INFORMATION TO USERS

Light scattering from binary optics

Ricks, Douglas Wayne, Ph.D.

The University of Arizona, 1993

LIGHT SCATTERING FROM BINARY OPTICS

by

Douglas Wayne Ricks

A Dissertation Submitted to the Faculty of the

OPTICAL SCIENCES CENTER

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1993

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have

read the dissertation prepared by_____Douglas Wayne Ricks_____

entitled ____Light Scattering from Binary Optics____

_____

_____

_____

_____

and recommend that it be accepted as fulfilling the dissertation

requirement for the Degree of ___Doctor of Philosophy___

Dr. Angus Macleod	_signature_	5/5/93
		Date
Dr. Tom Milster	_signature_	5/5/93
		Date
Dr. Robert Shannon	_signature_	5/5/93
		Date
		Date
		Date

Final approval and acceptance of this dissertation is contingent upon
the candidate's submission of the final copy of the dissertation to the
Graduate College.

I hereby certify that I have read this dissertation prepared under my
direction and recommend that it be accepted as fulfilling the dissertation
requirement.

signature   5/5/93
Dissertation Director   Date

Dr. Angus Macleod

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _Douglas W. Ricks_

ACKNOWLEDGEMENTS

I owe a debt of gratitude to many individuals for their support and encouragement. I would like to thank supervisors (Tom Loftus, Wayne Tanaka, and Joseph Cardani) and management (Robert Hintz, Dr. Edward Kutchma, Milton Burford, Dr. James Stanford, Dr. Ronald Atkins, and Dr. Ronald Derr) at the Naval Air Warfare Center Weapons Division (NAWCWD) for providing financial support. Some financial support was also given by Dr. George Mueller of the Naval Research Laboratory with the management support of Conrad Driussi (NAWCWD). The idea of studying scattering from binary optics was given to me by Wayne Tanaka.

Over the last nine years Dr. Hal Bennett has provided encouragement and guidance in my dissertation research. I have benefited from discussions on scattering with him and with Dr. Jean Bennett. Dr. Merle Elson was very helpful in discussions on diffraction theory and in deriving the roughness scattering formulas for binary optics. Van Hodgkin performed the Wyko profilometer measurements, and Dr. Jean Bennett the Talystep measurements. Van Hodgkin also provided other scatter measurement data. I had the assistance of Larry Chizek and Paul Works in making scattering measurements. Dennis Arndt, with whom I share an office, has been helpful in making me clarify my ideas.

I would also like to thank Dr. Hal Bennett, Dr. Merle Elson, and Dr. Lewis DeSandre for reading draft copies of this dissertation. Their comments on improving the dissertation are appreciated. Dr. Bennett's comments in particular were helpful in assisting the reader to understand throughout the paper where he has been, why he is there, and where he is going.

My dissertation committee should be acknowledged for the difficulty of working with me long distance. I always found them positive, polite, and encouraging. I don't think anyone could be more upbeat than Dr. Macleod.

Finally I would like to thank my family. My in-laws and my parents and seven brothers and sisters for their encouragement. My seven children for their patience and understanding. Saving the best for last, I would like to acknowledge the support of my wife, Janet Allyn Leake Ricks, during the twenty years of our marriage. She had to carry much of the load of raising the family while I was working on my degrees. She also proofread each draft of this dissertation. Without her willingness to let me pursue an advanced degree, and without her faith and confidence in me, this dissertation would not have been possible.

TABLE OF CONTENTS

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

TABLE OF CONTENTS - Continued

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS - Continued

LIST OF ILLUSTRATIONS - Continued

LIST OF ILLUSTRATIONS - Continued

LIST OF ILLUSTRATIONS - Continued

LIST OF ILLUSTRATIONS - Continued

LIST OF TABLES

ABSTRACT

Binary optics is a new technology that makes use of the principle of diffraction instead of reflection or refraction to change an incident wavefront. This technology takes advantage of the recent progress in microlithography. There are many new and exciting applications for binary optics, and we can also expect to see the replacement of some conventional optical elements with binary optics.

In many ways a binary optic behaves like a diffraction grating with a period that changes continually over the surface of the optic. We find that energy is scattered into different diffraction orders, and there is scattering similar to "grass", "ghosts", "errors of run", "accidental errors of amplitude", and diffuse scattering from surface roughness, just like there is from a diffraction grating.

There are vector theories and scalar theories of diffraction. In this dissertation we give the conditions under which the various theories are applicable. We derive a formula for scattering from binary optics with slightly rough surfaces. By comparing this theory to computer simulations of scattering from binary optics we show that the theory can account for random fabrication errors. Formulas are also derived to predict the scattering from systematic errors. The author designed and built an instrument to measure scattering at small angles, and we show that measured scattering from binary optics can be predicted by the theories developed.

1. INTRODUCTION

1.1 Overview

In this dissertation we shall be considering optics which diffract energy into different orders in a controlled, systematic way. Due to processing errors some energy will also be randomly distributed between diffraction orders. The distribution of energy may be intentional or unintentional, desirable or undesirable. We use the term "scattering" or "scatter" to refer to all undesirable distributions of energy.

To begin with we describe what a binary optic is, how it is fabricated, how it is used. We show that binary optics are the result of a long history of efforts to utilize the principle of diffraction. They are related to, but somewhat different from, diffraction gratings, zone plates, and holographic elements.

In Chapter 2 we briefly review the history and state-of-the-art of diffraction theory as it applies to binary optics. We explain the assumptions in the derivation of the scalar theory to better understand the conditions under which the vector (electromagnetic) and the scalar theory will give similar results. Binary optics are characterized by abrupt changes in the surface profile, and we review the current knowledge of EM theory as it applies to these types of structures.

Next, in Chapter 3, we review the current scattering theories for diffraction gratings. In the literature binary optics are modelled as diffraction gratings with a variable period. First the grating equation, then formulas for the diffraction efficiency of grating slits and blazed gratings are derived. We then discuss diffraction

efficiencies for Dammann gratings, Fresnel zone plates, and general binary optics. Several systematic and random processing errors are also reviewed.

The Rayleigh-Sommerfeld theory is the most general scalar theory of diffraction, but it is difficult to use. Fourier optics is much easier to work with, and many explicit analytic expressions have been derived which give valuable insight. Fourier optics is based upon Fraunhofer diffraction, however, which is much more restrictive than Rayleigh-Sommerfeld diffraction. Fourier optics is not correct for large off-axis angles. This limitation is removed in Chapter 4 with a new, original formulation of Fourier optics. The new Fourier optics retains the accuracy of the Rayleigh-Sommerfeld formula when the aperture is relatively small. The new formula agrees with vector theory in many situations where the old Fourier theory does not. Chapter 4 outlines when each vector and scalar theory is valid, as applied to the scattering theories of Chapter 3.

Chapter 5 applies the principle of scattering to systematic errors. The fundamental principle is that a surface can be regarded as consisting of two components; an error free, perfectly smooth part, and the part which causes scattering. When the scattering part consists of systematic errors it acts as a "ghost grating" and redistributes energy in the various diffraction orders. When the errors are random the scattering part acts like a rough surface. Several original formulas are derived and shown to explain the results of actual measurements. Since the energy goes into particular diffraction orders, the important parameter is diffraction

efficiency. Geometric optics can be used to determine how the energy will be distributed in space.

Random errors are discussed in Chapter 6. A new formula is derived that predicts scattering from random roughness and random processing errors. A comparison is made of computer simulations with the roughness formula. The simple roughness formula is shown to express very well the average scattering levels. Formulas are derived to give the effective "correlation length" and "roughness" for random process errors.

Finally in Chapter 7 the theory of scattering from random errors is compared to actual measurements. A novel scatterometer is described which is capable of measuring the very small angles needed to verify the theory for binary optics. The measurements of the surface which are necessary to predict the scattering are described. We also discuss the inverse problem of determining the surface characteristics by the scattering measurements.

In summary, Chapters 1, 2, and 3 review the background and current knowledge of scattering from binary optics. The rest of the dissertation breaks new theoretical ground, with references to the measurements and empirical works of others.

1.2 What is a Binary Optic?

In order to better understand binary optics we will now review some of the concepts and optical devices that preceded binary optics. This review will only touch

on a few of the high points since optics has a very long history. We will emphasize diffractive optical elements in this review.

Mirrors and lenses were known to the ancient Egyptians about 2000 B.C. Mention of the properties of reflection and refraction were made in the works of the Greeks several hundred years before Christ. The mathematical laws governing reflection and refraction were developed gradually by Arab and European researchers over the next two thousand years. Snell in 1621 discovered the law of refraction, though it was Rene' Descartes who a few years later first published the law of refraction in terms of sines.

The phenomenon of diffraction was first scientifically studied by Francesco Grimaldi in the seventeenth century. Grimaldi had observed bands of light in the shadow of a rod illuminated by a source. Until the development of the wave theory two centuries later, however, the laws of diffraction could not be clearly stated. Some early attempts at a wave theory were begun in the seventeenth century, but became unpopular in the latter half of that century because of the great influence of Isaac Newton. This changed at the beginning of the nineteenth century due to the work of Thomas Young, Augustin Fresnel and Dominique Arago. Fresnel's observations of diffraction led to a new theory of diffraction and at the same time helped to revive the wave theory of light.

At about the same time Fresnel was making his discoveries of diffraction, the German optician Joseph Fraunhofer was using a primitive diffraction grating to study the solar spectrum. In 1821 Fraunhofer published his discoveries and furthered both

the science of spectroscopy and our understanding of diffraction gratings. The earliest diffraction gratings consisted of multi-slit assemblies. Wire was wrapped around a pair of fine threaded screws. Light is diffracted as it passes through the slits, and the periodic nature of the slits gives a modulation to the amplitude of the wavefront. Another type of diffraction grating, developed at a later time, modulates the phase of the wavefront by periodic variations in the optical thickness across the grating.

The wave theory as developed by Fresnel assumed that light transverses a medium. Light was thought of in a mechanical sense as vibrations of the "ether". Towards the latter half of the nineteenth century James Maxwell developed a set of mathematical equations which established a relationship between light and electromagnetism. This put the wave theory on a more rigorous theoretical basis and today Maxwell's equations are believed to correctly predict the wave behavior of light. In most practical cases the Fresnel theory is adequate; Maxwell's equations are necessary when the size of apertures or surface features is comparable or smaller than the wavelength of light. Progress is continually being made in the application of these theories to optical problems. Anomalies in diffraction gratings could not be fully explained until the analysis of diffraction gratings was improved in the 1970's. Diffraction theory will be discussed in more detail in Chapter 2.

The diffraction grating was the first attempt to utilize diffraction instead of refraction or reflection to control wavefronts and ray directions. Another early diffractive optical element was the "zone plate", often called the Fresnel zone plate,

but invented by J.L. Soret around 1875. Fresnel originated the concept of dividing a wavefront into zones. Let P be some observation point, and b the shortest distance between the wavefront and the point P. The wavefront is assumed monochromatic with wavelength λ. The first zone is the set of all points on the wavefront which lie between b and $b + \lambda/2$ from P. The second zone contains all the points on the wavefront a distance between $b + \lambda/2$ and $b + \lambda$ from P. Additional zones are defined in a similar manner. A zone plate is an optical element which blocks the first, third, fifth, etc., zones of the wavefront but transmits the even numbered zones. In this manner the portions of the wavefront which will add constructively at P will be allowed to pass, while the destructive portions are blocked. As a result, a portion of the wavefront is focused by diffraction to the point P.

It was suggested by Lord Rayleigh that a more efficient zone plate could be achieved if the odd-numbered zones were not blocked off, but allowed to pass after adding an additional half wavelength to the optical path. Then all of the zones would reinforce each other. In 1898 R.W. Wood presented several methods of achieving Lord Rayleigh's suggestion. In one method alternate zones are etched to the proper depth on a plate of glass by means of hydrofluoric acid. This type of zone plate is sometimes called a "Wood lens", a "phase-reversal zone plate", or a "phase only zone plate". It is only recently that modern photolithographic technology has made the zone plate a practical and useful substitute for lenses.

Another predecessor to binary optics is holography. Some early work in holography was begun in 1947, but this technology was not developed much until the

invention of the laser in the 1960's. The first paper on the use of a hologram as a diffractive optical element dates back to 1957, though most of the pioneering work was in the mid-1960's by Adolf Lohmann of the University of Erlangen in West Germany[1]. The first holographic optical elements (HOEs) were constructed by recording in a medium such as film the interference pattern of two coherent wavefronts. The developed film acts as a grating. This grating can be illuminated by one of the two original wavefronts and the grating will diffract the wavefront into a wavefront identical to the other construction wavefront.

A subclass of holographic optical elements is the binary or two-phase holograms where the interference pattern is digitized as either "in-phase" or "out-of-phase". Computers can generate HOEs of this nature by black and white interference pattern drawings of two hypothetical incident wavefronts. The great advantage of this approach is that it is no longer necessary to physically create in the lab the two construction wavefronts. The ideal mathematical description of the wavefront may be used. In the case where the incident wavefronts are descriptions of a plane wavefront and a converging wavefront, a photograph of the resultant digitized interference pattern would be a Fresnel zone plate.

In 1977 a group at Massachusetts Institute of Technology, Lincoln Labs (MIT/LL), under the direction of Wilfrid Veldkamp began experimenting with a class of optical elements that Dr. Veldkamp labelled "binary optics". As defined by Dr. Veldkamp[2], binary optics are "a diffractive optics technology that uses computer generated designs of microscopic surface relief patterns and VLSI [very large scale

integration] ion-etching technology". The name is based on "(1) a binary electronic mask technology (chrome on glass), (2) a binary mask design code, 2**N phase steps for N masks, and (3) a two-level or high-low, ion etch process with uniform surface treatment". In this dissertation we shall not restrict binary optics to ion-etching technology, but will include binary optics fabricated by any technology which produces vertical walls and flat multiple level surface relief patterns.

Binary optics, then, function on the principle of diffraction. They are diffractive optical elements like diffraction gratings, Fresnel zone plates, and holographic optical elements. Binary optics, though, are constructed in a particular manner which gives the surface of the optic a pattern of multiple discrete flat levels. A diffraction grating with etched grooves of perpendicular walls and flat bottoms would be an example of a binary optic. A "phase-reversal zone plate" as described by R.W. Wood is another example of a binary optic. A binary holographic optical element can be used to create a binary optic; the element is used as an etching mask to transfer a pattern to a surface. These examples are two-level binary optics. Binary optics may have multiple levels using a process of repeated etching with a succession of masks. Multiple level binary optics are an approximation to the continuous surface profiles of the more general "kinoforms."

1.3 Fabrication of Binary Optics

The design of a binary optic usually begins with a determination of the desired surface relief pattern. In some cases this can be achieved by using a lens design

computer program (such as CODE V from Optical Research Associates) which can handle holographic optical elements. The number of levels used to approximate the desired surface is then specified and another program calculates the pattern that each mask must have. Each mask is then made using specialized equipment developed for VLSI technology. Often the mask consists of a thin pattern of chrome deposited on glass, and is made using electron-beam lithography.

In the ion-etching technology a thin layer of positive photoresist is deposited on an optical surface. The mask is placed in contact with the photoresist and exposed to ultraviolet light. The pattern of photoresist under the chrome is protected from the ultraviolet light. When the exposed photoresist is sufficiently broken down into smaller molecules by the ultraviolet light, it is chemically removed. The unexposed photoresist resists removal. The surface now has a pattern of photoresist on it. This surface is then exposed to gases which have been ionized and accelerated in a process called reactive ion-etching. The photoresist protects portions of the surface; the unprotected areas are etched away. When the etch pattern is sufficiently deep this process is stopped, and the remaining photoresist is chemically removed. This entire process of mask generation, covering and exposure of photoresist, etching, and resist stripping is repeated if the binary optic is to have 4 levels instead of 2 levels. The state of the art is a 16 level structure formed by repeating the process with 4 separate masks.

The mask generation step is usually time consuming and expensive, though the mask can be used repeatedly. The alignment of multiple masks may be a practical

limit on the technology. For this reason alternate fabrication processes are being investigated. One approach uses an electron beam gun to directly etch many levels in a one step process. Another approach uses a single mask with a number of gray scales.

An alternate technology to ion-etching uses the deposition of thin films to achieve the desired surface relief pattern. Single point diamond turning has also been utilized for circularly symmetric patterns. In some cases the diamond does not leave vertical walls and flat bottoms and actually produces kinoforms instead of binary optics.

1.4 Applications

Binary optics have been investigated for numerous applications, both conventional and unconventional. As conventional optical elements, binary optics compete with lenses, mirrors, diffraction gratings, and beam splitters. Unconventional applications include anti-reflective coatings, optical filters, beam steering, and coherent beam addition.

Binary optics can be designed to replace conventional refractive and reflective optics. As Snell's law is to refractive optics, the grating law is to diffractive optics. That law is

$$m \lambda = a (\sin \theta_1 + \sin \theta_2), \tag{1}$$

where λ is the wavelength, m is the order number, a is the grating spacing, θ_1 is the incident angle, and θ_2 is the diffracted angle. In a typical binary optic the "grating

spacing" will be different from place to place on the optic. We see from the grating law that the shorter the wavelength, the smaller the diffracted angle. This is in contrast to ordinary refractive materials where the index is larger for shorter wavelengths, so the refracted angle is larger for shorter wavelengths. It turns out that binary optic lenses have a large amount of negative dispersion so a system of individual binary optic lenses would be useful only in nearly monochromatic light. However, a hybrid system of both refractive and diffractive elements can be broadband. The positive dispersion of a positive refractive lens can be compensated by the negative dispersion of a positive diffractive lens etched onto the refractive lens, giving a simple one-element achromatic lens.

A broadband all-diffractive lens system can also be designed, contrary to what is widely believed to be true. The usual analysis considers only a system of individual binary optics lenses, as mentioned above. It has been shown, however, that a pair of separated diffractive optical elements can be designed to work together to form a broadband lens. Neither element forms a lens by itself. The design of this lens system does not follow the conventional approach of tracing rays from object to image, but the local grating spacing of each element is found by an iterative approach of tracing rays from object to image and back again.

A great advantage of binary optic lenses in conventional designs is that the lenses are equivalent to general aspheric elements. A binary optic can alter a given input wavefront to an arbitrary output wavefront. As a result binary optics are very useful for the final correction of the aberrations in the rest of an optical system. In

another approach to aberration control, aberrations can be corrected where they originate if every element is a hybrid diffractive-refractive element. This is of significant benefit to alignment and tolerancing in an optical system. Normally the aberrations of one element are corrected by opposite aberrations in other optical elements. This requires tight control of the relative positions of the optical elements in the refractive or reflective optical design, and also frequently requires many more elements than the diffractive-refractive design.

It is also seen in the grating law that a single incident wavefront results in multiple diffraction orders. Where the binary optic is used as a conventional lens, this may prove undesirable. The grating equation does not give the relative energy going into each diffraction order. The relative energy depends on the shape of the grating grooves, while the angles depend on the period of the grooves. It is possible to shape the surface within a grating period to control the distribution of the incident energy into the various diffraction orders. For a binary optic lens it is usually desirable to put one hundred percent of the energy into the first diffraction order.

For certain kinds of beam splitters and combiners it may be preferable to put equal amounts of energy into a specified number of diffraction orders. Optical multiplexers and applications for optical computers use these type of binary optics, sometimes referred to as Dammann gratings. In one application a number of diode lasers were positioned at angles corresponding to the diffraction orders of a Dammann grating. The separate incident beams were combined by the Dammann grating to obtain a single temporally and spatially coherent beam.

Arrays of "microlenses" can be conveniently fabricated as binary optics. An array of positive lenses can be positioned close to an array of negative lenses of equal and opposite power. A slight displacement of one array with respect to the other will result in a large change in the beam angle of an incident beam. Another beam steering technique uses a binary optic element which is like a blazed diffraction grating of uniformly varying grating spacing. As the binary optic is moved in a beam of light, the beam is steered. The functions of beam steering and focussing can also be combined. A third beam steering technique uses a pair of rotating binary optic elements that acts like a pair of (Risley) prisms.

The surface of a refractive optical element will reflect a portion of the incident beam because of the abrupt change in refractive index. To reduce these reflections a conventional refractive element is coated with a thin "anti-reflection" film. Another way to eliminate reflection losses is to etch the surface with a very fine diffraction grating. As seen from the grating law, equation (1), if the grating spacing is small enough, then the sine of the diffracted angle is greater than one for all nonzero diffraction orders. Every nonzero diffraction order in this case forms an evanescent wave which carries away no energy. As a result, all energy goes into the undeviated zero order, and the surface acts as an antireflection layer. This binary optic antireflection surface profile can be superimposed on other binary optics profiles for a combined effect. Using similar methods, binary optics can be formed to make optical filters to reflect or transmit particular wavelength bands.

Applications of binary optics appear to be limited at this time only by the knowledge and experience of the optical designer. Binary optics promise low cost fabrication and assembly with the additional benefits of reducing the size and weight of an optical system. Knowledge of binary optics is rapidly spreading as the Optical Society of America and the Society of Photo-Instrument Engineers (SPIE) sponsor several popular conferences each year on the subject. Applications which depend on the unique attributes of binary optics will probably be developed first; other applications will wait until optical designers become more comfortable and familiar with the technology. In some cases better analytical and manufacturing tools are also needed.

2. DIFFRACTION THEORY

Binary optics operate on the principle of diffraction. In addition, scattering can be best understood in terms of diffraction. In this chapter we review diffraction theory as it relates to scattering from binary optics. A vector theory based on Maxwell's electromagnetic wave equations gives the most accurate results, but there are exact solutions for only a few simple problems. Numerical solutions are very computation intensive and are generally restricted to simple periodic structures. Thus the rigorous vector theory is not well suited to the problem of scattering from binary optics, especially when the binary optics have random fabrication errors.

Fortunately the scalar theory can be made to yield accurate results for most practical problems. Much insight into the physics of scattering can be gained by examining the simple scalar solutions. Determining the appropriate diffraction theory is a question of current interest, and Chapter 4 sheds new light on this subject. In the present chapter we review the derivation of several famous and important scalar diffraction formulas which we will use many times in the rest of this dissertation. The derivation of the formulas helps us to understand the limitations of the scalar theory.

One of the limitations of the scalar theory is that it assumes that fields make abrupt changes at edges and walls. We know that physically this is not the case. Since binary optics have so many edges and walls it is important to know whether these features cause scattering. To answer this question we must turn to a more

rigorous vector theory. In the last part of this chapter we review what the rigorous EM vector theory says about edges, walls and corners.

2.1 Scalar Diffraction Theory

2.1.1 Early Theories

If we look at an aperture from just inside the geometric shadow, we see diffracted light that appears to come from the edge of the aperture. Sir Isaac Newton was probably influenced by this observation when he formulated his theory of the diffraction phenomena. Newton assumed that the diffracting edge exerts forces on the light corpuscles passing in its neighborhood.

Dr. Thomas Young proposed a wave theory version of Newton's theory. In Young's theory the incident light passes undisturbed through the diffracting aperture, with some light rays hitting the diffracting edge and being scattered. Interference of the diffracted wave (the light scattered from the edge) and the geometric-optical wave give the diffraction pattern. Augustin Jean Fresnel believed that diffraction came about as the propagation of "secondary wavelets" in the aperture, following the concepts of Christian Huygens, to which Fresnel added the interference principle. Fresnel convinced Young in 1818 to give up his theory and accept the Huygens principle. There does not seem to be any physical justification for the existence of secondary wavelets, especially if the ether, as assumed by Huygens and Fresnel, does not really exist.

2.1.2 Kirchhoff Theory

Fresnel's theory of diffraction contains a number of ad hoc assumptions about the effective amplitudes and phases of Huygens' secondary sources. In 1882 Gustav Kirchhoff put the theories of Huygens and Fresnel on a better mathematical basis. The "Fresnel-Kirchhoff" diffraction formula became very popular. Kirchhoff begins his development of diffraction theory with the Helmholtz equation which previously had been derived for acoustics. The Helmholtz equation is:

$$(\nabla^2 + k^2)U = 0 \tag{1}$$

where ∇^2 is the laplacian operator,

$$\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2, \tag{2}$$

U is a complex function of position and

$$k = 2\pi/\lambda. \tag{3}$$

Next Kirchhoff uses Green's theorem which relates the field in a certain volume to any surface surrounding that volume. If U and G are two complex-valued functions of position, S is a closed surface surrounding a volume V, and if U and G and their first and second partial derivatives are single-valued and continuous within and on S, then we can write Green's theorem as

$$\iiint_V (G\nabla^2 U - U\nabla^2 G)\,dv = \iint_S (G\partial U/\partial n - U\partial G/\partial n)\,ds. \tag{4}$$

The choice of the "Green's function," G, is somewhat arbitrary; Kirchhoff chose

$$G(P) = \exp(ikr)/r \tag{5}$$

where r is the distance between an arbitrary point P and the observation point. This Green's function is the expression of an expanding spherical wave.

Kirchhoff then assumes that the illuminating light field, $U(P)$, and its normal derivative within an aperture are exactly the same as they would be in the absence of the aperture. He also assumed that just behind the aperture, in the geometrical shadow, the light field and its normal derivative are exactly zero. The field U is assumed to be a spherical wave with a point of origin a distance r' from the point P. The field $U(P_o)$ at the observation point can then be written (after a fair amount of mathematical manipulation) as

$$U(P_o) = (1/4\pi) \iint U(P)G(P)[\cos(\mathbf{n,r'})(ik - 1/r') - \cos(\mathbf{n,r})(ik - 1/r)] \, ds \quad (6)$$

where $\cos(\mathbf{n,r'})$ is the cosine of the angle between the outward normal $\mathbf{n}$ of the aperture and the vector $\mathbf{r'}$ connecting the origin of the spherical wave $U(P)$ and the point P. The integral is taken over the aperture. If the distances r and r' are large compared to the wavelength λ, then equation (6) can be simplified to the Fresnel-Kirchhoff diffraction formula,

$$U(P_o) = (1/2i\lambda) \iint U(P)G(P)[\cos(\mathbf{n,r'}) - \cos(\mathbf{n,r})] \, ds. \quad (7)$$

2.1.3 Sommerfeld Theory

The Kirchhoff theory is not self-consistent as was pointed out by Sommerfeld in 1894. If a field and its normal derivative are both zero over some region, then the field is zero everywhere. Sommerfeld was able to remove the inconsistencies in the Kirchhoff theory by choosing a different Green's function. Then in Sommerfeld's theory no assumption of the normal derivative in the shadow region is necessary, but all the other assumptions remain. Sommerfeld chose a function $G'(P)$ representing

the difference between a spherical wave coming from the observation point and the mirror image of that point on the other side of the aperture,

$$G'(P) = \exp(ikr)/r - \exp(ikr_2)/r_2. \qquad (8)$$

For a point in the aperture

$$r = r_2,$$

$$\cos(\mathbf{n},\mathbf{r}) = -\cos(\mathbf{n},\mathbf{r}_2),$$

and therefore in the aperture

$$G'(P) = 0,$$

$$\partial G'(P)/\partial n = 2\cos(\mathbf{n},\mathbf{r})(ik - 1/r)G(P).$$

The Sommerfeld derivation does not require an assumption for $\partial U/\partial n$ because this term is multiplied by G', which is zero in the aperture. The final result is the general Rayleigh-Sommerfeld diffraction formula:

$$U(P_o) = (1/2\pi) \iint U(P)G(P)\cos(\mathbf{n},\mathbf{r})(1/r - ik) \, ds. \qquad (9)$$

This is sometimes simplified by assuming $r >> \lambda$, giving the approximate Rayleigh-Sommerfeld formula,

$$U(P_o) = (1/i\lambda) \iint U(P)G(P)\cos(\mathbf{n},\mathbf{r}) \, ds. \qquad (10)$$

The only difference between the resultant Kirchhoff diffraction formulas (6) or (7) and the Sommerfeld diffraction formulas (9) or (10) is the so-called "obliquity factor", the part of the formula giving the relative importance of the angle between the light source and a point in the aperture as seen from the observation point. It is not known which of the chosen Green's functions give the best results.

In the Fresnel approximation the obliquity factor is approximately one, that is,

$$\cos(n,r) = 1,$$

which is approximately true if the distance, z, between the aperture plane and the observation plane is much greater than the maximum dimension of the aperture and the distance from the z-axis to the observation point. A further Fresnel approximation, which is true under similar conditions, is that the distance r in the denominator of $G(P)$ is equal to z. In the exponent of $G(P)$ the approximation is

$$r = z + (x_0 - x_1)^2 / 2z + (y_0 - y_1)^2 / 2z,$$

where (x_0, y_0) are position coordinates in the plane of observation and (x_1, y_1) are position coordinates in the plane of the aperture. Applying these approximations gives the Fresnel diffraction formula,

$$U(x_0, y_0) = \exp(ikz)/iz\lambda \iint U(x_1, y_1) \exp\{ik[(x_0 - x_1)^2 + (y_0 - y_1)^2]/2z\} \, dx_1 dy_1. \quad (11)$$

In the Fraunhofer approximation the more stringent assumption is made that

$$z >> k(x_1^2 + y_1^2)/2.$$

Equation (11) can then be simplified to

$$U(x_0, y_0) = \exp(ikz) \exp[ik(x_0^2 + y_0^2)/2z]/iz\lambda$$

$$\iint U(x_1, y_1) \exp[-ik(x_0 x_1 + y_0 y_1)/z] \, dx_1 dy_1. \quad (12)$$

The integral in equation (12) can be regarded as the Fourier transform of $U(x_1, y_1)$, with the substitution of $x_0/\lambda z$ and $y_0/\lambda z$ for the frequency variables. Fourier optics is a powerful and relatively convenient and simple method to use to solve diffraction problems, but it is limited to the Fraunhofer approximation. In Chapter 4 a new

formulation of Fourier optics is derived which gives the accuracy of the Rayleigh-Sommerfeld formula but is just as simple to use as traditional Fourier optics.

