To the Graduate Council:

I am submitting herewith a dissertation written by Xiao Zhang entitled "New Electron Beam Methods for Materials Microcharacterization." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Metallurgical Engineering.

David C Joy.
David C. Joy, Major Professor

We have read this dissertation
and recommend its acceptance:

E. E. Stansbury

C R Brooks

Roland Gronsky

Accepted for the Council:

Associate Vice Chancellor
and Dean of The Graduate School

New Electron Beam Methods for Materials Microcharacterization

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Xiao Zhang

August 1992

DEDICATION

This dissertation is dedicated to my wife, Yiqun Pan. I hope she knows how much she is loved and appreciated.

ACKNOWLEDGMENTS

I would like to express my heartfelt thanks to my major professor, Prof. David C. Joy, for his patience, guidance, support, and encouragement which provided many insights and ideas for this work.

In the course of dissertation research, a number of other scholars have assisted me generously. In particular I am indebted to Professor R.M. Bagby, Professor R.A. Buchanan, and Professor C.R. Brooks who served as my dissertation committee member. Great appreciation is also expressed to Dr. L.F. Allard, Dr. T.A. Nolan, Dr. J. Dunlap, for their advice and guidance during my research work.

The help from the electron microscopy facility led by Prof. Joy and his excellent colleagues, Dr. J. Dunlap, Dr. C. Joy, Mr. R. Williams, Dr. Y. Zhang, and Mr. R. Bunn from the biological consortium, was essential throughout the entire course of this work. With their consultation, criticism, and encouragement, I was able to achieve results otherwise impossible.

I am grateful to Mr. R. Bunn for his help in making the EBIC amplifier and the power supply unit of electron biprism. Also the helps from Dr. T.V. Blalock and Mr. F. Dyer for discussions and technical assistance in the C-V measurements are very appreciated.

I would like to thank Mr. T.W. Ogren from Eastman Kodak for supporting photographic films in our electron holographic research.

Special gratitude is expressed to Mr. T. Hashimoto from Hitachi Ltd., for his advice and expertise of HF2000 TEM. Thanks also go to

Mr. T. Taguchi and Mr. Y. Yoda from Nissei Sangyo America, Ltd.

I gratefully acknowledge partial support of this work by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by Martin Marietta Energy Systems Inc., for the Department of Energy under contract DE-AC05-840R21400. In addition, teaching assistantships provided by the Department of Materials and Engineering are gratefully acknowledged.

I also would like to thank Prof. Joy's secretary Ms. J.F. Trollinger and the MSE office staff: Ms. Sandra Maples, Ms. Kay Davis, Ms. Betty Frazier, Ms. Sancy Hail, Ms. Phillis Klindt, Ms. Pat Houser, and Ms. Sue Turner for their encouragement and help.

I would like to thank my fellow graduate students for their help and support.

Finally, and most of all, I would like to express my most sincere gratefulness to my wife, Yiqun Pan, for her unceasing support, patience and encouragement.

ABSTRACT

Developments of modern technologies force materials scientists looking deeply into material inner structures to find answers for more and more high demands and challenges of superior material properties and possible sources of creating new materials. When the working scale shrinks to the sub-micrometer even atomic level, one has to rely on the microcharacterization techniques, especially the electron-optical methods, to perform the tasks. Electron microscopy is the most common means of material microcharacterization. Increasingly electron microscopy is becoming a quantitative rather than a qualitative science. In this dissertation two new quantitative methods have been developed for materials chemical and electrical microcharacterization. A new Monte Carlo method has been developed for studying the spatial resolution of X-ray thin film microanalysis. New experimental techniques and quantitative analysis methods have been developed for charge collection electron microscopy.

With the development of field-emission gun transmission electron microscope, electron holography becomes a reality to materials scientists, which opens a new window for materials study. The major work in this dissertation is devoted to the development of electron holography techniques for materials analysis with new application on study of ferroelectric domain wall structure. The technique being developed allows the direct imaging of domain wall structure and quantitative measurement of local polarization.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION.	1
II.	A NEW MONTE CARLO METHOD FOR STUDY OF X-RAY SPATIAL RESOLUTION OF THIN FILM MATERIALS. . . .	8
	X-ray Spatial Resolution of EDS Microanalysis. . . .	8
	Monte Carlo Techniques.	16
	Monte Carlo Simulation for Study of X-ray Spatial Resolution.	18
	Applications of the Monte Carlo Modeling to Materials Microanalysis.	30
	References.	42
III.	ELECTRICAL CHARACTERIZATION OF MATERIALS -EBIC METHOD AND MONTE CARLO SIMULATION.	46
	Introduction.	46
	A Simple Specimen Holder for EBIC Imaging on the Hitachi S800 SEM.	50
	Fabrication of Au/p-Si Schottky Barrier for EBIC study.	62
	EBIC Study Using Monte Carlo Method.	83
	References.	95
IV.	ELECTRON HOLOGRAPHY THEORY.	99
	Background-Why electron holography?.	99
	What is Electron Holography?.	105

CHAPTER PAGE

The Principle of Wave-Front Reconstruction. 107

What Can Electron Holography Provide Beyond Other Techniques?. 124

References. 129

V. PRACTICAL ELECTRON HOLOGRAPHY. 133

Introduction. 133

Field Emission Electron Source. 134

Electron Biprism. 142

Generating Electron Holograms. 162

Recording Electron Holograms. 170

Reconstruction of Electron Holograms. 183

References. 203

VI. APPLICATIONS OF ELECTRON HOLOGRAPHY IN MATERIALS SCIENCE. 206

Introduction. 206

Electron Holography Techniques for Study of Ferroelectric Domain Walls. 211

Abstract. 211

Introduction. 213

Principle of Electron Holographic Method. . . 214

Quantitative Method. 215

Experimental Methods. 219

Results and Discussion. 221

References. 233

CHAPTER PAGE

VITA. 236

LIST OF FIGURES

FIGURE PAGE

2.1. Schematic illustration of a electron beam with a Gaussian distribution. 13

2.2. Single scattering model of beam broadening. 14

2.3. Coordinate system for Monte Carlo simulation. 21

2.4. Flow chart of Monte Carlo simulation. 31

2.5. Monte Carlo simulation of (a) PE and FSE electron trajectories and X-ray photon dots, and (b) X-ray spatial resolution in a thin carbon film (the plots were directly adopted from the computer screen). 33

2.6. Comparison of experimental and Monte Carlo simulation results of X-ray spatial resolution, plotted as a function of film thickness. 36

2.7. Effects of FSEs on X-ray spatial resolution as a function of atomic number. 38

2.8. Effects of FSEs on X-ray spatial resolution as a function of film thickness for a carbon specimen. 39

2.9. FSE yield as function of film thickness. 40

2.10. Plots of ratio of X-ray yield generated by primary electrons and by FSEs as a function of film thickness, and plots of relative effects of FSEs on X-ray spatial resolution as a function of film thickness. The right side y-axis is the relative difference of X-ray spatial resolution with and without considering of FSEs. 41

FIGURE PAGE

3.1. Schematic diagram of Schottky barrier charge collection. 47

3.2. Schematic diagram of EBIC measuring and imaging system. 52

3.3. Schematic diagram of the design of the EBIC specimen holder. 53

3.4. Superposed SE and reversed EBIC micrographs of dislocations on polycrystalline silicon solar cells (light-receiving side). Bright contrast = low EBIC signal. 55

3.5. (a) Superposed SE and reversed EBIC micrograph of a dislocation in a silicon solar cell. Bright contrast = low EBIC signal, (b) Y-modulated reversed EBIC map of the same area, (c) higher magnification micrograph of the same dislocation of (a), and (d) the EBIC line scan trace across the dislocation in (c). 57

3.6. Depth resolution in EBIC. (a) SE micrograph of a silicon solar cell, (b) reversed EBIC micrograph of a dislocation in the same area of (a) at 25 keV beam energy, (c) at 20 keV beam, and (d) at 15 keV. It is almost invisible. . . . 58

3.7. Superposed SE and reversed EBIC micrograph of dark contrast (i.e. high EBIC signal) of defects. 59

3.8. (a) SE micrograph of grain boundaries of the polycrystalline silicon solar cell, and (b) Y-modulated EBIC map of the same area of (a). 61

FIGURE PAGE

3.9. SEM secondary electron micrograph showing surface morphology of a 165-Å-thick Au layer on p-Si. The sputtering was performed at 15 mA for 4 min. 68

3.10. Forward current density-voltage characteristics for an Au(165Å)/p-Si Schottky barrier. 69

3.11. The forward current density-voltage characteristics (shown in Figure 3.10) after correction for recombination in the depletion layer. 71

3.12. A plot of $1/C^2$ vs V for a Au(350Å)/p-Si Schottky barrier 72

3.13. A plot of net impurity distribution vs depletion depth superposed with a plot of EBIC gain vs maximum electron penetration depth. 75

3.14. An EBIC micrograph of dislocations in a Au(165Å)/p-Si Schottky barrier with no external bias (at 6 keV beam energy). The central line is the trace of a electron beam scan. The bottom contour is the correspond contrast variation. 76

3.15. Measured variation of EBIC contrast with beam energy for the dislocations shown in Fig.3.14. The dislocation depth is estimated about 0.2 μm. 77

3.16. A plot of dislocation width vs $E^{1.75}$ (E is beam energy in keV). The dislocation width is defined as the full width at half maximum of the change in contrast. 79

FIGURE PAGE

3.17. (a) 15 keV EBIC image showing the presence of the "clouds" in Au(165Å)/p-Si Schottky barrier. (b) Y-modulated EBIC signal of the same area. 80

3.18. A 15 keV EBIC image with a higher beam current (5 nA) showing some precipitate-like dark contrasts inside each "clouds". 81

3.19. Monte Carlo simulation of EBIC signal generation (adopted directly from the computer screen). The up-left part of the screen is the plot of electron trajectories and the up-right part is the volume distribution of electron-hole pairs generation. The low-left part is the calculation of EBIC gain for various combinations of diffusion length and depletion depth. . 89

3.20. Experimentally measured and calculated EBIC gain for various beam energy and diffusion length. 92

3.21. Comparison of experimentally measured and Monte Carlo simulation of the EBIC gain matched with different values diffusion length for a range of beam energy. 93

3.22. A plot of net impurity distribution vs depletion depth superposed with a plot of diffusion length matched with experimentally measured EBIC gain for the same range of the depth. 94

4.1. Schematic illustration of a field-emission TEM lens system, electron biprism, and formation of an electron hologram 100

FIGURE PAGE

4.2. Gabor's in-line hologram scheme (projection method): (a) hologram formation, and (b) hologram reconstruction (adopted from [Tonomura, 1986]). 109
4.3. Schematic diagram of the coordination system used by Gabor (adopted from [Gabor, 1949]) 112
4.4. In-line scheme with transmission method: (a) hologram formation, and (b) hologram reconstruction (adopted from [Tonomura, 1986]). 116
4.5. Off-axis electron holographic scheme: (a) hologram formation, and (b) hologram reconstruction (adopted from [Tonomura, 1990]). 121
4.6. Hologram of CeO_2 with the interference fringe spacing of 0.32 Å, which makes a point resolution below 1 Å achievable (adopted from [Völkl & Lichte, 1990]). 125
5.1. Illumination system to demonstrate contribution to illumination angle. 136
5.2. Schematic configuration of a field-emission electron gun (adopted from [Tonomura, 1992]). 140
5.3. Schematic configuration of the illumination lens system of the Hitachi HF2000 FE TEM (adopted from Hitachi Instrument News, 19th ed., 1990). 141
5.4. Design sketch for bi-prism (adopted from [Tonomura, 1992]). 143

FIGURE PAGE

5.5. Rotatable electron biprism used in the study (adopted from [Tonomura, 1992]). 145
5.6. The shadow image of the biprism before the application of a potential. 146
5.7. SEM image of a lab-made fiber with a diameter of 0.3 μm. 149
5.8. Schematic illustration of selection criteria of the side band. 150
5.9. Schematic illustration of the parameters used in calculations of interference fringe spacing and interference width. 152
5.10. Fringe spacing versus biprism voltage at 200 keV beam energy. 155
5.11. (a) Interference width versus biprism voltage at different b values, and (b) interference width versus fringe spacing at different b values. 157
5.12. Minimum required biprism voltage and pixel number versus desired final resolution. 158
5.13. Effect of shadow of biprism on interference width versus biprism voltage at different b values. 159
5.14. Fringe contrast versus fringe spacing. 161

FIGURE PAGE

5.15. (a) Divergent beam Fraunhofer hologram of MgO cubes on a thin carbon film, and (b) In-focus TEM bright field image of the same MgO cubes. 164
5.16. Schematic illustration of illumination condition for off-axis electron holography. 168
5.17. Off-axis hologram of graphite taken at 200 keV beam energy. 169
5.18. Monte Carlo simulation of electron spreading in a photographic emulsion of 12 μm thickness (AgBr in Gelatin) at 200 keV beam energy. 172
5.19. Comparison of the absolute square of the side bands of a hologram of carbon film; left before and right after the optical density correction (adopted from [Völkl, 1991]). 177
5.20. Modulation transfer functions of two Kodak films versus spatial frequency. 179
5.21. Average noise levels of two Kodak films versus spatial frequency. 181
5.22. Schematic illustration of optical reconstruction of in-line divergent beam Fraunhofer hologram. 184
5.23. Optical reconstructed image of the MgO cubes from Figure 5.15(a). 185
5.24. Schematic illustration of optical reconstruction of off-axis hologram (adopted from [Tonomura, 1986]). . . 187

FIGURE PAGE

5.25. FFT spectrum (in log scale) of the graphite hologram from Figure 5.17. 191
5.26. Numerically reconstructed images of the hologram from Figure 5.17. (a) Amplitude image, and (b) phase image. 193
5.27. Off-axis hologram of a MgO cube. Only a corner of the cube is emerged in the interference area. 195
5.28. Numerically reconstructed images of the hologram from Figure 5.27 with a spatial frequency of 1.04 nm^{-1}. (a) Amplitude image, and (b) phase image. 196
5.29. Reconstructed image from the same hologram (Figure 5.27) but with a spatial frequency of 1.56 nm^{-1}. 197
5.30. Phase distribution from Figure 5.28(b) visualized as the wave front. 198
5.31. 8-times amplified phase image of the MgO cube. 199
5.32. Two kinds of interferograms: (a) reference wave is parallel to reconstructed image wave (azimuth angle between reference wave and reconstructed phase image wave is zero), and (b) tilted reference wave (adopted from [Tonomura, 1992]). 201
5.33. Holographic interferograms from the reconstructed image image (Figure 5.28(b)) with different azimuth angles. 202
6.1. Schematic illustration of ferroelectric hysteresis loop (adopted from [Jona & Shirane, 1962]). 208

FIGURE PAGE

6.2. (a) The cubic crystal structure of $BaTiO_3$ above the Curie temperature. The prototype crystal is calcium titanate (perovskite), and (b) below the Curie temperature the structure is tetragonal (adopted from [Kittel, 1986]). . . 212

6.3. Bright field TEM image of 90° domain walls in $BaTiO_3$ at 200 keV incident beam energy. 220

6.4. Electron hologram across a 90° domain wall using fringe spacing by 3.7 Å. Edges of the wall are indicated by pairs of arrows. Inset shows possible configuration of changes of polarization vectors across the wall. 222

6.5. Schematic illustration of the measurement of phase shift from the hologram fringe pattern. 223

6.6. Geometric relation between interference fringes and domain wall. X-Y is the coordination of the domain wall; X'-Y' is the coordination of the fringes. 225

6.7. (a) Electron hologram of a wadge-shape domain wall with a fringe spacing of 4.6 Å, and (b) interferogram of the same region. 226

6.8. Experimentally determined variation of normalized polarization component P_z across 90° wall (open circles) compared with theory of Zhirnov (dotted line). 228

FIGURE PAGE

6.9. (a) Electron hologram across a 90° domain wall using fringe spaced by 1.5 Å, (b) numerically reconstructed phase image (amplified by 4 times), and (c) interferogram from the same region. 230

6.10. Anomalous fringe bifurcations from some domain walls. Fringe spacings are 3.7 Å. (a) Y-shape fringe splitting, and (b) X-shape fringe splitting. 231

LIST OF TABLES

FIGURE PAGE

3.1. Barrier heights Φ_b of Au/p-Si Schottky barriers. 70

CHAPTER ONE

INTRODUCTION

In recent years the size scale on which we perform our technologies has shrunk. This new microtechnology can be characterized by extremely small functional elements, high structural complexity, great diversity of the materials involved, and sophisticated technological processes. Semiconductor manufacturing represents both the reason for, and the predominant example, of microtechnology. But other examples can be found in areas as diverse as biomaterials and water filtration. Each of these examples is characterized by a requirement to reliably manufacture structures with dimensions approaching those of the atoms regarding depth or thickness (junctions, monolayer coverage, multilayers, etc.) or approaching sub-micrometer level in lateral extension (linewidth, quantum well, etc.), and consequently the techniques of materials science microcharacterization are essential. In comparison with other industrial procedures, therefore, microtechnology makes particularly high demands on analysis as components, devices and systems become increasingly sophisticated. Many different methods of analysis are required to characterize materials, processes, process agents, and ultimately complete components or systems. The analysis is required not only in research and development but also for failure analysis and quality control in production lines.

The relationship between structure and property is central to materials science. The properties of materials are controlled by

composition and structure ranging from atomic to macroscopic dimensions, and the material scientist needs to have microcharacterization techniques available that provide this compositional and structural information both for the interior and the surface of materials. In essence the modern materials scientists or engineers require an understanding of the properties of engineering materials through their microstructure since from this they are then able to make intelligent predictions of how improved properties may be achieved via microstructural control. To a very large extent the development of modern engineering materials has thus been paralleled by necessity by the very rapid evolution of microcharacterization techniques.

Effective microcharacterization techniques must offer high lateral and vertical resolution, high sensitivity, low detection limits and high quantitative accuracy, so methods which use electron beam as measuring probes are particularly well suited for this purpose. High depth resolution is achieved by choosing signals emitted from the specimen surface, and high lateral resolution is obtained by focusing the electron beam. The strong interaction effects of electrons with solid material also make these methods very sensitive and capable of attaining extremely low detection limits.

Electron optical methods are used for characterizing extremely fine structural details and inhomogeneities up to atomic scale. They provide qualitative and quantitative data about the material properties such as the crystal structure, atomic configuration as well as the chemical composition and its spatial distribution.

Investigations of this type are used to obtain a knowledge and thus control of the material effects on the performance of miniaturized components. They are also an indispensable aid for determining the effects of microscopic and submicroscopic inhomogeneities on macroscopic material properties and thus provide the basis for synthesizing new materials.

A large number of current research topics in materials science and engineering area involve surfaces, interfaces, and thin films. Hence the information that needs to be obtained should be confined to small spatial domains, either the outermost surface zones of a material or device with sampling depths typically up a few micrometers maximum, or interfaces with typical dimensions of from a tenth to a few nanometers. Elementary composition, atomic arrangements and structural information are of central interest. The microcharacterization of surfaces, interfaces, and thin films is a most elaborate and often difficult task. The appropriate techniques for this task include scanning electron microscopy (SEM) and transmission electron microscopy (TEM) (including analytical electron microscopy (AEM)).

The SEM is used for imaging the surfaces of materials with the aid of signal derived from the interaction between the probe electrons and the specimen. The SEM has provided large versatile in providing various modes for imaging with secondary or backscattered electrons for topographic and elemental composition contrast, voltage and magnetic field contrast, and with electron-beam-induced-current (EBIC) for electrical characterization and

defect analysis of semiconductors. Crystallographic information can be obtained from channeling patterns, composition from X-ray microanalysis, spectroscopic data from cathodoluminescene, etc. The typical lateral resolution is 1 μm for analysis, while for topographical imaging with a field emission gun (FEG) values as good as 0.7 nm can be achieved. The interpretation of SEM images, X-ray spectra or EBIC signals is often not as straight forward as it appears to be. So a quantitative understanding of the electron beam - specimen interactions is required.

In contrast to the SEM, in which the electrons probe bulk specimens, the TEM makes use of high-energy electrons to irradiate thin specimens around 0.1 μm or less in thickness. High resolution image of inner structures can be obtained with a lateral resolution as high as the order of 0.1 nm, while electron diffraction provides crystallographic information about sub-micron areas of the specimen. The ultimate goal of electron microscopy is to reveal atomic structure of the object in order to meet the challenges from the modern development of material science and engineering at atomic dimensions. The AEM is a TEM equipped with a scanning attachment. The specimen can be scanned by a focused electron probe as in the SEM. The TEM is then operated as a scanning TEM (STEM). The three most important methods of analytical electron microscopy are X-ray microanalysis, electron energy loss spectroscopy (EELS), and electron microdiffraction. Information obtained includes elemental nanoanalysis (in which the spatial resolution can reach the same order of magnitude as the probe diameter), high resolution imaging

of inner structures, crystal structure and lattice parameters of nanodomains. The requirement of high spatial resolution in TEM microanalysis makes even higher demand for a detailed understanding of interaction of electron beam and specimen.

The resolution improvement of the TEM is mainly limited by lens spherical aberrations. Although significant improvements have been achieved in electromagnetic lens systems of microscope over the past few years, a lens will never be free of spherical aberrations. The further reduction of spherical aberration is a dead end as far as improving the atomic resolution is concerned. Recognizing this, this dissertation describes one way to circumvent the difficulties and obtain full information about the object wave by using electron holography. It will show that with the development of field emission (FE) TEM, high coherence electron beam can be used to conduct electron holography, and opens an entire new world in microcharacterization ranging from basic physics to advanced technologies.

The main objectives of electron holography are high resolution microscopy and interference microscopy. In high resolution electron holography both intensity and phase of an electron wave are recorded simultaneously as an electron hologram without any loss of information, whereas in the TEM image only the intensity of the electron wave is recorded. The amplitude and the phase of the electron wave can be reconstructed from the hologram. Therefore, images can be obtained under arbitrary focusing conditions, and point resolution can be improved by compensating for lens

aberrations in the reconstruction process. Thus, high resolution electron holography has the ability to overcome the point resolution limitation of lens aberrations and makes the ultra-high resolution as low as subangstrom range possible. Because of the ability to use phase information, the visualization of phase image with interference microscopy makes electron holography an effective tool to study atom steps, electrical and magnetic fields generated by the thin film specimen as well as polymeric specimen (pure phase object).

Increasingly electron microscopy is becoming a quantitative rather than a qualitative science. As mentioned above, it is necessary to understand the essential interactions between electrons and specimen in order to correctly interpret images or spectra, especially in high resolution microanalysis. Chapter two of the dissertation is mainly dealing with X-ray chemical analysis of inhomogeneous materials. A new Monte Carlo method is developed for thin film microanalysis. Chapter three emphasizes on electrical characterization of materials, and includes Monte Carlo modeling of EBIC analysis and experimental developments of EBIC technique. The newly emerged electron holography is still in its early developing stage and it already exhibits promising results in wide areas of materials science and engineering. Currently only two to three research groups in the world have the ability to conduct the research in electron holography and none of them is in the U.S. This dissertation contributes to the development of new electron optical techniques which not only provide new theoretical base to quantify data from inhomogeneous (nanometer) scale but also develop new

microcharacterization methods to meet the challenge from modern development of materials. Chapter four to six are devoted to the topic of electron holography techniques.

CHAPTER TWO

A NEW MONTE CARLO METHOD FOR STUDY OF X-RAY SPATIAL RESOLUTION OF THIN FILM MATERIALS

2.1 X-ray Spatial Resolution of EDS Microanalysis

As dimension sizes decrease from the micrometer to the sub-micrometer, and even now to the nanometer-scale for materials such as nanocomposites, the spatial resolution of EDS microanalysis is of greater concern to material scientists who use AEM to quantitatively determine local inhomogeneities of chemical composition of thin film specimen. The spatial resolution of X-ray microanalysis in the thin film may be defined as that the size of the beam-specimen interaction volume from which 90% of the X-ray emanate, which is determined by the scattering of the electron beam within the specimen. The electron range at the beam energies used in EDS microanalysis is often greater than the dimensions of the features being characterized, and the deposition volume is broader than the beam spot. The resultant degradation of spatial resolution should be evaluated with consideration of both beam spot size and beam broadening with the specimen. The spatial resolution of X-ray microanalysis is an important parameter and has the same relevance to AEM as image resolution has to high resolution electron microscopy.

2.1.1 Electron beam-material interactions

The interaction of an electron beam with a thin film specimen is complex. Before ultimately losing its energy or escaping from the specimen, each incident electron may undergo hundreds or thousands of separate scattering events, distributed between elastic and a variety of inelastic processes. For elastic scattering the beam electron is deflected by the positive charges of atomic nuclei, as screened or reduced in value by the orbital electrons. Elastic scattering is primarily responsible for changing the direction of flight of the electron. The average value of the elastic scattering angle is typically of the order of 5° up to 180°. Elastic scattering events do not alter the energy of the beam electron. Inelastic scattering decreases the energy of the beam electron, but the scattering angle is very small, typically 2° or less. Inelastic scattering events lead to the generation of secondary electron (excitation of loosely bound electrons of conduction band), Auger electrons and characteristic X-rays (ionization of tightly bound inner-shell electrons), bremsstrahlung or continuum X-rays (generation of X-ray photons due to deceleration of beam electron in the Coulombic field of an atom), electron-hole pairs (mobile charge carriers), lattice vibrations (excitation of phonons, i.e., heat), and electron oscillations (excitation of plasmons, waves in the "free electron gas" existing between the ionic cores in a solid).

2.1.2 Fast secondary electron effects

In thin film studied in AEM high energy incident electron beam is involved and very few electrons are sufficiently reduced in energy. In this case most electrons suffer only a small number of inelastic collisions such that the volume of sample from which X-rays are excited by the incident electron beam is determined primarily by the elastic scattering events. However, for X-ray analysis of light elements in thin films another mechanism must be concerned which can lead to degradation of spatial resolution, namely the formation of "fast secondary electrons" (FSEs). When a primary electron (PE) enters the thin film, it transfers part of its energy to atomic electrons, resulting in ionization of excitation. An electron ionized with a large energy transfer will become a secondary electron with kinetic energy ranging from near zero to half of the incident primary electron energy. The slow secondary electrons (< 50 eV) are not as important because they do not travel far from their position of generation. FSEs, however, are electrons with keV energy. The inelastic scattering which generates the FSEs causes only a small deflection of the primary electron beam, but the FSE leaves the impact point at an angle nearly perpendicular to the incident election. For a primary electron energy of 100 keV and a FSE energy of 2 keV the angle is approximately 80° [Joy, 1982]. Although the calculated yield of FSEs is relatively small, their energy is such that their cross section for some ionization processes may be much grater than that of the primary electron. The FSEs deposit their energy along the path that nearly at right angles to the incident electron and

X-ray photons excited by FSE is the main reason of degradation of the spatial resolution of X-rays, particularly for light elements. Any quantitative study of X-ray spatial resolution for thin film analysis has to consider the FSE effect.

2.1.3 Electron beam size, beam broadening, and the spatial resolution

The spatial resolution (R) of X-ray microanalysis in a thin film is a function of the specimen thickness and composition, and of the characteristics of the incident electron beam (size and current distribution). The beam-specimen interaction volume which is essential to the spatial resolution is a combination of the incident electron beam size (d) and the beam broadening (b) due to elastic scatter and FSE generation within the specimen.

In reality an electron beam has a finite size. The beam diameter d is a variable which is important primarily for the characterization of small area inhomogeneities. It is usually assumed that an ideal finely defocused incident electron beam has a Gaussian intensity distribution along its diameter [Wells, 1974] although experimental studies have shown that this is the case only when the electron beam is properly limited by the second condenser aperture [Cliff & Kenway, 1982; Klein et al, 1987]. Assuming the intensity distribution in the incident electron beam to be Gaussian, it is important to specify what fraction of the total beam current is included in a specified beam diameter since a Gaussian distribution extends to infinity. One may choose a single standard deviation (σ) diameter which would represent about 40% of the total intensity. One also can choose an electron beam diameter containing 50% of the

total current, which is termed the full width at half maximum (FWHM) [Cliff & Kenway, 1982]. An electron beam diameter containing 90% of the beam current also can be used as a defined beam size which is the full width at tenth maximum (FWTM). The FWHM is 2.83 times larger than the σ and the FWTM is about 1.81 times larger than the FWHM (see Figure 2.1). Since all three diameters are all consistently related for a Gaussian distribution, any one of them can be used as a definition of the incident beam size provided the diameter used is specified.

Beam broadening within the specimen is usually assumed having a conical geometry. To constrain the cone dimensions, beam broadening is arbitrarily defined as the volume containing 90% of the beam electrons. There are several analytical beam broadening equations proposed by various authors [Goldstein et al, 1977; Cliff & Lorimer, 1981; and Doig et al, 1981]. The simplest equation based on cross section of single scattering was proposed by Goldstein [Goldstein et al, 1977]:

$$b = 6.25 \times 10^2 (\rho/A)^{0.5} (Z/E) t^{1.5} \qquad (2.1)$$

where b is the beam broadening (cm) as defined in Figure 2.2 experienced by a point electron beam of energy E (keV), ρ is the density, A is the atomic weight, Z is the atomic number, and t is the thickness of the film. The Goldstein's equation is applicable to light element thin films. For thin film of intermediate and high atomic number, a plural scattering [Cliff & Lorimer, 1981] or multiple scattering [Doig et al, 1981] model should be used.

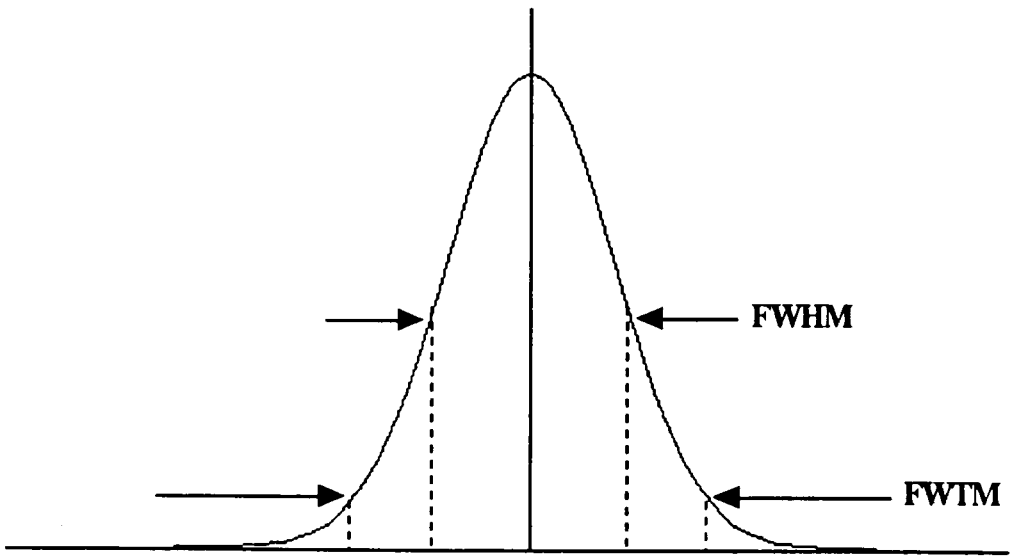


Figure 2.1. Schematic illustration of a electron beam with a Gaussian distribution.

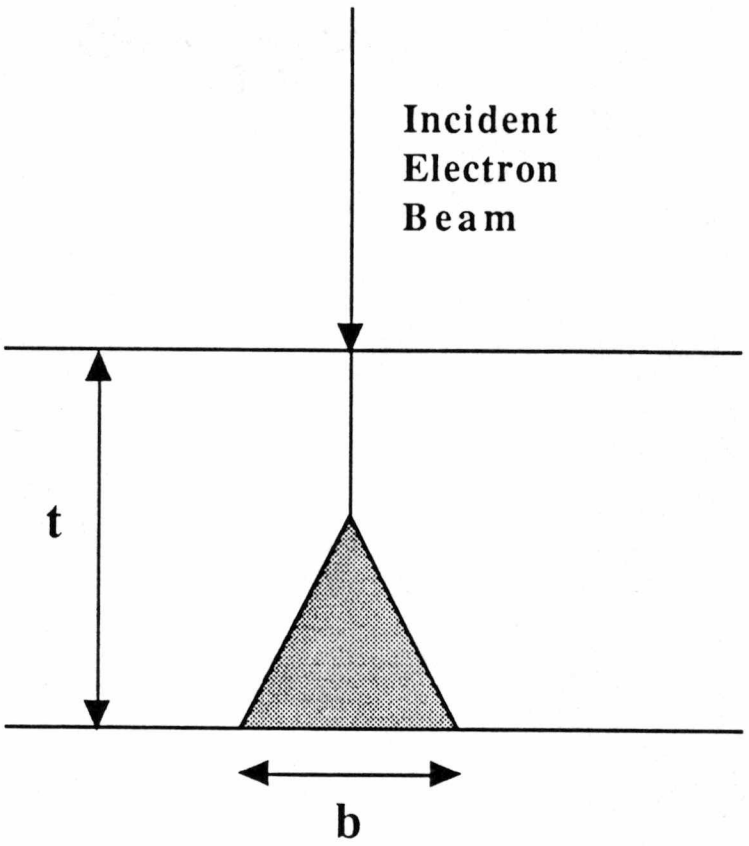


Figure 2.2. Single scattering model of beam broadening.

Goldstein [Goldstein, 1979] proposed a simple equation for the spatial resolution R as

$$R = b + d \qquad (2.2)$$

while Reed [Reed, 1982], arguing that the beam diameter was Gaussian, proposed that:

$$R = (b^2 + d^2)^{1/2} \qquad (2.3)$$

Both definitions suffered because they were all based on the dimensions of the beam emerging from the exit face of the thin film. Most recently Williams [Williams et al, 1991] proposed R as an average size of the interaction volume:

$$R = (d + R_{max})/2 \quad \Leftarrow \quad R_{max} = (d^2 + b^2)^{1/2} \qquad (2.4)$$

Like all previous definitions, this is arbitrary and ignores the effects of diffraction and FSE. None of above definition has been consistently applied in the literature, mainly because varied and inconsistent definitions of b and d have been used to determine R.

2.1.4 Previous research works on X-ray spatial resolution

Following the original work by Kyser and Geiss [Kyser & Geiss, 1977], there have been many studies of the X-ray spatial resolution that may be obtained in the AEM [Easterling, 1977; Rao & Lifshin, 1977; Hutchings et al, 1979; Stephenson et al, 1981; Joy, 1982; Romig, 1982; Michael & Williams, 1987; Michael, 1989; Goldstein et al, 1990; and Williams et al, 1991]. Most of these studies have been concerned with the effects of specimen thickness and beam voltage on the X-

ray spatial resolution and have concentrated their attention on comparing the results obtained in the AEM with the various theoretical calculations of beam broadening [Goldstein et al, 1977; Newbury & Myklebust, 1980; Cliff & Lorimer, 1981; Doig & Flewitt, 1981; and Reed, 1982]. Few of them have considered the incident electron beam size. Very few of them have considered the FSE effects. Meanwhile the definitions of R, d, and b are varied among different authors. Therefore, the results are often controversial and inconsistent.

Quantitative determination of X-ray spatial resolution of thin film requires detailed understanding of essential interactions of electron beam and specimen. In order to develop quantitative microanalytical techniques for thin film specimens, it is often necessary to formulate a model of the interaction of the electron beam with the specimen. Some of these models have been mentioned above. Such models can provide a quantitative description of one or more of the measurable products of the beam-specimen interactions: secondary electrons, backscattered electrons, characteristic and bremsstrahlung X-rays, absorbed current, and so on. One of these models is Monte Carlo electron trajectory simulation which is the most flexible of the modeling techniques and the most easily understood.

2.2 Monte Carlo Techniques

Monte Carlo simulations of electron beam-solid interactions are powerful tools for modern materials scientists and electron microscopists. The basic principle of Monte Carlo techniques is to

calculate the three-dimensional trajectory path of each incident primary electron (and any subsequent electron products) in a step wise manner as it decelerates within the solid target, accounting for both elastic and inelastic scattering [Joy, 1991]. The distance between scattering events, the scattering angles, and the rate of energy loss with distance are calculated from physical or empirical models. Random numbers (hence, the name "Monte Carlo") are used to select the scattering angles so that, for a large number of electron trajectories, the resulting probability histogram closely resembles the actual angular distribution for single scattering events. Because real electron beams are composed of many electrons, a large number of electron trajectory simulations are required to obtain physically meaningful results since no two electrons will have identical trajectories. The standard deviation of the statical result is given by $\sigma = \sqrt{N}$ where N is the number of trajectories simulated. Hence the relative standard deviation $\sigma_i = \sigma/N = 1/\sqrt{N}$. Thus results from any Monte Carlo simulation are subject to an error which is approximately inversely proportional to the square root of the number of events simulated [Joy, 1986]. To achieve an acceptable accuracy (i.e., as accurate as a typical experiment performed on an electron microscope) while at the same time remaining capable of generating statistically valid data in a reasonable time period on a personal computer, it is necessary to compute at least 5000-10000 trajectories.

The advantages using Monte Carlo calculation include at least the following five factors [Kyser, 1981]: (1) the capability to simulate

trajectories in any sample/geometry desired, (2) the large variety of signals which can be calculated such as backward and forward scattering electrons and secondary electrons with their angular/energy distributions, characteristic or Bremsstrahlung X-ray generation etc., (3) the "physical" insight and ease of data interpretation obtained, (4) the accessibility of entire data sets of the position and local energy of every incident electron or secondary electron, and (5) the ready availability of powerful computers and fast processors.

2.3 Monte Carlo Simulation for Study of X-ray Spatial Resolution

2.3.1 Definition of X-ray spatial resolution in Monte Carlo simulation

In this study we use the single scattering Monte Carlo model [Joy, 1988] to calculate the X-ray spatial resolution of thin film microanalysis with convolution of both beam size effects and FSE effects. In the following discussion, the X-ray spatial resolution (R) is defined as the radius within which 90% of the X-rays signal is generated. Since the FWHM of the electron beam intensity is normally quoted in the literature, we use the FWHM as the definition of the incident beam size in our Monte Carlo Calculation.

The spatial resolution of X-ray microanalysis calculated by Monte Carlo method is expected to be of greater precision since all of the incident electrons and resultant FSEs contribute to the generation of X-rays.

2.3.2 Basic assumptions

Before construction of our Monte Carlo model, it is instructive to give several important assumptions in discussing the interaction between the electron beam and solid:

(1) In any Monte Carlo electron trajectory simulation, the beam electron and resultant secondary electron are considered as discrete particles which undergo elastic and inelastic scattering with the atoms of the specimen, which is considered amorphous (i.e., non crystallography information will be considered in the simulation).

(2) Only elastic scattering events are considered in determining the path taken by any given incident electron as it moves through the thin film, since as discussed in section 2.1.1, elastic scattering events dominate the electron angular deflection. While in determining the spatial distribution of interaction we consider FSEs (inelastic scattering events) as well.

(3) The electron is assumed to lose energy continuously, at a rate determined by the modified Bethe relationship [Bethe, 1930; Joy & Luo, 1989], rather than as the result of discrete inelastic events. This simplification allows the net result of all possible inelastic events to be accounted for in a single expression.

The assumptions (2) and (3) are not essential to the construction of a Monte Carlo program but will greatly reduce the computation effects involved while introducing little systematic error into the results produced.

2.3.3 The single scattering Monte Carlo model

The Monte Carlo model used here is a modified version of a computer program developed by Joy and based on single scattering mechanisms [Joy, 1988] and a computer program developed by the same author considering the FSE effects [Joy, 1982].

elastic scattering

Within the assumptions made above the most accurate Monte Carlo simulation of the electron beam interaction is the one which attempts to account for each scattering event undergoes by the electron as it travels through the specimen [Newbury & Myklebust, 1981].

The electron is assumed to undergo an elastic scattering event at some point P_n represented by the coordinates (x,y,z) (Figure 2.3), after having traveled from its previous scattering event at P_{n-1}. The next point P_{n+1}(xn,yn,zn) to which the electron is scattered is needed to be calculated. The parameters which describe the instantaneous situation of the electron are its energy E and directional cosines cx, cy, and cz of the trajectory segment that brought the electron from its previous scattering location P_{n-1} to the present point P_n. These directional cosines are relative to a fixed set of axes attached to the specimen. With the convention the positive z-axis is set to be the same as the direction of the incident electron beam, i.e., normal and directed to the specimen surface, the x-axis lies on the specimen surface and is vertical to the beam direction, and the y-axis completes a right hand set of axes. To obtain the position of the new scattering point P_{n+1} one needs to first know the distance between

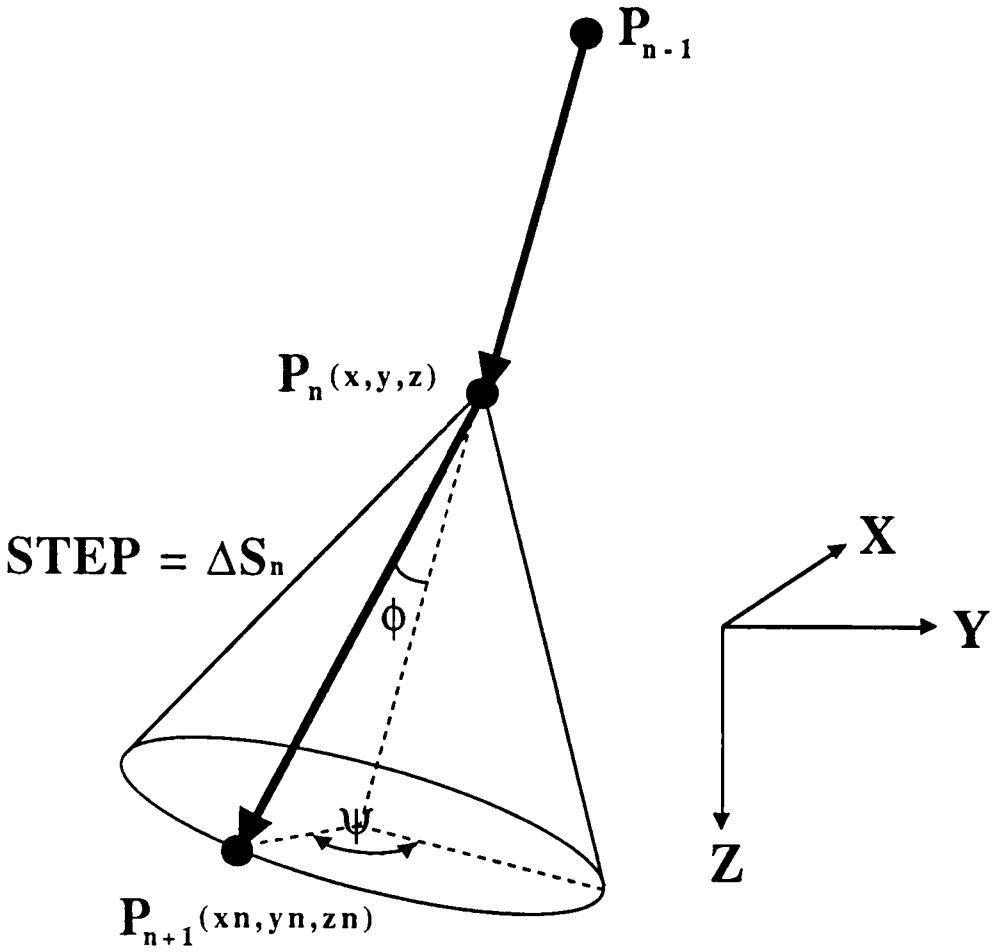


Figure 2.3. Coordinate system for Monte Carlo simulation.

P_{n+1} and the point P_n, and the elastic scattering angles.

The elastic cross-section is the probability that an elastic scattering event will occur and is given in terms of the effective size of the elastic scattering center with a unit of events/electron per (atom/cm^2). The relativistically corrected screened Rutherford elastic cross-section σ_E is given by [Joy, 1988]:

$$\sigma_E = 5.21\times10^{-21}\cdot\frac{Z^2}{E^2}\cdot\frac{4\pi}{\alpha(1+\alpha)}\cdot\left(\frac{E+511}{E+1024}\right)^2 \quad (cm^2) \qquad (2.5)$$

where E is the electron energy in keV, Z is the atomic number of the target atom, and α is a screen factor which accounts for the fact that the incident electron does not see all of the positive charges of the nucleus because of the orbital electrons and it has been estimated by the expression [Bishop, 1976]:

$$\alpha = 3.4 \times 10^{-3} \cdot \frac{Z^{0.67}}{E} \qquad (2.6)$$

The mean free path λ between elastic scattering event (the average distance the beam electron must travel through the specimen to undergo an average of one elastic scattering event) is defined by the above elastic cross-section:

$$\lambda = \frac{A}{N_a\rho\sigma_E} \quad (cm) \qquad (2.7)$$

where N_a is Avogadro's number, ρ is the density (g/cm^2), and A is the atomic weight (g/mol). Since the real electron travelling distance will vary in a random way about the values of λ, the variability is

introduced by calculating the distance (or step length) between the scattering events at P_n and P_{n+1} from:

$$\text{step } (\lambda_n) = -\lambda \cdot \log_e(\text{RND}) \quad \text{(cm)} \tag{2.8}$$

where RND is an equidistributed random number between 0 and 1 selected by the computer.

In the scattering event at P_n which marks the start of the step, the electron deflected through some angle ϕ relative to its previous direction (see Figure 2.3). The angle of scattering ϕ is determined from the angular differential form of the Rutherford cross-section and has the form [Newbury et al, 1976]:

$$\cos(\phi) = 1 - \frac{2\alpha \cdot \text{RND}}{(1 + \alpha - \text{RND})} \tag{2.9}$$

where α is the screening coefficient given in eqn. 2.6. This equation generates a unique angle in the range $0° < \phi < 180°$, producing an angular distribution which matches that obtained experimentally.

The electron can scatter freely to any point in the range of 0° to 360° on the base of the cone shown in Figure 2.3 so the azimuthal scattering angle ψ is then given as

$$\psi = 2\pi \cdot \text{RND} \tag{2.10}$$

where RND is an independent random number selected by the computer.

At this point all of the information needed to specify the scattering step of beam electron from P_n to P_{n+1} is now available.

inelastic scattering

As it travels through the specimen the incident electron loses its energy at a rate -(dE/dS) because of the inelastic scattering events. Although inelastic scattering is neglected as far as angular deviations and the path length are concerned, the continuous energy loss along a segment dS can be calculated using the modified version of the Bethe stopping power equation:

$$\frac{dE}{dS} = -78500 \cdot \frac{\rho Z}{AE} \cdot \log\left(\frac{1.66E}{J} + 1\right) \quad (keV/cm) \qquad (2.11)$$

where J is the mean ionization potential given by [Berger & Seltzer, 1964]

$$J = \left(9.76 \cdot Z + \frac{58.5}{Z^{0.19}}\right) \times 10^{-3} \quad (keV) \qquad (2.12)$$

The mean ionization energy J represents the average rate of energy transfer due to all possible inelastic events. Then, the energy loss ΔE along the step from P_n to P_{n+1} is approximated as

$$\Delta E = \text{step } (dE/dS) \qquad (2.13)$$

The production of secondary radiation along the electron path can be calculated from the appropriate cross-section, the specimen atom species parameters (concentration, atomic number, atomic weight, and density), and the path length. Calculation of resultant X-ray generation will discussed late in this section.

fast secondary electrons

In the events of generation of FSE, a knock-on collision model is used and the differential cross-section of the process is given by [Evens, 1955]:

$$\frac{d\sigma}{d\varepsilon} = \frac{\pi \cdot e^4}{E^2} \cdot \left\{ \left(\frac{1}{\varepsilon}\right)^2 + \left(\frac{1}{1-\varepsilon}\right)^2 \right\} \qquad (2.14)$$

where E is the incident beam energy, e is the electronic charge, and $\varepsilon = \Delta E/E$ is the normalized kinetic energy of the secondary electron relative to that of the primary electron. The incremental energy loss dE occurring along a segment dS of the primary electron trajectory due to knock-on collision is given from eqn. 2.14 as [Joy, 1983]

$$\frac{dE}{dS} = -\int_{\varepsilon_c}^{0.5} E \cdot \varepsilon \cdot \frac{d\sigma}{d\varepsilon}\, d\sigma \qquad (KeV/cm) \qquad (2.15)$$

where ε_c is the cut-off energy to avoid infinite cross-section at $\varepsilon = 0$. Here, $\varepsilon_c = 0.01$ has been used.

The inelastic scattering event causes a deflection α of the primary electron, which can be determined as

$$\sin^2(\alpha) = \frac{2\varepsilon}{(2 + k - k \cdot \varepsilon)} \qquad (2.16)$$

where k is the kinetic energy of the electron normalized by the electron rest mass, and k = E/511 (keV). The FSE leaves the impact point at an angle β to the original beam electron trajectory given by [Moller, 1931]:

$$\sin^2(\beta) = \frac{2(1-\varepsilon)}{(2+k\cdot\varepsilon)} \qquad (2.17)$$

incident electron beam size

A Gaussian form beam is usually assumed as follow for a point source [Murata & Kyser, 1987]:

$$g(r) = \frac{1}{\pi\,\delta^2}\cdot\exp\left(-\frac{r^2}{\delta^2}\right) \qquad (2.18)$$

where $\delta/\sqrt{2}$ is the standard deviation (σ). Here the beam diameter of 2δ is defined as the FWHM. In our program, the beam distribution is included at the initial stage of the Monte Carlo simulation, i.e., the beam distribution is formed using an independent computer-generated uniform random number (RND). The pole (r, ϕ) is determined for an incident point

$$r = \delta\sqrt{-\ln(RND)} \qquad (2.19)$$

here r serves to scale the electron impact points along a given diameter vector. The value for ϕ is distributed using another set of random number RND', $\phi = 2\pi$ RND'. For normal beam incidence on a flat specimen, the coordination of the beam impact point is given by:

$$x_0 = \cos(\phi)\cdot r \qquad y_0 = \sin(\phi)\cdot r \qquad z_0 = 0 \qquad (2.20)$$

All of the electrons are used in the calculations except that electron impacts at a distance from the center greater than 4σ are disregarded [Myklebust et al, 1976].

X-ray generation

When the incident electrons undergo elastic scattering events, simultaneously the various processes of inelastic scattering occur. In particular, the beam electron can cause inner shell ionization with the subsequent production of X-rays during deexcitation. For the excitation of characteristic X-ray, the Bethe cross-section can be used to calculate the inner shell ionization [Joy, 1988]:

$$\sigma_x = 6.51 \times 10^{-20} \left(\frac{1.8}{E \cdot E_c}\right) \log\left(\frac{0.65E}{E_c}\right) \quad \text{ionization/el/atom/cm}^2 \tag{2.21}$$

where E_c is the critical ionization energy (in keV) of the X-ray line of interest. The X-ray production, I_s (photons/electron) along the step segment is then

$$I_s = \sigma_x \cdot \frac{N_a}{A} \cdot \rho \cdot \text{step} \cdot \omega \cdot f \tag{2.22}$$

where ω is the fluorescent yield (photon/ionization) and f is the atomic fraction of the atom of interest in the compound. The X-ray yield from an element of interest can therefore now be calculated from eqns. 2.21-2.22 in the computation model given above.

2.3.4 Computer implementation

FSE and the incident electron beam diameter, as modelled above, are incorporated into the single scattering Monte Carlo simulation. The elastic mean free path λ_{el} is calculated by eqn. 2.7, and the inelastic mean free path λ_{in} can be obtained similarly using the knock-on collision inelastic cross-section σ_{in} which is found by

integrating eqn. (2.14) over the range $\varepsilon_c < \varepsilon < 0.5$. The total scattering mean free path λ_T is thus

$$(1/\lambda_T) = (1/\lambda_{el}) + (1/\lambda_{in}) \qquad (2.23)$$

The incident electron travels through the specimen by steps of various length given by relation eqn. 2.8 using λ_T as the mean free path. The electron is initially allowed to penetrate a distance equal to the "step," given by eqn. 2.8, into the thin film specimen without suffering either scattering or energy loss. The effect of a finite Gaussian source size is included using eqn. 2.19-2.20 around the nominal (0,0,0) coordinates.

At the end of the step, unless the value of "step" generated by eqn. 2.8 is greater than the total thickness of the specimen, in which case a new trajectory is started, the nature of the subsequent scattering events is determined by considering the relative probabilities P_{el} of an elastic, or P_{in} of an inelastic event, where

$$P_{el} = \lambda_T/\lambda_{el} \quad \text{and} \quad P_{in} = (1 - P_{el}) \qquad (2.24)$$

It follows naturally that $(P_{el} + P_{in} = 1)$. If the next random number generated by the computer is less than P_{el} then the scattering is taken to be elastic, otherwise it is inelastic.

The rest of the steps is then calculated depending on the selection of the scattering events from which an appropriate mean free path, scattering angle, and energy loss rate are selected from the previous model.

If the scattering is determined to be elastic then a new step length is obtained from eqn. 2.8 using λ_T. The scattering angle α is given by eqn. 2.9. The electron can scatter with equal probability in any azimuthal direction determined by eqn. 2.10. The energy loss along the segment is calculated by eqns. 2.11-2.13.

If, on the other hand, the scattering is determined to be inelastic then the coordinates of the primary electron are stored, and the energy of the FSE produced by the knock-on collision is determined by solving for ε the equation

$$\mathrm{RND} = \frac{\int_{\varepsilon_c}^{\varepsilon} (d\sigma/d\varepsilon)\, d\varepsilon}{\int_{\varepsilon_c}^{0.5} (d\sigma/d\varepsilon)\, d\varepsilon} \qquad (2.25)$$

where RND is another independent random number. The FSE leaves the impact point at the angle β given by eqn. 2.17 and with an angle found by an application of eqn. 2.10. The mean free path for the FSE with energy ε is found from integration of eqns. 2.14 and 2.7, and the FSE is then tracked through the specimen assuming elastic scattering parameter ϕ and ψ found from eqns. 2.9 and 2.10, and energy loss rate found from eqn. 2.11. The production of tertiary and higher-order knock-on electrons by the FSE is not considered here although such modification would be possible if sufficient computing time are available. This procedure is continued until the FSE leaves the thin film, or drops its energy below 50 eV. Tracking of the primary electron is then resumed from the previous stored coordinates until

it in turn passes out of the specimen, testing at each subsequent scattering step for an elastic or an inelastic event as described above.

