Computer Simulation of High Resolution Transmission Electron Micrographs: Theory and Analysis

Roar Kilaas

National Center for Electron Microscopy
Materials and Molecular Research Division
Lawrence Berkeley Laboratory
and
Department of Physics
University of California
Berkeley, CA 94720

LBL--19058

DE85 013927

Abstract

Computer simulation of electron micrographs is an invaluable aid in their proper interpretation and in defining optimum conditions for obtaining images experimentally. Since modern instruments are capable of atomic resolution, simulation techniques employing high precision are required. This thesis makes contributions to four specific areas of this field.

First, the validity of a new method for simulating high resolution electron microscope images has been critically examined. This method, which has been termed the real space method (RSP) since the entire calculation is performed without any Fourier transforms, offers a considerable reduction in computing time over the conventional multislice approach when identical sampling conditions are employed. However, for the same level of accuracy the real space method requires more sampling points and more computing time than the conventional multislice method. These characteristics are illustrated with calculated results using both methods to identify practical limitations.

Second, three different methods for computing scattering amplitudes in High Resolution Transmission Electron Microscopy (HRTEM) have been investigated as to their ability to include upper Laue layer (ULL) interaction. The conventional first order multislice method using fast Fourier transform (FFT) and the second order multislice (SOM) method are shown to yield calculated intensities of first order Laue reflections with the use of slice thicknesses smaller than the crystal periodicity along the incident electron beam direction. It is argued that

the calculated intensities of ULL reflections approach the correct values in the limiting case of vanishing slice thickness and electron wavelength. The third method, the improved phasegrating method (IPG) does also in principle include ULL effects, but is severely limited as to choice of slice thickness and sampling interval. A practical way to use slice thicknesses less than the crystal periodicity along the incident beam direction is shown for both the conventional FFT method and the second order multislice method and tested on a spinel structure. It is also shown that the IPG method does not easily allow for a slice thickness different from the crystal periodicity in the beam direction.

Third, a new method for computing scattering amplitudes in high resolution transmission electron microscopy has been examined. The method which is called the Improved Phasegrating (IPG) method is shown to produce reasonable results only for very small specimen thicknesses and diverges for thicknesses larger than 20 Å in [001] copper for accelerating voltages between 200kV and 1MV. The validity of the method is discussed and is shown to depend on electron wavelength, slice thickness,the number of reflections that are included in the calculation and the choice of specimen. It is also shown that the method does not readily allow for slice thicknesses smaller than the specimen periodicity along the incident electron beam direction.

Fourth, the effect of a surface layer of amorphous silicon dioxide on images of crystalline silicon has been investigated for a range of crystal thicknesses varying from zero to $2\frac{1}{2}$ times that of the surface layer. It is shown that an amorphous surface produces fluctuations in image contrast which introduces difficulties in the interpretation of defects in very thin specimens. These difficulties are less pronounced but still present in thicker crystals. It is also shown that an edge smoothly approaching zero thickness produces an image that changes gradually from crystalline to amorphous character.

Table of Contents

Chapter 1

Introduction

The current generation of electron microscopes makes possible the attainment of near atomic resolution in images of crystals [1]. However as the resolution of transmission electron microscopes increases so does the need for reliable computer simulation of electron micrographs [2]. The image contrast in high resolution electron micrographs varies rapidly with objective lens defocus and specimen thickness. Only for very thin specimens, typically less than 30–40 Å and for selective settings of defocus does the image contrast bear a one to one correspondence to the specimen structure [3–6]. To provide proper interpretation of the image it is usually necessary to match experimental images to computed images for a range of defocus values. Only then can one assume that the different parameters that enter the calculation, in particular the specimen thickness, are correct.

At lower resolution (typically greater than 20 Å), the observed image contrast is well understood and specific rules apply to the interpretation of these images. Thus there are rules that determine whether a stacking fault is intrinsic or extrinsic [7] and rules that allow the determination of dislocation Burgers vectors [8]. The image contrast in this regime is usually referred to as "amplitude contrast," and the image is formed by ONE beam of electrons, that being either the forward scattered beam or one scattered through a specific angle, normally a Bragg scattering angle in the specimen. Thus the variation of contrast in the image arises due to local variation of scattering amplitudes across the area of the specimen under observation.

In high resolution microscopy the image contrast is referred to as "phase contrast," that is, two or more diffracted beams recombine in the image plane to give essentially an interference pattern. The periodicities that are present in the image are determined by the total number of reflections that combine to form the image and to what degree the magnetic lenses in the microscope are able to preserve the phase relationships among scattered electrons. The

highest resolution microscope presently available, the Atomic Resolution Microscope at Lawrence Berkeley Laboratory can be tuned to introduce approximately a constant phase shift to scattered electrons associated with a shift in wave vectors of magnitude less than 0.63 $Å^{-1}$, corresponding to a point to point resolution of 1.6 Å. It is possible to get information beyond this limit [9], usually referred to as the Scherzer limit [10] or the structure image resolution limit, by proper adjustment of the defocus of the objective lens to reach the linear image resolution limit [11] or the information retrieval limit which is set by the objective lens chromatic aberration and voltage and current instabilities [12].

The simulation of an image can be considered to consist of three parts. The first is the calculation of the effective potential seen by the electron as it moves through the specimen. In principle the potential should include all scattering processes, both elastic and inelastic. However, in practice only elastic scattering is usually included. Complex lattice potentials accounting for certain inelastic scattering mechanisms, notably plasmon, phonon and core excitations, have been calculated [13-14], but in the rare cases that inelastic scattering is considered, the complex part of the potential is varied to provide the best match. The elastic part of the potential is computed from electron scattering factors which are calculated using either relativistic Hatree-Fock atomic wave functions [15] or relativistic Dirac-Slater wave functions [16], the former being the most common. The electron scattering factors can also be computed from experimental x-ray structure factors.

The second part of the calculation involves the actual propagation of the electron through the specimen. This means finding the electron distribution as a function of specimen thickness. The imaging electrons are assumed to be incoherent, the final image is a sum of one-electron images, and the calculation involves solving the one-electron Schrödinger equation where the relativistic nature of the electrons at accelerating voltages of several hundred thousand volts is incorporated in the relativistic mass. Because of the difficulty in solving Schrödinger's equation, several approximate solutions exist, each with its own regime of validity and advantages and disadvantages.

The third part of the calculation treats the interaction between the electrons and the imaging system of the microscope. The most important lens determining the resolution of the microscope is the objective lens, and the parameters that are considered at this stage include:

- Spherical abberation in the objective lens.
- Chromatic abberation in the objective lens.
- Any objective lens aperture.
- Objective lens defocus.
- Temporal and spatial incoherence of the imaging electrons.
- Any current and voltage instability of the microscope.

The part of computation that treats the interaction between the specimen and the electron beam are mostly based upon a dynamical multislice formulation proposed by Cowley and Moodie using physical optics [17]. The connection to quantum mechanics, notably the Schrödinger equation was provided by Van Dyck [18] and independently by Jap and Glaeser [19]. These methods are referred to as multislice methods because the specimen is divided into thin slices each normally having the same slice thickness Δz. The electron wave function is calculated at each slice from the wave function at the previous slice; beginning with the known electron distribution at the top of the specimen.

This thesis consists of four independent research projects, each designed to further knowledge in the field of computer simulation of high resolution electron micrographs. Chapter 2, 3 and 4 discuss different multislice methods for computing scattering amplitudes in High Resolution Transmission Electron Microscopy (HRTEM), while Chapter 5 shows how computer simulation can be applied to answer important questions regarding the interpretation of high resolution images.

In some sense the work presented in this thesis can be regarded as experiments carried out on an experimental system that took several years to build. Before any of this work could be undertaken, it was necessary to write a few thousands of line of computer code to implement the latest in the theory of image simulation. With no resident expert on image simulation in HRTEM and no previous existing software at LBL this was a long and often frustrating task. However, it all paid off and there are now programs residing on discs under PSS library MSLICE at LBL that will compute scattering factors, perform multislicing, simulate the effect of the microscope lenses, plot images as gray level plots, plot projected slice potentials, compute and plot diffraction patterns, plot amplitudes and phases of selected beams as a function of thickness and plot contrast transfer functions. These programs and some that have not been mentioned were all necessary in order to carry out the work that is presented in the following chapters.

References

[1] R. Gronsky, in 38th Ann. Proc. Electron Microscopy Sc. Amer., San Francisco, CA., 1980, G.W. Bailey (ed), p.2.

[2] W.O. Saxton, Advances in Electronics and Electron Physics, Suppl. 10, Academic press Inc., New York, L. Marton (ed.), p X1.

[3] J.C.H. Spence, Experimental High-Resolution Electron Microscopy, Clarendon Press, Oxford, p. 86.

[4] D.F. Lynch and M.A. O'Keefe, Acta Cryst. A28, (1972), p. 536

[5] G.R. Anstis, D.F. Lynch, A.F. Moodie and M.A. O'Keefe, Acta Cryst. A29, (1973), p. 138

[6] D.F. Lynch, A.F. Moodie and M.A. O'Keefe, Acta Cryst. A31, (1975), p. 300.

[7] G. Thomas and M.J. Goringe, Transmission Electron Microscopy of Materials, John Wiley & Sons, (1979), p. 268.

[8] G. Thomas and M.J. Goringe, Transmission Electron Microscopy of Materials, John Wiley & Sons, (1979), p. 280.

[9] J.L. Hutchison, Ultramicroscopy, 9, (1982), p. 191.

[10] O. Scherzer, Journal of Applied Physics, 20, (1949), p. 20.

[11] M.A. O'Keefe, Private Communication.

[12] J. Frank, Optik, 38, (1973), p.519.

[13] G. Radi Acta Cryst. A26, (1970), p. 41.

[14] Y. Kamiya, Y Nakai and E. Masuda, Acta Cryst. A37, (1981), p. 485.

[15] P.A. Doyle and P.S. Turner, Acta Cryst. A24, (1968), p. 390.

[16] D.T. Cramer and J.T. Waber, Acta Cryst. 18, (1965), p. 104.

[17] J.M. Cowley and A.F. Moodie, Acta Cryst. 10, (1957), p.609.

[18] D. Van Dyck, Acta Cryst. A34, (1978), p. 94.

[19] B. Jap and R. Glaeser, Acta Cryst. A34, (1978), p. 112.

Chapter 2

Real Space Image Simulation in High Resolution Electron Microscopy

Abstract

The validity of a new method for simulating high resolution electron microscope images has been critically examined. This method, which has been termed the real space method (RSP) since the entire calculation is performed without any Fourier transforms, offers a considerable reduction in computing time over the conventional multislice approach when identical sampling conditions are employed. However, for the same level of accuracy the real space method requires more sampling points and more computing time than the conventional multislice method. These characteristics are illustrated with calculated results using both methods to identify practical limitations.

1. Introduction

The current generation of electron microscopes makes possible the attainment of near-atomic resolution [1] in images of crystals; nevertheless there is still an urgent need for reliable computer simulation of these images in order that they might be correctly interpreted [2]. From a pragmatic point of view the full potential of computer simulation is realized only in an on-site, real-time system which affords immediate comparison between computed and experimental results. This in turn requires the development of more rapid and more accurate algorithms.

Most image calculation programs are based upon a dynamical multislice formulation originally proposed by Cowley and Moodie [3] using physical optics. Quantum mechanical arguments were later provided by Van Dyck [4] and independently by Jap and Glaeser [5]. The original calculation time of these early programs is proportional to N^2, N being the number of dynamical reflections included in the calculation. However, using fast Fourier transforms (FFT), Ishizuka and Uyeda [6] demonstrated that the time becomes proportional to N LogN.

More recently, a method derived by Van Dyck [7] promises to further reduce the calculation time such that it is directly proportional to N alone. This method, which treats the interaction between the electron beam and the specimen, will be referred to as the real space method since the entire calculation is done in real space without the use of Fourier transforms.

This paper examines the domain of validity of the real space method and presents a comparison between the real space method and the conventional multislice method. Specific emphasis is placed on the number of dynamical reflections that must be included, maximum slice thickness and calculation time.

2. Theory

2.1 General theory

The geometry of the problem is outlined in Fig. 2-1. An electron with wave-vector $\mathbf{k}_0$ is incident upon a thin region described by a potential U(r) and for simplicity the electron is assumed to be traveling in the z-direction. In the actual computation the specimen is represented by a sandwich of successive slices, each slice having a thickness ϵ.

The electron wave function $\Psi(\mathbf{r})$ is a solution to Schrödinger's equation

$$\left[-\frac{\mathbf{h}^2}{8\pi^2 \mathbf{m}} \nabla^2 - \mathbf{eU}(\mathbf{r}) \right] \Psi(\mathbf{r}) = \frac{\mathbf{h}^2 \mathbf{k}_0^2}{2\mathbf{m}} \Psi(\mathbf{r}) \ . \tag{2.1.1}$$

For high energy electrons $\mathbf{eU}(\mathbf{r}) \ll \frac{\mathbf{h}^2 \mathbf{k}_0^2}{2\mathbf{m}}$, therefore the potential U can be considered a perturbation and the total wave function can be written as a modulated wave function of the form

$$\Psi(\mathbf{r}) = \phi(\mathbf{r})\, e^{2\pi i \mathbf{k}_0 \cdot \mathbf{r}} = \phi(\mathbf{r}) e^{2\pi i k_0 z} \ . \tag{2.1.2}$$

By inserting the above expression into (2.1.1), one has

$$\left[\nabla^2 + 4\pi i k_0 \frac{\partial}{\partial z} + \frac{8\pi^2 \mathbf{me}}{\mathbf{h}^2} \mathbf{U}(\mathbf{r}) \right] \phi(\mathbf{r}) = 0 \ . \tag{2.1.3}$$

By involving the definition

$$V(r) = \frac{8\pi^2 me}{h^2} U(r) \tag{2.1.4}$$

equation (2.1.3) becomes

$$[\nabla^2 + 4\pi i k_0 \frac{\partial}{\partial z} + V(r)]\, \phi(r) = 0 \ . \tag{2.1.5}$$

At this point it is customary to ignore the second derivative with respect to z in (2.1.5) by making the assumption that $\phi(r)$ is a slowly varying function with respect to z, such that

$$\left|\frac{\partial^2 \phi}{\partial z^2}\right| \ll k_0 \left|\frac{\partial \phi}{\partial z}\right| \tag{2.1.6}$$

This amounts to ignoring backscattered electrons and a slight change in the electron wavevector as the electron traverses the potential. A more complete discussion has been given by Van Dyck [8].

Ignoring the second order derivative transforms (2.1.5) into a first order differential equation in z,

$$\frac{\partial}{\partial z} \phi(r) = \frac{i}{4\pi k_0} [\nabla_\perp^2 + V(r)]\, \phi(r) \ . \tag{2.1.7}$$

Formally the solution to (2.1.7) can be written

$$\phi(x,y,\epsilon) = e^{\frac{i}{4\pi k_0}[\epsilon \nabla_\perp^2 + \int_0^\epsilon V(r)dz]} \phi(x,y,0) \ .$$

Note that $e^{A}\Psi$ is defined through the power series

$$e^{A}\Psi = \sum_{n=0}^{\infty} (\frac{A^n}{n!})\, \Psi \ . \tag{2.1.8}$$

Defining $\Delta = \frac{i}{4\pi k_0} \nabla_\perp^2$ and $V_p = \frac{1}{\epsilon} \frac{i}{4\pi k_0} \int_0^\epsilon V(r)dz$

it follows that

$$\phi(x,y,\epsilon) = e^{\epsilon[\Delta + V_p]} \phi(x,y,0) \ . \tag{2.1.9}$$

2.2 Analytical Solutions to (2.1.9)

Unfortunately there are no closed analytical solutions to (2.1.9), requiring the use of various approximate solutions.

i) General Multislice

In the formation of the general multislice method one writes

$$\phi(x,y,\epsilon) = e^{\epsilon[\Delta+V_p]}\,\phi(x,y,0) \approx e^{\epsilon\Delta}e^{\epsilon V_p}\,\phi(x,y,0)\ . \tag{2.2.1}$$

Since Δ and V_p do not commute, (2.2.1) is correct to first order in ϵ, Δ and V_p with a resulting error of the order of $\frac{1}{2}[\Delta,V_p]\epsilon^2$ where [ ] denotes commutation. The advantage to using equation (2.1.1) is that it becomes possible to write down analytical solutions to the expressions

$$f_1(x,y,\epsilon) = e^{\epsilon V_p}\,f_1\,(x,y,0) \tag{2.2.2}$$

$$f_2(x,y,\epsilon) = e^{\epsilon\Delta}\,f_2\,(x,y,0) \tag{2.2.3}$$

of the form

$$f_1(x,y,\epsilon) = \exp\{\frac{i}{4\pi k_0}\int_0^{\epsilon} V\,(x,y,z')dz'\}f_1\,(x,y,0) \tag{2.2.4}$$

$$f_2(x,y,\epsilon) = \frac{k_0}{i\epsilon}\iint dx'dy'f_2(x',y',0)\exp\{\frac{i\pi k_0}{\epsilon}[(x-x')^2+(y-y')^2]\}\ . \tag{2.2.5}$$

Defining the functions h and g through the expressions

$$f_1(\epsilon) = h\cdot f_1\,(0)$$

$$f_2(\epsilon) = g*f_2\,(0)$$

equation (2.2.1) can be written in alternative form:

$$\phi(\epsilon) = g*[h\cdot\phi(0)]\ . \tag{2.2.6}$$

Since the specimen possesses a periodic potential of period ϵ in the z-direction over its total thickness $N\epsilon$, it is necessary to use N successive applications of (2.2.6). The quickest way to numerically solve the equation

$$\phi(N\epsilon) = g*[h\cdot[g*[h\cdot[g*[\ \ldots\ldots\ \phi(0)]\ \cdots\ ] \tag{2.2.7}$$

is to use Fourier transforms as shown by Ishizuka and Uyeda [6] who utilized the algorithm shown in Fig. 2-1. In the remaining part of this paper the general multislice method will be referred to as the FFT method.

ii) Real Space Method

A different approach to finding an approximate solution to (2.1.9) is to expand the exponential in powers of Δ and V_p (Van Dyck [7]), and construct a solution that can be written as a product of functions $f_{i,\epsilon}(\Delta)\, g_{i,\epsilon}(V_p)$ that when expanded in powers of Δ and V_p corresponds to the expansion of (2.1.9) to any desired order of Δ and V_p. A unique solution correct to second order in Δ and V_p was found to be

$$\phi(\epsilon) = \exp\{\tfrac{1}{2}\epsilon\,(1+\delta)V_p\}\,\exp\{\epsilon\Delta\}\,\exp\{\tfrac{1}{2}\epsilon\,(1-\delta)V_p\}\,\phi(0) \qquad (2.2.8)$$

where

$$\delta \equiv \frac{\overline{z}(x,y)-\epsilon/2}{\epsilon/2}\,,\quad \overline{z} \equiv \frac{\int_0^{\epsilon} z'V(x,y,z')dz'}{\epsilon V_p}\,. \qquad (2.2.9),(2.2.10)$$

The parameter $\delta(x,y)$ is a measure of potential eccentricity and is zero for $\overline{z} = \epsilon/2$. The major difference between the real space (RSP) method and the FFT method is that the RSP method uses an expansion of the propagator, $\exp\{\epsilon\Delta\}$, keeping only terms up to second order in ϵ. The argument is that (2.2.8) itself is correct only to second order in ϵ, and no accuracy should be lost by using an expansion of the propagator. The expression for the propagator thus becomes

$$\exp\{\epsilon\Delta\} \approx 1 + \epsilon\Delta + \tfrac{1}{2}\epsilon^2\Delta^2 = 1 + \frac{i\lambda\epsilon}{4\pi}\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right] - \frac{\lambda^2\epsilon^2}{32\pi^2}\left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right]^2 . \qquad (2.2.11)$$

Numerically one solves the equation

$$f(x,y,\epsilon) = \exp\{\epsilon\Delta\}\, f(x,y,0) \approx (1 + \epsilon\Delta + \tfrac{1}{2}\epsilon^2\Delta^2)\, f(x,y,0) \qquad (2.2.12)$$

by dividing up the x and y axes in intervals of δ and η respectively. Thus

$$f(x,y,\epsilon) \approx f(x,y,0) + \frac{i\lambda\epsilon}{4\pi}\{\frac{1}{\delta^2}[f(x+\delta,y,0) + f(x-\delta,y,0) - 2f(x,y,0)]$$

$$+\frac{1}{\eta^2}[f(x,y+\eta,0) + f(x,y-\eta,0) - 2f(x,y,0)]\}$$

$$-\frac{\lambda^2\epsilon^2}{32\pi^2}\{\frac{1}{\delta^4}[f(x+2\delta,y,0) + f(x-2\delta,y,0)$$

$$-4f(x+\delta,y,0) - 4f(x-\delta,y,0) - 6f(x,y,0)]\} +$$

$$+\frac{1}{\eta^4}[f(x,y+2\eta,0) + f(x,y-2\eta,0) - 4f(x,y+\eta,0)$$

$$-4f(x,y-\eta,0) - 6f(x,y,0)]$$

$$+\frac{2}{\eta^2\delta^2}[f(x+\delta,y+\eta,0) + f(x+\delta,y-\eta,0) + f(x-\delta,y+\eta,0)$$

$$+f(x-\delta,y-\eta,0)$$

$$-2f(x+\delta,y,0) - 2f(x-\delta,y,0) - 2f(x,y+\eta,0) - 2f(x,y-\eta,0)$$

$$-4f(x,y,0)] \quad . \tag{2.2.13}$$

The computation time for the RSP method becomes proportional to N, the number of sampling points, while it is proportional to NLogN for the FFT method. Another advantage to a real space approach is that it can eliminate the need to use periodic extension when simulating images from faulted crystals.

In the case of a potential having a mirror plane at $z = \epsilon/2$, one obtains

$$\phi(N\epsilon) = e^{\frac{1}{2}\epsilon V_p}(1+\epsilon\Delta+\tfrac{1}{2}\epsilon^2\Delta^2)e^{\epsilon V_p}(1+\epsilon\Delta+\tfrac{1}{2}\epsilon^2\Delta^2)e^{\epsilon V_p}\ldots.$$

$$\ldots(1+\epsilon\Delta+\tfrac{1}{2}\epsilon^2\Delta^2)e^{\frac{1}{2}\epsilon V_p}\phi(0) \quad . \tag{2.2.14}$$

Thus the general multislice calculation becomes accurate to second order for this particular case by simply beginning and ending with half a phasegrating.

