_c = \tan^{-1}(K_y/K_x) \tag{1.62}$$

where θ_c is the angle that the incident beam makes with the specimen normal and ϕ_c is the azimuth angle.

Since the electron gun is a point source that emits electrons isotropically (i.e., without dependence on ϕ_c), the impulse response of a microscope is not a function of $\underset{\sim}{K}_o$ for monochromatic radiation If we call $\psi_{ob}(\underset{\sim}{K}_o, \underset{\sim}{r}_o)$ the electron wave function at the object-exit surface for a given wave vector orientation, the corresponding image wave can be written as

$$\psi(\underset{\sim}{K}_o, \underset{\sim}{r}_i) = \int_{A_{obt}} \psi_{ob}(\underset{\sim}{K}_o, \underset{\sim}{r}_o)\ G(\underset{\sim}{r}_i - \underset{\sim}{r}_o)\ d\underset{\sim}{r}_o \tag{1.63}$$

We will now restrict our analysis to thin specimens that obey the single slice model of the phase-grating theory. Now, provided that $\theta_c t_o < \rho_s$, the oblique illumination just produces a phase shift in the object wave, given by

$$\psi_{ob}(\underset{\sim}{K}_o,\underset{\sim}{r}_o) = A_o\, \Lambda_t(\underset{\sim}{r}_o)\, \exp(2\pi i\, \underset{\sim}{K}_o\cdot\underset{\sim}{r}_o) \tag{1.64}$$

By introducing the object spectrum, in this case defined by

$$S_o(\underset{\sim}{k}) = \mathcal{F}\{A_o\Lambda_t(\underset{\sim}{r}_o)\}$$

and substituting Eq. (1.64) into Eq. (1.63), the image wave can be written as

$$\psi(\underset{\sim}{K}_o,\underset{\sim}{r}_i) = \int_{A_{obt}} \exp(2\pi i\underset{\sim}{K}_o\cdot\underset{\sim}{r}_o)d\underset{\sim}{r}_o \left\{ \iint S_o(\underset{\sim}{k})\, \exp(2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_o)T(\underset{\sim}{k}') \right.$$
$$\left. \cdot\ \exp[2\pi i\underset{\sim}{k}'\cdot(\underset{\sim}{r}_i-\underset{\sim}{r}_o)]d\underset{\sim}{k}\ d\underset{\sim}{k}'\right\}$$

$$= \iint S_o(\underset{\sim}{k})\, T(\underset{\sim}{k}')\, \exp(2\pi i\underset{\sim}{k}'\cdot\underset{\sim}{r}_i)d\underset{\sim}{k}\ d\underset{\sim}{k}'$$

$$\cdot\left\{\int_{A_{obt}} \exp[2\pi i\underset{\sim}{r}_o\ (\underset{\sim}{K}_o+\ \underset{\sim}{k}-\underset{\sim}{k}')]\ d\underset{\sim}{r}_o\right\} \tag{1.65}$$

where the object wave and the system impulse response have been substituted by their transforms. Finally, we can simplify Eq. (1.65) into the form[24]

$$\psi(\underset{\sim}{K}_o,\underset{\sim}{r}_i) = \int S_o(\underset{\sim}{k}-\underset{\sim}{K}_o)\, T(\underset{\sim}{k})\, \exp(2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)\, d\underset{\sim}{k} \tag{1.66}$$

and Eq. (1.60) for the image intensity into the form

$$|\psi(\underset{\sim}{r}_i)|^2 = \iiint S_o(\underset{\sim}{k} - \underset{\sim}{K}_o) S_o^*(\underset{\sim}{k}' - \underset{\sim}{K}_o)\, T(\underset{\sim}{k})\, T^*(\underset{\sim}{k}')\, \exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i]$$

$$\cdot\, Q(\underset{\sim}{K}_o)\, d\underset{\sim}{k}\, d\underset{\sim}{k}'\, d\underset{\sim}{K}_o \qquad (1.67)$$

This integration over $\underset{\sim}{K}_o$ represents a convolution of the object power spectrum $|S_o(\underset{\sim}{k})|^2$ with $Q(\underset{\sim}{K}_o)$. If the second-condenser-aperture half angle is small, the microscope is very nearly spatially coherent; so that $Q(\underset{\sim}{K}_o) \simeq \delta(\underset{\sim}{K}_o)$, and the image intensity becomes a coherent superposition of electron waves expressed as

$$|\psi(\underset{\sim}{r}_i)|^2 = \left| \int S_o(\underset{\sim}{k})\, T(\underset{\sim}{k})\, \exp(2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i)\, d\underset{\sim}{k} \right|^2 \qquad (1.68)$$

B) Chromatic incoherence

In this section we shall describe how the distribution in electron beam energies affect image intensities for a spatially coherent illumination. The chromatic incoherence caused by the thermal energy spread of the source and by time-dependent fluctuations in accelerating voltage and objective lens current will also be examined.

i) Energy spread

Since electrons with different energies are assumed to be incoherent with respect to phase, the image intensity is then a superposition of monochromatic partial-wave intensities. The object wave function is independent of the energy spread for elastic scattering events, hence the image wave amplitude for a given electron energy can be written as

$$\psi(\delta U, \underset{\sim}{r}_i) = \int_{A_{obt}} \psi_{ob}(\underset{\sim}{r}_o)\, G(\delta U, \underset{\sim}{r}_i - \underset{\sim}{r}_o) d\underset{\sim}{r}_o \qquad (1.69)$$

where δU is the energy spread about a mean value of U. If we express this convolution integral in terms of the system ATF, then an equivalent expression for $\psi(\delta U, \underset{\sim}{r}_i)$ can be written:

$$\psi(\delta U, \underset{\sim}{r}_i) = \int S_0(\underset{\sim}{k})\, T(\delta U, \underset{\sim}{k})\, \exp(2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_i)\, d\underset{\sim}{k} \tag{1.70}$$

where the ATF, $T(\delta U, \underset{\sim}{k})$, now includes a chromatic aberration term. By calling C_c the third-order chromatic-aberration constant of the objective lens, we can represent $T(\delta U, \underset{\sim}{k})$ by (cf. Eqs. 1.2, 1.59)[9,24]:

$$T(\delta U, \underset{\sim}{k}) = \frac{1}{M}\, b(\underset{\sim}{k})\, \exp\left[-\frac{\pi i}{\lambda}\left(2C_s\, \theta^4 - \Delta Z \theta^2 + C_c \frac{\delta U \cdot \theta^2}{U}\right)\right] \tag{1.71}$$

where $\lambda(\delta U)$ has been approximated by a constant wavelength for the high-energy electrons, and the axial astigmatism has been neglected.

For a given electron-gun geometry, the energy distribution, $N_e(\delta U)$, of the incident beam can be determined experimentally[25]. The image intensity, which is a weighted incoherent superposition of monochromatic partial-wave intensities, can then be evaluated from Eq. (1.70):

$$|\psi(\underset{\sim}{r}_i)|^2 = \int_0^\infty N_e(\delta U) d(\delta U) \{ \iint S_0(\underset{\sim}{k}) S_0^*(\underset{\sim}{k}') T(\delta U, \underset{\sim}{k}) T^*(\delta U, \underset{\sim}{k}') \cdot \exp[2\pi i (\underset{\sim}{k} - \underset{\sim}{k}') \cdot \underset{\sim}{r}_i] d\underset{\sim}{k}\, d\underset{\sim}{k}' \tag{1.72}$$

In order to characterize the effects of chromatic aberration on the image, we extract from Eq. (1.72) the factors which depend on δU. We then consider the integral

$$\int_0^\infty N_e(\delta U)\ T(\delta U,\underset{\sim}{k})\ T^*(\delta U,\underset{\sim}{k}')\ d(\delta U)$$

which, by excluding terms independent of δU, yields the quantity

$$L(\underset{\sim}{k},\underset{\sim}{k}') \simeq \int_{-\infty}^{\infty} \exp[-\frac{\pi i}{\lambda} C_c \frac{\delta U}{U} (\theta^2-\theta'^2)]\ N_e(\delta U)\ d(\delta U) \qquad (1.73)$$

where the limits of integration have been extended to $(-\infty,\infty)$ without introducing any appreciable errors. Thus, we reach the conclusion that the quantity $L(\underset{\sim}{k},\underset{\sim}{k}')$, which governs the overlapping of the partial-wave intensities at the image plane, is just equal to the Fourier transform of the energy-spread distribution[24]. By writing Eq. (1.72) as

$$|\psi(\underset{\sim}{r}_i)|^2 = \iint S_o(\underset{\sim}{k})S_o^*(\underset{\sim}{k}')\ T(\underset{\sim}{k})T^*(\underset{\sim}{k}')\ L(\underset{\sim}{k},\underset{\sim}{k}')$$

$$\cdot\ \exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i]d\underset{\sim}{k}\ d\underset{\sim}{k}' \qquad (1.74)$$

and then looking at the spatial-frequency dependence of $L(\underset{\sim}{k},\underset{\sim}{k}')$ in Eq. (1.73), we see that the energy-spread attenuates the high spatial frequencies of the image spectrum. Since the phase shift which was introduced by the chromatic aberration has the same angular dependence as the defocusing term, the attenuation of high-resolution detail can be partially cancelled by an appropriate underfocusing of the objective lens.

ii) Time-dependent fluctuations

Fluctuations in the accelerating voltage and the current power supply of the objective lens, produce an incoherent superposition of

the image intensities over the photographic exposure time t_r . Therefore, by analogy with Eq. (1.72), we can write the recorded image intensity as

$$|\psi(\underset{\sim}{r}_i)|^2 = \int_0^{t_r} dt' \iint S_o(\underset{\sim}{k}) S_o^*(\underset{\sim}{k}') T[\Delta U(t'), \Delta i_\ell(t'), \underset{\sim}{k}] \cdot T^*[\Delta U(t'), \Delta i_\ell(t'), \underset{\sim}{k}'] \exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i] d\underset{\sim}{k}\, d\underset{\sim}{k}' \qquad (1.75)$$

where $\Delta U(t')$ is the fluctuation in accelerating voltage, and $\Delta i_\ell(t')$ is the fluctuation in objective lens current. The instantaneous transfer function, $T[\Delta U(t'), \Delta i_\ell(t'), \underset{\sim}{k}]$, can then be linked to the chromatic aberration constant by the following expression[24]:

$$T[\Delta U(t'), \Delta i_\ell(t'), \underset{\sim}{k}] = \frac{b(\underset{\sim}{k})}{M} \exp\left\{-\frac{\pi i}{2\lambda}\left[C_s\theta^4 - 2\Delta Z\theta^2 + 2C_c\theta^2\left(\frac{\Delta U(t')}{U} - 2\frac{\Delta i_\ell(t')}{i_\ell}\right)\right]\right\} \qquad (1.76)$$

where i_ℓ is a time average value for the objective lens current. By inspecting Eq. (1.76), we conclude that these fluctuations have the same effect on the transfer function as a time fluctuating focus parameter, which would be specified by

$$\Delta Z(t') = \Delta Z + C_c\left[\frac{\Delta U(t')}{U} - 2\frac{\Delta i_\ell(t')}{i_\ell}\right] \qquad (1.77)$$

A time average of these fluctuations is normally evaluated. In high-resolution microscopes, these time averages have been reduced to about 2 parts in 10^6 over an exposure time of two minutes. Hence, the chromatic incoherence caused by these fluctuations is usually negligible

in comparison to the thermal energy spread.

C) Spatial and chromatic incoherence

When both the spatial and the chromatic incoherence of the incident beam are taken into consideration, we can still view the image as an incoherent superposition of appropriately weighted partial waves. Since electrons with different $\underset{\sim}{K}_o$ and δU are assumed to be mutually uncorrelated, this distribution in both angular and energy spread, $Q(\underset{\sim}{K}_o,\delta U)$, is a separable function of $\underset{\sim}{K}_o$ and δU ; so that it can be expressed as: $Q(\underset{\sim}{K}_o,\delta U) = Q(\underset{\sim}{K}_o)N_e(\delta U)$. If we ignore time-dependent chromatic-aberration factors, the image intensity can then be computed from the expression (cf. Eqs. (1.66), (1.74)):

$$|\psi(\underset{\sim}{r}_i)|^2 = \iint T(\underset{\sim}{k})T^*(\underset{\sim}{k}')\, L(\underset{\sim}{k},\underset{\sim}{k}')\, \exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i]$$

$$\cdot \{ \int S_o(\underset{\sim}{k}-\underset{\sim}{K}_o)S_o^*(\underset{\sim}{k}'-\underset{\sim}{K}_o)Q(\underset{\sim}{K}_o)\, d\underset{\sim}{K}_o\}\, d\underset{\sim}{k}\, d\underset{\sim}{k}' \qquad (1.78)$$

where we have assumed that the specimen can be viewed as a transmission function.

1.4 Image Recording

A transmission electron micrograph can be thought of as a two-dimensional field of information, in which each element is proportional to the number of electrons that impinge upon it during a finite period of time. The mechanical conversion of a flux of electrons into an optical density distribution can be achieved in various ways, such as: electron photography, electron-to-photon conversion via fluorescent materials, and charge transport devices. We shall limit the following

discussion to the photographic recording process[4].

1.4.1 Electron Statistics

The quantum nature of the electron wave represents the main contribution to the noise during the image recording. The electrons arriving at a finite area in the image plane obey Poisson statistics, which also are used to describe quantum noise. Let us denote by N_e the average number of electrons that have been recorded on an area A of the image per recording time. Since the corresponding standard deviation is $\sqrt{N_e}$, the signal-to-noise ratio varies as the square root of the electron intensity. Hence, the relative quantum or shot noise has a $1/\sqrt{N_e}$ dependence. We shall now evaluate the magnitude of this noise as a function of experimentally known parameters[4].

The electron beam current density j_e, the operating magnification M, and the incident electron flux N_e, are linked by the identity:

$$N_e = \frac{j_e t_r A}{M^2 e} \tag{1.79}$$

where e is the charge of the electron. Consequently, the amount of noise decreases as either the exposure time or the beam current density are increased. Since j_e is fixed by the electron gun geometry and the source brightness, and t_r cannot exceed a certain threshold beyond which contamination and radiation damage modify the specimen structure, the noise cannot be reduced below a certain level. For example, consider the case where $j_e = 0.01$ Amp/mm^2 (i.e., for a hairpin tungsten filament), M = 500,000X, and A = 50 x 50 μm^2; then, in order to limit statistical electron noise to about 3%, we need N_e = 1100 electrons

and an exposure time of $t_r = 1.8$ sec.

When the microscope operates at high resolution, an upper bound for the tolerable level of quantum noise is a function of the achievable image contrast. Let us define the image contrast, C_i, as

$$C_i = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{\Delta I}{I_{max} + I_{min}} \tag{1.80}$$

Then, contrast variations, ΔI, are only detectable in quantum noise if they exceed the random noise fluctuation by a given amount, that is,

$$\Delta I \geq f_c \sqrt{N_e} \tag{1.81}$$

where f_c is a factor which depends on the experimental conditions[4]. By recalling that N_e is the average intensity, we can express the image contrast as

$$C_i \simeq \frac{\Delta I}{2N_e} \geq \frac{f_c}{2\sqrt{N_e}} \Longrightarrow N_e \geq \frac{f_c^2}{4C_i^2} \tag{1.82}$$

For example, in order to detect a contrast of 8%, we need a minimum of 1100 electrons, if we use the detection criterion $f_c = 5$.

Finally, we shall consider the critical parameters for visualization of an atom if the micrograph is recorded on an emulsion that requires P_e electrons per μm^2 to give an optical density of one (e.g., P_e is about unity for a moderately fine grain emulsion); then the number of electrons available for the imaging of an atom can be written as

$$N_e = \pi r_a^2 M^2 P_e \cdot 10^{-8} \tag{1.83}$$

where r_a is the atomic radius in Angstroms. By combining Eqs. (1.82) and (1.83), we reach the conclusion that

$$M \geq \frac{f_c \cdot 10^4}{2r_a C_i \sqrt{\pi P_e}} \tag{1.84}$$

For instance, the minimum instrumental magnification necessary to resolve an atom of $r_a = 1\text{Å}$, with a contrast of 8% above the background, is 200,000X.

In the high-resolution range, we have therefore shown that the properties of the emulsion, radiation damage and source brightness must be evaluated in order to obtain atomic resolution. Consequently, the operating conditions must be carefully chosen, in order to achieve a given resolution at a reasonable contrast level.

1.4.2 Photographic Graininess

Photographic graininess can be treated as the product of two statistically random processes: one is the electron noise which has just been described, the other is the photographic noise. It has been shown that the emulsion is an almost perfect noiseless detector when exposed to high-energy electrons[4]. Thus, random fluctuations in optical density (i.e., graininess) are nearly accounted for by electron noise. An explanation of this phenomenon is that the electron beam is sufficiently ionizing to ensure that the exposure of an emulsion to electrons is a single hit process. As a result, the optical density is directly proportional to the exposure to within 5% for densities up to about 1.5.

Since the response of an emulsion to electrons is linear, the theory is greatly simplified and yields the following expression for the granularity[4]

$$\frac{\sigma_d}{\bar{d}} = \frac{1}{\sqrt{N_e}} \left(1 + \frac{2}{\bar{\eta}_{sg}}\right)^{1/2} \tag{1.85}$$

where σ_d is the variance of the optical density distribution, $\bar{d}$ is the mean of the optical density distribution, and $\bar{\eta}_{sg}$ is the average number of silver grains produced by a single electron. Equation (1.85) shows that the electron statistical noise is amplified by the photographic process by a factor $(1+2/\bar{\eta}_{sg})^{1/2}$. This factor is less than two for ordinary electron emulsions, since the mean number of developed grains is larger than one. Since graininess is mostly caused by quantum noise, the fine detail recorded above the graininess is determined by the number of electrons that pass through the specimen during the time of exposure.

The statistical properties of granularity can be measured from the transform of a micrograph recorded in the absence of specimen. An estimate of the noise correlation between adjacent areas in the image, and of the variance of the noise density distribution can be obtained from this transform. The Information Theory techniques presented in Chapter Two will assume that this random noise can be dealt with as a superposition of Gaussian-distributed independent sources, which are additive to the input signal[26].

CHAPTER TWO

IMAGE PROCESSING

CHAPTER TWO

IMAGE PROCESSING

2.1 Introduction

The electron microscopist often needs to further process his data when a quantitative interpretation of high-resolution micrographs is desired. This processing (e.g., image enhancement and the extraction of selective features) is best implemented by digital computer, which is more versatile in handling complex nonlinear problems than a coherent optical processor. In this chapter we shall discuss the preparation of images for computer analysis (i.e., the conversion of analog video data to and from digital form and the resulting image degradation and its impact on resolution. Finally, we shall discuss the processing algorithms that are used in this work.

2.2 The Digitization Process

2.2.1 Overview

The optical information from an electron micrograph is converted into digits by a film recorder/scanner (see Fig. 2-1)[27,28]. First a flying-spot video film converter (VFC) scans the transparency, and the transmittance is measured continuously during the scan by a photomultiplier tube. Part of the light produced by the cathode ray tube (CRT) is diverted by a beam splitter to monitor a feedback loop. A converter then divides the scan into incremental spatial elements and transforms the photographic measurements into digital form. The picture elements, or pixels as they are called, are quantized to 6 bits accuracy and recorded on a magnetic tape. The scanning is monitored by

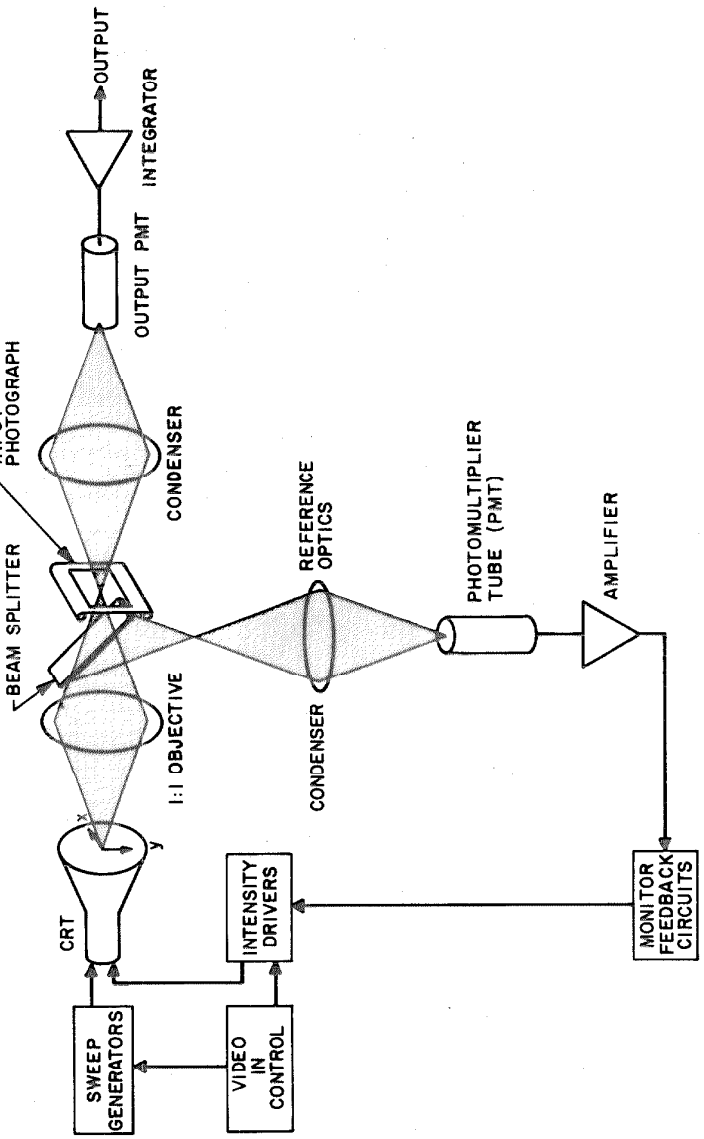


Fig. 2-1. Block diagram of the video film recorder/scanner

an operator who uses video displays to adjust the gain control in order to ensure that the dynamical range of optical densities is fully covered. A schematic diagram of the processing system is shown in Fig. 2-2.

2.2.2 Sampling

After digital transformation, the picture is represented as a discrete array of optical densities. In this section we shall examine the ensuing information loss during the conversion of the image from optical to binary form. We shall also discuss the relationship between the discrete and the continuous optical density distributions.

2.2.2.1 Data collection by the integrator

If $d(\underset{\sim}{r})$ is the continuous density distribution of the micrograph, the light intensity falling inside the photometer spot (i.e., the integrator in Fig. 2-1) is proportional to the film transmittance:

$$I(\underset{\sim}{r}) = I_o t(\underset{\sim}{r}) = I_o e^{-2.3d(\underset{\sim}{r})} \tag{2.1}$$

where I_o is the uniform illumination intensity incident upon the film, and $t(\underset{\sim}{r})$ is the transmittance of the film. The scanning aperture, which is defined by the CRT, dwells on a given element (i,j) of the CRT raster grid, while the transmitted signal is integrated in order to attenuate the noise fluctuations. The output of this process is

$$I_o \hat{t}_{ij} = \frac{1}{A_p} \int_{A_p} I_{ij}(\underset{\sim}{r})\, d\underset{\sim}{r} \tag{2.2}$$

where $\hat{t}_{ij}$ is the transmittance at the point (i,j) and A_p is the area of the scanning aperture. The integrator read out is just a

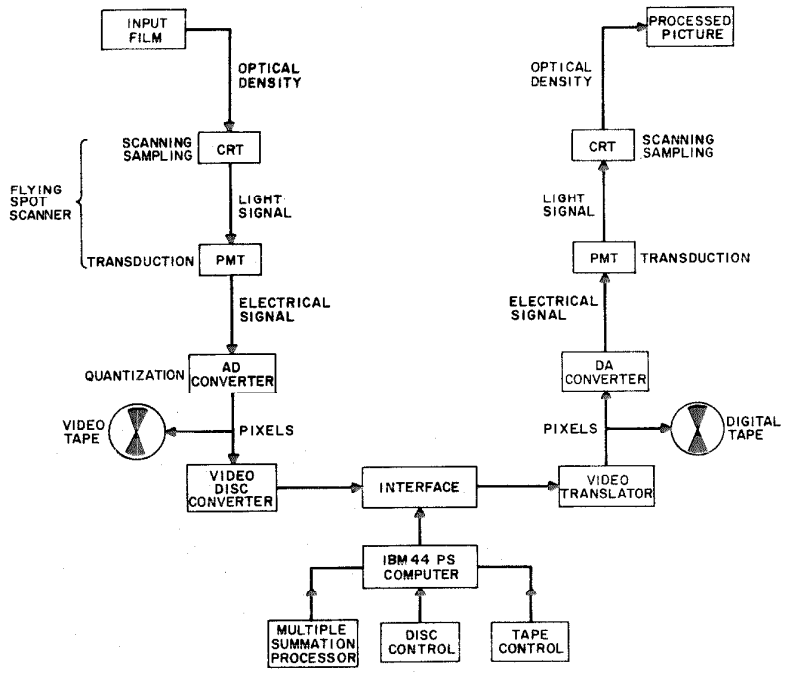


Fig. 2-2. Schematic diagram of an image processing system, showing the video film conversion steps between the original picture and its processed version

convolution of $t(\underset{\sim}{r})$ with the CRT scanning-spot aperture function $s(\underset{\sim}{r})$:

$$\hat{t}(\underset{\sim}{r}) = t(\underset{\sim}{r}) * s(\underset{\sim}{r}) \tag{2.3}$$

The spot is projected by an objective lens onto the film; so that $s(\underset{\sim}{r})$ is a convolution of the impulse response of the objective lens with the illumination function from the CRT, which is Gaussian. We may approximate $s(\underset{\sim}{r})$ by the Gaussian distribution

$$s(\underset{\sim}{r}) = e^{-2.4r^2/a^2} \tag{2.4}$$

where a is the half-amplitude diameter of the projected spot and is approximately equal to the spot size. The signal from the photomultiplier tube undergoes a logarithmic transformation, which yields an analog representation of the optical density along the path of the scanning spot:

$$\hat{d}_{ij} = -\log_{10}\hat{t}_{ij} = -\log_{10}\left[\frac{1}{A_p}\int_{A_p} e^{-2.3d_{ij}(\underset{\sim}{r})}\, d\underset{\sim}{r}\right] \tag{2.5}$$

where $\hat{d}_{ij}$ is the analog representation of the optical density at the position (i,j).

Let us now consider how $\hat{d}(\underset{\sim}{r})$ is related to the continuous film density $d(\underset{\sim}{r})$. The mean optical density $\overline{d}_{ij}$ over the spot located at the (i,j) pixel of the grid is defined as:

$$\overline{d}_{ij} = \frac{1}{A_p}\int_{A_p} d_{ij}(\underset{\sim}{r})\, d\underset{\sim}{r} \tag{2.6}$$

where the integration is performed over the aperture centered at (i,j). If the spot is small enough, the local variations in density within the spot are negligible, so that

$$d_{ij}(\underset{\sim}{r}) = \overline{d}_{ij} + \Delta d_{ij}(\underset{\sim}{r}) \tag{2.7}$$

where $|\Delta d_{ij}(\underset{\sim}{r})| << \overline{d}_{ij}$. In this approximation Eq. (2.5) becomes

$$\hat{d}_{ij} \simeq -\log_{10}[\frac{1}{A_p}\int_{A_p} e^{-2.3\overline{d}_{ij}(1 + \frac{\Delta d_{ij}(\underset{\sim}{r})}{\overline{d}_{ij}})}\, d\underset{\sim}{r}]$$

$$= \overline{d}_{ij} - \log_{10}[\frac{1}{A_p}\int_{A_p} d\underset{\sim}{r} + \frac{1}{\overline{d}_{ij}}\frac{1}{A_p}\int_{A_p} \Delta d_{ij}(\underset{\sim}{r})d\underset{\sim}{r}] = \overline{d}_{ij} \tag{2.8}$$

$$= 1 \qquad\qquad = 0$$

and therefore

$$\hat{d}_{ij} = -\log_{10}\hat{t}_{ij} = \frac{1}{A_p}\int_{A_p} d_{ij}(\underset{\sim}{r})\, d\underset{\sim}{r} \tag{2.9}$$

In a general formulation, the scanner output is a convolution of the original density distribution, $d(\underset{\sim}{r})$, with the CRT spot impulse response, $s(\underset{\sim}{r})$:

$$\hat{d}(\underset{\sim}{r}) = d(\underset{\sim}{r}) * s(\underset{\sim}{r}) \tag{2.10}$$

For a small CRT spot we find that by taking the logarithm of the measured transmittances we retrieve the same values that would be recorded by a photometer reading densities instead of transmittances. However, due to the finite size of the spot, the weighted averaging over the

aperture acts as a low-pass filter on the original picture. Thus, the Nyquist frequency of the data determines the projected spot size required to avoid any information loss.

2.2.2.2 Scanning on a grid

In practice a sample and hold device, which is driven by the horizontal sweep circuits does the actual sampling[27]. The digitizing time over which the signal is integrated is about 130 μsec, and the dead time between pulses for spot motion and digital read out is around 20 μsec. (A typical flying-spot scanning rate would be 6700 pixels per second.) Let us now consider the way in which sampling over an array of discrete points distorts the information content of the micrograph.

The sampling function, comb $(\underset{\sim}{r})$, is a regular square grid of delta functions spaced at intervals of Δ and is written[29]

$$\text{comb}(\underset{\sim}{r}/\Delta) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \delta\left(\frac{x}{\Delta} - n, \frac{y}{\Delta} - m\right) \tag{2.11}$$

where n and m are integers. We will first ignore the finite area of the picture and assume that it extends infinitely in both directions; then the discrete sampled density distribution is given by

$$d'(\underset{\sim}{r}) = \text{comb}(\underset{\sim}{r}/\Delta) \cdot \hat{d}(\underset{\sim}{r}) = \text{comb}(\underset{\sim}{r}/\Delta) \cdot (d * s) \tag{2.12}$$

The transform of the sampled function, $D'(\underset{\sim}{k})$, is a convolution of the transform of the continuous function with a square grid of points spaced at $1/\Delta$

$$D'(\underset{\sim}{k}) = \Delta \cdot \text{comb}(\Delta \cdot \underset{\sim}{k}) * [D(\underset{\sim}{k}) \cdot s(\underset{\sim}{k})] \tag{2.13}$$

where $\underset{\sim}{k}$ is the Fourier transform variable. Now if we assume that the original density distribution $d(\underset{\sim}{r})$ is band limited to $|\underset{\sim}{k}| \leq k_m$, then the Whittaker-Shannon sampling theorem[30] states that the original spectrum can be recovered from the discrete spectrum without introducing aliasing errors if

$$\Delta \leq \frac{1}{2k_m} \tag{2.14}$$

Then if M is the magnification of the microscope and ρ_s is the smallest resolvable detail, we have

$$k_m = 1/M\rho_s \tag{2.15a}$$

and

$$\Delta_m = M\rho_s/2 \tag{2.15b}$$

For example, when $M = 400{,}000X$ and $\rho_s = 3\mathring{A}$, then $\Delta_m = 60\mu$.

Another consideration in the choice of the sampling frequency is the correlation analysis, which is used to determine the translational and rotational positions of two pictures. Since the accuracy of the correlation analysis is proportional to the fineness of the mesh over which it is performed, we are forced to select a finer mesh for the sampling grid than would be prescribed by the sampling theorem. Therefore, in order to retrieve the full information and to avoid correlation between successive measurements, the sampling mesh size should be equal to the flying-spot diameter.

2.2.2.3 Boundary artifacts

The array densitometered for computation can be considered as the product of an infinite density array, $d(\underset{\sim}{r})$, with a box function,

$\Pi(r)$:

$$d_B(\underset{\sim}{r}) = d(\underset{\sim}{r}) \cdot \Pi(\tfrac{x}{X},\tfrac{y}{Y}) \tag{2.16}$$

where the parameters X and Y are the lateral dimensions of the input picture, and

$$\Pi(\tfrac{x}{X},\tfrac{y}{Y}) = \begin{cases} 1 & \text{if } \left|\tfrac{x}{X}\right| \leq \tfrac{1}{2} \text{ and } \left|\tfrac{y}{Y}\right| \leq \tfrac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

Consequently, when masking off an area, points lying outside of the box are assigned an optical density of zero. This results in boundaries with large fluctuations in density.

The truncation by $\Pi(\underset{\sim}{r})$ destroys the band limitedness of the distribution. This is because the effect of $\Pi(\underset{\sim}{r})$ is to convolve the infinite array transform, $D(\underset{\sim}{k})$, with the transform of the area, that is,

$$D_B(\underset{\sim}{k}) = D(\underset{\sim}{k}) * \left[\frac{\sin(\pi X k_x)}{\pi k_x} \quad \frac{\sin(\pi Y k_y)}{\pi k_y}\right] \tag{2.17}$$

The result of the convolution is to produce strong amplitude spikes running out from the transform origin. When these discontinuities are significant (i.e., for large values of the optical density around the perimeter of the box), apodization becomes necessary. (Usually, one smooths the boundary fluctuations with a Gaussian fall-off.) The extent to which image points near the edges are altered depends on the image-spread function of the microscope (e.g., the most favorable conditions occur at the Scherzer defocus for weak phase objects).

2.2.2.4 Reconstruction of the Continuous Distribution from the Discrete Distribution

Since the original signal $d(\underset{\sim}{r})$ is truncated, convolved with the spot window function, and sampled over a grid, the discrete image $d_c(\underset{\sim}{r})$, is given by

$$d_c(\underset{\sim}{r}) = \text{comb}(\underset{\sim}{r}/\Delta) \cdot [s(r) * (\Pi(\tfrac{x}{X},\tfrac{y}{Y}) \cdot d(\underset{\sim}{r}))] \tag{2.18}$$

The result of Fourier transforming the array of optical densities defined by Eq. (2.18) is

$$D_c(\underset{\sim}{k}) = \Delta \cdot \text{comb}(\Delta\underset{\sim}{k}) * \{S(\underset{\sim}{k})[(\frac{\sin(\pi X k_x)}{\pi k_x})(\frac{\sin(\pi Y k_y)}{\pi k_y})] * D(\underset{\sim}{k})\} \tag{2.19}$$

where $D_c(\underset{\sim}{k})$ is the computed transform, and $D(\underset{\sim}{k})$ the transform of the continuous distribution.

In order to reconstruct the signal, we must interpolate between the sampled points; this amounts to convolving $d_c(\underset{\sim}{r})$ with an interpolation function $\Gamma(\underset{\sim}{r})$:

$$d_{int}(\underset{\sim}{r}) = d_c(\underset{\sim}{r}) * \Gamma(\underset{\sim}{r}) = \{\text{comb}(\underset{\sim}{r}/\Delta)[s(\underset{\sim}{r}) * (\Pi(\tfrac{x}{X},\tfrac{y}{Y}) \cdot d(\underset{\sim}{r}))]\} * \Gamma(\underset{\sim}{r}) \tag{2.20}$$

In the 0^{th} order the interpolation function is

$$\Gamma_0(\underset{\sim}{r}) = \Pi(\tfrac{x}{2\Delta},\tfrac{y}{2\Delta}) \tag{2.21}$$

so that the interpolation has a smoothing effect that tends to band limit the density spectrum. If the signal were perfectly band limited, which truncation prevents it from being, the sampling theorem states

that exact recovery could be obtained from the sampled array by injecting at each sample point the first-order interpolation function $\Gamma_1(\underset{\sim}{r})$ where

$$\Gamma_1(\underset{\sim}{r}) = \frac{\sin(\frac{\pi x}{\Delta})}{\pi x} \; \frac{\sin(\frac{\pi y}{\Delta})}{\pi y} \tag{2.22}$$

A practical rule for digitizing images is to employ a spot size with a diameter $a \leq \Delta$, and a sampling frequency such that $\Delta \leq 1/2k_m$. Application of this criterion results in very little loss of information.

2.2.3 Quantization

The voltage stored in the integrator shown in Fig. 2-1 is converted into binary form by an analog to digital converter[27] (the output being formatted and written on magnetic tape as the scan proceeds.) During this conversion process, photometric nonlinearities bias the optical density measurements at both ends of the density range. These nonlinearities originate from the lens system, the photomultipliers, and the analog electronics. To compensate for these distortions, calibration measurements are taken during the scan in order to determine the scanning transfer functions of the input density versus output grey level. These transfer functions then enable one to make photometric corrections.

The quantizing accuracy is 6 bits, yielding 64 grey levels[27]. The transmittance at the first step away from zero is $t_1 = N_g^{-1}$ where N_g is the number of grey levels. This means that the peak optical density that can be recorded is $\log N_g$, if the original optical densities are on the linear portion of the Hurter-Driffield curve. In order to

preserve information, some precautions must be taken to reduce the peak density of the film to 1.8 or less. One must also consider the noise performance of the image processing system, since noise interacts with the visual appearance of the picture and perturbs the subsequent digital processing.

Sources of noise during scanning of the film include the scanning light, the detector, the amplifier system, and the analog to digital converter. In the presence of noise, it is the signal plus noise which is quantized. Consequently, in order to ensure that there is no ambiguity in the designation of a particular signal level as a certain digital number (DN), the root-mean-square noise should be kept within one-third of the digital step size[27].

2.3 Digital Processing Techniques

After the micrographs are digitized, corrections are performed by computer in order to overcome limitations imposed by the noise, the signal distortions, and the transfer function of the microscope[31,32,33]. These techniques are used to improve the signal-to-noise ratio and the image resolution, thereby rendering a quantitative image analysis tractable. In this section we describe the information retrieval techniques that will be encountered in the processing schemes used in Chapters Three and Four.

2.3.1 Contrast Manipulations

In pattern recognition problems, statistical analysis is often used to determine such parameters as the mean density, the variance of the density distribution, the area, and the perimeter of an image

feature. These parameters can be derived from the first-order probability density function or histogram. The histogram, H(D), of the grey level distribution values is computed from the input picture. It yields the parameters used for generating a transfer function on the domain of intensity values (pixels are eight bits long, their levels vary from 0 to 255), and statistical information about the image. For instance, the area A of a picture can be expressed as

$$\sum_{D=0}^{255} H(D) = NL \times NS = A \tag{2.23}$$

where NL is the number of lines and NS is the number of samples in the picture. The mean picture density, $\overline{D}$, is expressed in terms of the histogram as

$$\frac{1}{A} \sum_{D=0}^{255} D \times H(D) = \overline{D} \tag{2.24}$$

The histograms of input micrographs that contain little visible detail show a large concentration of pixels in a narrow portion of the grey scale. The image can be enhanced by expanding the grey scale. By analogy to the distribution transform in statistics, an expansion of the grey levels amounts to modifying the shape of the histogram. This process is called contrast stretching.

After digitizing, one stretches the grey scale over the full dynamic range (0 to 255) by applying an intensity transfer function to the input densities, yielding

$$\tilde{D} = f(D) \tag{2.25}$$

where $\tilde{D}$ is the stretched density, f is the transfer function, and D is the input density. In this process one allows for a saturation of the histogram of $L\%$ at 0 DN and $H\%$ at 255 DN. Then the tails, D_{min} and D_{max}, of the histogram shown in Fig. 2-3 are defined by

$$\frac{1}{A}\sum_{D=0}^{D_{min}} H(D) = L\%$$

$$\frac{1}{A}\sum_{D=D_{max}}^{255} H(D) = H\% \tag{2.26}$$

Thus, the function f maps $[D_{min},D_{max}]$ onto $[0,255]$. For example, a linear contrast enhancement would be:

$$\begin{aligned} \tilde{D} &= 0 && D \le D_{min} \\ \tilde{D} &= 255\,\frac{D-D_{min}}{D_{max}-D_{min}} = \bar{\alpha}D+\bar{\beta} && D_{min} \le D \le D_{max} \\ \tilde{D} &= 255 && D \ge D_{max} \end{aligned} \tag{2.27}$$

where $\bar{\alpha}$ and $\bar{\beta}$ are parameters which define the transformation.

After the stretching operation, the histogram is transformed into[26]

$$\tilde{H}(\tilde{D}) = H(D)/(df/dD) = H(f^{-1}(\tilde{D}))/(df/dD) \tag{2.28}$$

For instance, in the case of a linear stretch, Eq. (2.28) becomes

$$\tilde{H}(\tilde{D}) = \frac{1}{\bar{\alpha}} H(\frac{\tilde{D} - \bar{\beta}}{\bar{\alpha}}) \tag{2.29}$$

In Eq. (2.29) we see that enhancing the contrast (i.e., $\bar{\alpha} > 1$) flattens

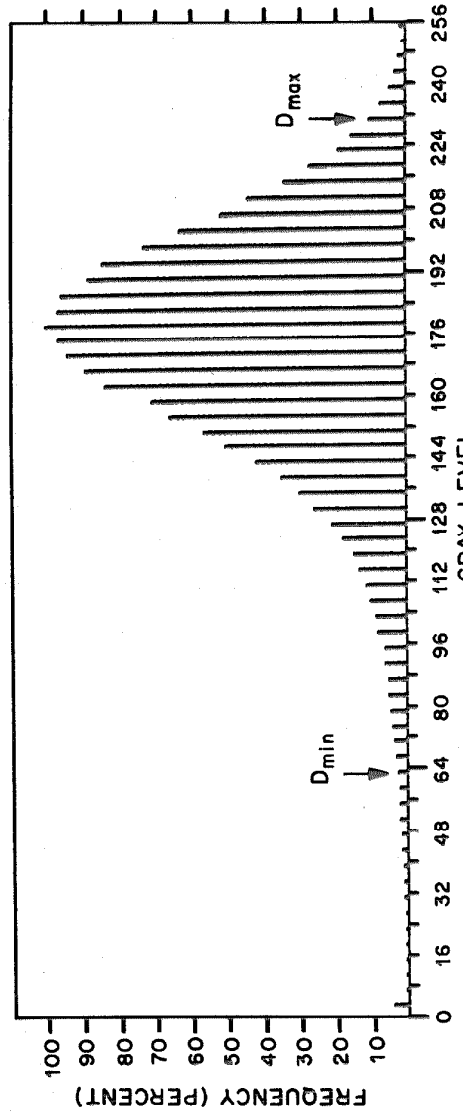


Fig. 2-3. Definition of the tails of the histogram, D_{min} and D_{max}. The vertical axis gives the frequency of the data normalized to a maximum frequency of 100%. The horizontal axis shows the gray levels of the input picture.

the DN frequency distribution. This selective contrast stretching of fine detail is equivalent to a high spatial frequency enhancement.

Contrast manipulation techniques are also used to remove shading effects and perform photometric corrections on non-uniformly illuminated pictures. Shading effects can be removed by finding a polynomial surface that fits the trending in the picture, and then removing the trend pixel by pixel by linear interpolation. Another procedure for the removal of large-scale slow variations in picture data numbers applies a Gaussian DC notch filter to the Fourier transform of the input data set[34]. The removal of trends in a picture with fairly uniform shading is a critical step in both correlation and spectral analysis.

2.3.2 Geometrical Operations

Adjustments in the translation, rotation, and magnification of a picture require a geometric distortion of the original sampling grid. The relationship between the desired image $\tilde{d}(x,y)$ and the original image $d(x,y)$ is given by

$$\tilde{d}(x,y) = d[a(x,y),b(x,y)] \tag{2.30}$$

where $a(x,y)$ and $b(x,y)$ define the transformed coordinates. For instance, a rotation of the picture through an angle θ, is specified by

$$a(x,y) = x \cos\theta - y \sin\theta$$

$$b(x,y) = x \sin\theta + y \cos\theta$$

We shall now consider the way that a geometric transformation is implemented on a discrete set of data points. If we are given the sample values $d(i\Delta x, j\Delta y)$ of $d(x,y)$ (where $i = 1,2,\cdots,N$; $j = 1,2,\cdots,M$, Δx and Δy are the spacings of the input grid) then we can determine the samples $\tilde{d}(k\Delta x', \ell\Delta y')$ (where $k = 1,2,\cdots,N'$ and $\ell = 1,2,\cdots,M'$, $\Delta x'$ and $\Delta y'$ define the output grid spacings) by bilinear interpolation. To see this let (K,L) be two fixed integers and (I,J) be two integers such that

$$\tilde{d}_{K,L} = \tilde{d}(K\Delta x', L\Delta y') = d(x,y) = d((I+\alpha)\Delta x, (J+\beta)\Delta y) \qquad (2.31)$$

for $0 \leq \alpha < 1$ and $0 \leq \beta < 1$. Then a bilinear interpolation yields[33]

$$\tilde{d}_{K,L} = d_{I,J}(1-\alpha)(1-\beta) + d_{I,J+1}(1-\alpha)\beta + d_{I+1,J}\,\alpha(1-\beta) + d_{I+1,J+1}\alpha\beta \qquad (2.32)$$

More elaborate interpolation procedures may be developed by using higher order polynomials for interpolating functions.

2.3.3 Noise Removal

Errors in the recorded signal can be caused by random noise additions or multiplications from various sources. If we assume that both the signal and the noise are additive and mutually independent, then we are able to perform a statistical analysis.

2.3.3.1 Periodic noise removal

A visual examination of the micrograph may disclose a systematic frequency superimposed upon the original image. This noise can

originate during the scanning of the film and appears in the transform as a series of frequencies located at right angles to the direction of the scan. This noise can be removed by employing a noise-removal algorithm which locates noise spikes in the spectrum and interpolates across them[35].

2.3.3.2 Snow removal

Random points of bad data cause substantial changes in the signal levels, which saturate pixel intensities and give rise to the appearance of "snow". This "salt and pepper" noise can be removed using majority logic techniques[36]. For example, pixels whose intensity deviations exceed a certain threshold difference, δ, are replaced with their local average.

When an edge is present in the picture, this averaging must be restricted to a direction along the edge. A simple algorithm for this processing employs a unidirectional nearest-neighbor average.

$$\text{If} \quad \frac{1}{3}\left|2d_{i,j} - d_{i-1,j} - d_{i+1,j}\right| \geq \delta \Longrightarrow \tilde{d}_{ij} = \frac{1}{3}(d_{i-1,j} + d_{i,j} + d_{i+1,j})$$

$$\text{If} \quad \frac{1}{3}\left|2d_{i,j} - d_{i-1,j} - d_{i+1,j}\right| < \delta \Longrightarrow \tilde{d}_{ij} = d_{i,j} \qquad (2.33)$$

This nonlinear process produces a net improvement in image quality for moderate noise levels.

2.3.3.3 Random noise removal

A) Averaging of multiple copies

If the signal has been degraded by quantum electron noise, photographic granularity, and quantization noise introduced during the

processing; then these noise sources can be assumed to be independent and additive with respect to the signal. (This approximation holds well in bright-field imaging where electron fluctuations and scattered-wave oscillations are small compared to the transmitted beam.) The additive noise $n(x,y)$ is assumed to be normally distributed with mean 0 and variance σ^2. Then if N copies of the picture are made in such a way that the noise samples in these copies are all independent, an averaging of the set results in a variance of the average picture-noise distribution equal to σ^2/N [35]. Consequently, one improves the signal-to-noise ratio by $\sqrt{N}$, when N pictures with overlapping areas are superimposed. Similarly, the image of a periodic structure can be enhanced by translating the image by an amount equal to the repeat distance, and then superimposing the image upon itself.

B) Bandpass spatial frequency filtering

Another way to smooth a picture is to attenuate the high-spatial-frequency content of the signal by applying mathematical apertures to its Fourier transform. This low-pass-spatial filtering reduces the film noise and the noise due to electron intensity fluctuations in the micrograph. In order to determine the optimum aperture, one must estimate the highest frequencies that carry information about the object (e.g., when imaging weak phase objects, one uses the Fourier resolution test). After the signal spectrum has been determined, one may then delete a selected band of spatial frequencies corresponding to random and quasi-periodic noise[35,37,38].

C) Wiener filtering

When both the signal and noise spectra overlap, a filter theory developed by Wiener gives a solution to the problem of separating the signal from the noise[39,40]. We assume that the received image, $d_{rec}(x,y)$, results from a linear superposition process

$$d_{rec}(\underset{\sim}{r}) = d(\underset{\sim}{r}) + n(\underset{\sim}{r}) \tag{2.34}$$

where $d(\underset{\sim}{r})$ is the object density distribution and $n(\underset{\sim}{r})$ is a spatially-stationary additive noise. The Wiener filter is designed in such a way as to minimize the mean-square discrepancy between the actual output, $d_{rec}(\underset{\sim}{r})$, and the desired output, $d(\underset{\sim}{r})$. Because the signal distributions are stationary, the mean-square-error integral, E_n, which normally is computed over an infinite domain, may be approximated by

$$E_n = \frac{1}{A} \int_A [d_{rec}(\underset{\sim}{r}) - d(\underset{\sim}{r})]^2 \, d\underset{\sim}{r} \tag{2.35}$$

This minimization problem introduces the concept of the cross-correlation function (CCF) between the density distributions $d(\underset{\sim}{r})$ and $d_{rec}(\underset{\sim}{r})$. If the image area A over which the CCF is computed is large enough, then we may write

$$\phi_{d,d_{rec}}(\underset{\sim}{r}) = \frac{1}{A} \int_A d(\underset{\sim}{r}+\underset{\sim}{r}') \, d_{rec}(\underset{\sim}{r}') \, d\underset{\sim}{r}' \tag{2.36}$$

where $\phi_{d,d_{rec}}(\underset{\sim}{r})$ is the CCF. Then, if we call $\Phi_{d,d_{rec}}(\underset{\sim}{k})$ the CCF Fourier transform, the Wiener restoration filter is found to be[39]

$$H_w(\underset{\sim}{k}) = \frac{\Phi_{d,d_{rec}}(\underset{\sim}{k})}{\Phi_{d_{rec},d_{rec}}(\underset{\sim}{k})} \tag{2.37}$$

If the input signal and the noise are assumed to be uncorrelated (i.e., $\phi_{d,n}(\underset{\sim}{r}) = 0$), then we may write

$$\phi_{d,d_{rec}}(\underset{\sim}{r}) = \phi_{d,d}(\underset{\sim}{r})$$

$$\phi_{d_{rec},d_{rec}}(\underset{\sim}{r}) = \phi_{d,d}(\underset{\sim}{r}) + \phi_{n,n}(\underset{\sim}{r}) \tag{2.38}$$

where $\phi_{d,d}(\underset{\sim}{r})$ is the autocorrelation function (ACF) of the input signal and $\phi_{n,n}(\underset{\sim}{r})$ is the ACF of the noise. Then in this approximation, the Wiener restoration filter becomes

$$H_w(\underset{\sim}{k}) = \frac{\Phi_{d,d}(\underset{\sim}{k})}{\Phi_{d,d}(\underset{\sim}{k}) + \Phi_{n,n}(\underset{\sim}{k})} = \frac{|D(\underset{\sim}{k})|^2}{|D(\underset{\sim}{k})|^2 + |N(\underset{\sim}{k})|^2} \tag{2.39}$$

where $|N(\underset{\sim}{k})|^2$ is the power spectrum of the noise and $\Phi_{d,d}(\underset{\sim}{k}) = |D(\underset{\sim}{k})|^2$ by the Wiener-Khintchine theorem. The problem now is to obtain the power spectra $|D(\underset{\sim}{k})|^2$ and $|N(\underset{\sim}{k})|^2$. If the noise is white (i.e., $|N(\underset{\sim}{k})| = |N(\underset{\sim}{k}_m)| =$ constant N_{wh}) and if the signal-to-noise ratio can be assumed to be zero at the highest frequency, $\underset{\sim}{k}_m$, carrying information, then the signal and the noise spectra can be estimated from

$$|D_{rec}(\underset{\sim}{k}_m)|^2 = |N(\underset{\sim}{k}_m)|^2$$

$$|D(\underset{\sim}{k})|^2 = |D_{rec}(\underset{\sim}{k})|^2 - |D_{rec}(\underset{\sim}{k}_m)|^2 \tag{2.40}$$

Another approach is to derive the optimum filter from an a priori knowledge of the statistics of the object and the assumption of a certain probability distribution for the signal power spectrum[41]. This filtering is only valid when both the random noise and the signal are uncorrelated and additive. In electron microscopy this condition is often not met in dark-field imaging, since both the quantum noise and the signal are comparable in magnitude.

2.3.4 Matched Filtering and Correlation Analysis

It is often necessary to find a part of one picture that matches a similar part in another picture. This can be achieved by the method of cross-correlation, which allows us to determine whether or not portions of two pictures are identical (except for translation and multiplication by a constant). An explanation of this technique is best made within the framework of signal detection theory and matched filtering analysis[30,42,43].

2.3.4.1 Matched filtering

When a known signal is embedded in a noise background, the problem of detecting this signal in the measured optical density distribution often arises. A class of filters, called energy matched filters, is used as a pattern recognition device. This filter, $H_{MF}(\underset{\sim}{k})$, is optimum in maximizing the signal-to-noise ratio for detection purposes. A simple derivation of the analytical expression for this filter will now be presented for the case where the input signal, $d(\underset{\sim}{r})$, is to be detected in the presence of stationary noise, $n(\underset{\sim}{r})$.

The result of applying this filter to the received signal, $d_{rec}(\underset{\sim}{r})$, is to alter the input-noise power spectrum, $|N(\underset{\sim}{k})|^2$, to[40]

$$|N_{MF}(\underset{\sim}{k})|^2 = |N(\underset{\sim}{k})|^2 \, |H_{MF}(\underset{\sim}{k})|^2 \tag{2.41}$$

where $|N_{MF}(\underset{\sim}{k})|^2$ is the noise spectrum of the filtered image. According to Parseval's theorem, we know that the energy of the output noise, P_N, may be expressed as

$$P_N = \int_{-\infty}^{\infty} |N_{MF}(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} \tag{2.42}$$

A similar argument holds for the transformation of the input signal. After the filtering operation, the signal energy becomes

$$P_s = |h_{MF} * d|^2 = \left| \int_{-\infty}^{\infty} H_{MF}(\underset{\sim}{k}) \, D(\underset{\sim}{k}) \, \exp(2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}) d\underset{\sim}{k} \right|^2 \tag{2.43}$$

where $h_{MF}(\underset{\sim}{r})$ is the filter point-spread function. Therefore, by convolving the density distribution, $d_{rec}(\underset{\sim}{r})$, with a matched filter, $h_{MF}(\underset{\sim}{r})$, one modifies the original signal-to-noise ratio into

$$\tau_{MF} = \frac{P_s}{P_N} = \frac{\left| \int_{-\infty}^{\infty} D(\underset{\sim}{k}) \, H_{MF}(\underset{\sim}{k}) \, \exp(2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}) \, d\underset{\sim}{k} \right|^2}{\int_{-\infty}^{\infty} |N(k)|^2 \, |H_{MF}(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}} \tag{2.44}$$

According to the Schwartz inequality an upper bound for Eq. (2.44) is given by

$$\tau \leq \tau_{max} = \int_{-\infty}^{\infty} \left| \frac{D(\underset{\sim}{k})}{N(\underset{\sim}{k})} \right|^2 d\underset{\sim}{k} \tag{2.45}$$

with the equality holding for

$$H_{MF}(\underset{\sim}{k}) = \frac{D^*(\underset{\sim}{k})\ \exp(-2\pi i(k_x\xi + k_y\eta))}{|N(\underset{\sim}{k})|^2} \tag{2.46}$$

where ξ and η are real space coordinates of a point in the filtered image. The energy matched filter impulse response is then

$$h_{MF}(x,y) = d^*(\xi - x,\ \eta - y) * \mathcal{F}^{-1}\{|N(k)|^{-2}\} \tag{2.47}$$

In the case of white noise the filter is a rotated and conjugated version of the signal to be detected:

$$h_{MF}(x,y) = \frac{1}{N_{wh}^2}\, d^*(\xi - x,\ \eta - y) \tag{2.48}$$

The operation of the matched filter is then equivalent to a cross-correlation of the measured signal, $d_{rec}(x,y)$, with the real signal $d(x,y)$, so that

$$\begin{aligned} d_f(\xi,\eta) &= \frac{1}{N_{wh}^2} \lim_{A \to \infty} \int_A d_{rec}(x,y)\ d(x-\xi, y-\eta)\, d\underset{\sim}{r} \\ &= \frac{1}{N_{wh}^2}\, \phi_{d_{rec},d}(-\xi,-\eta) \end{aligned} \tag{2.49}$$

where $d_f(\xi,\eta)$ is the convolution product $d_{rec} * h_{MF}$. The energy matched filter in the presence of white noise is therefore a correlator; that is, a detection device for determining all possible translations of a given input signal.

2.3.4.2 Correlation analysis

A) Definition of correlation functions

The problem of maximizing the signal-to-noise ratio of a received signal $d_{rec}(\underset{\sim}{r}) = d(\underset{\sim}{r}) + n(\underset{\sim}{r})$ which is submerged in white Gaussian noise, has led us to the concept of the correlation filter. By applying a maximum likelihood analysis we can also describe the correlator as an inference estimator. For instance, suppose the output of an information channel, $d_{rec}(\underset{\sim}{r})$, is the sum of an input signal, $d_o(\underset{\sim}{r}+\underset{\sim}{r}')$, and an independent zero-mean-Gaussian noise distribution, $n(\underset{\sim}{r})$. Given the output, $d_{rec}(\underset{\sim}{r})$, and assuming that the channel inputs are equally probable, the conditional probability density that $d_o(\underset{\sim}{r}+\underset{\sim}{r}')$ is being sent reaches its highest value at the maximum of[42,43]

$$\phi_{d_o,d_{rec}}(\underset{\sim}{r}') = \frac{1}{A}\int_A d_o(\underset{\sim}{r}+\underset{\sim}{r}')\, d_{rec}(\underset{\sim}{r}')\, d\underset{\sim}{r}' \qquad (2.50)$$

(i.e., at the peak of the cross-correlation between the two real density distributions). The value of this integral is a check of the validity of the assumption that $d_{rec}(\underset{\sim}{r})$ is a superposition of $d_o(\underset{\sim}{r})$ with additive white Gaussian noise. Equation (2.50) measures the statistical dependence between neighboring areas of two distributions, and should be evaluated in the limit, $A \to \infty$. In practice the integration domain is chosen large enough for this approximation to hold.

B) Determination of matching parameters

i) Translation parameters

In image processing one often encounters the problem of matching two pictures that are noisy versions of the same pattern, $d_o(\underset{\sim}{r})$,

translated by the amounts $\underset{\sim}{R}_1$ and $\underset{\sim}{R}_2$. These pictures can be expressed by

$$\begin{aligned} d_1(\underset{\sim}{r}) &= d_0(\underset{\sim}{r}-\underset{\sim}{R}_1) + n_1(\underset{\sim}{r}) \\ d_2(\underset{\sim}{r}) &= d_0(\underset{\sim}{r}-\underset{\sim}{R}_2) + n_2(\underset{\sim}{r}) \end{aligned} \tag{2.51}$$

The noise distributions $n_1(\underset{\sim}{r})$ and $n_2(\underset{\sim}{r})$ in Eqs. (2.51) are assumed to be zero-mean white Gaussian; hence they are spatially stationary and we may write

$$d_2(\underset{\sim}{r}) = d_1(\underset{\sim}{r} + \underset{\sim}{R}_1 - \underset{\sim}{R}_2) + n_2(\underset{\sim}{r}) - n_1(\underset{\sim}{r}) \tag{2.52}$$

Consequently we may consider one of the noisy images, $d_1(\underset{\sim}{r})$, as the input signal and the other image, $d_2(\underset{\sim}{r})$, as being derived from the first in a two-fold noise process with an intermediate state, $d_0(\underset{\sim}{r})$. The convolution theorem on probability densities tells us that the noise distribution, $[n_2(\underset{\sim}{r}) - n_1(\underset{\sim}{r})]$, is Gaussian with a variance

$$\sigma_n^2 = \sigma_{n_1}^2 + \sigma_{n_2}^2 \tag{2.53}$$

We also see in Eq. (2.52) that the peak of the CCF, $\phi_{d_1,d_2}(\underset{\sim}{r})$, is located at

$$\underset{\sim}{r}_D = \underset{\sim}{R}_2 - \underset{\sim}{R}_1 \tag{2.54}$$

The point $\underset{\sim}{r}_D$ yields the relative translational displacements of the two pictures. The angular orientation can be determined prior to the correlation analysis, by visual matching, using well defined image features as markers.

ii) Width of the correlation peak

The accuracy to which the matching parameters are determined depends upon the sharpness of the correlation peak. If we assume that the images are nondistorted representations of the same object structure, $d_o(\underset{\sim}{r})$, and are contaminated by uncorrelated additive noise, then

$$\phi_{d_2,d_1} = \phi_{d_o,d_o} * \delta(\underset{\sim}{r}-\underset{\sim}{r}_D) + \phi_{d_o,n_2} * \delta(\underset{\sim}{r}-\underset{\sim}{R}_1) + \phi_{d_o,n_1} * \delta(\underset{\sim}{r}-\underset{\sim}{R}_2) + \phi_{n_1,n_2} \qquad (2.55)$$

In Eq. (2.55) the noise terms all vanish, since they are statistically independent of the signal, so that

$$\phi_{d_2,d_1} = \phi_{d_o,d_o} * \delta(\underset{\sim}{r}-\underset{\sim}{r}_D) \qquad (2.56)$$

For an object consisting of point-like atoms, ϕ_{d_o,d_o} in Eq. (2.56) is a delta function. However, a more realistic description would take the shape of the atomic-scattering factors into account (e.g., for carbon atoms the result is a peak with 1Å halfwidth), as well as the contribution of the point-spread function of the microscope. We shall show in Chapter Three that there is also a direct correspondence between the Wiener spectrum of the object and the width of its correlation peak.

iii) Possible artifacts

The correlation peak may be obscured when the pictures have either steep traverse density gradients, or if there exist sharp image

features running in a preferential direction. In the latter case, one correlates selected regions in both pictures which show a large amount of uniformly spread detail. The transformation function that determines the mutual displacement of the two pictures is estimated from a polynomial surface fit, which is applied to the set of translation vectors of all the correlated regions.

If the pictures show a fairly uniform traverse shading, then the CCF peak may no longer correspond to the absolute CCF maximum. To illustrate how the gradient shows up in the ACF, let us consider the shading function given by

$$v(x,y) = \zeta x$$
$$|x| \leq X/2 \quad , \quad |y| \leq Y/2 \tag{2.57}$$

where $v(x,y)$ is the shading function and ζ is the slope of the shading. The array that is densitometered is a truncated superposition of the continuous trend function, $v(x,y)$, and the statistical image distribution, $d(x,y)$:

$$d^{\dagger}(x,y) = [d(x,y) + v(x,y)] \cdot \Pi\left(\frac{x}{X},\frac{y}{Y}\right) \tag{2.58}$$

If the correlation area is large enough, we can neglect the effect of the truncation window on the sampled picture, so that Eq. (2.58) becomes

$$d^{\dagger}(x,y) = d(x,y) + v(x,y) \cdot \Pi\left(\frac{x}{X},\frac{y}{Y}\right) \tag{2.59}$$

Since the input image and the trend function are uncorrelated,

$$\phi_{d^\dagger,d^\dagger} = \phi_{d,d} + \phi_{v\Pi,v\Pi} \tag{2.60}$$

where

$$\phi_{v\Pi,v\Pi}(x,y) = (Y - |y|)\left(\frac{X^3}{12} - \frac{X^2}{4}|x| + \frac{|x|^3}{6}\right)\frac{\zeta^2}{XY}\,\Pi\left(\frac{x}{2X},\frac{y}{2Y}\right) \tag{2.61}$$

For small $|x|$ the ACF falls off linearly with x from its peak value $\phi_{v\Pi,v\Pi}(0,0) = X^2\zeta^2/12$. Depending on the shading parameter, ζ , the shading ACF may completely obliterate the signal ACF. Correspondingly, the correlation peak of two signals will be obscured by the CCF peak of their respective trend functions. Then the Wiener spectrum of the picture is modified by strong spikes along the k_x axis, given by

$$\Phi_{v\Pi,v\Pi}(\underset{\sim}{k}) = \frac{\zeta^2X^3Y}{36}\,\mathrm{sinc}^2(Yk_y)\,\mathrm{sinc}^2(Xk_x/3) \tag{2.62}$$

In order to avoid artifacts in the cross-correlation step, one may remove the density gradient, $v(x,y) = \zeta x + \gamma y$, in the raw pictures in two different ways. In the first method one sets the Fourier coefficients to zero along $\gamma k_x - \zeta k_y = 0$, or one replaces them by a neighbor averaging with statistically random phase. In the second method one subtracts the trend from the raw pictures by a least-squares fitting process (see 2.3.1). The parameters ζ and γ are found at the minimum of the integral

$$\int_A [d(x,y) - \zeta x - \gamma y - \bar{d}]^2\, d\underset{\sim}{r}$$

where $\bar{d}$ is the mean optical density of the input picture.

C) Discrete estimate of the correlation integrals

Let us represent two pictures by their pixel matrices $(a_{k,\ell})$ and $(b_{k,\ell})$. Then the correlation integral between two areas of these pictures can be approximated by the summation

$$\phi_{ij}^{(ab)} = \frac{1}{KL} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} a_{k,\ell}\, b_{k+i,\ell+j} \tag{2.63}$$

where K and L define the size of the overlapping area in pixels, and $[i=0,1,\cdots I-1;\ j=0,1,\cdots,J-1]$ with I and J being the correlation matrix dimensions. There exist two methods for computing Eq. (2.63), one in Fourier space and the other in real space. We shall now consider their relative advantages and domains of applicability.

i) Fourier method

From the convolution theorem, we know that

$$\phi^{(ab)} = \mathcal{F}^{-1}\{A^*(k)\, B(k)\} \tag{2.64}$$

where $\phi^{(ab)}$ is the CCF between two pictures whose Fourier spectra are $A(\underset{\sim}{k})$ and $B(\underset{\sim}{k})$. The fast Fourier transform algorithm[44] (FFT) is a very efficient method for computing the discrete transform pair of a series of data samples:

$$A_{m,n} = \frac{1}{K} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} a_{k,\ell} \exp\left[-2\pi i\left(\frac{mk}{K} + \frac{n\ell}{L}\right)\right]$$

$$a_{k,\ell} = \frac{1}{L} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} A_{m,n} \exp\left[2\pi i\left(\frac{mk}{K} + \frac{n\ell}{L}\right)\right] \tag{2.65}$$

Then by converting Eq. (2.64) into its discrete equivalent we obtain

$$\phi_{i,j}^{(ab)} = \frac{1}{L} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} A_{m,n}^{*} B_{m,n} \exp[2\pi i(\frac{mi}{K} + \frac{nj}{L})] \qquad (2.66)$$

The correlation matrix, $\phi_{i,j}^{(ab)}$, is calculated for two infinite layer lattices with the finite images being unit cells[45] (see Sec. 2.3.5.1-B). Consequently, the picture size, $[(K-1)\Delta,(L-1)\Delta]$, must be chosen large enough to avoid aliasing errors. Furthermore, the FFT algorithm restricts K and L to powers of two. This technique, although faster than the direct summation method, is not flexible and may introduce errors for small input areas or if the CCF peak lies near an edge of the correlation matrix. However, by using Fourier relations it is easy to spatial filter the pictures during the cross-correlation step. For instance, density gradients across the pictures can be removed in the transform domain by setting to zero the Fourier coefficients along the gradient direction.

ii) Direct summation method

The discrete evaluation of the correlation integral given by Eq. (2.63), may be performed directly by multiplying the matrix arrays of data numbers at the point (i,j). A correlation program computes $\phi_{ij}^{(ab)}$ (this is done using the computer-linked IBM 2937 Multiplier-Summation processor), and then searches for the peak value of the correlation matrix. Prior to this operation, both input data sets must be filtered, in order to attenuate the background noise and to remove picture trends, which could obscure the peak location. The peak may be negative when correlating two images of a through-focus series (see Chapter Three). An indication of such a peak is an ill-defined

maximum, or the fact that features in its neighborhood do not show approximate centrosymmetry. The Fourier method is then preferable to the direct summation method, since inverse filtering can be performed during the correlation step.

To protect against negative correlation peaks, one calculates the sum of the squares of the differences between a rectangle in the primary input picture and a set of rectangles in the secondary input picture . The mutual displacement is found at the minimum of

$$\phi'^{(ab)}_{i,j} = \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} (a_{k,\ell} - b_{k+i,\ell+j})^2 \qquad (2.67)$$

The rectangle size, $[(K-1)\Delta,(L-1)\Delta]$, must exceed a threshold value which is consistent with the assumption that the noise is additive and uncorrelated to the signal.

iii) Differences between calculated and real correlation functions

The sequence of operations involved in sampling a picture modifies the statistics of the optical density distribution as follows. First the scanning aperture, which acts as a low pass filter, smooths the film granularity. If the sampling step is chosen to be larger than the halfwidth of the noise ACF, while still satisfying the Nyquist criterion, then a signal-to-noise ratio enhancement appears in the picture. Finally, by truncating the picture to a finite size, one may introduce boundary errors if the area over which the CCF is computed is not large enough. Thus, the computed correlation functions are only estimates, within the framework of the theory of a stationary process,

of the correlation integrals. The deviations are kept to a minimum by choosing a very fine sampling grid.

D) Correlation peak determination

The accuracy of the match between two pictures is estimated by correlating different areas and then observing the spread in the displacement values $\underset{\sim}{r}_D$. The correlation program finds the peak position to within one sampling unit. The actual maximum, however, might be somewhere between the sampling points of the correlation matrix. If the average spread in peak positions is less than a sampling-grid spacing, a more accurate estimate of its location may be found by either of the two processing schemes which we will now describe.

The first method[46] relies upon the fact that the ACF of a real optical-density distribution is centrosymmetric about its peak. Then according to Eq. (2.56), we see that the CCF of two pictures, which differ from each other by uncorrelated additive noise, will be centrosymmetric. The peak position will then coincide with the center of gravity of the correlation surface, $\phi_{d_1,d_2}(\underset{\sim}{r})$, given by the weighted average of CCF values:

$$\underset{\sim}{r}_p = \frac{\sum_{\text{peak}} \underset{\sim}{r} \cdot \phi_{d_1,d_2}(\underset{\sim}{r})}{\sum_{\text{peak}} \phi_{d_1,d_2}(\underset{\sim}{r})} \tag{2.68}$$

$\underset{\sim}{r}_p$ is the coordinate of the peak in fractions of a pixel, and the summation extends over the width, $2H_p\Delta$, of the peak. The coordinates of the CCF peak, (x_p,y_p), for discrete data samples, are given by

$$x_p = \frac{\sum_{i=-H_p}^{H_p} \sum_{j=-H_p}^{H_p} \phi_{i,j}^{(d_1,d_2)} x_{i,j}}{\sum_{i=-H_p}^{H_p} \sum_{j=-H_p}^{H_p} \phi_{i,j}^{(d_1,d_2)}}$$

$$y_p = \frac{\sum_{i=-H_p}^{H_p} \sum_{j=-H_p}^{H_p} \phi_{i,j}^{(d_1,d_2)} y_{i,j}}{\sum_{i=-H_p}^{H_p} \sum_{j=-H_p}^{H_p} \phi_{i,j}^{(d_1,d_2)}} \qquad (2.69)$$

The technique that we have just described relies on the assumption of centrosymmetry; therefore it is invalid if sharp features or density gradients exist in the original image.

In the second method one computes the coordinates of the CCF peak from a bilinear interpolation of the CCF values near the peak. In practice, a 3 x 3 matrix of CCF values in the neighborhood of the peak (see Fig. 2.4) is used to determine the peak coordinates. Analytically, we can express these coordinates as

$$x_p = \frac{\Delta}{2} \sum_{i=1}^{3} \left(\frac{\phi_{i,3} - \phi_{i,1}}{2\phi_{i,2} - \phi_{i,1} - \phi_{i,3}} \right)$$

$$y_p = \frac{\Delta}{2} \sum_{j=1}^{3} \left(\frac{\phi_{3,j} - \phi_{1,j}}{2\phi_{2,j} - \phi_{1,j} - \phi_{3,j}} \right) \qquad (2.70)$$

Both of these algorithms output the peak positions in fractions of pixels, since the pictures are sampled on a two-dimensional grid. It is therefore necessary to shift one of the sampling grids by an

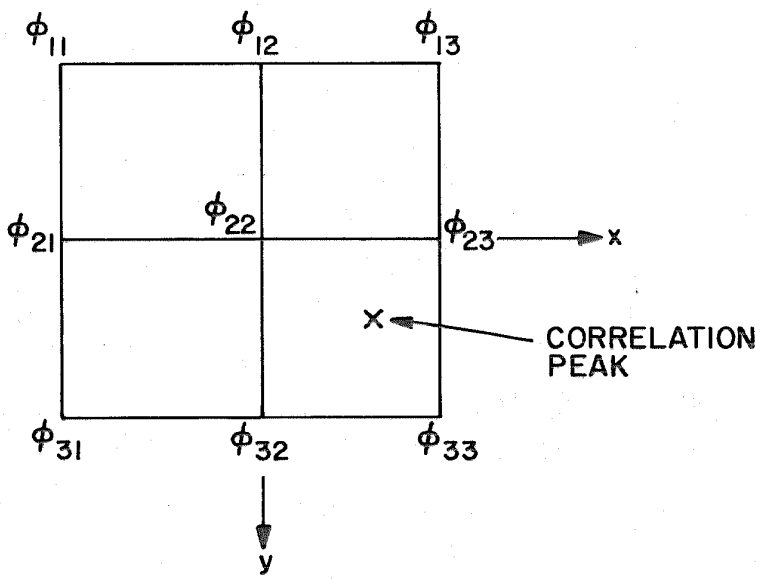


Fig. 2-4 Diagram of the CCF matrix used to determine the coordinates of the co-relation peak

amount equal to the fractional peak displacements, (x_p, y_p). This requires one to calculate the pixel values at intermediate grid positions by interpolation. If we let $\alpha = x_p/\Delta$ and $\beta = y_p/\Delta$, then Eq. (2.32) yields the transformation formula for the case of a bilinear interpolation.

E) Subtraction of images

A statistical analysis of the difference picture between two images provides additional information on the correctness of the registration. For example, let $d_1(x,y)$ and $d_2(x,y)$ be two noisy representations of the same object, $d_0(x,y)$, which is assumed to have white Gaussian noise with zero mean. In order to simplify the analysis, we will assume that a contrast stretch has been performed on the input data sets to equalize their mean values. Then, if these pictures are slightly mismatched with respect to each other in the x-direction, we may represent them by

$$d_1(x,y) = d_0(x,y) + n_1(x,y)$$
$$d_2(x,y) = d_0(x+\Delta x,y) + n_2(x,y) \qquad (2.71)$$

where Δx is the mismatch parameter. By forming the difference image, $d_s(x,y)$, we obtain

$$d_s(x,y) = d_2(x,y) - d_1(x,y) = \Delta x\, d_0'(x,y) + n(x,y) \qquad (2.72)$$

where $d_0'(x,y)$ is the derivative object picture and $n(x,y)$ is an uncorrelated Gaussian noise distribution with variance $\sqrt{\sigma_1^2 + \sigma_2^2}$. We will now examine how the parameter Δx is linked to the histogram

$H_{d_s}(d_s)$ of the difference image.

It can be shown that the probability density of a summation signal is equal to the convolution product of the probability densities of its component signals, when the component signals are mutually uncorrelated[26]. The histogram can then be obtained from Eq. (2.72) and is given by

$$H_{d_s}(d_s) = \frac{1}{\Delta x} H_{d_0'}\left(\frac{d_0'}{\Delta x}\right) * H_n(n) \tag{2.73}$$

where $H_{d_0'}(d_0')$ is the histogram of the derivation picture, and $H_n(n)$ is the histogram of the noise distribution. Most derivative pictures have a symmetrical Gaussian histogram centered at the origin, so that we may write

$$H_{d_0'}(d_0') \propto \exp[-d_0'^2/2\sigma_0^2]$$

$$H_n(n) \propto \exp[-n^2/2(\sigma_1^2 + \sigma_2^2)] \tag{2.74}$$

where σ_0 is a characteristic of the object. Consequently, from Eq. (2.73) we see that $H_{d_s}(d_s)$ is a Gaussian histogram with half-width

$$\sigma_s = [\sigma_1^2 + \sigma_2^2 + (\sigma_0\ \Delta x)^2]^{1/2} \tag{2.75}$$

Equation (2.75) implies that the narrower the histogram of the difference picture, the better the registration.

We have discussed correlation techniques in some detail; since they are an essential step in the image processing schemes described later in this work, and if carelessly applied they can introduce

systematic errors into the processed pictures.

2.3.5 Spatial Frequency Filtering and Fourier Transforms

The purpose of a variety of picture processing operations is to extract the information content of an image in the presence of degradation[47,48,49]. Digital filtering is a computer enhancement technique that brings out selected features of an image by: suppression of random noise and background shading, amplification of fine detail, and image restoration.

2.3.5.1 Design of linear non-recursive filters

A) Definition

A linear non-recursive filter is a position-invariant operator whose output is a weighted superposition of input image points[33]. We can also describe the output by a convolution product between the filter impulse response $h(\underset{\sim}{r})$ and the input picture $d(\underset{\sim}{r})$:

$$d_f(\underset{\sim}{r}) = h * d \tag{2.76}$$

A discrete approximation for this convolution integral is

$$d^f_{k,\ell} = \sum_{i=-\frac{I-1}{2}}^{\frac{I-1}{2}} \sum_{j=-\frac{J-1}{2}}^{\frac{J-1}{2}} h_{i,j}\, d_{k-i,\ell-j} \tag{2.77}$$

where the convolution kernel, $(h_{i,j})$, is defined by an $I \times J$ matrix. The matrix $(h_{i,j})$ is a correct representation of the continuous filter response when sampling artifacts are minimized. We must therefore select a sampling step which does not destroy the band-limitedness of the input data. In order to reduce the ringing phenomenon introduced

by the filter, we can smooth the edge discontinuities at the truncation by multiplying $h(\underset{\sim}{r})$ with a window function $w(\underset{\sim}{r})$. Often the two-dimensional window is chosen to be a circularly symmetric Gaussian fall-off described by

$$w(r) = \begin{cases} 1 & \text{if } r < r_G \\ \exp[-(r-r_G)^2/2\sigma_G^2] & \text{if } r_G < r < \text{Min}(X_f, Y_f) \end{cases} \tag{2.78}$$

where r_G and σ_G are adjustable parameters for the window, X_f and Y_f are the lateral dimensions of the truncation box.

Linear filtering can also be implemented in Fourier space. Filtering in this domain amounts to multiplying the Fourier transforms of the image, $D(\underset{\sim}{k})$, by the filter, $H(\underset{\sim}{k})$:

$$D_f(\underset{\sim}{k}) = H(\underset{\sim}{k}) \cdot D(\underset{\sim}{k}) \tag{2.79}$$

where $D_f(k)$ is the filtered transform. The processed output is then obtained by inverse transforming Eq. (2.79).

B) Discrete Fourier transform (DFT)

The advent of the FFT algorithm enables us to perform numerical operations, such as DFT of data samples, in a reasonable amount of time[44]. When this algorithm is used to evaluate Fourier integrals, it is advisable to examine the relationship of the continuous integral transform to the DFT[45].

In order to demonstrate the relationship between the integral transform and the DFT, let $d(x,y)$ be a continuous density distribution and $d_{k,\ell} = d(k\Delta,\ell\Delta)$ be the

corresponding sequence of Nyquist samples (i.e., $\Delta = 1/2k_m$). Then if $d(\underset{\sim}{r})$ and $D(\underset{\sim}{k})$ are a Fourier integral transform pair, it can be shown that $d^p_{k,\ell}$, [k=0,1,···,K-1; ℓ=0,1,···,L-1], and $D^p_{m,n}$, [m=0,1,···,K-1; n=0,1,···,L-1], are a DFT pair given by Eq. (2.65), with

$$(K-1)\Delta = 1/\delta k_x = X$$

$$(L-1)\Delta = 1/\delta k_y = Y$$

$$d^p_{k,\ell} = d^p(k\Delta,\ell\Delta) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} d(k\Delta + iX, \ell\Delta + jY)$$

$$D^p_{m,n} = D^p(m\,\delta k_x, n\delta k_y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} D(m\,\delta k_x + i/\Delta, n\delta k_y + j/\Delta)$$

The periodic function, $d^p(\underset{\sim}{r})$, is formed by the superposition of $d(\underset{\sim}{r})$ shifted by all multiples of the fundamental period in x and y. If we suppose that Δ is chosen so that the aliasing involved in constructing $D^p(\underset{\sim}{k})$ is negligible (i.e., $|D(\underset{\sim}{k})| \simeq 0$ for $|\underset{\sim}{k}| > k_m = 1/2\Delta$), then the continuous transform may be evaluated from the following identity:

$$D(k_x,k_y) = D^p(k'_x,k'_y) \tag{2.80}$$

where

$$k_{\substack{x\\y}} = \begin{cases} k'_{\substack{x\\y}} & \text{if} \quad 0 \le k'_{\substack{x\\y}} \le k_m \\ k'_{\substack{x\\y}} - 2k_m & \text{if} \quad k_m \le k'_{\substack{x\\y}} \le 2k_m \end{cases}$$

Consequently, the error in approximating $D(\underset{\sim}{k})$ by $D^p(\underset{\sim}{k})$ in the range $|k_x| \le k_m$ and $|k_y| \le k_m$ results from the overlapping of adjacent frequency spectra. The truncation of $d(\underset{\sim}{r})$ to a finite input size allows

the infinite summation involved in obtaining $d^p(\underset{\sim}{r})$ to be computed. This truncation introduces side lobes in the estimated spectrum because the box transform is convolved with $D^p(\underset{\sim}{k})$. Consequently, the DFT show streaks, which can be removed by apodizing the sampled input with a window function that vanishes at the picture boundaries (i.e., a Gaussian fall-off).