The calculation of the X-ray generation is incorporated in the above computation loop. If after first testing at each step that the electron energy E (for either an incident primary electron or a FSE) is greater than E_c, then a photon is produced and the corespondent position of the excitation is then individually stored in the form of dot density (all photons generated at the same location are accumulated), and is displayed simultaneously on the screen as a single dot. During the inelastic events, the photons generated by primary and those generated by FSE are separately stored and are displayed in different color on the screen. The X-ray yield generated by primary electrons and those generated by FSEs can be calculated, respectively, using eqns. 2.21-2.22. The integration of the X-ray intensity distribution is then carried out through out the interaction volume. Finally, the program plots the cumulative X-ray yield as a function of position relative to the beam impact point and by measuring the lateral distance over which this rises from 5% to 95% of its normalized maximum value a reliable measure of the X-ray spatial resolution is obtained.

The details of the entire calculation are shown in the flow chart (Figure 2.4).

2.4 Applications of the Monte Carlo Modeling to Materials Microanalysis

The complete program as outlined here is written in Microsoft® QuickBasic and is implemented on a Macintosh II class computer

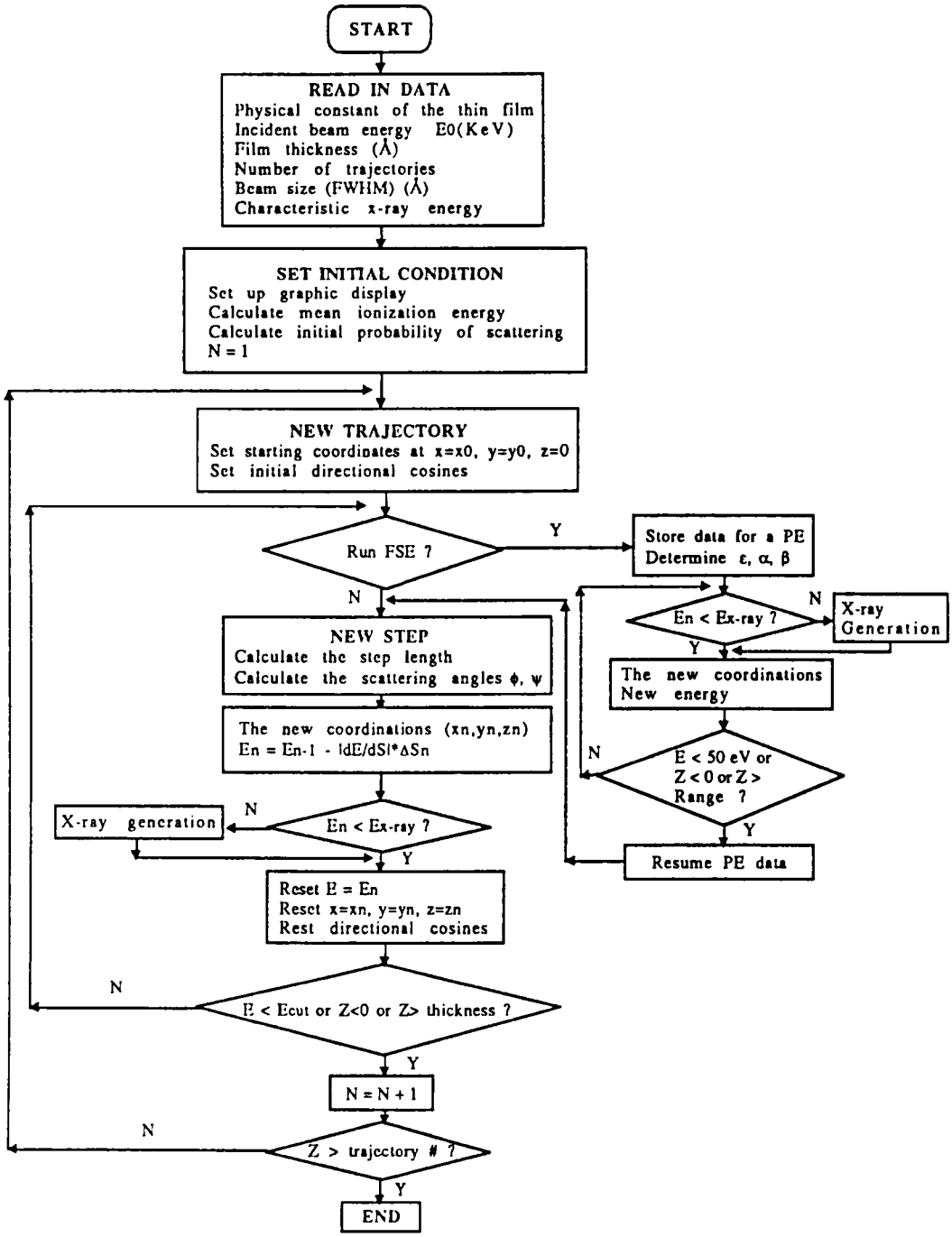
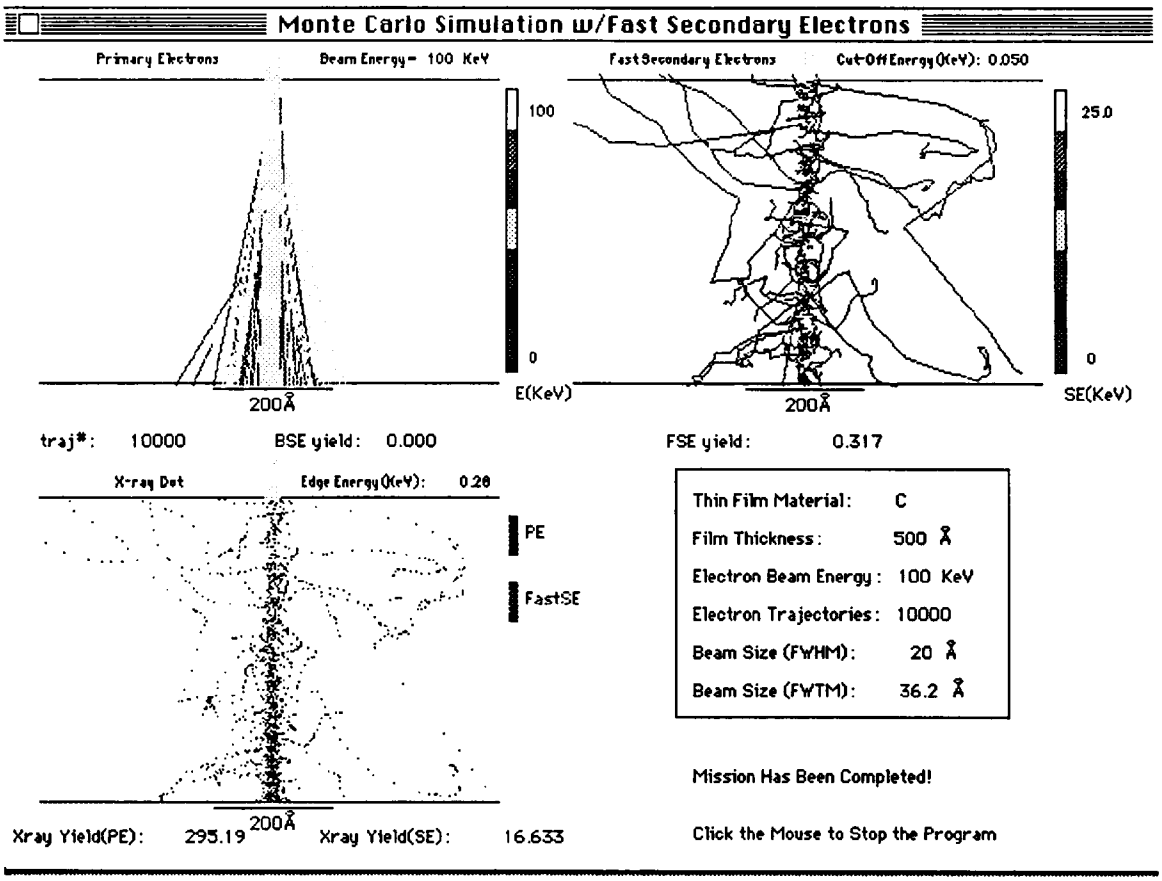


Figure 2.4. Flow chart of Monte Carlo simulation.

which has excellent Macintosh toolbox functions that allow high resolution graphic displays to plot in real time the primary and FSE trajectories and X-ray photon "dots" on the screen. After the simulation, the second page of the display shows the plot of X-ray spatial resolution calculated by the integration of the X-ray profile with and without consideration of FSE. All the specimen and the electron beam parameters are displayed on the screen. The program runs very fast with those on screen real-time trajectory display (three sets of trajectories are displayed simultaneously) producing about 600 primary trajectories per minute with a Mac IIci equipped with a math co-processor. Figure 2.5 shows an example of the trajectories and the X-ray spatial resolution computed by the above program for electrons travelling through 500 Å thick carbon film at 100 keV. The color change of each trajectory is due to the loss of energy during inelastic scattering events. The X-ray dot map shows two different colors, one being generated by primary electrons (PE) and the other being generated by FSEs. The second page (Figure 2.5b) shows the calculated X-ray spatial resolution.

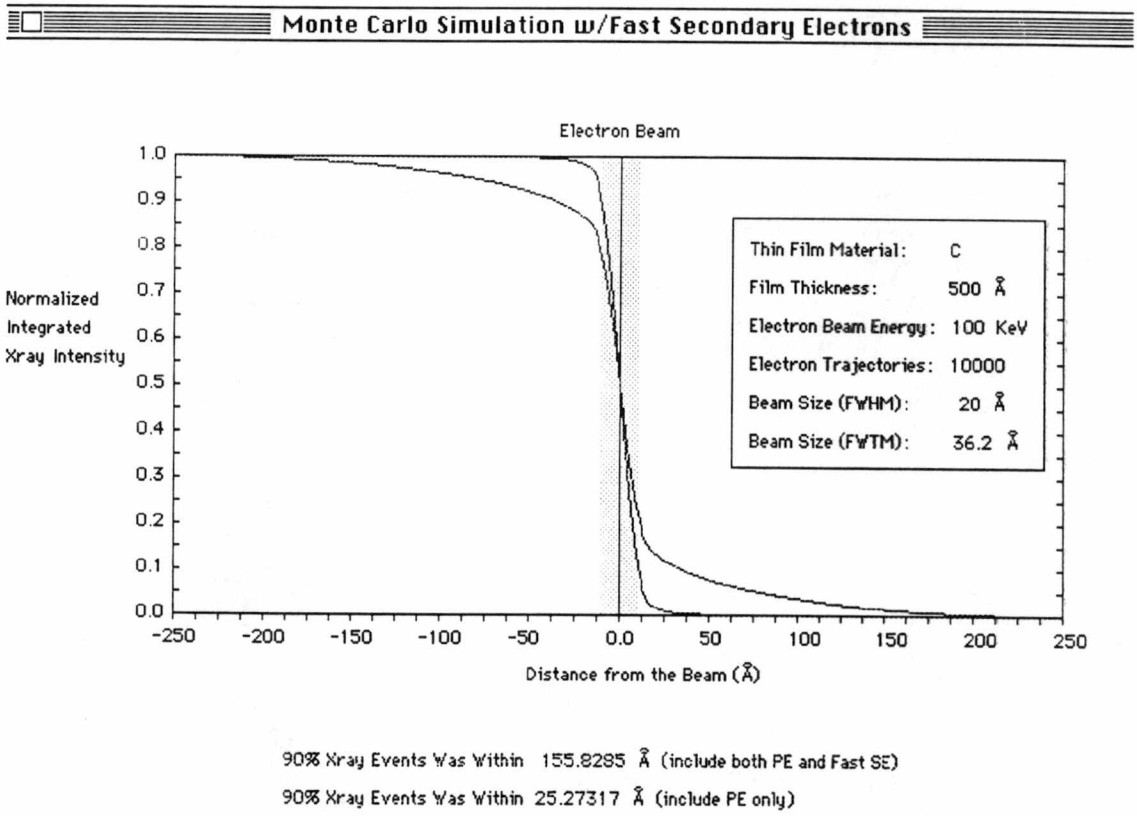
2.4.1 Applications (experimental verification)

The program is first tested against the current available experimental results to insure the accuracy of our approach to the problem. The only experimental data available now are from the experiment conducted in a VG HB-501 STEM with a cold field emission source operated at 100 keV [Michael, 1989]. The microscope was equipped with a Link Analytical AN10/95 X-ray system and a Link LZ5 series E windowless X-ray detector. The incident beam size



a

Figure 2.5. Monte Carlo simulation of (a) PE and FSE electron trajectories and x-ray photon dots, and (b) x-ray spatial resolution in a thin carbon film (the plots were directly adopted from the computer screen).


Monte Carlo Simulation w/Fast Secondary Electrons
Electron Beam
Normalized
Integrated
Xray Intensity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-250
-200
-150
-100
-50
0.0
50
100
150
200
250
Distance from the Beam (Å)
Thin Film Material: C
Film Thickness: 500 Å
Electron Beam Energy: 100 KeV
Electron Trajectories: 10000
Beam Size (FWHM): 20 Å
Beam Size (FWTM): 36.2 Å
90% Xray Events Was Within 155.8285 Å (include both PE and Fast SE)
90% Xray Events Was Within 25.27317 Å (include PE only)

b

was 1.8 nm (FWTM) with 0.8-1.0 nA probe current. The chromium composition profiles from which the spatial resolution could be determined were obtained from 40wt%Ni-30wt%Cr-Fe specimens that had been held at 1000°C for 100 hours. The matrix phase (γ) contained 25-29wt%Cr and the precipitate phase (α) contained 80-90wt%Cr. The concentration profiles were obtained by stepping the electron probe across interface between γ and α phases. The specimen draft during microanalysis is compensated by the on-line computer program. The specimen thicknesses were determined by convergent beam electron diffraction (CBED) and through the EELS log ratio technique with a maximum error of 10%. The experimental results of X-ray spatial resolution (FWHM) were determined from the composition profiles obtained from a step change in composition by measurement of the distance represented by 76% of the total composition change, and the spatial resolution FWTM (90% of the incident electrons used to define the spatial resolution) was obtained by multiplying the FWHM by 1.81.

The Monte Carlo simulations for Cr K-line X-ray excitation are carried out using 2500 electron trajectories with an estimated error of 2%. The calculated results and the corespondent experimental results are shown in Figure 2.6. Comparison of the experimental and the calculated spatial resolutions shows reasonably good agreements with as low as 4% discrepancy. The differences are larger for the thinner films due to the lower X-ray count rates [Michael, 1989]. The discrepancies may be induced by the experimental error, non-Gaussian current distribution, and possible oversimplification of the

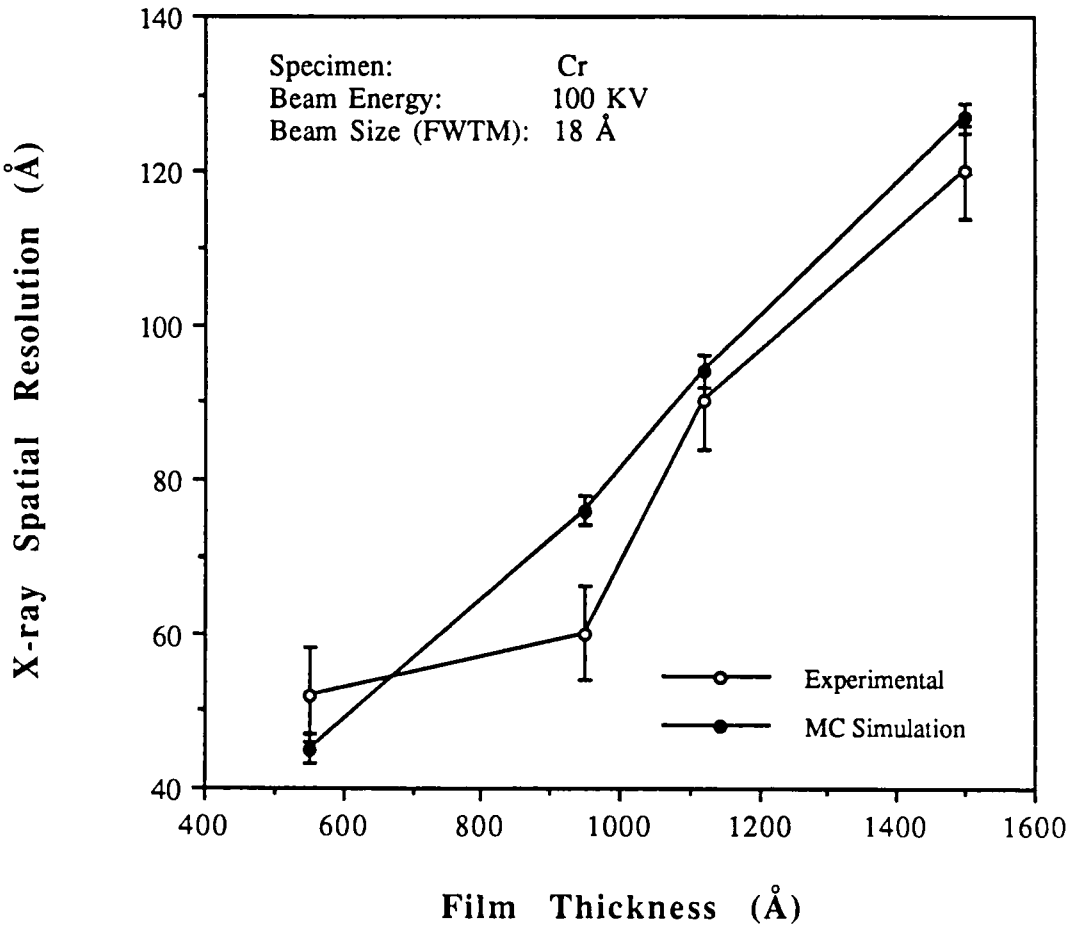


Figure 2.6. Comparison of experimental and Monte Carlo simulation results of x-ray spatial resolution, plotted as a function of film thickness.

modelling. But when considering the uncertainty of the measured spatial resolution (± 0.5 nm) and ± 2% statistic error of the calculation, the differences are all close to or within the error range. Therefore, it is reasonable to say that the developed modelling is successful as far as the X-ray spatial resolution is concerned.

2.4.2 Fast secondary electron effects

As we discussed above the degradation of X-ray spatial resolution due to FSE production is significant for light elements. Figure 2.7 shows the calculated results of thin film materials from atomic number 5 (boron) to atomic number 30 (zinc). The incident beam size is assumed to be 20 Å (FWHM) and the beam energy is 100 KV. In Figure 2.7, Delta R = R_{PE+FSE} - R_{PE}, i.e., the difference of the X-ray spatial resolutions with and without consideration of FSE effect. It is evident from the results shown in Figure 2.7 that one has to consider the FSE degradation on X-ray spatial resolution when the subject of analysis is of atomic number 11 (sodium) or lower.

The resolution of X-ray analysis decreases (volume of 90% X-ray generated increases) with increase of film thickness (shown in Figure 2.8). The best spatial resolution is obtained for the smallest film thickness. It must also be kept in mind that attempts to improve spatial resolution reduce the X-ray intensity and hence decrease the analytical sensitivity [Romig et al, 1990]. Figure 2.9 shows the calculated results of FSE yield for carbon films with thickness changes from 100 Å to 3000 Å. The FSE yield increases almost linearly with the increase of specimen thickness due to the increasing of inelastic scattering events. As a result, the degree of FSE

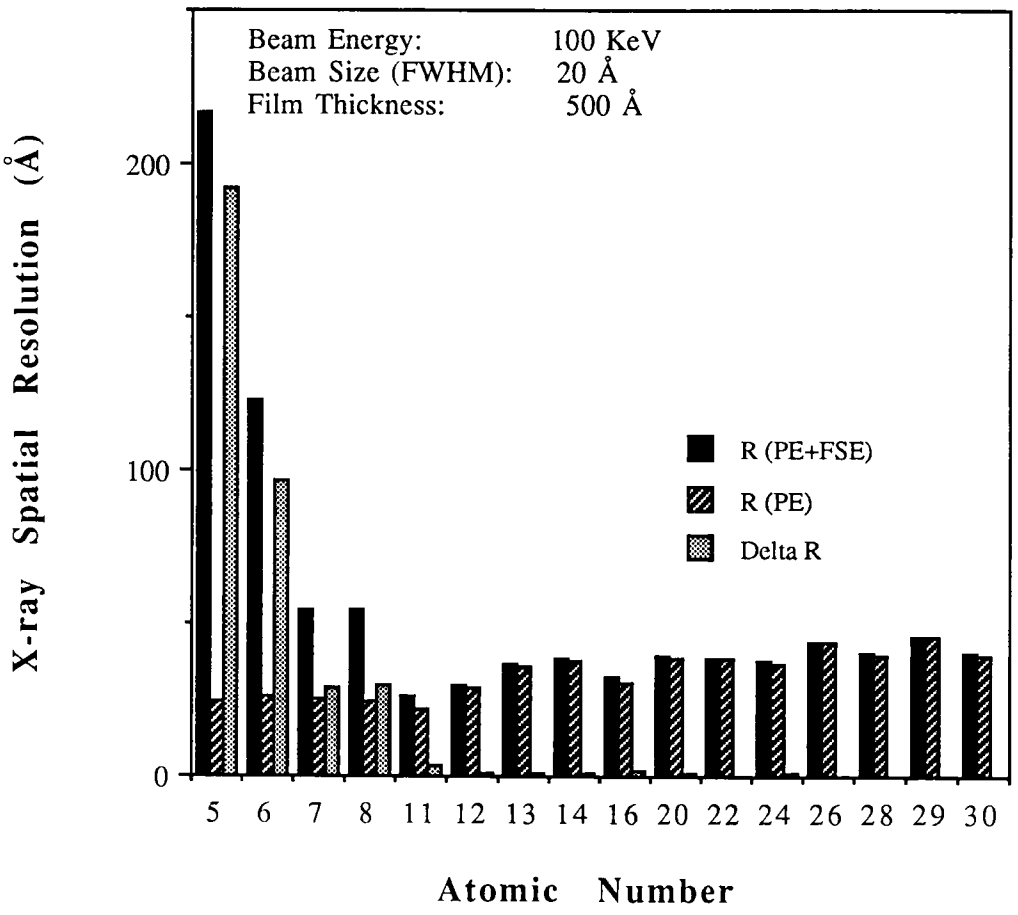


Figure 2.7. Effects of FSEs on X-ray spatial resolution as a function of atomic number.

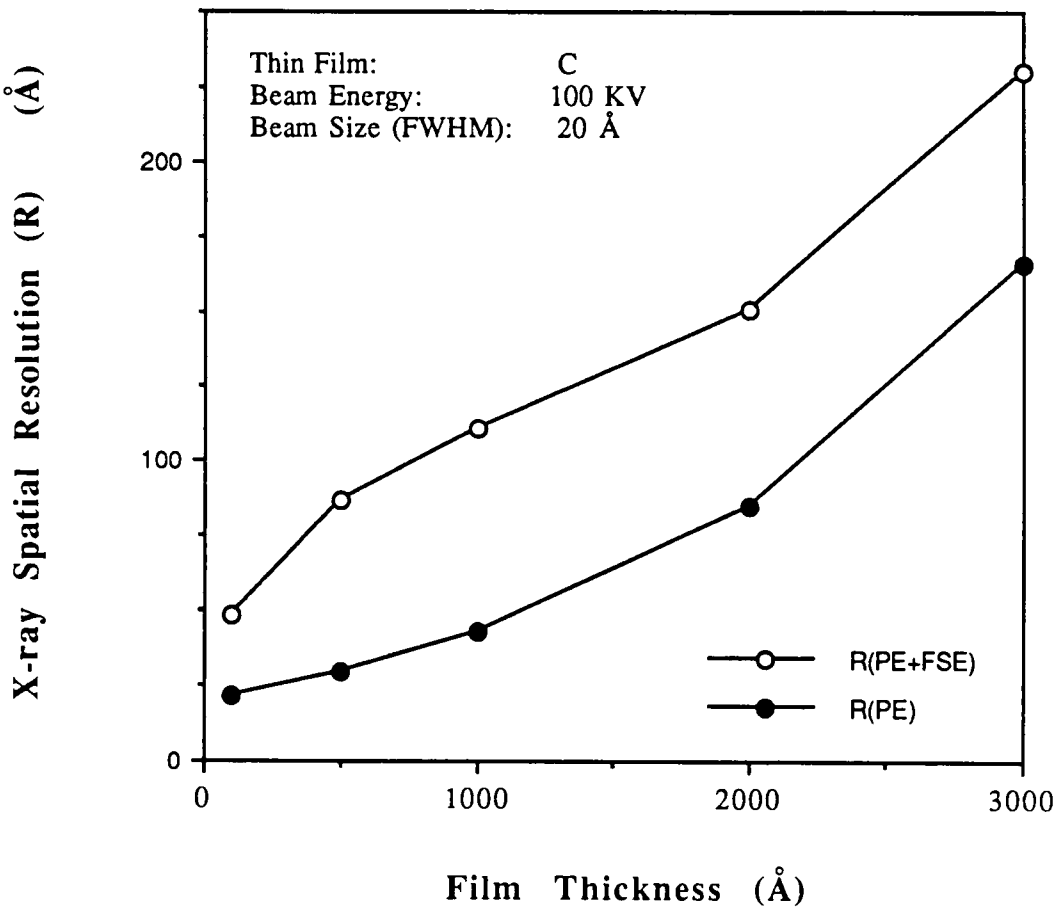


Figure 2.8. Effects of FSEs on x-ray spatial resolution as a function of film thickness for a carbon specimen.

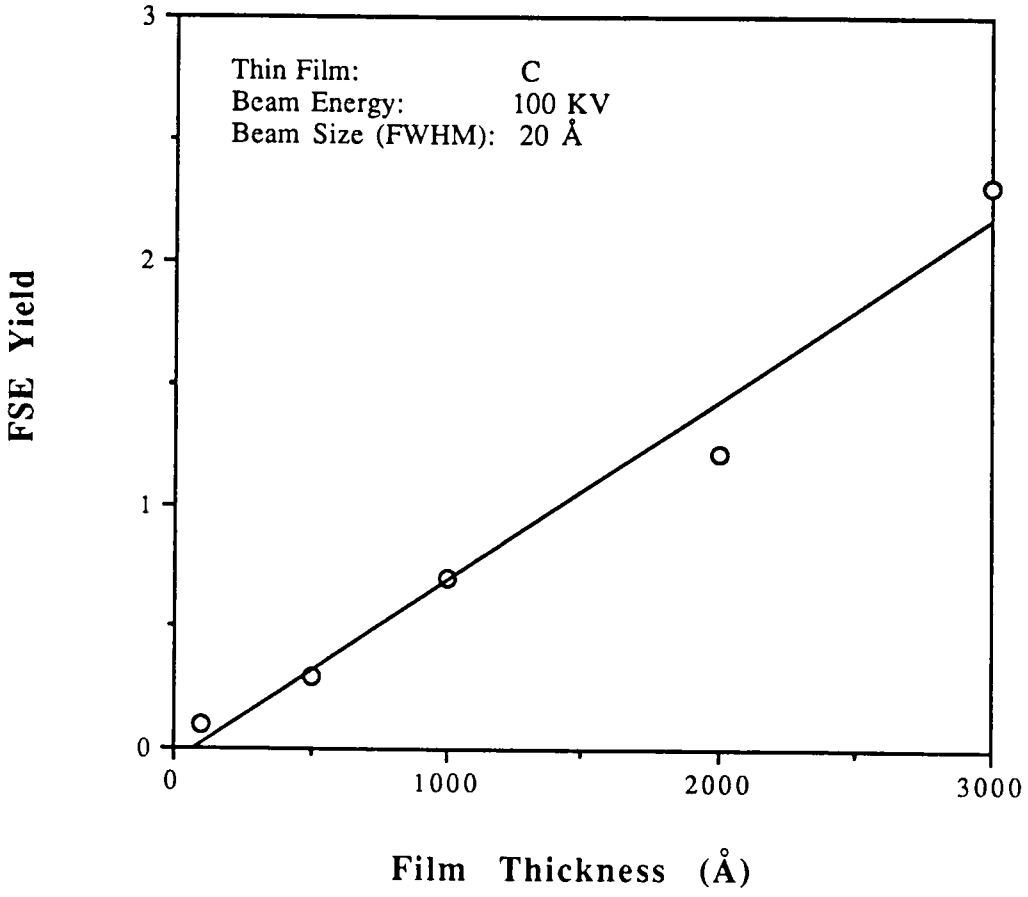


Figure 2.9. FSE yield as function of film thickness.

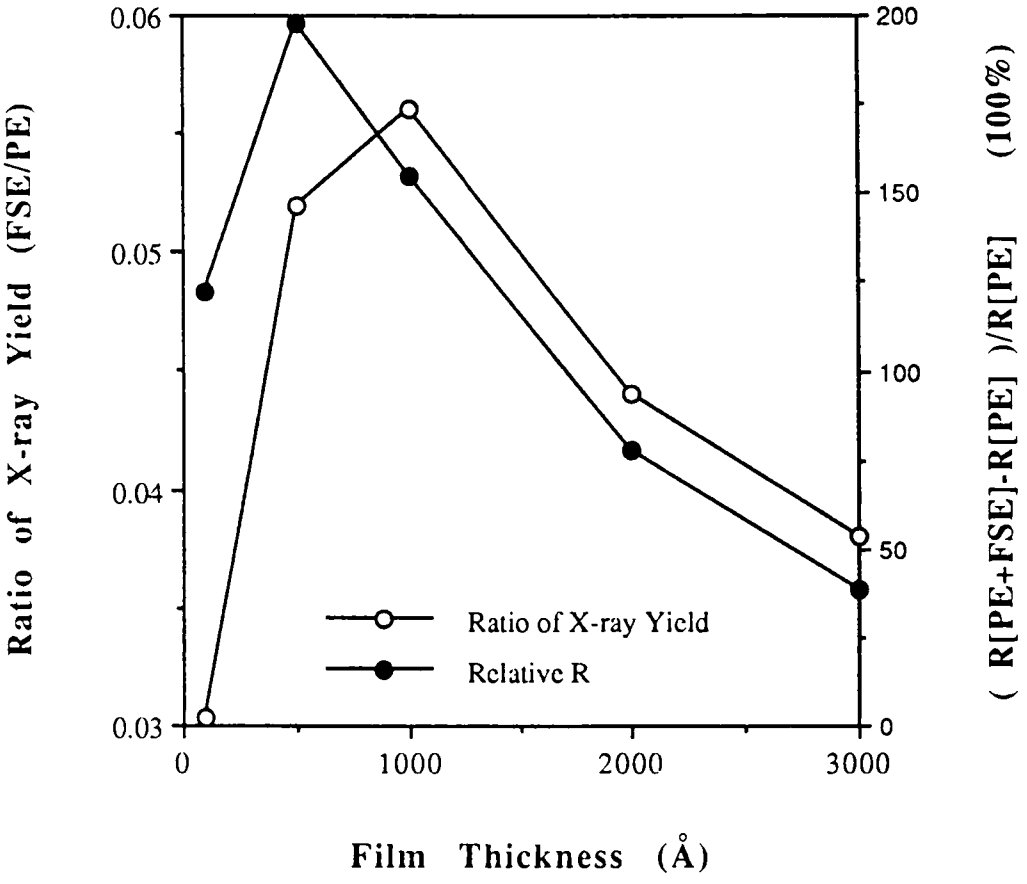


Figure 2.10. Plots of ratio of x-ray yield generated by primary electrons and by FSEs as a function of film thickness, and plots of relative effects of FSEs on x-ray spatial resolution as a function of film thickness. The right y-axis is the relative difference of x-ray spatial resolution with and without considering of FSEs.

degradation on X-ray spatial resolution increases with increase of specimen thickness. Figure 2.10 shows the plots of relative difference of R vs specimen thickness and the ratio of X-ray yield (FSE/PE) vs specimen thickness. The degree of FSE degradation increases rapidly for specimen thickness below 1000 Å for the case of carbon film.

REFERENCES

Bethe, H., Annalen de Phys. (Leipzig), **5**(1930), p. 325-333

Bishop, H.E., in NBS Special Publication 460, (ed. by Heinrich, K.F.J. et al), 1976, p. 5-15

Cliff, G. and Kenway, P.G., in Microbeam Analysis 1982, (ed. by Heinrich, K.F.J.), p. 107, 1982, San Francisco Press, San Francisco, CA.

Cliff, G. and Lorimer, G.W., in Quantitative Microanalysis with High Spatial Resolution, (ed. by Jacobs, M.H., Lorimer, G.W., and Doig, P.) Book 277, Metals Society (London), 1981, p. 47-52

Doig, P., Lonsdale, D., Flewitt, P.E.J., in Quantitative Microanalysis with High Spatial Resolution, Book 277, Metals Society (London), 1981, p. 41-46

Easterling, K.E., J. Mat. Sci., **12**(1977), p. 857

Evens, R.D., The Atomic Nucleus, (McGraw Hill, New York, 1955), p. 576

Goldstein, J.I., in Introduction to Analytical Electron Microscopy, (ed. by Hren, J.J., Goldstein, J.I., and Joy, D.C.), p. 102, 1979, Plenum Press, New York

Goldstein, J.I., Lyman, C.E., and J. Zhang, Proc. 12th Inter. Cong. Electron Microscopy, p. 450, 1990, San Francisco Press, San Francisco, CA.

Goldstein, J.I., Costley, J.L., Lorimer, G.W., and Reed, S.J.B., SEM/1977, I, p. 315-324

Hutchings, R., Loretto, M.H., Jones, I.P., and Smallman, R.E., Ultramicroscopy, **3**(1979), p. 401

Joy, D.C., J. Microscopy, **128**(1982), Pt2, RP1-RP2

Joy, D.C., Microelectronic Engineering, **1**(1983), p. 103-119

Joy, D.C. J. Microscopy, **143**(1986), Pt3, p. 233-248

Joy, D.C., Inst. Phys. Conf. Ser. No. 93: Volume 1, 1988, p. 23-32

Joy, D.C., Scanning Microscopy, **5**(1991), p. 329-337

Joy, D.C. and Luo, S., Scanning, **11**(1989), p. 176-180

Klein, C., Ayer, R. and Williams, D.B., in Intermediate Voltage Microscopy and Its Application to Materials Science, 1987, Electron Optics Publishing Group, Mahwah, N.J.

Kyser, D.F., Scanning Electron Microscopy, 1981, I., p. 47-62

Kyser, D.F. and Geiss, R.H., Proc. 12th Ann. Conf. Microbeam Analysis Society, 1977, paper 110, San Francisco Press, San Francisco, CA.

Michael, J.R., Proc. 47th Annual Meeting of EMSA, p. 202, 1989, San Francisco Press, San Francisco, CA.

Michael, J.R. and Williams, D.B., J. Microscopy, **147**(1987), p. 289

Murata, K. and Kyser, D.F., Advances in Electronics and Electron Physics, **69**(1987), p. 195

Myklebust, R.L., Newbury, D.E., and Yakowitz, H., in NBS Special Publication 460, (ed. by Heinrich, K.F.J. et al), 1976, p. 106

Newbury, D.E., in Microbeam Analysis 1982, (ed. by Heinrich, K.F.J.), p. 79, 1982, San Francisco Press, San Francisco, CA.

Newbury, D.E., Yakowitz, H., and Myklebust, R.L., in NBS Special Publication 460, (ed. by Heinrich, K.F.J. et al), 1976, p. 151-164

Newbury, D.E. and Myklebust, R.L., in Microbeam Analysis 1981, (ed. by Wittry, D.B.), p. 173, 1980, San Francisco Press, San Francisco, CA.

Newbury, D.E. and Myklebust, R.L., in Analytical Electron Microscopy 1981, (ed. by Geiss, R.H.), p. 91-98, 1981, San Francisco Press, San Francisco, CA.

Rao, P. and Lifshin, E., Proc. 12th Annual Conf. MAS, p. 118

Reed, S.J.B., Ultramicroscopy, 7(1982), p. 405

Romig, A.D., Newbury, D.E., and Myklebust, R.L., in Microbeam Analysis 1982, (ed. by Heinrich, K.F.J.), p. 72, 1982, San Francisco Press, San Francisco, CA.

Romig, A.D., et al, in Microbeam Analysis 1990, (ed. by Michael, J.R. and Ingram, P.), p. 275, 1982, San Francisco Press, San Francisco, CA.

Stephenson, T.A., Loretto, M.H., Jones, I.P., in Quantitative Microanalysis with High Spatial Resolution, (ed. by Jacobs, M.H., Lorimer, G.W., and Doig, P.) Book 277, Metals Society (London), 1981, p. 35

Wells, O.C., Scanning Electron Microscopy, p. 71, (McGraw Hill, New York, 1974)

Williams, D.B., Michael, J.R., Goldstein, J.I., and Romig Jr., A.D., Proc. 49th Annual Meeting of EMSA, p. 472, 1991, San Francisco Press, San Francisco, CA.

CHAPTER THREE

ELECTRICAL CHARACTERIZATION OF MATERIALS - EBIC METHOD AND MONTE CARLO SIMULATION

3.1 Introduction

Crystal defects are known to affect the performance of semiconductor devices, and the methods allowing a detailed analysis of structural and electrical characteristics of individual defects are considered with increasing interest for materials scientists and solid state physicists. Charge collection microscopy, usually known by the acronym EBIC (Electron Beam Induced Current) imaging, is a widely used powerful technique for the observation and characterization of semiconductor materials and devices in the scanning electron microscope (SEM) [Leamy, 1982; Holt, 1989]. The impurities and defects which influence the electrical parameters of the semiconductors can be characterized using EBIC method with a lateral resolution of about 1 μm [Czyzewski & Joy, 1990]

The principle of charge collection microscopy is shown in Figure 3.1. This technique is specific to semiconductors (and some nonmetallic materials) because the incident high-energy electron deposits its energy in the materials in a cascade process which eventually leads to the significant increase in the number of mobile charge carriers. Electrons are promoted from the filled valence band, where they are free to move around the lattice, and a correspondingly positively charged hole is left in the valence band, which is also mobile, forming a so-called electron-hole pair. Each

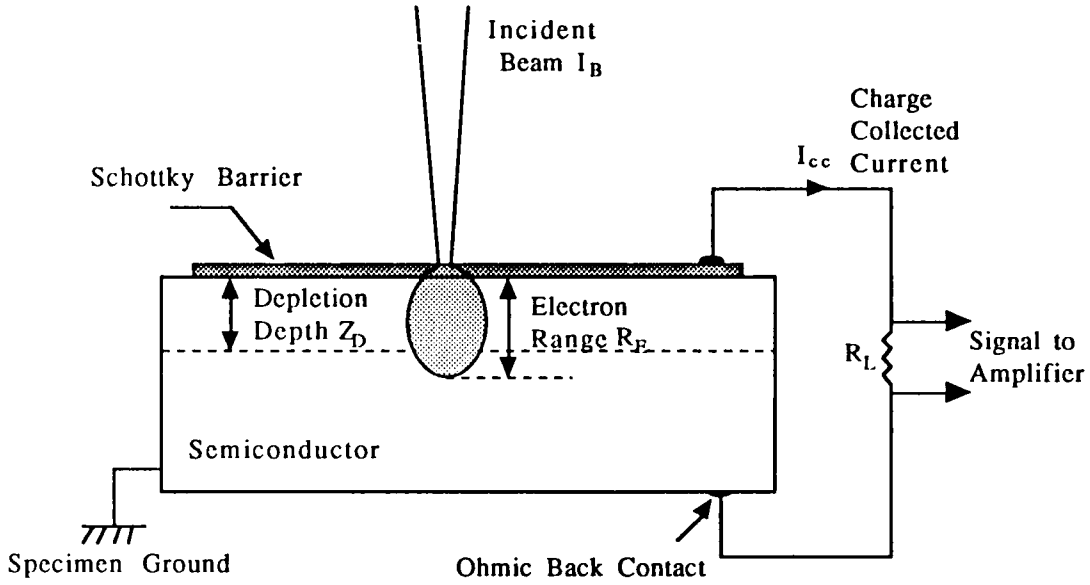


Figure 3.1. Schematic diagram of Schottky barrier charge collection.

incident electron can generate a large number of such electron-hole pairs because the energy required to produce each one is small, typically about three times the band-gap in the semiconductor [Klein, 1968] and thus a few electron volts, so that one 10 keV electron hitting a silicon target produces nearly 3000 electron-hole pairs. These electron-hole pairs are created throughout the interaction volume of the primary electrons. The electrons and holes tend to mutually attract and annihilate. However, if these carriers are generated in a region containing a superimposed electric field, such as is found in the space charge depleted region surrounding a p-n junction or Schottky barrier, then the electrons and holes are swept apart before recombination occurs and move in opposite directions. The internal motion of the carriers across the space-charge depleted region causes an equal amount of charge to flow round the external circuit connected to electrodes on the front and back surfaces of the specimen. If the carriers are deposited in neutral (field free) material then only that fraction which can diffuse back to the depleted region will result in external current flow, since the rest of the electrons and holes will recombine.

In the SEM the current flowing around the external loop is measured and displayed as a function of the beam position to produce the charge-collected, EBIC, image. Contrast arises at defects where the local recombination rate differs from the bulk rate for perfect material. In EBIC experiments, the specimen itself is used as the detector of the signal. The most common mode of EBIC operation relies on the bias from p-n junctions or Schottky barriers with the

current detected by amplifiers of relatively low input impedance. The electric field required for EBIC experiments may be produced by fabricating a p-n junction within the specimen, or a Schottky barrier at its surface, or by making the specimen the dielectric medium of a capacitor [Leamy et al, 1978]. Among them, the Schottky barrier is best suited to the characterization of semiconductor materials [Leamy et al, 1978] since the entire region covered by the barrier film can be observed.

A Schottky barrier is a thin metal film evaporated onto the surface of a semiconductor. At the junction between the metal and the semiconductor, a potential exists, typically 0.5 V, which depletes the region beneath the barrier in just the same way as at a p-n junction. The most important application of the Schottky barrier technique is to study electrically active defects in semiconductor materials. If a dislocation or stacking fault is electrically active, then it can trap a carrier for sufficient time to allow it to recombine with a passing carrier of the opposite sign. Such activity occurs at such features as line and point defects and grain boundaries, and its presence can have a devastating effect on the performance of any device fabricated over the top of the defect. The ability of the EBIC technique to observe these active defects is therefore an important tool for the study of materials.

In this chapter, the studies of electrical characterization of materials are discussed in three separate sections. Section 3.2 and section 3.3 are two manuscripts being submitted for publication. Section 3.2 is focused on the design of a specimen holder for EBIC

imaging on the Hitachi S800 SEM. Section 3.3 is focused on the development of a new way to fabricate Schottky barrier for EBIC study. Section 3.4 is focused on quantitative study of EBIC results using Monte Carlo simulation.

3.2 A Simple Specimen Holder for EBIC Imaging on the Hitachi S800 SEM*

3.2.1 Abstract

A design of a simple specimen holder to implement EBIC experiments on the Hitachi S800 SEM has been described. The quality and usefulness of the specimen holder is illustrated by EBIC images of polycrystalline silicon solar cell.

3.2.2 Introduction

Electron beam induced current (EBIC) technique has become widely used in the characterization of semiconductor materials and microelectronic devices. EBIC signals are measured in the form of a short circuit current, and this requires that both the front and the back surfaces of the specimen be electrically connected. We have developed a simple specimen holder to implement EBIC experiments on SEMs such as the Hitachi S800 which do not possess isolated electrical lead-outs. Here we will discuss the design of the specimen holder and illustrate the results by EBIC observations of polycrystalline silicon solar cells.

---

* This part is a manuscript of a paper being submitted for publication. The co-author is D.C. Joy, Electron Microscopy Facility, The University of Tennessee, and High Temperature Materials Lab, Oak Ridge National Laboratory

3.2.3 Structure of the specimen holder

The schematic diagram of an EBIC imaging circuit is shown in Figure 3.2. Carriers injected by the incident electron beam are collected by the depletion field and flow through the external circuit to give a current that can be used to produce an image by modulating the brightness of a CRT display scanned in synchronism with the probe. The short circuit current is amplified by a modified high-gain, large-bandwidth amplifier [Maher & Rossouw, 1983]. To collect the signal the amplifier must be connected to both sides of the junction.

The EBIC specimen holder is designed to be used with the standard specimen holder of the Hitachi S800 scanning electron microscope (SEM) or other similarly configured machines (see Figure 3.3). In this instrument the specimen is normally electrically attached to the stage mount which, in turn, is grounded through an external BNC coaxial plug. In order to obtain an EBIC image the stage is designed so as to ground one side of the sample internally, and then use the external coaxial connector to provide the other contact.

The specimen holder consists of an isolated stub on top of a base stage. A nylon screw is used to insulate the specimen from the holder. In one end of the base, two holes are drilled to support the contacts to the sample. Nylon screws and plastic washers hold the wires which are used to connect to the top specimen surface. Surface contacts to the device to be examined are made by gold wires held in position with gold paste in the front and with silver paste in the back. The wire from the top surface is further connected to the tilt

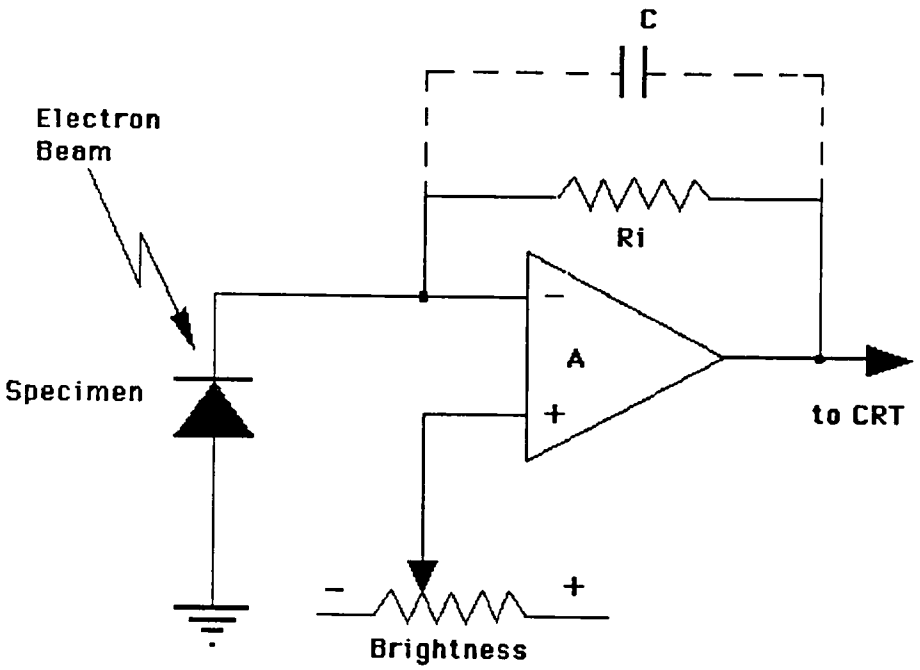


Figure 3.2. Schematic diagram of EBIC measuring and imaging system.

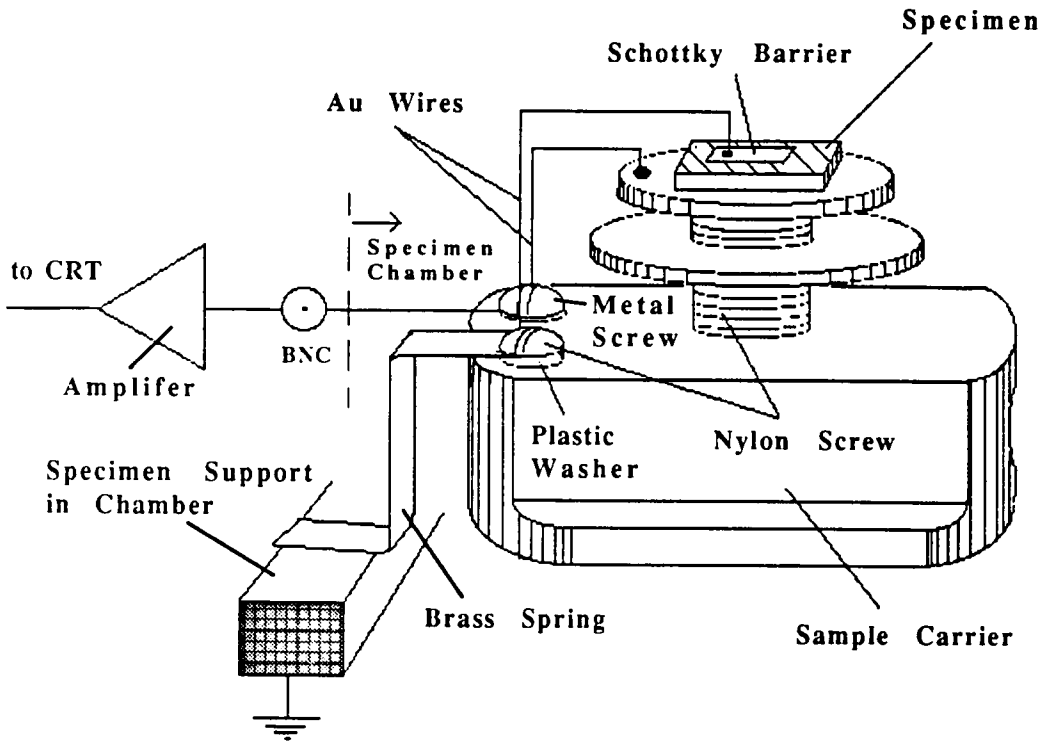


Figure 3.3. Schematic diagram of the design of the EBIC specimen holder.

screw bar (which is isolated from the main specimen stage) by a brass spring. As the module is inserted into position the spring contact grounds one side of the sample to the stage support in the specimen chamber while the other connects to the specimen holder and hence to the external BNC connector. A volt-ohm meter applied across the BNC plug will allow the device to be tested. The forward resistance should be 10% or less of the back resistance for successful EBIC imaging.

3.2.4 Results

Using this simple EBIC specimen holder, we have been able to obtain good EBIC signals. The following EBIC images are taken from a commercial polycrystalline silicon solar cell. The top surface (light-receiving) is polycrystalline silicon with silver contact grids and the backside is a deposited aluminum layer.

A major advantage of the EBIC technique is that it is sensitive to electrically active localized defects such as dislocations and grain boundaries. The most common type of EBIC contrast is interpreted as due to varying rates of enhanced recombination of electron-hole pairs at, or near, the defect. Figure 3.4 is two superposed images of secondary electrons (SE) and reversed EBIC of top surface of silicon solar cells. In reversed EBIC mode, brighter contrast means lower charge collection signals and hence a higher recombination rate. Figure 3.4(a) shows several electrically active dislocations (bright lines), one of which is seen to shift direction as it crosses a twin boundary. Figure 3.4(b) shows two "Y" shaped dislocations which are possibly partial dislocations associated with stacking faults. Figure

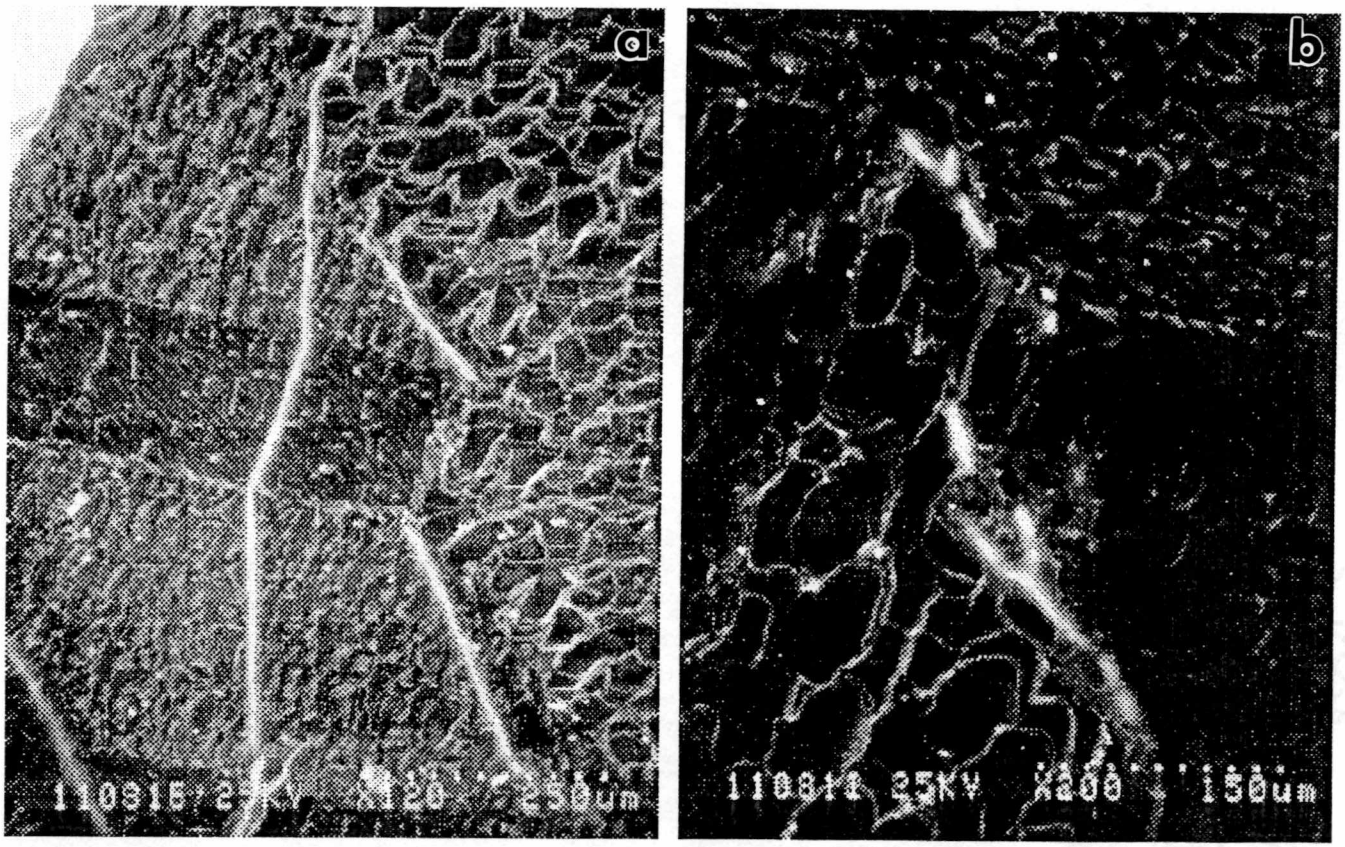


Figure 3.4. Superposed SE and reversed EBIC micrographs of dislocations on polycrystalline silicon solar cells (light-receiving side). Bright contrast = low EBIC signal.

3.5 shows one dislocation in a variety of display modes. Figure 3.5(a) is a superposed image of SE and reversed EBIC. Figure 3.5(b) is Y-modulated EBIC map of the same area. Y-modulated EBIC involves the use of the EBIC signal to modulate the vertical deflection of the CRT such that a more graphic presentation of spatial variation in the EBIC signal is given. Figure 3.5(c) is a higher magnification of the same area, while Figure 3.5(d) is a line scan EBIC signal of the dislocation as a function of distance. The profile of the EBIC signal is not symmetrical about the dislocation line, which may imply different recombination rates on the two sides.