2.3 The validity of the Real Space Method

Compared to the FFT method, the usefulness of the RSP method depends on the effect of throwing away terms of order (ϵ^3, Δ^3) in (2.2.11). The error depends on the slice thickness ϵ as well as the magnitude of the derivatives. By studying the effect of the operator exp $\{\epsilon\Delta\}$ on the function $\phi(x,y,0)$, one notices that in reciprocal space the effect is that of a pure phase-operator.

$$\Phi(\mathbf{g},0) = \iint dxdy\phi(\rho,0)e^{-2\pi i\mathbf{g}\cdot\rho} \tag{2.3.1}$$

$$\Phi(\mathbf{g},\epsilon) = \iint dxdy[e^{\epsilon\Delta}\phi(\rho,0)]e^{-2\pi i\mathbf{g}\cdot\rho} = \Phi(\mathbf{g},0)e^{-i\pi\lambda\epsilon g^2} \tag{2.3.2}$$

The expansion of the propagator to second order in ϵ and Δ is equivalent to writing

$$\Phi(\mathbf{g},\epsilon) \approx (1 - i\pi\lambda\epsilon g^2 + \tfrac{1}{2}\pi^2\lambda^2\epsilon^2 g^4)\,\Phi(\mathbf{g},0) \tag{2.3.3}$$

such that the intensity of the corresponding reflection after the electron has traveled the distance ϵ, is

$$I(\mathbf{g},\epsilon) \equiv |\Phi(\mathbf{g},\epsilon)|^2 = [1 + \tfrac{1}{4}(\pi\lambda\epsilon g^2)^4]\, I(\mathbf{g},0) \ . \tag{2.3.4}$$

To make sure that high order reflections (large g-vector) are not significantly amplified through the action of the propagator, it is necessary to use a slice thickness ϵ and an effective g_{max} such that

$$\frac{\pi\lambda\epsilon}{\sqrt{2}}\, g_{max}^2 \ll 1 \ \text{ or } \ \lambda\epsilon g_{max}^2 \ll \frac{\sqrt{2}}{\pi} = 0.45 \ . \tag{2.3.5}$$

For a periodic potential, period a, of cubic symmetry, the only g-vectors allowed are of the type $g_x = \frac{h}{a}$, $g_y = \frac{k}{a}$; h,k integers. For a numerical calculation with N sampling points in the x- and y-direction, $\delta = a/N$ and $\eta = a/N$. The equivalent expression to (2.3.4) is obtained by inserting (2.2.13) into the expression

$$\Phi(u,v,\epsilon) = \sum_{x,y} \phi(x,y,\epsilon)e^{-2\pi iux - 2\pi ivy}$$

and letting $u = h/a$ and $v = k/a$.

This gives

$$I(h,k,\epsilon) = I(h,k,0)\cdot\{1 + (\frac{\lambda\epsilon N^2}{4\pi a^2})^4 [\cos 4\pi h/N + \cos 4\pi k/N -$$

$$8\cos 2\pi h/N - 8\cos 2\pi k/N + 2\cos 2\pi(h+k)/N +$$

$$2\cos 2\pi(h-k)/N + 10]^2\} \ . \qquad (2.3.6)$$

For the special case of h = k,

$$I(h,h,\epsilon) = I(h,h,0)\cdot\{1 + (\frac{\lambda\epsilon N^2}{4\pi a^2})^4 [\cos 4\pi h/N - 16\cos 2\pi h/N$$

$$+ 12]^2\} \ . \qquad (2.3.7)$$

By dividing the a-axis up to N intervals, one is limited to $g_{max} = \frac{h_{max}}{a} = \frac{N}{2a}$. This gives

$$I(h_{max},h_{max},\epsilon) = I(h_{max},h_{max},0)\cdot\{1 + (\frac{4\sqrt{2}\lambda\epsilon}{\pi} g_{max}^2)^4\} \qquad (2.3.8)$$

and correspondingly one must impose

$$\lambda\epsilon g_{max}^2 \ll \frac{\pi}{4\sqrt{2}} = 0.56 \ . \qquad (2.3.9)$$

Equations (2.3.5) and (2.3.9) set an upper limit on the slice thickness and the number of reflections that can be included in the calculation. The slice thickness ϵ and the number of sampling points in each direction x and y must be chosen so as to satisfy

$$K_{max} \equiv \lambda\epsilon g_{max}^2 \ll \tfrac{1}{2} \ . \qquad (2.3.10)$$

3. Results of Computer Calculations

In order to compare the real space method with the conventional multislice method using fast Fourier transforms, programs were written that could be run in either FFT or RSP mode. To make it possible to use different values for the slice thickness, a three dimensional potential was calculated through a 3 dimensional Fourier transform. The specimen is copper (lattice const. 3.6 Å) and the c-axis (z-dir.) was divided into 16 intervals such that a slice thickness of either 3.6 Å, 1.8 Å, 0.9 Å, or 0.45 Å could be used. To be able to compare the results of the two methods under various conditions, amplitudes and phases of selected reflections were plot-

ted as a function of thickness.

Figure 2-3 shows what happens when $K \equiv \lambda\epsilon g_{max}^2$ increases beyond the critical value of 1/2. The reflections are 000, 200, 220,and 440 and the solid line is the result of the FFT method while the broken line represents the PSP method. Only amplitude vs. thickness is plotted and the slice thickness is kept constant at 3.6 Å. The maximum reciprocal lattice vector g_{max} takes on the values of 1.4 $Å^{-1}$, 1.9 $Å^{-1}$, and 2.5 $Å^{-1}$ to give a value for K of 0.18, 0.33, and 0.57 respectively.

As K increases the discrepancy between the two methods decreases, and significantly, when K increases beyond its critical value the RSP method starts to diverge. For this particular value of K the divergence sets in at about 100 Å and the intensity of the reflection 400 is seen to start growing almost exponentially. At about 100 Å there is enough intensity in the 440 reflection for it to be affected by the action of the propagator. The low order reflections are not affected directly by the propagator, although they are influenced by the interaction with higher order reflections through the crystal potential.

Figure 2-4 shows the amplitudes and phases of various reflections for 3 different values of ϵ and g_{max} keeping K constant at 0.25. Notice that while in Figure 2-1 the accelerating voltage is 200 kV, it is now set at 1 MV. The FFT calculation is almost unaffected by changes in ϵ and g_{max} (the results for ϵ = 3.6 Å and g_{max} = 2.8 Å are shown here) indicating that for $g_{max} > 2.8$ $Å^{-1}$, no appreciable aliasing effects are introduced. However, the results of the RSP calculation vary significantly as g_{max} increases (ϵ decreases), but the results of the RSP method approach that of the FFT method as the number of reflections included in the calculation increases.

Figure 2-5 shows the result of the RSP calculation for three different values of the slice thickness, at constant g_{max} equal to 2.8 $Å^{-1}$. Although the results vary somewhat depending on the slice thickness, reducing the slice thickness does not have a major effect; i.e., it does not cause the result of the RSP method to approach that of the FFT calculation.

Finally, Table 1 shows some computation times for the FFT and RSP methods. The times that are given are the computational times per slice for a slice thickness of 3.6 Å at three different numbers of sampling points. Using $\epsilon = 3.6$ Å obviously results in the fastest calculation since it is only necessary to calculate one $V_p(x,y)$. The programs were all run on a CDC 7600 computer.

4. Discussion

The primary motive behind the formulation of the real space method as an alternative way to do computer simulation of electron microscope images is that the RSP method appears to offer the following advantages:

1) There is no error due to aliasing which might occur when using Fourier transforms.

2) There is a possibility of eliminating the need to use periodic extension in faulted crystals.

3) The method might allow using a larger slice thickness, being correct to second order in ϵ.

4) A reduction in computer time is possible for the RSP method, the time per slice being proportional to N, the number of sampling points, rather than NLogN as for the FFT method.

As to the first claim, it is true that there is no aliasing associated with the RSP method. However, when all the physically relevant reflections are taken into account in an FFT calculation, aliasing should not be a problem. Rather one can argue that if aliasing ever does affect the result, not enough reflections have been included to give a meaningful result anyway. Although there has been no attention given here to determining when aliasing begins to affect the FFT calculation, the above results indicate that for an accelerating voltage of 1 MV, no such effect occurs as long as $g_{max} > 2.8$ Å^{-1}.

In most cases it will be possible to avoid using a period unit cell in an RSP calculation and thus avoid the need to use a periodic continuation when simulating images of defective crystals. The choice to use periodic continuation or not depends on how one decides to calculate the derivatives at the boundary of the x-y plane.

With respect to the third possible advantage, it is instructive to examine the asymptotic behavior of the RSP method as the slice thickness is decreased and the number of sampling points is increased. The first condition imposed on the RSP method is that $K = \lambda\epsilon g_{max}^2 \ll 1/2$. Similarly, a limit on K is also imposed in the conventional multislice method. For example, Ishizuka and Uyeda [6] using a stationary phase method in deriving the multislice formula arrive at the condition $K \ll 1$. Lynch and O'Keefe [9] argue that to avoid upper layer line reinforcement, the parameter for the pseudo-layer interaction $\alpha(g) = \frac{1}{2}\lambda\epsilon g^2$ should be less than 0.5. This again corresponds to $K < 1$, however for safety, a value of $K_{max} = 0.2$ was used. Thus the major difference between the RSP method and the FFT method in this respect is that while an RSP calculation for $K_{max} > 1/2$ beings to diverge, the corresponding FFT calculation does not. In either case, care should be taken with respect to the size of the slice thickness and the number of reflections needed to satisfy the condition $K < 1/2$.

However, in spite of similar conditions imposed on the two methods, it is seen from the results of the computer calculations that there are important differences between the RSP and FFT methods. Only as the slice thickness decreases and the number of reflections increases, does the result of the RSP calculation approach that of the FFT calculation. Furthermore while the FFT method is barely affected by changes in ϵ and g_{max} (as long as $\epsilon \leq 3.6$ Å and $g_{max} \geq 2.8$ Å^{-1} for $V_o = 1$ MV), the RSP calculation is strongly affected. Reducing the slice thickness gives only minor changes in the result of the RSP calculation, which means that the number of reflections included in the calculation has the strongest influence on the result. Although the need to incorporate reflections beyond 2.8 Å^{-1} is not indicated for the FFT

approach, this may be necessary in the RSP approach to ensure that none of the relevant reflections become artificially amplified through the action of the expanded propagator. Thus instead of being able to use a larger slice thickness in the RSP method compared to the FFT method, it seems more likely that a smaller slice thickness is essential in order to accommodate the inclusion of a larger number of reflections.

Finally, under identical conditions the RSP method offers a significant reduction in computer time. For the range of commonly used N (number of sampling points), the reduction in computer time per slice amounts to a factor of 3–5. This represents a significant savings in computer time and could prove to be a great value when using smaller and slower computers. It must however be noted that the saving is in calculation time per slice for the same N and is only effective if the same slice thickness and the same number of reflections can be used in the two methods.

5. Conclusion

The RSP method gives results similar to the conventional multislice calculation when care is taken to include enough reflections. To keep within the domain of validity of the RSP method, it might be necessary to reduce the slice thickness as the number of reflections increases, as needed to maintain $\lambda\epsilon g_{max}^2 < 1/2$. If this condition is not satisfied, the RSP method will begin to diverge due to a near-exponential growth of higher order reflections. The divergence is due to the amplification effect of the expanded propagator and does not set in until a nominally low intensity has been scattered into those reflections having g-vectors with magnitude close to g_{max}. Although a similar boundary condition is imposed on the validity of the FFT method, going beyond the domain of validity does not cause any divergence. In order to obtain reliable results from the RSP method it might be necessary to include more reflections than required with the FFT method which consequently also requires smaller slice thicknesses and therefore increased computational time. Investigations into further improvements of, and extended applications of the real space method are currently under way.

6. References

[1] R. Gronsky, in 38th Ann. Proc. Electron Microscopy Soc. Amer., San Francisco, CA., 1980, G.W. Bailey (ed.), p.2.

[2] W.O. Saxton, Advances in Electronics and Electron Physics, Suppl. 10, Academic Press Inc., New York, L. Marton (ed.), p. X1.

[3] J.M. Cowley and A.F. Moodie, Acta Cryst. 10, (1957), p. 609.

[4] D. Van Dyck, Acta Cryst. A34, (1978), p. 94.

[5] B. Jap and R. Glaeser, Acta Cryst. A34, (1978), p. 112.

[6] K. Ishizuka and N. Uyeda, Acta Cryst. A33, (1977), p. 740.

[7] D. Van Dyck, Journal of Microscopy, 119, (1980), p. 141.

[8] D. Van Dyck, Phys. Stat. Sol., B72, (1976), p. 321.

[9] D.F. Lynch and M.A. O'Keefe, Acta Cryst. A28, (1972), p. 536.

N_x,N_y	$N = N_x \cdot N_y$	t_{RSP} [sec]	t_{FFT} [sec]
20	400	0.017	0.063
28	784	0.034	0.12
40	1600	0.066	0.23

Table 1. Computational times per slice, slice thickness 3.6 Å, for the FFT method and the RSP method at three different values of the number of sampling points N.

FIGURE CAPTIONS

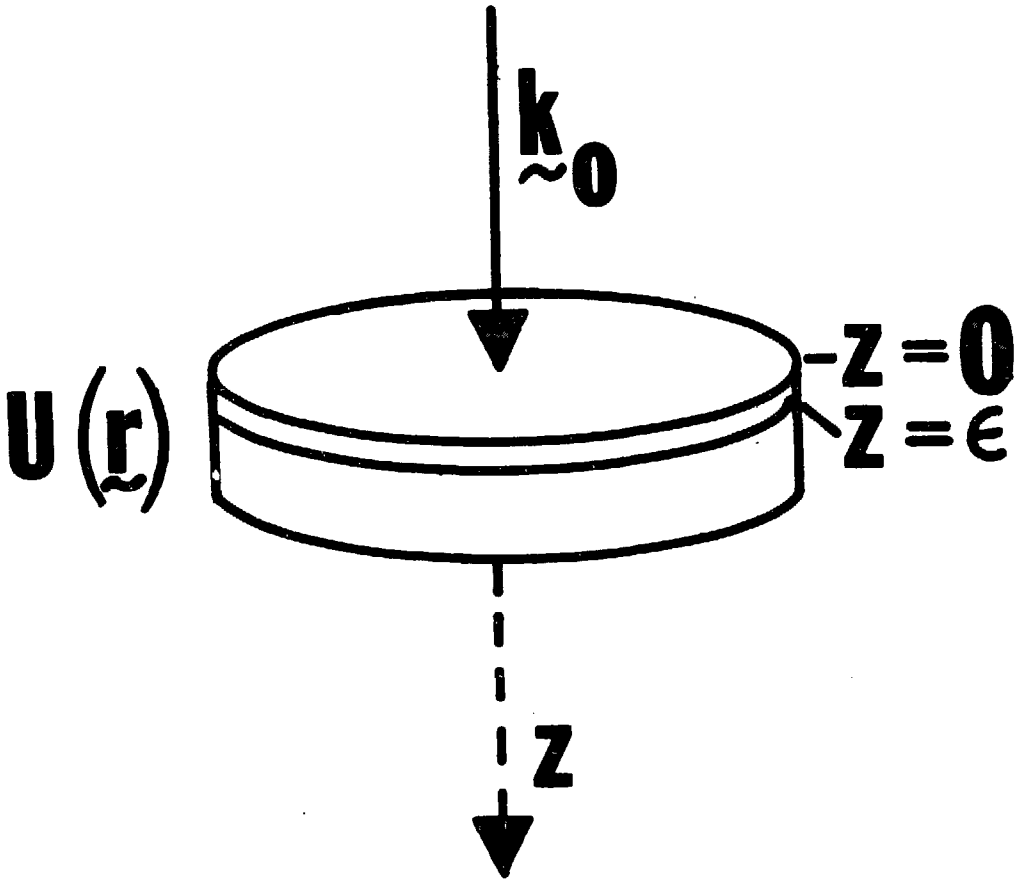
Fig. 2-1. Schematic representation of the scattering problem. The specimen is depicted as a potential distribution of U(r) which may be divided into a series of slices having thickness ϵ.

Fig. 2-2. Schematic representation of the fast Fourier transform (FFT) algorithm used by Ishizuka and Uyeda [6].

Fig. 2-3. Amplitude vs. thickness for the reflections 000, 200, 220, and 440 for copper [001]. Accelerating voltage is 200 kV and the slice thickness is 3.6 Å. The result from the FFT calculation is shown by the solid line, and the broken line represents the RSP calculation. In the first column $g_{max} = 1.4$ Å^{-1} (K = 0.18), in the second column $g_{max} = 1.9$ Å^{-1} (K = 0.33) and in the third column $g_{max} = 2.5$ Å^{-1} (K = 0.57).

Fig. 2-4. Amplitude and phase (in units of π) vs. thickness for 3 sets of values of the slice thickness ϵ and g_{max}. The values are (1) $\epsilon = 3.6$ Å, $g_{max} = 2.8$ Å^{-1}; (2) $\epsilon = 1.8$ Å, $g_{max} = 3.9$ Å^{-1}; (3) $\epsilon = 0.9$ Å, $g_{max} = 5.5$ Å^{-1} labeled separately for the RSP method. These values gave essentially the same results for the FFT method, plotted as the single FFT curve.

Fig. 2-5. Amplitude and phase (in units of π) vs. thickness for 3 different values of the slice thickness ϵ; g_{max} is kept constant at 2.8 Å^{-1}, ϵ takes the values 3.6 Å (-), 1.8 Å (—) and 0.9 Å (++). Calculation is by the RSP method.



XBL 8212-12231

Fig. 2-1.

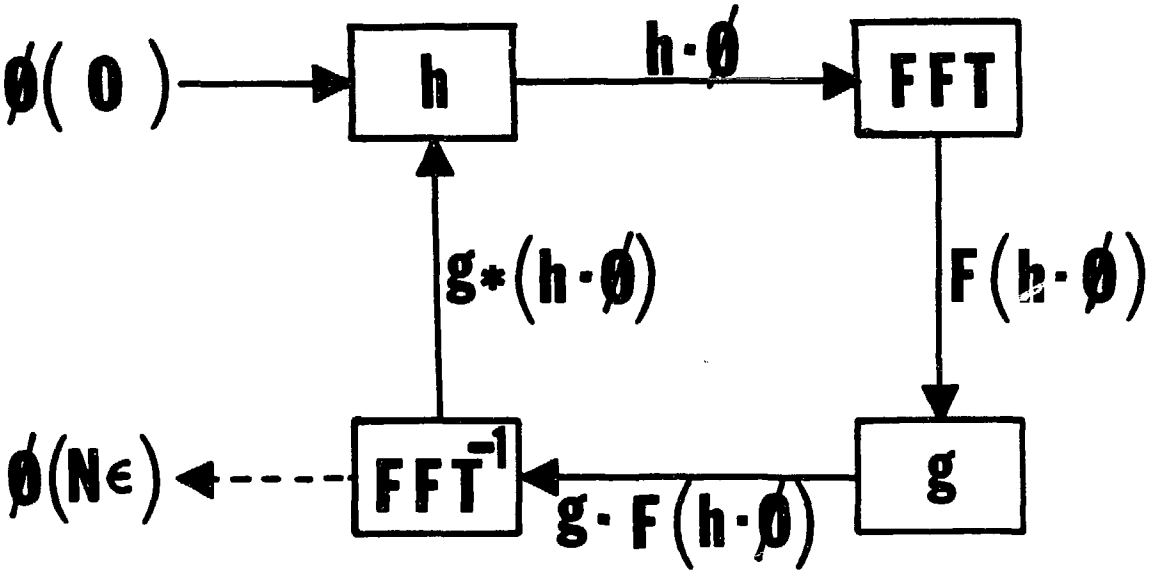


Fig. 2-2.

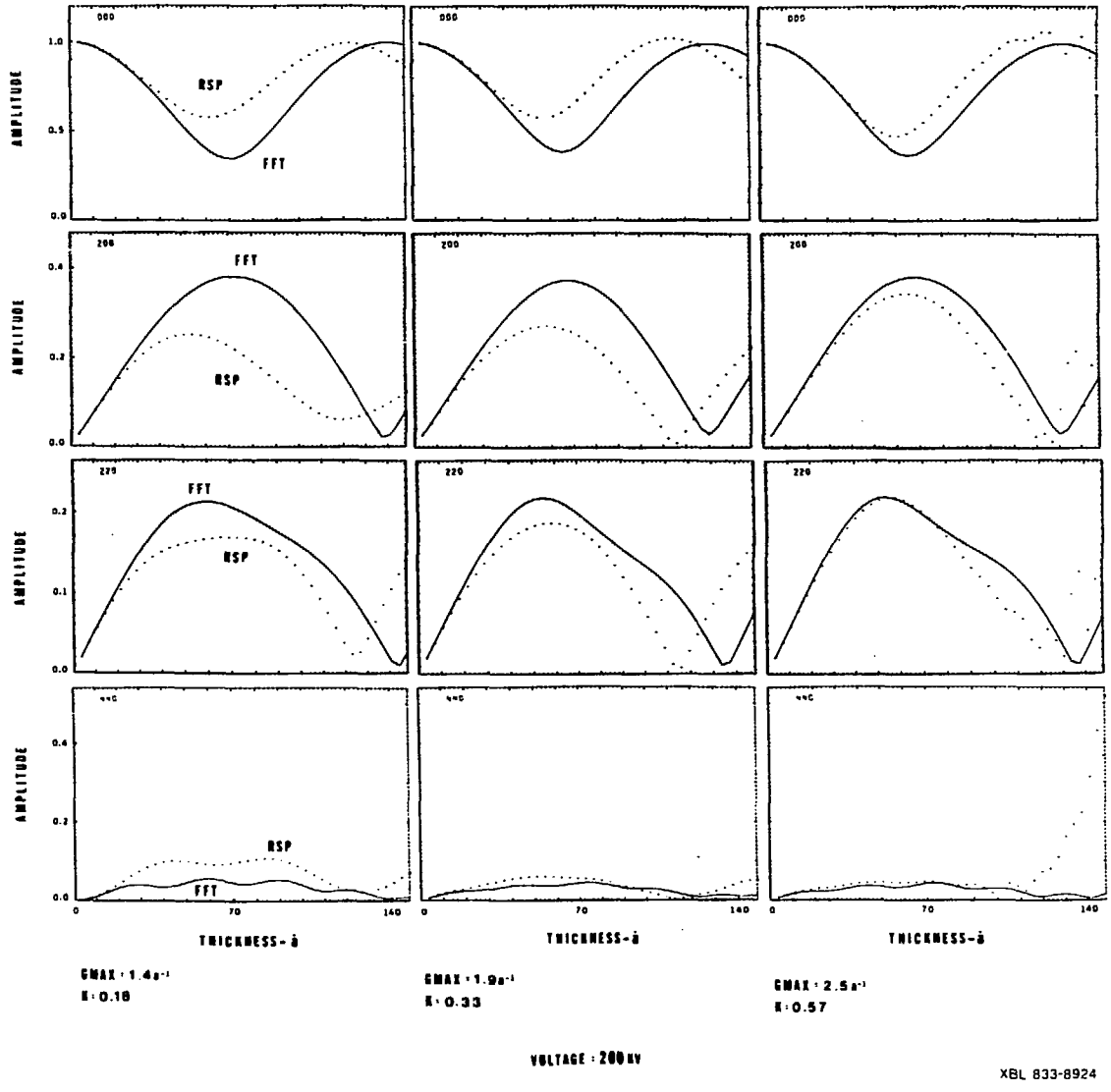

AMPLITUDE
RSP
FFT
THICKNESS-Å
GMAX : 1.4 Å⁻¹
K : 0.18
GMAX : 1.9 Å⁻¹
K : 0.33
GMAX : 2.5 Å⁻¹
K : 0.57
VOLTAGE : 200 KV
XBL 833-8924

Fig. 2-3.