2.1.4 Application to Binary Optics

Suppose we wish to determine the light field at some point of observation behind a binary optic. In order to use the Kirchhoff or the Sommerfeld diffraction formula we must specify the field on some surface representing the binary optic. In practice the binary optic is assumed to have zero thickness; the incident wavefront phase profile is altered by the addition of a perfect replication of the surface relief profile. We ignore all of the things that might happen within the grooves: diffraction of the field in or out of the grooves; multiple scattering; shadowing; and matching of boundaries conditions along edges and walls.

The assumption of zero thickness is not necessary. One could define the diffracting aperture using the actual binary optic surface profile. The incident field would be assumed to propagate undisturbed to all points on the binary optics surface. This approach to Kirchhoff or Sommerfeld diffraction would lead to a more complicated integrand to evaluate, and would probably be no more accurate. There still remains a fundamental limitation of the Kirchhoff theory: the assumption that there is no disturbance of the field near the edges of a surface discontinuity.

2.1.5 Boundary Waves or Rays

It is convenient to think of Kirchhoff's formula (or Sommerfeld's) as the contribution of each point on the aperture as a source of "secondary wavelets".

However, Kirchhoff's formula can be rewritten to be the sum of two terms, one of which is the undisturbed illuminating wavefront. When this illuminating wavefront is a plane wave, or an isotropically divergent or convergent spherical wave, then the second term is a "diffracted wave" or "boundary wave" emanating from the edge of the aperture. This concept was originally developed by Gian Antonio Maggi in 1888, was further developed by Rubinowicz in 1917, and finally by Miyamoto and Wolf in 1962. Young's physical view of scattering of the wave from the edge of the aperture seems to have been justified.

Unfortunately this simple intuitive concept does not appear to be correct in all cases. For an arbitrary illuminating wavefront the second term of the rewritten Kirchhoff formula does not originate from the edge of the aperture, but from secondary wavelets originating all along the path from a point on the aperture edge to the observation point[1]. These boundary wave approaches are not more accurate in calculating the edge effects of apertures; they only present Kirchhoff's formula in a way which shows the relationship between edges and the diffraction pattern. In some cases using edge waves makes diffraction calculations easier. A 1962 Russian paper called the "Method of Edge Waves in the Physical Theory of Diffraction," reportedly made a 30-40% contribution to a Lockheed computer program used to design a stealth aircraft[2].

If the aperture edges "cause" the diffraction by reflection or scattering of some sort, then an infinite aperture should show no diffraction. In a recent paper by D.C. Bertilone the diffraction pattern from an infinite aperture was calculated[3]. The

illuminating wavefront was assumed to be converging, and at the "aperture plane" the beam had exact, finite boundaries. Bertilone's work shows that the Sommerfeld and Kirchhoff diffraction theories predict diffraction even when no physical aperture is present.

J.B. Keller has developed a geometrical theory of diffraction[4]. In Keller's theory "diffracted rays" are assumed to be generated by edges, walls and corners of an aperture or obstacle. The field at some observation point is determined by the combination of diffracted and undiffracted rays that reach the point. Apparently no one working with binary optics uses Keller's theory.

2.2 Electromagnetic Theory

2.2.1 Analytical Solutions

The preceding were "scalar" diffraction theories; they do not use Maxwell's equations relating the various EM vector components. In fact, Kirchhoff's assumptions contradict EM theory which says that there must be boundary conditions met along the edge of the aperture which would not be required in the absence of the aperture. Also EM theory says that some fields will extend behind the aperture for a few wavelengths. The first rigorous EM solution to a diffraction problem was given in 1896 by Sommerfeld[5]. He considered the case of an illuminating plane wave incident on a perfectly conducting, zero thickness, half-plane aperture. Interestingly, the solution in the illuminated region is the sum of two terms: the original incident wave, and a cylindrical wave that appears to proceed from the edge of the aperture.

In the shadow region only the cylindrical "boundary wave" propagates. Again we see evidence that Young's views contain an "element of truth."

It is possible to generalize Kirchhoff's theory to an EM theory of diffraction at a black screen, as shown by Friedrich Kottler[6]. Kottler essentially makes the same assumptions as Kirchhoff, but justifies it on the basis of solving a *saltus* or discontinuous problem at a black (i.e, a non-conducting, non-reflecting) aperture. Kottler finds pulsating line densities at the aperture edge, which explains the phenomenon of the apparent luminous line. In Kottler's scalar version of the theory, the field at the edge points cannot be calculated. He guesses that the edge points "must exhibit a large curvature and consequently a *large*, but *not infinite*, field intensity".

2.2.2 Numerical Solutions

There are several rigorous ways to numerically solve Maxwell's equations for diffraction gratings. Often a binary optic can be analyzed by treating it as a diffraction grating with a period that varies from point to point on the surface. So a review of the vector theory of diffraction gratings should be helpful in understanding the application of rigorous vector theory to binary optics.

One rigorous approach, the full-vector coupled wave analysis, divides the grating into several layers parallel to the grating plane:

> Each layer contains an electric field composed as an infinite sum of plane waves which is truncated at an appropriate high-frequency cutoff. The electric fields of each layer are then subject to boundary conditions imposed by Maxwell's equations.[7]

Since the emerging field is periodic, it is expressed as a Fourier series. Each term in the Fourier series is associated with a diffraction order. If the sum of the energy in the propagating diffraction orders does not equal the illuminating energy, then the calculation is assumed to be inaccurate. There can be no scattering of energy between diffraction orders as long as it is true that every periodic field can be expressed as a Fourier series. If the period is short compared to a wavelength, then only a few diffraction orders will propagate. The rest of the Fourier terms correspond to evanescent waves which carry away no energy.

The edges, which characterize a binary optic, sometimes present a difficulty for the vector theory. In a rigorous approach taken by Chandezon and others[8] there is a mathematical difficulty with edges. They get around this by approximating the profile by a truncated Fourier series. "We know that the new grating so defined, which has no edges, gives practically the same efficiency as the echelete [laminar] one, as soon as the number of Fourier coefficients exceeds about 10."[8] Their approach is based on coordinate translations and partial differential equations. A comparison with the integral method gave the same results.

In the coupled wave approach it is said that only "one layer fully characterizes a binary (two-level) grating".[7] This implies that the field does not change as it propagates from the substrate up between the walls of the grooves. But this is "in contrast to the actual situation", according to Whitman and others.[9] For ease of calculation they assume that the field is a continuous function along the lateral dimensions, with no "variations in current density along vertical parts of the surface

profile". They observe, however, that a moderate change in the shape of the surface profile has little effect on the distribution of energy in the far field.

2.2.3 Surface Plasmons

In metallic gratings an EM theory predicts some enhanced scattering when the grating period is approximately the same as the wavelength.[10] Evidently, surface plasmons are excited resonantly, particularly in "the presence of sharp edges in the profile of the grooves". Another study of metallic gratings concluded:

> What we have shown here is that the mere inclusion of roughness in an electromagnetic formalism of diffraction gratings allows us to predict the existence of one of the most interesting (from a theoretical point of view) spectral impurities [scattering]: that which is related with the absorption and re-radiation of a photon via surface plasmons.[11]

2.2.4 Questions Remain

In a recent panel discussion on grating technology the question was asked:

> "I'd like to echo Mike Hutley's comments about knowing what's happening at the grating surfaces. I would like to ask the French contingent, because their work is so good, if they can begin to tell us what's happening in some of these corners. One knows that there has to be some very large currents because of discontinuity. We would love to know what's happening."[12]

M. Neviere from the University of Marseille answered:

> "We have no physical explanation, but the theory indeed predicts, for example, that the bottom of a groove of a blazed grating has no influence on the efficiency curve while the top of the groove has a very important influence."

A couple of years later Mike Hutley was still wondering:

The description of anomalies in terms of surface plasmons and in terms of Brewster angle effects has given us some considerable insight into this but, at the time of writing, there still remain some gaps in our understanding. We still cannot pinpoint which features of a grating (symmetry of profile, straightness of facets, sharpness of corners, roughness of surface, etc.) are likely to induce anomalies.[13]

2.2.5 Comparison: Scalar and EM Theories

A comparison between the results of scalar and EM diffraction theory ought to indicate the effect of edges, walls and corners. There have been some comparisons made in the literature, but usually the scalar theory utilized is Fraunhofer or Fourier optics.[14,15] When a difference is noted between the vector theory and the "scalar" theory the author has set up an example where the assumptions of the Fraunhofer theory are not met. The authors should have used the Rayleigh-Sommerfeld theory rather than the Fraunhofer. These comparisons should therefore be taken with a grain of salt.

In general when the smallest feature size approaches the wavelength of light then Fourier optics theory no longer gives results consistent with EM theory. This is also supported by a recent study that shows that both the field amplitude and phase are severely altered within a half wavelength of the edges and walls.[16] In the study by Vasara et. al., the Fourier optics theory, when compared to a rigorous EM theory, gave a different distribution of energy among diffraction orders for binary optic gratings with sufficiently small features.[14] The EM theory did not predict, however, a loss of energy or scattering of energy. In fact, it was possible to design the binary optic to put nearly 100 percent of the incident energy into the desired

diffraction orders if the EM theory was used in the design. In Chapter 4 we explore in depth the relationship between the various theories, going beyond the work of other described in this chapter.

2.3 Summary

Scalar diffraction theory is a simplification of electromagnetic vector theory, where each component of the vector is assumed to act independently of the other components. The order of exactness among the scalar diffraction formulas, from the most to the least exact, is Rayleigh-Sommerfeld, Fresnel-Kirchhoff, Fresnel, and Fraunhofer. Conventional Fourier optics is based on the Fraunhofer theory; we assumed that (1) the distance from the aperture to the observation plane is much larger than the aperture diameter, and (2) that we are only interested in the field close to the optical axis. Typically for binary optics, the assumption is also made that the field as it exits the binary optic is an exact copy of the surface profile.

In this chapter we have seen that further research is needed for the following problems:

1. Find a way to express the general diffraction problem in terms of aperture or edge waves.

2. Compare the results of Fresnel-Kirchhoff, Rayleigh-Sommerfeld, edge waves, Keller's theory, and EM theories to experiments.

3. Compare the results using the zero thickness assumption to the finite thickness diffraction problem for binary optics.

4. Determine whether the field changes as it propagates up between the walls of a binary optic.

5. Determine if post polishing to remove corners decreases the scattering from a binary optic.

6. Investigate further the role of evanescent waves in scattering.

Electromagnetic theory shows that edges, walls, and corners affect the phase and amplitude of nearby fields, contrary to the assumptions of scalar diffraction theory. This disruption of the field will merely redistribute energy among diffraction orders for periodic surfaces. The finite size of an aperture or beam causes a disruption of the field leading to broadening of each diffraction order. Scattering between orders exists, according to the EM theory of surface plasmons, when the periodic edges in metallic gratings concentrate light energy which emerges from non-periodic features (i.e. random roughness). The theory of surface plasmons has been developed for metallic gratings with free electrons at the surface which can propagate energy along the surface. We shall see in Chapters 3 and 6 that the key to understanding interorder scattering in binary optics is to understand the random roughness or the random misplacements of edges or other surface features. Many questions remain about the effects of certain features in a grating.

3. REVIEW OF SCATTERING THEORY

For a conventional optical element, such as a mirror, scattering comes from surface roughness or inhomogeneities of surface constants. These surface variations are usually random and are best treated by statistical analysis. In some cases the method of manufacture leads to periodic or quasi-periodic surface features. Single-point diamond turning is an example of a manufacturing process that leaves the surface with accidental periodic surface heights. In these cases the surface bears some resemblance to a diffraction grating and there will be distinct diffraction orders. A randomly rough surface can be regarded as the superposition of an infinite number of diffraction gratings, each with a different grating spacing and hence a different first order diffraction angle. If the *rms* roughness of the surface is much smaller than the wavelength of light, then only the first diffraction order is of any consequence.

In a binary optic the second and higher diffraction orders are important because the depth of the pattern is about one wavelength. In some cases these higher orders of diffraction are desirable; but in many cases anything except the first order can be considered "scattered" light. Scattering from binary optics will also come from the inevitable random surface roughness, just as it does from conventional optics. The fabrication of binary optics calls for extremely precise positioning and control of etching to insure that each groove is the proper width and depth, hence limitations in the accuracy of fabrication are another source of scattering. Errors of

fabrication might be either systematic (deterministic) or random. In this chapter we review existing scattering theories as they might be applied to binary optics.

3.1 Diffraction Efficiency, Design Limitations

An understanding of diffraction gratings is valuable in understanding binary optics. In this section we review the derivation of the grating equation, equation (1.1), to illustrate the origin of diffraction orders. Then we derive approximate equations that give the relative amount of energy in each diffraction order for a number of types of binary optics.

3.1.1 Derivation of the Grating Equation

Consider a grating consisting of a number of thin slits spaced a distance a apart. A plane wave is incident on the grating at an angle θ_1 between the normal to the wavefront and the normal to the grating, as shown in figure (3.1). Diffraction spreads the wavefront after passing through a slit. We consider an observation point a great distance from the grating, so that rays from adjacent slits A and B reach the observation point after being diffracted by essentially the same angle θ_2. The ray from B travels a greater distance than the ray from A. That distance is given by $a \sin\theta_1 + a \sin\theta_2$. The relative phase difference ϕ will therefore be

$$\phi = 2\pi \, (a \, \sin\theta_1 + a \, \sin\theta_2)/\lambda. \qquad (1)$$

These wavefronts will constructively interfere if ϕ is equal to an integer m times 2π. Substituting this value for ϕ gives the grating equation (1.1). The values of m give the diffraction orders.

3.1.2 Slit Grating

The amount of energy in each diffraction order depends on the way that each slit "scatters" the energy, but not on the spacing of the grating. A small slit should diffract a higher proportion of energy into large diffraction orders than a large slit. In a given slit a ray that passes near one slit edge will arrive somewhat out of phase with a ray that passes near the other slit edge, though it will be in phase with corresponding rays in every other slit. Suppose now that each slit is filled with a substance that equalizes the optical path lengths within the slit for a particular diffraction order, say order p. A prism that refracts in the direction of the pth diffraction order would be an example of such a substance. All the rays now arrive in phase at the distant observation point corresponding to order p. The result is a "blazed" grating. In a diffraction order m', where $m' \neq p$, the rays across the slit tend to be out of phase with each other and destructively interfere. As a result, most of the transmitted energy is found in the pth diffraction order.

To make these physical ideas more mathematical we use the Fraunhofer diffraction integral, equation (2.12), from Chapter 2. (The development here is similar to that of Born and Wolf[1], pp. 401-405). We consider a grating with N identical slits, spaced a distance a apart. If the grating is illuminated by a plane wave with angle of incidence θ_1, then the phase of the wavefront in corresponding points in an adjacent slit will be different by the constant term $\exp[-ik(a \sin\theta_1)]$. In general there will be a phase altering substance in the slit which produces the field $U(x_1,y_1)$

across each slit. The grating is in the x_1y_1 plane, with the slits parallel to the y_1 axis. We can then write the field $U(x_0, y_0)$ at the observation point as

$$U(x_0, y_0, z) = C \, \Sigma_j \iint U(x_1, y_1)$$

$$\exp\{-ik[ja \sin\theta_1 + (x_0(x_1 + ja) + y_0y_1)/z]\} \, dx_1 dy_1, \qquad (2)$$

where C is the constant in front of the integral in equation (2.12), and the integration is over any one slit aperture.

Equation (2) can be rewritten as

$$U(x_0, y_0, z) = U^0(x_0, y_0, z) \, \Sigma_j \exp(-ikja(\sin\theta_1 + x_0/z))$$

$$= U^0(x_0, y_0, z) \{1 - \exp[-ikNa(\sin\theta_1 + x_0/z)]\}/\{1 - \exp[-ika(\sin\theta_1 + x_0/z)]\}, \quad (3)$$

where $U^0(x_0, y_0, z)$ is the field from a single slit and is given by

$$U^0(x_0, y_0, z) = C \iint U(x_1, y_1) \exp[-ik(x_0x_1 + y_0y_1)/z] \, dx_1 dy_1$$

$$= C \, F\{A(x_1, y_1)\}^{**}F\{U(x_1, y_1)\}|_{\xi = x0/\lambda z, \, \eta = y0/\lambda z}, \qquad (4)$$

where $F\{U\}$ is the Fourier transform of U; ξ and η are the spatial frequencies associated with x_1 and y_1 respectively. $A(x_1, y_1)$ is the aperture function; it has the value of one for a point (x_1, y_1) in the slit aperture, and a value of zero elsewhere. The operation "**" is convolution.

Since the intensity is the product of the field and its complex conjugate, we take the result from equation (3) and obtain

$$I(x_0, y_0, z) = I^0(x_0, y_0, z) \, S(N, ka[\sin\theta_1 + x_0/z]/2), \qquad (5)$$

where the function $S(N, x)$ is defined by

$$S(N, x) \equiv (\sin Nx/\sin x)^2. \qquad (6)$$

It is to be noted that equation (5) expresses the intensity as the product of two functions: I^0 represents the intensity from a single slit, and $S(N, ka[\sin\theta_1 + x_0/z]/2)$ represents the interference of light from different slits. $S(N, ka[\sin\theta_1 + x_0/z]/2)$ has a maximum when the denominator goes to zero, or where

$$ka(\sin\theta_1 + x_0/z)/2 = m\pi \tag{7}$$

when m is integer. Since $x_0/z = \sin\theta_2$ in the approximation we are considering, equation (7) is equivalent to the grating equation, (1.1).

3.1.3 Blazed Gratings

To represent a blazed grating we consider in a slit the linear phase function

$$U(x_1, y_1) = \exp[-ik(\sin\theta_1 - p\lambda/a)x_1], \tag{8}$$

where p is an integer. Substituting this in equation (4) gives

$$U^0(x_0, y_0, z) = C \ \delta(x_0/\lambda z + \sin\theta_1/\lambda - p/a)**F\{A(x_1, y_1)\}\big|_{\xi = x0/\lambda z, \ \eta = y0/\lambda z} \tag{9}$$

where $\delta()$ is Dirac's delta function. Hence $U^0(x_0, y_0, z)$ is the Fourier transform of the aperture function but shifted so that the center of that function is in the direction given by

$$\sin\theta_1 - p\lambda/a + x_0/z = 0, \tag{10}$$

which is the direction of the pth diffraction order. All of the energy will be in the pth diffraction order and the only scattered light will be the diffraction spreading due to the slit aperture, $F\{A(x_1, y_1)\}$.

3.1.4 Dammann Gratings

We now consider a binary optic designed to put a specified amount of energy into particular diffraction orders. Energy going into other orders is scattered light. A Dammann grating has a simple two level phase profile that diffracts most of the incident energy into the first few diffraction orders. The grating can generate one dimensional or two dimensional arrays of equal intensity spots. The name comes from H. Dammann who first proposed this type of grating in 1971.[2] We consider a one-dimensional symmetrical Dammann grating. We could proceed in an analysis that parallels that of the slit grating, but an alternate method will be instructional.

The M (per half period) transition points, where the pattern changes abruptly from an upper level to a lower level or from a lower level to an upper level, are labelled $\{a_1,...,a_M\}$. We use the convention that $a_0 = 0$ and $a_{M+1} = a/2$ where a is the period length. The phase function can be written as

$$U(x_1) = \Sigma_{j=0} \exp(i\phi(x_1)) \, \text{rect}\{[x_1 - (a_{j+1} + a_j)/2]/(a_{j+1} - a_j)\}. \qquad (11)$$

The phase term $\phi(x)$ is given by

$$\phi(x) = \begin{cases} 0 & \text{if } a_j \leq x \leq a_{j+1} \text{ and } j \text{ is even} \\ \phi_0 \equiv 2\pi H(n-1)/\lambda & \text{otherwise.} \end{cases} \qquad (12)$$

The depth of the grating is H, λ is the wavelength of light, and n is the index of refraction.

We can determine the relative amplitude A_m of the wavefront in each diffraction order m if we assume that the grating is infinite in the x_1-direction and write $U(x_1)$ in a Fourier series:

$$U(x_1) = \Sigma_{m=-\infty} A_m \exp(2\pi i m x_1/a) . \qquad (13)$$

Solving for A_m,

$$A_m = (2/a) \int_0^{a/2} U(x_1) \cos(2\pi m x_1/a) \, dx_1 . \qquad (14)$$

By direct integration we find that

$$A_0 = \begin{cases} \exp(i\phi_0) + (2/a)[1 - \exp(i\phi_0)] \, \Sigma_{j=1} \, (-1)^{j+1} \, a_j & M \text{ odd} \\ 1 + (2/a)[1 - \exp(i\phi_0)] \, \Sigma_{j=1}(-1)^{j+1} \, a_j & M \text{ even} \end{cases} \qquad (15)$$

$$A_m = (1/\pi m)[1 - \exp(i\phi_0)] \, \Sigma_{j=1} \, (-1)^{j+1} \, \sin(2\pi m a_j/a) \quad m \neq 0. \qquad (16)$$

Equations (15) and (16) agree with the results of Krackhardt and Streibl[3] in the case when $a = 1$ and $\phi_0 = \pi$, but contradict the results of others.[4-6]

Normally a Dammann grating is designed by numerically solving equations (15) and (16) for the transition points $\{a_j\}$ so that the intensities $I_m = |A_m|^2$ are equal for $m = 0, \pm1, \pm2$ and up to 25 or more. Only occasionally is ϕ_0 used as a variable. The amount of "scattered" energy in any diffraction order can be calculated from equations (15) and (16), once the transition points and etch depth have been determined. In a Dammann grating the scattered light results in an overall efficiency loss, but at least it is diffracted into angles that do not overlap. In the phase-only Fresnel zone plate that is not the case.

3.1.5 Fresnel Zone Plates

We consider a phase-only Fresnel zone plate which is designed to convert a normally incident plane wave to a converging spherical wave. The wave converges to a point a distance f from the zone plate. We may think of the Fresnel zone plate

as a diffraction grating with a grating period a that decreases as a function of x, the distance from the optical axis. For the first order diffracted light to pass through the focus at f, we use the grating equation to obtain

$$\lambda = a(x) \sin\theta_2$$

or

$$a(x) = \lambda/\sin\theta_2 = \lambda f/x \qquad (17)$$

using the paraxial approximation. The angle of diffraction of the mth order also depends on the distance x. From the grating equation and equation (17) we have

$$\sin\theta_{2,m} = m\lambda/a(x) = x/(f/m). \qquad (18)$$

We see from equation (18) that the mth diffraction order comes to a focus at a distance f/m from the zone plate. Each higher order expands, after passing through its focus, and contributes to the scattered light in the first order focal plane.

The amount of scattered light depends on the relative amplitudes of the diffraction orders, which in turn depends on the "groove" profile within a grating period. We consider an M-level Fresnel zone plate. Within each grating period or zone there are M steps in a "staircase". Between the $(j-1)$th and the jth step the phase changes by $2\pi/M$ radians. The surface changes phase at the transition points $\{a_j\}$, hence the width of the jth step is $a_j - a_{j-1}$. For a Fresnel zone plate of focal length f and wavelength λ, the jth transition point is located where the distance to the focal point is $\lambda j/M + f$, to approximate a converging spherical wavefront. Hence

$$a_j^2 + f^2 = (\lambda j/M + f)^2$$

or

$$a_j^2 = 2\lambda jf/M + (\lambda j/M)^2 \ . \tag{19}$$

For a Fresnel phase-only zone plate the fraction of light that is not diffracted into the desired order (usually the first order) is equivalent to the total integrated scatter.

3.1.6 General Binary Optics

The diffraction efficiency, or fraction of light in a particular order, will now be derived for binary optics in general. We must make a number of assumptions:

1. The diffraction can be calculated by the Fraunhofer diffraction formula, using the paraxial approximations.

2. The diffraction efficiency is calculated for a single groove or zone and the rest of the binary optic is supposed to extend to infinity with the same groove form and "grating spacing".

3. Within a single "period" or zone all step widths and heights are equal. Considering all the approximations, it is amazing that the formula works as well as it does.

The field in a single grating "slit", zone, period, or groove is

$$U(x_1, y_1) = \Sigma_j \ \text{Rect}\{[x_1 - (a_{j+1} + a_j)/2]/[a_{j+1} - a_j]\} \ \exp(\phi ij/M), \tag{20}$$

or, since $a_{j+1} - a_j = a/M$ is assumed to be a constant (assumption 3 above),

$$U(x_1, y_1) = \Sigma_j \ \text{Rect}[Mx_1/a - j - 1/2] \ \exp(\phi ij/M), \tag{21}$$

where ϕ/M is the phase change per step. The far-field pattern is the Fourier transform of $U(x_1, y_1)$; in other words, using equations (4) and (21), the field from a single period is

$$U^0(x_0, y_0, z) = C \, a/M \, \text{sinc}(a x_0/Mz\lambda) \, \Sigma_j \, \exp[-2\pi i (j + 1/2) a x_0/Mz\lambda + \phi ij/M]. \quad (22)$$

Hence

$$I^0(x_0, y_0, z) = [C \, a/M \, \text{sinc}(a x_0/Mz\lambda)]^2 \, S(M, \pi a x_0/Mz\lambda - \phi/2M) \quad (23)$$

where S was defined by equation (6).

In the pth diffraction order, paraxial approximation,

$$p\lambda = a x_0/z$$

so using equations (5) and (23) the intensity in the pth order, summed over the whole

grating, is

$$I(p) = [C \, a/M \, \text{sinc}(p/M)]^2 \, S(M, \pi p/M - \phi/2M) \, S(N, p\pi), \quad (24)$$

where we have used the assumption that this zone is repeated N times with period

a. As N goes to infinity the function $S(N, p\pi)$ becomes narrower and narrower until

we can say that the relative energy in a diffraction order is the same as the relative

intensity. The relative diffraction efficiency $\eta(p, M)$ of order p and with M levels is

therefore,

$$\eta(p, M) = \{[\sin(\pi p/M) \sin(\pi p - \phi/2)]/[\pi p \, \sin(\pi p/M - \phi/2M)]\}^2. \quad (25)$$

This result has been derived by others.[7] We note that the highest efficiency for order

p occurs when $\phi/2 = \pi p$.

3.2 Systematic Fabrication Errors

So far we have considered scattering that comes from energy diffracted into

unwanted orders as a result of the design. To be eliminated one must find a new

design. There are other sources of scattering in binary optics that depend on the

exactness of the manufacturing process. There are periodic or repeated fabrication errors that could be called systematic, and there are randomly generated fabrication errors. We now discuss the systematic type of error.

3.2.1 Gratings

Several types of systematic errors have been investigated in the past in connection with diffraction gratings. When a groove position in the grating varies from the correct position, some energy will be scattered. *Periodic* variations result in concentrating scattering into spurious diffraction orders called *ghosts*. Ghosts are common in ruled gratings because of periodic errors in turning screws and so forth. An approximate value of the intensity of such ghosts relative to the main diffracted order was given by Rowland in 1893. As presented by Hutley[8]

$$I_{ghost}/I(m) = (\pi m e/a)^2 \qquad (26)$$

where m is the diffraction order, e is the maximum departure from the correct position, and a is the period of the grating. The location of the ghost will depend on the period of the error.

As an example of equation (26) if $e = 0.1\mu m$, $a = 1.0\mu m$, $m = 1$, then the ghost intensity is approximately ten percent of the main diffraction order. We note in equation (26) that the ghost intensity is proportional to the square of the diffraction order. There should be no ghosts around the zero or central order. Random variations in the position of the grooves result in "grass," a broad spectrum of ghosts of random amplitudes in the plane of the diffraction orders. Scattering

from surface roughness, on the other hand, has no preferred orientation and is just as intense out of the plane of incidence as in the plane of incidence.

Another type of systematic error in a grating is called fan-out error, where the grooves are not parallel to each other but the spacing between grooves increases in the direction of the grooves. In a sense we have a series of gratings each with a different grating period. The effect will be to spread out each diffraction order. A small amount of fan-out could greatly reduce the peak intensity of the diffraction order.

3.2.2 Dammann Gratings

The formula for variations in the groove position assumes that the form of the groove remains unchanged. A uniform change in the groove form changes the distribution of energy among the diffraction orders as we have already seen. As an example, we note from equations (15) and (16) that a change in ϕ_0 will change A_0/A_m but does *not* change A_j/A_m when $j,m \neq 0$. If the Dammann grating is accidentally etched uniformly deeper than called for by the design, then the central diffraction order will have a different amplitude than the other diffraction orders, but the other diffraction orders will keep the same relative intensities with respect to each other.

3.2.3 Quantization Error

A systematic error in binary optics, not found in ruled or holographic gratings, is the quantization error. This error occurs because many mask-making machines and digital graphing devices are capable of making only rectangles of discrete size

and position. The approximation of the rectangles to the desired surface relief pattern leads to scattering. The effect of the quantization error can be estimated by modeling the quantization as additive noise.[9] The root mean square phase error is given by

$$\phi_{rms} = (\pi/\sqrt{3})\ e/a, \qquad (27)$$

where a is the period in the direction of the sides of the rectangle, and e is the quantization step size in one dimension of the rectangle, or the minimum flash box size in the orthogonal dimension. With present technology the minimum step size is 0.1μm and the minimum flash box size is about 1 μm.