With the change of beam voltage, it is possible to estimate the depth of a dislocation. An example is shown in Figure 3.6. Figure 3.6(a) is the SE image of an area where a dislocation is revealed under EBIC mode. Figure 3.6(b)-(d) show a series EBIC images of the dislocation with different beam voltages (from 25 keV down to 15 keV). The EBIC signal of the defect almost vanishes at 15 keV. Using Monte Carlo simulation [Joy, 1983], the depth of this dislocation can be estimated as between 2 and 4 μm. Although the localized high recombination rate at defects is well recognized, occasionally "bright" defect contrast in silicon is observed [Kittler, 1989]. Figure 3.7 (reversed EBIC image) is an example in which the interpretation of the contrast complicated. There are many suggested mechanisms for "bright" contrast, however, a unique theoretical modelling has not yet been performed [Kittler, 1989]. One possible reason for enhanced defect signal could be an additional charge collection within the space charge region around defects resulting from an

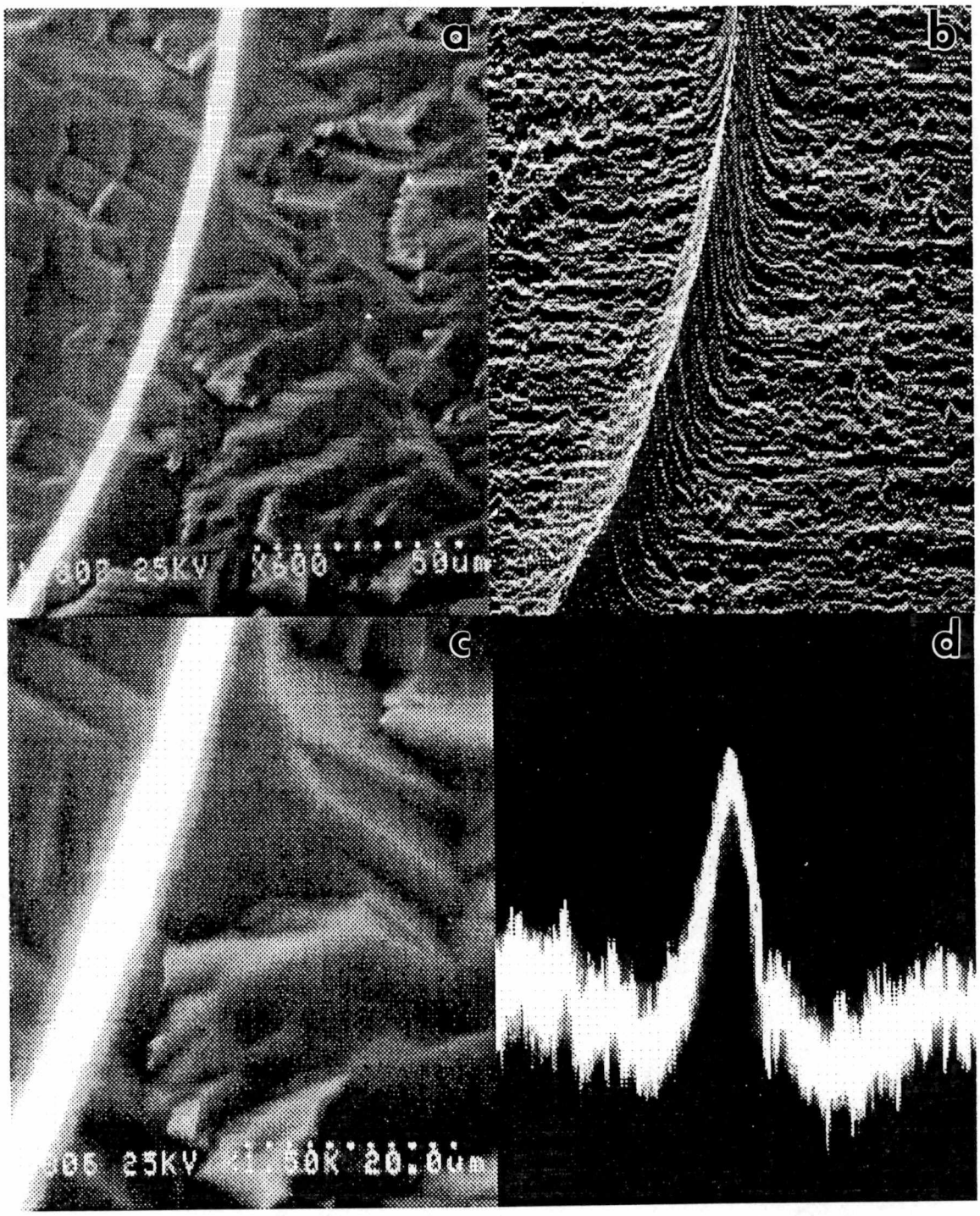


Figure 3.5. (a) Superposed SE and reversed EBIC micrograph of a dislocation in a silicon solar cell. Bright contrast = low EBIC signal,(b) Y-modulated reversed EBIC map of the same area, (c) higher magnification micrograph of the same dislocation of (a), and(d) the EBIC line scan trace across the dislocation in (c).

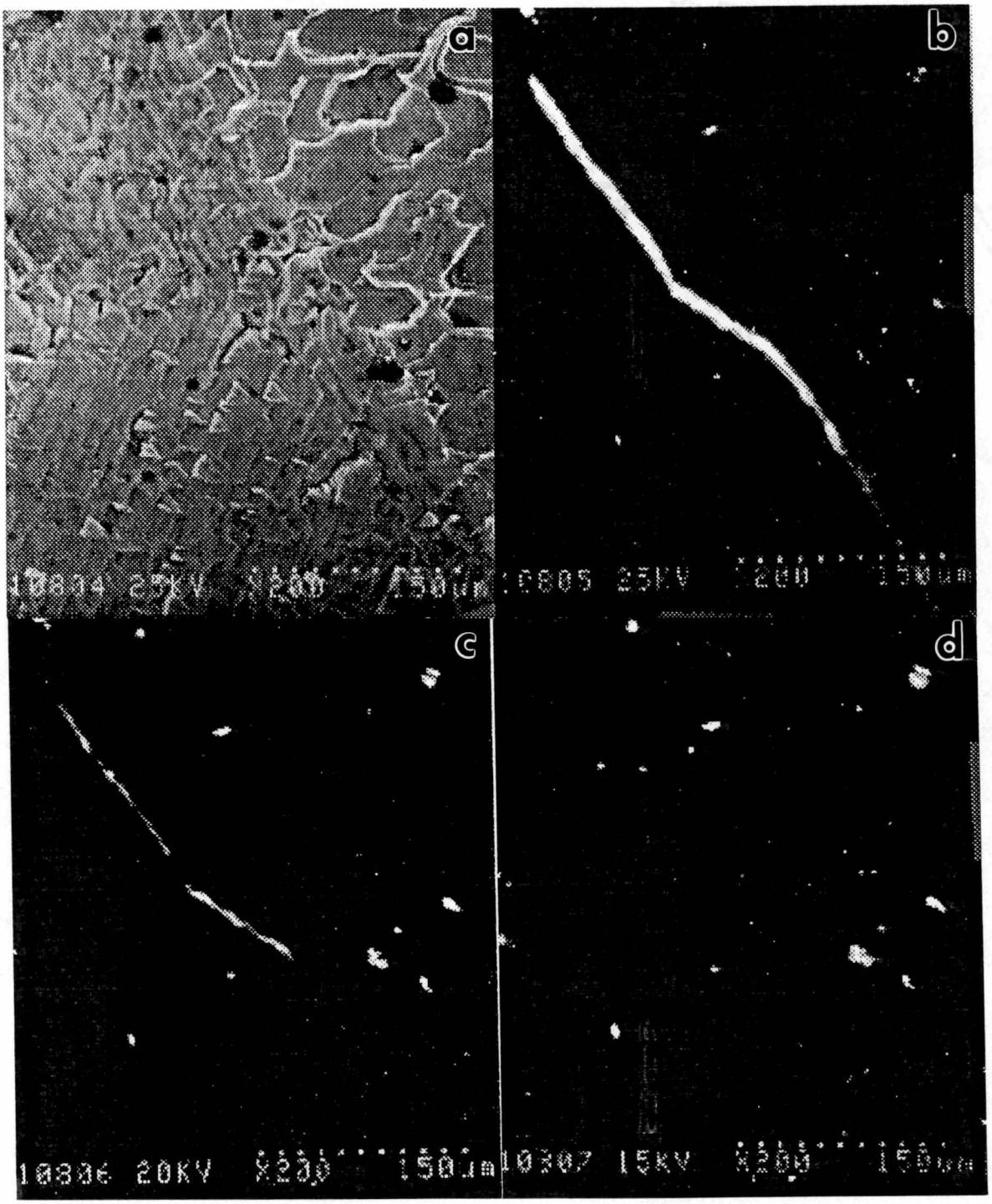


Figure 3.6. Depth resolution in EBIC. (a) SE micrograph of a silicon solar cell, (b) reversed EBIC micrograph of a dislocation in the same area of (a) at 25 keV beam energy, (c) at 20 keV beam, and (d) at 15 keV. It is almost invisible.

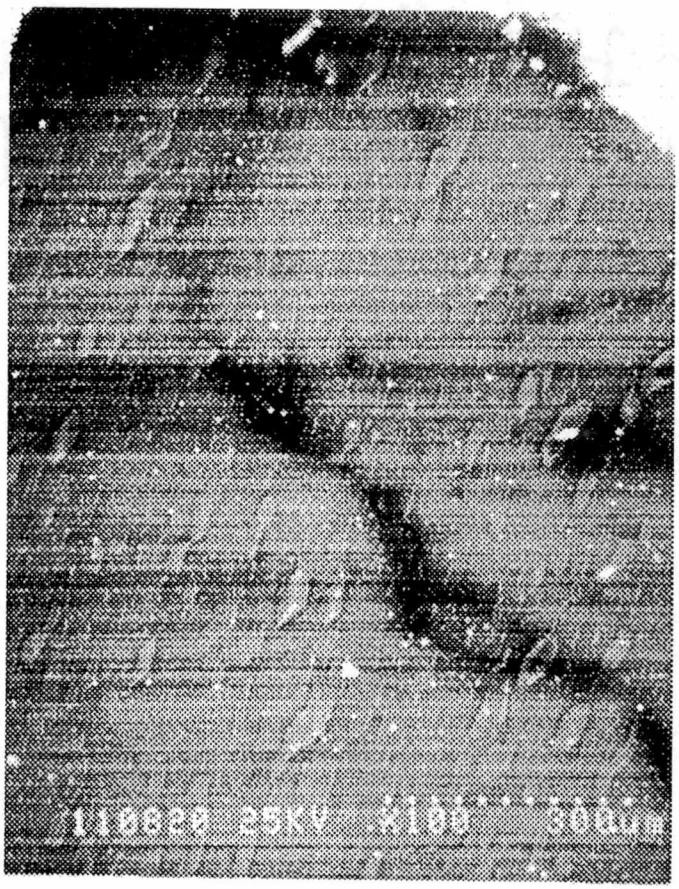


Figure 3.7. Superposed SE and reversed EBIC micrograph of dark contrast (i.e. high EBIC signal) of defects.

occupation of defect centers by majority carriers [Blumtritt et al, 1988]. It has been realized that the "bright" contrasts result mostly from defect-impurity interactions.

In polycrystalline solar cells, the grain boundary recombination can represent a significant loss in short circuit current and consequently in output power of the solar cell [Hanoka & Bell, 1981]. Figure 3.8 shows the secondary electron image of grain boundaries of a polycrystalline silicon solar cell, and the reversed EBIC (Y-modulated) image of the same area. The grain boundaries in the reversed EBIC image mode show as high contrast (reversed) which indicates high recombination rates along the boundaries, but note that the contrast varies along some of the boundaries indicating variation of recombination rate.

As a conclusion, the simple EBIC specimen holder described is a useful accessory for SEMs, such as the Hitachi S800, which does not possess isolated electrical lead-outs, and provides micrographs of excellent quality. The images of the solar cell illustrate the variety of detail available using this mode of imaging.

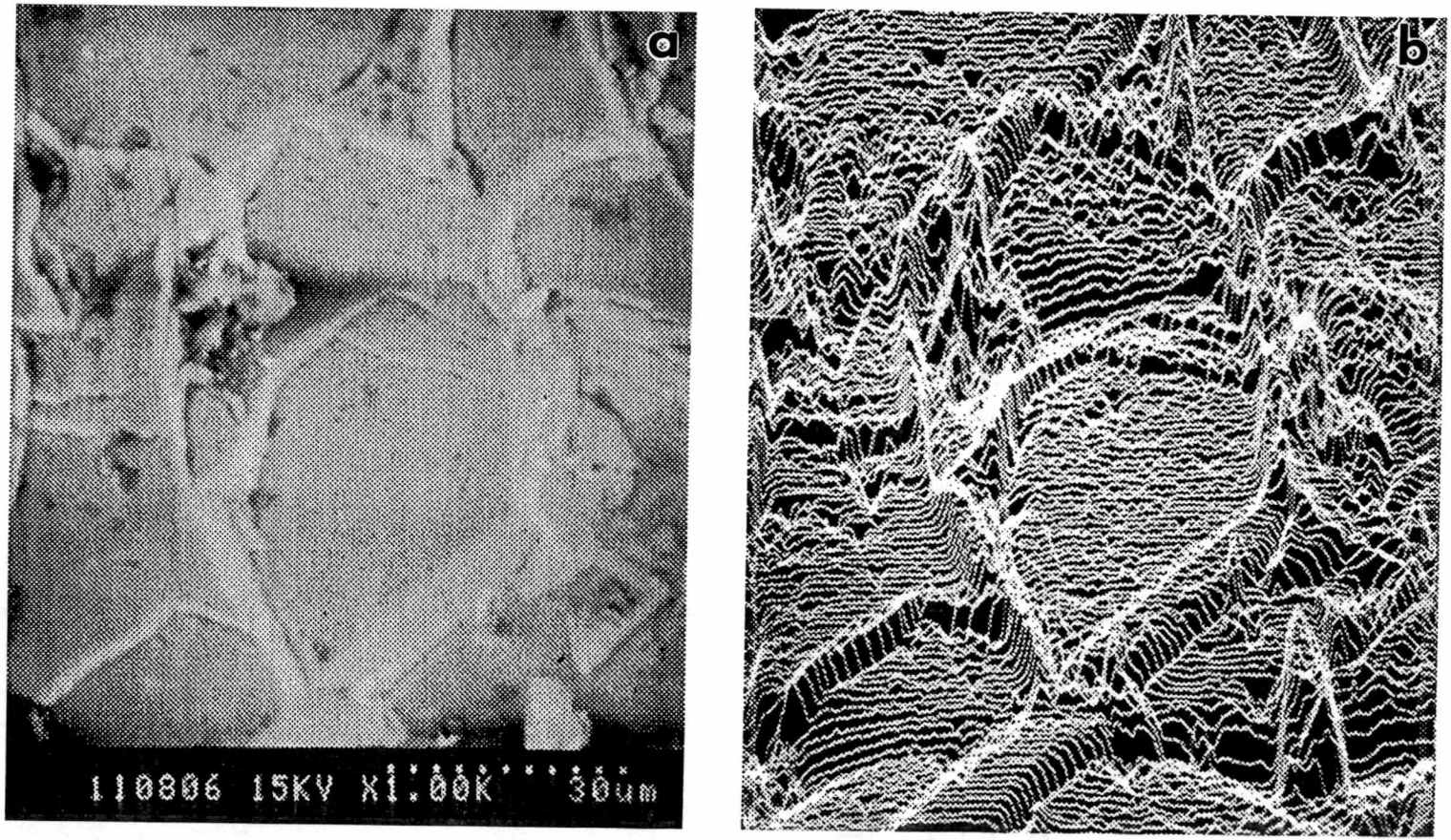


Figure 3.8. (a) SE micrograph of grain boundaries of the polycrystalline silicon solar cell, and (b) Y-modulated EBIC map of the same area of (a).

3.3. Fabrication of Au/p-Si Schottky Barrier for EBIC Study*

3.3.1 Abstract

A simple method for fabrication of a Au/p-Si Schottky barrier suitable for EBIC study has been developed. The fabricated Au/p-Si Schottky barriers have been tested in terms of their mechanical and electrical properties. EBIC measurements of dislocation contrast has been conducted using the fabricated Schottky barriers.

3.3.2 Introduction

Rectifying metal-semiconductor contact, known as Schottky barrier, is important not only for semiconductor device manufacture but also for microcharacterization of semiconductor materials. Schottky barrier charge collection microscopy is a widely applicable technique for the study of defects in semiconductors, since the entire area covered by the barrier film can be observed. Schottky barrier is comparatively easy to fabricate, does not exhibit minority carrier effects that is similar to those observed in p-n junction devices [Sharma, 1984], and is sensitive to the electrical properties of the semiconductor defect [Leamy et al, 1976]. Various metal-semiconductor systems have been studied for making Schottky barrier, within which Si Schottky barrier is of primary interest in this study.

Making a Si Schottky barrier is usually involved in evaporating a metal thin film on an etched Si surface in a vacuum chamber

*This part is a manuscript of a paper being submitted for publication. The co-author is D.C. Joy, Electron Microscopy Facility, The University of Tennessee, and High Temperature Materials Lab, Oak Ridge National Laboratory

having a pressure of about 10^{-6} Torr or evaporating the metal on *in situ* cleaved Si surface in an ultrahigh vacuum ($<10^{-8}$ Torr) [Rhoderick & Williams, 1988]. Sputtering methods have also been reported [Gutknecht & Strutt, 1971; Mullins & Brunnschweiler, 1976; Finetti *et al*, 1984]. One of the advantages of the sputtering method is the good mechanical adhesion of metal films to the Si surface [Rhoderick & Williams, 1988]. The purpose of this study is to develop an alternative fabrication method for Schottky barrier which does not need UHV evaporation, instead only basic chemical lab equipment and a routine low vacuum sputter coater are required.

It has been reported [McGee, 1982; Ponjee et al, 1984] that 3-mercaptopropytrimethoxysilane (MPS) can be used to improve the adhesion of Ag and Au evaporated films to glass and SiO_2 substrates. It has been proposed that the silane moieties covalently bind to the oxide surface through siloxane bonds [Goss et al, 1989]. Here we will discuss a new simple method of making a Au/p-Si Schottky barrier for EBIC study with a low vacuum requirement and a better mechanical adhesion. The characteristics of this Schottky barrier have been studied. Also some preliminary EBIC results of the Schottky barrier will be presented.

3.3.3 Experiments

The fabrication method is based on a method suggested by Goss et al [1989] with a few modifications. A commercially obtained p-type Si wafer with etched surface was used as a substrate. Visible residues on the wafer surfaces were removed by wiping with lens paper. The wafers were then chemically cleaned by immersion in 1:4

H_2O_2/H_2SO_4 at 70°C for about 20 min., rinsed with distilled H_2O, and stripped off visible surface water with a stream of Ar. In order to make the wafer surfaces completely hydrophobic, the surfaces were further cleaned in 3:7 HF/H_2O solution for 1 min. After above cleaning procedures, the wafers were then placed in an oven at 105°C for about 10 min.

The silanization solution consisted of 10 g MPS, 10 g H_2O, and 400 g 2-propanol. The MPS solution was placed in a reflux. The cleaned and dried wafers were immersed in the boiling solution for 10 min., then carefully rinsed with 2-propanol, blown dry with a jet of Ar, and cured in a drying oven at 100-107°C for 8 min. The procedure of immersion, rinsing, and curing was carried out twice more. The estimated thickness of the MPS layer from the x-ray photoelectron spectroscopy data is about 100 Å.

Gold sputtering-deposition was carried out in a Emitech K550 Sputter Coater at room temperature. The sputter coater equipped with a Au target (with 60 mm diameter and 0.1 mm thick) and operated at a base pressure of 0.1 mbar ($7.5x10^{-2}$ Torr) with 4 p.s.i. Ar gas supply. Stainless steel masks were used to define the Schottky barrier area. The Au film thickness was measured by weighting method. The active barrier areas were in the range of 0.02-0.15 cm^2. The thickness of the Au layers was calibrated by electron energy loss spectroscopy (EELS) data measured in a Hitachi H800 scanning transmission electron microscope (STEM). Different sputtering parameter setting were used. The resultant Au film thickness was in the range of 100-350 Å.

A simple qualitative test was used to examine the adhesion of the gold films. A piece of Scotch tape was firmly placed over the Au film and the surrounding Si substrate, and was then removed. The fraction of the film that was transferred to the tape was then a relative measure of the film adhesion.

After fabricating the barriers, current density-voltage (J-V) and capacitance-voltage (C-V) measurements were subsequently performed in atmospheric condition. Surface ohmic contacts are made by Au wires with Au paste on the top of the Au film and with silver paste on the back. The J-V characteristics of Au/p-Si Schottky barrier were measured using a curve tracer and a combination of a DC power supply, a Fluke37 digital multimeter and a Fluke75 digital multimeter. The C-V measurements were carried out by a combination of a power supply, a HP4332A capacitance meter, a Fluke37 digital multimeter, several 1 MΩ resistors and a 480 pF capacitance. Parasitic circuit capacitance in series with the barrier under test were measured prior to making electrical contact to the barrier, and was subsequently subtracted from the measured capacitance to infer the barrier capacitance.

EBIC performance can be characterized by measuring the current gain of the diode at various accelerating voltages, $g = I_{cc}/I_b$, where I_{cc} is the measured short-circuit current collected for a given incident beam current I_b [Joy, 1986]. EBIC study of the plastically deformed Au/p-Si Schottky barrier was performed in a Hitachi S800 SEM using a newly designed specimen holder (see Section 3.2 of this chapter) and a modified high-gain, large-bandwidth amplifier [Maher

& Rossouw, 1983]. The short-circuit currents were measured by an Tektronix 5103N oscilloscope with a variable resistor. The incident electron beam currents were measured on to a metal strip using a Keithley 602 electrometer, and the correction for electron backscattering was made from Monte Carlo calculations. The contrast of the defect (C%) is defined by $C\% = 100 \bullet (I_B - I_d)/I_B$ [Joy, 1974], where I_B and I_d are the EBIC current far from any defect, and at the defect respectively.

3.3.4 Results and discussion

The adhesion of the barrier film to the semiconductor is important since the barrier behaves as an diode only when a proper adhesion exists. In adhesion test we found that when Au was coated directly onto the cleaned Si substrate without MPS treatment, the gold was completely removed by the tape. No transfer to the tape was observed when a MPS underlayer was used. It has been inferred from the similar test that the yield strength for the MPS/Au assembly is greater than 80 g/cm^2 [Goss et al, 1989].

The surface morphology of the sputtered Au layer is studied by SEM. Figure 3.9 shows an SEM micrograph of 165-Å-thick Au layer on p-type Si. A flat surface with fine Au grains and a lot of microcracks is observed. Ideally one needs a very smooth surface of the barrier film but since the EBIC contrast is strong, some roughness of the film is acceptable.

The sputtered Au/p-Si Schottky barriers have diode characteristics in terms of the changes in forward-backward resistance ratio when they respond to the light. The changes in the



Figure 3.9. SEM secondary electron micrograph showing surface morphology of a 165-Å-thick Au layer on p-Si. The sputtering was performed at 15 mA for 4 min.

forward-backward resistance ratio are in the order of 10 (dark)-200 (light). Usually a Schottky barriers with a ratio of 10-100 is considered as a good one .

The barrier heights, $\Phi_{b,}$ of the Au/P-Si Schottky barriers were obtained by the J-V and the C-V measurements. Figure 3.10 shows a typical forward J-V characteristics for the sputtered Au/p-Si barrier. In all cases an exponential J-V behavior is observed, with an ideality factor, n, in the range of 2-3, and the current density was truly exponential with voltage only over a short range. In such cases, the effective barrier heights can only be deduced from the saturation current density values (J_s) by extrapolating the value of J towards zero voltage. The J_s has the relation with the Φ_b by:

$$J_s = A^* \cdot T^2 \cdot \exp\left(-\frac{q \cdot \Phi_b}{K \cdot T} \right) \qquad (3.1)$$

where A* is the effective Richardson constant ($A^*=32\ cm^{-2}K^{-2}$ for p-Si), T is the absolute temperature, q the electron charge, and K the Boltzmann constant. Table 3.1 shows the Φ_b data deduced from the J-V measurements.

Table 3.1 shows that the barrier heights are much higher than the reported values of 0.34 eV for Au/p-Si Schottky diode made by evaporation method [Smith & Rhoderic, 1971]. This effect has been well documented as the sputtering damage due to the generation of donor-like defects closed to the semiconductor surface [Mullins & Brunnschweiler, 1976; Grusell *et al*, 1980]. It produces a potential minimum in the valence band so that the effective barrier height of p-type Si is increased [Finetti et al, 1984; Adegboyega,1989].

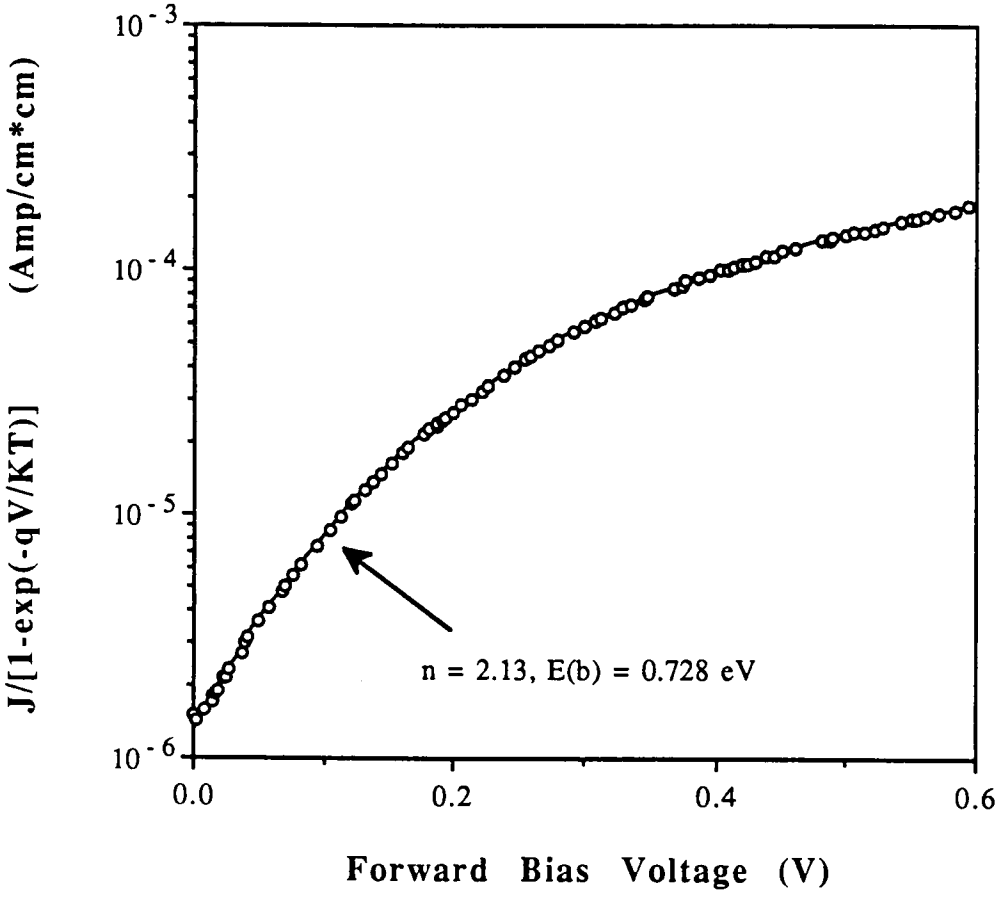


Figure 3.10. Forward current density-voltage characteristics for an Au(165Å)/p-Si Schottky barrier.

Table 3.1. Barrier Heights of Au/p-Si Schottky Barriers

Au/ p-Si Schottky barrier	J - V Measurement		C - V Measurement	
Thickness (Å)	n	Φ_b (eV)	Φ_b (eV)	N_a ($1/m^3$)
102	2.69	0.781	0.552	4.30×10^{19}
165	2.13	0.728	1.456	3.41×10^{20}
245	2.35	0.847	0.875	1.30×10^{26}
350	3.46	0.752	0.401	8.12×10^{19}

One possible reason of the short exponential ranges of the J-V plots is the recombination in the depletion region. If one adds the recombination term suggested by [Rhoderick & Williams, 1988] into the ideal thermionic-emission term, a conventional logarithmic plot of J-V is well fitted over four decade (Figure 3.11) and an ideality factor of 1.064 was obtained. Recombination current may therefore cause apparent departure of n from unity. The deviation from the ideal thermionic-emission behavior may also induced by interfacial SiO_2 layer and the MPS layer.

The barrier heights also show some variation with different thickness of Au layer (see Table 3.1). But they are not consisted with the values from the C-V measurements. Figure 3.12 shows a typical C^{-2}-V characteristics of the barrier. For C-V characteristics of the barrier, the relation

$$C^{-2} = (2/A^2\, q\, N_a\, \varepsilon_s)\cdot(\Phi_b - \xi + V_R - KT/q + \Delta\phi) \qquad (3.2)$$

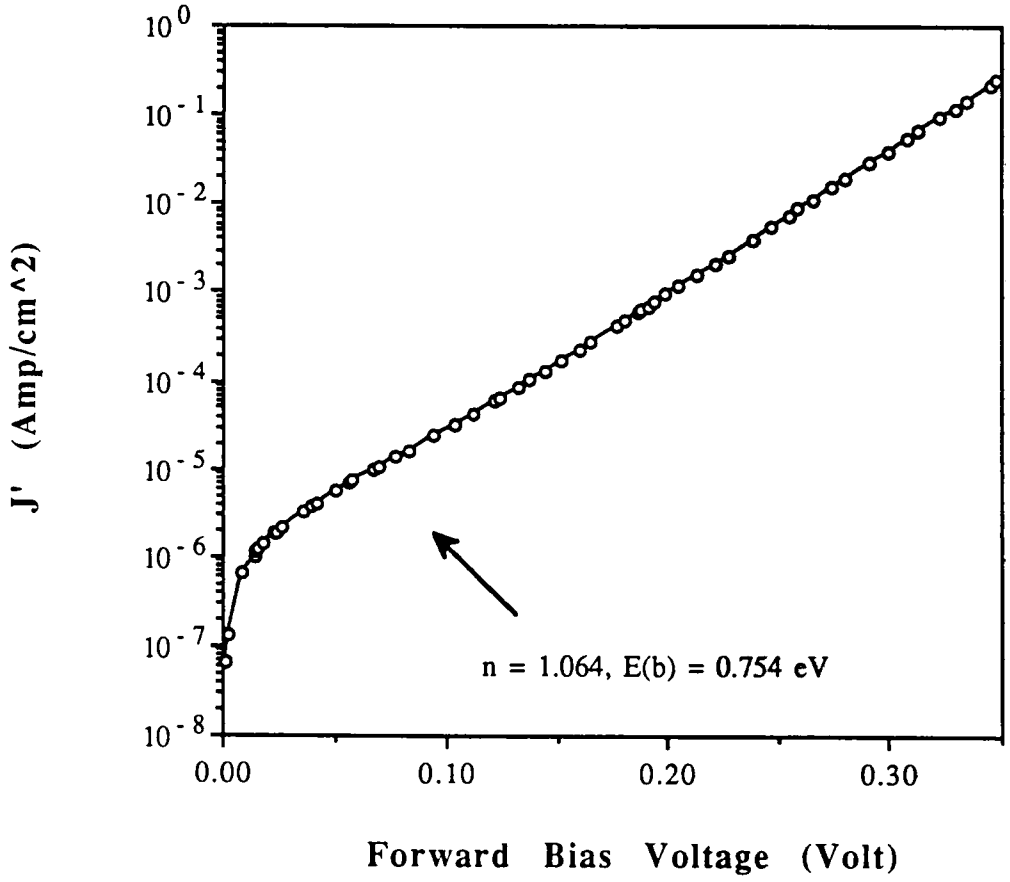


Figure 3.11. The forward current density-voltage characteristics (shown in Fig. 3.10) after correction for recombination in the depletion layer.

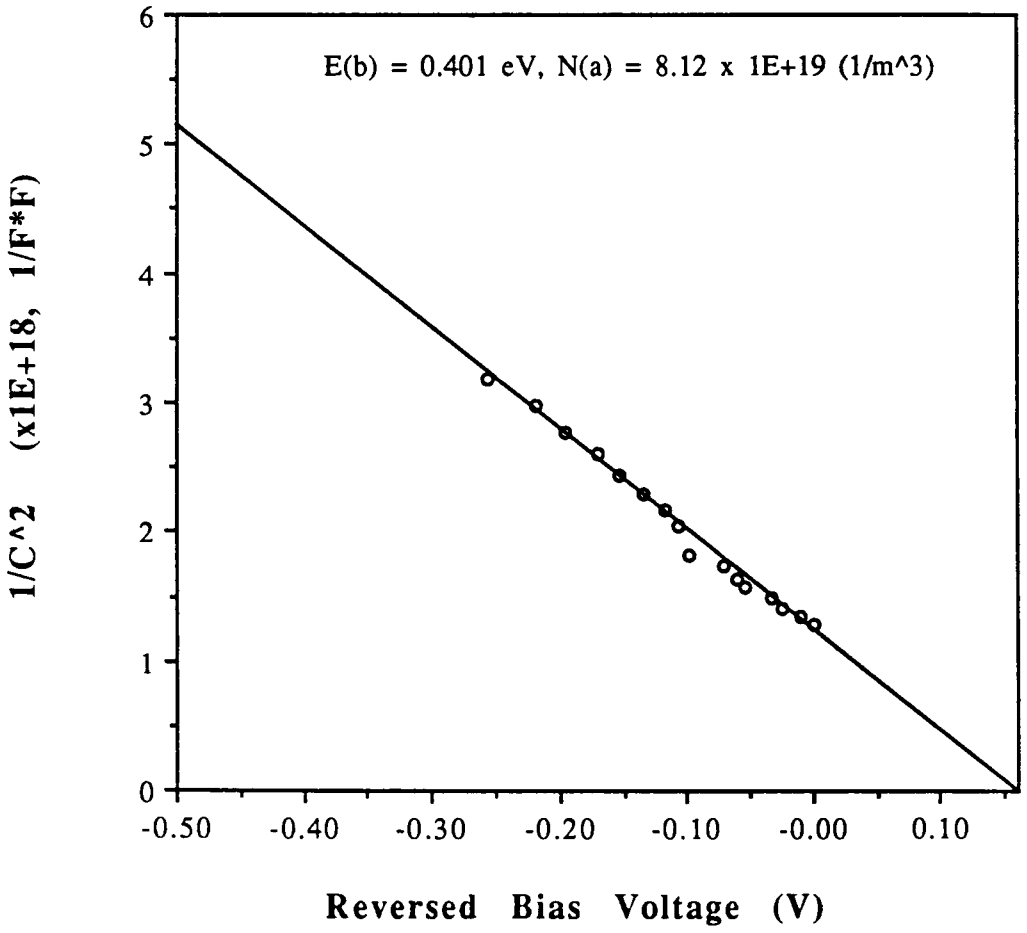


Figure 3.12. A plot of $1/C^2$ vs V for a Au(350Å)/p-Si Schottky barrier.

holds [Rhoderick & Williams, 1988], where C is the differential capacitance, A the barrier area, N_a the acceptor concentration, ε_s the permittivity of Si, V_R is the applied reverse bias, ξ the energy difference between the bulk Fermi level and the edge of the valence band (0.235 eV) [Charlson & Lien, 1975], and $\Delta\phi$ is the image force barrier lowering (0.0138 eV) [Charlson & Lien, 1975]. The Φ_b were derived from the intercept voltage V_I at $1/C^{-2} = 0$ of the plot C^{-2}-V_R if a straight line is observed. The barrier height is then given by:

$$\Phi_b = (V_I + \xi + KT/q - \Delta\phi) \qquad (3.3)$$

The values of Φ_b derived by using this method are also reported in Table 3.1. It is noted that the barrier heights derived from the C-V measurements are generally lower than that for the J-V measurements, except the case of 250 Å-thick Au film, which is closed to the value from the J-V measurement. This discrepancy has also been reported [Finetti et al, 1984] and has also been attributed to the donor-like traps created at the Si/metal interface during the sputtering process. The low Φ_b values from the C-V measurements may also induced by the insulating interfacial SiO_2 layer and the MPS layer, which add another negative term in the Φ_b equation [Goodman, 1963].

Under fixed V_R the depletion layer width W below the surface is obtained from the relation [Sharma, 1984]:

$$W = (2\cdot\varepsilon_s/qN_a\cdot|\Phi_b - \xi + V_R|)^{1/2} \qquad (3.4)$$

On the other hand, it can be shown that the local net impurity density N(x) at x is given by

$$N(x) = - C^3/(\varepsilon_s q A^2)\cdot(dC/dV)^{-1} \qquad (3.5)$$

for the case of non-uniform impurity distribution [Blood & Orton, 1978], where A is the barrier area. Using the data of 165 Å-thick Au/p-Si Schottky barrier, a plot of N(x) - W is shown in Figure 3.13. A plot of EBIC gain vs maximum electron penetration depth is superposed on it.

The following section will show some imaging results of our EBIC study of the Au/p-Si Schottky barrier.

Localized defects such as dislocations and precipitates are easily revealed by EBIC technique under Schottky barrier geometry. EBIC contrast generally arises from recombination of beam generated carriers at the defect. Figure 3.14 shows a EBIC micrograph of dislocations in Au (165Å)/p-Si Schottky barrier at 6 keV beam voltage. The contrast variations shown in the bottom were obtained by the beam scan along the straight line which across the dislocations. The contrast is only visible from the dislocation when the beam voltage is at 3 keV, which indicates that the dislocations lie at a depth equal to, or less than, 0.2 μm, the calculated electron range [Kanaya & Okayma, 1972; Joy *et al*, 1986]. It is found that the magnitude of the contrast from the dislocation is a sensitive function of its depth relative to the electron range. Figure 3.15 shows the variation of the contrast from the dislocations indicated in Figure 3.14. The contrasts arise at 3 keV when the dislocation is started to

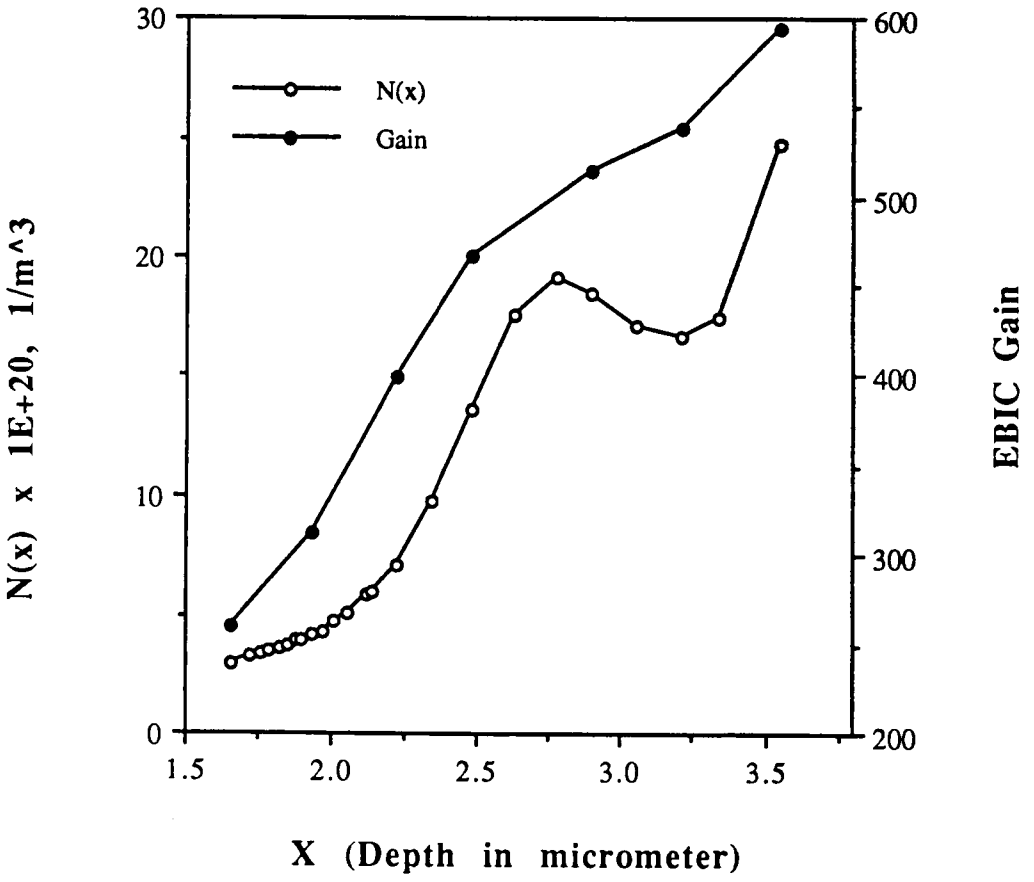


Figure 3.13. A plot of net impurity distribution vs depletion depth superposed with a plot of EBIC gain vs maximum electron penetration depth.

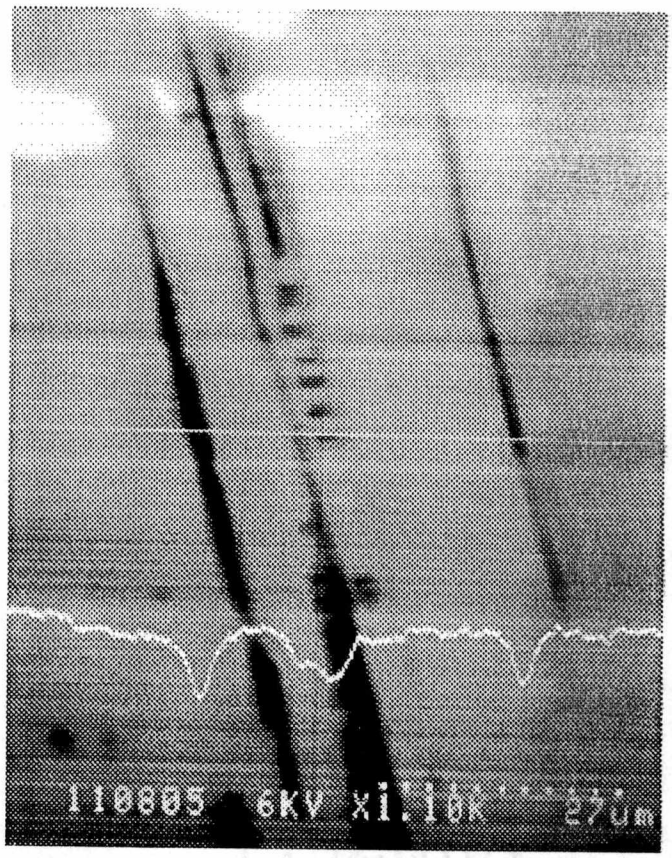


Figure 3.14. An EBIC micrograph of dislocations in a Au(165Å)/p-Si Schottky barrier with no external bias (at 6 keV beam energy). The central line is the trace of a electron beam scan. The bottom contour is the correspond contrast variation.

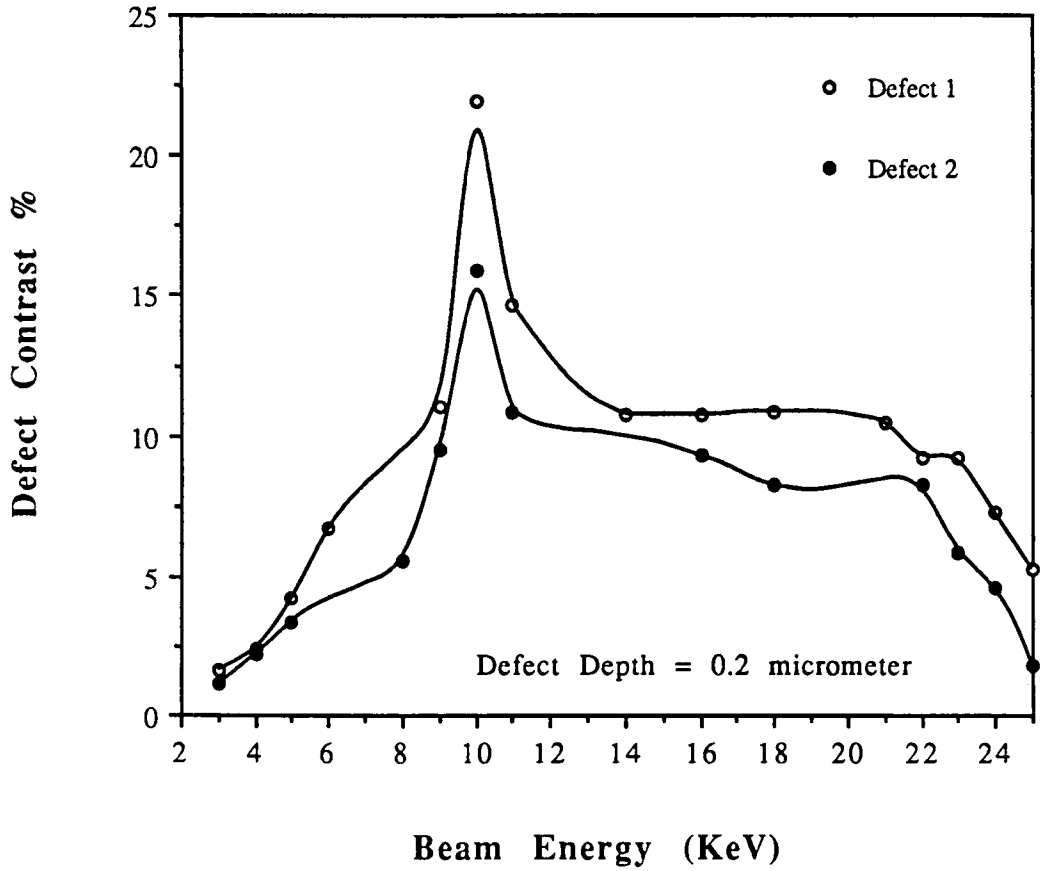


Figure 3.15. Measured variation of EBIC contrast with beam energy for the dislocations shown in Fig.3.14. The dislocation depth is estimated about 0.2 μm.

be revealed. The contrasts increase rapidly with the increase of the beam voltage, and reach a maximum at about 10 keV. Further increase in beam energy leads to a steady fall in the contrasts and a broadening of the dislocation profile. Apparently, two dislocations have different recombination rates in terms of the contrast values. It generally agrees with the results from the Monte Carlo simulation [Joy et al, 1986]. The maximum contrast occurs at the point where the electron range is about 3 times the dislocation depth, which is predicted by the Monte Carlo simulation.

The resolution of defect depends on N(x, y, z), the density of beam generated electron-hole pairs per unit energy, on the defect shape and position with respect to the surface, on the electric field strength, E(z), as well as on the local lifetime and its carrier density dependence [Leamy et al, 1976]. The expected shape of N(x, y, z) is reflected in the broadening of the image with the depth. It is expected that N(x, y, z) scales approximately with electron range. If we assume $E^{1.75}$ relation is valid for the dependence of electron range on the beam voltage, a $E^{1.75}$ dependence of the dislocation width (defined as a FWHM width) at a given depth is observed (Figure 3.16).

EBIC observations on some of the Au/p-Si Schottky barriers have revealed prominent bright "clouds" (plasma sites) as shown in Figure 3.17. These "clouds" do not correspond to any visible surface features in secondary electron image. When increasing the beam current to increase the excitation density, we found that inside each "cloud," there are some precipitate-like dark contrasts (Figure 3.18).

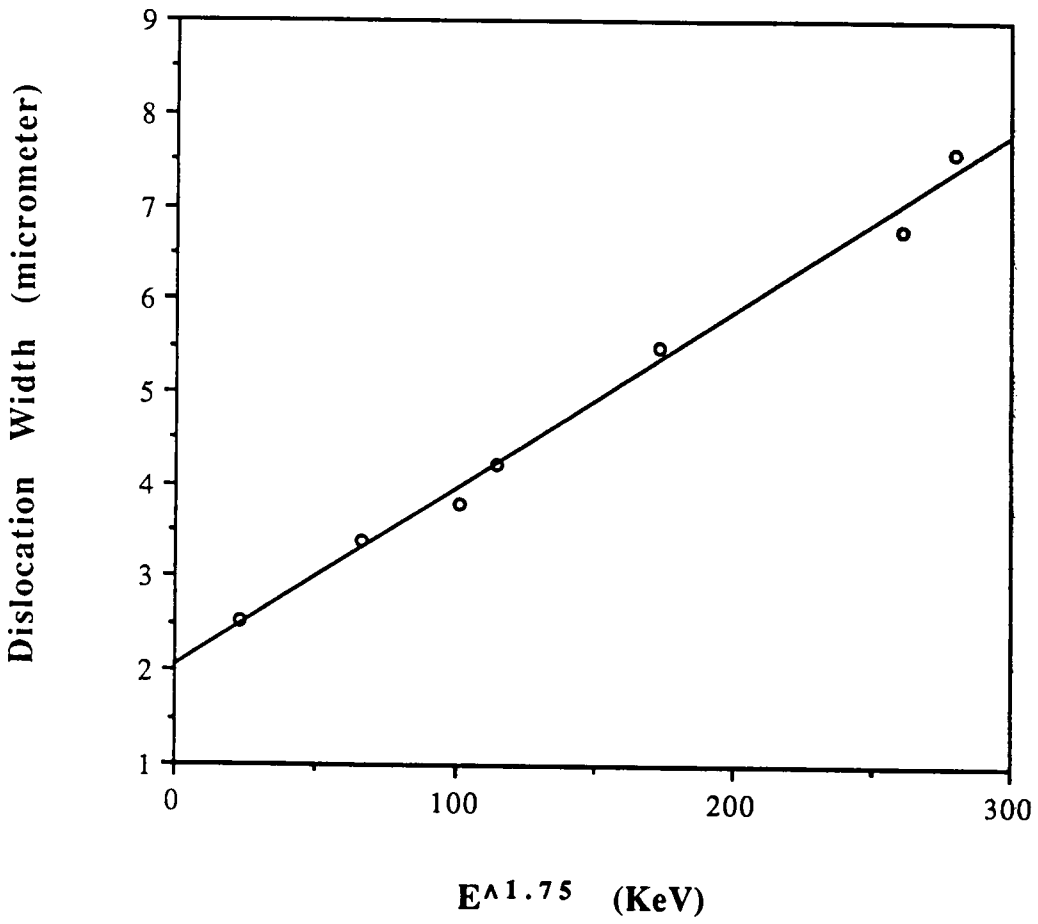


Figure 3.16. A plot of dislocation width vs $E^{1.75}$ (E is beam energy in keV). The dislocation width is defined as the full width at half maximum of the change in contrast.

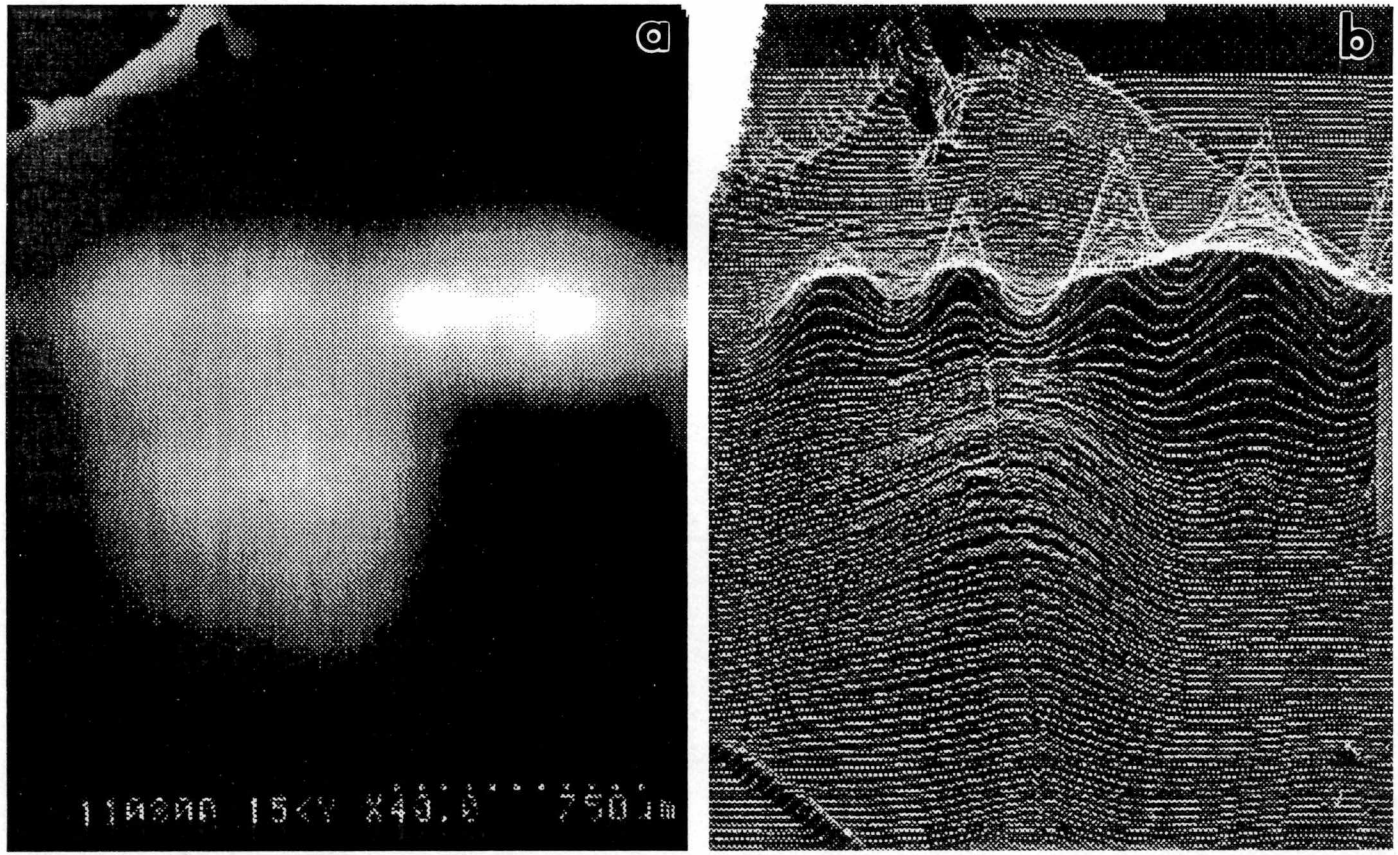


Figure 3.17. (a) 15 KeV EBIC image showing the presence of the "clouds" in Au(165Å)/p-Si Schottky barrier. (b) Y-modulated EBIC signal of the same area.