Once the sampling step Δ has been chosen so as to satisfy the Nyquist criterion, the picture size to be transformed is then selected (K and L are restricted to be powers of two by the FFT algorithm). The sampling step determines the intervals δk_x and δk_y, at which one samples $D^p(\underset{\sim}{k})$. Then with these choices of K and L , we form the aliased sequence $d^p_{k,\ell}$, and apply the algorithm.

Most properties of the DFT are in agreement with the corresponding properties of the Fourier integral transform; one such property is the convolution relationship,

$$d^{p\ f}_{k,\ell} = \sum_{i=-\frac{I-1}{2}}^{\frac{I-1}{2}} \sum_{j=-\frac{J-1}{2}}^{\frac{J-1}{2}} h^p_{i,j}\ d^p_{k-i,\ell-j}$$

which becomes

$$D^{p\ f}_{m,n} = H^p_{m,n} \cdot D^p_{m,n} \tag{2.81}$$

Equations (2.81) are valid everywhere, except at the picture boundaries where they introduce a wrap-around error. However, their validity is preserved if the following identities are correct:

$$d_{k,\ell} = d_{k+K,\ell} \qquad \text{for } 0 \le k \le \frac{I-1}{2} \text{ and } K-\frac{I-1}{2} \le k \le K$$

$$d_{k,\ell} = d_{k,\ell+L} \qquad \text{for } 0 \le \ell \le \frac{J-1}{2} \text{ and } L-\frac{J-1}{2} \le \ell \le L$$

If the above conditions are met, then the filter output is

$$d^f_{k,\ell} = \frac{1}{L} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} H^p_{m,n} D^p_{m,n} \exp[2\pi i(\frac{km}{K} + \frac{\ell n}{L})] \qquad (2.82)$$

2.3.5.2 Spatial filtering applications

A) Low-pass filters

Los-pass filters are used to reduce the effect of random picture noise, which can obscure large low-contrast features, and to remove unwanted high-frequency structures such as sharp edges[34,35,50]. These filters smooth the picture by replacing each image point with a weighted neighbor average. In the case of a box filter, these weights are equal to

$$h^{LP}_{i,j} = \frac{1}{I \times J} \qquad (2.83)$$

and the corresponding filter transfer function is given by

$$H^{LP}(k_x,k_y) = \frac{\sin[\pi\Delta(I-1)k_x]}{\pi k_x} \cdot \frac{\sin[\pi\Delta(J-1)k_y]}{\pi k_y} \qquad (2.84)$$

This kernel generates undesirable sidelobes in the discrete filter transform for $|k_x| > \delta k_x(K-1)/(I-1)$ and for $|k_y| > \delta k_y(L-1)/(J-1)$ (see Eq. (2.84)). In order to reduce the ringing effect caused by sharp discontinuities in weight values at the truncation, one can smooth the boundary by applying a polynomial fit to the truncated sequence of the weights.

Low-pass filters can also be implemented in Fourier space, where high-spatial frequencies corresponding to noise can be attenuated.

Here a selective enhancement of the signal-to-noise ratio is achieved by filtering out the frequencies beyond the Nyquist frequency of the input data. For example, a low-pass isotropic filter can be expressed as:

$$H^{LP}(\underset{\sim}{k}) = \begin{cases} \exp[-(|\underset{\sim}{k}| - k_m)^2/2\sigma_{LP}^2] & \text{if } |\underset{\sim}{k}| > k_m \\ 1 & \text{if } |\underset{\sim}{k}| \leq k_m \end{cases} \tag{2.85}$$

So far we have restricted our discussion to linear filters. We note that an example of a nonlinear low-pass filter was described in Section 2.3.3.2. This filter applies a local averaging, with thresholding, in order to suppress "salt and pepper" noise.

B) High-pass filters

A major use for high-pass filters is to reveal small low-contrast features that are superimposed on uniform backgrounds[34,35]. These filters suppress undesirable large-scale intensity variations that are composed mostly of low-spatial-frequency components. A high-pass filter can be specified in terms of its low-pass equivalent by writing

$$H^{HP}(k) = 1 - H^{LP}(k) \tag{2.86}$$

If the desired high-pass filter is the complement of a known low-pass filter, an obvious way to construct such a filter is to build its sequence of weights by letting

$$h^{HP}_{i,j} = \delta_{i,j} - h^{LP}_{i,j}$$

where $h^{P}_{i,j}$ is the low-pass kernel matrix.

A filter which partially suppresses the lower frequencies while enhancing the higher frequencies is called a high-emphasis filter. It can be formed from a high-pass filter by multiplying every value of $H^{HP}(\underset{\sim}{k})$ by a constant, ζ^{HE}, and then shifting the DC component:

$$H^{HE}(\underset{\sim}{k}) = \zeta^{HE} H^{HP}(\underset{\sim}{k}) + 1 \tag{2.87}$$

where $\zeta^{HE} > 1$. Consequently, if $(h^{HP}_{i,j})$ is a high-pass-filter matrix, then we can build a high-emphasis filter by letting

$$h^{HE}_{i,j} = \zeta^{HE} h^{HP}_{i,j} + \delta_{i,j}$$

This filter creates dark bands surrounding the sharper high-contrast-ratio edges in an image (i.e., it overemphasizes the edges thereby reinforcing the Mack phenomenon[40]).

C) Restoration filters

i) Introduction

Imaging systems degrade the information content of an object by distorting phase, amplitude, and by superimposing noise onto its recorded image. For an incoherently illuminated imaging system that obeys the isoplanatic approximation, the image formation process may be described by a linear mapping of the object distribution, $d_o(\underset{\sim}{r})$, onto an image, $d_i(\underset{\sim}{r})$:

$$d_i(\underset{\sim}{r}) = t_s * d_o + n(\underset{\sim}{r}) \tag{2.88}$$

where $t_s(\underset{\sim}{r})$ is the incoherent point-spread function, and $n(\underset{\sim}{r})$ is

additive noise. An indication of the performance of the system is provided by the optical transfer function (OTF), which assesses the frequency response of an imaging device. When weakly scattering objects are observed with an electron microscope (assuming partially coherent illumination and the isoplanatic approximation), the image properties can be described by an OTF called the contrast transfer function.

If the OTF is known (e.g., by calibration measurements, or analytically as a function of such parameters as defocusing, astigmatism, and image motion), it is possible to restore the object spectrum by image processing techniques.

We shall now examine the procedures used to extract the object information from an image which is formed by a noise-free system. Then, we shall describe image restoration schemes for systems which contain random noise that cannot be separated from the image[47,48,49,51].

ii) Noise-free systems

Let us call $d_{NF}(\underset{\sim}{r})$ the degraded image; then, assuming that the noise is negligible, we may describe the imaging process by

$$d_{NF}(\underset{\sim}{r}) = t_s * d_o \tag{2.89}$$

An equivalent description can be obtained by Fourier transforming Eq. (2.89):

$$D_{NF}(\underset{\sim}{k}) = T_s(\underset{\sim}{k}) \cdot D_o(\underset{\sim}{k}) \tag{2.90}$$

where $T_s(\underset{\sim}{k})$ is the imaging OTF.

An obvious compensating filter for this system is

$$H_{NF}(\underset{\sim}{k}) = 1/T_s(\underset{\sim}{k}) \tag{2.91}$$

which results in a restored image, $h * d_{NF} = d_o(\underset{\sim}{r})$, that is an exact representation of the object.

In practice, the transfer function may have zeros in the frequency ranges that are of interest. At the spatial frequencies where $T_s(\underset{\sim}{k})$ vanishes, the restored object spectrum is undetermined and the filter, $H_{NF}(\underset{\sim}{k})$, would take infinite values. In the design of the compensating filter we must also consider the noise, which always degrades an image. By using an inverse filter such as $H_{NF}(\underset{\sim}{k})$, we would amplify the noise in the restored image at the zeros of the transfer function, and destroy the validity of the restoration.

We shall now describe two alternate restoration procedures that can be implemented in the case of noisy systems.

iii) Wiener optimum filter

Consider a signal $d_i(\underset{\sim}{r})$ that is the superposition of a noise-free degraded image $d_{NF}(\underset{\sim}{r})$ with some statistically uncorrelated noise. We know from Section 2.3.3.3C that an optimum filter minimizing the mean-square error,

$$E_n = \frac{1}{A} \int_A [d_f(\underset{\sim}{r}) - d_{NF}(\underset{\sim}{r})]^2 \, d\underset{\sim}{r}$$

where $d_f(\underset{\sim}{r})$ is the filtered image, has the frequency response

$$H_W(\underset{\sim}{k}) = \frac{|D_{NF}(\underset{\sim}{k})|^2}{|D_{NF}(\underset{\sim}{k})|^2 + |N(\underset{\sim}{k})|^2} = \frac{|T_s(\underset{\sim}{k})|^2 \cdot |D_o(\underset{\sim}{k})|^2}{|T_s(\underset{\sim}{k})|^2 \cdot |D_o(\underset{\sim}{k})|^2 + |N(\underset{\sim}{k})|^2} \quad (2.92)$$

We can express the filtered image $d_f(\underset{\sim}{r})$ as

$$d_f(\underset{\sim}{r}) = h_W * d_i = h_W * (t_s * d_o) + h_W * n \quad (2.93)$$

Now since, $d_{NF}(\underset{\sim}{r}) = t_s * d_o$, the optimum filter $H_{WD}(\underset{\sim}{k})$ yielding the least mean-square discrepancy between the restored image and the object, is derived from $H_W(\underset{\sim}{k})$ by dividing by $T_s(\underset{\sim}{k})$, that is,

$$H_{WD}(\underset{\sim}{k}) = \frac{H_W(\underset{\sim}{k})}{T_s(\underset{\sim}{k})} = \frac{T_s^*(\underset{\sim}{k}) \cdot |D_o(\underset{\sim}{k})|^2}{|T_s(\underset{\sim}{k})|^2 \cdot |D_o(\underset{\sim}{k})|^2 + |N(\underset{\sim}{k})|^2} \quad (2.94)$$

The quality of this restoration scheme may be assessed by computing the resulting minimum of the mean-square error,

$$E_{min} = \frac{1}{B} \int_B \frac{|D_o(\underset{\sim}{k})|^2 \cdot |N(\underset{\sim}{k})|^2}{|T_s(\underset{\sim}{k})|^2 \cdot |D_o(\underset{\sim}{k})|^2 + |N(\underset{\sim}{k})|^2} \, d\underset{\sim}{k} \quad (2.95)$$

where B is the spatial-frequency domain over which the restoration is performed.

Practical implementation of the optimum filter requires a priori knowledge of the noise statistics and the object structure[52]. Electron micrographs have, for instance, been restored by assuming a white Gaussian noise and a Gaussian probability distribution for the object spectrum[41].

iv) Modified inverse filter

We saw in Sec. 2.3.5.2C.ii that noise amplification at the zeroes of $T_s(\underset{\sim}{k})$ jeopardize the restoration efficiency of the inverse filter $H_{NF}(\underset{\sim}{k})$[47]. In order to avoid these artifacts, we can modify this filter by a weight function, $g(\underset{\sim}{k})$, having the same zeroes as $T_s(\underset{\sim}{k})$ so that the modified filter

$$H_{mi}(\underset{\sim}{k}) = \frac{g(\underset{\sim}{k})}{T_s(\underset{\sim}{k})} \tag{2.96}$$

remains finite. We can determine the correct weight function by studying the signal-to-noise ratio before and after restoration.

According to Parseval's theorem, the variance of the filtered noise distribution $n'(\underset{\sim}{r}) = h_{mi} * n$, can be written as

$$\sigma_{n'} = \frac{1}{B}\int_B |N'(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} = \frac{1}{B}\int_B |H_{mi}(\underset{\sim}{k})|^2 |N(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} \tag{2.97}$$

where B describes the frequency range of the filter, and $N'(\underset{\sim}{k})$ is the noise spectrum in the filtered image. Since the filter and the noise power spectra are uncorrelated, we may approximate Eq. (2.97) by:

$$\sigma_{n'} \simeq \frac{1}{B}\int_B |H_{mi}(\underset{\sim}{k})|^2 d\underset{\sim}{k} \cdot \frac{1}{B}\int |N(\underset{\sim}{k})|^2 d\underset{\sim}{k} = \sigma_n \cdot \frac{1}{B}\int_B |H_{mi}(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} \tag{2.98}$$

The noise amplification factor

$$f_n = \sigma_{n'}/\sigma_n \simeq \frac{1}{B}\int_B |H_{mi}(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} \tag{2.99}$$

is a critical parameter in the design of a filter. An esimate of f_n

as a function of both B and $g(\underset{\sim}{k})$, provides a useful guide for the determination of an adequate inverse filter. For instance, in Eq. (2.99) we see that f_n will blow up at the zeroes of $T_s(\underset{\sim}{k})$ if we set $H_{mi}(\underset{\sim}{k}) = 1/T_s(\underset{\sim}{k})$.

We will now consider the changes in the variance of the noise-free degraded image, $d_{NF}(\underset{\sim}{r})$, that were introduced during the restoration process. The variance of $d_{NF}(\underset{\sim}{r})$ before the reconstruction can be expressed by

$$\sigma_{d_{NF}} = \sigma_o \cdot \frac{1}{B} \int_B |T_s(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} \tag{2.100}$$

where the variance of the original object distribution is given by

$$\sigma_o = \frac{1}{B} \int_B |D_o(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}$$

After filtering, $\sigma_{d_{NF}}$ is converted into

$$\sigma'_{d_{NF}} = \sigma_o \cdot \frac{1}{B} \int_B |T_s(\underset{\sim}{k})|^2 \cdot |H_{mi}(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} \tag{2.101}$$

where $\sigma'_{d_{NF}}$ measures the variance of the degraded object distribution, which has been altered by the reconstruction process. Finally, let us define a quantity f_d that correlates the variances of signal distributions before and after filtering. This parameter, which is called the relative structural content, is defined by

$$f_d = \sigma'_{d_{NF}}/\sigma_{d_{NF}} = \frac{\frac{1}{B}\int_B |T_s(\underset{\sim}{k})|^2 \cdot |H_{mi}(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}}{\frac{1}{B}\int_B |T_s(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}} \tag{2.102}$$

A comparison of $\sigma'_{d_{NF}}$ with σ_o provides a quality measure for the restoration. When these parameters are approximately equal, a correlation between the object and the restored image is assured. Normally, $\sigma_{d_{NF}}$ is less than σ_o, hence a figure of merit for the reconstruction is provided by f_d. This parameter should be larger than unity, if an improvement in the signal is attained.

Another parameter to consider before filtering an image is τ_r, the estimate of the quotient f_d/f_n. This parameter is an estimate of the improvement in the signal-to-noise ratio when the filter is applied to an image. We can write τ_r as

$$\tau_r = \frac{\sigma_{d'_{NF}}/\sigma_{n'}}{\sigma_{d_{NF}}/\sigma_n} = \frac{f_d}{f_n} = \frac{\frac{1}{B}\int_B |H_{mi}(\underset{\sim}{k})|^2 \, |T_s(\underset{\sim}{k})|^2 d\underset{\sim}{k}}{\frac{1}{B}\int_B |H_{mi}(\underset{\sim}{k})|^2 d\underset{\sim}{k} \cdot \frac{1}{B}\int |T_s(\underset{\sim}{k})|^2 d\underset{\sim}{k}} \tag{2.103}$$

The quantities τ_r, f_n and f_d are critical parameters in the choice of an appropriate weight function $g(\underset{\sim}{k})$. For instance, by using Eq. (2.96), we may evaluate τ_r for a particular weight function, from the identity:

$$\tau_r = \frac{\frac{1}{B}\int_B |g(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}}{\frac{1}{B}\int_B \left| \frac{g(\underset{\sim}{k})}{T_s(\underset{\sim}{k})} \right|^2 d\underset{\sim}{k} \cdot \frac{1}{B}\int_B |T_s(\underset{\sim}{k})|^2 d\underset{\sim}{k}} \tag{2.104}$$

A simple heuristic method for constructing a modified compensating filter is demonstrated in the following discussion. One prevents the inverse filter from exceeding a given threshold, $1/\overline{\delta}$, by using the weight function

$$g(\underset{\sim}{k}) = \begin{cases} 1 & \text{for } |T_s(\underset{\sim}{k})| > \overline{\delta} \\ |T_s(\underset{\sim}{k})| & \text{otherwise} \end{cases} \tag{2.105}$$

where $\overline{\delta}$ is an empirically chosen truncation factor. The initial signal-to-noise ratio, τ, of the picture to be processed determines the optimum value for $\overline{\delta}$. If τ assumes values close to unity in certain regions, one should select a threshold value large enough to minimize the noise amplification resulting from the filtering.

For instance, in order to illustrate the design of a filter, let us call $B_{\overline{\delta}}$ the subset of B where $|T_s(\underset{\sim}{k})|$ exceeds a lower bound $\overline{\delta}$ ($B_{\overline{\delta}}$ may be disjoint for an oscillating transfer function). By using this notation the noise and signal amplification factors that are given by Eqs. (2.99) and (2.102) become

$$f_n = \frac{1}{B} \int_{B_{\overline{\delta}}} \frac{1}{|T_s(\underset{\sim}{k})|^2} \, d\underset{\sim}{k} + 1 - \frac{B_{\overline{\delta}}}{B} \tag{2.106a}$$

$$f_d = \frac{1}{\frac{1}{B} \int_B |T_s(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}} \left[\frac{B_{\overline{\delta}}}{B} + \frac{1}{B} \int_{B - B_{\overline{\delta}}} |T_s(\underset{\sim}{k})|^2 \, d\underset{\sim}{k} \right] \tag{2.106b}$$

For an electron microscope f_n and f_d can be plotted against the objective aperture angle (i.e., the Fourier domain B), and against the threshold $\overline{\delta}$. Linfoot's image evaluation criteria can also be

computed for the filter, since they give a quantitative assessment of the validity of the reconstruction (see Sec. 2.3.6.2).

D) Feature-selective filters

Edges or lines in a particular direction can be emphasized by applying feature-selective filters to an image[35]. Some practical operators, which selectively enhance image features, will now be presented.

i) Directional derivatives

Differential operators extract edges by cross-correlating the picture with a matrix of weights that geometrically resembles the features to be enhanced. For example, in order to extract lines at a slope $\overline{\nu}$ in the image, one can convolve the picture with a template having a positive ridge centered on the line $y = \overline{\nu}x$, with a negative valley on each side of the ridges. The spatial invariance of the filter requires the weight values to be chosen in such a way that their summation over the template vanishes.

ii) Gradient operator

One can deblur a picture in every angular orientation by taking the derivative at each point in the direction that the grey level changes the fastest. This nonlinear operator extracts edges by calling any point an edge point, if the local gradient exceeds a given threshold.

An approximate representation of an isotropic gradient filter, for discrete data, is given by the finite differences expression for a squared gradient:

$$|\nabla d|^2_{i,j} \simeq (d_{i+1,j+1} - d_{i,j})^2 + (d_{i+1,j} - d_{i,j+1})^2 \qquad (2.107)$$

This expression is not symmetrical, and therefore partially destroys the rotational invariance of the operator. More symmetrical expressions for the gradient operator can be devised when the input data require their use.

iii) Laplacian operator

The Laplacian operator is a matched filter that detects objects on an edge criterion, since it outputs peaks at image points where the second derivative is non-zero. For a digital picture, we can approximate the Laplacian by:

$$(\nabla^2 d)_{i,j} \simeq (d_{i+i,j} + d_{i,j+1} + d_{i-1,j} + d_{i,j-1}) - 4d_{i,j} \qquad (2.108)$$

The Laplacian operator is therefore equivalent to a convolution filter kernel whose matrix is

$$\begin{Vmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{Vmatrix}$$

The output of this operator is proportional to the difference between the grey level at a point and the average grey level in an annulus centered at the point.

2.3.6 Image Evaluation

The information capacity of an image has limited use since its assessment depends on an a priori knowledge of the noise and signal statistics[43]. However, various criteria have been proposed, based on

linear filtering theory to evaluate the quality of optical imaging systems[53,54,55,56]. In this section we will assume that the image, $d_i(\underset{\sim}{r})$, may be related to the object by

$$d_i(r) = t_s * d_o + n(r)$$

where $t_s(\underset{\sim}{r})$ is the system impulse response, and $n(\underset{\sim}{r})$ is signal-independent additive noise.

2.3.6.1 Mean square error criterion

The mean square error criterion measures the deviation of a channel output from its corresponding input signal. It yields a statistical estimate of the variance of the difference distribution

$$E = \frac{1}{A}\int_A [d_i(\underset{\sim}{r}) - d_o(\underset{\sim}{r})]^2 \, d\underset{\sim}{r}$$

where A is the area of the image. The mean-square error can also be expressed in terms of the signal and the noise power spectra:

$$E = \frac{1}{B}\int_B [|D_o(\underset{\sim}{k})|^2 \cdot (|T_s(\underset{\sim}{k})|^2 - 1) + |N(\underset{\sim}{k})|^2] \, dk \qquad (2.109)$$

Minimization of this error is an often used criterion for decoding a received signal, and it generates a family of optimum filters.

2.3.6.2 Linfoot's figures of merit

It is often desirable to characterize an imaging system in a manner that is independent of its noise properties. Criteria have therefore been generated in order to describe the properties of the image point-spread function of the optical system when the noise spectrum

is unknown[55]. These criteria, called Linfoot's figures of merit, indicate how the image formation process alters the resemblance between an object and its corresponding noise-free image.

By analogy with the mean-square error criterion, we may introduce a parameter which measures the closeness of the correspondence between an object and its noise-free image. The only deviations which are taken into consideration originate from the system transfer function. These deviations are specified by the normalized mean-square discrepancy, called the fidelity, which is defined by

$$\Phi_r = 1 - \frac{\frac{1}{A}\int_A [d_{NF}(\underset{\sim}{r}) - d_o(\underset{\sim}{r})]^2 \, d\underset{\sim}{r}}{\frac{1}{A}\int_A |d_o(\underset{\sim}{r})|^2 \, d\underset{\sim}{r}} = \frac{\frac{1}{B}\int_B |D_o(\underset{\sim}{k})|^2 \cdot [2T_s(\underset{\sim}{k}) - |T_s(\underset{\sim}{k})|^2] \, d\underset{\sim}{k}}{\frac{1}{B}\int_B |D_o(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}} \qquad (2.110)$$

We have shown in Sec. 2.3.4.2 that the CCF of two density distributions could be regarded as an inference estimator. The CCF integral may therefore yield a correlation measure between object and image. For a noise-free system, the normalized CCF is named the correlation quality, and is expressed as:

$$\Psi_r = \frac{\frac{1}{A}\int_A d_{NF}(\underset{\sim}{r}) \, d_o(\underset{\sim}{r}) d\underset{\sim}{r}}{\frac{1}{A}\int_A |d_o(\underset{\sim}{r})|^2 \, d\underset{\sim}{r}} = \frac{\frac{1}{B}\int_B |D_o(\underset{\sim}{k})|^2 \; T_s(\underset{\sim}{k}) \, d\underset{\sim}{k}}{\frac{1}{B}\int_B |D_o(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}} \qquad (2.111)$$

Finally, we can describe how the imaging device blurs the object contrast by damping the intensity fluctuations in the object distribution. This contrast attenuation can be described by

$$\Xi_r = \frac{\frac{1}{A}\int_A |d_{NF}(\underset{\sim}{r})|^2 \, d\underset{\sim}{r}}{\frac{1}{A}\int_A |d_o(\underset{\sim}{r})|^2 \, d\underset{\sim}{r}} = \frac{\frac{1}{B}\int_B |D_o(\underset{\sim}{k})|^2 \; |T_s(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}}{\frac{1}{B}\int_B |D_o(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}} \qquad (2.112)$$

where Ξ_r is called the relative structural constant. This parameter measures the image power relative to the object power and is linked to Φ_r and Ψ_r by the relationship:

$$\Xi_r = 2\Psi_r - \Phi_r$$

2.3.6.3 Signal-to-noise ratio

The signal-to-noise ratio is a measure of the detectability of a signal that has been contaminated by additive noise. In the general case where the system transfer function distorts the object representation, the signal-to-noise ratio, τ , is defined by:

$$\tau = \frac{\frac{1}{A}\int_A |d_{NF}(\underset{\sim}{r})|^2 \, d\underset{\sim}{r}}{\frac{1}{A}\int_A |n(\underset{\sim}{r})|^2 \, d\underset{\sim}{r}} = \frac{\frac{1}{B}\int_B |D_o(\underset{\sim}{k})|^2 \; |T_s(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}}{\frac{1}{B}\int_B |N(\underset{\sim}{k})|^2 \, d\underset{\sim}{k}} \qquad (2.113)$$

An evaluation of τ from the recorded image provides an indication of the quality of the image. However, the most popular indicator for image quality, when designing optical systems, is the mean square error

criterion. It agrees reasonably well with subjective evaluation in many cases and gives a quantitative assessment of an image restoration procedure.

CHAPTER THREE

FOURIER PROCESSING OF BRIGHT-FIELD ELECTRON MICROGRAPHS OF WEAKLY SCATTERING OBJECTS

CHAPTER THREE

FOURIER PROCESSING OF BRIGHT-FIELD ELECTRON MICROGRAPHS OF WEAKLY SCATTERING OBJECTS

3.1 Introduction

An electron microscope can theoretically operate as a diffractometer that is capable of performing structure analysis at atomic resolution. However, due to objective lens aberrations which distort the phases of the diffracted waves, the ultimate resolution of the instrument is limited. It is possible to perform a posteriori corrections to electron micrographs in order to regain lost resolution if the amplitudes and phases of the scattered electron waves are known. The phase information can be retrieved, as in holography, by using a reference wave. In electron microscopy, bright-field images of weakly scattering specimens can be viewed as in line holography[14,21].

A theory of image formation has been developed for weak phase and amplitude objects viewed in bright field; in this theory the image intensity is linearly related to the object phase and amplitude modulations. As in the formation of a single-atom image (see Sec. 1.2.2), contrast transfer functions govern the contrast mechanisms and the ultimate resolution of the instrument. Various reconstruction schemes, based on image filtering theory, have been devised to retrieve the complex object transmittance and to enhance the resolution by correcting the pupil function. One method recombines micrographs, taken at different defocusing values by computer processing[57]. Another scheme uses

complementary semicircular apertures (cf. single side-band holography) and processes the picture optically[58]. Using these techniques it is possible to make both a full object restoration and a heavy/light atom discrimination at the same time.

In this chapter we shall present the formalism of the linear theory of image formation for weakly scattering objects. Then we shall describe the preliminary analysis of optical diffractograms which yields the experimental parameters used in the image restoration programs described in this chapter. A heavy/light atom discrimination technique is then discussed, and its potential is then examined experimentally[59]. Next, the validity of using linear transfer theory to interpret the contrast from periodic objects, such as bovine liver catalase, is considered[60]. A plot of the overall transfer function is obtained from a through-focus series of catalase, and then a partial image reconstruction is performed[61]. Finally, a study of the effectiveness of the selective contrast enhancement method for heavy atoms is presented, and side-effect contributions are discussed.

3.2 Linear Theory of Image Formation for Weak-Phase and Amplitude Objects

In Eq. (1.56) we have seen that the image wave amplitude is linked to the object wave function at the exit surface of the specimen by a convolution integral. When the specimen obeys the single slice approximation, we can write this equation as

$$\psi(\underset{\sim}{r}_i) = A_o \int_{A_{obt}} \Lambda_t(\underset{\sim}{r}_o)\, G(\underset{\sim}{r}_i - \underset{\sim}{r}_o)\, d\underset{\sim}{r}_o \qquad (3.1)$$

where $G(\underset{\sim}{r})$ is the microscope impulse response for coherent illumination

in the isoplanatic case. The central problem of electron microscopy is to invert this equation in order to recover information that is contained in $\Lambda_t(\underset{\sim}{r}_o)$ about the specimen structure. Experimentally, we measure the image intensity, $|\psi(\underset{\sim}{r}_i)|^2$, so that only $|\psi(\underset{\sim}{r}_i)|$ is known. For weakly scattering specimens we know that the transmission function is given by Eq. (1.43), so that we can write

$$\Lambda_t(\underset{\sim}{r}_o) = 1 + i\phi(\underset{\sim}{r}_o) - \mu(\underset{\sim}{r}_o)$$

$$\text{for} \quad |\mu(\underset{\sim}{r}_o)| < |\phi(\underset{\sim}{r}_o) << 1 \tag{3.2}$$

We shall now show that it is possible to evaluate $\Lambda_t(\underset{\sim}{r}_o)$ from $|\psi(\underset{\sim}{r}_i)|$. In order to reconstruct the phase distribution, $\phi(\underset{\sim}{r}_o)$, and the amplitude distribution, $\mu(\underset{\sim}{r}_o)$, we shall assume in the following analysis that the object transmission function satisfies Eq. (3.2) to first order.

3.2.1 Contrast Transfer Function of the Microscope

3.2.1.1 Basic equations

We obtain the object spectrum by transforming Eq. (3.2):

$$S_o(\underset{\sim}{k}) = \mathcal{F}\{A_o\Lambda_t(\underset{\sim}{r}_o)\} = A_o[\delta(\underset{\sim}{k}) + i\Phi(\underset{\sim}{k}) - M(\underset{\sim}{k})] \tag{3.3}$$

The image wave function is then given by the inverse transform of $S_o(\underset{\sim}{k})$ phase shifted by the pupil function $P(\underset{\sim}{k})$:

$$\psi(\underset{\sim}{r}_i) = \frac{1}{M}\mathcal{F}^{-1}\{P(\underset{\sim}{k})\ S_o(\underset{\sim}{k})\} \tag{3.4}$$

We will assume that the microscope operates in the bright-field mode

with the objective aperture axially centered on the optic axis, so that the objective aperture function satisfies the symmetry rule:

$$b(\underset{\sim}{k}) = b(-\underset{\sim}{k}) = \begin{cases} 1 & \text{within the aperture} \\ 0 & \text{outside} \end{cases} \tag{3.5}$$

Combining Eq. (3.3) and Eq. (3.4) leads to the following expressions for the image function:

$$\begin{aligned} \psi(\underset{\sim}{r}_i) &= \frac{A_o}{M} + \frac{A_o}{M} \int [i\Phi(\underset{\sim}{k}) - M(\underset{\sim}{k})]\, b(\underset{\sim}{k}) \exp[-i\gamma(\underset{\sim}{k}) + 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i]\, d\underset{\sim}{k} \\ &= \frac{A_o}{M} + \psi_{diff}(\underset{\sim}{r}_i) \end{aligned} \tag{3.6}$$

where $\psi_{diff}(\underset{\sim}{r}_i)$ is the elastically scattered wave at the image plane. The recorded image intensity can then be expressed as

$$|\psi(\underset{\sim}{r}_i)|^2 = \frac{A_o^2}{M^2} + \frac{2A_o}{M} \mathrm{Re}[\psi_{diff}(\underset{\sim}{r}_i)] + |\psi_{diff}(\underset{\sim}{r}_i)|^2 \tag{3.7}$$

where $|\psi_{diff}(\underset{\sim}{r}_i)|^2$ is called the dark field term. In the first-order treatment, the dark field term is assumed to be negligible in comparison with the linear terms (we will later examine the impact of this assumption on our restoration schemes, and how one may compensate for it by iteration procedures[62]). Expanding Eq. (3.7) yields

$$\begin{aligned} |\psi(\underset{\sim}{r}_i)|^2 = \frac{A_o^2}{M^2} \{1 &+ \int [i\Phi(\underset{\sim}{k}) - M(\underset{\sim}{k})] b(\underset{\sim}{k}) \exp[-i\gamma(\underset{\sim}{k}) + 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i] d\underset{\sim}{k} \\ &+ \int [-i\Phi^*(\underset{\sim}{k}) - M^*(\underset{\sim}{k})] b(\underset{\sim}{k}) \exp[i\gamma(\underset{\sim}{k}) - 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i] d\underset{\sim}{k}\} \end{aligned} \tag{3.8}$$

Now, since $\phi(\underset{\sim}{r}_o)$ and $\mu(\underset{\sim}{r}_o)$ are real valued functions, we can write

$$\Phi^*(\underset{\sim}{k}) = \Phi(-\underset{\sim}{k})$$

$$M^*(\underset{\sim}{k}) = M(-\underset{\sim}{k}) \tag{3.9}$$

In conventional bright-field microscopy, the aberration phase shift $\gamma(\underset{\sim}{k})$ has two-fold symmetry; so that if the assumption of isoplanacy is fulfilled[6] we have

$$\gamma(\underset{\sim}{k}) = \gamma(-\underset{\sim}{k}) \tag{3.10}$$

Substituting Eq. (3.9) and Eq. (3.10) into the second integral of Eq. (3.8), replacing $\underset{\sim}{k}$ by $-\underset{\sim}{k}$, and then reversing the limits of integration, the image intensity reduces to:

$$|\psi(\underset{\sim}{r}_i)|^2 = \frac{A_o^2}{M^2}\{1 + \int [2 \sin \gamma(\underset{\sim}{k})\, \Phi(\underset{\sim}{k}) - 2 \cos \gamma(\underset{\sim}{k})\, M(\underset{\sim}{k})]\, b(\underset{\sim}{k}) \cdot \exp(2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i) d\underset{\sim}{k}\} \tag{3.11}$$

If we now define the contrast in the bright-field image as

$$Con(\underset{\sim}{r}_i) = \frac{|\psi(\underset{\sim}{r}_i)|^2 - A_o^2/M^2}{A_o^2/M^2} = \frac{M^2}{A_o^2}|\psi(\underset{\sim}{r}_i)|^2 - 1 \tag{3.12}$$

then the image transform will be defined as the Fourier transform of the image contrast[60]:

$$j(\underset{\sim}{k}) = \mathcal{F}\{Con(\underset{\sim}{r}_i)\} = [2 \sin \gamma(\underset{\sim}{k})\Phi(\underset{\sim}{k}) - 2 \cos \gamma(\underset{\sim}{k})M(\underset{\sim}{k})]b(\underset{\sim}{k}) \tag{3.13}$$

The interpretation of the image in terms of its transform is most easily understood by considering the simpler cases of either a pure

phase object or an amplitude object.

3.2.1.2 Phase contrast transfer function

A) Definition

The phase contrast transfer function describes the response of the instrument at various frequencies when examining a weak-phase object. Thin biological specimens or amorphous specimens, such as thin carbon films, are almost pure weak-phase objects. In considering these specimens, we shall ignore the inelastic scattering events that contribute to the image intensity[63,64]. In this approximation, the image transform given by Eq. (3.13) becomes

$$j_{ph}(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \sin \gamma(\underset{\sim}{k}) \Phi(\underset{\sim}{k}) = Ph(\underset{\sim}{k}) \Phi(\underset{\sim}{k}) \qquad (3.14)$$

where $Ph(\underset{\sim}{k}) = 2 \sin \gamma(\underset{\sim}{k})\, b(\underset{\sim}{k})$, and is called the PCTF. Equation (3.14) defines then a linear relationship between image contrast and object-phase distribution:

$$Con(\underset{\sim}{r}_i) = \phi(\underset{\sim}{r}_o) * \mathcal{F}^{-1}\{Ph(\underset{\sim}{k})\} \qquad (3.15)$$

The image transform is thus identical to an object transform that has been modulated by the transfer function $Ph(\underset{\sim}{k})$. The latter function reduces the amplitudes of the Fourier components by $2 \sin \gamma(\underset{\sim}{k})$. Consequently an aberration-free lens would give no contrast at any spatial frequency (this statement is only true if we neglect the curvature of the Ewald sphere[14], see 3.2.3.1). The physical significance of the PCTF is best illustrated with reference to Figure 3-1 showing graphs of $2 \sin \gamma(\underset{\sim}{k}) = 2 \sin(\frac{\pi}{2\lambda} C_s \theta^4 - \frac{\pi}{\lambda} \Delta Z \theta^2)$ versus θ at various

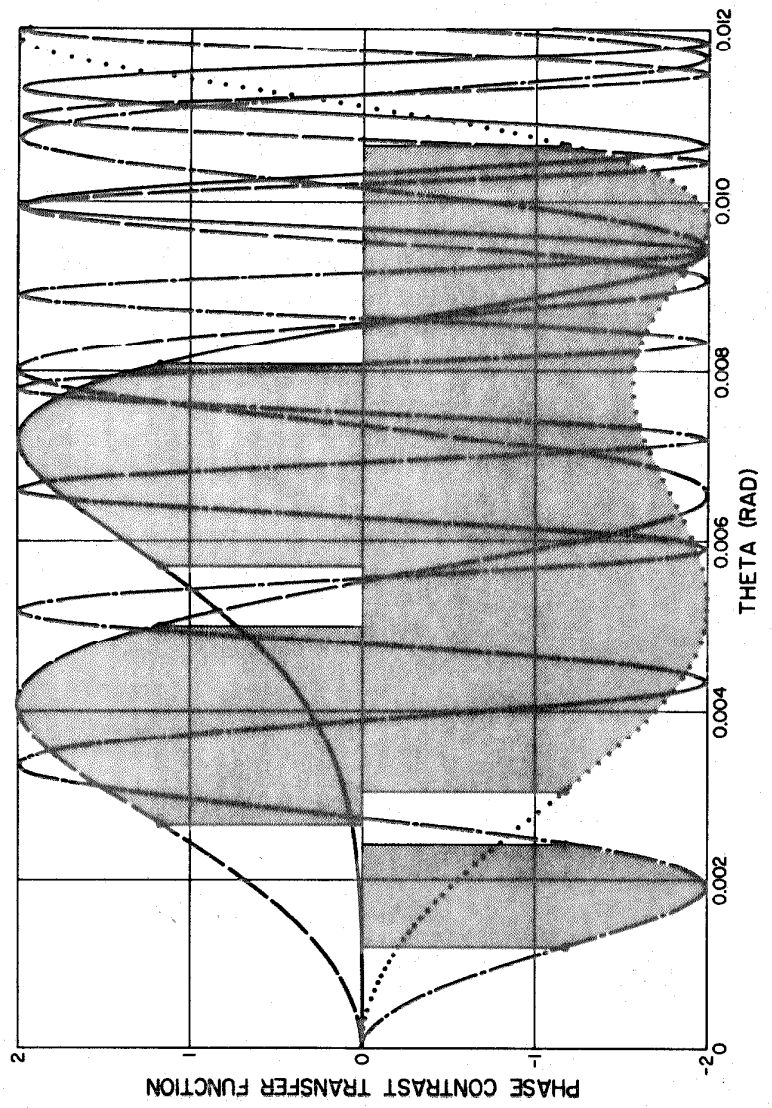


Fig. 3-1. Transfer functions, 2 sin $\gamma(\theta)$, for a weak-phase object. The curves were calculated for $\lambda = 0.037$Å and $C_s = 1.44$ mm. The dash-dot curve corresponds to $\Delta Z = 5000$Å, the dotted curve corresponds to $\Delta Z_{opt} = 860$Å, the solid curve to $\Delta Z = 0$ Å, and the dashed curve to $\Delta Z = -1000$Å. The shaded areas indicate the bandwidth of the signal that is transmitted above the noise in the first contrast zone of the transforms. The threshold used in calculating the bandwidths corresponds to $\gamma(\theta)$ deviating from $(2n-1)\pi/2$ by $\pm 0.3\pi$.

defocusing values ΔZ . At Gaussian focus (i.e., $\Delta Z = 0$) low-resolution components of the object transform are obliterated, since $\sin \gamma(\underset{\sim}{k})$ is solely determined by the spherical aberration term which becomes significant at higher spatial frequencies. Contrast reversals occur where the transfer function goes through zero and changes sign; thus indicating a phase change of 180^{o} for the corresponding spectral components. These phase changes limit the resolution to a value defined by the first zero crossing of $\sin \gamma(\underset{\sim}{k})$. Beyond this limit, object details are imaged with rapidly alternating contrast as a function of spatial frequency. The optimum defocusing, ΔZ_{opt}, is attained when the image distribution is a faithful representation of the object structure; that is, when the transfer function is close to its maximum amplitude over a broad range of spatial frequencies. The conditions for optimum phase contrast are given by Eq. (1.38), and they correspond to a phase shift at the optimal objective aperture half-angle of $\gamma(\theta_{obj}^{opt} = -0.71\pi)$[10]. For example, the optimum defocusing is $\Delta Z^{opt} = 860\text{Å}$ for a microscope operating at 100 keV with $C_s = 1.4$ mm.

In Fig. 3.1 the curve corresponding to $\Delta Z^{opt} = 860\text{Å}$ shows that the image and object transforms are identical over the range $0.082\text{Å}^{-1} \leq |\underset{\sim}{k}| \leq 0.29\ \text{Å}^{-1}$, where details are being imaged with "true" phase contrast. At larger ΔZ values, the range of normal phase contrast shrinks down and moves to lower $|\underset{\sim}{k}|$ values (i.e., at $\Delta Z = 5000\text{Å}$ the range $0.032\text{Å}^{-1} \leq |\underset{\sim}{k}| \leq 0.065\text{Å}^{-1}$ is imaged with enhanced phase contrast). Selective contrast enhancement of object details can thus be achieved with an appropriate setting of the defocusing parameter. Thon has developed a useful method[65] for determining

the defocusing dependence of the PCTF on ΔZ , by plotting the phase-contrast characteristic curves.

B) Phase transfer characteristics

Maximum contrast in weak-phase objects occurs for phase shifts which are odd multiples of $\pi/2$:

$$\gamma(\theta) = \frac{\pi}{2\lambda} C_s\theta^4 - \frac{\pi}{\lambda} \Delta Z\theta^2 = (2n-1)\frac{\pi}{2} \tag{3.16}$$

Thus, for any given defocus, the object details which are transmitted with maximum phase contrast are:

$$\Lambda = \lambda/\theta = \lambda\{\frac{\Delta Z}{C_s} \pm [(\frac{\Delta Z}{C_s})^2 + \frac{(2n-1)\lambda}{C_s}]^{1/2}\}^{-1/2} \tag{3.17}$$

For example, one can read from the characteristic curves shown in Fig. 3-2 which spatial frequencies are transferred at a given defocus ΔZ.

The characteristic curves shown in Fig. 3-2 can be determined experimentally by taking a through focus series of a thin carbon film which has an almost white frequency spectrum. For such an object $|\Phi(\underset{\sim}{k})| = 1$, so that

$$|j_{ph}(\underset{\sim}{k})| = 2b(\underset{\sim}{k})|\sin \gamma(\underset{\sim}{k})| \tag{3.18}$$

The Fourier transforms of these micrographs show ring patterns corresponding to the transmitted frequency bands, so that the defocusing and axial astigmatism can be measured (see Sec. 3.3.3.2).

3.2.1.3 Amplitude contrast transfer function

A pure weak-amplitude object has a transmission function that is given by[21]

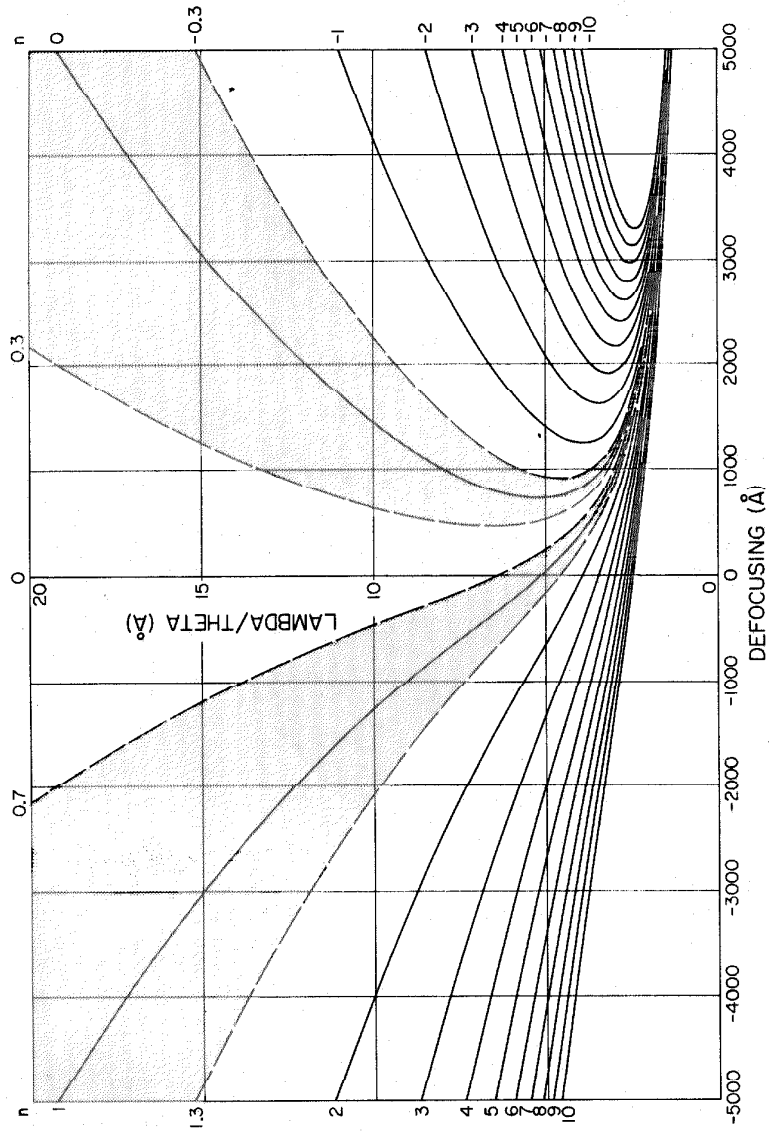


Fig. 3-2. Phase contrast characteristic curves showing the defocus dependence of $\Lambda(=\lambda/\theta)$ on ΔZ, according to Eq. (3.17). The shaded areas indicate the bandwidth of the first contrast zone that is experimentally observed to be transferred above noise. The threshold used in calculating the bandwidths corresponds to $\gamma(\theta)$ deviating from $(2n-1)\pi/2$ by $\pm 0.3\pi$.

$$\Lambda_{amp}(\underset{\sim}{r}_o) = 1 - \mu(\underset{\sim}{r}_o) \quad \text{for} \quad |\mu(\underset{\sim}{r}_o)| << 1 \tag{3.19}$$

Using Eq. (3.13) we can write the image transform of a weak-amplitude object as

$$j_{amp}(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \cos \gamma(\underset{\sim}{k}) \, M(\underset{\sim}{k}) = -Amp(\underset{\sim}{k}) \, M(\underset{\sim}{k}) \tag{3.20}$$

where $Amp(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \cos \gamma(\underset{\sim}{k})$ and is called the amplitude contrast transfer function (ACTF). For an aberration-free lens, the ACTF is constant within the open region of the aperture; hence all frequencies within this range are imaged with the same contrast as the object. Figure 3-3 shows plots of $Amp(\underset{\sim}{k})$ for various defocusing values, and it illustrates the fact that only low frequencies in amplitude modulation are transferred to the image.

Amplitude modulation is always coupled with some phase modulation, and it arises in different ways for real specimens. The amplitude of the primary wave can be modulated by the combined effects of the absorption of electrons by the specimen, removal of elastically scattered electrons from the image forming beam by the objective aperture, contributions from the imaginary part of the scattering amplitudes, and inelastic scattering events. In the section on single atom imaging, a phenomenological equivalence between some of these contrast mechanisms was demonstrated. We showed that the electron loss resulting from scattering outside the objective aperture amounts to an absorption of the electron wave by the object. For an extended cluster of atoms, however, we must also investigate the effects of plural elastic scattering events on the intensity of the electron wave[66]. If the specimen

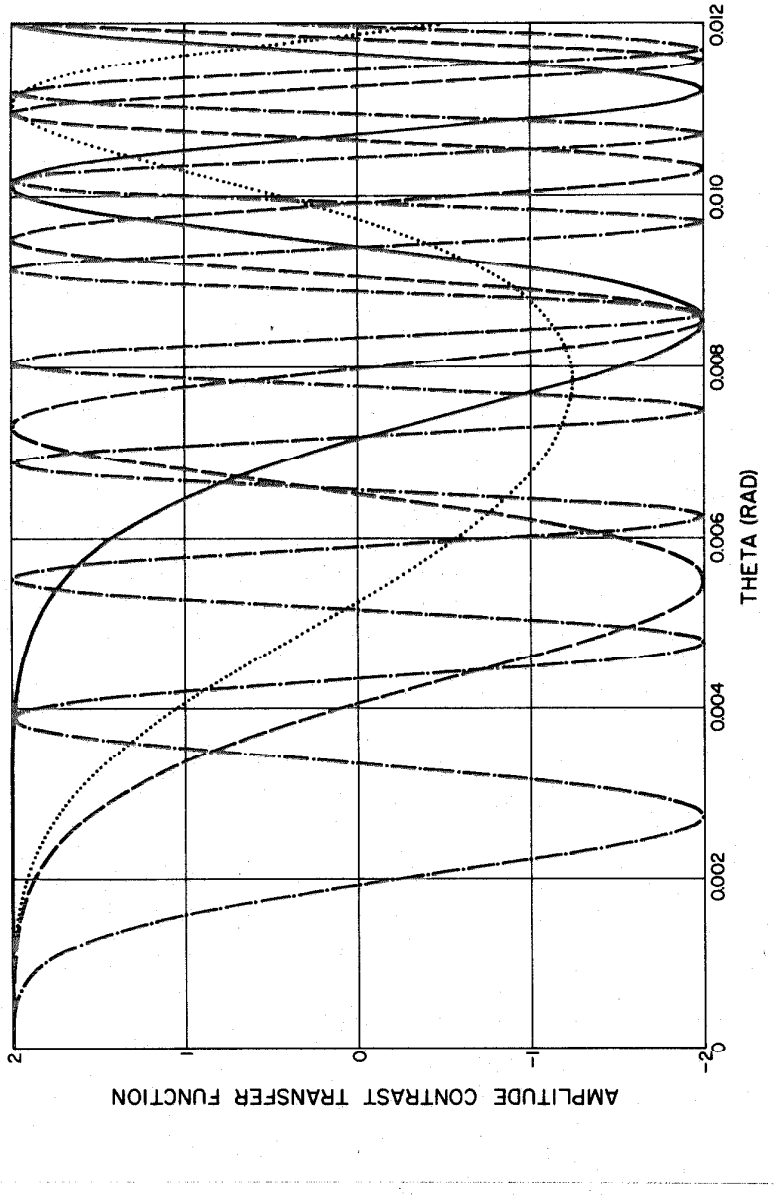


Fig. 3-3. Transfer functions, 2 cos $\gamma(\theta)$, for a weak amplitude object. These waves were calculated for λ = 0.037Å and C_s = 1.4 mm. The dash-dot curve corresponds to ΔZ = 5000Å, the dotted curve to ΔZ^{opt} = 860Å, the solid curve to ΔZ = 0Å and the dashed curve to ΔZ = -1000Å.

exhibits local lattice periodicities, the sizeable Bragg diffracted beams will also affect the validity of our first-order expansion of Λ_t . In order to account for these additional scattering phenomena, second-order terms in the expansion of the phase transmittance, $\exp[i\phi(\underline{r}_o)]$, must then be retained in the analysis (see Sec. 3.2.3.2). These diffraction effects are in fact negligible in the case of thin amorphous structures.

Inelastic scattering introduces a low-resolution amplitude term in the object wave, due to chromatic aberration. In the bright-field mode, these mutually incoherent electrons yield a blurred image for small ΔZ , so that their contribution to the image can be considered to be noise. It was shown, however, that the inelastic image cannot be neglected at larger defocusing values[67] (see 3.2.3.3).

Amplitude contrast, although it plays a minor role in high-resolution bright-field microscopy, provides a means of discriminating between atoms of different atomic number. Image processing schemes, which reconstruct the imaginary part of the structure factor, have indeed demonstrated a selective contrast enhancement[59] (see 3.3.4). In these schemes it is necessary to maximize the amplitude modulation in the image. The appropriate defocusing values can be read directly from the amplitude transfer characteristics, which indicate the object details that are transmitted with maximum amplitude contrast. Maximum contrast for weak amplitude objects is achieved when the phase shifts are multiples of π :

$$\gamma(\theta) = \frac{\pi}{2\lambda} C_s\theta^4 - \frac{\pi}{\lambda} \Delta Z\theta^2 = n\pi \qquad (3.21)$$

so that the characteristic curves are described by

$$\Lambda = \lambda/\theta = \lambda\{\frac{\Delta Z}{C_s} \pm [(\frac{\Delta Z}{C_s})^2 + \frac{2n\lambda}{C_s}]^{1/2}\}^{-1/2} \qquad (3.22)$$

In Fig. 3-4, characteristic curves have been plotted for various values of n. These curves clearly show that the range of ΔZ near Gaussian focus transfers the high spatial frequencies with optimal amplitude contrast.

3.2.1.4 Combined phase and amplitude contrast transfer function

In the first-order treatment the dark field contrast and plural scattering are ignored; so the image transform of a weak phase and amplitude object is, by virtue of Eq. (3.13), equal to

$$j(\underset{\sim}{k}) = \Phi(\underset{\sim}{k})\ Ph(\underset{\sim}{k}) - M(\underset{\sim}{k})\ Amp(\underset{\sim}{k}) \qquad (3.23)$$

We have established, in our analysis of the imaging of a single atom, that the relative contribution of scattering contrast versus phase contrast has the same order of magnitude as the anomalous phase shift angle, so that

$$|\Xi(\underset{\sim}{k})| = \left|-\frac{M(\underset{\sim}{k})}{\Phi(\underset{\sim}{k})}\right| \simeq \tan\eta(\theta) << 1 \qquad (3.24)$$

A similar result holds for a weakly scattering specimen, if the scattering properties of the object can be described by a complex structure factor, previously defined in Eq. (1.14) as

$$F(\underset{\sim}{k}) = \sum_{j=1}^{N_a} |f_j(\theta)|\ \exp(i\eta_j(\theta) - 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j)$$

In this relationship we have neglected thickness effects and assumed that the potential field of the atom is not appreciably altered by molecular binding. In these approximations the ratio of amplitude

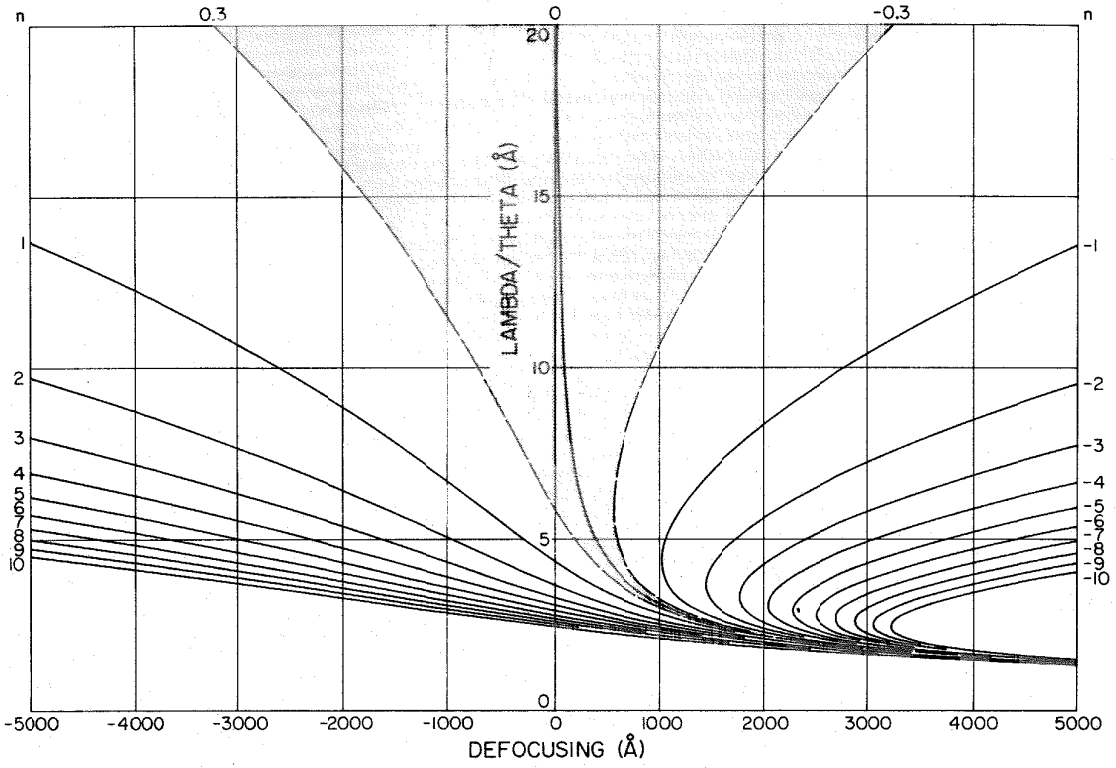


Fig. 3-4. Amplitude contrast characteristic curves showing the defocusing dependence of $\Lambda(=\lambda/\theta)$ on ΔZ, according to Eq. (3.22). The shaded area indicates the bandwidth of the central contrast zone that is experimentally observed to be transferred above noise. The threshold used in calculating the bandwidths corresponds to $\gamma(\theta)$ deviating from $n\pi$ by $\pm 0.3\pi$.

contrast to phase contrast is approximately proportional to $\eta(\theta)$, and therefore increases linearly with the atomic number Z. Consequently, the image transform of a thin amorphous specimen, or of a stained biological molecule, can be written as

$$j(\underset{\sim}{k}) = 2b(\underset{\sim}{k})\,\Phi(\underset{\sim}{k})[\sin\gamma(\underset{\sim}{k}) + \Xi(\underset{\sim}{k})\cos\gamma(\underset{\sim}{k})] \tag{3.25}$$

where $|\Xi(\underset{\sim}{k})| << 1$. The overall instrument contrast transfer function is then defined by[61]

$$C(\underset{\sim}{k}) = \frac{2b(\underset{\sim}{k})}{\cos\nu(\underset{\sim}{k})}\sin[\gamma(\underset{\sim}{k}) + \nu(\underset{\sim}{k})] \tag{3.26}$$

where

$$\nu(\underset{\sim}{k}) = \tan^{-1}\Xi(\underset{\sim}{k})$$

and $1/\cos\nu(\underset{\sim}{k})$ is a normalizing factor close to unity, since $|\Xi(\underset{\sim}{k})| << 1$. We see that the effect of a small fraction of amplitude contrast is to introduce an additional phase shift, $\nu(\underset{\sim}{k})$, in the PCTF. The modified PCTF, $2b(\underset{\sim}{k})\sin[\gamma(\underset{\sim}{k}) + \nu(\underset{\sim}{k})]$, yields slightly different optimal contrast conditions in comparison to the case of a pure weak phase object. The discrepancy between the two optimal microscope settings varies as a function of the imaged atomic species.

The dependence of $\Xi(\underset{\sim}{k})$ on Z provides a useful tool for element discrimination. A separation of the superimposed images of heavy and light atoms can be achieved by combining the transforms of two or more micrographs taken under different transfer conditions (see Sec. 3.3). If the restoration procedure compensates directly for the spatial frequency distortions induced by $C(\underset{\sim}{k})$, an experimental

determination of $\Xi(\underset{\sim}{k})$ becomes necessary. A theoretical evaluation of $\Xi(\underset{\sim}{k})$ is difficult in practice since real specimens are built from atoms with different atomic numbers. An example of this experimental determination of $C(\underset{\sim}{k})$ is shown in the Fourier analysis of a through-focus series of uranium stained catalase[61] (see Sec. 3.4).

3.2.2 Effects of Spatial Incoherence and Chromatic Aberration on the Contrast Transfer Function

Spatial incoherence and energy spread were shown in Sec. 1.3.2.2 to destroy the linearity between object and image wave amplitudes[68]. Provided that the object contrast is small, we shall prove that the linear transfer theory of image formation remains valid for spatially and chromatically incoherent illumination. In this case, the phase and amplitude transfer functions can be represented as a product of an envelope function and of the corresponding coherent contrast transfer function[69,70,71]. Therefore, including partial coherence in the analysis still allows us to restore the complex object wave function from the microscope image.

3.2.2.1 Spatial incoherence

The imaging characteristics of an electron microscope are those of a nearly coherent system, since the illumination angle is much smaller than the objective aperture angle. We have previously come to the conclusion that the second condenser aperture can be considered the effective source of illumination and is characterized by the angular distribution $Q(\underset{\sim}{K}_o)$. For weakly scattering objects, the effects of partial coherence are accounted for by replacing $\exp[-i\gamma(\underset{\sim}{k})]$ in the

formalism by the integral (see Appendix AI)

$$\int_{B_{con}} \exp\{-i[\gamma(\underset{\sim}{k}+\underset{\sim}{K}_o) - \gamma(\underset{\sim}{K}_o)]\}Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o \tag{3.27}$$

where B_{con} is the open area of the condenser aperture and

$$|\underset{\sim}{K}_o| + |\underset{\sim}{k}| < k_m = \frac{\theta_{obj}}{\lambda}$$

If we restrict our analysis to high spatial frequencies satisfying $|\underset{\sim}{k}| >> |\underset{\sim}{K}_o|$, the above substitution amounts to multiplying the constant transfer functions for coherent illumination by an envelope function $E_{si}(\underset{\sim}{k})$.[71] For a uniformly illuminated circular source, this envelope function can then be written in analytical form (cf. Eq.(A.21)) as

$$E_{si}(\theta) = \frac{2J_1[2\pi\theta_{con}(C_s\theta^3 - \Delta Z\theta)/\lambda]}{2\pi\theta_{con}(C_s\theta^3 - \Delta Z\theta)/\lambda} \tag{3.28}$$

where θ_{con} is the condenser aperture half angle and axial astigmatism is assumed to be negligible. By inspecting Eq. (3.28), we notice that the envelope function contains a section around $\theta = \sqrt{\Delta Z/C_s}$ where $E_{si}(\theta)$ leaves the coherent transfer functions unchanged at underfocusing. This section is clearly visible in the plot of $E_{si}(\theta)$ shown in Fig. 3-5.

The corresponding PCTF, corrected for partial coherence, is plotted in Fig. 3-6d. This figure reveals a decrease in the effective resolution, caused by a sharp attenuation of the high spatial frequencies.

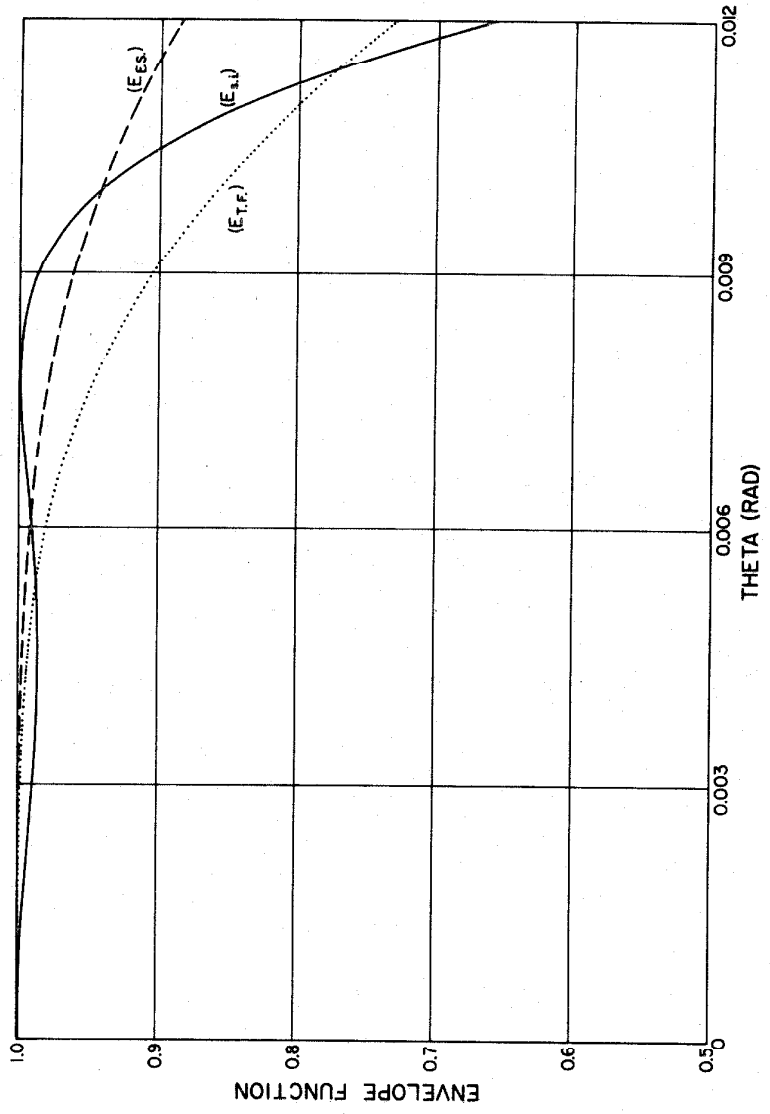


Fig. 3.5 Envelope functions of the electron microscopic contrast transfer functions calculated for 100 keV electrons, C_s = 1.4 mm, C_c = 1.4 mm, θ_{con} = 7.5×10^{-4}rad, ΔZ = 860Å, σ_{tf} = 65Å and T_{fil} = 2800°K. For partially coherent illumination, $E_{si}(\theta)$ is given by Eq. (3.28). For chromatic incoherence, $E_{es}(\theta)$ and $E_{tf}(\theta)$ are given by Eq. (3.29) and Eq. (3.32), respectively.

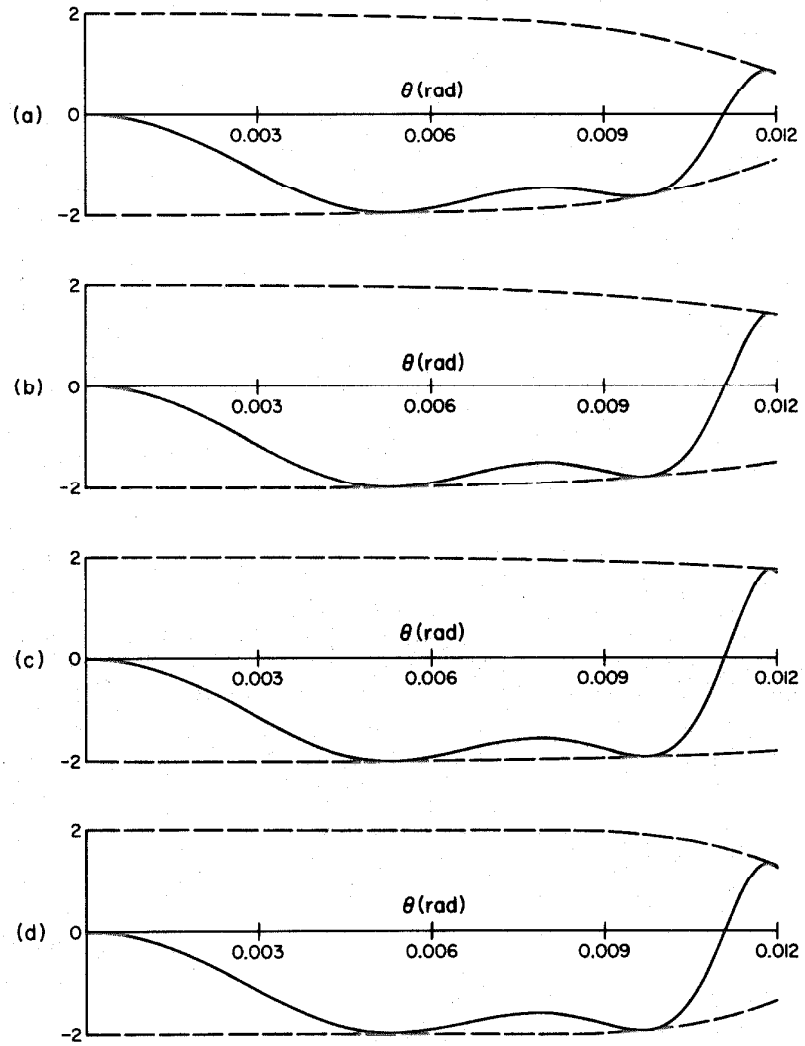


Fig. 3-6 In this figure the corrected phase contrast transfer functions, 2 sin $\gamma(\theta)$ $E(\theta)$, are shown as solid lines. The PCTF has been corrected for spatial incoherence and chromatic aberration. The envelope functions, $E(\theta)$, are shown as dashed lines and have been plotted for identical electron optical parameters as were used in Fig. 3-5: (a) $E(\theta) = E_{si}(\theta)E_{es}(\theta) \cdot E_{tf}(\theta)$; (b) $E(\theta) = E_{tf}(\theta)$; (c) $E(\theta) = E_{es}(\theta)$; (d) $E(\theta) = E_{si}(\theta)$

3.2.2.2 Chromatic incoherence

A) Energy spread

If $N_e(\delta U)$ is the energy distribution of the incident electron beam about the mean, we have seen in Sec. 1.3.2.2B that the Fourier transform of $N_e(\delta U)$

$$L(\underset{\sim}{k},0) \simeq \int_{-\infty}^{\infty} \exp(-\pi i \lambda \frac{\delta U}{U} C_c |\underset{\sim}{k}|^2)\, N_e(\delta U)\, d(\delta U)$$

completely defines the effect of chromatic aberration on the image. If $N_e(\delta U)$ is symmetrical about its mean value, an energy spread can be characterized by an envelope function $E_{es}(\theta)$[70] (see Appendix A2). Experimental measurements have led to the conclusion that $N_e(\delta U)$ satisfies a Maxwellian distribution which may, in practice, be approximated by a Gaussian of the same halfwidth[25] (i.e., δU_{es}).

In the assumption that the object scatters weakly, the effects of energy spread on the contrast transfer functions can be described by the envelope function (cf. Eq. (A.36)):

$$E_{es}(\theta) = \exp[-(\frac{1.22 k_B T_{fil}}{U} \frac{C_c}{\lambda} \frac{\pi\theta^2}{\sqrt{2}})] \qquad (3.29)$$

where T_{fil} is the tungsten filament temperature, k_B the Boltzmann constant, and C_c the chromatic aberration constant. A plot of $E_{es}(\theta)$ is given in Figure 3-5 for $T_{fil} = 2800^{\circ}K$ (i.e., a hairpin filament), and for $C_c = 1.4$ mm. Figure 3-6c illustrates how the attenuation term $E_{es}(\theta)$ mainly affects the high-resolution frequencies of the PCTF corrected for energy-spread effects.

B) Time fluctuations

Fluctuations in the objective lens current and in the accelerating potential introduce an additional phase shift in the aberration function (cf. Eq. (1.77)):

$$\frac{\pi}{\lambda} C_c \left[\frac{\Delta U(t')}{U} - 2 \frac{\Delta i_\ell(t')}{i_\ell}\right]\theta^2 = \frac{\pi}{\lambda} \Delta Z(t')\theta^2$$

which can be interpreted as a time-dependent defocusing error term. Provided that the distribution $H_z(\delta z)$ of δz values is symmetrical about their mean during the exposure time, the random fluctuations multiply the contrast transfer functions by an envelope function[70] (see Appendix A2):

$$E_{tf}(\theta) = \int_{-\infty}^{\infty} \exp(-\pi i\, \delta z \theta^2/\lambda)\, H_z(\delta z)\, d(\delta z) \qquad (3.30)$$

For instance, a Gaussian distributed $H_z(\delta z)$ with a halfwidth given by

$$\sigma_{tf} = C_c\left[\left(\frac{<\Delta U>}{U}\right)^2 + \left(\frac{2<\Delta i_\ell>}{i_\ell}\right)^2\right]^{1/2} \qquad (3.31)$$

leads to an envelope function

$$E_{tf}(\theta) = \exp\left[-\left(\frac{\sigma_{tf}}{\lambda} \frac{\pi}{\sqrt{2}} \theta^2\right)^2\right] \qquad (3.32)$$

In a modern high-resolution microscope, lens and current supplies are regulated in such a way that their average values lie in the range $(<\Delta U>/U) < 2 \times 10^{-6}/\text{min}$ and $(<\Delta i_\ell>/i_\ell) < 2 \times 10^{-6}/\text{min}$. Therefore, the magnitude of the defocus standard deviation is $\sigma_{tf} \simeq 65\text{Å}$. The

corresponding envelope function is plotted in Fig. 3-5 and its associated PCTF, corrected for chromatic incoherence due to time fluctuations, is shown in Fig. 3-6b. An examination of Fig. 3-5 reveals that these fluctuations result in a loss of contrast at high spatial frequencies where the width of a contrast zone becomes comparable to the width σ_{tf} of the range of defocus fluctuations. Consequently, these instabilities place a limit on the point resolution of the microscope.

3.2.2.3 Combined effects

The properties of the electron microscope contrast functions for partially coherent illumination, both spatially and chromatically, have been examined in detail in the case of weakly scattering objects. We have reached the result that the linear theory of image formation still holds, as long as the object contrast is small. Contrast transfer functions can then be represented as a product of an ideal transfer function for coherent illumination and an envelope function, $E(\underset{\sim}{k})$, resulting from the combined action of the resolution limiting effects[71]. Therefore, we can define the partially coherent transfer functions as:

$$T_s^P(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \sin \gamma(\underset{\sim}{k})\, E(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \sin \gamma(\underset{\sim}{k}) E_{si}(\underset{\sim}{k}) E_{es}(\underset{\sim}{k}) E_{tf}(\underset{\sim}{k}) \tag{3.33}$$

and

$$T_s^A(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \cos \gamma(\underset{\sim}{k})\, E(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \cos \gamma(\underset{\sim}{k}) E_{si}(\underset{\sim}{k}) E_{es}(\underset{\sim}{k}) E_{tf}(\underset{\sim}{k}) \tag{3.34}$$

where $T_s^P(\underset{\sim}{k})$ is the modified PCTF and $T_s^A(\underset{\sim}{k})$ is the modified ACTF. A plot of $T_s^P(\underset{\sim}{k})$ with the same electron optical parameters as in the previous examples, is given in Fig. 3-6a.

3.2.3 Applicability of the Linear Theory of Image Formation

For weakly scattering specimens, the output signal of a microscope, (i.e., the recorded image intensity) can be regarded as a linear superposition of elementary images derived from weighted object points. The microscope imaging properties are then described in a simple way by its phase and amplitude contrast transfer functions. Image reconstruction schemes, which correct for the phase errors from the pupil function, can be implemented according to the theorems of linear filtering theory.

Various approximations were made in Sec. 3.2.1.1, which enabled us to write the image transform, $j(\underset{\sim}{k})$, as a linear combination of $\Phi(\underset{\sim}{k})$ and $M(\underset{\sim}{k})$ (cf. Eq. (3.13)). The validity of these approximations will now be discussed, assuming the illumination to be perfectly coherent.

3.2.3.1 Influence of specimen thickness

The elastic scattering of electrons has been described by a single slice model, where the phase and amplitude changes of the electron wave are considered to be localized in a plane perpendicular to the electron beam. We have seen that, for this approximation to hold, the effects of Fresnel diffraction within the specimen must be negligible (i.e., $o_{obj}^2 t_o/\lambda = \lambda t/\rho_s^2 \ll 1$). An upper bound for thickness

values, above which Fresnel diffraction cannot be neglected, has been calculated for bright-field imaging to be[20]

$$t_M^{bf} = 0.5\ \rho_s^2/\lambda \tag{3.35}$$

For instance, at a resolution of $\rho_s = 3\text{Å}$, the upper bound for 100 keV electrons is $t_M^{bf} = 120\text{Å}$.

We shall now examine how thickness alters the phase and the amplitude components of the complex structure factor. It can be shown that even in the first Born approximation, phase contrast is not the only contrast mechanism in the bright-field imaging of atoms[14,62] (see Appendix B). If we neglect the anomalous phase shift contribution, the transform of the object phase modulation then becomes (cf. Eq. (B.10))

$$\Phi(\underset{\sim}{k}) = \lambda \sum_{j=1}^{N_a} f_j'(\theta)\ e^{-2\pi i \underset{\sim}{k}\cdot \underset{\sim}{r}_o^j} \cos\left(\frac{\pi Z_o^j \theta^2}{\lambda}\right) \tag{3.36}$$

or, equivalently,

$$\phi(\underset{\sim}{r}_o) = \lambda \sum_{j=1}^{N_a} \mathcal{F}^{-1}\{f_j'(\theta)\} * \delta(\underset{\sim}{r}_o - \underset{\sim}{r}_o^j) * \mathcal{F}^{-1}\{\cos\left(\frac{\pi Z_o^j \theta^2}{\lambda}\right)\} \tag{3.37}$$

Now for the sake of simplicity, we will assume that our specimen consists of identical atoms. Then, in this approximation Eq. (3.37) can be written

$$\phi(\underset{\sim}{r}_o) = \lambda\, \mathcal{F}^{-1}\{f'(\theta)\} * \left[\sum_{j=1}^{N_a} \delta(\underset{\sim}{r}_o - \underset{\sim}{r}_o^j) * \mathcal{F}^{-1}\{\cos\left(\frac{\pi Z_o^j \theta^2}{\lambda}\right)\}\right] \tag{3.38}$$

Equation (3.38) can be compared to Eq. (1.49) that was derived under

the assumption that the specimen could be assumed to be two-dimensional:

$$\phi_{2d}(\underset{\sim}{r}_o) = \lambda\, \mathcal{F}^{-1}\{f'(\theta)\} * \sum_{j=1}^{N_a} \delta(\underset{\sim}{r}_o - \underset{\sim}{r}_o^j)$$

From this comparison we conclude that atomic resolution is impaired by the overlapping effect of atomic layers. This resolution loss is caused by the blurring function, $\mathcal{F}^{-1}\{\cos(\pi Z_o^j \theta^2/\lambda)\}$, which is injected at each point of the projected structure.

When anomalous scattering is taken into account, the relationship between $\phi(\underset{\sim}{r}_o)$ and the projected density function becomes extremely complex. The phase modulation term may be expressed, using Eq. (B.10), as

$$\phi(\underset{\sim}{r}_o) = \lambda \sum_{j=1}^{N_a} [\mathcal{F}^{-1}\{f_j'(\theta)\} * \mathcal{F}^{-1}\{\cos(\frac{\pi Z_o^j \theta^2}{\lambda})\} - \mathcal{F}^{-1}\{f''(\theta)\} * \mathcal{F}^{-1}\{\sin(\frac{\pi Z_o^j \theta^2}{\lambda})\}] * \delta(\underset{\sim}{r}_o - \underset{\sim}{r}_o^j) \} \qquad (3.39)$$

Consequently, both the real and imaginary parts of the atomic scattering amplitudes contribute simultaneously to both the phase and the amplitude contrast images. This phenomenon obviously counteracts the heavy/light atom discrimination effect, which relies on the Z dependence of the anomalous phase shift[62].

3.2.3.2 Effects of plural scattering and contribution from the dark-field term

The object wave function at the exit surface of a thin specimen, assuming a coherent incident beam, can be described by the transmission function, $\Lambda_t(\underset{\sim}{r}_o)$, given by Eq. (1.42). For weakly scattering objects we showed that the exponential in Eq. (1.42) can be replaced by a series

taken to first-order terms (see Eq. 3.2). If the crystal exceeds the threshold thickness defined by Eq. (1.44), then the first-order theory must be corrected.

In order to correct the first-order theory, we can retain higher-order terms in the exponential expansion of Eq. (1.42). These higher-order terms describe plural scattering events[60], and to second-order Eq. (3.2) becomes

$$\Lambda_t(\underset{\sim}{r}_o) = 1 + i\phi(\underset{\sim}{r}_o) - \frac{1}{2}\phi^2(\underset{\sim}{r}_o) - \mu(\underset{\sim}{r}_o)$$

$$\text{for} \quad |\mu(\underset{\sim}{r}_o)| < |\phi(\underset{\sim}{r}_o)| < 1 \tag{3.40}$$

An additional amplitude contrast term arises from the interaction of the objective aperture with second-order term $-0.5\ \phi^2(\underset{\sim}{r}_o)$.