3.3 Random Fabrication Errors

From the theory of gratings a random change in form from groove to groove is called "accidental error of amplitude", and is a form of random error independent of grating order.[10] A random error in the position of the groove causes "grass", as mentioned previously. The proportion of the energy scattered into grass is given by[8]

$$I_{grass}/I(m) = (4\pi m e_{rms}/a)^2, \qquad (28)$$

where e_{rms} is the *rms* error in groove position. According to this formula grass increases as the square of the diffraction order, just like the intensity for ghosts.

In practice no optical surface is perfectly smooth; there will be some roughness. The scattering in reflection from a slightly rough surface is given by the two-dimensional bidirectional reflection distribution function (BRDF). For normal

incidence and assuming an exponential autocorrelation function for the statistics of the random roughness, the BRDF is given by[11]

$$\text{BRDF} = (2k^4/\pi)\,(\sigma\tau)^2\,\cos\theta_s\,F(\theta_s)\,/[1 + p^2\tau^2]^{3/2}, \qquad (29)$$

where $F(\theta_s)$ is approximately one for small angles, and is exactly one in some theories. The *rms* roughness is σ, the correlation length is τ. The term p is the spatial frequency variable defined by

$$p \equiv k(\sin\theta_s - \sin\theta_2).$$

The specular angle is θ_2, the scattering angle is θ_s.

Equation (29) was derived for slightly rough reflective surfaces. This existing state-of-the-art formula is not really appropriate for binary optics. How is the scattered light distributed around each diffraction order? How do we modify the formula for transmissive optics? Do the large step heights in a binary optic alter the scattering from surface roughness? Can random processing errors be thought of as random roughness and the scattering easily calculated? These questions will be answered in Chapter 6 when a new formula will be derived that is appropriate for binary optics.

3.4 Summary

In this review we have seen that several formulas have been derived to predict the diffraction efficiency of gratings. The most important of these formulas is equation (25). It is assumed in the derivation of these formulas that the Fraunhofer diffraction formula is adequate, that the grating has a constant period that extends

to infinity, and that all step widths and heights are equal. These formulas assume that the diffraction grating is fabricated exactly as designed.

For systematic fabrication errors there have been only two formulas: (1) a formula which gives the intensity for ghosts when there is a periodic variation in the groove position, and (2) a formula which gives the *rms* phase error of quantization errors. It is relatively easy to determine the scattering from systematic errors; the energy is distributed in the various diffraction orders. A knowledge of the diffraction efficiency, the grating equation, and geometrical optics will completely determine the scattering. For random errors we have a formula for slightly rough flat surfaces.

The question of how well these grating formulas apply to binary optics will be addressed in Chapters 4 and 5. The random roughness formula currently in use to predict scattering is not adequate for binary optics. A new formula will be derived in Chapter 6.

62

FIGURE 3.1, Derivation of the grating equation.

4. COMPARISON OF DIFFRACTION THEORIES

It is important to determine the conditions under which the various diffraction theories give valid results. To do this we will compare the rigorous vector theory to the Rayleigh-Sommerfeld theory for diffraction gratings. We are limited to diffraction gratings in this comparison because the vector theory is too difficult to calculate for more complicated structures. We then compare the various scalar formulas for a conventional lens and a Fresnel zone plate.

The diffraction efficiency will be the parameter of interest in this chapter. For an infinite, perfect grating, the diffraction efficiency is just the square of the field amplitude in a particular diffraction order. When the size of the grating is finite there is some ambiguity; do we mean the fraction of incident energy which is spread around a diffraction order, or do we mean $I(m)/I_0$, where $I(m)$ is the peak intensity in the mth diffraction order, and I_0 is the peak intensity assuming far-field approximations and 100 percent of the incident energy is diffracted into that order? In this chapter we shall be using this relative peak intensity.

In the far-field, nearly on-axis case, the intensity profile of a beam diffracted by a grating is just a scaled down version of the intensity profile that results from diffraction of the aperture. In this case, the relative peak intensity is a good measure of the fraction of energy in the diffraction order. Off-axis the intensity profile is distorted by the obliquity factor, and close to the binary optic the intensity profile of

higher diffraction orders is distorted by aberrations. In these cases the relative peak intensity is not a good measure of energy.

Until now we have reviewed the works of others. We have seen that the existing theories are inadequate for binary optics. Beginning with this chapter original formulas will be derived. To simplify the calculations we first derive a novel extension of Fourier optics which retains the accuracy of the Rayleigh-Sommerfeld formula. This formula is compared to vector theory calculations as calculated by several different researchers.

4.1 Extended Fourier Optics

4.1.1 Derivation of New Formula

We now derive an extended Fresnel or Fraunhofer formula that removes the limitation that the observation point must be close to the normal to the aperture. We can then use the much simpler formulas of Fourier optics and retain the accuracy of the Rayleigh-Sommerfeld. A comparison of this new formula with numerical integration of the Rayleigh-Sommerfeld diffraction formula shows it to be extremely accurate in all cases where the aperture is small compared to the distance to the observation point. To derive the formula we begin with the Rayleigh-Sommerfeld formula [equation (2.9)],

$$U(x_0, y_0, z) = (1/2\pi) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1, y_1) \cos(n,r) (1/r - ik) \exp(ikr)/r \, dx_1 \, dy_1, \quad (1)$$

where cos(**n**,**r**) is the cosine of the angle between the normal to the aperture and the vector **r** which goes from a point in the aperture to the observation point (x_0, y_0). The z-axis is taken as the normal to the aperture. The magnitude of **r** is r given by

$$r^2 = (x_0 - x_1)^2 + (y_0 - y_1)^2 + z^2 \ . \qquad (2)$$

We now make the assumption that $r >> \lambda$, and that the aperture is relatively small,

$$r_0^2 \equiv x_0^2 + y_0^2 + z^2 >> (x_1^2 + y_1^2)_{max} \ . \qquad (3)$$

We can substitute r_0 for r in the denominator of equation (1) giving,

$$U(x_0, y_0, z) = (1/ir_0\lambda) \cos(\mathbf{n},\mathbf{r}_0) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1, y_1) \exp(ikr) \, dx_1 \, dy_1 \ . \qquad (4)$$

In the phase term $\exp(ikr)$ we have to be more careful since a small change in r results in a large phase change. We can write r as

$$r = r_0[1 + (x_1^2 - 2x_0x_1 + y_1^2 - 2y_0y_1)/r_0^2]^{1/2}. \qquad (5)$$

We next use the binomial expansion of the square root, $i.e.$,

$$(1 + b)^{1/2} = 1 + b/2 - b^2/8 + \cdots \qquad |b| < 1. \qquad (6)$$

If we keep the first two terms of the binomial expansion we have for equation (5),

$$r \approx r_0 + (x_1^2 - 2x_0x_1 + y_1^2 - 2y_0y_1)/2r_0. \qquad (7)$$

We have used approximations assuming a relatively small aperture size, but the observation point can be as far from the z-axis as we choose. We now have for equation (4),

$$U(x_0, y_0, z) = (1/ir_0\lambda) \exp(ikr_0) \cos(\mathbf{n},\mathbf{r}_0) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1, y_1)$$

$$\exp[ik(x_1^2 + y_1^2)/2r_0] \exp[-ik(x_0x_1 + y_0y_1)/r_0] \, dx_1 \, dy_1. \qquad (8)$$

If we further assume that

$$r_0 >> k(x_1^2 + y_1^2)_{max} , \qquad (9)$$

then we can write for equation (8)

$$U(x_0, y_0, z) = (1/ir_0\lambda) \exp(ikr_0) \cos(\mathbf{n},\mathbf{r}_0) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(x_1, y_1)$$

$$\exp[-ik(x_0 x_1 + y_0 y_1)/r_0] \, dx_1 \, dy_1. \qquad (10)$$

The standard Fresnel and Fraunhofer diffraction formulas can be found from equations (8) and (10), respectively, by making the appropriate approximations to r_0 when x_0 and y_0 are small compared to z.

4.1.2 Application to Diffraction Efficiency

The integral of equation (10) is the Fourier Transform of $U(x_1, y_1)$ where we would make the substitutions $x_0/\lambda r_0$ and $y_0/\lambda r_0$ for the spatial frequency components, instead of $x_0/\lambda z$ and $y_0/\lambda z$ which are made in the standard Fourier optics. Now all the power and convenience of Fourier optics can be applied to large angle problems.

We note at this point that Shack and Harvey[1] observed that the diffracted wave field can be described in terms of the direction cosines (instead of the spatial variables x_0 and y_0). The aperture is assumed to be illuminated by a spherical wave converging to a point on a hemispherical observation surface. When the radius of the observation surface is large compared to the aperture size, then the field over the entire hemisphere is accurately described by the scaled Fourier transform of the amplitude transmittance function of the aperture. However, the derivation given in 4.1.1 does not require a spherical incident wavefront; the incident wavefront and the transmittance function are arbitrary.

The results of the last chapter can now be incorporated in the extended Fourier optics theory. For example, consider the derivation of equation (3.25). Instead of the paraxial approximation for the sine of an angle we use the exact relationship in the grating equation (normal incidence angle),

$$p\lambda/a = \sin\theta_2 = x_0/r_0 \tag{11}$$

for grooves parallel to the y-axis.

We cannot assume that the grating is infinite, as we did in Chapter 3; what we calculate represents the relative peak intensity in the direction of the diffraction angle. For a lens we are actually calculating the Strehl ratio. With this new interpretation we use equation (10) to derive the far-field diffraction efficiency for the extended Fourier optics case; the result is

$$\eta_{extended}(p, M) = \cos^2(\mathbf{n},\mathbf{r}_0)\ \eta_{standard}(p, M)$$

$$= (1 - \sin^2\theta_2)\ \eta_{standard}(p, M)$$

$$= (1 - (p\lambda/a)^2)\ \eta_{standard}(p, M), \tag{12}$$

where $\eta_{standard}(p, M)$ is given by equation (3.25). We see from equation (12) that the diffraction efficiency decreases with increasing diffraction angle because of the obliquity factor $\cos(\mathbf{n},\mathbf{r}_0)$.

4.2 Rigorous Vector Theory and the Extended Fourier Optics Theory

Several comparisons of the rigorous vector theory have been made in the literature with what the authors term the "scalar" theory of diffraction. Invariably they mean the Fraunhofer diffraction approximation. From the approximations made

in the derivation of the Rayleigh-Sommerfeld diffraction formula we expect that the scalar theory will become inaccurate as the period of a grating approaches the wavelength of light. Small grating periods result in large diffraction angles. But the Fresnel and Fraunhofer formulas are only valid when the aperture is small (compared to the observation distance) and the observation point is close to the normal to the grating. An observation point close to the normal corresponds to a small diffraction angle, which represents a large period. Therefore comparisons of the vector theory and the Fresnel or Fraunhofer diffraction formulas will show the limitations of the Fresnel or Fraunhofer approximations rather than the limitations of the scalar theory.

4.2.1 Extended Fourier Optics Theory and Swanson's Calculations

We shall now compare the diffraction efficiency calculations using the rigorous vector theory with calculations using equation (12). Recently G.J. Swanson proposed an extended scalar theory of diffraction efficiency.[2] His theory attempts to account for the finite thickness of a binary optic by assuming that energy will be lost by "shadowing" and that the optimum depth intuitively should depend on Snell's law. Figure (4.1), from Reference (4.2), is a plot of efficiency as a function of the wavelength to period ratio for a 16-level grating on a substrate with index of refraction $n = 4$. In the limit of small wavelength to period, the efficiency is one because of an assumed antireflective coating on this grating . I have added a plot of equation (12) to the comparison of Swanson's extended scalar theory, and the

electromagnetic (vector) theory calculations for TE and TM modes. The vector theory calculations use the DIFFRACT program which is based on Moharam and Gaylord's coupled wave theory approach to solving Maxwell's equations.[3] Equation (12) agrees well with the vector theory. It appears that the obliquity factor explains the drop in efficiency as the wavelength/period ratio is increased.

However figure (4.2), from Reference (4.2), indicates that the rigorous theory predicts that the diffraction efficiency depends on the substrate index. In this example there is no anti-reflective coating and the efficiency is reduced by the reflection losses. The reflection loss R is given by

$$R = (n - 1)^2/(n + 1)^2. \tag{13}$$

Therefore the far-field diffraction efficiency, according to the extended Fourier optics theory, is

$$\eta_{extended}(p, M) = (n - 1)^2/(n + 1)^2 [1 - (p\lambda/a)^2] \, \eta_{standard}(p, M) \tag{14}$$

when we take into account the reflection losses.

According to Swanson's calculations the diffraction efficiency is much lower when $n = 1.5$ than when $n = 4$. Contrary to the vector theory, equation (14) says that scalar diffraction efficiency does not depend on the substrate refractive index, except for the reflection loss. In figure (4.2) the agreement of equation (14) and the vector theory is not good, even for periods several times the wavelength. It appears therefore that the extended Fourier optics theory can not account for the index of refraction dependence, as the vector theory and Swanson's theory do to some extent.

4.2.2 Extended Fourier Optics and Elson's Calculations

J.M. Elson also calculated the relative intensity in the first diffraction order using the DIFFRACT program.[4] Figure (4.3) is a comparison of equation (14) with Elson's calculations for a four-level binary grating of step height $1.647/4$ μm. The refractive index was $n = 3.43$. The agreement is quite good. In figure (4.4) we compare equation (14) with the vector theory for an eight-level grating of step height $1.647/8$ μm. Again the agreement is very good.

In figures (4.5), (4.6), and (4.7) the relative peak intensity of a 16-level grating is calculated as a function of the grating height (combined height of the 16 levels). Both the vector theory (as calculated by Elson) and equation (14) are plotted. The wavelength is 0.6328 μm and the index of refraction is $n = 1.5$. In figure (4.5) the period is 3.0 μm, in figure (4.6) the period is 2.0 μm, and in figure (4.7) the period is 1.0 μm. According to the scalar theory the optimum grating height is determined solely by the optimum phase difference. We see from Elson's calculations of the vector theory that the optimum grating height also depends on the grating period. When the wavelength to period ratio is large, then evidently the optimum grating depth is no longer given by the scalar theory. The efficiency as determined by equation (14) does not compare well to the vector theory for small periods when the grating height is more than twice the wavelength. Elson[4], like Swanson[2], calculates an optimum depth based on the assumption that the diffraction angle should agree with Snell's law. According to this theory the optimum heights are 1.21, 1.15, and

0.87 μm for the cases shown in figures (4.5), (4.6), and (4.7). The agreement of Elson's or Swanson's theory with the vector theory is not good for small periods.

Elson also calculated the relative intensity as a function of period for a 16-level grating with wavelength λ = 4.0 μm, n = 3.43, and grating height 1.646 μm. A comparison of this vector calculation and equation (14) is shown in figure (4.8). The same grating with an antireflective coating is shown in figure (4.9). The agreement is good, especially for periods three or more times the wavelength.

4.2.3 Extended Fourier Optics and Johnson and Kathman's Calculations

Johnson and Kathman[5] used the coupled mode approach to solving the rigorous vector calculations for a multistep phase grating. They considered periods from 1 to 20 wavelengths. The substrate refractive index was 2.69. The diffraction efficiency in the first order is calculated for a 2-level, 4-level, and 8-level grating as shown in figures (4.10), (4.11), and (4.12) respectively. Equation (14) is plotted for comparison. The agreement is good for periods greater than about five wavelengths. It can also be seen that the more levels in the grating, the longer it takes to rise to the maximum efficiency according to the vector theory. Evidently the efficiency is affected by the small step widths of the grating, and not just the period size.

The Rayleigh-Sommerfeld theory predicts that the peak intensity is proportional to $\cos^2(n,r_0)$ and it can be shown that the diffraction pattern on a sphere of radius r_0 is approximately proportional to $1/\cos(n,r_0)$. The larger the aperture, the better the approximation. Therefore the energy in the diffraction order is

proportional to $\cos(\mathbf{n},\mathbf{r}_0)$. This is confirmed using numerical integration. If we had used the relative energy instead of the Strehl ratio, the agreement with the vector theory in figures (4.1) through (4.12) would not have been as good. In either the rigorous vector theory or the Rayleigh-Sommerfeld theory the energy in the 1st order decreases as the period decreases. According to the vector theory, energy is conserved in a lossless grating. The energy lost in the first order must go into another order; when the period of the grating is less than the wavelength, all the energy is reflected or goes into the zero order. The Rayleigh-Sommerfeld theory predicts no energy in the zero order. What is needed is a comparison with experiment to better determine regions of validity for each theory.

4.3 Scalar Theory - Conventional Lens

4.3.1 Rayleigh-Sommerfeld Versus Fraunhofer Diffraction Formulas

We now compare the various scalar diffraction formulas. We pick the case of a perfect lens of radius L and focal length f. The Fraunhofer diffraction formula predicts an intensity I at the focal point given by

$$I_{\text{Fh}} = (\pi L^2 / \lambda f)^2. \tag{15}$$

The field at the focal point, using the approximate Rayleigh-Sommerfeld formula [equation (2.10)] when the field in the aperture is $\exp(-ikr_1)$, can be found from

$$U(0,0,f) = (1/i\lambda) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-ikr_1)\, f/r_1^2 \, \exp(ikr_1) \, dx_1 \, dy_1$$

$$= (f/i\lambda) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1/r_1^2 \, dx_1 \, dy_1 \,, \tag{16}$$

where

$$r_1^2 = f^2 + x_1^2 + y_1^2. \tag{17}$$

Changing to polar coordinates we have for equation (16)

$$U(0,0,f) = 2\pi f/i\lambda \int_0^L r/(r^2 + f^2) \, dr \tag{18}$$

$$= \pi f/i\lambda \ln[1 + (L/f)^2]. \tag{19}$$

The intensity using the Rayleigh-Sommerfeld formula is therefore

$$I_{RS} = (\pi f/\lambda)^2 \{\ln[1 + (L/f)^2]\}^2. \tag{20}$$

The ratio of equation (20) to equation (15) gives the relative peak intensity of the Rayleigh-Sommerfeld formula at the focal point. We have

$$I_{RS}/I_{Fh} = (f/L)^4 \{\ln[1 + (L/f)^2]\}^2. \tag{21}$$

We see that even for a perfect lens there is a decline in peak intensity as the f-number decreases. Rays of light entering a lens near the edge are refracted at greater angles and therefore do not contribute as much to the intensity at the focus.

4.3.2 Fresnel Versus Fraunhofer Diffraction Formulas

The peak intensity at the focus, as derived from the Fresnel diffraction formula, is

$$I_{Fn}(0,0,f) = (k/f)^2 \{\int_0^L \exp[ikr^2/2f - ik(f^2 + r^2)^{\frac{1}{2}}] \, r \, dr\}^2, \tag{22}$$

which we will have to solve numerically. The peak intensity, as predicted by the Fresnel formula, divided by the Fraunhofer formula [equation (15)] is

$$I_{Fn}/I_{Fh} = 4/L^4 \{\int_0^L \exp[ikr^2/2f - ik(f^2 + r^2)^{\frac{1}{2}}] \, r \, dr\}^2. \tag{23}$$

Figure (4.13) is plot of equations (21) and (23) for $\lambda = 10 \ \mu m$ and $f = 50$ mm. We see that the Fraunhofer formula is a better approximation to the Rayleigh-

Sommerfeld formula in this case. Good agreement between the formulas is achieved when the f-number ($f/2L$) is greater than about 5. The Fraunhofer formula is reasonably close for f-numbers as low as 2.

In figure (4.14) we are again comparing equations (21) and (23), but this time the focal length is much shorter; f = 0.1 mm. Equation (21) depends only on the ratio of f/L and so equation (21) looks the same in figure (4.14) as it did in figure (4.13). The Fresnel diffraction formula [equation (23)] does depend on λ and f as well as f/L. For this short focal length the Fresnel and Fraunhofer formulas are essentially the same for f-numbers greater than 1.

In the binomial expansion of the square root [equation (5)], the first two terms are kept for the Fresnel approximation. The assumption is that succeeding terms are small. In the present case, for the next term to be small we must have

$$\lambda f^3/L^4 >> \pi/4 . \tag{24}$$

In figures (4.15) and (4.16) we have plotted equations (21) and (23) as a function of the parameter $\lambda f^3/L^4$. A comparison of the two figures shows that the parameter of equation (24) makes equation (23) look the same in each plot. When the parameter is greater than $\pi/4$, the difference between the Fresnel and Fraunhofer is less than 10 percent, and is negligible when the parameter is greater than π. However both the Fresnel and Fraunhofer will be significantly different from the Rayleigh-Sommerfeld formula when the focal length is short, unless the parameter is much greater than π.

4.4 Scalar Theories - Fresnel Zone Plate

4.4.1 Diffraction Gratings and Zone Plates

There is a close relationship between a grating and a Fresnel zone plate. Each zone in the zone plate acts like a grating to diffract the light toward the focal point. Let a_j be the distance from the center of the zone plate to the end of the jth zone. By definition the distance from the focal point to a point at the end of the jth zone is $f + j\lambda$, where f is the focal length of the zone plate. Then

$$a_j^2 = (f + j\lambda)^2 - f^2 \qquad (25)$$

by the Pythagorean relation.

The distance from the center of the zone plate to the center of the jth zone is $(a_j + a_{j-1})/2$. The average distance from the focal point to the jth zone is $(f + j\lambda + f + (j - 1)\lambda)/2$. Let $\theta_{j,FZ}$ represent the angle a ray is deflected in the jth zone. Then

$$\sin\theta_{j,FZ} = (a_j + a_{j-1})/(f + j\lambda + f + (j - 1)\lambda)$$

$$= (a_j + a_{j-1})\lambda/(a_j^2 - a_{j-1}^2)$$

$$= \lambda/(a_j - a_{j-1})$$

$$= \sin\theta_{j,grating}, \qquad (26)$$

where $\theta_{j,grating}$ is the first-order diffraction angle of a grating with period $(a_j - a_{j-1})$. We used equation (25) in the derivation of equation (26). Each zone in the zone plate therefore diffracts light into exactly the same angle as an infinite grating with a period the width of the zone, even though the width of a zone changes from zone to zone.

As j increases, $(a_j - a_{j-1}) \rightarrow \lambda$, so the first-order diffraction angle approaches 90 degrees. For the pth-order, the sine of the diffraction angle is equal to $p\lambda/(a_j - a_{j-1})$. When $(a_i - a_{i-1}) < p\lambda$ the angle will be imaginary and no energy from the ith zone, or larger, will reach the pth order "focal point." We see that in the $j+1$st zone the pth-order diffracted ray will intersect the z-axis closer to the zone plate than the jth zone. In the paraxial approximation, the pth-order focal length f_p is equal to f_1/p, but in general the focal length is a function of the radial position in the zone plate for $p > 1$. We may call this spherical aberration for higher orders of diffraction.

4.4.2 Local Diffraction Efficiency

As the zone radius increases, the spacing of the zones decreases and the efficiency will decrease. We might assume that the efficiency will decrease as $\cos^2(\mathbf{n},\mathbf{r})$, as it does according to the Rayleigh-Sommerfeld scalar theory when the aperture is small compared to the distance to the observation point. But we are interested in distances which are comparable to the aperture diameter. Equation (21) gives the overall maximum efficiency a Fresnel zone plate could have (assuming an infinite number of levels; i.e., $M \rightarrow \infty$ within each zone). To compare the grating efficiency to the Fresnel zone plate efficiency, we need to calculate the local efficiency in each zone, when the focal point is close to the aperture. We do this by taking the differential of the intensity as given by the Rayleigh-Sommerfeld formula, equation (20), with respect to the radius L of the aperture. Then we ratio this to the

differential of the Fraunhofer formula, equation (15). The result is

$$\eta_{local,max,FZ} = \ln[1 + (L/f)^2]/[(L/f)^2 + (L/f)^4] . \qquad (27)$$

The diffraction efficiency in a Fresnel zone is *not* the same as the efficiency of a grating with the same local period. Figure (4.17) compares the grating efficiency [upper curve, from equation (12)] to the local Fresnel efficiency [equation (27)]. The *overall* efficiency curve is very close to the grating efficiency at the extreme edge of the zone plate. Also shown in figure (4.17) is an approximate relation Buralli and Morris[6] found to express the drop in grating efficiency, as predicted by the DIFFRACT program using the rigorous vector theory. Buralli and Morris assumed a classical blazed grating (hence $M \rightarrow \infty$) with substrate refractive index $n = 4$, and wavelength $\lambda = 10.6$ μm. Their formula is

$$\eta_{local} = 1 - 0.32\lambda/(a_j - a_{j-1}). \qquad (28)$$

Equation (28) agrees reasonably well with the Rayleigh-Sommerfeld grating efficiency formula, but differs significantly from equation (27) for small f-numbers (large L/f ratios). All of these formulas depart from the conventional Fraunhofer or Fourier optics formula which does not change at all with the f-number.

4.4.3 First Diffraction Order Efficiency

A plot of the diffraction efficiency for a phase-only Fresnel zone plate will also help to determine the regions of validity of the various diffraction formulas. Figure (4.18) shows the intensity in the first diffraction order for a Fresnel zone plate designed to maximize the intensity in the 1st order. The intensity is taken relative

to the Fraunhofer peak intensity for a perfect lens, equation (15). The zone plate is designed using equation (3.19) to have a focal length of 2 mm when the wavelength is 0.6328 μm. The intensity was calculated numerically according to the Rayleigh-Sommerfeld theory using equation (2.9). The far-field Fraunhofer formula plotted is $\eta_{standard}(p, M)$ from equation (3.25). Using the Fraunhofer formula directly is not appropriate for binary optics because the focal point is much closer to the aperture than the Fraunhofer region, even though for a perfect lens the focal point is said to be in the Fraunhofer region. An attempt to use the Fraunhofer formula with a binary optic predicts, incorrectly, that the relative intensity on-axis is nearly zero.

The Fresnel diffraction formula can be solved for points along the z-axis for a circular Fresnel zone plate. The result is

$$I_{FZ}(0,0, z) = \{\Sigma_j \exp(-kij/M)[\exp(kia_j^2/2z) - \exp(kia_{j-i}^2/2z)]\}^2, \qquad (29)$$

where $\{a_j\}$ are the transition points given by equation (3.19), and M is the number of levels. The relative peak intensity is I_{FZ}/I_{FH} and this ratio is calculated to give the Fresnel formula in figures (4.18), (4.19), and (4.20). The location of the peaks are found by plotting intensity along the z-axis. The Fresnel formula, equation (29), shows a peak at 2.031 mm. The Rayleigh-Sommerfeld formula indicates the focus is at the paraxial focus, $f = 2.00$ mm. Figure (4.18) shows that for an f-number of f/2 the Fresnel formula does not accurately predict the peak intensity. The far-field Fraunhofer equation and the Rayleigh-Sommerfeld formula are in better agreement with each other. For a small f-number the standard diffraction efficiency equation

[equation (3.25)] predicts a higher efficiency than the more accurate Rayleigh-Sommerfeld formula.

4.4.4 Third Diffraction Order Efficiency

For a Fresnel zone plate with an f-number of f/10, the first order diffraction efficiency, as determined by either the Rayleigh-Sommerfeld or the Fresnel formulas gives the same result as the far-field Fraunhofer. There are some differences in the 3rd and 5th diffraction orders if the design is made to maximize the first order efficiency. Figure (4.19) is a plot of the peak intensity in the 3rd order relative to the value we obtain from equation (15), when f_3 = 10/3 mm. The Fresnel formula shows a peak approximately at the 3rd order paraxial focus f_3. But the Rayleigh-Sommerfeld maximum intensity is at z = 3.317 mm. The Fresnel and the Fraunhofer formulas predict the same values. The Rayleigh-Sommerfeld formula predicts a slightly lower peak due to the obliquity factor and spherical aberration.

4.4.5 Fifth Diffraction Order Efficiency

Figure (4.20) compares the various formulas for the 5th-order. The maximum intensity according to the Rayleigh-Sommerfeld formula was located at z = 1.97 mm. A narrower peak was found using equation (29) at z = 2.01 mm. The paraxial focus, f_5 = 10/5 = 2.00 mm, was used in equation (15) to give the relative peak intensity. It can be seen that the Fresnel and Fraunhofer formulas agree, but they are not accurate even for an f/10 lens if we are interested in the intensity at a point much closer to the lens than the first-order focal point. If we had plotted the peak

intensity in the 1st, 3rd, and 5th orders relative to the 1st order then we would have seen that for a 2-level Fresnel zone plate the peak intensities for each order are equal according to the Fresnel and Fraunhofer formulas, but not according to the Rayleigh-Sommerfeld formula.

4.4.6 Diffraction Efficiency Formulas

We have not used the extended Fourier optics formula for the Fresnel zone plate because it was derived for distances large compared to the aperture. On-axis all the theories agree well at large distances. It would be interesting to try the extended Fourier optics theory integrating over the lens aperture.

The far-field Fraunhofer efficiency formula $\eta_{standard}(p, M)$ compares well to the Rayleigh-Sommerfeld formula when the f-number is not small. Equation (21) relates the Rayleigh-Sommerfeld formula to the far-field Fraunhofer formula for the intensity at the focal point of a perfect lens. It seems reasonable then to suppose that $\eta_{RS}(1,M)$, the diffraction efficiency in the first-order as determined by the Rayleigh-Sommerfeld formula, is given by

$$\eta_{RS}(1, M) = (f/L)^4 \{\ln[1 + (L/f)^2]\}^2 \, \eta_{standard}(1, M). \qquad (30)$$

Comparison with the exact numerical integration shows this to be very accurate near the paraxial region. The formula does not work as well for higher diffraction orders where we are no longer in the paraxial region (i.e., where $\sin\theta \approx \tan\theta$), and where spherical aberration in higher diffraction orders makes the lens less than perfect. In figure (4.20) the peak is lower than would be expected by equation (30).