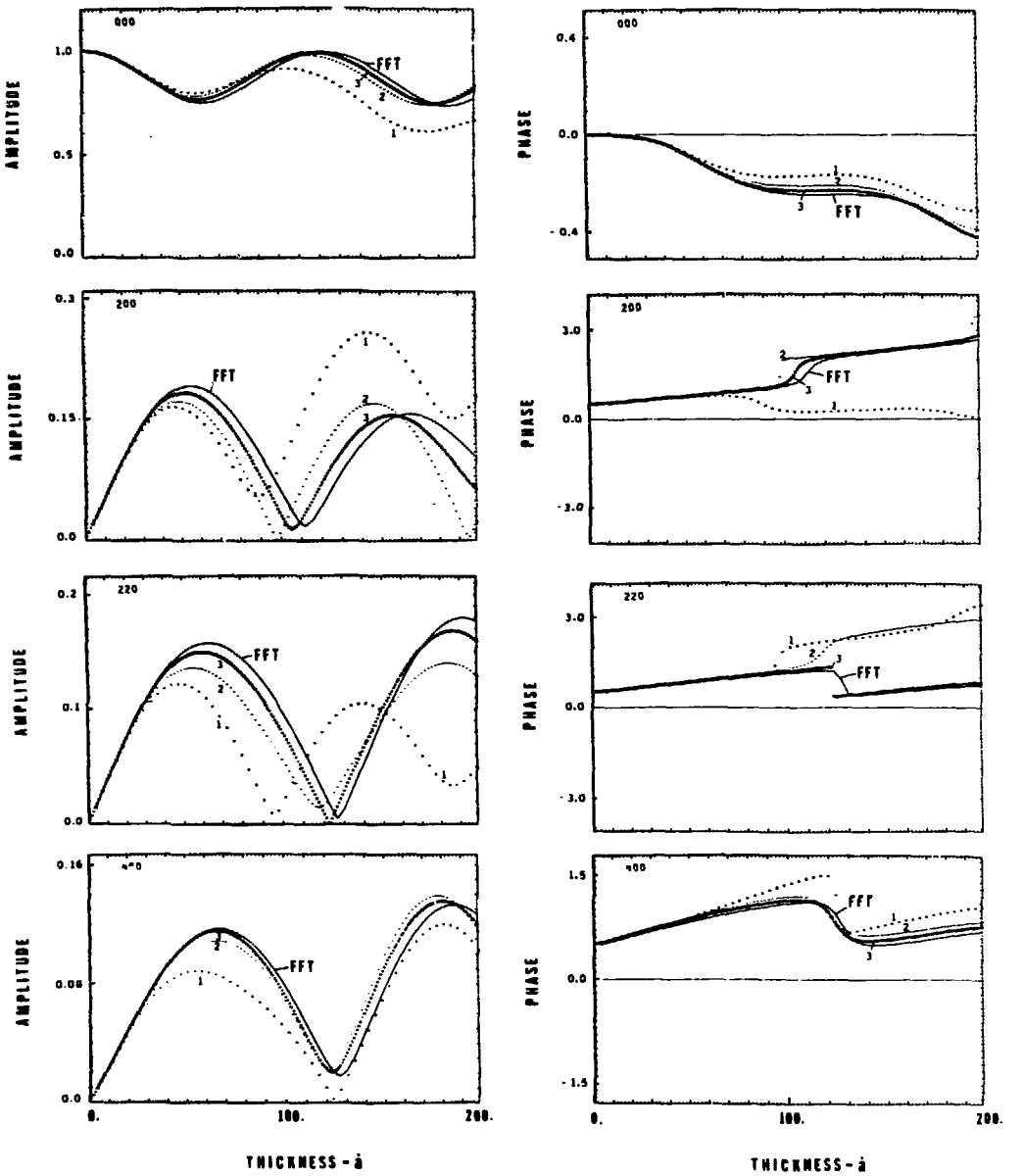

000
200
220
400
AMPLITUDE
PHASE
FFT
1.0
0.5
0.0
0.4
-0.4
0.3
0.15
3.0
-3.0
0.2
0.1
0.16
0.08
1.5
-1.5
0.
100.
200.
THICKNESS - Å
K=0.25
VOLTAGE : 1000 KV
XBL 833-8922

Fig. 2-4.

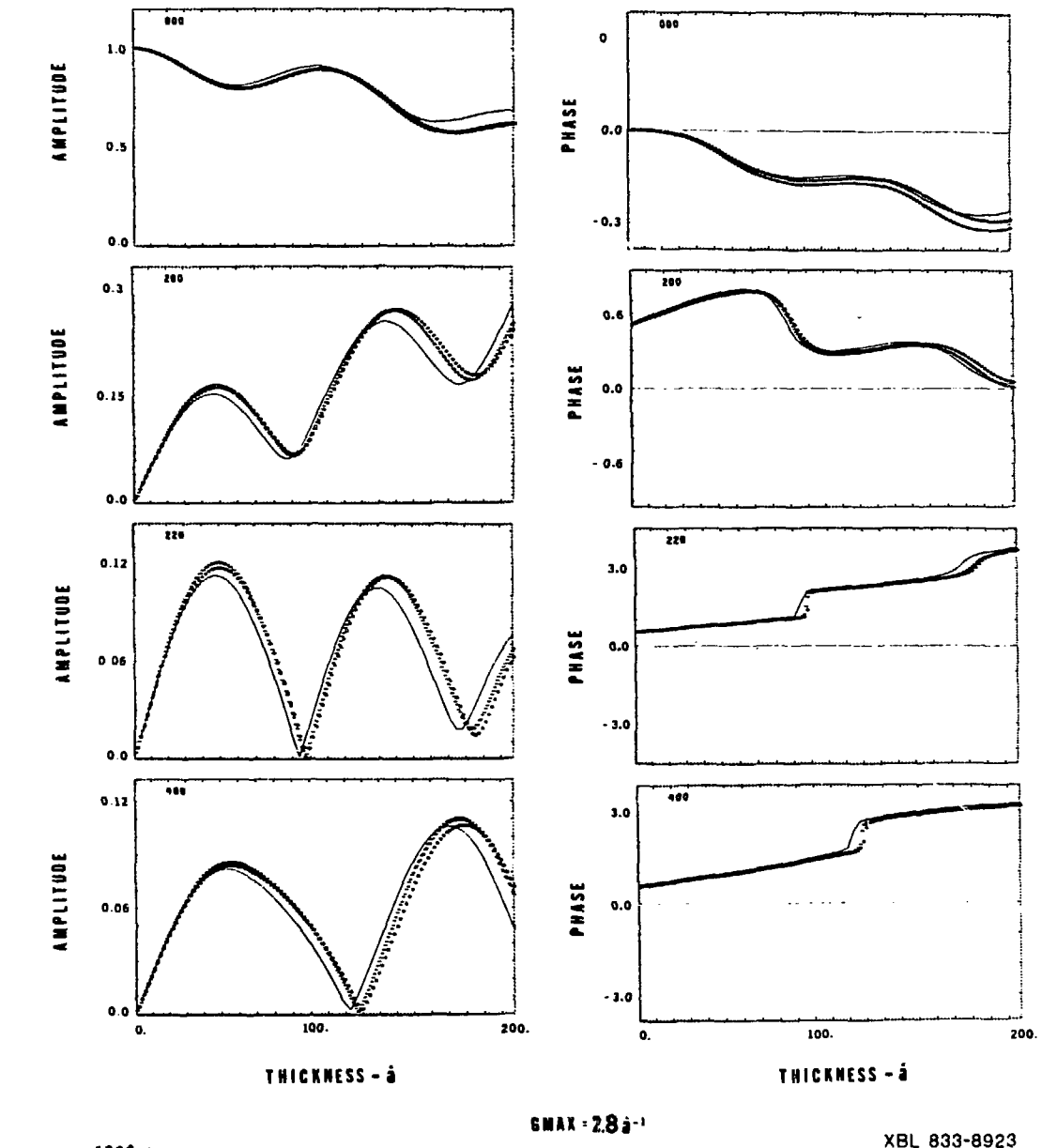

COPPER 001
AMPLITUDE
PHASE
THICKNESS - Å
VOLTAGE : 1000KV
GMAX : 2.8 Å⁻¹
XBL 833-8923

Fig. 2-5.

Chapter 3

On the Inclusion of Upper Laue Layers in Computational Methods In High Resolution Transmission Electron Microscopy

Abstract

Three different methods for computing scattering amplitudes in High Resolution Transmission Electron Microscopy (HRTEM) have been investigated as to their ability to include upper Laue layer (ULL) interaction. The conventional first order multislice method using fast Fourier transform (FFT) and the second order multislice method (SOM method) are shown to yield calculated intensities of first order Laue reflections with the use of slice thicknesses smaller than the crystal periodicity along the incident electron beam direction. It is argued that the calculated intensities of ULL reflections approach the correct values in the limiting case of vanishing slice thickness and electron wavelength. The third method, the improved phasegrating method (IPG) does also in principle include ULL effects, but is severely limited as to choice of slice thickness and sampling interval.

A practical way to use slice thicknesses less than the crystal periodicity along the incident beam direction is shown for both the conventional FFT method and the second order multislice method and tested on a spinel structure. It is also shown that the IPG method does not easily allow for a slice thickness different from the crystal periodicity in the beam direction.

1. Introduction

Because of the small curvature of the Ewald sphere most electrons scatter into directions given by the reciprocal lattice points lying in the zero order Laue zone [ZOLZ], see Fig. 3-1. Diffraction into upper Laue layers is a small effect, but is easily observed in Convergent Beam Electron Diffraction and High Resolution Transmission Electron Microscopy. The ULL reflec-

tions do not contribute directly to the image in HRTEM since the objective aperture, whether a real aperture or a virtual aperture defined by the Envelope Function [1], exclude the contribution to the image from the ULL beams. However, because of dynamical scattering the ULL reflections will modify the intensity of the zero order Laue reflections and consequently affect the image. Where ULL interactions are no longer negligible they must be included in the computation of electron micrographs to give correct results.

For practical reasons most image simulations in HRTEM employ a method based on multislicing. The specimen is sectioned into slices perpendicular to the electron beam and the electron wave function is calculated at every slice in a recursive way starting from the known electron distribution at the beginning of the first slice. Again for practical reasons, the crystal periodicity parallel to the electron beam is invariably chosen as the slice thickness provided it does not violate the criteria of validity for these multislice methods, see Chapter 2.

Various methods to include the effect of the ULL have been suggested. These are:

1. The use of slices smaller than the crystal periodicity parallel to the direction of the incoming electron beam [2],[3].
2. Second order multislice, using potential eccentricity within the slice [4].
3. Improved phase grating method [5].

So far the success of these methods in including ULL interactions has not been proven. It is generally believed that as the variation of the crystal potential along the incident electron beam direction is taken into account, the ULL reflections are automatically included. The degree to which this is correct is the topic of this paper.

The 200 reflections in $MgAl_2O_4$ [001] are not allowed by the structure (spinel). However Steeds [6] has shown that they appear in experimental diffraction patterns and accredit their presence to scattering from the first order Laue layer (FOLZ). This being the case, the ability to account for their presence would serve as a test as to the inclusion of ULL effects in present multislice calculations.

Additionally, with the use of small slice thicknesses it is of utmost importance that one correctly takes into account the potential for each slice in each of the three multislice calculations. Different authors approach this problem in different ways, not all equivalent and unfortunately not all correct, and this paper will address this question.

2. Theory

The three multislice formulations that will be discussed are all approximations to the solution of the modified Schrodinger equation below [4]

$$\frac{\partial \phi}{\partial z} = i\sigma V\phi + \frac{i\lambda}{4\pi} \nabla_{\perp}^2 \phi \tag{2.1}$$

where

$$\sigma = \frac{2\pi m e \lambda}{h^2} \tag{2.2}$$

and

$$\nabla_{\perp}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{2.3}$$

V is the crystal potential in volts.

2.1 Conventional Multislice with Small Slice Thicknesses

The conventional multislice formulation involves a recursive application of the following equation:

$$\phi(x,y,z_m + \Delta z) = P(x,y,\Delta z) * [Q(x,y,z_m,\Delta z) \cdot \phi(x,y,z_m)] \tag{2.1.1}$$

where P(x,y,z) is the free space propagator, and Q(x,y,z,Δz) is called the phasegrating. The expressions for P and Q are:

$$Q(x,y,z_m,\Delta z) = \exp\left\{ i\sigma \int_{z_m}^{z_m+\Delta z} V(x,y,z)dz \right\} \tag{2.1.2}$$

$$P(x,y,\Delta z) = -\frac{i}{\lambda \Delta z} \exp\left\{ \frac{i\pi}{\lambda \Delta z} (x^2 + y^2) \right\} \tag{2.1.3}$$

All information about the scattering potential is contained in Q and only reflections allowed by the Fourier transform of Q are possible. Similarly all the information about the curvature of the Ewald sphere is contained in the propagator which keep track of the excitation errors of each reflection (in the zero order Laue zone), see Appendix A.

If the crystal periodicity (c) parallel to the electron beam is used as a slice thickness i.e. $\Delta z = c$, then only reciprocal lattice points in the ZOLZ together with the corresponding structure factors determine the allowed reflections, as shown below.

It is customary to define a "projected" potential as follows :

$$V_p(x,y,z_m,\Delta z) \equiv \frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z} V(x,y,z)dz \qquad (2.1.4)$$

$$= \frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}\sum_{h,k,\ell} V(h,k,\ell)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b}+\frac{\ell z}{c})}dz$$

where V has been expressed as a Fourier series. The $V(h,k,\ell)$'s are calculated by performing a sum over all atoms within the unit cell

$$V(h,k,\ell) = \frac{h^2}{2\pi meV_c}\sum_i \mathcal{F}_i(h,k,\ell)e^{-2\pi i(\frac{hx_i}{a}+\frac{ky_i}{b}+\frac{\ell z_i}{c})}dz \qquad (2.1.5)$$

where (x_i,y_i,z_i) is the position of atom no. i with an electron scattering factor $\mathcal{F}_i$. V_c is the volume of the unit cell. Performing the integral gives ,as also pointed out by Self et al. [7] :

$$V_p(x,y,z_m,\Delta z) = \frac{1}{\Delta z}\sum_{h,k,\ell} V(h,k,\ell)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\frac{\sin\pi\ell\Delta z/c}{\pi\ell/c}e^{2\pi i\ell z_{mo}/c} \qquad (2.1.6)$$

where $z_{mo} \equiv z + \Delta z/2$

Setting $\Delta z = c/n$ gives the following expression for the projected potential

$$V_p(z_m,\Delta z=c/n) = \sum_{h,k,\ell} V(h,k,\ell)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\frac{\sin\pi\ell/n}{\pi\ell/n}e^{2\pi i\ell z_{mo}/c} \qquad (2.1.8)$$

With $n = 1$, (2.1.8) reduces to

$$V_p(z_m,\Delta z=c) = \sum_{h,k} V(h,k,o)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})} \qquad (2.1.9)$$

As can be seen from the expression above, only the components of the potential with a reciprocal vector lying in the ZOLZ contribute to the scattering. Thus no ULL effects can enter in the calculation. If, however n is not 1, the expression for the projected potential is a sum over all reciprocal vectors with a weighting factor proportional to the Fourier coefficient of the crystal potential. In principle this should allow for dynamical scattering between all (hkℓ) reflections and thus automatically account for ULL effects. Whether this is accurate or not will be addressed later in paragraph 5.

Previously the projected potential has been calculated in several ways:

i) Calculate the projected potential for a slice of thickness c. If a slice thickness of c/n is to be used, then simply divide the previously calculated potential by n and use this as the projected potential for each slice.

ii) Divide the unit cell into n volume elements which may include atom sections. The projected potential for each slice is calculated from the projection of the atoms contained within the slice according to (2.1.5). The sum is carried out over the atoms contained within the slice.

iii) Calculate a three dimensional crystal potential by summing over V(hkℓ) and using a 3 d. Fourier transform to find V(x,y,z). With modern computers the integral over V from z to z + Δz can easily be performed.

The first approach is obviously incorrect. The second method would be correct if it where not for the fact that the crystal potential is periodic in c and not in c/n. The third method is correct, but is impractical in cases where large unit cells are considered because of the huge amount of computer memory (128^3 = 8Mbytes) that is needed to store a three dimensional potential.

A practical method to calculate the projected potential for each slice is based on (2.1.6) and thus doesn't require additional memory. Starting from (2.1.6) one calculates the projected potential by first summing over ℓ. One can write

$$V_p(x,y,z_m,n) = \sum_{h,k} V'_{z_m,n}(h,k) e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})} \tag{2.1.10}$$

where

$$V'_{z_m,n}(h,k) \equiv \sum_{\ell} V(h,k,\ell) \frac{\sin\pi\ell/n}{\pi\ell/n} e^{2\pi i\ell z_m/c} \tag{2.1.11}$$

This amounts to modifying all V(h,k,0) by adding in a contribution from the ULL, as shown in Fig. 3-1. Thus reflections that were forbidden by V(h,k,0) may now be allowed by V'(h,k,0). The intensity of forbidden reflections will be zero for thicknesses corresponding to a multiple of unit cell distances c, only if there is a complete cancellation from the contents of the unit cell.

This yields an effective algorithm for dividing the specimen up into slices smaller than c. First, the coefficients $V(h,k,\ell)$ should be formed by summing over all atoms within a bona fide unit cell with a periodicity c along the incident beam direction. Secondly, if there are n "sub-slices", it is necessary to calculate n separate phasegratings or projected potentials according to (2.1.8), where for a given n, only Z_{mo} will change from sub-slice to sub-slice. The most efficient way to generate the n phasegratings depends on available computer memory.

2.2 Second Order Multislice

This method goes one step further and aspires to include ULL effects within the slice. Developed by Van Dyck [5] it introduces the concept of potential eccentricity within each slice. The equivalent to (2.1.1) is the recursive operation (see Chapter 2, equation 2.2.8):

$$\phi(x,y,z_m+\Delta z) = e^{\frac{1}{2}i\sigma\Delta z(1+\delta)V_p} e^{\frac{i\lambda\Delta z}{4\pi}\nabla_t^2} e^{\frac{1}{2}i\sigma\Delta z(1-\delta)V_p} \phi(x,y,z_m) \tag{2.2.1}$$

where the potential eccentricity δ is given as

$$\delta = \delta_m(x,y) \equiv \frac{\overline{z}_m(x,y) - \Delta z/2}{\Delta z/2} \tag{2.2.2}$$

The average quantity z is defined as

$$\overline{z}_m(x,y) \equiv \int_{z_m}^{z_m+\Delta z} (z - z_m)\, V(x,y,z)dz \tag{2.2.3}$$

In the evaluation of the real space method (Chapter 2), the above recursive operation was used to calculate both projected potentials and potential eccentricities from the 3-dim. crystal potential as outlined in 2.1 iii). However, the expressions $1/2(1+\delta)V_p$ and $1/2(1-\delta)V_p$ can be calculated in a way similar to the procedure described in 2.1 for calculating the projected potential as shown below:

It is useful to first calculate the quantity

$$\frac{1}{\Delta z}V_p\bar{z}_m \equiv \frac{1}{\Delta z^2}\int_{z_m}^{z_m+\Delta z}(z-z_m)\,V(x,y,z)dz$$

$$= \frac{1}{\Delta z^2}\int_{z_m}^{z_m+\Delta z}\sum_{h,k,\ell}(z-z_m)\,V(h,k,\ell)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b}+\frac{\ell z}{c})}dz \qquad (2.2.4)$$

$$= \frac{1}{\Delta z^2}\sum_{h,k}e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\sum_{\ell}V(h,k,\ell)\frac{e^{2\pi i\ell z_m/c}}{2\pi i\ell/c}\left(\Delta z\, e^{\pi i\ell\Delta z/c} - \frac{\sin\pi\ell\Delta z/c}{\pi\ell/c}\right)$$

Setting $\Delta z = c/n$ gives

$$\frac{V_p\bar{z}_m}{\Delta z} = \sum_{h,k}e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\sum_{\ell}V(h,k,\ell)\frac{n}{2\pi i\ell}e^{2\pi i\ell z_m/c}\left(e^{\pi i\ell z_m/c} - \frac{\sin\pi\ell/n}{\pi\ell/n}\right) \qquad (2.2.5)$$

$$\equiv \sum_{h,k}V_{\bar{z}}(h,k)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}$$

One can now write:

$$\tfrac{1}{2}(1+\delta)V_p = \frac{\bar{z}V_p}{\Delta z} = \sum_{h,k}V_{\bar{z}}(h,k)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})} \qquad (2.2.6)$$

$$\tfrac{1}{2}(1-\delta)V_p = V_p - \frac{\bar{z}V_p}{\Delta z} = \sum_{h,k}[V'(h,k) - V_{\bar{z}}(h,k)]\,e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})} \qquad (2.2.7)$$

2.3 Improved Phasegrating Method

This method, again suggested by Van Dyck [5], is based upon a modulated phasegrating. The effect of the potential is considered larger than the effect of the propagator which is treated as a perturbation. In this case the wavefunction is written

$$\phi(x,y,z) = \exp\left\{i\sigma\int_0^z V(x,y,z')dz'\right\}\theta(x,y,z) \qquad (2.3.1)$$

Substituting the above expression into the equation for ϕ (2.1) gives

$$\theta(x,y,z) = \theta(x,y,o) + \frac{i\lambda}{4\pi}\int_o^z dz' \left\{ \nabla_\perp^2\theta(z') + i\sigma\Delta z[\nabla_\perp^2 V_p(z')]\theta(z') \right. \tag{2.3.2}$$

$$\left. + 2i\sigma\Delta z \nabla_\perp V_p(z') \cdot \nabla_\perp\theta(z') + (i\sigma\Delta z)^2 \theta(z')[\nabla_\perp V_p(z')]^2 \right\}$$

which yields a first order perturbation result for theta

$$\theta(z) \approx \theta(o) + \frac{i\lambda\Delta z}{4\pi}\left\{ \nabla_\perp^2\theta(o) + i\sigma\Delta z\theta(o)\int_o^z dz' \right. \tag{2.3.3}$$

$$\left. [\frac{1}{\Delta z}\nabla_\perp^2 V_p + i\sigma\Delta z\frac{1}{\Delta z}(\nabla_\perp V_p)^2] + 2i\sigma\Delta z\nabla_\perp\theta(o) \cdot \int_o^z dz' \frac{1}{\Delta z}\nabla_\perp V_p \right\}$$

A practical application of (2.3.3) can be formulated as follows: One considers a multislice approach where the specimen is divided into N slices perpendicular to the incident electron beam. The wavefunction after the first slice of thickness Δz is written

$$\phi(x,y,\Delta z) = \exp\{i\sigma\Delta z V_p(\Delta z)\}\theta(\Delta z) \tag{2.3.4}$$

For N slices of thickness Δz (2.3.4) becomes

$$\phi(x,y,N\Delta z) = \exp\left\{ i\sigma \int_o^{N\Delta z} V(x,y,z') \right\}\theta(N\Delta z) \tag{2.3.5}$$

where θ (NΔz) is a recursive application of (2.3.3).