By analogy with the derivation of Eq. (3.4), the wave amplitude at the image plane, modified by second-order terms, can be expressed as

$$\psi(\underset{\sim}{r}_i) = \frac{A_o}{M}\int S_o^{II}(\underset{\sim}{k})b(\underset{\sim}{k})\ \exp[-i\gamma(\underset{\sim}{k}) + 2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i]d\underset{\sim}{k} \tag{3.41}$$

The modified object spectrum, $S_o^{II}(\underset{\sim}{k})$ is now given by

$$S_o^{II}(\underset{\sim}{k}) = \delta(\underset{\sim}{k}) + i\Phi(\underset{\sim}{k}) - \frac{1}{2}\Phi^{II}(\underset{\sim}{k}) - M(\underset{\sim}{k}) \tag{3.42}$$

where $\Phi^{II}(\underset{\sim}{k})$ is the Fourier transform of $\phi^2(\underset{\sim}{r}_o)$ and can be defined by

$$\Phi^{II}(\underset{\sim}{k}) = \int_{-\infty}^{\infty}\Phi(\underset{\sim}{k}')\ \Phi^*(\underset{\sim}{k}-\underset{\sim}{k}')d\underset{\sim}{k}' = \Phi^{II*}(-\underset{\sim}{k}') \tag{3.43}$$

The image transform can then be computed from

$$j(\underset{\sim}{k}) = \mathcal{F}\{Con(\underset{\sim}{r}_i)\} = \int_{-\infty}^{\infty} S_0^{II}(\underset{\sim}{k}')S_0^{II*}(\underset{\sim}{k}'-\underset{\sim}{k})\, b(\underset{\sim}{k}')\, b(\underset{\sim}{k}'-\underset{\sim}{k})$$

$$\cdot \exp\{-i[\gamma(\underset{\sim}{k}') - \gamma(\underset{\sim}{k}'-\underset{\sim}{k})]\}\, d\underset{\sim}{k}' \qquad (3.44)$$

After substituting Eq. (3.42) into the expression for $j(k)$, we shall neglect terms of higher order than two. In Eq.(3.43) we notice that the object spectrum still obeys the symmetry rule $S_0^{II}(k) = S_0^{II*}(-k)$. Hence, a derivation exactly similar to the one given in the first-order treatment (see Sec. 3.2.1.1) will yield for the image transform:

$$j^{II}(\underset{\sim}{k}) = 2b(\underset{\sim}{k}) \sin \gamma(\underset{\sim}{k})\, \Phi(\underset{\sim}{k}) - 2b(\underset{\sim}{k}) \cos \gamma(\underset{\sim}{k})$$

$$\cdot \left[M(\underset{\sim}{k}) + \frac{1}{2}\Phi^{II}(\underset{\sim}{k})\right] + D_f(\underset{\sim}{k}) \qquad (3.45)$$

Here, $D_f(\underset{\sim}{k})$ is the transform of the dark-field image and is approximately equal to

$$D_f(\underset{\sim}{k}) = \int_{-\infty}^{\infty} \Phi(\underset{\sim}{k}')\, \Phi(\underset{\sim}{k}-\underset{\sim}{k}')b(\underset{\sim}{k}')b(\underset{\sim}{k}'-\underset{\sim}{k})\exp\{-i[\gamma(\underset{\sim}{k}') - \gamma(\underset{\sim}{k}'-\underset{\sim}{k})]\}d\underset{\sim}{k}' \qquad (3.46)$$

The second-order theory of image formation introduces two additional terms: a dark-field term, $D_f(k)$, and a term stemming from double scattering within the specimen, $D_{ds}(k)$, which is equal to

$$D_{ds}(\underset{\sim}{k}) = -b(\underset{\sim}{k}) \cos \gamma(\underset{\sim}{k}) \int_{-\infty}^{\infty} \Phi(\underset{\sim}{k}')\, \Phi^*(\underset{\sim}{k}-\underset{\sim}{k}')\, d\underset{\sim}{k}' \qquad (3.47)$$

This latter term can be separated into two parts; the first is a contribution from $\underset{\sim}{k}'$ values falling outside the objective aperture, and

the second part consists of a second-order correction to the phase contrast image[60].

Second-order effects will be important near the zeros of the first-order PCTF. In the case of a crystal, these additional contrast terms can be attributed to the mutual interaction between two of the strongly diffracted beams and also to the existence of a significant dark-field term. It follows that their contribution to the image must be evaluated for each particular case, whenever resolution extending schemes are implemented[61].

3.2.3.3 Contribution from inelastic scattering

The overall effects of inelastically scattered electrons are to reduce image contrast, thermally heat the specimen, and to cause radiation damage. These combined effects will limit the information content of a microscope image. We know that for weakly scattering objects, plasmon excitation is the predominant inelastic scattering process[63].

If we call δE_p the plasmon energy loss, we can show that the inelastic angular distribution decreases as $(\theta^2 + \theta_e^2)^{-1}$, where $\theta_e = \delta E_p/2U$. Typical values of a single plasmon loss are a few eV, so that the distribution is sharply peaked in the incident beam direction. The total inelastic cross section, σ_{inel}, for this process has been computed by Lenz, who used a Thomas Fermi model for the atom in order to demonstrate that $\sigma_{inel} \propto Z^{1/3}$.[17] In contrast to this, the elastic cross section, σ_{el}, varies as $Z^{4/3}$. We see, therefore, that inelastic scattering predominates at low Z, and that the linear

dependence of $\sigma_{el}/\sigma_{inel}$ on the atomic number could be used to discriminate between heavy and light atoms[72].

In conventional electron microscopy the image formed with inelastically scattered electrons is blurred by the chromatic aberration of the objective lens. Contrast calculations have been made as a function of defocusing, which indicate that in the normally used defocus range (i.e., $0 \leq |\Delta Z| \leq 1000\text{Å}$) chromatic aberration limits the inelastic image resolution to about 20Å.[67] However, the inelastic image displays a much higher resolution at very large underfocus values. In this range of ΔZ it becomes comparable in intensity to the elastic term and consequently cannot be ignored in the analysis. The effect of inelastic scattering is to modulate the amplitude of the primary wave, thus appearing as an additional contribution to $\mu(\underset{\sim}{r}_o)$. In our treatment, inelastic scattering will be neglected (although it is particularly strong for carbon) and assumed to yield an unstructured background image blurred by chromatic aberration.

3.3 Image Restoration Methods

3.3.1 Principles

If the interaction between the object and the illuminating beam is weak, we have shown in the previous section that the electron current distribution in the bright-field image is linearly related to both the phase shift and the amplitude reduction of the electron beam. This relationship remains linear even under conditions of partial coherence.

In Eq. (3.13) the formation of the final image is described as a product of two separate entities: the scattering properties of the specimen, and the transfer properties of the imaging system. If the transfer functions are known, it should then be possible to unambiguously reconstruct the object from at least two micrographs taken under different imaging conditions. These two micrographs are taken so as to fill the information gap corresponding to the zeros of the transfer functions. In order to obtain this information the micrographs can be recorded at either different ΔZ, different conditions of illumination, or using different apertures[73,74]. In these latter cases, however, the image transform is no longer expressed by Eq. (3.13), since $b(k) \neq b(-k)$.

Single side-band holography is one of the methods which has been implemented optically to compensate for the effects of the oscillating transfer function. This method consists of taking bright-field images of the same object with two complementary semi-circular apertures and then extracting separately the phase and amplitude components of the micrograph by an a posteriori optical reconstruction. However, the charge build-up on the apertures introduce spurious phase shifts, which must be compensated for, and limit the practicability of this resolution extending scheme[58].

The image reconstruction process used in our research combines micrographs with different defocusing in order to correct the information reduction and to retrieve the phase and amplitude contrast contributions[59]. This scheme requires for its correct execution, an accurate determination of the pupil function and the ability to perform a perfect registration between the micrographs of a defocus series. It

then calculates the object Fourier transform from a weighted sum of the transforms of the micrographs. The complexity of the procedure necessitates the use of a digital computer to perform these operations. A preliminary analysis of the optical transforms is, however, necessary in order to judge the quality of the input series with regard to specimen drift and objective lens astigmatism, since these effects limit the performance of the restoration. Optical transforms and their interpretation will now be discussed.

3.3.2 Analysis of Optical Diffractograms

3.3.2.1 Optical bench set-up

The optical diffractometer used in this work is illustrated schematically in Fig. 3-7. It consists of a Spectra Physics 5 mW He-Ne laser, a beam expanding lens with a 12 μm pinhole that removes the stray light, a collimating lens forming a parallel beam incident on the micrograph, a diffraction lens of long focal length, a magnifying lens enlarging the diffraction pattern by a factor of M_p, and a camera recording the power spectrum of the transparency. For a given spatial frequency in the object, $1/\Lambda$, the radius r_ℓ at which the diffraction spot appears in the transform plane of the optical diffractometer is given by

$$r_\ell = \frac{\lambda_\ell f M_p}{\Lambda M} \tag{3.48}$$

where λ_ℓ is the wavelength of the laser light. The diffractometer is calibrated using a diffraction grating of known spacing. This calibration yields the quantity, $\lambda_\ell f M_p$, in Eq. (3.48).

The optical diffraction pattern formed from an electron micrograph is directly related to the structural detail recorded on the

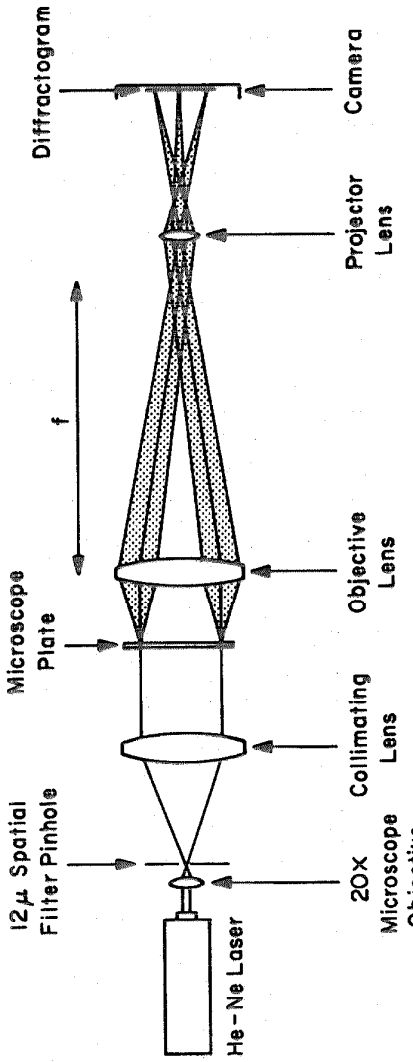


Fig. 3-7 Schematic diagram of the optical set-up used for obtaining optical diffractograms

plate. If the image contains periodic detail, discrete diffraction spots will be observed in the optical diffractogram. If a periodic specimen is supported by an amorphous substrate the diffractogram will consist of diffraction spots superimposed with a diffuse noise spectrum. The noise spectrum is from the substrate and it is altered in a characteristic fashion by such imaging defects as astigmatism, specimen drift, and out of focus effects[65]. A quantitative analysis of the noise spectrum is possible in the case of bright-field images of thin substrates.

3.3.2.2 Evaluation of defocusing and astigmatism

In bright-field microscopy, phase contrast yields the strongest contribution to high-resolution image structure. Thin biological specimens rarely exhibit diffraction maxima at the atomic resolution level because of local distortions in their periodicity. However, they are often supported by a thin amorphous film, which allows one to make a determination of the electron microscope parameters necessary for image processing. In the high spatial-frequency range, these optical transforms correspond to the power spectrum of a weak-phase object, which according to Eq. (3.14) is proportional to

$$|j_{ph}(\underset{\sim}{k})|^2 = |2b(\underset{\sim}{k}) \sin \gamma(\underset{\sim}{k}) \Phi(\underset{\sim}{k})|^2$$

in the case of a coherent illumination. In this frequency range the PCTF, $2 \sin \gamma(\underset{\sim}{k})$, is wildly oscillating and transmits only certain frequency bands to the image. Consequently, a diffractogram will show the transferred frequency bands as bright rings when the corresponding frequencies are present in the specimen phase modulation. A thin layer

of amorphous carbon constitutes an ideal specimen for checking the transfer conditions of a microscope. The frequency spectrum of amorphous carbon is approximately white, since the carbon grains vary in size from very small (a few angstroms) to quite large (several nanometers). This wide frequency spectrum accounts for the speckled, uninformative appearance seen in a bright-field micrograph of such a specimen.

In general, the optical transform of a high-resolution bright-field micrograph will yield a map of $\sin^2\gamma(\underset{\sim}{k})$, and therefore allow an evaluation of the defocusing and the axial astigmatism. As an example of such a determination, consider Fig. 3-8, which illustrates the process of extracting ΔZ, ΔZ_a and ϕ_a from an optical transform. Since the outermost ring corresponds to phase contrast we can write

$$|j(\underset{\sim}{k})|^2 \propto \sin^2\left(\frac{\pi}{2\lambda} C_s\theta^4 - \frac{\pi}{\lambda}\Delta\tilde{Z}\theta^2\right) \tag{3.49}$$

where $\Delta\tilde{Z} = \Delta Z + 0.5\,\Delta Z_a \sin 2(\phi-\phi_o)$ is the effective defocusing at the position $\underset{\sim}{k}$. An examination of the elliptical pattern yields ϕ_o, the reference angle of the axial astigmatism, and the radii r_{max} and r_{min} that correspond to the intensity maxima along the principal axes of the ellipse. We can then calculate Λ_{min} and Λ_{max} from Eq. (3.48). These Λ values are then read into the phase contrast transfer characteristics (plotted for the appropriate spherical aberration constant) from which one can determine the defocus values $\Delta\tilde{Z}_{min}$ and $\Delta\tilde{Z}_{max}$ corresponding to the transferred frequency band. If the measured Λ values fall on either the $n = 0$ or the $n = 1$ characteristic curves, the edges of the visible contrast zone are defined by:

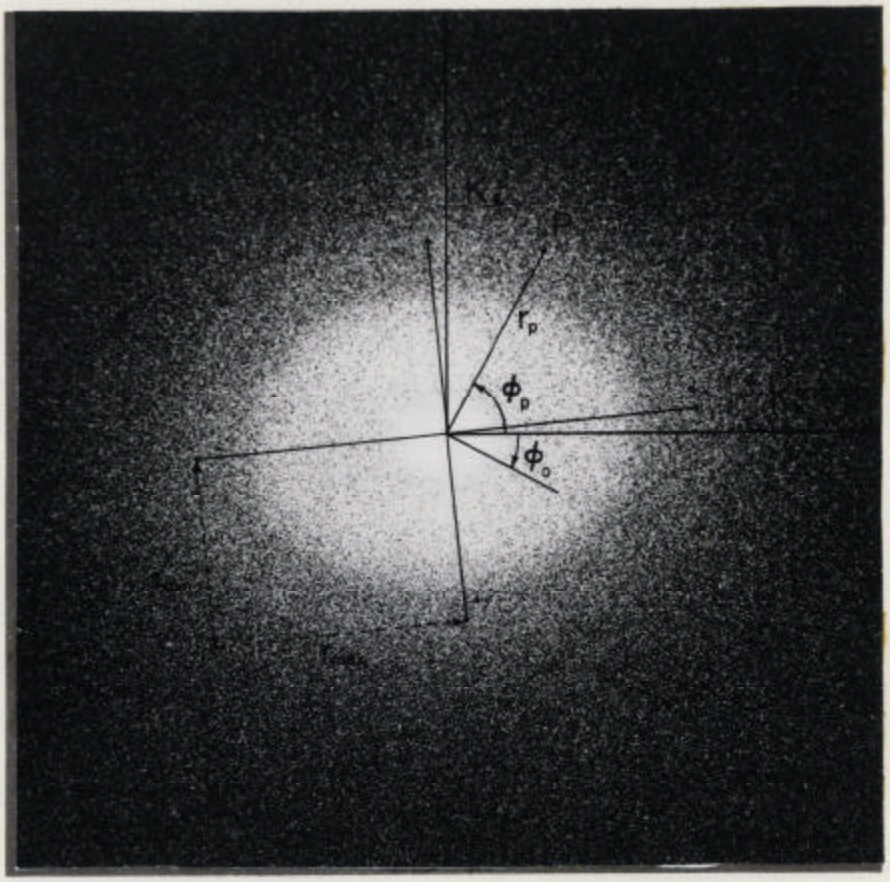


Fig. 3-8. Optical transform of micrograph (a) of the catalase series shown in Fig. 3-20. This transform illustrates how one determines ΔZ, ΔZ_a, and ϕ_o from the principal radii, r_{min} and r_{max}, of the outermost elliptical ring. This outermost ring corresponds to phase contrast.

$\gamma = [(2n-1)\pi/2 \pm 0.3\pi]$ [65]. Therefore, a measurement of the observed zone bandwidth will provide an alternative way of estimating $\Delta\tilde{Z}$ and ΔZ_a, which can be a more accurate value than can be determined by measuring the peak location if the peak is not easily discerned. For example, by using this procedure, we have analyzed Fig. 3-8 and determined the following parameters: $\Delta\tilde{Z}_{min} \simeq -825\text{Å}$, $\Delta\tilde{Z}_{max} \simeq -475\text{Å}$, $\Delta Z \simeq -650\text{Å}$, and $\Delta Z_a \simeq 350\text{Å}$. In this case the amplitude contrast that is transferred is clearly visible in the innermost ellipse shown in Fig. 3-8. The optical transform is then proportional to

$$|j(\underset{\sim}{k})|^2 \propto \cos^2\left(\frac{\pi}{2\lambda} C_s\theta^4 - \frac{\pi}{\lambda}\Delta\tilde{Z}\theta^2\right) \tag{3.50}$$

which allows an evaluation of ΔZ, ΔZ_a and ϕ_o from the amplitude contrast transfer characteristics (this determination is less accurate since $r_\ell \propto 1/\Lambda$).

3.3.2.3 Detection of specimen drift

The impinging electron beam induces local thermal gradients in the specimen, which cause specimen motion during the photographic recording of the image. The recorded image is therefore an incoherent superposition of partial intensities:

$$|\psi_{dr}(\underset{\sim}{r}_i)|^2 = \frac{1}{t_r}\int_0^{t_r} |\psi(\underset{\sim}{r}_i)|^2 * \delta[\underset{\sim}{r}_i - \underset{\sim}{D}_r(t')]\, dt' \tag{3.51}$$

where $\underset{\sim}{D}_r(t')$ is a time dependent function describing the image motion. The image motion can also be described as a convolution of the ideal image with a blurring function, $b_d(\underset{\sim}{r}_i)$, given by

$$b_d(\underset{\sim}{r}_i) = \frac{1}{t_r}\int_0^{t_r} \delta[\underset{\sim}{r}_i - \underset{\sim}{D}_r(t')]\, dt' \tag{3.52}$$

If motion has occurred during image recording, we see from Eq. (1.51) that the resulting image will be multiplied by an image motion OTF, $B_d(\underset{\sim}{k})$, defined by

$$B_d(\underset{\sim}{k}) = \mathcal{F}\{b_d(\underset{\sim}{r}_i)\} = \frac{1}{t_r}\int_0^{t_r} \exp[-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{D}_r(t')]\, dt' \tag{3.53}$$

Consequently, the observed intensity distribution in the optical diffractogram becomes proportional to $|j(\underset{\sim}{k})\ B_d(\underset{\sim}{k})|^2$.

The transform will show fringes if image motion has occurred during the recording[75]. These fringes correspond to a modulation of the contrast transfer function by the image motion modulation transfer function, $|B_d(\underset{\sim}{k})|$. For example, if the specimen drift has a constant velocity, $\underset{\sim}{v}_d$, the modulation function is expressed as

$$|B_d(\underset{\sim}{k})| = \frac{\sin(\pi t_r \underset{\sim}{k}\cdot\underset{\sim}{v}_d)}{\pi t_r \underset{\sim}{k}\cdot\underset{\sim}{v}_d} \tag{3.54}$$

This drift will show up in the transform as a fringe pattern overlaid upon the phase contrast zones; the resulting fringes are equidistant and their spacing is inversely proportional to the drift velocity.

It is possible to design an image restoration filter to deblur an image that has undergone motion during the recording, if the specimen drift function $D_r(t')$ is known. However, it is more practical to re-record the image if the optical transforms indicate that specimen motion has taken place.

3.3.3 Restoration by Deconvolution

The oscillations of the contrast transfer functions and the limited band of spatial frequencies that are imaged with nearly maximum contrast bound the valid resolution of the instrument. Therefore, only a single broadened frequency band of optimal contrast can be obtained at an appropriate defocus. This would be sufficient if one were imaging periodic objects, but very few high-resolution specimens contain just one frequency component. Thus, for a general object, artifacts that are produced by image reversals and by missing frequencies at the zeros of the transfer functions make the image interpretation difficult. Therefore, an increase in the resolution of an electron microscope can be gained by an a posteriori compensation for its lens aberrations.

In the following discussion, we shall assume that one deals with a weak phase object (i.e., $M(\underset{\sim}{k}) = 0$) that is imaged under partially coherent illumination. According to Eq. (3.33), the image degradation can then be considered as the convolution product of an ideal image with a point spread function $t_s^P(\underset{\sim}{r}_i)$, superimposed onto a noise background $n(\underset{\sim}{r}_i)$. (The noise originates from photographic graininess and from inelastic scattering blurred by chromatic aberrations). We can then write the image contrast as

$$\mathrm{Con}(\underset{\sim}{r}_i) = \int_{-\infty}^{\infty} t_s^P(\underset{\sim}{r}_i - \underset{\sim}{r}_o)\, \phi(\underset{\sim}{r}_o) d\underset{\sim}{r}_o + n(\underset{\sim}{r}_i) \qquad (3.55)$$

If we neglect diffraction effects caused by the objective aperture, then the point spread function can be written

$$t_s^P(\underset{\sim}{r}_i) = \mathcal{F}^{-1}\{2b(\underset{\sim}{k}) \sin \gamma(\underset{\sim}{k})\, E(\underset{\sim}{k})\} \simeq \mathcal{F}^{-1}\{2 \sin \gamma(\underset{\sim}{k})\, E(\underset{\sim}{k})\} \quad (3.56)$$

The design of an inverse filter that will partially restore the projected object phase distribution from its associated image contrast function has been discussed in Section 2.3.5.2C. The practical implementation of an inverse filter requires a knowledge of $t_s^P(\underset{\sim}{r}_i)$; therefore electron optical parameters must be accurately determined before the restoration procedure. Determination of these electron microscope parameters can be achieved by the least-squares fit procedure presented in the next section. We shall then derive criteria for the validity of a reconstruction, and we shall illustrate the practical limitations of this method[76,77].

3.3.3.1 Determination of the electron optical parameters

A) Principles

Optical diffractograms yield values for defocusing and axial astigmatism that are subject to a significant experimental error. However, the transfer function $T_s^P(\underset{\sim}{k})$ and the position of its zeros must be known very accurately for the design of an inverse filter. In order to improve the determination of ΔZ and ΔZ_a, a least squares procedure can be used which fits the transform of an amorphous film with the function, $g_r(\theta)|\sin \gamma(\theta,\phi)|$, where $g_r(\theta)$ represents the radial dependence of the spatial frequency spectrum of the film[77].

A carbon support film is ideal for this procedure, since it has an almost white frequency spectrum. In this procedure a portion of the image containing this film is Fourier transformed digitally. The electron microscope parameters are then found at the minimum of

the least-squares sum

$$\Delta_{em} = \sum_{i=1}^{N_p} [g_r(\theta_i) \sin \gamma(\theta_i,\phi_i,\Delta Z,\Delta Z_a,\phi_o)| - |j_c(\theta_i,\phi_i)|]^2 \quad (3.57)$$

where $|j_c(\theta_i,\phi_i)|$ is the amplitude of the carbon film Fourier coefficient at $\underset{\sim}{k}_i$. The sum in Eq. (3.57) is evaluated over an annulus-shaped frequency band containing N_p pixels that is designed in such a way as to exclude the amplitude transferred central zone.

Optical diffractograms of bright field images of amorphous carbon show that the power spectrum has a radial dependence of $1/\theta^2$ [78]. Consequently, we can express the function $g_r(\theta)$ as

$$g_r(\theta) = \frac{A_r}{\theta} \quad (3.58)$$

where A_r is a constant to be determined. This radially dependent weight-function accounts for the shape of the atomic scattering amplitude and for partial coherence effects that attenuate high-spatial frequencies.

B) Computer algorithm

The autocorrelation integral, $\phi_{d,d}(\underset{\sim}{r}_i)$, of an area of a micrograph that shows only the carbon substrate is first calculated by the relation

$$\phi_{d,d}(\underset{\sim}{r}_i) = \mathcal{F}^{-1}\{D(\underset{\sim}{k})\ D^*(\underset{\sim}{k})\}$$

where $D(\underset{\sim}{k})$ denotes the discrete transform of a selected carbon area (refer to Sec. 2.3.4.2C). A central portion of the ACF, $\phi_{d,d}(\underset{\sim}{r}_i)$, is then cut out and multiplied by a radial function $w(\underset{\sim}{r})$, where

$$w(\underset{\sim}{r}) = \begin{cases} 1 & \text{if } r < r_G \\ \exp[-(r-r_G)^2/2\sigma_G^2] & \text{if } r > r_G \end{cases} \tag{3.59}$$

in order to smooth the discontinuities of the window boundaries. The multiplication of the ACF by a window function amounts to convolving the power spectrum with the function $\mathcal{F}\{w(\underset{\sim}{r})\}$. Such an apodization will provide a smoothed power spectrum, described by

$$|j_c(\underset{\sim}{k})|^2 = \mathcal{F}\{\phi_{d,d}(r_i)\, w(\underset{\sim}{r}_i)\} \tag{3.60}$$

Minimizing the least squares sum given by Eq. (3.57) yields four equations for the parameters A_r, ΔZ, ΔZ_a, and ϕ_o :

$$\sum_{i=1}^{N_p} \frac{|\sin\ \gamma_i|}{\theta_i} \left[\frac{A_r}{\theta_i} |\sin \gamma_i| - |j_{ci}|\right] = 0 \tag{3.61}$$

$$\sum_{i=1}^{N} \frac{1}{\theta_i} \frac{\partial|\sin \gamma_i|}{\partial q_n} \left[\frac{A_r}{\theta_i} |\sin \gamma_i| - |j_{ci}|\right] = 0,\ n=1,2,3 \tag{3.62}$$

where q_n stands for either ΔZ, ΔZ_a, or ϕ_o . A matrix inversion of Eq. (3.61) will yield the parameter A_r . The other equations, however, are nonlinear in q_n and must be solved by Newton's method. The starting values for the parameters q_n are obtained from the optical diffractograms. These values are then fed into the least-squares program which re-evaluates the electron microscope parameters by iteration until a convergence criterion is satisfied. With the least-squares fit, the electron microscope parameters can be determined to an accuracy of 5%. The main advantage of this technique is to furnish the best

possible representation for the function $|\sin \gamma|$; thereby allowing a match of the inverse filter with the input data.

3.3.3.2 Inverse filter design

A) Overview

The electron microscope transfers information about the structure of a weak-phase object that is band limited by the function $\sin \gamma(\underset{\sim}{k})\ E(\underset{\sim}{k})$. At the Scherzer focusing conditions this transfer exhibits an almost perfect behavior over a wide range of spatial frequencies (see Fig. 3-6a). We therefore expect to achieve nearly optimal contrast at ΔZ^{opt}. The strong attenuation of low frequency and high frequency components in the image at ΔZ^{opt} can be partially compensated for by inverse filtering. In addition, we can also extend the valid resolution of the microscope by correcting for the phase reversals and for the amplitude attenuation of Fourier components lying beyond the Scherzer domain (but within the instrument resolution limit).

In order to assess the validity of an object reconstruction by deconvolution, we have introduced in Sec. 2.3.5.2C such criteria as the noise and signal amplification factors, and the relative structural content. We shall now apply this analysis to the case of the modified filter given by

$$H(\underset{\sim}{k}) = \frac{g(\underset{\sim}{k})}{\sin \gamma(\underset{\sim}{k})\ E(\underset{\sim}{k})} \tag{3.63}$$

where $g(\underset{\sim}{k})$ is designed in such a way as to prevent noise amplification at the zeros of $\sin \gamma(\underset{\sim}{k})$ in the restored image. We must also ensure that the noise spectrum does not become amplified, at a certain spatial frequency, beyond a given threshold Q_n. This requires that

the following inequality be fulfilled (cf. Eq. 2.99):

$$|H(\underset{\sim}{k})|^2 = \left|\frac{g(\underset{\sim}{k})}{\sin \gamma(\underset{\sim}{k})\ E(\underset{\sim}{k})}\right|^2 \leq Q_n \qquad \text{for all } \underset{\sim}{k} \in B_{obj} \tag{3.64}$$

Linfoot's image evaluation criteria provide a quantitative assessment of the success of a restoration. These criteria were introduced in Sec. 2.3.6.2 in order to evaluate the similitude between an object and its noisy degraded image. If we assume that the specimen exhibits a white power spectrum (i.e., $|\Phi(\underset{\sim}{k})| = 1$), and if we neglect axial astigmatism, we can then express Linfoot's criteria in the following analytical form (cf. Eqs. 2.110, 2.111, 2.112):

$$\Psi_r = \frac{2}{\theta_{obj}^2} \int_0^{\theta_{obj}} E(\theta) \sin \gamma(\theta)\theta \, d\theta \tag{3.65}$$

$$\Xi_r = \frac{2}{\theta_{obj}^2} \int_0^{\theta_{obj}} E^2(\theta) \sin^2\gamma(\theta)\theta \, d\theta \tag{3.66}$$

$$\Phi_r = 2\Psi_r - \Xi_r \tag{3.67}$$

We shall calculate these figures of merit as a function of θ_{obj} in order to estimate the gain in mutual information between the object and the restored image. In this case we want to compare $\phi(\underset{\sim}{r}_o)$ with the output of the restoration $Con_f(\underset{\sim}{r}_i)$ which can be expressed as

$$Con_f(\underset{\sim}{r}_i) = h * Con \tag{3.68}$$

where $Con(\underset{\sim}{r}_i)$ is described by Eq. (3.55). Linfoot's criteria then become:

$$\Psi_r^{res} = \frac{2}{\theta_{obj}^2} \int_0^{\theta_{obj}} g(\theta)\theta \, d\theta \tag{3.69}$$

$$\Xi_r^{res} = \frac{2}{\theta_{obj}^2} \int_0^{\theta_{obj}} g^2(\theta)\theta \, d\theta \tag{3.70}$$

$$\Phi_r^{res} = 2\Psi_r^{res} - \Xi_r^{res} \tag{3.71}$$

B) Practical implementation

In order to illustrate the problems which one faces in implementing inverse filters we shall choose a weight function specified by (cf. Eq. 2.105)

$$g(\theta) = \begin{cases} 1 & \text{for } |\sin \gamma(\theta)E(\theta)| > \bar{\delta} \\ |\sin \gamma(\theta)\, E(\theta)| & \text{otherwise} \end{cases} \tag{3.72}$$

where $\bar{\delta}$ is a threshold value that is related to the noise-amplification threshold by $\bar{\delta} = Q_n^{-1/2}$. This particular inverse filter will be applied to one picture of a through focus series of gold on carbon, so that we shall study the behavior of this filter for the corresponding electron microscope parameters of the picture to be processed. We now want to examine the effects of $H(\underset{\sim}{k})$ on noise propagation and on image quality assuming that axial astigmatism can be ignored.

The PCTF corrected for partially coherent illumination, the inverse filter, and its weight function are plotted against θ_{obj} in Fig. 3-9. We observe sharp discontinuities in the inverse filter profile when $|H(\underset{\sim}{k})| = \bar{\delta}$; these discontinuities could have been smoothed

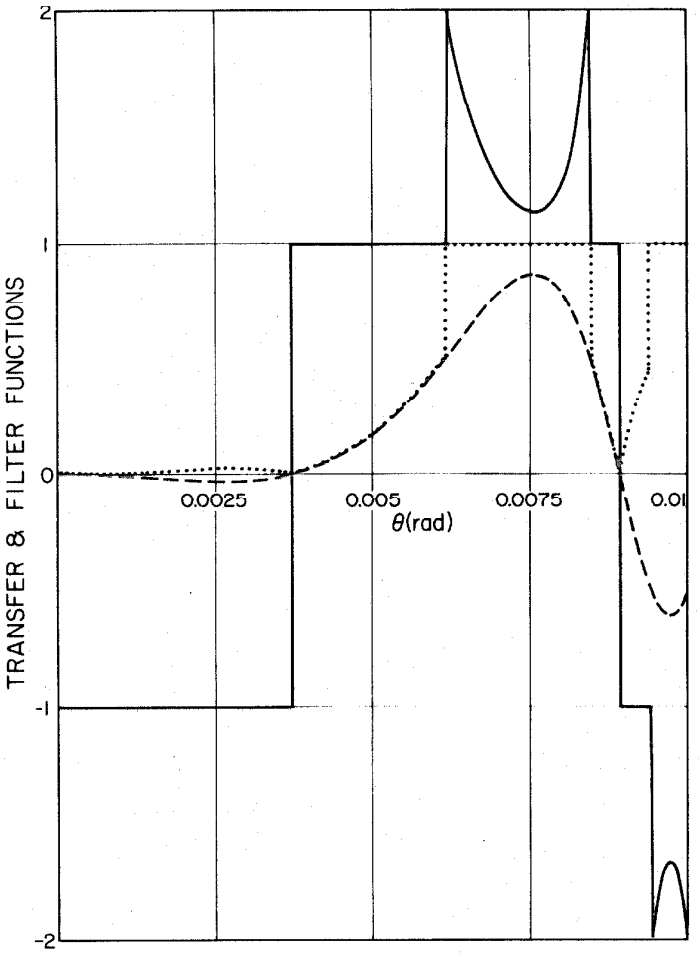


Fig. 3-9 Plots of the phase contrast transfer function corrected for partial coherence, of the inverse filter, and of its associated weight function. The electron microscope parameters used are identical to those employed in Fig. 3-6a and $\overline{\delta} = 0.5$. The PCTF, 2 sin $\gamma(\theta)E(\theta)$, is shown as a dashed curve, the filter $H(\theta)$ as a solid curve, and the weight $g(\theta)$ as a dotted curve.

by a more judicious choice for the weight function [77] (e.g., $g(\theta) = \exp[\alpha_r - \alpha_r/|\sin \gamma(\theta)|]$. We also see that the PCTF only transmits valid information in a narrow frequency band, $3.8 \times 10^{-3}\text{rad} \leq \theta \leq 8.9 \times 10^{-3}\text{rad}$; outside this band phase reversals occur. The effects of the restoration on the transfer of spatial frequencies is best illustrated by examining plots of Linfoot's criteria before and after the filtering (see Fig. 3.10a,b). Before the reconstruction, the correlation quality is negative at low spatial frequencies, indicating a contrast reversal; we also notice a sharp fall-off in Ψ_r beyond the second zero crossing of the PCTF. After filtering, this parameter remains positive and assumes larger values without reaching a value of one which corresponds to an ideal reconstruction. Most of the gain in resolution occurs at high frequencies beyond 8.5×10^{-3}rad. In this spectral range correlation quality and image fidelity remain approximately constant in the filtered image, in contrast to their steep fall-off in the original image. By selecting a threshold value of $\overline{\delta} = 0.5$, we have set a limit to the achievable improvement in resolution. A smaller $\overline{\delta}$ would improve the resolution at the expense of amplifying the noise at certain frequencies, since $\overline{\delta}$ and Q_n obey inverse relationships. Consequently, one must find a compromise between noise propagation and quality of the reconstruction[76].

Another consideration that enters into the design of an inverse filter is the influence of the filter on the signal and the noise spectra. In order to study this influence, the signal and the noise amplification factors, as well as the relative signal-to-noise ratio, were calculated at different θ_{obj}. Analytical expressions for f_n,

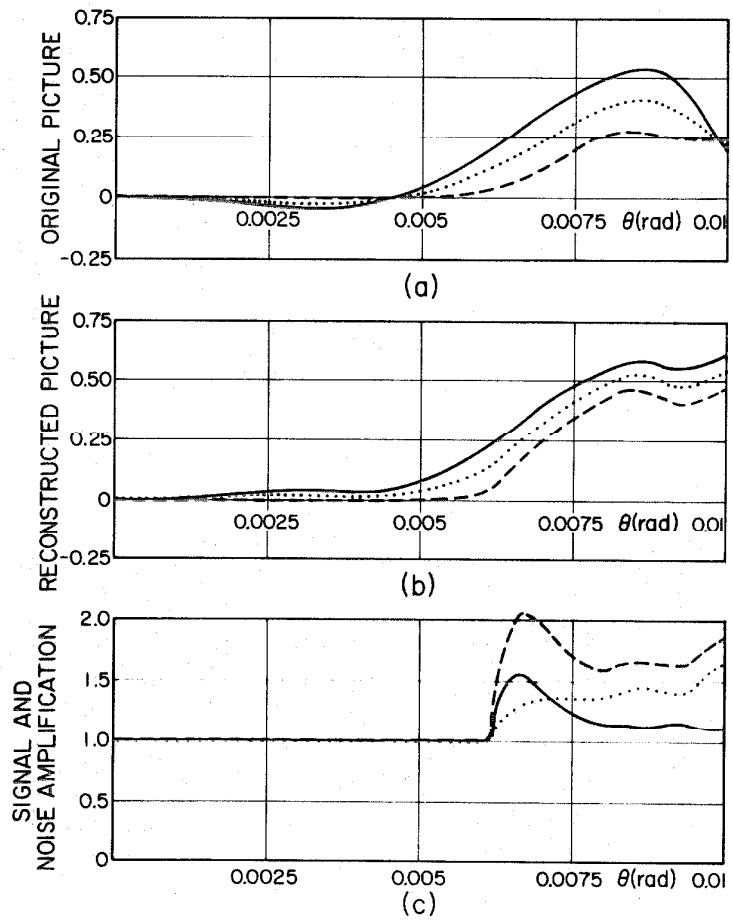


Fig. 3-10 Linfoot's image evaluation criteria are plotted against θ_{obj} for $\Delta Z = 860\text{Å}$ and $\overline{\delta} = 0.5$ in (a) the original picture, (b) the reconstructed picture. Ψ_r is shown as a dotted line, Ξ_r as a dashed line, and Φ_r as a solid line. The signal and noise amplification factors are plotted against θ_{obj} in (c): $\tilde{f}_n$ is shown as a dotted line, $\tilde{f}_d$ as a dashed line, and $\tilde{\tau}_r$ as a solid line. Inspection of $\tilde{\tau}_r$ reveals that a gain in contrast has been obtained in a narrow band of frequencies, which is caused by the large threshold $\overline{\delta}$.

f_d, and τ_r were previously derived for a general filter (see Sec. 2.3.5.2C). Now for a circularly symmetric filter, we can convert these expressions into

$$\tilde{f}_n = \frac{2}{\theta_{obj}^2} \int_0^{\theta_{obj}} \frac{g^2(\theta)}{\sin^2\gamma(\theta)\ E^2(\theta)}\ \theta\ d\theta \tag{3.73}$$

$$\tilde{f}_d = \frac{\int_0^{\theta_{obj}} g^2(\theta)\theta\ d\theta}{\int_0^{\theta_{obj}} \sin^2\gamma(\theta)\ E^2(\theta)\theta\ d\theta} \tag{3.74}$$

$$\tilde{\tau}_r = \tilde{f}_d / \tilde{f}_n \tag{3.75}$$

These three quantities have been plotted in Fig. 3.10c as a function of θ_{obj}. An examination of these plots reveals that $\tilde{\tau}_r$ increases markedly in a narrow frequency band near 6.6×10^{-3} rad, where the PCTF does not reach its maximum. At higher frequencies, $\tilde{\tau}_r$ then tapers off gradually; a fact which indicates that no significant gain in contrast is achieved. Consequently, an increase in $\tilde{\tau}_r$ in selected ranges of frequencies is governed by the threshold $\overline{\delta}$ used in the restoration.

The success of an object reconstruction will depend mostly on the choice of an appropriate value for $\overline{\delta}$. This choice will depend on the noise level when the transferred signal is hardly detectable above background. If our weak-phase object assumption is valid, an inverse filter will then extend the instrumental resolution.

3.3.3.3 Computer algorithm

The block diagram in Fig. 3-11 shows the computer algorithm used in inverse filtering. First, the brightness levels of the row data are spread over the full dynamic range by applying a linear stretch to the original gray scale (see Sec. 2.3.1). Then, in order to determine the contrast of the specimen with respect to the substrate background, such statistical properties as the mean and the variance are listed locally. For instance, for the gold on carbon specimen the signal-to-noise ratio, SNR_i, of the i^{th} cluster can be defined as

$$SNR_i = \sqrt{\frac{\sigma_{d_i}^2 + \overline{d}_i^2}{\sigma_{sub}^2 + \overline{d}_{sub}^2}} \tag{3.76}$$

where d_i is the intensity of the i^{th} cluster and d_{sub} is the intensity of the substrate. Next, the area of the picture where the substrate is imaged is Fourier transformed, and its frequency components are then fitted to $|\sin \gamma(\underset{\sim}{k})|$ (see Sec. 3.3.3.1). If this area contains periodic information, spikes in the spectrum will alter the convergence of the iteration loop during the least squares fit. In order to remove these spurious spikes, we can interpolate the Fourier components across the spikes in such a way as to suppress the periodic pattern (this procedure can be compared to the coherent noise removal algorithm of Sec. 2.3.3.1). Finally, by changing the amount of smoothing on the raw spectrum and the size of the frequency domain over which the fit is calculated, one determines the optimum electron optical parameters. The optimum electron optical parameters are subsequently fed into the program that calculates the inverse filter function.

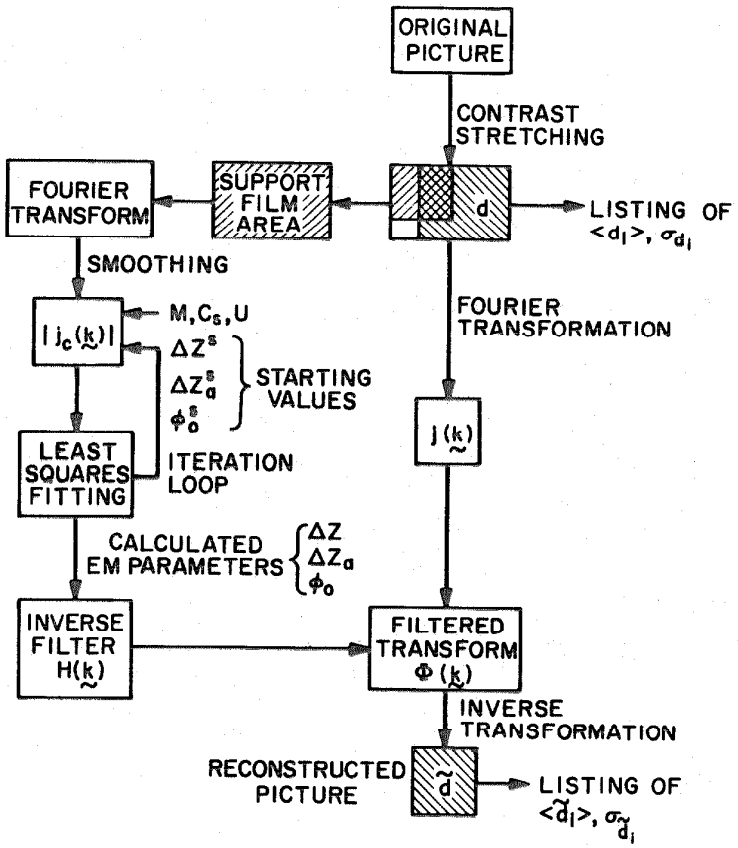


Fig. 3-11 Block diagram illustrating the sequence of processing steps involved in restoring a weak-phase object

Multiplication of the original picture transform by $H(k)$, followed by an inverse transformation then yields a reconstructed picture. In order to evaluate the resulting gain in information in a quantitative fashion, the SNR's of various regions in the original picture are compared to their equivalent in the processed picture.

The inverse filtering scheme we have just presented suffers from many limitations in practice. For instance the object phase distribution can only be reconstructed over a frequency band whose limits are a function of the noise background and of the electron microscope parameters. Therefore, parts of the transform which lie near the zeros of the transfer function will be missing in the processed image. This procedure also neglects the amplitude modulation terms; therefore, it is invalid for specimens that contain heavy atoms. In such specimens, amplitude contrast mechanisms contribute significantly to the overall contrast. Consequently, to obtain a complete description of a weakly scattering specimen, its complex structure factor should be retrieved over the complete resolution range. We shall now present a technique which combines electron micrographs recorded at various defocus values, in order to overcome the above limitations. By separating the phase and the amplitude contrast, this technique also allows one to discriminate between heavy and light elements.

3.3.4 Heavy/Light Atom Discrimination Technique

3.3.4.1 Theory

A) Introduction

The problem of restoring the complex object transmission function amounts to extracting $\Phi(\underset{\sim}{k})$ and $M(\underset{\sim}{k})$ from the transform of a

bright-field image (cf. Eq.(3.3)). A minimum of two micrographs that have been taken under different conditions are required to solve the two unknowns of this problem. A processing scheme, invented by P. Schiske[57], restores the sine- and the cosine-transferred part of the object transform by combining the transforms of N bright-field pictures taken under different defocus conditions. In this formalism the specimen is described using complex scattering factors and the spatial locations of its constituent atoms. Within certain approximations $\Phi(\underset{\sim}{k})$ is found to be linearly related to the real part, and M(k) to the imaginary part of the scattering factor. The imaginary part of the scattering factor is often referred to as the anomalous scattering. The anomalous scattering exhibits a strong dependence on atomic number, and thus can be used as an atom discrimination tool. In the following discussion we shall derive an explicit expression for the image transform in terms of $f_j(\theta)$ and $\underset{\sim}{r}^j_{o}$. Next we shall describe the algorithm that calculates the complex structure factor.

B) General formalism

The diffracted wave which has undergone a single elastic scattering event in a three-dimensional structure has been derived in Eq. (1.8). In the bright-field image this wave is recombined at the image plane with the primary beam that has been attenuated by both elastic and inelastic scattering as it propagates through the specimen, and with an inelastic scattered wave, $\psi_{inel}(\underset{\sim}{r}_i)$. Plural scattering events which introduce second-order terms will be neglected in this discussion. We therefore assume that the specimen thickness does not exceed the mean-free path for a single elastic scattering event. Let

us now examine the various factors that describe the image wave function.

The incident wave is spatially modulated by scattering events, so that we can write $\psi_o'(\underset{\sim}{r}_o)$ at the exit surface of the specimen as

$$\psi_o'(\underset{\sim}{r}_o) = A_o[1 - q_\sigma(\underset{\sim}{r}_o)t_o]^{1/2} \simeq A_o[1 - \frac{1}{2} t_o q_\sigma(\underset{\sim}{r}_o)] \tag{3.77}$$

where $q_\sigma(\underset{\sim}{r}_o)$ is the total cross section for elastic and inelastic events. Aberrations in the objective lens then modify the modulated primary beam, which at the image plane becomes

$$\psi_o'(\underset{\sim}{r}_i) = \frac{A_o}{M}[1 - \frac{1}{2} t_o \int b(\underset{\sim}{k})Q_\sigma(\underset{\sim}{k})\exp(-i\gamma(\underset{\sim}{k}) + 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k}] \tag{3.78}$$

where $Q_\sigma(\underset{\sim}{k})$ denotes the Fourier transform of $q_\sigma(\underset{\sim}{r}_o)$. In this fashion we can account for the occurrence of slight deviations from purely kinematical scattering, as long as single scattering remains the dominant mechanism.

It is interesting to note that spatial modulations in the incident beam, caused by inelastic collisions, are transferred to the image plane even in the case where inelastically scattered electrons are filtered out by an energy analyzer[62]. Since $\psi_{inel}(\underset{\sim}{r}_i)$ is incoherent with respect to $\psi_s(\underset{\sim}{r}_i)$ and is blurred by chromatic aberration, we shall ignore $\psi_{inel}(\underset{\sim}{r}_i)$ in the ensuing analysis. The wave function at the image plane can then be written as

$$\psi(\underset{\sim}{r}_i) = \psi_o'(\underset{\sim}{r}_i) + \psi_s(\underset{\sim}{r}_i) \tag{3.79}$$

so that the recorded bright-field intensity is now expressed as

$$|\psi(\underset{\sim}{r}_i)|^2 = \frac{A_o^2}{M^2}\left|\{1+\int b(\underset{\sim}{k})\,[i\lambda \sum_{j=1}^{N_a} |f_j(\theta)| \cdot \exp(i\eta_j(\theta) - i\pi z_o^j\theta^2/\lambda - 2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_o^j) - \frac{t_o}{2}Q_\sigma(\underset{\sim}{k})] \cdot \exp(-i\gamma(\underset{\sim}{k}) + 2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k}\}\right|^2 \qquad (3.80)$$

The derivation of the image transform is analogous to that presented in Appendix B, and leads to the result

$$j(\underset{\sim}{k}) = 2b(\underset{\sim}{k})\,E(\underset{\sim}{k})[\lambda \sum_{j=1}^{N_a} |f_j(\theta)| e^{-2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} \sin[\gamma(\theta,\phi) - \eta_j(\theta) - \gamma z_o^j\theta^2/\lambda] - \frac{t_o}{2}Q_\sigma(\underset{\sim}{k})\cos\gamma(\underset{\sim}{k})] \qquad (3.81)$$

where the dark-field contribution to image contrast is neglected. This term is a quadratic function of $\Phi(\underset{\sim}{k})$ and $M(\underset{\sim}{k})$; therefore it destroys the linearity of $j(\underset{\sim}{k})$. When the dark-field intensity is not negligible, an iteration procedure presented in Sec. 2.3.3.4 can be used to evaluate its magnitude with respect to the other contrast terms.

If we now group the factors in Eq. (3.81) that describe the phase and the amplitude contrast, we are led to the following identities

$$\tilde{\Phi}(\underset{\sim}{k}) = \lambda \sum_{j=1}^{N_a} |f_j(\theta)| \cos[\eta_j(\theta) + \pi z_o^j\theta^2/\lambda] e^{-2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} \qquad (3.82a)$$

$$\tilde{M}(\underset{\sim}{k}) = \lambda \sum_{j=1}^{N_a} |f_j(\theta)| \sin[\eta_j(\theta) + \pi z_o^j\theta^2/\lambda] e^{-2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} - \frac{t_o}{2}Q_\sigma(\underset{\sim}{k}) \qquad (3.82b)$$

The above expressions can be compared to the complex structure factor defined by Eq. (1.14). We shall show in the next section that $\lambda F(\underset{\sim}{k})$ is approximately equal to $\tilde{\Phi}(\underset{\sim}{k}) + i\tilde{M}(\underset{\sim}{k})$. Therefore, by reconstructing the complex object transmission function we also attain selective contrast enhancement.

C) Discrimination effect

i) Principles

Let us now assume that dynamic effects responsible for the incident beam amplitude modulation are negligible. We shall also assume that the specimen thickness is such that the phase shifts, $\pi z_o^j \theta^2/\lambda$, due to the deviations of the j^{th} atom from the mean defocus, can be ignored. Then we can approximate $\tilde{\Phi}(\underset{\sim}{k})$ and $\tilde{M}(\underset{\sim}{k})$ by

$$\Phi(\underset{\sim}{k}) = \lambda \sum_{j=1}^{N_a} |f_j(\theta)| \cos \eta_j(\theta) e^{-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j} = \lambda F_r(\underset{\sim}{k}) \tag{3.83}$$

and

$$M(\underset{\sim}{k}) = \lambda \sum_{j=1}^{N_a} |f_j(\theta)| \sin \eta_j(\theta) e^{-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j} = \lambda F_i(\underset{\sim}{k}) \tag{3.84}$$

The phase and amplitude object modulations can now be envisioned as being a convolution of a set of delta functions representing the projected locations of the atoms with the atomic image functions due to the real or to the anomalous scattering:

$$\phi(\underset{\sim}{r}_o) = \lambda \sum_{j=1}^{N_a} \delta(\underset{\sim}{r}_o - \underset{\sim}{r}_o^j) * \mathcal{F}^{-1}\{f_j'(\theta)\} \tag{3.85}$$

$$\mu(\underset{\sim}{r}_o) = \lambda \sum_{j=1}^{N_a} \delta(\underset{\sim}{r}_o - \underset{\sim}{r}_o^j) * \mathcal{F}^{-1}\{f_j''(\theta)\} \tag{3.86}$$

Consequently, the restoration will produce two images of the same two-dimensional structure, whose differing contrast dependence on the atomic number can be used to separate the images of heavy atoms from light atoms.

We shall now estimate the relative contributions of the real and the imaginary image (i.e., $\phi(\underset{\sim}{r}_0)$ and $\mu(\underset{\sim}{r}_0)$) to the overall contrast of a single atom. The predicted contrast will be based upon available calculations of the atomic scattering amplitudes for an isolated atom. Since we shall ignore the alterations of the atomic inner potential by molecular binding, these contrast estimates provide only an approximate measure of the potential discrimination effect that can be expected in the restored image.

ii) Discrimination effect for single atoms

We have shown that $\phi(\underset{\sim}{r}_0)$ is imaged with optimum contrast when the aberration phase shift is an odd multiple of $\pi/2$. The atomic contrast is then entirely due to the real part of the scattering amplitude and so can be described by (cf. Eq. (1.37))

$$C_a^P(\rho_i) = 2\lambda \int_0^{\theta_{obj}} f_j'(\theta)\exp(2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k} =$$

$$= \frac{4\pi}{\lambda} \int_0^{\theta_{obj}} f_j'(\theta)\, J_0(2\pi\theta\rho_i/\lambda)\, \sin\theta\, d\theta$$

Correspondingly, we have already mentioned that, when $\gamma(\underset{\sim}{k})$ is an integer multiple of π, the observed image contrast is a direct map of the amplitude modulation term and can be represented by (cf. Eq. (1.22)):

$$C_a^A(\rho_i) = 2\lambda \int_0^{\theta_{obj}} f_j''(\theta)\, \exp(2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i)\, d\underset{\sim}{k}$$

$$= \frac{4\pi}{\lambda} \int_0^{\theta_{obj}} f_j''(0)\, J_0(2\pi\theta\rho_i/\lambda) \sin\theta\, d\theta$$

It is apparent that the Z dependence of both contrast mechanisms is governed simultaneously by $|f_j(\theta)|$ and by $\eta_j(\theta)$. In the first Born approximation, calculations have led to a $Z^{4/3}$ dependence of the elastic scattering cross section. The elastically scattered wave appears linearly in the image contrast and therefore obeys only a $Z^{2/3}$ dependence (i.e., $|f_j(\theta)| \propto Z^{2/3}$). In order to determine the variations in anomalous phase shift angle with atomic number, polynomial expressions for $|f_j(\theta)|$ and $\eta_j(\theta)$ were fitted to Thomas-Fermi-Dirac calculations of the complex atomic scattering amplitudes[79]. The resulting polynomials are

$$|f_j(\theta)| = \exp\{\sum_{i=0}^{4} \alpha_{i,j}^{\dagger} [\frac{4\pi}{\lambda} \sin(\theta/2)]^i\} \qquad (3.87a)$$

and

$$\eta_j(\theta) = \sum_{i=0}^{4} \beta_{i,j}^{\dagger} [\frac{4\pi}{\lambda} \sin(\theta/2)]^i \qquad (3.87b)$$

where the coefficients $\alpha_{i,j}^{\dagger}$ and $\beta_{i,j}^{\dagger}$ have been tabulated by Haase for atomic numbers ranging from 6 to 92. These calculations indicate a Z^n $(2/3 \leq n \leq 1)$ relationship for the sine of the anomalous phase shift angle. Consequently, we can expect a $Z^{n+2/3}$ increase in the anomalous scattering contribution with increasing atomic number (this can be compared to a Z discrimination in scanning transmission microscopy that is achieved by imaging separately the elastic and the

inelastic waves[72], and to a $Z^{4/3}$ separation effect in conventional dark-field microscopy).

In order to illustrate the dependence of $f_j'(\theta)$ and $f_j''(\theta)$ on atomic number, we shall plot these quantities and their relative ratio against θ for atoms of gold and carbon. The polynomial expressions of Eq. (3.87a,b) were employed to calculate the curves shown in Fig. 3-12. However, a more realistic estimate of the separation effect can be achieved by comparing, as a function of θ_{obj}, the maxima of $C_a^P(\rho_i)$, $C_a^A(\rho_i)$, and the image contrast that is observed. By neglecting axial astigmatism, the latter can be evaluated from:

$$C_a^T(\rho_i) = \frac{4\pi}{\lambda} \int_0^{\theta_{obj}} [f_j'(\theta)\sin\gamma(\theta) - f_j''(\theta)\cos\gamma(\theta)] \cdot E(\theta) J_1(2\pi\theta\rho_i/\lambda)\sin\theta\, d\theta \tag{3.88}$$

We then calculate the integrals defining C_a^P, C_a^A, and C_a^T, at the atomic image center for various objective aperture half-angles. The resulting curves are shown in Fig. 3-13 for gold and carbon. The ratio of the integrals calculated for gold to those calculated for carbon gives the Z sensitivity of the respective images. For instance, at $\theta_{obj} = 10^{-2}$rad, one finds a ratio of 5:1 for the phase image, 43:1 for the amplitude image, as compared to 7.5:1 in the original image. These numbers are only estimates of a potential discrimination that can ideally be achieved in a specimen composed of bound atoms. In addition, the restoration procedure imposes practical limitations to the selective enhancement of different atomic species. We have previously listed the various contributions to the cosine-transferred part of the

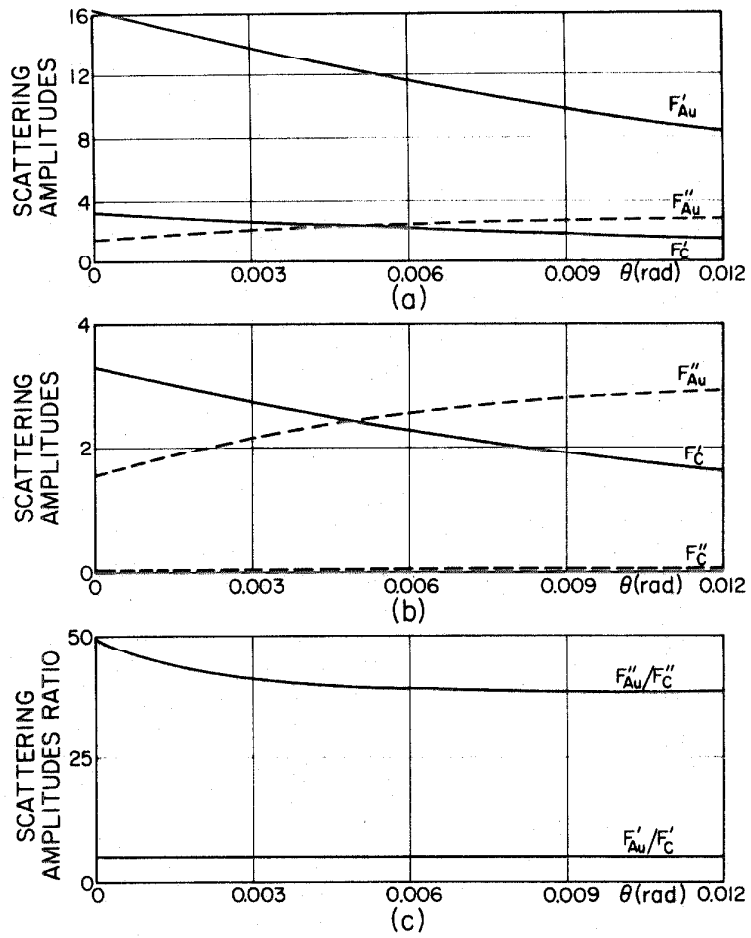


Fig. 3-12 Real and imaginary parts of the atomic scattering amplitudes for a gold atom and for a carbon atom. The vertical scale in Å has been enlarged from (a) to (b) to illustrate f''_C. The relative magnitude of the potential discrimination effect is shown in (c).

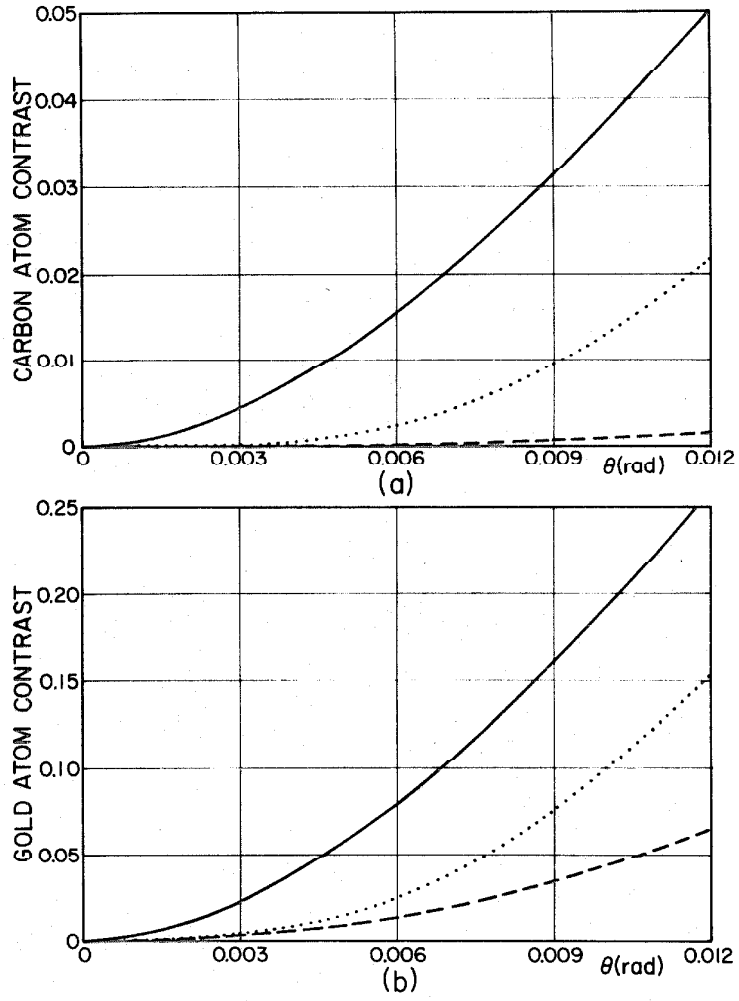


Fig. 3-13 Phase contrast, amplitude contrast, and image contrast observed at $\Delta Z = 95\text{Å}$ for (a) a carbon atom and (b) a gold atom. The solid lines correspond to $C_a^P(0)$, the dotted lines to $C_a^T(0)$, and the dashed lines to $C_a^A(0)$. A marked increase in $C_a^A(0)$ is visible for the gold atom.

transform (cf. Eq. (3.82b)), and shown that the partial breakdown of the kinematic approximation introduces an additional amplitude contrast term. This contribution contains low resolution information that can be filtered out during the processing. We will now investigate the way in which the specimen thickness affects the atomic discrimination technique.

iii) Effect of the curvature of the Ewald sphere

It can be shown that the curvature of the Ewald sphere introduces a thickness dependent phase angle, $\pi z_o^j \theta^2/\lambda$, in addition to the anomalous phase angle (see Appendix B). This is because the real part and the imaginary part of the scattering amplitudes are scrambled up in the amplitude contrast term, which now transfers information on an effective anomalous scattering amplitude given by[62] (cf. Eq. (B.13)):

$$f''_{j\,eff}(\theta) = f''_j(\theta)\cos(\pi z_o^j \theta^2/\lambda) + f'_j(\theta)\sin(\pi z_o^j \theta^2/\lambda) \qquad (3.89)$$

The additive term in Eq. (3.89) that depends on $f'_j(\theta)$ will then counteract the effectiveness of the method, since f''_j/f'_j varies as Z^n where $2/3 \leq n \leq 1$. Moreover, this additional contribution is highest at the top and at the bottom layers of the specimen, where $|z_o^j| = t_o/2$. In order to illustrate how the imaginary part of the scattering amplitude is affected, the original and the effective anomalous scattering terms are plotted in Fig. 3-14 as a function of scattering angle θ. An examination of Fig. 3-14 reveals that at an aperture half-angle of 10^{-2}rad, the contribution from the Ewald sphere (i.e., $\simeq f'_j(\theta)\sin(\pi z_o^j \theta^2/\lambda)$ exceeds the anomalous term $f''_j(\theta)$ by a

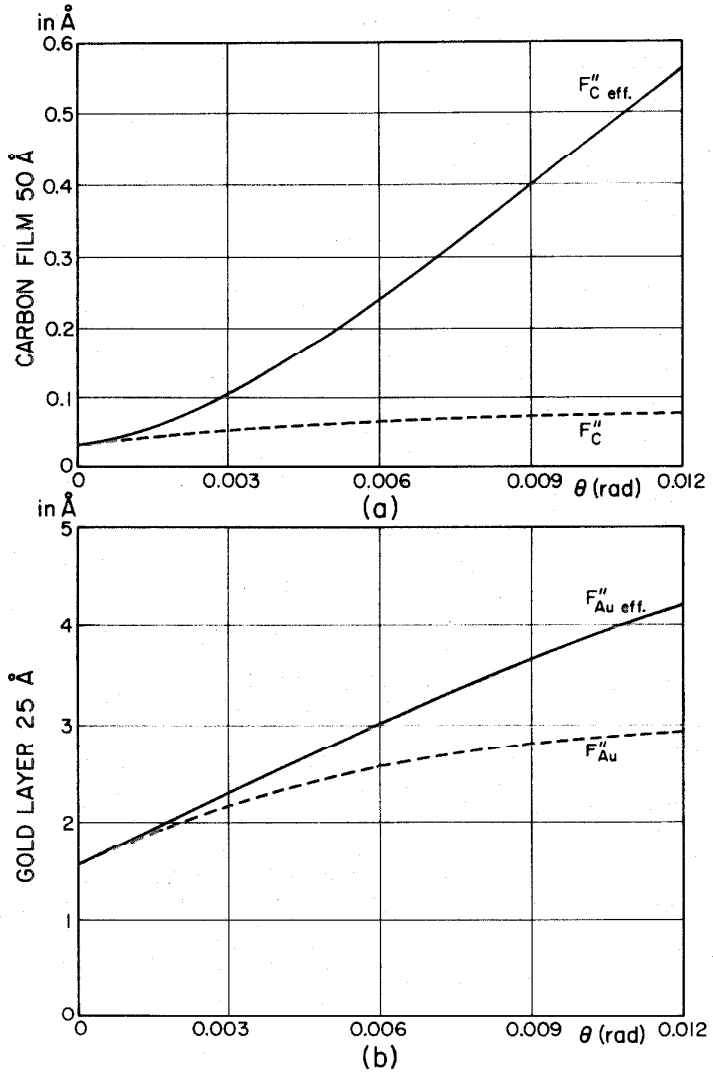


Fig. 3-14 Anomalous scattering amplitudes shown as dashed lines for $f''_j(\theta)$, and as solid lines for $f''^{eff}_j(\theta)$ where the curvature of the Ewald sphere adds a correction factor. They are plotted as a function of the scattering angle for (a) a carbon film of 50Å, (b) a gold layer of 25Å. In the case of carbon, we notice the strong influence of thickness effects on the anomalous scattering image.

factor of 5 for carbon and a factor of 0.34 for gold. It is therefore imperative to work with ultra-thin support films in order to achieve selective constrast enhancement at very high resolutions. (In Sec. 3.5.1.1 a technique will be described for preparing carbon support films approximately 15Å thick.)

On several occasions we have stated that specimen thickness limits the ultimate resolution of a microscopic image; at the same time it attenuates the selective enhancement of different atomic species. In order to overcome this limitation, a reconstruction method has been proposed by Hoppe, where the primary beam is tilted through a series of angles and the micrographs are combined in order to obtain a restoration that is free of Ewald sphere artifacts[73]. Such three-dimensional reconstruction schemes have been successfully implemented at the medium resolution range for amplitude objects. Their extension to the atomic resolution level is feasible, although the complexity in the data analysis and specimen contamination represent major obstacles to such processing.

3.3.4.2 Schiske restoration method

A) Principle

We shall now describe an algorithm that restores both the phase and the amplitude of the transmission function of a weakly scattering object. In order to simplify the notation, we shall describe the object by a complex object function, $o(\underset{\sim}{r}_o)$, defined as

$$o(\underset{\sim}{r}_o) = \phi(\underset{\sim}{r}_o) + i\mu(\underset{\sim}{r}_o) \tag{3.90}$$

The object spectrum can then be written as

$$S_0(\underset{\sim}{k}) = \delta(\underset{\sim}{k}) + i\, O(\underset{\sim}{k}) = \delta(\underset{\sim}{k}) + i[\Phi(\underset{\sim}{k}) + iM(\underset{\sim}{k})] \tag{3.91}$$

so that the image transform is now expressed as

$$j_\ell(\underset{\sim}{k}) = ib(\underset{\sim}{k})\, E(\underset{\sim}{k})[O(\underset{\sim}{k})e^{-i\gamma_\ell(\underset{\sim}{k})} - O^*(-\underset{\sim}{k})e^{i\gamma_\ell(\underset{\sim}{k})}] \tag{3.92}$$

It is clear from the above equation that both $O(\underset{\sim}{k})$ and $O^*(-\underset{\sim}{k})$ can be determined if more than one measurement of $[j_\ell(\underset{\sim}{k}),\gamma_\ell(\underset{\sim}{k})]$ is made for different $\gamma_\ell(\underset{\sim}{k})$. In practice, these measurements require a spatial registration of micrographs taken under different conditions, and an evaluation of the operating electron optical parameters. Cross-correlation methods will therefore be examined in the next section, because they are subject to various anomalies in the case of bright-field images. The determination of the phase functions has already been examined in Sec. 3.3.3.1, where a computer least-squares filtering procedure was described. In the following discussion we shall analyze a reconstruction scheme that was first proposed by Schiske and combines micrographs taken at different defocusings[57].

In the Schiske restoration method, a set of data $[j_\ell(\underset{\sim}{k}),\gamma_\ell(\underset{\sim}{k})]$, $(\ell=1,2,\cdots,N)$, is recorded as a focus series, and then combined by a least-squares method that solves the overdetermined system of Eq. (3.92) for the two unknowns, $O(\underset{\sim}{k})$ and $O^*(-\underset{\sim}{k})$. In order to introduce certain problems associated with this restoration scheme, we shall first consider the case where only two pictures are combined.

Two pictures taken at different defocus are sufficient, in theory, to solve the system given by Eq. (3.92) for the complex object

transform. Solving Eq. (3.92) for $O(\underset{\sim}{k})$ yields

$$O(\underset{\sim}{k}) = \frac{j_1(\underset{\sim}{k})\, e^{i\gamma_2(\underset{\sim}{k})} - j_2(\underset{\sim}{k})\, e^{i\gamma_1(\underset{\sim}{k})}}{2E(\theta)\, \sin[\gamma_1(\underset{\sim}{k}) - \gamma_2(\underset{\sim}{k})]} \tag{3.93}$$

for $\theta < \theta_{obj}$ and $\gamma_2 \neq \gamma_1 + m\pi$. A singularity will occur in the reconstructed picture when the denominator of Eq. (3.93) vanishes, that is, at

$$\theta_{sing} = \left(\frac{\lambda m}{\Delta Z_{st}}\right)^{1/2}, \quad m=1,2,\cdots \tag{3.94}$$

where ΔZ_{st} is the defocus step between the two pictures (we are assuming that ΔZ_a and ϕ_o are nearly identical in the two pictures). The first discontinuity (i.e., m=1) does not appear in the spatial frequency range of interest for small enough ΔZ_{st}. For example, in the region $\theta \leq 10^{-2}$rad, this singularity will only occur if $\Delta Z_{st} \geq 370\text{Å}$. It therefore suffices to select smaller defocus step than this critical value. In practice, however, a restoration with only two pictures creates overwhelming noise artifacts near the zeros of the combined contrast transfer function.

When radiation damage to the specimen is not a critical parameter, it is advantageous to record a large number of pictures for use in the reconstruction, in order to reduce the noise amplification. A least squares procedure is then used to solve the over-determined system of equations for $O(\underset{\sim}{k})$. By writing Eq. (3.92) as

$$e^{i\gamma_\ell(\underset{\sim}{k})}\, j_\ell(\underset{\sim}{k}) = [iE(\underset{\sim}{k})\, O(\underset{\sim}{k})] - [iE(\underset{\sim}{k})O^*(-\underset{\sim}{k})]e^{2i\gamma_\ell(\underset{\sim}{k})} \qquad (3.95)$$

where $(\ell=1,2,\cdots,N)$, we have converted our system into a form: $y_\ell = \tilde{A} + \tilde{B}\, x_\ell$ for which the least-squares technique can easily be implemented. The least-squares criterion determines $\tilde{A}$ and $\tilde{B}$ at the minimum of

$$\Delta_{sc} = \sum_{\ell=1}^{N} |\tilde{y}_\ell - \tilde{A} - \tilde{B}\, x_\ell|^2$$

by setting in turn the partial derivatives of Δ_{sc} with respect to $\tilde{A}^*$ and $\tilde{B}^*$ to zero. We than obtain two normal equations in the two unknowns $\tilde{A}$ and $\tilde{B}$ that yield for $\tilde{A}$ the following expression:

$$\tilde{A} = \frac{(\sum_{\ell=1}^{N} |x_\ell|^2)(\sum_{\ell=1}^{N} y_\ell) - (\sum_{\ell=1}^{N} x_\ell)(\sum_{\ell=1}^{N} x_\ell^* y_\ell)}{N \sum_{\ell=1}^{N} |x_\ell|^2 - |\sum_{\ell=1}^{N} x_\ell|^2} \qquad (3.96)$$

After replacing the values $\tilde{A}$, x_ℓ , and y_ℓ by their equivalent values in Eq. (3.95), we can transform Eq. (3.96) into Schiske's formula for the complex object transform:

$$O(k) = -\frac{i}{E(\theta)} \frac{\sum_{\ell=1}^{N} j_\ell(\underset{\sim}{k})\, e^{i\gamma_\ell(\underset{\sim}{k})} \{N - \sum_{m=1}^{N} e^{2i[\gamma_m(\underset{\sim}{k}) - \gamma_\ell(\underset{\sim}{k})]}\}}{N^2 - \left|\sum_{\ell=1}^{N} e^{2i\gamma_\ell(\underset{\sim}{k})}\right|^2} \qquad (3.97)$$

For $N=2$ one retrieves, as expected, the formula shown in Eq. (3.93).

When the curvature of the Ewald sphere and plural scattering events are neglected, the above restoration scheme yields the complex structure factor, since

$$O(\underset{\sim}{k}) = \lambda F(\underset{\sim}{k}) = \lambda[F_r(\underset{\sim}{k}) + iF_i(\underset{\sim}{k})] = \lambda \sum_{j=1}^{N_a} |f_j(\theta)| \exp(i\eta_j(\theta) - 2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_o^j) \tag{3.98}$$

By inverse transforming $O(\underset{\sim}{k})$, we retrieve the complex image:

$$o(\underset{\sim}{r}_o) = \mathcal{F}^{-1} \{b(\underset{\sim}{k})\, O(\underset{\sim}{k})\} \tag{3.99}$$

whose imaginary part is the anomalous scattering image that should display selective contrast enhancement at heavy-atom locations.

We have already examined the effect of specimen thickness and of the partial breakdown of the kinematical approximation on the restoration. We shall now analyze the effects of experimental errors on the evaluation of Schiske's formula, and the impact of these errors on the contrast interpretation of the restored image.

B) Evaluation of Schiske's formula

Prior to the retrieval of $O(\underset{\sim}{k})$, as computed in Eq. (3.97), the mutual positions of the pictures in the through-focus series and the respective phase factors $\gamma_\ell(\underset{\sim}{k})$ must be found. We shall now investigate the sensitivity of the calculated object transform to errors in positioning the images and in evaluating $\gamma_\ell(\underset{\sim}{k})$.

i) Cross-correlation step

The mutual translational positions of two micrographs are determined by calculating the cross-correlation matrices between two areas of the input pictures. The location of the correlation peak, however, is subject to uncertainties which will be discussed in Sec. 3.3.4.3. These uncertainties result in the picture being shifted from its true matching position by an amount $\underset{\sim}{r}_D$; so that according to the

shift theorem, the corresponding image transform will undergo a phase shift

$$\exp(-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_D) = \exp[-\frac{2\pi i}{\lambda}|\underset{\sim}{r}_D|\theta\cos(\phi - \phi_D)] \qquad (3.100)$$

where ϕ_D is the azimuthal angle of the translation vector $\underset{\sim}{r}_D$. We see from Eq. (3.100) that high spatial frequencies in the restored image will be significantly distorted if $\underset{\sim}{r}_D$ is not kept to a minimum. In order to minimize this spurious artifact, a very fine scanning step size is chosen in the early stages of the processing, when the pictures are being spatially registered.

ii) Determination of the phase factors

If the specimen is supported by a carbon substrate, a least-squares fit, between the power spectrum of the substrate and $\sin^2\gamma_\ell(\underset{\sim}{k})$ weighted by a radial function, allows us to determine the phase factors of each picture of the series (see Sec. 3.3.3.1). The least-squares program finds the optimal electron microscope parameters, $(\Delta Z_\ell, \Delta Z_{a\ell}, \phi_{o\ell})$ for each image, over an annulus-shaped frequency band of the diffractogram, whose radii can be varied at will. By using different radii and different starting parameter values, the dispersion in the fitting process has been evaluated to be $\sigma_{\Delta Z} \simeq 60\text{Å}$ for pictures of a through focus series. A decrease in this dispersion can be achieved, once the absolute defocus value of one micrograph has been determined (the focus steps are calibrated to a high accuracy in an electron microscope). For instance, in our experimental work on gold clusters supported by a carbon film, the estimated dispersion was only $\sigma_{\Delta Z} \simeq 30\text{Å}$.

We shall next analyze the sensitivity of the restoration scheme to an error in the determination of the phase factors[59]. Since the defocus difference between successive pictures is a known parameter, we suppose that all phase factor estimates, $\tilde{\gamma}_\ell(\underset{\sim}{k})$, deviate from their true values by the same error, $\Delta\gamma(\theta)$, that is,

$$\tilde{\gamma}_\ell - \gamma_\ell = \Delta\gamma(\underset{\sim}{k}) = \sigma_{\Delta Z}\pi\theta^2/\lambda \quad , \qquad \ell=1,2,\cdots,N \tag{3.101}$$

These experimental uncertainties in the phase factors affect the validity of the restoration, since the calculated object transform, $\tilde{O}(\underset{\sim}{k})$, now deviates from its true value, $O(\underset{\sim}{k})$, by a phase error term. The correspondence between $\tilde{O}(\underset{\sim}{k})$ and $O(\underset{\sim}{k})$ is easily established if we express Eq. (3.97) in the following way:

$$\tilde{O}(\underset{\sim}{k}) = -\frac{i}{E(\theta)} e^{i(\gamma_1+\Delta\gamma)} \frac{\sum_{\ell=1}^{N} j_\ell(\underset{\sim}{k})\, e^{i(\tilde{\gamma}_\ell-\tilde{\gamma}_1)}\left[N-\sum_{m=1}^{N} e^{2i(\tilde{\gamma}_m-\tilde{\gamma}_\ell)}\right]}{N^2-\left|\sum_{\ell=1}^{N} e^{2i(\tilde{\gamma}_\ell-\tilde{\gamma}_1)}\right|^2} \tag{3.102}$$

According to our previous assumption, the phase differences are not subject to an experimental error (i.e., $\tilde{\gamma}_m - \tilde{\gamma}_\ell = \gamma_m - \gamma_\ell$), so that we can write

$$\tilde{O}(\underset{\sim}{k}) = e^{i\Delta\gamma(\underset{\sim}{k})}\, O(\underset{\sim}{k}) \tag{3.103}$$

Consequently, the phase error term $\exp[i\Delta\gamma(\underset{\sim}{k})]$ redistributes the phase and amplitude contributions as

$$\tilde{\Phi}(\underset{\sim}{k}) = \tfrac{1}{2}[\tilde{O}(\underset{\sim}{k}) + \tilde{O}^*(-\underset{\sim}{k})] = \Phi(\underset{\sim}{k}) \cos\Delta\gamma(\underset{\sim}{k}) - M(\underset{\sim}{k})\sin\Delta\gamma(\underset{\sim}{k}) \tag{3.104}$$

and

$$\tilde{M}(\underset{\sim}{k}) = \frac{1}{2}[\tilde{O}(\underset{\sim}{k}) - \tilde{O}^*(-\underset{\sim}{k})] = \Phi(\underset{\sim}{k})\sin \Delta\gamma(\underset{\sim}{k}) + \tilde{M}(\underset{\sim}{k}) \cos \Delta\gamma(\underset{\sim}{k}) \quad (3.105)$$

If we assume that anomalous scattering is the only contribution to the cosine transferred part of the transform, the restoration scheme therefore calculates an imaginary image transform given by (cf. Eq. (3.80))

$$\tilde{F}_i(\underset{\sim}{k}) = \sum_{j=1}^{N_a} |f_j(\theta)| \sin[\eta_j(\theta) + \Delta\gamma(\theta)] e^{-2\pi i \underset{\sim}{k}\cdot \underset{\sim}{r}_o^j} \quad (3.106)$$

where the effective imaginary part of the scattering amplitude has become

$$\tilde{f}''_{j\ eff}(\theta) = f''_j(\theta) \cos \Delta\gamma(\theta) + f'_j(\theta)\sin \Delta\gamma(\theta) \quad (3.107)$$

By comparing Eq. (3.107) with Eq. (3.89), we see that a phase error $\Delta\gamma(\theta)$ has the same adverse effects on the discrimination efficiency as the curvature of the Ewald sphere. The higher-order spatial frequency components in the restored anomalous-scattering image will also be sensitive to an incorrect measurement of the absolute focus values ΔZ_ℓ .

iii) Singularities in Schiske's formula

When the denominator of Eq. (3.97) vanishes, artifacts are introduced in the restoration of the complex object function[59]. These artifacts occur either in the vicinity of $\theta = 0$, or if the focus step assumes a critical value equal to $\Delta Z_{crit} = \lambda/\theta^2_{obj}$ (see Appendix C). In the latter case, the singularity takes place when the difference between the phase factors of two consecutive pictures of the series is

equal to π . We have already pointed out this singularity in Eq. (3.94) when the restoration procedure combined only two micrographs. In order to avoid this singularity, it suffices to select a small enough focus step such that ΔZ_{st} never reaches the critical value ΔZ_{crit} within the object spectrum. For instance, in our experimental analysis we did not encounter this difficulty during the processing, since the focus step size was only 80Å.

When θ approaches zero, it can be shown that $M(\underset{\sim}{k})$ remains finite, whereas $\Phi(\underset{\sim}{k})$ is discontinuous at the origin and goes to infinity as $1/\theta^2$. Hence, in restoring the phase contrast image, we must remove the low-frequency artifacts in the calculated phase transform. This can be achieved by applying a high-pass filter to $\Phi(\underset{\sim}{k})$, and then inverse transforming to obtain $\phi(\underset{\sim}{r}_o)$.