4.5 Summary

A new formula has been derived which greatly extends Fourier optics. This formula removes the restriction that the observation point must be close to the optical axis. A comparison of the extended Fourier optics theory and rigorous vector theory shows good agreement for substrates with high refractive indices. The agreement is not as good for low refractive indices when the wavelength to period ratio is more than about 0.1. Apparently when the refractive index is low the required step heights are large, which distorts the EM fields, contrary to the assumptions of scalar theory.

The extended Fourier optics theory shows good agreement with vector theory for some 4-level and 8-level gratings, with a wavelength to period ratio less than 0.3. The diffraction efficiency versus step height of a 16-level grating shows some significant differences between theories; and the larger the wavelength to period ratio, the stronger the differences. Again higher refractive indices lessened the differences. In all cases the extended Fourier optics theory shows a decrease in diffraction efficiency at large angles, which is in agreement with the rigorous vector theory, but disagrees with the results of the conventional Fourier optics theory.

Comparing scalar theories for conventional lenses we find that an f-number greater than about four gives good agreement between Fraunhofer and Rayleigh-Sommerfeld formulas. The appropriate parameter for comparing Fresnel and

Fraunhofer formulas is the parameter $\lambda f^3/L^4$ where f is the focal length and L is the lens radius. For good agreement $\lambda f^3/L^4$ should be greater than one.

Contrary to what is generally believed:

(1) The local efficiency of a Fresnel zone plate is not the same as the efficiency of a grating with the same period. The differences are relatively small for f-numbers greater than about two.

(2) The Fraunhofer formula does not give correct results for binary optics in the Fresnel region, even when the Fourier optics results are correct and the conventional lens results are correct.

(3) The Fourier optics formula predicts the diffraction efficiency of Fresnel zone plates better than the Fresnel diffraction formula does.

(4) The Fresnel diffraction formula can be significantly wrong for a small (say $f/2$ or less) f-number Fresnel zone plate.
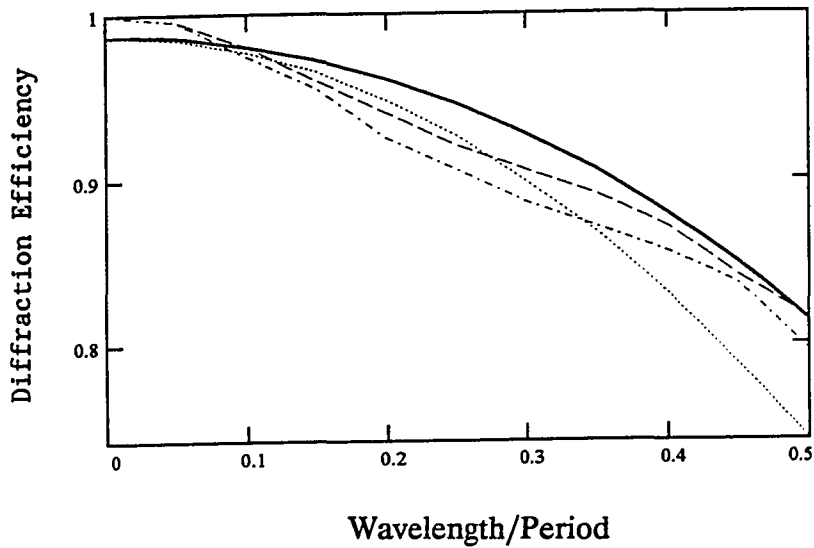
FIGURE 4.1, Predicted first-order diffraction efficiency for a grating on a substrate with $n = 4$. Comparing equation (4.12, ···) with vector theory (TM --, TE - -) and Swanson's theory (solid line) (Ref. 4.2).
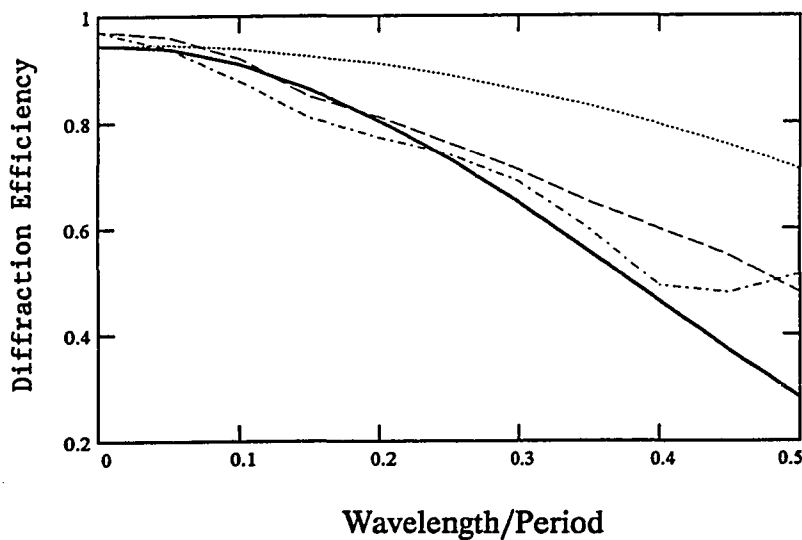
FIGURE 4.2, Predicted first-order diffraction efficiency for a grating on a substrate with $n = 1.5$. Comparing equation (4.14, ···) with vector theory (TM --, TE - -) and Swanson's theory (line, Ref. 4.2).

Diffraction Efficiency
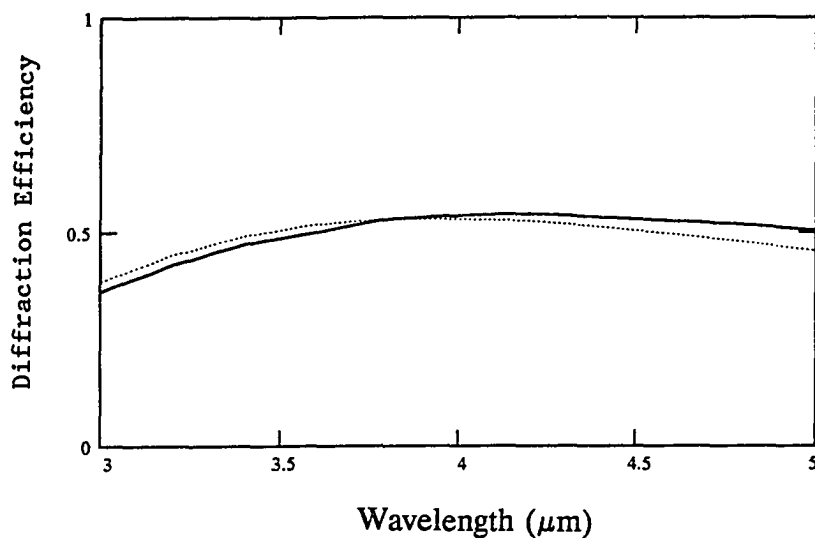
Wavelength (μm)

FIGURE 4.3, Predicted first-order diffraction efficiency versus wavelength. Four-level grating with a period of 16.5 μm. Comparing equation (4.14, ···) with vector theory (solid curve, Ref. 4.4).

Diffraction Efficiency
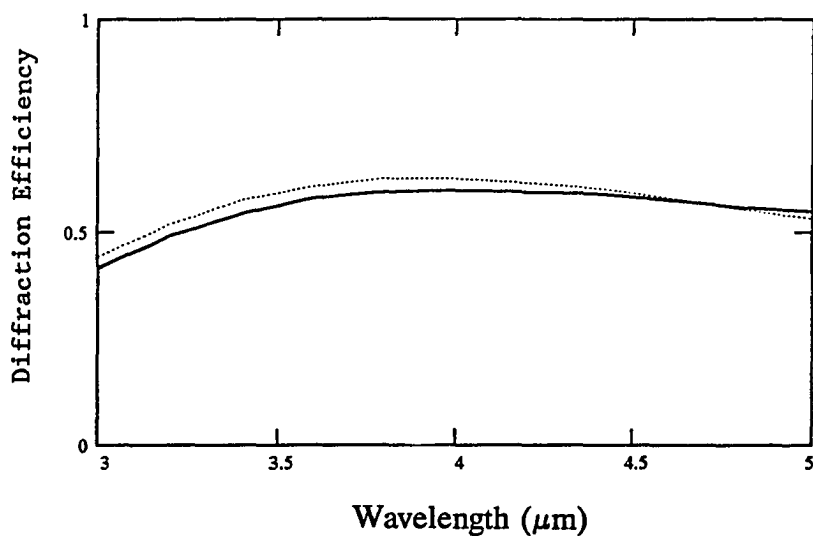
Wavelength (μm)

FIGURE 4.4, Predicted first-order diffraction efficiency versus wavelength. Eight-level grating with a period of 16.5 μm. Comparing equation (4.14, ···) with vector theory (solid curve, Ref. 4.4).
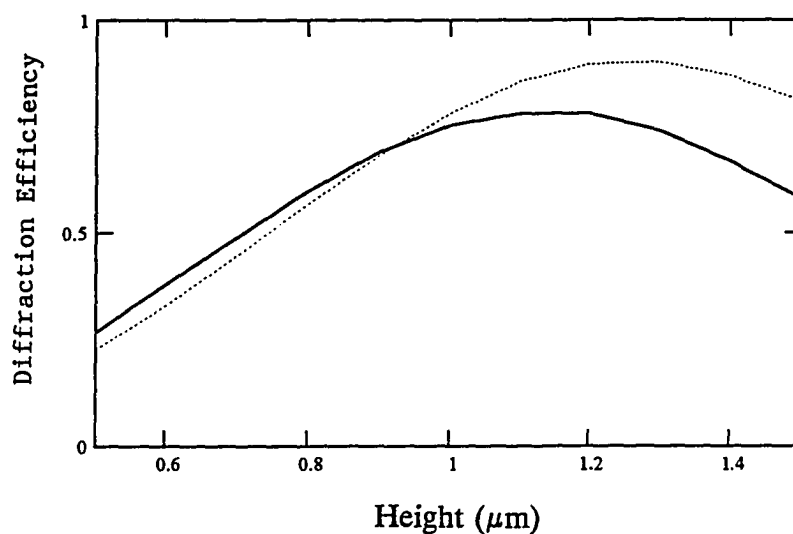
FIGURE 4.5, Predicted first-order diffraction efficiency versus grating height. Sixteen-level, period is 3.0 μm, λ = 0.6328 μm, and n = 1.5. Comparing equation (4.14, ···) with vector theory (line, Ref. 4.4).
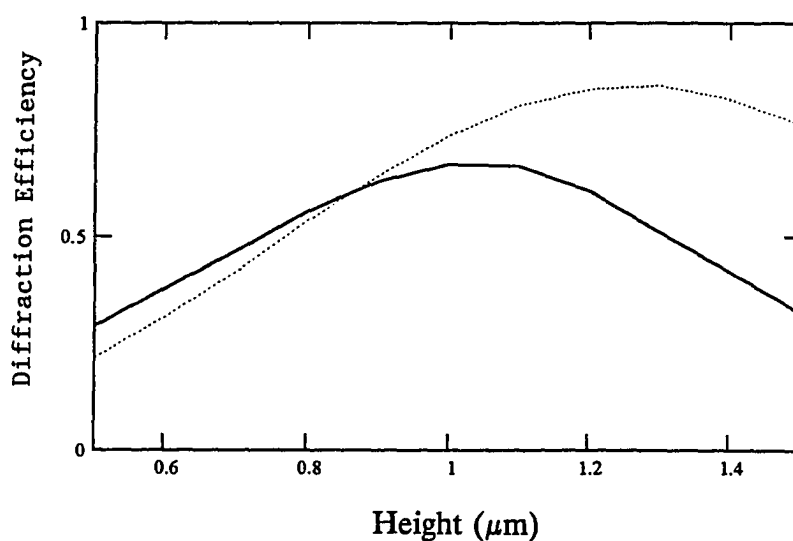
FIGURE 4.6, Predicted first-order diffraction efficiency versus grating height. Sixteen-level, period is 2.0 μm, λ = 0.6328 μm, and n = 1.5. Comparing equation (4.14, ···) with vector theory (line, Ref. 4.4).
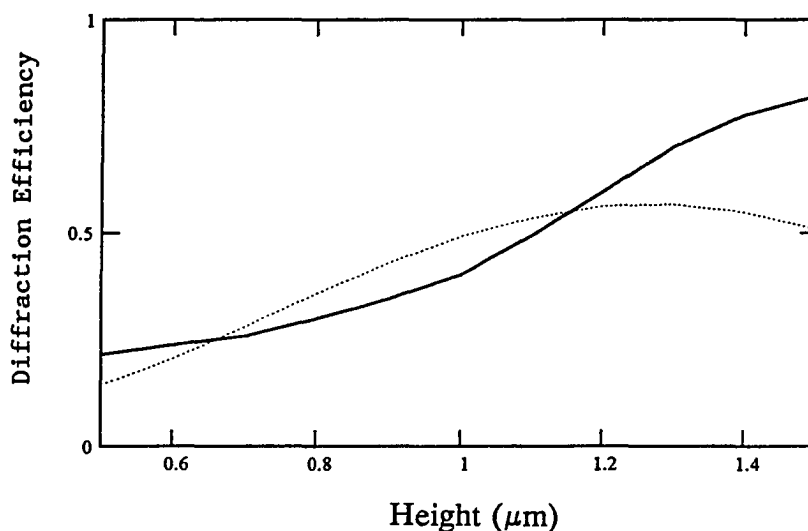
FIGURE 4.7, Predicted first-order diffraction efficiency versus grating height. Sixteen-level, period 1.0 μm, λ = 0.6328 μm, and n = 1.5. Comparing equation (4.14, ⋯) with vector theory (line, Ref. 4.4).
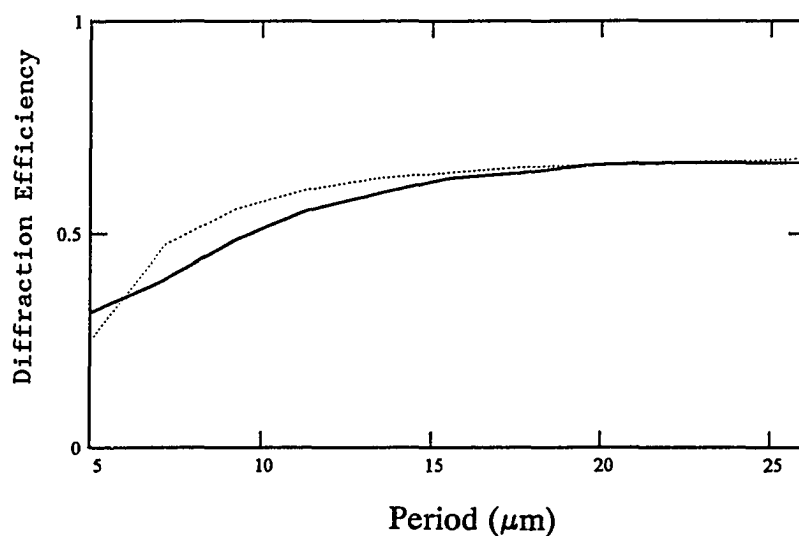
FIGURE 4.8, Predicted first-order diffraction efficiency versus period. Grating height = 1.646 μm, 16-level, λ = 4.0 μm, and n = 3.43. Comparing equation (4.14, ⋯) with vector theory (line, Ref. 4.4).
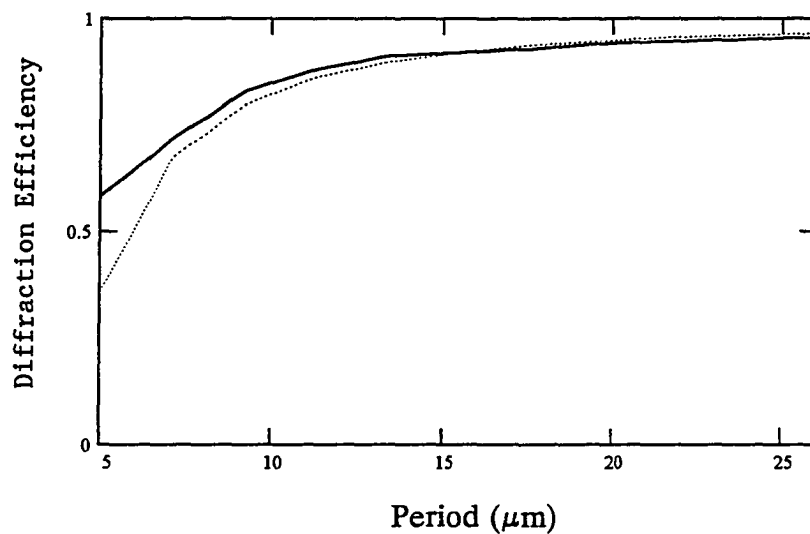
FIGURE 4.9, Predicted first-order diffraction efficiency versus period. Grating height 1.646 μm, 16-level, λ = 4.0 μm, n = 3.43, with AR coating. Equation (4.14, $\cdots$) and vector theory (line, Ref. 4.4).
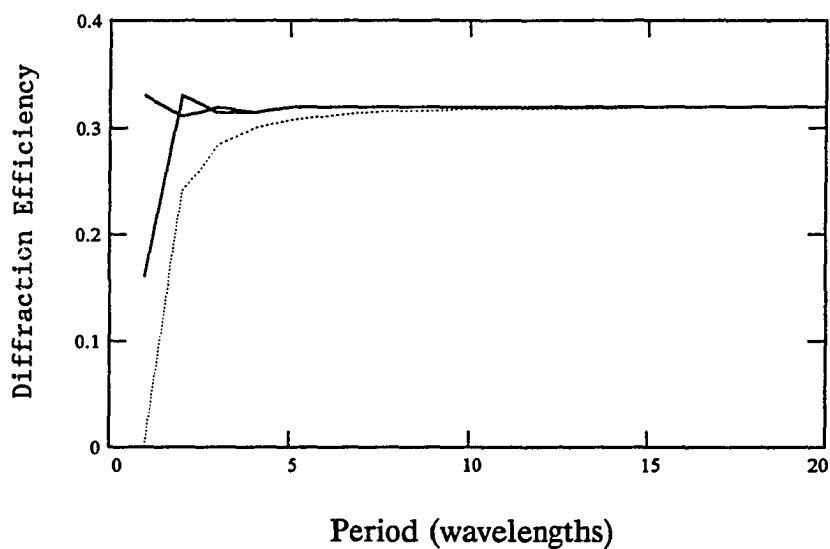
FIGURE 4.10, Predicted first-order diffraction efficiency versus period. Two-level, n = 2.69. Comparing equation (4.14, $\cdots$) with vector theory (line, Ref. 4.5).
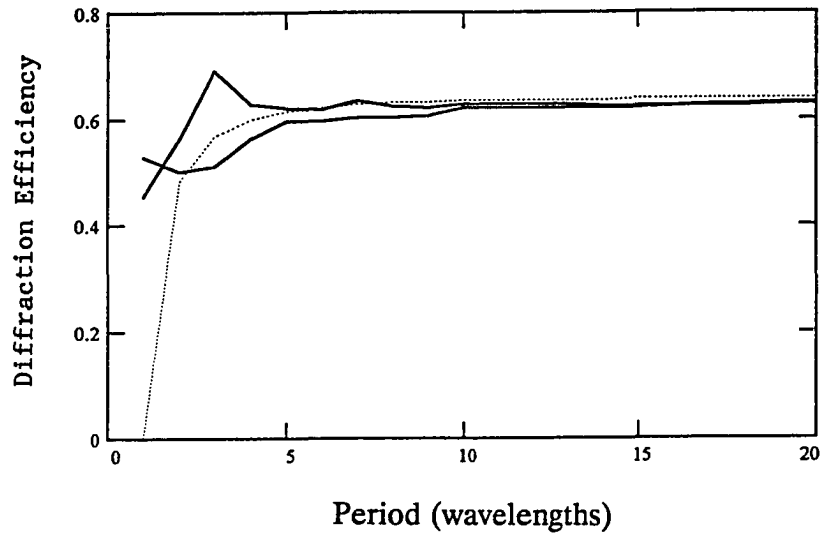
FIGURE 4.11, Predicted first-order diffraction efficiency versus period. Four-level grating, on a substrate with n = 2.69. Comparing equation (4.14, ···) with vector theory (line, Ref. 4.5).
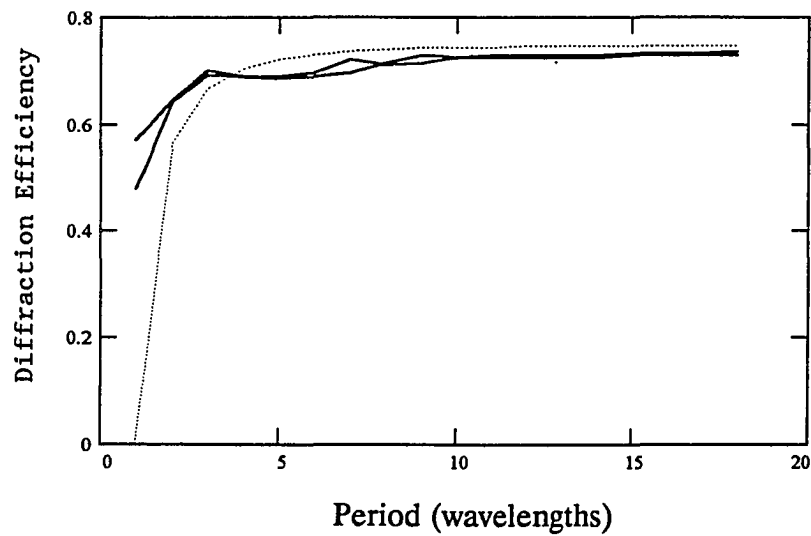
FIGURE 4.12, Predicted first-order diffraction efficiency versus period. Eight-level, on a substrate with n = 1.5. Comparing equation (4.14, ···) with vector theory (line, Ref. 4.5).
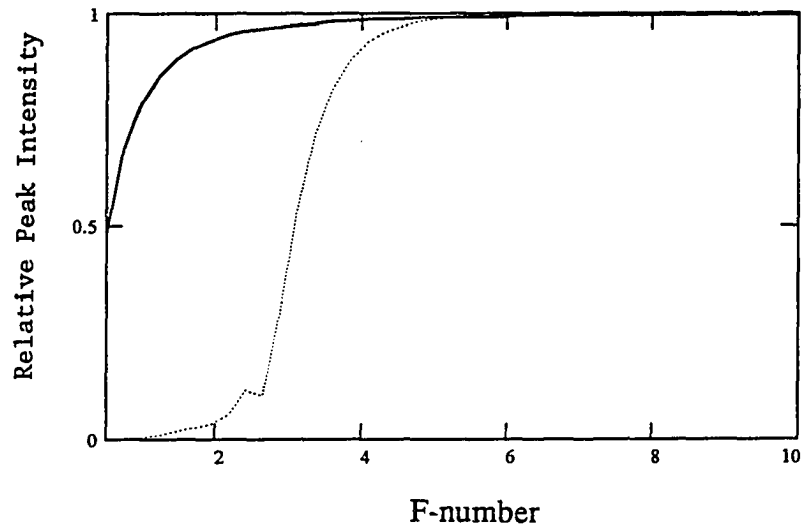
FIGURE 4.13, Peak intensity relative to Fraunhofer peak intensity versus f-number. λ = 10 μm, f = 50 mm. Comparing Rayleigh-Sommerfeld (line) and Fresnel formulas (···).
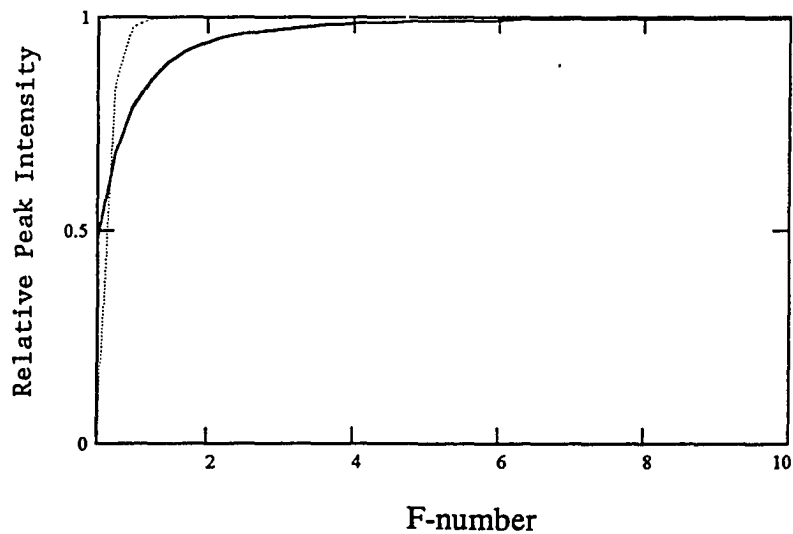
FIGURE 4.14, Peak intensity relative to Fraunhofer peak intensity versus f-number. λ = 10 μm, f = 0.1 mm. Comparing Rayleigh-Sommerfeld (line) and Fresnel formulas (···).
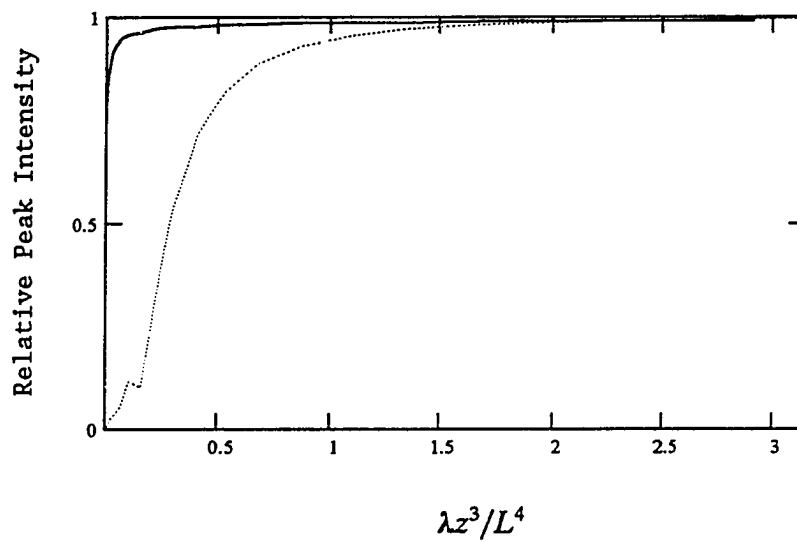
FIGURE 4.15, Peak intensity relative to Fraunhofer peak intensity. $\lambda = 10$ μm, $f = 50$ mm. Comparing Rayleigh-Sommerfeld (line) and Fresnel formulas (···).
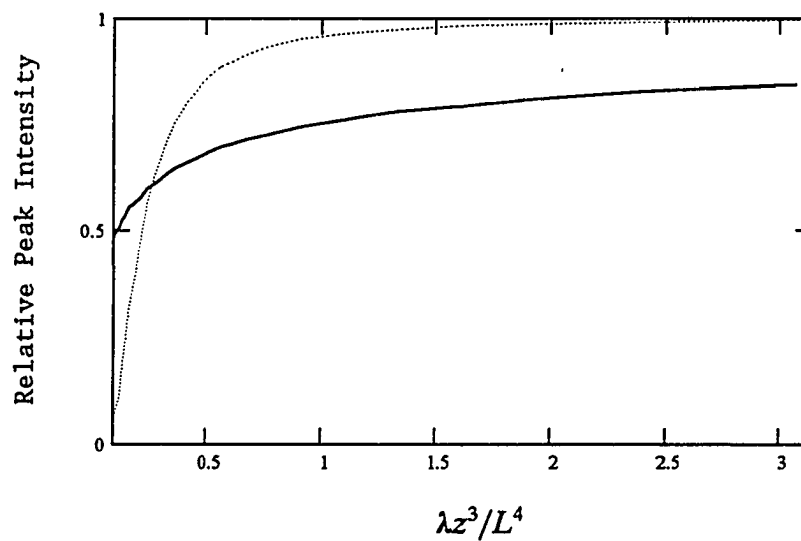
FIGURE 4.16, Peak intensity relative to Fraunhofer peak intensity. $\lambda = 10$ μm, $f = 0.1$ mm. Comparing Rayleigh-Sommerfeld (line) and Fresnel formulas (···).
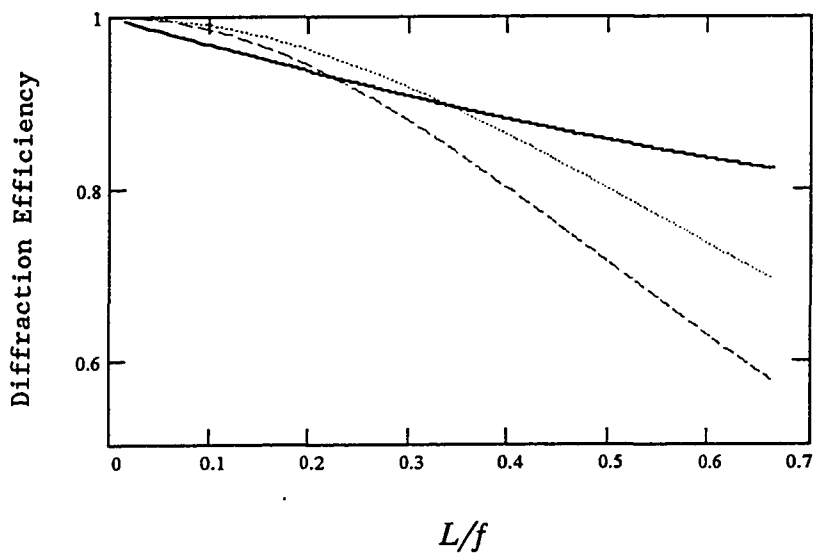
FIGURE 4.17, Maximum efficiency versus relative position L/f on a Fresnel zone plate. $f = 20$ mm, $\lambda = 10.6$ μm. Vector theory (–), Rayleigh-Sommerfeld (RS) grating (···) and RS zone plate (- -).
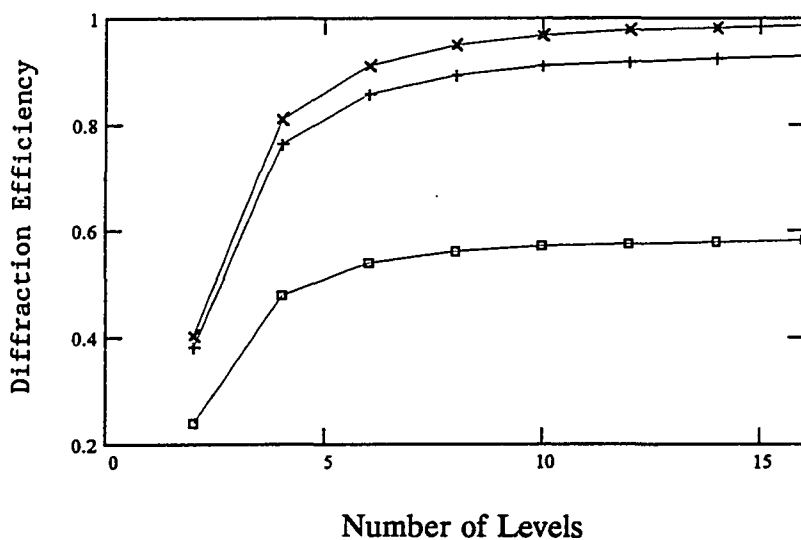
FIGURE 4.18, Predicted first-order diffraction efficiency versus number of levels. Zone plate, f/2, $f = 2$ mm, $\lambda = 0.6328$ μm. Comparing equation (3.25) (X) with RS (+) and Fresnel (0).
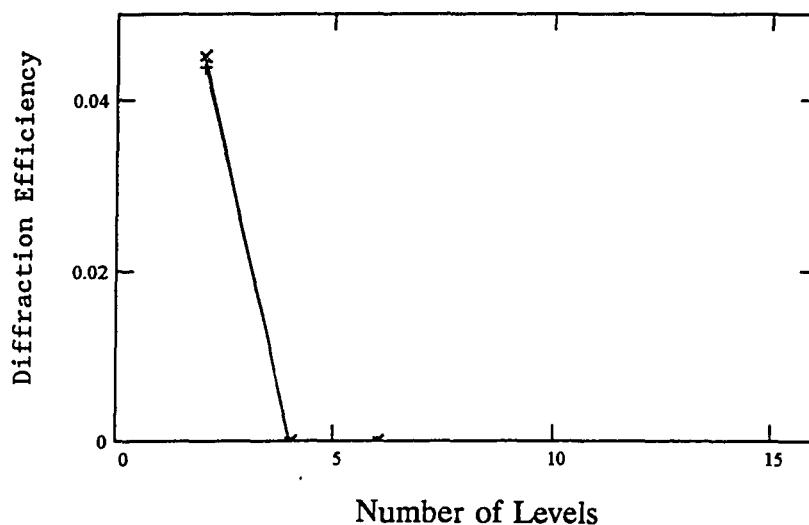
FIGURE 4.19, Predicted 3rd-order diffraction efficiency for an f/10 zone plate versus number of levels. $f = 10$ mm, $\lambda = 0.6328$ μm. Comparing equation (3.25) (X) with RS (+) and Fresnel (line).
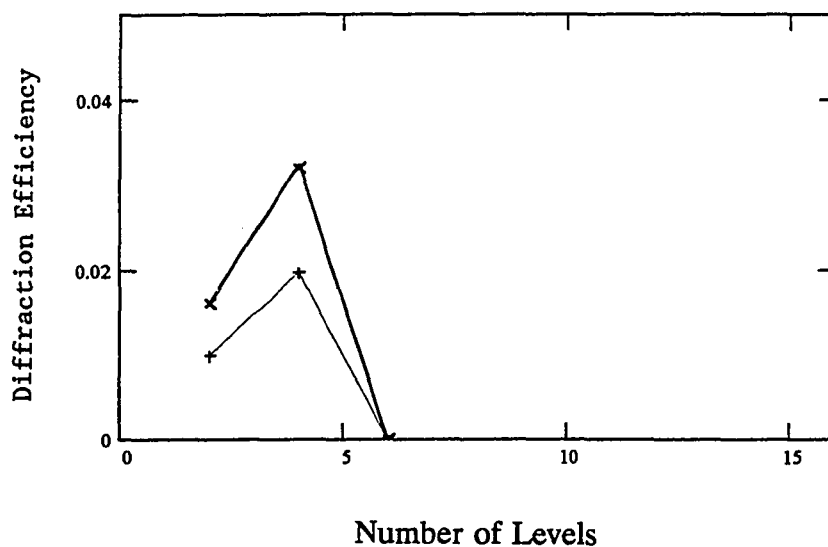
FIGURE 4.20, Predicted 5th-order diffraction efficiency for a zone plate versus number of levels. $f = 10$ mm, $\lambda = 0.6328$ μm. Comparing equation (3.25) (X) with RS (+) and Fresnel (upper line).