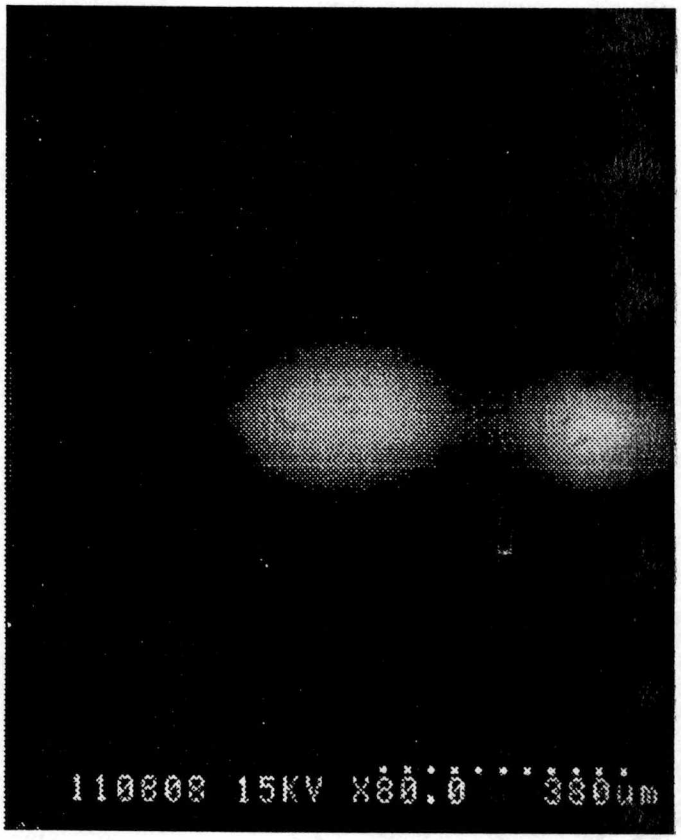


Figure 3.18. A 15 keV EBIC image with a higher beam current (5 nA) showing some precipitate-like dark contrasts inside each "clouds".

These plasmas have a characteristic "turn-on" voltage at 4 keV. When we measured the EBIC gain at one of the plasma site, and compared with the net impurity distribution N(x) obtained by the C-V measurement, we found that there are some correlation between the gain and the impurity density (Figure 3.13). It may imply that the plasmas are induced by some precipitates or impurities which concentrate the field lines such that beam injected carriers are preferentially accelerated and gain energy [Leamy et al, 1978]. Charge multiplication contrast arises when this energy gain exceeds the local ionization energy of the p-Si. More evidence of the impurity effect on the plasma generation will be discussed in section 3.4 of Monte Carlo simulation study.

As a conclusion, we have investigated a simple method for fabrication of a Au/p-Si Schottky barrier suitable for EBIC study. The Au/p-Si Schottky barrier strongly adheres to the Si substrate and is comparable on quality to typical properties of a conventional Schottky diode in terms of the optical, the J-V and the C-V characteristics. The sputtering damage is observed in the electric property measurements. In EBIC studies, we have measured the contrasts from the dislocations. The contrast dependence on beam voltage agrees well with the results calculated by Monte Carlo simulation. The defect width has a linear relation with the electron range. Several plasma sites are found in EBIC images. They may be induced by some precipitates.

3.4 EBIC Study Using Monte Carlo Method

As discussed in Chapter Two, Monte Carlo simulation is the most important method available for treating the interactions of the electron beam with solid specimens [Joy, 1989]. Quantitative analysis of EBIC defect image needs detailed understanding of beam-specimen interaction, and again, Monte Carlo simulation is an indispensable tool for the study.

The information necessary to calculate an EBIC image is firstly the volume density of production of electron-hole pairs produced by electron bombardment [Newbury, 1989; Holt, 1989]. This calculation involves three steps [Joy, 1986]. Firstly, modelling the motion of the incident electron through the specimen, secondly calculating the distribution and number of the charge carriers so produced, and thirdly computing the efficiency with which these carriers are subsequently collected to contribute to the measured current.

3.4.1 Monte Carlo modelling

The single scattering model discussed in Chapter Two is not appropriate to apply for bulk specimen in SEM characterization, since even on a fast computer, the calculation is slow because a large number (typically 250 to 400) of interactions must be computed for each trajectory before the energy of electron has fallen to a low enough energy ($\leq$ 500 eV), for the residual range to be ignored [Joy, 1991]. Therefore, for bulk specimens in SEM, a less rigorous, but substantially faster, model is applied and it is so-called plural scattering model.

The basic assumptions of the plural scattering model are the same as those for the single scattering model discussed in Chapter Two, but the implementation is markedly different. The total electron range R_E for a given incident beam energy is taken to be the Bethe range R_B by using Simpson's rule to numerically evaluate the integral [Joy, 1991]:

$$R_B = \int_0^E \left(\frac{-1}{dE/dS}\right) dE \qquad (3.6)$$

where dE/dS is the stopping power given by eqn. 2.11. The Bethe range is then divided into fifty segments of equal length which is assumed to be the step of each scattering event. This number is a compromise between the requirements for highest accuracy and smallest computing time [Joy, 1986]. E_n, the electron energy at the start of nth step of the trajectory, is found by

$$E_n = E_{n-1} - \int_{step} \left(\frac{dE}{dS}\right) dS \qquad (3.7)$$

where dE/dS is again obtained by using eqn. 2.11. E_1 is set to be incident energy E_0, and E_{51} is set to be zero. The azimuthal scattering angle is given by eqn. 2.10 again. The electron scattering angle ϕ is found by [Curgenven & Duncumb, 1971; Love et al, 1977]

$$\tan\left(\frac{\phi}{2}\right) = \tan\left(\frac{\phi_0}{2}\right)\cdot\left(\frac{E_0}{E}\right)\cdot\left(\frac{1}{\sqrt{RND}} - 1\right) \qquad (3.8)$$

where

$$\tan\left(\frac{\phi_0}{2}\right) = \frac{0.0144\ Z}{2\ p\ E_0} \qquad (3.9)$$

where p is the impact parameter determined experimentally and ϕ_0 gives the minimum scattering angle for the incident energy E_0, and it could be obtained from the backscattering coefficient η [Joy, 1991]:

$$\tan\left(\frac{\phi_0}{2}\right) = 0.016697 + 0.55108\eta - 0.96777\eta^2 + 1.8846\eta^3 \qquad (3.10)$$

where the backscattering coefficient η is found for a material of atomic Z at incident beam energy E_0 as [Hunger & Küchler, 1979]:

$$\eta(Z, E) = E_0^m \cdot C \Rightarrow m = 0.1382 - \frac{0.9211}{\sqrt{Z}}$$

$$\Rightarrow C = 0.1904 - 0.2235\ (\ln Z) + 0.1292\ (\ln Z)^2 + 0.01491\ (\ln Z)^3 \qquad (3.11)$$

Despite the apparent simplicity of this plural scattering approach the agreement between predictions made using the plural scattering model and the single scattering model is generally excellent except for a few cases where the granularity (i.e., the size of the step length) of the plural scattering model is too high to permit a model to be realistic [Joy, 1988]. For the case of EBIC imaging the model has been found to be perfectly adequate, probably because the phenomena of interest rely on the integrated properties (i.e., size, shape, depth distribution) of the interaction volume, and these quantities are relatively insensitive to the details of the model chosen [Joy, 1986].

charge generation

The generation of the electron-hole pairs of charge carriers follows directly from the trajectory discussed above. Along the nth step the energy deposited is

$$\Delta E = E_n - E_{n-1}. \qquad (3.12)$$

Assuming all of this energy is ultimately used for carrier generation, then the total number of electron-hole pairs N produced along the step is

$$N = \Delta E/E_{eh}. \qquad (3.13)$$

where energy E_{eh} necessary to create an electron-hole pair is typically three times the bandgap energy for the semiconductor [Klein, 1968]; for silicon it is 3.6 eV. Because of the small length of each trajectory step (e.g., for 20 keV beam the electron range in silicon is about 5 μm), the electron-hole pairs generated along it can be treated as being generated by a point-like source placed at the step center with the source efficiency described by eqn. 3.13. The N carriers generated along each trajectory step are recorded at the exact location of the point-like source, which enable one to calculate the distribution of carrier pairs at the end of the program.

charge collection

As mentioned above, the Schottky barrier is a widely used tool for semiconductor studies. In this study, the Monte Carlo simulation is applied to EBIC imaging in the planar Schottky barrier geometry.

With the distribution of carrier pairs as a starting point, the next step is to calculate the measurable current gain (G) of the external current. The current gain G of the Schottky barrier is defined as a ratio of current collected to current injected:

$$G = I_{cc}/I_b \qquad (3.14)$$

where I_{cc} is the measured short-circuit current collected for a given incident beam current I_b. The in-built potential of the Schottky barrier depletes the material beneath it to a depth Z_D, the depletion depth. All carriers generated by the beam at depth $Z < Z_D$ will be collected with unity efficiency. Thus for a beam energy such that the electron range R_E is less than Z_D the gain G (ignoring the effect of backscattering) should equal the average number of electron-hole pairs produced per incident electron:

$$G = E_0/E_{eh} \qquad (3.15)$$

where E_0 is the incident beam energy as usual. However, the effect of backscattering is to reduce this gain, since part of the incident energy is effectively lost. The Monte Carlo simulation directly accounts for this loss in the calculation of the original distribution of generation of carrier pairs , since each trajectory is followed to completion, whether it is fully absorbed or backscatters. Carrier pairs produced at depths $Z > Z_D$ are in neutral material, and in the absence of a drift field the electrons and holes will recombine except for a fraction, η, which will diffused back to the depletion region and contribute to the

signal. This fraction η, for a point source at Z, is [Wittry & Kyser, 1964]

$$\eta = \exp\{-(Z - Z_D)/L\} \qquad (3.16)$$

where L is the minority carrier diffusion length which represents the distance that minority carriers (in the case of p-type semiconductor, electrons) can diffuse on the average before their number decrease by a factor of 1/e due to recombination.

For each step of the trajectory the number of carriers produced and collected can then be computed since the point of generation relative to the edge of the depletion volume and the diffusion length are both known. Since the simulation separates the carrier collection from the carrier generation calculation, each set of trajectories can be used to find gain values for many different values of Z_D and L. This both minimizes the computing time required and provides improved statistical accuracy between data points obtained with different parameters but the same trajectory set.

3.4.2 Applications

carrier generation profiles

Figure 3.19 shows an example of computation for an Au(165Å)/p-silicon Schottky barrier (discussed in section 3.3) at 15 keV. The computer screen display includes plots of incident electron trajectories with various color which corresponds to different energy loss, volume distribution of carrier pairs generated inside the specimen as a direct result of incident beam current injection, and the variation of gain with the depletion depth and diffusion length of

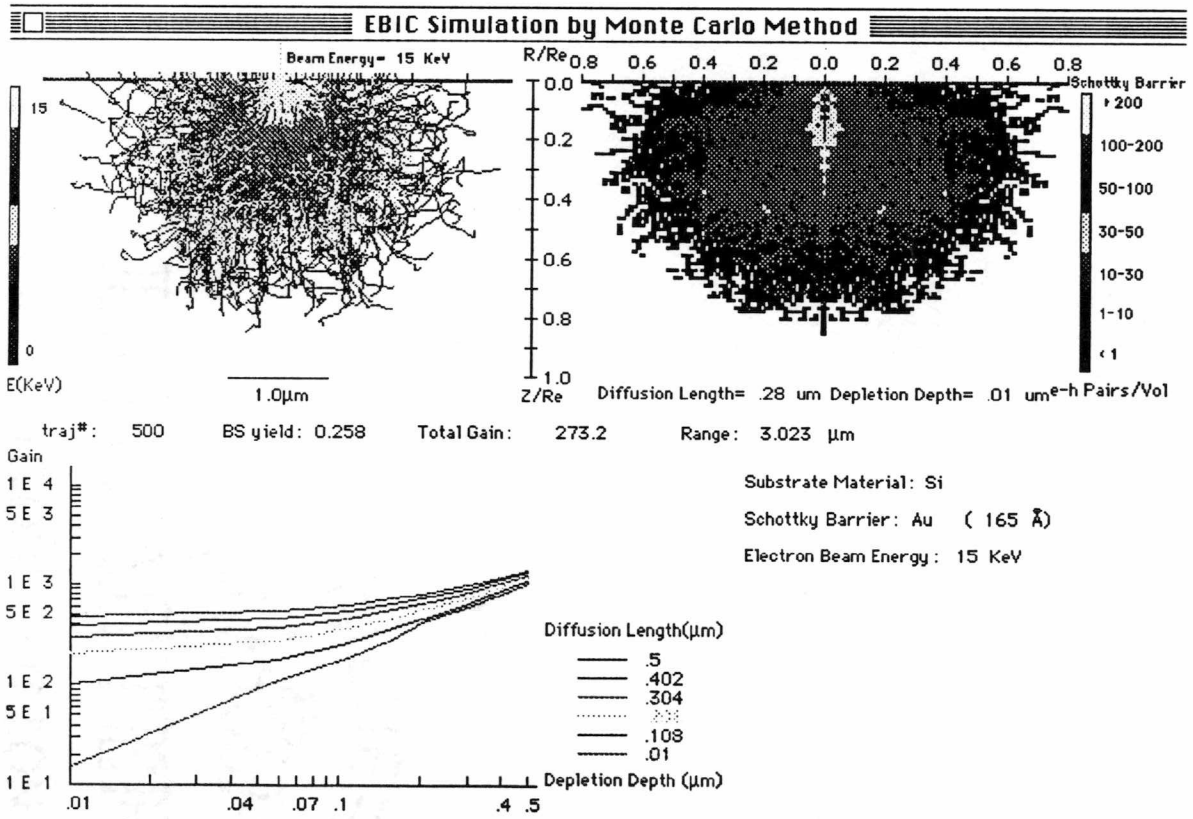


Figure 3.19. Monte Carlo simulation of EBIC signal generation (adopted directly from the computer screen). The up-left part of the screen is the plot of electron trajectories and the up-right part is the volume distribution of electron-hole pairs generation. The low-left part is the calculation of EBIC gain for various combinations of diffusion length and depletion depth.

the Schottky barrier. The color carrier generation profiles have the typical "tear-drop" shape, with a diameter of the order of the electron range, and with a maximum depth of penetration of about 0.8 of the range. The distribution of carrier generation is far from uniform and large percent (about 20%) of the carriers is generated in a volume that is only about 0.1 of the range in diameter. The wide variation in carrier generation within the beam interaction volume has important consequences for the form of image profiles, particularly the defect imaging profile which is determined by the form of the carrier generation function [Joy, 1986].

measurements of local minority diffusion length

Because the gain variation is both substantial in magnitude and monotonic in form, a comparison of experimental results at several incident beam energies with corresponding computed data provides an accurate and direct method for deducing L and Z_D in the Schottky geometry [Joy, 1986].

As an example, it has been discussed in section 3.3 the EBIC imaging of plasma sites at the Au(165Å)/p-Si Schottky barrier. Figure 3.13 shows the correlation between the EBIC gain and the net impurity density. Now one can use the Monte Carlo simulation to study the change of minority carrier diffusion length L at the plasma generation site. From the measurements of Schottky barrier height (Table 3.1) it was found that in the case of Au(165Å)/p-Si Schottky barrier (gold film thickness is 165 Å), the acceptor concentration is 3.41×10^{20} ($1/m^3$). Then from eqn. 3.4 the depletion depth for this barrier was calculated to be 2.2 μm. Using this data one should be

able to calculate the gain with varies diffusion length. But it was found that Z_D = 0.108 μm was the best fit to the experimental data. This low values of the depletion depth is possibly due to the large acceptor concentration at the plasma site. The deduced acceptor concentration from the new Z_D using eqn. 3.4 is about three orders higher than the measured bulk concentration which is conceivable due to micro-inhomogeneity. Comparison between the calculated gain and the experimentally measured one at different beam energy is shown in Figure 3.20. It is obvious that with different beam energy (and hence different beam penetration depth) the diffusion length changes substantially. The best fitted results are shown in Figure 3.21. The deduced diffusion lengths change from 0.01 μm to 0.5 μm with beam energy increasing from 10 keV to 16 keV. When one plots the depth distribution of net impurity density at the barrier and the diffusion length against the maximum electron penetration depth, it was found that changes of diffusion length is directly related to the impurity distribution (Figure 3.22). The higher impurity concentration leads to more efficient extraction of carriers from the plasma and greater gain [Leamy et al, 1978]. When more ionized dopant impurities (donor or acceptor) exist, the minority carriers, in this case, electrons, are more easily to be scattered and its mobility decreases. The charge carrier path will be deflected due to Coulomb force interaction [Sze, 1986] and hence the diffusion length increases. At room temperature, in the case of p-type silicon, when impurity concentration changes from 10^{20} to 10^{23} (m^{-3}), the mobility and diffusivity of holes (majority carriers) are decreased by about two

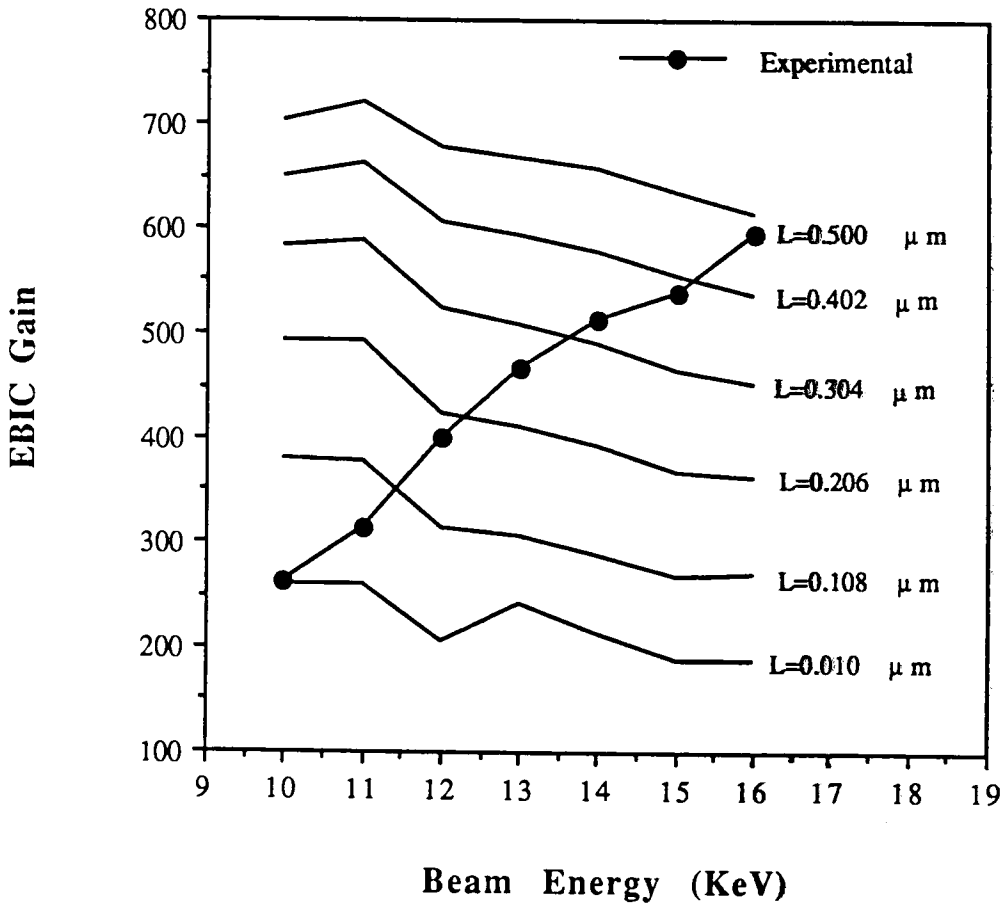


Figure 3.20. Experimentally measured and calculated EBIC gain for various beam energy and diffusion length.

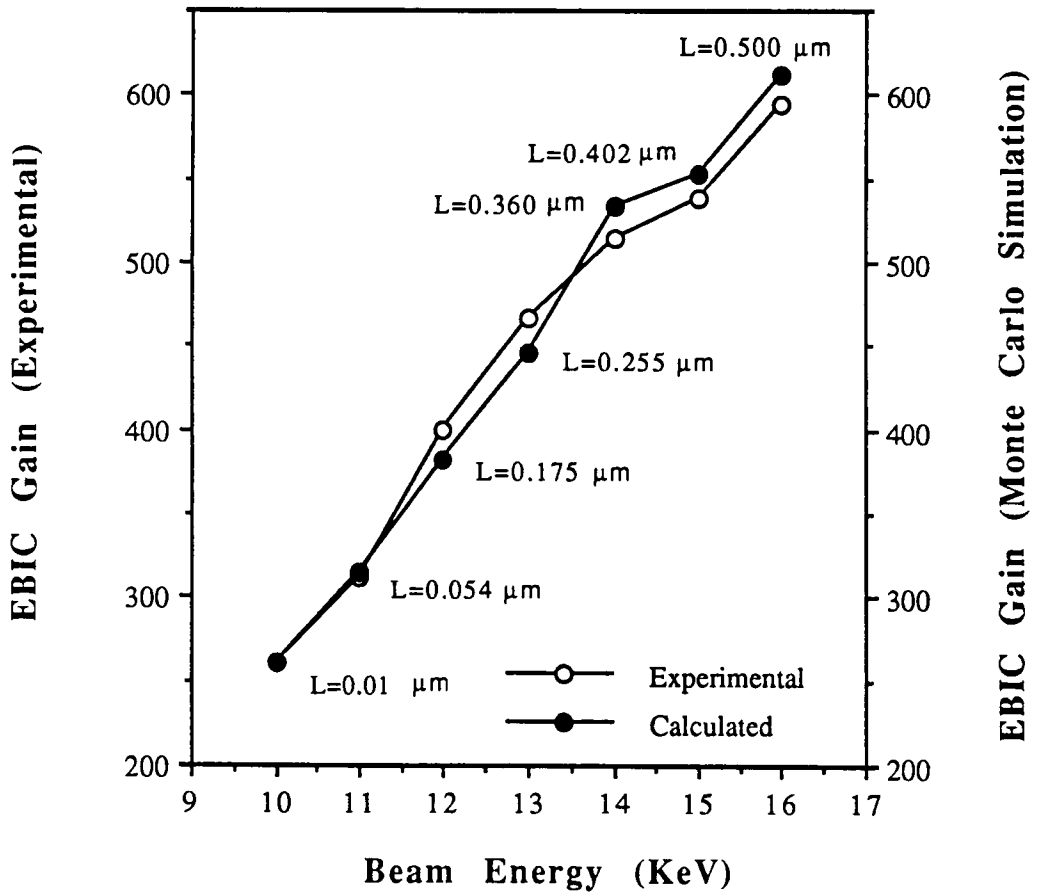


Figure 3.21. Comparison of experimentally measured and Monte Carlo simulation of the EBIC gain matched with different values diffusion length for a range of beam energy.

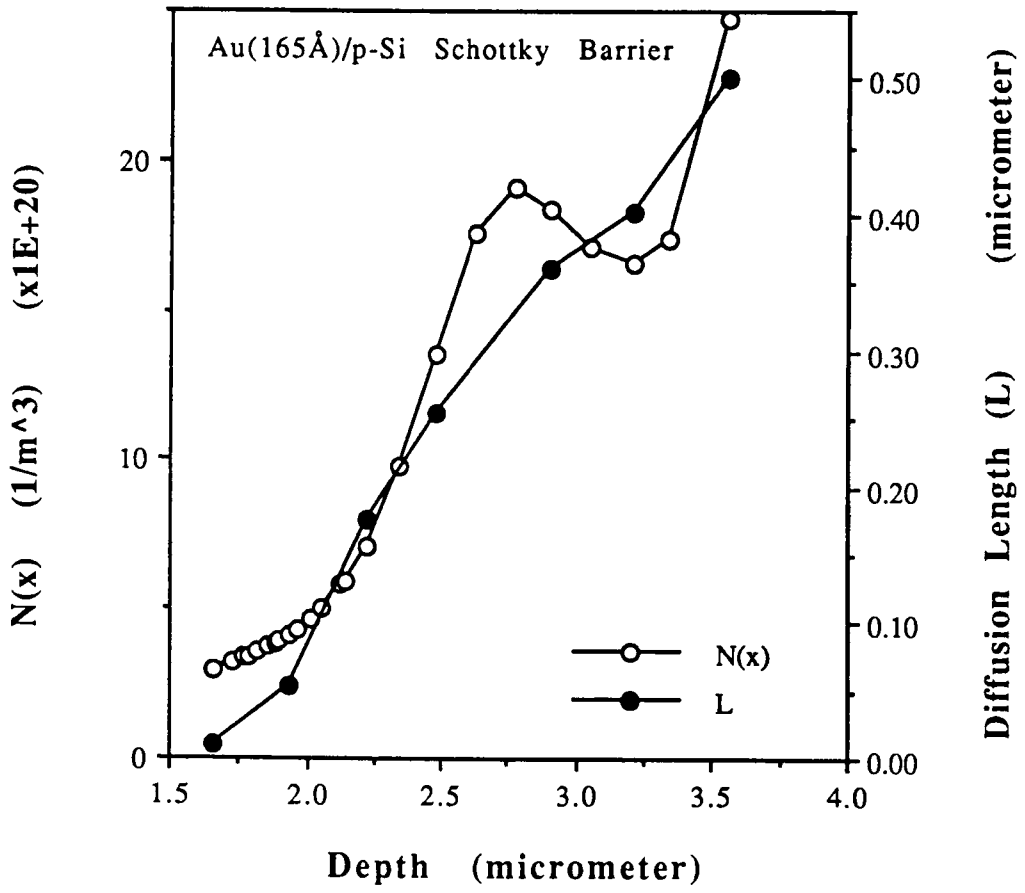


Figure 3.22. A plot of net impurity distribution vs depletion depth superposed with a plot of diffusion length matched with experimentally measured EBIC gain for the same range of the depth.

times [Sze, 1986], which again decreases the recombination rate of electron-hole pairs and thus increases the diffusion length of electrons.

As a conclusion, Monte Carlo simulation plays an important role in quantitative analysis of EBIC data. The future work will integrate the effects of electron beam-defects interaction into the Monte Carlo model.

References

Adegboyega, G.A., Phys. Stat. Sol., (a)**111**(1989), K31-K35

Blood, P. and Orton, J.W., Rep. Prog. Phys., **41**(1978), p. 11-257

Blumtritt, H., Gleichmann, R., and Sullivan, T., Poster presentation at 12th Tagung "Elektronenmikroskopie," Dresden (GDR), Jan., 1988

Charlson, E.J. and Lien, J.C., J. Appl. Phys. **46**(1975), p. 3982-3987

Curgenven, L, and Duncumb, P., 1971, Tube Investments Research Laboratories Report 303

Czyzewski, Z. and Joy, D.C., Scanning, **12**(1990), p. 5-12

Finetti, M., Sumi, I., Bartur, M., Banwell, T. and Nicolet, M.-A., Solid-St. Electron. **27**(1984), p. 617.

Goodman, A.M., J. Appl. Phys., **34**(1963), p. 329-338.

Goss, C.A., Charych, D.H. and Majda, M., private communication, 1989

Grusell, E., Berg, S. and Anderson, L.P., J. Electrochem. Soc. **127**(1980), p. 1573

Gutknecht, P. and Strutt, M.J.O., Electron Lett. **7**(1971), p. 298

Hanoka, J.I. and Bell, R.O., Ann. Rev. Mater. Sci., **11**(1981), p. 353-380

Holt, D.B., in SEM Microcharacterization of Semiconductors, (Academic Press Ltd., 1989), p. 241-338

Hunger, H.-J. and Küchler, L., Phys. Stat. Sol., (a), **56**(1979), K45-48

Joy, D.C., Scanning Electron. Microsc. (1974), p. 328-334

Joy, D.C., Conference of Microscopy of Semiconductor Materials, Oxford, March 21-23, 1983

Joy, D.C., J. Microscopy, **143**(1986), Pt3, p. 233-248

Joy, D.C., Inst. Phys. Conf. Ser. No. 93: Volume 1, 1988, p. 23-32

Joy, D.C., in SEM Microcharacterization of Semiconductors, (Academic Press Ltd., 1989), p. 63

Joy, D.C., Scanning Microscopy, **5**(1991)2, p. 329-337

Joy, D.C., Maher, D.M. and Farrow, R.C., Mat. Res. Soc. Symp. Proc., **69**(1986), p. 171-176

Kanaya, K. and Okayma, S., J. Phys. D: Appl. Phys., **5**(1972), p. 43-58.

Kittler, M., Solid State Phenomena, **6&7**(1989), p. 367-382

Klein, C.A., J. Appl. Phys., **39**(1968), p. 2029

Leamy, H.J., J. Appl. Phys., **53**(1982), R51-80

Leamy, H.J., Kimerling, L.C. and Ferris, S.D., Scanning Electron Microsc. 1976, **IV**, p. 530-538.

Leamy, H.J., kimerling, L.C., and Ferris, S.D., Scanning Electron Microscopy, 1978, I, p. 717-726

Love, G. Cox, M.G.C., and Scott, V.D. J. Phys. D:Appl. Phys., **10**(1977), p. 7-23

Maher, F.J. and Rossouw, C.J., J. Phys. E: Sci. Instrum. **16**(1983), p. 1238-1241

McGee, J.B., U.S. Patent , 1982, US 4,315,970

Mullins, F.H. and Brunnschweiler, A., Solid-St. Electron. **9**(1976), p. 47

Newbury, D.E., in SEM Microcharacterization of Semiconductors, (Academic Press Ltd., 1989), p. 57

Ponjee, J.J., Nelissen, J.W.A. and Verwijlen, C.J.A., Eur. Pat. Appl., 1984, EP 111,957

Rhoderick, E.H. and Williams, R.H., Metal-Semiconductor Contacts, 2nd ed., (Clarendon Press, Oxford, 1988)

Sharma, B.L., in Metal-Semiconductor Schottky Barrier Junctions and Their Applications, (ed. by Sharma, B.L.), (Plenum Press, New York, 1984)

Smith, B.L. and Rhoderic, E.H., Solid-St. Electron., **14**(1971), p. 71-75

Sze, S. M., Semiconductor Devices Physics and Technology, (John Wiley & Sons, New York, 1986), p. 32

Wittry, D.B. and Kyser, D.F., J. Appl. Phys., **35**(1964), p. 2439

CHAPTER FOUR

ELECTRON HOLOGRAPHY THEORY

4.1 Background-Why Electron Holography?

The ultimate goal of electron microscopy is to reveal atomic structure of the object in order to meet the challenges from the modern development of material science and engineering in atomic dimensions. When a plane electron wave U_0 illuminates a thin specimen inside a transmission electron microscope (see Figure 4.1), after the interaction of the beam electrons with the specimen the wave function (the object wave as it exits the object) takes the form:

$$U_o(\mathbf{r}) = A_o(\mathbf{r}) \exp[i\Phi_o(\mathbf{r})] \qquad (4.1)$$

where $\mathbf{r} = (x, y)$ is the position vector, $A_o(\mathbf{r})$ and $\Phi_o(\mathbf{r})$ are the amplitude and phase, respectively. This means that the amplitude $A_o(\mathbf{r})$ and the phase $\Phi_o(\mathbf{r})$ are locally modulated in some manner corresponding to the object structure. A knowledge of the two functions A_o and Φ_o is essential for a full description of the object. For the determination of the specimen structure, the object wave U_o is imaged into an image wave U_I in the image plane at a higher magnification. Unfortunately, the image wave $U_I(\mathbf{r}) = A_I(\mathbf{r}) \exp[i\Phi_I(\mathbf{r})]$ does not agree with the objective wave, mainly due to the aberration function $\chi(\mathbf{R})$ of the objective lens that acts like a phase plate in the back focal plane:

$$\chi(\mathbf{R}) = 2\pi k\,[0.25\ C_s\,(\mathbf{R}/k)^4 + 0.5\ D_z\,(\mathbf{R}/k)^2] \qquad (4.2)$$

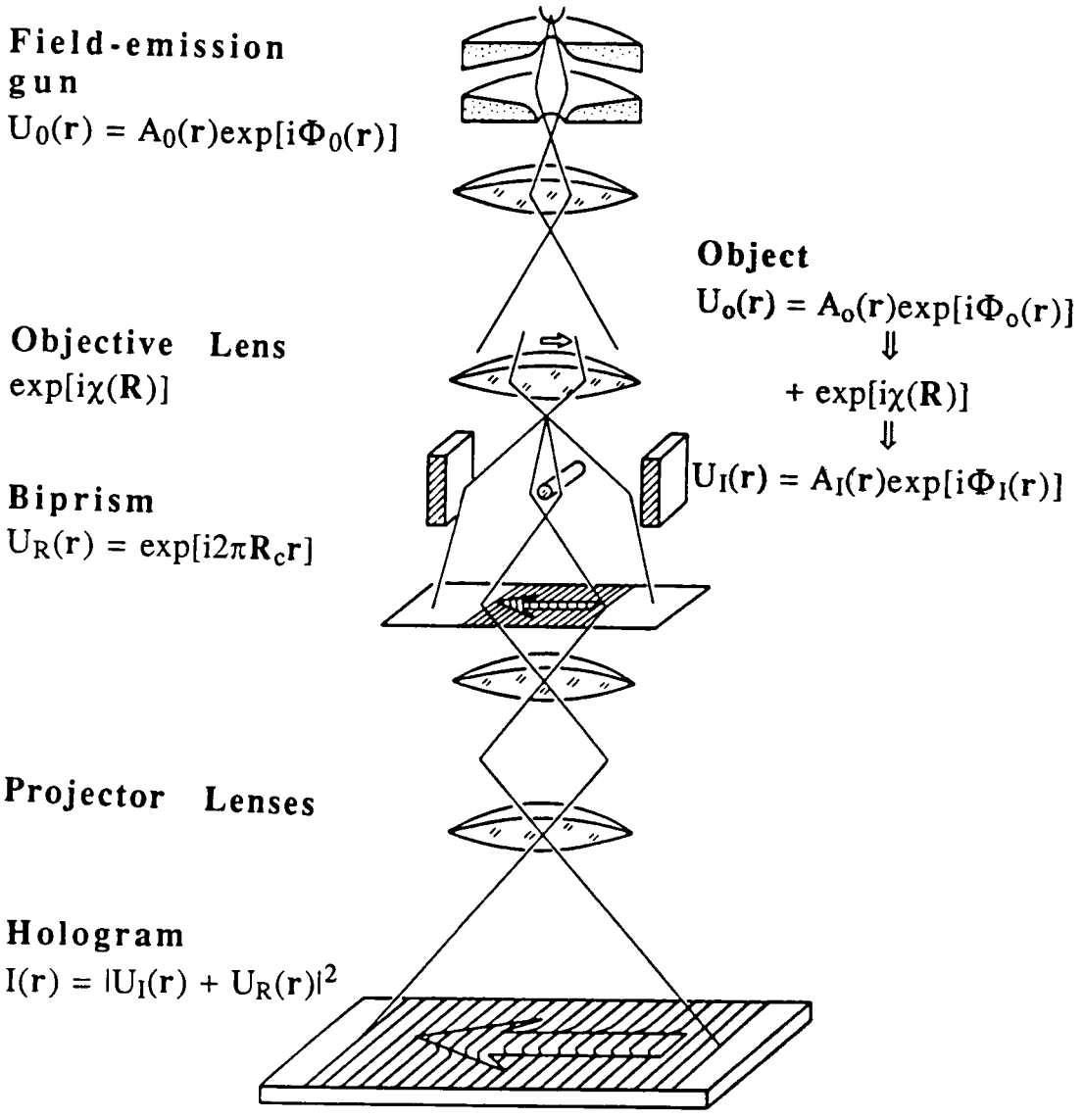


Figure 4.1. Schematic illustration of a field-emission TEM lens system, electron biprism, and formation of an electron hologram.

where k is the wave number, **R** the spatial frequency (**R** = 1/S, S, a periodic spacing in the specimen), C_s the coefficient of spherical aberration which has the dimension of length and D_z the defocus. In the back focal plane of the objective lens, the wave aberration $\chi(\mathbf{R})$ is convoluted with the object wave U_o as an exponential phase factor through Fourier transform:

$$FT[U_I(\mathbf{r})] = FT[U_o(\mathbf{r})]\ \exp[-i\chi(\mathbf{R})] \qquad (4.3)$$

and consequently, the image wave U_I becomes

$$U_I(\mathbf{r}) = FT^{-1}\{FT[U_o(\mathbf{r})]\ \exp[i\chi(\mathbf{R})]\} = A_I(\mathbf{r})\ \exp[i\Phi_I(\mathbf{r})] \qquad (4.4)$$

with amplitude $A_I(\mathbf{r})$ and phase $\Phi_I(\mathbf{r})$, which in general differs from the original object wave. This inverse Fourier transform may also include the influence of the finite illumination aperture and energy spread of the electron gun, which further degrades the image quality. The result may be expressed in terms of a contrast-transfer function (CTF) for the different spatial frequencies. All of these make visual interpretation of the image impossible. Only in the case of perfect imaging ($\chi(\mathbf{R}) = 0$) would the image wave be identical with the object wave. Finally, in ordinary electron microscopy only the intensity $\langle A^2 \rangle$ can be determined from a TEM micrograph, and the phase information $\Phi(\mathbf{r})$ is not retrievable.

As early as 1936, Scherzer proved theoretically that cylindrically symmetric electro-optical lenses can never be free of spherical aberrations since axially symmetric electron optical components always have a converging lens action [Scherzer, 1936].

Further in 1949 Scherzer found that through proper balancing of spherical aberration and defocus (so called "Scherzer focus," $D_z = -2.5(C_s\lambda/2\pi)^{1/2}$) the hitherto assumed limit could be improved by a factor of about three. This improvement results from the so-called phase contrast, which arises when the atoms illuminated with parallel electron waves become the origin of spherical waves [Scherzer, 1949]. Nevertheless, even then the image still suffers from the spherical aberration and one angstrom image is impossible. Structure-image resolution (or "Scherzer resolution") of TEM can be approximately defined as [Reimer, 1984]

$$\delta_{min} = 0.43\ (C_s\lambda^3)^{1/4} \qquad (4.5)$$

where δ_{min} is the half-width of the image-intensity distribution passes through a minimum at Scherzer focus. As a numerical example, for a Hitachi H800 TEM the theoretical Scherzer resolution is 2.26 Å. Further image resolution relies on the improvement of spherical aberration and other microscopic properties.

Although over the last several years intermediate voltage TEM has delivered interpretable images of crystal structures with a better than 2 Å resolution, the performance of the electron microscope is still unsatisfactory as compared with what it could be; further advancement can be expected. Many important problems still can not be studied because conventional TEM imaging techniques have specific important limitations:

(1) Further improvements in spatial resolution can be obtained by decreasing the electron wavelength λ using accelerating voltages

higher than 200 keV (see eqn. 4.5). For crystal structure image, it is an advantage that the Bragg angles decreases at the same rate as λ, since more reflections can contribute to the image with the correct phase unperturbed by spherical aberration [Reimer, 1984]. However, there are two problems that are of concern. First, a high-voltage electron microscope needs a larger pole-piece system and the values of C_s is correspondingly greater. Second, the high-energy beam will exceed the threshold energy for knock-on damage in many materials of interest [Reimer, 1984]. This is therefore not considered a practical means for resolution improvement.

(2) Further substantial improvements in resolution using current techniques would require a dramatic improvement in both the C_s and chromatic (C_c) aberration coefficients of the objective lens. The current common values of C_s and C_c are in the range of 0.7-1.4 mm. Since both C_s and C_c scale with the physical size of the lens, an improvement of the order that might be useful would require a lens design in which the lens gap into which the specimen is inserted to be so small that it would seriously restrict both specimen size and specimen tilt capabilities. Since the resolution is proportional to the fourth root of the spherical aberration (see eqn. 4.5) there then seemed little hope of obtaining further significant improvements in the resolution capabilities of electron microscopes [Hillier, 1947].

(3) As mentioned above, the ordinary TEM micrograph records only the intensity distribution $I(\mathbf{r}) = U_I(\mathbf{r})U_I(\mathbf{r})^* = A^2$, i.e., the modulus of the transmitted object wave convoluted with the wave aberration, the phase information is lost; here $U_I(\mathbf{r})^*$ is the complex conjugate of

$U_I(\mathbf{r})$. Therefore, phase and amplitude of the object wave are not accessible from conventional TEM imaging process. The so-called "phase object," which includes several major classes of materials and problems such as "amorphous," biological, and polymeric materials, can only be examined in current electron microscopes if the aberrations of the lenses are used to convert phase changes to corresponding amplitude variation.

For example, only in the case of "weak" pure phase or amplitude objects (a TEM specimen thinner than 100 Å and of low atomic number) is there a way to determine the object structure, and even then only approximately. Assume a weak phase object given by

$$U_o(\mathbf{r}) = \exp[i\Phi(\mathbf{r})] \approx 1 + i\Phi(\mathbf{r}),\ \Phi \ll 2\pi \qquad (4.6)$$

where quadratic and higher order terms in $\Phi(\mathbf{r})$ have been neglected. In the image plane one has the intensity distribution [Lichte, 1991a]

$$I(\mathbf{r}) = 1 - \Phi(\mathbf{r}) * G(\mathbf{r}) \qquad (4.7)$$

where the asterisk denotes the convolution of $\Phi(\mathbf{r})$ with the Fourier transform $G(\mathbf{r})$ of the phase CTF, $PCTF(\mathbf{R}) = 2\sin[\chi(\mathbf{R})]$. The values of $PCTF(\mathbf{R})$ give the transfer of a pure weak sinusoidal phase grating of spatial frequency $\mathbf{R}$ into the contrast of the final image. Only when the wave aberration $\chi(\mathbf{R})$ of the objective lens equals to zero is $I(\mathbf{r}) = 1 - 2\Phi(\mathbf{r})$, hence the phase shift $\Phi(\mathbf{r})$ due to the object is determined uniquely by the intensity in the image. From eqn. 4.5 one may find that $\chi(\mathbf{R}) = 0$ can be obtained with a combination of the C_s and an

appropriate defocusing D_z. Since this process relies on poorly defined and even casual properties of the lens and imaging conditions, the interpretation of such images is hazardous at best.

The another type of phase object includes phenomena such as those associated with magnetic or electric fields. In conventional TEM the magnetic field can only be imaged by Lorentz microscopy method which is basically a defocused image and is an indirect, qualitative, and low resolution image. The imaging of an electrostatic field was never realized. The field studies of materials like p-n junction of semiconductor and superconductor is almost impossible using conventional electron microscopy methods. The modern development of microtechnology requires new ways to study magnetic, ferroelectric, semiconductor, and superconductor materials.

One way to circumvent the improvement difficulties and obtain full information of the object wave is to use electron holography, and it is the topic of the next section.

4.2 What is Electron Holography ?

Electron holography, invented in 1948 by the German (naturalized British) physicist Denis Gabor, and which won him the 1971 Nobel Prize for Physics, is a procedure of "electron wavefront reconstruction" [Gabor, 1949, 1951]. It involves two-step imaging process in which both the amplitude and phase variations of an object wavefront are photographically recorded in an electron microscope as an interference pattern which is produced by the interference between a diffracted wave (Fresnel or Fraunhofer diffraction pattern, produced by elastic scattering) emanating from

the object and a homogeneous coherent background. This was to be followed by an optical imaging stage called *reconstruction* which consists of placing the photographic transparency film into a coherent beam of light and producing an image of the original object. The resulting record of the interference pattern obtained in the first step is called a *hologram* derived from the Greek and meaning total recording (*holos*, whole or total; *gramma*, record) because the photographic record contains information concerning both the amplitude and phase of a wave transmitted by or reflected from the object under examination.

Gabor's initial conception of holography was as a way to get around obstacles limiting the resolving power of electron microscopes (with optimum combination of diffraction and of spherical aberration in the electron microscope objectives, resolution was limited in the range of 5 to 10 Å at that time). Gabor hoped to use a deliberately aberrated optical lens to cancel the electron lens aberrations and thus produce an unaberrated, high-resolution image of the object. With the development of the laser, a high coherence light source, holography rapidly grew into a major branch of light optics.

Electron holography has attracted renewed interest in the electron microscopy world with the development of field-emission guns (FEGs) of high brightness (1000 times higher than conventional tungsten thermal source TEM) and coherence. Though better resolution than in the conventional bright field mode is already available because of the reduced damping of the contrast transfer

function, electron holography is becoming of increasing interest for quantitative studies of phase image of the object. The advantage of electron holography stems from the complete recording of amplitude and phase of the object wave, thereby avoiding loss of information. Consequently, without being restricted to pure phase or weak phase, it is possible to determine phase Φ and amplitude A of the object wave individually.

Although high resolution electron holography is an intrinsic part of electron holography, this study will concentrate on the development of new electron holography techniques for phase imaging which is a newly added and fast growing area in electron holography.

4.3 The Principle of Wave-Front Reconstruction

The essence of holography is the formation of an interference pattern, the hologram, between a reference wave and a wave modulated by the specimen. Therefore, holography is basically a lensless imaging method which relies only upon the interference phenomena of waves. Thus, holography is applicable to all kinds of waves--whether light (laser), X-ray, sound, electron, photoelectron, or neutron waves--irrespective of the existence of a lens for each wave. The detection of any object depends on the degree to which it scatters the illuminating irradiation.

4.3.1 Gabor's in-line holographic scheme

In order to fully appreciate electron holography, it is necessary to review the original scheme of Gabor holography. In his so-called

"projection method" the microscope object is illuminated by a beam of electrons which issue from a small aperture, either directly or through a lens system without any further lenses between the object and the photographic plate, where it forms a hologram (see Figure 4.2). This set-up was also called "lensless microscope" which has been used later in X-ray holography. If the hologram, after suitable processed, is illuminated by a replica of the original wave (which does not have to be the same electron wave), the original object is reconstructed in space, together with a "conjugate object" or "twin image," which is in general diffuse and distorted if a suitable type of object is examined.

Mathematically, in an electron microscope, if an aberrated complex imaging wavefront is $U_I = A_I \exp(i\Phi_I)$ at the observation plane beyond the specimen, the visually observed or photographically recorded quantity in that plane is the intensity $I = |U_I^2|$. This square law detection has eliminated the phase information which was present in the wavefront. The addition of a background wave U_B (or called a reference wave) onto the object-dependent wave yields an hologram with the intensity at the observation plane given by

$$I = |U_I + U_B|^2 = |U_I|^2 + |U_B|^2 + U_I^* \cdot U_B + U_I \cdot U_B^* \qquad (4.8)$$

where U_I^* and U_B^* denote the conjugate parts of U_I and U_B, respectively. The first two terms on the right hand side depend only on the intensities associated with the object and background waves while the last two terms depend on both their amplitudes and

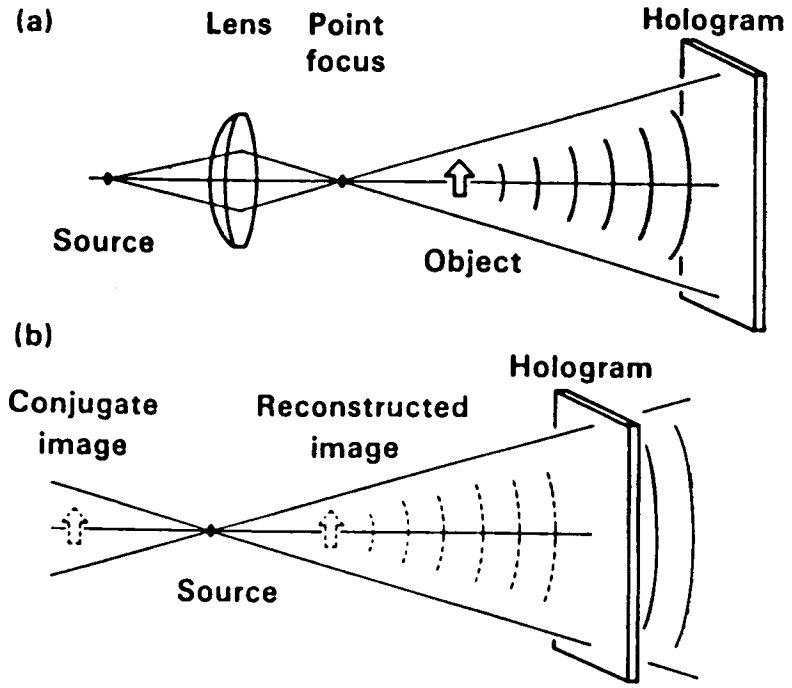


Figure 4.2. Gabor's in-line hologram scheme (projection method): (a) hologram formation, and (b) hologram reconstruction (adopted from [Tonomura, 1986]).

relative phases.

Consider a monochromatic, coherent electron wave $U = A\exp(i\Phi)$ illuminating a photographic film. Let us assume that the photographic film (the hologram) is developed by reversal, or printed with some overall Γ, where Γ (gamma) is the slope of the Hurter-Driffield curve (optical density vs exposure). On the "straight branch" of the Hurter-Driffield curve its amplitude transmission t will be $t = KA^{\Gamma} = K(UU^*)^{\Gamma/2}$, where K is a constant. Imagine now that U is split into a background wave U_B and an aberrated object wave U_I:

$$\begin{aligned} U &= U_B + U_I = A_B \cdot \exp(i\Phi_B) + A_I \cdot \exp(i\Phi_I) \\ &= \exp(i\Phi_B)\{A_B + A_I \cdot \exp[i(\Phi_I - \Phi_B)]\} \end{aligned} \qquad (4.9)$$

In the reconstruction stage, if one illuminates the hologram with the background wave U_B alone, one obtains immediately behind the emulsion a "substitutional wave" $U_s = tU_B$. Then

$$U_s = A_B \exp(i\Phi_B)\{K[A_B^2 + A_I^2 + 2A_BA_I\cos(\Phi_I - \Phi_B)]^{\Gamma/2}\} \qquad (4.10)$$

The simplest and also the most advantageous case is $\Gamma = 2$. Assuming this value,

$$U_s = KA_B^2\exp(i\Phi_B)\{A_B + A_I^2/A_B + A_I\exp[i(\Phi_I - \Phi_B)] + A_I\exp[-i(\Phi_I - \Phi_B)]\} \qquad (4.11)$$

For simplicity assume that the background wave is uniform in intensity, i.e., A_B is a constant, and compare eqns. 4.9 and 4.11, i.e., assuming background is uniform. The factors before the brackets are identical, apart from a constant factor. Inside the bracket, one will

see that $A_B[1 + A_I^2/A_B]$ has been substituted for the background amplitude A_B. If one assumes - and this is very essential condition - that the reference wave is strong relative to the amplitudes scattered by the object, i.e., a weakly scattered object, the change is insignificant. The next term $A_I \exp[i(\Phi_I - \Phi_B)]$ is identical in the two equations, and the term has almost the same phase with initial object wave except a constant shift Φ_B. This means that the original object wave has been reconstructed, with the correct phase, and, neglecting the small term A_I^2/A_B, with the relative amplitude. This term is the so-called "real" image. But the last term $A_I \exp[-i(\Phi_I - \Phi_B)]$ represents a wave with the same amplitude as the object wave U_I, but with a phase of opposite sign relative to the reference wave. This term is the so-called "conjugate" image.

The following section uses diffraction theory to describe the holography process. With the conditions of the Fresnel-Kirchhoff diffraction formula, and using the coordination notations suggested by Gabor [Gabor, 1949] (see Figure 4.3), when a homocentric wave illuminates an object, the wave U_Q at an observation point Q far away from the object P(x,y), and with angles α and β with respective to the source is:

$$U(\alpha,\beta) = \iint t(x,y)\cdot\exp\{ik[(x^2 + y^2]/2z_0 - (x\cdot\cos\alpha + y\cdot\cos\beta)]\}\ dxdy \qquad (4.12)$$

where t(x,y) is the object function at P(x,y). Introduce "Fourier variables" ξ, η instead of the direction cosines by the substitution $\xi = 1/\lambda \cos\alpha$ and $\eta = 1/\lambda \cos\beta$. If the illuminating cone is narrow enough,

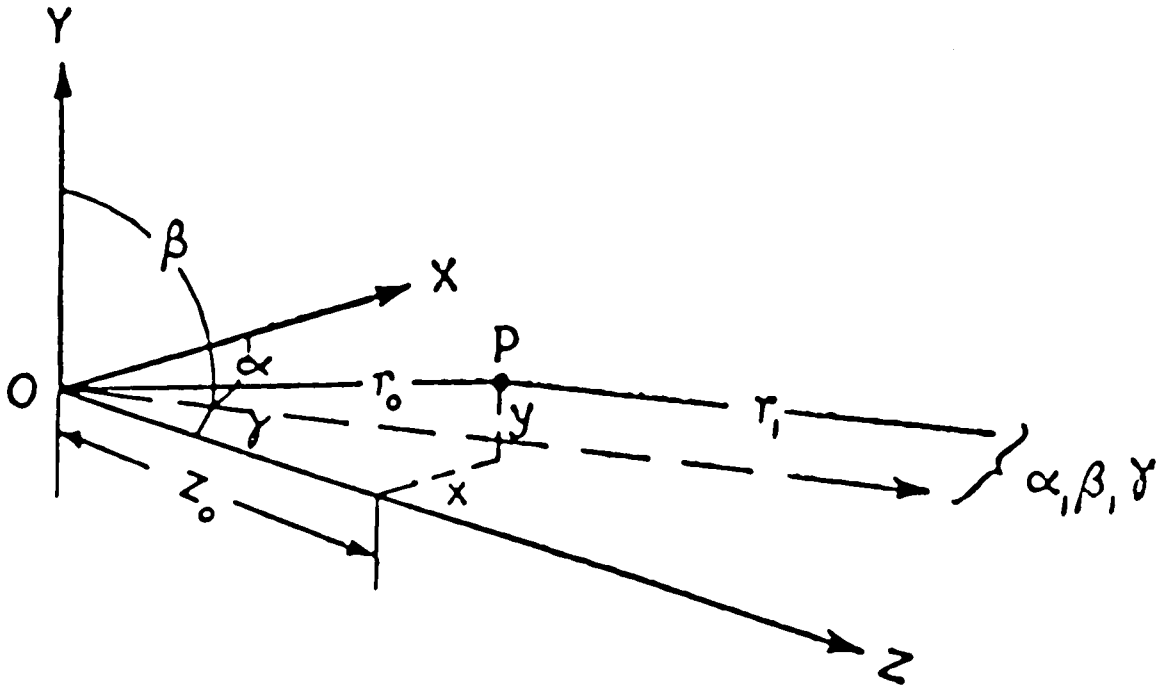


Figure 4.3. Schematic diagram of the coordination system used by Gabor (adopted from [Gabor, 1949]).