3.3.4.3 Cross-correlation analysis

A) Mutual translational positioning of a focus series

Restoring the object wave function from micrographs taken at different focus conditions requires that the pictures be spatially matched to a high accuracy. In order to find the mutual positions of the pictures of the series, we shall now examine the problems that one encounters in implementing the correlation techniques developed in Sec. 2.3.4.2 for the case of bright-field micrographs. We shall see that the correlation method can lead to false CCF maxima, that are predicted by the linear transfer theory of image formation[80]. In order to simplify the analysis, we shall assume that the specimen structure remains unaltered during successive exposures, and that it only modulates the phase distribution of the incident beam.

i) Theory

Under the above assumptions we are faced with the problem of matching two pictures which are noisy versions of the same weak-phase object, imaged at different transfer conditions. It is also assumed that the noise is additive and uncorrelated to the signal, so that we can describe the image contrast of two pictures of the series by

$$d_1(\underset{\sim}{r}) = t^P_{s1} * \phi_o + n_1(\underset{\sim}{r}) \tag{3.108}$$

and

$$d_2(\underset{\sim}{r}) = (t^P_{s2} * \phi_o) * \delta(\underset{\sim}{r}-\underset{\sim}{r}_D) + n_2(\underset{\sim}{r}) \tag{3.109}$$

where $\underset{\sim}{r}_D$ is the relative translational displacement, $t^P_{s1}(\underset{\sim}{r})$ and $t^P_{s2}(\underset{\sim}{r})$ are the instrument point-spread functions. The transforms of the point-spread functions are (cf. Eq. (3.56))

$$T^P_{s1}(\underset{\sim}{k}) = 2 \sin \gamma_1(\underset{\sim}{k})\, E_1(\underset{\sim}{k}) \tag{3.110}$$

and

$$T^P_{s2}(\underset{\sim}{k}) = 2 \sin \gamma_2(\underset{\sim}{k})\, E_2(\underset{\sim}{k}) \tag{3.111}$$

If we denote by $D_1(\underset{\sim}{k})$ and $D_2(\underset{\sim}{k})$ the two image transforms, then according to the convolution theorem the CCF of the two image distributions can be expressed as

$$\phi_{d_1,d_2}(\underset{\sim}{r}) = \mathcal{F}^{-1}\{D_1(\underset{\sim}{k})\, D_2^*(\underset{\sim}{k})\}$$

Transforming Eq. (3.108) and Eq. (3.109), and then inserting the transforms into the above expression for $\phi_{d_1,d_2}(\underset{\sim}{r})$, we obtain

$$\mathcal{F}\{\phi_{d_1,d_2}(\underset{\sim}{r})\} = [\Phi_o(\underset{\sim}{k})\ T^P_{s1}(\underset{\sim}{k}) + N_1(\underset{\sim}{k})][\Phi^*_o(\underset{\sim}{k})\ T^{*P}_{s2}(\underset{\sim}{k}) e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_D} + N^*_2(\underset{\sim}{k})]$$

$$= |\Phi_o(\underset{\sim}{k})|^2\ T^P_{s1}(\underset{\sim}{k})\ T^{*P}_{s2}(\underset{\sim}{k})\ e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_D} + N^\dagger(\underset{\sim}{k}) \qquad (3.112)$$

The noise transform $N^\dagger(\underset{\sim}{k})$ in Eq. (3.112) consists of three terms whose phases are randomly distributed over the frequency domain B. Provided that B extends far enough so that $\mathcal{F}^{-1}\{N^\dagger(\underset{\sim}{k})\}$ averages out to zero, we can neglect the influence of the noisy background in the digital evaluation of $\phi_{d_1,d_2}(\underset{\sim}{r})$. Discrete Fourier transform algorithms, however, restrict the usable size of B so that noise artifacts can impede the detection of the CCF peak.

If we neglect the noise factor $N^\dagger(k)$ and inverse transform Eq. (3.112), the expression for the CCF reduces to

$$\phi_{d_1,d_2}(\underset{\sim}{r}_i) = (\phi_{\phi_o,\phi_o} * \phi_{t_{s1},t_{s2}}) * \delta(\underset{\sim}{r}_i - \underset{\sim}{r}_D) \qquad (3.113)$$

where ϕ_{ϕ_o,ϕ_o} is the autocorrelation function of the object. Therefore, the CCF of the micrographs is a convolution product of the object ACF with the CCF of the point-spread functions, $\phi_{t_{s1},t_{s2}}(\underset{\sim}{r}_i)$ [80]. We can estimate the half-width of an object ACF by looking at the case of a single atom, for which the ACF becomes

$$\phi_{\phi_o,\phi_o}(\rho_i) = 2\pi \int_0^{\theta_{obj}} f'^2(\theta)\ J_o(2\pi\theta\rho_i/\lambda)\sin\theta\, d\theta \qquad (3.114)$$

Carbon atoms, for instance, possess a very narrow ACF peak ($1\,\text{Å}$ in half-width), since their Wiener spectrum is almost white. In general,

the half-width of a specimen ACF is much narrower than the half-width of $\phi_{t_{s1},s2}(\underset{\sim}{r}_i)$, so that the latter determines the accuracy of the registration. For a non-astigmatic lens, $\phi_{t_{s1},t_{s2}}(\underset{\sim}{r}_i)$ is defined by

$$\phi_{t_{s1},t_{s2}}(\rho_i) = 2\pi \int_0^{\theta_{obj}} 4 \sin \gamma_1(\theta) \sin \gamma_2(\theta) E_1(\theta) E_2(\theta) J_0(2\pi\theta\rho_i/\lambda) \cdot \sin\theta \, d\theta \qquad (3.115)$$

This integral was calculated for the ΔZ_ℓ combinations, measured experimentally in the through-focus series of gold on carbon, after having been normalized with respect to

$$\phi_{t_o,t_o}(0) - 2\pi \int_0^{\theta_{obj}} 4J_0(0) \sin\theta \, d\theta \simeq 4\pi\theta_{obj}^2 \qquad (3.116)$$

which is the maximum value of the ACF of an aberration-free transfer function. Figure 3-15 shows plots of $\phi_{t_{s1},t_{s2}}(\rho_i)$ against ρ_i for several picture pairs. We observe variations in the half-widths of the CCF with the defocus step size between two ΔZ_ℓ. Sign reversals, that transform the peak into a minimum at the origin, occur when the micrographs that are cross-correlated lie on opposite sides of Gaussian focus.

We conclude then that the peak of $\phi_{d_1,d_2}(\underset{\sim}{r}_i)$ can be very broad or even negative for certain combinations of ΔZ_ℓ. The occurrences of possible artifacts in the CCF peak determination must be investigated, especially in the case where no conspicuous image features exist that could be used to confirm the matching positions found by the correlation program. Practical methods to check the

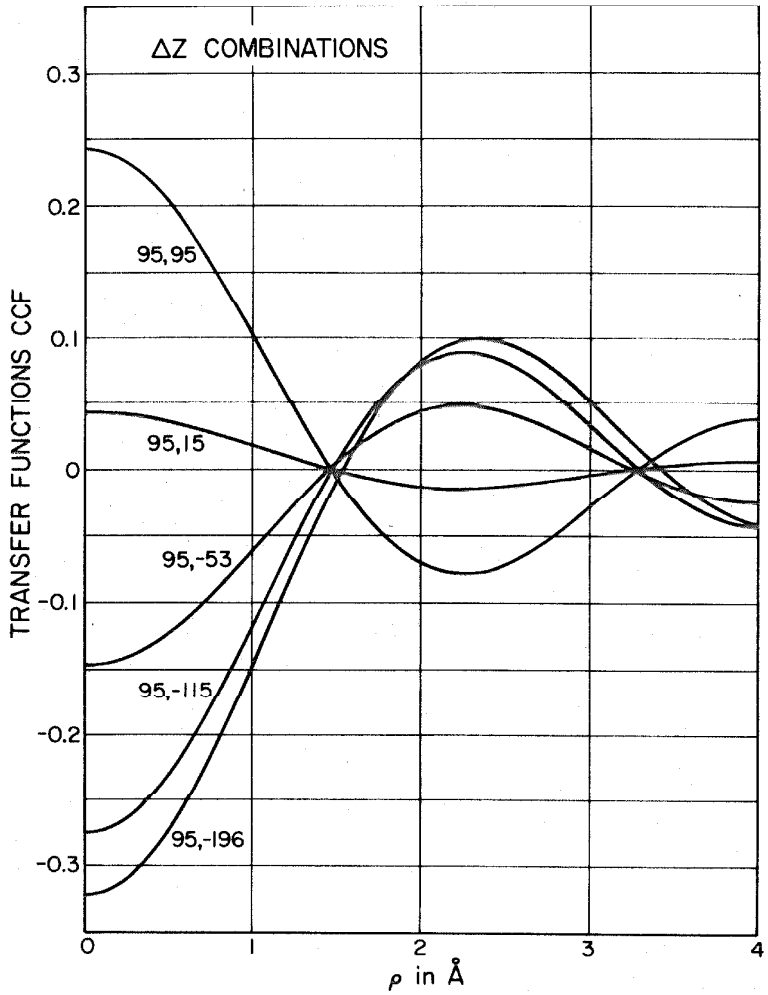


Fig. 3-15 Cross-correlation function between successive pairs of micrographs of the gold on carbon specimen. This figure was determined using Eq. (3.139) and the different ΔZ combinations are indicated.

validity of a picture registration will now be discussed.

ii) Digital implementation

Before cross-correlating two micrographs, a preliminary measurement of their respective phase factors has told us whether or not to expect a sign reversal in their CCF. The discrete evaluation of the correlation integrals can be performed either in real space or in Fourier space, and we shall present the procedures employed to characterize a negative peak in both spaces.

In the direct summation method, maximum correlation is achieved at the peak of (see Eq. (2.63))

$$\phi_{i,j}^{(d_1,d_2)} = \frac{1}{KL} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} d_{k,\ell}^{(1)} \, d_{k+i,\ell+j}^{(2)}$$

One then searches for a minimum peak when it is known that $\phi_{t_{s1},t_{s2}}(\underset{\sim}{r}_i)$ takes negative values for a given objective aperture half-angle.

The Fourier method allows us to optimize the detectability of the CCF peak. The spatial frequency zones whose $T_{s1}^{P}(\underset{\sim}{k}) \, T_{s2}^{P}(\underset{\sim}{k})$ is negative, pull the CCF peak down towards the negative side. Once we know both transfer functions, these zones can be eliminated from the correlation integral. A simple procedure consists in reversing the sign of $D_1(\underset{\sim}{k}) \, D_2^{*}(\underset{\sim}{k})$ in these zones. This amounts to computing an estimate of the CCF equal to

$$\phi_{d_1,d_2}(\underset{\sim}{r}) = \frac{1}{B} \int_B |\Phi_o(\underset{\sim}{k})|^2 \, |T_{s1}^{P}(\underset{\sim}{k})| \, |T_{s2}^{P}(\underset{\sim}{k})| \, e^{2\pi i \underset{\sim}{k} \cdot (\underset{\sim}{r} - \underset{\sim}{r}_D)} \, d\underset{\sim}{k} \qquad (3.117)$$

The technique mentioned above offers two advantages; it prevents sign reversals of the CCF and narrows down the half-width of the CCF peak. However, the finite size of the discrete transform domain can induce aliasing errors and transform artifacts. If the transform domain over which $\phi_{d_1,d_2}(\underset{\sim}{r}_i)$ is computed is too small, the contribution from the noise terms to the CCF integral will not cancel out and will obscure the detection of the CCF peak. Unfortunately, the algorithm which we used to compute

$$\phi_{i,j}^{(d_1,d_2)} = \frac{1}{L} \sum_{m=0}^{K-1} \sum_{n=0}^{L-1} D_{m,n}^{(1)*} D_{m,n}^{(2)} \exp[2\pi i(\frac{mi}{K} + \frac{nj}{L})]$$

required that the complete transform arrays $D_{m,n}^{(1)}$ and $D_{m,n}^{(2)}$ be stored in the core. In our work the core availability limited the size of the input arrays to 64 x 64, so that transform artifacts could not be totally eliminated. In order to check the accuracy of the method, the peak is determined from all possible pairs of pictures in the focus series.

B) Fourier resolution test

The transform of a micrograph indicates which spatial frequencies have been transferred by the instrument to the image plane. Since high-frequency noise components are superimposed upon the signal, it is often difficult to determine the highest resolved detail that has been transmitted to the image. By modulating the signal transform, we are able to overcome inaccuracies in determining $\underset{\sim}{k}_{max}$. This determination is achieved in practice on an area of the specimen which behaves

as a weak-phase object (e.g., a carbon substrate). In this assumption, a simple analysis of the observed modulations in the transform becomes possible.

Let us now study the ACF of the specimen. The half-width of an ACF is a measure of the ultimate instrument resolution, once phase reversals of the transfer function are corrected. According to Eq. (3.112) we can express the power spectrum $|D_1(\underset{\sim}{k})|^2$ as

$$|D_1(\underset{\sim}{k})|^2 = |\Phi_o(\underset{\sim}{k})|^2|T^P_{s1}(\underset{\sim}{k})|^2 + |N_1(\underset{\sim}{k})|^2 + 2\mathrm{Re}[\Phi_o(\underset{\sim}{k})T^P_{s1}(\underset{\sim}{k})N^*_1(\underset{\sim}{k})] \quad (3.118)$$

If we invoke the Wiener-Khinchine theorem, we can retrieve the ACF of the substrate by inverse transforming Eq. (3.118). Because noise and signal are uncorrelated, the last term of this equation averages out to zero in the back transformation, so that

$$\phi_{d_1,d_1}(\underset{\sim}{r}_i) = \phi_{\phi_o,\phi_o} * \phi_{t_{s1},t_{s2}} + \phi_{n_1,n_1}(\underset{\sim}{r}_i) \quad (3.119)$$

We see from Eq. (3.119) that the half-width of the overall ACF is broadened by a noise ACF, and that the peak value of $\phi_{d_1,d_1}(\underset{\sim}{r}_i)$ is modified by this noise contribution. The noise background will mask the outer contrast zones of the diffractogram. These contrast zones lie near the resolution limit set by such instrumental instabilities as mechanical and chromatic aberration.

It is therefore advantageous to modulate the image transform, in order to determine the extension of the Fourier domain that is transferred by the microscope. This modulation is easily achieved by superimposing two pictures, displaced from one another by a vector

$\underset{\sim}{r}_D$ whose magnitude is larger than the minimum resolvable detail[81]. The picture sum is subsequently transformed, either optically or digitally. A system of equidistant fringes which modulate the image transform will then appear in the diffractogram $|D_1(\underset{\sim}{k}) + D_2(\underset{\sim}{k})|^2$. Using Eq. (3.118), we can describe this modulated transform by

$$|D_1(\underset{\sim}{k}) + D_2(\underset{\sim}{k})|^2 = \mathcal{F}\{\phi_{d_1,d_1} + \phi_{d_2,d_2} + \phi_{d_1,d_2} + \phi_{d_2,d_1}\}$$

$$\simeq |\Phi_o(k)|^2[|T^P_{s1}(\underset{\sim}{k})|^2 + |T^P_{s2}(\underset{\sim}{k})|^2 + 2T^P_{s1}(\underset{\sim}{k})T^P_{s2}(\underset{\sim}{k})\cos(2\pi\underset{\sim}{k}\cdot\underset{\sim}{r}_D)]$$

$$+ |N_1(\underset{\sim}{k})|^2 + |N_2(\underset{\sim}{k})|^2 \qquad (3.120)$$

where uncorrelated noise terms have been neglected in the derivation. We notice from Eq. (3.120) that the modulation only affects the CCF transform, $\mathcal{F}\{\phi_{d_1,d_2}\}$, and leaves the noise spectrum unaltered. The fringe pattern runs perpendicular to the displacement vector $\underset{\sim}{r}_D$. Its spatial extent gives a measure of the CCF half-width, and therefore the resolution of the instrument.

This resolution test can also be used to determine the relative phases of the contrast zones in the transforms of bright-field micrographs taken at different defocus[81]. We see from Eq. (3.120) that the fringe system is shifted by half a period whenever the phase contrast transfer functions $\sin\gamma_1(\underset{\sim}{k})$ and $\sin\gamma_2(\underset{\sim}{k})$ have opposite signs. The occurrence of the phase shift in the fringe pattern is seen as a reversal of the fringe contrast in the corresponding zone of the diffractogram. In the normal position, the fringes exhibit a cosine maximum at the origin of the transform. As we shall demonstrate, a shift in the

innermost contrast zone indicates that the two micrographs have defocus values on either side of the Gaussian setting. In this contrast zone we can ignore the spherical aberration term, which is only significant for large scattering angles, and we can eliminate the influence of axial astigmatism by choosing $\underset{\sim}{r}_D$ perpendicular to the reference angle ϕ_o. Under these conditions, the following relationship holds:

$$\text{sign}\{\sin \gamma_i(\underset{\sim}{k})\} = -\text{sign}\{\Delta Z_i\} , \quad i=1,2 \tag{3.121}$$

Therefore, a shift of the fringes will be observed in the innermost zone for $\text{sign}(\Delta Z_1) = -\text{sign}(\Delta Z_2)$. This modulation technique provides a means of visualizing phase reversals directly from a diffractogram.

Figure 3-16 shows the superposition transform of pairs of micrographs from the focus series of gold clusters on carbon. The streaking on the axes of the digital transforms is caused by boundary discontinuities which could be eliminated by apodization. In Fig. 3-16a the spatial extent of the modulations can be measured without ambiguity, since the fringes display a much higher contrast than the noise background. $N\sigma$ shift of the fringe patterns appears in the diffractogram, a fact which one would expect, since the defocusing values are $\Delta Z_1 = 95\mathring{A}$ and $\Delta Z_2 = 15\mathring{A}$. The azimuthal dependence of the fringes shown in Fig. 3-16b is responsible for their poor visibility. In this example, $\Delta Z_1 = 95\mathring{A}$, $\Delta Z_2 = -53\mathring{A}$, and $\Delta Z_a = 145\mathring{A}$; so that contrast reversals will occur when

$$\text{sign}\{\Delta Z_1 + \frac{\Delta Z_a}{2} \sin 2(\phi-\phi_o)\} = -\text{sign}\{\Delta Z_2 + \frac{\Delta Z_a}{2} \sin 2(\phi-\phi_o)\} \tag{3.122}$$

For the above electron optical parameters, Eq. (3.122) is satisfied at

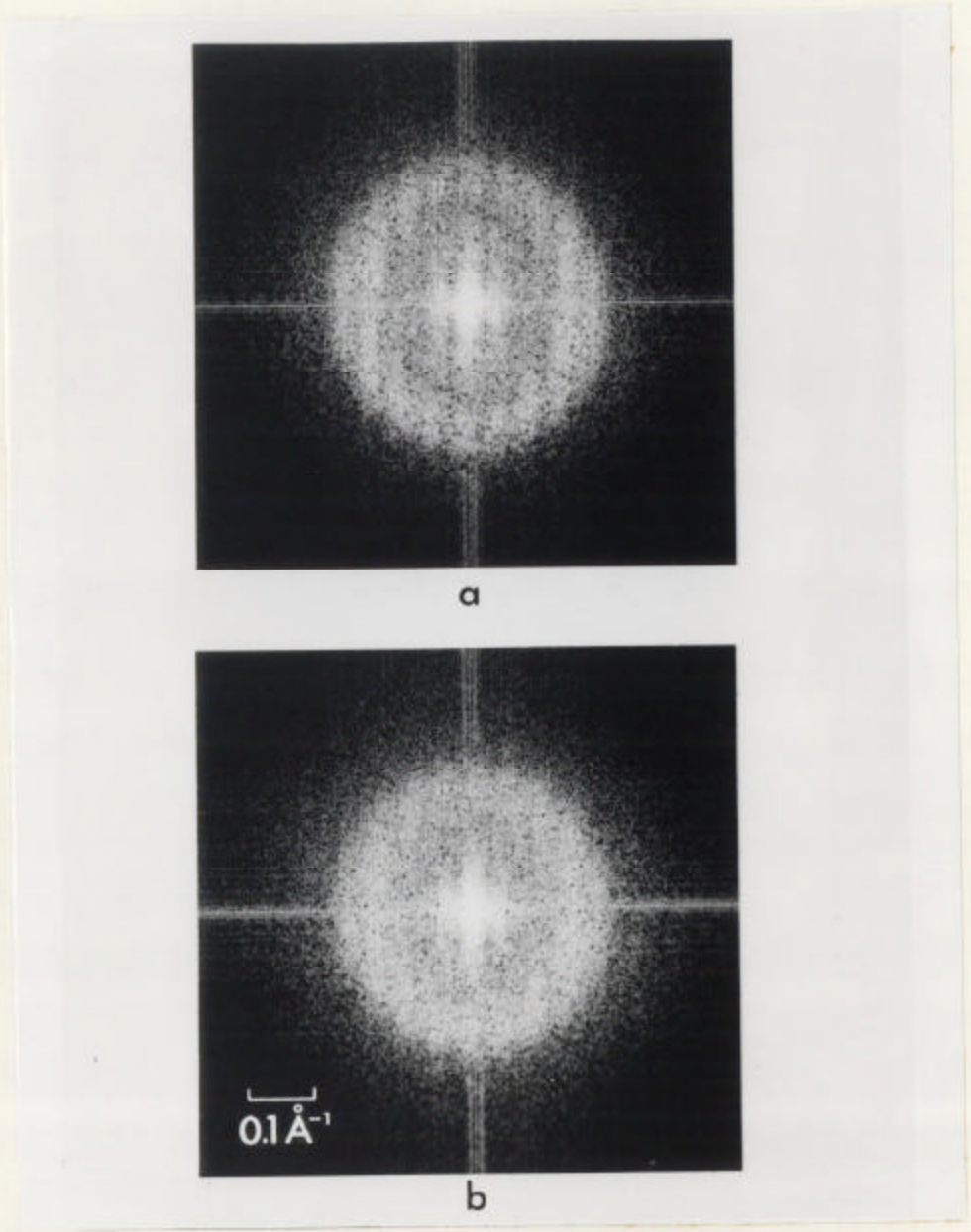


Fig. 3-16 Fourier resolution test using two images of the focus series of gold on carbon. The pair of micrographs have the defocus values: (a) $\Delta Z_1 = 95\text{Å}$, $\Delta Z_2 = 15\text{Å}$, and (b) $\Delta Z_1 = 95\text{Å}$, $\Delta Z_2 = -53\text{Å}$. The modulation fringes have a periodicity of $1/r_D$, where r_D is the relative displacement between the two pictures and is equal to $r_D = 20\text{Å}$. The modulation extends to 0.25Å^{-1}, indicating a point resolution of 4Å.

various zones of the diffractogram. The multiplicity of the resulting fringe contrast reversals is responsible for the observed blurring of the fringe visibility.

The excellent agreement reached between experiment and theory confirms the validity of the linear transfer theory. This theory has enabled us to spatially register high-resolution bright-field micrographs, where defocusing effects have destroyed all visual similarities. We are now capable of implementing the restoration scheme that combines bright-field micrographs of a focus series.

3.3.4.4 Computer processing

We shall now present an outline of the processing algorithm that restores the object wave function by combining pictures of a focus series. The details of its implementation will be discussed in the experimental study of Sec. 3.5. In the following analysis, we assume that anomalous scattering is the main amplitude contrast mechanism; so that the object transform and the complex structure factor are equal within a multiplication constant (i.e., $O(\underset{\sim}{k}) = \lambda F(\underset{\sim}{k})$).

The block diagram of the processing scheme, shown in Fig. 3-17, serves as a guideline for the procedure. The main tasks of the program are threefold: matching the positions of the focus series pictures, evaluating the electron optical parameters for every picture, and computing the complex object structure factor $F(\underset{\sim}{k})$. We shall now describe how these tasks are implemented by computer.

Once scanning with video film converter is completed, a linear stretch of the brightness levels is applied so that they cover the full dynamic range from 0 to 255. An area of the picture showing conspicuous

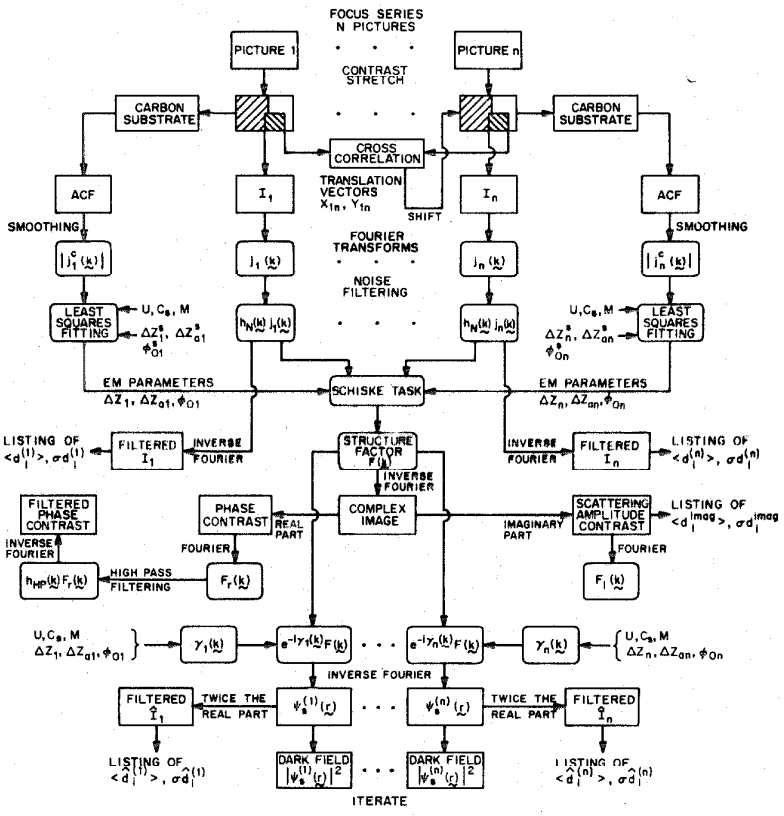


Fig. 3-17 Block diagram of the computer system used in the restoration of the structure factor, $F(\underset{\sim}{k})$, from a defocus series. The least-squares fitting, which yields the electron microscope parameters, was performed on an area containing 128x128 pixels. The cross-correlation matrix was calculated over a portion of the image containing 64x64 pixels.

details is then selected for cross-correlation purposes, after having been adjusted for angular orientation. The resulting matching displacement vectors are then fed into a program which calculates pixel values at intermediate positions of the sampling grid, by bilinear interpolation (see Sec. 2.3.4.2). This latter procedure is repeated for each picture of the series, which is now spatially registered.

The phase factors of every picture are found by an iteration process based on a least-squares fit between $\sin|\gamma_\ell(\underset{\sim}{k})|$ and the transform of an area of the substrate $|j_{\ell c}(\underset{\sim}{k})|$. Starting values for the electron microscope parameters Δz_ℓ^s, $\Delta z_{a\ell}^s$, and $\phi_{o\ell}^s$ are estimated from a preliminary analysis of the optical diffractograms. By applying the Fourier resolution test, one measures the highest spatial frequency, k_{max}, that carries information about the object structure. The phase factors $\gamma_\ell(\underset{\sim}{k})$ are now determined from an optimal least-squares fit computed within the frequency band $|\underset{\sim}{k}| \leq k_{max}$.

Next, the Fourier transforms of the images, $j_\ell(\underset{\sim}{k})$, are calculated with the FFT algorithm. Scanning noise may be apparent on the transform axes, in which case the Fourier coefficients along the axes are set to zero. (This does not cause an appreciable information loss because we are dealing with aperiodic objects). One then applies a bandpass filter to all image transforms of the series. The low-pass component of the filter alternates high frequency noise beyond the resolution limit k_{max}. Since inelastic scattering is mainly confined to low frequencies, the high-pass component will boost up the anomalous contribution to the cosine transferred part of the transform. The desired filter, $h_N(\underset{\sim}{k})$, is defined as:

$$h_N(\underset{\sim}{k}) = \begin{cases} \exp[-(k_{min} - |\underset{\sim}{k}|)^2/2\sigma_{min}^2] & \text{for } 0 < |\underset{\sim}{k}| \leq k_{min} \\ 1 & \text{for } k_{min} < |\underset{\sim}{k}| < k_{max} \\ \exp[-(|\underset{\sim}{k}| - k_{max})^2/2\sigma_{max}^2] & \text{for } k_{max} \leq |\underset{\sim}{k}| \leq k_m \end{cases} \tag{3.123}$$

where σ_{min}, σ_{max} are the Gaussian fall-off half-widths at both ends of the bandpass. The transforms $j_\ell(\underset{\sim}{k})$ are modified by this spatial filter to

$$j'_\ell(\underset{\sim}{k}) = h_N(\underset{\sim}{k})\, j_\ell(\underset{\sim}{k}) \quad , \qquad \ell=1,2,\cdots,N \tag{3.124}$$

An inverse Fourier transformation of $j'_\ell(\underset{\sim}{k})$ yields the noise-filtered series, I'_ℓ, which serves as a reference set for subsequent processing transformations. In order to evaluate the signal-to-noise ratios in various areas of the picture I'_ℓ, mean and standard deviation are listed (cf. Eq. (3.76)).

Finally, the phase factors $\gamma_\ell(\underset{\sim}{k})$ and the modified transforms $j'_\ell(\underset{\sim}{k})$ of every picture of the focus series are combined by the program "Schiske", which outputs the complex object transform $F(\underset{\sim}{k})$. The images of the series are in turn calculated by multiplying $F(\underset{\sim}{k})$ with $ie^{-i\gamma_\ell(\underset{\sim}{k})}$, inverse transforming, and then taking twice the real part of the result (cf. Eq. (1.8) Eq. (3.98)):

$$\hat{I}_\ell = 2\mathrm{Re}(\mathcal{F}^{-1}\{ie^{-i\gamma_\ell(\underset{\sim}{k})} F(\underset{\sim}{k})\}) = 2\mathrm{Re}[\psi_s^{(\ell)}(\underset{\sim}{r})] \tag{3.125}$$

A comparison of these images with the original pictures I'_ℓ provides a check of the validity of the reconstruction. For this purpose, the

difference pictures, $(\hat{I}'_{\ell} - \hat{I}_{\ell})$, are computed, displayed, and their histograms are listed. The average variance, σ_{rm}, gives a reliability measure of the whole scheme; this procedure is analogous to a similar procedure used in X-ray crystallography (see Appendix D). If the match between the reference series and the reconstructed series is satisfactory, one proceeds to compute the inverse transform of $F(k)$, which yields the complex object function

$$o(\underset{\sim}{r}_o) = \phi(\underset{\sim}{r}_o) + i\mu(\underset{\sim}{r}_o)$$

The imaginary part, $\mu(\underset{\sim}{r}_o)$, can be displayed directly, since its transform, $F_i(\underset{\sim}{k})$, is not singular at $\theta = 0$. In this anomalous scattering image one should observe selective contrast enhancement, at heavy atom locations since

$$\mu(\underset{\sim}{r}_o) = \lambda \sum_{j=1}^{N_a} \delta(\underset{\sim}{r}_o - \underset{\sim}{r}_o^j) * \mathcal{F}^{-1}\{f_j''(\theta)\}$$

As for the phase term, $\phi(\underset{\sim}{r}_o)$, it was previously pointed out that $\Phi(\underset{\sim}{k})$ is discontinuous at the origin $\theta = 0$. This singularity can be removed by high-pass filtering of $F_r(\underset{\sim}{k})$ with the Gaussian function

$$h^{HP}(\underset{\sim}{k}) = \begin{cases} \exp[-(|\underset{\sim}{k}| - k_{HP})^2/2\sigma_{HP}^2] & \text{for } |\underset{\sim}{k}| < k_{HP} \\ 1 & \text{elsewhere} \end{cases} \qquad (3.126)$$

where the halfwidth, σ_{HP}, is chosen in such a way as to avoid ringing effects.

An improvement in the restoration scheme can be achieved if one takes the dark field term $|\psi_s^{(\ell)}(\underset{\sim}{r})|^2$ into account (see Appendix D).

The restored object transform $F(\underset{\sim}{k})$ gives a first estimate of $|\psi_s^{(\ell)}(\underset{\sim}{r})|^2$:

$$|\psi_s^{(\ell)}(\underset{\sim}{r})|^2 = \left|\mathcal{F}^{-1}\{ie^{-i\gamma_\ell(\underset{\sim}{k})}F(\underset{\sim}{k})\}\right|^2 \tag{3.127}$$

The above estimate of the dark field term is subsequently subtracted from the corresponding images of the input series. In practice, one iteration is sufficient to eliminate the perturbing influence of this quadratic term in the discrimination program.

3.4 Analysis of a Through-Focus Series of Catalase

3.4.1 Introduction

If the specimen is weakly scattering, we have shown that the complete object wave function can be determined from two micrographs taken under different conditions. The linear theory of image formation involves, however, two important assumptions which limit its applicability. It is assumed in this theory that the specimen is sufficiently thin to be treated as a two-dimensional distribution of atoms, and that the phase shift of the electrons does not vary by more than about $\pi/4$ within a coherently imaged region. Several restoration schemes, based on the linear relationship between object transmittance and image contrast, have been devised in order to compensate for defocusing effects and for contrast artifacts caused by aberrations in the objective lens. The extension of these schemes to thin crystalline specimens will fail if second order effects become dominant at the zeros of the phase contrast transfer function. These effects introduce a nonlinear

relationship between image intensity and projected potential, and their strength depends on the crystallinity of the specimen and on the atomic number of its constituent atoms.

In order to analyze how this theory of image formation applies to thin crystalline specimens, we conducted an experimental study of the variations in the image transform that are induced by changing the phase factors. A negatively stained specimen of bovine liver catalase was selected for this study (its morphology is essentially two-dimensional and it is not overly sensitive to radiation damage). It was found that the variations in image transform with ΔZ could be described reasonably well if one represented the transform by[61] (cf. Eq. (3.25)):

$$j(\underset{\sim}{k}) = 2b(\underset{\sim}{k})\,\Phi(\underset{\sim}{k})[\sin\gamma(\underset{\sim}{k}) + \Xi(\underset{\sim}{k})\cos\gamma(\underset{\sim}{k})]$$

where $\Xi(\underset{\sim}{k})$ is the ratio of scattering versus phase contrast at a spatial frequency $\underset{\sim}{k}$. Unfortunately, specimen preparation technique limits the meaningful structural information to about 20Å resolution, when one observes ordered protein crystals in the dry state[82]. As a result, defocusing effects were the predominant influence in this medium resolution experiment. In the following discussion, we shall investigate the effects of ΔZ on the phase and the amplitude of the diffracted beams for each micrograph of the catalase focus series. Finally, we shall present an image restoration scheme.

3.4.2 Experimental Procedure

3.4.2.1 Specimen preparation

A suspension of bovine liver catalase crystals, fixed in glutaraldehyde, was obtained from Ladd Research Industries. In order to support the catalase specimens, a method for preparing microgrids containing extremely fine holes (0.5 μm in diameter) was developed. In this method, one deposits a thin plastic film (biodene, 0.5% wt/vol in methyl acetate) upon a previously cooled glass slide. Temperature gradients existing during condensation produce a cloud of ultra-fine water particles, which pierce holes through the plastic film. The film is lifted from the water using a 400 mesh copper grid, and then coated with a thin carbon layer ($\simeq$ 200Å). In order to dissolve any plastic remaining on the film, it is steamed in methyl acetate vapor. Finally, drops of the catalase suspension were deposited on these grids. Then, just before the suspension dries, a drop of negative stain (1% uranyl acetate) is added, blotted, and allowed to dry[83].

3.4.2.2 Electron microscopy

Bright-field micrographs of catalase were taken on a JEOL 100B operating at 100 keV and a magnification of 75,000 X, using a 200 μm second condenser aperture and a 120 μm objective aperture. The preparation technique produces large and very thin platelets, some of which run across microholes. The area of interest was chosen and a through focus series of twelve pictures, with focus steps of ΔZ_{st} = 400Å was then recorded.

3.4.2.3 Optical transforms

The electron optical parameters were determined for each micrograph from their optical diffractograms shown in Fig. 3-18. The absence of streaks and of fringe systems in these diffractograms indicates that no specimen drift has occurred during the exposure. In addition to the noisy background, the diffractograms showed diffraction spots that corresponded to the catalase crystals. The defocusing and astigmatism were then obtained by measuring the radii of the outermost rings, which correspond to phase contrast, in the diffractogram. Next, inserting these values into the phase contrast characteristics of the instrument, we directly read the parameters ΔZ_{ℓ} and $\Delta Z_{a\ell}$. The relative focus difference between two successive pictures is known from ΔZ_{st}. An absolute estimate of the defocus values is then achieved by finding the best fit for the sequence of ΔZ_{ℓ}. As an additional check, we can also estimate ΔZ_{ℓ} and $\Delta Z_{a\ell}$ from an analysis of the innermost elliptical zone. This zone corresponds to amplitude contrast transfer. The measured values of ΔZ_{ℓ} are listed in Table 3-1 and are accurate to within 100Å. The average-axial-astigmatism focus difference and the reference angle were found to be: $\Delta Z_a = 380$Å and $\phi_o = -62^o$.

A qualitative understanding of the effects of defocusing on the diffraction pattern can be obtained by examining the intensity spikes in each transform. Near Gaussian focus, phase contrast vanishes in the resolution range of interest; the image is due mainly to amplitude contrast and exhibits low frequency details. Most of the high-resolution components are missing, so that the corresponding micrographs show little morphological information and weak contrast. At higher defocusing

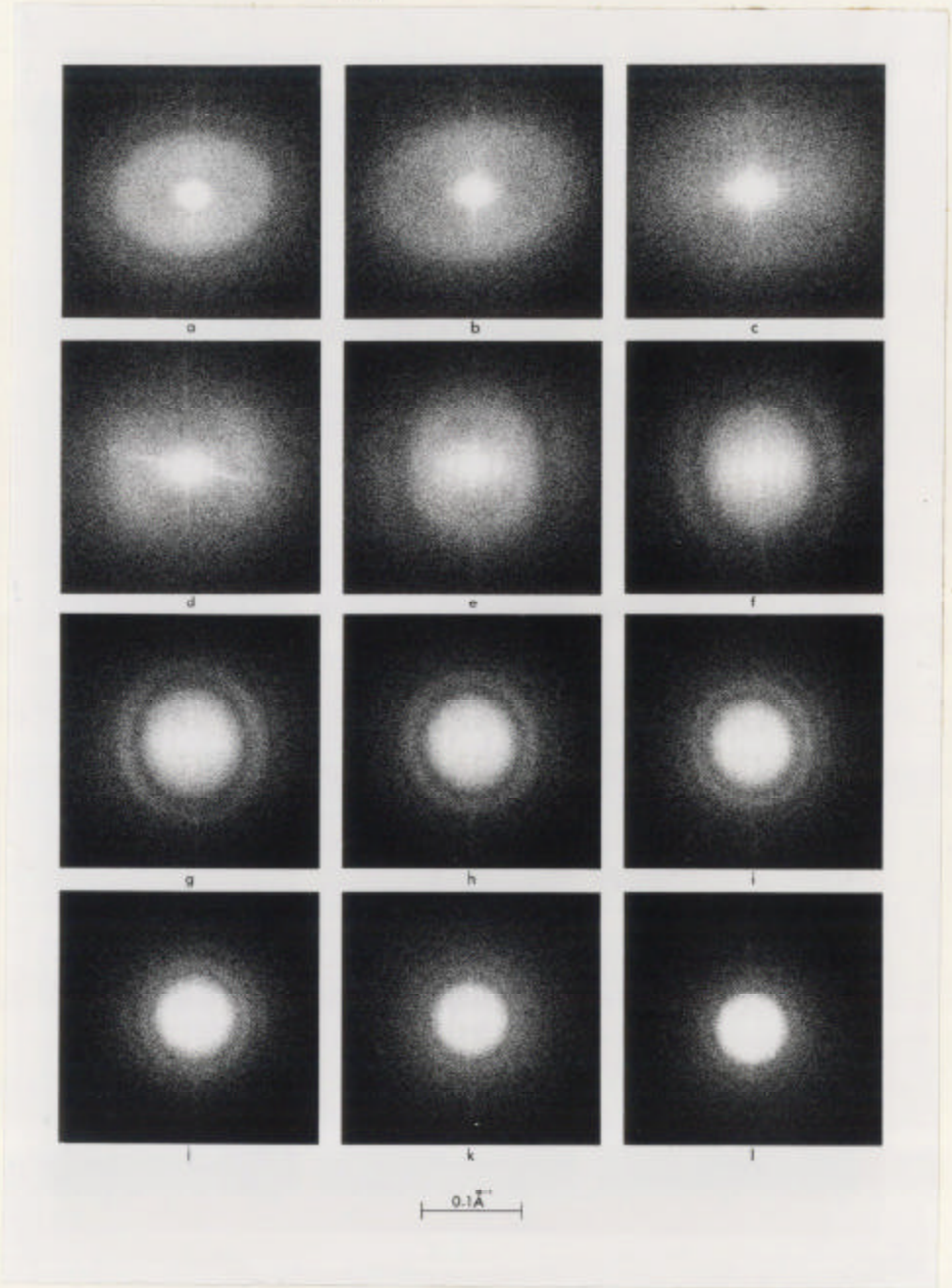


Fig. 3-18 Optical diffractograms of the through-focus series of bright-field micrographs of uranium-stained ox-liver catalase. Diffraction spots can be seen in the amplitude transferred central contrast zone. The outermost phase contrast zones are used to evaluate the electron optical parameters (ΔZ_ℓ, $\Delta Z_{a\ell}$, $\phi_{o\ell}$) .

Table 3-1 Estimates of the defocusing values for the micrographs of catalase shown in Fig. (3-20). The defocusing values were obtained from the optical transforms shown in Fig. (3-18). The axial astigmatism of the micrographs is described by $\Delta Z_a \simeq 380\text{Å}$ and $\phi_o \simeq -62^o$.

Micrograph	ΔZ_ℓ (Å)
a	-650
b	-250
c	150
d	550
e	950
f	1350
g	1750
h	2150
i	2550
j	2950
k	3350
l	3750

enhancement in phase contrast brings out the higher-order diffraction spots, which were previously hidden in the noise background of the diffractogram. The specimen contains no visible diffraction spot beyond 0.0348Å^{-1}; so that the smallest visible structural detail in these underfocused micrographs is 29Å. Underfocus contrast enhancement yields a sharp and pleasing image for $\Delta Z = 3750$Å (see Fig. 3-20ℓ). Here little attenuation of the high frequency components takes place. These visual observations are indeed confirmed by a quantitative analysis of the numerical transforms.

3.4.3 Quantitative Analysis of the Transforms

3.4.3.1 Video input-output

The bright-field micrographs of the catalase series were copied from the electron microscope plates onto fine grain film. An area of 12.8 x 12.8 mm corresponding to 1700 x 1700Å on the object scale was scanned with a VFC at a sampling density of $\Delta = 25$ μm for each micrograph of the series. The equivalent scanning distance is 3.3Å at 75,000X magnification, and thus does not impair the resolution. This very fine step size was chosen in order to ensure proper alignment of the pictures during the early stage of the processing. During the processing the calculated images were displayed at a resolution of 50 μm by the VFC. In the case of the Fourier transforms, the amplitudes $|D(\underset{\sim}{k})|$ were displayed logarithmically, since their range of values spans several orders of magnitude. This representation yields the same intensity distribution as an optical diffractogram, because of the logarithmic response of a film emulsion to light.

3.4.3.2 Processing algorithm

The block diagram shown in Fig. 3-19 illustrates the various stages of the processing. An area, 512 x 512 pixels, was scanned at a sampling density of Δ = 25 μm for each micrograph of the series. The histograms were saturated to 1% on either tail by linearly stretching the pixels about the mean. A box holding five principal lattice lines was cut out and converted into a square area. During this process local lattice distortions and angular mis-orientations were compensated for by applying a geometrical operator. The resulting pictures were then matched translationally by cross-correlating over an area equal to 128 x 128 pixels. We have previously mentioned that artifacts can obscure the correlation peak of bright-field electron micrographs taken at various defocusing. In our work, the peak halfwidth was strongly dependent upon the respective ΔZ_ℓ of the picture pairs, and reached its minimum for consecutive pictures of the series. The correlation peak was then found by bilinear interpolation; the matching displacements that we obtained proved to be accurate to within half a pixel despite the lack of conspicuous details in certain pictures of the series.

After completion of the correlation step, the sampling step size was doubled to Δ = 50 μm (equivalent to 6.7Å on the object scale) without causing any aliasing error, since the object resolution is limited to 29Å. A box, 128 x 128 pixels, was cut out for further processing. This box contained 5 horizontal principal-lattice lines, each separated by 172Å; each line consisted of 13 unit cells (each cell is 67Å wide). Figure 3-20 shows this area after geometrical alignment

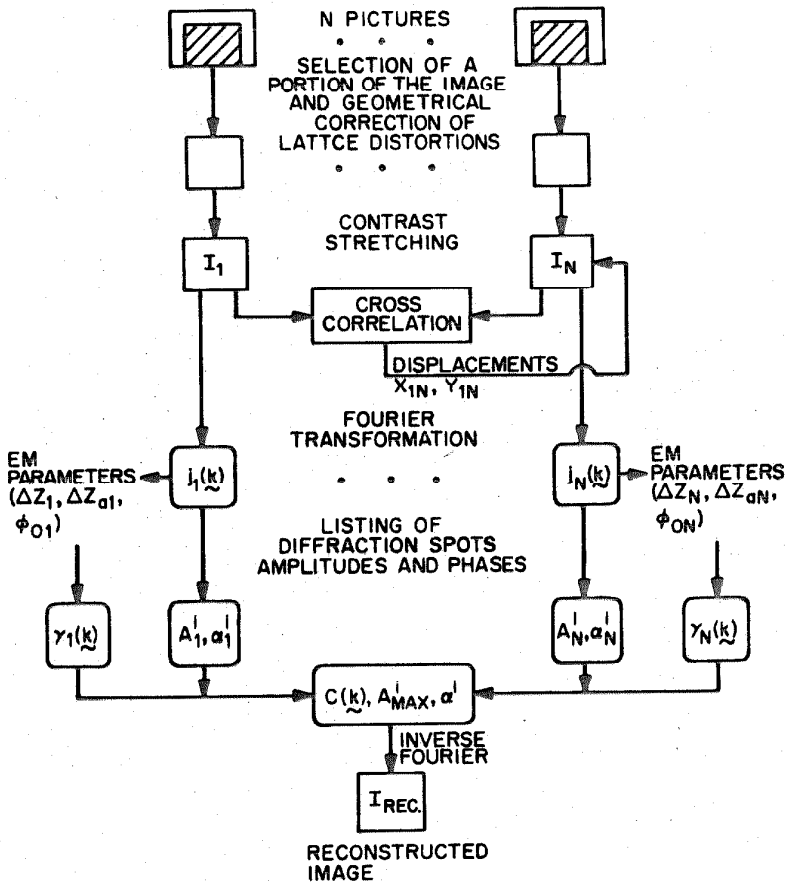


Fig. 3-19 Block diagram of the computer system used to experimentally determine the variations of the catalase transform with changes in the defocusing. The cross-correlation matrix was computed over an area equal to 64x64 pixels, and the Fourier transform over an area 128x128 pixels. Next the amplitudes A^i_n and the phases α^i_n of the diffraction spots are listed. The scaled amplitudes yield an experimental transfer function that is used to reconstruct an image free from defocusing artifacts.

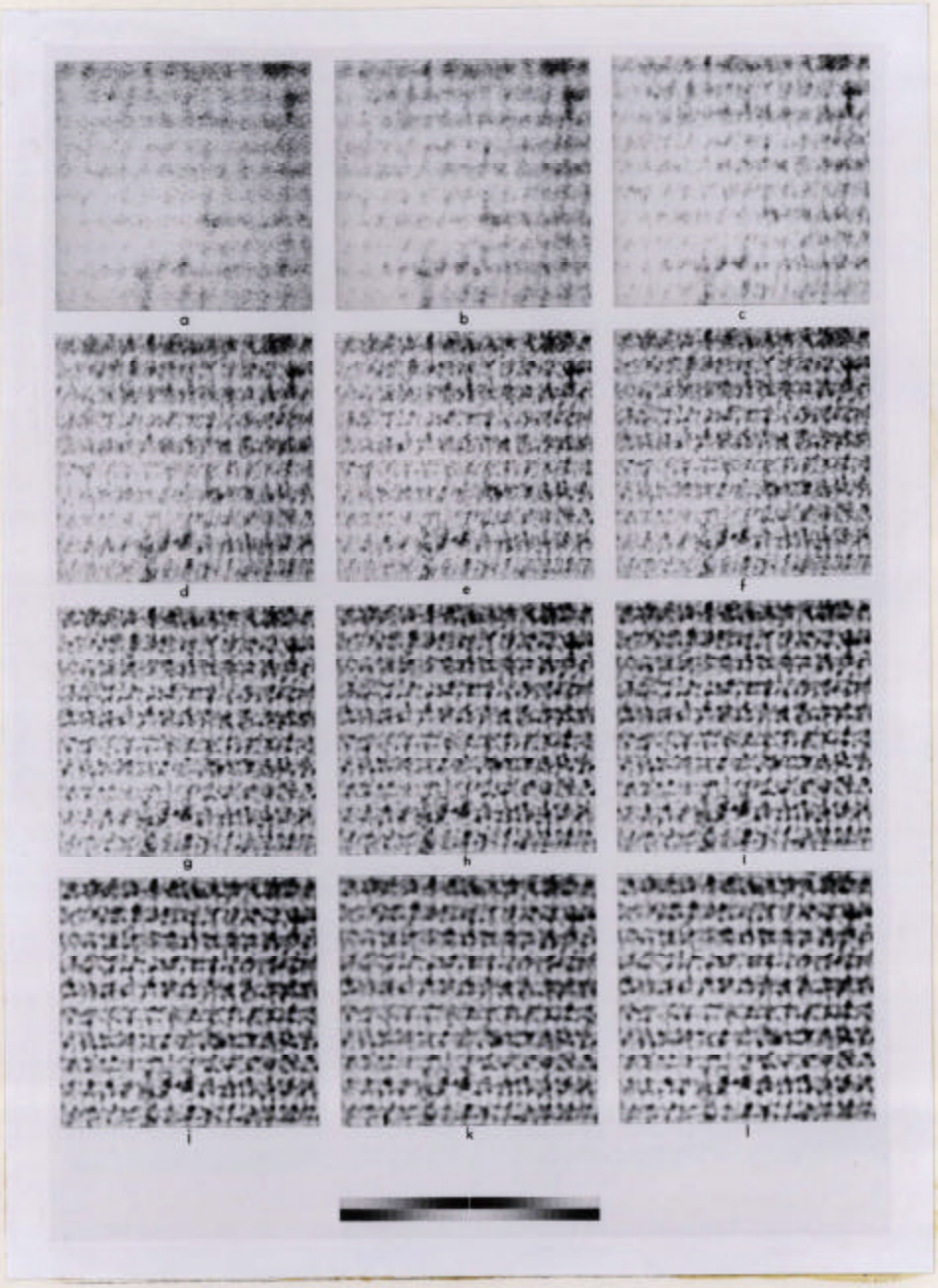


Fig. 3-20 Micrographs of the catalase focus series after contrast stretching and mutual positioning by cross-correlation. The apparent lattice spacings in the micrographs are 67Å in the horizontal and 86Å in the vertical direction. The defocusing values corresponding to each picture are listed in Table 3-1.

for each picture of the focus series. The molecules along any one row alternate in orientation, so that the repeat distance is twice the intermolecular separation along this direction (i.e., 2 x 86Å). The observed morphological unit of the crystal agrees well with the proposed structure of four subunits with 222 symmetry[83,84]. The structure of catalase in the dry state is orthorhombic with pseudo-hexagonal axes lying across the rectangles in the direction of the rows of molecules[85].

The periodicity in the image is apparent in the computed Fourier transform. This transform yields the phase, as well as the amplitude, of the diffraction spots; thus it allows us to make a quantitative analysis of the effects of defocusing on the image transform. Boundary artifacts were minimized in the discrete transform by setting the average optical density around the perimeter of the input array to zero. This procedure prevented possible truncation spikes from interfering with the diffraction spots. A digital transform of a catalase picture, shown in Fig. 3-21, demonstrates the absence of these spikes on both axes of the transform; we also notice the close resemblance between this discrete transform and the corresponding continuous optical transform shown in Fig. 3-18ℓ.

3.4.3.3 Analysis of the transforms

In computing these transforms, the phase origin was chosen to be approximately at the intersection of the glide axis and the mirror line of the catalase unit cell. Local bending in the periodic array manifest itself as an asymmetry with respect to the meridian line. This twisting of the lattice can be seen in Fig. 3-22, which is a

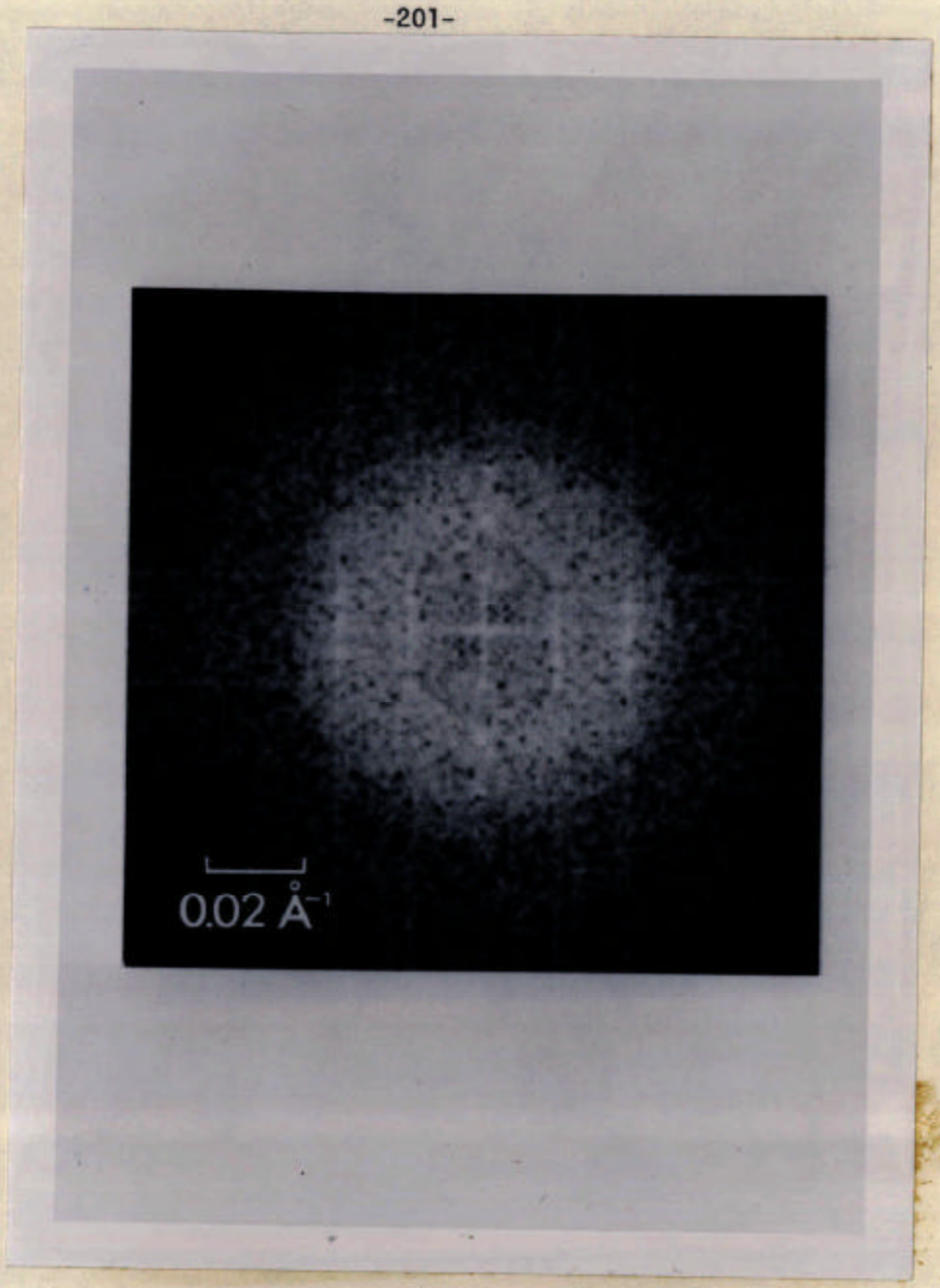


Fig. 3-21 Digital transform of Fig. 3-20ℓ of the catalase focus series, computed over an area of 128x128 pixels. Granularity in the transform is caused by the small size of the input array. No transform artifacts, induced by the truncation, are visible. In order to show the diffuse ring corresponding to phase contrast, this transform is displayed in a logarithmic representation.

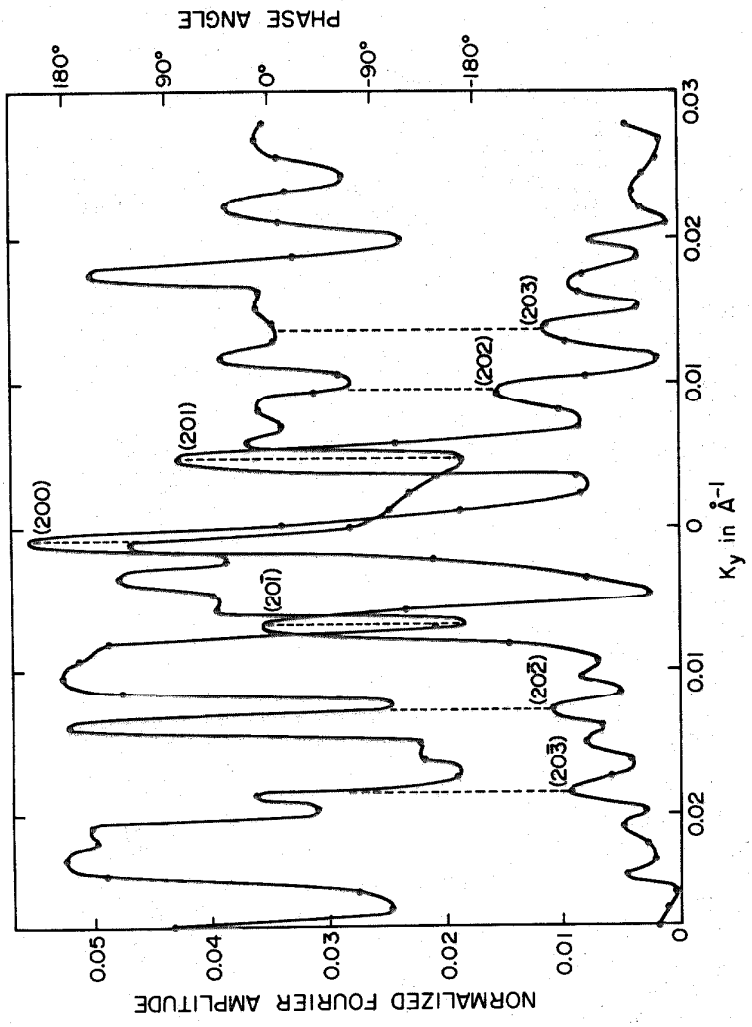


Fig. 3-22 Section of the computer generated transform of Fig. 3-21 along the line $K_x = 2 \times 1/67 Å^{-1}$ (i.e., h=2 in Fig. 3-23). The Fourier amplitude is shown in the bottom curve, and the accompanying phase angle is shown in the top curve. Notice the asymmetry of the profiles with respect to the origin; this indicates a strong lattice distortion of the input pictures.

plot of the Fourier amplitude, A^i_ℓ, along the line $h = 2$. In this plot we also observe rapid oscillations of the accompanying phase angle. Since the numerical transform is discretely sampled on a grid, the amplitude maxima and the respective phases of every diffracted beam were determined by bilinear interpolation across a grid spacing.

As theoretically predicted, we did not detect sizeable variations in the phases of the diffracted beams in the range of ΔZ_ℓ selected. The observed fluctuations of the phase angles (less than 15 degrees) of the beams are mainly caused by the limited accuracy of the translational matching. A relative displacement of $\underset{\sim}{r}_D$ will introduce a phase shift of $\exp(-2\pi i \underset{\sim}{k} \cdot \underset{\sim}{r}_D)$ in the transform; this spurious phase term affects mostly the higher order Fourier components. The phase angles were therefore extrapolated to multiples of $\pi/2$ because of the known plane group symmetry of the crystal. The amount by which the phase origin is displaced from its ideal location can be determined by examining the relative phases of the symmetrically placed Fourier coefficients[86].

The amplitudes of the diffraction spots, after interpolating between sampled points to obtain the true peak values, are listed in Table 3-2. The numerical change in the amplitudes reflects closely the gradual degradation of image contrast as one approaches Gaussian focus. Despite the fact that these Fourier components did not reach their maximum amplitudes at any of the ΔZ_ℓ values, we shall demonstrate that these observations agree with predictions from the linear theory of image formation. In Fig. 3-23 we have sketched the diffraction pattern, corrected for lattice distortions[61,87]. The observed

Table 3.2 Amplitudes of the diffracted beams in each micrograph of the catalase series shown in Fig. 3-20, after normalization by the amplitude of the transmitted beam ($|D(0)| = 10000$). The numbers with asterisks indicate the maxima, A^i_{max}, used in the partial reconstruction of Fig. 3-25c.

Micrograph / Diffracted Beam	a	b	c	d	e	f	g	h	i	j	k	ℓ
(001)	370	435	436*	384	380	349	308	300	300	302	278	309
(002)	699	937	949	1080	1036	1058	1092	1101	1169	1252	1250	1431*
(004)	45	86	149	294	379	510	595	678	786	852	789	892*
(006)	35	25	25	42	75	92	138	156	199	201	174	213*
(100)	303	425	420	452	406	454	405	426	493	451	438	516*
(101)	61	86	79	103	70	94	94	105	125*	113	114	116
$(10\bar{1})$	131	211	236	246	237	228	293	01	292	314	321*	293
(102)	57	68	81	124	101	112	108	124	129	127	132*	126
$(10\bar{2})$	63	125	181	198	205	214	212	222	285	286	256	330*
(200)	25	102	186	284	338	325	435	469*	367	448	446	409
(201)	47	79	133	212	244	246	324	344	292	354*	344	318
$(20\bar{1})$	38	118	130	207	262	224	319	347*	258	317	307	264

Table 3.2 (Continued)

	a	b	c	d	e	f	g	h	i	j	k	ℓ
(202)	15	43	53	79	88	100	88	81	109	94	70	115*
(20$\bar{2}$)	16	31	22	55	84	92	98	108	86	95	94	104*
(203)	38	39	49	62	78	84	100	108*	92	83	56	65
(20$\bar{3}$)	23	33	49	75	93	76	123	124*	87	105	88	70

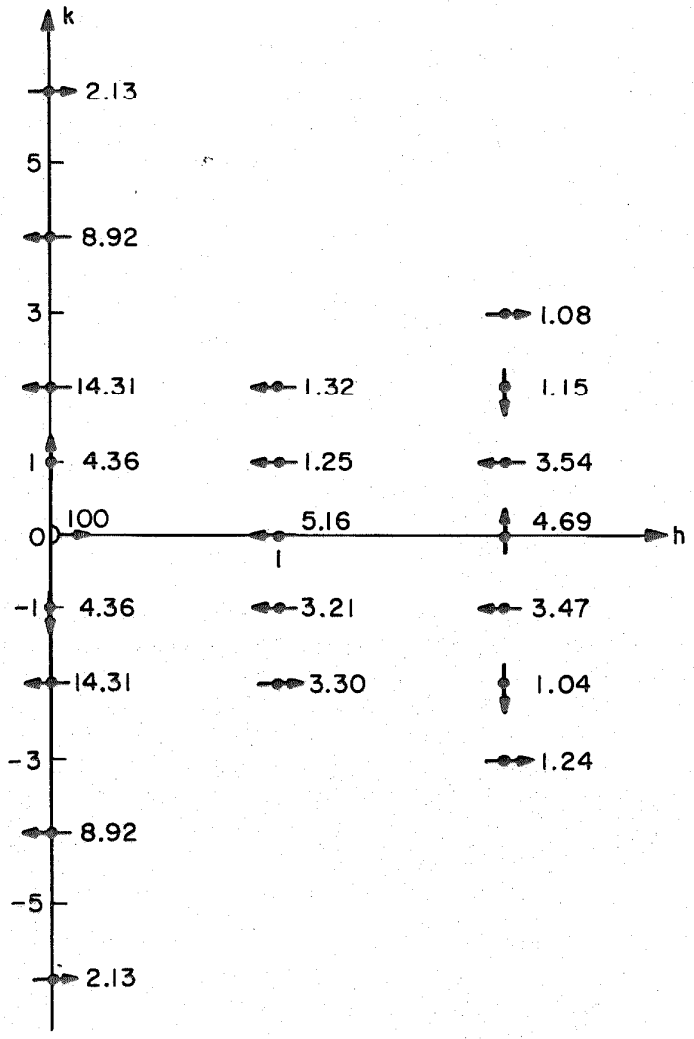


Fig. 3-23 Diffraction pattern of a catalase crystal reconstructed from the computed transforms of a focal series. Compensations have been made for asymmetries in the input transforms. Phase angles have been approximated to their nearest integer multiples of $\pi/2$, from their preliminary estimates shown in Fig. 3-22. The direction of the arrow indicates the phase angle and the number gives the maximum amplitude, observed in the focal series, of the diffraction spot relative to the zero order component. Back transformation of this pattern yields the reconstructed image shown in Fig. 3-25.

amplitude maxima of the beams, and their associated phase angles, are indicated in this figure.

We shall now consider certain problems associated with the normalization of the amplitudes of the diffracted beams with respect to the DC term. The zero order Fourier component, $|D_\ell(0)|$, is proportional to the mean optical density, $\langle d_\ell \rangle$, of any picture ℓ. If one calls $\tilde{d}_\ell(\underset{\sim}{r})$ the picture obtained from $d_\ell(\underset{\sim}{r})$ by a linear contrast enhancement which saturates the tails of the histogram to a fixed percentage, the following relationships hold

$$\tilde{d}_\ell(\underset{\sim}{r}) = \overline{\alpha}_\ell \, d_\ell(\underset{\sim}{r}) + \overline{\beta}_\ell \tag{3.128a}$$

and

$$\tilde{D}_\ell(\underset{\sim}{k}) = \overline{\alpha}_\ell \, D_\ell(\underset{\sim}{k}) + \overline{\beta}_\ell \delta(\underset{\sim}{k}) \tag{3.128b}$$

The normalization process is thus strongly dependent upon this preliminary contrast manipulation, which converts $D_\ell(\underset{\sim}{k})/D_\ell(0)$ into

$$\frac{\tilde{D}_\ell(\underset{\sim}{k})}{\tilde{D}_\ell(0)} = \frac{D_\ell(\underset{\sim}{k})}{D_\ell(0) + \overline{\beta}_\ell/\overline{\alpha}_\ell} \qquad \text{for} \quad |\underset{\sim}{k}| \neq 0 \tag{3.129}$$

Since the exposure times of each micrograph were nearly identical, the total electron flux impinging upon each plate is a constant. One therefore expects the mean optical density of the pictures that have undergone a contrast stretch to be similar (i.e., $\langle \tilde{d}_\ell \rangle \simeq$ constant).

3.4.3.4 Combined transfer function

If the predictions of the linear transfer theory are valid, then the scaled amplitudes of the diffraction spots in each micrograph

of the series give a pointwise determination of the combined transfer function,

$$C(\underset{\sim}{k}) = [1 - \Xi^2(\underset{\sim}{k})]^{1/2} \sin \gamma(\underset{\sim}{k}) + \Xi(\underset{\sim}{k})\cos \gamma(\underset{\sim}{k}) \qquad (3.130)$$

where $\Xi(\underset{\sim}{k}) = -M(\underset{\sim}{k})/\Phi(\underset{\sim}{k})$ is the relative contribution of amplitude contrast over phase contrast. The relative amount of amplitude contrast depends on the spatial frequency θ/λ, on the objective aperture half-angle, and on the atomic composition of the specimen (i.e., on the staining process for a biological specimen). In the case of negatively stained catalase, $|\Xi(\underset{\sim}{k})|$ has been measured experimentally[61] and found to decrease linearly from 0.4 at very low spatial frequencies to 0.1 at $\theta/\lambda = 0.04\mathring{A}^{-1}$. In our experiment, this a priori knowledge of $|\Xi(\underset{\sim}{k})|$ provided a guideline for scaling the diffracted beam amplitudes. The amplitude maxima were not observed in the experimental range of ΔZ, thus they were estimated by matching the experimental data with $C(\underset{\sim}{k})$. The scaled amplitudes were then plotted for the complete focal series (see Fig. 3-24) as a function of the reduced variable : $u = \text{sign}\{\Delta\tilde{Z}\}\ \theta\sqrt{|\Delta\tilde{Z}|}/\lambda$ where $\Delta\tilde{Z}$ stands for the focus parameter modified by axial astigmatism. Table 3-3 lists the azimuth angles ϕ^i and the frequencies θ^i/λ of the various diffraction spots ($\Delta\tilde{Z}$ can then be computed locally at each diffraction spot). We see from Fig. 3-24 that most of the experimental points lie near a curve that is described by

$$C(u) = -[1 - \Xi^2(\theta)]^{1/2}\sin \pi\lambda u^2 + \Xi(\theta)\cos \pi\lambda u^2 \qquad (3.131)$$

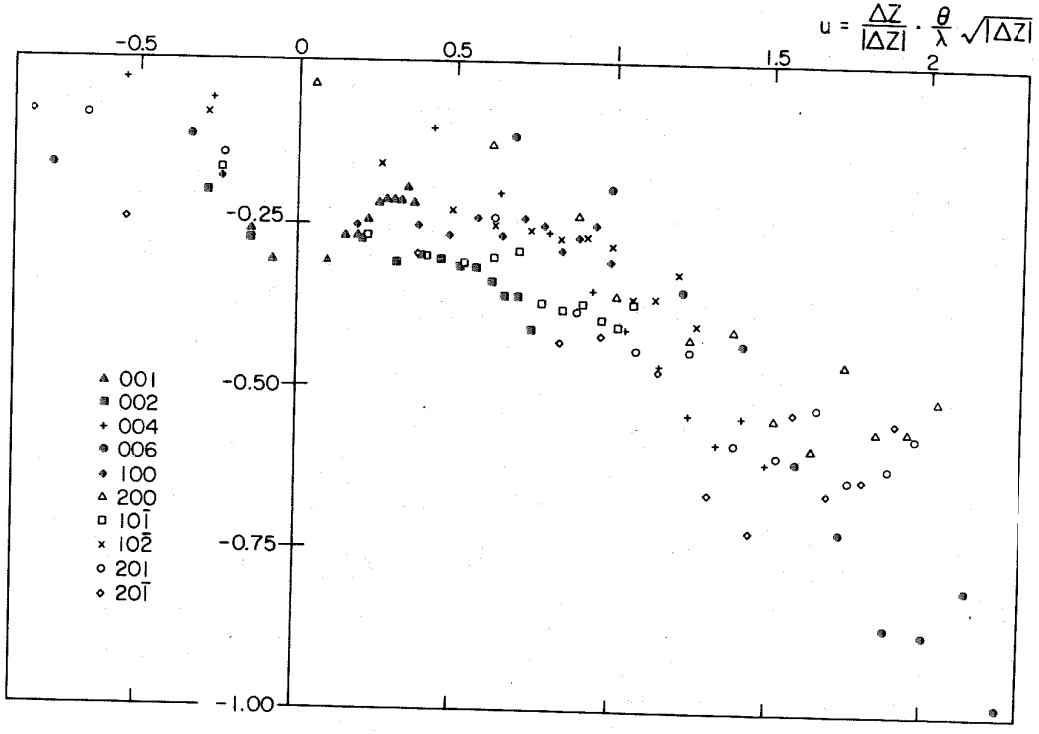


Fig. 3-24 Scaled amplitudes of the diffraction spots measured from the complete focal series of catalase and plotted as a function of the reduced variable $u = \text{sign}(\Delta\tilde{Z})\ \theta\sqrt{|\Delta\tilde{Z}|}/\lambda$. The spread of the experimental points is due to the frequency dependence of amplitude contrast, and to errors stemming from the registration of the pictures. Most of the points lie between the phase contrast curve, $-\sin \pi\lambda u^2$, and the 40% amplitude contrast curve, $-\sin(\pi\lambda u^2 + 0.41)$.

Table 3.3 Reciprocal lattice spacings and azimuthal angles associated with each diffracted beam of the catalase transform shown in Fig. 3-18. The azimuthal angles, ϕ^i, have been corrected for lattice distortions.

Diffracted beam	θ^i/λ (in $\mathring{A}^{-1}$)	ϕ^i in degrees
(001)	0.0058	90
(002)	0.0116	90
(004)	0.0232	90
(006)	0.0348	90
(100)	0.0149	0
(101)	0.0160	23
(10$\bar{1}$)	0.0160	-23
(102)	0.0189	44
(10$\bar{2}$)	0.0189	-44
(200)	0.0298	0
(201)	0.0304	11
(20$\bar{1}$)	0.0304	-11
(202)	0.0320	23
(20$\bar{2}$)	0.0320	-23
(203)	0.0345	32
(20$\bar{3}$)	0.0345	-32

Although the agreement with the theory is only qualitative, the theoretical transfer functions for each individual reflection describe the observed contrast variations with defocusing. The variations of amplitude contrast with θ, however, cannot be included in a single transfer function[61]. A complete restoration scheme would therefore combine two or more micrographs at different focus to separate the two contrast mechanisms.

3.4.3.5 Partial reconstruction

An image transform that is partially compensated for defocusing effects is reconstructed by assigning to each diffraction spot its maximum experimental amplitude and its phase corrected for lattice distortions (see Fig. 3-23). In order to eliminate random noise that is superimposed on the image, we only include discrete diffraction spots in the image transform. Inverse transforming this diffraction pattern yields an image which is derived from the unprocessed image by a twofold operation. First, amplitude attenuation of the spots due to the contrast transfer function are partially compensated, then the unit cell is averaged over all the repeat unit cells by superposition. The reconstructed image is shown in Fig. 3-25c, and it displays the same morphological units as the highly underfocused original pictures (ℓ) of the series (see Fig. 3-25a). To accentuate the comparison, the latter image was noise filtered by inserting square Fourier windows (0.0012Å^{-1} a side) around the diffraction spots, and subsequently inverse transforming. The result of this filtering, shown in Fig. 3-25b, gives a pleasing image where one easily discerns the warping of the lattice. Figure 3-26 is a magnified version of Fig. 3-25, displaying two vertical subunits of the catalase. Little resolution has been

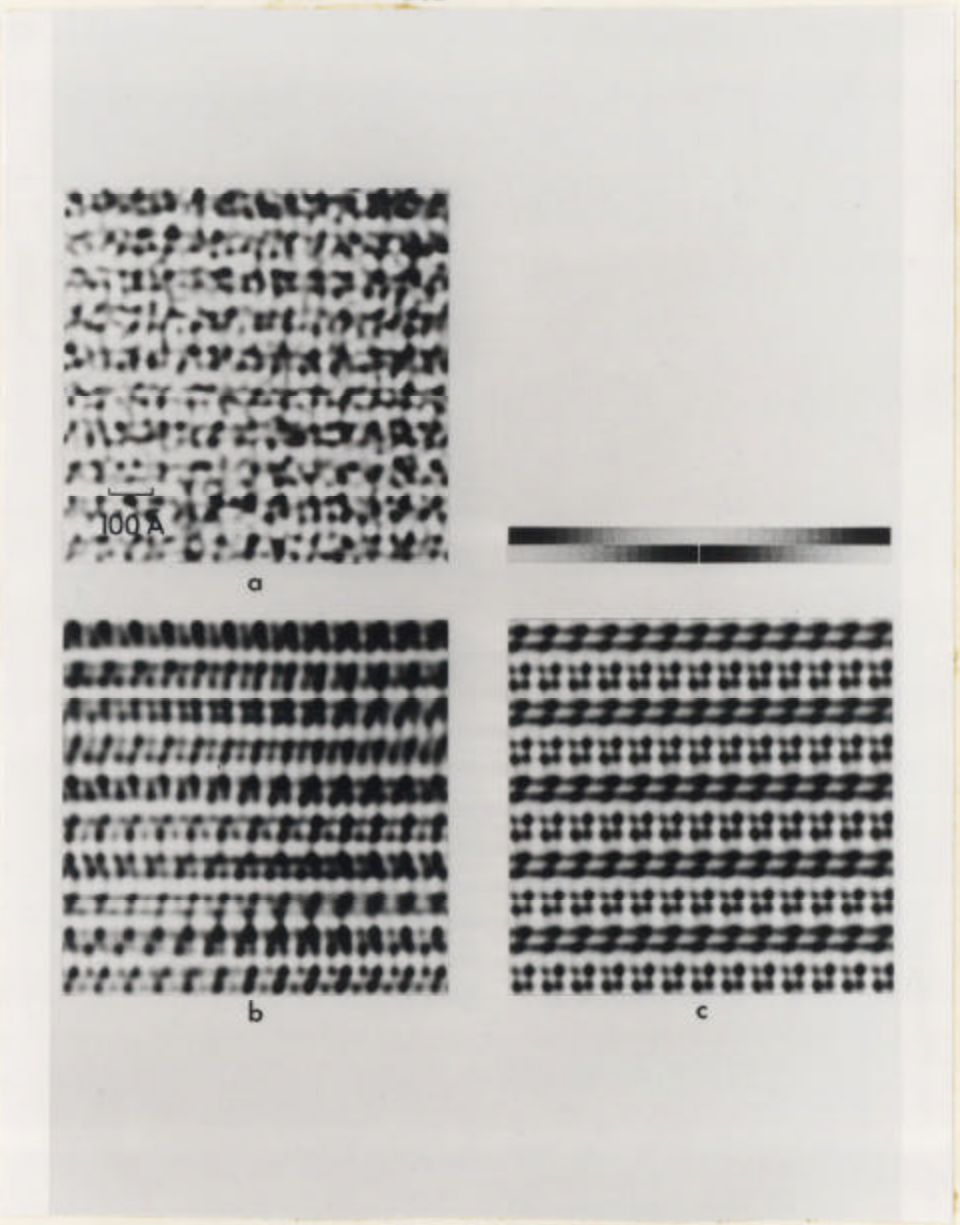


Fig. 3-25 Comparison of the most underfocused micrograph, Fig. 3-20ℓ, of the catalase series with noise-filtered images reconstructed from the catalase transforms. The images presented show (a) Fig. 3-20ℓ of the series after contrast stretching, (b) Fig. 3-20ℓ after noise filtering, and averaging by insertion of Fourier windows around its diffraction spots, and (c) the reconstructed image partially corrected for defocusing artifacts.

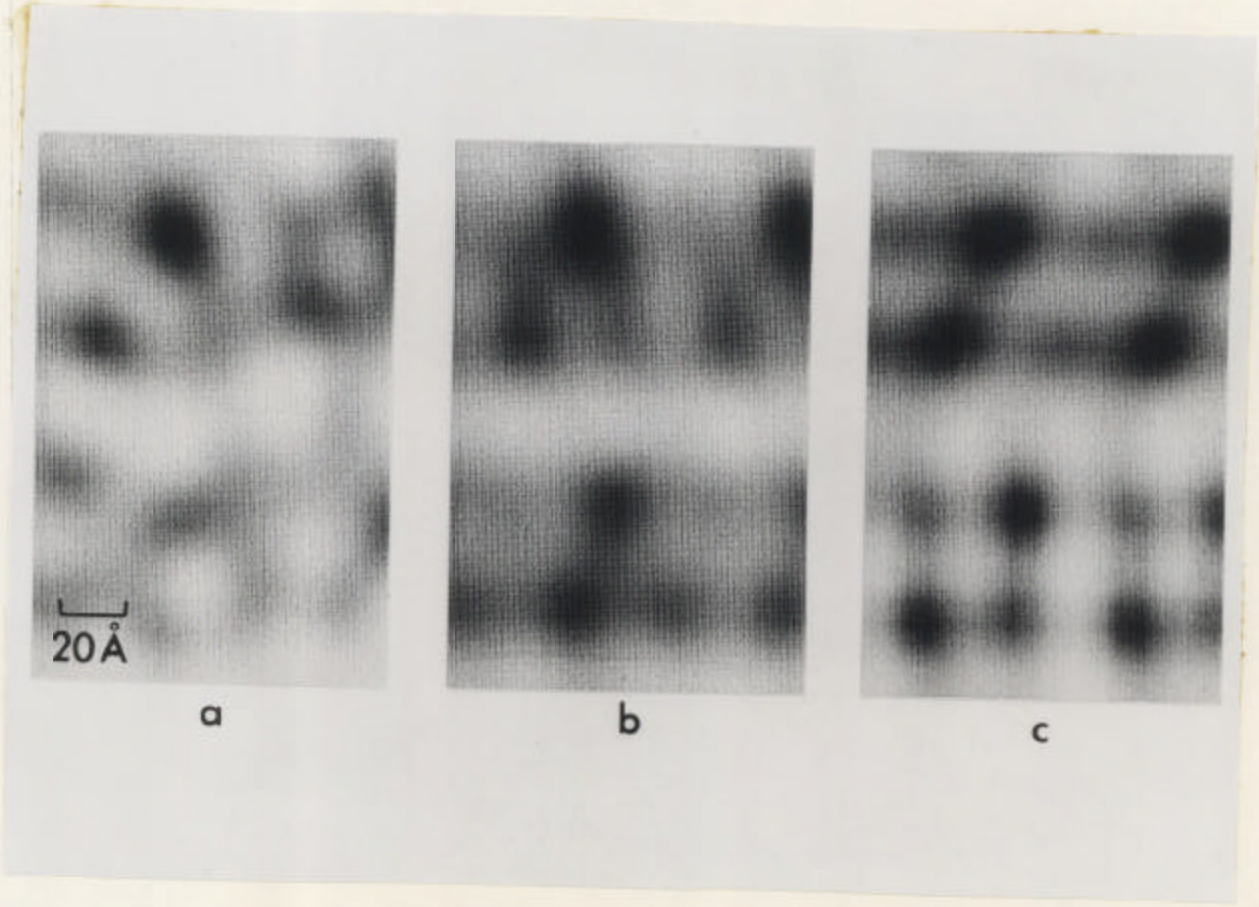


Fig. 3-26 Magnified version of Fig. 3-25 showing two vertical subunits (2x86Å) of the catalase molecule at various stages of the processing: (a) Fig. 3-20ℓ of the focus series after contrast stretching; (b) Fig. 3-20ℓ after noise filtering and averaging by insertion of Fourier windows around its diffraction spots; and (c) the reconstructed image partially corrected for defocusing artifacts.

gained along the processing because of the lattice distortions of the specimen that were introduced during the drying stage of the crystallization.