5. SYSTEMATIC ERRORS

The principle of scattering which we employ is this: that a real-life binary optic is the sum of an ideal surface and a scattering surface. Systematic errors make this scattering surface behave as another grating which interferes with the field distribution of the ideal surface. The two surfaces have the same local period, so the result is a redistribution of energy among the diffraction orders. There is no interorder scattering. Once we know the new diffraction efficiency for a systematic error, we can use the grating equation and geometrical optics to determine the scattering.

Systematic errors, or deterministic errors, include: etch depth errors, line edge errors, mask alignment errors, quantization errors, and the shortcomings of an *M*-level binary optic approximation to the ideal continuous phase profile. We shall compare computer simulations and diffraction efficiency measurements to original analytical formulas which we will derive in this chapter. The computer simulations were performed on a personal computer using the software tool Mathcad 2.5.

5.1 Etch Depth Errors

Systematic errors are deterministic; a known etch depth error that varies gradually from center to edge, or an exposure error that results in each groove being narrow by a fixed amount. We first consider etch depth errors. In this chapter, and

the next, we will be discussing a number of Fresnel zone plate designs; Table 5.1 summarizes these designs.

Table 5.1. Fresnel zone plate designs

Design	$\lambda(\mu m)$	Levels	T	f(mm)	f#	n	b(mm)	$\lambda f^3/a_T^4$	I_{RS}/I_{Fh}
FZP1	0.6328	16	316	10	10	1.457	0.35	3.2π	0.998
FZP2A	0.6328	4	79	10	10	1.457	0.35	3.2π	0.998
FZP2B	0.6328	4	2005	254	10	1.457	∞	0.13π	0.998
FZP3A	0.6328	2	40	10	10	1.457	∞	3.2π	0.998
FZP3B	0.6328	2	40	10	10	1.457	0.35	3.2π	0.998
FZP3C	0.6328	2	395	100	10	1.457	3.50	0.32π	0.998
FZP4	10.6	16	19	10	10	4.0	0.35	54π	0.998
FZP5	10	16	3360	152.4	3	4.0	∞	0.03π	0.973
FZP6	10	16	300	12.7	1	4.0	∞	0.0005π	0.797
FZP7	10	8	1680	152.4	3	4.0	∞	0.03π	0.973
FZP8	10	8	300	12.7	1	4.0	∞	0.0005π	0.797

5.1.1 Appropriate Diffraction Formula

In Table 5.1 the factor I_{RS}/I_{Fh} indicates how closely the Fraunhofer formula approximates the Rayleigh-Sommerfeld formula in the focal plane of a lens. It is calculated using equation (4.21). We expect the conventional Fourier analysis results (based on the Fraunhofer formula) for designs FZP5 and FZP7 to predict peak

intensities about 2.8 percent more than the Rayleigh-Sommerfeld formula. The Rayleigh-Sommerfeld formula predicts an intensity only 79.7 percent of that predicted by the Fourier analysis for designs FZP6 and FZP8. For the other designs in Table 5.1 the results of the Fourier analysis should be the same as the Rayleigh-Sommerfeld results.

Table 5.1 also shows the factor $\lambda f^3/a_T^4$, which is a measure of how close the Fresnel formula is to the Fraunhofer formula, as shown in the Chapter 4. Since the formulas are nearly equal when this factor is greater than about $\pi/2$, the Fresnel formula results for designs FZP2B, FZP5, FZP6, FZP7, and FZP8 can be expect to deviate significantly from the Fraunhofer formula. For these designs we cannot use the Fresnel formula with much accuracy. Design FZP3C is marginal; the Fresnel formula is about five percent under the Fraunhofer value for design FZP3C according to figure (4.15). For the other designs we conclude that if we are reasonably close to the axis we can use the results of either Fourier analysis (Fraunhofer based) or the Fresnel formula.

5.1.2 Formula for Systematic Etch Depth Errors

The relative loss in diffraction efficiency with etch depth error can be obtained from equation (3.25). Let $\eta_{opt}(p,M)$ be the diffraction efficiency optimized for an order p and with M steps. Let ϵ be the relative etch depth error, then the phase is given by

$$\phi = 2\pi p(1 + \epsilon). \tag{1}$$

Combining equation (1) with equation (3.25) we obtain a new formula for the diffraction efficiency in the presence of systematic etch depth errors:

$$\eta(p, M) = \eta_{opt}(p, M) \{[\sin(p\epsilon\pi)]/[M \sin(p\epsilon\pi/M)]\}^2. \qquad (2)$$

It can be thought of as the interference of a "ghost grating" with the systematic errors and the optimized, or ideal grating with M levels.

5.1.3 Uniform Etch Depth Error

Figure (5.1) is a plot of the diffraction efficiency for Fresnel zone plate design FZP3A as a function of percent error in etch depth. The data points were calculated using the approximate Rayleigh-Sommerfeld equation (2.10) which for a circularly symmetric Fresnel zone plate, on axis, can be written as

$$U(0,0, z) = \Sigma_j (2\pi z/\lambda i) \phi_j \int_{a_j}^{a_{j+1}} r/(r^2 + z^2) \exp[ik(z^2 + r^2)^{1/2}] dr \qquad (3)$$

where,

$$a_j = [2fj\lambda/M + (\lambda j/M)^2]^{1/2}, \qquad (4)$$

$$\phi_j = \exp\{-2\pi i[1 + e]\mathrm{mod}(j, M)/M\}, \qquad (5)$$

and z is the distance from zone plate to observation point, f is the focal length, T is the maximum number of transition points $\{a_j\}$, M is the number of levels, and e is the fractional etch depth error. The function $\mathrm{mod}(j, M)$ is the remainder when j is divided by M. For this example the illuminating beam was assumed to be uniform, so the gaussian beam radius b (at the $1/e^2$ points of intensity) is infinite.

For an etch depth error that is uniform across the surface we can use equation (2) for the far-field, paraxial case. In a 2-level Fresnel zone plate an etch depth

error of *e* will alter the phase over half the surface of the optic (in the grooves only), so we use *e*/2 for the error term in equation (2). Our new analytical expression, for etch depth error, shown in figure (5.1), matches very well the results found using the Rayleigh-Sommerfeld integral. A computer simulation result similar to figure (5.1) was calculated by Cox, et. al.[1]

5.1.4 Linearly Varying Etch 95XDepthror

Uniform etch depth error is one type of systematic error that can occur in a binary optic. The etch depth can also vary across the binary optic; for example, each etch depth may be 5 percent deeper than it should be near the center, gradually becoming 5 percent shallower near the edges. This might occur when the narrower grooves near the edges reduce the etching rate. We can model this type of error by writing

$$\phi_j = \exp[-2\pi i(1 + e(1 - 2a_j/a_T))\mathrm{mod}(j, M)/M] \tag{6}$$

for the phase term ϕ_j in equation (3). In computer simulations using the Huygens-Fresnel integral [the approximate Rayleigh-Sommerfeld equation, (2.10)], Goodman and Farn[2] found empirically that the diffraction efficiency of a Fresnel zone plate falls off as the square of the etch error, the effect of etch errors is only a weak function of the number of masks, and it is independent of the f-number and radius. They considered Fresnel zone plates with f-numbers $f/1$ and $f/3$ and radii of one and three inches. We shall show that our equation (2) predicts the same empirical results obtained by Goodman and Farn.

Figure (5.2) shows results identical to Farn and Goodman's data for a 16-level and an 8-level Fresnel zone plate. The relative diffraction efficiency is the peak intensity [equations (3), (4) and (6)] divided by the product of the peak intensity of a perfect lens [equation (4.20)] and the maximum efficiency of an M-level zone plate [equation (3.25)]. Designs FZP5, FZP6, FZP7, and FZP8 from Table 5.1 are used in the calculations. The results do not depend on the f-number, and depend only slightly on the number of levels (hence on the number of masks in the fabrication process). Also shown in figure (5.2) are curves calculated from equation (2), where $\epsilon = 7/8\, e$ for an 8-level zone plate, and $\epsilon = 15/16\, e$ for a 16-level zone plate. The relative etch depth error ϵ appears to be the average systematic etch depth error along an axis of the Fresnel zone plate.

A systematic etch depth error decreasing from center to edge should shift the focal position. When the efficiency is calculated at the optimum focal position then the diffraction efficiency decrease is not nearly as severe as calculated by Farn and Goodman in reference (2). The upper data points in figure (5.2) show our calculations for diffraction efficiency at best focus (empirically determined). A good fit to the data is obtained by using equation (2) with $\epsilon = e/\sqrt{3}$, which is the *rms* value of the systematic etch depth error e along an axis of the Fresnel zone plate. It is found that at the best focus the diffraction efficiency is essentially independent of the number of masks, the f-number, and the radius.

5.2 Line Edge Errors

Over- or under-exposure of the photoresist during the pattern transfer process can result in systematic line edge errors. Under-exposure results in the grooves being narrower than they should be. Farn and Goodman found empirically through computer simulations that diffraction efficiency (defined as the Strehl ratio) of a Fresnel zone plate falls off linearly with the line edge error. The effect of line edge error is a strong function of the f-number, a weak function of the number of masks and is independent of radius.[2] In section 5.2.1 we will compare measured results with our own computer simulations that verify the results of Farn and Goodman. In section 5.2.2 we derive a simple analytical function which agrees well with the previous results.

5.2.1 Local Diffraction Efficiency

Cox et. al. measured the diffraction efficiency of an $f/10$ Fresnel zone plate in ten "zones" of equal area.[1] The first zone is the central portion of the zone plate, covering ten percent of the entire zone plate. Two zone plates were deliberately under-exposed so that the etched portions were 1.35 μm narrower than they should be. A plot of the measured local diffraction efficiency (fraction of power) is shown in figure (5.3) for the average of the two zone plates. Also shown is a plot of the calculated local diffraction efficiency in each of the ten "zones" as defined by Cox, et. al.

To calculate the local diffraction efficiency, we first calculate the peak intensities I_1, I_2, . . ., I_{10} for zone plates with radius a_4, a_8, . . ., a_{40} using design FZP3A, which has a smaller diameter but the same $f^{\#} = 10$ as the Cox data, and the same average period in each zone. We use equation (1) and

$$a_j = [2fj\lambda/M + (\lambda j/M)^2]^{\frac{1}{2}} + e \bmod(j, M) \tag{7}$$

$$\phi_j = \exp(-2\pi ij/M). \tag{8}$$

To compare with the Cox data the line edge error e is 1.35 μm, and $M = 2$ for a 2-level zone plate. The relative peak intensities are obtain by dividing by the peak intensity I_{RS} of a perfect lens of radius a_4, a_8, . . ., a_{40} . The local diffraction efficiencies are therefore

$$\eta_1 = I_1/I_{RS1} \qquad\qquad \text{1st zone}$$

$$\eta_2 = 2{*}I_2/I_{RS2} - I_1/I_{RS1} \qquad\qquad \text{2nd zone}$$

$$\cdots$$

$$\eta_{10} = 10{*}I_{10}/I_{RS10} - 9{*}I_9/I_{RS9} \qquad\qquad \text{10th zone.} \tag{9}$$

5.2.2 Formula for Systematic Line Edge Errors

We expect the diffraction efficiency to decrease with increasing zone number since the error e becomes relatively more significant as the spacing decreases. In a 2-level zone plate a line edge error represents an area of the zone plate that does not contribute at all to the peak intensity. We therefore subtract the error fraction of a period from the efficiency. In a 2-level zone plate a period is every two transition points. In design FZP3A there are forty transition points, hence four

transition points in each of ten "zones" as defined by Cox, et. al. If d_i is the average period in the ith "zone",

$$d_i = (a_{4i+2} - a_{4i} + a_{4i+4} - a_{4i+2})/2, \tag{10}$$

then the local diffraction efficiency η_i is approximately given by

$$\eta_i(e) = (1 - e/d_i) \, [M/\pi \, \sin(\pi/M)]^2, \tag{11}$$

where e is the line edge error. When $e = 0$ equation (11) reduces to the standard formula for diffraction efficiency with no errors [equation (3.25)]. To compare with the Cox data we use the value of $e = 1.35 \, \mu$m measured by Cox, and we find a good fit using equation (11), as shown in figure (5.3). The diffraction efficiency is seen in equation (11) to decrease linearly with line edge error and is a strong function of the f-number (the smaller the f-number, the smaller the value of d in the outer zones) as observed by Farn and Goodman.[2] The diffraction efficiency reduction is independent of the number of masks and the radius.

5.2.3 Uniform Expansion

If a binary optic pattern is expanded or contracted by a factor B, we have another type of systematic error. In a Dammann grating we see from the design equations (3.15) and (3.16) that there is no change in the relative magnitudes of the diffraction orders. The efficiency depends on the ratio of a_j/a, and $Ba_j/Ba = a_j/a$. The grating period will be increased by the factor B, so the diffraction angles will decrease according to the grating equation. There will be no change in scattered light.

For the Fresnel zone plate the situation is a little different. We write the focal length f from equation (4) in terms of the jth transition point a_j, the wavelength λ, and the number of etch levels M,

$$f = a_j^2 M/2j\lambda - j\lambda/2M. \tag{12}$$

A uniform expansion of the transition points $\{a_j\}$ by the factor B results in a new focal length f' given by

$$f' = B^2 a_j^2 M/2j\lambda - j\lambda/2M$$

$$f' = B^2 f + (B^2 - 1)\lambda j/2M. \tag{13}$$

When j is small (the paraxial case) the second term of equation (13) can be ignored and the new focal length is just B^2 times the old focal length. In general the focal length depends on j, which means there is spherical aberration. The scattering in this case is a spreading of the energy around each diffraction order, and the larger the aperture, the larger the spreading.

5.3 Mask Alignment Errors

The experimental and theoretical work of Cox[1,3] shows that a fixed lateral error, or a mask alignment error, leads to a fall off in diffraction efficiency that is more severe in the outer zones of the zone plate than zones near the center. In the outer zones a fixed error represents a greater percentage of the pattern and we expect it to cause the greater efficiency decrease. By measuring the local diffraction efficiency of a number of zone plates with design FZP2B, Cox found that the fall off in efficiency is approximately linear from zone to zone. We have plotted in figure

(5.4) the diffraction efficiency in the outer most (i.e. 10th) zone as a function of the mask alignment error. The data points with a small alignment error represent typical fabrication errors, the large alignment error values were intentional.

Making use of our principle of scattering we see that mask misalignment introduces a ghost grating which is out of phase with the first grating. The result will be that the fraction of energy diffracted out of the intended diffraction order is the fraction of mask misalignment. We can derive a simple new formula for the diffraction efficiency in the jth zone when there is a mask misalignment of e:

$$\eta_j(e) = (1 - e/\Delta a_{j,ave}) \, [M/\pi \, \sin(\pi/M)]^2. \qquad (14)$$

Figure (5.4) compares Cox's measured data with equation (14), when $M = 4$ for this 4-level design, and $\Delta a_{10,ave} = 3.25$ μm for the mean separation of transition points in the tenth zone. Plots (not shown here) of Cox's data for other zones shows equally good agreement with equation (14). Note that the mean separation Δa is much smaller than the average period d used in equation (11) for line edge errors. Cox concludes that mask alignment error is more significant than etch depth error or line edge error and our formulas prove it.

Knife-edge scan measurements by Cox et. al. showed that typical edge depth, line edge and mask alignment errors do not noticeably affect the MTF.[3] The scattered light is very diffuse and very low in intensity at the first-order focal point. When the mask alignment error is significant, say 1.4 μm for the 4-level $f/10$ design, then the MTF curves will show significant astigmatism. A small f-number Fresnel zone plate could be expected to show a greater sensitivity to processing errors than

the $f/10$ design considered by Cox. On the other hand a longer wavelength would increase the zone spacing and reduce the effect of a given line edge or mask alignment error.

5.4 Quantization Errors

We now consider quantization effects. This mechanism of scattering has a degree of randomness to it, depending on the method by which the binary optic pattern is approximated by rectangles. Chen and Osborne[4] have studied the effects of quantization on cylindrical zone plates of the amplitude type. These are the old-fashioned type with alternating transparent and opaque strips, with the boundary points given by equation (4). Chen and Osborne use the Fresnel approximation but consider only cases where the results agree with the Rayleigh-Sommerfeld formula. They conclude that the effect of quantization is mainly on the intensity of the secondary maxima. The central peak intensity is not affected very much by quantization.

The method of quantization can have a considerable impact on the "sidelobe" levels, according to Chen and Osborne. For example, suppose that the device which makes the binary optic pattern can only make rectangles in increments of 0.1 μm. The ideal pattern may require one edge of the pattern to begin at 1000.04 μm and end at 1010.26 μm. One could quantize by truncating, with the result that these two transition points would be 1000.00 μm and 1010.2 μm. Or one could find the nearest increment and determine transition points 1000.0 μm and 1010.3 μm. The nearest

increment method results in much less scattered light. The smaller the f-number, the closer the spacing between transition points, and the greater the scattering from quantization errors. The larger the increment size of the device generating pattern, the greater the scattering.

In another series of computer simulations using the Fresnel formula, Kajanto et. al. investigated quantization effects on circular Fresnel zone plates.[5] They found, as did Chen and Osborne, that even rough quantization does not significantly change the spot size in the focal plane, but it does increase the amount of energy around the main "lobe" or Airy disk. They also found that the method of quantization affected the intensity of ghosts. By adding randomness to the start position of the approximating rectangles, they were able to avoid the unwanted periodicities and could greatly reduce the ghosts intensities. Reducing the ghosts results in "scattering light to form a weak widespread halolike background".[5]

One way to design a zone plate is to determine a maximum tolerable phase error $\Delta\Phi$, then approximate the ideal zone plate pattern with rectangles such that the maximum lateral deviation e of each rectangle from the ideal position gives a phase error of

$$\Delta\Phi = 2\pi e/a, \tag{15}$$

where a is the local period. With Fresnel zone plates designed this way Kajanto et. al. obtain empirically the following approximation to the diffraction efficiency in the first-order for a 2-level Fresnel zone plate:

$$\eta = \eta_0[1 - 18.5(e/a)^2] \qquad \text{if } e/a < 1/6 \tag{16}$$

where $\eta_0 = 0.101$ for an amplitude zone plate and $\eta_0 = 0.405$ for a phase-only zone plate.

In Chapter 3, it was pointed out that quantization error could be estimated by modeling it as additive noise. Using our principle of scattering we assume that the diffraction efficiency decreases linearly with additive noise;

$$\eta = \eta_0[1 - (\pi/\sqrt{3})(e/a)]. \tag{17}$$

Figure (5.5) is a comparison of the two formulas, normalized by dividing out η_0. For $e/a < 1/6$, the range of validity for equation (16), equation (17) predicts similar diffraction efficiency values.

5.5 M-level Approximation Errors

A 100 percent efficient Fresnel zone plate would have a continuously curving profile. We have seen that a 16-level binary optic approaches that level of efficiency. How is the energy distributed that is not diffracted into the first order? We will first discuss a method of calculating the angle resolved scattering, then show how geometrical optics can help to interpret the results.

5.5.1 Calculation Method

To calculate the scattering we first reduce the diffraction "ringing," caused by the finite size of the Fresnel zone plate. The amount of energy outside of the central lobe can be minimized by assuming that the Fresnel zone plate is illuminated by a gaussian beam that has low intensity at the edges of the aperture.

The Fresnel diffraction formula for cylindrical Fresnel zone plates can be quickly solved using the Fast Fourier Transform (FFT) algorithm. The details are given in Appendix A. Figure (5.6) compares the intensity as a function of angle for cylindrical Fresnel zone plates with 2-, 4-, and 16-levels. We are using designs FZP3B, FZP2A, and FZP1. The intensity is relative to the peak intensity I_p of a perfect lens with the same f-number. The data has been smoothed by averaging over sixteen successive data points. The first point plotted is the mean of the first sixteen computer generated data points, I_1 to I_{16} divided by I_p. The second point plotted is the mean of I_2 to I_{17} divided by I_p, and so forth. There are 2048 data points for each curve, and the points are spaced at intervals of λ.

5.5.2 Geometrical Optics

We notice in figure (5.6) that the intensity of the scattered light from the 2-level zone plate seems to drop to the intensity of the 4-level zone plate at about 5.7 degrees. This can be explained as follows. The diffraction efficiency equation (3.25) indicates that for a 2-level binary optic about 40.5 percent of the energy goes into the 1st-order, 40.5 percent into the minus 1st-order, 4.5 percent into the 3rd-order, and 4.5 percent into the minus 3rd-order. No energy goes into even orders for error-free binary optics optimized for the 1st-order. The 4-level (and higher) binary optic has no energy in the ± 3rd-order and no energy in the minus 1st-order. Negative orders are diverging and contribute a very weak intensity in the 1st-order focal plane.

In the 1st-order focal plane the biggest difference between the 2-level scattering and the 4-level scattering will be the contribution of the 3rd-order energy. The paraxial focus of the 3rd-order is at $f/3$, where f is the 1st-order focal length. The marginal ray (3rd-order) passes the 1st-order focal plane at (x_0', f), hence

$$x_0'/(f - f/3) = (f/2f^{\#})/(f/3) \tag{18}$$

where $f^{\#}$ is the 1st-order f-number of the zone plate. See figure (5.7). Therefore

$$x_0' = f/f^{\#}. \tag{19}$$

This is the geometric edge of the 3rd-order diffraction in the 1st-order focal plane. As seen from the center of the zone plate this edge makes an angle θ', given by

$$\tan\theta' = x_0'/f$$

$$= 1/f^{\#}. \tag{20}$$

For an $f/10$ zone plate $\theta' = 5.7$ degrees, in good agreement with our observation in figure (5.6).

5.5.3 Dependance on Focal Length

We should also point out that the relative intensity of the scattering from all diffraction orders decreases with increasing focal length. The location of the point x_0', for example, as given by equation (19), increases linearly with focal length. Figure (5.8) is a plot which compares the relative scattering intensities of designs FZP3A and FZP3B. In this figure the intensities shown are relative to their peak intensities. These are 2-level, $f/10$ cylindrical zone plates, differing only in focal length. The focal length of FZP3B is ten times the focal length of FZP3A, and the

relative scattering intensity of FZP3A is ten times the scattering intensity of FZP3B. For a cylindrical zone plate the geometrical spot width increases linearly with distance, but for a circular zone plate the spot radius increases as the square of the distance. We therefore expect for a circular zone plate that the relative intensity decreases as the square of the focal length, if the f-number is held constant. Presenting our results in terms of BTDF would normalize the f dependence.

5.6 Summary

We have seen that a binary optic can be thought of as the sum of an ideal grating, and a "ghost grating", when there are systematic errors. The ghost grating interferes with the diffraction efficiency of the ideal grating, and the energy will be redistributed among the diffraction orders. With the grating equation and geometrical optics the distribution of the scattered light can be determined.

A number of systematic errors have been discussed; uniform etch depth error, linearly varying etch depth error, line edge error, uniform expansion of the line edges, mask alignment error, quantization error, and M-level approximation errors. Previously published work to understand these errors has been forced to rely on empirical results from computer simulations. In this chapter and in Appendix A we have discussed several simplifying calculation methods. In addition we have derived some simple analytical expressions that not only agree with our own computer simulations and the results of others, but also agree with actual measurements.
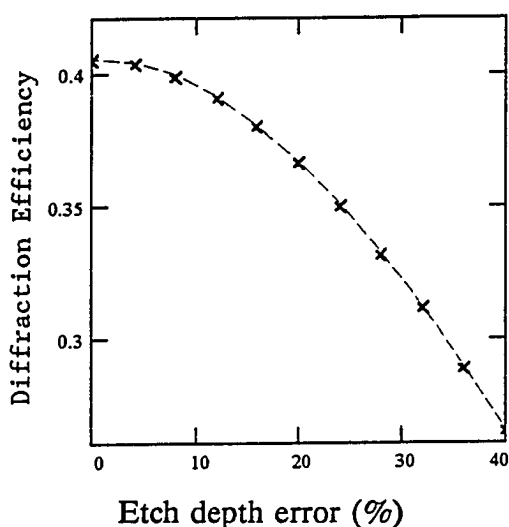
Etch depth error (%)

FIGURE 5.1, Diffraction efficiency versus uniform etch depth error. Zone plate (FZP3A). Computer simulation (X) and eq. (5.2).