ξ and η can be taken to represent the coordinates in the "physical shadow" of a plane object, which includes the diffraction effects, and is to be distinguished from the "geometrical shadow" into which it merges at vanishing wave length. The geometrical shadow of a point P(x,y) has the Fourier coordinates $\xi = x/\lambda z_0$, $\eta = y/\lambda z_0$, where z_0 is the distance of the point P in the object plane. The quantity

$$\mu = \lambda z_0 \tag{4.13}$$

is the only parameter of the diffraction problem. $\sqrt{\mu}$ can be considered as the characteristic length. Details on the object larger than $\sqrt{\mu}$ will be shown to have shadows more or less similar to themselves, fine details lose all likeness by diffraction. Using the above notations and the abbreviation $x^2 + y^2 = r^2$, the eqn. 4.12 can be written as:

$$U(\xi,\eta) = \iint [t(x,y)\exp(\pi i r^2/\mu)] \exp[-2\pi i(x\xi + y\eta)]\, dxdy \tag{4.14}$$

So from the Fourier transformation,

$$t(x,y) = \exp(-\pi i r^2/\mu) \iint U(\xi,\eta) \exp[+2\pi i(x\xi + y\eta)]\, d\xi d\eta \tag{4.15}$$

Now by imagining the amplitude $U(\xi, \eta)$ as produced by the passage of the illumination wave U_0 through a "shadow object" in the plane ξ, η with a transmission $\tau(\xi,\eta)$, i.e., one puts $U(\xi,\eta) = U_0(\xi,\eta)\tau(\xi,\eta)$. Then the background U_B can be obtained directly from eqn. 4.14 by putting $t = 1$ and with the abbreviation $\xi^2 + \eta^2 = \rho^2$:

$$U_B(\xi,\eta) = i\mu \cdot \exp(-i\pi\mu\rho^2) \tag{4.16}$$

Finally one obtains the symmetrical transformation formulas:

$$\tau(\xi,\eta) = 1/i\mu\exp(\pi i\mu\rho^2) \iint t(x,y)\ \exp(-\pi r^2/i\mu)\ \exp[-2\pi i(x\xi + y\eta)]\ dxdy \tag{4.17}$$

and

$$t(x,y) = i\mu\exp(\pi r^2/i\mu) \iint \tau(\xi,\eta)\ \exp(-\pi i\mu\rho^2)\ \exp[+2\pi i(x\xi + y\eta)]\ d\xi d\eta \tag{4.18}$$

These were called by Gabor the "shadow transformation," and $t(x,y)$, $\tau(\xi,\eta)$ a pair of "shadow transforms." Eqns. 4.17-18 in succession restore the original. Physically this means that if instead of a photograph one could produce a "shadow object" with the absorbing and refractive properties of $\tau(\xi,\eta)$, and illuminate this with the background, one should exactly restore the object $t(x,y)$ in its original position. As a photograph cannot imitate the imaginary part of τ, a certain residual wave arise, i.e., twin images, exist.

Thus it can be concluded that an object-dependent imaging wave can indeed be reconstructed from the hologram. The practical viability of the electron holographic method will mainly depend on whether the real imaging wave can be efficiently separated from its background and conjugate waves.

Although the feasibility of holographic image formation was optically investigated by Gabor himself [Gabor, 1949; 1951], electron holography experiments were initially carried out by Haine, Dyson and Mulvey in England over a period of three years or so beginning in 1950 [Haine & Dyson, 1950; Haine & Mulvey, 1952] to study the

"projection" and the "transmission" holographic methods which were proposed by Gabor and by Haine and Dyson.

Eqn. 4.11 represents the scheme originally proposed by Gabor [1949] which become known as "projection method" or in-line holography (see Figure 4.2) because the background and scattered (object) waves propagate along the same axis. The projection method involved the formation of a point focus using an electron beam. The diameter of the focus determines the resolution of the reconstructed image. The spherical wave from the point focus is disturbed by an object located near the focus. Most of the wave pass through it undisturbed. Both the disturbed (object wave) and the undisturbed (reference wave) part are projected onto the hologram plane, and recorded on film as a hologram.

The projection method, however, was not practical for electron microscopy then since it was not easy to form an electron beam with a focal size of a few Å diameter. Hence the "transmission method" was proposed [Haine & Dyson, 1950]. As can be seen in Figure 4.4, an object is illuminated with a collimated electron beam. Since the effect of an object is assumed as small as with the projection method, waves tend to pass the object virtually without being affected in any way. Object and background waves overlap to form an interference pattern. The pattern is enlarged by electron lenses, and recorded on film as hologram. So this is a method of taking an out-of-focus electron microscopic image (a Fresnel hologram) using collimated illumination. Experiments regarding this method were conducted by Haine and Mulvey [1952] and Hibi [1956].

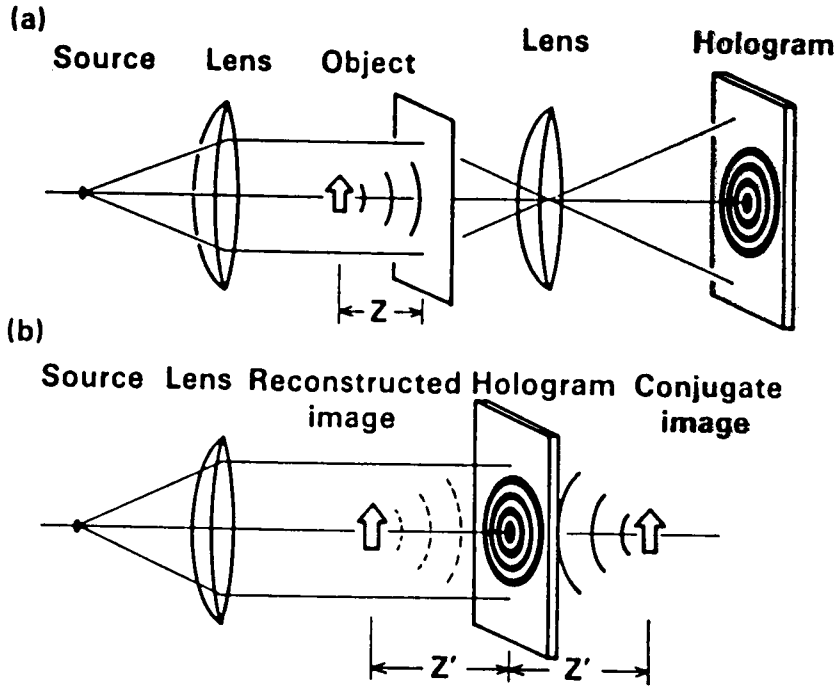


Figure 4.4. In-line scheme with transmission method: (a) hologram formation, and (b) hologram reconstruction (adopted from [Tonomura, 1986]).

The main disadvantages of the In-line holography method are: (1) it is restricted to weakly scattering objects or small objects (since the background wave has to function as a reference wave); and (2) the reconstructed image is always degraded by the presence of the aberrated twin image which placed a further restrictions on the object size. Although many methods has been tried to eliminate the effect of the conjugate image, the most effective way found so far is Fraunhofer holography invented by Thompson [1965]. With this method, a hologram is formed in the Fraunhofer or far-field diffraction region of an object (the hologram formed far from the image plane) given by

$$a_{max}^2 \ll \lambda D_z \qquad (4.19)$$

where a is the size of the object, D_z the defocus of the hologram plane relative to the Gaussian image plane, referred back to the object. Under this condition, twin reconstructed images have no influence on each other since the twin image is spread over a very much greater area than the main image and hence has very much lower contrast [Thompson, 1965; Develis et al, 1966]. If one of the images is observed, the other image is completely blurred in this plane. Eqn. 4.19 can be satisfied for the individual particles or objects while the overall object surface is still in the Fresnel or near-field diffraction region. The electron version of this experiment was conducted by Tonomura et al in 1968 with the resolution of a few tens of Å at best [Tonomura et al, 1968]. The resolution in this case is limited by the spatial coherence length $\lambda/2\alpha$ of the illuminating electron beam (2α

is the illumination angle of the beam) [Tonomura, 1986]. If the total angular spread of the beam incident on the specimen is 2α, the smallest detail that can be resolved S_{rmax} will be

$$S_{rmax} \approx 3.2\alpha D_z \qquad (4.20)$$

provided that D_z is not too small. Then, a specimen structure large than S_{rmax} and situated in a large structure-free area can be reconstructed without any disturbance from the twin image. For a given field of view s_m, if a resolution of S_{rmax} is desired, one must satisfy the condition

$$\alpha < C\, S_{rmax}\lambda/(2s_m^2) \qquad (4.21)$$

where C is a constant between 1 and 1.5 [Hawkes, 1978]. Based on eqns. 4.20-21, for a resolution of 10 Å in an object field of 1000 Å across at 200 keV, one finds $\alpha < 0.2$ micro-rads, showing the limitation of in-line holography. Only a small specimen area can be imaged and an extremely low aperture α is necessary, which can be obtained only with a field-emission gun. The electron source coherence is increased at the expense of beam current; in consequence, the exposure time needed to record the hologram is long and the mechanical and electrical stabilities of the electron microscope needs to be very good. It is clear that the advent of reliable field-emission sources of very high brightness is indispensable to the development of in-line electron holography.

A unique advantage of Fraunhofer electron holography is the ideal nature of the transfer function [Bonnet et al, 1978]. This is due

to the fact that the microscope objective lens is strongly defocused (far-field region) so that the two diffracted beams $\pm\theta$ do not interfere. The ideal transfer function after reconstruction is then unity and the twin image do not overlap on the reconstructed one. This actually provides a solution to solve the problems in high resolution electron microscopy: the existence of gaps in the transfer function and the difficulty to find complex amplitude of the diffracted image wave from registered intensity. The contamination and beam damage problem encountered in biological specimen may be solved using Fraunhofer holography in medium resolution range [Bonnet et al, 1978; Zhang & Joy, 1991].

4.3.2 Off-axis electron holographic scheme

Fraunhofer holography seeks to put one of the twin images into infinity. Another way to separate the twin images is to put one of the twin images out of sight, which is the subject of this section: the off-axis scheme.

Since 1962, the twin effect has been dealt with in laser holography by the use of an off-axis mode that eliminates the overlap between images [Leith & Upatnieks, 1962]. Now let us assume a tilted reference wave $U_R = \exp[i2\pi \mathbf{R}_c\mathbf{r})]$ where $\mathbf{R}_c$ is spatial carrier frequency, using eqn. 4.8, then the intensity distribution at the hologram plane is given by

$$I = |U_I + U_R|^2 = 1 + A_I^2(\mathbf{r}) + 2A_I(\mathbf{r})\cos[2\pi\mathbf{R}_c\mathbf{r} - \Phi_I(\mathbf{r})]$$
$$= (1 + A_I^2) + A_I \exp[i\Phi_I(\mathbf{r}) - 2\pi\mathbf{R}_c\mathbf{r}] + A_I \exp[-i\Phi_I(\mathbf{r}) + 2\pi\mathbf{R}_c\mathbf{r}] \quad (4.22)$$

where $|\mathbf{R}_c| = \alpha/\lambda$, α is the tilting angle from the imaging wave U_I. The fringe pattern consists of a cosinusoidal fringe system that is modulated in contrast by the amplitude modulation $A_I(\mathbf{r})$ of the specimen wave and in position by the local phase shift $\Phi(\mathbf{r})$. The first term of eqn. 4.22 represents a plane wave travelling down the incident direction attenuated by the intensity distribution of the object wave. The second term generates a virtual image of the object at a suitable distance, while the exponential factor $\exp(-i2\pi \mathbf{R}_c\mathbf{r})$ indicates that this virtual image is deflected off the incident direction at the angle α. Similarly, the last term is proportional to the conjugate wave which generates the real image deflected at the angle $-\alpha$ from the incident direction. Therefore, this mode of holography (Figure 4.5) separately reconstructs twin images into two diffracted beams propagating in different directions with the angle of 2α. The image wave U_I can then be reconstructed from the hologram in both amplitude and phase in the following way. Consider the Fourier transform of the intensity distribution I of the hologram (see eqn. 4.20) [Lichte, 1991a]

$$\begin{aligned} FT[I(\mathbf{r})] = {} & \delta(\mathbf{R}) && \text{"zero beam"} \\ & + FT[A_I^2(\mathbf{r})] && \text{autocorrelation} \\ & + FT\{A_I(\mathbf{r})\exp[i\Phi_I(\mathbf{r})]\} * \delta(\mathbf{R} + \mathbf{R}_c) && \text{sideband (+1)} \\ & + FT\{A_I(\mathbf{r})\exp[-i\Phi_I(\mathbf{r})]\} * \delta(\mathbf{R} - \mathbf{R}_c) && \text{sideband (−1)} \end{aligned} \qquad (4.23)$$

In addition to the zero beam and the autocorrelation function (power spectrum of the amplitude), one finds two sidebands in diffraction

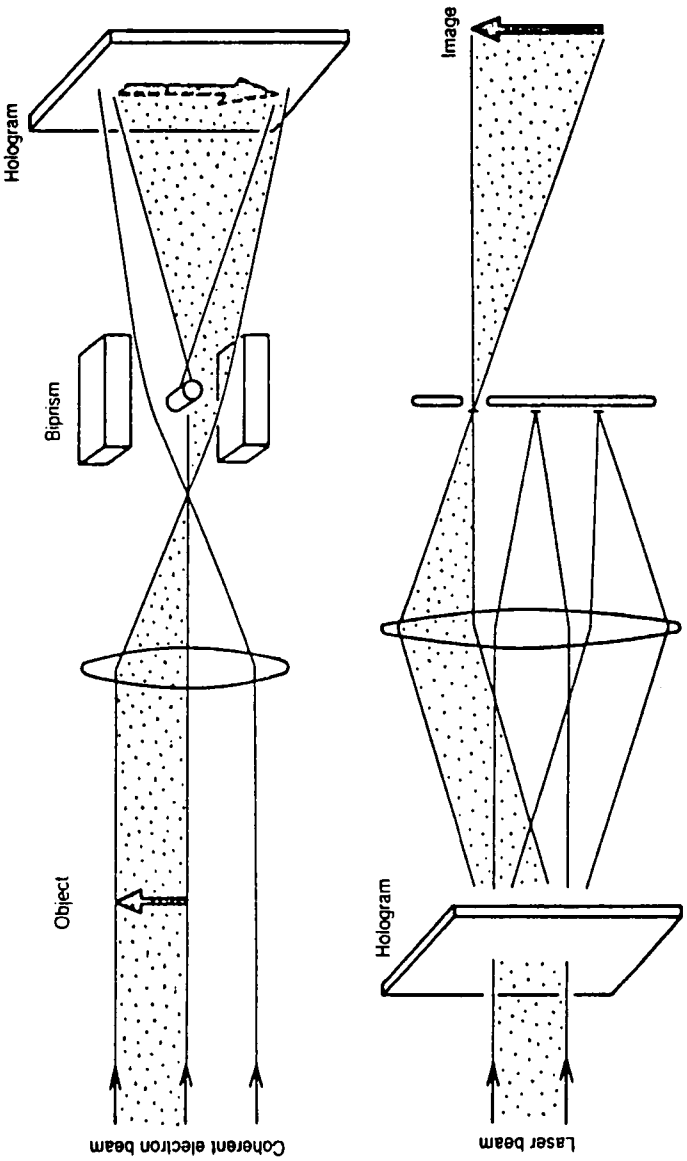


Figure 4.5. Off-axis electron holographic scheme: (a) hologram formation, and (b) hologram reconstruction (adopted from [Tonomura, 1990]).

plane (Fourier space) representing the Fourier spectrum of the image wave and its conjugate, respectively. If the carrier frequency $\mathbf{R}_c$ is chosen sufficiently high, the autocorrelation does not overlap with the sidebands. Then one sideband can be completely isolated from the remainder. Centering this isolated spectrum and performing the inverse Fourier transform yields the appropriate image wave U_I. Remember from eqns. 4.3-4, another form of Fourier transform of the hologram is given by

$$\begin{aligned} FT[I(\mathbf{r})] = {} & \delta(\mathbf{R}) && \text{"zero beam"} \\ & + FT[A_I^2(\mathbf{r})] && \text{autocorrelation} \\ & + FT[U_o(\mathbf{r})]\ \exp[i\chi(\mathbf{R})] * \delta(\mathbf{R} + \mathbf{R}_c) && \text{sideband (+1)} \\ & + FT[U_o(\mathbf{r})]^*\ \exp[-i\chi(\mathbf{R})] * \delta(\mathbf{R} - \mathbf{R}_c) && \text{sideband (-1)} \end{aligned} \qquad (4.24)$$

which gives a direct picture of the aberration effects on the object wave. When the carrier frequency $\mathbf{R}_c$ is three times higher than the spatial frequency constructing of the object wave, the third and forth terms are separated completely from the other two terms.

Selecting the third or forth term of eqn. 4.24 with an aperture and shifting the coordinate by $\pm\mathbf{R}_c$, respectively, one obtains a wave function on the image plane after inverse Fourier transformation as $U_I(\mathbf{r}) = A_I(\mathbf{r})\ \exp[i\Phi_I(\mathbf{r})]$. Thus the electron microscope image, which includes effects of aberrations, is reconstructed. If all parameters of aberrations are known, then the aberration function can be calculated by eqn. 4.2. Then aberration of lens can be corrected during the inverse Fourier transformation by multiplying an

exponential term $\exp[+i\chi(\mathbf{R})]$ to the selected sideband. The amplitude and the phase distributions can be calculated separately, and those data are stored as two terms, such as real part ($A_o\cos\Phi_o$) and imaginary part ($A_o\sin\Phi_o$) in numerical reconstruction [Harada et al, 1990]. The reconstruction also can be performed by optical bench using laser light or by numerical calculation using digitized hologram.

Although the development of off-axis holography by Leith and Upatineiks [Leith & Upatineiks, 1962] was epochal, the possibility was already mentioned in Gabor's original paper [Gabor, 1949], but was not fully detailed. The electron version of the off-axis scheme was finally realized when Möllenstedt and Düker invented an electron biprism [Möllenstedt & Düker, 1954]. Off-axis image-plane electron holography experiments were subsequently carried out by Möllenstedt and Wahl [Möllenstedt & Wahl, 1868] who used the electron biprism, as well as Tonomura who used a single-crystalline film as a beam splitter [Tonomura, 1969]. Since the invention of the electron biprism, the off-axis image plane electron holography has dominated the direction of the development of electron holography techniques, and the electron biprism has become an indispensable mean to perform electron holography experiments. The method was successfully applied in high resolution imaging for the first time by Lichte [Lichte, 1987].

4.4 What Can Electron Holography Provide Beyond Other Techniques ?

(1) The original objective of holography was to break through the resolution limit of electron microscopes by compensating for the

aberrations of the objective lens in the optical or numerical reconstruction stage of electron holography. The phase and amplitude components of the hologram can be individually compensated for the errors of the lens to form a final hologram from which a corrected image having higher spatial resolution than that attainable without the use of holographic techniques can be reconstructed. A conventional high resolution micrograph cannot be "deconvoluted" to remove the effects of the instrument response function because the phase and amplitude information are convoluted together in a nonlinear way. In the case of the hologram, however, the necessary corrections are mathematically straightforward, and reversible. From eqns. 4.3-4 and 4.22, one can perform a "spatial filtering" process on a re-centered isolated sideband, i.e., multiplying the sideband by $\exp(-i\chi(\mathbf{R}))$ and then through inverse Fourier transform to get back aberration-free objective wave $U_o(\mathbf{r})$ [Franke et al, 1988].

Preliminary work by Lichte and co-workers [1990] (Figure 4.6) suggests that spatial resolution below 1 Å may already be possible from images reconstructed from holograms generated on currently available instruments capable of 2 Å resolution using conventional imaging techniques. A reconstructed image may be different from the electron microscopic image in some cases. Since the object information in a hologram is recorded in a form of interference fringes, only coherently scattered electrons can contribute to the reconstructed image, which can be regarded as a "zero-loss" image not blurred by ineleastically scattered electrons

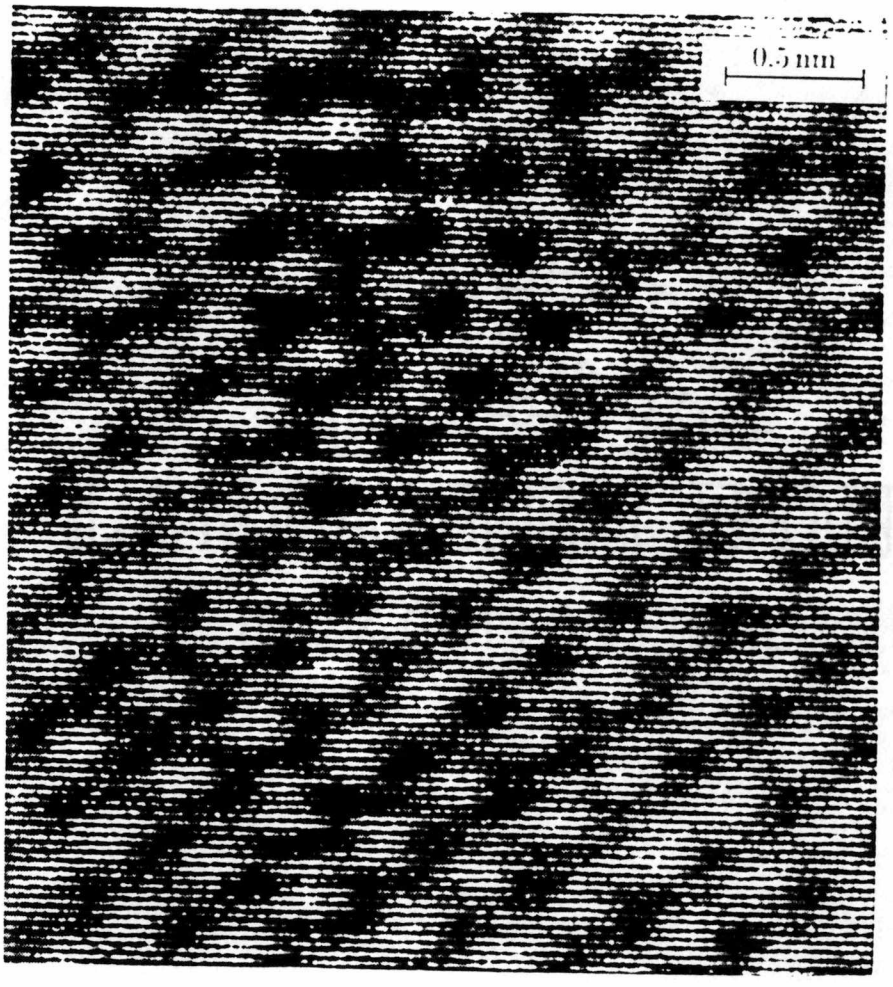


Figure 4.6. Hologram of CeO_2 with the interference fringe spacing of 0.32 Å, which makes a point resolution below 1 Å achievable (adopted from [Völkl & Lichte, 1990]).

since inelastic scattered electrons only add a amplitude term on the left hand side of eqn. 4.22. Therefore, an image reconstructed from an off-axis electron hologram is an "energy-filtered" image which can sometimes looks very different from a "normal" TEM image.

(2) Because both pure phase and amplitude images can be reconstructed, new forms of imaging can be envisaged via so-called interference electron microscopy which is mainly a phase shift visualization method. Even Gabor did not anticipate in this new field of application electron holography. A phase shift of the electron beam due to the object can be described as [Yanzi, 1984]

$$\Phi(\mathbf{r}) = \frac{\pi}{\lambda E}\int V(\mathbf{r}, z)dz \qquad (4.25)$$

where z is distance of electron travelled along the direction of electron beam, E is kinetic energy of the electron, and V is internal potential of the object, which includes mean inner potential and the potential induced by either magnetic field or electric field. The interference electron microscopy is unique in the sense that any phase change due to the internal potential change can be imaged at high sensitivity and at high spatial resolution.

Recent developments in this area include direct in-focus images of magnetic fields (magnetic particles, recording medium, and magnetic domain walls) [Tonomura, 1987], electrostatic fields (p-n junction and ferroelectric domain walls) [Pozzi & Missiroli, 1973; Zhang et al, 1992], and quantized fluxons in superconductors [Matsuda et al, 1991] etc., which not only reveal their spatial distribution but permit their magnitude to be determined from a

measurement of their phase shift. Also, highly accurate measurements of surface topography on an atomic scale, again using precise phase shift measurements, has been carried out [Osakabe et al, 1989] since an electron passing through one atom is delayed by about 1/100 of a wavelength, and phase shifts a factor of one hundred smaller than this are detectable. Finally, images of polymeric and unstained biological materials (analogous to phase microscopy in optical instruments) may be obtained in which contrast is produced by relative phase shifts rather than by the creation of defocus or chemical amplitude contrast [Kawasaki et al, 1986; Joy et al, 1992].

The other possible applications in the form of interference electron microscopy include defect studies in semiconductor materials (such as core structures of dislocations), electric field studies of Schottky barrier, electric imaging of quantum well structures, and interface studies of multilayer structures. It may also be possible to perform atomic level microanalysis by observing the enhanced phase shift and scattering associated with different atoms since each kind of atom has its unique inner potential.

(3) For analysis of the crystal structures, selected area microdiffraction method is very helpful. However, due to the physical limitation of available aperture size, the diameter of the selected area usually cannot be smaller than 100 nm. On the other hand, a conventional diffractogram from an electron micrograph does not provide the needed information, since in a diffractogram the intensities of reflections symmetrical about the optical axis are

always the same and the phase information is not complete [Lichte, 1991b]. The complex Fourier spectrum determined from the hologram reveals the very same information as the electron diffraction pattern in the microscope but with arbitrary aperture shape and size, and both amplitude and phase of it are accessible. Nano area microdiffraction [Lichte & Völkl, 1991], diffraction contrast, and weak beam defects analysis can all be performed with large dynamic area of parameter changes (such as aperture size and positions and phase shifting plate) in a single hologram on the vary same nano area of the specimen!

(4) Since images reconstructed from corrected holograms will maintain their phase and amplitude information intact, it will be possible to refine structure calculations by comparing both experimental amplitudes and phases with the computed image simulations rather than, as at present, comparing only intensities. It is also possible to directly and accurately use Patterson function method [Cowley, 1975] to study the structure of disordered materials (for example, amorphous and polymer materials).

(5) Optical or computer-reconstructed holograms also can be used to present three dimensional images of the specimen for observation rather than using only the two-dimensional projections available at present. Since the reconstructed image is three dimensional, an in-focus image can be obtained optically or numerically even if the original image was out of focus in forming the hologram.

REFERENCES

Ade, G. and Lauer, R., Optik, **88**(1991)3, p. 103-108

Bonnet, N. Troyon, M. and Gallion, P., Proc. 9th Int. Congr. Electron Micros., Toronto, **1**(1978), p. 222

Cowley, J.M., Diffraction Physics, (North-Holland Pub. Co., 1975), p. 125

DeVelis, J.B., Parrent, G.B. and Thompson, B.J., J. Opt. Soc. Am. , **56**(1966), p. 423-427

Franke, F.J., Herrmann, K.-H., and Lichte, H., Scanning Micros. Suppl., **2**(1988), p. 59

Gabor, D., Proc. Roy. Soc. London, Ser. **A197**(1949), p. 45

Gabor, D., Proc. Roy. Soc. London, Ser. **B64**(1951), p. 449

Haine, M.E. and Dyson, J., Nature, **166**(1950), p. 315

Haine, M.E. and Mulvey, T., J. Opt. Soc. Am., **42**(1952), p. 763

Hanszen, K-J., in Adv. Electron. Elect. Phys., **59**(1982), p. 1

Harada, K., Ogai, K., and Shimizu, R., J. Electron. Microsc., **39**(1990), p. 465-469

Hawkes, P.W., Adv. in Opt. and Elect. Micro., 7(1978), p. 101

Hibi, T., J. Electron. Micros., **4**(1956), p. 10

Hillier, H. and Ramberg, E.G., J. Appl. Phys., **18**(1947), p. 48

Joy, D.C., Zhang, X., Zhang, Y., and Hashimoto, T., "Electron Holography for Polymer Microscopy", presented in the annual meeting of American Chemistry Society, San Francisco, CA, April, (1992)

Kawasaki, T., Endo, J., Matsuda, T., Osakabe, N., and Tonomura, A., J. Electron Microsc., **35**(1986), p. 211

Leith, E.N. and Upatnieks, J., J. Opt. Soc. Am., **52**(1962), p. 1123

Lichte, H., Bildenen off axis Elektronenholographie atomarer Strukturen. Habilitationsschrift Tübingen, 1987

Lichte, H., Adv. in Opt. and Elect. Micro., **12**(1991a), p. 25

Lichte, H., Inst. Phys. Conf. Ser. No.119: Section 11, (1991b), p. 465

Lichte, H. and Völkl, E., Proc. 49th Annual Meeting of EMSA, San Jose, (1991), p. 670

Matsuda, M., Fukuhara, A., Yoshida, T., Hasegawa, S., Tonomura, A., and Ru, Q., Phys. Rev. Lett., **66**(1991), p. 457

Möllenstedt G. and Bayh, W., Phys. Bl., **18**(1962), p. 299

Möllenstedt G. and Düker, H., Naturwiss, **42**(1954), p. 41

Möllenstedt G. and Wahl, H., Naturwissenschaften, **55**(1968), p. 340

Osakabe, N., Endo, J., Matsuda, T., Tonomura, A., and Fukuhara, A., Phys. Rev.Lett., **62**(1989), p. 29

Pozzi, G. and Missiroli, G.F., J. Microsc. (Paris), **18**(1973), p. 103

Reimer, L., Transmission Electron Microscopy Physics of Image Formation and Microanalysis, (Springer-Verlag, 1984)

Scherzer, O., Z. Phys. **101**(1936), p. 593

Scherzer, O., JAP, **20**(1949), p. 20

Thompson, B.J., Jap. J. Appl. Phys., **4**, Suppl. **1**(1965), p. 302

Tonomura, A., J. Electron. Micros., **18**(1969), p. 77

Tonomura, A., Prog. in Opt., XXII, (1986), p. 201

Tonomura, A., Rev. Modern Phys., **59**(1987), p. 639

Tonomura, A., Phys. Today, (1990)4, p. 22

Tonomura, A., Fukuhara, A., Watanabe, H., and Komoda, T., Jap. J. Appl. Phys., **7**(1968), p. 295

Völkl, E. and Lichte, H., Ultramicros., **32**(1990), p. 177

Yanzi, M., Optik, **68**(1984), p. 319

Zhang, X. and Joy, D.C., Proc. 49th Annual Meeting of EMSA, San Jose, (1991), p. 678

Zhang, X., Hashimoto, T., and Joy, D.C., Appl. Phys. Lett., **60**(1992), p. 784

CHAPTER FIVE

PRACTICAL ELECTRON HOLOGRAPHY

5.1 Introduction

As discussed in the last chapter, electron holography distinguishes itself from conventional transmission electron microscopy by the promised capability of improving the ultimate resolution of the electron microscopy image and by the ability of completely retrieving both amplitude and phase information of the specimen and hence the formation of phase image. It has been realized that a TEM with a high brightness, high coherent field-emission electron source is a basic requirement to perform electron holography studies [Tonomura, 1989; Lichte, 1990]. The Hitachi HF2000 FE TEM is the first commercial TEM which permits the routine generation of high resolution electron holograms. The other essential apparatus is the electrostatic biprism [Möllenstedt & Düker, 1954]. In this chapter, major experimental apparatus such as the field-emission electron microscope and the electron biprism, and key experimental techniques for the in-line and the off-axis electron holography will be discussed in detail. The discussion will be focused on the implementation of electron holography on the Hitachi HF2000 FE TEM.

Eqn. 4.22 has shown the intensity distribution of an off-axis electron hologram. If one considers the effect of finite electron beam coherence then eqn. 4.22 becomes [Lichte, 1991]

$$I = 1 + A_I^2(\mathbf{r}) + 2V \cdot A_I(\mathbf{r}) \cos[2\pi \mathbf{R}_c \mathbf{r} - \Phi_I(\mathbf{r})] \qquad (5.1)$$

where V is the contrast of the hologram fringes ($0 \leq V \leq 1$) and is a function of degree of spatial coherence, the instabilities of microscope, a.c. stray fields, and recording medium transfer properties. Changing contrast V from 1 (perfect coherence) towards the minimum of 0 (incoherence) will damp, and eventually destroy, the contrast of the interference fringes. Therefore, the most important requirement for electron holography is a beam of electrons of high enough coherence to produce fringes of good contrast. Secondly, this coherent beam must keep a long enough coherent length extended over a large area to pass both through the specimen as well as through a suitable hole on the adjacent specimen area so that an image and a reference beam are formed. Thirdly, an electrostatic biprism is needed to cause the two beams to overlap and form an interferogram. Finally, a suitable recording medium has to be chosen to record the fringe patterns (holograms) on the medium as linearly as possible with highest signal to noise ratio.

5.2 Field-Emission Electron Source

The coherence of a wave field refers to its ability to produce interference effects [Spence, 1988]. The principle of electron holography relies on the basic interference and diffraction properties of an electron beam. Coherent waves are a fundamental requirement for interference imaging. The recent rapid development of practical electron holography is mainly due to the development of the field-emission electron source [Crewe, 1968; Tonomura, 1987].

5.2.1 Coherence length

A single electron corresponds to a spatially localized wave packet [Tonomura, 1986]. The phase of a wave function is coherent only within a wave packet. Therefore, an electron wave-packet can only interfere with itself, thus the ideal illuminating beam consists of a perfectly plane, mono-energetic wave with a unique wave vector **k** (where $|\mathbf{k}| = 1/\lambda$). Strictly speaking, interference fringes are observed as the time average of a probability density $|\Psi|^2$ formed by each successively arriving electron.

Practically any physically realizable electron beam has a finite source size and hence a non-zero illumination angle 2α (see Figure 5.1). The illumination angle 2α is the sum of the convergence angle α_c of the beam, and of a term α_p representing the contribution from a finite probe size. This finite beam size results in a beam that has partial coherence ($V \leq 1$) extending over a region of lateral extent $\lambda/2\alpha$. This lateral extent is called spatial (or transverse) coherence length (R) which represents the dimensions of the coherent region in a plane perpendicular to the propagation direction.

In addition of finite source size, if the electron beam also has an energy spread ΔE (eV) then, parallel to the incident direction of the electrons, there is also partial coherence extending longitudinally for a distance of the order of $2\lambda E/\Delta E$, where E is the kinetic energy of the electron. This distance is called time (or longitudinal) coherence length and is the maximum difference between two beams travelling along two separate paths, under which interference is possible [Tonomura, 1986]. The energy spread of the cold FE gun on the

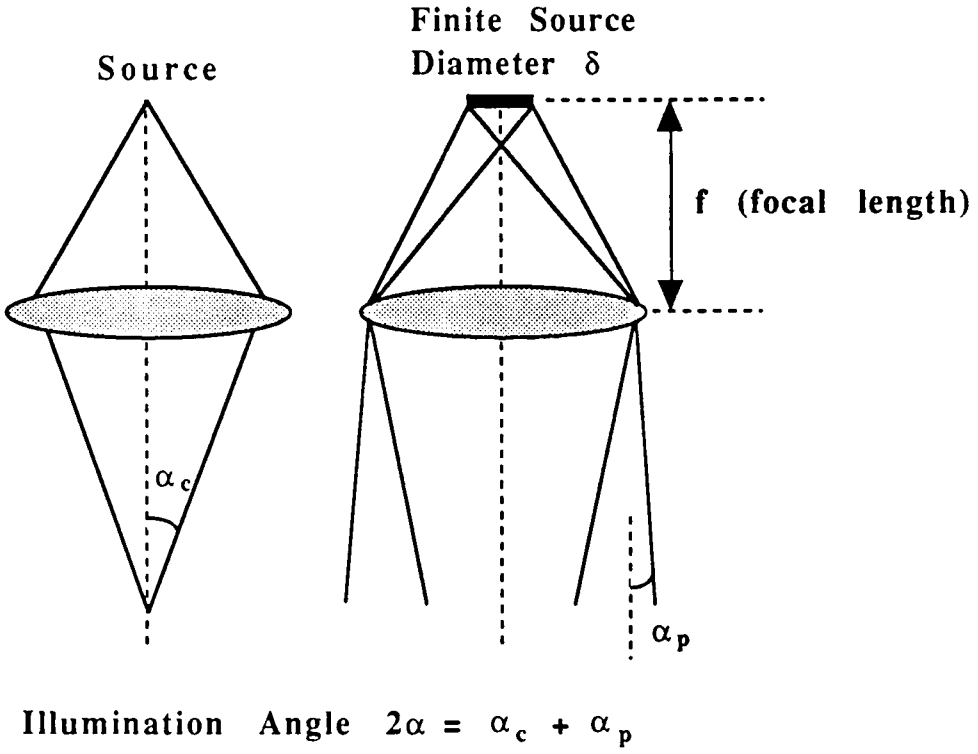


Figure 5.1. Illumination system to demonstrate contribution to illumination angle.

HF2000 FE TEM is measured as 0.4 eV at 200 keV (from all sources including noise and drift) using Gatan PEELS™ system at 8.9 μA emission current and with 50 μm condensor aperture and 1 mm spectrometer entrance aperture [Joy, 1991]. The calculated time coherence length is therefore about 1.8 μm, compared with about a half micrometer for a conventional LaB_6 electron gun.

Although time coherence length will affect the contrast of the interference fringes, the spatial coherence length imposes more stringent limits on the observability of interference fringes [Lichte, 1991; Tonomura, 1987]. The spatial coherence length may seem to become arbitrarily large with the selection of a small illumination angle 2α. Actually, though, it is controlled by the brightness of the electron gun.

5.2.2 Brightness

The brightness β of an electron beam is defined as:

$$\beta = \frac{d^2 I}{d\Omega dA} \tag{5.2}$$

where I is the beam current, Ω is the solid angle defining the "cone" of the electrons, and A is the area in one of the beam crossovers [Kasper, 1980]. Since β is conserved in any plane of an electro-optical system and since brightness is virtually uniform with angle α, in the small angle approximation, the brightness is then

$$\beta = \frac{J}{\pi\alpha^2} \tag{5.3}$$

where J is the current density (amps/cm^2). When illumination angle 2α is decreased in an attempt to increase the spatial coherence length, current density J of the electron beam is also reduced. The available current density is given $J = \beta\pi\alpha^2$. Therefore, the spatial coherence length increases with beam brightness and the relation is given by

$$R = \frac{\lambda}{2}\sqrt{\frac{\pi\cdot\beta}{J}} \qquad (5.4)$$

Spatial coherence length is, therefore, proportional to $\sqrt{\beta}$. For example, arbitrarily assuming a minimum acceptable current density at the specimen of 0.1 amps/cm^2, a source of brightness 10^6 amps/cm^2/sr at 100 keV (a typical thermionic tungsten gun) would be restricted to a minimum practical collimation angle 2α of 0.36 mrad, and a corresponding spatial coherence length of only about 10 Å; while a source of brightness 10^9 amps/cm^2/sr (a typical field-emission gun) could be collimated to 10 micro-rads and given a spatial coherent length of about 3000 Å in size. A field-emission gun is thus a prerequisite for electron holography if useful imaging currents and reasonable recording times are to be attained.

The maximum size of the convergence angle α_c is determined by the second condensor aperture near the point of condensor focus. When the second condensor lens is defocused, the convergence angle is defined not by the size of condensor aperture but by the size of the crossover image and its distance from the specimen [Agar et al,

1974]. The smaller α_c can be achieved by spreading the beam as far as enough brightness is maintained for holographic imaging.

The probe size contribution α_p to the illumination angle 2α is a function of the apparent source size and the design of the illumination system. With the source placed at the focus of a condensor lens then α_p is equal to d/f_0, where d is the size of the illumination source and f_0 is the focal length of the lens. For a typical case where f_0 is of the order of 1 cm then a thermionic source, for which a typical source size is of the order of few tens of microns, would have an α_p of about 1 mrad. However, a field-emission source, for which the effective source size is of the order of 50 Å, would give an α_p of 0.5 micro-rads much better than a thermionic source.

5.2.3 Field-emission electron gun and HF2000 FE TEM

In Hitachi HF2000 cold FE TEM, a field-emission electron gun consists of a pointed cathode and two anodes (Figure 5.2). The cathode is made of a single crystal tungsten tip polished to 1000 Å radius of curvature. Electrons are emitted from a <310> oriented tungsten tip when a high voltage (V_1) is applied across the cathode and the first anode. Another high voltage (V_2) is applied across the cathode and the second anode. The first and second anode voltages (V_1 and V_2) constitute an electrostatic lens which focuses the electrons and forms a beam of about 100 Å in diameter. The cold field-emission electron source is built in a compact gun, maintained at a vacuum of the order of 10^{-11} torr to ensure stable operation.

In the HF2000 FE TEM, the electron beam is accelerated in a 6-stage ceramic accelerator up to 200 keV (Figure 5.3). The microscope

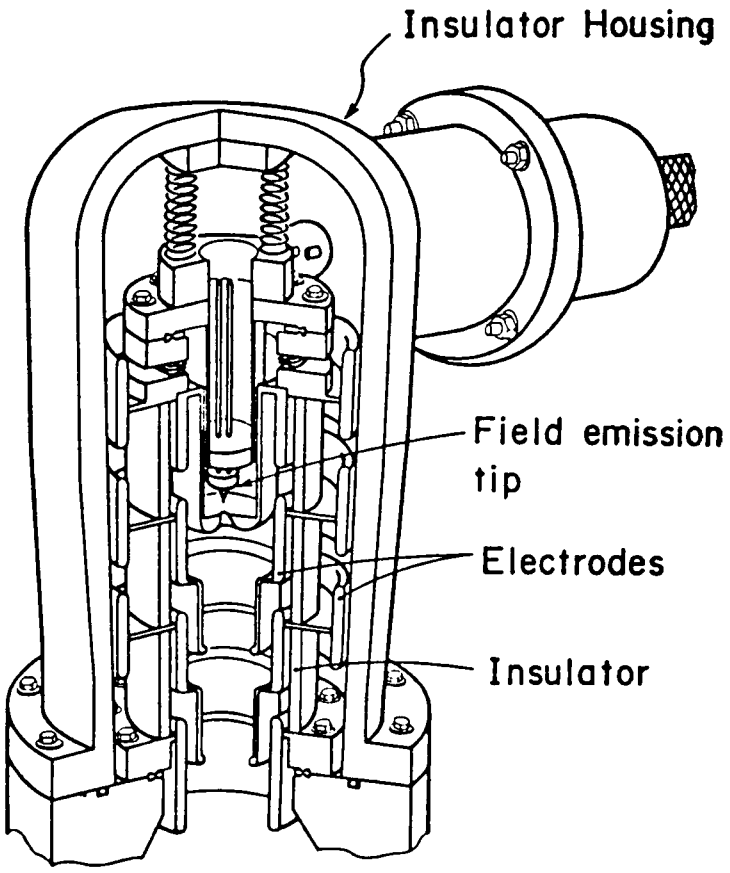


Figure 5.2. Schematic configuration of a field-emission electron gun (adopted from [Tonomura, 1992]).

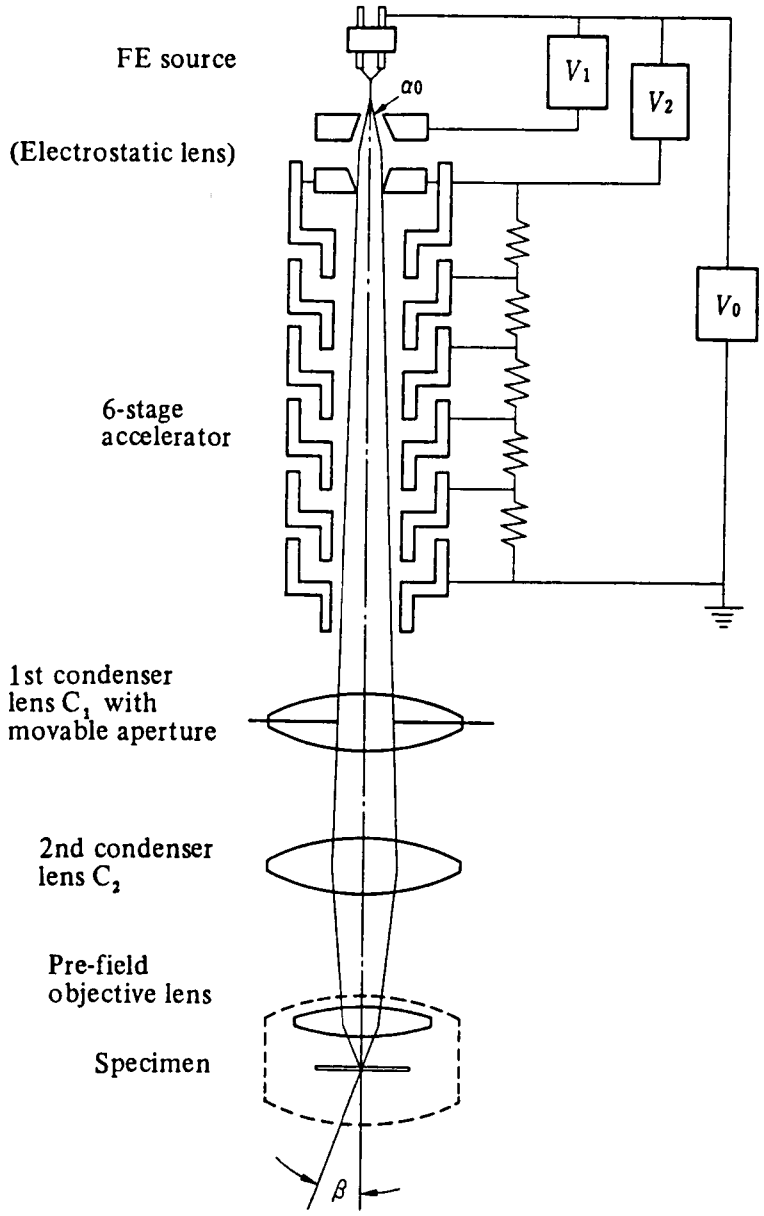


Figure 5.3. Schematic configuration of the illumination lens system of the Hitachi HF2000 FE TEM (adopted from Hitachi Instrument News, 19th ed., 1990).

is equipped with a cold FE electron gun of a brightness of about $3x10^8$ amps/cm^2/sr. The actual illuminating optics consist of a pair of condensor lenses and a condensor-objective lens. The condensor lenses permit the source size to be demagnified to minimize while the condensor-objective lens forms a demagnified image of the condensor aperture at the specimen plane, giving a small beam area with nearly parallel illumination, i.e., for an illuminated area 20 Å in diameter the effective total illumination angle is measured to be 5 micro-rads. With the aid of the second and the third condensor stigmators, the electron source could be changed from a round shape to an elliptical shape with suitable eccentricity and orientation to optimize mutually the contrast and the intensity of the hologram. The object lens has a spherical aberration coefficient C_s = 1.2 mm and chromatic aberration coefficient C_c = 1.49 mm. By combination of intermediate lenses and projection lenses at suitable excitations the magnification of the image could be varied from 50x to 1000,000x. The magnification selections are controlled by a digital computer.

5.3 Electron Biprism

5.3.1 Configuration

In these days the most effective way to generate a high quality electron hologram is to employ the off-axis electron holography technique. In this case, the hologram is produced by the interference of the object and reference waves which are both deflected towards the optical axis of the microscope by the action of the electrostatic biprism [Möllenstedt & Düker, 1954]. A schematic drawing of the biprism design is shown in Figure 5.4. The biprism consists of a

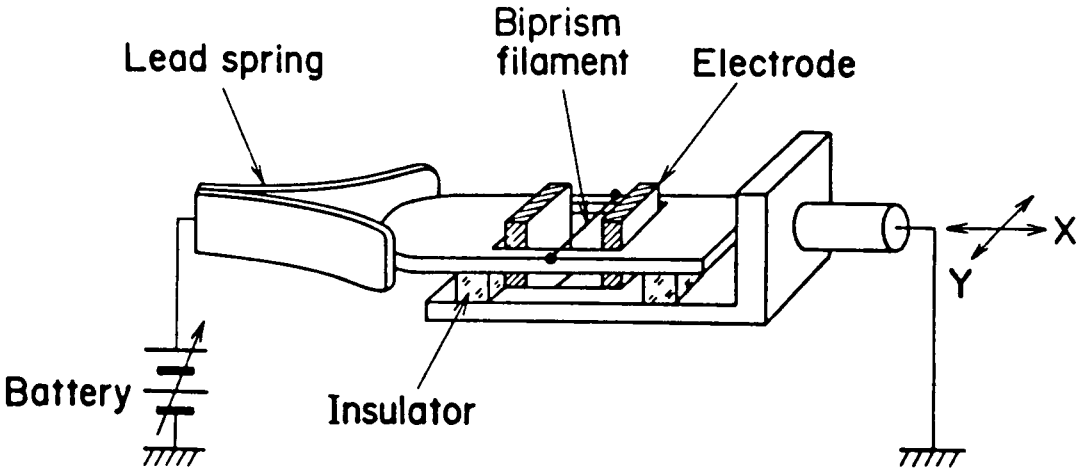


Figure 5.4. Design sketch for biprism (adopted from [Tonomura, 1992]).

conductive central fine wire, carrying -300 to +300 volts, placed between two grounded parallel electrode plates. The electrostatic potential is applied through a spring lead from outside of the microscope. On the HF2000 the wire is actually rotatable in the biprism holder so as to position the fringes in any desired direction relative to a feature of interest, and it can be inserted as required like an aperture (Figure 5.5). The biprism is positioned between the back focal plane of the object lens and the image plane (another location is between the first and the second intermediate apertures for generating ultra-fine fringe spacing). For ordinary microscopy the biprism can easily be removed from the optical axis.

At a positive electrostatic potential the paths of the electrons passing on both sides of the central wire are deflected towards the optical axis of the microscope; while at negative potential the split electron beams will diverge from the optical axis. In general a positive potential is applied on the biprism. Figure 5.6 shows the shadow image of the central wire before the application of a potential together with the Fresnel fringes formed from scattering at the edge of the wire. When a positive potential is applied to the biprism then the interference fringes fill the field of view of a finite width. The stability of potential applied to the biprism should be in the order of 10^{-6}/min [Tonomura, 1992]. Since the wire draws no current, any stable, noise free voltage source such as standard dry cell batteries feeding a resistive potential divider, or a digitally controlled electronic supply, is suitable [Joy et al, 1992; Tonomura, 1992]. It is desirable to have a digital voltmeter in parallel with the

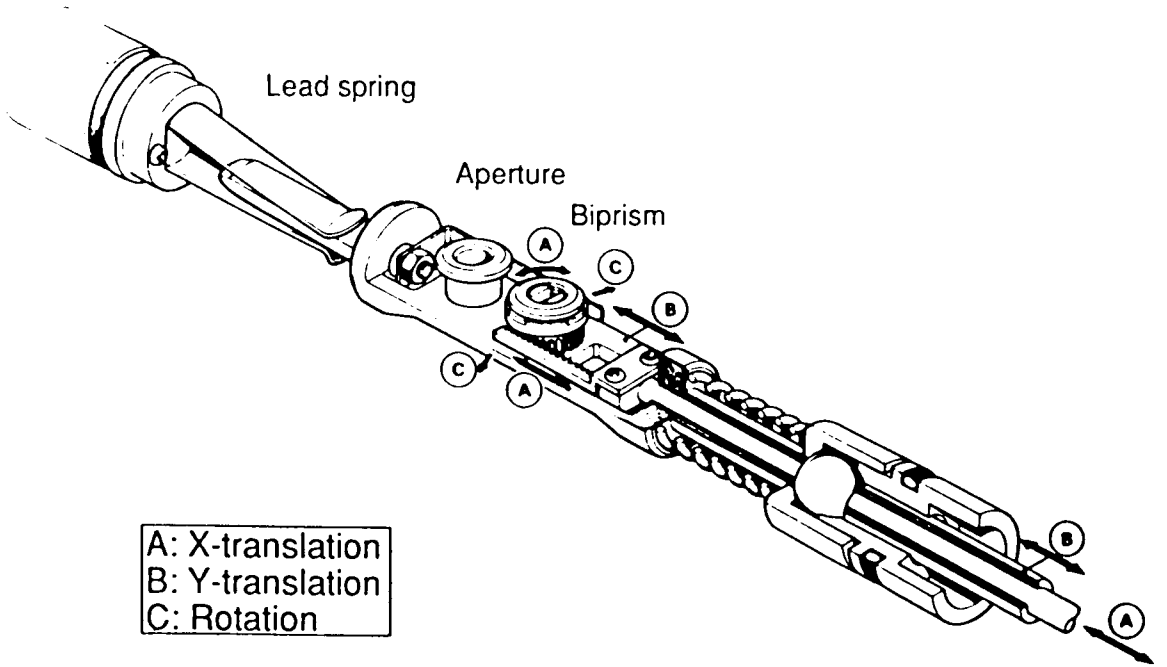


Figure 5.5. Rotatable electron biprism used in the study (adopted from [Tonomura, 1992]).

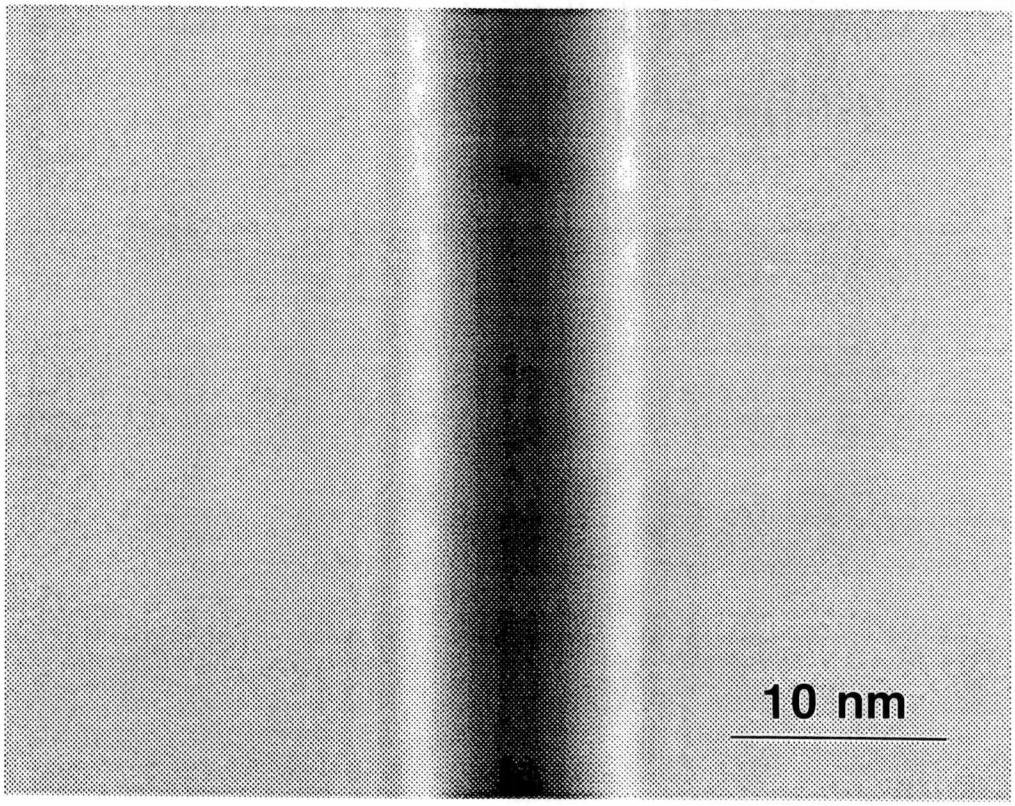


Figure 5.6. The shadow image of the biprism before the application of a potential.

wire so that the exact potential applied can be determined independently.