In the resolution range of interest, $0.0058\text{Å}^{-1} \leq \theta/\lambda \leq 0.0348\text{Å}^{-1}$, the overall contrast transfer function $C(u)$ will be maximized at a certain defocus ΔZ_{op}. Since $C(u)$ is bordered on one side by the pure phase contrast function $-\sin \pi\lambda u^2$, and on the other side by the 40% amplitude contrast function $-0.915 \sin \pi\lambda u^2 - 0.4 \cos \pi\lambda u^2 = -\sin(\pi\lambda u^2 + 0.41)$, the optimum ΔZ_{op} lies between the optimal settings $\Delta Z_{op}^{(1)}$ and $\Delta Z_{op}^{(2)}$ corresponding to these two transfer functions. The highest frequency $\theta/\lambda = 0.0348\text{Å}$, is thus transferred without attenuation between $\Delta Z_{op}^{(1)} = 11100\text{Å}$ and $\Delta Z_{op}^{(2)} = 8200\text{Å}$, so that an optimal focus for this catalase specimen would be around $\Delta Z_{op} = 9500\text{Å}$. At this choice of ΔZ, we expect the underfocus enhancement to yield a high-contrast image.

An analysis of the optical diffractogram will determine the spatial-frequency spectrum of the specimen. If the amplitude contrast contribution is approximately known, we can adjust the defocusing to the range of spatial frequencies in the object so as to achieve optimal contrast. Artifacts occuring beyond the first zero of the transfer function can easily be detected by inspection of the diffractogram.

3.4.4 Conclusion

The linear theory of image formation, which relates the image transform to the object structure, gives a valid interpretation of the contributing contrast mechanisms at the medium resolution level. Provided that one deals with a weakly scattering object, this theory

also offers the possibility of compensating for the effects of spherical aberration and partial coherence in order to extend the resolution of the electron microscope.

At optimum conditions, the microscope is capable of achieving a valid resolution of about 4.5Å, because the main high-resolution mechanism is phase contrast (see Eq. (1.38)). Resolution extending schemes will then be useful whenever the specimen contains meaningful structural information at this level. These schemes are currently limited by instrumental instabilities and radiation damage of the specimen.

Several solutions to overcome the above limitations seem promising. For instance, by observing protein crystals in environmental wet cells, one can preserve the intimate morphological structure and allow the imaging of its fine details[82]. Radiation damage can be minimized by cooling the specimen to very low temperatures; such temperatures also reduce the specimen contamination rate[88]. The total exposure time can be minimized by means of image intensifying devices, and by applying image processing techniques to retrieve the data from statistically noisy images[89].

Under carefully controlled experimental conditions, such compensating schemes as the Schiske method, which restores the complex object function, might open the way to the visualization of atoms. We have examined in Sec. 3.3 the processing scheme and the various factors which must be considered in the interpretation of the restored image. We shall now present an experimental study to check the validity of the approximations upon which the processing algorithm is based.

3.5 Study of the Z Discrimination Effect on Gold Clusters Lying on a Carbon Substrate

3.5.1 Experimental Procedure

3.5.1.1 Specimen preparation

The need for ultra-thin support films to attain the ultimate in high-resolution electron microscopy has witnessed the development of several techniques for producing films thinner than 20Å [90]. We have developed an efficient method for producing thin carbon films that exhibit excellent thermal stability under the electron beam[91].

In our technique, a thin layer of carbon is evaporated upon the surface of a parlodion film(pyroxilin, cellulose nitrate 1% solution in amyl acetate) lying on a copper grid, placed on top of a rotating disc (See Fig. 3-27) to achieve a better thickness uniformity. Carbon is simultaneously deposited upon a glass-slide indicator, with a thin wire (4 mils) across the slide. This control slide is later examined under an optical interferometer, using thallium radiation of 5370Å, which yields a thickness measurement accurate to within 30Å. Two pyrex tubes, lightly smeared with silicone grease, enclose the indicator system and the target grids. The diameters of both tubes and the distance between grids, carbon rods, and glass slide are determined so as to ensure that the vaporation of carbon obeys an inverse square law. Carbon atoms that are partially reflected by the tubes hit the substrate and the slide at random angles of incidence, thereby improving the nondirectionality of the structure. A microcontact was incorporated in the electric circuit of the evaporator to permit short-pulsed evaporations. In general, two or three pulses were sufficient to obtain

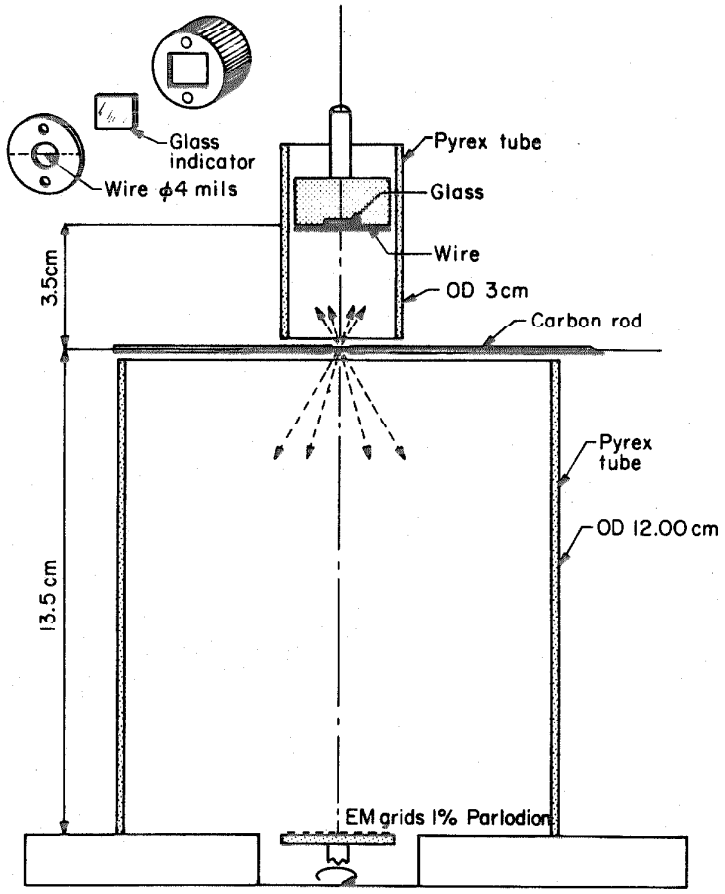


Fig. 3-27 Diagram of the evaporation system used to prepare ultra-thin carbon support films. The glass indicator is shown in perspective.

about 150Å of carbon on the glass slide. The calculated fraction of carbon deposited on the grid, in this geometry, is 10% of that deposited on the indicator slide.

After the evaporation the grids are placed on a screen suspended above a solution of isoamyl acetate, whose vapor dissolves the parlodion film. The grids should only be exposed to this vapor for about two hours at room temperature; if this time is exceeded copper carbonate builds up and contaminates the film. In order to ensure that the parlodion film has been completely dissolved by the isoamyl acetate vapor, control grids covered with a 2% parlodion film and no carbon are placed in the vapor and periodically examined with the electron microscope.

The gold on carbon specimen is prepared by first spraying the surface of the support film with carbon-black spheres, followed by the deposition of gold by shallow-angle evaporation. This specimen was selected since its chemical composition is well defined and it is insensitive to radiation damage. The sample was examined at the tail of the shadow and revealed gold clusters with a wide range of sizes.

3.5.1.2 Electron microscopy

A through-focus series of the gold on carbon specimen was taken with a JEOL 100B microscope operating at 100 keV with an electron optical magnification of 400,000X. The objective lens astigmatism was corrected prior to the recording of a focus series, and the specimen contamination rate was minimized by using a liquid-nitrogen cold finger. Once thermal drift of the specimen had settled to a tolerable level, the electron micrographs were rapidly recorded.

Despite all these precautions, several trials were necessary to minimize the effects of axial astigmatism. This astigmatism is most

strikingly displayed as elliptical rings in the optical diffractograms. Specimen drift was minimal during the exposure, since no fringe system was visible in the optical transform. The best focus series was then selected for further processing. The operational characteristics of the instrument are listed in Table 3.4, and Table 3.5 lists the optically determined electron microscope parameters of the series.

3.5.2 Computer processing steps

3.5.2.1 Video input-output

The bright-field micrographs of the focus series were copied onto fine grain film, without any appreciable loss in resolution within the spatial frequency range of interest. An area of 15 x 15 mm, corresponding to 375 x 375Å on the object scale, was scanned with the VFC at a sampling density of Δ = 25 µm. The equivalent scanning distance is 0.625Å at 400,000X magnification (this value is much smaller than the inverse of the Nyquist frequency for these data). The picture elements were then saturated to 1% on both sides of the gray scale, by stretching the histogram to achieve a full dynamic range of the data between 0 and 255.

3.5.3.2 Processing algorithm

A detailed account of the processing algorithm has already been presented in Sec. 3.3.4.4 and a block diagram of the procedure shown in Fig. 3-17. We shall discuss in this section the problems which were encountered during the processing of the gold on carbon series, and the accuracy of the restoration.

The first step of the processing consists in finding the mutual positions of all pictures in the series. Since the defocusing values

Table 3.4 Operating characteristics of the JEOL 100B transmission electron microscope, during the recording of the micrographs of gold on carbon shown in Fig. 3-30.

Specifications of the JEOL 100B TEM

Second condenser aperture	100 μm in dia. $\Longrightarrow$ $\theta_{con} = 7.5 \times 10^{-4}$ rad
Objective Lens	Focal length: F = 1.6 mm Spherical aberration: C_s = 1.4 min Chromatic aberration: C_c = 1.4 min
Objective aperture	60 μm in dia. $\Longrightarrow$ $\theta_{obj} = 1.9 \times 10^{-2}$ rad
Instrument stability	High-voltage supply: $\langle \Delta U \rangle / U \leq 2 \times 10^{-6}$/min Objective lens current: $\langle \Delta i_{\ell} \rangle / i_{\ell} \leq 2 \times 10^{-6}$/min
Focus step of the series	$\Delta Z_{st} = 80$ Å
Contamination of the specimen	Less than 0.1 Å/min

Table 3.5 Estimates of the defocusing values for the micrographs of gold on carbon shown in Fig. 3-30. These defocusing values were obtained from the optical transforms of the micrographs. The axial astigmatism of the micrographs is described by $\Delta Z_a \simeq 165\text{Å}$ and $\phi_o \simeq 45^o$.

Micrograph	ΔZ_ℓ (Å)
a	-245
b	-175
c	- 80
d	15
e	70

ΔZ_ℓ are lying on either side of the Gaussian position (see Table 3.5) we expect the occurrence of artifacts in evaluating the correlation peaks of two pictures with ΔZ_ℓ of opposite signs. When a negative peak is predicted by the calculated CCF of the corresponding transfer functions, the pictures must be complemented prior to the correlation step in order to avoid errors in the peak determination (the program searches for the maximum of the correlation matrix). The sums:

$$\phi_{i,j}^{(d_1,d_2)} = \frac{1}{KL} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} d_{k,\ell}^{(1)} \, d_{k+i,\ell+j}^{(2)}$$

and

$$\phi'^{(d_1,d_2)}_{i,j} = \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} (d_{k,\ell}^{(1)} - d_{k+i,\ell+j}^{(2)})^2$$

were computed in various overlapping portions 88 x 88 pixels of the pictures. Prior to the computation, the angular orientation was adjusted by matching conspicuous image features. We observed a strong dependence of the CCF peak halfwidth on the particular correlation pair. An example of the peak definition with respect to its neighborhood is shown in Fig. 3-28. The symmetry of the peak indicates that picture gradients are minimal and that artifacts caused by the discrete evaluation of the correlation integral are negligible. A subsequent fit about the peak value of the correlation surface $\phi_{i,j}^{(d_1,d_2)}$ with a quadratic surface yields the mutual displacement vectors to within a fraction of a pixel accuracy. Errors in the registration of the pictures are negligible if ill-defined peaks or trending artifacts are eliminated. A better registration would, however, be obtained by

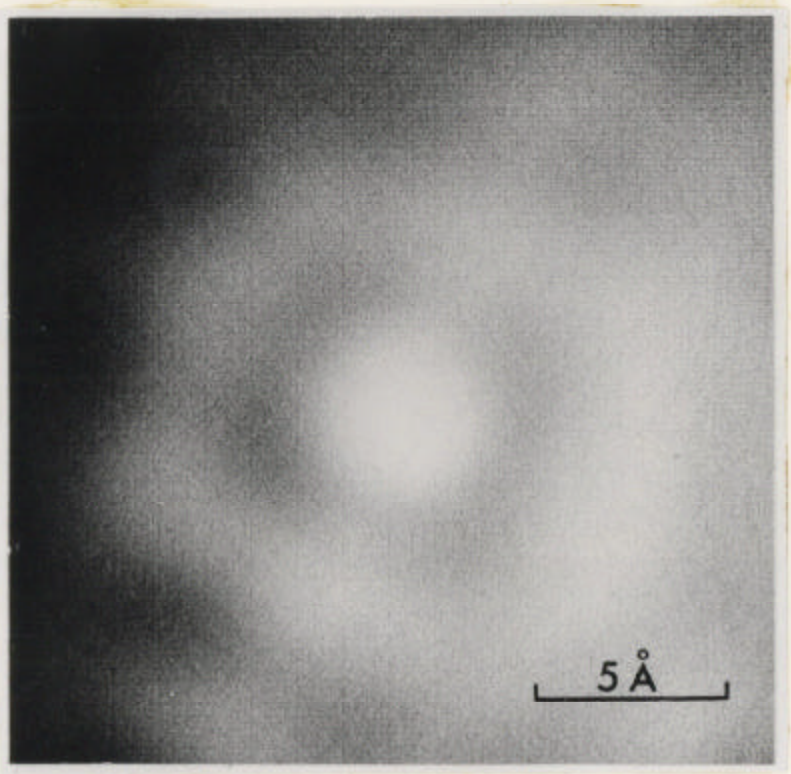


Fig. 3-28 Cross-correlation function between two micrographs, Fig. 3-30d and e, of the through-focus series of gold on carbon. In order to avoid the occurrence of subsidiary maxima in the CCF, density gradients in the original pictures were removed. The peak is well defined from its neighborhood; therefore the two corresponding pictures can be spatially registered to high accuracy.

computing the correlation matrices by the Fourier method. Phase reversals due to the contrast transfer function, which pull the CCF peak down towards the negative side, can then be filtered out (see Sec. 3.3.4.3). Truncation errors introduced by the limited size of the input arrays (the size of the input array is governed by core storage requirements) have overthrown the advantages of this method. The focus series, after contrast stretching and positioning, is illustrated in Fig. 3-29.

In the second step of the processing, the sampling density was halved by averaging over four pixels in sequence. The Nyquist frequency corresponding to the new scanning step ($\Delta = 50\ \mu m \Rightarrow k_m = 0.4\text{Å}^{-1}$) is still smaller than the resolution of the instrument. In order to determine the highest frequency that carries information, a Fourier resolution test was performed. In this test, the superposition transform of two micrographs of the series is computed. The superposition transform shown in Fig. 3-16a exhibits modulation fringes which extend up to 0.25Å^{-1}, and therefore indicate an instrumental resolution of 4Å. Thus a slightly larger value, $k_{max} = 0.28\text{Å}^{-1}$, was chosen as the aperture radius of the noise-filter function $h_N(\underset{\sim}{k})$ (see Eq. 3-123)). We then selected in each picture of the series a 160 x 160Å area of carbon film, and determined the autocorrelation matrix using the Wiener-Khinchine relationship. An 80 x 80Å central portion of the ACF array was then cut out and multiplied by the radial function $w(\underset{\sim}{r})$ (see Eq. (3.59)) in order to smooth fluctuations in the carbon-film power spectrum. Values for the cutoff radius and the Gaussian fall-off halfwidth were chosen to be $r_G = 22.5\text{Å}$ and $\sigma_G = 6.25\text{Å}$ respectively. The optimal phase

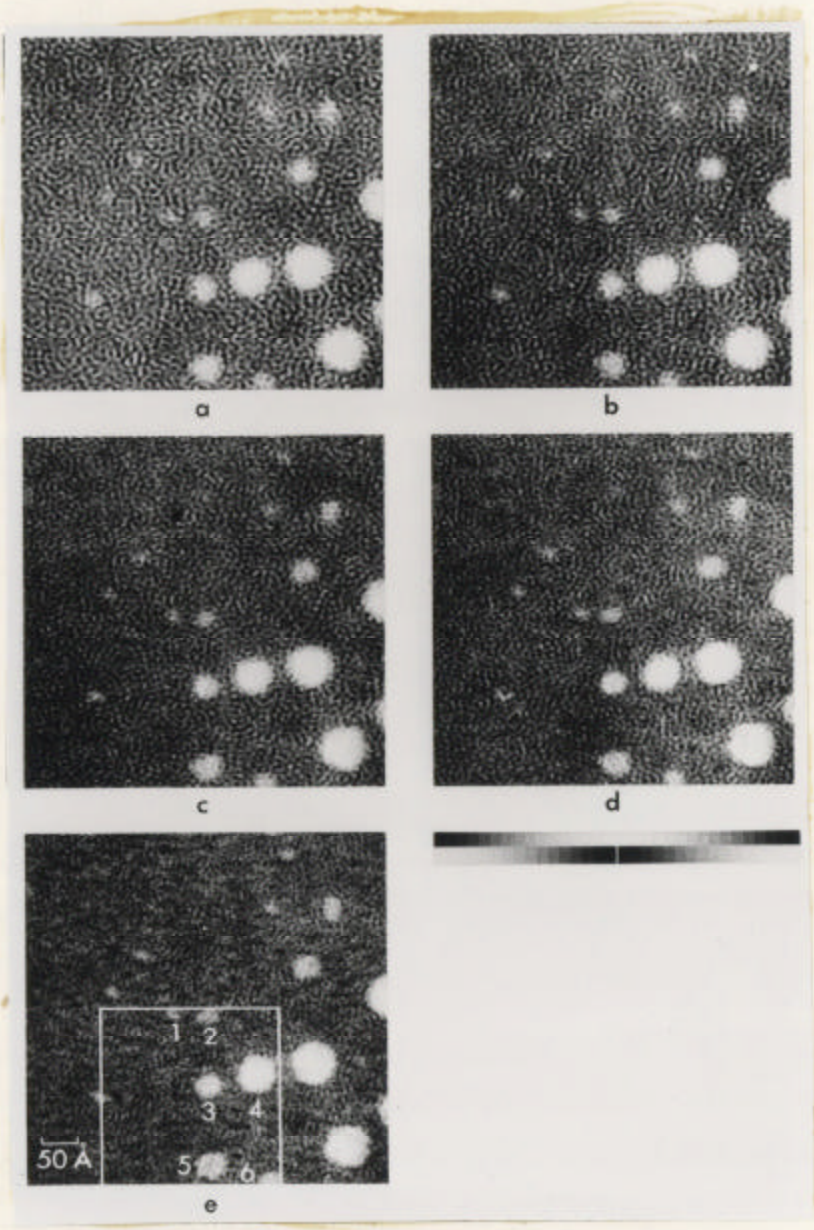


Fig. 3-29 Focus series of gold on carbon after contrast stretching and positioning. The enclosed box in Fig. e, 320x320Å, shows the area used in the subsequent processing.

factors $\gamma(\underset{\sim}{k})$ were determined by a least-squares fit between $|j_c(\underset{\sim}{k})|$ and $g_r(\theta)\,|\sin\gamma(\underset{\sim}{k})|$ within the range $k_{max}/4 < |\underset{\sim}{k}| < k_{max}$. A measure of the divergence in the successive iteration runs is obtained by varying the parameters: r_G, σ_G, k_{max}, and the starting electron microscope values. It could be ascertained that the output electron microscope parameters furnished the best possible representation for the function $\sin\gamma(\underset{\sim}{k})$, since fluctuations in their measurements were negligible. The results of the fitting program are listed in Table 3-6; they match remarkably well with the preliminary estimates.

The object transform is computed from the transforms of the micrographs using Schiske's formula. An area, 128 x 128 pixels, in which several gold clusters could be observed, was then cut out of each frame (the small area was chosen in order to reduce the total amount of computer time). The focus series used in the reconstruction is shown in Fig. 3-30 and the associated numerical transforms are shown in Fig. 3-31. Since no transform artifacts, due to either scanning noise or edge discontinuities, were visible in Fig. 3-31, the noise filter function $h_N(\underset{\sim}{k})$ did not set the Fourier components along the principal axes to zero. The bandwidth of the filter was determined by k_{max} and by the largest hole radius, k_{min}, which would not cause ringing around the clusters. The filter parameters were chosen as follows (see Eq.(3.123)): $k_{min} = 0.025\text{Å}^{-1}$, $k_{max} = 0.28\text{Å}^{-1}$, $\sigma_{min} = 0.0125\text{Å}^{-1}$, and $\sigma_{max} = 0.0312\text{Å}^{-1}$.

By combining the modified Fourier coefficients, the complex structure factor, $F(\underset{\sim}{k})$, was determined according to Eq. (3.97). The effects of partial coherence were ignored in this processing, although they are quite significant under these data recording conditions. (These

Table 3.6 Estimates of the defocusing values for the micrographs of gold on carbon shown in Fig. 3-30. These defocusing values were computed from a least-squares fit between $|j_{\ell}^{c}(\underset{\sim}{k})|$ and $g_r(\theta)|\sin \gamma_{\ell}(\underset{\sim}{k})|$ in each micrograph of the through-focus series. The axial astigmatism of the micrographs was determined to be $\Delta Z_a \simeq 155\text{Å}$ and $\phi_o \simeq 46^{o}$.

Micrograph	ΔZ_{ℓ}(Å)
a	-196
b	-115
c	- 53
d	15
e	95

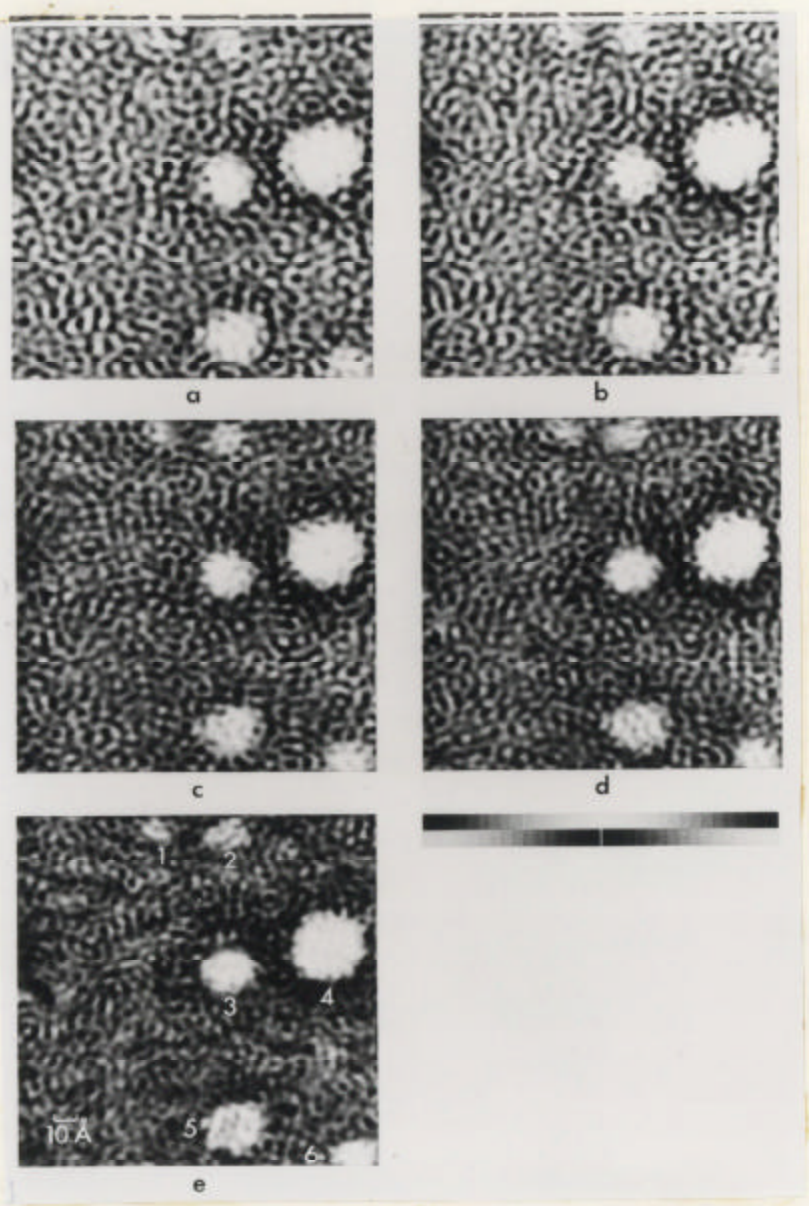


Fig. 3-30 Focus series of gold on carbon after positioning, and noise filtering using the bandpass filter defined in Eq. (3.123). The filter characteristics are: $k_{min} = 0.025\text{Å}^{-1}$, $k_{max} = 0.28\text{Å}^{-1}$, $\sigma_{min} = 0.0125\text{Å}^{-1}$ and $\sigma_{max} = 0.0312\text{Å}^{-1}$.

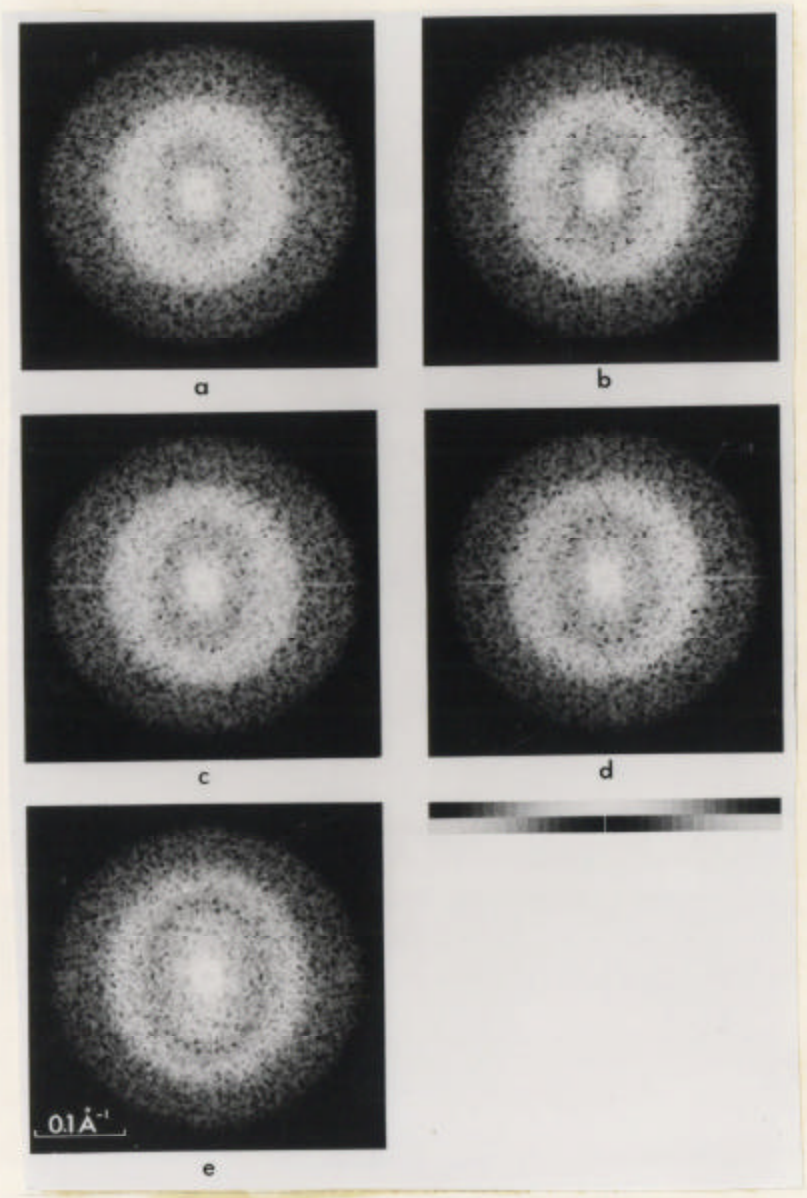


Fig. 3-31 Numerical transforms of the focus series shown in Fig. 3-30 computed over an area containing 128x128 pixels. The small size of the input array accounts for the observed granularity. Scanning noise and truncation box artifacts are not visible, so that apodization is not necessary at the picture boundaries. The transforms are displayed in a logarithmic representation.

conditions are the same as those for which the envelope functions were plotted in Fig. 3-5.) Various numbers of pictures were used in the restoration procedure and as expected the noise is quite amplified near the zeros of the contrast transfer function when only two pictures are used. The noise perturbation ceases to be a limiting factor when the number of pictures in the least-squares combination is sufficiently large. The real part of $F(\underset{\sim}{k})$ exhibits a discontinuity near the origin which can be removed by high-pass filtering. The imaginary part, on the contrary, can be displayed directly, and its selective contrast enhancement is brought out progressively as the noise amplification is reduced. The whole reconstruction is then assessed according to an algorithm presented in Appendix D.

If the reconstructed images match closely enough with the originals, the various dark field terms $|\psi_s^{(\ell)}(\underset{\sim}{r})|^2$ are evaluated and subtracted from the original series. One iteration is generally sufficient to minimize the influence of this quadratic term on the restored complex-object wave. The results of this procedure are now discussed.

3.5.3 Results of the Restoration

3.5.3.1 Input images

The interesting features of the changes in cluster contrast with defocus are seen in the original focus series shown in Fig. 3-29. In the overfocus region, only details of medium size are apparent, the smaller clusters being almost hidden by the speckle-like structure of the carbon background. As one reaches underfocus conditions finer details are enhanced, and simultaneously the contrast of the gold

clusters against the substrate is improved.

A study of the behavior of the phase and amplitude contrast transfer functions, modified for partial coherence, confirms this visual assessment. Figures 3-32 and 3-33 illustrate these functions, calculated for the various electron microscope parameters that were obtained by the least-squares fitting procedure. The coarse variations in the background contrast with ΔZ_ℓ is easily interpreted, since we know that a carbon film behaves as a pure weak-phase object with an almost white power spectrum. The smaller gold clusters exhibit a decrease in their diameter as one crosses from overfocus to underfocus, a fact which is accounted for by an inspection of the scattering contrast transfer curves shown in Fig. 3-33. The larger attenuation of low-frequency components at $\Delta Z = -196\text{Å}$ results in a correspondingly broader impulse response in this range of resolution. It is important, however, to be sure that the anomalous-scattering contrast transfer does not change significantly within this small focus range, since this ensures that the frequency gaps, near the zero crossing of each transfer function, are filled in the restored complex-object wave. On the other hand, the high-resolution details within the clusters are extremely sensitive to the changes in the transfer conditions with ΔZ_ℓ.

Although not intentional, the carbon substrate was shadowed with gold in such a way that the average cluster size exceeded 30Å even at the rim of the shadow. As a consequence, most clusters exhibited very pronounced contrast features which led to a saturation of the optical densities at their sites during the scan. This proved to be one of the limiting factors of the success of the whole restoration,

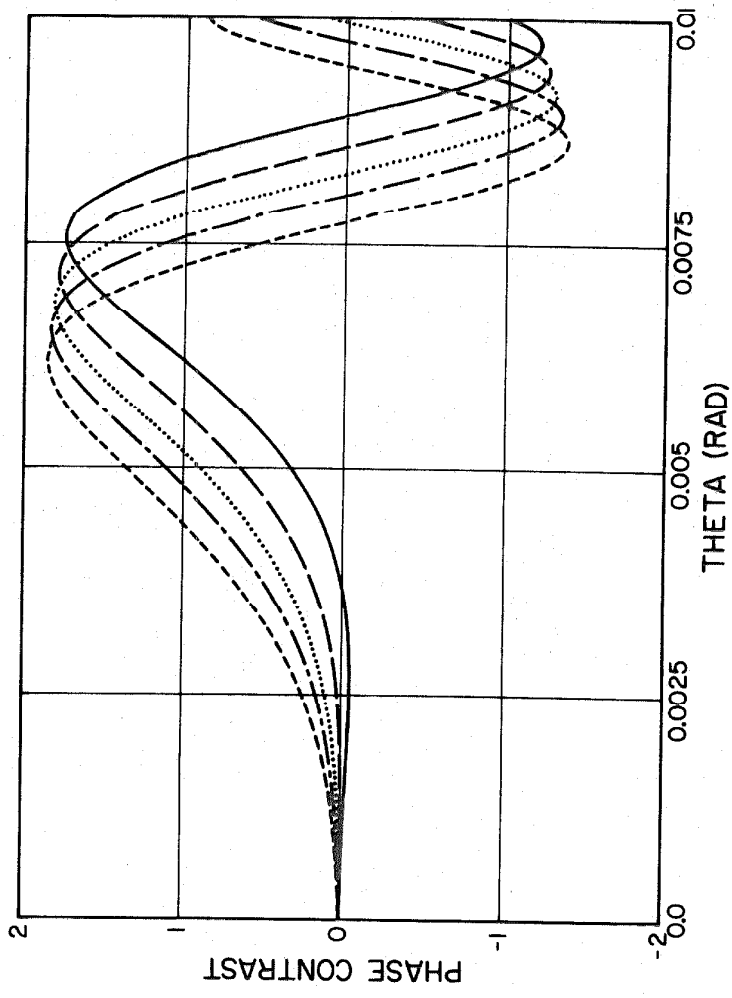


Fig. 3-32 Phase contrast transfer functions, 2 sin $\gamma(\theta)$ $E(\theta)$, corrected for the effects of partial coherence. The envelope function $E(\theta)$ is identical to the one shown in Fig. 3-6 when spatial and chromatic incoherence are considered. The fine-dashed curve corresponds to $\Delta Z = -196\text{Å}$, the dash/dot curve to $\Delta Z = -115\text{Å}$, the dotted curve to $\Delta Z = -53\text{Å}$, the coarse-dashed curve to $\Delta Z = 15\text{Å}$, and the solid curve to $\Delta Z = 95\text{Å}$.

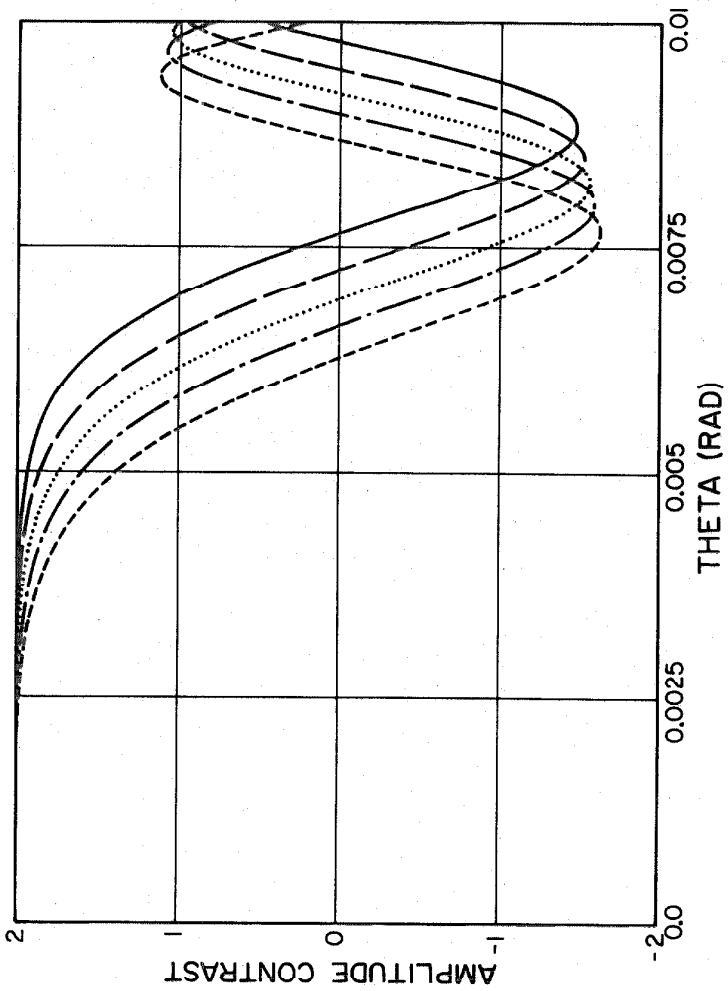


Fig. 3-33 Amplitude contrast transfer functions, 2 cos $\gamma(\theta)$ $E(\theta)$, corrected for the effects of partial coherence. The envelope function $E(\theta)$ is identical to the one shown in Fig. 3-6 when spatial and chromatic incoherence are considered. The fine-dashed curve corresponds to $\Delta Z = -196\text{Å}$, the dash/dot curve to $\Delta Z = -115\text{Å}$, the dotted curve to $\Delta Z = -53\text{Å}$, the coarse-dashed curve to $\Delta Z = 15\text{Å}$, and the solid curve to $\Delta Z = 95\text{Å}$.

because the original pictures displayed a high signal-to-noise ratio. Furthermore, at the gold sites, the kinematical scattering approximation is only partially valid, since the mean free path for single scattering in gold is 25Å. Another experimental deficiency lay in the amount of axial astigmatism ($\Delta Z_a \simeq 155$Å) which riddles the micrographs and causes an angular variation of information transfer far more significant than the variations within a single focus step.

In spite of all these difficulties, the restoration scheme was implemented on an area of the input images where low contrast gold clusters were visible. A box containing 128 x 128 pixel elements was cut out and noise filtered in the five micrographs, which had been previously brought into mutual alignment by cross correlation. These pictures are shown in Fig. 3-30 and their associated diffractograms are shown in Fig. 3-31. In the diffractograms the innermost elliptical zone corresponds to amplitude transfer, where the predominant contribution stems from carbon's large cross-section for inelastic scattering at small angles. The inelastic scattering in the original pictures can be reduced with high-pass filtering.

A check of the performance of a pure phase-contrast inverse filter (the design of this filter has been studied in Sec. 3.3.3.2) was examined on a picture of the series. Two different threshold values were used to obtain the final reconstructed images shown in Fig. 3-34; we notice that no information gain has been achieved. On the contrary, one expects a few artifacts in the restored image around the gold sites where anomalous scattering becomes sizeable. A more quantitative evaluation of the reconstruction is given by the relative signal-to-noise ratio (SNR_i), at the cluster locations, before and after applying

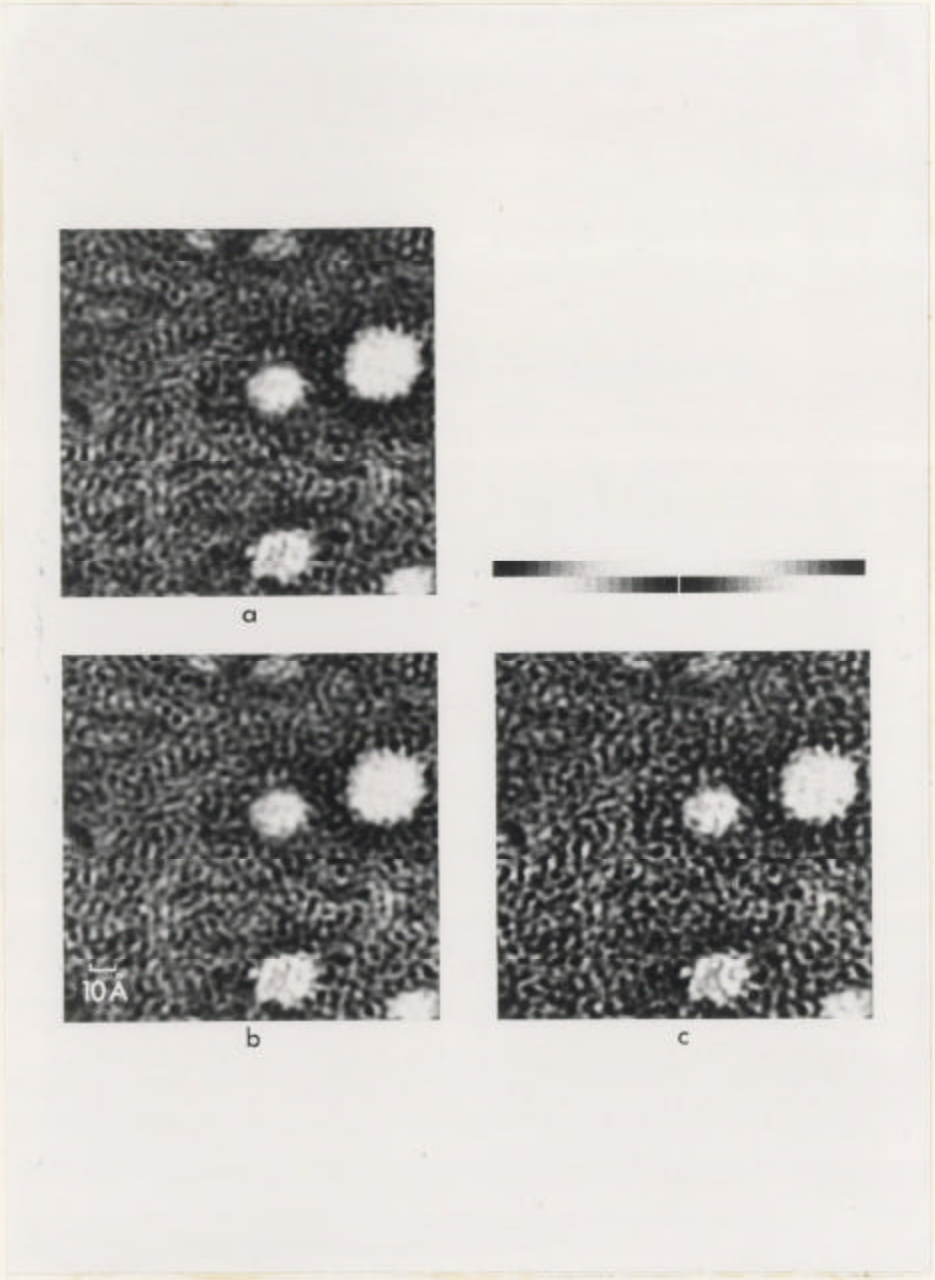


Fig. 3-34 Comparison of an underfocused image of the series of gold on carbon with its inverse filtered versions. The above pictures correspond to (a) Fig. 3-30e, (b) inverse filtered version of (a) with $\overline{\delta} = 0.5$, (c) inverse filtered version of (a) with $\overline{\delta} = 0.2$. Figure c shows the appearance of noise artifacts.

the filter. The results listed in Table 3-7 prove the inefficiency of an inverse filter algorithm to restore the complex-object wave function.

3.5.3.2 The anomalous scattering image

The amplitude contrast image, $\mathcal{F}^{-1}\{F_i(\underset{\sim}{k})\}$, is illustrated in Fig. 3-35 where N = 2,3,4 and 5 pictures of the series are used in the restoration. As the number of pictures increases, a gradual improvement in the discrimination effect is brought out; this phenomenon is expected if the noise propagation is taken into account. Some features can be discerned in the image of the smaller gold clusters, which are related to their structure, although these features are obscured by the adjacent noise pattern.

The selective contrast enhancement yields an imaginary image, $\mu(\underset{\sim}{r}_0)$, with slightly better contrast at the heavy atom sites than the best input image of the series (Fig. 3-30e). In Sec. 3.3.4.1, the atomic image contrast integrals were calculated for both phase and amplitude contrast, as well as for the electron microscope conditions under which Fig. 3-30e was recorded. The expected improvement in the separation effect, going from the original to the anomalous scattering image, was found to be about 7. These results were derived for isolated atoms, thus neglecting the influence of bonding on the atomic scattering potential. We have emphasized that specimen thickness limits the potential of the discrimination technique, by shuffling the real and the imaginary parts of the scattering amplitudes. This effect becomes all the more degrading as one proceeds to extend the resolution of the microscope. Inaccuracies in the measurement of the phase

Table 3.7 Listing of the signal-to-noise ratios of each gold cluster, at various stages of the processing. The SNR_i are given in the originals and the reconstructed series, then in $F_i(k)$, for two iteration loops successively. The SNR_i are also listed for the inverse filtering step.

Stage of the Processing	Picture	SNR_1	SNR_2	SNR_3	SNR_4	SNR_5	SNR_6
Originals shown in Fig. 3-30	a	1.315	1.199	1.559	1.562	1.445	1.443
	b	1.380	1.368	1.661	1.739	1.454	1.585
	c	1.353	1.342	1.798	1.800	1.569	1.708
	d	1.391	1.350	1.791	1.800	1.500	1.688
	e	1.479	1.399	1.852	1.900	1.638	1.781
Reconstructed images shown in Fig. 3-38	$\hat{a}$	1.373	1.282	1.593	1.638	1.478	1.513
	$\hat{b}$	1.444	1.354	1.754	1.787	1.574	1.677
	$\hat{c}$	1.483	1.406	1.867	1.892	1.638	1.790
	$\hat{d}$	1.508	1.446	1.966	1.973	1.681	1.883
	$\hat{e}$	1.487	1.441	1.948	1.944	1.647	1.877

Table 3.7 (Continued)

Stage of the Processing	Picture	SNR_1	SNR_2	SNR_3	SNR_4	SNR_5	SNR_6
Anomalous scattering images shown in Fig. 3-35	$F_i(\underset{\sim}{k})$ N=2	1.271	1.191	1.344	1.311	1.235	1.285
	$F_i(k)$ N=3	1.354	1.250	1.566	1.535	1.398	1.500
	$F_i(k)$ N=4	1.383	1.300	1.615	1.613	1.433	1.562
	$F_i(k)$ N=5	1.483	1.405	1.873	1.880	1.621	1.781
Original series minus the dark-field term	a_{it}	1.324	1.204	1.520	1.550	1.449	1.404
	b_{it}	1.400	1.388	1.634	1.741	1.472	1.540
	c_{it}	1.375	1.355	1.786	1.793	1.584	1.672
	d_{it}	1.424	1.370	1.808	1.803	1.523	1.655
	e_{it}	1.516	1.419	1.863	1.900	1.669	1.747
Images reconstructed from the above iterated series	$\hat{a}_{it}$	1.364	1.266	1.576	1.605	1.459	1.479
	$\hat{b}_{it}$	1.430	1.333	1.729	1.744	1.549	1.626
	$\hat{c}_{it}$	1.469	1.382	1.824	1.838	1.610	1.722
	$\hat{d}_{it}$	1.485	1.413	1.900	1.894	1.641	1.788
	$\hat{e}_{it}$	1.476	1.416	1.909	1.890	1.623	1.800

Table 3.7 (Continued)

Stage of the Processing	Picture	SNR_1	SNR_2	SNR_3	SNR_4	SNR_5	SNR_6
Anomalous Image	$F_i(\underset{\sim}{k})_{it}$ N = 5	1.477	1.387	1.841	1.837	1.604	1.718
Inverse Filtering Shown in Fig. 3-34	e_f $\overline{\delta}$ = 0.2	1.469	1.371	1.815	1.857	1.604	1.745
	e_f $\overline{\delta}$ = 0.5	1.484	1.394	1.838	1.890	1.631	1.768

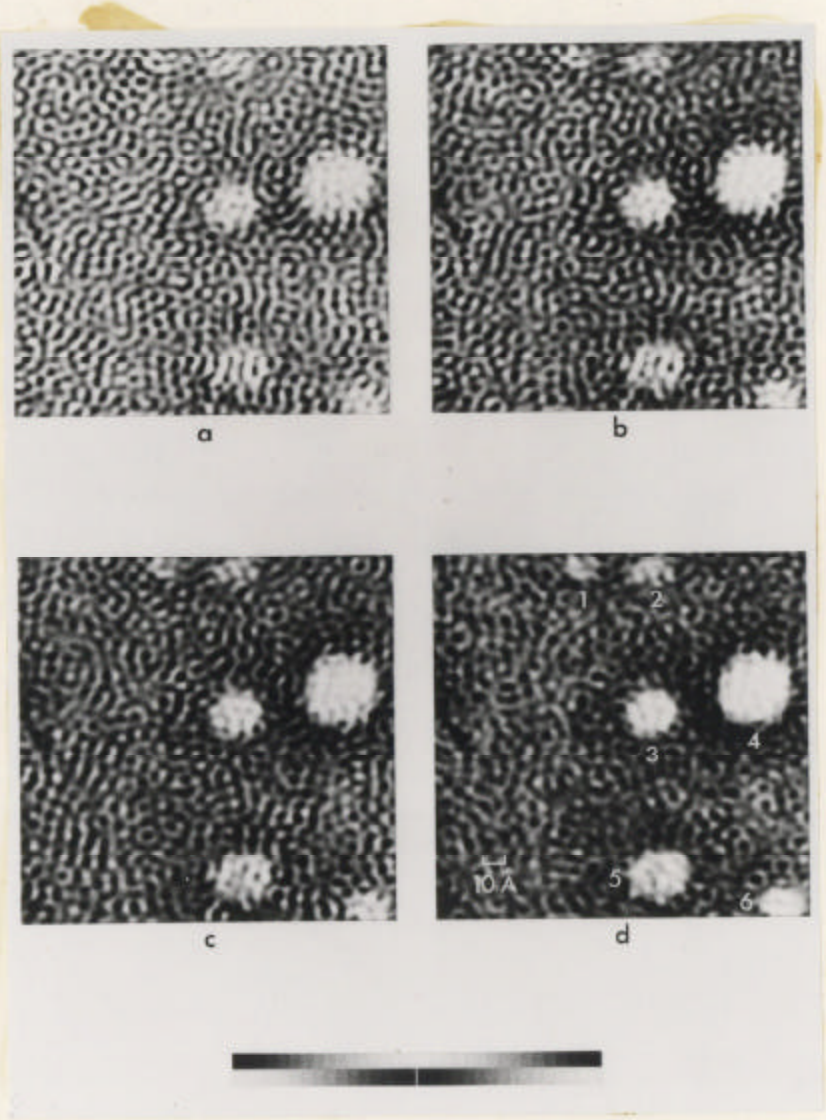


Fig. 3-35 Anomalous scattering image, $\mu(\underset{\sim}{r}_0)$, calculated by the Schiske restoration procedure from (a) N=2, (b) N=3, (c) N=4, and (d) N=5 pictures of the series of gold on carbon. Noise contributions near the zeros of the transfer function create artifacts, if N is too small, that reduce the selective contrast enhancement.

factors were shown (see Eq. (3.107)) to lead to similar adverse effects. All in all, these perturbations set major obstacles to the success of this technique, which only reaches its full potential at very high resolution. In our case, the gold clusters were visualized as very strong scattering centers, a fact which reflects itself clearly in the numerical estimates of the SNR of the input images (see Table 3-7; in Fig. 3-20e the SNR range from 1.4 to 1.9). A contrast enhancement was therefore difficult to demonstrate, although the restored complex image is a more valid representation of the object, and displays a broader frequency band.

By studying the Fourier transform of $\mu(\underset{\sim}{r}_o)$ (see Fig. 3-36a), one notices an increase in $|F_i(\underset{\sim}{k})|$ within the frequency band, a fact which one expects from the radial dependence of $f''_j(\theta)$ with the scattering angle. This transform exhibits a gap in the spatial frequency spectrum,where structural information has not been transferred in any pictures of the series. The restoration covers a spectral range corresponding to a superposition of the input spectra. As a consequence, the defocus step should be chosen in such a way that the Fourier spectrum, after reconstruction, is fully covered by the series. An inspection of the transform $F_i(\underset{\sim}{k})$ shows little gain in bandwidth, when compared to the original pictures. The defocus step that was selected is too small to enable a detectable resolution enhancement, since the contrast zones in the original pictures overlap quite significantly. Another difficulty lay in the residual axial astigmatism of the objective lens. Some resolution gain, however, is visible in the anomalous contrast image (see Fig. 3-35d) at the location of the small gold

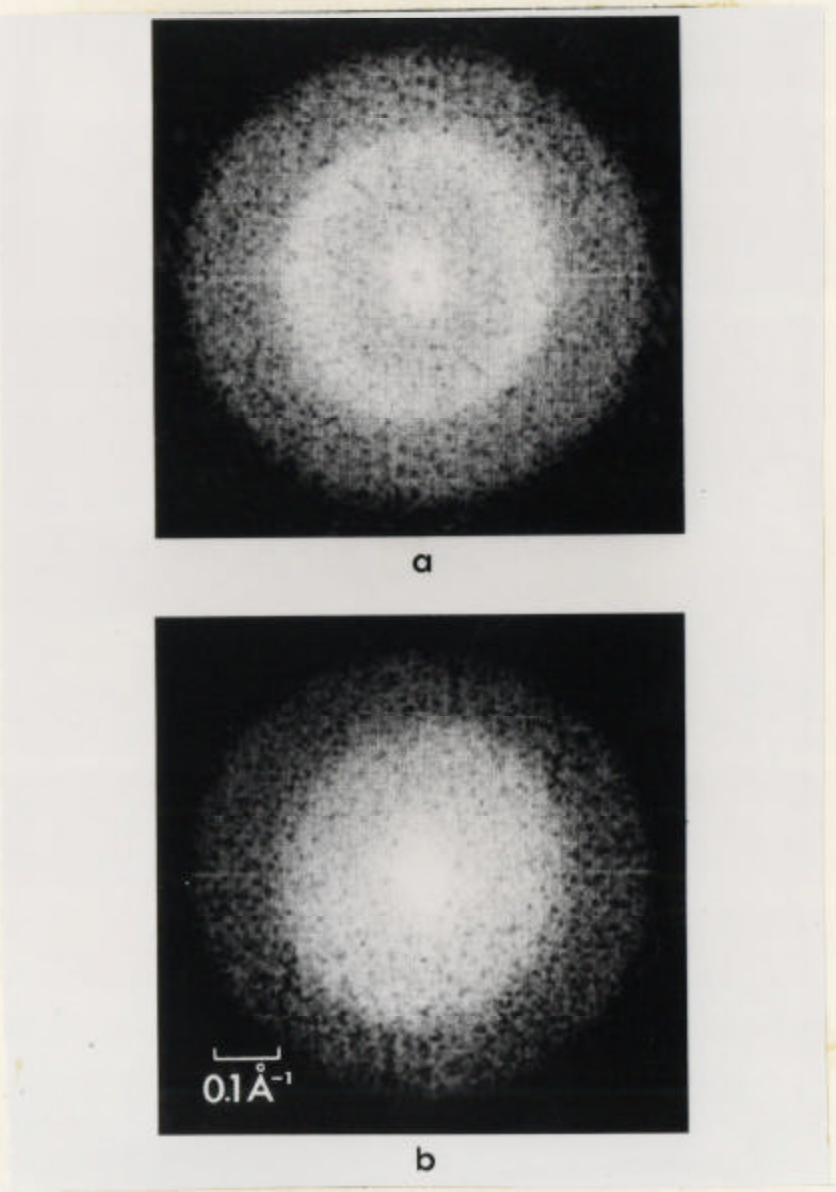


Fig. 3-36 (a) Imaginary part of the structure factor, $F_i(\underset{\sim}{k})$;
(b) Complex structure factor, $F(\underset{\sim}{k})$, which has been computed from N = 5 pictures from the original series of gold on carbon.

clusters (labeled i=1,2,6). Unfortunately, the low-order Fourier coefficients came out very strong and obliterated the high-frequency details contained in the amplitude-contrast contribution of the object. A similar high-frequency zone is visible in the transform $F_i(\underset{\sim}{k})$ of an isolated carbon area; this indicates a contribution from the real scattering factors through the curvature of the Ewald sphere. Thus, we conclude that specimen preparation and examination techniques set very stringent limits to the resolution and contrast enhancement of the Schiske scheme. The most restricting experimental factors, in our study, were the large spatial extent of the gold clusters (see Table 3-8) and their thickness, which accounted for their initial high contrast on the substrate background.

The phase contrast image $\mathcal{F}^{-1}\{F_r(\underset{\sim}{k})\}$ shows the predominance of low resolution details, due to its analytical singularity near the origin (see Appendix C). The cloudy intensity distribution shown in Fig. 3-37b can be visualized as a pictorial representation of the variations in specimen thickness, thus confirming the sizeable thickness of the large gold clusters. A high-pass filter makes the real image appear granular (see Fig. 3-37a), without any discernible structural details. When the low frequencies that image the outlines of the clusters are suppressed, the real images of the gold atoms are totally obscured by the carbon film. A theoretical contrast ratio of 5:1 was calculated for the phase image of isolated atoms, when all scattering angles are included in the image formation. The fact that no selective contrast is visible may be attributed to the low-resolution content of these clusters that was previously mentioned.

Table 3.8 Size of the gold clusters indicated in the micrographs of Fig. (3-30).

Gold cluster number	Average diameter (Å)
1	12
2	17
3	22
4	32
5	25
6	18

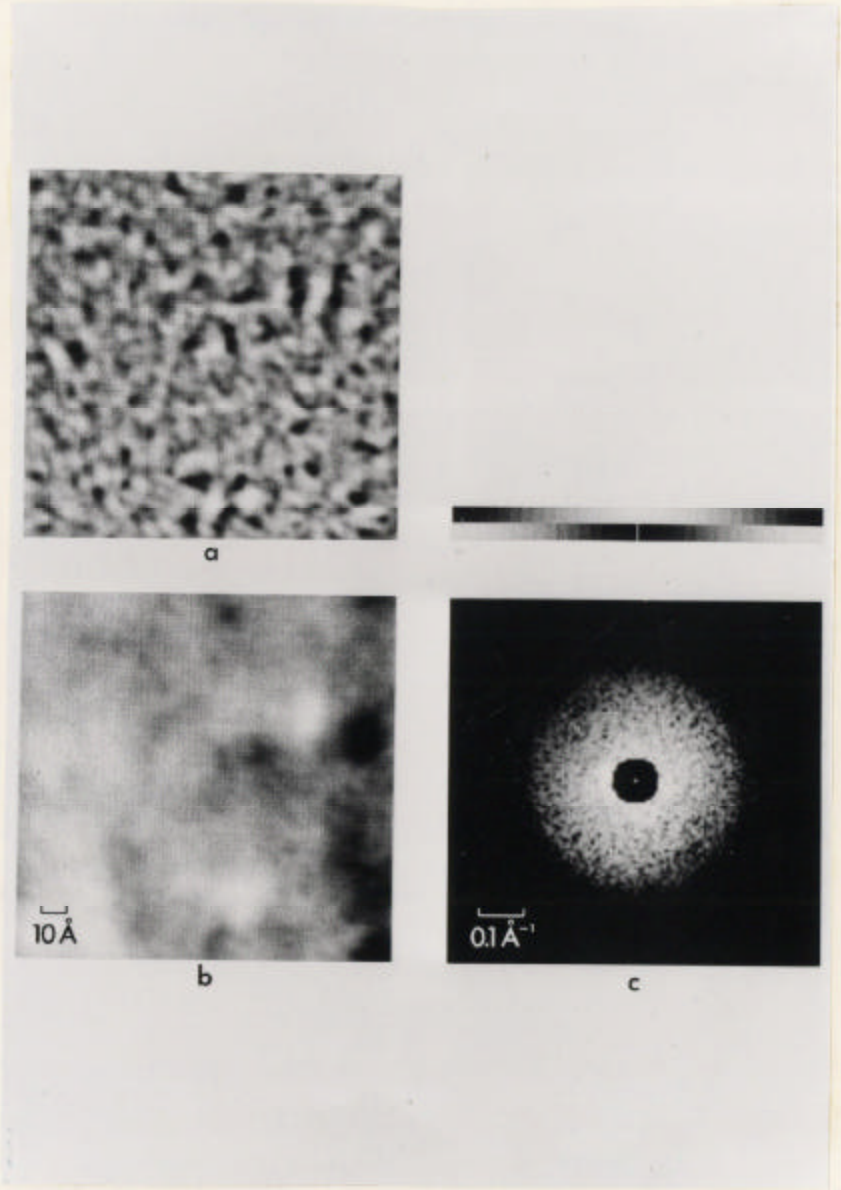


Fig. 3-37 (a) Phase contrast image $\phi(\underset{\sim}{r}_0)$, after high-pass filtering. (b) Unfiltered phase contrast image representing the contributions due to the real part of the scattering amplitude, to the curvature of the Ewald sphere, and to inelastic scattering. (c) High-pass filter $h^{HP}(\underset{\sim}{k})$ used to remove low-frequency artifacts.

To check the validity of the reconstruction, the focus series was calculated form the complex structure factor by choosing the electron microscope parameters equal to their values found by the least-squares fit (see Fig. 3-38). Normalization problems associated with the comparison of these pictures with the original series are examined in Appendix D. An excellent agreement was found between the respective images; this agreement could be measured quantitatively by the average least-squares deviations of the difference pictures. Such a figure of merit, defined by Eq. (D.5), was evaluated to be $\sigma_{I'-\hat{I}} = 19.5$ on a scale of 0 to 255, as opposed to an average distribution halfwidth in the original pictures of $\sigma_{I'} = 44.7$. Figure 3-39b illustrates the intensity distribution of a difference picture after contrast stretching; this difference picture exhibits almost no structural information. This lack of structural detail indicates that the dark-field term plays a minor role in the overall reconstruction scheme. To confirm this the quadratic term $|\psi_s^{(\ell)}(\underset{\sim}{r})|^2$ was evaluated for every phase factor from Eq. (D.6). Such a dark-field image, shown in Fig. 3-39a, illustrates the minor contribution to the bright-field image played by $|\psi_s^{(\ell)}(\underset{\sim}{r})|^2$. These dark-field images were then subtracted from the original series, and the restoration was implemented on this new focal series. The results of this iteration (see Table 3-7), indicate that the iterated amplitude-contrast image yields no improvement in the discrimination effect when compared to the first iteration loop.

3.5.4 Conclusion

By assuming a wave-optical theory of image formation, we have reached a quantitative understanding of the various contrast

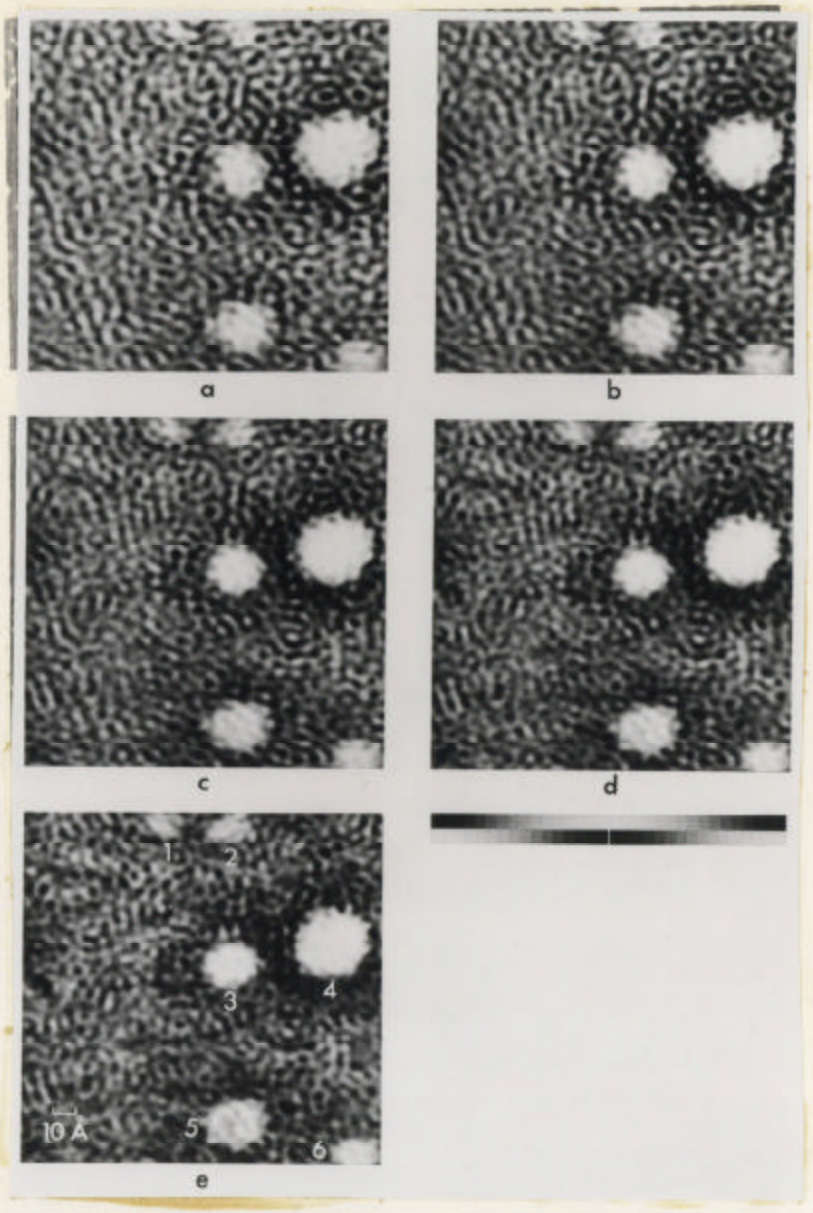
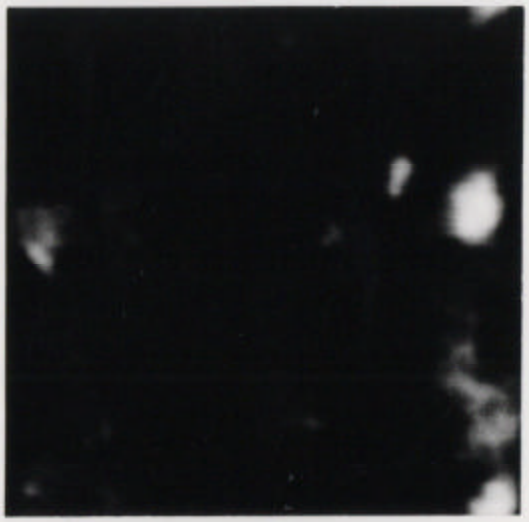
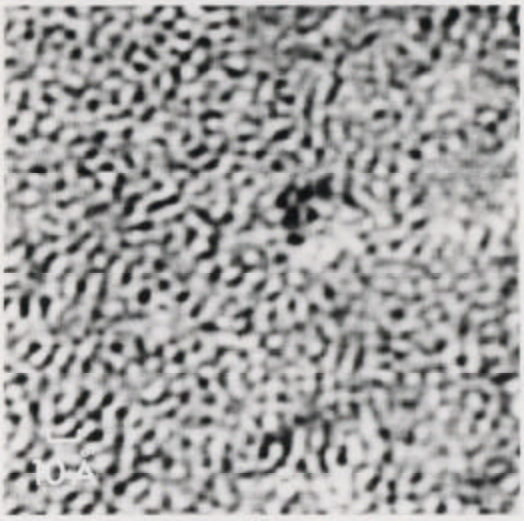


Fig. 3-38 Focus series of gold on carbon reconstructed from the complex structure factor. A comparison of this figure with Fig. 3-30 confirms the validity of the restoration.



a



b

Fig. 3-39 (a) Dark-field image calculated from the complex structure factor, at the same electron microscope parameters as Fig. 3-29e; (b) Difference image, $I_e - \hat{I}_e$, between two corresponding pictures of the original and the reconstructed focus series. This image was contrast stretched and DC shifted before being displayed.

contributions to the image of a weakly scattering object. Criteria have been used, at various stages of the processing, to ensure that the procedure yields the correct object function. Inadequacies in the specimen preparation and in the data recording, however, were shown to have significant effects on the success of the restoration. Moreover, the algorithm was implemented under the assumption that the incident illumination is coherent and monochromatic. Consequently, we ignored the high-frequency fall-off due to the partial coherence of the source. The correctness in the determination of the phase factors was checked by retrieving the input pictures from the computed structure factor, and comparing with the original series. This test does not provide, however, an absolute guarantee that the $\gamma_\ell(\underset{\sim}{k})$ are correct, because the algorithm is designed in such a way that an optimal solution to a given set of $\gamma_\ell(\underset{\sim}{k})$ and input pictures is always found. In addition, the restored imaginary image is relatively insensitive, within a certain range, to the absolute focus values.

Currently, there does not exist a method for checking the validity of the contrast assumptions upon which the whole reconstruction scheme is based. It is very likely that in our restoration, the kinematical scattering approximation broke down at the heavy atom locations. The resulting modulation of the primary wave would appear in the cosine-transferred part of the transform and thus further attenuate the selective enhancement effect. One can only claim the existence of such perturbations to interpret the experimental results. Their assessment requires additional experimental evidence, showing the relative strength of all sources of high-resolution contrast in a

particular defocus range.

Better results would have been achieved if the clusters had contained only a few atomic layers of gold. An improved high vacuum in the evaporator and a shallower shadowing angle would result in a deposition of very fine gold particles near the edges of the shadow. We would expect a twofold improvement in the reconstruction, with a decrease in the size of the clusters.

On the one hand, the predominance of low-resolution detail, which prevented a visible resolution enhancement, would be attenuated. The high spatial frequency content of the image, which was obscured by low resolution detail in our work, would show a noticeable gain in resolution after completion of the scheme. It would also permit a reduction of the inelastic scattering term by high-pass filtering. On the other hand, by starting from small clusters which are hardly visible on a granulated background, the selective enhancement would prove its efficiency by bringing up strong peaks at heavy atom sites. The thickness dependence of the discrimination method would play a minor role and thus the imaging conditions would more closely meet the required assumptions of the existing theories.

What is needed is a calculated model with which to check the restoration program and appraise the various contrast sources. The importance of inelastic scattering, radiation damage, quantum electron effects, inaccuracies due to the density measurement, and errors in the determination of the electron optical parameters should be assessed.

As a conclusion, we point out that object restoration methods may misrepresent the input data, if a careful experimental verification of the underlying theoretical assumptions is lacking. The Z discrimination technique will only show its potential when adequate specimen preparation techniques are evolved. Instrumental instabilities and specimen damage make electron microscopy at the atomic resolution range very unreliable. Ultra-high vacuums, superconducting objective lenses, ultra-sensitive recording sensors, in conjunction with restoration schemes, certainly will open new avenues of research in the pursuit of atomic resolution. It is therefore not inconceivable that energy filtering microscopes, such as the scanning transmission electron microscope[72], equipped with the latest advances in technology, will be capable of visualizing molecules at the atomic level. The tolerable radiation damage threshold will govern the maximum exposure time and hence the initial image contrast for a particular imaging mode. At present it would seem that a scanning transmission electron microscope, with energy filtering, would yield the best solution to the radiation damage problem. This is because the scanning transmission mode submits the specimen to less radiation exposure, since electron detectors and amplifiers are used to record the image.

This chapter has stressed the problems involved in extending the existing microscope resolution by image processing techniques. In the next chapter, we shall concern ourselves with the numerical extraction of a signal from a noisy background. As the examples will demonstrate, the processed images allow a quantitative contrast analysis when observing an imperfect crystal at high resolution.

CHAPTER FOUR

APPLICATION OF COMPUTER TECHNIQUES FOR NOISE REMOVAL TO THE QUANTITATIVE INTERPRETATION OF HIGH-RESOLUTION MICROGRAPHS

CHAPTER FOUR

APPLICATION OF COMPUTER TECHNIQUES FOR NOISE REMOVAL TO THE QUANTITATIVE INTERPRETATION OF HIGH-RESOLUTION MICROGRAPHS

4.1 Introduction

Various sources of noise limit the amount of structural information that can be extracted from an electron micrograph. These sources of noise include: statistically random noise originating from the quantum nature of the signal and from the granularity of the film, "snow" that is introduced during the VFC scan, and spurious periodic noise. In the following analysis we shall consider only random noise since the other noise sources can be eliminated in the preprocessing stage.

In order to simplify the analysis, we shall assume that this statistically random noise is additive and uncorrelated to the signal. These conditions are experimentally fulfilled when the contrast variations in the image have the same order of magnitude as noise fluctuations. The detectability of a weak signal embedded in a noise background has been discussed in Sec. 1.4.1, where a relationship between the number of electrons impinging upon a given area of the recording sensor and the contrast level C_i was derived.

Let us now consider two image points (d_1 and d_2) whose relative intensities differ by the detection threshold: $d_1 - d_2 = 2\bar{d}C_i$, where $\bar{d}$ is the picture mean. Then, according to Eq. (1.82), the ratio of their respective noise fluctuations equals

$$\frac{\sigma_{n_1}}{\sigma_{n_2}} = \sqrt{\frac{d_1}{d_2}} \simeq 1 + C_i \simeq 1 + \frac{f_c}{2\sqrt{\bar{d}}} \qquad (4.1)$$

Consequently in the case when the illumination level $\bar{d}$ is high and the signal modulation is comparable to the noise background, the noise statistics are spatially invariant and are approximately independent of the signal. These assumptions are valid when the number of electrons recorded during the exposure is large enough, and when the image does not exhibit sharp contrast variations. These assumptions are not likely to hold when the microscope is operated in the dark-field mode, unless long exposure times are employed. In this case, specimen stability and contamination rates become the limiting resolution factors.

The above analysis has led us to the conclusion that the noise is reasonably uncorrelated to the signal. If we now assume that the noise obeys white-Gaussian statistics, we can expect an increase in the signal-to-noise ratio of $\sqrt{N}$ by superimposing N separate micrographs of the same object. The computer is employed to carry out the registration (translation and rotation) of the micrographs, before performing a linear averaging. This averaging is a useful method for increasing the visibility of detail in electron micrographs that contain regular arrays of like units. After this processing we obtain an enhanced version of the average object, since the variability in the appearance of individual image units has been attenuated. We shall now present specific examples of the enhancement of lattice images by computer processing.

4.2 Digital Filtering of Periodic Images

4.2.1 Processing Algorithm

Although high-resolution lattice images are often partially obscured by artifactual detail, filtering procedures have been evolved which allow a precise determination of the repeat spacing. This is accomplished by attenuating random perturbations, which are brought about by local specimen buckling or by statistical noise. By filtering one can then convert an image that is quite lacking in apparent periodicity into an image that contains a regularly repeating structure.[92]

Two alternative processing schemes for enhancing a periodic signal in the presence of noise are available. One method, carried out in real space, is the computer analogue of a photographic integration technique[93]. The other method, which can also be performed optically, consists of selectively masking the image transform so that only certain Fourier components contribute to the final reconstruction[94]. The relationship between these two methods, when they are digitally implemented, is presented in Appendix E. We shall now discuss the relative merits of each technique.

4.2.1.1 Fourier filtering

The Fourier transform of a micrograph can be used to detect periodicities that would be invisible to visual inspection. A periodic image transform shows maxima, which can be separated from the transform of the noise by using windows of variable size. The portion of the transform which lies outside the window is then set to zero, and the

periodic image is obtained by inverse transformation.

Several remarks, regarding possible transform artifacts, are worth making. Firstly, if the number of fringes in the box is not an integer, the various orders of the discrete Fourier decomposition of the fringe pattern will not fall on the sampling grid. One must therefore look for the maxima by interpolation across a sampling mesh, and assign the correct phase angles to the maxima. This problem is not critical, however, when one is dealing with a lattice image that is mostly sinusoidal and thus contains only one strong first-order maximum in its transform. Another reason for including, as nearly as possible, an integer number of fringes in the processed image, is to avoid edge discontinuities in the discrete transform. These discontinuities at the truncation of the image bring up spikes on the transform axes that could overlap the diffraction spots. Lastly, one should remove strong optical density gradients running across the picture, since these gradients induce spikes in the transform in a direction perpendicular to the gradient.

Once the picture has been scanned and has undergone a contrast stretch, one proceeds to geometrically align the translational unit vectors in the periodic image with the picture boundaries. A box containing an integer number of repeat units is then cut out (the number of repeat units is determined by visual inspection if the periodicity is visible, or by an a priori knowledge of the selected signal). The dimensions of the box are then modified so that the number of pixels along one edge is a multiple of two; this requirement is imposed by the FFT algorithm. The Fourier transform is then multiplied by a

template that is transparent to the detectable periodic signal and inverted, thus yielding an enhanced lattice image. One advantage of this method in comparison to the real space processing scheme is that it permits a variable number of components to enter into the reconstruction. This proves to be very valuable when one deals with nearly periodic images, since local lattice distortions can be preserved by merely opening the Fourier windows. If the filter windows are chosen to be one pixel wide, the analysis presented in Appendix E tells us that the two processing schemes are equivalent. The fact that this equivalence is only partially true is a consequence of the numerous artifacts that can be introduced during the computation.

4.2.1.2 Convolution averaging

Convolution averaging is particularly useful for the enhancement of periodic images that contain random noise. If we assume that the noise is statistically random and additive to the signal, a signal-to-noise ratio increase of $\sqrt{N}$ is obtained by translating the image by the amount of the periodicity and then superimposing the image N times upon itself[26]. This method, however, is restricted to processing lattice images from perfect crystals, since any variation in lattice periodicity will cause a phase error in the final image.

When implementing this technique one follows the same preliminary steps as in the previous method for aligning the input area and determining the spatial extent of the repeat unit. In the average picture, any ill-defined periodicity will appear as sharp discontinuities at the boundaries of the repeat unit. This scheme fails entirely if strong picture trending obscures the periodic features.

4.2.2 Enhancement of Lattice Images Embedded into Noise

4.2.2.1 Silicon

A) Specimen preparation and data collection

A [111] crystal wafer of silicon was subjected to an oxidizing atmosphere at 1050°C, producing the following reaction:

$$3Si + xO_2 \rightarrow Si + 2SiO_x$$

This reaction produce must be considered to be less than desirable for lattice imaging experiments. Electron transparent specimens of the reaction product were obtained by chemical thinning in a 90% HNO_3 - 10% HF solution. These specimens were then examined in a JEOL 100B electron microscope operating at 100 keV.

In general, the selected-area-diffraction patterns obtained from these specimens were typical of an amorphous medium, but in one case evidence of weak Bragg maxima was found (see Fig. 4-1b). These weak maxima were used to obtain a lattice image from the remnant Si crystal. The lattice image was obtained by allowing the weak ±220 maxima to interfere with the transmitted beam in the final-image plane (see Fig. 4-1a); the working magnification of the electron microscope was 415,000X.

In Fig. 4-1a we see a very faint fringe pattern superimposed upon a structureless background. We shall now discuss the recovery and enhancement of this weak periodic signal[95].

B) Detection of the periodic information

The electron micrograph containing the weak Si fringes was

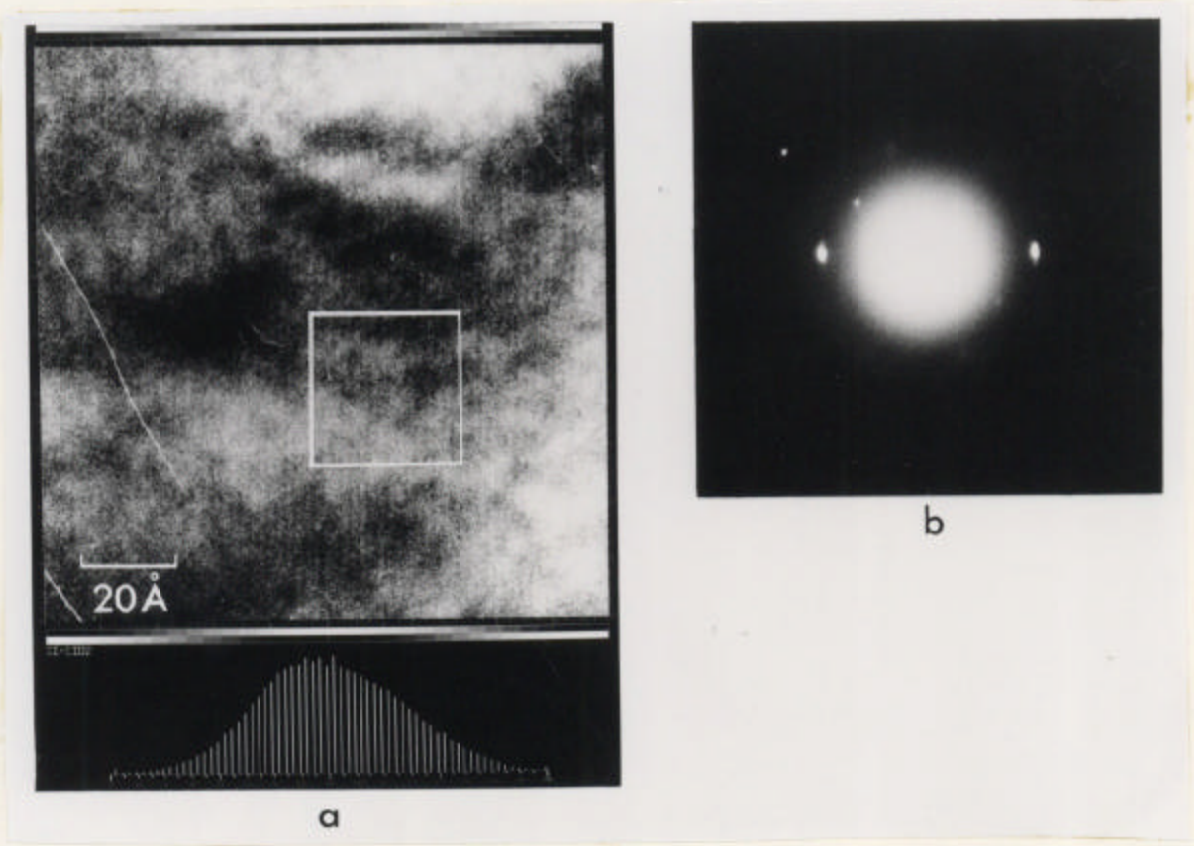


Fig. 4-1. (a) Computer displayed electron micrograph of Si-SiO_x film, with its associated histogram of optical densities. The area enclosed in the box is 32.6 x 32.6Å and was selected for further processing (this area contains 17 lattice fringes). (b) Selected area diffraction pattern corresponding to the thin film imaged in (a). The interference of the transmitted beam and the ±220 maxima at the image plane produced the observed lattice image.

first copied onto fine-grain film at a magnification of 5X in order to improve the fringe visibility with respect to the substrate. This photographic enhancement was necessary, because the CRT scanner was not designed for low-contrast input pictures. Next, the optical densities of Fig. 4-1a were measured and encoded at a sampling step size of $\Delta = 25\mu m$. This step size corresponds to $0.12\mathring{A}$ on the object scale. A visual appraisal of the angular orientation of the fringes indicated that they were running at a slight tilt.