The major problem using (2.3.3) above comes from solving for the integrals over $V_p(z)$. However, proceeding as before, one gets

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z} \nabla_\perp^2 V_p dz' = -(2\pi)^2 \sum_{h,k} e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right) \tag{2.3.6}$$

$$\times \sum_\ell V(h,k,\ell)\frac{n}{2\pi i\ell}\left(e^{2\pi i\ell z_m/c}\frac{\sin \pi\ell/n}{\pi\ell/n} - 1 \right)$$

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z} \frac{\partial V_p}{\partial x} dz' = 2\pi i \sum_{h,k} e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}(\frac{h}{a}) \sum_\ell V(h,k,\ell)\frac{n}{2\pi i\ell}\left(e^{2\pi i\ell z_m/c}\frac{\sin \pi\ell/n}{\pi\ell/n} - 1 \right) \tag{2.3.7}$$

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z} \frac{\partial V_p}{\partial y} dz' = 2\pi i \sum_{h,k} e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}(\frac{k}{b}) \sum_\ell V(h,k,\ell)\frac{n}{2\pi i\ell}\left(e^{2\pi i\ell z_m/c}\frac{\sin \pi\ell/n}{\pi\ell/n} - 1 \right) \tag{2.3.8}$$

The expression for $\int_{z_m}^{z_m+\Delta z} [(\frac{\partial V_p}{\partial x})^2 + (\frac{\partial V_p}{\partial y})^2]dz$ is given as

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}[(\frac{\partial V_p}{\partial x})^2+(\frac{\partial V_p}{\partial y})^2]\,dz' =$$

$$-(2\pi)^2\sum_{\substack{h,k\\h',k'}}(\frac{hh'}{a^2}+\frac{kk'}{b^2})\,e^{2\pi i(\frac{h+h'}{a}x+\frac{k+k'}{b}y)}$$

$$\times\sum_{\ell\ell'}V(h,k,\ell)V(h',k',\ell')\frac{n}{2\pi i\ell}\,\frac{n}{2\pi i\ell'}\left[e^{2\pi i(\ell+\ell')Z_m/c}\,\frac{\sin\pi(\ell+\ell')/n}{\pi(\ell+\ell')/n}\right.$$

$$\left.-e^{2\pi i\ell Z_m/c}\,\frac{\sin\pi\ell/n}{\pi\ell/n}-e^{2\pi i\ell' Z_m/c}\,\frac{\sin\pi\ell'/n}{\pi\ell'/n}+1\right] \qquad (2.3.9)$$

Equation 2.3.9 is complicated by the crossterms contained within the [ ]. In the limiting case $\Delta z/c$ goes to zero (2.3.9) simplifies to

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}[(\frac{\partial V_p}{\partial x})^2+(\frac{\partial V_p}{\partial y})^2]\,dz' = -\frac{(2\pi)^2}{3}\left[\left(\sum_{h,k}\frac{h}{a}V(h,k,0)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\right)^2\right.$$

$$\left.+\left(\sum_{h,k}\frac{k}{b}V(h,k,0)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\right)\right] \qquad (2.3.10)$$

The other simplifying case is $\Delta z = c$. In this case one obtains as follows

$$\frac{1}{c}\int_{z_m}^{z_m+c}\nabla_\perp^2 V_p dz' = -(2\pi)^2\sum_{h,k}e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right) \qquad (2.3.11)$$

$$\times\left[(\tfrac{1}{2}+\frac{z_m}{c})V(h,k,0)-\sum_{\ell\neq 0}V(h,k,\ell)\frac{n}{2\pi i\ell}\right]$$

$$\frac{1}{c}\int_{z_m}^{z_m+c}\frac{\partial V_p}{\partial x}dz' = 2\pi i\sum_{h,k}e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}(\frac{h}{a})\left[(\tfrac{1}{2}+\frac{z_m}{c})V(h,k,0)-\sum_{\ell\neq 0}V(h,k,\ell)\frac{n}{2\pi i\ell}\right] \qquad (2.3.12)$$

$$\frac{1}{c}\int_{z_m}^{z_m+c}\frac{\partial V_p}{\partial y}dz' = 2\pi i\sum_{h,k}e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}(\frac{k}{b})\left[(\tfrac{1}{2}+\frac{z_m}{c})V(h,k,0)-\sum_{\ell\neq 0}V(h,k,\ell)\frac{n}{2\pi i\ell}\right] \qquad (2.3.13)$$

$$\frac{1}{c}\int_{z_m}^{z_m+c}(\frac{\partial V_p}{\partial x})^2dz' = [\tfrac{1}{3}+\frac{z_m}{c}(1+\frac{z_m}{c})]\left[\sum_{h,k}2\pi i(\frac{h}{a})V(h,k,0)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\right]^2 \qquad (2.3.14)$$

$$+2\left[\sum_{h,k}2\pi i(\frac{h}{a})V(h,k,0)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\right]\left[\sum_{h,k}2\pi i(\frac{h}{a})e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\sum_{\ell\neq 0}(\frac{1}{2\pi i\ell})^2V(h,k,\ell)\right]$$

$$+ (\tfrac{1}{2}+\frac{z_m}{c})\left[\sum_{h,k} 2\pi i(\frac{h}{a})V(h,k,0)e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\right]$$

$$\times \left[\sum_{h,k} 2\pi i(\frac{h}{a})e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\sum_{\ell\neq 0}\frac{i}{2\pi\ell}V(h,k,\ell)\right]$$

$$+ \left[\sum_{h,k} 2\pi i(\frac{h}{a})e^{2\pi i(\frac{hx}{a}+\frac{ky}{b})}\sum_{\ell\neq 0}\frac{1}{2\pi i\ell}V(h,k,\ell)\right]^2$$

The expression for the integral $\int(\frac{\partial V_p}{\partial y})^2 dz$ follows from (2.3.14).

3. Procedure

Computer programs to implement the various methods were written according to the theory outlined in paragraph 2. and applied to the test case of a crystal of $MgAl_2O_4$. In the case of the improved phasegrating method only $\Delta z = c$ was considered since this was the only case that allowed the expression given by (2.3.9) to be calculated in a reasonable time. As with the conventional multislice method and the second order multislice method, slice thicknesses of 8.08 Å, 4.04 Å, 2.02 Å and 1.01 Å were used, corresponding to n = 1,2,4 and 8 respectively.

4. Results

The results of the computations are given in Figs. 3-2 through 3-5. Figure 3-2 shows amplitudes for the central beam and the reflections 110, 200, 220 and 400 for a slice thickness of 8.08 Å, 2.02 Å and 1.01 Å calculated by the conventional multislice (FFT) method, while Fig. 3-3 shows the corresponding results calculated by the second order multislice method. The accelerating voltage is 200 kV and all reflections out to 4.0 $Å^{-1}$ were included in the calculations. The improved phasegrating method failed to produce reasonable results for this structure in the case of $\Delta z = c$ and $g_{max} = 4.0$ $Å^{-1}$ (see Discussion).

The 110 reflection is forbidden in the classical sense by the FCC structure and has zero amplitude for any thickness when $\Delta z = c$. The 200 is forbidden by reflections within the ZOLZ, but is claimed to be allowed through scattering from the first order Laue layer [6]. The calculations however do not show any significant amount of scattering into the 200 reflection

(compared to the 110 reflection). Within the unit cell both the amplitudes of the 110 and the 200 reflection deviates from zero, (n = 4,8) and only at thicknesses corresponding to multiples of c, do they become negligibly small. The reflections are "forbidden" because scattering from one part of the unit cell interferes destructively with scattering from another part of the unit cell. Only if the cancellation is complete, which would require that the electron wavefunction for all scattering purposes remains the same throughout the cell, does the amplitude go to zero.

Figures 3-4 and 3-5 show four diffraction patterns calculated from a 300 Å thick crystal of $MgAl_2O_4$. In a) the slice thickness is c, while in b), c) and d) the slice thickness is c/2, c/4 and c/8 respectively. The result in Fig. 3-4 are for the FFT method while the results in Fig. 3-5 are for the SOM method.

5. Discussion

The failure to give reasonable results by the improved phasegrating method is caused by the large slice thickness required to perform the calculation and is discussed in Chapter 4. Because of the necessary large g_{max} that must be used, the computation cannot be performed within the domain of validity for this method. The results show in Chapter 4 that the criteria for validity of the IPG are more severe than that of the real space method.

Both the first order FFT method and the second order multislice method fail to indicate any scattering into the 200 reflections from out of the ZOLZ. This does not mean that the FFT or the SOM methods do not include upper Laue layer effects, only that they fail to account for the experimentally observed intensities of the 200 reflections. It is not clear whether these intensities are due to multiple scattering off the first order zone (i.e. $[1,25,1] + [1,\overline{25},\overline{1}]$) or due to the effect of the higher order zones directly through the modified Fourier coefficient V'(200) of the crystal potential. A systematic study of the behavior of the diffracted beams with beam tilt is under way and should hopefully clarify this.

The circle of excited reflections showing up in Figs. 3-4a and 3-5a are not first order Laue reflections as their location in reciprocal space would indicate, but ZOLZ that are excited

because the excitation error associated with these reflections become equal to 1/c corresponding to the Ewald sphere cutting through the first order Laue zone. At this point the phase in the propagator becomes 2π, equivalent to an excitation error of 0. This occurs whenever $\xi(h,k,0) = 1/\Delta z$. In the case of n = 2,4,and 8 the "pseudo" ULL reflections correspond to scattering vectors larger than the maximum reciprocal vector included in the calculations and are no longer visible. However a new set of reflections located on the same circle in reciprocal space now start to appear and this time they correspond to actual first order Laue reflections, Figs. 3-4c,d, and 3-5b,c,d. Figure 3-4b does not show any ULL reflections, indicating that in the case of the FFT method a slice thickness of half a unit cell is insufficient to give ULL effects. This is not true for the SOM method which show ULL reflections even for n = 2 (Fig. 3-5b). This can only be attributed to the use of potential eccentricity which allow for modulations within the slice. Thus even in the case of n = 4 and n = 8, where both methods show the presence of ULL reflections, it must be concluded that the SOM method is the more accurate of the two.

The degree of accuracy to which the intensities of higher order reflections have been calculated still remains to be discussed. Only as the wavelength and the slice thickness approach zero do the first order and the second order method accurately include the interactions of upper Laue layers. As pointed out in paragraph 2 the information about the scattering potential is contained in the phasegrating while the propagator keeps track of the excitation errors. Physically, ULL scattering occurs when the Ewald sphere approaches the first order Laue layer as shown in Fig. 3-1, that is when the excitation error for the corresponding reflection becomes small. From (2.1.11) and (2.2.5) it is clear that the contribution of a particular (hkℓ) reflection is proportional to V(hkℓ), the Fourier coefficient of the potential, and does not depend upon its excitation error.In the case where the Ewald sphere passes through an ULL reflection, say (1,25,1) , V(1,25,1) would be summed into V'(1,25,0) with a weighting factor which does not depend on ξ(1,25,1). Additionally the remaining V(1,25,1) are all summed into V'(1,25,0) regardless of their excitation error. In effect the phasegrating "sees" a flat Ewald sphere cutting through every section of the reciprocal space. When the phasegrating is convoluted with the

propagator the contribution of V'(1,25,0) depends on ξ(1,25,0) and not on ξ(1,25,1). The error in the phase of the propagator for a first order Laue reflection depends on the slice thickness and the electron wavelength and is shown in Appendix A to be $2\pi\lambda^2\Delta z g^2/c$. As the wavelength and slice thickness decreases the propagator approaches its correct value and the accuracy to which the ULL are included in the FFT method and the SOM method increases.

The improved phasegrating method does not separate the effect of the potential and the Ewald sphere (through the wavelength) and should thus better allow for the inclusion of ULL interactions. However, the restriction on wavelength, slice thickness and sampling interval associated with the IPG, see Chapter 4, excludes the use of this method on the given problem. The Ewald sphere cuts through the first ULL at approx. 3.1 Å^{-1} which sets a lower limit on g_{max}, the maximum reciprocal scattering vector that must be included in the calculations. In order to produce reasonable results, it is necessary to use a slice thickness less than 1 Å, but the complexity of the method does not readily allow for a slice thickness less than c (8.08 Å).

6. Conclusion

Of the three methods that are discussed in this paper, the second order multislice method is the most suited for inclusion of ULL reflections. The use of potential eccentricity permits the use of larger slice thicknesses without sacrificing the inclusion of ULL effects. The accuracy of the calculation of higher order reflections depends on both slice thickness and electron wavelength and increases as the thickness of the slice and the wavelength approach zero. This is true for both the FFT method and the SOM method. The IPG method contains 3 dimensional information even when the slice thickness equals the crystal periodicity in the electron beam direction, but the method is impractical under most conditions because of severe restrictions on sampling interval and slice thickness.

7. References

[1] J. Frank, Optik 38 (1973) 519

[2] P. Goodman and A.F. Moodie, Acta Cryst A30 (1974) 280

[3] B. F. Buxton, M.D. Shannon and J. A. Eades, Proc. 42nd EMSA Meeting (1984), 524

[4] D. Van Dyck, J. Microscopy 119 (1980) 141

[5] D. Van Dyck, J. Microscopy 132 (1983) 31

[6] J. W. Steeds and N. S. Evans, Proc. 38th EMSA Meeting, (1980), 188

[7] P. Self, M.A. O'Keefe, P.R. Buseck and A.E.C. Spargo, Ultramicroscopy 11 (1983) 35

FIGURE CAPTIONS

Fig. 3-1. Schematic drawing showing a segment of the Ewald sphere and its relation to the reciprocal lattice. The zero and the first order Laue zones are indicated in the figure. Large open circles represent reflections lying in the zero order zone, while small open circles indicate the "column" of reciprocal points whose Fourier coefficients of the potential are summed into the Fourier coefficient of the corresponding zero order reflection to give a new effective potential.

Fig. 3-2. Amplitude vs. thickness for the reflections 000, 110, 200, 220, and 400 for $MgAl_2O_4$ [001]. Accelerating voltage is 200 kV and the slice thickness is 8.08 Å, 2.02 Å and 1.01 Å corresponding to n = 1, 4, and 8 respectively. Calculation is by the FFT method.

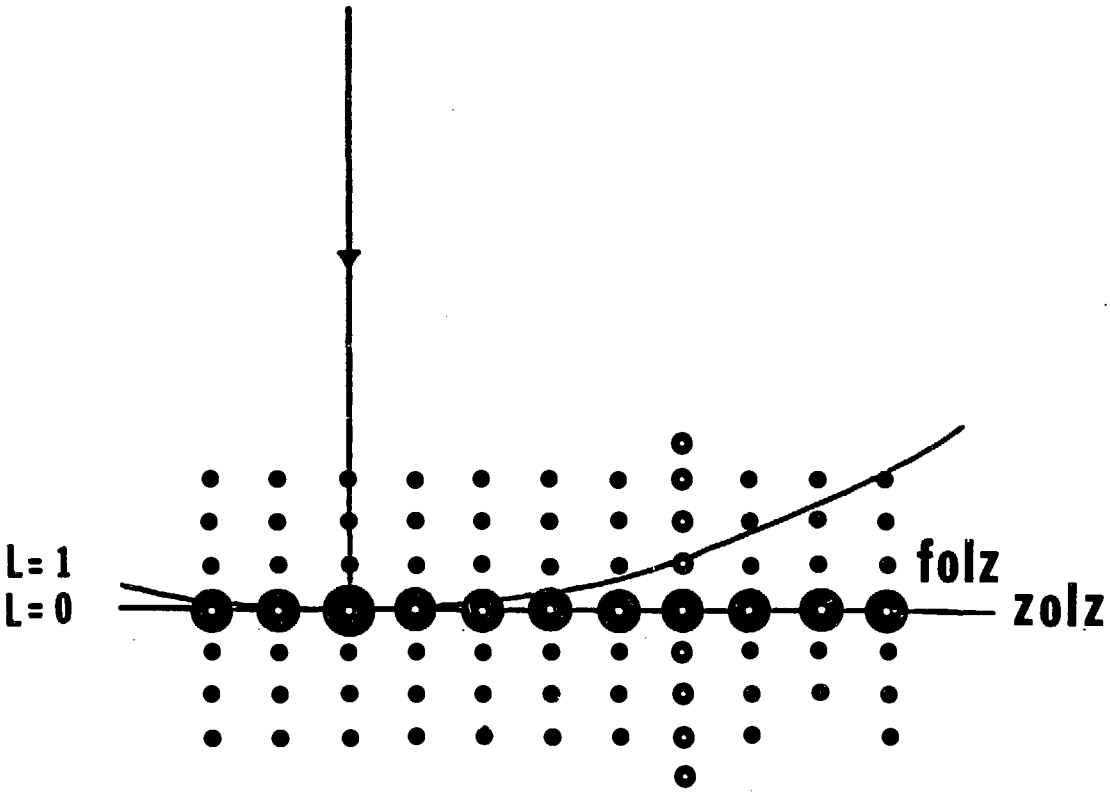
Fig. 3-3. Same as for Fig. 3-2 except that the calculation is by the SOM method.

Fig. 3-4. Computed diffraction patterns for a 300 Å thick specimen of $MgAl_2O_4$. The calculation is by the FFT method and the slice thickness is indicated by the value of n (Δz = 8.08 Å/n).

Fig. 3-5. The same as for Fig. 3-4 except that the calculation is by the SOM method.

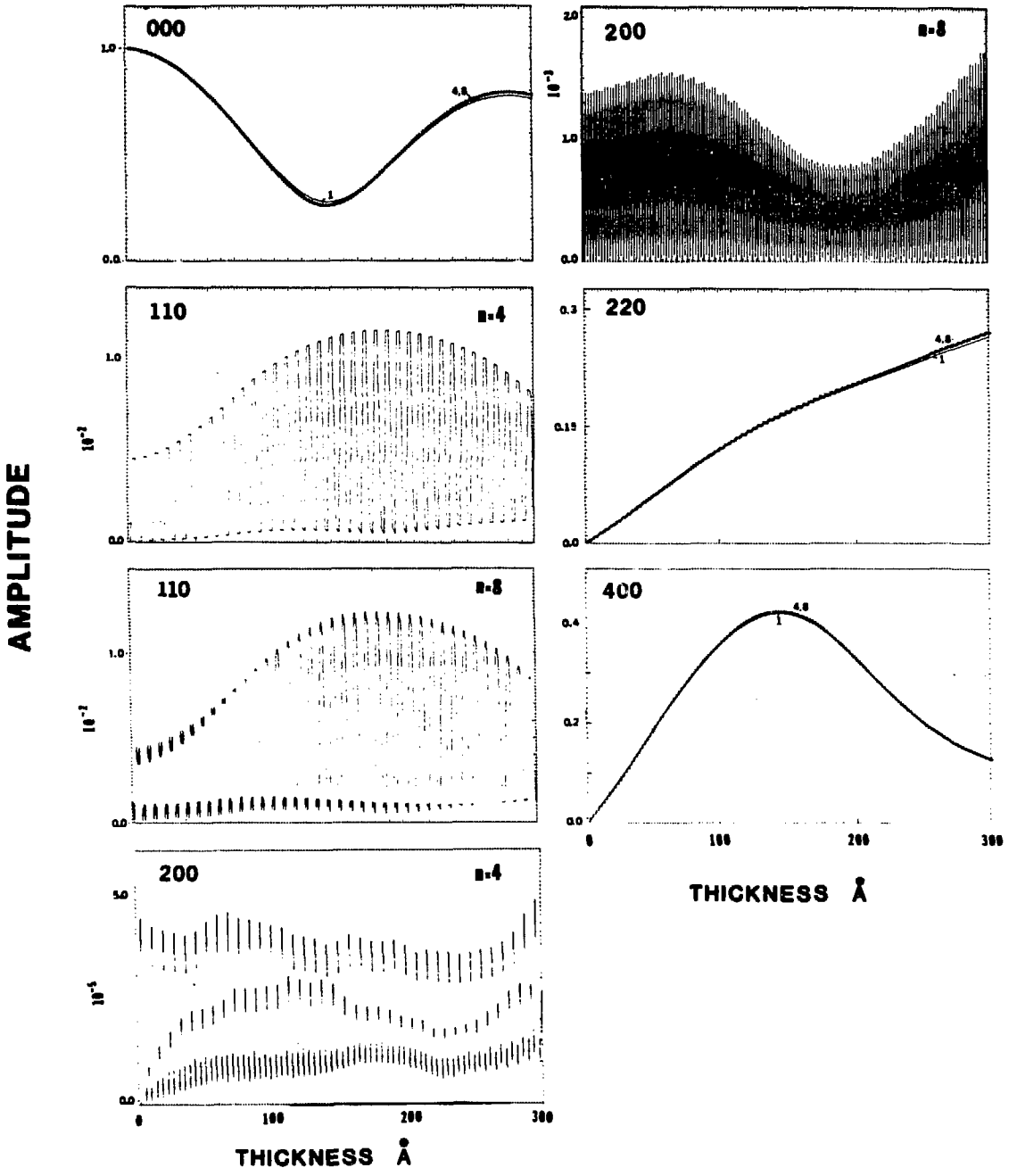
Fig. 3-6. Schematic drawing illustrating the central beam ($\mathbf{k}_0$) and a scattered beam ($\mathbf{k}_1$) traversing a slice of thickness Δz.

Fig. 3-7. Schematic drawing showing two scattered beams; one corresponding to scattering in the zero order zone and one corresponding to scattering into the first order zone.



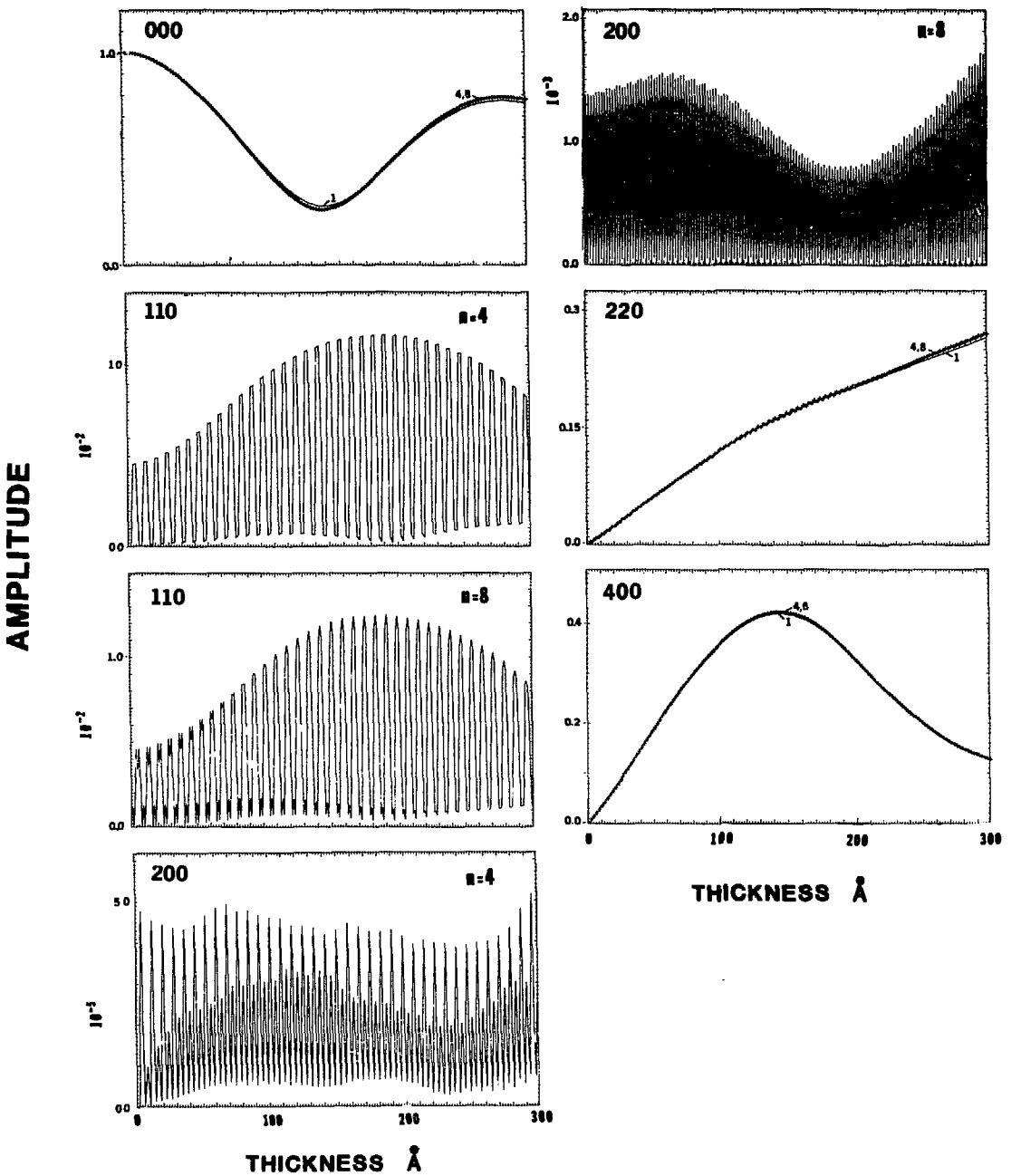
XBL 851-980

Fig. 3-1.