5.3.2 Making an electron biprism

The central wire should be no more than about 0.5 μm in diameter to avoiding shadowing the field of view, since the extremely short electron spatial coherence length is further decreased by the shadow. The wire should be able to be stretched taut across an aperture 5 mm in size. Based on these requirements, the most reasonable candidate material for the wire is quartz fiber [Strong et al, 1938; Tonomura, 1986], made conducting by the application of an evaporated gold film (300 Å).

The wires are made by melting a 5 mm diameter quartz rod in a gas flame. Using two hands to pull the rod apart very quickly and allowing the rising gas jet to pull the wire up and away from the tip, one can easily get a piece of fine fiber joining the tip of the rod with a length between 5 to 12 inches and a diameter of 0.2 to 0.5 μm in the front end. The resultant fine fiber is too small to be directly visible in most cases, and it must be observed by scattered light. From the color of scattered light one can get a first degree guess of the fiber diameter (i.e., based on the wavelength of the light). The fibers are picked up on metal wire forks coated with an adhesive, and then positioned across the biprism holder, fastened into place with care (to keep it centered and tensioned) and fixed at the two ends using either silver paint or 1% collodion solution in amyl acetate. The mounted fiber is then coated by evaporating a 300 Å gold film on it. The fibers made by this way are straight and of quite

uniform diameter for some distance up to several inches. Figure 5.7 is a SEM image of one of the fiber we made in the lab. The diameter of the wire is determined as 0.30 μm (FWTM). With reasonable care a fiber lasts for many months when kept in the microscope.

The size of the resulting fine fiber will depend on a number of factors. Chief among these are the size of original fiber, the flame temperature, and the speed of pulling it apart. Another key factor is the cleanliness of the fiber. Much of the dust on objects around a lab and floating in the air is inorganic. If a fiber is heated where a piece of dust has settled, the metallic salts form silicates and in general completely spoil the surface and that is to say then, the fiber is useless no matter how thin it is.

5.3.3 Interference fringe spacing

In the diffraction plane, the maximum radius of the central autocorrelation is $2|\mathbf{R}_{max}|$ where $|\mathbf{R}_{max}| = 1/S_{rmax}$ is the maximum spatial frequency of the object, S_{rmax} is the smallest feature of the object to be resolved (Figure 5.8). The side band width is therefore $2|\mathbf{R}_{max}|$. Consequently, the spectrum of the image wave can only be isolated completely by an aperture $2|\mathbf{R}_{max}|$ if the carrier spatial frequency $|\mathbf{R}_c| = 1/S_f$ is higher than $3|\mathbf{R}_{max}|$, where S_f is the interference spacing at the object plane. As a general rule the resolution obtainable in a reconstructed hologram (S_{rmax}) is about three times the fringe spacing S_f [Lichte, 1991]. Thus the ability to be able set and alter the fringe spacing is of importance since different applications will make different demands.

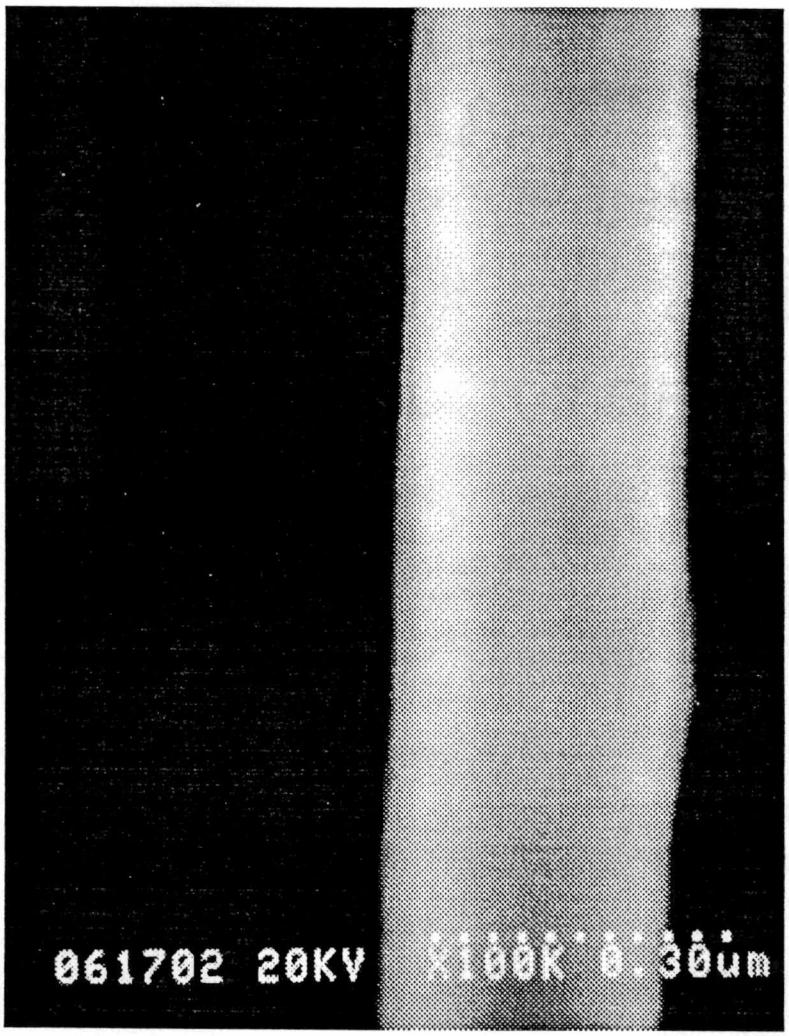


Figure 5.7. SEM image of a lab-made fiber with a diameter of 0.3 μm.

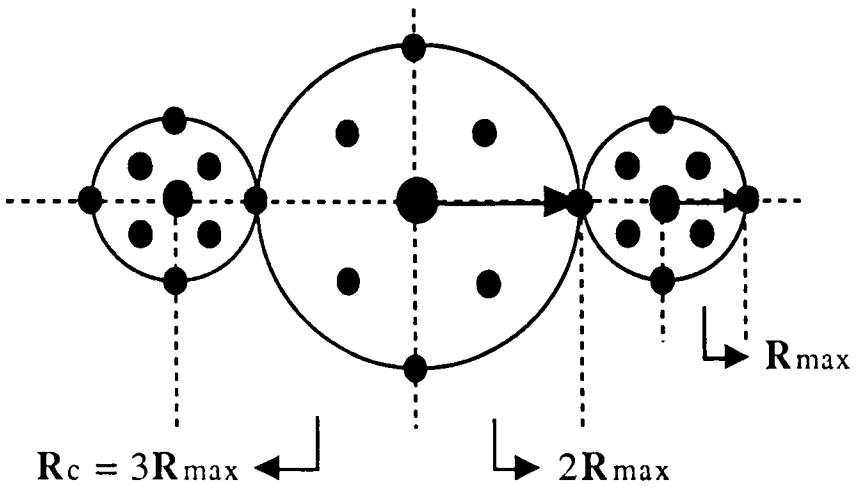


Figure 5.8. Schematic illustration of selection criteria of the side band.

Experimentally, the interference fringe spacing is determined by the field generated around the central wire (Figure 5.9). When a positive electric potential V_b is applied to the wire, the electric potential V(r) around the neighborhood of the wire can be approximated by [Tonomura, 1987]:

$$V(r) = V_b \cdot \frac{\ln(r/r_1)}{\ln(r_1/r_2)} \tag{5.5}$$

where r_1 and r_2 are the wire radius and the distance from the wire center to the electrode, respectively. Since electrons will be deflected when they pass through the biprism, the deflection angle can be calculated as [Tonomura, 1987]

$$\delta = \frac{\pi e V_b}{2E \cdot \ln(r_1/r_2)} \tag{5.6}$$

where E is the kinetic energy of an electron which is given by

$$E = E_0 \cdot \left(1 + \frac{1}{2} \cdot \frac{eE_0}{m_0 C^2}\right)\left(1 + \frac{eE_0}{m_0 C^2}\right)^{-1} \tag{5.7}$$

where E_0, m_0, e, and C are electron accelerating voltage, electron rest mass, electron charge, and velocity of light. It is obvious from eqn. 5.6 that the deflection angle is constant, independent of the incident position of an electron, and is proportional to the applied voltage.

Let us denote the distance between the effective electron source (in this case the back focal plane of the object lens) and the central wire as "a," and the distance between the wire and the intermediate image plane as "b." Based on the wave theory, let us

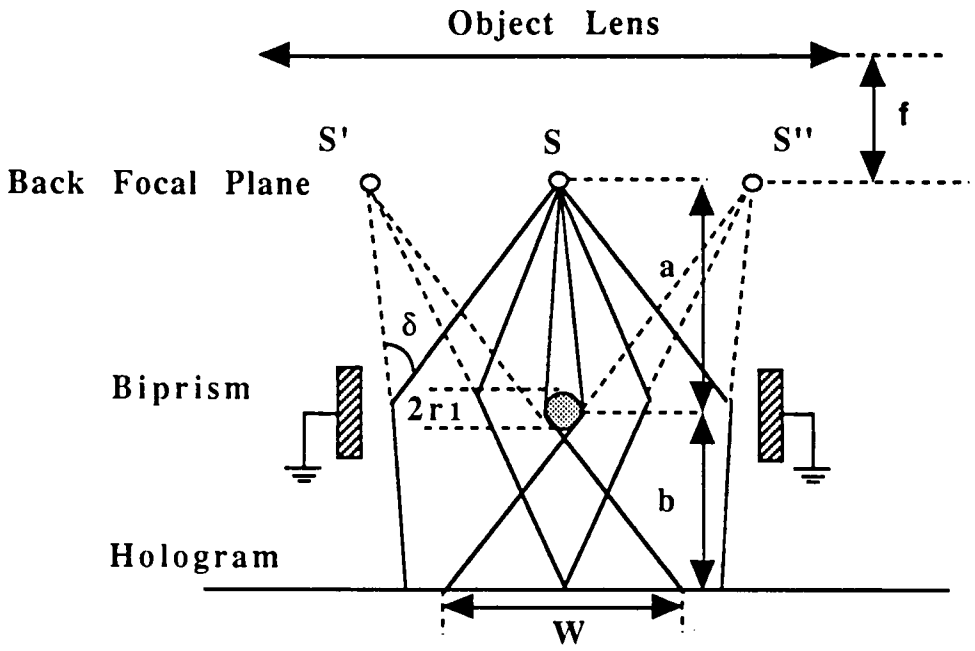


Figure 5.9. Schematic illustration of the parameters used in calculations of interference fringe spacing and interference width.

assume one coherent electron beam is split into two waves passing through two sides of the electron biprism, and then the two beams interference each other. From the density distribution of fringe patterns within the interference region, the interference fringe spacing is given by [Chen, 1986; Lichte, 1986]

$$S_f = \frac{\pi\cdot(a+b)\cdot\lambda}{2a\cdot\delta} \qquad \text{at the image plane}$$

$$= \frac{\pi\cdot\lambda\cdot f}{2a\cdot\delta} \qquad \text{at the object plane } \{M = (a+b)/f\} \qquad (5.8)$$

where f is the focal length of the object length, M is the magnification of the object lens. Therefore, combining eqns. 5.6 and 5.8 we can calculate the fringe spacing (S_f) as

$$S_f = \frac{E\cdot f\cdot |\ln(r_1/r_2)|}{a\cdot eV_b} \qquad \text{at the object plane} \qquad (5.9)$$

Meantime, the overlapping width of the two waves, and hence the interference width (W) is given by

$$W = 2\delta\cdot b - \frac{2(a+b)\cdot r_1}{a} \qquad \text{at the image plane}$$

$$= \frac{2b\cdot f\cdot\delta}{a+b} - 2r_1\cdot\frac{f}{a} \qquad \text{at the object plane}$$

$$= \frac{2\pi\cdot f\cdot b\cdot eV_b}{E\cdot(a+b)\cdot|\ln(r_1/r_2)|} - 2r_1\cdot\frac{f}{a} \qquad \text{at the object plane}$$

$$(5.10)$$

The number of interference fringe is therefore can be calculated as N = W/S_f. Eqn. 5.9 indicates clearly that the fringe spacing is inversely

proportional to the applied voltage and is logarithmically related to the radius of the center wire. The interference width is directly related to the distance b and the radius r_1 since the voltage V_b is determined by the fringe spacing S_f. Experimentally, as far as the fringe spacing and the interference width are concerned one has these two parameters to play with: the applied voltage V_b and the distance b which is controlled by the intermediate lens. From eqn. 5.8, the condition $|\mathbf{R}_c| > 3|\mathbf{R}_{max}|$ for separation of the sidebands from the autocorrelation can be written in another form:

$$\delta \geq 3 \cdot \frac{\pi \cdot f \cdot \lambda}{2a \cdot S_{rmax}} \qquad (5.11)$$

where S_{rmax} is the resolution limit to be resolved. This leads to the minimum values of applied voltage V_b as

$$V_b \geq \frac{3f \cdot \lambda \cdot E \cdot |\ln(r_1/r_2)|}{e \cdot S_{rmax} \cdot a} \quad \text{(Volts)} \qquad (5.12)$$

Evidently, from eqns. 5.8-9 the fringe number can be made arbitrarily large by increasing the deflection angle δ, i.e., increasing the applied voltage V_b on the biprism. However, the experimental results show that interference fringes will disappear when V_b reaches some finite value. This is due to the degree of spatial coherence damping the contrast of the interference fringes.

Using our experimental data: λ = 0.0251 Å, E_0 = 200 keV, a = 180 mm, r_1 = 0.25 μm, r_2 = 5 mm, and f = 1.96 mm, we find $S_{f(obj)} = 140.73/V_b(V)$ (Å), which is very close to our experimentally measured data. Figure 5.10 shows the relation between $S_{f(obj)}$ and

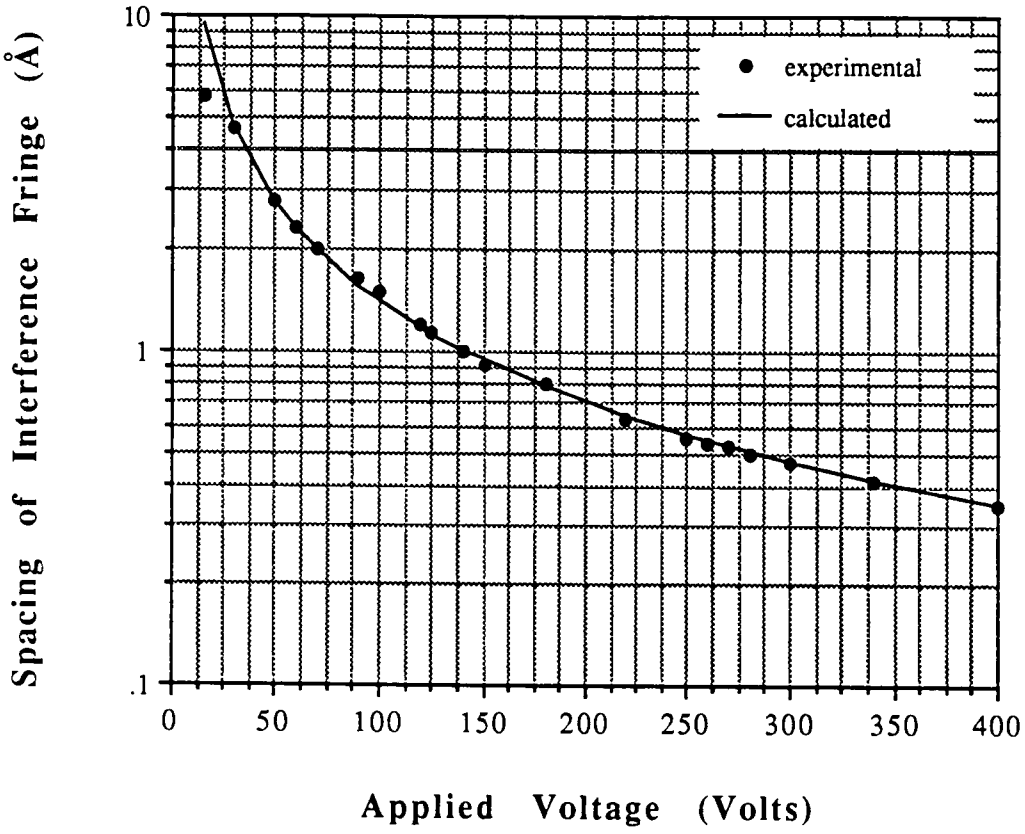


Figure 5.10. Fringe spacing versus biprism voltage at 200 keV beam energy.

applied voltage (V_b) under above conditions. Since the applied voltage is determined by the choice of the fringe spacing, the width of the hologram can only be controlled by the distance b using the intermediate lens. Figure 5.11(a) shows the calculated results (from eqn. 5.10) of the interference width changes with the change of applied voltage at different b values. Figure 5.11(b) shows the width changes with desired fringe spacing at different b values. In order to satisfy the condition $|\mathbf{R}_c| > 3|\mathbf{R}_{max}|$ to avoid overlapping of the three areas in Fourier space, the minimum requirement of the applied voltage is calculated based on eqn. 5.12. The result is shown in Figure 5.12. Also shown in the figure is the corresponding minimum pixel number required to reconstruct the desired spatial resolution. This will be discussed in section 5.6.2.

The shadow of the biprism imposes another limitation on the interference width. The shadow produces a region from the optical axis through which no part of the image wave will pass [Völkl, 1991]. Therefore, the usable part of the interference width at the image plane is only

$$W_{eff} = W - r_1 \cdot \frac{a + b}{b} - f \cdot \lambda \cdot \frac{b}{a \cdot S_{rmax}} \qquad (5.13)$$

where f is the focal length of the objective lens. Considering the limitation of the shadow image of the central wire, the relation between effective and its original interference width vs applied voltage is shown in Figure 5.13. It is obvious from the results that above 100 V of applied voltage, the relative differences between

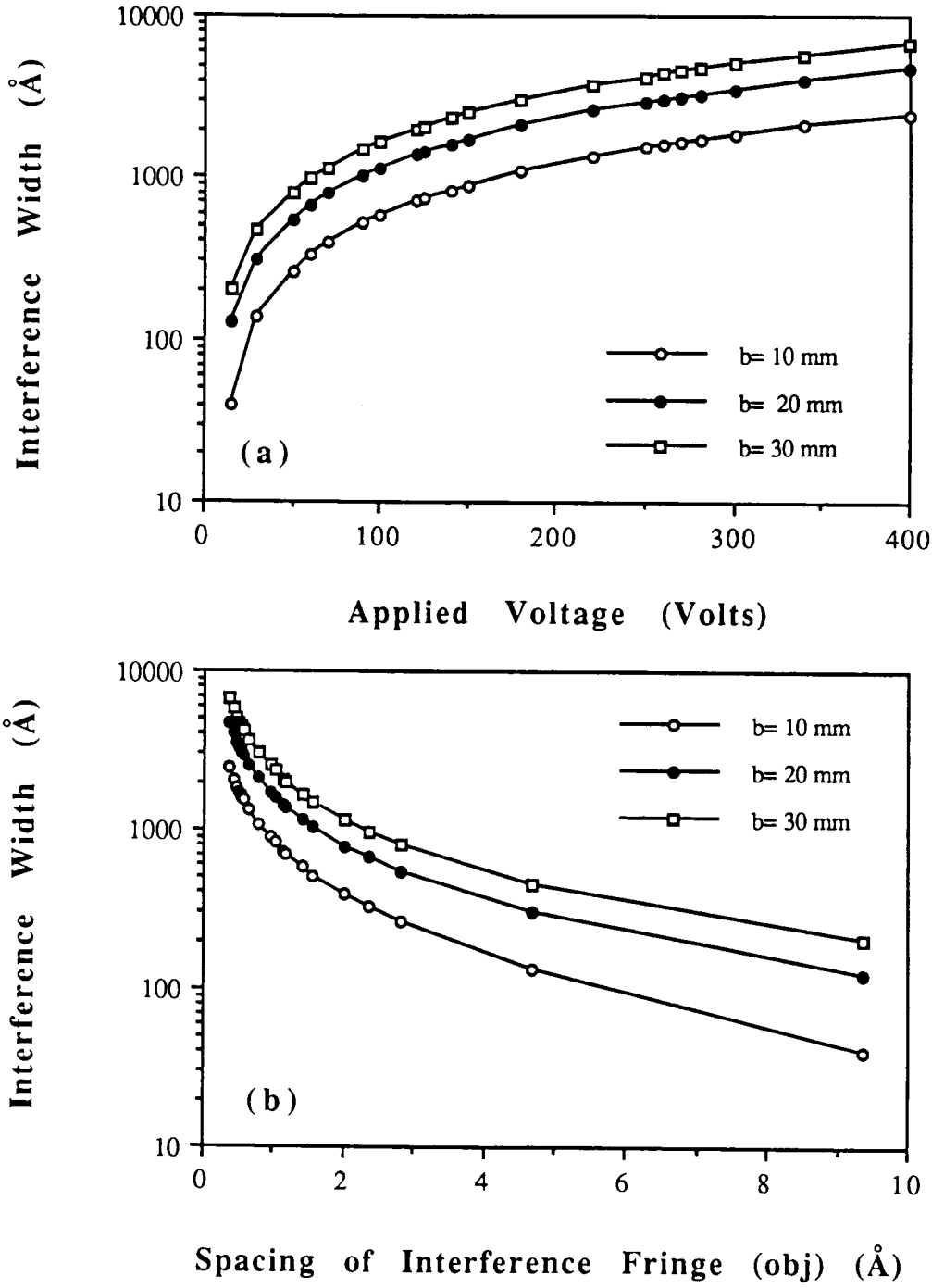


Figure 5.11. (a) Interference width versus biprism voltage at different b values, and (b) interference width versus fringe spacing at different b values.

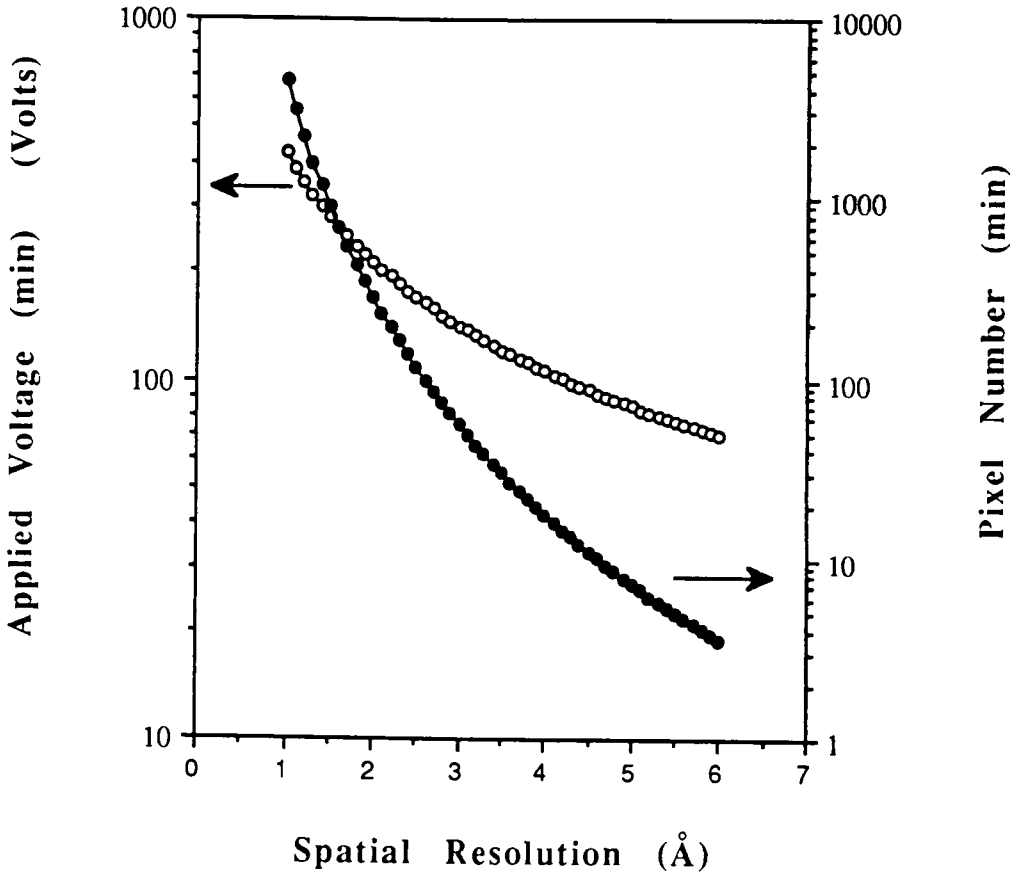


Figure 5.12. Minimum required biprism voltage and pixel number versus desired final resolution.

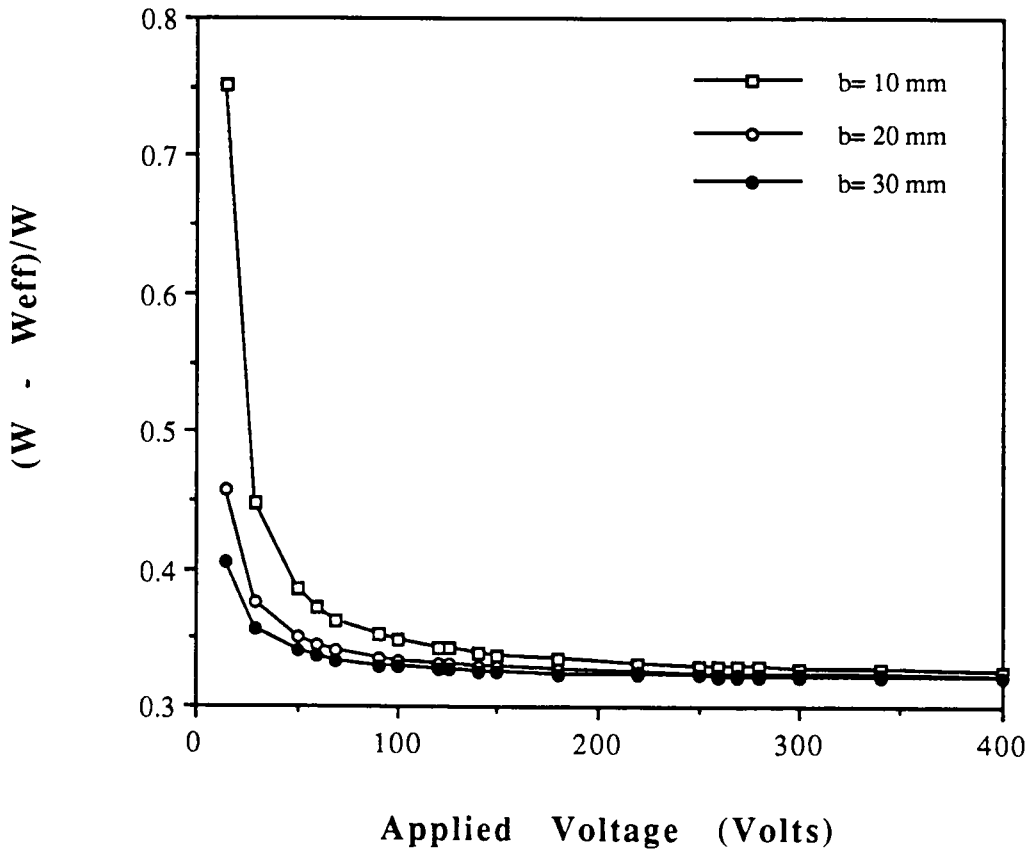


Figure 5.13. Effect of shadow of biprism on interference width versus biprism voltage at different b values.

these two are in the range of 30 to 35%. The highest difference is about 75%.

5.3.4 Complex degree of coherence

The electron beam generated by field-emission gun is considered as a partial coherence beam. The real part of the complex degree of the coherence [Born & Wolf, 1959] of the electron beam (γ_{12}) can be measured experimentally as $\gamma_{12} = (I_1 - I_2)/(I_1 + I_2)$, where I_1 is the value of maximum intensity of the fringes and I_2 is that of the minimum intensity [Hibi & Takahashi, 1969]. When $|\gamma_{12}| = 1$, the beam is said to be coherent; while $\gamma_{12} = 0$, the beam is said to be incoherent. If $0 < |\gamma_{12}| < 1$, the beam is said to be partially coherent. If the two electron beams that are to be interfered are of equal intensity, the measured values of contrast V of the interference fringes (without specimen) can be used to represent the $|\gamma_{12}|$ [Born & Wolf, 1959] from which one is able to evaluate the coherence of the electron beam. As an example, Figure 5.14 shows the measured the average fringe contrast for three photographic films, plotted as a function of the interference fringe spacing on the object plane. The fringe contrast also can be calculated as

$$\gamma = \frac{\sin\left(2\pi\cdot\frac{S}{\sigma}\right)}{2\pi\cdot\frac{S}{\sigma}} \quad \Leftarrow \quad \frac{S}{\sigma} = \frac{2b\cdot S}{a\cdot(a+b)\cdot\lambda}\cdot(r_1 + a\cdot\delta) \qquad (5.14)$$

where S is the size of virtual electron source, all other parameter are of their usual meanings. The calculated contrast shown in Figure 5.14 is actually the product of the contrast from eqn. 5.14 and the MTF of

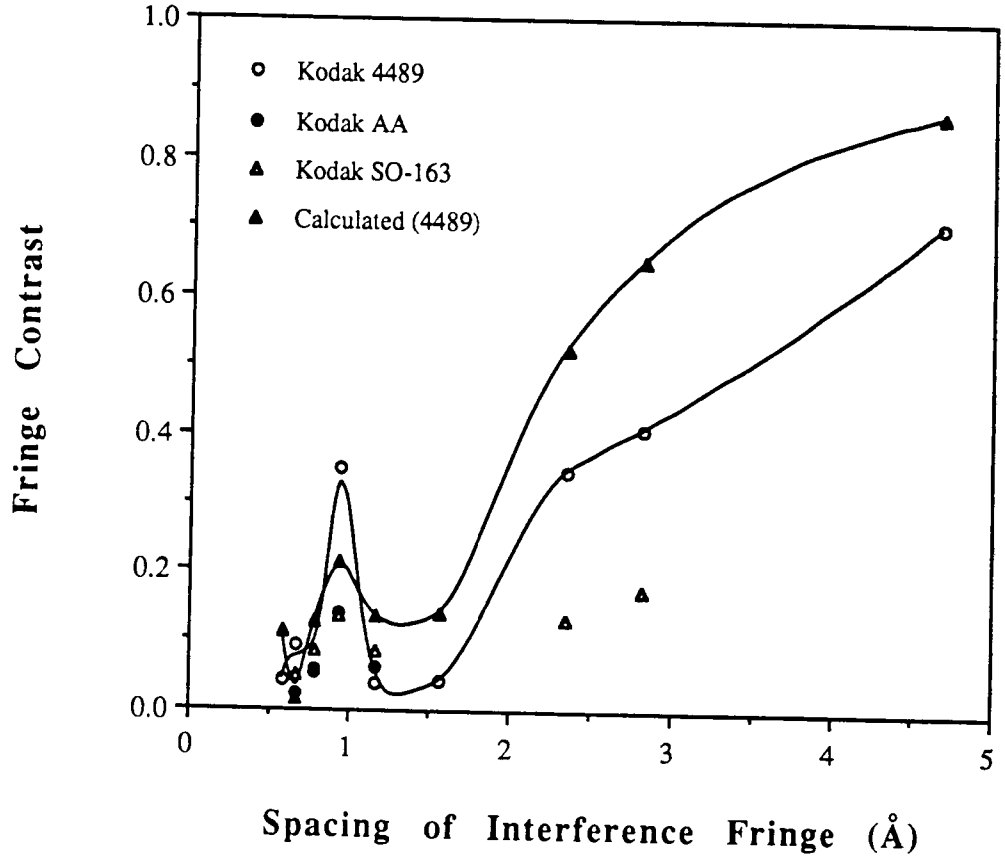


Figure 5.14. Fringe contrast versus fringe spacing.

Kodak 4489 film from eqn. 5.18 which includes the effect of the film. The calculated contrast fits the trend of the contrast changes from all three films. The discrepancy stems from the fact that calculated contrast does not include the effects of instability of the microscope, a.c. stray field, non-linear transformation of the films, and many other practical factors. It is shown in Figure 5.14 that when the virtual source width increases with increase of applied voltage on the biprism, the contrasts of the fringes gradually decrease up to an almost complete vanishing and reappear with the contrast associated with contrast inversion [Drahos & Delong, 1986].

5.4 Generating Electron Holograms

5.4.1 In-line Fraunhofer electron hologram

The in-line Fraunhofer electron hologram is no more than a highly defocused TEM image. The key factor that controls the quality of the hologram is the beam coherence. The Fraunhofer far field diffraction is determined by eqn. 4.19 and the resolution of the hologram is controlled by eqns. 4.20.

There are several factors needs to be considered when one chooses the defocusing distance D_{zc} for condensor lens and D_{zo} for object lens. (1) The spatial coherence length. The spatial coherence length is controlled by illumination angle 2α which, in turn, is controlled by the defocus (D_{zc}) of the condensor lens (see section 5.2.2). The higher the defocus, the smaller the illumination angle, and the larger the spatial coherence length, but the lower the current density (see eqn. 5.3). With a very low current density the exposure time increases which induces the problems such as the stabilities of

beam current, energy spread, and mechanical vibration. This also limits the magnification of taking hologram. (2) The information limits. The information that an in-line hologram carried is dependent on the number and the contrast of the Fresnel fringes [Gabor, 1951]. Let us define a clear space of average width, W, which is the distance between the objects. The best choice of the W is so that the fringe just covers both sides (just cover it). Using Abble's relation [Born & Wolf, 1959], the defocus of the object lens is given by

$$D_{zo} \approx W \cdot \frac{d_A}{\lambda} \tag{5.15}$$

where d_A is the Abble resolution limit, and is determined by the size of the object to be resolved. (3) Consideration of reconstruction. Since the conjugate image must disturb the reconstructed image as little as possible, Gabor found that the underfocus of the object lens is the best choice [Gabor, 1951]. Further, it is found that the minimum defocus of the object lens D_{zo} has to satisfy the condition:

$$-D_{zo} > \frac{3}{4} C_s \cdot \left(\frac{\lambda}{d_A} \right)^2 \tag{5.16}$$

where C_s is the spherical aberration of the object lens.

In our in-line holography experiments on HF2000, the illumination is overfocused to give a small crossover above the specimen. The measured illumination angle is about 3 micro-rads which gives the spatial coherence length of about 4000 Å. Under this condition the allowed exposure time is about 100 seconds. Figure 5.15(a) is an divergent beam in-line hologram of MgO crystals taken

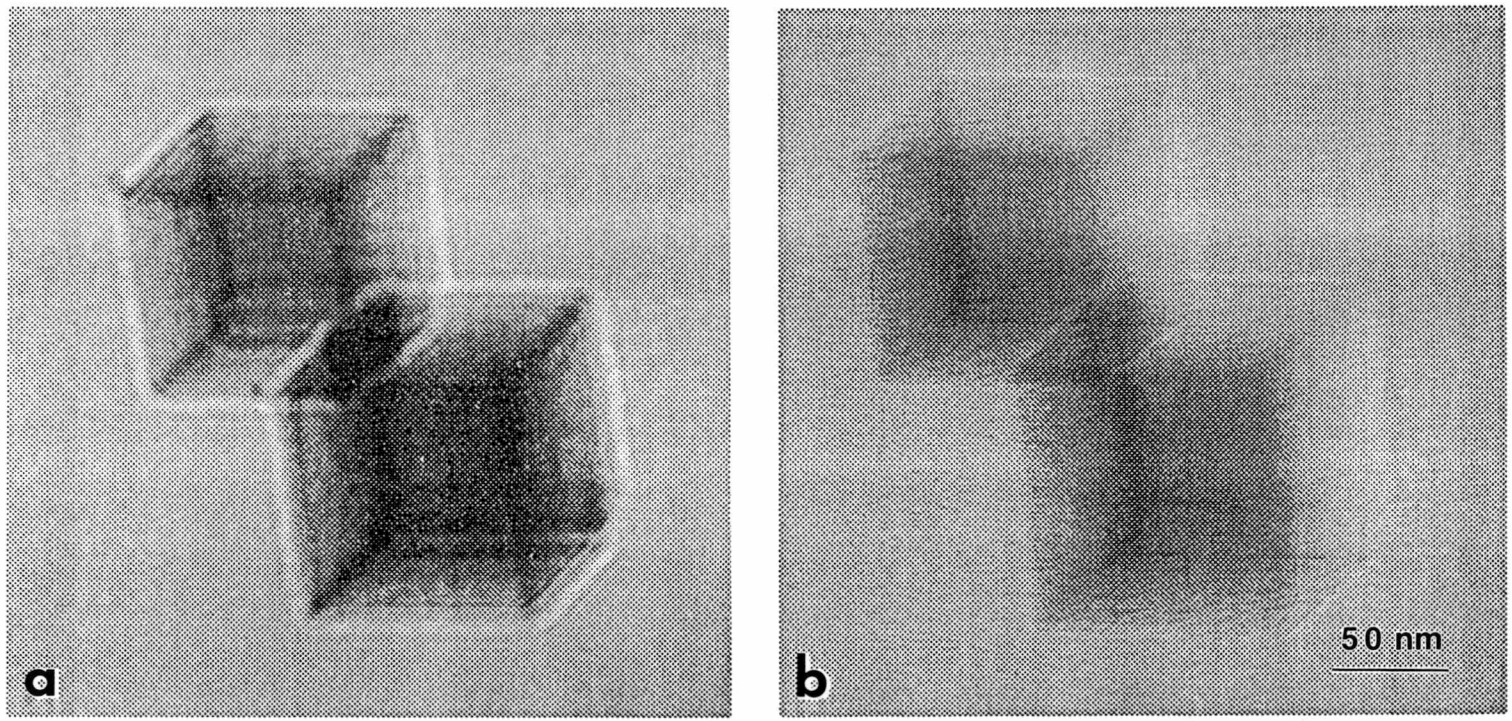


Figure 5.15. (a) Divergent beam Fraunhofer hologram of MgO cubes on a thin carbon film, and (b) In-focus TEM bright field image of the same MgO cubes.

under above conditions. Figure 5.15(b) is an corresponding in-focus image of MgO. The width of the thin edge of the MgO cube is about 15 Å (d_A for eqn. 5.15, a_{max} for eqn. 4.19). The image was taken under 50 μm defocus which satisfies the Fraunhofer condition (eqn. 4.19). D_{zo} calculated from eqn. 5.16 is 25 Å. The clear space W in this case is about 1000 Å (edge length of the cube), D_{zo} calculated from eqn. 5.15 is then 60 μm. Therefore the defocus condition is also satisfied. The calculated resolution of the hologram is about 5 Å (eqn. 4.20) which is in agreement with the resolution estimated from the finest visible Fresnel fringe spacing.

5.4.2 Off-axis electron hologram

The procedure for generating an off-axis electron hologram is similar to that for a conventional TEM micrograph. The area of interest on the specimen must include a hole large enough to let reference beam pass through. The size of the hole must be at least equal to or larger than the spatial coherence length. The specimen is then focused at "Gabor focus" (0.56 times of Scherzer focus) if the hologram is reconstructed at low resolution levels without correction of wave aberration [Lichte, 1991a; 1991b]. At the "Gabor focus" the object information is put into the image phase, which decreases the central autocorrelation intensity at the diffraction plane, and hence reduces in some degree of the strict limitation of choosing S_{rmax} and S_f (Note that for atomic resolution hologram the choice of focus condition is very important and is determined by the degree of collection of maximum information from the hologram [Lichte, 1991b]). Then the biprism is positioned in the direction that fits the

best interest of the operator. The potential on the biprism can then be switched on and increased to produce a fringe spacing about 3 ~ 4 Å wide in order to be seen clearly on the TEM fluorescent screen at the magnification of 300 ~ 500,000x. Next the illumination of the area must be optimized for the best utilization of the beam coherence. The coherent current I(V) available for recording a hologram from a source with a degree of contrast V is [Lichte, 1991a]:

$$I(V) = -B \cdot \ln(V)/k^2 \qquad (5.17)$$

where $k = 1/\lambda$. Thus, the coherent current is completely determined by the brightness of the electron source. Since the total coherent current is invariant under optical imaging, the current density in the hologram plane can only be optimized by a more favorable distribution [Lichte, 1991a]. Forming the normal rotationally symmetric illumination beam spot at the sample is not necessary because a high degree of coherence is only required between points subsequently superimposed by the action of the biprism.

Since the beginning of experiments in off-axis electron holography produced by means of the electrostatic biprism [Möllenstedt & Düker, 1954], people have used a slit aperture in the condensor lens with its main axis parallel to the biprism axis to achieve higher spatial coherence in the direction perpendicular to the biprism axis and still to have reasonable exposure time [Möllenstedt & Wahl, 1968]. Another more common means to achieve an elliptical illumination is using the stigmator control of the condensor lens

[Lauer, 1982; Hanszen, 1986; Lichte; 1991a]. In this way, the beam is shaped, by the condensor stigmators and focus, to produce an illumination patch which is as wide as required in the direction parallel to the biprism but much smaller in the perpendicular direction (Figure 5.16). If the lateral extension (parallel to the biprism axis) is r_L and the extension in the perpendicular direction is r_P, then the resulting interference fringes have the same contrast but with a current density higher by r_L/r_P, as compared with one recorded with a rotational symmetric source with radius r_P [Lichte, 1991a]. During the adjustment of the illumination shape, care should be excised to position the long axis of the elliptical beam exactly parallel to the biprism axis to achieve the best contrast of the interference fringes.

If the beam is adequately coherent then, the potential on the biprism can be increased to produce the desired fringe spacing. Using HF2000 FE TEM, the typical beam current at the object is about 100 pico-amps (at 200 keV) and the hologram can usually be seen directly on the screen when using binoculars. The incident dose at the object is between 1000 and 10,000 electrons/Å^2 which, while no problem for metals and semiconductors, is far in excess of the upper safe dose for polymer and biological tissue [Joy et al, 1992].

Figure 5.17 shows a high resolution hologram recorded from graphite using the condition discussed above. The fringe spacing is 1.2 Å and the exposure time was about 50 seconds. In this case the hologram contains relatively strong amplitude contrast (i.e., effectively the normal bright field image) as well as a strong fringe

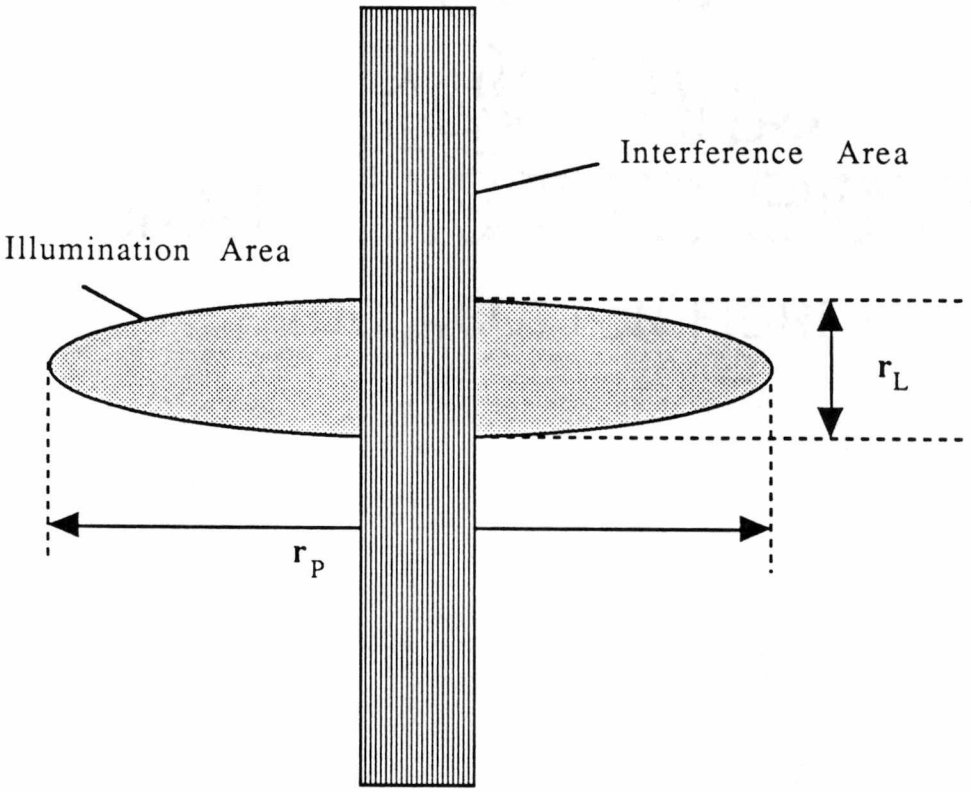


Figure 5.16. Schematic illustration of illumination condition for off-axis electron holography.

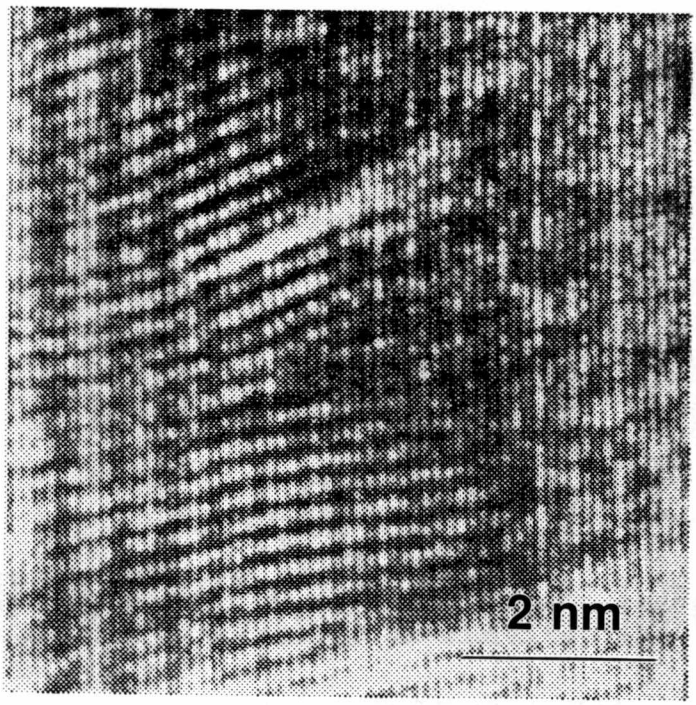


Figure 5.17. Off-axis hologram of graphite taken at 200 keV beam energy.

pattern. In the cases of weak phase specimen the fringe pattern may be the only visible feature

5.5 Recording Electron Holograms

Electron hologram can be recorded by two recording mediums, namely, photographic films and a TV camera or a charge coupled device (CCD) array.

5.5.1 Photographic films

Photographic films are still the most common recording medium in the electron holography community, although the non-linear transformation is their intrinsic problem [Lichte, 1991a]. Because of the limited brightness of electron beam a highly sensitive film is required. After a hologram is recorded on a photographic film, the film is developed and subsequently, either it is put on an optical bench for optical reconstruction. or is digitized by a flat band scanner for subsequent numerical reconstruction.

The shortcomings stem from the fact that electrons are multiply scattered in the emulsion. An ordinary black-and-white emulsion layer of the type used in conventional electron microscopy is simply a suspension of small crystals of silver halide (mostly silver bromide, grain sizes ranging from 0.1 to 1 μm) in gelatin with a typical thickness of 12 μm [Farnell & Flint, 1973]. Each incident electron may pass through perhaps 30 or more silver halide grains before it is stopped completely or pass through the emulsion where it is either absorbed by the supporting material or deflected back into the emulsion. The inelastic scattering events lead to the energy

transfer to the silver halide, which is responsible for the formation of specks of silver atoms in the affected grains. If the aggregate of silver atoms formed is sufficiently large--believed to be between 3 and 6 atoms--the entire grain will be capable of being converted to metallic silver during photographic development [Eastman Kodak, 1978]. Figure 5.18 shows the Monte Carlo simulation from of 200 keV incident electron scattering inside a 12 μm thick silver bromide-gelatin mixture layer (see Chapter Two). The primary and secondary electron trajectories extend several tens of grains, which significantly increases the speed of electron exposure, but considerately worsens the lateral resolution.

modulation transfer function (MTF)

The electron-scattering volume, which defines the emulsion point spread function (PSF), depends on the grain sizes, grain density, incident electron energy, and film developing procedures. The Fourier transformation of this point spread function is the modulation transfer function (MTF) of the emulsion, which is also defined for a given spatial frequency as the contrast in a developed image divided by the contrast in the pattern used to expose the image, measured at that spatial frequency [Downing & Grano, 1982]. The analytical representation of the MTF has been found as

$$MTF(R) = \frac{1}{1 + (R/S_0)^2} \qquad (5.18)$$

where R is the spatial frequency and S_0 is defined as the frequency at which the MTF has a value of FWHM ($S_0 \ln 2/\pi$) [Lichte et al, 1987].

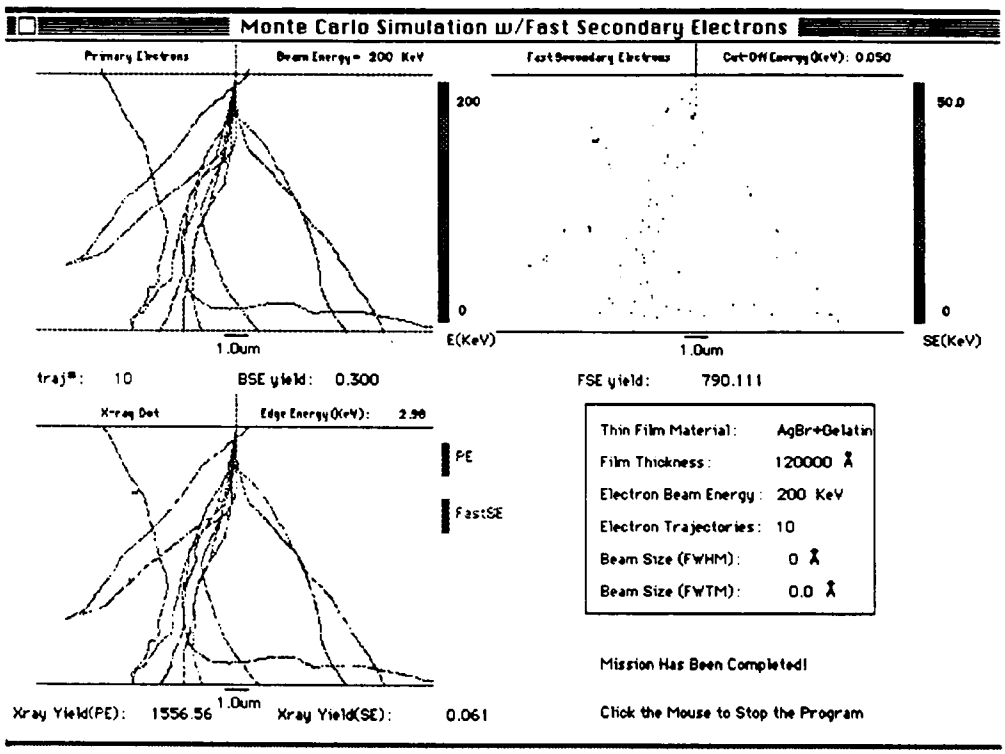


Figure 5.18. Monte Carlo simulation of electron spreading in a photographic emulsion of 12 μm thickness (AgBr in Gelatin) at 200 keV beam energy.

In fact, S_0 is a parameter describing the width of the PSF due to the electron scattering in the emulsion:

$$PSF(x_1) = (\pi/S_0)\cdot\exp(-2\pi|x_1|/S_0) \qquad (5.19)$$

where $x_1 = Mx$ is the coordinate in the image plane, M is the electron optical magnification. In general, emulsions with a wide MTF are less sensitive than emulsions with a narrow MTF. If taking a cosinusoidal electron current distribution $J(\mathbf{r}) = J_0(1 + \cos(2\pi\mathbf{R}_0\mathbf{r}))$, the response of the photographic film is a cosinusoidal distribution of optical density [Lichte, 1991a]:

$$D(r) = D_0\cdot[1 + MTF(\mathbf{R}_0)\cdot\cos(2\pi\mathbf{R}_0\cdot\mathbf{r})] \qquad (5.12)$$

where the contrast of the hologram is decreased by the MTF of the emulsion. Let S_{ph} be the spatial frequency in the hologram $(1/S_f)$. When S_{ph} is chosen to be close to S_0, the contrast is greatly reduced; while for $S_{ph} >> S_0$ the number of electrons per pixel (picture element) must be increased to achieve a sufficient optical density (OD) of the photographic emulsion [Lichte, 1986].

optical density (OD)

Optical density (D) denotes quantitatively the ability of the developed emulsion to attenuate light. It is defined as $D = \log_{10}(1/T)$, where T is the fraction of incident light transmitted by the emulsion (transparency). A value D = 1 means one tenth transmission. Let $E = Jt$ as an exposure ($amps\bullet sec/cm^2$), where J is the current density (also intensity at the surface of the emulsion layer) and t is the exposure time. When a photographic emulsion is exposed to the

electron, at the initial stage one or more grains can be rendered developable by each incident electron, which results in a linear relation between D and E since D is proportional to the number of developed grains per unit area, and E is proportional to the number of electrons incident per unit area. With further increase of exposure, the probability of repeatedly hitting the same grain increase, which induces the non-linearlity between D and E. The non-linear relation is given by [Farnell & Flint, 1973]

$$D = D_s \cdot (1 - e^{-CJt}) = D_s \cdot (1 - e^{-CE}) \qquad (5.21)$$

where D_s is the saturation density, C is the speed of the emulsion and depends on the size and sensitivity of the grains. If $D << D_s$ then $D = C D_s E$, i.e., the linear relation exhibited in the initial portion of the D - E curve.