A box containing $m = 17$ fringes was subsequently cut from Fig. 4-1a for further processing. The number of fringes in this box was first estimated visually; this estimation was later improved by an examination of the discrete Fourier transform. One expects the first-order diffraction maximum to be m pixels away from the DC term if m fringes have been selected (see Appendix E). If the maximum falls on a sampled point of the transform, the number of fringes included in the box is an integer. Otherwise, the dimensions of the box must be altered until m becomes an integer.

An examination of the discrete transform shown in Fig. 4-2b reveals the presence of vertical spikes. These spikes are caused by the sharp picture trending in the area that was selected for processing. Since these spike artifacts did not interfere with the diffraction spots, the spatial-frequency filter which separates the desired periodic image from the random background led to a reconstructed lattice image in which picture trending has been suppressed.

The effect or progressively opening the Fourier windows is illustrated in Fig. 4-2. By gradually allowing local lattice

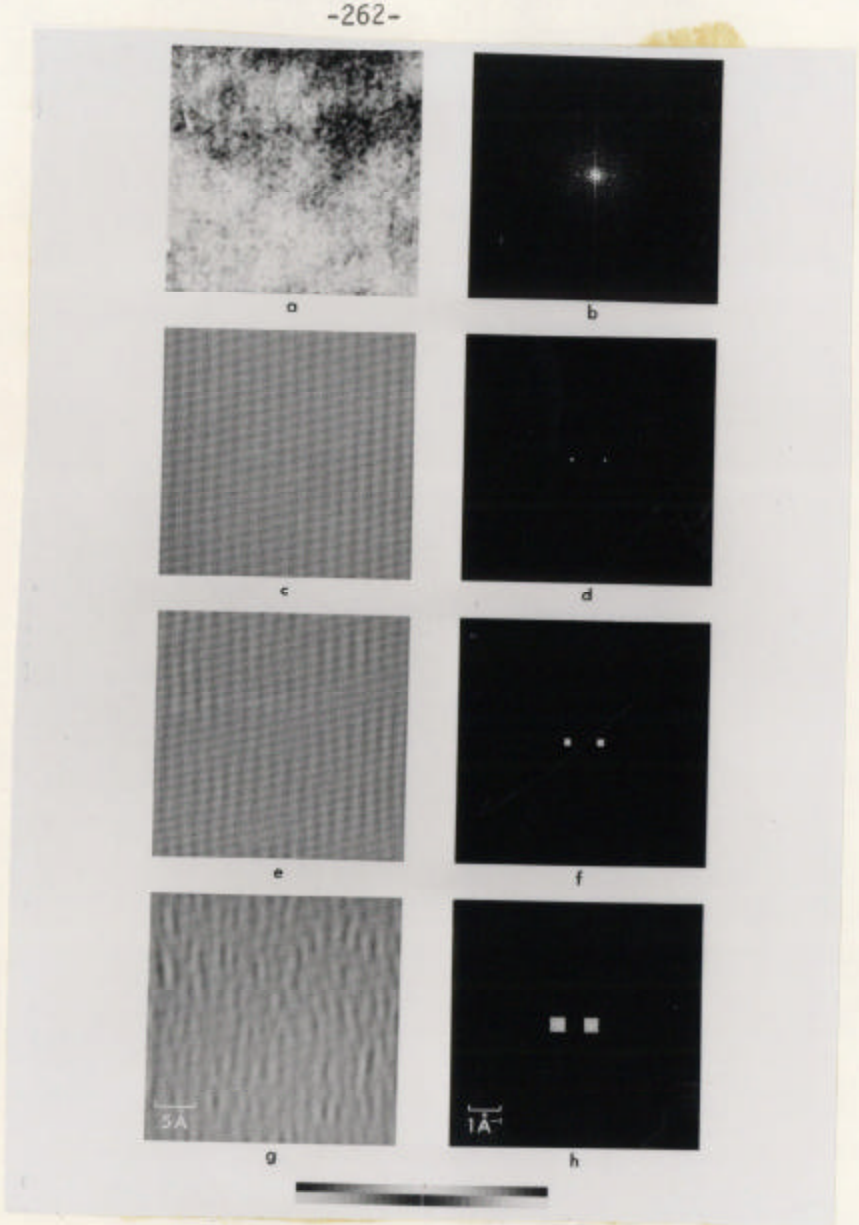


Fig. 4-2. Effects of varying the size of the Fourier windows on the reconstruction of the silicon lattice image. The pictures presented show (a) the 32.6 x 32.6Å box of Fig. 4-1a after contrast stretching, (b) the Fourier transform of the digitized area where the vertical streak is caused by density gradients visible in the input. The remaining picture pairs show the spatial-frequency filter on the right column and the corresponding three-beam reconstructed image on the left column. The lateral size of a Fourier window is: (d) 0.098Å^{-1}, (f) 0.228Å^{-1}, and (h) 0.49Å^{-1}. By opening the window, we clearly destroy the periodicity of the fringe pattern by superimposing noise from diffuse scattering.

irregularities to contribute to the reconstruction, we can evaluate the amount of distortions that exists in the original image (e.g., if ΔX is the lattice periodicity in $\mathring{A}$ and w the size of the Fourier window in $\mathring{A}^{-1}$, then it can be shown that the periodicities that are transmitted by the window lie between $(1/\Delta X + w/2)^{-1}$ and $(1/\Delta X - w/2)^{-1}$). In the filtered image the (220) atomic planes in silicon can easily be resolved[95].

The convolution averaging method cannot be applied with much success to the Si lattice image (see Fig. 4-3). In contrast to the preceding technique where density gradients are automatically removed by the filtering step, one observes distortions in the averaged picture, which are due to the nonuniform grey level in the input picture. The artifacts which are induced by this picture trending destroy the clarity of the processed image. Furthermore, this averaging method proves its inapplicability by significantly amplifying the lattice distortions in the output image.

4.2.2.2 Lattice images of gold

A) Experimental procedure

Electron-transparent gold films with <001> orientation were prepared by vacuum deposition onto a cleaved NaCl crystal. The films were removed by dissolving the sodium chloride; they were then mounted on a 400 mesh microscope grid. These specimens were observed with a JEOL 100B operating at 100 keV and a magnification of 400,000X. The electron gun was mechanically tilted (to reduce the effect of objective lens aberrations) so that several operating reflections fell onto a circle centered at the optic axis[96].

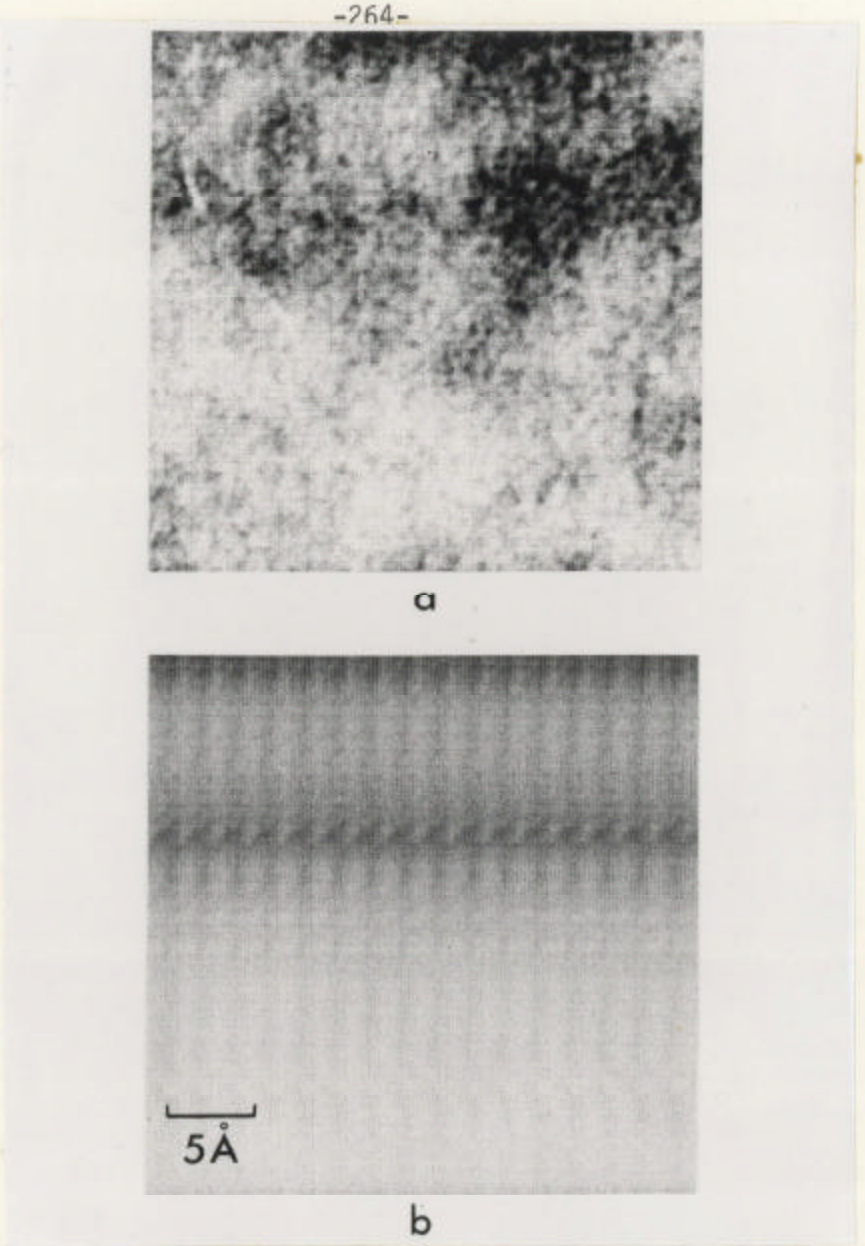


Fig. 4-3. Convolution averaging of a weak periodic signal that is obscured by overlapping density gradients. The pictures show (a) the 32.6 x 32.6Å box of Fig. 4-1a after contrast stretching. (b) the result of superimposing the 17 unit cells and taking the linear average. Artifacts in the reconstruction are induced by the non-uniform contrast of the input image, low signal-to-noise ratio, and by the fact that the original fringe system is distorted.

Lattice images were obtained from the gold film by allowing the transmitted beam and the (200), (020), and (220) diffracted beams to pass through a 120 μm objective aperture and interfere in the final image plane. The resulting electron micrograph showed crossed-lattice planes corresponding to the (200) and the (020) atomic planes (see Fig. 4-4). The absence of the (220) lattice planes in the final image was due to thermal bending, which locally tilted the (220) maxima away from the Bragg condition. Consequently, the interference image formed with the strong transmitted beam and the very weak (220) beam could not be recognized in the micrograph. Since the intensity of the diffraction maxima are extremely sensitive to crystal thickness and orientation, we expect a lattice image to vary accordingly throughout the specimen.

B) Results of the filtering step

An electron micrograph showing the (200) and (020) atomic planes in gold was copied onto film at a magnification of 4X . It was then digitized every 50 μm; this sampling step corresponds to 0.314Å on the object scale. Next, a 44.8 x 44.8Å box was cut from the digitized micrograph shown in Fig. 4-4. The dimensions of the box were chosen so as to include an integer number of fringes (i.e., m = n = 22).

After noise filtering Fig. 4-4, a more clear picture of the intersecting lattice planes was obtained (see Fig. 4-5). Since areas with good lattice periodicity were visible from the start, convolution averaging was performed on a 12.2 x 12.2Å area of Fig. 4-4. The result of this processing is shown in Fig. 4-6, where atom locations are

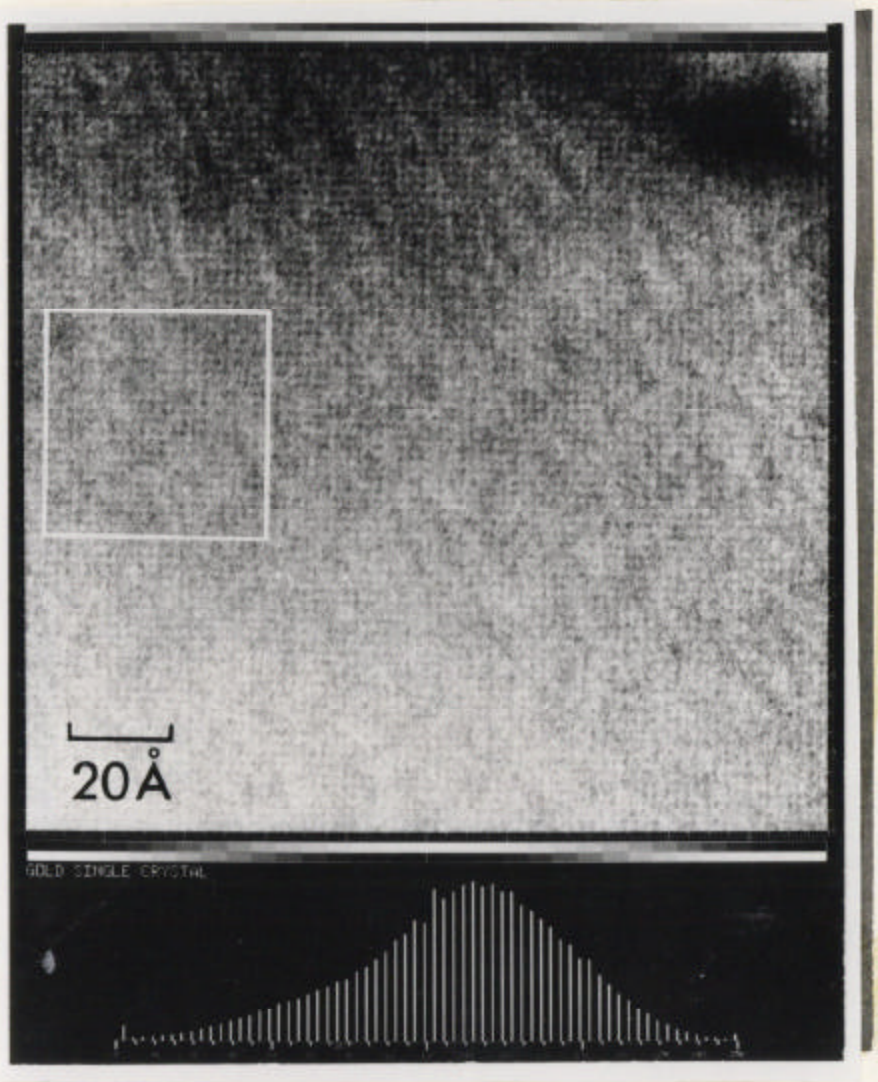


Fig. 4-4. Computer displayed crossed-lattice image of the (200) planes in a single crystal of gold and its associated histogram of optical densities. The area selected for further processing is enclosed in a 44.8 x 44.8Å box. This area contains 22 crossed-lattice fringes.

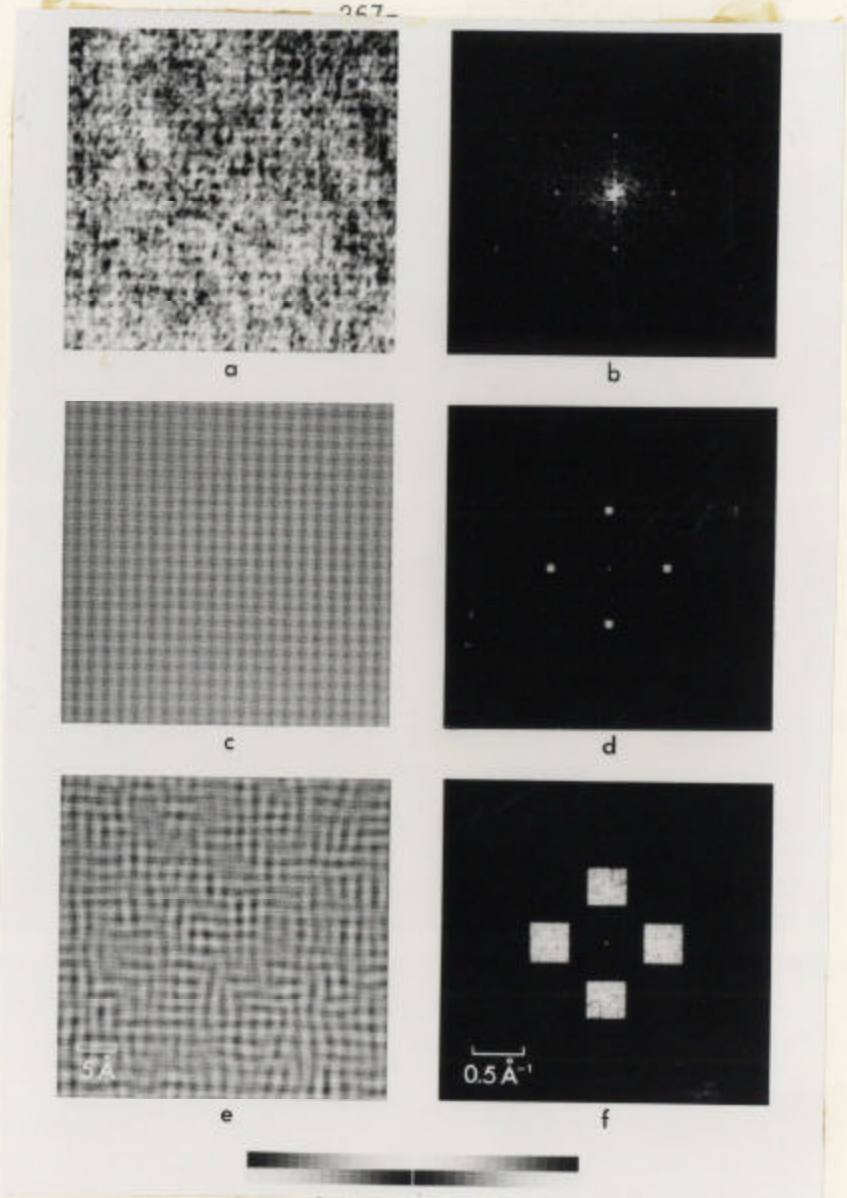


Fig 4-5. Effects of varying the size of the Fourier windows on the reconstruction of the crossed-lattice image of gold. The pictures presented show (a) the 44.8 x 44.8Å box of Fig. 4-4 after contrast stretching, (b) the Fourier transform of the digitized area where the streaks near the origin are caused by the large scale density gradients that are visible in the input picture. The remaining picture pairs show the spatial-frequency filter on the right column and the corresponding three-beam reconstructed image on the left column. The lateral size of the Fourier window is successively (d) 0.075Å^{-1}, and (f) 0.373Å^{-1}. By opening the window we allow local lattice distortions to contribute to the reconstructed image.

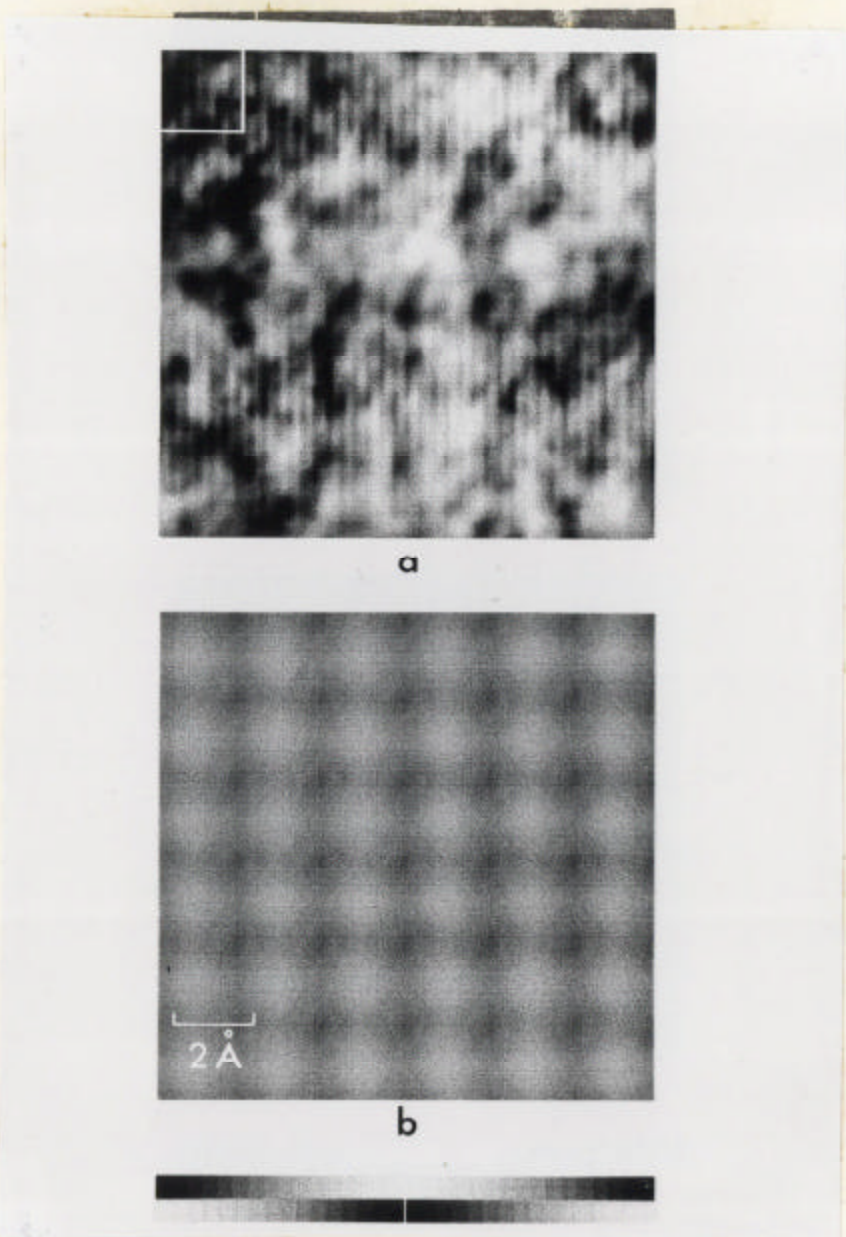


Fig. 4-6. Convolution averaging of a crossed-lattice image of gold. The pictures show (a) a portion of Fig. 4-5a that contains 6 lattice fringes both horizontally and vertically; the box in the upper left-hand corner represents a single atom of gold. (b) the result of superimposing 6 x 6 unit cells and taking the linear average. The 2.04Å spacing of the (200) planes in gold can now be seen.

now becoming visible. By comparing Fig. 4-5 to Fig. 4-6 we see that local lattice distortions that are averaged out by the convolution averaging scheme can be maintained in the Fourier reconstruction scheme by enlarging the windows[97].

A promising area of application for these processing schemes is in the study of the influence that crystal defects have on lattice images. By subtracting the perfect periodicity from the original micrograph, local fringe bending and terminations will be revealed. Unfortunately there does not exist a one-to-one correspondence between a fringe pattern and the atomic planes because of the sensitivity of the image to the diffraction geometry[98]. Lattice imaging experiments conducted under known diffracting conditions, coupled with both image processing and theoretical contrast analysis, should provide a powerful means for investigating crystalline defects.

We shall now present an analysis of lattice image contrast based on the many-beam dynamical theory of electron diffraction[96,98]. This analysis will serve as an introduction to the more difficult subject of contrast interpretation of images of crystal imperfections.

4.2.3 Interpretation of Filtered Lattice Images; Dynamical Theory for Perfect Crystals

4.2.3.1 Outline of the theory of high-energy-electron diffraction from perfect crystals

In the elastic scattering approximation, the propagation of a fast electron in a crystal can be described using the time-independent Schrödinger equation:

$$\nabla^2\psi(\underset{\sim}{r}) + \frac{2m_e}{\hbar^2}[E + V(\underset{\sim}{r})]\,\psi(\underset{\sim}{r}) = 0 \tag{4.2}$$

where E is the energy of the fast electron, m_e is the relativistically corrected electron mass, and $V(\underset{\sim}{r})$ is the potential energy of the electron in the crystal. We can expand $V(\underset{\sim}{r})$ in a Fourier series as

$$V(\underset{\sim}{r}) = \frac{\hbar^2}{2m_e}\sum_g U_g \exp(2\pi i \underset{\sim}{g}\cdot\underset{\sim}{r}) \tag{4.3}$$

where $\underset{\sim}{g}$ is the reciprocal lattice vector corresponding to the atomic planes $(hk\ell)$. The Fourier coefficient of Eq. (4.3) can be written as

$$U_g = \frac{m_e}{m_{eo}}\frac{e^{-M_g}}{\pi\Omega}\sum_j f_j(\theta)\exp(-2\pi i \underset{\sim}{g}\cdot\underset{\sim}{r}_j) \tag{4.4}$$

where m_{eo} is the rest mass of the electron, M_g is the Debye-Waller factor which takes account of the thermal vibrations of the crystal, Ω is the volume of the unit cell, the summation is over the j atoms of the unit cell, $f_j(\theta)$ is the atomic scattering factor for electrons calculated on the first Born approximation, and r_j is a vector from the origin of the unit cell to the j^{th} atom of the unit cell.

If we assume for the moment that we are dealing with the propagation of a fast electron in an infinite crystal, then our wave function can be written in the Bloch representation as

$$\psi(\underset{\sim}{r}) = \sum_j\sum_g C_g(j)\exp[2\pi i \underset{\sim}{k}_g(j)\cdot\underset{\sim}{r}] = \sum_j B_j(\underset{\sim}{r}) \tag{4.5}$$

where the summation over j accounts for the Bloch-wave degeneracy in energy and $\underset{\sim}{k}_g(j) = \underset{\sim}{k}_j + \underset{\sim}{g}$. In the case of a semi-infinite crystal the Bloch wave function can be written as

$$\psi(\underset{\sim}{r}) = \sum_j \sum_g \Theta_j \, C_g(j) \exp[2\pi i \underset{\sim}{k}_g(j) \cdot \underset{\sim}{r}] = \sum_j \Theta_j \, B_j(\underset{\sim}{r}) \qquad (4.6)$$

where the excitation coefficients Θ_j are determined by wave matching at the entrance surface of the crystal.

Substituting Eqs. (4.5) and (4.3) into Eq. (4.2) and Fourier transforming, we obtain the infinite set of equations

$$\begin{matrix} \vdots \\ [K^2 - |\underset{\sim}{k}_g(j)|^2] \, C_g(j) + \sum_h{}' \, U_h C_{g-h}(j) = 0 \\ \vdots \end{matrix} \qquad (4.7)$$

where $K^2 = \chi^2 + U_o$, χ is the vacuum wave vector of the incident monochromatic beam of electrons, and the prime on the summation indicates that the summation over $h = 0$ is omitted.

If we now assume that we are in the Laue diffraction geometry (z is the coordinate parallel to the zone axis of the diffracting planes), and we assume that only reflections on the zeroth Laue zone are excited, then Eq. (4.7) can be written in matrix form as

$$\underline{\underline{A}} \, \underline{C}(j) - \gamma(j) \, \underline{C}(j) = 0 \qquad (4.8)$$

where $\underline{\underline{A}}$ is an $\hat{N} \times \hat{N}$ matrix (in the $\hat{N}$ beam approximation with elements

$$A_{gh} = \begin{cases} s_g & \text{for } g = h \\ U_{g-h/2K} & \text{for } g \neq h \end{cases}$$

s_g is a parameter which describes the angular deviation of the incident beam from the classical Bragg angle for the reflections g, $\underline{C}(j)$ is a column vector with elements $C_g(j)$, and $\gamma(j) = k_z(j) - K_z$.

By setting the secular determinant equal to zero:

$$\det[\underline{\underline{A}} - \gamma(j)\,\underline{\underline{I}}] = 0$$

where $\underline{\underline{I}}$ is the identity matrix, we obtain a formula, called the dispersion equation, that relates the possible wave vectors $\underset{\sim}{k}(j)$ that can be excited in the crystal to the total electron energy. The solutions to the dispersion equation map in reciprocal space as dispersion surfaces. In the $\hat{N}$ beam approximation there are $\hat{N}$ of these surfaces, which correspond to the $\hat{N}$ different kinetic energy states, $\hbar^2k^2(j)/2m_e$, of an electron of fixed total energy. Figure 4-7 shows a six-beam dispersion surface with its associated Ewald sphere construction.

The remaining unknowns of the theory are the Bloch-wave excitation coefficients Θ_j. By wave matching at the entrance surface of the crystal, and noting that in the high-energy approximation the amplitudes of elastically reflected waves are negligible, we can write

$$\underline{\underline{C}}\,\underline{\Theta} = \underline{\Phi}(0) \tag{4.9}$$

where $\underline{\underline{C}}$ is a matrix whose columns are the eigenvectors $\underline{C}(j)$, $\underline{\Theta}$ is a column vector with elements Θ_j , and $\underline{\Phi}(0)$ is a column vector with elements $\Phi_g(0) = \delta_{go}$. In the case of a centrosymmetric crystal we know $\underline{\underline{\tilde{C}}}\,\underline{\underline{C}} = \underline{\underline{I}}$, so that Eq. (4.9) can be written

$$\underline{\Theta} = \underline{\underline{\tilde{C}}}\,\underline{\Phi}(0) \tag{4.10}$$

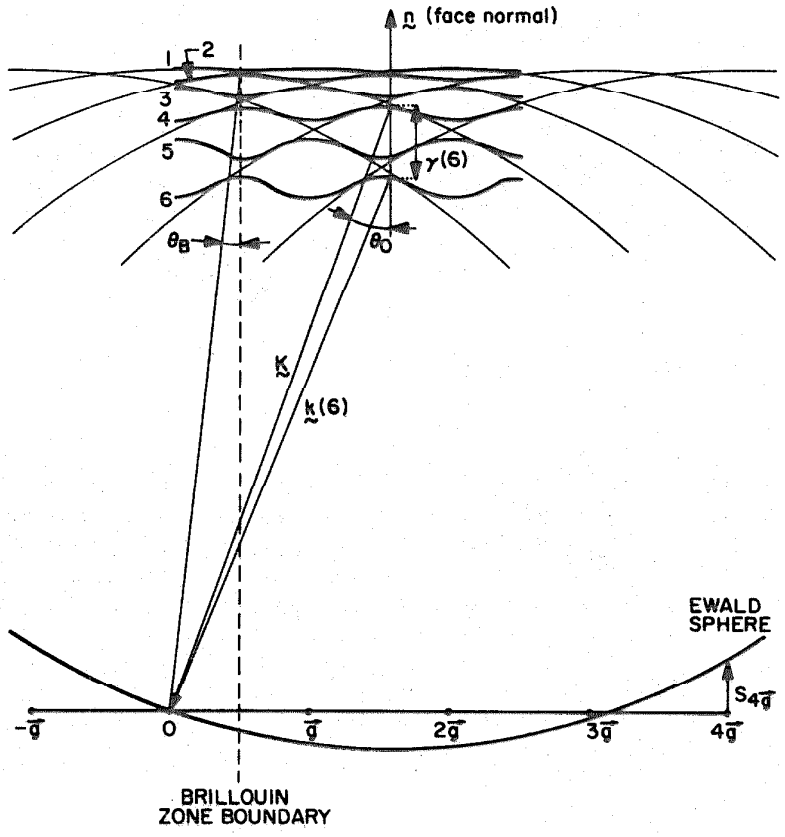


Fig. 4-7. Diagram (not to scale) of the dispersion surface and the associated Ewald sphere construction for diffracting conditions in which six systematic reflections are excited. Note that the dispersion surface has six branches since six beams are excited. This diagram also shows the deviation parameter S_{4g}, which is negative since the 4g beam lies outside the Ewald sphere. The eigenvalue $\gamma(6)$ and its associated eigenvector $\underset{\sim}{k}(6)$ are also shown.

Now, in order to obtain the wave function at the exit surface of the crystal, we shall first rewrite $\psi(\underset{\sim}{r})$ in the Darwin representation:

$$\psi(\underset{\sim}{r}) = \sum_g \phi_g(\underset{\sim}{r})\, D_g(\underset{\sim}{r}) = \sum_g \phi_g(\underset{\sim}{r})\, \exp(2\pi i \underset{\sim}{k}_g \cdot \underset{\sim}{r}) \tag{4.11}$$

where $D_g(\underset{\sim}{r})$ is a plane wave traveling in the $\underset{\sim}{k}_g$ direction. Then, equating Eq. (4.6) and Eq. (4.11) we obtain

$$\phi_g(\underset{\sim}{r}) = \sum_j \Theta_j \; C_g(j)\, \exp\{2\pi i[\underset{\sim}{k}(j) - \underset{\sim}{K}]\cdot\underset{\sim}{r}\}$$

$$= \sum_j \Theta_j \; C_g(j)\, \exp[2\pi i \gamma(j) z] \tag{4.12}$$

In matrix form we can write Eq. (4.12) as

$$\underline{\Phi}(z) = \underline{\underline{C}}\;\underline{\underline{E}}\;\underline{\Theta} = \underline{\underline{C}}\;\underline{\underline{E}}\;\underline{\underline{\tilde{C}}}\;\underline{\Phi}(0) \tag{4.13}$$

where $\underline{\underline{E}}$ is a diagonal matrix with elements

$$E_{ij} = \{\exp[2\pi i \gamma(j) z]\}\; \delta_{ij}$$

By differentiating Eq. (4.13) with respect to z, and combining the result with Eq. (4.8) we obtain

$$\frac{d}{dz}\,\underline{\Phi}(z) = 2\pi i\, \underline{\underline{A}}\,\underline{\Phi}(z) \tag{4.14}$$

which has the formal solution

$$\underline{\Phi}(z) = e^{2\pi i\, \underline{\underline{A}} z}\, \underline{\Phi}(0) \tag{4.15}$$

Equation (4.15) defines the amplitudes of the various waves at the

exit surface of the crystal.

4.2.3.2 Absorption

In high-energy-electron diffraction absorption is considered to be the effect of inelastic scattering on the elastic scattering. Yoshioka has shown that the equation which describes this process (i.e., the damping of elastically scattered Bloch waves) can be written[100]

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V(\underset{\sim}{r}) + iV'(\underset{\sim}{r})\right] B_j(\underset{\sim}{r}) = E_o B_j(\underset{\sim}{r}) \tag{4.16}$$

where $V'(\underset{\sim}{r})$ has lattice periodicity and is the interaction term which describes the effect of inelastic scattering on the elastically scattered wave. The damped Bloch waves are given by

$$B_j(\underset{\sim}{r}) = \sum_g C_g(j) \exp\{2\pi i[\underset{\sim}{k}(j) + i\underset{\sim}{k}'(j) + \underset{\sim}{g}]\cdot\underset{\sim}{r}\} \tag{4.17}$$

where the imaginary part of the wave vector is an amplitude absorption coefficient and the $C_g(j)$ are always complex. In the Laue diffraction geometry the direction of current flow is along the z axis; hence the inelastic scattering is peaked in this direction and we can write for Eq. (4.17):

$$B_j(\underset{\sim}{r}) = \sum_g \exp[-2\, k_z'(j)z]\, C_g(j) \exp[2\pi i\, \underset{\sim}{k}_g(j)\cdot\underset{\sim}{r}] \tag{4.18}$$

The unknowns of Eq. (4.18) are the $k_z'(j)$ and the complex $C_g(j)$. These quantities can be calculated using perturbation theory[101]. In particular, if there is no multiplicity in $\underset{\sim}{k}(j)$ we can write,

$$k_z'(j) = \frac{1}{2K} \sum_g \sum_h C_g^*(j)\, C_h^*(j)\, U'_{g-h} \tag{4.19}$$

Finally, if $V'(\underset{\sim}{r}) << V(\underset{\sim}{r})$ we can neglect the imaginary part of $C_g(j)$.

4.2.3.3 Computation scheme

In the Bloch formalism, the determination of the various amplitude functions, $\phi_g(z)$, reduces to a matrix inversion problem. Since absorption may be treated as a perturbation, the common procedure is to solve the eigenvalue equation (see Eq. (4.8)) by first neglecting inelastic scattering effects. Altorithms for finding the eigenvectors and eigenvalues of a hermitian matrix are very efficient in comparison to the general non-hermitian case. One then obtains the imaginary part of the complex eigenvalue from Eq. (4.19). The various beams are finally computed at the desired depth z by inserting the computed values for eigenvalues and eigenvectors into Eq. (4.12).

4.2.3.4 Lattice image formation

The wave function at the point $\underset{\sim}{r}_o$ on the exit surface of a crystal of thickness t_o may be written as

$$\psi(\underset{\sim}{r}_o) = \sum_g \phi_g(t_o) \exp(2\pi i \, \underset{\sim}{k}_g \cdot \underset{\sim}{r}_o) \tag{4.20}$$

In lattice imaging, one selects appropriate diffracted beams, ϕ_g, with an objective aperture and lets them recombine at the image plane to form an interference pattern. The wave function, $\psi_i(\underset{\sim}{r}_i)$, in the image plane is related to $\psi(\underset{\sim}{r}_o)$ by the instrument point spread function, or equivalently by the amplitude transfer function $\exp[-i\gamma(\theta,\phi)]/M$, where M is the magnification and $\gamma(\theta,\phi)$ arises from the lens aberrations and focal conditions (see Eq. (1.59)). If one designates by $\zeta_g(t_o)$ the phase angle of the g^{th} Darwin coefficient (i.e.,

$\phi_g(t_o) = |\phi_g(t_o)|\exp[i\zeta_g(t_o)]$, the image wave in reduced coordinates is

$$\psi(\underset{\sim}{r}_i) = \frac{1}{M}\exp(2\pi i \underset{\sim}{K}\cdot\underset{\sim}{r}_i)\sum_g |\phi_g(t_o)|\exp\{i[2\pi \underset{\sim}{g}\cdot\underset{\sim}{r}_i + \zeta_g(t_o) - \gamma_g]\} \tag{4.21}$$

where γ_g is the aberration phase shift suffered by the g^{th} reflection.

If only the transmitted beam and one diffracted beam (i.e., $g = 0$ and $g = 1$) are allowed to recombine at the image plane, then the intensity distribution in the image is given by

$$|\psi(x_i)|^2 = \frac{1}{M^2}\{|\phi_o|^2 + |\phi_1|^2 + 2|\phi_o||\phi_1|\cos[2\pi g\, x_i + \zeta_1 - \zeta_o - \gamma_o + \gamma_1] \tag{4.22}$$

where the coordinate x_i is chosen parallel to g. The relative phase $(\zeta_1-\zeta_o)$ of the beams is a function of both the foil thickness and the deviation parameter; this accounts for the observed fringe bending in the images of both wedge shaped and bent crystals.

The computer enhancement of experimental fringe patterns revealed local deviations from perfect periodicity, which are now seen to stem from spatial variations of the relative phases of the beams (i.e., $\zeta_g - \zeta_h$) due to a buckling of the specimen.

To extend this analysis to deformed crystals, the amplitudes $\phi_g(t_o)$ must be determined from the dynamical theory of electron diffraction from imperfect crystals. Recent calculations have shown that the fringe images are sensitive to both focus and variations in the diffraction geometry[98]. Consequently, lattice imaging seems to be a powerful tool for investigating dislocation structure since the fringes

display good contrast where the fringe bending is large. The validity of mapping the lattice distribution onto the observed fringe distribution must be questioned, since diffraction effects can produce fringe artifacts.

The weak-beam method of electron microscopy has proven to be a powerful technique for investigating lattice defects[102,103]. In the next section a systematic analysis of a complex defect geometry is carried out with this observation technique. By enhancing the weak-beam images, a quantitative comparison with theoretical calculations can be achieved. This example will be used to illustrate the problems encountered in the interpretation of high-resolution electron micrographs from crystals containing defects.

4.3 Analysis of Weak-Beam Images of a Dissociated Dislocation Dipole

4.3.1 The Theory of the Weak-Beam Method

4.3.1.1 Introduction

Although the underlying principle of the weak-beam imaging technique is rather simple, a detailed contrast analysis requires a critical evaluation of the various approximations generally assumed in the formalism. In this technique an increase in the resolution of closely spaced defects is obtained from a narrowing of the individual defect image widths. This narrowing of the image width also permits a more accurate determination of the defect position. Schematically, this effect arises because weakly excited beams are brought into Bragg condition by large local tilts of the atomic planes near the defect. The relationship between the defect image and the defect

geometry depends on the localization of the strain field, the diffraction geometry, and structural parameters such as the foil thickness and the depth of the defect[103]. In this work a dissociated dipole was selected for study. This defect was chosen for two reasons. Firstly, this defect had not been analyzed previously, and therefore we hoped to reach a quantitative agreement between image simulations and computer enhanced experimental data. Secondly, such an experiment could test the applicability of this method for investigating the atomic structure of defects.

4.3.1.2 Dynamical theory of high-energy-electron diffraction from imperfect crystals

The potential field of an imperfect crystal may be written as a modified Fourier series of the form

$$V(\underset{\sim}{r}) = \frac{h^2}{2m_e} \sum_g U_g(\underset{\sim}{r}) \exp(2\pi i \underset{\sim}{g}\cdot\underset{\sim}{r}) \qquad (4.23)$$

where the potential coefficient $U_g(\underset{\sim}{r})$ account for the quasi-periodic nature of the lattice. In the deformable ion approximation, the Fourier coefficients of potential can be written[96]

$$U_g(\underset{\sim}{r}) = U_g \exp(-2\pi i \underset{\sim}{g}\cdot\underset{\sim}{R}) \qquad (4.24)$$

where $\underset{\sim}{R}$ is a vector function that describes the local atomic displacements in the imperfect crystal.

By substituting the Darwin wave function, Eq. (4.14), and the imperfect crystal potential into the time-independent Schrödinger equation, we obtain[102]

$$\sum \{\frac{-i}{4\pi(K_z+g_z)} \nabla^2\phi_g(\underset{\sim}{r}) + \frac{\partial\phi_g(\underset{\sim}{r})}{\partial z} + \frac{K_x+g_x}{K_z+g_z}\frac{\partial\phi_g(\underset{\sim}{r})}{\partial x} + \frac{K_y+g_y}{K_z+g_z}\frac{\partial\phi_g(\underset{\sim}{r})}{\partial y}$$

$$- i\pi\left(\frac{\chi^2 - |\underset{\sim}{K}+\underset{\sim}{g}|^2}{K_z+g_z}\right)\phi_g(\underset{\sim}{r}) - i\pi \sum_h \frac{U_{g-h}\phi_h(\underset{\sim}{r})}{K_z+g_z} \exp[-2\pi i(\underset{\sim}{g}-\underset{\sim}{h})\cdot\underset{\sim}{R}]\}$$

$$\cdot \exp[2\pi i(\underset{\sim}{K}+\underset{\sim}{g})\cdot\underset{\sim}{r}] = 0 \qquad (4.25)$$

If we now assume that $\phi_g(\underset{\sim}{r})$ and $U_g(\underset{\sim}{r})$ are slowly varying over the dimensions of a unit cell, such that the Fourier transform of the product term $U_{g-h}(\underset{\sim}{r})\phi_g(\underset{\sim}{r})$ has support in the first Brillouin zone, we may separate Eq. (4.25) into a set of algebraic equations. This orthogonality condition breaks down when the deformation of the crystal varies very rapidly with position[102].

We shall further restrict our discussion to diffraction from crystals with only x-z dependence: this condition is met experimentally by tilting the specimen so that only one row of systematic reflections is excited. The $\underset{\sim}{g}$ vectors are assumed to lie along the x direction of the x-y plane (i.e., the zeroth Laue zone) in reciprocal space. In these restrictions the separated equations become

$$\frac{-i}{4\pi K_z}\left(\frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial x^2}\right)\phi_g(\underset{\sim}{r}) + \frac{\partial\phi_g(\underset{\sim}{r})}{\partial z} + \frac{K_x+g}{K_z}\frac{\partial\phi_g(\underset{\sim}{r})}{\partial x} - i\pi\frac{K^2 - |\underset{\sim}{K}+\underset{\sim}{g}|^2}{K_z}\phi_g(\underset{\sim}{r})$$

$$\frac{-i\pi}{K_z}\sum_{h\neq g} U_{g-h}\,\phi_h(\underset{\sim}{r}) \exp[-2\pi i(\underset{\sim}{g}-\underset{\sim}{h})\cdot\underset{\sim}{R}] = 0 \qquad (4.26)$$

Since the z direction is approximately the direction of current flow in the crystal, we expect that the variations in the $\phi_g(x,z)$ can become rapid along the x axis but not in the direction of the current flow, so

that we may neglect the term involving $\partial^2\phi_g/\partial z^2$ in Eq. (4.26). Using this approximation and the following identities:

$$K_z = K \cos\theta_o \simeq K$$

$$K^2 - |\underset{\sim}{K} + \underset{\sim}{g}|^2 \simeq 2K\, s_g \tag{4.27}$$

which were obtained using the geometry of the dispersion surface shown in Fig. 4-7 , Eq. (4.26) becomes the following second-order parabolic system

$$[-i\underline{\underline{C}}\frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial z} + \underline{\underline{B}}\frac{\partial}{\partial x}]\,\underline{\Phi}(x,z) = 2\pi i\,\underline{\underline{A}}(x,z)\,\underline{\Phi}(x,z) \tag{4.28}$$

where the various matrices are defined as

$$\underline{\Phi}(x,z) = [\phi_{g(1)}(x,z),\cdots,\phi_{g(\hat{N})}(x,z)]^\dagger$$

$$[A]_{gh} = \begin{cases} s_g & \text{for } g = h \\ U_{g-h}\exp[-2\pi i(\underset{\sim}{g}-\underset{\sim}{h})\cdot\underset{\sim}{R}]/2K & \text{for } g \neq h \end{cases}$$

$$[B]_{gh} = [K_x + g]\,\delta_{g,h}/K$$

$$[C]_{gh} = \delta_{g,h}/4\pi K$$

Ordinarily one drops the second-order derivative in this last equation to yield a first-order hyperbolic system. This is a reasonable approximation because the elements of $\underline{\Phi}$ are normally slowly varying for near-perfect crystals and the elements of $\underline{\underline{C}}$ are small in high-energy-electron diffraction. However, it has been reported that these second-order terms are of importance when the displacement field

$\underset{\sim}{R}(x,z)$ exhibits rapid x-dependent variations over a distance of a few angstroms[104] (higher order reflections are more sensitive to this perturbation.) We therefore retained these terms in our calculations of weak-beam images. The numerical scheme used to solve Eq. (4.28) is given in Appendix F.

We can gain valuable information about the physical behavior of the solutions to Eq. (4.28) by neglecting the Laplacian term. One can then transform the partial differential equation into an ordinary differential equation by introducing the directional derivative operators[99]

$$\partial/\partial n_g = \partial/\partial z + [(K_x + g)/K_z]\, \partial/\partial x \tag{4.29}$$

which differentiate the variables $\phi_g(\underset{\sim}{r})$ along the $\hat{N}$ characteristic directions. Equation (4.28) is then converted into

$$\underline{\underline{n}}\, \underline{\Phi}(x,z) = 2\pi i\, \underline{\underline{A}}(x,z)\, \underline{\Phi}(x,z) \tag{4.30}$$

where $\underline{\underline{n}}$ is a diagonal matrix with elements $[n]_{g,h} = (\partial/\partial n_g)\delta_{g,h}$. Given the solution $\underline{\Phi}(0)$ at the entrance surface of the crystal, the solution at any point of the exit surface can be constructed by iterated integration of Eq. (4.30) on a triangular mesh defined by the $\hat{N}$ characteristics. As Takagi has pointed out,[105] these characteristics provide a theoretical basis for the column approximation in which one neglects all x derivatives in Eq. (4.28). This assumption amounts to ignoring the triangular area sampled by the transmitted and diffracted beams: for a crystal of thickness t_o and for two-beam diffraction, the base of the triangular mesh is $t_o\theta_g$. It is therefore

invalid when calculating image intensities for higher-order reflections in thick crystals, particularly near regions of strong localized disorder such as a dislocation core.

Despite the stringent requirements imposed by the column approximation, it has been consistently used in the analysis of weak-beam images[103]. Certain criteria were subsequently derived for optimizing the resolution of weak-beam images. In this approximation scattering processes can be described by

$$\frac{d}{dz}\,\underset{\sim}{\Phi}(x,z) = 2\pi i\,\underline{\underline{A}}(x,z)\,\underset{\sim}{\Phi}(x,z) \tag{4.31}$$

If we now make the phase transformation

$$\hat{\phi}_g(\underset{\sim}{r}) = \phi_g(\underset{\sim}{r})\,\exp[2\pi i \underset{\sim}{g}\cdot\underset{\sim}{R}]$$

we can express Eq. (4.31) as

$$\frac{d}{dz}\,\hat{\underset{\sim}{\Phi}}(x,z) = 2\pi i\,\hat{\underline{\underline{A}}}(x,z)\,\hat{\underset{\sim}{\Phi}}(x,z) \tag{4.32}$$

where

$$[\hat{A}]_{g,h} = \begin{cases} s_g + \dfrac{d}{dz}[\underset{\sim}{g}\cdot\underset{\sim}{R}] & \text{for } g = h \\ U_{g-h} & \text{for } g \neq h \end{cases}$$

Equation (4.32) is the most straightforward of the column approximation equations in terms of computational ease, since all that must be done is to locally vary the diagonal elements of the $\hat{\underline{\underline{A}}}$ matrix as one numerically integrates along the z direction.

The various diffraction approximations are most easily discussed in terms of a perturbation treatment of high-energy-electron

diffraction from imperfect crystals[99]. The wave function in this theory is written in a modified Bloch wave representation:

$$\psi(\underset{\sim}{r}) = \sum_j \Theta_j(\underset{\sim}{r})\ B_j(\underset{\sim}{r}) \tag{4.33}$$

The excitation coefficients are changed from the constants of the perfect crystal case into slowly varying functions of position, $\Theta_j(\underset{\sim}{r})$, in the imperfect crystal. At the entrance surface of the foil perfect crystal Bloch states are excited, and as these Bloch waves propagate through the crystal, scattering transitions to new Bloch states occur. These transitions are controlled by the magnitude of the perturbing potential, $V_d(\underset{\sim}{r})$, which is the difference between the perfect and the imperfect crystal potentials. An insight into the resulting image contrast is obtained by looking at the transitions from one state to another. The transition rate is controlled by the sum $\sum_j (B_j, V_d B_j)$, which bears a close relation to the Fourier transform of $V_d(\underset{\sim}{r})$. Therefore defects with highly localized strain fields will induce a large lateral redistribution of the excited Bloch states, thus rendering the column approximation invalid. The Bloch formalism offers a physical picture of how the defect redistributes the excited states. It also allows a numerical estimate of the transition probability amplitudes for the column approximation. Unfortunately, the column approximation can only be used reliably in regions where the strain field is weak. However, it has proved to be useful in suggesting contrast experiments, such as studying defects with weak beams[106].

4.3.1.3 The weak-beam technique

It has been found experimentally that beams which are weak in regions of perfect crystal because of their large deviation parameters could attain an appreciable magnitude near regions where a strain field is large. In the vicinity of dislocation lines, for instance, the weak-beam intensity maxima have a very narrow halfwidth (≃15A). This technique thus allows an accurate determination of the dislocation position and of the strain field interactions.

Several weak-beam diffraction geometries may be selected; the optimal geometry is a function of both the defect geometry and the required resolution. For example, one may excite the low-order reflection $\underset{\sim}{g}$ and form the image with the weak-systematic $-\underset{\sim}{g}$, or excite a higher-order reflection like $3\underset{\sim}{g}$ and use the weak-beam $\underset{\sim}{g}$, or finally tilt the crystal so that no zeroth-order Laue reflections are strongly excited and image with $\underset{\sim}{g}$. This latter situation will be analyzed since it lends itself to a simple analysis of the potential of this imaging method.

We can predict the shapes and positions of weak-beam peaks.using Eq. (4.32) where only two beams are considered. Then the relationship between the amplitudes ϕ_g of the diffracted beam and ϕ_o of the transmitted beam is

$$\frac{d\phi_o}{dz} = \frac{\pi i}{\xi_g}\phi_g$$

$$\frac{d\phi_g}{dz} = \frac{\pi i}{\xi_g}\phi_o + 2\pi i[s_g + \frac{d}{dz}(\underset{\sim}{g}\cdot\underset{\sim}{R})] \qquad (4.34)$$

where $\xi_g = K/U_g$ is the extinction distance. The coupling between ϕ_o and ϕ_g will be greatest when

$$s_g + \frac{d}{dz}(\underset{\sim}{g}\cdot\underset{\sim}{R}) = 0 \qquad (4.35)$$

which is the condition that the local displacement $\underset{\sim}{R}$ twists the atomic planes into the Bragg condition. Consequently, one expects a weak-beam peak at a turning point of $dR(z)/dz$. The peaks have narrow halfwidth, their position is relatively thickness independent and lies close to the position predicted by Eq. (4.35). The above criterion yields the displacement between the observed image peak and the crystal defect: its accuracy has been confirmed by many-beam calculations for simple defect geometries[106].

The image halfwidth Δx can be estimated from the kinematical theory in the limit $|s_g| \to \infty$: for a single edge dislocation an approximate value for the image halfwidth is[96]

$$\Delta x \simeq \frac{0.28}{|s_g|}\left[1 + \frac{1}{2(1-\nu)}\right] \qquad (4.36)$$

where ν is Poisson's ratio. For a weak-beam image of an edge dislocation, in which detail on the order of 25Å is desired, a minimum value of $|s_g|$ is 0.02Å^{-1}. The advantages of operating with $|s_g|$ as large as possible are twofold. Firstly, one can make a reasonably accurate determination of the dislocation position without knowing the dislocation depth and the foil thickness. Secondly, this diffraction geometry ensures that the weak beam only interacts strongly with the primary beam in regions of large strain, resulting in an image

which gives a qualitative picture of the defect geometry.

If the Ewald sphere passes close to any reciprocal point, the image becomes dependent on the defect depth and the foil thickness. Our experimental study indicates that such effects are advantageous when the weak-beam method is extended to studies of dislocation detail approaching atomic dimensions. Dynamical scattering from the weak beam to the strongly excited waves are then equally important in determining the contrast, which is now thickness dependent. When defect detail below 25Å is to be investigated, the column approximation cannot be invoked since the variations in the displacement field $\underset{\sim}{R}$ are not negligible over a distance $t_o\theta_g$. Our calculations indicate that this requirement becomes all the more stringent as one excites higher order reflections. In the following study, we shall attempt to match the positions as well as the magnitudes of the weak-beam peaks to computer simulated profiles; the difficulties encountered in interpreting the observed contrast at this high-resolution range, and the inadequacies of the existing theories to predict the contrast behavior, will be discussed.

4.3.2 Geometrical Analysis of a Weak-Beam Image

4.3.2.1 Experimental procedure

The specimen used in this experiment was a high-purity germanium single crystal (specific resistivity to 6Ω-cm) doped with arsenic to a concentration of 8×10^{14} atoms/cm^3. The single crystal was deformed in compression for 75 sec at 600°C, by a static load of 1.5 kg/mm^2 along the <123> axis. The crystal was air cooled, and then sliced parallel to the (111) glide plane. Electron microscope specimens, 3 mm

in diameter were then ultrasonically cut from these (111) slices. Electron transparent specimens were prepared by: chemically polishing the 3 mm disks in a rotating bath of 90% HNO_3 -10% HF, and then final polishing with an ion mill using 6 keV argon ions.

The microscopy was performed with a 200 keV JEOL electron microscope operating at a magnification of 88,500X. In order to photographically record the weak-beam images, exposure times of 10 to 20 sec were required. Electron micrographs of a dislocation dipole were recorded at the diffracting conditions listed in Table 4-1. These diffracting conditions were selected in order to identify the dipole and study the sensitivity of the weak-beam method. The depth of the dipole was determined by stereoscopy.

4.3.2.2 Determination of the dislocation dipole geometry

A dislocation dipole can be identified by first taking a two-beam bright-field image using $+\underset{\sim}{g}$ and then $-\underset{\sim}{g}$, holding s_g constant in both cases. If we suspect that our image represents a dipole, then for one $\underset{\sim}{g}$ the dislocation images will be separated and for the other $\underset{\sim}{g}$ they will collapse (see Fig. 4-8a,b), since the bright-field image is strongly antisymmetric about a dislocation line. The dipole segment AB will be the subject of the remaining analysis.

In order to determine the Burgers vectors of the dipole segment AB we must correlate the observed dark-field images with their respective reflections. For a general dislocation, $\underset{\sim}{R}$ can be written according to isotropic elasticity theory, as[107]

Table 4-1

Summary of the diffraction geometries for which the micrographs of the dipole in germanium were recorded.

Figure	Beam used to form the images	Number of systematic beams	Intersection of the row of systematics with the Ewald sphere in units of $\underset{\sim}{g}$	Deviation parameter of the imaging beam ($\underset{\sim}{g}\cdot\underset{\sim}{s}$) ($\AA^{-1}$)	Dipole contrast	Observations
4-8a	[000]	2	$\simeq$ -1.1		outside	Existence of a dislocation dipole
4-8b	[000]	2	$\simeq$ 1.1		inside	
4-9a	[$\bar{2}$02]	6	$\gtrsim$ 3.0	0.0063	inside	Intrinsic stacking fault
4-9b	[0$\bar{2}\bar{2}$]	6	$\gtrsim$ 3.0	0.0063	outside	
4-10	[$\bar{2}$20]	6	3.016	0.0063	inside	Dissociation into partials
4-12	[$\bar{2}$20]	2	$\simeq$ 1.0	0		Interstitial type
4-22a	[$\bar{2}$20]	6	$\gtrsim$ 3.0		inside	
4-22b	[$\bar{2}$20]	6	3.165	0.0068	inside	
4-22c	[$\bar{2}$20]	6	2.89	0.0059	inside	
4-22d	[$\bar{4}$40]	6	$\simeq$ 3.0	0.0063	inside	
4-23a	[$\bar{2}$20]	4	$\simeq$ -1.1	-0.0066	outside	
4-23b	[$\bar{4}$40]	4	$\lesssim$ -1.0	-0.0063	outside	
4-23c	[$\bar{2}$20]	4	-1.127	-0.0067	outside	

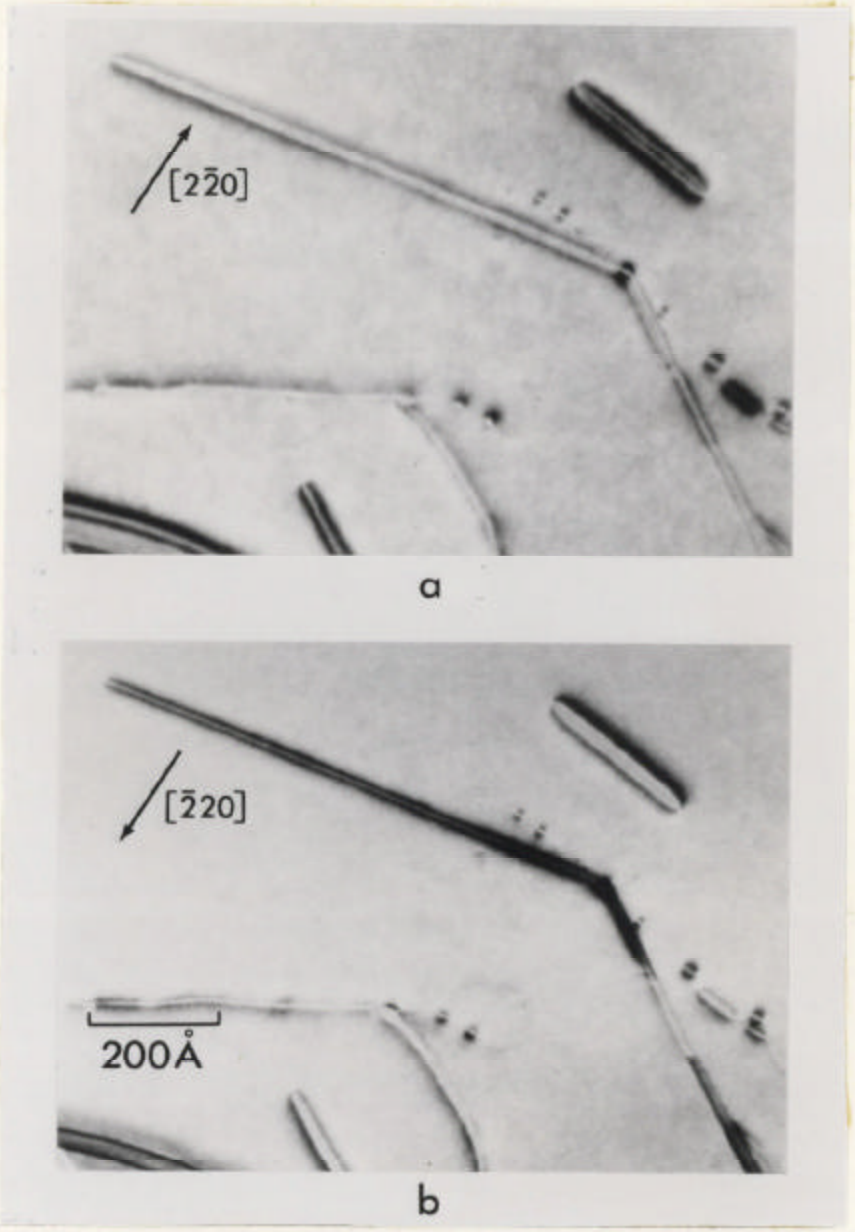


Fig. 4-8 Two-beam bright-field images of dislocations in germanium: (a) In this picture the operating reflection is $\underset{\sim}{g} = [2\bar{2}0]$ with $s_g \simeq 0$. Note that the large dipole exhibits outside contrast. (b) In this picture the operating reflection is $\underset{\sim}{g} = [\bar{2}20]$ with $s_g \simeq 0$. Here the dipole exhibits inside contrast. The dipole segment AB will be the subject of the remaining analysis.

$$\underset{\sim}{R}(\underset{\sim}{r}) = \frac{1}{2\pi}\left\{\underset{\sim}{b}\theta + \underset{\sim}{b}_e \frac{\sin 2\theta}{4(1-\nu)} + \underset{\sim}{b} \times \underset{\sim}{u}\left[\frac{1-2\nu}{2(1-\nu)}\right]\log|\underset{\sim}{r}| + \frac{\cos 2\theta}{4(1-\nu)}\right] \qquad (4.37)$$

where $\underset{\sim}{b}_e$ is the Burgers vector of the edge component of the dislocation, θ is an angular coordinate in the plane normal to the dislocation line measured from the slip plane, and $\underset{\sim}{u}$ is a unit vector along the dislocation line. In Eq. (4.37) the vector $\underset{\sim}{b} \times \underset{\sim}{u}$ is approximately parallel to the incident beam (the incident beam is approximately parallel to $\langle 111 \rangle$ and the glide dislocations have $\frac{1}{2}\langle 110 \rangle$ Burgers vectors). In this case the contrast term $\underset{\sim}{g}\cdot\underset{\sim}{R}$ of Eq. (4.34) is approximately proportional to $\underset{\sim}{g}\cdot\underset{\sim}{b}$. The Burgers vectors can thus be determined by finding the reflection for which the dark-field image of the dipole disappears. This invisibility criterion is illustrated in Fig. 4-9a where the operating reflection $(\bar{2}02)$ causes only the segment AB of the dipole to vanish. This contrast behavior provides experimental evidence that the dipole is dissociated into Shockley partial dislocation with Burgers vectors of the type $\frac{1}{2}\langle 112 \rangle$.

According to Eq. (4.35) the image of a dislocation line lies on the side of the dislocation where $\underset{\sim}{g}\cdot\underset{\sim}{R}$ is of opposite sign to s_g. Consequently, by looking at the inside-outside contrast of the dipole OA in Fig. 4-8 it was found that the unit vector of the left dislocation is $\underset{\sim}{u}_1 = [10\bar{1}]/\sqrt{2}$, and the unit vector of the right dislocation is $\underset{\sim}{u}_2 = [\bar{1}01]/\sqrt{2}$.

In Fig. 4-9a we see that the dipole segment OA is not pure edge since it does not vanish when imaged with the $(\bar{2}02)$ reflection. The Burgers vectors for this dipole segment can thus be oriented at integer multiples of 30^o with respect to the dislocation line. By

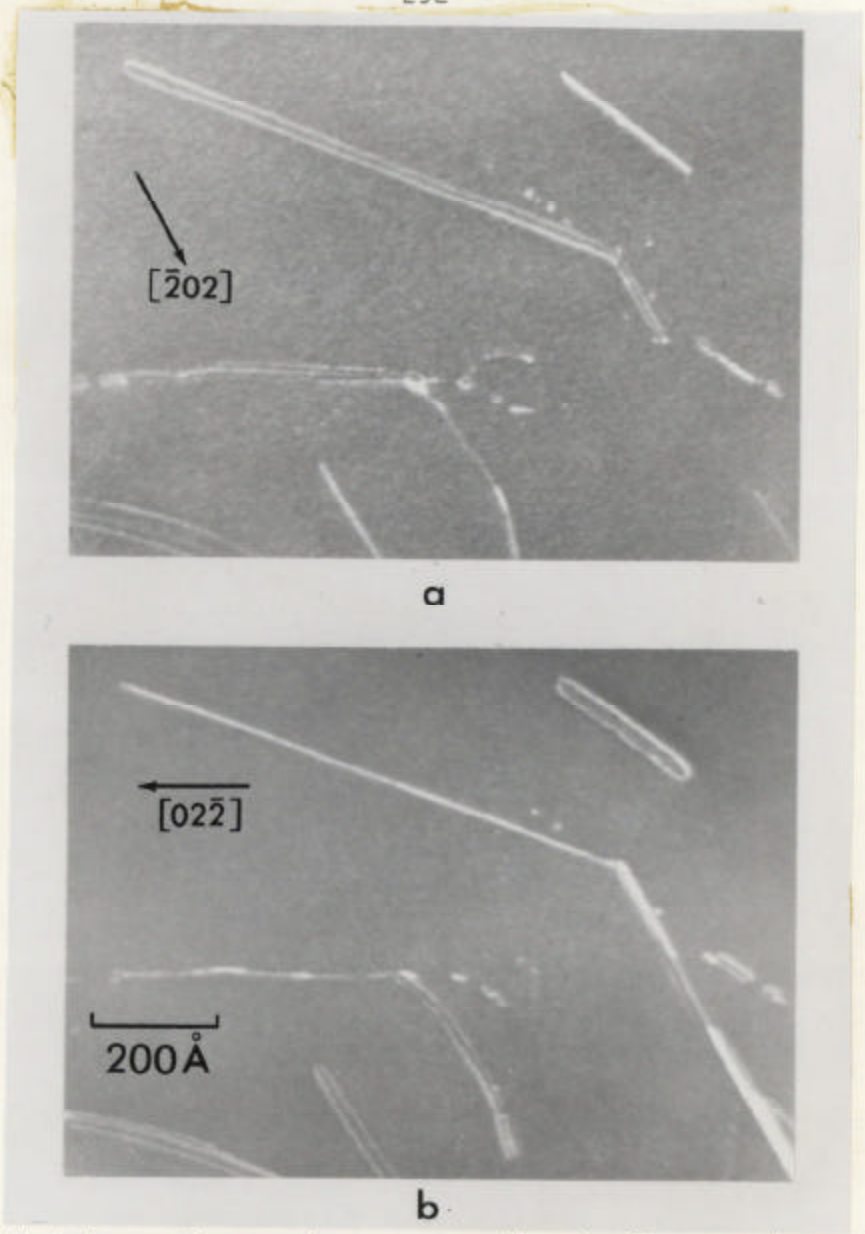


Fig. 4-9 Weak-beam micrographs corresponding to the case where different rows of systematics are excited. (a) The reflection [$\bar{2}$02] extinguishes the partial dislocations whereas the undissociated segment of the dipole is still visible. This phenomenon allows one to determine the Burgers vectors of the partials. (b) The reflection [02$\bar{2}$] resolves the partials although the two innermost peaks are almost overlapping which suggests that $\underset{\sim}{g}\cdot\underset{\sim}{b}_i^p$ for the partials take different values.

inspecting the weak-beam images shown in Figs. 4-9 and 4-10 where the $(\bar{2}02)$, $(02\bar{2})$ and $(\bar{2}20)$ reflections are successively operating, the non-vanishing contrast of the undissociated dipole segment OA reduces the possible orientation angles for $(\underset{\sim}{b},\underset{\sim}{u})$ to 60 and 120 degrees. If one now considers the decomposition of $\underset{\sim}{b}$ into partials, $\underset{\sim}{b}^p$, and notices that the $(\bar{2}20)$ reflection resolves the dipole splitting with higher contrast than the $(02\bar{2})$ reflection, an analysis of the products $\underset{\sim}{g}\cdot\underset{\sim}{b}^p$ (see Table 4-2) is consistent with a dissociated 120^o dipole, for which

$$\begin{aligned} \underset{\sim}{b}^{(1)} &= \underset{\sim}{b}^p_1 + \underset{\sim}{b}^p_2 = \frac{1}{2}[\bar{1}10] = \frac{1}{6}[\bar{1}2\bar{1}] + \frac{1}{6}[\bar{2}11] \\ \underset{\sim}{b}^{(2)} &= \underset{\sim}{b}^p_3 + \underset{\sim}{b}^p_4 = \frac{1}{2}[1\bar{1}0] = \frac{1}{6}[1\bar{2}1] + \frac{1}{6}[2\bar{1}\bar{1}] \end{aligned} \tag{4.38}$$

There seems to be some discrepancy between this model and the weak-beam image of Fig. 4-9a. Here the products $\underset{\sim}{g}\cdot\underset{\sim}{b}^p$ assume values of 0 and ± 1 and an empirical rule tells us that partials are effectively invisible for $|\underset{\sim}{g}\cdot\underset{\sim}{b}^p| \leq \frac{1}{3}$ when the $\underset{\sim}{g}\cdot\underset{\sim}{b}^p \times u$ term is unimportant[96]. However, the fact that the partial dislocations are completely extinguished by the $(\bar{2}02)$ beam is due to the vanishing of $\underset{\sim}{g}\times\underset{\sim}{u}$ and $\underset{\sim}{g}\cdot\underset{\sim}{b}^p_e$. The latter term is quite significant for the $(02\bar{2})$ case; since it accounts for the presence of the fringes seen in Fig. 4-9b.

The Burgers vectors of the partial dislocations can be determined if the nature of their associated stacking faults is known. The first results for silicon have been reported by Ray[108], who observed a large number of dislocation segments dissociated into Shockley

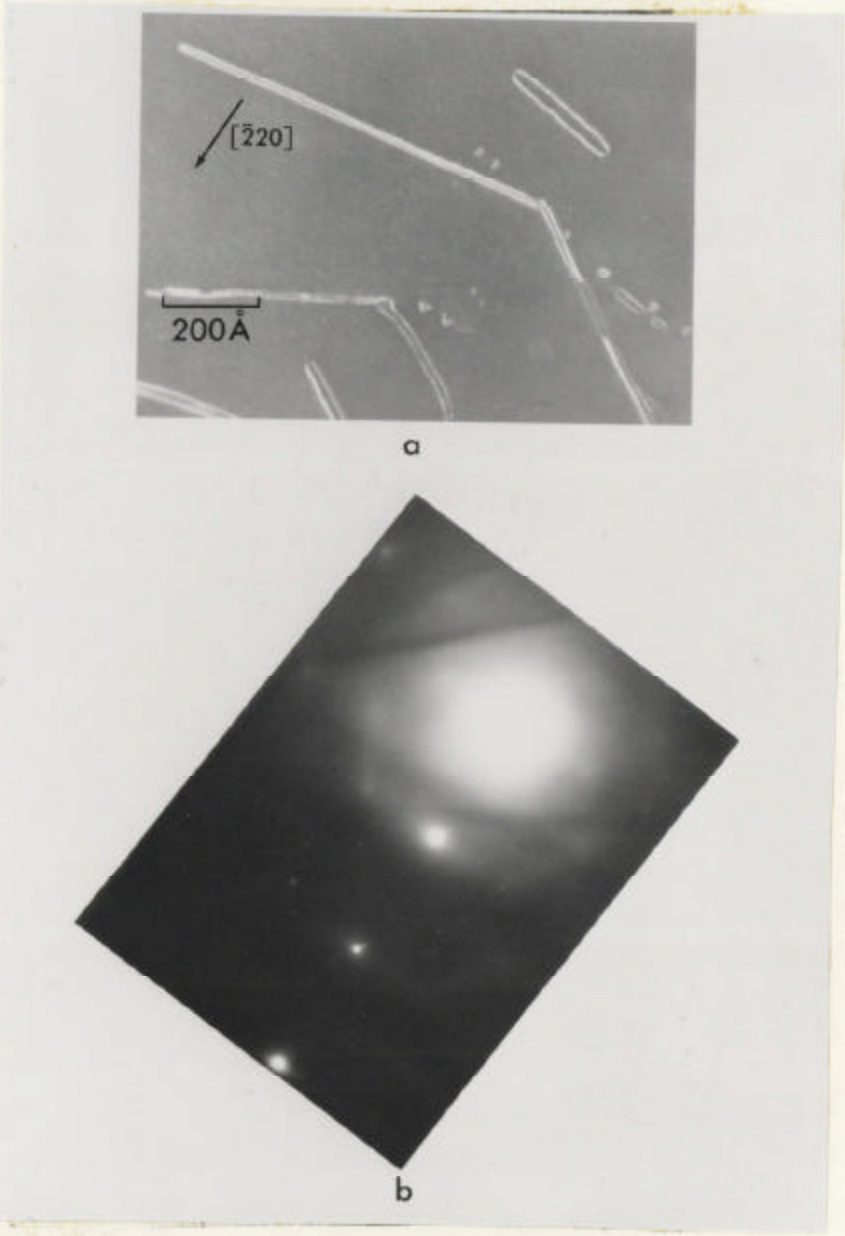


Fig. 4-10(a) Weak-beam image using $\underset{\sim}{g} = [\bar{2}20]$. Note that the segment AB of the dipole is dissociated into four partial dislocations which are well resolved on the background.
(b) Selected area diffraction pattern of (a).

Table 4-2

List of the edge components $\underset{\sim}{b}_e$, of the cross products $\underset{\sim}{b} \times \underset{\sim}{u}_1$, and of the products $\underset{\sim}{g} \cdot \underset{\sim}{b}$ for the undissociated Burgers vector $\underset{\sim}{b}^{(1)}$ and its two associated partial Burgers vectors. The $\underset{\sim}{g} \cdot \underset{\sim}{b}$ terms are calculated for all reflections used experimentally, both for the total $\underset{\sim}{b}$ and for its edge component $\underset{\sim}{b}_e$.

Burgers vector $\underset{\sim}{b}$	$\underset{\sim}{b}^{(1)} = \frac{1}{2}[\bar{1}10]$	$\underset{\sim}{b}_1^p = \frac{1}{6}[\bar{1}2\bar{1}]$	$\underset{\sim}{b}_2^p = \frac{1}{6}[\bar{2}11]$
Edge component $\underset{\sim}{b}_e$	$\frac{1}{4}[\bar{1}2\bar{1}]$	$\frac{1}{6}[\bar{1}2\bar{1}]$	$\frac{1}{12}[\bar{1}2\bar{1}]$
$\underset{\sim}{b} \times \underset{\sim}{u}_1$	$-\frac{1}{2\sqrt{2}}[111]$	$-\frac{1}{3\sqrt{2}}[111]$	$-\frac{1}{6\sqrt{2}}[111]$
$\underset{\sim}{b} \cdot (\bar{2}20)$	2	1	1
$\underset{\sim}{b} \cdot (\bar{2}02)$	1	0	1
$\underset{\sim}{b} \cdot (0\bar{2}\bar{2})$	-1	-1	0
$\underset{\sim}{b}_e \cdot (\bar{2}20)$	$\frac{3}{2}$	1	$\frac{1}{3}$
$\underset{\sim}{b}_e \cdot (\bar{2}02)$	0	0	0
$\underset{\sim}{b}_e \cdot (0\bar{2}\bar{2})$	$-\frac{3}{2}$	-1	$-\frac{1}{2}$

partials bounding an intrinsic stacking fault. In this light we shall tentatively select the partial dislocation geometry to be that given in Fig. 4-11 [107]. The angle between the dislocation lines of the dipole was determined by stereoscopy.

The last question to be answered about the dipole geometry is whether it is a vacancy or an interstitial type. The dislocation dipole may be approximated to a prismatic loop[109], whose black-white image in dark-field can be used to determine whether it is an interstitial or vacancy loop. The Ashby - Brown criterion[110] predicts an interstitial defect if the black side of the black-white image of the defect is towards the direction of the imaged $\underset{\sim}{g}$ (as long as the defect lies at a depth larger than one-half extinction distance from either foil surface). Since stereoscopy indicated that the defect was close to the foil center, we conclude from Fig. 4-12 that the dipole sketched in Fig. 4-11 is of interstitial type.

4.3.2.3 Determination of the foil thickness

In order to make contrast calculations, the thickness of the specimen must be known. In this section we shall present a new method for determining the specimen thickness.

The intensities of diffracted beams can be calculated to a high accuracy if the Fourier coefficients of potential (corrected for absorption) are known. The calculation of accurate intensities also requires a precise knowledge of the experimental diffracting conditions. If an accurate calculation of the intensity of the diffracted beams can be made, then the specimen thickness can be estimated by

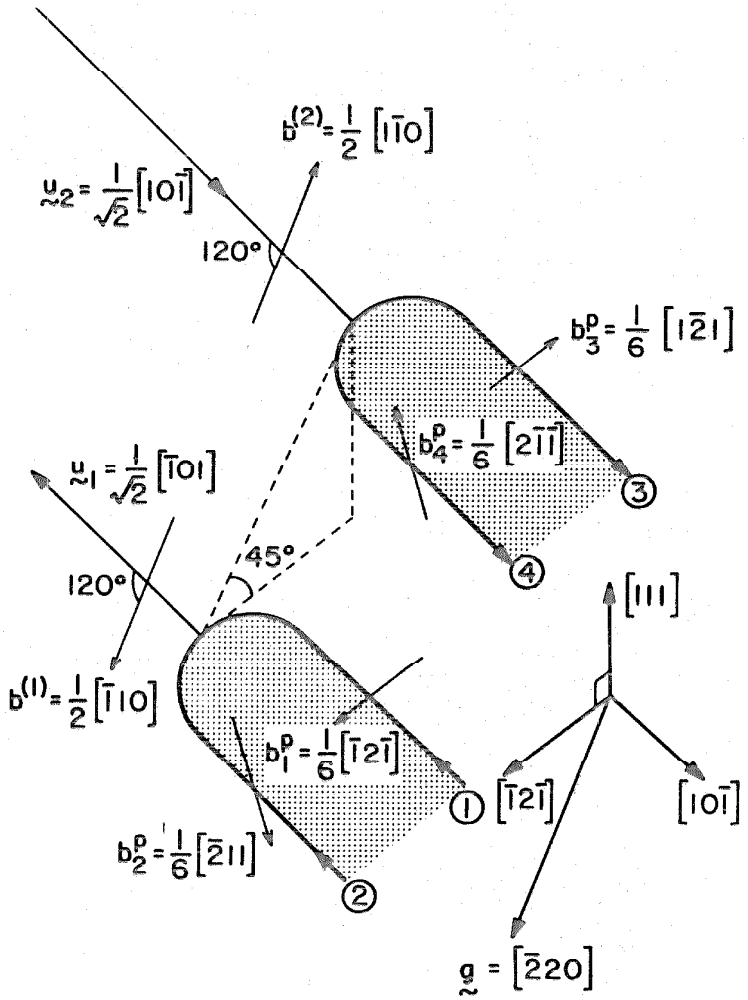


Fig. 4-11. Diagram showing the configuration of the dissociated interstitial dipole used in this work. The notation used to number the partial dislocations is consistent with Fig. 4-17, which shows the dislocation geometry in the $(\bar{1}\bar{1}2)$ plane.