XBL 851-982

Fig. 3-2.



XBL 851-983

Fig. 3-3.

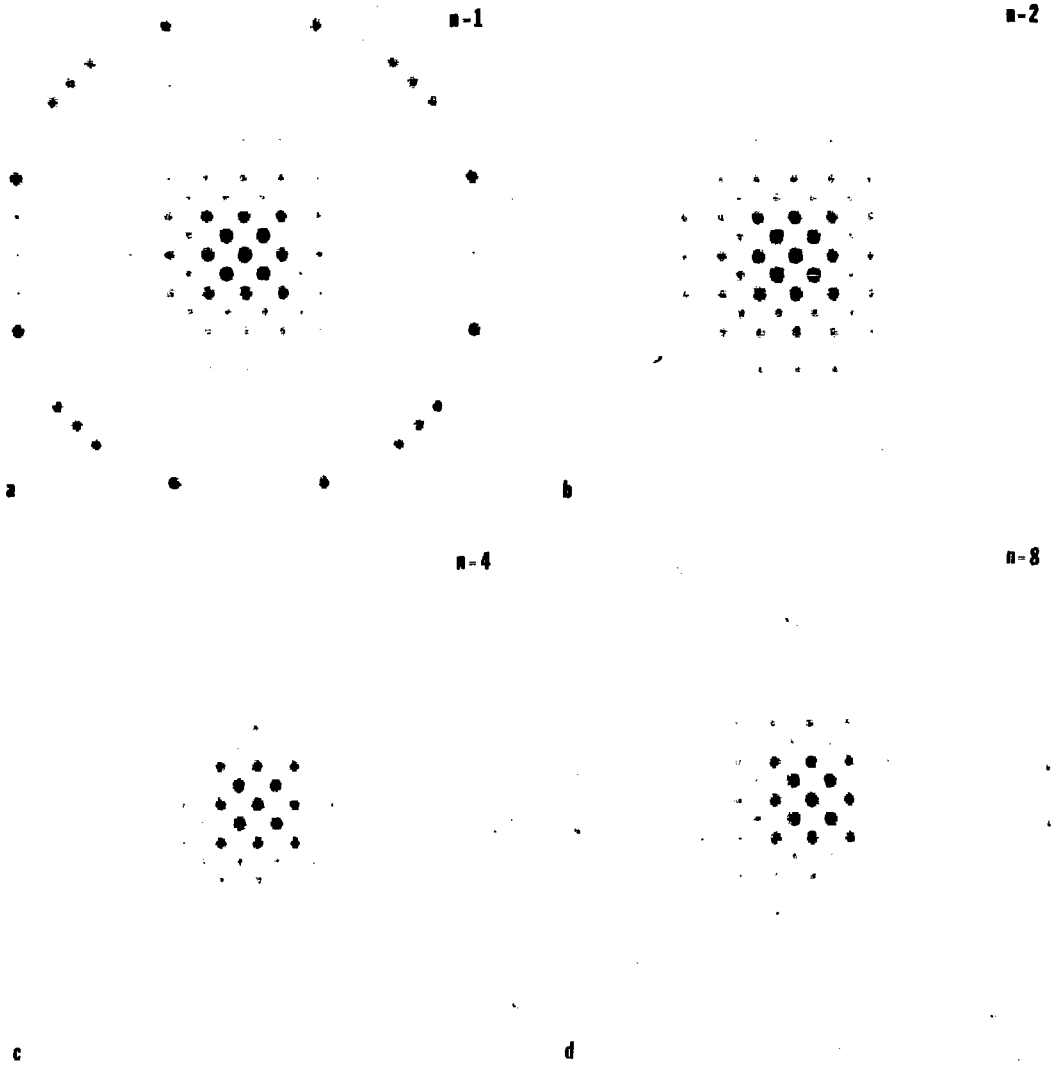
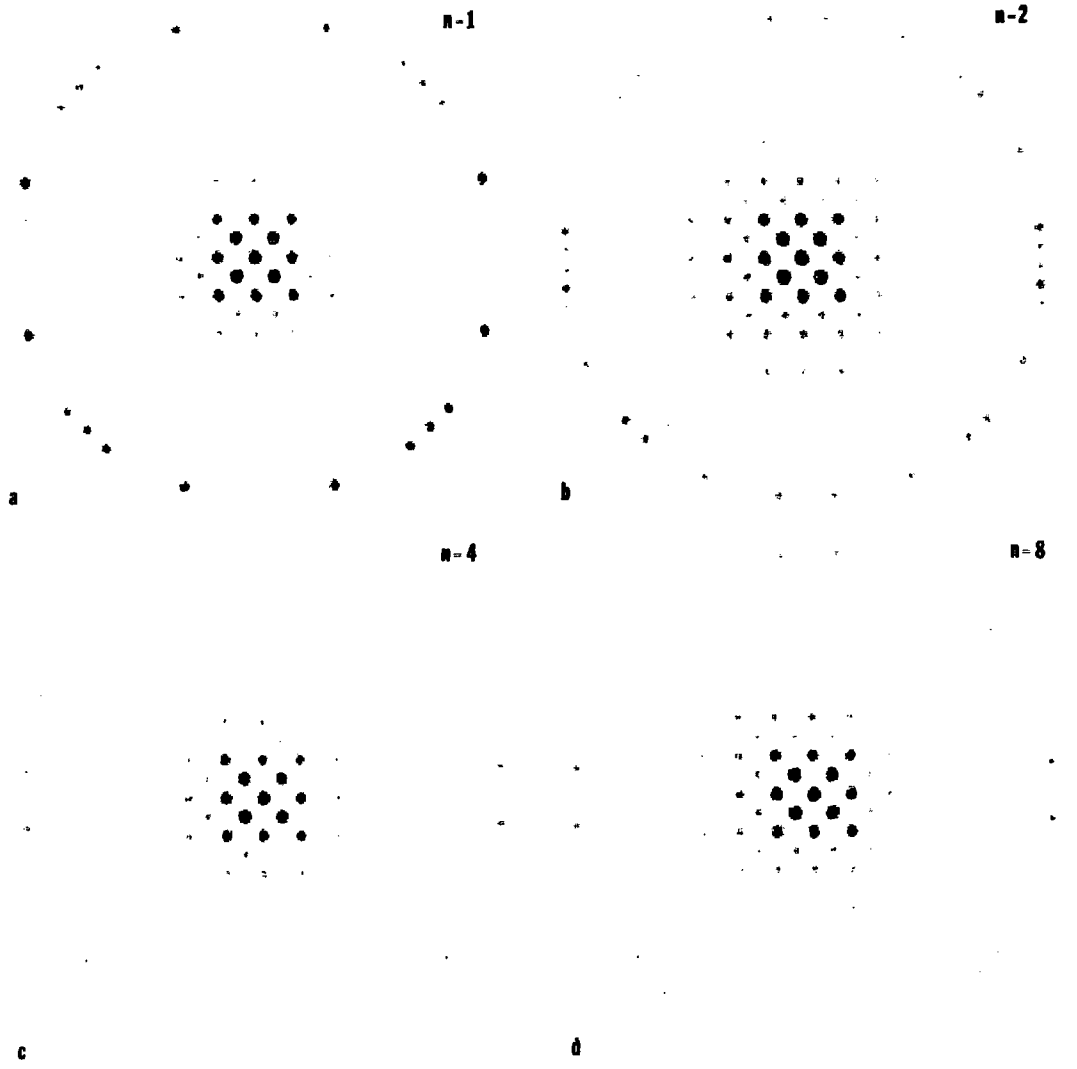
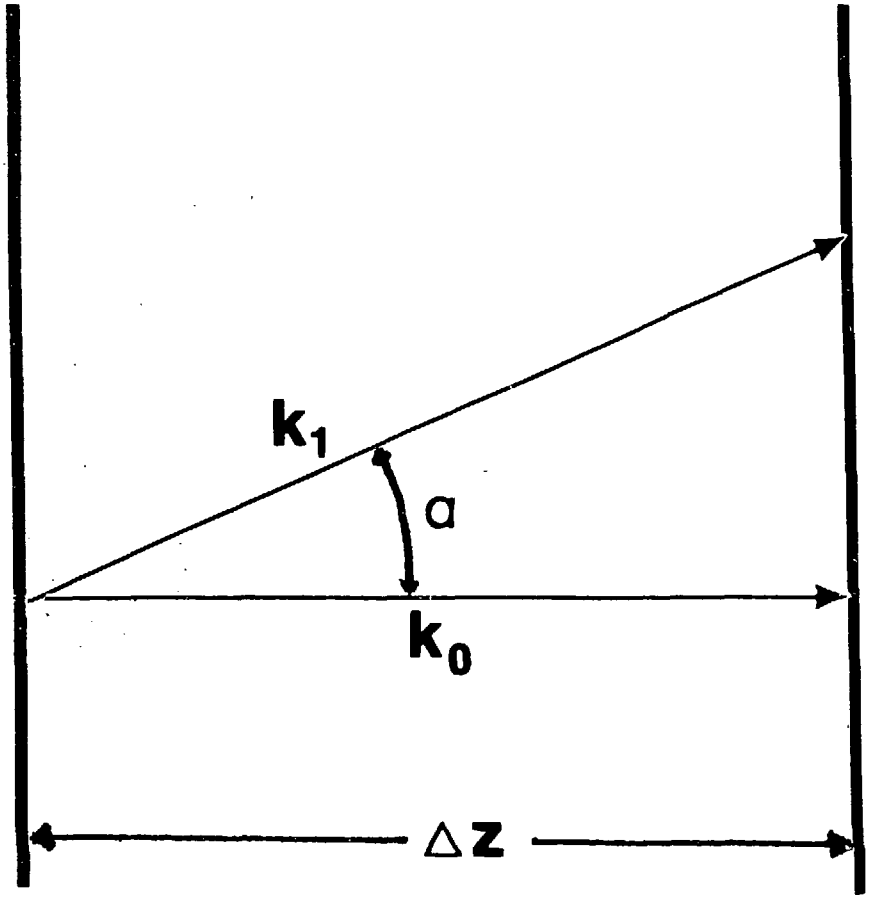


Fig. 3-4.



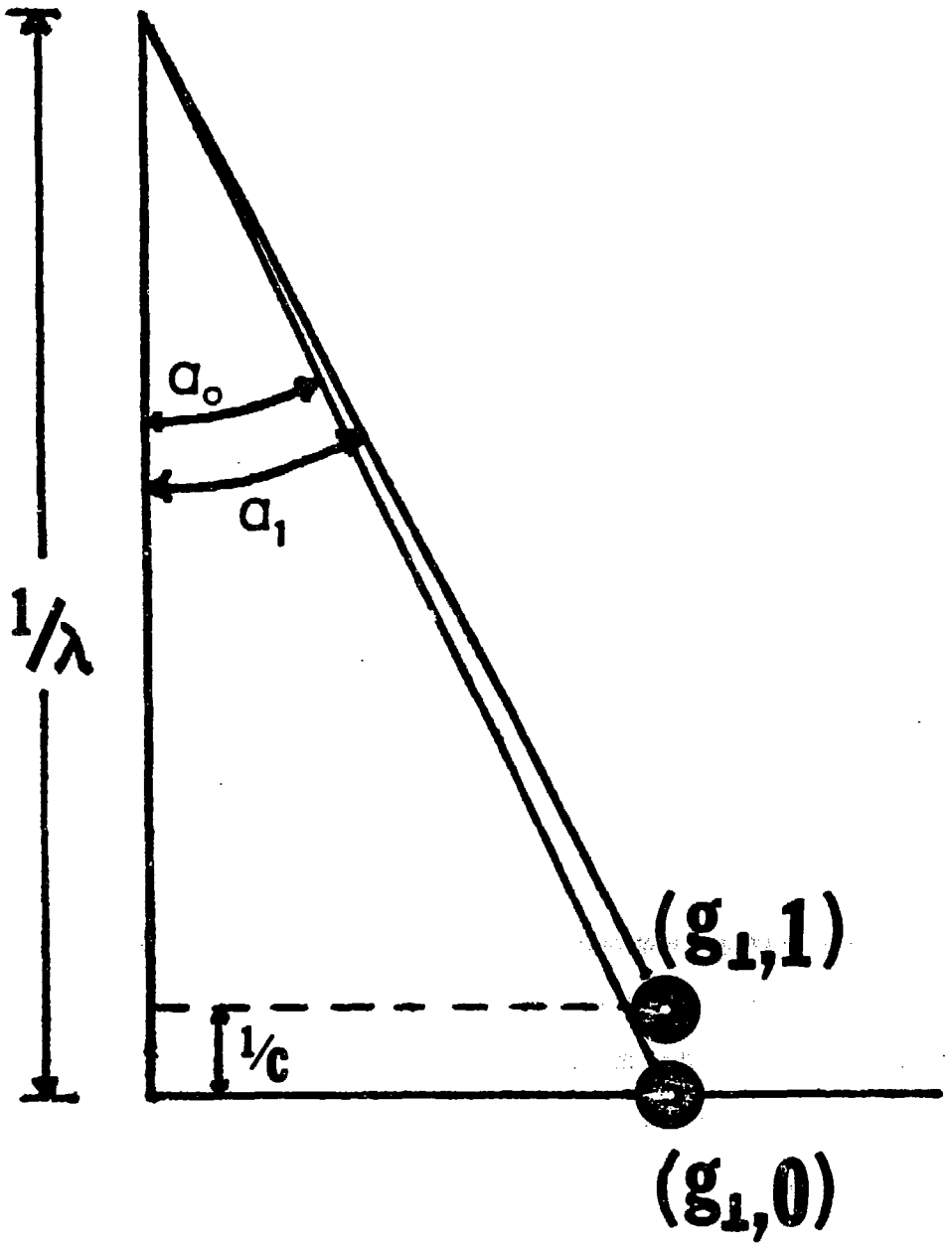
XBB 851-792

Fig. 3-5.



XBL 851-979

Fig. 3-6.



XBL 851-981

Fig. 3-7.

Chapter 4

On The Improved Phasegrating Method

Abstract

A new method for computing scattering amplitudes in High Resolution Transmission Electron Microscopy has been examined. The method which is called the Improved Phasegrating (IPG) method is shown to produce reasonable results only for very small specimen thicknesses and diverges for thicknesses larger than 20 Å – 40 Å in copper [001] for accelerating voltages between 200kV – 1MV. The validity of the method is discussed and is shown to depend on electron wavelength, slice thickness,the number of reflections that are included in the calculation and the choice of specimen. It is also shown that the method does not readily allow for slice thicknesses smaller than the specimen periodicity along the incident electron beam direction.

1. Introduction

The ability of present multislice calculations to include upper Laue layer interactions were studied in chapter 3 and it was shown that the Second Order Multislice (SOM) method allows for a larger slice thickness than the conventional first order multislice method employing fast Fourier transforms (the FFT method) while still including ULL effects. In order to include ULL effects into the SOM method and the FFT method it is necessary to use slice thicknesses smaller than the crystal periodicity (c) in the electron beam direction. The improved phasegrating method proposed by Van Dyck [1] allows for the inclusion of higher order zones even when the slice thickness is equal to c. However, no results using this method have been published and it is not clear that the IPG method presents an alternative to existing methods. This work was undertaken in a hope to shed light on the applicability of the method. The model system is a specimen of Cu [001], and the formulation presented in Chapter 3 is used. In order to facili-

tate the reading, some of the theory in Chapter 3 is repeated below.

2. Theory

The improved phasegrating method is an approximate solution of the following modified Schrodinger equation [2]:

$$\frac{\partial \phi}{\partial z} = i\sigma V\phi + \frac{i\lambda}{4\pi}\nabla_\perp^2\phi = i\sigma V\phi + \Delta\phi \tag{2.1}$$

where

$$\sigma = \frac{2\pi m e \lambda}{h^2} \tag{2.2}$$

and

$$\nabla_\perp^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{2.3}$$

and the wave function ψ has been written as a modulated plane wave of the form:

$$\psi(r) = \phi(r)\, e^{ik\cdot r} \tag{2.4}$$

V is the crystal potential in volts. If the effect of the potential is larger than that of Δ, it is appropriate to start from an exact solution in V and treat Δ as a perturbation. Van Dyck [1] suggests writing the wave function as a modulated phasegrating of the form:

$$\phi(x,y,z) = \exp\left\{ i\sigma \int_0^z V(x,y,z')dz' \right\} \theta(x,y,z) \tag{2.5}$$

which, after substitution into (2.1) yields:

$$\theta(x,y,z) = \theta(x,y,o) + \frac{i\lambda}{4\pi}\int_0^z dz' \left\{ \nabla_\perp^2\theta(z') + i\sigma\Delta z[\nabla_\perp^2 V_p(z')]\theta(z') \right. \tag{2.6}$$

$$\left. + \; 2i\sigma\Delta z \nabla_\perp V_p(z') \cdot \nabla_\perp \theta(z') + (i\sigma\Delta z)^2\, \theta(z')[\nabla_\perp V_p(z')]^2 \right\}$$

The first order perturbation result for theta becomes

$$\theta(z) \approx \theta(o) + \frac{i\lambda\Delta z}{4\pi}\left\{ \nabla_\perp^2\theta(o) + i\sigma\Delta z\theta(o)\int_0^z dz' \right. \tag{2.7}$$

$$\left[\frac{1}{\Delta z}\nabla_\perp^2 V_p + i\sigma\Delta z\frac{1}{\Delta z}(\nabla_\perp V_p)^2\right] + 2i\sigma\Delta z\nabla_\perp\theta(o)\cdot\int_o^z dz'\,\frac{1}{\Delta z}\nabla_\perp V_p\Bigg\}$$

It was shown in Chapter 3 that the integrals appearing in (2.7) can be expressed as follows:

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}\nabla_\perp^2 V_p dz' = -(2\pi)^2\sum_{h,k} e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right) \tag{2.8}$$

$$\times\sum_\ell V(h,k,\ell)\frac{n}{2\pi i\ell}\left(e^{2\pi i\ell z_m/c}\frac{\sin\pi\ell/n}{\pi\ell/n}-1\right)$$

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}\frac{\partial V_p}{\partial x}dz' = 2\pi i\sum_{h,k} e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\left(\frac{h}{a}\right)\sum_\ell V(h,k,\ell)\frac{n}{2\pi i\ell}\left(e^{2\pi i\ell z_m/c}\frac{\sin\pi\ell/n}{\pi\ell/n}-1\right) \tag{2.9}$$

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}\frac{\partial V_p}{\partial y}dz' = 2\pi i\sum_{h,k} e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\left(\frac{k}{b}\right)\sum_\ell V(h,k,\ell)\frac{n}{2\pi i\ell}\left(e^{2\pi i\ell z_m/c}\frac{\sin\pi\ell/n}{\pi\ell/n}-1\right) \tag{2.10}$$

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}\left[\left(\frac{\partial V_p}{\partial x}\right)^2+\left(\frac{\partial V_p}{\partial y}\right)^2\right]dz' =$$

$$-(2\pi)^2\sum_{\substack{h,k\\h',k'}}\left(\frac{hh'}{a^2}+\frac{kk'}{b^2}\right)e^{2\pi i\left(\frac{h+h'}{a}x+\frac{k+k'}{b}y\right)}$$

$$\times\sum_{\ell\ell'}V(h,k,\ell)V(h',k',\ell')\frac{n}{2\pi i\ell}\,\frac{n}{2\pi i\ell'}\Bigg[e^{2\pi i(\ell+\ell')z_m/c}\,\frac{\sin\pi(\ell+\ell')/n}{\pi(\ell+\ell')/n}$$

$$-e^{2\pi i\ell z_m/c}\,\frac{\sin\pi\ell/n}{\pi\ell/n}-e^{2\pi i\ell' z_m/c}\,\frac{\sin\pi\ell'/n}{\pi\ell'/n}+1\Bigg] \tag{2.11}$$

where the slice thickness Δz is equal to c/n. Equation (2.11) simplifies in two cases :

i) the limit $\Delta z \to 0$

$$\frac{1}{\Delta z}\int_{z_m}^{z_m+\Delta z}\left[\left(\frac{\partial V_p}{\partial x}\right)^2+\left(\frac{\partial V_p}{\partial y}\right)^2\right]dz' = -\frac{(2\pi)^2}{3}\left[\left(\sum_{h,k}\frac{h}{a}V(h,k,0)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\right)^2\right.$$

$$\left.+\left(\sum_{h,k}\frac{k}{b}V(h,k,0)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\right)\right] \tag{2.12}$$

ii) $n = 1$

$$\frac{1}{c}\int_{z_m}^{z_m+c}\left(\frac{\partial V_p}{\partial x}\right)^2 dz' = \left[\frac{1}{3}+\frac{z_m}{c}\left(1+\frac{z_m}{c}\right)\right]\left[\sum_{h,k} 2\pi i\left(\frac{h}{a}\right)V(h,k,0)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\right]^2 \qquad (2.13)$$

$$+ 2\left[\sum_{h,k} 2\pi i\left(\frac{h}{a}\right)V(h,k,0)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\right]\left[\sum_{h,k} 2\pi i\left(\frac{h}{a}\right)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\sum_{\ell\neq 0}\left(\frac{1}{2\pi i\ell}\right)^2 V(h,k,\ell)\right]$$

$$+ \left(\frac{1}{2}+\frac{z_m}{c}\right)\left[\sum_{h,k} 2\pi i\left(\frac{h}{a}\right)V(h,k,0)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\right]$$

$$\times\left[\sum_{h,k} 2\pi i\left(\frac{h}{a}\right)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\sum_{\ell\neq 0}\frac{i}{2\pi\ell}V(h,k,\ell)\right]$$

$$+ \left[\sum_{h,k} 2\pi i\left(\frac{h}{a}\right)e^{2\pi i\left(\frac{hx}{a}+\frac{ky}{b}\right)}\sum_{\ell\neq 0}\frac{1}{2\pi i\ell}V(h,k,\ell)\right]^2$$

The equivalent expression for the derivative with respect to y follows from (2.13).

2.1 Validity of the Improved Phasegrating Method

Equation (2.7) represents the second term in a series expansion for theta, the first being $\theta_o = \theta(o)$. In order to get a feeling for the error introduced by the truncation of the series it is instructive to look at the Fourier transform of (2.7). To simplify the expressions one can include only terms where $\ell = 0$ and ignore terms with $\ell \neq 0$, that is ignore the effect of higher order zones. Using (2.8) through (2.11) one obtains the following for the Fourier transform of theta.

$$\theta(\mathbf{g},z_m+\Delta z) = \theta(\mathbf{g},z_m) - i\pi\lambda\Delta z\{\mathbf{g}^2\theta(\mathbf{g},z_m) \qquad (2.1.1)$$

$$+ (i\sigma\Delta z)\left(\tfrac{1}{2}+\frac{z_m}{\Delta z}\right)\sum_{\mathbf{g}'}\mathbf{g}'^2\theta(\mathbf{g}-\mathbf{g}',z_m)\,V(\mathbf{g}',0)$$

$$+ (i\sigma\Delta z)^2\left[\tfrac{1}{3}+\frac{z_m}{\Delta z}\left(1+\frac{z_m}{\Delta z}\right)\right]\sum_{\mathbf{g}',\mathbf{g}''}\mathbf{g}''\cdot(\mathbf{g}'-\mathbf{g}'')\,\theta(\mathbf{g}-\mathbf{g}',z_m)$$

$$\times V(\mathbf{g}'',0)\,V(\mathbf{g}'-\mathbf{g}'',0)$$

$$+ 2(i\sigma\Delta z)\left(\tfrac{1}{2}+\frac{z_m}{\Delta z}\right)\sum_{\mathbf{g}'}\mathbf{g}'\cdot(\mathbf{g}-\mathbf{g}')\,\theta(\mathbf{g}-\mathbf{g}',z_m)\,V(\mathbf{g}',0)\}$$

As seen from the expression above, the convergence of the series depends on the wavelength, slice thickness, strength of the crystal potential and the maximum reciprocal wavevector (g_{max})

included in the calculation. By inserting the first order perturbation result for theta into (2.6) it is possible to get a second order result for theta and obtain the equivalent of (2.1.1). However the higher order expressions quickly become very complicated and it is difficult to obtain a useful criterion for the validity of the expansion. The part of the expansion that does not involve the potential goes as

$$\theta(g,z_m+\Delta z) = (1-i\pi\lambda\Delta z g^2 + \frac{(i\pi\lambda\Delta z)^2}{2} g^4)\,\theta(g,z_m) \qquad (2.1.2)$$

and this corresponds to the expansion of the propagator that shows up in the formulation of the real space (RSP) method in Chapter 2. However since there is no such requirement that the intensity of each $\theta(h,k)$ remains unchanged from slice to slice, only that $\sum_{h,k} |\theta(h,k)|^2 = 1$, one cannot apply the same condition as for the RSP method. Also there are terms involving the strength of the crystal potential which complicate matters. In the limit that the interaction parameter σ or the strength of the potential go to zero the criterion of validity becomes

$$K \equiv \lambda\Delta z g^2 \ll 1/\pi \qquad (2.1.3)$$

That this is not a sufficient restriction will become evident from the results in paragraph 4.