The slope γ of D - $\log_{10}(E)$ curve is loosely referred to as the contrast of the emulsion [Valentine, 1966]. Since $\gamma = dD/d(\log_{10}E) = 2.3E\ dD/dE$, at the initial stage $\gamma = 2.3D$, therefore the contrast of an emulsion exposed to electrons depends only on the density and when exposed to the same density, all emulsions will have the same contrast. In more general case, the contrast γ is given by

$$\gamma = 2.3(D_s - D) \cdot \ln \frac{D_s}{D_s - D} \qquad (5.22)$$

Differentiating eqn. 5.22 one can find the maximum contrast value of an emulsion is

$$\gamma_{max} = (2.3/e)D_s = 0.85\ D_s \quad \text{at } D = \frac{e-1}{e} D_s = 0.65\ D_s \qquad (5.23)$$

Unfortunately, the γ_{max} is located at the non-linear part of the D - E curve (linear relation is usually valid at the condition $D < D_s/4$). The intensity transparency of the developed emulsion can be defined as [Lichte, 1991a]

$$T = 10 - D \tag{5.24}$$

and the amplitude transparency τ is defined as

$$\tau = 10^{D/2} \tag{5.25}$$

non-linearity transformation

In eqn. 4.24, one of the sidebands (+1) is given by

$$F(\mathbf{r}) = FT[U_o(\mathbf{r})] \exp[i(\chi(\mathbf{R})]*\delta(\mathbf{R}+\mathbf{R}_c) \tag{5.26}$$

where $F(\mathbf{r})$ is the Fourier transform of true image current density $J(\mathbf{r})$. If D - E has a linear relation, assuming a pure weak phase object, one should observe

$$|F(\mathbf{r})F(\mathbf{r})^*| = |FT[U_o(\mathbf{r})]FT[U_o(\mathbf{r})]^*| \tag{5.27}$$

When the film is numerically digitized, the digitized version of current density $J(\mathbf{r})$ is $P(\mathbf{r})$. The digitized pixel value distribution $P(\mathbf{r})$ can be expressed as [Völkl, 1991]

$$P(\mathbf{r}) = a\cdot\exp[-D(\mathbf{r})\cdot\ln 10] = a\cdot\exp\{-D_s\cdot\ln 10\cdot[1-\exp(-CtJ(\mathbf{r}))]\} \tag{5.28}$$

where a is a constant to account for the illumination intensity. Because of the non-linear relation between $J(\mathbf{r})$ and $P(\mathbf{r})$, eqn. 5.27, in

general, is not true,. so in order to get precise reconstruction of atomic image wave, the non-linearity has to be corrected.

For laser optical reconstruction the amplitude transparency is decisive. The non-linearity of photographic film sometimes leads to a falsification of the amplitude of the reconstructed wave [Ade, 1982]. Although several attempts have been made [Tokarski, 1968; Downing & Grano, 1982], so far the compensation of non-linearity is very difficult to achieve in light optics [Lichte, 1991a].

For numerical reconstruction, the problem can be solved by measuring the intensity transparency (T) using a densitometer or by measuring the manufacturer supported gray wage using a scanner. The parameter D_s and C are then determined by fitting the measured density data. Knowing the non-linear relation between D and E, one can now calculate the true current density J of the hologram by

$$J(\mathbf{r}) = \frac{-1}{Ct} \cdot \ln\left(1 + \frac{1}{D_s} \cdot \log\frac{P(\mathbf{r})}{a}\right) \qquad (5.29)$$

An example of successful correction [Völkl, 1991] is shown in Figure 5.19.

emulsion noise

Emulsion noise is the random statistical deviations of the optical density about the mean. It is a signal-independent noise which includes shot noise (the statistical variation in the number of electrons incident within segments of a uniformly exposed area), fog (grains developed without having been struck by electrons), and variations in the number of grains made developable by each

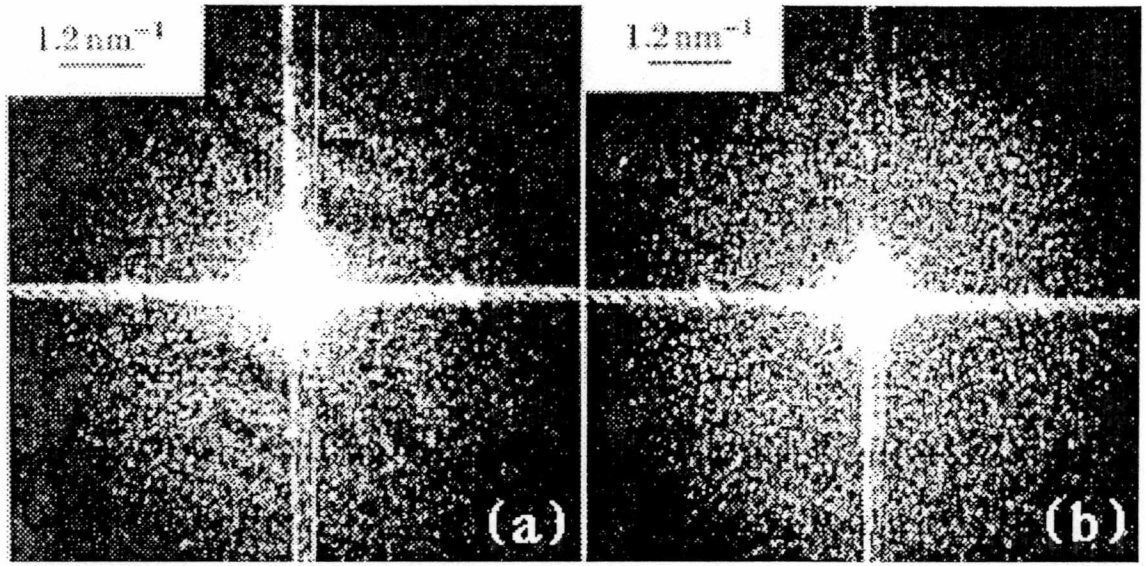


Figure 5.19. Comparison of the absolute square of the side bands of a hologram of carbon film; left before and right after the optical density correction (adopted from [Völkl, 1991]).

electron [Downing & Grano, 1982]. Deconvolution of the noise from the optical density is essential for evaluation of the dependence of the transfer function on exposure. The amplitude of the Fourier transform of an area of an emulsion uniformly exposed to electrons is given by [Downing & Grano, 1982]:

$$n(R) = (a + b\sqrt{E})\cdot\exp(mR) \qquad (5.30)$$

where R is the spatial frequency, a, b, and m are experimentally determined constant, and E is the exposure. The effects of n(s) on an amplitude of Fourier transform is then given by [Hayward & Stroud, 1981]

$$P(v) = \frac{\pi}{2[n(R)]^2}\cdot v\cdot\exp\left[-\frac{\pi}{4}\left(\frac{v}{n(R)}\right)^2\right] \qquad (5.31)$$

where P(v) is the probability of the transform having the value v. The constants of eqn. 5.30 is provided by Hayward for Kodak AA X-ray film (a = 540, b = 4600, m = -1.3, and S_0 = 0.042 μm^{-1}) and Kodak 4489 EM film (a = 380, b = 400, m = -0.2, and S_0 = 0.082 μm^{-1}).

choice of photographic films

Two different kind of Kodak photographic films have been used in the holographic experiments: the Kodak AA X-ray film which has double side emulsions and the Kodak 4489 electron microscopy (EM) film. The calculated MTFs of these two film are shown in Figure 5.20. The Kodak 4489 EM film seems to be a better one, with an MTF better than 0.4 down to a period of 10 μm which corresponds to a 0.1 Å fringe spacing at the object plane recorded at a magnification

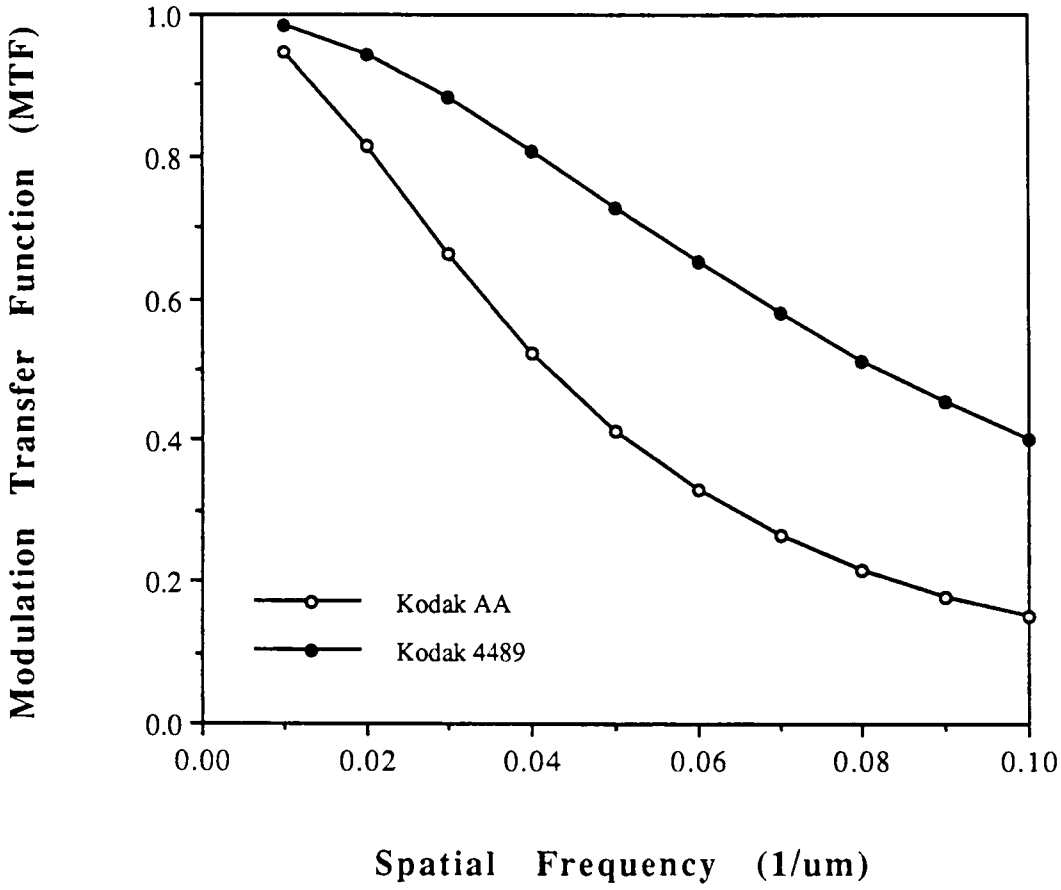


Figure 5.20. Modulation transfer functions of two Kodak films versus spatial frequency.

of 1000,000 times. The MTF of the Kodak AA film is only 0.15 at the same conditions. The average noise levels of these two films are also calculated and the results are shown in Figure 5.21. The results are also in favour of the Kodak 4489 film, the noise of which is 10 times lower than that of the Kodak AA film. Therefore, the fog (OD measured on unexposed film which was developed and fixed, minus the OD measured on film which was fixed but not developed) of the Kodak 4489 film is about 50% lower than that of the Kodak AA film [Downing & Grano, 1982].

The Kodak AA film has a narrow MTF due to the thick emulsion and thus is more sensitive than the Kodak 4489 film. The speed (reciprocal of the electron exposure in $e/\mu m^2$ required to produce an OD of one above the fog) of the Kodak AA film is 4.2 while the speed of the Kodak 4489 film is only 0.35, which leads to the conclusion that the Kodak AA film is 12 times faster than the Kodak 4489 film. This is significant when the limited brightness, microscope stability, and specimen drift are concerned. If one recorded 0.3 Å fringe spacing on the film at the magnification of 1000,000 times, the MTF is 0.62 for the Kodak AA film and 0.86 for the Kodak 4489 film which is about 30% higher.

Based on the above information and our experimental experience, when enough brightness is available, the Kodak 4489 film is certainly a better choice; while if one considers the limited brightness, microscope stability, and specimen drift, the best choice becomes the Kodak AA film. For low magnification and medium resolution electron holography and for a specimen that can afford a

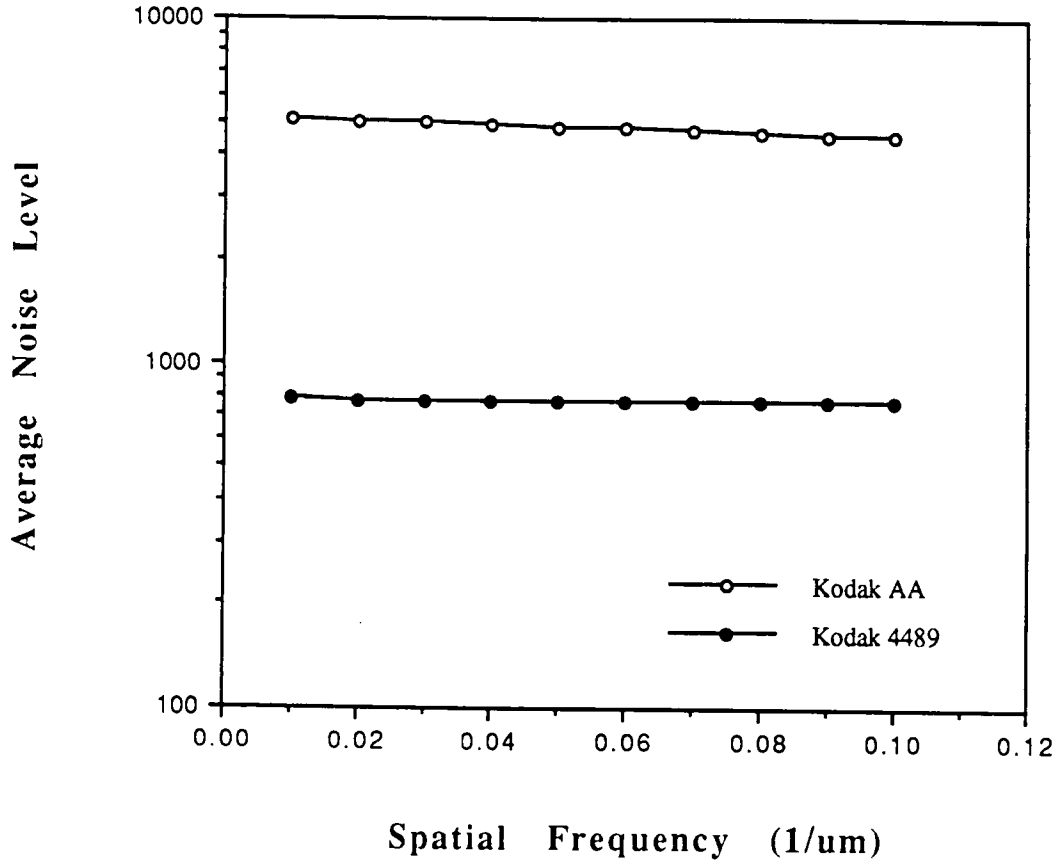


Figure 5.21. Average noise levels of two Kodak films versus spatial frequency.

high current density, we have routinely used the Kodak 4489 film. For the work requiring high magnification, high resolution, quick exposure, and low dose, we have used the Kodak AA film and the results are very satisfied. It also be reported that AGFA Scientia 23D56 emulsion is another good choice (speed is 1.1) [Lichte, 1991a] but it was not available to us at the time.

5.5.2 Digital recording

Another way to directly record the hologram is to use a TV camera, or a CCD array. It would provide a sensitive approaching that of film while providing a direct and very linear readout of the hologram. However, the number of carrier fringes in the hologram is limited due to the number of scanning lines for a regular TV camera. The more favorable digital recording system is a TV camera with a large number of CCD arrays.

An electronic device like this will be especially invaluable for real-time computer reconstruction although at present there is not enough pixel numbers (at least 2048x2048 pixels for a final reconstructed image with 512x512 pixels) for a complete information extraction. The current available Gatan slow-scan CCD camera (1kx1k arrays) is good enough for normal high resolution electron microscopy, but it is not enough for high resolution electron holography [Allard et al, 1992]. Further development is expected for a 2kx2k CCD system.

5.6 Reconstruction of Electron Holograms

Holography is a two-step image process. After the generation of an electron hologram, the hologram must then either be interpreted directly (sometimes in off-axis holography) or reconstructed to produce an image which can be interpreted.

5.6.1 Reconstruction of in-Line Fraunhofer electron hologram

So far all reconstruction work of in-line hologram is done by optical means [Tonomura, 1987]. Optical reconstruction can be carried out using a divergent He-Ne laser beam (λ' = 6328 Å) with a ray path analogous to that of the original TEM (Figure 5.22), and the reconstructed image is recorded directly onto Polaroid films. Figure 5.23 shows the reconstructed image of MgO crystals. The initial hologram (Figure 5.15(a)) is taken in a defocus of 50 μm at 150,000 times magnification with a 3 micro-rads illumination angle. and the image is found at Z = 40 cm away from the hologram. The image shows excellent resolution, detail, and contrast which is significant higher than that of the in-focus image (Figure 5.15(b)). Note that the reconstructed image is a photographic negative compared to the original image. A faint background of coherent noise from the reconstruction optics is visible across the image but does not constitute a significant problem. The twin image is almost invisible. The resolution d of reconstructed image is estimated to be better than 7 Å as judged from the finest visible fringe spacing [Tonomura, 1986], which is agree with the calculated resolution d' (eqn. 4.20).

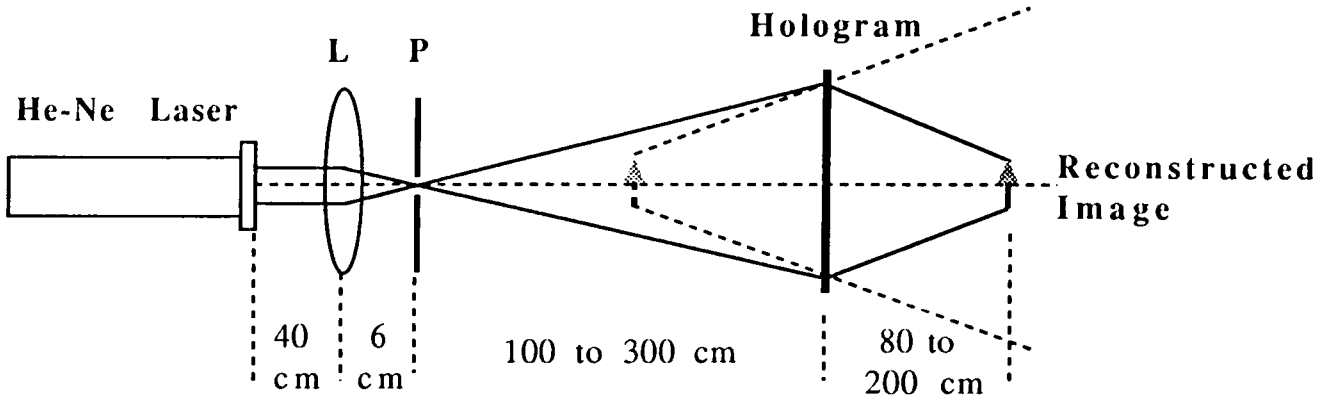


Figure 5.22. Schematic illustration of optical reconstruction of in-line divergent beam Fraunhofer hologram.

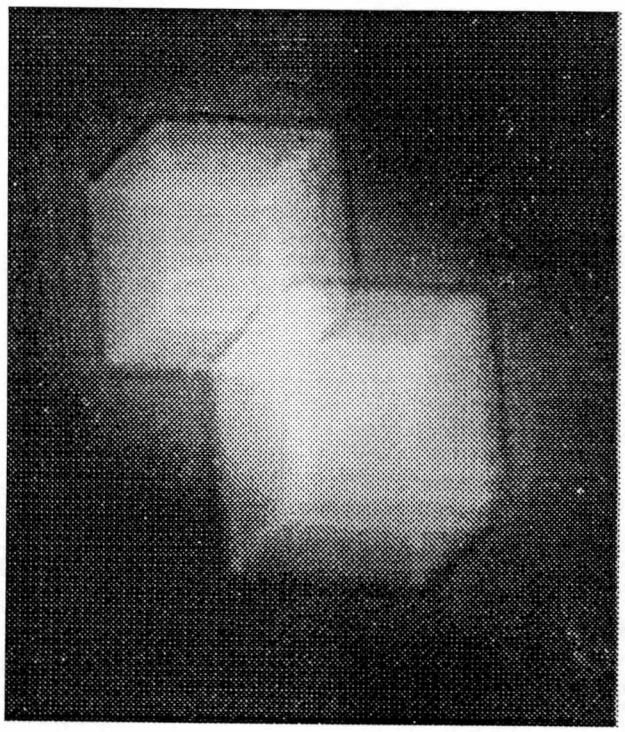


Figure 5.23. Optical reconstructed image of the MgO cubes from Figure 5.15(a).

5.6.2 Reconstruction of off-axis electron hologram

In a few cases, especially when only the phase information is concerned, useful information can be obtained without reconstruction (see Chapter Six). The direct interpretation of a hologram is based on the fact that fringe shifting on holograms is the direct result of the local phase change. By measuring the degree of the fringe shift one can directly obtain the quantitative phase information from the hologram [Zhang et al, 1992]. In most cases, however, one needs to reconstruct the hologram to obtain amplitude and phase information at high resolution. There are two ways to reconstruct an hologram: optically or numerically.

optical reconstruction

All early work was carried out by optical means using systems in which the original electron-optical wavefronts are replaced by optical waves, and the specimen is replaced by the hologram [Tonomura, 1987]. An experimental arrangement is shown in Figure 5.24. An electron hologram is illuminated with a collimated laser beam and the resultant light distribution in the hologram plane is Fourier transformed by means of a optical lens. The desired diffraction orders are selected with an aid of a diaphragm, and the corresponding reconstructed image is recorded in the image plane of the hologram.

Although optical reconstruction methods are easy to apply and visually instructive, and are still a valuable tool for use in electron holography, the limited flexibility and lack of quantification makes them inferior to numerical reconstruction.

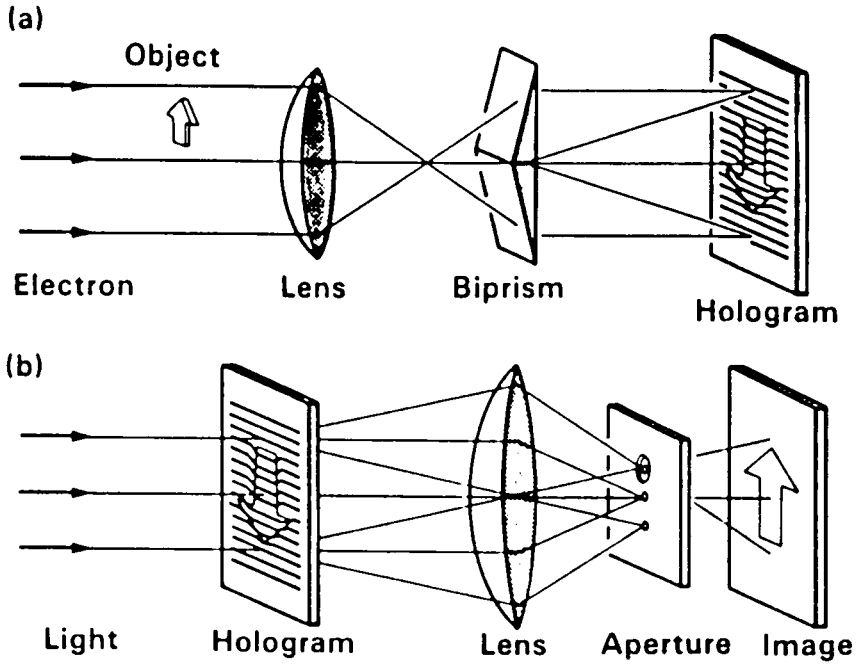


Figure 5.24. Schematic illustration of optical reconstruction of off-axis hologram (adopted from [Tonomura, 1986]).

numerical reconstruction

For numerical reconstruction, the hologram is first digitized, either from a photographic film or from a direct CCD array readout. The optical density of each pixel is then fed into a computer. All reconstruction steps, forward Fourier transform, sideband selection and spectrum re-centering, and inverse Fourier transform including correction of spherical aberration, are done numerically. The resultant numerical image wave is stored in the computer as a two-dimensional array of complex numbers, real part of which representing amplitude and imaginary part representing phase. Then, the images of amplitude and phase of the reconstructed objective wave can be displayed on the computer monitor with 256 gray levels. The subsequent operations, such as phase amplification and interferogram, are also done numerically by re-scaling the phase and adding a reference wave, respectively.

The advantages of numerical reconstruction stem from the fact that all the information is stored in individual pixel with high numerical quantification and great flexibility. The non-linearity of photographic film can be corrected before the numerical operation [Völkl, 1991; Lichte, 1991a], which is difficult to do by optical means. Manipulation of image data in both image plane and Fourier spaces is conveniently available for all pixels and various image processing operations are performed easily. Any intermediate results are given quantitatively.

When one digitizes a hologram the total number of fringes has to be considered. Because the sideband has to be cut during the

reconstruction, the reconstructed pixel number is considerably smaller than the original hologram size. If one starts at 1024x1024 pixel density of original hologram, after reconstruction the reconstructed image has only 256x256 pixel density. The minimum required sampling rate per hologram fringe is 4 [Lenz & Völkl, 1990]. If a total number of pixels of digitized image is N, then the image field is $(NS_f/4)$. The maximum spatial frequency that can be processed for a certain number of pixels is [Lichte, 1992]:

$$R_{max} \leq [(N/24)\cdot k^3/C_s]^{1/4} \qquad (5.33)$$

where k is the wave number $(1/\lambda)$. The total number of fringes is therefore determined by both the image field and the maximum spatial frequency. The calculated results are shown in Figure 5.12. Note that for reconstruction of a 1 Å resolution, a total of 4554 pixels are needed which is beyond the capability of any current available commercial CCD array.

Procedures of numerical reconstruction is following:

(1) Forward Fourier transform of the hologram in two dimensions. The resultant Fourier spectrum is represented by eqn. 4.24.

(2) Select a sideband from the spectrum. As discussed in section 5.3.3, if the carrier spatial frequency $\mathbf{R}_c$ is chosen at the condition of $|\mathbf{R}_c| > 3|\mathbf{R}_{max}|$, then the three components, zero order + autocorrelation, left and right sidebands, are completely separated. Either one of the sideband can be chosen using a numerical aperture of width of $2|\mathbf{R}_{max}|$. The selected sideband is then shift to the origin of

the Fourier space by an amount $\delta(\mathbf{R} \pm \mathbf{R}_c)$ depending on which sideband being chosen.

(3) Inverse Fourier transform of the selected sideband in two dimensions. The results is a complex array, the real part of which is the "amplitude" image and the imaginary part is the "phase" image. If correction of spherical aberration of the object lens is required [Franke et al, 1988], then the sideband is multiplied by the factor $\exp[-i\chi(\mathbf{r})]$ being re-transformed to give the desired object wave $U_o(\mathbf{r})$.

Figure 5.25 shows the result of forward Fourier transforming the hologram of Figure 5.17. The figure shows a gray map of the power spectrum of the transform (scaled logarithmically to make the detail easier to see). The feature of the transform computed in eqn. 4.24, the zero beam and autocorrelation function, and the two sidebands are clearly visible. One of the sideband can be chosen, re-centered at the region of the coordination system, and then inversely transformed. The diameter of the sideband region that is to be selected must be chosen carefully. If the region is too large then it will interfere with the central carrier and autocorrelation function zone; if the region is too small then the quality of the reconstruction will suffer because of the loss of high spatial frequency information.

A computer program has been developed [Zhang et al, 1992; Joy et al, 1992b] to implement the above procedures with a Macintosh II computer. The program is written in C running under the Macintosh Programmers Workshop (MPW) system. After numerical reconstruction, a 256-gray-level map of the real part of

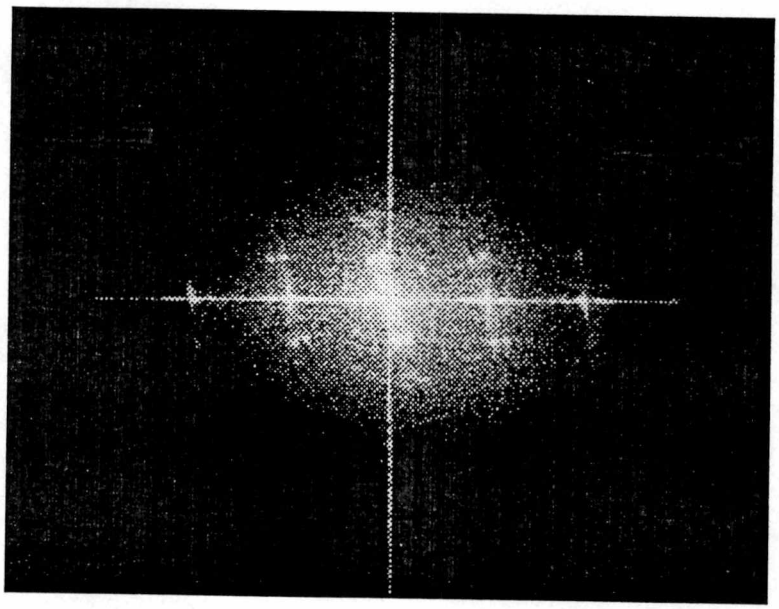


Figure 5.25. FFT spectrum (in log scale) of the graphite hologram from Figure 5.17.

the complex array is displayed as the amplitude image on the monitor. A gray map formed from the imaginary part of the array by displaying the phase shift (0 to 2π radians) in different shades of gray comprises the reconstructed phase image from the hologram.

Figure 5.26 shows the reconstructed amplitude image (a) and the phase image (b) of the graphite hologram (Figure 5.17) using the developed computer program. Graphite is a strong amplitude object and the reconstructed amplitude image has a very high contrast; while the phase image reveals some details of the lattice fringes but the contrast is low. Fresnel fringes from the biprism edges are clearly visible in both images suggested that the spatial frequency of the carrier is still too low which makes center part of the spectrum and the sideband overlap in some degree.

Electrons will change their phases due to the interaction of image wave and reference wave. In some cases, the electron phase shift is less than a wavelength [Tonomura, 1987], which makes the observation almost impossible. The problem can be solved by further enhancing the phase image by changing the plotting methods. Instead of allowing the phase changes 0 to 2π to have the image brightness change from black and white, the range 0 to $2\pi/n$--where n is 2, 4, 8 etc--is used instead [Joy et al, 1992a]. If the phase shift exceeds the value $2\pi/n$, then the brightness scale simply repeats. This phase amplification procedure then produces contours of constant phase on the image which could map out the projected phase distribution of the sample. By making n large the phase change between the contours are reduced giving a mode detailed map.

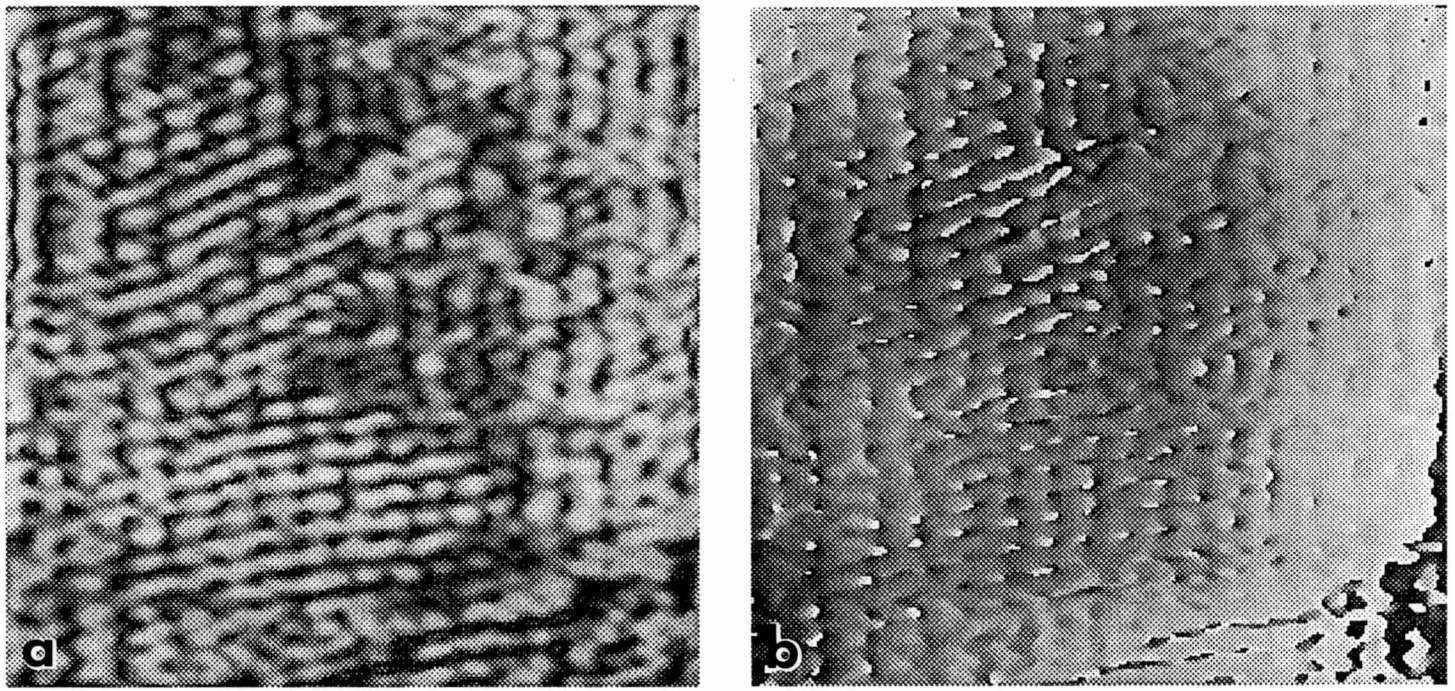


Figure 5.26. Numerically reconstructed images of the hologram from Figure 5.17. (a) Amplitude image, and (b) phase image.

Figure 5.27 shows a hologram of a single crystal MgO cube. The fringe bendings are clearly visible in the vicinity of the edge of the cube due to the phase shift. The reconstructed amplitude image and the phase image using a spatial frequency of 1.04 nm^{-1} are shown in Figure 5.28 (a) and (b), respectively. The reconstructed phase image has a higher contrast than the amplitude image since MgO crystal is a weak amplitude object [Harada & Shimizu, 1991], and in both images the Fresnel fringes are almost invisible. If one uses a spatial frequency of 1.56 nm^{-1} to reconstruct the image, Fresnel fringes are inevitably showing up in the final reconstructed images (Figure 5.29) and the image quality is degradated. This is due to the selection of a large sideband which overlaps with the center part of the spectrum. A three-dimensional plot of the phase distribution of the MgO cube visualized as the wave front is shown in Figure 5.30. The projected phase shift is directly proportional to the thickness change of the MgO cube, therefore, Figure 5.30 displays the 3-D profile of the thickness map of the MgO cube. The phase image with 8 times phase amplification is shown in Figure 5.31 which reveals more detail thickness changes on the edge of the MgO cube. The phase change between two adjacent contours is $\pi/4$ which corresponds to about 90 Å thickness change.

5.6.3 Electron interferometry

The reconstructed phase image can be further utilized to generate an interferogram. An interferogram is an interference pattern between a reconstructed phase image and a plane reference wave. The re-interference generates a contour map which enhances

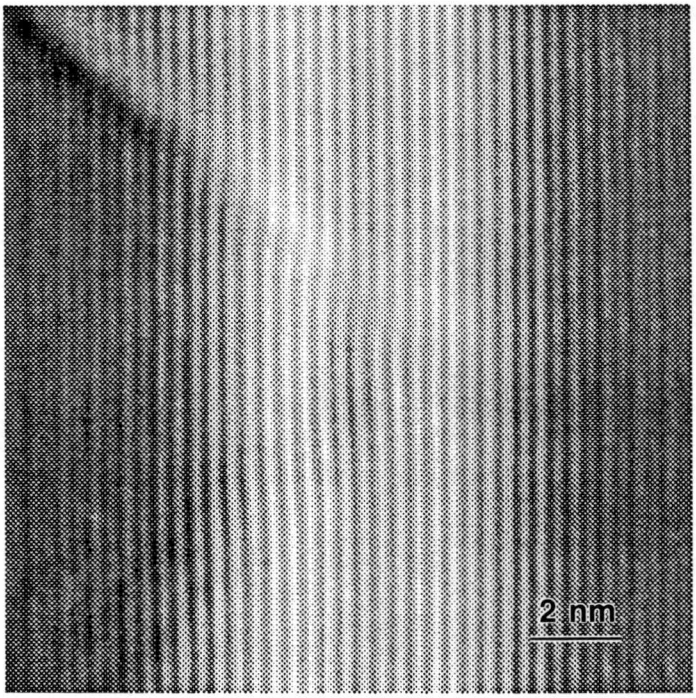


Figure 5.27. Off-axis hologram of a MgO cube. Only a corner of the cube is emerged in the interference area.

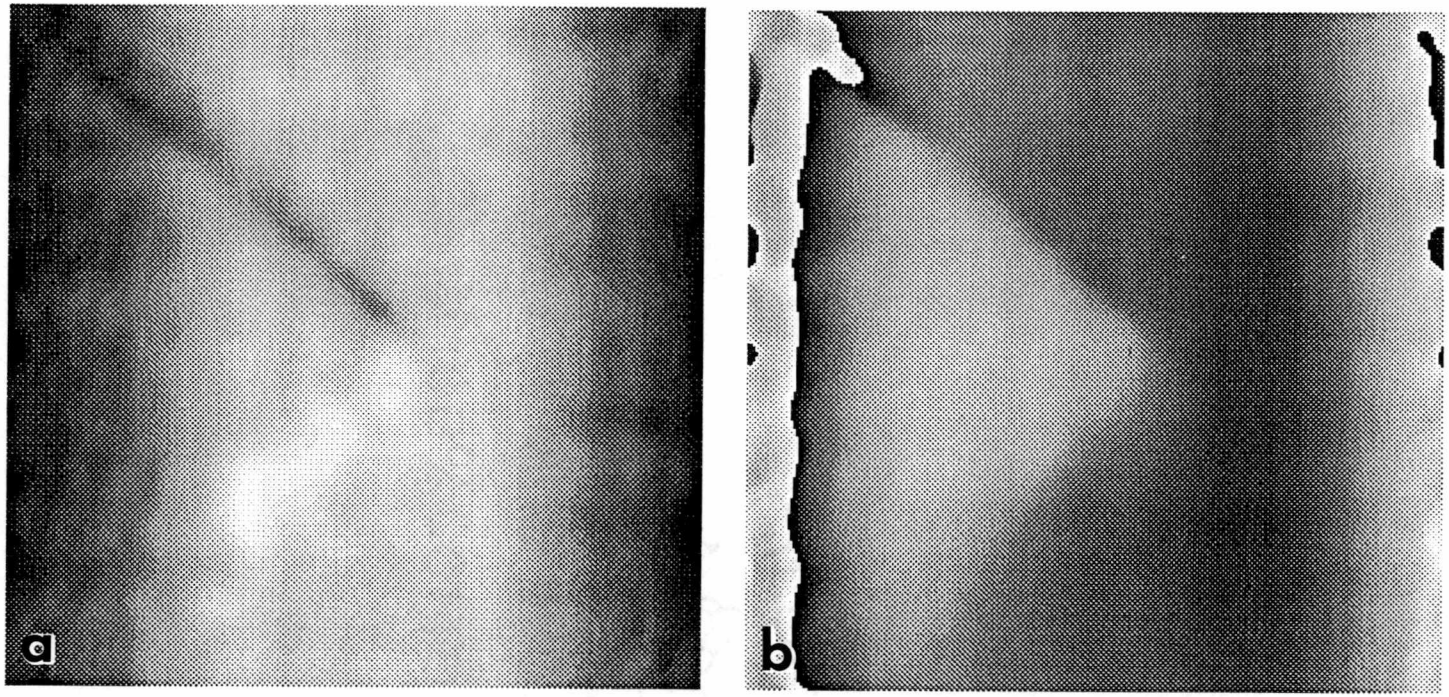


Figure 5.28. Numerically reconstructed images of the hologram from Figure 5.27 with a spatial frequency of 1.04 nm^{-1}. (a) Amplitude image, and (b) phase image.

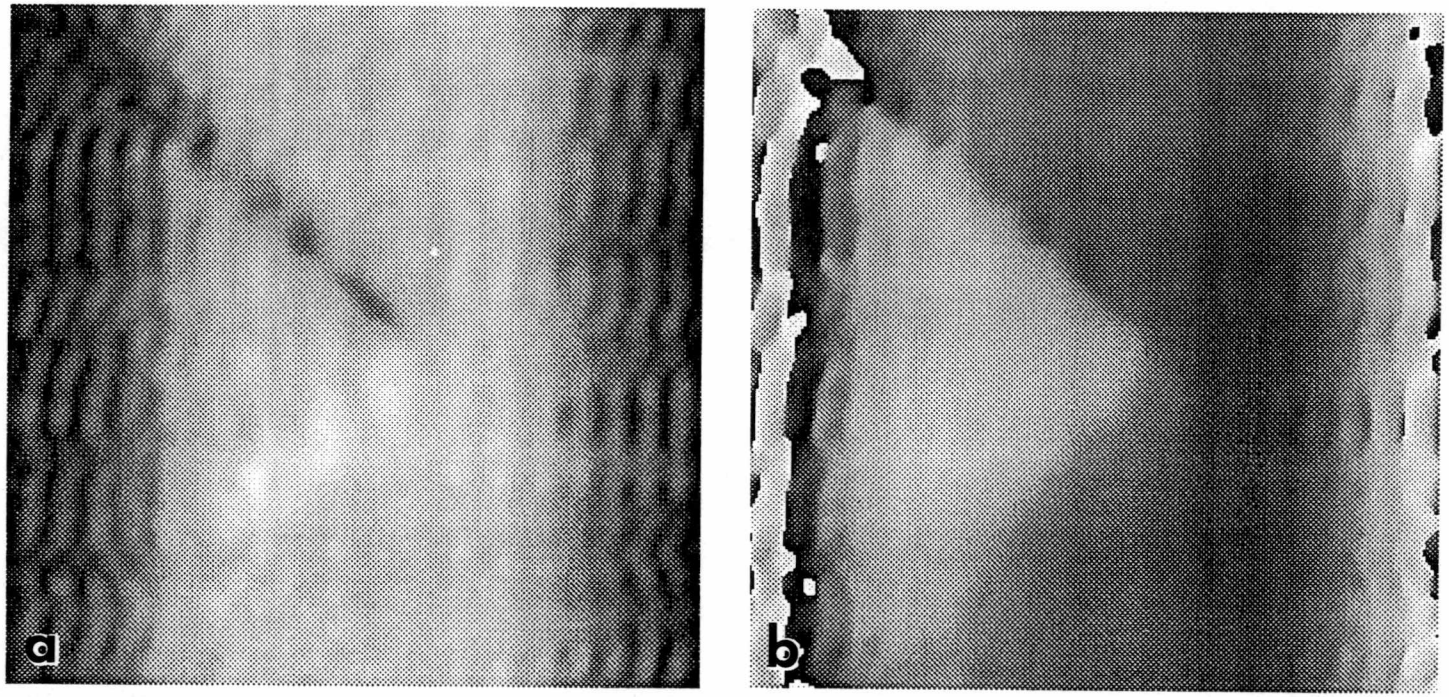


Figure 5.29. Reconstructed image from the same hologram (Figure 5.27) but with a spatial frequency of 1.56 nm^{-1}.

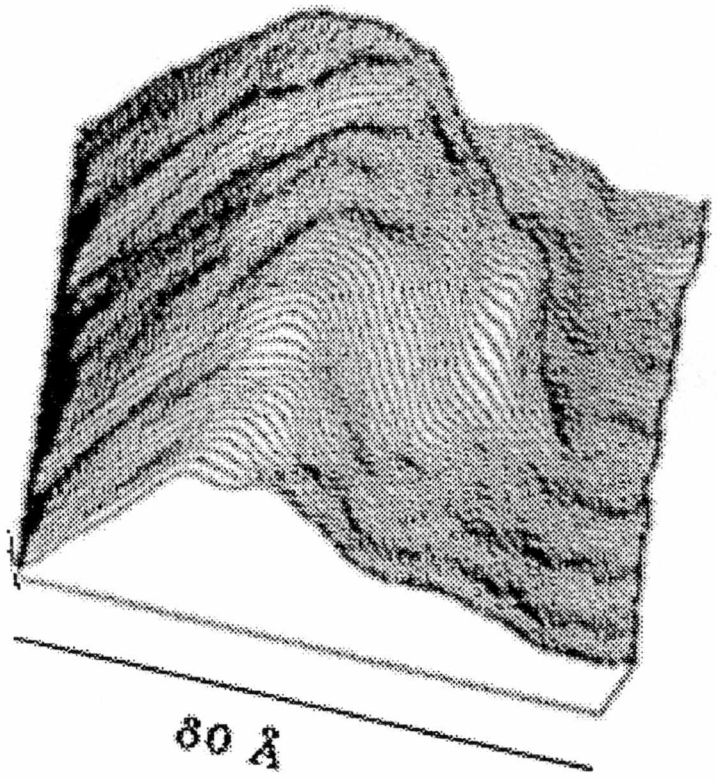


Figure 5.30. Phase distribution from Figure 5.28(b) visualized as the wave front.

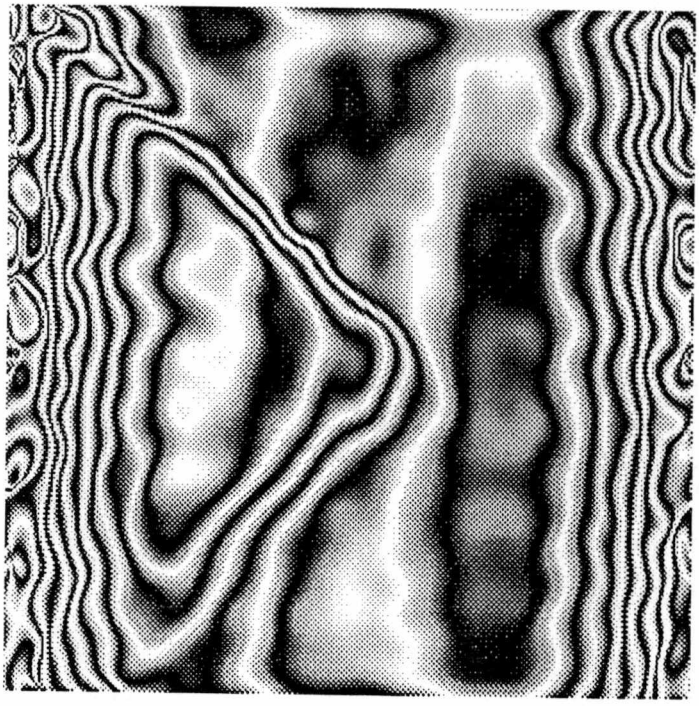


Figure 5.31. 8-times amplified phase image of the MgO cube.

the visibility of the phase change of the specimen. If the reconstructed phase image wave is $\Psi_o = \exp[i\Phi(\mathbf{r})]$, and a reference wave $\Psi_R = \exp[i2\pi\mathbf{R}_c\mathbf{r}]$, where $\mathbf{R}_c$ is the carrier frequency and $|\mathbf{R}_c| = \theta/\lambda$, $\mathbf{R}_c\mathbf{r} = (1/\lambda)(x\cos\theta, y\sin\theta)$, θ is the azimuth angle of inclination between reference wave and reconstructed phase image wave, x and y are the coordinates in the image plane. Consequently, the resultant intensity distribution of the interferogram is

$$I_{int} = |\Psi_o + \Psi_R|^2 = 2 + 2\cos[2\pi\cdot\mathbf{R}_c\cdot\mathbf{r} - \Phi(\mathbf{r})] \qquad (5.34)$$

The contour line in the interferogram has a constant contrast and the position and the orientation are modulated by both the azimuth angle θ and the local phase shift $\Phi(\mathbf{r})$.

If the reference wave is parallel with the reconstructed phase image (i.e., $\theta = 0$), the resultant interferogram is similar to the phase amplification contour except the phase change between the contour is still 2π. This kind of interferogram in sometimes lacks some wavefront information as shown in Figure 5.32(a). One cannot distinguish that contour map of a mountain from that of a valley. The alternative way to obtain this information is to tilt the reference beam with an azimuth angle θ [Tonomura, 1992] as shown in Figure 5.32(b).

Comparing eqn. 5.34 with eqn. 4.22, the right hand side of eqn. 5.34 is almost the same as the hologram excepts the amplitude modulation is eliminated. It means that an interferogram provides more clear phase information than a hologram, and it facilitates the visual interpretation and extraction of the desired information of the

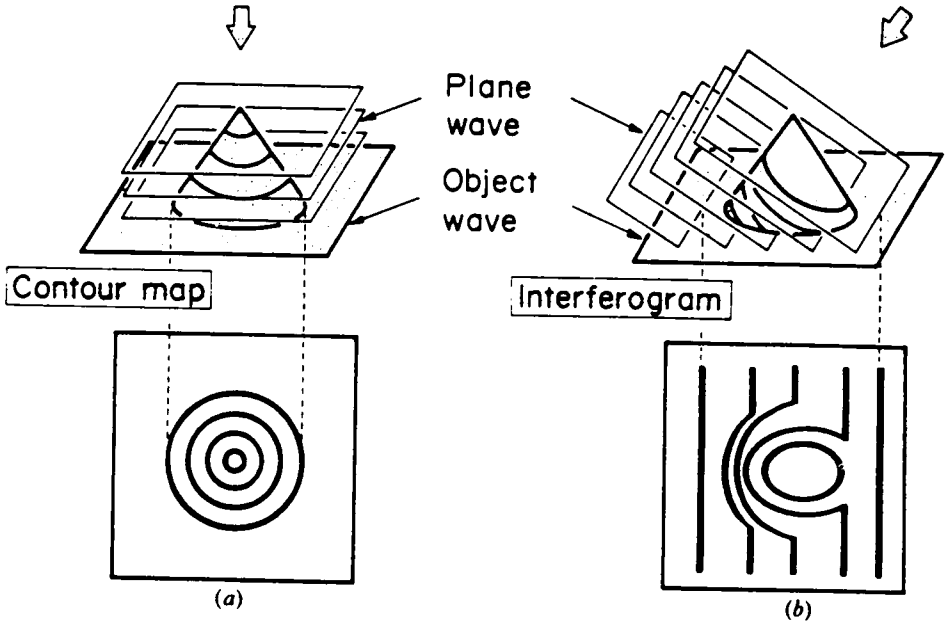


Figure 5.32. Two kinds of interferograms: (a) reference wave is parallel to reconstructed image wave (azimuth angle between reference wave and reconstructed phase image wave is zero), and (b) tilted reference wave (adopted from [Tonomura, 1992]).



Figure 5.33. Holographic interferograms from the reconstructed image image (Figure 5.28(b)) with different azimuth angles.

object. Beyond that, an interferogram and a hologram are basically identical.

Figure 5.33 shows a set of interferograms with different azimuth angles of inclination of reference wave. Fringes of different orientation show up the phase modulation, which enhances the 3-D vision of the MgO cube.

REFERENCES

Ade, G., Optik, **62**(1982), p. 67

Agar, A.W., Alderson, R.H., and Chescoe, D., Principle and Practice of Electron Microscope Operation, (North-Holland Pub. Co., 1974), p. 24

Allard, L.F., Nolan, T.A., Joy, D.C., Hashimoto, T., Zhang, X., and Zhang, Y., submitted to Ultramicros., 1992

Born, M. and Wolf, E., Princ. of Optics, (Pergamon Press, 1959)

Chen, J., Opt. Acta, **33**(1986)6, p. 733-744

Downing, K.H. and Grano, D.A., Ultramicros., **7**(1982), p. 381

Drahos, V. and Delong, A., Opt. Acta, **33**(1986), p. 173

Eastman Kodak Co., Kodak Electron Image Plates, P-116, 1978

Farnell, G.C. and Flint, R.B., J. Micros., **97**(1973), p. 271

Franke, F.J., Herrman, K.H., and Lichte, H., Scanning Micros. **Suppl.2**(1988), p. 59

Frieser, H., Klein, E., and Zeitler, E., Z. angew. Phys., **11**(1959), p. 190

Gabor, D., Proc. 1st Int. Cong. Electron Micros., Paris, 1951, p. 9

Hanszen, K.J., J. Phys., **D19**(1986), p. 373

Harada, K. and R. Shimizu, J. Electron Microsc., **40**(1991), p. 92

Hayward, S.B. and Stroud, R.M., J. Mol. Biol., **151**(1981), p. 491

Hibi, T. and Takahashi, S., Z. angew. Phys., **27**(1969)3, p. 132

Joy, D.C., Proc. 49th Annual Meeting of EMSA, San Jose, (1991), p. 700

Joy, D.C., Zhang, X., Zhang, Y., Hashimoto, T., Bunn, R., Allard, L., and Nolan, T.A., submitted to Ultramicros., 1992a

Joy, D.C., Zhang, Y, Bunn, R., to be submitted to Ultramicros, 1992b

Kasper, E., in Adv. Opt. Elect. Microsc., **8**(1980), p. 207

Lauer, R., Proc. 10th Int. Cong. Electron Micros., Hamburg, **1**(1982), p. 427

Lenz, F. and Völkl, E., Proc. 12th Int. Cong. Electro. Micros., **1**(1990), p. 228

Lichte, H., Adv. Opt. Elect., **12**(1991a), p. 25

Lichte, H., Ultramicro., **38**(1991b), p. 13

Lichte, H., submitted to Ultramicros., 1992

Lichte, H., Herrmann, K.-H., and Lenz, F., Optik, **77**(1987), p. 135

Lichte, H. and Völkl, E., Inst. Phys. Conf. Ser. No. 93: **1**(1988), p. 191

Möllenstedt G. and Düker, H., Naturwiss, **42**(1954), p.41

Möllenstedt, G. and H. Wahl, Naturwissenschaften, **55**(1968), p. 340

Strong, J., Neher, H.V., Whitford, A.E., Cartwright, C.H., and Hayward, R., Proc. in Exp. Phys., (Prentice Hall: New York, 1938), p. 188

Tokarski, J.M.J., Appl. Optics, **7**(1968), p. 989

Tonomura, A., Prog. in Optics, XXIII, (Elsevier Sci. Pub., 1986), p. 185

Tonomura, A., Rev. Mod. Phys., **59**(1992)3, p. 639

Valentine, R.C., in Adv. Opt. Electron Micros., **1**(1966), p. 180

Völkl, E., Ph.D. Dissertation, Eberhard-Karls-Universität, Tübingen, 1991

Zhang, X., Hashimoto, T., and Joy, D.C., Appl. Phys. Lett., **60**(1992), p. 784

Zhang, Y, Bunn, R., Joy, D.C., Zhang, X., Allard, L.F., and Nolan, T.A., presented at 4th Conf. Frontiers of Electro. Micros. in Mater. Sci., Oakland, CA, April 21-24, 1992

CHAPTER SIX

APPLICATION OF ELECTRON HOLOGRAPHY TO MATERIALS SCIENCE

6.1 Introduction

The goals of development of electron holography techniques are: (1) the pursuit of high resolution, and (2) imaging of pure phase objects, either ones being materials (such as macromolecules, voids, etc.) or ones being immaterials (such as magnetic or electrostatic potential fields). The efforts of breaking the limitation of microscope resolution is Gabor's initial motivation of invention of holography. The 1 Å resolution of TEM image is no longer a microscopist's dream [Völkl & Lichte, 1990]. The possibility of finally reaching the 1 Å level is expected to be in near future. Except the FE TEM, the necessary instrumentation for the high resolution electron holography is a high-pixel-number CCD arrays (at least 4M pixels) optimally coupled to the microscope [Allard et al, 1992], which even now is not commercially available. Hence, the current work is now focusing on the development of the application of electron holography to materials science using phase imaging techniques.