Etch depth error (%)
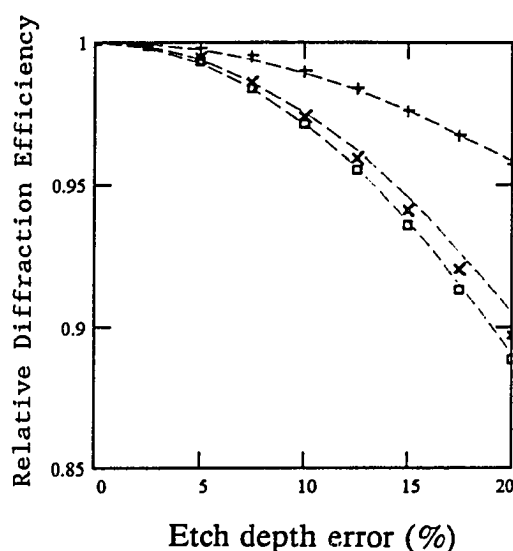
FIGURE 5.2, Relative diffraction efficiency versus systematic error (%). 16-level (0), 8-level (X), and best focus (+); compared to eq. (5.2).
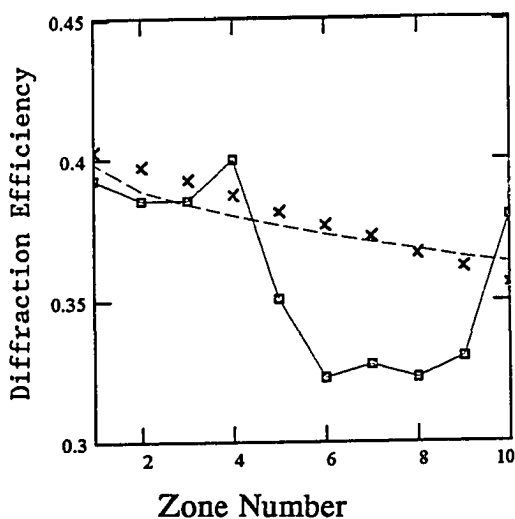
Zone Number

FIGURE 5.3, Diffraction efficiency versus zone number for a fixed (1.35 μm) line edge error. Comparing Rayleigh-Sommerfeld (X), eq. (5.11, - -), and data (-0-0-, Ref. 5.1)
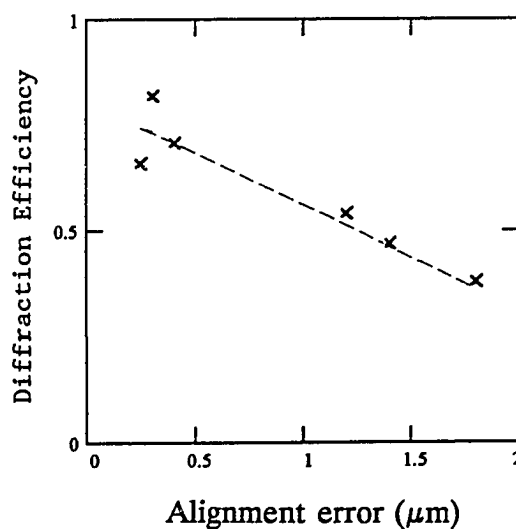
Alignment error (μm)

FIGURE 5.4, Diffraction efficiency versus mask alignment error (μm). Eq. (5.14,- -) and measured data (X, Ref. 5.3).
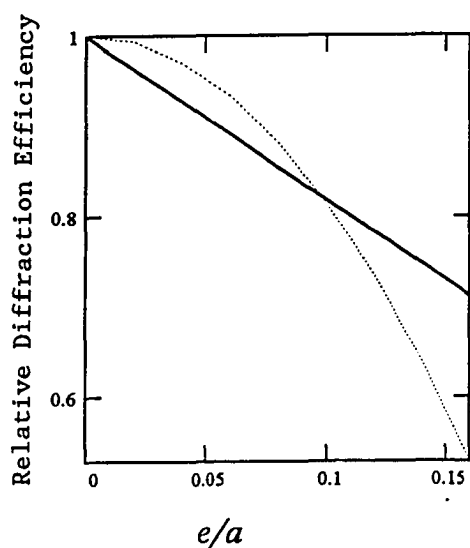
FIGURE 5.5, Relative diffraction
efficiency versus quantization error
per local period. Equation (5.17)
and equation (5.16,···).
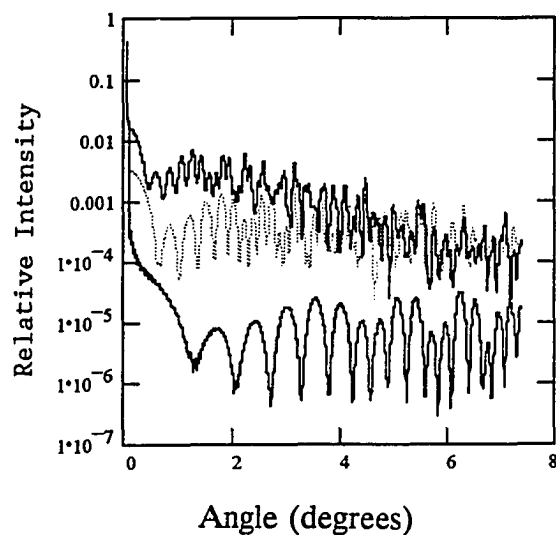
FIGURE 5.6, Relative intensity versus
angle. Cylindrical zone plate; 16-level
(FZP1, bottom), 4-level (FZP2A,···), and
2-level (FZP3B, top).
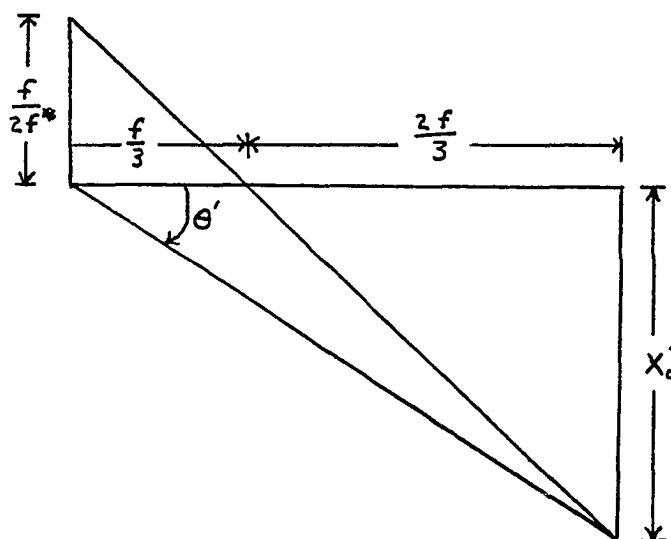
FIGURE 5.7, Geometry of 3rd order energy in the 1st-order focal plane.
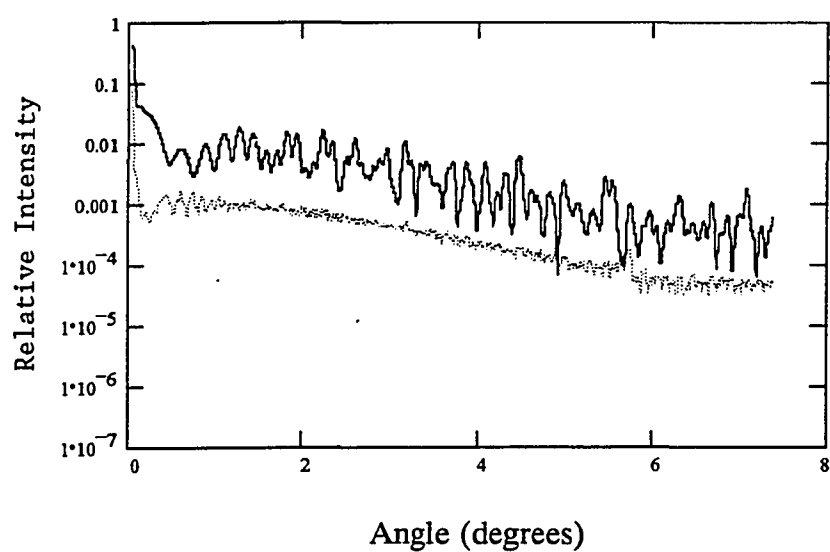
FIGURE 5.8, Relative intensity versus scattering angle. Cylindrical
Fresnel zone plate; 2-level, λ = 0.6328, f/10, f = 10 mm (top), and f
= 100 mm (bottom curve).

6. RANDOM ERRORS

Random errors are processing errors that are statistical in nature. In this chapter we will derive formulas for scattering from a binary optic with two-dimensional or one-dimensional random roughness. We shall show that these formulas can predict the scattering from Dammann gratings and Fresnel zone plates with random etch depth and random line edge errors. These formulas will be shown to agree with the average results of computer simulations. The principle of scattering which we use is that when random errors are present, then the binary optic is the sum of an ideal grating and a randomly rough surface.

6.1 Random Error Formulas

6.1.1 Two-Dimensional Random Roughness Formula

We shall now derive a formula that predicts the angular spread of scattering in the presence of random surface perturbations.[1] This formula will provide insight into the characteristics of a binary optic that cause scattering. It will enable us to quantitatively determine how scattering is dependent on the wavelength, the index of refraction of the substrate, the number of levels in the design, the step or groove width, the step height, the diffraction efficiency of each order, and the standard deviation of the random error.

We begin with the Fraunhofer diffraction integral and use an approach that parallels the traditional approach taken for flat surfaces with a slight amount of

random roughness. The roughness is assumed to perturb slightly the phase of the incident wavefront. For our work we assume a gaussian intensity profile to reduce the effects of diffraction from the binary optic aperture. A uniform intensity profile leads to a similar result.

The field in the aperture is expressed as a Fourier series, and the binary optic is modelled as a grating. We assume that the distribution of surface heights is random with an exponential autocorrelation function. The surface is then seen as the sum of an ideal (or smooth) surface, with no scattering, and a rough surface. The details are given in Appendix B. The result is

$$<I(x_0, y_0, z)> \ = \ <I_{smooth}(x_0, y_0, z)> \ + \ <I_{rough}(x_0, y_0, z)> \tag{1}$$

where $<I(x_0, y_0, z)>$ is the average intensity at the observation point. The rough and smooth components are given by

$$<I_{rough}(x_0, y_0, z)> \ = \ [\pi \sigma \tau b k(n-1)/\lambda z]^2$$

$$\Sigma_m \ |A_m|^2 \ \{1 \ + \ [(2\pi m/a - x_0 k/z)^2 + (y_0 k/z)^2]\tau^2\}^{-3/2}, \tag{2}$$

and

$$<I_{smooth}(x_0, y_0, z)> \ = \ (b^2\pi/\lambda z)^2 \ \{1 - [k(n-1)\sigma]^2\}$$

$$\Sigma_m \ |A_m|^2 \ \exp\{-2b^2\pi^2[(x_0/\lambda z - m/a)^2 + (y_0/\lambda z)^2]\}, \tag{3}$$

where σ is the *rms* roughness, τ is the correlation length, $k = 2\pi/\lambda$, b is the radius of the incident gaussian beam, n is the refractive index of the substrate, a is the period of the grating, and $|A_m|^2$ is the diffraction efficiency of the mth order. These formulas can be used for random roughness on the surface of the binary optic, or, as we shall later see, for random fabrication errors on circular Fresnel zone plates.

6.1.2 One-Dimensional Random Roughness Formula

The one-dimensional random roughness formula is derived in a fashion similar to the two-dimensional case. We assume that the diffraction equations are independent of the y_1-axis. Instead of a zero order Hankel Transform of the autocorrelation function, which we used in Appendix B for the two-dimensional case, we have the Fourier transform of the one-dimensional autocorrelation function. The end result is

$$<I(x_0, y_0, z)> = <I_{rough}(x_0, y_0, z)> + <I_{smooth}(x_0, y_0, z)>, \qquad (4)$$

where

$$<I_{rough}(x_0, y_0, z)> = [k(n - 1)\sigma/\lambda z]^2 b\tau \sqrt{(2\pi)}$$

$$\Sigma_m |A_m|^2 [1 + \tau^2(kx_0/z - 2\pi m/a)^2]^{-1}, \qquad (5)$$

and

$$<I_{smooth}(x_0, y_0, z)> = (b/\lambda z)^2 \pi \{1 - [k(n - 1)\sigma]^2\}$$

$$\Sigma_m |A_m|^2 \exp\{-2b^2\pi^2[x_0/\lambda z - m/a]^2\}. \qquad (6)$$

We shall use equations (4), (5), and (6) for Dammann gratings and cylindrical Fresnel zone plates.

6.2 Etch Depth Errors

We first consider etch depth errors where the etch depth varies randomly from groove to groove. The Dammann grating will be our first example, then we will consider a cylindrical Fresnel zone plate, and finally a circular Fresnel zone plate.

A cylindrical zone plate has straight grooves which cause an incident beam to come to a line focus. The grooves in a circular zone plate are circular.

6.2.1 Dammann Gratings

A Dammann grating is a simple, 2-level binary optic where the diffraction orders are usually well separated. This type of binary optic is relatively easy to model and interpret the results. We will compare our analytic formula (4) to computer simulations using random numbers and integral equations.

6.2.1.1 Computer Simulations

The Dammann grating does not come to a focus; we may evaluate its performance at a convenient location. Hence we can use Fourier optics analysis. We consider a number of symmetric grating designs, given in Table 6.1.

Table 6.1. Dammann Grating Designs

Design	Transition points		depth	period	λ	Equal intensity
	$a_1(\mu m)$	$a_2(\mu m)$	$h(\mu m)$	$a(\mu m)$	(μm)	(orders)
DG1	43.02	86.96	0.545	200	0.633	0, ±1, ±2, ±3
DG2	43.02	86.96	9.129	200	10.6	0, ±1, ±2, ±3
DG3	25	75	0.692	200	0.633	±2
DG4	50	100	0.692	200	0.633	±1
DG5	10.75	21.74	0.545	50	0.633	0, ±1, ±2, ±3

For a Dammann grating we need only consider the one-dimensional problem in x_1, as we did for the cylindrical zone plate in Chapter 5. Ignoring some unimportant phase terms in front of the Fraunhofer formula (2.12) the field at the distance z from the grating is given by

$$U(x_0, z) = 1/z\lambda \int_{-\infty}^{\infty} U(x_1)\, G(x_1)\, \exp(-ikx_0x_1/z)\, dx_1 \qquad (6)$$

where $G(x_1)$ is the incident gaussian field, and $U(x_1)$ is the phase profile imparted by the Dammann grating. Equation (6) is the Fourier transform of the product of $G(x_1)$ and $U(x_1)$, so we use FFTs to find the convolution of the Fourier transform of each term, as we did for the cylindrical Fresnel zone plate in Appendix A. The details for Dammann gratings with random etch depth errors will be found in Appendix C.

Figure (6.1) shows the relative intensity of Dammann grating design DG1 with a random etch depth error. In addition to the values in Table 6.1, $b = 4.5$ mm, $f = 1.82$ m, and $n = 1.45702$. The values for the A_m are calculated from equations (3.15) and (3.16). The standard deviation of the etch depth error, δ, is 0.01 μm for the case shown in figure (6.1). The intensity profile was calculated with four sets of random numbers and the results were averaged together. The results were also smoothed by averaging over sixteen data points, and normalized by dividing by the peak intensity in the error-free case. In figure (6.1) the zero diffraction order is at zero degrees, and the next four positive diffraction orders are also shown. The negative diffraction orders would be similar and are not plotted.

6.2.1.2 Roughness and Correlation Length

Also shown in figure (6.1) is a plot of equation (4). The correlation length τ we estimate to be approximately equal to the width of the etched portion of the Dammann grating. In this portion the correlation is unity, falling off to zero outside of the etched area. For design DG1 there are two etched portions per period; the width of each is $86.96 - 43.02 = 43.94$, so we set $\tau = 44$. The *rms* roughness is

$$\sigma = (\delta^2\tau/a + 0 + \delta^2\tau/a + 0)^{\frac{1}{2}}$$

$$= \delta\ (2\tau/a)^{\frac{1}{2}}. \tag{7}$$

6.2.1.3 Standard Deviation Dependence

In figure (6.2) the standard deviation of the etch error has been increased to $\delta = 0.1\ \mu$m. The average of four computer simulations is shown together with the 1-D roughness equation (4). The value of τ is again 44 μm, and we use equation (7) to calculate the *rms* roughness. It appears to be a good fit. We see from equation (5), and it is confirmed in the computer simulations illustrated in figures (6.1) and (6.2), that the scattering is proportional to δ^2, since δ and σ are linearly related through equation (7).

The diffraction efficiency can be given by equation (6), which indicates that for random etch depth errors with two etched surfaces per period, the diffraction efficiency decreases by

$$\eta = \eta_0\{1 - [k(n-1)\sigma]^2\}$$

$$= \eta_0\{1 - (2\tau/a)[k(n-1)\delta]^2\} \tag{8}$$

using equation (7). The diffraction efficiency with no errors is η_0. The diffraction efficiency is seen to fall off with the square of the etch depth error.

6.2.1.4 Probability Function Dependence

Does it matter whether the random etch depth errors are distributed uniformly, or would an exponential or gaussian probability density function make a difference? We calculate the relative intensity of Dammann grating DG1 with a random etch depth error of $\delta = 0.1$ μm using uniform, gaussian, and exponential probability density functions. Figures (6.2), (6.3), and (6.4) show very little difference in scattering levels, and this illustrates the fact that it does not matter which probability density function is used.

6.2.1.5 Wavelength Dependence

What effect does wavelength have on the scattering? Consider design DG2, which has the same transition points (hence the same pattern), but differs in the design wavelength and etch depth. The relative intensity profile is shown in figure (6.5), along with a plot of the one-dimensional scattering formula (4). The agreement of formula (4) with the computer simulations is as good for the 10.6 μm wavelength design as it is for the 0.6328 μm wavelength design. The etch depth error has a standard deviation of 0.1 μm, the same as the error illustrated in figures (6.2), (6.3), or (6.4), but the relative scattering is much less. This is in keeping with the prediction of equations (5) and (6) that the relative scattering is

$$< I_{rough}(x_0, y_0)> / I_{peak} =$$

$$[k(n - 1)\sigma/\lambda z]^2 b \tau \sqrt{(2\pi)} \; \Sigma_m |A_m|^2 [1 + \tau^2(kx_0/z - 2\pi m/a)^2]^{-1}/(b/\lambda z)^2 \pi |A_{peak}|^2$$

$$= [k(n - 1)\sigma]^2 \tau \sqrt{2} \; \Sigma_m |A_m/A_{peak}|^2 [b\sqrt{\pi} + b\sqrt{\pi} \tau^2(kx_0/z - 2\pi m/a)^2]^{-1}, \qquad (9)$$

assuming there is not much overlap of energy between diffraction orders. We see from equation (9) that the relative scattering is proportional to k^2 and therefore proportional to $1/\lambda^2$, as we can also see by comparing figures (6.5) and (6.2).

If design DG2 is illuminated by light of wavelength λ = 0.4971 μm, the beam will still be divided into seven diffraction orders of equal intensity, as before. But the relative scattering for the same etch depth error is much greater at the shorter wavelength, as illustrated by comparing figure (6.5) with figure (6.6). Again equation (4) correctly predicts the intensity levels. We see again that the relative scattering is proportional to $1/\lambda^2$.

6.2.1.6 Groove Width Dependence

Do the results of designs DG1, and DG2 apply to other designs? Consider figure (6.7) which illustrates the relative intensity of Dammann grating DG3 with a random etch depth error of δ = 0.1 μm. In this design there are two etched areas in each period, the width of each etched area is 50 μm. We therefore set τ = 50 μm. The *rms* error is again found by equation (7). The one-dimensional roughness scattering formula (4) is reasonably accurate, though it fails to predict the rapid fall off in intensity near the 4th-diffraction order (0.72 degrees). Figure (6.9), which illustrates the case of design DG4, also tends to predict a smooth scattering level and ignores the very low intensities. Design DG4 is somewhat degenerate, only one

transition point is needed. There is an unetched portion of width 100 μm and an etch portion of 100 μm in each period. In this case $\tau = 100$ μm and $\sigma = \delta \sqrt{(\tau/a)}$.

6.2.2 Cylindrical Fresnel Zone Plates

Continuing with our examination of random etch depth errors, we now look at cylindrical Fresnel zone plates. This type of zone plate will show one-dimensional scattering characteristics. The cylindrical Fresnel zone plate creates a line focus (at $x_0 = 0$) for each diffraction order at a distance from the zone plate that depends on the diffraction order.

6.2.2.1 Computer Simulations

We calculate the relative intensity of the cylindrical Fresnel zone plate by the FFT method, using Appendix A and a random phase term, as in Appendix B. The relative *scatter* intensity is the difference between the intensity with an etch depth error and the intensity with no error, divided by the peak intensity with no error. Figure (6.9) shows the relative scatter intensity as a function of angle for the case δ = 0.1 μm and design FZP1 (from Table 5.1). The intensities were averaged for four simulations of random etch depth error. The data has been smoothed over sixteen data points, as usual.

6.2.2.2 Relative Scattering From Roughness

We shall concern ourselves only with the scattering from roughness in the first diffraction order; in the 1st-order focal plane scattering from other diffraction orders

is insignificant. The relative scatter intensity from roughness is therefore [using equation (9)],

$$<I_{rough}(x_0, y_0)>/I_{peak} = [k(n - 1)\sigma]^2\tau\sqrt{2}\ [b\sqrt{\pi} + b\sqrt{\pi}(\tau kx_0/z)^2]^{-1} \qquad (10)$$

for the cylindrical Fresnel zone plate. As before, the correlation length τ is the mean width of the etched portions;

$$\tau = \text{mean}\{a_{j+1} - a_j\}, \qquad (11)$$

which for design FZP1 gives $\tau = 1.58\ \mu m$. For the Fresnel zone plates all levels were assumed to have some etch depth error, so the mean is taken over all intervals. This also implies that $\sigma = \delta$. Figure (6.9) shows formula (10) compared to the computer simulation, and the agreement is very good.

6.2.2.3 Standard Deviation Dependence

Figure (6.9) should be compared to figure (6.10) where this time the standard deviation of the etch depth error is $\delta = 0.01\ \mu m$. The scattering drops two orders of magnitude with one order of magnitude decrease in the error. Again equation (10) provides a good prediction of the average value of the scattering at a particular angle. Equation (10) predicts that the scattering is proportional to σ^2, and $\sigma^2 = \delta^2$ in this case.

6.2.2.4 Dependence on Number of Levels

A 4-level Fresnel zone plate (FZP2A) shows behavior similar to the 16-level zone plates just considered. Figure (6.11) illustrates this case when the standard deviation of the etch depth error is $\delta = 0.1\ \mu m$. A plot of equation (10) is shown

for comparison. Equation (11) gives τ = 6.33 μm; and σ = δ. The random scatter in this case is less than the systematic error for angles greater than about three degrees; averaging is necessary to reveal the scattering. The gaps in the curve indicate where the average with etch depth errors is less than the intensity without etch depth errors. Further averaging should make the relative scatter intensity converge to the plot of equation (10).

6.2.2.5 Wavelength and Refractive Index Dependence

Figure (6.12) is the relative scatter intensity plot for design FZP4. At this longer wavelength (10.6 μm) the plot is much smoother than the 0.6328 μm case. Also shown in figure (6.12) is a plot of equation (10). For design FZP4, τ = 26.42 μm. Comparing with figure (6.9) we see that the relative scattering levels are about the same for the two wavelengths with the same average etch depth error. While the longer wavelength lowers the scattering levels, the higher index of refraction and the longer correlation length increase the scattering, as seen in equation (10).

6.2.3 Circular Fresnel Zone Plates

We now consider random etch depth errors in circular Fresnel zone plates.

6.2.3.1 Computer Simulations

The scattering from random etch depth errors was calculated using the Fresnel formula (4.12), which for a circular Fresnel zone plate can be simplified to

$$I(r) = |(2\pi/\lambda z)\Sigma_j \phi_j \int_{aj}^{aj+1} \alpha J_0(\alpha kr/z)\exp[(i\pi/\lambda z - 1/b^2)\alpha^2] d\alpha|^2, \quad (12)$$

where J_0 is the zero-order Bessel function, and the limits of the integration for each term are a_j to a_{j+1}. The variable r is the distance from the focal point $(0,0, z)$ in the observation plane to the point (x_0, y_0, z). The phase function ϕ_j has a random error term.

6.2.3.2 Roughness Formula

For the circular Fresnel zone plate we must use the two-dimensional roughness scattering formula (1). For our proposes here we want the scatter intensity of the "rough" part divided by the peak intensity of the smooth part in the absence of roughness. Using equations (2) and (3) we find

$$<I_{rough}(x_0, y_0)>/I_{peak} = [k(n-1)\sigma\tau/b]^2 [1 + (\tau kx_0/z)^2 + (\tau ky_0/z)^2]^{-3/2} \quad (13)$$

for the two-dimensional or circular Fresnel zone plate. The *rms* roughness will be as before, $\sigma = \delta$; the correlation length will again be given by equation (11).

Figure (6.13) is a plot of the relative scatter intensity for design FZP4 with random etch depth error characterized by the standard deviation $\delta = 0.1 \mu m$. Also shown in figure (6.13) is a plot of equation (13). When we compare figure (6.13) to figure (6.12) we see that there is less scattering relative to the peak intensity in the two-dimensional case, because the peak intensity is higher.

6.2.3.3 One-Dimensional Versus Two-Dimensional

The relative scattering for the circular Fresnel zone plate with design FZP1 is shown in figure (6.14). The relative scattering in the two-dimensional case is much

less than the relative scattering in the one-dimensional case [figure (6.9)] for the same average etch depth error.

6.2.3.4 Wavelength, Refractive Index, F-Number, and Number of Levels Dependence

Design FZP1 also has much less relative scatter than design FZP4. Considering only the terms which change from design FZP1 to FZP4, the magnitude of the scattering is proportional to $[k(n - 1)\tau]^2$. The correlation length τ is actually a function of the wavelength λ, the number of levels M, and the f-number $f^{\#}$. To show this we derive an approximate value for τ. If the f-number is not too small we can approximate the transition point a_j from equation (5.4) by using the paraxial approximation,

$$a_j^2 \approx 2\lambda f j/M.$$

Hence,

$$a_{j+1}^2 - a_j^2 \approx 2\lambda f/M$$

and therefore

$$a_{j+1} - a_j \approx 2\lambda f/(Ma_{j+1} + Ma_j).$$

We have then that

$$\tau = \text{mean}(a_{j+1} - a_j) \approx (2\lambda f/M) \text{ mean}[1/(a_{j+1} + a_j)]$$

$$\tau \approx 4\lambda f^{\#}/M. \tag{14}$$

Using equation (14) we see that $k\tau$ has the same value in both design FZP1 and FZP4, and therefore the difference in relative scattering levels is due solely to the difference in refractive index of the two designs.

6.2.3.5 Standard Deviation Dependence

Figure (6.15) illustrates the fact that a random etch depth error with standard deviation $\delta = 0.01$ μm has 1/100th the scatter intensity of an error of $\delta = 0.1$ μm. The two-dimensional roughness formula (13) again predicts well the average scattering level.

6.3 Line Edge Errors

We now examine the effects of random line edge error in Dammann gratings and Fresnel zone plates. We use the FFT formulas for the Dammann grating and the cylindrical zone plate (Appendices A and C), and the Fresnel diffraction integral (12) for the circular zone plate. The location of the line edge is made random by adding e_j to the transition point a_j, where e_j is a random number with mean of zero and standard deviation δ. Our random roughness formulas were derived under the assumption of a small random phase shift to the surface, but we will attempt to use these formulas for the random line edge errors.

6.3.1 Dammann Gratings

6.3.1.1 Roughness and Correlation Length

A random line edge error could be regarded as a perturbation from the ideal pattern. If we subtract the ideal grating from the real life grating with line edge errors we are left with a surface that has random width surface roughness located approximately at the transition points. The height of the perturbation is the step height, h. The *rms* roughness σ is given by

$$\sigma = [m <w> h^2/a]^{1/2}, \tag{15}$$

where m is the number of edges which vary from their proper positions in a period a, and $<w>$ is the mean width of the perturbations. For a uniform probability over either the interval $[0, \ \delta\sqrt{3}]$ or $[-\delta\sqrt{3}, \ 0]$,

$$<w> = (\delta\sqrt{3})/2, \tag{16}$$

where δ is the standard deviation of the error. We can calculate σ from (15), (16), and a knowledge of the standard deviation of the line edge error.

In the derivation of the roughness formula (4), we assumed a two-sided exponential probability density function, $(1/2\tau) \exp(-|x|/\tau)$. Hence δ is related to the correlation length τ by

$$
\begin{aligned}
\delta \quad &= [(1/2\tau) \int_{-\infty}^{\infty} x^2 \exp(-|x|/\tau) \, dx]^{1/2} \\
&= [(1/\tau) \int_{0}^{\infty} x^2 \exp(-|x|/\tau) \, dx]^{1/2} \\
&= [(1/\tau) \, 2\tau^3]^{1/2} \\
&= \tau\sqrt{2}. \tag{17}
\end{aligned}
$$

With σ and τ now given by equations (15), (16), and (17), we can compare the computer simulations to a plot of the roughness formulas.