3. Procedure

The theory outlined in the previous paragraph was implemented in computer programs and run on a CDC 7600. The model system is copper in the [001] orientation and the calculation was performed for an accelerating voltage of 200 kV and 1000kV. Because of the practical problems associated with the calculation of the expression in (2.11), the only slice thickness considered was 3.6 Å, which corresponds to the specimen periodicity in the incident electron beam direction. The difficulties associated with use of a smaller slice thickness are discussed in paragraph 5.

4. Results

The results consist of a series of comparisons between the conventional (FFT) multislice method, the phasegrating method and the improved phasegrating method and are shown in Figs. 4-1 through 4-3. In Fig. 4-1 the accelerating voltage is 200 kV while in Figs. 4-2 and 4-3 the voltage is 1 MV. As seen from the figures the amplitudes calculated by the IPG method begin to diverge after approximately 20 – 40 Å depending on λ and g_{max}. In Fig. 4-1 and 4-2 the slice thickness is equal to c allowing the inclusion of higher order zones into (2.11). The contribution by terms given by $\ell \neq 0$ is essential for including out of the zone effects, but is small compared to the term $\ell, \ell' = 0$ and was ignored in (2.11) such that amplitudes and phases in the case of $\Delta z < c$ could be computed. Thus Fig. 4-3 shows amplitudes calculated for three different values of g_{max} (2.0 $Å^{-1}$, 2.8 $Å^{-1}$ and 4.0 $Å^{-1}$) while varying Δz (n) as to keep the value of K constant (0.126 and 0.063). The results indicate that varying the slice thickness while keeping the wavelength and g_{max} constant has little effect.

5. Discussion

It is clear from Figs. 4-1 through 4-3 that the improved phasegrating method fails to give reasonable results beyond a thickness of 20 Å – 40 Å for the combinations of wavelength, slice thickness and sampling interval that were used. As expected the method works better for higher voltages where the wavelength gets smaller and the propagator becomes less important. Surprisingly, reducing the slice thickness does not appear to increase the accuracy of the method as (2.1.1) would indicate, although the results in Fig. 4-3 are slightly inaccurate since the contribution of higher order zones were ignored in the term involving the square of the derivatives. Apart from wavelength and slice thickness, the convergence of the expansion depends also on sampling interval (g_{max}) and on the strength of the crystal potential. The dependence on sampling is clearly seen in the results, where extending the calculation further into reciprocal space causes the amplitudes of diffracted beams to diverge at a decreasing thickness. The Fourier coefficients of the potential are determined by choice of specimen and was

not varied. Copper with an atomic number of 29 represents a compromise between heavy and light elements and the results serve as a useful guide for other elements.

By neglecting the terms involving the potential it is possible to set an upper limit on the value of K for which the expansion for theta converges. This limit corresponds to $K_{max} = 1/3$, but it is clear from the calculations that when the potential is included more severe restrictions are imposed, restrictions that now also depends on the the crystal potential. However it is very difficult to find a useful expression in this case.

Because the method rapidly diverges, it might only be of academic importance to consider the extent to which upper Laue layer effects are included in the improved phasegrating method (IPG). Compared to commonly used multislice methods that rely on small slice thickness to include higher order interactions, see Chapter 2, the IPG method includes higher order effects also in the case where the slice thickness is equal to the crystal periodicity along the incident electron beam direction. However because of the crossterms that appear in (2.11), the method becomes impractical when n is different from 1. In the case of $n \neq 1$ the calculation of (2.11) requires a minimum of 10^9 operations (convolution over 6 indices) for 32^3 sampling points, and needs to be repeated n times. Even with the use of modern day super computers this is hardly a small calculation.

6. Conclusion

The results show that due the limited range of validity of the improved phasegrating method it is not suitable for computation of scattering amplitudes in HRTEM. Even though the results in Fig. 4-3 show very little dependence on Δz, the validity of the method depends on the slice thickness and the calculation should improve with smaller slice thicknesses. However computational considerations prohibits the proper use of arbitrary slice thickness and prevents further investigation into thickness dependence. An upper limit of $K = \lambda \Delta z g_{max}^2 = 1/\pi$ is necessary to give convergence to the series expansion for theta, but it is clear that the sampling interval in combination with the strength of the crystal potential is more important in

determining the conditions for convergence. However an exact expression for convergence was not found.

7. References

[1] D. Van Dyck, J. Microscopy, 132, (1983), p. 31.

[2] D. Van Dyck, J. Microscopy, 119, (1980), p. 141.

FIGURE CAPTIONS

Fig. 4-1. Amplitude and phase (in units of π) vs. thickness for the reflections 000, 200 and 440 in copper [001]. Calculations are performed by the conventional (FFT) multislice method (–), the phasegrating (PG) approximation (++) and the improved phasegrating (IPG) method (**). Accelerating potential is 200 kV and the crystal potential has been sampled out to 2.0 $Å^{-1}$. The slice thickness is 3.6 Å corresponding to n = 1.

Fig. 4-2. As in Fig. 4-1 except that the accelerating potential has been set to 1 MV.

Fig. 4-3. Amplitude vs thickness for the reflection 200 for two sets of values of K. In the first column K = 0.126 and in the second column K = 0.063. The value of g_{max} has been set to 2.0 $Å^{-1}$, 2.8 $Å^{-1}$ and 4.0 $Å^{-1}$ in the first, second and third row respectively. The slice thickness required to keep K constant is indicated by the value of n (Δz = 3.6 Å/n).

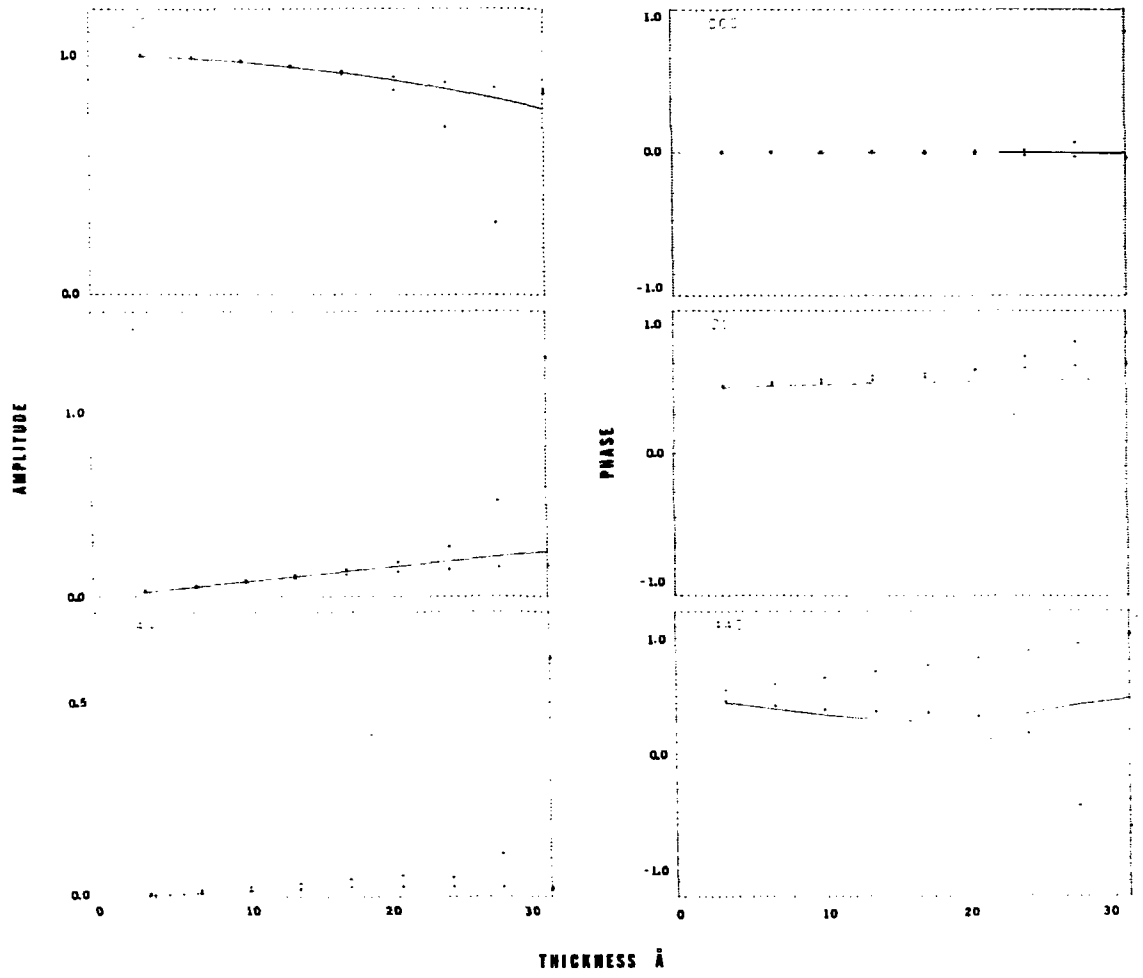
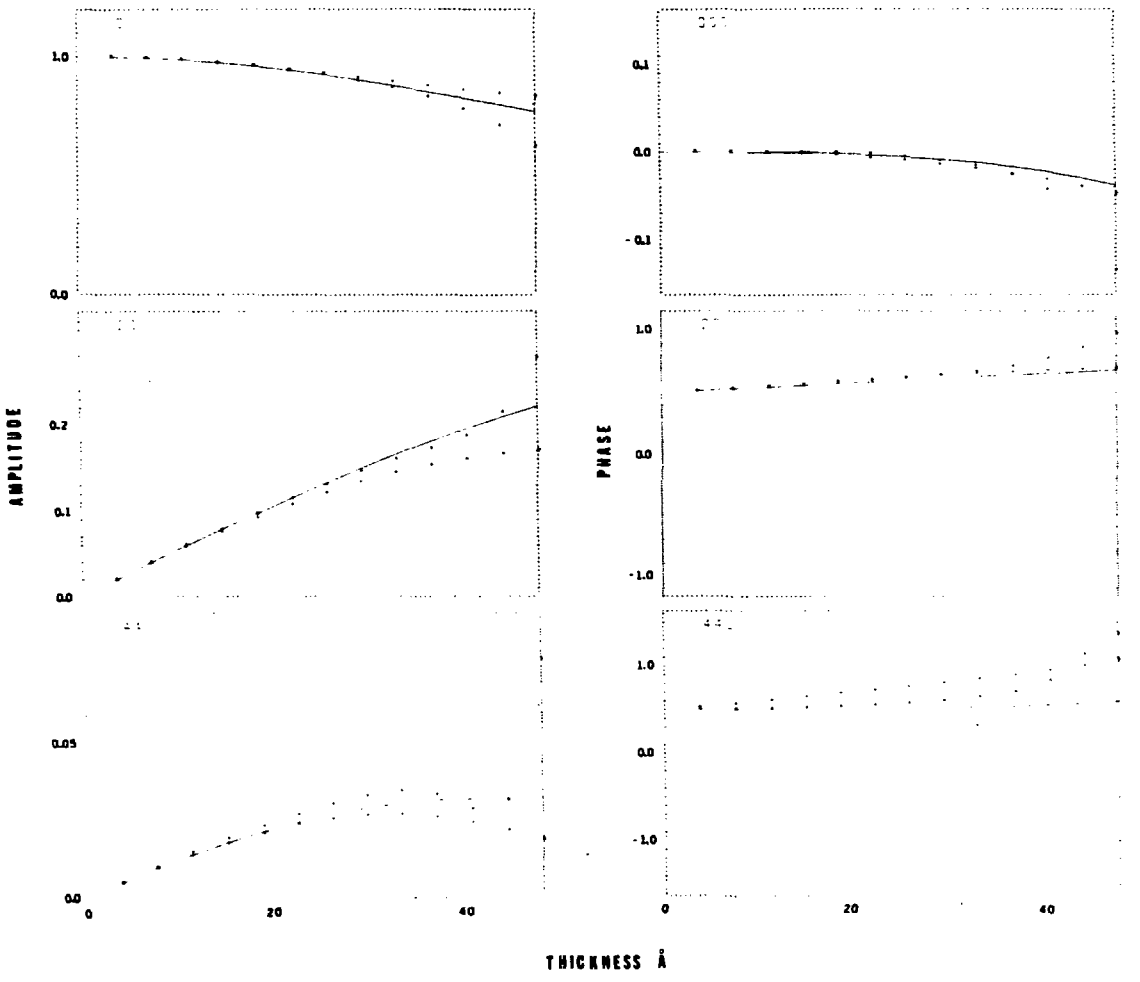
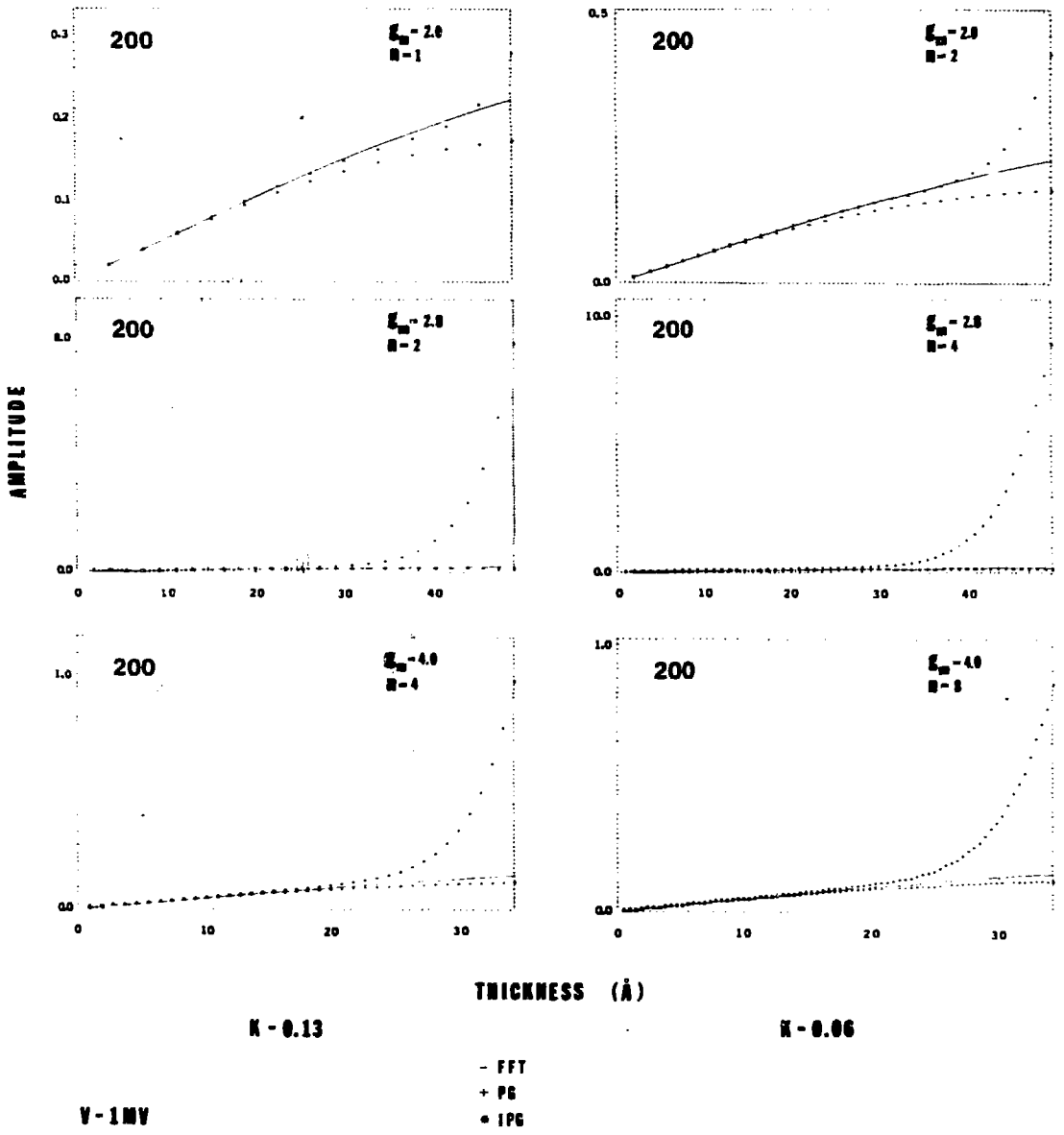

AMPLITUDE
PHASE
THICKNESS Å
V - 200 kV
G_{max} - 2. Å^{-1}
- FFT
+ PG
* IPG
XBL 852-1173

Fig. 4-1.



XBL 852-1172

Fig. 4-2.



XBL 852-1171

Fig. 4-3.

Chapter 5

The Effect of Amorphous Surface Layers on Images of Crystals In High Resolution Transmission Electron Microscopy

Abstract

The effect of a surface layer of amorphous silicon dioxide on images of crystalline silicon has been investigated for a range of crystal thicknesses varying from zero to $2\frac{1}{2}$ times that of the surface layer. It is shown that an amorphous surface produces fluctuations in image contrast which introduces difficulties in the interpretation of defects in very thin specimens. These difficulties are less pronounced but still present in thicker crystals. It is also shown that an edge smoothly approaching zero thickness produces an image that changes gradually from crystalline to amorphous character.

Introduction

It is well known that most materials under investigation in the electron microscope form a surface oxide layer or an amorphous layer which can affect the resulting image, see Fig. 5-1 and Fig. 5-2. In fact, a noisy image is often attributed to surface effects without further explanation. Furthermore, very little work has been done to evaluate the extent to which an amorphous surface actually influences the nature of the image.

Krakow [1] calculated the image of crystalline gold showing reasonable agreement with an experimental image when the top gold layer was substituted with a layer of randomly arranged gold atoms. However, when the same substitution was made for the bottom layer the computed image lost its crystalline appearance completely. In another study, Anstis et. al. [2] calculated the effect of an amorphous surface layer on the images of a 30 degrees partial dislocation in silicon. They showed that the amorphous surface makes it impossible to distinguish between

the shuffle and glide models of the dislocation core in a 28 Å thick crystal when the surface layer was 20 Å thick. Similar results have been quoted by Bourret et al. [3].

The object of the present research program is to fully understand the contribution of such surface layers to high resolution image contrast, examining in particular the effect of amorphous layer thickness relative to the thickness of the crystal on which it resides. The problem of specifying suitable atomic coordinates in an amorphous material is addressed by choice of a model system, amorphous silicon dioxide on silicon, for which atomic coordinate data exists [4]. Calculations are carried out which simulate high resolution electron micrographs of a single crystal of silicon and then a $\Sigma = 9$ bicrystal of silicon, the latter having a fully specified periodic defect structure [5]. Atom coordinates for the $\Sigma = 9$ grain boundary were obtained by relaxation of a coincident Site Lattice [6]. Images of crystals with different thicknesses are simulated for a constant thickness amorphous oxide film and compared to the case where no oxide is present.

Image Calculation Procedure

The images presented in this work were simulated by a multislice computation routine [7] written to run on a CDC 7600 such that any number of different slices and slice thicknesses could be used. In addition, any number of sampling points up to 256^2 can be used and these can be arranged in any square or rectangular array. All adjustable microscope parameters were set to values characteristic of a JEOL JEM 200CX at Scherzer defocus ; $c_s = 1.2$ mm, $\Delta f = -660$ Å, delta = 50 Å, alpha = 0.5 mrad.

Atomic coordinate data for the amorphous silicon dioxide layers was taken from the work of Bell and Dean [4] who used a random network theory model with a mean Si-O-Si bond angle of 153 degrees. They generated the positions of 614 atoms within a volume element of $31 \times 23 \times 22$ cubic Angstroms; this model closely matched the experimental radial distribution function to 8 Å detail. These atom positions were used in the calculations by dividing the above volume element into 7 slices of 3.14 Å thickness each, and using only the middle 5 slices for a total layer thickness of 15.7 Å.

An oxide film was then hypothetically attached to a single crystal of silicon in [110] orientation by covering an area of 5×5 crystalline unit cells (Fig. 5-3). This required the selection of a rectangular slab of oxide with dimensions 27.1 Å by 19.2 Å which retained the same thickness of 15.7 Å. Oxide films on the top and bottom surfaces were positioned such that they did not artificially superimpose; this was accomplished by simple translation of the bottom layer by a/2[111] with respect to the top layer.

The potential from each layer of material in this sandwich was then calculated from its corresponding electron scattering factors out to a maximum reciprocal space dimension of 3.0 $Å^{-1}$. This corresponds to an array of 182×114 sampling points such that every fifth sampling point along any direction was a Bragg reflection from the silicon matrix.

The oxide film attachment to the silicon bicrystal was accomplished in a slightly different way. The model of the silicon bicrystal shown in Fig. 5-4 covers an area larger than the size of the oxide such that a rectangular slab with oxide in the center had to be used. Two surface oxide layers were created, one 15.7 Å thick and another 9.4 Å thick. With the 9.4 Å thick surface oxide, unique top and bottom surface layers were constructed from the model of the oxide by simply using different slices for the two layers, whereas in the case of the 15.7 Å thick oxide the top and the bottom layers were rotationally displaced 180 degrees. Calculations of the resulting sandwiches were performed using 256×128 sampling points.

Results

The results of the calculations are shown in Figs. 5-5 through 5-9. In Fig. 5-5 a set of computed images of the amorphous silicon dioxide is shown together with the corresponding projected potentials. Figure 5-6(a) shows a set of images calculated for perfect silicon with thickness varying from 0. to 38.4 Å and a constant total thickness of amorphous top and bottom surface layers of 31.4 Å. The image of perfect crystal with no surface oxide varies negligibly with thickness in the range 8 – 38 Å and is shown in Fig. 5-6(b) for comparison.

Figure 5-7 shows the projected potential for the silicon bicrystal, while Fig. 5-8 and Fig. 5-9 show computed images. The $\Sigma = 9$ grainboundary is a pure tilt-boundary with a rotation angle between the two grains of 38.9 degrees. The boundary plane is (122). Figure 5-4 shows a schematic model of the unit cell used in the calculation. The thickness of the silicon varies from 8 – 38 Å and the total thickness of amorphous top and bottom surface layers is 31.4 Å and 18.8 Å respectively. These images should be compared to the set of images calculated for silicon $\Sigma = 9$ without surface oxide shown in Fig. 5-8. The rotational displacement used in the construction of the bottom surface oxide layer 15.7 Å thick produces an artificial mirror-symmetry when the top and bottom layer is viewed in projection, which can be noticed when the bicrystal is very thin (7.7 Å), but otherwise has no effect on the results. Because of the smaller size of the oxide ,only the center portion of the images was considered.