Electron holography provides a direct means to record pure phase objects, which even Gabor did not anticipate in this new field of application. As mentioned in Chapter Four, recent developments in this area mainly include imaging of magnetic fields [Tonomura, 1987], electrostatic field (p-n junctions) [Pozzi et al, 1973], quantized fluxons in superconductors [Matsuda et al, 1991], and atomic step

surface topography [Osakabe et al, 1989]. In this chapter the development a new holographic method for direct imaging of ferroelectric domain wall structure will be discussed.

6.1.1 Ferroelectric materials

Materials that are both polarized and able to reverse polarity under the influence of an electric field are known as ferroelectrics and have many applications in electronic devices. The discovery of ferroelectric phenomenon was made in 1921 by [Valasek, 1921] in an investigation of Rochelle salt (sodium potassium tartrate terahydrate). If one applies an electric field E across a pair of parallel plates of ferroelectric crystal with electrodes connected to the surfaces, and the plates are in series with a condenser, one will find that the measured electric displacement **D** (measured by the charge on a condenser), and hence the polarization **P** ($\mathbf{D} = \mathbf{E} + \mathbf{P}/\varepsilon_0$), is not determined uniquely by the applied field **E**, but depends also on its previous values. The plot of polarization P versus field E curve shows a hysteresis loop (see Figure 6.1). The existence of the hysteresis loop implies that the material possesses a spontaneous polarization, that is, a polarization which persists when the applied field is zero. In Figure 6.1, the length OE represents the spontaneous polarization P_s; OF represents the coercive field E_c, the minimum field needed to reverse the direction of polarization. The spontaneous polarization can be defined in an alternative way as the electric dipole moment per unit volume (in the unit of C/cm^2) [Kittel, 1986]. The hysteresis feature of a ferroelectric material will disappear above a certain temperature called ferroelectric Curie point of that material, and the

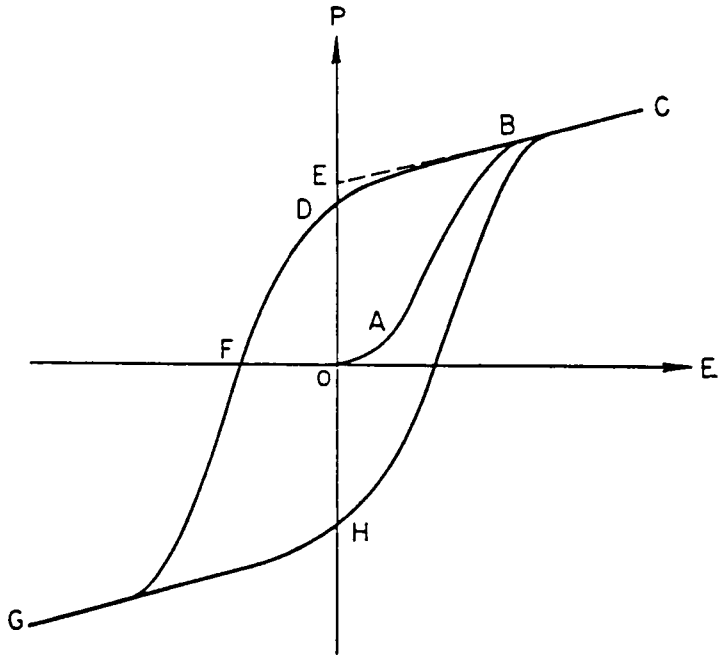


Figure 6.1. Schematic illustration of ferroelectric hysteresis loop (adopted from [Jona & Shirane, 1962]).

materials is said in a paraelectric state. A ferroelectrics exhibits a spontaneous polarization even in the absence of an external electric field. In the ferroelectric state the center of positive charge of the crystal does not coincide with the center of negative charge [Kittel, 1986], and this induce a dipole moment which is the source of spontaneous polarization.

Even in a single crystal ferroelectric material, the polarization properties are not identical at all points. There are homogeneous regions (domain) but the polarization direction differs from one region to another, without in any destroying the coherence of the lattice structure--that is, the crystal structure remains unbroken across a domain boundary (domain wall). Domain walls play an important role in dynamic properties of ferroelectric materials. Domains whose c axis is perpendicular to the principle faces of the crystal wafer are called "c domains"; the "a domain" have the c axis in the plane of the wafer.

Many of the properties of ferroelectrics such as hysteresis loops, frequency responses, dielectric constant, aging, dielectric loss, etc., are related to the motion of domain walls [Little, 1955; Plessner, 1956]. It is therefore of great importance to be able to examine and accurately interpret ferroelectric domain structure [Hu et al, 1986]. Observation of domain structures are conducted by optical polarizing microscope (depending on the changes of refractive index) [Mitsui & Furuichi, 1953], etching (depending on different etching rates) [Burfoot & Taylor, 1979], and powder deposition methods analogous to the well known Bitter technique for magnetic domains [Pearson &

Feldmann, 1959]. Above mentioned methods are all indirect, low spatial resolution methods. The better resolution can be achieved by using electron microscopy methods including both SEM [Le Bihan & Maussion, 1974] and TEM [Blank & Amelinckx, 1963]. The SEM technique depends on changes of surface electric potential from domain to domain [Le Bahin, 1989]; while TEM techniques depend on differences in deviation from the Bragg condition or depend on failure of Friedel's law [Tanaka & Honjo, 1964]. Recently, it was reported that atomic force microscopy (AFM) also can be used to imaging ferroelectric domain walls [Saurenbach & Terris, 1990], which is sensitive to the stray electric field emanating from the surface.

Among many ferroelectric crystals, the domain structures of barium titanate are of particular interest because of its extensive applications in electronic components [Cross, 1984].

6.1.2 Barium titanate ($BaTiO_3$)

One ferroelectric-the synthetic compound barium titanate ($BaTiO_3$)-is perhaps the best-known electroceramic in commercial use. $BaTiO_3$ was found as a ferroelectrics by von Hippel et al, [1946] and independently by Vul & Vereshtschagin [1945]. Since then it has been extensively studied and found a lots of applications in many electronic components such as capacitors, positive temperature coefficient (PTC) thermistors, and piezoelectric (PZT) transducers.

$BaTiO_3$ belongs to the same structure family of 1-2-3 high temperature superconductor, namely, the perovskite structure [Hazen, 1988]. From 5°C to 120°C, $BaTiO_3$ is tetragonal (space group

P4mm) with lattice parameter a = 3.99Å, c = 4.03Å. The barium sits at (1/2, 1/2, 1/2+0.018), titanium at (0, 0, 0.32), and oxygen at (0, 0, 1/2-0.007), (1/2, 0, 0.003), and (0, 1/2, 0.003) [Fan & Bursill, 1988]. Figure 6.2 shows the crystal structure of $BaTiO_3$. Below 5°C to -90 °C, $BaTiO_3$ is orthorhombic (Amm2), below -90°C rhombohedral (R3m), and above 120°C (the Curie point) it is cubic (Pm3m) [Burfoot & Taylor, 1979].

Below the Curie temperature $BaTiO_3$ deformed from the cubic structure by Ba^{++} and Ti^{4+} ions displaced relative to the O^{2-} ions, thereby developing a dipole moment (Figure 6.2), and this dipole moment is the source of spontaneous polarization of $BaTiO_3$. Therefore, the direction of spontaneous polarization of $BaTiO_3$ is along the c axis and parallel to any one of the three original [100] cube axes.

Since $BaTiO_3$ has the most simple perovskite structure in its group materials, and has been widely used in the electronic components, it is then selected in the following holographic study.

6.2 Electron Holography Techniques for Study of Ferroelectric Domain Walls*

6.2.1 Abstract

An electron holographic method allows not only the direct imaging of the domain wall but also the measurement of local spontaneous polarization. Preliminary results of first high resolution

This part is a manuscript of a paper submitted for publication. The co-author is D.C. Joy, Y. Zhang, T. Hashimoto, L.F. Allard*, and T.A. Nolan*, Electron Microscopy Facility, The University of Tennessee, and *High Temperature Materials Lab, Oak Ridge National Laboratory.

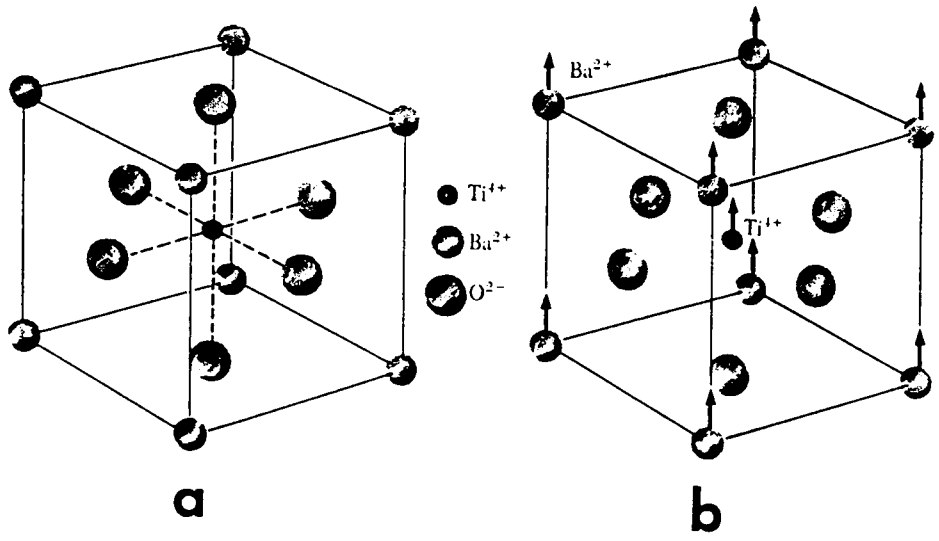


Figure 6.2. (a) The cubic crystal structure of $BaTiO_3$ above the Curie temperature. The prototype crystal is calcium titanate (perovskite), and (b) below the Curie temperature the structure is tetragonal (adopted from [Kittel, 1986]).

imaging of ferroelectric domain walls by using this technique in $BaTiO_3$ and quantitative measurements of electrostatic field distribution across domain walls have been obtained.

6.2.2 Introduction

The unique advantage of recording both amplitude and phase of the object wave makes electron holography superior to the conventional electron microscopy in which only the modulus of the amplitude of the object wave is recorded. The visibility of the phase shifts of an electron wave affected by a local scalar or/and a vector potential makes it possible to directly image the distributions of an electric or a magnetic field at high resolution. Therefore, electron holography has the ability to be a valuable tool in the study of magnetic and superconductor materials [Tonomura, 1987].

Domain wall structures play an important role in determining the characteristics of ferroelectric materials. The walls are regions where dipoles change their directions by 90° and 180° and both the width of the walls and the way in which the polarization changes across them are of interest and importance. Ferroelectric domain walls were initially observed in the $BaTiO_3$ by the standard TEM imaging techniques of diffraction contrast reviewed by [Hu et al, 1986]. In the case of a 90° wall which is a twin boundary on the (101) plane of the tetragonal crystal, the diffraction contrast stems from the strain field around the domain wall. For a 180° wall the contrast relies on the failure of Friedel's law, which makes the contrast weak and observation and image interpretation difficult. The estimation of domain wall width using the residual contrast due

to the strain in the wall regions is an indirect measurement and also encounters a difficulty in evaluating the effect of non-common reflections [Tanaka & Honjo, 1964]. So while this method suffices to show the position of the domain walls, it provides no direct visualization of the electrostatic field itself and hence of the width of the wall, nor can it be used to measure the spontaneous polarization. But using electron holography the phase changes resulting from the polarization changes across the domain wall provide both a direct measurement of domain wall thickness and of the spontaneous polarization [Zhang et al, 1992].

6.2.3 Principle of electron holographic method

In off-axis electron holography techniques the FE TEM is equipped with an electron bi-prism [Möllenstadt & Wahl, 1955] to split the incident electron beam so that one half of the beam travels outside the specimen (a homogeneous coherent background) and combines with the other half of the beam passing through the specimen (a scattered or diffracted wave) to form the electron interference pattern which is the hologram. When electron beams pass through a local electric field, they suffer differential phase shifts and as a result the interference fringes are bent. These bendings are associated with the distribution of the object field so that it is possible to measure the field strength from the distortion of the interference pattern. Since the phase change between one fringe maximum and the next is 2π, a measurement of the fringe shift permits a quantitative determination of the phase change can be made. Hence the magnitude of the local polarization vector can be

found. In this paper an electron holographic method to image and measure ferroelectric domain walls is discussed, and then we will present preliminary results of first high resolution imaging of ferroelectric domain walls by using this technique in $BaTiO_3$ and quantitative measurements of electrostatic field distribution across domain walls.

6.2.4 Quantitative method

Let us calculate the phase difference of the electron waves for the interference electron beams caused by above effect. In the case of a thin film specimen with the presence of a "weak" potential field (comparing with a 200 keV electron beam energy), if aberrations of the objective lens are neglected, a phase shift of the electron beam due to the effect of a local electrostatic field can be described as [Yanzi, 1984]:

$$\Phi(x_0,y_0) = \frac{\pi}{\lambda E}\cdot\int V(x_0,y_0,z)dz \qquad (6.1)$$

where x_0, y_0 define a point on the object, λ is electron wavelength, z is distance of electron travelled along the direction of motion of electron, and E is kinetic energy of the electron, which is given by:

$$E = V_b\left(1 + \frac{1}{2}\cdot\frac{eV_b}{m_0C^2}\right)\left(1 + \frac{eV_b}{m_0C^2}\right)^{-1} \qquad (6.2)$$

here V_b, m_0, e, and C are electron accelerating voltage, electron rest mass, electron charge, and velocity of light, respectively. From eqn. 6.1 one can find that the phase is proportional to the potential

averaged along the electron path, therefore, referred to as "projected potential," its knowledge leads to the determination of electrostatic field distribution.

The integration in eqn. 6.1 is taken along an electron path parallel to the optical axis z, inside and outside the specimen to include stray fields. The total phase shift can be divided into two contributions, one arising from the potential distribution within the specimen, which includes also the mean inner potential, and the other arising from the external fields both above and below the specimen. In the case of $BaTiO_3$, there has been some controversy about the existence of an anomalous surface layer or a polarized surface layer, especially at temperatures near to and above the Curie point (120°C), which may induce surface space charges [Tanaka & Honjo, 1964; Jona & Shirane, 1962; Lines & Glass, 1977; Mitsui et al, 1976]. If there is an anomalous surface layer any measurement of the domain structure will not truly represent the bulk specimen. In this work, we neglect any surface effects of $BaTiO_3$ since at approximately room temperature the material is well below its Curie point and this assumption seems reasonable. Only internal electric fields will be considered in the vicinity of the domain wall.

Domain configurations observed in the electron microscope are frequently changed by nucleation and disappearance of domains, which presumably due to electric and thermal stresses caused by electron irradiation [Tanaka & Honjo, 1964]. In the present experiments, many wage shaped domain nuclei formed at the specimen surface which has a smaller domain size than the regular

one. Many of the holographic measurements are conducted on these "fresh" domains where div $\mathbf{P} \neq 0$.

The internal potential V(x,y,z) includes two terms: V_e for electrostatic field and V_m for mean inner potential. When the external field is zero, the spontaneous polarization $\mathbf{P}_s$ is given by:

$$\mathbf{P}_s = 3\varepsilon_0(\varepsilon\text{-}1)\mathbf{E} \tag{6.3}$$

after Lorentz correction [Kittel, 1986], where $\mathbf{E}$ is the effective local field, ε_0 is the dielectric constant of free space and ε is the dielectric constant of the specimen, and V_e is given by

$$V_e = \frac{\int \mathbf{P}_s \cdot d\mathbf{Z}}{3\varepsilon_0(\varepsilon - 1)} \tag{6.4}$$

where a thin ferroelectric film is approximated as an infinite parallel plate capacitor (lateral length : thickness > 60000). Therefore, assuming constant projected polarization throughout a foil of thickness t, one has

$$\Phi(x_0,y_0) = \frac{\pi \cdot \cos\theta \cdot P_s\,(x_0,y_0) \cdot t^2}{6\lambda E \cdot \varepsilon_0(\varepsilon - 1)} + \frac{\pi}{\lambda E} \cdot V_m t \tag{6.5}$$

where θ is the angle between the electron beam and the polarization vector $\mathbf{P}_s$.

The value of Φ cannot be experimentally determined. The observable quantity is the phase difference $\Delta\Phi$ between two points P_0 and P_1 in object plane [Lenz, 1972]. $\Delta\Phi$ is given by the equation

$$\Delta\Phi = \frac{\pi}{6\lambda E\cdot\varepsilon_0(\varepsilon - 1)}[\cos\theta_0\cdot P_s(x_0,y_0)\cdot t_0^2 - \cos\theta_1\cdot P_s(x_1,y_1)\cdot t_1^2] + \frac{\pi}{\lambda E}V_m(t_0 - t_1) \qquad (6.6)$$

In the case of a 90° domain wall where the polarization vectors make angle θ_0 and θ_1 with the beam then the maximum phase change encountered on crossing the wall will be:

$$\Delta\Phi_{max} = \frac{\pi\cdot P_s\cdot t^2}{6\lambda E\cdot\varepsilon_0(\varepsilon - 1)}\cdot(\cos\theta_0 - \cos\theta_1) \qquad (6.7)$$

where we are assuming a uniform sample thickness t across the domain wall. Note that from the experiment one measures the magnitude of the fringe bending $\Delta X/X$ rather than the maximum phase difference $\Delta\Phi_{max}$ which are related by a simple expression $\Delta\Phi_{max} = 2\pi(\Delta X/X)$. Eqn. 6.7 shows that we can experimentally measure local spontaneous polarization vector $\mathbf{P}_s$ from the value of $\Delta\Phi_{max}$. The spontaneous polarization can only be measured previously by the pyroelectric method [Lines & Glass, 1977] which is a kind of bulk measurement without any spatial resolution. From the slope of the fringe bending one also can determine the distribution of the polarization vector at each point on crossing the domain wall. It should be noted from eqn. 6.7 that if all the components of a polarization vector are perpendicular to the electron beam, no electric field effect on the phase shift would be observable. This is also expected from local field calculation when polarization axis is in plane [Kittel, 1986].

Electron holography is a two step process since each hologram must be reconstructed on an optical bench, or digitized and then

numerically reconstructed to yield the desired final image. Since pure phase image is expected to be a real mapping of electrostatic potential distribution, provided the specimen has a uniform thickness, we also performed numerical reconstruction of a pure phase image with phase difference amplification up to 8 times. An interferogram is also obtained from the reconstruction. The contour lines in the interferogram represent the equipotential lines of the projected potential.

6.2.5 Experimental method

specimen

A high purity, tetragonal $BaTiO_3$ single crystal oriented on a (001) axis was used in this study. The specimen was prepared by mechanical polishing, dimpling, and finally ion milling to produce a central thinned region for TEM observation by 200 keV electrons. As mentioned above, 90° domain walls may be observed in $BaTiO_3$ by standard TEM bright field imaging techniques using diffraction contrast [Hu et al, 1986; Tanaka & Honjo, 1964]. Figure 6.3 shows a typical image of this type from this specimen with the domain walls showing in high contrast. This method was used to select the area for subsequent holographic study.

reconstruction of electron holograms

Electron holograms recorded on photographic films are digitized using a flat bed scanner with a resolution of 70 dpi at 256 grey levels. The digitized hologram is then reconstructed numerically by a developed MPW C program running on the Apple Macintosh II

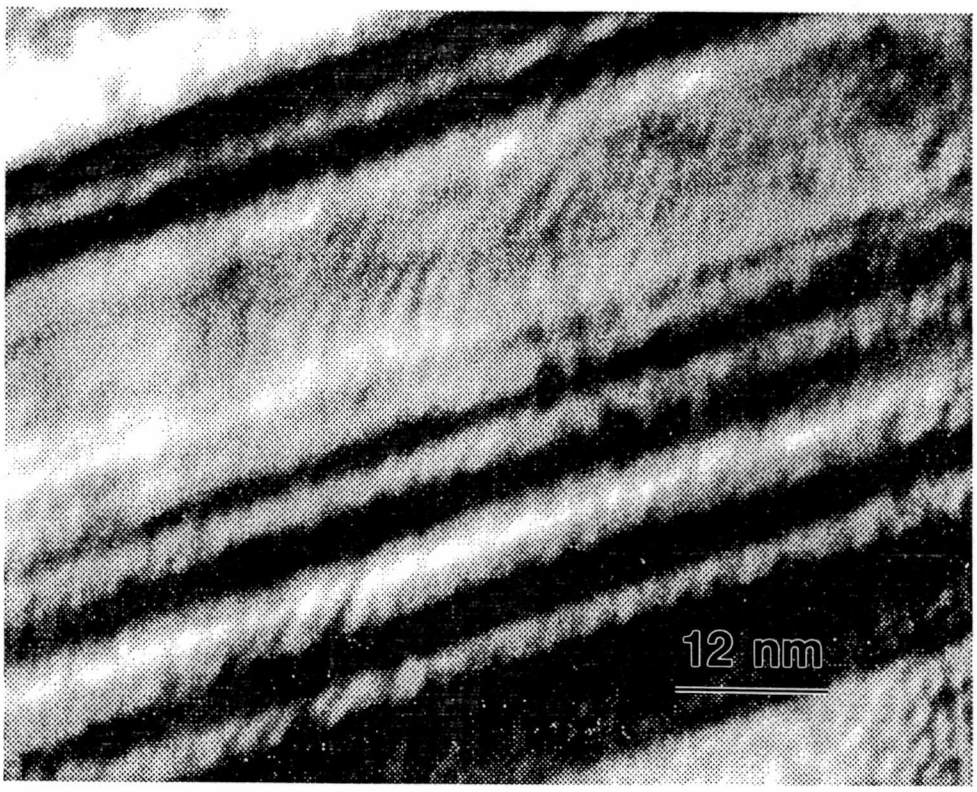


Figure 6.3. Bright field TEM image of 90° domain walls in $BaTiO_3$ at 200 keV incident beam energy.

computer [Zhang et al, 1992]. The maximum size of data file is 512x512 pixels because of the limitation of computer memory. After reconstruction several different data files are obtained, which includes amplitude, phase, phase amplification, and interferograms calculated from the initial hologram. The typical computing time for a 512x512 file is about 20 minutes when using an unmodified Mac II computer.

6.2.6 Results and discussion

Figure 6.4 shows the electron hologram obtained over a 90° domain wall region found as in Figure 6.3. The specimen was so oriented so as to place the plane of the domain wall, as well as a component of the spontaneous polarization, parallel to the electron beam. On either side of the wall the interference fringes are straight but in the vicinity of the wall the fringes can be seen to undergo bending. If we extend lines along the original position of the fringes (Figure 6.5) then the lateral offset of the fringe is directly proportional to the differential phase shift encountered by the electron as it passes through the specimen at different positions. This phase shift is made up of two components, one due to the inner potential of the specimen and one due to the polarization. In both cases the total phase shift will depend on the thickness of the region traversed by the beam. Because the areas under observation were only a few tens of angstroms in width the thickness of the sample was constant, so variations in the phase are due only to variations in the electrostatic potential sampled by the beam as it passes through the sample. The width of the domain wall, therefore, can be

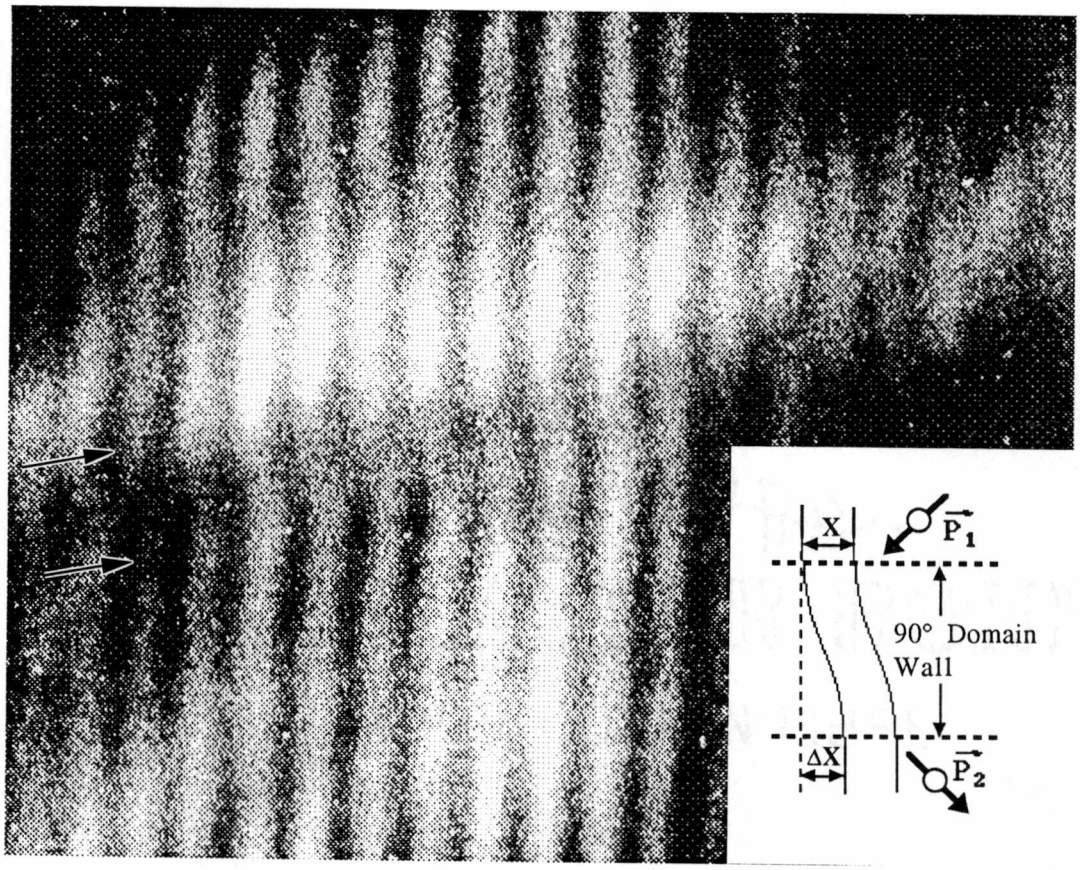


Figure 6.4. Electron hologram across a 90° domain wall using fringe spacing by 3.7 Å. Edges of the wall are indicated by pairs of arrows. Inset shows possible configuration of changes of polarization vectors across the wall.

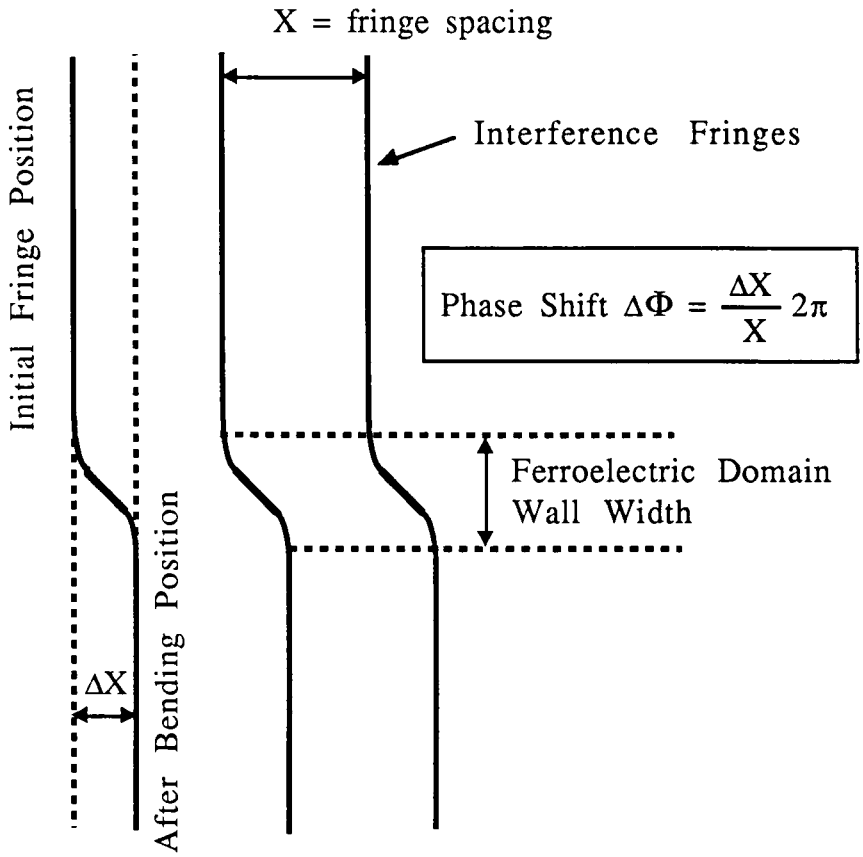


Figure 6.5. Schematic illustration of the measurement of phase shift from the hologram fringe pattern.

measured directly from the offset to the finish of bending for each of the fringes as it crosses the nominal wall position (see Figure 6.5). When interference fringes are not perpendicular to the projected domain wall as in the case in Figure 6.6, the measured wall thickness has to be corrected to the right direction as shown in the figure, namely, $W=W_m/Sin\alpha$, here W is corrected wall thickness, W_m a measured wall thickness, and α an angle between the fringe and the wall as defined in Figure 6.6. The inset in Figure 6.4 shows a schematic configuration of changes of polarization vector across the domain wall.

Measurements on a number of 90° domain walls show that the projected wall thickness is between 2 to 5 nm. As expected, these values are lower than those estimated by conventional TEM methods since these give only an indirect estimate from strain field, but are in good agreement with values predicted from calculations by various authors [Jona & Shirane, 1962; Lines & Glass, 1977; Zhirnov, 1959]. The width of each individual domain wall was found to be constant along its length, but there were differences between walls due to small changes of orientation encountered when moving from one domain to another.

Figure 6.7(a) shows a sinusoidal type fringe bending which occurs near the tip of a wadge-shape domain. The numerically calculated interferogram of the same area is shown in Figure 6.7(b). The centers of the domain wall positions have been marked by dash lines. Since all the points along one contour line in the interferogram

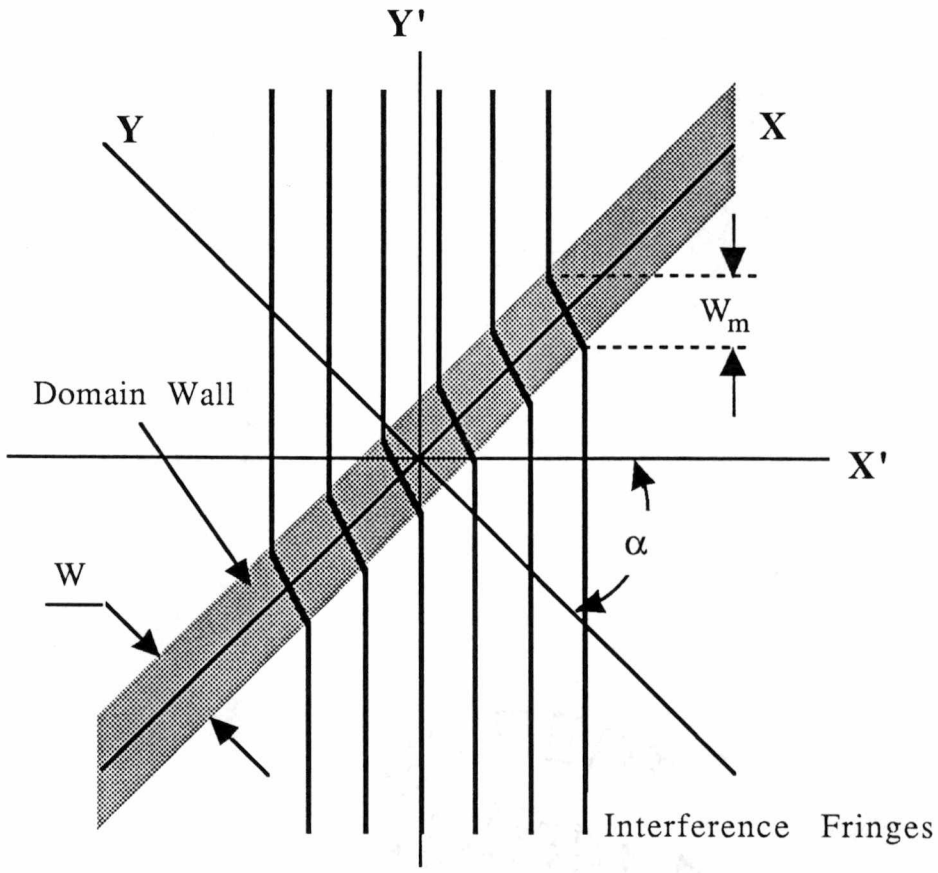


Figure 6.6. Geometric relation between interference fringes and domain wall. X-Y is the coordination of the domain wall; X'-Y' is the coordination of the fringes.

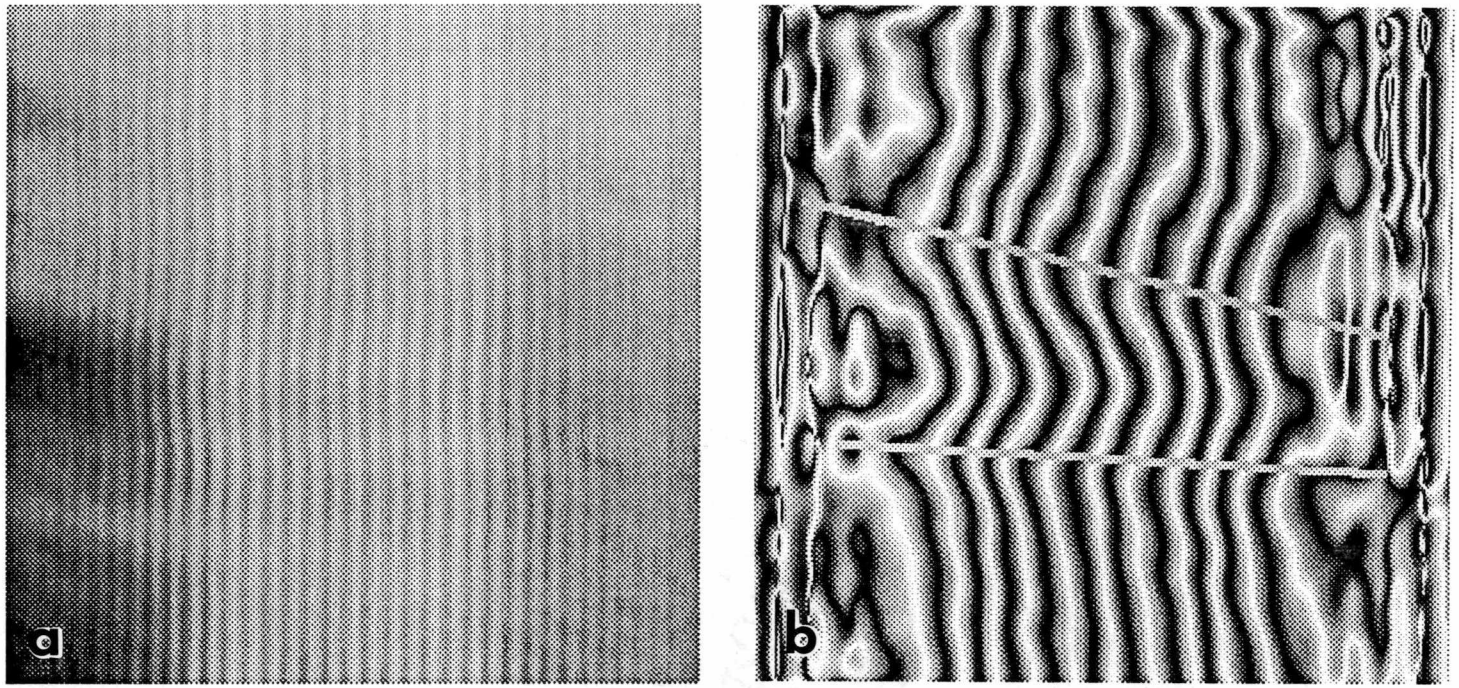


Figure 6.7. (a) Electron hologram of a wadge-shape domain wall with a fringe spacing of 4.6 Å, and (b) interferogram of the same region.

are of equal potential, the interferogram reveals significant changes of electrostatic field across the two sides of the domain.

An electron hologram of ferroelectric domain wall enables us to measure the maximum fringe bending $\Delta X/X$, and hence $\Delta\Phi_{max}$. Using eqn. 6.7, we are able to calculate the magnitude of local spontaneous polarization $\mathbf{P}_s$. As an example, for the hologram shown in Figure 6.5, the electron beam is travelling along the [111] zone axis, with the polarization vectors along [-100] and [001]. Thus $\cos\theta_1 = -1/\sqrt{3}$ and $\cos\theta_0 = +1/\sqrt{3}$; λ for 200 keV electrons is 0.0251 Å. For this orientation $\varepsilon \approx \varepsilon_c$, where ε_c is the dielectric constant along the c-axis, thus, at 300 K ε is about 160 [Mitsui et al, 1976], and ε_0 is 8.85×10^{-12} C^2/J m. The maximum fringe offset is in average of 0.32 fringe spacing on crossing the wall, so the maximum phase shift is 0.65π. Substituting these values into eqn. 6.6 we find that $P_s = 3.1\times10^{-5}$ C/cm^2, a value close to the measured average bulk value of $P_s = 2.6\times10^{-5}$ C/cm^2 [Kittel, 1986].

Eqn. 6.7 also shows that phase shift at any point across the domain wall is directly related to P_z, the component of polarization parallel to the electron beam direction. The shape of the fringe bending, therefore, quantitatively displays the variation of $P_z(x)$ as a function of position x across the wall. Figure 6.8 plots the normalized variation of P_z, averaged from measurements on four fringes. According to the Zhirnov model [Zhirnov, 1959], the spontaneous polarization vector changes gradually, $P_z = P_0\tanh(x/d)$, where P_0 is the equilibrium polarization and 2d is the wall width. The agreement between the experimental and predicted variation is close, but the

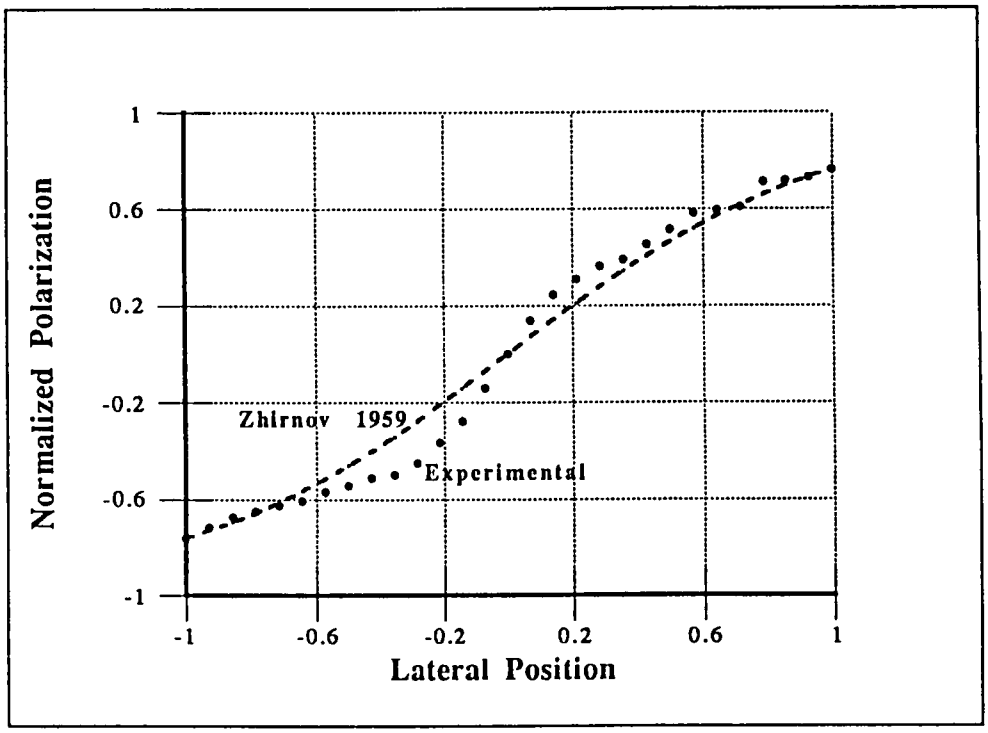


Figure 6.8. Experimentally determined variation of normalized polarization component Pz across 90° wall (open circles) compared with theory of Zhirnov (dotted line).

experimental curve appears to vary more rapidly in the central region than expected.

Figure 6.9(a) shows a high resolution electron hologram of a 90° domain wall using fringes spaced 1.5 Å as well as the numerically reconstructed phase image with four time phase amplification and the interferogram from the same region. In the phase image (Figure 6.9(b)) there are significant phase changes across the domain wall. The phase difference between two adjacent phase contour is $\pi/2$. The interferogram (Figure 6.9(c)) provides a very intuitive view of the phase changes, and hence the potential field changes, across the domain wall. From both the phase image and the interferogram one can find that the polarization changes its direction in arcs.

In addition to the 90° domain walls shown in above, we have also observed anomalous effects in holograms from some domain walls. Figure 6.10 shows two examples of this type. Instead of a simple bend in the fringes as they cross the domain wall, it can be seen that in several instances a fringe bifurcates (Figure 6.10(a)) so that the number of fringes per unit length on either side of the wall is not the same, and the fringe pattern appears similar to crystal planes around an edge dislocation, or two fringes bifurcate and diverge again as shown in Figure 6.10(b). Since each of the interference fringe represents a contour of constant phase this implies that there is a discontinuity in the phase and hence in the potential sampled by the electron. A possible explanation is that these discontinuities correspond to a region of charge. As shown in

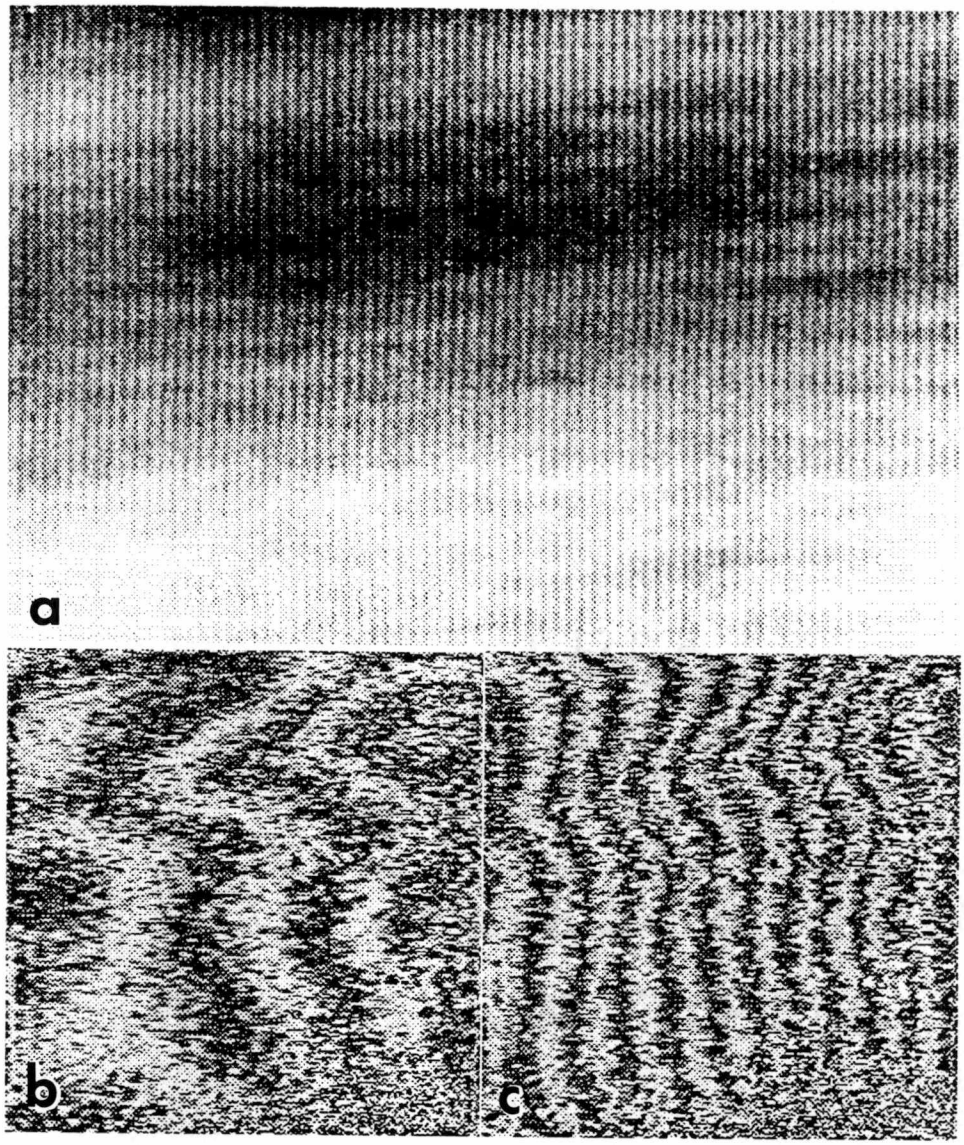


Figure 6.9. (a) Electron hologram across a 90° domain wall using fringe spaced by 1.5 Å, (b) numerically reconstructed phase image (amplified by 4 times), and (c) interferogram from the same region.

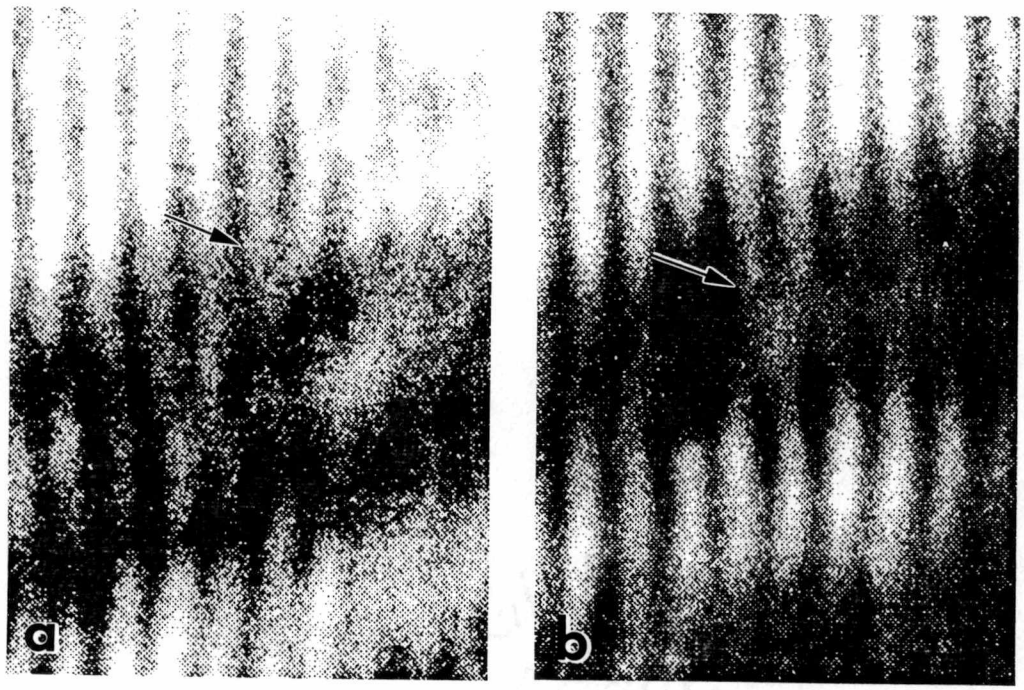


Figure 6.10. Anomalous fringe bifurcations from some domain walls. Fringe spacings are 3.7 Å. (a) Y-shape fringe splitting, and (b) X-shape fringe splitting.

Figure 6.5 of Fukuhara et al [1983] a magnetic monopole would produce a hologram fringe splitting of this type, the number of extra fringes being a function of the strength of the monopole. The authors suggest that a localized region of electrostatic charge behaves similarly. Such a region could result from the attraction of charged oxygen vacancies to the domain wall or a head to head 180° wall is possible which induces a charged wall [Tanaka & Honjo, 1964]. The maximum phase shift of the order of 2.5π in those phase discontinuous regions (Figure 10(a)) is a too high value for the built in potential only and it is about 4 times higher than the value measured from Figure 6.4, confirming the presence of a significant local field.

As a conclusion, in this work a new electron holographic method to study the domain wall structures of ferroelectric materials has been developed. It provides not only a mean to directly image the ferroelectric domain wall but also a mean to quantitatively measure the domain well thickness as well as local spontaneous polarization. Examples of the application of our method to the studies of domain wall structures of $BaTiO_3$ have been given. Measured domain wall width for 90° walls is between 2 to 5 nm. Reconstructed phase image and interferogram display vividly the electrostatic potential changes across the domain walls. The measured values of local spontaneous polarization are very closed to the measured bulk value. Experimentally measured polarization changes across the domain wall is closed to the prediction based on the Zhirnov model. Some anomalous fringe bifurcations were observed.

References

Allard, L.F., Nolan, T.A., Joy, D.C., Hashimoto, T., Zhang, X., and Zhang, Y., submitted to Ultramicros., 1992

Blank, H. and Amelinckx, S., Appl. Phys. Lett., **2**(1963), p. 140

Burfoot, J.C. and Taylor, G.W., Polar Dielectrics and Their Applications, (Macmillan Press, Ltd., 1979)

Cross, L.E., Amer. Ceram. Soc. Bull., **63**(1984), p. 586

Fan, X. and Bursill, L.A., Optik, **78**(1988)2, p. 67

Fukuhara, A., Shinagawa, K., Tonomura, A., and Fujiwara, H., Phys. Rev., **B27**(1983). p. 1839

Hazen, R.M., Scienti. Amer., (1988)6, p. 74

Hu, Y.H., Chen, H.M., Wen, Z.X., and Harmer, M.P., J. Amer. Ceram. Soc., **69**(1986), P. 594

Jona, F. and Shirane, G., Ferroelectric Crystals, (Macmillan Co., New York, 1962).

Kittel, C., Introduction to Solid State Physics, 6th ed., (John Wiley & Sons, Inc., New York, 1986)

Le Bahin, R., Ferroelectrics, **97**(1989), p. 19

Le Bihan, R. and Maussion, M., Ferroelectrics, 7(1974), p. 307

Lenz, F., Z. Physik, **249**(1972), p. 462

Lines, M.E. and Glass, A.M., Principles and Applications of Ferroelectrics and Related Materials, (Clarendon Press, Oxford, 1977)

Matsuda, M., Fukuhara, A., Yoshida, T., Hasegawa, S., Tonomura, A., and Ru, Q., Phys. Rev. Lett., **66**(1991), p. 457

Mitsui, T. and Furuichi, J., Phys. Rev., **90**(1953), p. 193

Mitsui, T., Tatsuzaki, I., and Nakamura, E., ed., An Introduction to the Physics of Ferroelectrics, (Gordon and Breach Sci. Pub., London, 1976)

Möllenstadt, G. and Wahl, H., Naturwissenschaften, **42**(1954), p. 41

Osakabe, N., Endo, J., Matsuda, T., Tonomura, A., and Fukuhara, A., Phys. Rev. Lett., **62**(1989), p. 29

Pearson, G.L. and Feldmann, W.L., J. Phys. Chem. Sol., **9**(1959), p. 28

Plessner, K., Proc. Phys. Soc., London, **69**(1956), p. 1261

Pozzi, G. and Missiroli, G.F., J. Microsc. (Paris), **18**(1973), p. 103

Tanaka, M. and Honjo, G., J. Phys. Soc. Japan, **19**(1964), p. 954

Tonomura, A., Rev. Modern Phys., **59**(1987)3, p. 639

Valasek, J., Phys. Rev., **17**(1921), p. 475

Vul, B.M. and Vereschtschagin, L.F., Daklady Akad. Nauk SSSR, **49**(1945), p. 662

von Hippel, A., Breckenridge, R.G., Chesley, F.G., and Tisza, L., Ind. Eng. Chem., **38**(1946), p. 1097

Völkl, E. and Lichte, H., Ultramicros., **32**(1990), p. 177

Yanzi, M., Optik, **68**(1984), p. 319

Zhang, Y., Bunn, R.D., Joy, D.C., Zhang, X., Allard, L.F., and Nolan, T.A., presented at 4th Conf. Frontiers of Electro. Micros. in Mater. Sci., Oakland, CA, April 21-24, 1992

Zhang, X., Hashimoto, T., Joy, D.C., Appl. Phys. Lett., **60**(1992)6, P. 784

Zhirnov, V.A., Sov. Phys., JETP, (1959), p. 882

VITA

Xiao Zhang was born in Shanghai, China on September 25, 1961. He attended elementary and secondary schools in that city and was graduated from City West High School in June 1980. The following August he entered Shanghai University of Technology, and in October 1984 he received a Bachelor of Engineering degree in Metallurgical Engineering.

In the fall of that year he was employed as a lecturer in Shanghai University of Science and Technology.

After two and one-half years working in the university, in March 1987, he entered the graduate school of the University of Tennessee, Knoxville. He received the Master of Science degree with a major in Materials Science and Engineering in August 1989. After following three years study in the same university, he received the degree of Doctor of Philosophy in August 1992, with a major in Metallurgical Engineering.

The author is a student member of ASM International, The Minerals, Metals & Materials Society, Materials Research Society, Electron Microscopy Society of American, Microbeam Analysis Society, The American Ceramic Society, Society for the Advancement of Materials and Process Engineering, The Society of Photo-Optical Instrumentation Engineers, The Electrochemical Society, and National Association of Corrosion Engineers. He is also a member of Tau Beta Pi (National Engineering Honor Society).