6.3.1.2 Standard Deviation Dependence

Figure (6.16) shows the relative intensity versus angle for Dammann grating DG1 with line edge error $\delta = 1 \ \mu$m. A plot of the roughness formula (4) is also shown (it is almost constant at 10^{-4}). There are four edges per period in design DG1, so $m = 4$ in equation (15). The step height h is given in Table 6.1. The scattering

predicted by equation (4) appears to be a little high compared to the computer simulation, possibly because the assumption made in the derivation that h is small compared to λ is not very good.

Figure (6.17) shows the case for $\delta = 0.1$ μm. Equation (5) correctly predicts an increase in scattering of magnitude 100 when δ is increased by a factor of 10. To see this in equation (5) we note that

$$<I_{rough}(x_0, y_0)> \propto \sigma^2 \tau$$

$$\propto <w> \tau$$

$$\propto \delta^2 \tag{18}$$

using equations (15), (16), and (17).

6.3.1.3 Period Dependence

When the period is decreased by a factor of four (design DG5), the scattering increases by a factor of four, as illustrated in figure (6.18). The scattering is proportional to σ^2 in equation (5), and therefore proportional to $1/a$, according to equation (15).

6.3.1.4 Wavelength Dependence

When the wavelength is increased, and the step height is increased proportionally (design DG2), the scattering intensity remains the same for random line edge errors. This is illustrated by comparing figure (6.19) to figure (6.16). The roughness formula (9) predicts this since the relative scattering is proportional to $(\sigma k)^2$, which is proportional to $(h/\lambda)^2$ by equation (15).

Figure (6.20) is a plot of the relative intensity from Dammann grating DG2 when illuminated by a beam of wavelength $\lambda = 0.4971\ \mu$m. The scattering levels are the same at this wavelength as they are when the illumination is 10.6 μm, as illustrated in figure (6.19). From this it might be erroneously concluded that line edge errors are not dependent on wavelength. If the wavelength is gradually changed from 10.6 μm to 0.4971 μm we see systematic errors as well as random errors with a wavelength dependence. At 0.4971 μm the etch depth is more than one wavelength deep; the important parameter is the relative phase difference between the etched and unetched portions of the surface. For purposes of computing the roughness scattering we should use the effective height h_{eff},

$$h_{\text{eff}} = \text{mod}[(n - 1)h, \lambda]/(n - 1), \qquad (19)$$

where h is the actual etch depth, n is the refractive index, and $\text{mod}[(n - 1)h, \lambda]$ is the remainder when $(n - 1)h$ is divided by λ. The value of h_{eff} from equation (19) was used in equation (15) and a plot of the roughness scattering formula (4) is also shown in figure (6.20). The scattering levels are the same in figures (6.19) and (6.20) because at 0.4971 μm and at 10.6 μm the phase difference is effectively the same.

6.3.1.5 Etch Depth and Diffraction Efficiency Dependence

Figure (6.21) illustrates the relative intensity (with peak in the second diffraction order) and scattering for design DG3 when the standard deviation of the line edge error is 0.1 μm. The relative scattering levels are somewhat lower for design DG3 than for design DG1, as we see by comparing figure (6.21) to figure

(6.17). While the increased etched depth of design DG3 increases the scattering by a little, the peak intensity is much higher for design DG3 [$\eta(2)$ = 0.405] than for design DG1 [$\eta(0)$ = 0.121], using equation (3.16). Since the scattering intensity is plotted relative to the peak intensity, the scattering in design DG3 appears to be less than the scattering in design DG1.

6.3.1.6 Edges per Period Dependence

The relative intensity of the scattering of design DG4 with line edge errors, δ = 0.1 μm is shown in figure (6.22). The relative scattering is a little lower in design DG4 than design DG3 [compare figure (6.22) to figure (6.21)]. The peak intensity for design DG4 is in the first diffraction order and has a diffraction efficiency of 0.405, the same as the diffraction efficiency of design DG3 in the second order. In design DG4 there are only two edges per period, so in equation (15), m = 2. In the other Dammann grating designs there are four edges per period so the effective roughness is greater in the other designs, and therefore the scattering is greater.

6.3.2 Cylindrical Fresnel Zone Plates

6.3.2.1 Computer Simulations

We now consider line edge errors in Fresnel zone plates. Figure (6.23) is a plot of the relative scatter intensity of a cylindrical zone plate, design FZP1. The systematic error part of a 16-level zone plate has been subtracted out, leaving some holes in the data where the random line edge scatter has reduced the scattering to

less than the systematic error. We use the FFT formulas (Appendix C) and add a random number e_j to each transition point a_j.

6.3.2.2 Roughness and Correlation Length

To plot the one-dimensional roughness formula for cylindrical Fresnel zone plates, equation (10), we use the same formulas for τ and σ as we did for Dammann gratings. We could interpret equation (15) for the roughness using $m = T$ for the number of transition points or edges in the half-width or radius a of the design. We then write

$$\sigma = h[\delta\sqrt{3}Tf^{\#}/f]^{1/2} \tag{20}$$

using equations (15) and (16). A plot of equation (10) is also shown in figure (6.23).

6.3.2.3 Standard Deviation Dependence

The relative scattering increases by a factor of 100 when the line edge errors increase by a factor of 10, as can be seen by comparing figure (6.24) to figure (6.23). The roughness formula (5) predicts that $I_{rough} \propto \delta^2$ as discussed in relation to Dammann gratings [see equation (18)].

6.3.2.4 Number of Levels Dependence

Figure (6.25) is a plot of the relative scatter intensity for design FZP2, a four-level Fresnel zone plate. This design has one fourth the transition points T of design FZP1, which is a sixteen-level design, but the step heights h are four times as high, since

$$h = \lambda/[(n - 1)M]. \tag{21}$$

Roughness is proportional to $h\sqrt{T}$, as we see from equation (20), so design FZP2 has twice the roughness of design FZP1, and therefore four times the scattering. This can be seen by comparing figure (6.25) to figure (6.24).

6.3.2.5 Wavelength Dependence

The scattering from line edge errors is small compared to the systematic errors of design FZP4. In figure (6.26) the relative scatter intensity has a number of gaps where the random plus systematic error is less than the systematic error alone. Comparing figure (6.26) to figure (6.24) we see that there is less scattering from design FZP4. It appears to be a result of differences in the factors λ, σ, and n. We have from equation (5) or (10)

$$I_{rough} \propto [(n - 1)\sigma/\lambda]^2,$$

which using equation (20) is

$$\propto (n - 1)^2 \, h^2 \, T/\lambda^2.$$

T is roughly proportional to $1/\lambda$, and using equation (21) we have

$$I_{rough} \propto (n - 1)^2 \, [\lambda/(n - 1)]^2 \, 1/\lambda^3,$$

hence

$$I_{rough} \propto 1/\lambda. \tag{22}$$

We see that the refractive index n does not really affect the scattering intensity for line edge errors, and the net result is that the scattering is inversely proportional to

the wavelength. If a particular design were used at two wavelengths, then T and σ would be constant, and the scattering would be proportional to $1/\lambda^2$.

6.3.3 Circular Fresnel Zone Plates

Finally we consider designs FZP1 and FZP4 for circular Fresnel zone plates. The intensity is found using equation (12) with random numbers e_j added to the transition points a_j. These results are compared to the roughness formula (13). As before, the correlation length is $\delta/\sqrt{2}$, and the *rms* roughness σ is the *rms* surface height with respect to the design step height. For a circular Fresnel zone plate

$$\sigma = \{\Sigma_j[(a_j + <w>)^2 - a_j^2](h/a_T)^2\}^{1/2},$$

$$= (h/a_T)[T<w>^2 + 2<w>\Sigma_j a_j]^{1/2}, \text{ and}$$

$$= (h/a_T)[T\delta^2 3/4 + \delta\sqrt{3}\ \Sigma_j a_j]^{1/2}. \tag{23}$$

Figure (6.27) shows that the relative intensity of the scatter is down six or seven orders of magnitude from the peak intensity, even for a fairly large line edge error of 1 μm standard deviation. If we compare this to figure (6.28) for design FZP4 we see even less scattering. The roughness scattering formula does not seem to be a good fit to the computer simulation in this case, but there are also many gaps in the net scattering plot because the random scattering is only about 1/100 the systematic scattering in this case. As in the one-dimensional case, the roughness formula predicts a decrease in relative scattering proportional to $1/\lambda$ for optimized designs.

6.4 Summary

We have derived two basic formulas for roughness in binary optics, one formula for two-dimensional roughness, and another formula where the surface height variations are constant along one axis. We compared these formulas with computer simulations for a number of cases. The cases considered show the dependence of scattering levels on the wavelength, the standard deviation of the error, the number of levels in the design, the step width and height, the refractive index of the substrate, the number of edges per period, the diffraction efficiency, the f-number, the period length, whether errors are one-dimensional or two-dimensional, and the probability density function of the error. We conclude that in all cases the roughness formulas predict well the scattering from random etch depth and random line edge errors.

The parameters which describe the roughness of a surface are the *rms* roughness σ and the correlation length τ. We have calculated σ and τ for a number of examples, and have derived several formulas to assist in determining these parameters. We have also derived several formulas that simplify the numerical evaluation of the computer simulations.

FIGURE 6.1, Relative intensity, Dammann grating (DG1). Etch depth error δ = 0.01 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.4, ⋯).

FIGURE 6.2, Relative intensity, Dammann grating (DG1). Etch depth error δ = 0.1 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.4, ⋯).
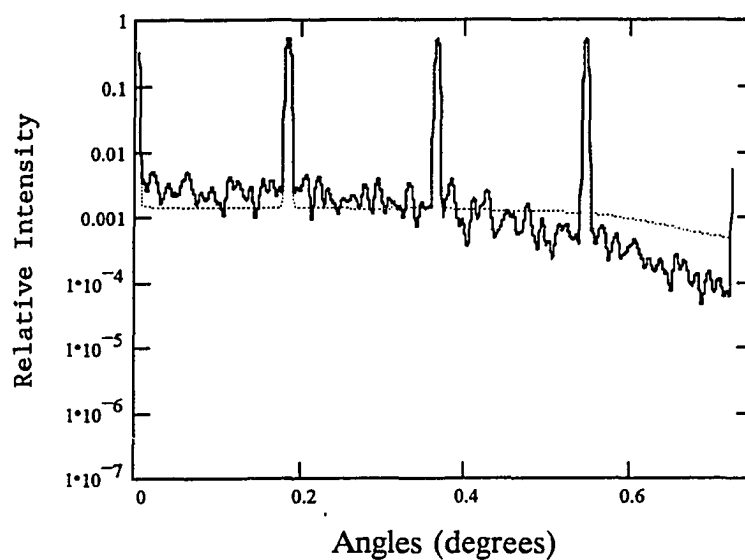
FIGURE 6.3, Relative intensity, Dammann grating (DG1). Etch depth error δ = 0.1 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.4, ⋯). Exponential probability.

FIGURE 6.4, Relative intensity, Dammann grating (DG1). Etch depth error δ = 0.1 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.4, ⋯). Normal probability.
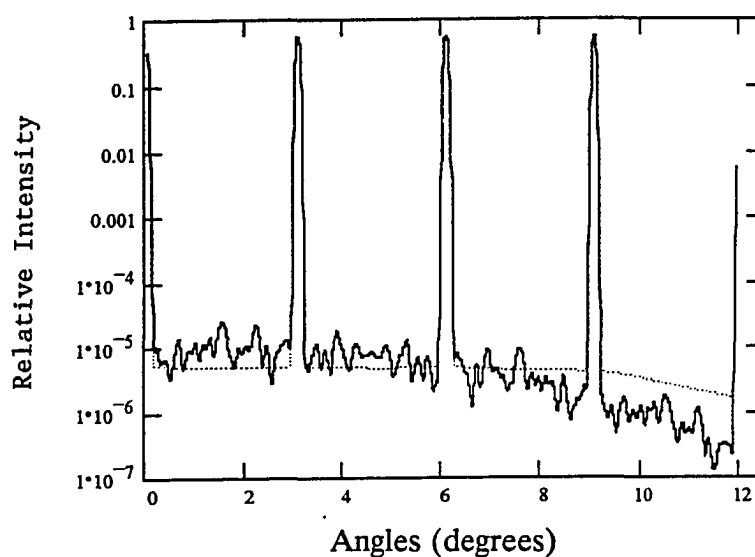
FIGURE 6.5, Relative intensity, Dammann grating (DG2). Etch depth error δ = 0.1 μm. Average of 4. Compared to equation (6.4, ···). Incident wavelength λ = 10.6 μm.
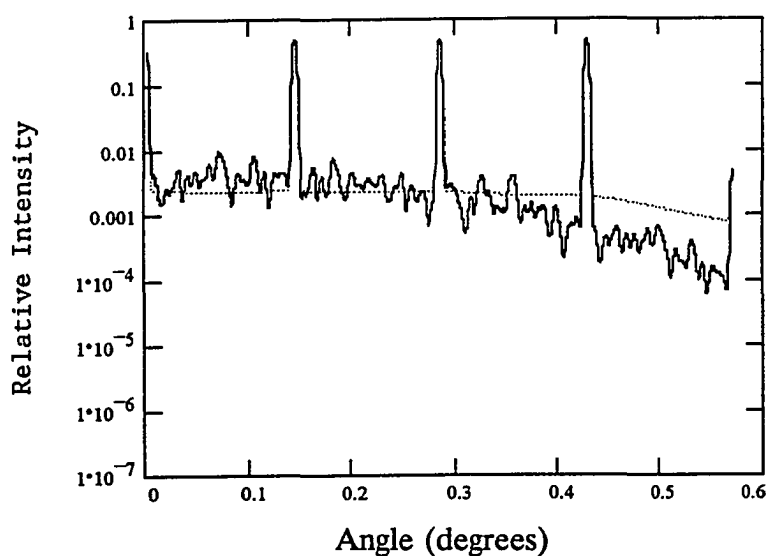
FIGURE 6.6, Relative intensity, Dammann grating (DG2). Etch depth error δ = 0.1 μm. Average of 4. Compared to equation (6.4, ···). Incident wavelength λ = 0.4971 μm.

FIGURE 6.7, Relative intensity, Dammann grating (DG3). Etch depth error δ = 0.1 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.4, ⋯).
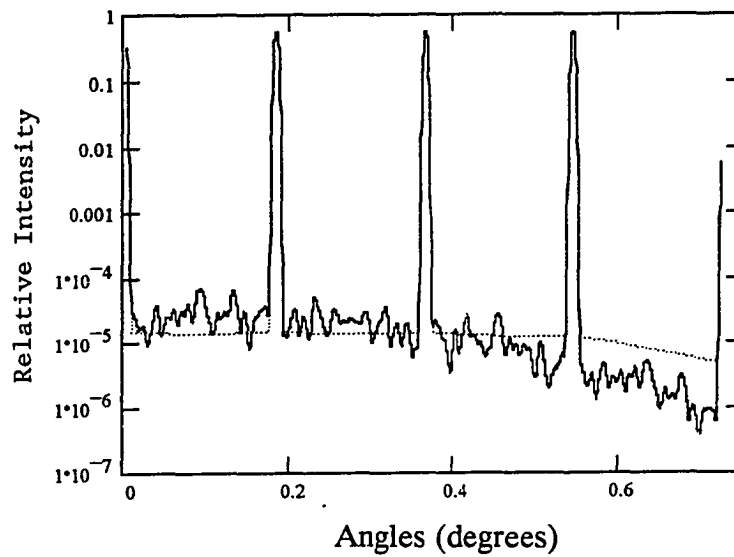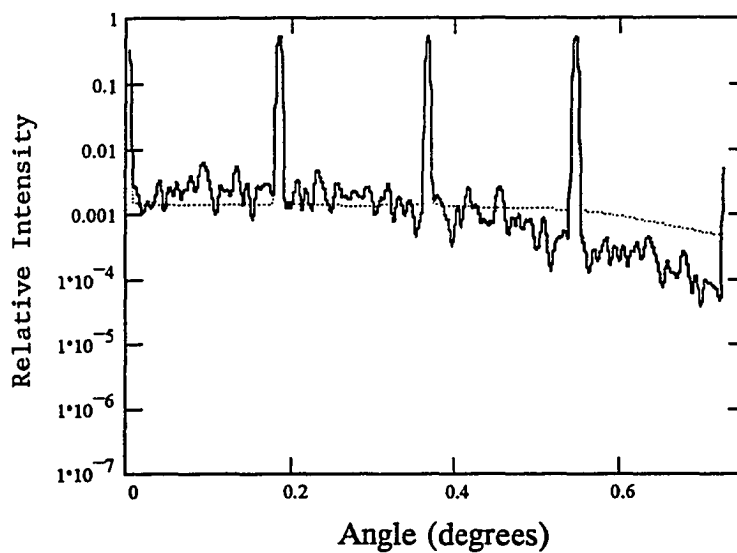

FIGURE 6.8, Relative intensity, Dammann grating (DG4). Etch depth error δ = 0.1 μm. Average of 2. Compared to 1-D roughness scattering formula, equation (6.4, ⋯).

FIGURE 6.9, Relative scatter intensity, cylindrical Fresnel zone plate (FZP1). Etch depth error δ = 0.1 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.10, ···).

FIGURE 6.10, Relative scatter intensity, cylindrical Fresnel zone plate (FZP1). Etch depth error δ = 0.01 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.10, ···).

FIGURE 6.11, Relative scatter intensity, cylindrical Fresnel zone plate (FZP2A). Etch depth error δ = 0.1 μm. Average of 16. Compared to 1-D roughness scattering formula, equation (6.10, ⋯).

FIGURE 6.12, Relative scatter intensity, cylindrical Fresnel zone plate (FZP4). Etch depth error δ = 0.1 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.10, ⋯).

FIGURE 6.13, Relative scatter intensity, circular Fresnel zone plate (FZP4). Etch depth error δ = 0.1 μm. Average of 8. Compared to 2-D roughness scattering formula, equation (6.13, ⋯).

FIGURE 6.14, Relative scatter intensity, circular Fresnel zone plate (FZP1). Etch depth error δ = 0.1 μm. Compared to 2-D roughness scattering formula, equation (6.13, ⋯).

FIGURE 6.15, Relative scatter intensity, circular Fresnel zone plate (FZP1). Etch depth error δ = 0.01 μm. Average of 3. Compared to 2-D roughness scattering formula, equation (6.13, $\cdots$).
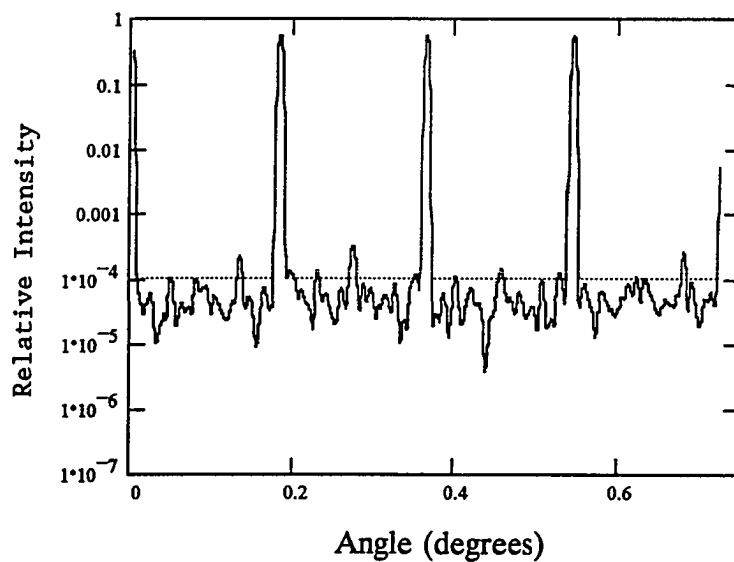
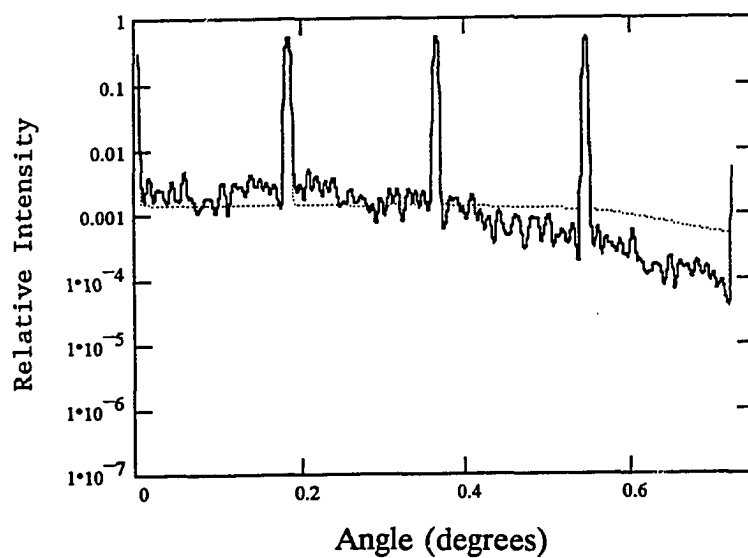
FIGURE 6.16, Relative intensity, Dammann grating (DG1). Line edge error δ = 1 μm. Average of 2. Compared to 1-D roughness scattering formula, equation (6.4, $\cdots$).
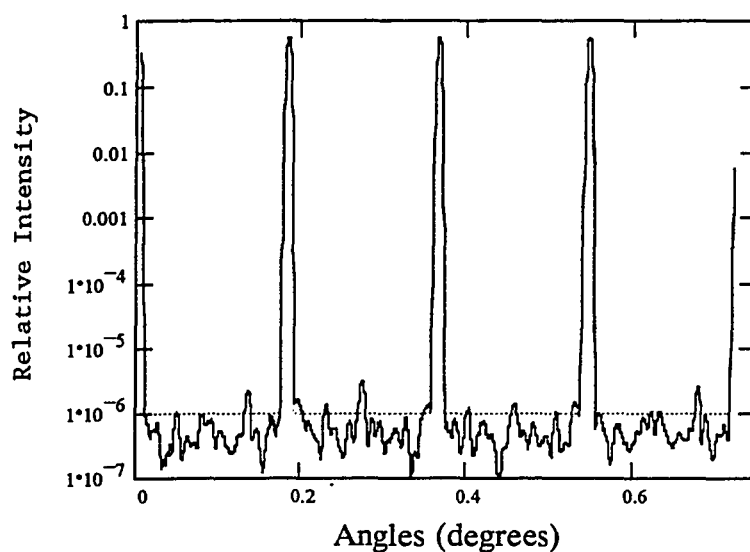
FIGURE 6.17, Relative intensity, Dammann grating (DG1). Line edge error δ = 0.1 μm. Average of 2. Compared to 1-D roughness scattering formula, equation (6.4, ⋯).
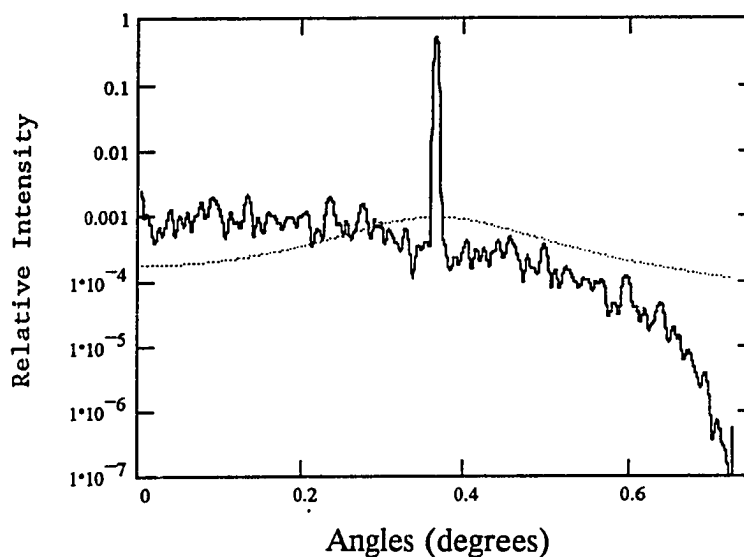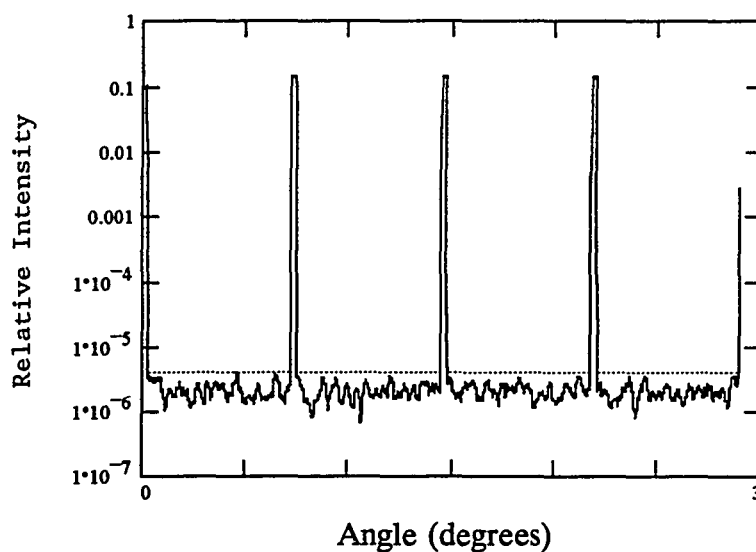

FIGURE 6.18, Relative intensity, Dammann grating (DG5). Line edge error δ = 0.1 μm. Average of 2. Compared to 1-D roughness scattering formula, equation (6.4, ⋯).
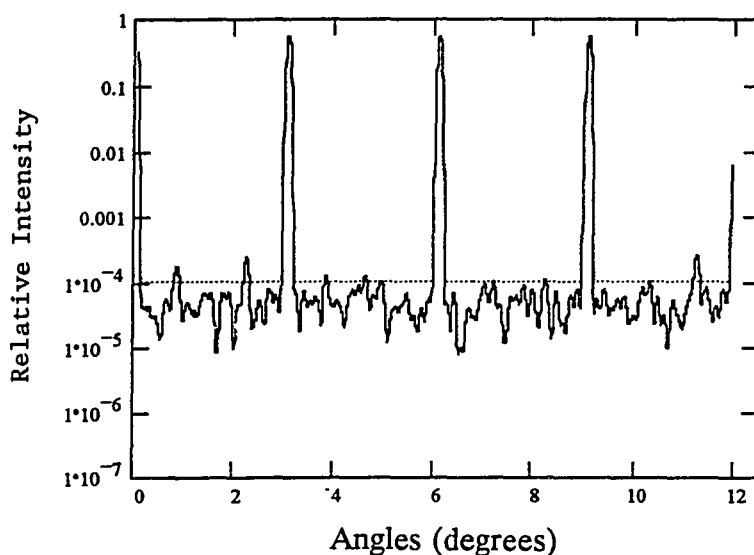
FIGURE 6.19, Relative intensity, Dammann grating (DG2). Line edge error δ = 1 μm. Average of 2. Compared to 1-D roughness scattering formula, equation (6.4, ···). Illumination λ = 10.6 μm.
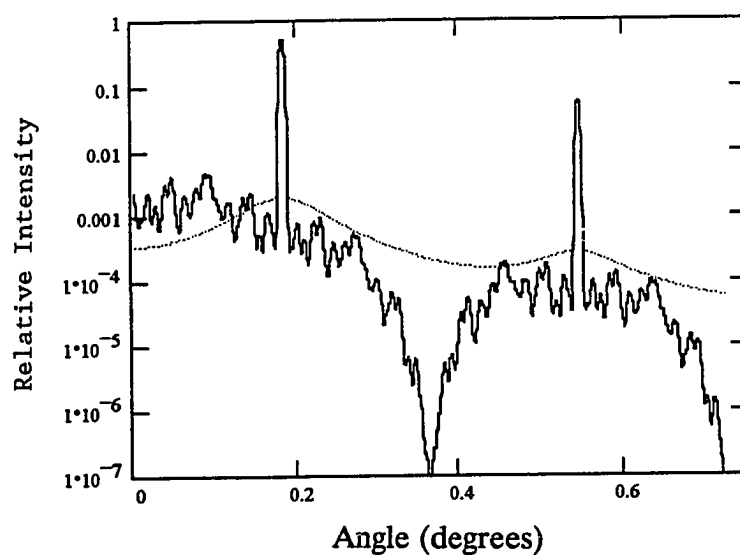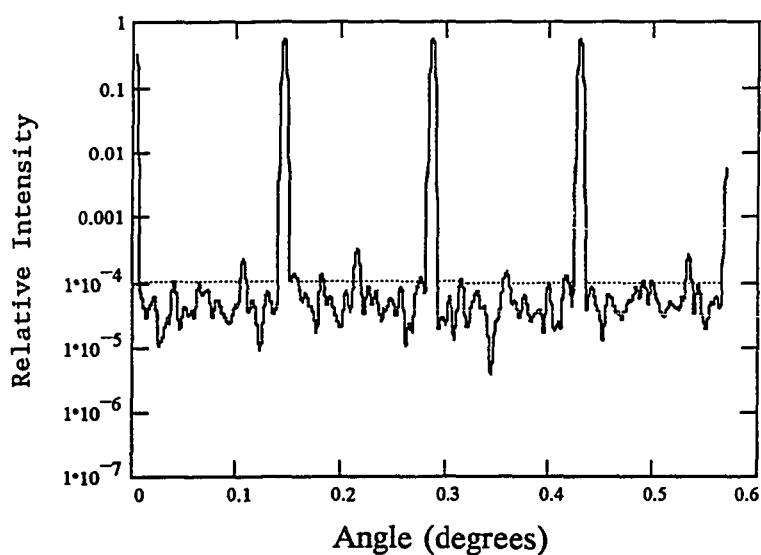