Discussion

i) Perfect Crystal

It is clear from Fig. 5-6(a) that the image changes from amorphous to crystalline in a gradual way. In fact, in the thin region of silicon, 8 – 16 Angstroms, parts of the image appear amorphous, while another part appears more crystalline. Thus "islands" of crystalline material appearing in an apparently amorphous area near the edge of a foil could possibly be due to variations in the surface oxide. For "thicker" areas, the surface layers produce irregularities in the image, such as wavy lines of atoms and fluctuating contrast. As the thickness of the silicon increases beyond the total thickness of the surface oxide, the image is very closely that of a pure crystal with no oxide. In the case of silicon this corresponds experimentally [8] to a thickness of approx. 40 Angstroms, see Fig. 5-2.

ii) Silicon Bicrystal

In this case, the most interesting result is the appearance of displaced atoms at the boundary plane which has significant implications for the general analysis of atomic positions near any such defect. Various models are usually possible for any given type of defect and the goal of HREM is to determine the correct model by matching computed images to experimental data. Sometimes the change in the image from model to model is subtle and "noise" in the image can make it impossible to distinguish between two models with any confidence. Unfortunately the presence of a surface layer will add to the difficulty of determining atom positions. As can be seen from Fig. 5-8, atoms may or may not show up due to surface contamination. The model of the $\Sigma = 9$ boundary has a periodicity of 11.5 Angstroms in the boundary direction, and there are two "unit cells" shown in Fig. 5-4. However, due to the presence of the surface oxide, the image does not reflect this periodicity and certain atoms could easily be thought missing, thus preventing a correct interpretation of the image. Naturally as the crystal gets thicker, the surface becomes less important, and for thicknesses slightly greater than the total surface-thickness, the surface effects have become negligible except at the dislocation cores where there still are small differences in contrast.

Conclusion

Because of dynamical interaction between Bragg scattered reflections HREM images are usually considered directly interpretable only for very thin specimens, typically less than 50 Angstroms. For thicker crystals computer matching of images are required to get detailed information down to 2 – 4 Angstroms resolution, but inelastic scattering may make even this approach unreliable.

In the presence of an amorphous surface layer, it is clear that the image of a very thin specimen is adversely affected by the surface. This may not represent a severe problem in imaging a perfectly crystalline structure, since one can always average over a number of identical unit cells to obtain a less noisy image. However the correct interpretation of a single

isolated defect may be rendered impossible by the presence of an amorphous surface layer. With such surface layers of the order of 20 Ångstrom thickness, the crystalline material should be at least 40 Ångstroms to provide an interpretable image. It should also be noted that a uniform wedge-shaped edge produces an image that gradually changes from crystalline to amorphous character. A sharp change from crystalline to amorphous indicates an edge that is approaching zero thickness in a discontinuous way.

References

[1] W. Krakow, Inst. Phys. Conf. Ser. No. 61; Electron Microscopy and Analysis, (1981), p. 365.

[2] G.R. Anstis, M.J. Goringe, J.L. Hutchinson and B.J. Muggridge, Inst. Phys. Conf. Ser. No. 68; Electron Microscopy and Analysis, (1983), p. 169.

[3] A. Bourret, J. Desseaux and C. D'Anterroches, Inst. Phys. Conf. Ser. No. 60; Microscopy of Semiconducting Materials, (1981), p. 9.

[4] R. J. Bell and P. Dean, Phil. Mag. 25, (1972), p. 1381.

[5] W. Bollmann, *Crystal Defects and Crystalline Interfaces*, Springer-Verlag, New York, (1970), p. 215.

[6] J.H. Rose, Private Communication.

[7] J. Mazur, R. Gronsky and J. Washburn, Inst. Phys. Conf. Ser. No. 67; Microscopy of Semiconducting Materials, (1983), p. 77.

FIGURE CAPTIONS

Fig. 5-1. High resolution electron micrograph of silicon in [001] orientation showing transition from crystalline to amorphous structure at a thin edge.

Fig. 5-2. Native oxide on 3 degrees off (111) Si surface; thickness 20 ± 3 Å.

Fig. 5-3. Projected unit cell of silicon in [110] orientation.

Fig. 5-4. Model of $\Sigma = 9$ grainboundary in silicon. The model shows the structural periodicity along the boundary plane.

Fig. 5-5. Computed electron micrographs of amorphous silicon dioxide. The projected potentials are shown for comparison.

Fig. 5-6(a). Computed electron micrographs of perfect silicon with native oxide. The thickness of amorphous top and bottom surface oxide is 31.4 Å. The thickness of the silicon matrix varies from 0. to 38.4 Å.

Fig. 5-6(b). Computed electron micrograph of perfect silicon,thickness 23 Å.

Fig. 5-7. Computed projected potential for silicon bicrystal from model in Fig. 5-4.

Fig. 5-8. Computed electron micrographs of silicon bicrystal with native surface oxide. The total thickness of the surface layers is held constant at 31.4 Å, while the thickness of the bicrystal varies from 8 – 38 Å.

Fig. 5-9. Computed electron micrographs of silicon bicrystal with native surface oxide. The total thickness of the surface layers is held constant at 18.8 Å, while the thickness of the bicrystal varies from 8 – 38 Å.

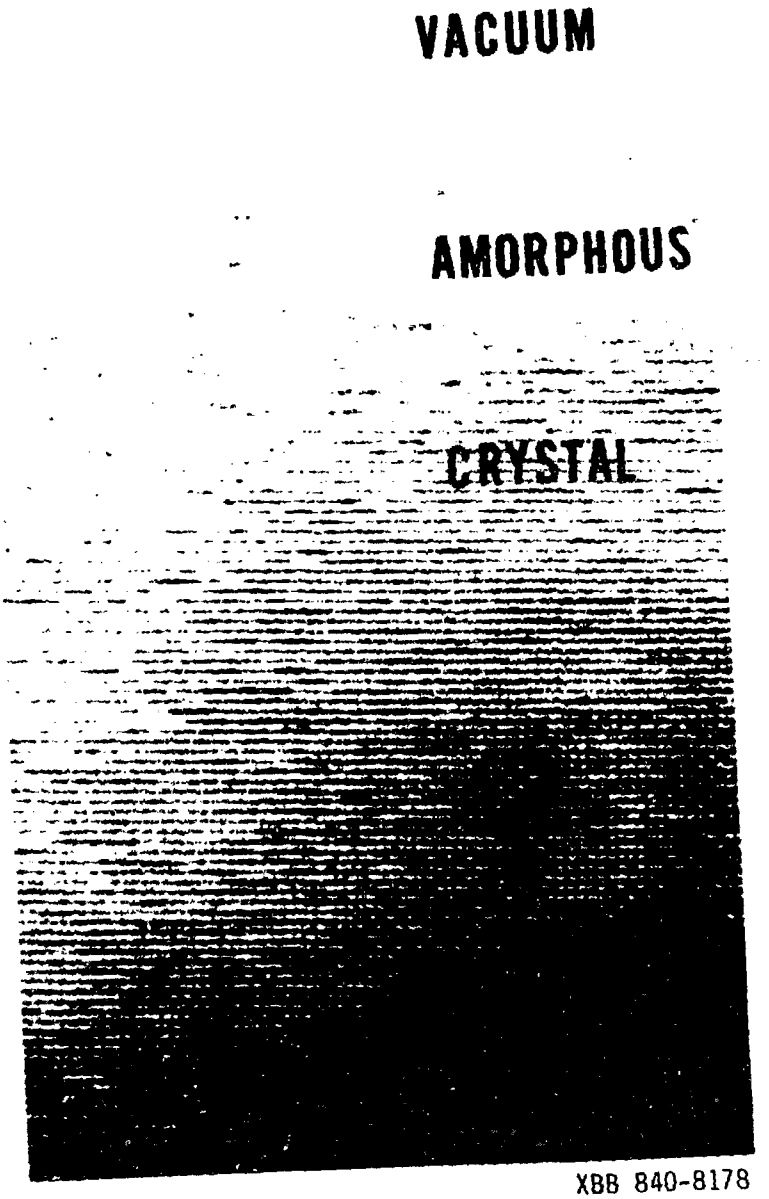
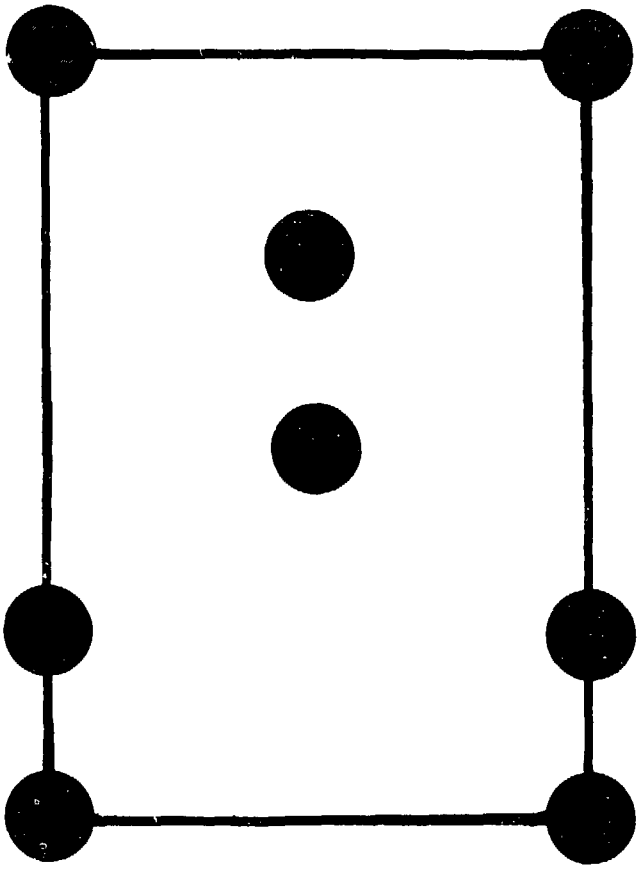


Fig. 5-1.

XBB 834-3022

Fig. 5-2.

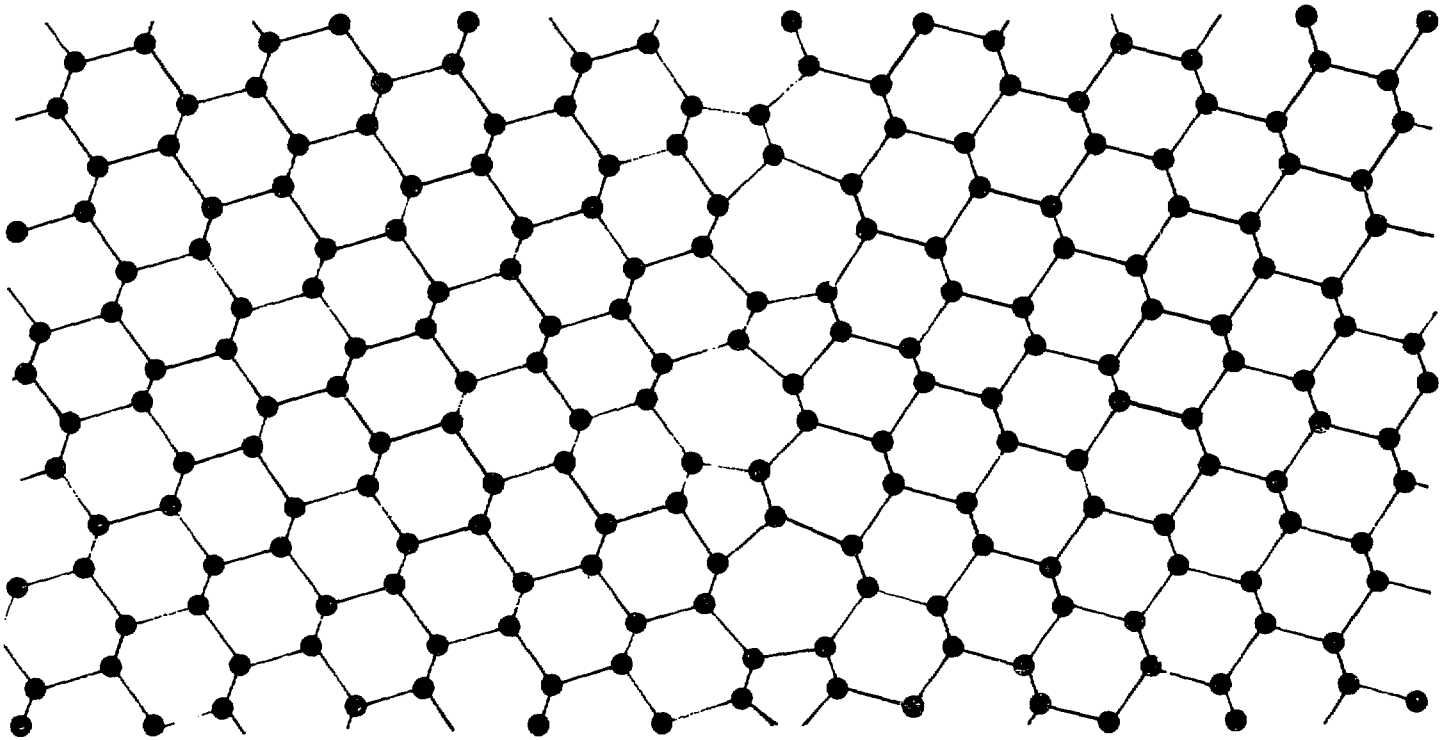


Si 110

XBL 8410-4105

Fig. 5-3.

$\Sigma=9$



XBL 8410-4106

Fig. 5-4.

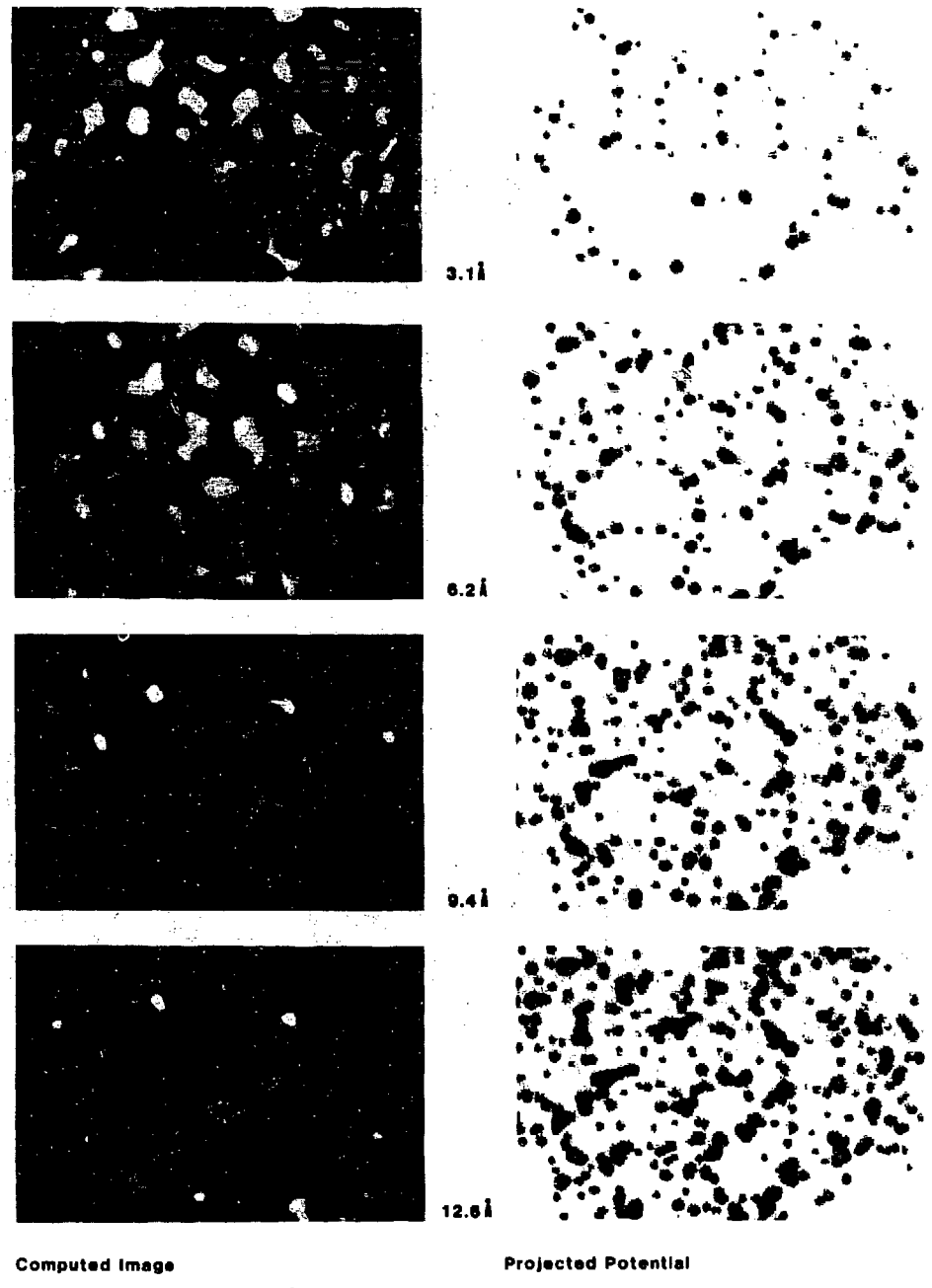

AMORPHOUS SILICON DIOXIDE
3.1Å
6.2Å
9.4Å
12.6Å
Computed Image
Projected Potential
XBB 847-5399A
Roar Kilaas

Fig. 5-5.

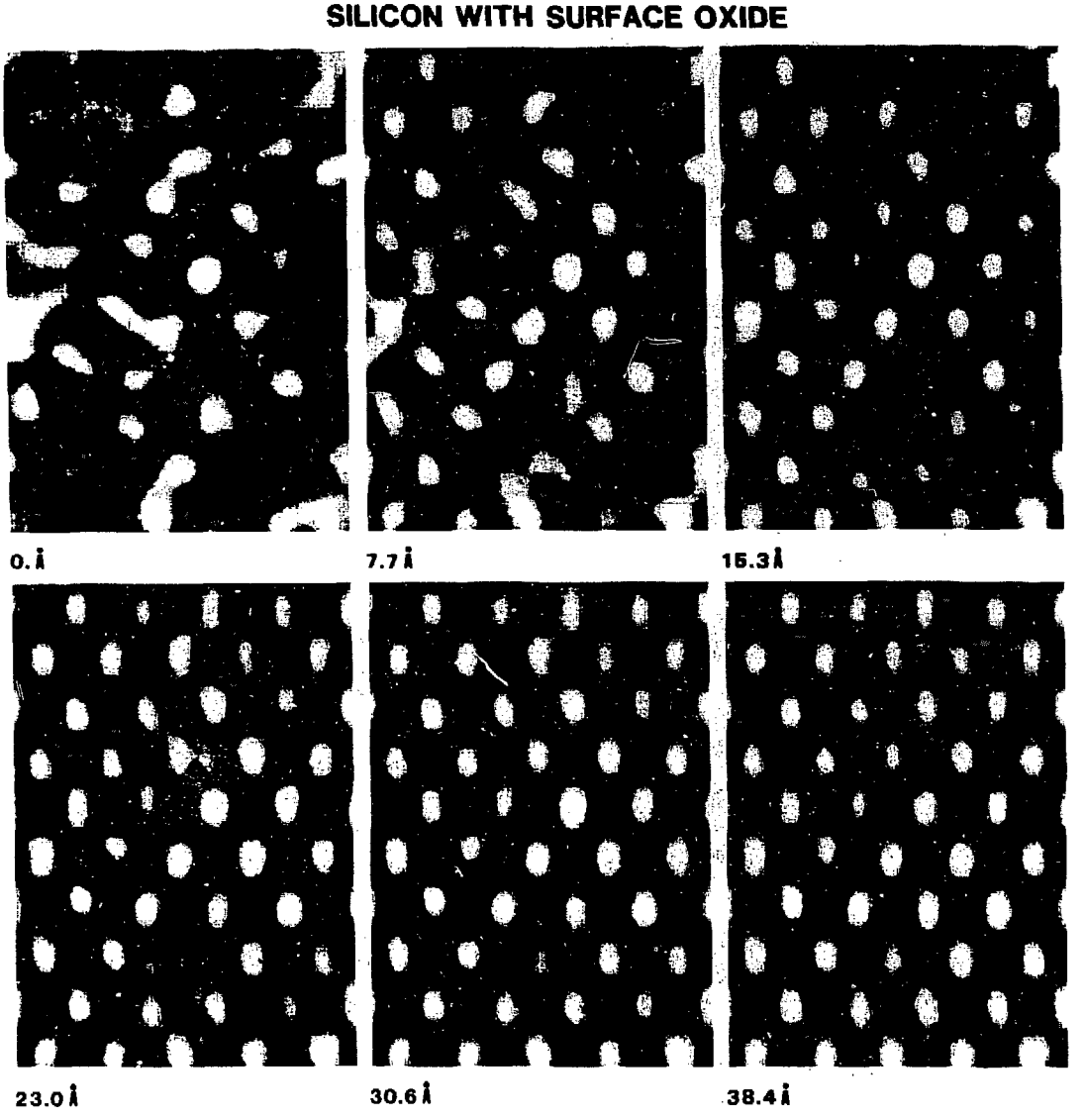
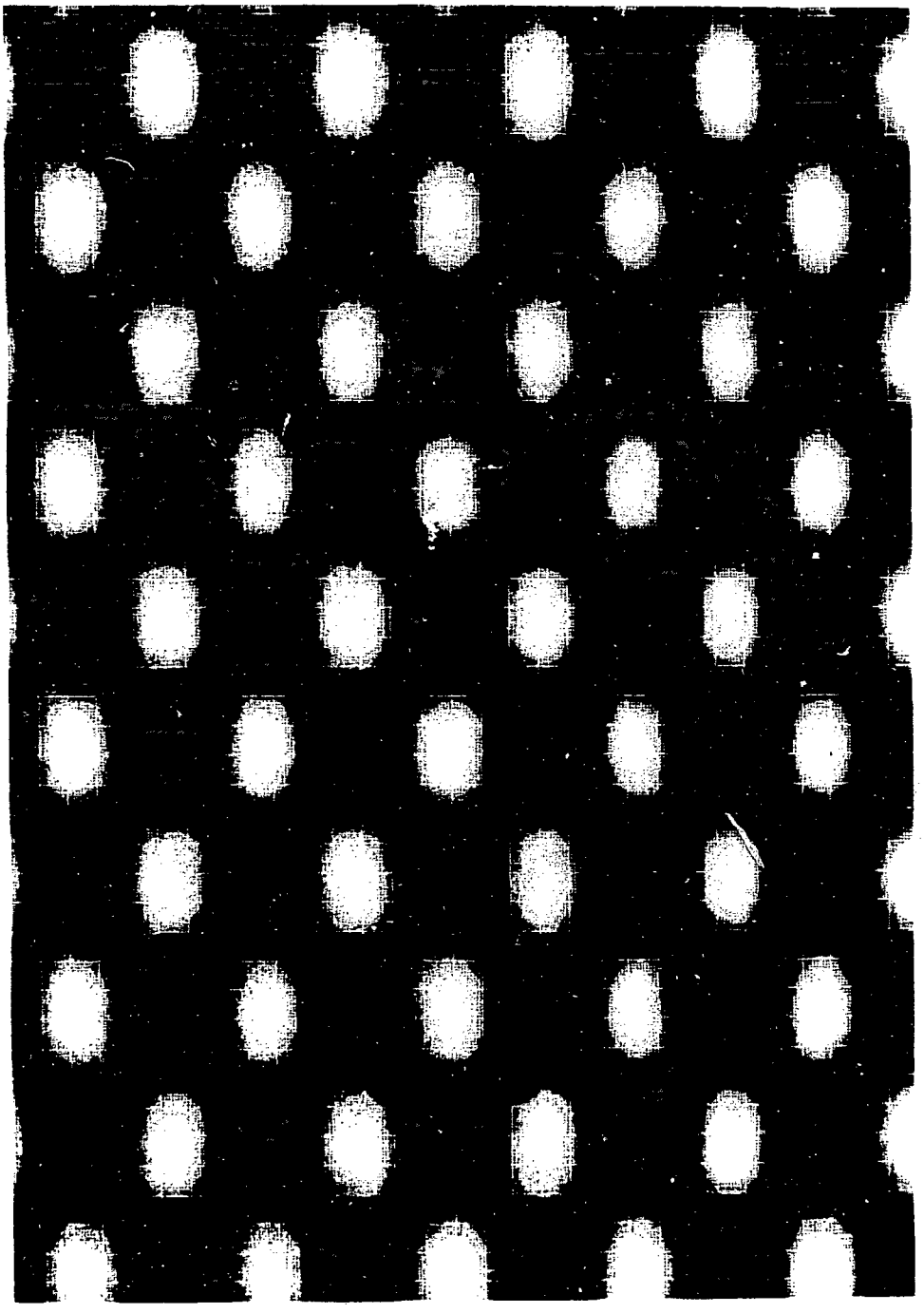

SILICON WITH SURFACE OXIDE
0.Å
7.7Å
15.3Å
23.0Å
30.6Å
38.4Å
XBB 840-8177

Fig. 5-6(a).