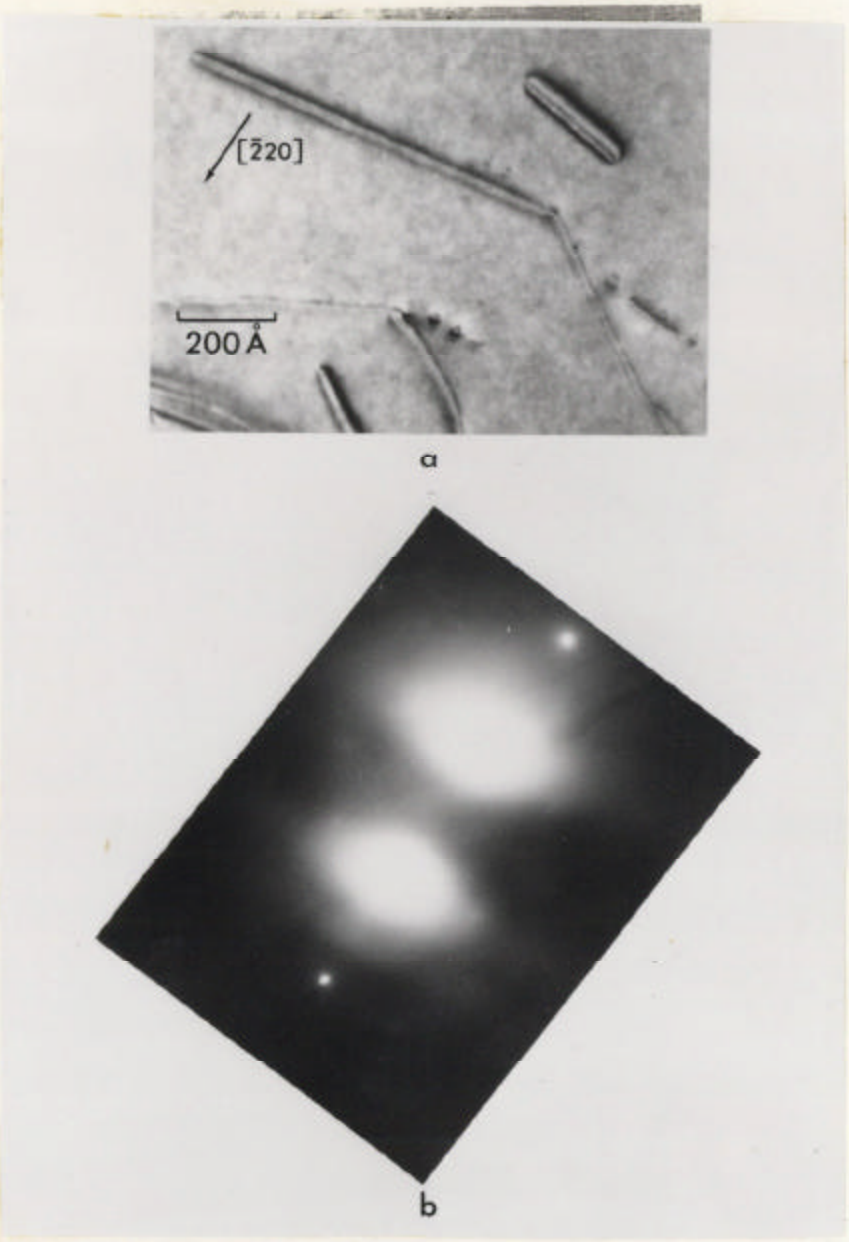


Fig. 4-12 (a) Dark-field micrograph using $\underset{\sim}{g} = [\bar{2}20]$
(b) Selected area diffraction pattern of (a)

comparing the thickness profiles to the intensities of each reflection in the diffraction pattern.

For example, consider the diffraction pattern shown in Fig. 4-10b. The thickness profile of each reflection in this pattern was calculated using Eq. (4.8). (The U_g were calculated using Eq. (4.4) for 200 keV electrons and a specimen temperature of $300^{o}K$. Values of U_g' were obtained from the theoretical estimates of Humphrey and Hirsch[111].) The resulting thickness profiles are shown in Fig. 4-13. The thickness of the specimen was obtained by comparing the thickness profiles of the g and 3g reflections. Since the intensity of these reflections is similar we conclude that their intensities are consistent with a thickness threshold of $t_o \gtrsim 3500\text{Å}$.

It is obvious that this method is very sensitive to the magnitude of the elements of the $\underline{\underline{A}}$ matrix. Therefore, we suggest a systematic study of such effects as specimen orientation and absorption on the thickness profiles.

4.3.3 Computer Enhancement of Weak-Beam Images

The appearance of well defined maxima in Fig. 4-10 is a consequence of a significant decrease in the dislocation image halfwidth. However, this reduction in image width is accompanied by a decrease in the signal-to-noise ratio, which renders a quantitative image-contrast analysis impractical. We shall now present the computer processing algorithm that was implemented on the weak-beam micrograph shown in Fig. 4-10, in order to enhance the fringe visibility[112].

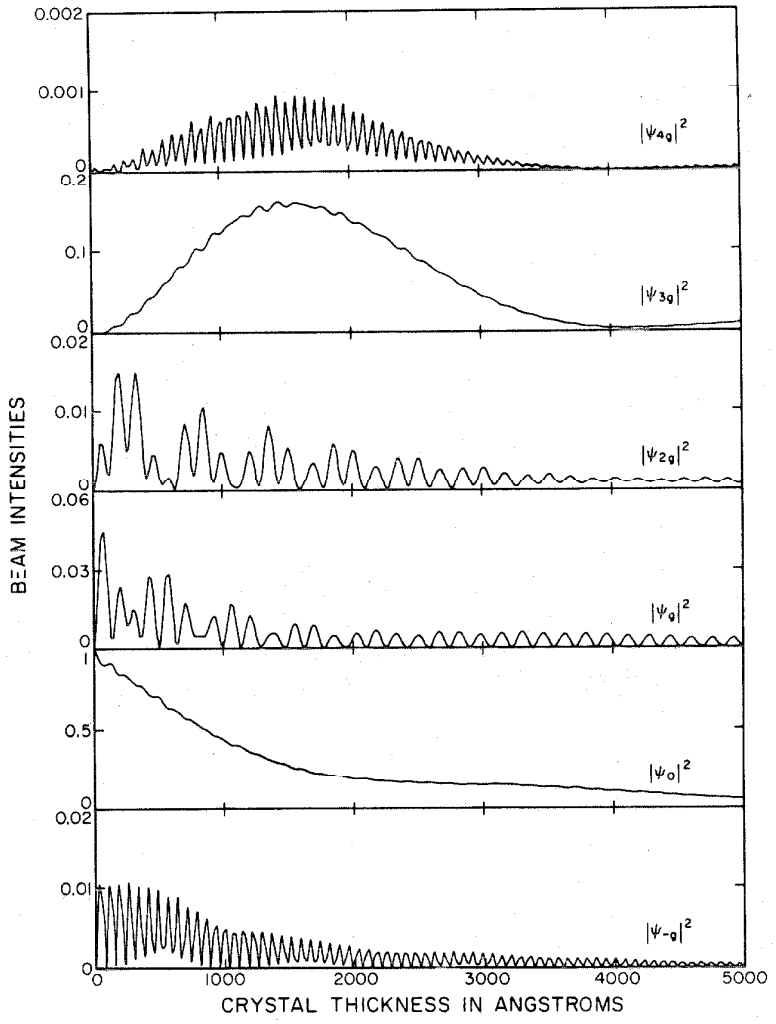


Fig. 4-13. Theoretical profiles of thickness fringes for a crystal of germanium. The calculations were made for 200 keV electrons and six operating reflections ($2\bar{2}0 \ldots \bar{8}80$; with the intersection of the Ewald sphere at $\simeq 3.016g$). Comparing the profiles for $|\psi_g|^2$ and $|\psi_{3g}|^2$ to Fig. 4-10b yields a lower bound of about 3500Å for the foil thickness.

4.3.3.1 Processing algorithm

Three weak-beam images of the same area were recorded in succession using the same electron microscope conditions. These micrographs were then copied onto fine-grain film, and digitized at a high sampling density (Δ = 25 μm). Figure 4-14 shows the rectangular area that was selected for further processing. The optical densities were then spread linearly over the full dynamic range by saturating both tails of the histogram to 1%. The main tasks of the processing algorithm were: preprocessing the images, spatial registration of the images, linear averaging of the images, and Fourier smoothing the optical density profiles.

The long exposure times required in weak-beam imaging are responsible for a contamination of the micrograph by islands of bad data (i.e., salt and pepper noise). An algorithm described by Eq. (2.33) replaces a datum if its local average exceeds a given threshold value δ. Since there is a preferential orientation of the image features in the vertical direction, this algorithm was implemented on a vertical box centered on each pixel in the image. Using a box containing 3 x 1 pixels we were able to eliminate most of the noise artifacts (typical values for δ were δ = 15 on a scale 0 to 255).

The translational and rotational registration of each picture was found by cross-correlating 113 x 254Å areas at various locations of the input images. During this process the CCF peak could only be determined to within one pixel because the image features exhibit an almost uniformly high contrast along the vertical direction. Recalling also that the CCF is an inference estimator only when the noise is

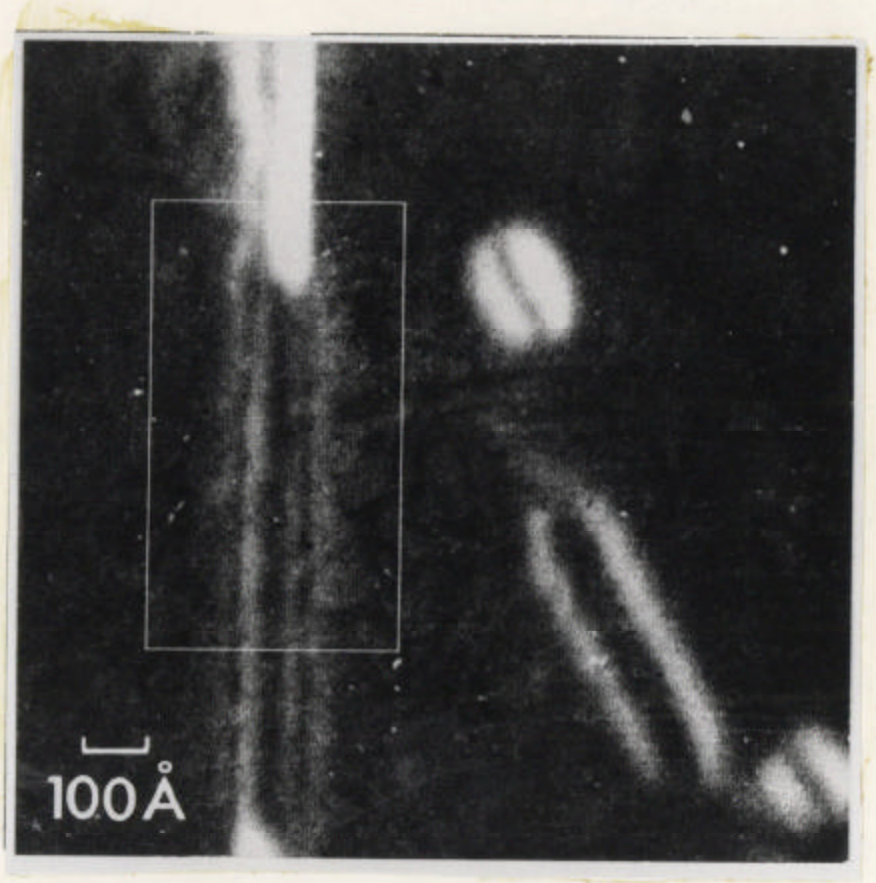


Fig. 4-14 Computer displayed section of the weak-beam image shown in Fig. 4-10a. The 762 x 365Å rectangular box that encloses a portion of the dissociated dipole was selected for further processing.

additive and uncorrelated to the signal; we must question the validity of this latter condition. It was mentioned in Sec. 4.1 that the noise is approximately independent of the signal only if the number of electrons that reach the image plane is sufficiently large. In weak-beam microscopy we might expect this condition to be only partially fulfilled (a fact which explains the difficulties we encountered during the matching process).

After achieving spatial registration, three pictures were superimposed and their average calculated. This averaging should attenuate the statistically random noise by a factor of about $\sqrt{3}$. An accurate estimate of the gain in the signal-to-noise ratio cannot be made, since the validity of our assumption that the noise is uncorrelated to the signal is questionable.

In order to better visualize the contrast enhancement obtained by this processing, optical density profiles were recorded and subsequently smoothed by a finite Fourier series expansion. (It can be shown that the best least-squares trigonometric approximation of order n to a periodic function is a Fourier series of order n). Discontinuities at both ends of the density profile induce overshoot effects in the Fourier transform which can be minimized by equalizing their density values. The technique we used amounted to rotating the profile about one of its ends, until both ends became equal. The discrete Fourier coefficients were then computed and the continuous trigonometric sum plotted for the corresponding coefficients. The implementation of the whole scheme is now presented.

4.3.3.2 Intensity profiles

A picture of the original series is shown in Fig. 4-15a after it has undergone contrast stretching. We notice the appearance of snow and the presence of light clusters spread randomly over the input array. Line profiles across this image were taken at five positions and are shown in Fig. 4-16a. The large fluctuations in intensity level obscure completely the general features of the image. We note that the visual clarity of the total image is due to the integrating property of human vision.

Figure 4-15b is the result of averaging three identical pictures by superposition, after a partial elimination of their snow pattern. In this figure we clearly see the four intensity maxima of the dissociated dipole. These maxima are strikingly displayed in the intensity profiles shown in Fig. 4-16b,c. We see that the relative spacing of each peak remains identical in contrast to sharp variations in the peak amplitudes. The spacing of the two innermost peaks is $2\Delta_1^{obs} = 65\mathring{A}$ and the spacing of the two outermost peaks is $2\Delta_2^{obs} = 155\mathring{A}$.

In order to further attenuate spurious high-frequency details, a low-pass box filter was applied to the average image (cf. Eq. (2.83)). The kernel matrix of this filter was ten pixels long and one pixel wide. The discontinuities at the boundaries of this filter produced the artifactual image shown in Fig. 4-15c (Note that vertical streaks are clearly visible). The corresponding image profiles shown in Fig. 4-16d do not exhibit any improvement in contrast with respect to the unfiltered image. We conclude then, that by using data averaging and majority logic techniques in conjunction with Fourier smoothing, we have obtained

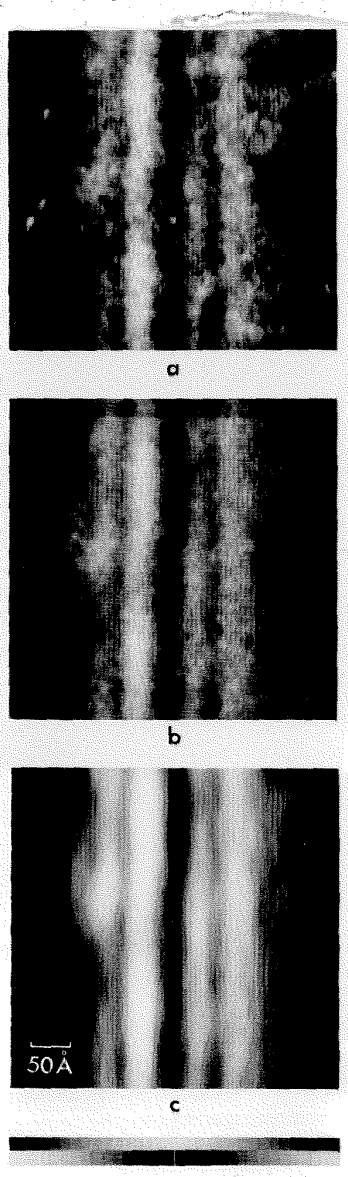


Fig. 4-15 (a) A 395 x 395Å portion of the dissociated-dipole image of Fig. 4-14 after positioning and linear contrast stretch. (b) Averaged picture resulting from the superposition of three micrographs in which "snow" noise was partially removed. (c) Low-pass box filtered version of (b). The artifacts that appear in the vertical direction are induced by the excessive vertical size of the kernel. The vertical striations that are visible in all the pictures are produced by the VFC output raster.

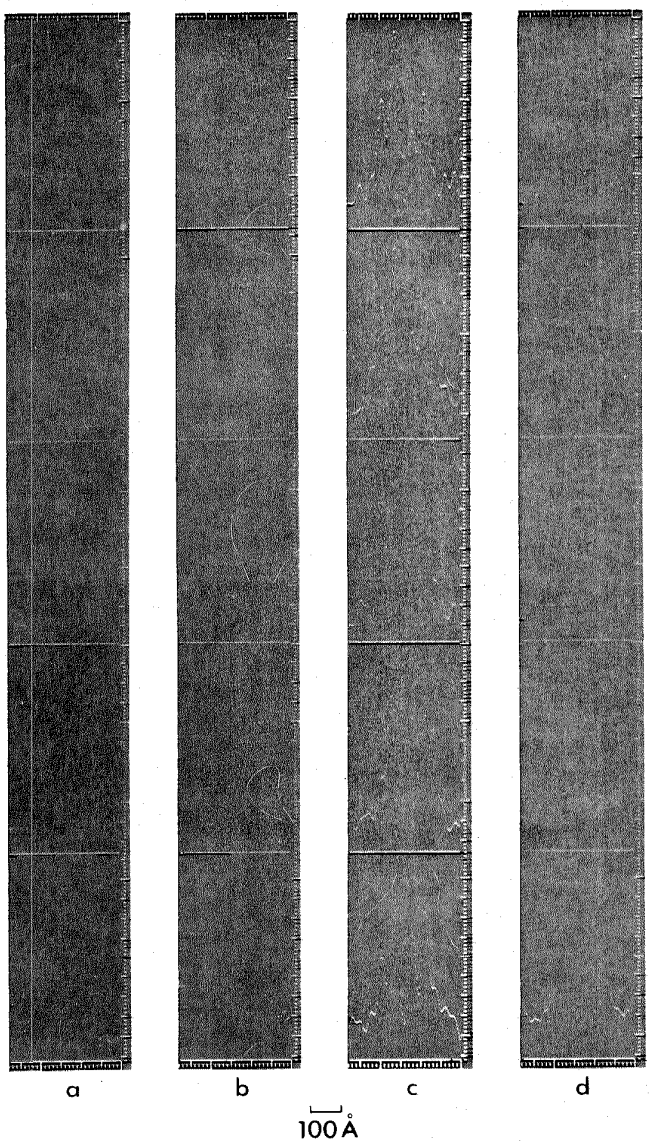


Fig. 4-16 Intensity profiles of Fig. 4-15 taken every 70Å from the top of the picture: (a) profiles across Fig. 4-15a; (b) profiles across the averaged picture of Fig. 4-15b; (c) Fourier smoothed version of (b); (d) Fourier smoothed profiles of Fig. 4-15c.

profiles that are suitable for comparison to theoretical image simulations.

4.3.4 Calculation of the Weak-Beam Contrast

4.3.4.1 Preliminary investigation of the weak-beam contrast from a dissociated edge dislocation

In order to check the performance of the computer program given in Appendix F and the sensitivity of the computed images to such parameters as the depth of the defect in the foil, the diffraction geometry, and the strain field of the defect, we conducted a preliminary study on a defect for which theoretical intensity profiles had been published. We shall first derive analytical relationships for the predicted image contrast, and then discuss the applicability of these relationships to more complex strain fields.

A) Survey of previous work

The theory of Sec. 4.3.1.3 has established that weak-beam images of dislocation lines show single intensity peaks when kinematical scattering conditions are satisfied. These peaks have very narrow halfwidths, and they identify the position of the dislocation line. The displacement of the image peak of the dislocation can be computed from (cf. Eq. (4.35)):[103]

$$s_g + \frac{d}{dz}(\underset{\sim}{g}\cdot\underset{\sim}{R}) = 0$$

$$\frac{d^2}{dz^2}(\underset{\sim}{g}\cdot\underset{\sim}{R}) = 0 \qquad (4.39)$$

In this section the validity of the above criteria for determining the separation of partial dislocations from the experimental observations

is discussed. The accuracy of the method will be verified by comparison with computed many-beam images.

Consider a dissociated edge dislocation ($\underset{\sim}{u} = [\bar{1}\bar{1}2]/\sqrt{6}$) in a [111] germanium crystal where

$$\underset{\sim}{b}^{(1)} = \frac{1}{2}[\bar{1}10] \rightarrow \frac{1}{6}[\bar{1}2\bar{1}] + \frac{1}{6}[\bar{2}11] = \underset{\sim}{b}_1^p + \underset{\sim}{b}_2^p$$

If the operating reflection is $\underset{\sim}{g} = \bar{2}20$ there is no stacking-fault contrast, since $\underset{\sim}{g}$ lies in the fault plane. In this case we have $\underset{\sim}{g}\cdot\underset{\sim}{b}_1^p = \underset{\sim}{g}\cdot\underset{\sim}{b}_2^p = 1$ for both partials. A kinematical estimate of the positions of the image peaks can then be found by applying Eq. (4.39). The predicted separation of the two weak-beam peaks is then[106]

$$\Delta_{kin} = \{\Delta^2 + \frac{16\pi^2}{s_g^2}(\underset{\sim}{g}\cdot\underset{\sim}{b}^p)^2 [1 + \frac{1}{2(1-\nu)}]\}^{1/2} \qquad (4.40)$$

Many-beam calculations were performed for this geometry for the systematic reflections ($\bar{2}20\cdots\bar{8}80$) in germanium and 200 keV electrons. The parameters which define the defect geometry are similar to those given in Fig. 4-17. We shall now examine how the image contrast varies with crystal thickness and with the operating beam. Then we shall compare the results of a dynamical calculation to the kinematical predictions for the spacing of the peaks.

B) Calculated weak-beam image of a dissociated edge dislocation

The image contrast calculations for a dissociated edge dislocation were carried out using the dynamical equations. The results are shown in Figs. 4-18 to 4-20 (Table 3-3 describes the operating conditions in each case). In general, one partial is imaged with higher

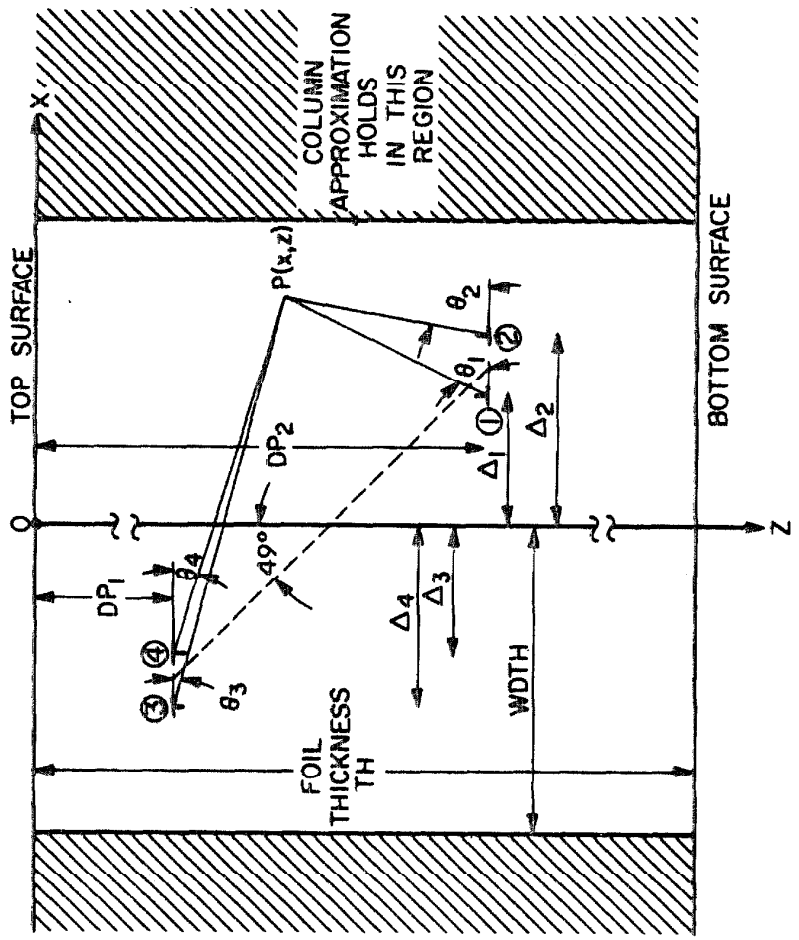


Fig. 4-17. Diagram showing the geometric parameters that were used in calculating the image of the dissociated interstitial dipole. The parameters indicated in this figure are consistent with those used in the program listed in Appendix F. The plane of the diagram is $[\bar{1}\,\bar{1}2]$

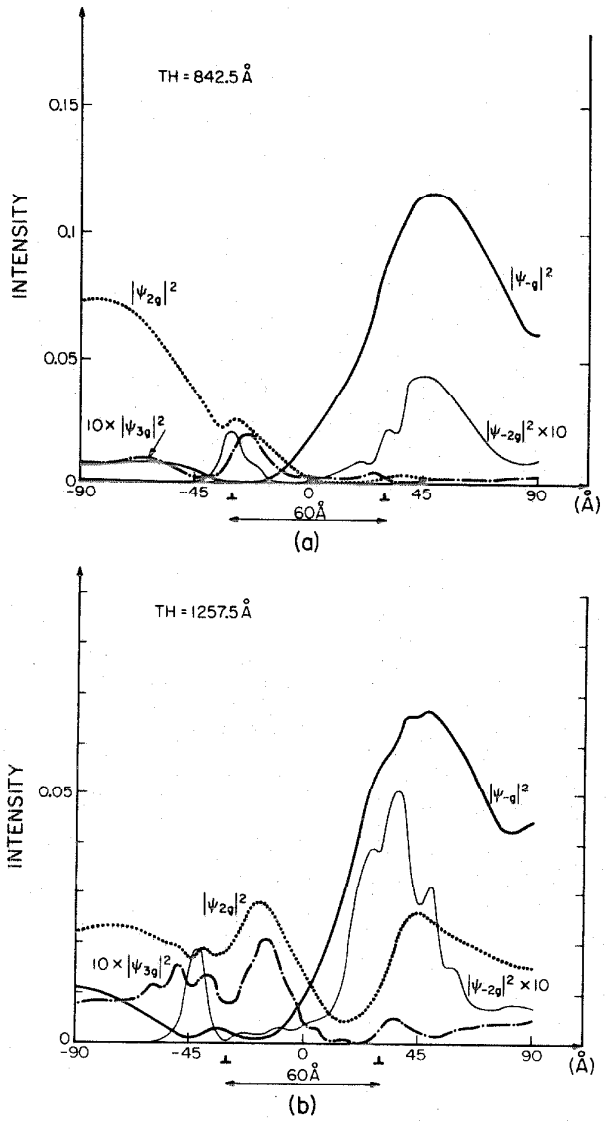


Fig. 4-18 Calculated weak-beam intensity profiles for a dissociated edge dislocation in germanium. The weakly excited beams resolve two dislocation peaks, whose relative magnitude varies when the specimen thickness changes from (a) t_o = 842.5Å to (b) t_o = 1257.5Å. In Table 4-3 a quantitative assessment of the peak spacings is given.

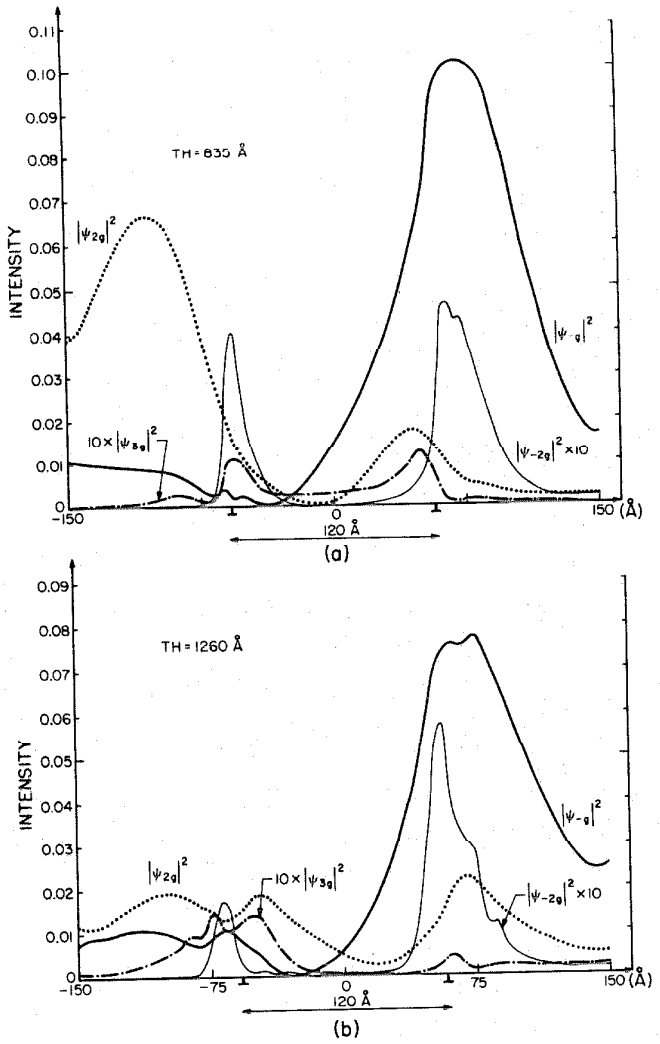


Fig. 4-19 Calculated weak-beam intensity profiles for a dissociated edge dislocation in germanium. The conditions are identical to the ones in Fig. 4-18, but the splitting of the partials is now $\Delta = 120\text{Å}$. Note that two dislocation peaks are resolved by the semi-weak beam $|\psi_{2g}|^2$.

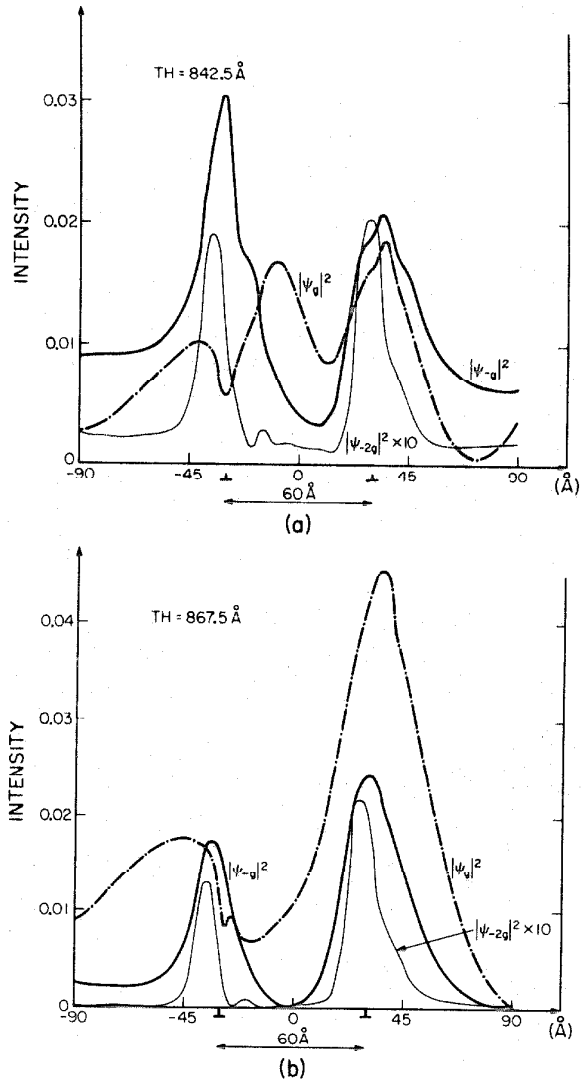


Fig. 4-20 Calculated weak-beam intensity profiles for a dissociated edge dislocation in germanium with $3\underset{\sim}{g}$ at the Bragg condition. Note the marked improvement in resolution when compared to Fig. 4-18. As expected, the beams with the largest deviation parameter produce an image peak with narrowest halfwidth.

Table 4-3

Summary of the diffracting conditions used in the calculation of the profiles in Figs. 4-18 to 4-20. All calculations were made using 6 systematic reflections.

Figure	DP_2 (Å)	Diffraction Geometry (FBRAGG)	Operating Beam	Deviation Parameter ($Å^{-1}$)	Δ (Å)	Δ_{kin} (Å)	Δ_{dyn} (Å)	Foil Thickness TH(Å)
4-18	510	1.0	-2g	-0.0188	60	67	76 79	842.5 1257.5
			- g	-0.0063	60	107		
			2g	-0.0063	60	107	62	1257.5
			3g	-0.0188	60	67	51 50	842.5 1257.5
4-19	510	1.0	-2g	-0.0188	120	123.5	122 123	835 1260
			- g	-0.0063	120	149		
			2g	-0.0063	120	149	149	835
			3g	-0.0188	120	123.5	105 112	835 1260
4-20	510	3.0	-2g	-0.0314	60	62.5	64 62	842.5 867.5
			- g	-0.0125	60	74.5	65 64	842.5 867.5
			g	0.0063	60	107	44 80	842.5 867.5
			2g	0.0063	60	107		

intensity than the other, due to the asymmetric strain field inside and outside of the extended dislocation. On changing the sign of $\underset{\sim}{g}\cdot\underset{\sim}{R}$ we therefore expect that the stronger image will shift from one partial dislocation to the other; this effect is clearly visible in the contrast curves which we shall now describe in detail.

In Fig. 4-18a,b, the computed profiles are plotted for two-foil thicknesses, and $\underset{\sim}{g}$ at the Bragg position. The dislocation line is positioned at a depth $DP_2 = 510\mathring{A}$ which nearly corresponds to a minimum in the intensity of $|\psi_g|^2$ in perfect crystal (see Fig. 4-13). The interaction between the weak beam, the strongly diffracting beam $\underset{\sim}{g}$, and the incident beam accounts for the variations in intensity with thickness. Since dynamical scattering is very strong the weak-beam images exhibit spurious side peaks which are thickness and depth dependent. An inspection of Table 4-3 reveals that the kinematical approach is invalid in predicting the spacing of the partials. However, the kinematical approach provides a qualitative understanding of the observation that the peak halfwidth is inversely proportional to the deviation parameter of the operating beam. We notice that the two dislocation peaks are best resolved by imaging the kinematical beams $-2\underset{\sim}{g}$ or $3\underset{\sim}{g}$. The contrast from these beams shows that thickness dependence only affects the position and the amplitude of the peaks, but not their mutual spacing. On the other hand, it was observed that the strong beams resolve the two peaks at certain crystal thicknesses. Since the deviation parameters of these beams is so small, the perturbing influence of the defect is only seen at discrete thickness intervals, approximately equal to their extinction distance.

These beams produce rather broad peaks, when compared to the peaks obtained with very weak beams.

The effect of separating the two partial dislocations was next investigated and the results are shown in Fig. 4-19. Observations similar to those made in the previous case, can be made about the sensitivity of the contrast to foil thickness. We also notice in Table 4-3 that a better agreement is obtained between the kinematical and the dynamical predictions. We attribute this to the fact that the strain field of each partial can be considered independently. Indeed, we shall show in the next section that a highly localized defect is best visualized under dynamical scattering conditions.

We shall now study a weak diffraction geometry in which $3\underset{\sim}{g}$ is at the Bragg position. One does not expect the column approximation to hold in this case, because the large curvature in the dispersion surface delocalizes the excitation points. Since the intensity of $3\underset{\sim}{g}$ in the crystal is much smaller than that of $\underset{\sim}{g}$ (the reason for this is the larger extinction distance $\xi_{3g} > \xi_g$ combined with an absorption of the waves in the crystal), the weak beams at the defect location will be mostly coupled with the incident beam. Consequently, one expects a disappearance of the spurious peaks, caused by multiple wave coupling in the weak-beam profiles. The computed image profiles shown in Fig. 4-20 confirm these predictions. We conclude that for a given defect configuration, a set of characteristic dynamical conditions must be selected in order to achieve an optimum weak-beam image contrast.

The weak-beam technique is easily interpreted when the main scattering comes from the incident beam into the weak diffracted beam. It is sometimes advantageous to excite many strong Bloch waves when the defect is very localized, as we shall demonstrate. However, contrast interpretation becomes challenging in these conditions, since the weak-beam image now depends on the diffraction geometry, on the depth of the defect, and on the foil thickness. By examining a simpler configuration, we have reached certain conclusions which will serve as guidelines in the following discussion.

4.3.4.2 Computer simulation of a dissociated-dipole weak-beam image

A series of calculations were undertaken to simulate the weak-beam contrast of the dissociated dipole studied in Sec. 4.3.3.2. We aimed at achieving a match between the profile of the computer processed micrograph shown in Fig. 4-16c and a theoretical image profile. The column approximation could not be used for this particular defect because of its localization. We know that the lateral spread of the wave points that are excited during the scattering process depends on the Fourier transform of the perturbing potential. This lateral spreading will be sizeable for defects with localized strain fields. Indeed, our calculations indicated an attenuation of the peaks, or even their complete disappearance, when the column approximation was made.

The computed image profiles were plotted for all the diffracting conditions for which experimental evidence was available (see Table 4-1). We found that the resolution capability of the weak-beam

method depended on a multiplicity of interacting factors. For instance, in kinematical conditions (i.e., when no systematic reflection lies on the Ewald sphere) only two image peaks could be detected. We can obtain a qualitative understanding of this phenomenon by plotting $\beta_g' = \frac{\partial}{\partial z}(\underset{\sim}{g}\cdot\underset{\sim}{R})$ at various spatial positions, as a function of depth z (see Fig. 4-21). By applying the criterion of Eq. (4.39), we expect an image peak when $\beta_g' + s_g = 0$. An inspection of Fig. 4-21 reveals that this latter condition is fulfilled at only two positions: when $s_g = 0.0063\text{Å}^{-1}$, predicted peak locations are at a distance of $\simeq \pm 45\text{Å}$ from the center of the dipole. This conclusion is demonstrated in Fig. 4-22a.

In dynamical conditions four peaks could only be observed distinctly if $3\underset{\sim}{g}$ was moderately excited and a semi-weak beam used for the imaging. By semi-weak beam, we mean a weakly excited beam that does not deviate too far from the Bragg condition (e.g., we imaged Fig. 4-10 with $s_g = 0.0063\text{Å}^{-1}$). Furthermore, the splitting of the edge dislocations into partial components is only visible in a narrow range of z values occurring in multiples of ξ_g. This range never exceeded 40Å for the geometry under which Fig. 4-10 was recorded, and corresponded to a minimum of the average beam intensity.

The depth of the defect proved to be the most critical parameter in our ability to resolve four peaks. We have indicated that once the weak beam has gained intensity in the region of the defect, it does not lose this information while passing through the remainder of the foil. Consequently, the imaging beam will be able to resolve

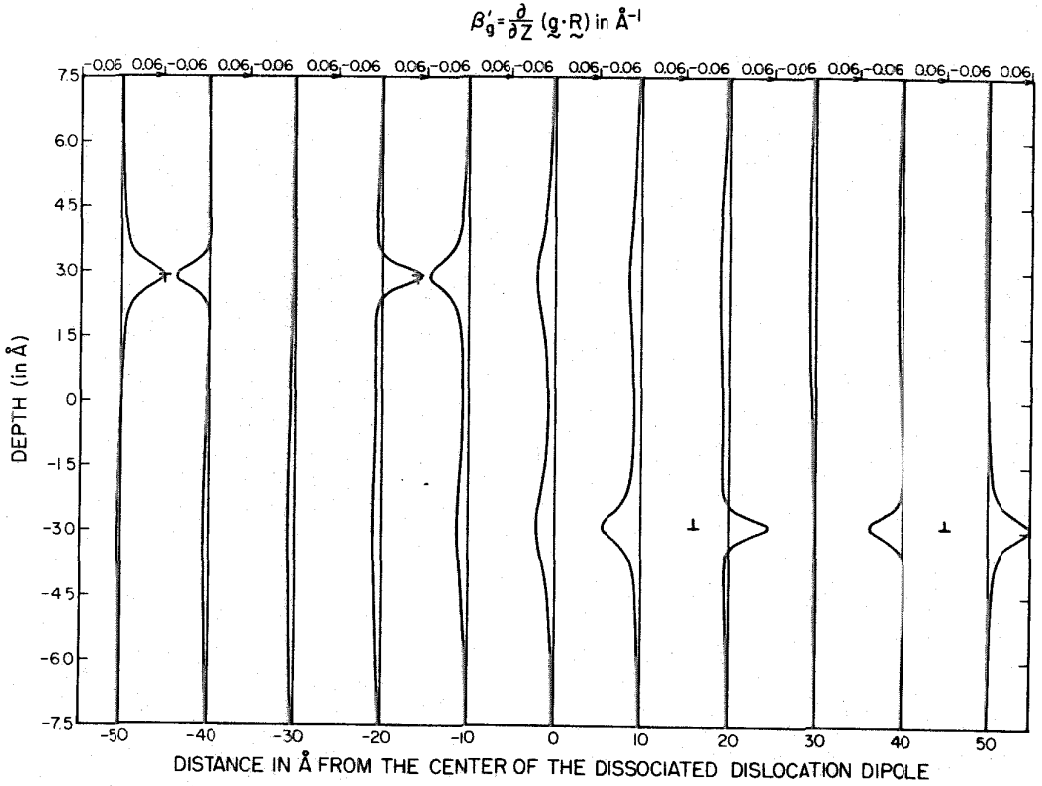


Fig. 4-21 The value of $\beta'_g = \frac{\partial}{\partial z}(\underset{\sim}{g}\cdot\underset{\sim}{R})$ down columns at various distances from a dissociated dislocation dipole lying parallel to the surface of a germanium foil. Referring to Fig. 4-17 the parameters are: $\Delta_1 = \Delta_3 = 16\text{Å}$, $\Delta_2 = \Delta_4 = 45\text{Å}$, $DP_1 = 1943\text{Å}$, $DP_2 = 1997\text{Å}$, and $\underset{\sim}{g} = [\bar{2}20]$. The weak-beam peaks are expected to occur for the column in which $s_g = 0.0063\text{Å}^{-1} = -\beta'_g$. This condition is only satisfied at two columns located at $\simeq \pm 45\text{Å}$ from the center of the dipole, indicating that the localized strain field can only be resolved by dynamically scattered electrons.

the structure only if it interacts strongly with the strain field of the defect; this implies that its magnitude should reach a minimum in a corresponding depth of perfect crystal in order to maximize the interaction. We have verified this depth dependent "resolving power" of a beam by varying the depth of the defect about a minimum of its intensity vs. thickness profile (see Fig. 4-13). If we refer to a sketch of the defect geometry shown in Fig. 4-17, the sensitivity of the imaging technique to such parameters as DP_1 and DP_2 was analyzed. When the two components of the dipole lay nearly symmetrical with respect to a minimum of the rocking curve, we achieved the best resolution. Otherwise, one defect interacted with the imaging beam at the expense of the other defect which exhibited only one image peak throughout the foil.

Lastly, the imaging mode for which the best resolution could be attained was investigated. It has been demonstrated theoretically that by exciting $3\underset{\sim}{g}$ and by imaging either the $\underset{\sim}{g}$ or $2\underset{\sim}{g}$ reflections, weak-beam peaks are obtained for each partial dislocation at periodic depth intervals. Calculations have also shown that only three peaks can be resolved by exciting $-\underset{\sim}{g}$ and imaging with $\underset{\sim}{g}$; this is experimentally verified in Fig. 4-23.

Despite the sensitivity of this imaging mode to depth, diffraction geometry, and foil thickness, a comparison of the experimental profile of Fig. 4-16c with a simulated profile was possible. The image was calculated for the following dipole parameters (cf. Fig. 4-17): $\Delta_1 = \Delta_3 = 16\text{Å}$, $\Delta_2 = \Delta_4 = 45\text{Å}$, $DP_1 = 1943\text{Å}$, $DP_2 = 1997\text{Å}$, TH = 3412.5Å, and $\lambda = 0.0251\text{Å}$. Both the theoretical and the experimental profiles

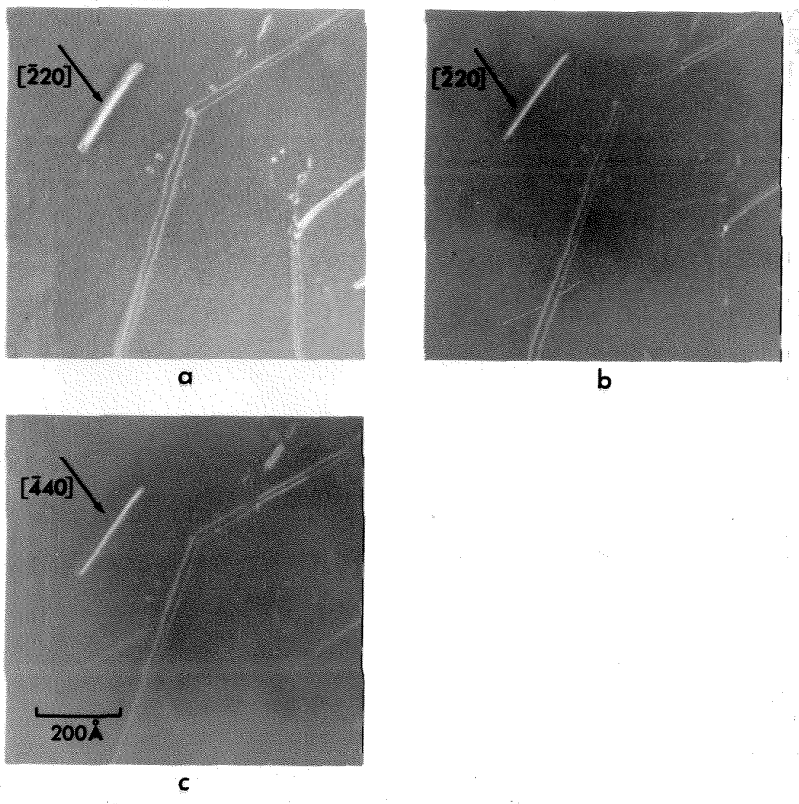


Fig. 4-23 Weak-beam images corresponding to: (a) $s_{-g} < 0$; (b) $s_{-g} \lesssim 0$, and (c) $s_{-g} < 0$.

shown in Fig. 4-24 exhibit similar behaviors. However, due to the rapid fluctuations in peak magnitude with thickness level, a one to one correspondence could not be achieved. A comparison of the mutual peak spacings will yield values for the splitting of the partials if we rely on the validity of a continuum elasticity model. By measuring the separation of the peaks, we obtained values: $\Delta_1^{th} \simeq \Delta_3^{th} \simeq 30\mathring{A}$, $\Delta_2^{th} \simeq \Delta_4^{th} \simeq 56\mathring{A}$, which were approximately thickness independent. If we recall from Sec. 4.3.3.2 that $\Delta_1^{obs} = 32.5\mathring{A}$ and $\Delta_2^{obs} = 77.5\mathring{A}$, we may conclude that the dipole parameters that were selected are too small. An estimate of their real values may be found by linear interpolation and leads to the result: $\Delta_1 = \Delta_3 \simeq 17\mathring{A}$, $\Delta_2 = \Delta_4 \simeq 62\mathring{A}$.

The sensitivity of the calculated image to various parameters (e.g., dislocation depth, foil thickness) renders a quantitative analysis at this high-resolution range difficult. A direct correspondence between image and structure will be reached when all the experimental and the theoretical uncertainties are analyzed. Further investigations are thus needed to predict the experimental conditions that would yield the best resolution for a given defect. For instance, in order to investigate a core structure, one might perform a high-resolution weak-beam experiment whose operational characteristics would have been suggested by the theory. A contrast analysis at the atomic level would require the use of a discrete atomic model for a dislocation core[113].

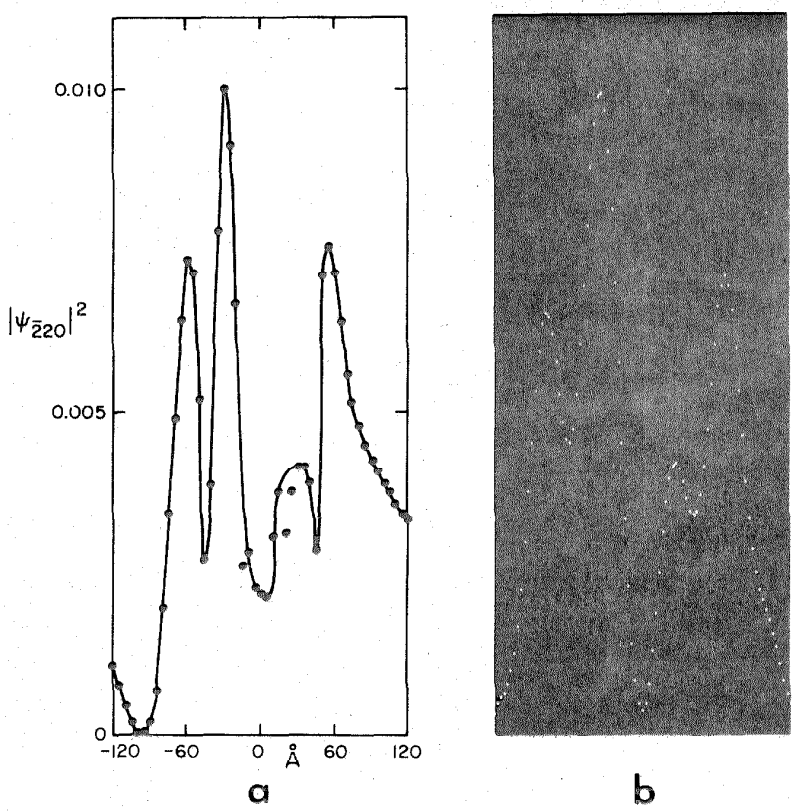


Fig. 4-24 (a) Theoretical profile of the dissociated edge dipole in germanium; (b) Experimental profile from the top of Fig. 4-16c.

4.4 Conclusion

Provided that the underlying assumptions concerning the statistical independence and additiveness of the noise to the signal are valid, an improvement in the signal-to-noise ratio can be attained by using image processing techniques. In periodic images this improvement is obtained by either translating the repeat unit and superimposing the image upon itself, or by filtering out the non-periodic information using Fourier masks. This latter method offers the possibility of extracting a signal which differs from the perfect periodicity. Consequently, crystal imperfections such as dislocation arrangements can be readily detected in a near-periodic lattice image.

Another application of image processing techniques lies in their capability of processing the raw data in such a way as to permit a quantitative comparison with the predictions of image formation theories. An example of this potential was given in the analysis of weak-beam images of a dissociated dislocation dipole. The weak-beam imaging technique yields a micrograph containing high-resolution details which are embedded in a background noise. By processing the micrograph, we were able to compare a theoretical image to the enhanced image.

The variety of problems to which an electron microscopist is exposed during the interpretation of micrographs would be greatly alleviated by interfacing the microscope with a computer. Such tasks as identifying defects or processing the signal would be performed automatically, and open the way to a better understanding of the structure of imperfect specimens at the atomic level.

APPENDIX A

Effects of Partial Coherence on the Contrast Transfer Functions

A-1 Spatial Incoherence

We start with the expression for the image intensity, calculated from an incoherent superposition of elastically scattered waves with different $\underset{\sim}{K}_o$, obtained in Sec. 1.3.2.2A

$$|\psi(\underset{\sim}{r}_i)|^2 = \iiint S_o(\underset{\sim}{k}-\underset{\sim}{K}_o)S_o^*(\underset{\sim}{k}'-\underset{\sim}{K}_o)T(\underset{\sim}{k})T^*(\underset{\sim}{k}')\exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i] \cdot Q(\underset{\sim}{K}_o)\, d\underset{\sim}{k}\, d\underset{\sim}{k}'d\underset{\sim}{K}_o \qquad (A.1)$$

where $Q(\underset{\sim}{K}_o)$ is the angular distribution of the incident beam at the effective source plane.

Let us now consider a weakly scattering object, whose Fourier transform is

$$S_o(\underset{\sim}{k}) = \delta(\underset{\sim}{k}) + i\Phi(\underset{\sim}{k}) - M(\underset{\sim}{k}) \qquad (A.2)$$

Quadratic terms in the integral $|\psi(\underset{\sim}{r}_i)|^2$ involving $[i\Phi(\underset{\sim}{k}-\underset{\sim}{K}_o) - M(\underset{\sim}{k}-\underset{\sim}{K}_o)]\,[-i\Phi^*(\underset{\sim}{k}'-\underset{\sim}{K}_o) - M^*(\underset{\sim}{k}'-\underset{\sim}{K}_o)]$ will be neglected in this analysis since they represent dark field image contrast which makes a minor contribution to the overall contrast. We then obtain, from Eq. (A.1), the following equation for the image intensity:

$$|\psi(\underset{\sim}{r}_i)|^2 = \iiint \delta(\underset{\sim}{k}-\underset{\sim}{K}_o)\delta(\underset{\sim}{k}'-\underset{\sim}{K}_o)T(\underset{\sim}{k})T^*(\underset{\sim}{k}')\exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i]Q(\underset{\sim}{K}_o)d\underset{\sim}{k}\, d\underset{\sim}{k}'d\underset{\sim}{K}_o$$

$$+ \iiint [i\Phi(\underset{\sim}{k}-\underset{\sim}{K}_o)\delta(\underset{\sim}{k}'-\underset{\sim}{K}_o) - i\Phi^*(\underset{\sim}{k}'-\underset{\sim}{K}_o)\delta(\underset{\sim}{k}-\underset{\sim}{K}_o)]T(\underset{\sim}{k})T^*(\underset{\sim}{k}')$$
$$\cdot \exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i]\, Q(\underset{\sim}{K}_o)d\underset{\sim}{k}\, d\underset{\sim}{k}'d\underset{\sim}{K}_o$$

$$- \iiint [M(\underset{\sim}{k}-\underset{\sim}{K}_o)\delta(\underset{\sim}{k}'-\underset{\sim}{K}_o) + M^*(\underset{\sim}{k}'-\underset{\sim}{K}_o)\delta(\underset{\sim}{k}-\underset{\sim}{K}_o]T(\underset{\sim}{k})T^*(\underset{\sim}{k}')$$
$$\cdot \exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i]\, Q(\underset{\sim}{K}_o)d\underset{\sim}{k}\, d\underset{\sim}{k}'d\underset{\sim}{K}_o \qquad (A.3)$$

After substituting the expression

$$T(\underset{\sim}{k}) = \frac{1}{M}\, b(\underset{\sim}{k})\, e^{-i\gamma(\underset{\sim}{k})} \qquad (A.4)$$

into Eq. (A.3), we shall examine the resulting three integrals

For the first term of Eq. (A.3) we can write

$$I_{si}^{(1)} = \int_{B_{con}} |T(\underset{\sim}{K}_o)|^2\, Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o = \frac{1}{M^2} \qquad (A.5)$$

because the distribution $Q(\underset{\sim}{K}_o)$ is normalized and the effective source aperture is smaller than the exit pupil (i.e., $b(\underset{\sim}{k}) = 1$ in the area of integration). This term is the primary wave and contributes to the background intensity.

The second term in Eq. (A.3) is responsible for phase contrast and can be expressed as

$$I_{si}^{(2)} = i \iint \Phi(\underset{\sim}{k}-\underset{\sim}{K}_o)T(\underset{\sim}{k})T^*(\underset{\sim}{K}_o)\, \exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{K}_o)\cdot\underset{\sim}{r}_i]\, Q(\underset{\sim}{K}_o)d\underset{\sim}{k}\, d\underset{\sim}{K}_o$$
$$-i \iint \Phi^*(\underset{\sim}{k}-\underset{\sim}{K}_o)T(\underset{\sim}{K}_o)T^*(\underset{\sim}{k})\, \exp[2\pi i(\underset{\sim}{K}_o-\underset{\sim}{k})\cdot\underset{\sim}{r}_i]\, Q(\underset{\sim}{K}_o)d\underset{\sim}{k}\, d\underset{\sim}{K}_o \qquad (A.6)$$

The pupil function, the effective source, and the phase transform obey the following symmetry properties:

$$\Phi^*(\underset{\sim}{k}) = \Phi(-\underset{\sim}{k})$$

$$Q(\underset{\sim}{K}_o) = Q(-\underset{\sim}{K}_o)$$

$$b(\underset{\sim}{k}) = b(-\underset{\sim}{k})$$

$$\gamma(\underset{\sim}{k}) = \gamma(-\underset{\sim}{k}) \qquad \text{(A.7)}$$

Observing these symmetry rules and making the change of variables, $\underset{\sim}{k}' = \underset{\sim}{k}-\underset{\sim}{K}_o$, in Eq. (A.6) leads to a simplified phase contrast term:

$$I_{si}^{(2)} = \frac{i}{M^2}\iint \Phi(\underset{\sim}{k}')b(\underset{\sim}{k}'+\underset{\sim}{K}_o)b(\underset{\sim}{K}_o)e^{-i[\gamma(\underset{\sim}{k}'+\underset{\sim}{K}_o)-\gamma(\underset{\sim}{K}_o)]}$$

$$\cdot \exp(2\pi i\underset{\sim}{k}'\cdot\underset{\sim}{r}_i)Q(\underset{\sim}{K}_o)d\underset{\sim}{k}'d\underset{\sim}{K}_o$$

$$-\frac{i}{M^2}\iint \Phi(\underset{\sim}{k}')b(\underset{\sim}{K}_o)b(\underset{\sim}{K}_o+\underset{\sim}{k}')e^{i[\gamma(\underset{\sim}{k}'+\underset{\sim}{K}_o)-\gamma(\underset{\sim}{K}_o)]}$$

$$\cdot \exp(2\pi i\underset{\sim}{k}'\cdot\underset{\sim}{r}_i)Q(\underset{\sim}{K}_o)d\underset{\sim}{k}'d\underset{\sim}{K}_o \qquad \text{(A.8)}$$

The phase contrast transfer function for partially coherent illumination is then represented by:

$$Ph_{si}(\underset{\sim}{k}) = 2\int_{B_{con}} b(\underset{\sim}{K}_o)b(\underset{\sim}{k}+\underset{\sim}{K}_o)\sin[\gamma(\underset{\sim}{k}+\underset{\sim}{K}_o)-\gamma(\underset{\sim}{K}_o)]Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o \qquad \text{(A.9)}$$

The highest spatial frequency transmitted by the objective aperture, or exit pupil, is

$$k_m = \theta_{obj}/\lambda \qquad \text{(A.10)}$$

The effective area of integration in Eq. (A.9) is the overlapping area

of the source and exit pupil; this remains the full source area only if

$$|\underset{\sim}{k}| + |\underset{\sim}{K}_o| < k_m \tag{A.11}$$

This condition is satisfied in conventional electron microscopy, so that the terms $b(\underset{\sim}{K}_o)$ and $b(\underset{\sim}{k}+\underset{\sim}{K}_o)$ can be omitted for small source apertures. In this approximation, the phase transfer function reduces to

$$Ph_{si}(\underset{\sim}{k}) = 2 \int_{B_{con}} \sin[\gamma(\underset{\sim}{k}+\underset{\sim}{K}_o) - \gamma(\underset{\sim}{K}_o)]Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o \tag{A.12}$$

A similar analysis for the third term, $I_{si}^{(3)}$, reveals that the amplitude contrast transfer function for partially coherent illumination is described by:

$$Amp_{si}(\underset{\sim}{k}) = 2 \int_{B_{con}} \cos[\gamma(\underset{\sim}{k}+\underset{\sim}{K}_o) - \gamma(\underset{\sim}{K}_o)]Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o \tag{A.13}$$

Now let us expand $\gamma(\underset{\sim}{k}+\underset{\sim}{K}_o)$, in the neighborhood of $\underset{\sim}{k}$, to the first order in $\underset{\sim}{K}_o$ [71]. This approximation is valid for the small condenser apertures currently used at high resolution, where the spatial frequencies of interest are restricted to $|\underset{\sim}{k}| >> |\underset{\sim}{K}_o|$ and yields:

$$\gamma(\underset{\sim}{k}+\underset{\sim}{K}_o) \simeq \gamma(\underset{\sim}{k}) + \underset{\sim}{K}_o \cdot \nabla\gamma(\underset{\sim}{k}) \tag{A.14}$$

where $\nabla\gamma(\underset{\sim}{k})$ is the gradient of the objective lens aberration function. Neglecting $\underset{\sim}{K}_o$ terms of high order allows us to omit $\gamma(\underset{\sim}{K}_o)$, since it does not contain terms of lower order than two. Within these approximations, Eq. (A.12) may be written

$$Ph_{si}(\underset{\sim}{k}) = 2 \sin \gamma(\underset{\sim}{k})\{\int \cos[\underset{\sim}{K}_o \cdot \nabla\gamma(\underset{\sim}{k})]Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o$$

$$+ 2 \cos \gamma(\underset{\sim}{k})\{ \int \sin[\underset{\sim}{K}_o \cdot \nabla\gamma(\underset{\sim}{k})]Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o \qquad (A.15)$$

and since the source distribution $Q(\underset{\sim}{K}_o)$ is symmetric the second integral vanishes. The transfer function is a product of the transfer function for coherent illumination with an envelope function:

$$Ph_{si}(\underset{\sim}{k}) = Ph(\underset{\sim}{k})\ E_{si}(\underset{\sim}{k}) \qquad (A.16)$$

where

$$E_{si}(\underset{\sim}{k}) = \int \cos[\underset{\sim}{K}_o \cdot \nabla\gamma(\underset{\sim}{k})]Q(\underset{\sim}{K}_o)d\underset{\sim}{K}_o = \int Q(\underset{\sim}{K}_o)e^{i\underset{\sim}{K}_o \cdot \nabla\gamma(\underset{\sim}{k})}d\underset{\sim}{K}_o \qquad (A.17)$$

and the envelope function is the Fourier transform of the effective source distribution, estimated at $\nabla\gamma(\underset{\sim}{k})/2\pi$.

Consider, for example, a disk source with rotational symmetry:

$$Q(\underset{\sim}{K}_o) = \begin{cases} \lambda^2/\pi\theta_{con}^2 & \text{if } |\underset{\sim}{K}_o| < \theta_{con}/\lambda \\ 0 & \text{elsewhere} \end{cases} \qquad (A.18)$$

where θ_{con} is condenser aperture half angle. The resulting envelope function has the form

$$E_{si}(\underset{\sim}{k}) = 2\,\frac{J_1[|\nabla\gamma(\underset{\sim}{k})|\theta_{con}/\lambda]}{|\nabla\gamma(\underset{\sim}{k})|\ \theta_{con}/\lambda} \qquad (A.19)$$

The above expression can be further simplified by assuming axial astigmatism to be negligible. The magnitude of the gradient of $\gamma(\underset{\sim}{k})$

becomes

$$|\nabla\gamma(\underset{\sim}{k})| = \lambda \left|\frac{d\gamma(\theta)}{d\theta}\right| = 2\pi(C_s\theta^3 - \Delta Z\theta) \tag{A.20}$$

The specific envelope function for a disk-shaped source is then[69,71],

$$E_{si}(\theta) = 2\,\frac{J_1[2\pi\theta_{con}(C_s\theta^3 - \Delta Z\theta)/\lambda]}{2\pi\theta_{con}(C_s\theta^3 - \Delta Z\theta)/\lambda} \tag{A.21}$$

An analogous treatment for the amplitude transfer function yields an identical product representation, with the same envelope function $E_{si}(\underset{\sim}{k})$. This representation has proven to be adequate in the range of illumination apertures used in high-resolution experiments. The approximations used have been checked by comparing exact and approximate computations of the phase transfer functions given by (A.12) and (A.16).

A-2 Chromatic Incoherence

Chromatic aberration arises if there is an energy spread in the incident beam or when time fluctuations of the objective lens current and of the accelerating voltage are taken into account. In Sec. 1.3.2.2B an expression for the resulting image intensity was shown to be

$$|\psi(\underset{\sim}{r}_i)|^2 = \iint S_o(\underset{\sim}{k})S_o^*(\underset{\sim}{k}')T(\underset{\sim}{k})T^*(\underset{\sim}{k}')L(\underset{\sim}{k},\underset{\sim}{k}')\exp[2\pi i(\underset{\sim}{k}-\underset{\sim}{k}')\cdot\underset{\sim}{r}_i]d\underset{\sim}{k}\,d\underset{\sim}{k}' \tag{A.22}$$

For the term $L(\underset{\sim}{k},\underset{\sim}{k}')$ in Eq. (A.22), when the energy spread is considered, we can write

$$L_{es}(\underset{\sim}{k},\underset{\sim}{k}') = \int_{-\infty}^{\infty} \exp[-i\pi\lambda C_c(k^2-k'^2)\,\frac{\delta U}{U}]\, N(\delta U)\, d(\delta U) \qquad (A.23)$$

When time fluctuations in lens current and voltage are considered, $L(\underset{\sim}{k},\underset{\sim}{k}')$ becomes

$$L_{tf}(\underset{\sim}{k},\underset{\sim}{k}') = \int_{0}^{t_r} \exp[-i\pi\lambda C_c(k^2-k'^2)(\frac{\Delta U(t')}{U} - 2\,\frac{\Delta i_\ell(t')}{i_\ell})]dt' \qquad (A.24)$$

where $\Delta i_\ell(t')$ is the time fluctuation in lens current and $\Delta U(t')$ is the time fluctuation in accelerating voltage.

For a weakly scattering object, we shall prove that the image transform and the object spectrum obey a linear relationship. We shall also derive the contrast transfer functions when the illumination is not chromatically coherent. The object transform is

$$S_o(\underset{\sim}{k}) = \delta(\underset{\sim}{k}) + i\Phi(\underset{\sim}{k}) - M(\underset{\sim}{k}) \qquad (A.25)$$

and, as in the previous section, quadratic terms $[i\Phi(\underset{\sim}{k}) - M(\underset{\sim}{k})][-i\Phi^*(\underset{\sim}{k}') - M^*(\underset{\sim}{k}')]$ in the expansion of $S_o(\underset{\sim}{k})\, S_o^*(\underset{\sim}{k}')$ will be ignored. The image intensity given by Eq. (A.22) becomes

$$\begin{aligned}|\psi(\underset{\sim}{r}_i)|^2 = {} & |T(0)|^2\, L(0,0) + i\{\int \Phi(\underset{\sim}{k})T(\underset{\sim}{k})T^*(0)L(\underset{\sim}{k},0)\exp(2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k} \\ & - \int \Phi^*(\underset{\sim}{k})T(0)T^*(\underset{\sim}{k})L^*(0,\underset{\sim}{k})\exp(-2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k}\} \\ & - \{\int M(\underset{\sim}{k})T(\underset{\sim}{k})T^*(0)L(\underset{\sim}{k},0)\exp(2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)\, d\underset{\sim}{k} \\ & + \int M^*(\underset{\sim}{k})T(0)T^*(\underset{\sim}{k})L^*(0,\underset{\sim}{k})\exp(-2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k}\} \end{aligned} \qquad (A.26)$$

Now making use of the following symmetry properties

$$T(\underset{\sim}{k}) = T(-\underset{\sim}{k})$$

$$L^*(0,-\underset{\sim}{k}) = L^*(\underset{\sim}{k},0) \tag{A.27}$$

we can further simplify Eq. (A.26):

$$|\psi(\underset{\sim}{r}_i)|^2 = \frac{1}{M^2} + \frac{i}{M^2}\int \Phi(\underset{\sim}{k})[e^{-i\gamma(\underset{\sim}{k})}L(\underset{\sim}{k},0) - e^{i\gamma(\underset{\sim}{k})}L^*(\underset{\sim}{k},0)]\exp(2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k}$$

$$- \frac{1}{M^2}\int M(\underset{\sim}{k})[e^{-i\gamma(\underset{\sim}{k})}L(\underset{\sim}{k},0) + e^{i\gamma(\underset{\sim}{k})}L^*(\underset{\sim}{k},0)]\exp(2\pi i\underset{\sim}{k}\cdot\underset{\sim}{r}_i)d\underset{\sim}{k} \tag{A.28}$$

It follows that the phase and amplitude contrast transfer functions may be written in the form

$$Ph_{ci}(\underset{\sim}{k}) = 2\sin\gamma(\underset{\sim}{k})\,\mathrm{Re}\{L(\underset{\sim}{k},0)\} - 2\cos\gamma(\underset{\sim}{k})\,\mathrm{Im}\{L(\underset{\sim}{k},0)\} \tag{A.29}$$

and

$$Amp_{ci}(\underset{\sim}{k}) = 2\cos\gamma(\underset{\sim}{k})\,\mathrm{Re}\{L(\underset{\sim}{k},0)\} + 2\sin\gamma(\underset{\sim}{k})\,\mathrm{Im}\{L(\underset{\sim}{k},0)\} \tag{A.30}$$

By treating independently the finite energy width of the electrons, and the time fluctuations, one obtains in both cases

$$\mathrm{Im}\{L(\underset{\sim}{k},0)\} \simeq 0 \tag{A.31}$$

for most practical purposes. Consequently, the transfer functions may be represented by a product of the ideal transfer function for chromatically coherent illumination and an envelope function given by

$$E_{ci}(\underset{\sim}{k}) = L(\underset{\sim}{k},0) \tag{A.32}$$

The energy spread, when measured experimentally at various filament bias voltages and emitter temperatures T_{fil} is found to fit a Maxwellian distribution[25,70]:

$$N_e(\delta U)\, d(\delta U) = \frac{\delta U + k_B T_{fil}}{(k_B T_{fil})^2} \exp[-(\delta U + 2k_B T_{fil})/k_B T_{fil}]d(\delta U) \qquad (A.33)$$

where $U \geq -2k_B T_{fil}$. The Maxwellian distribution has a halfwidth

$$\delta U_{es} = 1.22\ k_B T_{fil} \qquad (A.34)$$

This distribution shows little asymmetry about its peak value and may be approximated by a Gaussian of the same halfwidth

$$N_e(\delta U)d(\delta U) \simeq \frac{1}{\delta U_{es}\sqrt{2\pi}} \exp[-\delta U^2/2\delta U_{es}^2] \qquad (A.35)$$

Inserting Eq. (A.35) into Eq. (A.23), we note that $\mathrm{Im}\{L_{es}(\underset{\sim}{k},0)\}$ vanishes, so that we obtain for the energy-spread envelope function

$$E_{es}(\theta) = L_{es}(\underset{\sim}{k},0) = \exp[-(\frac{\delta U_{es}}{U}\frac{C_c}{\lambda}\frac{\pi}{\sqrt{2}}\theta^2)^2] \qquad (A.36)$$

Time-dependent fluctuations in lens current and accelerating voltage are responsible for the random rippling of the effective defocusing about its mean value, ΔZ, (cf. Eq. (1.77)):

$$\delta Z(t') = C_c[\frac{\Delta U(t')}{U} - 2\frac{\Delta i_\ell(t')}{i_\ell}] \qquad (A.37)$$

One may characterize $\delta Z(t')$ by the probability density function $H_z(\delta Z)$ that, at a given time t' less than the exposure time t_r,

δZ takes values between δZ and $\delta Z + d(\delta Z)$. The time integration in Eq. (A.24) is substituted by a spatial integration

$$L_{tf}(\underset{\sim}{k},0) = \int_{-\infty}^{\infty} \exp(-\pi i \lambda \underset{\sim}{k}^2 \delta Z)\, H_z(\delta Z)\, d(\delta Z) \tag{A.38}$$

Now, assuming $H_z(\delta Z)$ to be a Gaussian distribution with halfwidth σ_{tf}, yields

$$L_{tf}(\underset{\sim}{k},0) = \exp[-(\frac{\sigma_{tf}}{\lambda} \frac{\pi\theta^2}{\sqrt{2}})^2] \tag{A.39}$$

Let us define ΔU as the standard deviation in accelerating voltage and Δi_ℓ as the standard deviation in objective lens current. Since $\delta Z(t')$ is a sum of two Gaussian distributed variables, we have the identity:

$$\sigma_{tf} = C_c[(\frac{\Delta U}{U})^{1/2} + (\frac{2\Delta i_\ell}{i_\ell})^{1/2}] \tag{A.40}$$

The resulting envelope function can be written

$$E_{tf}(\theta) = \exp\{-[(\frac{\Delta U}{U})^{1/2} + (\frac{2\Delta i_\ell}{i_\ell})^{1/2}) \frac{C_c}{\lambda} \frac{\pi}{\sqrt{2}} \theta^2]^2\} \tag{A.41}$$

APPENDIX B

Influence of Specimen Thickness on the Linear Theory of Image Formation

The three-dimensional "structure factor" of an aperiodic array of atoms, sampled by the Ewald sphere, yields the object Fresnel transform

$$F_f(\underset{\sim}{k}) = \sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) + i\pi z_o^j \theta^2/\lambda - 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j] \qquad (B.1)$$

for kinematically scattered electrons. Treating the specimen as a two-dimensional density distribution is equivalent to neglecting the curvature of the Ewald sphere. By approximating the sphere by its tangent plane, one may characterize the specimen by its plane structure factor

$$F(\underset{\sim}{k}) = \sum_{j=1}^{N_a} (f_j'(\theta) + f_j''(\theta))\exp(-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j) \qquad (B.2)$$

where $f_j'(\theta)$ and $f_j''(\theta)$ are the real and imaginary parts, respectively, of the atomic scattering amplitudes. According to Eq. (1.8), the elastically scattered wave at the image plane can be written as

$$\psi_s(\underset{\sim}{r}_i) = \frac{iA_o\lambda}{M} \int b(\underset{\sim}{k}) e^{-i\gamma(\underset{\sim}{k})} \left\{\sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) + i\pi z_o^j\theta^2/\lambda - 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j]\right\} e^{2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i} \, d\underset{\sim}{k} \qquad (B.3)$$

If we neglect the dark field contribution, the bright field image contrast is expressed as:

$$\mathrm{Con}(\underset{\sim}{r}_i) = \frac{|\psi_s(\underset{\sim}{r}_i) + A_o/M|^2 - A_o^2/M^2}{A_o^2/M^2} \simeq 2\,\frac{M}{A_o}\,\mathrm{Re}\{\psi_s(\underset{\sim}{r}_i)\} \qquad (B.4)$$

This equation can be further simplified if we apply the symmetry relations

$$b(\underset{\sim}{k}) = b(-\underset{\sim}{k})$$

$$\gamma(\underset{\sim}{k}) = \gamma(-\underset{\sim}{k}) \tag{B.5}$$

when combining Eqs. (B.3) and (B.4), so that

$$\mathrm{Con}(\underset{\sim}{r}_i) = i\lambda \int \{e^{-i\gamma(\underset{\sim}{k})} \sum_{j=1}^{N_a} |f_j(\theta)| \exp[i\eta_j(\theta) + i\pi z_o^j \theta^2/\lambda - 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j]$$

$$- e^{i\gamma(\underset{\sim}{k})} \sum_{j=1}^{N_a} |f_j(\theta)| \exp[-\eta_j(\theta) - i\pi z_o^j \theta^2/\lambda - 2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j]\} b(\underset{\sim}{k}) e^{2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_i} \, d\underset{\sim}{k} \tag{B.6}$$

The image transform, previously defined in Eq. (3.13) as being the Fourier transform of the image contrast $\mathrm{Con}(r_i)$, is by virtue of Eq. (B.6)

$$j(\underset{\sim}{k}) = 2\lambda b(k) \sum_{j=1}^{N_a} |f_j(\theta)| e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} \sin[\gamma(\underset{\sim}{k}) - \eta_j(\theta) - \pi z_o^j \theta^2/\lambda] \tag{B.7}$$

Several interesting properties of the contrast mechanisms can be inferred from Eq. (B.7) when it is compared to the image transform of a weakly scattering object:

$$j(\underset{\sim}{k}) = b(\underset{\sim}{k})[2 \sin \gamma(\underset{\sim}{k}) \, \Phi(\underset{\sim}{k}) - 2 \cos \gamma(\underset{\sim}{k}) \, M(\underset{\sim}{k})] \tag{B.8}$$

We notice that a three-dimensional arrangement of atoms produces an amplitude contrast image, even when anomalous scattering is neglected[62]. Letting $\eta_j(\theta) = 0$ in Eq. (B.7) yields

$$j(\underset{\sim}{k}) = b(\underset{\sim}{k})\{2 \sin \gamma(\underset{\sim}{k})[\lambda \sum_{j=1}^{N_a} |f_j(\theta)| e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} \cos(\pi z_o^j \theta^2/\lambda)]$$

$$- 2 \cos \gamma(\underset{\sim}{k})[\lambda \sum_{j=1}^{N_a} |f_j(\theta)| e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} \sin(\pi z_o^j \theta^2/\lambda)] \qquad (B.9)$$

The spurious amplitude contrast term vanishes only if the object is centrosymmetric, or equivalently, if the structure factor sampled by the Ewald sphere obeys Friedel's law[14].

Friedel's law is always violated when the anomalous phase shift, $\eta_j(\theta)$, is taken into account. The Fourier transforms of the object phase and amplitude modulation functions are

$$\Phi(\underset{\sim}{k}) = \lambda \sum_{j=1}^{N_a} |f_j(\theta)| \cos[\eta_j(\theta) + \pi z_o^j \theta^2/\lambda] e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j}$$

$$= \lambda \sum_{j=1}^{N_a} f'_{j\ eff}(\theta)\ e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} \qquad (B.10)$$

and

$$M(\underset{\sim}{k}) = \lambda \sum_{j=1}^{N_a} |f_j(\theta)| \sin[\eta_j(\theta) + \pi z_o^j \theta^2/\lambda] e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j}$$

$$= \lambda \sum_{j=1}^{N_a} f''_{j\ eff}(\theta) e^{-2\pi i \underset{\sim}{k}\cdot\underset{\sim}{r}_o^j} \qquad (B.11)$$

We can define the effective real part of the atomic scattering amplitudes as

$$f'_{j\ eff}(\theta) = f'_j(\theta) \cos(\pi z_o^j \theta^2/\lambda) - f''_j(\theta) \sin(\pi z_o^j \theta^2/\lambda) \qquad (B.12)$$

and the imaginary part as[62]

$$f''_{j\ \mathrm{eff}}(\theta) = f''_j(\theta)\cos(\pi z_o^j \theta^2/\lambda) + f'_j(\theta)\sin(\pi z_o^j \theta^2/\lambda) \quad \text{(B.13)}$$

The purpose of object reconstruction schemes is to retrieve the information contained in $\Phi(\underset{\sim}{k})$ and $M(\underset{\sim}{k})$. If thickness effects are included, the simple relationship between these quantities and the complex atomic scattering amplitudes is lost; therefore it is necessary to keep specimens as thin as possible in order to gain full advantage of the restoration technique.

APPENDIX C

Singularities in Schiske's Formula

Singularities occur when the denominator of Eq. (3.97) vanishes, that is, when[59]

$$N^2 - \left| \sum_{\ell=1}^{N} e^{2i\gamma_\ell} \right|^2 = 0 \tag{C.1}$$

Two separate cases will be considered where this happens in practice. As both numerator and denominator go simultaneously to zero, Taylor expansions are used to remove the indeterminacy.

The first singularity occurs near $\theta = 0$, where the phase factors $\gamma_\ell(\theta,\phi)$ approach zero. If we approximate the exponentials by: $e^{i\gamma_\ell(\theta,\phi)} \simeq 1 + i\gamma_\ell(\theta,\phi)$, then Eq. (3.97) for the complex object transform is expressed as

$$O(\underset{\sim}{k}) \underset{\theta\to 0}{\simeq} -\frac{i}{E(\theta)} \frac{\sum_{\ell=1}^{N} j_\ell(\underset{\sim}{k})(1+i\gamma_\ell)\{N - \sum_{m=1}^{N}[1+2i(\gamma_m-\gamma_\ell)]\}}{N^2 - \left|\sum_{\ell=1}^{N}(1+2i\gamma_\ell)\right|^2} \tag{C.2}$$

We notice the following simplifications

$$\sum_{\ell=1}^{N}\sum_{m=1}^{N}(\gamma_m-\gamma_\ell) = \sum_{m=1}^{N}\sum_{\ell=1}^{N}(\gamma_\ell-\gamma_m) = 0$$

$$\left|\sum_{\ell=1}^{N}(1+2i\gamma_\ell)\right|^2 = N^2 + 4\sum_{\ell=1}^{N}\sum_{m=1}^{N}\gamma_\ell\gamma_m + 2i\sum_{\ell=1}^{N}\sum_{m=1}^{N}(\gamma_\ell-\gamma_m) \tag{C.3}$$

$$= 0$$

that reduce Eq. (C.2) to

$$O(\underset{\sim}{k}) \underset{\theta\to 0}{\simeq} -\frac{1}{E(\theta)} \frac{\sum_{\ell=1}^{N} j_\ell(\underset{\sim}{k})(1 + i\gamma_\ell)[\sum_{m=1}^{N} (\gamma_m-\gamma_\ell)]}{2(\sum_{\ell=1}^{N} \gamma_\ell)^2} \tag{C.4}$$

The phase and amplitude contrast transform may now be written, in the limit where $\theta \to 0$, as

$$\Phi(\underset{\sim}{k}) = \tfrac{1}{2}[O(\underset{\sim}{k}) + O^*(-\underset{\sim}{k})] \simeq -\frac{1}{E(\theta)} \frac{\sum_{\ell=1}^{N} j_\ell(\underset{\sim}{k}) \sum_{m=1}^{N} (\gamma_m-\gamma_\ell)}{2(\sum_{\ell=1}^{N} \gamma_\ell)^2} \tag{C.5}$$

and

$$M(\underset{\sim}{k}) = \tfrac{1}{2}[O(\underset{\sim}{k}) - O^*(-\underset{\sim}{k})] \simeq -\frac{1}{E(\theta)} \frac{\sum_{\ell=1}^{N} j_\ell(\underset{\sim}{k})\gamma_\ell \sum_{m=1}^{N} (\gamma_m-\gamma_\ell)}{2(\sum_{\ell=1}^{N} \gamma_\ell)^2} \tag{C.6}$$

It is therefore apparent that $\Phi(\underset{\sim}{k})$ diverges as $1/\theta^2$ when $\theta \to 0$, whereas the cosine transferred part, $M(\underset{\sim}{k})$, containing the anomalous scattering contribution, remains finite. The computer program "Schiske" actually sets the DC component of $O(\underset{\sim}{k})$ equal to the DC component of one of the input pictures, to avoid this artifact.

The second singularity occurs when the difference between successive $\gamma_\ell(\underset{\sim}{k})$ approaches π :

$$\gamma_\ell = \gamma_1 + (\ell-1)\pi + \varepsilon_\ell \tag{C.7}$$

where $|\varepsilon_\ell| \ll 1$. A first-order expansion of the exponentials, $e^{i\varepsilon_\ell} \simeq 1 + i\varepsilon_\ell$, yields the following identities:

$$\sum_{m=1}^{N} e^{2i(\gamma_m-\gamma_\ell)} = \sum_{m=1}^{N} e^{2i(\varepsilon_m-\varepsilon_\ell)} \simeq N + 2i \sum_{m=1}^{N} (\varepsilon_m-\varepsilon_\ell) = N+2i\varepsilon$$

and

$$\left|\sum_{m=1}^{N} e^{2i\gamma_m}\right|^2 = \left|\sum_{m=1}^{N} e^{2i(\gamma_m-\gamma_\ell)}\right|^2 = N^2+ 4\epsilon^2 \qquad (C.8)$$

Substituting these equations into the expression for the complex object transform, Eq. (3.97) leads to

$$O(\underset{\sim}{k}_s) \simeq \frac{1}{E(\theta)} \frac{\sum_{\ell=1}^{N} j_\ell(\underset{\sim}{k}_s)\, e^{i\gamma_\ell}}{2\varepsilon} \qquad (C.9)$$

Equation (C.9) blows up in the frequency zone around the singularity $\underset{\sim}{k}_s$; fortunately this artifact can be avoided by choosing a focus step smaller than the critical value where Eq. (C.7) is satisfied:

$$\Delta Z_{st} \leq \lambda/\theta_{obj}^2 \qquad (C.10)$$

APPENDIX D

Quantitative Assessment of the Schiske Restoration

A quantitative assessment of the reliability of the object reconstruction is obtained by calculating, from the complex object function, $O(\underset{\sim}{k})$, images with corresponding electron microscope parameters, and then comparing them to the original micrographs. In the first iteration, the Fourier transforms of the successive focus series are expressed as

$$D_f^{(\ell)}(\underset{\sim}{k}) = \mathcal{F}\{ \frac{A_{o\ell}^2}{M^2} + \frac{2A_{o\ell}}{M} \mathrm{Re}[\psi_s^{(\ell)}(\underset{\sim}{r}_i)]\}$$

$$= \frac{A_{o\ell}^2}{M} [\delta(\underset{\sim}{k}) + j_\ell(\underset{\sim}{k})]h_N(\underset{\sim}{k}) \qquad (D.1)$$

where the dark field contribution is ignored and $A_{o\ell}$ denotes the primary beam intensity of the ℓ^{th} micrograph. Since the exposures are approximately equal, $A_{o\ell}$ varies little with ℓ . During contrast stretching of the preliminary pictures, the mean value of the respective input images are equalized, so that $[A_{o\ell} = A_o;\ \ell=1,2,\cdots,N]$. The Schiske program then outputs a complex object transform which may be written as

$$O_T(\underset{\sim}{k}) = \frac{A_o^2}{M^2} [\delta(\underset{\sim}{k}) + O(\underset{\sim}{k})]\ h_N(\underset{\sim}{k}) \qquad (D.2)$$

where $h_N(\underset{\sim}{k})$ is the noise filter defined by Eq. (3.125).

From Eq. (B.3), we know that the elastically scattered wave is given by

$$\psi_s^{(\ell)}(r_i) = \frac{A_o}{M} \mathcal{F}^{-1} \{ie^{-i\gamma_\ell(\underset{\sim}{k})} O(\underset{\sim}{k})h_N(\underset{\sim}{k})\} \qquad (D.3)$$

Thus, the reconstructed focus series, calculated from the structure factor, is simply obtained by computing

$$\hat{\psi}_s^{(\ell)}(\underset{\sim}{r}_i) = \mathcal{F}^{-1}\{\hat{D}_f^{(\ell)}(\underset{\sim}{k})\} = \mathrm{Re}\{\mathcal{F}^{-1}[2ie^{-i\gamma_\ell(\underset{\sim}{k})}(O_T(\underset{\sim}{k}) - A_0^2\delta(\underset{\sim}{k})/2M^2)]\} \tag{D.4}$$

for all phase factors $\gamma_\ell(\underset{\sim}{k})$. An inverse fast Fourier transform program takes the real part of the inverse in one of its operating modes, then it converts the data from complex numbers into bytes without scaling the output during the conversion. A direct comparison of the output from Eq. (D.4) with the original pictures, $\mathcal{F}^{-1}\{D_f^{(\ell)}(\underset{\sim}{k})\}$, is therefore possible.

The difference pictures, $\mathcal{F}^{-1}\{D_f^{(\ell)}(\underset{\sim}{k}) - \hat{D}_f^{(\ell)}(\underset{\sim}{k})\}$, are subsequently formed and displayed. If the reconstruction is valid, a listing of the histograms yields narrow Gaussian-like distributions, whose mean square deviations, $\sigma_{\Delta Z_\ell}$, give a reliability measure for each defocusing. The whole reconstruction may therefore be judged by evaluating[62]

$$\sigma_{rm} = \frac{1}{N}\sum_{\ell=1}^{N}\sigma_{\Delta Z_\ell} \tag{D.5}$$

The display of the difference picture is a nice way to see how the deviations from the linear theory are distributed over the whole field of view. We might expect large contributions to the difference image from places where the dark field term, or plural scattering effects, become significant.

In order to calculate the quadratic term, the following relationship is used

$$|\psi_s^{(\ell)}(\underset{\sim}{r}_i)|^2 = \frac{M^2}{A_o^2} |\mathcal{F}^{-1} \{ie^{-i\gamma_\ell(\underset{\sim}{k})}[O_T(\underset{\sim}{k}) - A_o^2\delta(\underset{\sim}{k})/M^2]\}|^2 \qquad (D.6)$$

In practice the complex inverse transform is computed, and its intensity is later computed after appropriate scaling. One must therefore correct for possible scaling distortions before subtracting the result from the original series, $\mathcal{F}^{-1}\{D_f^{(\ell)}(\underset{\sim}{k})\}$. The same procedure, repeated to two iterations, should yield a better agreement in the reliability test.

As an aside, a method of checking the magnitude of the respective dark-field terms is now presented. We know from Eq. (3.46) that the dark-field transform takes a value at the origin equal to

$$D_f^{(\ell)}(0) = \int_{-\infty}^{\infty} O(\underset{\sim}{k})\, O(-\underset{\sim}{k})\, b(\underset{\sim}{k})\, d\underset{\sim}{k} \qquad (D.7)$$

which is independent of $\gamma_\ell(\underset{\sim}{k})$. Consequently, the various dark-field pictures should possess an identical mean optical density.

APPENDIX E

Relation between Fourier Filtering and Convolution Averaging

E.1 Introduction

For every image processing method in reciprocal space there is a counterpart in real space; however, the methods differ to a large extent in the practicability of their implementation. We shall now show that there is an exact relationship between convolution averaging (or linear integration by translational superposition of a repeat unit) and window filtering on the diffraction pattern. The derivation will be carried out on sampled data, where the scanning step has been chosen fine enough not to introduce aliasing errors.

Let $d(\underset{\sim}{r})$ be the continuous optical density distribution with m unit cells in the horizontal direction and n unit cells in the vertical direction of the array; then[114]

$$d(\underset{\sim}{r}) = \sum_{i=1}^{m} \sum_{j=1}^{n} d^{(i,j)} [x - (i-1)\Delta X, y - (j-1)\Delta Y] \qquad (E.1)$$

where

$$d^{(i,j)}(\underset{\sim}{r}) = \begin{cases} d(\underset{\sim}{r}) & \text{if } 0 \leq x \leq \Delta X \text{ and } 0 \leq y \leq \Delta Y \\ 0 & \text{elsewhere} \end{cases} \qquad (E.2)$$

and $d^{(i,j)}(\underset{\sim}{r})$ is the local intensity function characterizing the (i,j) unit cell with lateral dimensions ΔX and ΔY. Local deviations from exact periodicity and the superimposed noise background cause slight variations from one $d^{(i,j)}(\underset{\sim}{r})$ to another.

If the noise is additive to the signal, a linear average of the members of the array will result in an increase in the signal-to-noise

ratio of $\sqrt{N}$, where N is the number of superpositions or, equivalently, the number of unit cells $(N=m \times n)$. The average unit cell is described by the intensity pattern

$$\bar{d}(\underset{\sim}{r}) = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} d^{(i,j)}(\underset{\sim}{r}) \quad \text{if} \quad 0 \leq x \leq \Delta X \quad \text{and} \quad 0 \leq y \leq \Delta Y \tag{E.3}$$

One can then define the deviation of the (i,j)th member of the array from the mean as

$$\varepsilon^{(i,j)}(\underset{\sim}{r}) = d^{(i,j)}(\underset{\sim}{r}) - \bar{d}(\underset{\sim}{r}) \tag{E.4}$$

which satisfies the condition

$$\sum_{i=1}^{m} \sum_{j=1}^{n} \varepsilon^{(i,j)}(\underset{\sim}{r}) = 0 \tag{E.5}$$

The two-dimensional array of unit cells has a density distribution given by

$$d(\underset{\sim}{r}) = \bar{d}(\underset{\sim}{r}) * comb^{\dagger}(\underset{\sim}{r}) + \sum_{i=1}^{m} \sum_{j=1}^{n} \varepsilon^{(i,j)}[x - (i-1)\Delta X, y - (j-1)\Delta Y] \tag{E.6}$$

where $comb^{\dagger}(\underset{\sim}{r})$ stands for a planar arrangement of delta functions sampling the unit cells. We can write $comb^{\dagger}$ explicitly as

$$comb^{\dagger}(\underset{\sim}{r}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \delta[x - (i-1)\Delta X, \; y - (j-1)\Delta Y] \tag{E.7}$$

Since one deals with discrete arrays on the computer, let us define the following symbols:

Δ = sampling interval (assumed equal in both vertical and horizontal directions)

L = total number of pixels along one line or along one column (one can always geometrically distort the input array into a square in order to facilitate the Fourier transformation)

μ = number of sampled points in a unit cell in the horizontal direction

ν = number of sampled points in a unit cell in the vertical direction.

From these definitions we can construct the relationships

$$\begin{aligned} L &= m\mu = n\nu \\ \Delta X &= \mu\Delta \\ \Delta Y &= \nu\Delta \end{aligned} \tag{E.8}$$

Before proceeding to demonstrate mathematically the picture manipulations involved in both processing schemes, let us define the discrete optical density distribution by the array:

$$d_{k,\ell} = d(k\Delta, \ell\Delta) \tag{E.9}$$

E.2 Convolution Averaging

Convolution averaging is exactly analogous to the photographic integration method developed by Markham, where one repeatedly translates the electron micrograph in the unit vector directions, and then makes a linear superposition of the resulting set. One obtains the average unit cell from this operation. A convolution of this average unit cell with the lattice distribution, $\text{comb}^{\dagger}(\underset{\sim}{r})$, yields the average picture

$$d^{cA}(\underset{\sim}{r}) = \overline{d}(\underset{\sim}{r}) * \text{comb}^{\dagger}(\underset{\sim}{r}) \tag{E.10}$$

The continuous distributions in Eq. (E.10), when transposed to their discrete equivalents, take the values

$$\overline{d}_{k,\ell} = \overline{d}_{p+i\mu,q+j\nu} = \begin{cases} \overline{d}_{p,q} & \text{if } i=j=0 \\ 0 & \text{elsewhere} \end{cases} \tag{E.11}$$

and

$$\text{comb}^{\dagger}_{k,\ell} = \delta_{k-i\mu,\ell-j\nu} \tag{E.12}$$

where the following inequalities hold

$$\begin{aligned} 0 \le p \le \mu-1 \quad ; \quad 0 \le q \le \nu-1 \\ 0 \le i \le m-1 \quad ; \quad 0 \le j \le n-1 \end{aligned} \tag{E.13}$$

From Eq. (E.3), one sees that $\overline{d}_{p,q}$ is computed from the sampled array, $d_{k,\ell}$, by the formula

$$\overline{d}_{p,q} = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} d^{(i,j)}_{p,q} = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} d_{p+i\mu,q+j\nu} \tag{E.14}$$

Therefore, at a point (k,ℓ) of the grid, such that k is congruent to $p(\text{mod } \mu)$ and ℓ congruent to q (mod ν), the average picture intensity is expressed, according to Eq. (E.10), as

$$d^{cA}_{k,\ell} = \sum_{\tilde{p}=0}^{\mu-1} \sum_{\tilde{q}=0}^{\nu-1} \overline{d}_{\tilde{p},\tilde{q}}\, \delta_{p-\tilde{p},q-\tilde{q}} = \overline{d}_{p,q} \tag{E.15}$$

E.3 Fourier Filtering

The discrete Fourier transform was defined in Sec. 2.3.5.1B. It was shown that most properties (e.g., the convolution relationship) of

the DFT are in agreement with the corresponding properties of the Fourier integral transform.

If $D^{cA}(\underset{\sim}{k})$, $\overline{D}(\underset{\sim}{k})$, and $Comb^{\dagger}(\underset{\sim}{k})$ designate the respective DFT's of $d^{cA}(\underset{\sim}{r})$, $\overline{d}(\underset{\sim}{r})$, and $comb^{\dagger}(\underset{\sim}{r})$, then Eq. (E.10) becomes

$$D^{cA}(\underset{\sim}{k}) = \overline{D}(\underset{\sim}{k})\ Comb^{\dagger}(\underset{\sim}{k}) \tag{E.16}$$

Let us evaluate these transforms successively so as to be able to design an appropriate Fourier window.

The discrete transform of the average unit cell is described by:

$$\overline{D}_{r,s} = \frac{1}{L} \sum_{k=0}^{L-1} \sum_{\ell=0}^{L-1} \overline{d}_{k,\ell} \exp[-2\pi i(rk + s\ell)/L] \tag{E.17}$$

Then by referring to the properties of $\overline{d}_{k,\ell}$ mentioned in Eq. (E.11), we can write

$$\overline{D}_{r,s} = \frac{1}{L} \sum_{p=0}^{\mu-1} \sum_{q=0}^{\nu-1} \overline{d}_{p,q} \exp[-2\pi i(rp + sq)/L] \tag{E.18}$$

For instance, crossed lattice images of atomic planes would be visualized in the average picture as an almost perfect sinusoidal fringe pattern defined in complex notation by

$$\overline{d}_{p,q} = \exp[2\pi i(\frac{p}{\mu} + \frac{q}{\nu})] \tag{E.19}$$

Consequently, we have

$$\overline{D}_{r,s} = \frac{1}{L} \{\sum_{p=0}^{\mu-1} \exp[-2\pi i(\frac{r}{m} - 1)\frac{p}{\mu}]\}\{\sum_{q=0}^{\nu-1} \exp[-2\pi i(\frac{s}{n} - 1)\frac{q}{\nu}]\}$$

$$= \begin{cases} \frac{\mu\nu}{L} = \frac{L}{mn} & \text{if} \quad (r/m-1) \equiv 0 \pmod{\mu} \\ & \quad\; (s/n-1) \equiv 0 \pmod{\nu} \\ 0 & \text{elsewhere} \end{cases}$$

$$= \begin{cases} \frac{L}{mn} & \text{if} \quad r \equiv m \pmod{L} \\ & \quad\; s \equiv n \pmod{L} \\ 0 & \text{elsewhere} \end{cases} \qquad \text{(E.20)}$$

One therefore expects four diffraction peaks, corresponding to $\overline{D}_{o,o}$, $\overline{D}_{m,o}$, $\overline{D}_{o,n}$, $\overline{D}_{m,n}$ to appear in the transform.

In a similar fashion, the lattice transform may be written, from Eq. (E.12), as

$$\text{Comb}^{\dagger}_{r,s} = \frac{1}{L} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \exp[-2\pi i(ri\mu + sj\nu)/L]$$

$$= \frac{1}{L} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \exp[-2\pi i(\frac{ri}{m} + \frac{sj}{n})]$$

$$= \begin{cases} \frac{mn}{L} & \text{if} \quad r \equiv 0 \pmod{m} \\ & \quad\; s \equiv 0 \pmod{n} \\ 0 & \text{elsewhere} \end{cases} \qquad \text{(E.21)}$$

Thus $D^{cA}_{r,s} = \overline{D}_{r,s}\, \text{Comb}^{\dagger}_{r,s}$ vanishes unless

$$\begin{aligned} r &= um \qquad 0 \le u \le \mu-1 \\ s &= vn \qquad 0 \le v \le \nu-1 \end{aligned} \qquad \text{(E.22)}$$

It remains for us to prove that by selecting only the Fourier components satisfying Eq. (E.22) in $D(\underset{\sim}{k})$, and subsequently inverse transforming,

one indeed retrieves an image identical to $d^{cA}(\underset{\sim}{r})$.

The discrete image transform, $D(\underset{\sim}{k})$, is a sum of the average picture transform and of a term which accounts for: local deviations from periodicity, thermal diffuse scattering, and superimposed background noise. By converting Eq. (E.6) to the case of discrete data, one arrives at

$$d_{k,\ell} = d^{cA}_{k,\ell} + \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \varepsilon^{(i,j)}_{k-i\mu,\ell-j\nu}$$

$$= d^{cA}_{k,\ell} + \varepsilon^{(i,j)}_{p,q} = d^{cA}_{k,\ell} + \varepsilon_{p+i\mu,q+j\nu} \qquad \text{(E.23)}$$

where

$$k \equiv p \pmod{\mu} \qquad 0 \leq p \leq \mu-1$$

$$\ell \equiv q \pmod{\nu} \qquad 0 \leq q \leq \nu-1$$

Since the local deviations of $d(\underset{\sim}{r})$ from the mean $d^{cA}(\underset{\sim}{r})$ obey a condition equivalent to Eq. (E.5), the following relationship holds:

$$\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \varepsilon_{p+i\mu,q+j\nu} = 0 \qquad \text{for all} \quad p,q \qquad \text{(E.24)}$$

Use will be made of this property during the computation of the discrete transform of the image:

$$D_{r,s} = D^{cA}_{r,s} + \frac{1}{L} \sum_{k=0}^{L-1} \sum_{\ell=0}^{L-1} \varepsilon_{k,\ell} \exp[-2\pi i(rk + s\ell)/L]$$

$$= D^{cA}_{r,s} + \frac{1}{L} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \sum_{p=0}^{\mu-1} \sum_{q=0}^{\nu-1} \varepsilon_{p+i\mu,q+j\nu}$$

$$\cdot \exp[-2\pi i(rp + sq + ri\mu + sj\nu)/L] \qquad \text{(E.25)}$$

It now suffices to invert the orders of the summation in Eq. (E.25) to obtain

$$D_{r,s} = D^{cA}_{r,s} + \frac{1}{L}\sum_{p=0}^{\mu-1}\sum_{q=0}^{\nu-1} \exp[-2\pi i(rp+sq)/L]\{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} \varepsilon_{p+i\mu,q+j\nu} \cdot \exp[-2\pi i(\frac{ri}{m} + \frac{sj}{n})]\} \qquad (E.26)$$

A filter which leaves $D^{cA}_{r,s}$ unaltered is readily derived from Eq. (E.22):

$$H_{r,s} = \begin{cases} 1 & \text{if } r = um \quad 0 \le u \le \mu-1 \\ & \quad s = vn \quad 0 \le v \le \nu-1 \\ 0 & \text{elsewhere} \end{cases} \qquad (E.27)$$

By applying such a mask on the diffraction pattern, the transmitted Fourier components simplify to

$$D^{FF}_{r,s} = D_{r,s}\, H_{r,s} = D^{cA}_{r,s} + \frac{1}{L}\sum_{p=0}^{\mu-1}\sum_{q=0}^{\nu-1} \exp[-2\pi i(rp+sq)/L] \cdot \{\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} \varepsilon_{p+i\mu,q+j\nu}\}$$

$$= D^{cA}_{r,s} \qquad (E.28)$$

Consequently, an inverse transformation of the modified transform yields the average picture

$$d^{FF}_{k,\ell} = \frac{1}{L}\sum_{r=0}^{L-1}\sum_{s=0}^{L-1} D_{r,s}\, H_{r,s} \exp[2\pi i(rk+s\ell)/L] = d^{cA}_{k,\ell} \qquad (E.29)$$

By multiplying the picture transform by a window selecting only

successive diffraction maxima, one retrieves in the inverse transformation the picture which would result from a linear superposition of the unit cell intensity distribution.

The Fourier method has the advantage that one may allow perturbations of the periodic structure to contribute to the reconstruction by enlarging the Fourier windows. In practice, one selects only the diffraction spots which emerge from the noise spectrum in order to avoid the occurrence of possible noise artifacts.

APPENDIX F

Calculation of Weak-Beam Images from Imperfect Crystals

F.1 Analytical Foundations

The derivation of the equations describing electron diffraction from imperfect crystals was described in Sec. 4.3.1.2. Recalling that the Darwin coefficient, $\phi_g(\underset{\sim}{r})$, had to vary as slowly as possible over a unit cell, in order to break the summation over the reciprocal lattice vectors, we construct a slightly different wave function[102]

$$\psi(\underset{\sim}{r}) = \sum_g \phi'_g(\underset{\sim}{r}) \exp[2\pi i(\underset{\sim}{K}+\underset{\sim}{g}) + \underset{\sim}{s}_g)\cdot\underset{\sim}{r}] \qquad (F.1)$$

The set of equations obtained by substituting Eq. (F.1) into the Schrödinger equation is, by analogy with Eq. (4.28):

$$[-i\underline{C}\frac{\partial^2}{\partial x^2}+\frac{\partial}{\partial z}+\underline{B}\frac{\partial}{\partial x}]\ \underline{\Phi}'(x,z) = 2\pi i\underline{A}'(x,z)\ \underline{\Phi}'(x,z) \qquad (F.2)$$

where the elements of the matrices are given by

$$\begin{aligned}
[A']_{gh} &= U_{g-h} \exp[-2\pi i(\underset{\sim}{g}-\underset{\sim}{h})\cdot\underset{\sim}{R}] \exp[2\pi i(s_h-s_g)z]/2K \\
[B]_{gh} &= [K_x+g]\ \delta_{g,h}/K \\
[C]_{gh} &= \delta_{g,h}/4\pi K
\end{aligned} \qquad (F.3)$$

The numerical values of the U_g were computed from the polynomial expansions for the scattering amplitudes given by Smith and Burge[115] and were relativistically corrected for 200 keV electrons

(cf. Eq.(4.4)). The effects of absorption were included by allowing the potential field of the crystal to become complex; the imaginary parts of the potential were obtained from the values of $\mathrm{Im}(U_g)/\mathrm{Re}(U_g)$ calculated for liquid nitrogen temperature by Humphreys and Hirsch[111].

By deriving the displacements $\underset{\sim}{R}$ from the isotropic elastic continuum theory, the product $\underset{\sim}{g}\cdot\underset{\sim}{R}$ for the coordinate system shown in Fig. 4-17 is given by

$$\underset{\sim}{g}\cdot\underset{\sim}{R} = \frac{1}{2\pi}\sum_{i=1}^{4}\left[(\underset{\sim}{g}\cdot\underset{\sim}{b}_i)\theta_i + (\underset{\sim}{g}\cdot\underset{\sim}{b}_{ei})\frac{\sin 2\theta_i}{4(1-\nu)}\right] \qquad (F.4)$$

where $\underset{\sim}{b}_i$ is the Burger's vector of one of the partial dislocations and ν is Poisson's ratio. Equation (F.4) is, however, not realistic near the dislocation cores; thus the displacement fields inside a cylinder of radius $2.5\mathring{A}$ were computed by linear interpolation across the discontinuity. Surface effects were not included, since they are negligible for dislocations at depths larger than half an extinction distance from the surface.

F.2 Numerical Scheme

In order to integrate the parabolic partial-differential equations, it was necessary to develop viable numerical methods for this initial value problem. Initially, a Crank-Nicholson finite-difference implicit scheme[116] was tried; however, it was found to converge too slowly for practical purposes. We came to the realization that terms involving partial derivatives in x in Eq. (F.3) could be treated as a perturbation to the corresponding column-approximation equations[117].

The column-approximation equations are:

$$\frac{d}{dz}\,\underline{\Phi}'(x,z) = 2\pi i\,\underline{\underline{A}}'(x,z)\,\underline{\Phi}'(x,z) \qquad (F.5)$$

A combined multi-step algorithm was next developed which proved to be quite practical. In this technique the integration is carried out on a rectangular grid; using a modified Runge-Kutta explicit algorithm[118] that determines the x derivatives of $\phi_g'(x,z)$ iteratively within a single step. At each z these derivatives are evaluated for each column using a fifth-order interpolation formula involving the neighboring columns. To prevent spurious oscillations far away from the defect the two outermost columns are computed in the column approximation, which imposes a zero slope at the boundaries and thus induces a small edge discontinuity. The truncation error is evaluated by repeating the calculations with a different number of columns and new values for Δx and Δz. In addition, the validity of the whole numerical process is verified at selected levels by estimating its convergence in a single Runge-Kutta step[119]. The analytical complexity of the scheme hindered any attempt to find an upper bound for truncation and round-off errors, so that the appropriate step sizes must be matched to the strength of the displacement field. In the vicinity of the defect, Δx and Δz are halved to ensure that the local strain field is "seen" by the traveling electron wave. The phase, amplitude, and intensity distributions of the wave function are listed at various intervals in order to observe the image-contrast depth dependence in different diffracting geometries.

A schematic of the integration procedure is now described for a system of $\hat{N}$ equations, where the subscripts m and n designate respectively the discrete variables x_m and z_m.

(1) Initialize the variables: $\underline{\Phi}_{m,n}$, $\frac{\partial}{\partial x}\underline{\Phi}_{m,n}$, $\frac{\partial^2}{\partial x^2}\underline{\Phi}_{m,n}$, $\underline{\underline{A}}'_{m,n}$, Δx, and Δz.

(2) Compute: $$\underline{K}^{(1)}_{m,n} = \Delta z\left[i\underline{\underline{C}}\,\frac{\partial^2}{\partial x^2}\underline{\Phi}_{m,n} - \underline{\underline{B}}\,\frac{\partial}{\partial x}\underline{\Phi}_{m,n} + 2\pi i\,\underline{\underline{A}}'_{m,n}\,\underline{\Phi}_{m,n}\right]$$

$$\underline{\Phi}^{(1)}_{m,n+\frac{1}{2}} = \underline{\Phi}_{m,n} + \frac{1}{2}\underline{K}^{(1)}_{m,n}$$

$$\underline{\underline{A}}'_{m,n+\frac{1}{2}}$$

(3) Interpolate the x derivatives: $\frac{\partial}{\partial x}\underline{\Phi}^{(1)}_{m,n+\frac{1}{2}}$, $\frac{\partial^2}{\partial x^2}\underline{\Phi}^{(1)}_{m,n+\frac{1}{2}}$

(4) Compute: $$\underline{K}^{(2)}_{m,n} = \Delta z\left[i\underline{\underline{C}}\,\frac{\partial^2}{\partial x^2}\underline{\Phi}^{(1)}_{m,n+\frac{1}{2}} - \underline{\underline{B}}\,\frac{\partial}{\partial x}\underline{\Phi}^{(1)}_{m,n+\frac{1}{2}} + 2\pi i\,\underline{\underline{A}}'_{m,n+\frac{1}{2}}\underline{\Phi}^{(1)}_{m,n+\frac{1}{2}}\right]$$

$$\underline{\Phi}^{(2)}_{m,n+\frac{1}{2}} = \underline{\Phi}_{m,n} + \frac{1}{2}\underline{K}^{(2)}_{m,n}$$

(5) Interpolate the x derivatives: $\frac{\partial}{\partial x}\underline{\Phi}^{(2)}_{m,n+\frac{1}{2}}$, $\frac{\partial^2}{\partial x^2}\underline{\Phi}^{(2)}_{m,n+\frac{1}{2}}$

(6) Calculate: $$\underline{K}^{(3)}_{m,n} = \Delta z\left[i\underline{\underline{C}}\,\frac{\partial^2}{\partial x^2}\underline{\Phi}^{(2)}_{m,n+\frac{1}{2}} - \underline{\underline{B}}\,\frac{\partial}{\partial x}\underline{\Phi}^{(2)}_{m,n+\frac{1}{2}} + 2\pi i\underline{\underline{A}}'_{m,n+\frac{1}{2}}\,\underline{\Phi}^{(2)}_{m,n+\frac{1}{2}}\right]$$

(7) Calculate the "Collatz estimator"[119] at level n:

$$\theta_n = \max_{\substack{m=1,\cdots,M \\ i=1,\cdots,\hat{N}}} \left| \frac{\underline{K}^{(3)}_{m,n} - \underline{K}^{(2)}_{m,n}}{\underline{K}^{(2)}_{m,n} - \underline{K}^{(1)}_{m,n}} \right|$$

(8) If $\theta_n < \theta_n^{min}$ $\Delta x \leftarrow 2\Delta x$

$\Delta z \leftarrow 2\Delta z$ go to step (1)

$M \leftarrow (M+1)/2$

If $\theta_n^{min} < \theta_n < \theta_n^{max}$ proceed with the calculation

If $\theta_n > \theta_n^{max}$ $\Delta x \leftarrow \Delta x/2$

$\Delta z \leftarrow \Delta z/2$ go to step (1)

$M \leftarrow (2M-1)$

(9) Compute: $\underline{\Phi}^{(3)}_{m,n+1} = \underline{\Phi}_{m,n} + \underline{K}^{(3)}_{m,n}$

$\underline{\underline{A}}'_{m,n+1}$

(10) Interpolate the x derivatives: $\frac{\partial}{\partial x} \underline{\Phi}^{(3)}_{m,n+1}$, $\frac{\partial^2}{\partial x^2} \underline{\Phi}^{(3)}_{m,n+1}$

(11) Calculate: $\underline{K}^{(4)}_{m,n} = \Delta z[i\underline{\underline{C}} \frac{\partial^2}{\partial x^2} \underline{\Phi}^{(3)}_{m,n+1} - \underline{\underline{B}} \frac{\partial}{\partial x} \underline{\Phi}^{(3)}_{m,n+1} + 2\pi i \underline{\underline{A}}'_{m,n+1} \underline{\Phi}^{(3)}_{m,n+1}]$

$$\underline{\Phi}_{m,n+1} = \underline{\Phi}_{m,n} + \frac{1}{6}[\underline{K}^{(1)}_{m,n} + 2\underline{K}^{(2)}_{m,n} + 2\underline{K}^{(3)}_{m,n} + \underline{K}^{(4)}_{m,n}]$$

(12) Interpolate the x derivatives: $\frac{\partial}{\partial x} \underline{\Phi}_{m,n+1}$, $\frac{\partial^2}{\partial x^2} \underline{\Phi}_{m,n+1}$

(13) Go to step (1).

$$\theta_n = \max_{\substack{m=1,\cdots,M \\ i=1,\cdots,\hat{N}}} \left| \frac{\underline{K}^{(3)}_{m,n} - \underline{K}^{(2)}_{m,n}}{\underline{K}^{(2)}_{m,n} - \underline{K}^{(1)}_{m,n}} \right|$$

(8) If $\theta_n < \theta_n^{min}$ $\quad \Delta x \leftarrow 2\Delta x$

$\Delta z \leftarrow 2\Delta z$ go to step (1)

$M \leftarrow (M+1)/2$

If $\theta_n^{min} < \theta_n < \theta_n^{max}$ proceed with the calculation

If $\theta_n > \theta_n^{max}$ $\quad \Delta x \leftarrow \Delta x/2$

$\Delta z \leftarrow \Delta z/2$ go to step (1)

$M \leftarrow (2M-1)$

(9) Compute: $\underline{\Phi}^{(3)}_{m,n+1} = \underline{\Phi}_{m,n} + \underline{K}^{(3)}_{m,n}$

$\underline{\underline{A}}'_{m,n+1}$

(10) Interpolate the x derivatives: $\frac{\partial}{\partial x} \underline{\Phi}^{(3)}_{m,n+1}$, $\frac{\partial^2}{\partial x^2} \underline{\Phi}^{(3)}_{m,n+1}$

(11) Calculate: $\underline{K}^{(4)}_{m,n} = \Delta z[i\underline{\underline{C}} \frac{\partial^2}{\partial x^2} \underline{\Phi}^{(3)}_{m,n+1} - \underline{\underline{B}} \frac{\partial}{\partial x} \underline{\Phi}^{(3)}_{m,n+1} + 2\pi i \underline{\underline{A}}'_{m,n+1} \underline{\Phi}^{(3)}_{m,n+1}]$

$$\underline{\Phi}_{m,n+1} = \underline{\Phi}_{m,n} + \frac{1}{6}[\underline{K}^{(1)}_{m,n} + 2\underline{K}^{(2)}_{m,n} + 2\underline{K}^{(3)}_{m,n} + \underline{K}^{(4)}_{m,n}]$$

(12) Interpolate the x derivatives: $\frac{\partial}{\partial x} \underline{\Phi}_{m,n+1}$, $\frac{\partial^2}{\partial x^2} \underline{\Phi}_{m,n+1}$

(13) Go to step (1).

F.3 Fortran Implementation

A list of the program symbols and of their definitions is now presented. The displacement field $\underset{\sim}{R}(x,z)$ is calculated by the function RHO, which outputs the dot product $2\pi \underset{\sim}{g} \cdot \underset{\sim}{R}$ at every point of the mesh. It corresponds in this listing to a dissociated dipole geometry (it suffices therefore to modify RHO in order to compute the contrast from other defects). The parameters defining the defect geometry are illustrated in Fig. 4-17.

PROGRAM SYMBOLS	DEFINITION
PHI0,PHI1,DPHI,DDPHI	Matrices $\underline{\Phi}_{m,n}, \underline{\Phi}_{m,n+1}, \frac{\partial}{\partial x}\underline{\Phi}_{m,n}, \frac{\partial^2}{\partial x^2}\underline{\Phi}_{m,n}$
K1,K2,K3,AA	Matrices $\underline{K}^{(1)}_{m,n}, \underline{K}^{(2)}_{m,n}, \underline{K}^{(3)}_{m,n}, \underline{A}'_{m,n}$
G	Reciprocal vector $\lvert\underset{\sim}{g}\rvert = 1/d_{hk\ell}$ where $d_{hk\ell}$ is the interplanar spacing
FBRAGG	This determines in units of $\lvert\underset{\sim}{g}\rvert$ the intersection of the Ewald sphere with the row of systematics (i.e., = 3 if $3\underset{\sim}{g}$ is excited.)
LAMBDA	Electron wavelength in Å
TH	Foil thickness in Å
DPTH	Depth intervals in Å elapsing between sucsive printings
WDTH	Half the width in Å of the entrance surface (see Fig.4-17)
N	Number of operating beams
M	Number of columns into which the foil is divided

PROGRAM SYMBOLS	DEFINITION
MIN	This determines in units of $\lvert\underset{\sim}{g}\rvert$ the lowest systematic excited (i.e., = -3 if the lowest systematic in the calculation is $-3\underset{\sim}{g}$)
NOUT	Number of steps between successive printings of the intensities
MUPP	Maximum number of columns allowed by the matrix sizes
COL	This logical flag indicates whether or not a column approximation is used.
PRINT	This logical flag controls the printing of θ_n at regular intervals
COLLATZ	If this logical flag is true, θ_n changes the step sizes automatically
U	Complex Fourier potential coefficients U_o, U_g, $\cdots$, $U_{(\hat{N}-1)g}$
BURG	It equals $\lvert\underset{\sim}{g}\cdot\underset{\sim}{b}_i\rvert$ where $\underset{\sim}{b}_i$ is the partial dislocation Burgers vector
POISS	Poisson's ratio of the crystal
DP1,DP2	Depths in Å of the two pairs of partials (see Fig. 4-17)
COR(1),COR(2)	Equal respectively $\Delta 1$ and $\Delta 2$ in Å (see Fig. 4-17)
COR(3)$\cdots$COR(6)	This indicates the sign of the $\underset{\sim}{b}_i$ with respect to x .
COR(7)$\cdots$COR(10)	Products $\underset{\sim}{g}\cdot\underset{\sim}{b}_{ei}$ of the reciprocal vector $\underset{\sim}{g}$ with the edge components of the four partials.
ZONEB,ZONEE	Depths in Å at which one successively halves and doubles the step sizes to account for the rapid strain variations around the defect

PROGRAM SYMBOLS	DEFINITION
VMAMX,VMAN	Upper and lower bounds for θ_n: θ_n^{max} and θ_n^{min}
ZZZ	Depth in $\mathring{A}$ at which the listing starts.

F.4 Fortran Listing

```
      EXTERNAL RHO
      COMPLEX*16 PHI0(6,97),PHI1(6,97)
      COMPLEX DPHI(6,97),DDPHI(6,97),K1(6,97),K2(6,97),K3(6,97)
      COMPLEX AA(6,6),BB(6),C
      COMPLEX BUFF(6)
      REAL VAR(6,97),YNTENS(97),ABSC(97)
      REAL CC(6,6),B(6),SG(6)
      REAL LAMBDA,U(12)
      REAL A(97)
      REAL*8 AMPES(97),PHAES(97)
      INTEGER COUNT
      LOGICAL COL,PRINT,DEFECT
      LOGICAL COLLTZ,PLOTG
      COMMON/DISLOC/DEFECT,BURG,POISS,WDTH,TH,DP1,DP2,COR(10)
      COMMON/TUFF/C,PRINT,MIN1,XS,ZS
      COMMON/ART/M
      COMMON/BOUL/Z
      EQUIVALENCE (U0,U(1))
      EQUIVALENCE (K1(1),AMPES(1)),(K1(100),PHAES(1))
      EQUIVALENCE (K3(1),YNTENS(1)),(K2(1),ABSC(1))
      DATA COLLTZ/.TRUE./,PLOTG/.FALSE./,NOUT/10/
      Z=0.
      INSI=0
      INSA=1
      LAB=1
      PI=3.141592
      VMA=0.
    5 FORMAT(5F10.0)
    6 FORMAT(6F10.0)
    7 FORMAT(6I3)
    8 FORMAT(5L1)
   10 FORMAT(///,10X,'   OPERATING G ',(220)',/,10X,'GMIN = ',I2,/'   BR
     1AGG POSITION ',F6.2,10X,'  WAVELENGTH =',F10.4,/,10X,'THICKNESS O
     2F THE FOIL ',F10.4,' ANGSTROMS')
```

```
 11   FORMAT(///,10X,' NUMBER OF BEAMS =',I2,10X,'NUMBER OF COLUMNS =',I
     13)
 135  FORMAT(1H0,///,20X,'DEPTH =',F10.4,'ANGSTROMS')
 150  FORMAT(/,30X,'INTENSITY OF',I2,'TH BEAM',/)
 151  FORMAT(5X,8F12.5)
 152  FORMAT(/,20X,'AMPLITUDE AND PHASE OF',I2,'TH BEAM',/)
 161  FORMAT(1H0,///,10X,'G,B =',F6.2,/,10X,'POISSON RATIO =',F6.2,/,10X
     1,'DEPTH OF TOP DIPOLE =',F9.1,/,10X,'DEPTH OF BOTTOM DIPOLE =',F9.
     21,/,10X,'DELTA1 =',F8.4,/,10X,'DELTA2 =',F8.4,/,10X,'SIGNE1 =',F6.
     32,F6.2,/,10X,'SIGNE2 =',F6.2,F6.2)
 166  FORMAT(//,20X,'ABNORMAL TERMINATION : IERR =',I2)
 311  FORMAT(///,10X,'  POTENTIAL COEFFICIENTS',/,10X,6(F10.6,2X),/,10X,
     &6(F10.6,2X))
 313  FORMAT(///,10X,' DEVIATIONS PARAMETERS =',/,10X,6(F10.6,2X))
 503  FORMAT(//,'   THE NON COLUMN EQUATIONS ARE USED IN THIS CALCULATIO
     &N ')
 504  FORMAT(//,'   THE COLUMN APPROXIMATION IS USED IN THIS CALCULATION
     &')
      READ(5,6) G,FBRAGG,LAMBDA,TH,DPTH,WDTH
      READ(5,7) N,M,MIN,NOUT,IWEAK,MUPP
      READ(5,8) COL,PRINT,DEFECT,COLLTZ,PLOTG
      NN=2*N
      ZONEB=TH
      ZONEE=TH
      READ(5,6)(U(I),I=1,NN)
      IF(.NOT.DEFECT) GO TO 3
      READ(5,5) BURG,POISS,DP1,DP2,COR
      READ(5,5) ZONEB,ZONEE,VMAMX,VMAMN,ZZZ
      WRITE(6,161) BURG,POISS,DP1,DP2,(COR(J),J=1,6)
      POISS=2.0*(1.0-POISS)
3     WRITE(6,10) MIN,FBRAGG,LAMBDA,TH
      WRITE(6,11) N,M
      WRITE(6,311)(U(I),I=1,NN)
      WVEC=SQRT(U0+1.0/(LAMBDA**2))
      IF(.NOT.DEFECT) GO TO 9
      XS=2.0*WDTH/FLOAT(M-1)
```

```
      ZS=XS
      GO TO 17
    9 XS=0.
      ZS=1.
   17 MIN1=MIN+1
      C=(0.,1.)/(4.*PI*WVEC)
      IF(COL) GO TO 501
      WRITE(6,503)
      GO TO 502
  501 WRITE(6,504)
  502 ILOOP=0
      COUNT=0
      DO 28 J=1,M
      J1=J-1
      X=XS*FLOAT(J1)
   28 A(J)=RHO(Z,X)
      DO 20 I=1,N
      IN=I+N
      BB(I)=(0.,0.)
      BUFF(I)=(0.,0.)
      DO 863 J=1,N
  863 AA(I,J)=(0.,0.)
      U(I)=PI*U(I)/WVEC
      U(IN)=PI*U(IN)/WVEC
      B(I)=G*(FLOAT(I-MIN1)-0.5*FBRAGG)/WVEC
      IF(I.EQ.MIN1) GO TO 21
      DO 22 J=1,M
      PHII(I,J)=(0.,0.)
   22 PHIO(I,J)=(0.,0.)
      GO TO 20
   21 DO 23 J=1,M
      PHII(I,J)=(1.,0.)
   23 PHIO(I,J)=(1.,0.)
   20 CONTINUE
      DO 31 I=1,N
      DO 32 J=1,MUPP
```

```
      DPHI(I,J)=(0.,0.)
   32 DDPHI(I,J)=(0.,0.)
   31 CONTINUE
      XCENT=FBRAGG/2.*G
      ZCENT=SQRT(WVEC**2-XCENT**2)
      DO 12 I=1,N
   12 SG(I)=-ZCENT+SQRT(WVEC**2-(G*FLOAT(I-MIN1)-XCENT)**2)
      WRITE(6,313)(SG(I),I=1,N)
      DO 14 I=1,N
      DO 15 K=1,N
   15 CC(I,K)=(SG(I)-SG(K))*2.*PI
   14 CONTINUE
C BEGIN ITERATION OF RUNGE KUTTA SCHEME
C**************************************
  100 COUNT=COUNT+1
      IF(COUNT.GT.10) COUNT=1
      ZZ=Z+0.5*ZS
      M2=(M+1)/2
      CALL RUNGE(PHI0,PHI1,BUFF,DPHI,DDPHI,K1,A,AA,BB,CC,U,B,Z,0.5,N,MUP
     &P)
C END OF FIRST STEP
C ******************
      IF(COL) GO TO 45
      CALL DET5(XS,PHI1,DPHI,DDPHI,N,MUPP)
   45 CALL TUNGE(PHI0,PHI1,BUFF,DPHI,DDPHI,K1,K2,K2,A,AA,BB,CC,U,B,ZZ,2.
     &RHO,N,MUPP)
C END OF SECOND STEP
C *******************
      IF(COL) GO TO 55
      CALL DET5(XS,PHI1,DPHI,DDPHI,N,MUPP)
   55 CALL RUNGE(PHI0,PHI1,BUFF,DPHI,DDPHI,K3,A,AA,BB,CC,U,B,ZZ,1.0,N,MU
     &PP)
      IF((Z.LT.ZONEB).OR.(Z.GT.ZONEE)) GO TO 51
      INSA=0
      INSI=INSI+1
      IF(INSI.EQ.1) GO TO 52
```

```
      GO TO 70
   51 INSA=INSA+1
      IF(INSA.EQ.1) GO TO 361
      IF((COUNT.NE.1).OR.(.NOT.COLLTZ)) GO TO 70
      DO 96 I=1,N
      DO 96 J=1,M
   96 VAR(I,J)=CABS((K3(I,J)-K2(I,J))/(K2(I,J)-K1(I,J)))
      CALL VALMAX(VAR,VMA,N,MUPP)
      IF(VMA.LT.VMAMX) GO TO 61
      IF(ILOOP.EQ.1) GO TO 70
C HALVING THE STEP SIZE
C **********************
   52 IF(COL) GO TO 81
      M=2*M-1
      IF(M.GT.MUPP) GO TO 998
      ZS=ZS/2.0
      XS=XS/2.0
      DO 62 J=1,M
      J1=J+1
      J4=J-1
      X=XS*FLOAT(J4)
      A(J)=RHO(Z,X)
      IF(2*(J1/2)-J1)64,63,64
   63 J2=J1/2
      DO 65 I=1,N
      PHI0(I,J)=PHI0(I,J2)
   65 PHI1(I,J)=PHI0(I,J)
      GO TO 62
   64 J2=(J1-1)/2
      J3=J2+1
      DO 66 I=1,N
      PHI0(I,J)=0.5*(PHI0(I,J2)+PHI0(I,J3))
   66 PHI1(I,J)=PHI0(I,J)
   62 CONTINUE
      COUNT=0
      ILOOP=1
```

```
      CALL DET5(XS,PHI0,DPHI,DDPHI,N,MUPP)
      GO TO 100
   81 ZS=ZS/2.0
      DO 82 J=1,M
      J4=J-1
      X=XS*FLOAT(J4)
      A(J)=RHO(Z,X)
      DO 83 I=1,N
   83 PHI1(I,J)=PHI0(I,J)
   82 CONTINUE
      COUNT=0
      ILOOP=1
      GO TO 100
61    IF(VMA.GT.VMAMN) GO TO 70
      IF(ILOOP.EQ.1) GO TO 70
C DOUBLING THE STEP SIZE
C **********************
  361 IF(COL) GO TO 85
      M=(M+1)/2
      ZS=2.0*ZS
      XS=2.0*XS
      DO 72 J=1,M
      J1=2*J-1
      J2=J-1
      X=XS*FLOAT(J2)
      A(J)=RHO(Z,X)
      DO 72 I=1,N
      PHI0(I,J)=PHI0(I,J1)
      PHI1(I,J)=PHI0(I,J)
   72 CONTINUE
      COUNT=0
      CALL DET5(XS,PHI0,DPHI,DDPHI,N,MUPP)
      IF(INSA.EQ.1) GO TO 100
      ILOOP=1
      GO TO 100
   85 ZS=2.0*ZS
```

```
      DO 84 J=1,M
      J2=J-1
      X=XS*FLOAT(J2)
      A(J)=RHO(Z,X)
      DO 86 I=1,N
   86 PHI1(I,J)=PHI0(I,J)
   84 CONTINUE
      COUNT=0
      IF(INSA.EQ.1) GO TO 100
      ILOOP=1
      GO TO 100
C START THE THIRD STEP
C ********************
   70 Z=Z+ZS
      ZZZ=ZZZ-ZS
      ILOOP=0
      IF(COL) GO TO 75
      CALL DETS(XS,PHI1,DPHI,DDPHI,N,MUPP)
   75 CALL TUNGE(PHI0,PHI1,BUFF,DPHI,DDPHI,K1,K2,K3,A,AA,BB,CC,U,B,Z,1,R
     &HO,N,MUPP)
C END OF ONE RUNGE KUTTA STEP
C ***************************
      IF((-ZZZ.LT.DPTH).AND.(Z.LT.TH)) GO TO 100
      IF((COUNT/NOUT*NOUT).NE.COUNT) GO TO 144
      WRITE(6,135) Z
      DO 223 I=1,N
      DO 114 J=1,M
  114 YNTENS(J)=REAL(PHI1(I,J))**2+AIMAG(PHI1(I,J))**2
      WRITE(6,150)I
      WRITE(6,151)(YNTENS(J),J=1,M)
  223 CONTINUE
  144 IF((.NOT.DEFECT).OR.(.NOT.PLOTG)) GO TO 725
      ISYS=0
      DO 105 J=1,M
  105 ABSC(J)=FLOAT(-M2+J)*XS
      WRITE(6,135) Z
      DO 110 I=1,N
```

```
      DO 111 J=1,M
      AMPES(J)=REAL(PHI1(I,J))
      PHAES(J)=AIMAG(PHI1(I,J))
  111 YNTENS(J)=AMPES(J)**2+PHAES(J)**2
      WRITE(6,152) I
      WRITE(6,151)(AMPES(J),J=1,M)
      WRITE(6,151)(PHAES(J),J=1,M)
  110 CONTINUE
  725 ZZZ=ZZZ+DPTH
      IF(Z.LT.TH) GO TO 100
      GO TO 840
  998 IERR=1
      WRITE(6,146) IERR
      GO TO 840
  840 STOP
      END


      SUBROUTINE VALMAX(X,VMA,N,MUPP)
      COMMON/BOUL/Z
      COMMON/TUFF/C,PRINT
      COMMON/ART/M
      COMPLEX C
      DIMENSION X(N,MUPP)
      LOGICAL PRINT
      VMA=0.
      DO 10 K=1,N
      DO 10 I=1,M
      PIX=X(K,I)
      IF(PIX.LE.VMA) GO TO 10
      VMA=PIX
      IN=K
      IM=I
   10 CONTINUE
      IF(.NOT.PRINT) GO TO 11
```

```
      WRITE(6,95) IN,IM,VMA
      WRITE(6,230) Z
   95 FORMAT(1H0,'  THE MAXIMUM COLLATZ DEVIATION CORRESPONDS TO N = ',I
     12,'TH BEAM AND',I3,'COLUMNS ACROSS',/,10X,'ITS VALUE IS :',F10.4)
  230 FORMAT(///,20X,'THE DEPTH OF THE CALCULATION IS :',F10.4)
   11 RETURN
      END

      SUBROUTINE RUNGE(T,T1,BUFF,V,Q,S,A,AA,BB,CC,U,B,Z,STP,N,MUPP)
      COMMON/TUFF/C,PRINT,MIN1,XS,ZS
      COMMON/ART/M
      COMPLEX*16 T(N,MUPP),T1(N,MUPP)
      COMPLEX V(N,MUPP),Q(N,MUPP),S(N,MUPP)
      COMPLEX AA(N,N)
      COMPLEX C
      COMPLEX BUFF(1)
      COMPLEX BB(1)
      DIMENSION B(1),U(1),A(1)
      REAL CC(N,N)
      LOGICAL PRINT
      DO 30 J=1,M
      DO 34 I=1,N
      BUFF(I)=T1(I,J)
      DO 32 K=1,N
      JR=I-K+1
      IF(JR.LE.0) JR=K-I+1
      JRN=JR+N
      R=Z*CC(K,I)+A(J)*FLOAT(K-I)
      CR=COS(R)
      SR=SIN(R)
   32 AA(K,I)=CMPLX(-U(JRN)*CR+U(JR)*SR,U(JR)*CR+U(JRN)*SR)
   34 CONTINUE
      CALL GMPRD(AA,BUFF,BB,N)
      DO 35 I=1,N
      S(I,J)=ZS*(C*Q(I,J)-B(I)*V(I,J)+BB(I))
   35 T1(I,J)=T(I,J)+STP*S(I,J)
```

```
 30   CONTINUE
      RETURN
      END

      SUBROUTINE TUNGE(T,T1,BUFF,V,Q,S,W,W1,A,AA,BB,CC,U,B,Z,IND,RHO,N,M
     &UPP)
      COMMON/TUFF/C,PRINT,MIN1,XS,ZS
      COMMON/DISLOC/DEFECT,BURG,POISS,WDTH,TH,DP1,DP2,COR(10)
      COMMON/ART/M
      COMPLEX*16 T(N,MUPP),T1(N,MUPP)
      COMPLEX V(N,MUPP),Q(N,MUPP),S(N,MUPP),W(N,MUPP),W1(N,MUPP)
      COMPLEX AA(N,N)
      COMPLEX BUFF(1)
      COMPLEX BB(1)
      COMPLEX C
      DIMENSION B(1),U(1),A(1)
      REAL CC(N,N)
      LOGICAL DEFECT,PRINT
      DO 30 J=1,M
      J1=J-1
      X=XS*FLOAT(J1)
      A(J)=RHO(Z,X)
      DO 34 I=1,N
      BUFF(I)=T1(I,J)
      DO 32 K=1,N
      JR=I-K+1
      IF(JR.LE.0) JR=K-I+1
      JRN=JR+N
      R=Z*CC(K,I)+A(J)*FLOAT(K-I)
      CR=COS(R)
      SR=SIN(R)
 32   AA(K,I)=CMPLX(-U(JRN)*CR+U(JR)*SR,U(JR)*CR+U(JRN)*SR)
 34   CONTINUE
      CALL GMPRD(AA,BUFF,BB,N)
      GO TO (60,61),IND
 60   DO 50 I=1,N
```

```
      T1(I,J)=T(I,J)+ZS/6.0*(C*Q(I,J)-B(I)*V(I,J)+BB(I))+S(I,J)/6.0+(W(I
     1,J)+W1(I,J))/3.0
   50 T(I,J)=T1(I,J)
      GO TO 30
   61 DO 51 I=1,N
      W(I,J)=ZS*(C*Q(I,J)-B(I)*V(I,J)+BB(I))
   51 T1(I,J)=T(I,J)+0.5*W(I,J)
   30 CONTINUE
      RETURN
      END

      SUBROUTINE DET5(H,Y,Z,W,N,MUPP)
      COMMON/ART/M
      COMPLEX*16 Y(N,MUPP)
      COMPLEX Z(N,MUPP),W(N,MUPP)
      COMPLEX A,B,C
      HH=0.08333333/H
      DO 18 J=1,N
      B=(0.,0.)
      C=HH*(-3.*Y(J,1)-10.*Y(J,2)+18.*Y(J,3)-6.*Y(J,4)+Y(J,5))
      DO 3 I=5,M
      A=B
      B=C
      C=HH*(Y(J,I-4)-Y(J,I)+8.*(Y(J,I-1)-Y(J,I-3)))
    3 Z(J,I-4)=A
      A=HH*(-Y(J,M-4)+6.*Y(J,M-3)-18.*Y(J,M-2)+10.*Y(J,M-1)+3.*Y(J,M))
      Z(J,M)=(0.,0.)
      Z(J,M-1)=A
      Z(J,M-2)=C
      Z(J,M-3)=B
      B=(0.,0.)
      C=HH*(-3.*Z(J,1)-10.*Z(J,2)+18.*Z(J,3)-6.*Z(J,4)+Z(J,5))
      DO 5 I=5,M
      A=B
      B=C
      C=HH*(Z(J,I-4)-Z(J,I-1)+8.*(Z(J,I-1)-Z(J,I-3)))
```

```
5     W(J,I-4)=A
      A=HH*(-Z(J,M-4)+6.*Z(J,M-3)-18.*Z(J,M-2)+10.*Z(J,M-1)+3.*Z(J,M))
      W(J,M)=(0.,0.)
      W(J,M-1)=A
      W(J,M-2)=C
      W(J,M-3)=B
18    CONTINUE
      RETURN
      END

      SUBROUTINE CMPRD(A,B,R,N)
      COMPLEX A(N,N)
      COMPLEX R(1)
      COMPLEX B(1)
      DO 10 J=1,N
      R(J)=(0.,0.)
      DO 11 K=1,N
11    R(J)=R(J)+A(J,K)*B(K)
10    CONTINUE
      RETURN
      END

      FUNCTION RHO(Z,X)
      COMMON/DISLOC/DEFECT,BURG,POISS,WDTH,TH,DP1,DP2,COR(10)
      LOGICAL DEFECT
      PI=3.141592
      IF(DEFECT) GO TO 101
      RHO=0.
      RETURN
101   X=X-WDTH
      DEL1=COR(1)
      DEL2=COR(2)
      SN1=COR(3)
      SN2=COR(4)
      SN3=COR(5)
      SN4=COR(6)
```

```
      ED1=COR(7)
      ED2=COR(8)
      ED3=COR(9)
      ED4=COR(10)
      ZR=-Z+DP2
      ZZR=-Z+DP1
      XR1=X-DEL1
      XR2=X-DEL2
      XXR1=X+DEL1
      XXR2=X+DEL2
      QR=ZR**2
      QQR=ZZR**2
      IF((ABS(ZR).LT.5.0).OR.(ABS(ZZR).LT.5.0)) GO TO 1
   10 IF(XR1.NE.0.) GO TO 11
      ARCT1=SIGN(PI/2.,ZR)
      GO TO 12
   11 ARCT1=ATAN2(ZR,XR1)
   12 IF(XR2.NE.0.) GO TO 14
      ARCT2=SIGN(PI/2.,ZR)
      GO TO 13
   14 ARCT2=ATAN2(ZR,XR2)
   13 IF(XXR1.NE.0.) GO TO 15
      ARCT3=SIGN(PI/2.,ZZR)
      GO TO 16
   15 ARCT3=ATAN2(ZZR,XXR1)
   16 IF(XXR2.NE.0.) GO TO 17
      ARCT4=SIGN(PI/2.,ZZR)
      GO TO 18
   17 ARCT4=ATAN2(ZZR,XXR2)
   18 R1=SN1*BURG*(ARCT1+(XR1*ZR)/(POISS*(QR+XR1**2))*ED1)
      R2=SN2*BURG*(ARCT2+(XR2*ZR)/(POISS*(QR+XR2**2))*ED2)
      R3=SN3*BURG*(ARCT3+(XXR1*ZZR)/(POISS*(QQR+XXR1**2))*ED3)
      R4=SN4*BURG*(ARCT4+(XXR2*ZZR)/(POISS*(QQR+XXR2**2))*ED4)
      RHO=R1+R2+R3+R4
      RETURN
    1 IF(ABS(ZR).GE.5.0) GO TO 2
```

```
      PR1=QR+XR1**2
      PR2=QR+XR2**2
      IF(PR1.LT.25.) GO TO 3
      IF(PR2.LT.25.) GO TO 4
    2 PPR1=QQR+XXR1**2
      PPR2=QQR+XXR2**2
      IF(PPR1.LT.25.) GO TO 5
      IF(PPR2.LT.25.) GO TO 6
      GO TO 10
    3 IND=1
      QR=XR1
      XR1=SQRT(25.-QR)
      ARCT1=ATAN2(ZR,XR1)
      ARCTT=PI-ARCT1
      W=SN1*BURG*(ARCT1+XR1*ZR/(POISS*(QR+XR1**2))*ED1)
      W1=SN1*BURG*(ARCTT-XR1*ZR/(POISS*(QR+XR1**2))*ED1)
      R1=W+(W1-W)*(XR1-QR)/(2.*XR1)
      GO TO 30
    4 IND=2
      QR=XR2
      XR2=SQRT(25.-QR)
      ARCT2=ATAN2(ZR,XR2)
      ARCTT=PI-ARCT2
      W=SN2*BURG*(ARCT2+XR2*ZR/(POISS*(QR+XR2**2))*ED2)
      W1=SN2*BURG*(ARCTT-XR2*ZR/(POISS*(QR+XR2**2))*ED2)
      R2=W+(W1-W)*(XR2-QR)/(2.*XR2)
      GO TO 30
    5 IND=3
      QR=XXR1
      XXR1=SQRT(25.-QQR)
      ARCT3=ATAN2(ZZR,XXR1)
      ARCTT=PI-ARCT3
      W=SN3*BURG*(ARCT3+XXR1*ZZR/(POISS*(QQR+XXR1**2))*ED3)
      W1=SN3*BURG*(ARCTT-XXR1*ZZR/(POISS*(QQR+XXR1**2))*ED3)
      R3=W+(W1-W)*(XXR1-QR)/(2.*XXR1)
      GO TO 30
```

```
 6  IND=4
    QR=XXR2
    XXR2=SQRT(25.-QQR)
    ARCT4=ATAN2(ZZR,XXR2)
    ARCTT=PI-ARCT4
    W=SN4*BURG*(ARCT4+XXR2*ZZR/(POISS*(QQR+XXR2**2))*ED4)
    W1=SN4*BURG*(ARCTT-XXR2*ZZR/(POISS*(QQR+XXR2**2))*ED4)
    R4=W+(W1-W)*(XXR2-QR)/(2.*XXR2)
30  IF(IND.EQ.1) GO TO 31
    ARCT1=ATAN2(ZR,XR1)
    R1=SN1*BURG*(ARCT1+XR1*ZR/(POISS*(QR+XR1**2))*ED1)
31  IF(IND.EQ.2) GO TO 32
    ARCT2=ATAN2(ZR,XR2)
    R2=SN2*BURG*(ARCT2+XR2*ZR/(POISS*(QR+XR2**2))*ED2)
32  IF(IND.EQ.3) GO TO 33
    ARCT3=ATAN2(ZZR,XXR1)
    R3=SN3*BURG*(ARCT3+XXR1*ZZR/(POISS*(QQR+XXR1**2))*ED3)
33  IF(IND.EQ.4) GO TO 34
    ARCT4=ATAN2(ZZR,XXR2)
    R4=SN4*BURG*(ARCT4+XXR2*ZZR/(POISS*(QQR+XXR2**2))*ED4)
34  RHO=R1+R2+R3+R4
    IND=0
    RETURN
    END
```

REFERENCES

1. RUSKA, E., "Past and present attempts to attain the resolution limit of the transmission electron microscope", Adv. Opt. Electron Microscopy 1, 116-179 (1966).

2. ZEITLER, E., "Resolution in electron microscopy", Adv. Electron. Electron Phys. 25, 277-332 (1968).

3. GALLAGER, R. G., Information Theory and Reliable Communication, John Wiley & Sons, Inc., 1968.

4. VALENTINE, R. C., "The response of photographic emulsions to electrons", Adv. Opt. Electron Microscopy 1, 180-203 (1966).

5. COWLEY, J. M., and MOODIE, A. F., "The scattering of electrons by atoms and crystals. I. A new theoretical approach", Acta Cryst. 10, 609-619 (1957).

6. LENZ, F., "The influence of lens imperfections on image formation", Lab. Invest. 14, 808-818 (1965).

7. HEIDENREICH, R. D., Fundamentals of Transmission Electron Microscopy, Wiley Interscience, Vol. XIII, New York, 1964.

8. ZEITLER, E., and OLSEN, H., "Complex scattering amplitudes in elastic electron scattering", Phys. Rev. 162, 1439-1447 (1967).

9. BORN, M., and WOLF, E., Principles of Optics, Pergamon Press, 1970.

10. SCHERZER, O., "The theoretical resolution limit of the electron microscope", J. Appl. Phys. 21, 20-28 (1948).

11. EISENHANDLER, C. B., and SIEGEL, B. M., "Imaging of single atoms with the electron microscope by phase contrast", J. Appl. Phys. 37, 1613-1620 (1966).

12. HEIDENREICH, R. D., "Electron phase contrast images of molecular detail", Siemens Review XXXIV, Special issue, "X-Ray and Electron Microscopy News", 4-13 (1967).

13. REIMER, L., "Elektronenoptischer Phasenkontrast I. Ansatz für eine quantitative Theorie", Z. Naturforschg. 21a, 1489-1499 (1966).

14. HOPPE, W., "Principles of electron structure research at atomic resolution using conventional electron microscopes for the measurement of amplitudes and phases", Acta Cryst. A26, 414-425 (1970).

15. REIMER, L., "Elektronenoptischer Phasenkontrast II. Berechnung mit komplexen Atomstreuamplituden für Atome und Atomgruppen", Z. Naturforschg. 24a, 377-389 (1969).

16. NIEHRS, H., "Optimale Abbildungsbedingungen und Bildintensitäts verlauf bei einer Elektronenmikroskopie auf Atomen. I", Optik 30, 273-293 (1969).

17. LENZ, F., "Zur Streuung mittelschneller Elektronen in kleinste Winkel", Z. Naturforsch. 9a, 185-204 (1954).

18. MESSIAH, A., Quantum Mechanics, John Wiley & Sons, Inc., 1966.

19. UYEDA, R., "A theory on image formation of electron microscope, J. Phys. Soc. Japan 10, 256-264 (1955).

20. RIECKE, W. D., "Prospects for high resolution microscopy", Phil. Trans. Roy. Soc. Lond. B 261, 15-34 (1971).

21. HANSZEN, K. J., "The optical transfer theory of the electron microscope: fundamental principles and applications", Adv. Opt. Electron Microscopy 4, 1-84 (1971).

22. COWLEY, J. M., and IJIMA, S., "Electron microscope image contrast for thin crystals", Z. Naturforsch. 27a, 445-451 (1972).

23. GRINTON, G. R., and COWLEY, J. M., "Phase and amplitude contrast in electron micrographs of biological material", Optik 34, 221-233 (1971).

24. MISELL, D. L., "Image formation in the electron microscope. I. The application of transfer theory to a consideration of elastic electron scattering", J. Phys. A: Gen. Phys. 4, 782-797 (1971).

25. ANDERSEN, W.H.J., and MOL, A., "Simultaneous measurements of the brightness and the energy distribution of electrons emitted from a triode gun", J. Phys. D: Appl. Phys. 3, 965-979 (1970).

26. BENDAT, J., Principles and Applications of Random Noise Theory, John Wiley & Sons, Inc., 1958.

27. BILLINGSLEY, F. C., "Image processing for electron microscopy: II. A digital system", Adv. Opt. Electron Microscopy 4, 127-159 (1971).

28. MENDELSOHN, M. L., MAYALL, B. H., PREWITT, J.M.S., BOSTROM, R.C., and HOLCOMB, W. G., "Digital transformation and computer analysis of microscopic images", Adv. Opt. Electron Microscopy 2, 77-150 (1968).

29. GOODMAN, J. W., Introduction to Fourier Optics, McGraw-Hill, Inc., 1968.

30. GOLDMAN, S., Information Theory, Dover, Inc., New York, 1968.

31. BILLINGSLEY, F. C., "Applications of digital image processing", Appl. Opt. 9, 289-299 (1970).

32. NATHAN, R., "Image processing for electron microscopy: I. Enhancement procedures", Adv. Opt. Electron Microscopy 4, 85-126 (1971).

33. HUANG, T. S., SCHREIBER, W. F., and TRETIAK, O. J., "Image processing", Proc. IEEE 59, 1586-1609 (1971).

34. SELZER, R., "Improving biomedical image quality with computers", JPL Technical Report 32-1336 (1968).

35. ROSENFELD, A., Picture Processing by Computer, Academic Press, New York, 1969.

36. GRAHAM, R. E., "Snow removal--A noise stripping process for picture signals", IRE Trans. Inform. Theory IT-18, 129-144 (1962).

37. THIRY, H., "Some qualitative and quantitative results on spatial filtering of granularity", Appl. Opt. 3, 39-43 (1964).

38. ROETLING, P. G., "Effects of signal-dependent granularity, J. Opt. Soc. Amer. 55, 67-71 (1965).

39. WIENER, N., Extrapolation, Interpolation and Smoothing of Stationary Time Series, MIT Press, Cambridge, 1970.

40. RÖHLER, R., Informationstheorie in der Optik, Optik und Feinmechanik in Einzeldarstellungen, Band 6, ed. N. Günther, Wiss. Verlagsgesellshaft, Stuttgart, 1967.

41. WELTON, T. A., "Electron-optical factors limiting resolution in transmission electron microscopes", in Proc. Workshop Conference on Microscopy of Cluster Nuclei in Defected Crystals, Chalk River Nuclear Laboratories, CRNL-622-1, pp. 125-157 (Sept. 8-10, 1971).

42. ANDREWS, H. Computer Techniques in Image Processing, Academic Press, New York, 1970.

43. PAPOULIS, A., Systems and Transforms with Application in Optics, McGraw-Hill, Inc., 1968.

44. BINGHAM, C., GODFREY, M. D., and TUKEY, J. W., "Modern techniques of power spectrum estimation", IEEE Trans. Audio Electroacous. AU-15, 56-66 (1967).

45. COOLEY, J. M., LEWIS, P. L., and WELCH, P. D., "Application of the fast Fourier transform to computation of Fourier integrals, Fourier series, and convolution integrals", IEEE Trans. Audio Electroacous. AU-15, 79-84 (1967).

46. LANGER, R., FRANK, J., FELTYNOWSKI, A., and HOPPE, W., "Anwendung des Bilddifferenzverfahrens auf die Untersuchung von Strukturänderungen dünner Kohlefohlien bei Elektronenbestrahlung", Ber. Bunsenges. physik. Chem. 74, 1120-1126 (1970).

47. HARRIS, J. L., "Image evaluation and restoration", J. Opt. Soc. Amer. 56, 569-574 (1966).

48. RUSHFORTH, C. K., and HARRIS, R. W., "Restoration, resolution and noise", J. Opt. Soc. Amer. 58, 539-545 (1968).

49. SLEPIAN, D., "Linear least-squares filtering of distorted images", J. Opt. Soc. Amer. 57, 918-922 (1967).

50. HALL, E. L., KRUGER, R. P., DWYER, III, S. J., HALL, D. L., McLAREN, R. W., and LODWICK, G. S., "A survey of preprocessing and feature extraction techniques for radiographic images", IEEE Trans. Computers C-20, 1032-1044 (1971).

51. LOHMANN, A. W., and PARIS, D. P., "Computer generated spatial filters for coherent optical data processing", Appl. Opt. 7, 651-655 (1968).

52. IINUMA, T. A., and NAGAI, T., "Image restoration in radioisotope imaging systems", Phys. Med. Biol. 12, 501-509 (1967).

53. GABOR, D., "Information theory in electron microscopy", Lab. Invest. 14, 63-69 (1965).

54. LEVI, L., "On image evaluation and enhancement", Optica Acta 17, 59-76 (1970).

55. LINFOOT, E. H., "Transmission factors and optical design", J. Opt. Soc. Amer. 46, 740-752 (1956).

56. TORALDO DI FRANCIA, G., "Resolving power and information", J. Opt. Soc. Amer. 45, 497-501 (1955).

57. SCHISKE, P., "Zur Frage der Bildrekonstruktion durch Focusreihen", Proc. Fourth Eur. Reg. Conf. Electron Microscopy, Rome, Vol. 1, 145-146 (1968).

58. DOWNING, K. H., and SIEGEL, B. M., "Object wave determination by single-sideband methods", Proc. 31th Ann. Meeting EMSA, 266-267, New Orleans (1973).

59. FRANK, J., "A study on heavy/light atom discrimination in bright-field electron microscopy using the computer", Biophys. J. 12, 484-511 (1972).

60. ERICKSON, H. P., "The Fourier transform of an electron micrograph--First order and second order theory of image formation", Adv. Opt. Electron Microscopy 5, 1-51 (1972).

61. ERICKSON, H. P., and KLUG, A., "Measurement and compensation of defocusing and aberrations by Fourier processing of electron micrographs", Phil. Trans. Roy. Soc. Lond. B 261, 105-118 (1971).

62. FRANK, J., "Use of anomalous scattering for element discrimination", Image Processing and Computer-Aided Design in Electron Optics, P. W. Hawkes, ed., Academic Press, 1973, pp. 196-211.

63. MISELL, D. L., and CRICK, R. A., "Inelastic electron scattering in carbon and carbonaceous materials and chromatic aberration in electron microscopy", J. Phys. C: Solid St. Phys., Ser. 2 2, 2290-2296 (1969).

64. MISELL, D. L., and ATKINS, A. J., "Image resolution and image contrast in the electron microscope III. Inelastic scattering and coherent illumination", J. Phys. A:Math, Nucl. Gen. 6, 218-235 (1973).

65. THON, F., "Elektronenmikroskopische Untersuchungen an dünnen Kohlefolien", Z. Naturforsch. 20a, 154-155 (1965).

66. CHILDS, P. A., and MISELL, D. L., "Some aspects of the elastic plural scattering of electrons by atoms", J. Phys. D: Appl. Phys. 5, 2095-2104 (1972).

67. MISELL, D. L., "On the electron microscopy of biological objects", J. Phys. D: Appl. Phys. 6, L1-L5 (1973).

68. MISELL, D. L., "Image resolution and image contrast in the electron microscope. II. Elastic scattering and incoherent illumination", J. Phys. A: Math., Nucl. Gen. 6, 205-217 (1973).

69. HANSZEN, K. J., and TREPTE, L., "Die Kontrastübertragung im Elektronenmikroskop bei partiell kohärenter Beleuchtung. Teil B: Ausgedehnte scheibenförmige Strahlquelle und Zweiseitenbandübertragung", Optik 33, 182-198 (1971).

70. HANSZEN, K. J., and TREPTE, L., "Der Einfluss von Strom und Spannungsschwankungen sowie der Energiebreite der Strahlelektronen auf Kontrastübertragung und Auflösung des Elektronenmikroskops", Optik 32, 519-538 (1971).

71. FRANK, J., "The envelope of electron microscopic transfer functions for partially coherent illumination", Optik 38, 519-536 (1973).

72. CREWE, A. V., "High resolution scanning microscopy of biological specimens", Phil. Trans. Roy. Soc. Lond. B 261, 61-70 (1971).

73. HOPPE, W., "Zur Abbildung komplexer Bildfunktionen in der Elektronenmikroscopie", Z. Naturforsch 26a, 1155-1168 (1971).

74. HOPPE, W., "Recording, processing, and correction of electron microscope images", Proc. 5th Europ. Congr. El. Micr., Manchester, 612-617 (1972).

75. FRANK, J., "Nachweis von Objektbewegungen im Lichtoptischen Diffraktogramm von elektronenmikroskopischen Aufnahmen", in Optik 30, 171-180 (1969).

76. MISELL, D. L., and CHILDS, P. A., "Deconvolution in two dimensions with particular reference to the electron microscope", J. Phys. D: Appl. Phys. 5, 1760-1768 (1972).

77. FRANK, J., BUSSLER, P., LANGER, R., and HOPPE, W., "Einige Erfahrungen mit der rechnerischen Analyse und Synthese von elektronenmikroskopischen Bildern hoher Auflösung", Ber. Bunsenges. Phys. Chem. 74, 1105-1115 (1970).

78. REIMER, L., HEINE, H. G., and AJEIAN, R., "Optimal bedingungen für den Beugungsnachweis von Defokussierungsstrukturen in elektronenmikroskopischen Aufnahmen", Z. Naturforsch. 24a, 1846-1848 (1969).

79. HAASE, J., "Zusammenstellung der Koeffizienten für die Anpassung komplexer Streufaktoren für schnelle Elektronen durch Polynome. 2 Mitteilung: Thomas-Fermi-Dirac-Fall", Z. Naturforsch. 25a, 1219-1235 (1970).

80. FRANK, J., "Two dimensional correlation functions in electron microscope image analysis", Proc. 5th Europ. Congr. El. Micr., Manchester, 622-623 (1972).

81. FRANK, J., "Observation of the relative phases of electron microscopic phase contrast zones with the aid of the optical diffractometer", Optik 35, 608-612 (1972).

82. PARSONS, D. F., "Environmental wet cells for biological medium voltage and high voltage microscopy", Proc. 31st Ann. EMSA Meeting, New Orleans, pp. 14-15 (1973).

83. WRIGLEY, N. G., "The lattice spacing of crystalline catalase as an internal standard of length in electron microscopy", J. Ultrastruct. Res. 24, 454-464 (1968).

84. VALENTINE, R. C., "Sub-units of the catalase molecule seen by electron microscopy", Nature 204, 1262-1264 (1964).

85. LONGLEY, W., "The crystal structure of bovine liver catalase: a combined study by X-ray diffraction and electron microscopy", J. Mol. Biol. 30, 323-327 (1967).

86. DeROSIER, D. J., and MOORE, P. B., "Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry", J. Mol. Biol. 52, 355-369 (1970).

87. GERCHBERG, R. W., "Holography without fringes in the electron microscope", Nature 20, 404-406 (1972).

88. GLAESER, R. M., "Radiation damage and high resolution biological electron microscopy", Proc. 31st Ann. EMSA Meeting, New Orleans, pp. 226-227 (1973).

89. PARSONS, D. F., "Radiation damage in biological materials", Proc. 31st Ann. EMSA Meeting, New Orleans, pp. 482-483 (1973).

90. LANGMORE, J., WALL, J., ISAACSON, M., and CREWE, A. V., "Carbon support films for high resolution electron microscopy", Proc. 31st Ann. EMSA Meeting, New Orleans, pp. 76-77 (1973).

91. ARNAL, H., and HORGEN, H. M., "A simple technique for the preparation of ultrathin carbon films for electron microscope", AEC Report, CALT-767-P3-33 (1974).

92. NATHAN, R., "Computer enhancement of electron micrographs", Proc. 28th Ann. EMSA Meeting, Houston, pp. 28-29 (1970).

93. MARKHAM, R., HITCHBORN, J. H., HILLS, G. J., and FREY, S., "The anatomy of the tobacco mosaic virus", Virology 22, 342-359 (1964).

94. LIPSON, H., Optical Transforms, Academic Press, New York, 1972.

95. HORGEN, H. M., VILLAGRANA, R. E., and MAHER, D. M., "Computer enenhancement of high resolution electron micrographs containing weak periodic information", Proc. Fifth Eur. Congr. El. Micr., Manchester, pp. 582-583 (1972).

96. HIRSCH, P. B., HOWIE, A., NICHOLSON, R. B., PASHLEY, D. W., and WHELAN, M. J., Electron Microscopy of Thin Crystals, Butterworths, London, 1969.

97. HORGEN, H. M., VILLAGRANA, R. E., and MAHER, D. M., "Comments on the processing of periodic and near-periodic images by computer," Proc. Eighth Int. Cong. El. Micr., Canberra, 1974.

98. COCKAYNE, D.J.H., PARSONS, J. R., and HOELKE, C. W., "A study of the relationship between lattice fringes and lattice planes in electron microscope images of crystals containing defects", Phil. Mag. 20, 139-153 (1971).

99. LEWIS, A. L., and VILLAGRANA, R. E., "Perturbation treatment of high energy electron diffraction from imperfect crystals", Phys. Rev. B 6, 4382-4392 (1972).

100. YOSHIOKA, H., "Effect of inelastic waves on electron diffraction", J. Phys. Soc. Japan 12, 618-628 (1957).

101. METHERELL, A.J.F., and VILLAGRANA, R. E., "A perturbation treatment of absorption in electron diffraction", submitted to Acta Cryst.

102. HOWIE, A. and BASINKI, Z. S., "Approximations of the dynamical theory of diffraction contrast", Phil. Mag. 17, 1039-1063 (1968).

103. COCKAYNE, D.J.H., "A theoretical analysis of the weak-beam method of electron microscopy", Z. Naturforsch. 27a, 452-460 (1972).

104. HOWIE, A., and SWORN, C. H., "Column approximation effects in high resolution electron microscopy using weak diffracted beams", Phil. Mag. 22, 861-864 (1970).

105. TAKAGI, S., "Dynamical theory of diffraction applicable to crystals with any kind of small distortion", Acta Cryst. 15, 1311-1312 (1962).

106. COCKAYNE, D.J.H., RAY, I.L.F., and WHELAN, J. J., "Investigation of dislocation strain fields using weak beams," Phil. Mag. 20, 1265-1270 (1969).

107. HIRTH, J. P., and LOTHE, J., Theory of Dislocations, McGraw-Hill 1968.

108. RAY, I.L.F., and COCKAYNE, D.J.H., "The dissociation of dislocations in silicon," Proc. R. Soc. Lond. A 325, 543-554 (1971).

109. CASS, T. R., "Dislocation dipoles in deformed magnesium oxide", University of California UCRL-11996, Berkeley (1965).

110. ASHBY, M. F., and BROWN, L. M., "Diffraction contrast from inclusions", Phil. Mag. 8, 1649-1672 (1963).

111. HUMPHREYS, C. J., and HIRSCH, P. B., "Absorption parameters in electron diffraction theory", Phil. Mag. 18, 115-122 (1968).

112. HORGEN, H. M., VILLAGRANA, R. E., and MAHER, D. M., "Computer enhancement of weak-beam images", Proc. 31st Ann. EMSA Meeting, New-Orleans, pp. 274-275 (1973).

113. PERRIN, R. C., and SAVINO, E. J., "Computer simulation of weak beam images of extended dislocation in copper," T.P.507, A.E.R.E. Harwell, Berks, 1972.

114. MOORE, P. B., and DeROSIER, D. J., "Three-dimensional reconstruction of F-actin. Appendix: deconvolution", J. Mol. Biol. 50, 293-295 (1970).

115. SMITH, G. H., and BURGE, R. E., "The analytical representation of atomic scattering amplitudes for electrons", Acta Cryst. 15, 182-185 (1962).

116. CRANK, J., and NICHOLSON, P., "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conducting type", Proc. Cambridge Phil. Soc. 43, 50-67 (1947).

117. HAMMOND, R. B., "The effects of several scattering approximations on the calculation of weak-beam contrast for high-resolution mic-

roscopy", Bell Telephone Techn. Rep. (1972).

118. ISAACSON, E., and KELLER, H. B., Analysis of Numerical Methods John Wiley & Sons, Inc., 1966.

119. COLLATZ, L., The Numerical Treatment of Differential Equations, Springer-Verlag, Berlin, 1960.