XBB 840-7433A

Fig. 5-6(b).

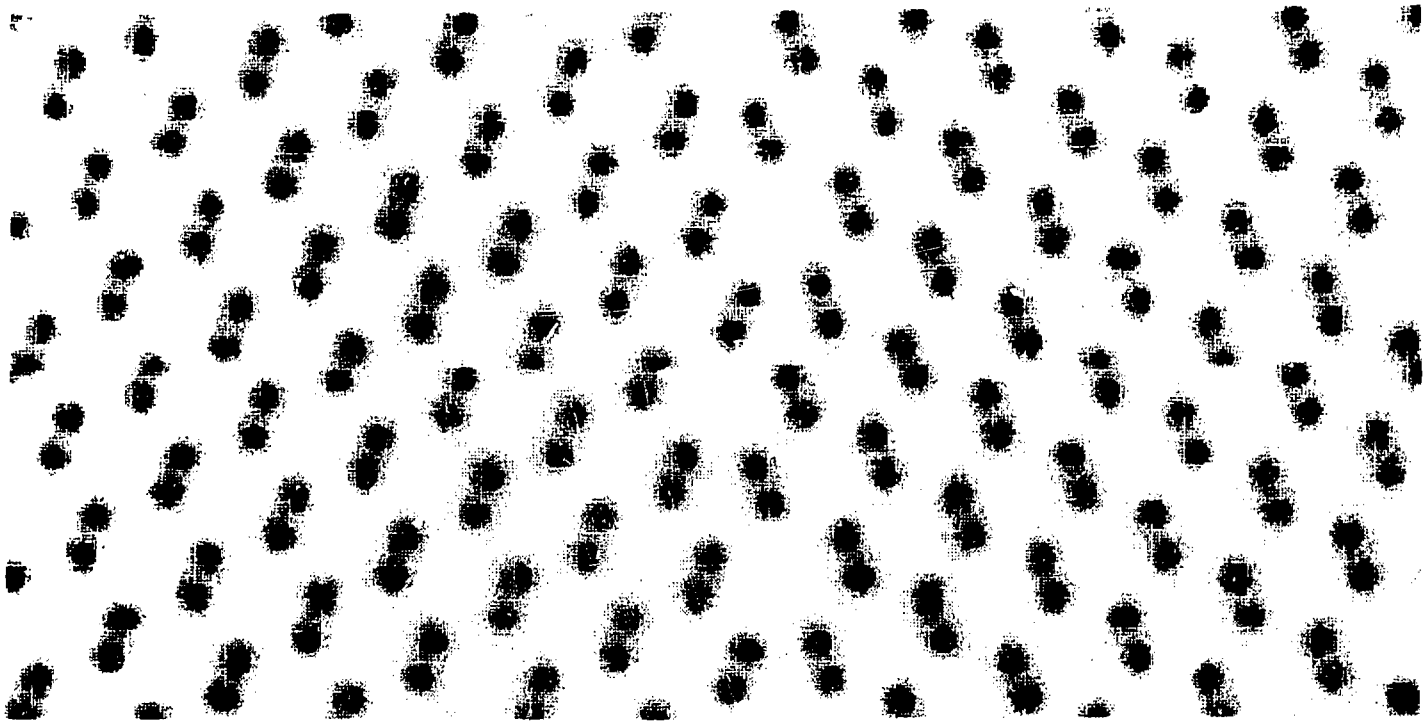


Fig. 5-7.

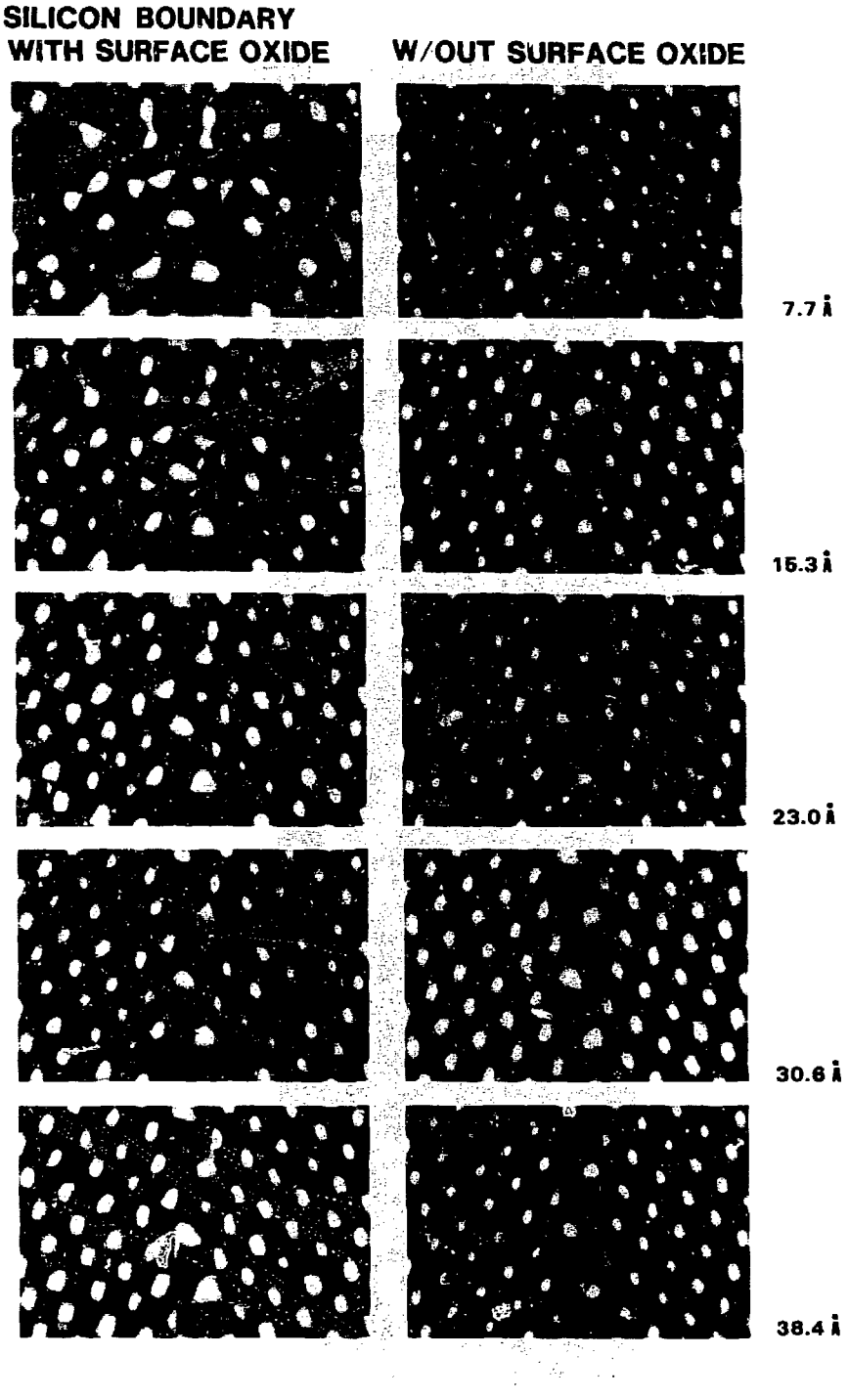

SILICON BOUNDARY
WITH SURFACE OXIDE
W/OUT SURFACE OXIDE
7.7 Å
15.3 Å
23.0 Å
30.6 Å
38.4 Å
XBB 840-7432A

Fig. 5-8.

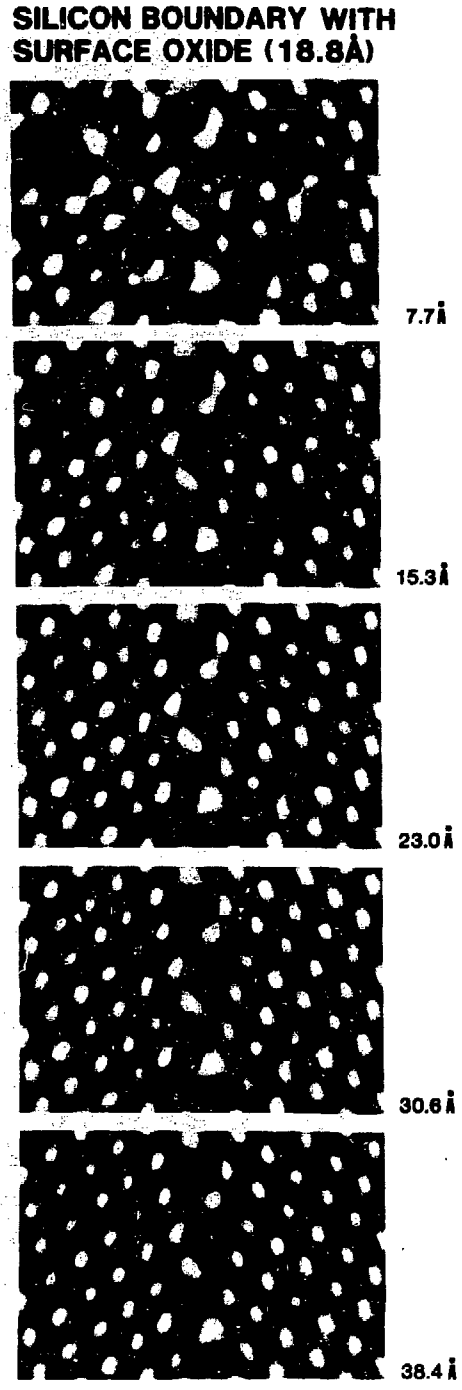

SILICON BOUNDARY WITH
SURFACE OXIDE (18.8Å)
7.7Å
15.3Å
23.0Å
30.6Å
38.4Å
XBB 840-7593A

Fig. 5-9.

Chapter 6

Summary

At this point it is of some interest to look back at the work presented in this thesis and to see where it stands with respect to other work in the area of high resolution transmission electron microscopy. I will take the time to review briefly the development of computer simulation of electron micrographs so the reader will gain an understanding of the current status of the field. I will also try to point out areas that I feel need further refinement and indicate what I believe to be the path of the future.

Because of the apparent ease with which electron microscopes are able to yield structural information to better than 3 Ångstrom resolution, high resolution transmission electron microscopy has gained an immense popularity during the last few years. However, it is very important to understand clearly both the strengths and weaknesses of HRTEM to know what it can and what it cannot do. It is not possible to take any material oriented in some arbitrary way, thin it down, insert it in the microscope and expect to obtain useful high resolution information. Very little three dimensional information can be obtained in HRTEM and the image is mostly a rendition of a two dimensional projected structure where all the atoms are projected onto a plane perpendicular to the incoming electron beam. Because of this, the specimen must be oriented such that the atoms superimpose in columns spaced far enough apart such that they can be resolved in the microscope. This eliminates all but a few low index orientations. Many close packed materials have planar separations that cannot be resolved in current microscopes. As an example only the [110] orientation of diamond structure materials could be resolved prior to the emergence of the Atomic Resolution Microscope at Lawrence Berkeley laboratory. Another difficulty is the preparation of the specimen itself. Very often it is not the perfect structure that is of interest, but rather some defect in the structure and usually it is the atomic arrangement at the core of the defect that one hopes to determine. However, both the defect

and the matrix must now be oriented such that they can be resolved. Finding just the right combination is often a very frustrating task. Another difficulty is obtaining the material itself suitably grown along a specific crystal direction. Sometimes projects that required maybe a few days to a week of microscope time needed years of specimen preparation time before the correct combination of specimen thickness and orientation was achieved. On top of all this comes the final interpretation of the image itself. Any given image might bear no resemblance to the actual structure of the specimen. Electrons will scatter and rescatter as they propagate through the specimen and will be acted upon by magnetic fields that eventually will focus them on a film. The image will change its character as one adjust the various knobs on the microscope and given a complicated structure it is no telling which setting results in an interpretable image. It is only for very specific settings of the objective lens defocus (depending on spherical astigmatism) and for specimen thickness less than 20 Å to 40 Å (depending on the scattering factor of the atoms in the structure) that one is fairly sure the image resembles the projected structure. This becomes more and more of a problem as the resolution increases and many images taken on today's microscopes can not be interpreted without computer matching. This entails computing images for various values of objective lens defocus, specimen thickness and sometimes the spatial and temporal coherence of the imaging electrons to match with the experimental images. Because of the many parameters that affect the image it is not sufficient to match just one computed image with one experimental image, but one must obtain a match for a set of images taken at different values of objective defocus, usually referred to as a through focus series. Thus computer simulation of electron micrographs has become a valuable tool in the field of high resolution electron microscopy. An example of how computer simulation can tell us how an image is affected by factors that are not usually considered, in this case surface layers of amorphous material, is shown in Chapter 5 of this thesis.

Even though the origin of present multislice calculations go back to 1957 when Cowley [Acta Cryst. 10, (1957), p.609] first formulated a theory based on physical optics, the first practical computations did not start to appear until the early seventies when a series of articles on

lattice imaging called the n-beam lattice image series were published by the Australian group consisting of Allpress, Hewat, Lynch, Moodie, O'Keefe and Sanders [Acta Cryst. A28, (1972), p. 528 and p. 536 ; Acta Cryst. A29, (1973), p. 138 and p. 389; Acta Cryst. A31, (1975), p. 300 and p. 307]. The computer programs that would grow to calculate 2-dimensional lattice images started off as a 1-dimensional multislice routine designed to compute 1-d electron diffraction patterns and did not deal with imaging. Imaging was introduced by O'Keefe who in his Ph.D. work incorporated the effect of objective lens defocus, spherical aberration and beam divergence to create the first practical algorithm for computation of 2-d lattice images based on the multislice formulation. The first 1-d lattice image program used reflections out to 15'th order and the first 2-d version used 31×31 beams in the calculation. At the time this was a huge calculation which took up most of the memory of the available computer. As the computing power increased over the next few years, the number of beams was also increased to get better accuracy, but at the expense of computing time. Because it is necessary to perform many calculations over different thicknesses and defocus values to obtain a good match, the computing time soon became prohibitive for many problems. It was not until Ishizuka and Uyeda in 1977 [Acta Cryst. A33, (1977), p. 740] suggested using fast Fourier transforms that a major improvement in computing time was made. Together with the improvement in computers this allowed the inclusion of many more beams and presently the larger calculations will include up to 256×256 beams.

Today, even though much work has gone into the search of faster and more accurate methods most programs still use the method of Ishizuka and Uyeda. The reasons are several and many can be found in the results of my work. The real space method which I discussed in Chapter 2 promised to reduce the calculation time and to allow the use of a larger slice thickness. However as I showed in Chapter 2, the method will diverge for too large slice thicknesses and in many cases will require more computing time for the same level of accuracy. The method is still useful for computing images from regions containing defects, but the conventional FFT method is better as an all around method. With special purpose hardware for

performing Fourier transforms becoming available, the time it takes to calculate a 2-d FFT might actually be less than the time required for the real space computation. The other method proposed during the last few years, the improved phasegrating method also does not present itself as a viable option because of its complexity (increasing computing time) and limitations imposed on the slice thickness as shown in Chapter 4. This is not to say that there has been no progress in the field of computer simulation of high resolution electron micrographs. Most of the progress has been in increasing the accuracy of the methods rather than making them faster. In terms of the multislice calculation, I showed in Chapter 3 how to properly use slice thicknesses smaller than the crystal periodicity along the incident beam direction for 3 different computational methods. I also showed how both the conventional first order and the second order multislice methods incorporate upper Laue layer interactions and showed that the accuracy is increased by using potential eccentricity in the second order method.

The multislice calculation only treats the interaction between the electrons and the specimen and a complete simulation of an image also includes the effect of defocus, aberrations, electron incoherence, microscope instabilities and the characteristics of the recording medium, usually film. The major part of the work in these areas has concentrated on the importance of temporal and spatial incoherence of the imaging electrons. Special "contrast transfer functions" that include the effect of incoherence, aberrations and defocus have been formulated under certain assumptions. Very often these contrast transfer functions are used even when the assumptions are no longer valid. This is mainly because of the increase in complexity and computing time that would go with a more accurate calculation, but also because the exact condition in the microscope is not known. It is clear that further refinement of the theory is needed. It is desirable to be able to vary the degree of coherence, something that is not presently possible. The contribution to the image by inelastically scattered electrons also needs to be better understood. Sometimes a match between the computed and experimental image is not obtained due to lack of proper treatment of coherence and inelastic scattering.

As the programs grew over the years the tendency was to use larger and faster computers such that the calculations could still be performed in a reasonable time. The ultimate goal in computer simulation of electron micrographs is to reach a situation where one can compute images in times of seconds, view them on a monitor and compare them to the experimental image while working on the microscope. A change in the model could be implemented by interactively "picking" up atoms, moving them around in the matrix and within seconds compute a new image. This could be done until one reaches a model that results in a fit between experimental and computed images. However, this goal can be achieved by going to smaller and more specialized hardware rather than timesharing on large mainframe computers. This trend is already starting. New systems will be built around a fairly small central processing unit whose job mainly will be to organize the flow of data. The actual computation of the image is suited for array processors that are uniquely designed to handle large vector operations such as FFTs. Because the multislice calculation is a recursive operation involving a few vector multiplications and a few FFTs, the entire calculation is easily performed on the array processor. The effect of the microscope is also easily included and the image can then be routed through an image processor. The image processor can be set up to simulate the recording device of the experimental image such that a straight comparison between experimental and computed image can be carried out. The image processor can also improve the experimental image by averaging over time to reduce noise and perform histogram equalization to enhance features. It can also be set up to quantify the mismatch between images. A laser filmwriter is an ideal recording device of computed images and can come with a scanning device for digitizing experimental images. In this way it is possible to build a specialized system that is ideally suited for high resolution electron microscopy. Such a system is planned for the National Center for Electron Microscopy here at Lawrence Berkeley Laboratory. It is in its final stages of planning at the time of writing and it is hoped that within a year it is a fully integrated and operative system.

Acknowledgements

This thesis would not exist if it were not for the help of many people to whom I owe gratitude. I wish to thank the following people that in one way or another contributed to the success of my stay here at Berkeley:

— B.J. Slagsvold without whom this thesis would never exist. Without his encouragement I never would have attempted to pursue a Ph.D. here at Berkeley.

— Ron Gronsky who has been my project supervisor for the last 4-1/2 years and whose guidance and helpful assistance I could not have done without. Thank you Ron for all the help you've given me.

— M.L. Cohen who I am sure will draw a sigh of relief now that my thesis is finally completed. I know it hasn't been easy being a liason for someone working outside the department. Thank you Marvin for not giving up on me.

— Alan Portis who cared enough to take the time to assist me in my preparations for my Qualifying Exam and who has been of tremendous help both academically and emotionally. The world would be a better place if there were more people like him.

— The entire group of friends in the trailers of 77F and 77G. In particular I would like to thank Eduardo, Jamie, Roseann and Mike for invaluable friendship and a lot of fun (you too Eileen). Thanks for all the good parties.

— M.A. O'Keefe who has been a good friend and a much needed source of knowledge and information. Thanks Mike for all the useful discussions we have had. By the way, where were you when I really needed you?

— Kathy and Madeline who have always been ready to assist me and who gave me a home down in building 72 when no one else wanted me. In particular I would like to thank Madeline for all the stimulating non-academic discussions.

— I also would like to thank Karla who put in a lot of time and effort in preparing this thesis.

— Finally I would like to thank Charlene who stood by me all this time and who made sure I never gave up along the way. Thank you Char for all the love you have given me. Maybe now I can make enough money to take YOU out to dinner for a change.

— This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Science Division of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Appendix A

The phase that appears in the exponent of the propagator is the difference in phase associated with electrons traveling a distance Δz having a wavevector $\mathbf{k}_0$ relative to those having a wavevector $\mathbf{k}_1$ as shown in Fig. 3-6. This phase difference is equal to

$$\Delta\phi = |\mathbf{k}_1| \frac{\Delta z}{\cos\alpha} - |\mathbf{k}_0| \Delta z = k_0 \Delta z\left(\frac{1}{\cos\alpha} - 1\right) \approx \tfrac{1}{2} k_0 \Delta z \alpha^2 \tag{A1}$$

For electrons reflected into the first order Laue zone the angle α_0 is equal to, see Fig. 3-7.

$$\alpha_0 = \frac{g_\perp}{1/\lambda} = g_\perp \lambda \tag{A2}$$

while for electrons scattered into the first order zone, the angle is

$$\alpha_1 = \frac{g_\perp}{1/\lambda - 1/c} \approx g_\perp \lambda\left(1 + \frac{\lambda}{c}\right) \tag{A3}$$

This gives a phase change of

$$\Delta\phi_0 = \tfrac{1}{2} k_0 \Delta z (g_\perp \lambda)^2 = \pi\lambda\Delta z g_\perp^2 = 2\pi\Delta z \xi(g_\perp) \tag{A4}$$

for electrons scattered into $(g_\perp,0)$, and a phase change of

$$\Delta\phi_1 = \tfrac{1}{2} k_0 \Delta z (g_\perp \lambda)^2 (1 + \frac{\lambda}{c})^2 = \pi\lambda\Delta z g_\perp^2 (1 + \frac{\lambda}{c})^2 \tag{A5}$$

for electrons scattered into $(g_\perp,1)$. The quantity $\xi(g_\perp) = 1/2\lambda g_\perp^2$ is the excitation error of the reflection $(g_\perp,0)$ and corresponds to the distance between the reciprocal lattice point $(g_\perp,0)$ and the Ewald sphere measured along the line connecting the center of the sphere with the reciprocal lattice point, see Fig. 3-1. The error in the phase of the propagator by using the excitation error of the zero order reflection instead of the the excitation of the first order reflection is thus

$$\Delta\phi_1 - \Delta\phi_0 = 2\pi x^2 \Delta z g_\perp^2 / c \tag{A6}$$

As can be seen from (A6) the error depends on slice thickness and electron wavelength and goes to zero as Δz and λ goes to zero.