FIGURE 6.20, Relative intensity, Dammann grating (DG2). Line edge error δ = 1 μm. Average of 2. Compared to 1-D roughness scattering formula, equation (6.4, ···). Illumination λ = 0.4971 μm.
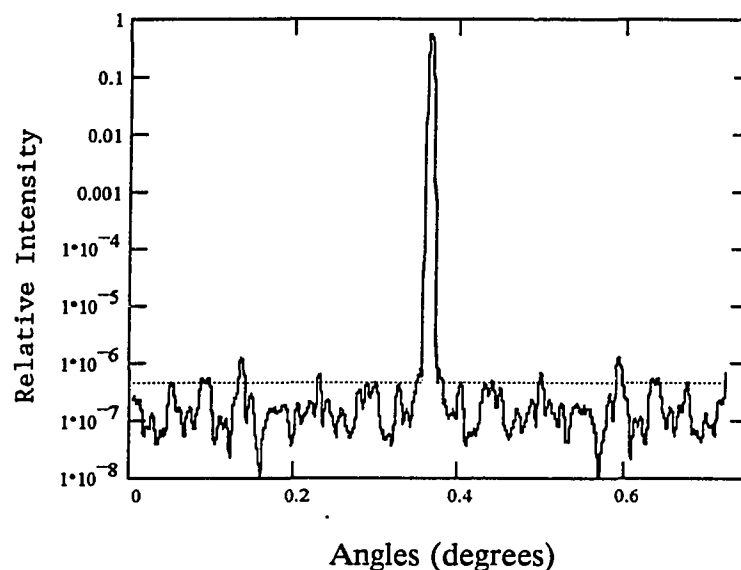
FIGURE 6.21, Relative intensity, Dammann grating (DG3). Line edge error δ = 0.1 μm. Compared to 1-D roughness scattering formula, equation (6.4, ...).
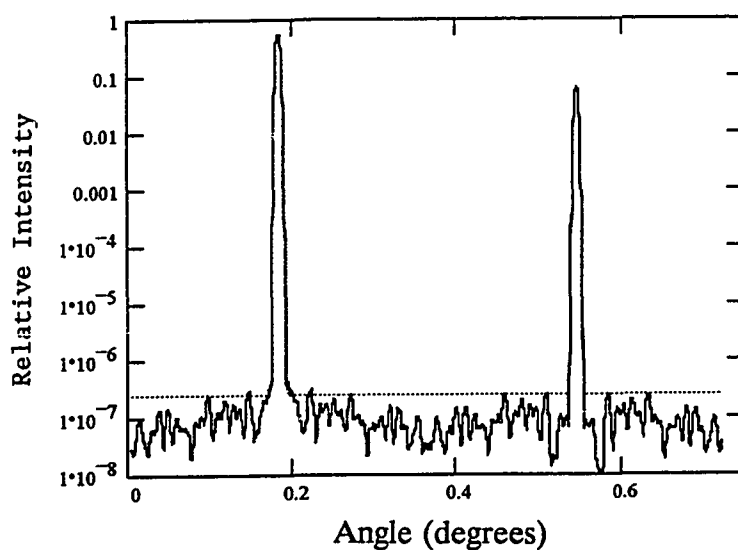
FIGURE 6.22, Relative intensity, Dammann grating (DG4). Line edge error δ = 0.1 μm. Compared to 1-D roughness scattering formula, equation (6.4, ...).
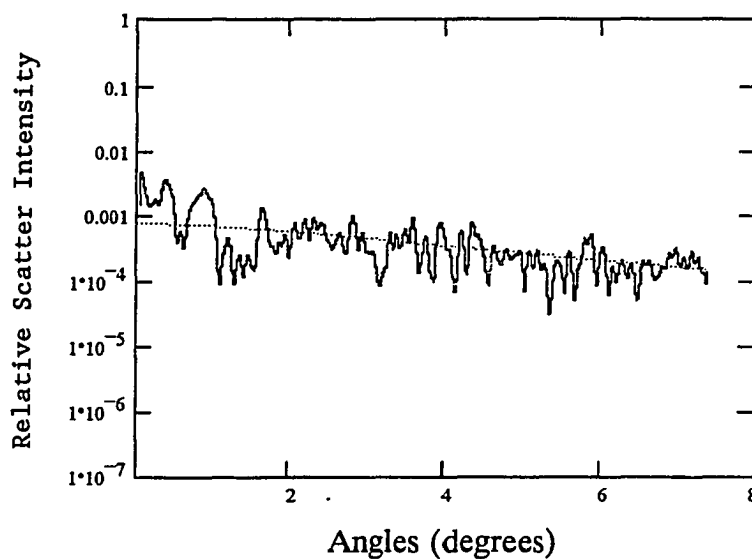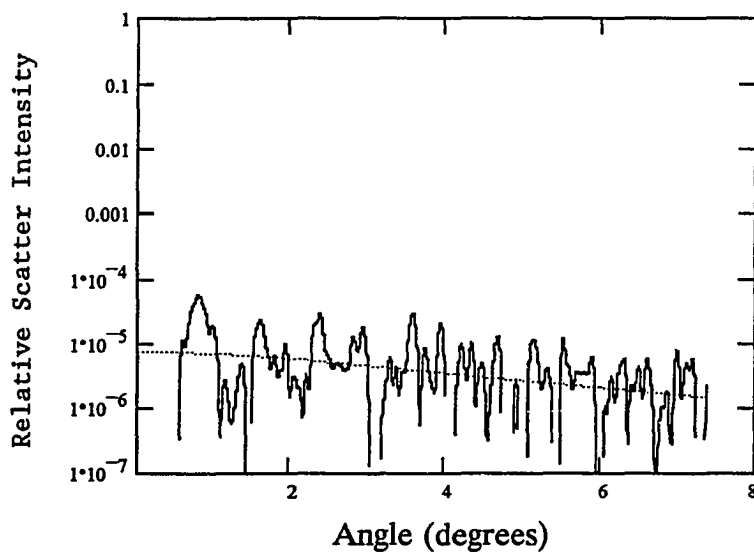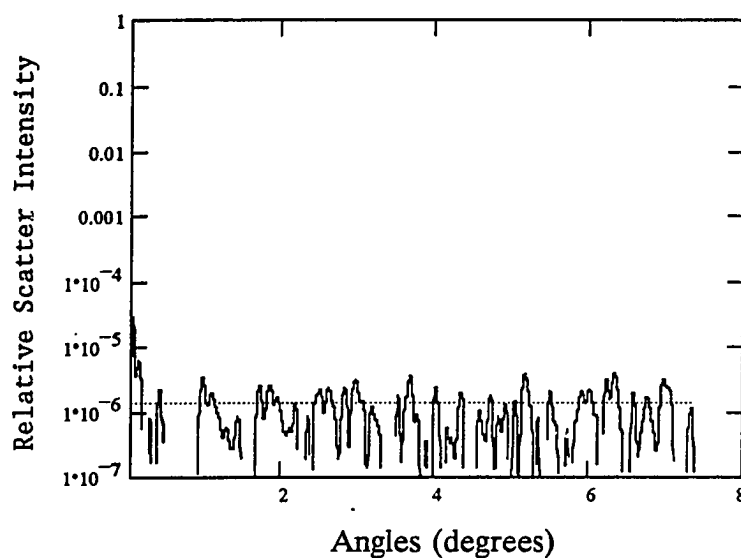
FIGURE 6.23, Relative scatter intensity, cylindrical Fresnel zone plate (FZP1). Line edge error δ = 0.1 μm. Average of 16. Compared to 1-D roughness scattering formula, equation (6.10, ···).
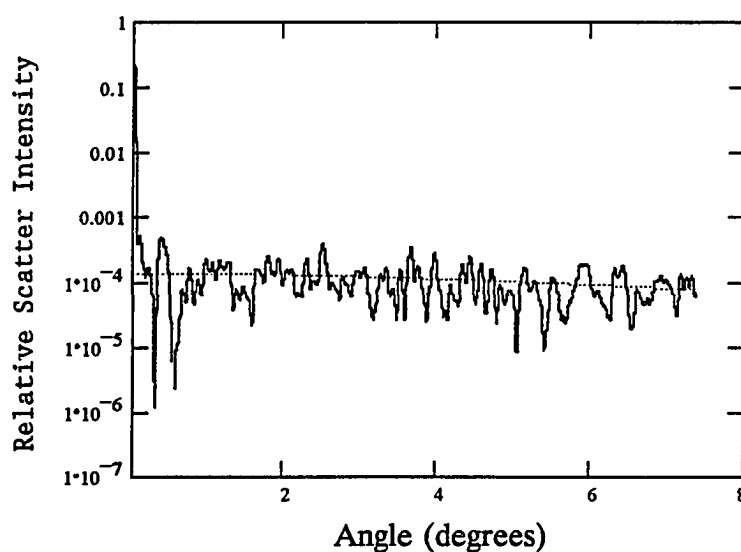
FIGURE 6.24, Relative scatter intensity, cylindrical Fresnel zone plate (FZP1). Line edge error δ = 1 μm. Average of 4. Compared to 1-D roughness scattering formula, equation (6.10, ···).
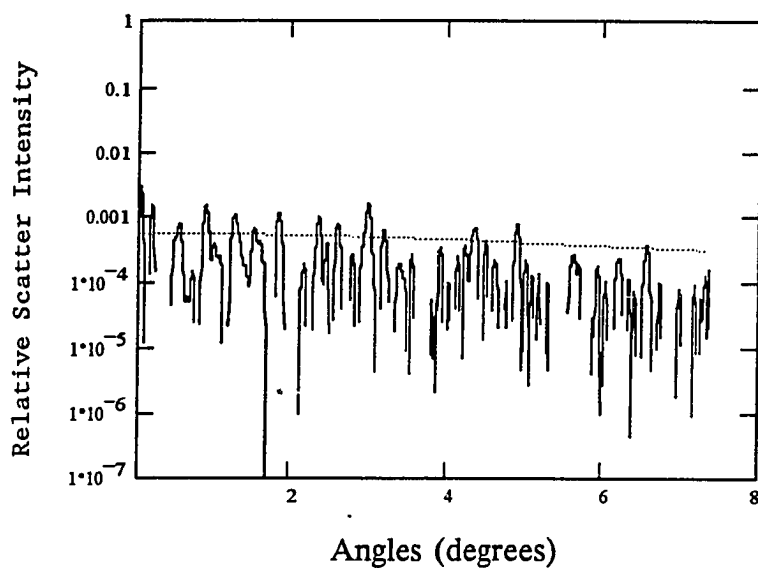
FIGURE 6.25, Relative scatter intensity, cylindrical Fresnel zone plate (FZP2). Line edge error $\delta = 1\,\mu\text{m}$. Average of 4. Compared to 1-D roughness scattering formula, equation (6.10, ⋯).
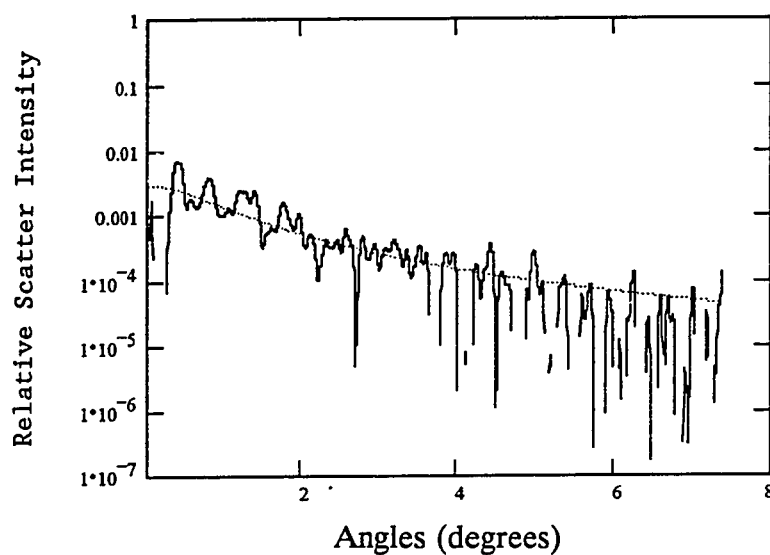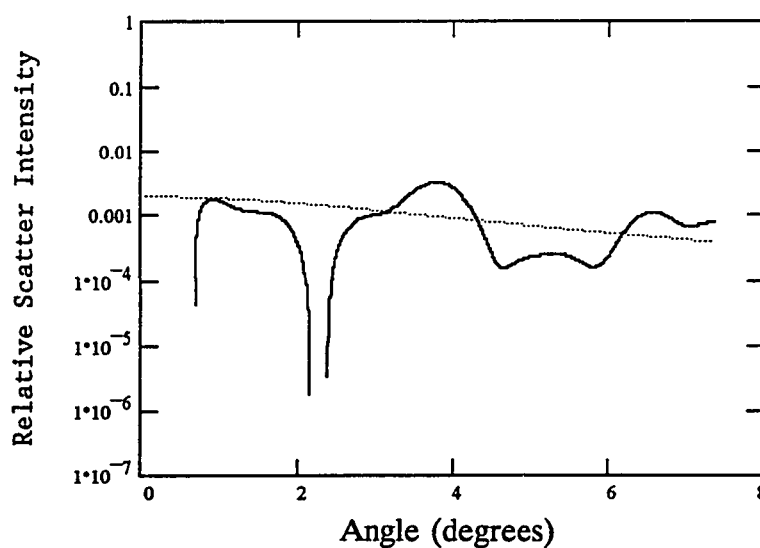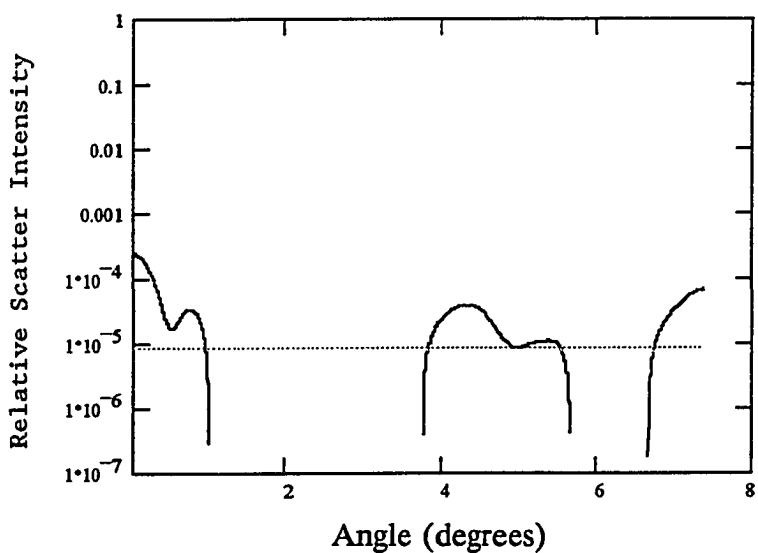

FIGURE 6.26, Relative scatter intensity, cylindrical Fresnel zone plate (FZP4). Line edge error $\delta = 1\,\mu\text{m}$. Average of 20. Compared to 1-D roughness scattering formula, equation (6.10, ⋯).
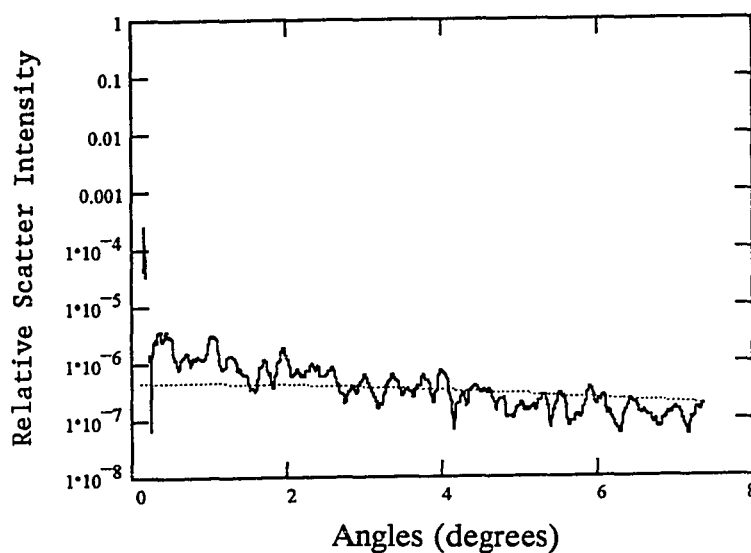
FIGURE 6.27, Relative scatter intensity, circular Fresnel zone plate (FZP1). Line edge error $\delta = 1$ μm. Average of 4. Compared to 2-D roughness scattering formula, equation (6.13, ⋯).
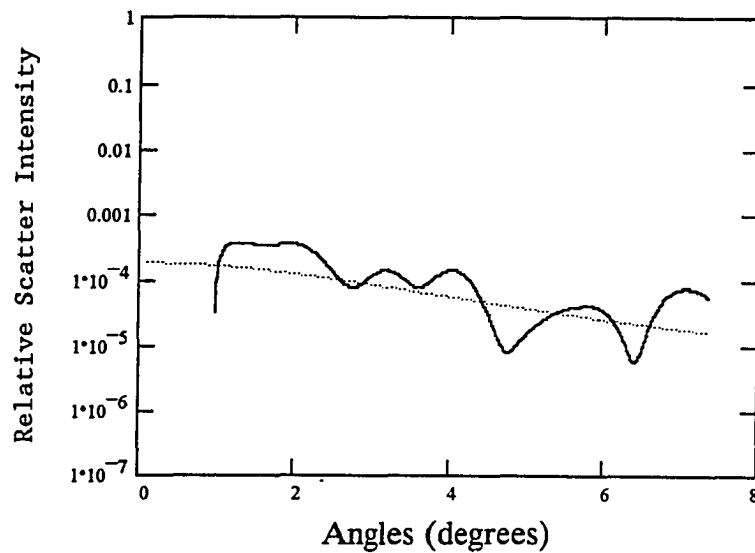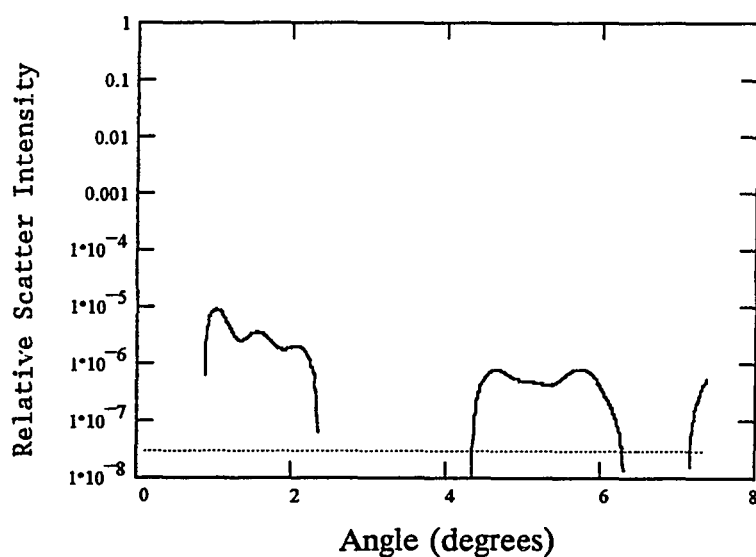

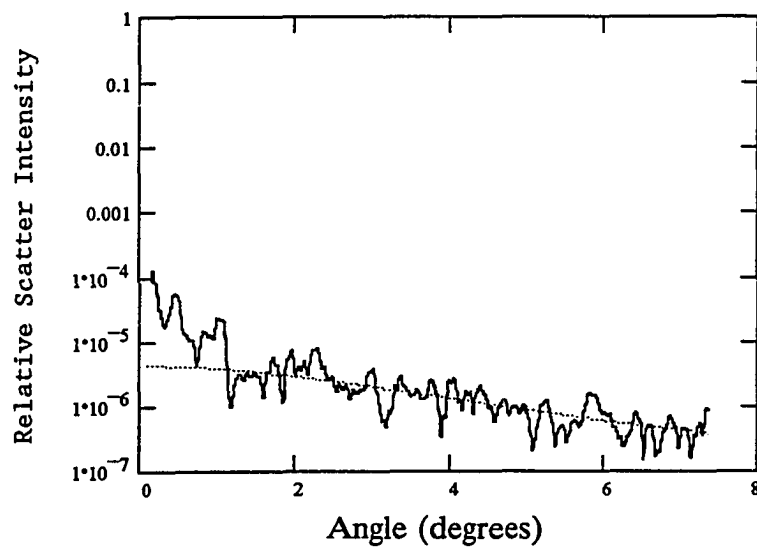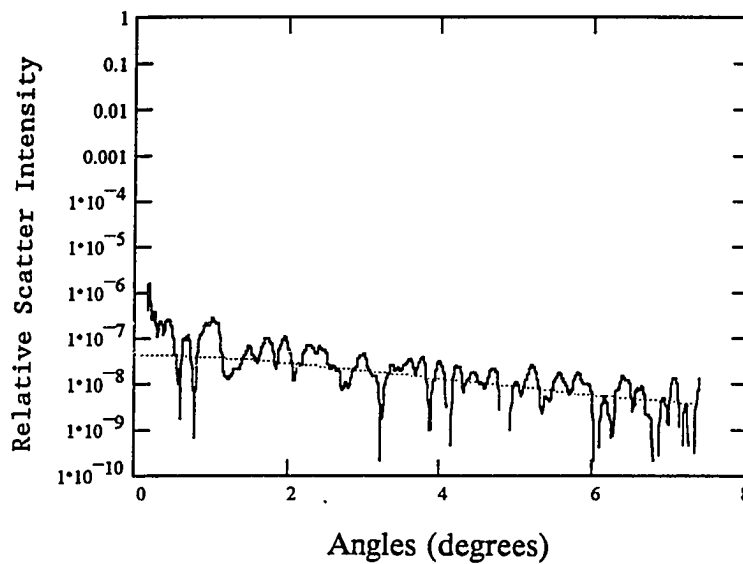FIGURE 6.28, Relative scatter intensity, circular Fresnel zone plate (FZP4). Line edge error $\delta = 1$ μm. Average of 52. Compared to 2-D roughness scattering formula, equation (6.13, ⋯).

7. SURFACE PROFILE AND SCATTERING MEASUREMENTS

In Chapter 5 we derived a new theory of scattering from systematic errors and we presented diffraction efficiency measurements to compare the theory to the experimental evidence. In Chapter 6 we derived a new theory of scattering from random errors and in this chapter we will compare this theory to measurements of angle resolved scattering. First we will discuss surface profile measurements made on two binary optic elements. Both of these elements are of the Dammann grating type; one (G1) was designed for use with a helium-neon laser ($\lambda = 0.6328$ μm), and the other (G2) for a carbon dioxide laser ($\lambda = 10.6$ μm). We will then discuss scattering measurements. A novel scatterometer will be described which is capable of measuring at angles very close to the unscattered portion of the beam. Finally the theory will be compared to the measurements, and then the possibility of determining the surface errors from the scattered light measurements will be discussed.

7.1 Surface Profile Measurements

7.1.1 Measurements of Grating G1

The binary optic for use at $\lambda = 0.6328$ μm , G1, was fabricated on silicon dioxide (SiO_2) using reactive ion etching. The surface profile was measured using two separate instruments; a Talystep and a Wyko two wavelength surface profilometer.

7.1.1.1 Talystep Measurements

The Talystep uses a diamond tipped probe in contact with the surface. The Talystep can measure roughness less than 0.1 nm *rms*, with a horizontal resolution determined by the diameter of the probe (0.6 μm for the measurement reported here). Measurements using the Talystep were made by Dr. Jean Bennett, Naval Air Warfare Center, Weapons Division (NAWCWD), in May 1989. Figure (7.1) shows the binary optic profile using 13,098 points on a 500 μm long trace. The sampling interval was 0.0382 μm. The period of the binary optic is 200 μm, so this trace was two and a half periods long. The instrument noise was determined to be 0.05 nm *rms*. The average etch depth of this binary optic is 655.4 nm according to the Talystep measurement.

Additional Talystep traces show the roughness of the lands (on the surface), and the grooves (bottom of the etched part). These are shown in figures (7.2) and (7.3). The surface is quite smooth, with only 0.35 nm *rms* roughness on the lands and 0.4 nm *rms* roughness in the grooves. The surface profile data was used to calculate the autocovariance functions for the lands, figure (7.4), and the grooves, figure (7.5). The correlation length is 1 μm for the lands and 0.2 μm for the grooves.

7.1.1.2 Wyko Profilometer Measurements

The Wyko profilometer uses phase-shifting interferometry to determine surface heights. Wyko measurements at multiple wavelengths were made by Van Hodgkin of NAWCWD. Making measurements at two separate wavelengths removes

the ambiguity involved with step heights more than $\lambda/4$. Figure (7.6) is a plot of the Wyko data. The etch depth was determined by taking the mean value of one section of the lands and subtracting the mean value of the adjacent etched portion. The mean etch depth is 661.9 nm with a standard deviation of 1.03 nm.

The transition points a_1 and a_2 given by the Wyko profilometer are 43.1 μm (33 pixels) and 86.2 μm (66 pixels), with a period of 199.8 μm (153 pixels). The values are 42.9 μm, 86.5 μm, and 200 μm determined by comparing a photograph of a ruler and the binary optic seen under a Nomarski microscope. The design values are 43.02 μm, 86.96 μm, and 200 μm.

7.1.1.3 Comparison of Talystep to Wyko Measurements

The Wyko measurement of the roughness of the lands and the grooves is somewhat greater than the values given by the Talystep. The mean *rms* of the lands is 0.79 nm, and the mean *rms* of the grooves is 0.66 nm. These averages ignore one or two pixels near the edges where the profile is changing rapidly. The Talystep measurements were taken in the flattest 70 μm portion of the lands, and the flattest 30 μm of the grooves. If we do the same thing with the Wyko data we obtain 0.43 nm as the mean *rms* for the lands and 0.43 nm as the mean *rms* for the grooves, values which are much closer to the Talystep values.

7.1.2 Measurements of Grating G2

The second binary optic, G2, for use at 10.6 μm, was fabricated on silicon (Si). The surface profile was measured by Van Hodgkin using the Wyko profilometer.

Figure (7.7) illustrates typical data for this binary optic. The mean etch depth (subtracting the mean of the groove from the mean of an adjacent land) is 1.23 μm with a standard deviation of 1.8 nm. There are numerous spikes in the data where transitions occur. Ignoring these spikes we find that the mean *rms* roughness of the lands is 1.05 nm, and the mean *rms* "roughness" of the grooves is 27.6 nm. The etch depth of the grooves is not constant, but changes gradually. Each groove has essentially the same profile, so the result is a systematic etch depth error rather than a true random roughness.

Resolution of Wyko Profilometer

The pixel spacing in the Wyko measurements is 1.306 μm which significantly limits the lateral resolution. The lateral resolution could be increased by using a higher magnification in the imaging system. There would be fewer transition points visible, which would reduce the statistical significance of measured line edge errors. The electronic noise limits the etch depth measurement accuracy to 0.1 nm *rms* at best. Measurements of a flat reference plate at this magnification show an error of about 0.4 to 0.6 nm *rms* across the surface, with most of the error consisting of a gradual or smooth variation of low surface spatial frequency. Hence the *rms* measurement error in the difference between two adjacent small portions of the surface should be closer to the electronic noise limit.

7.2 Scattering Measurements

Scattering measurements at 0.6328 μm were made in the author's lab; the 10.6 μm measurements were made by John Stover of Toomay, Mathis & Associates (TMA) on the CASI Scatterometer. We shall first describe the scatterometer in the author's lab.

7.2.1 Novel Small Angle Scatterometer

The author designed the scatterometer in his lab to measure small angle scattering, i.e. scattering near the specular or on-axis beam. The difficulty with measuring small angle scatter is the presence of the unscattered beam. Also scattering from all sources within the instrument contributes to the small-angle scattering.

7.2.1.1 Control of Diffraction and Aberrations

To minimize the effects of the unscattered beam, we use a laser with a good gaussian beam profile. In order to maintain a good gaussian beam through a system the beam must be truncated as little as possible. This requires optical elements which are at least four times larger in diameter than the beam (at the $1/e^2$ points of intensity). A uniform beam, or a truncated gaussian beam, will diffract energy into broad wings which will mask the small angle scattered light from the sample.

It is important to control aberrations. Spherical aberration is controlled by large f-number optics. The odd-order aberrations, such as coma, can be cancelled by symmetry. Figure (7.8) is a diagram of the scatterometer. Each mirror has a

focal length of 71.6 inches. When the angle of incidence on the second mirror is equal in magnitude, but opposite in sign, to the angle of incidence on the first mirror, then the odd-order aberrations will cancel. This still leaves astigmatism, but astigmatism does not limit resolution when the detector pinhole is scanned in the same plane as the plane of incidence on the mirrors.

The sample is placed between the two mirrors so that the incident beam may be collimated. Lack of collimation has a surprisingly large effect on scattering measurements, particularly at large scattering angles. It is a source of error not usually accounted for in typical scatterometers.

7.2.1.2 Control of Stray Light

Instrument scattering is controlled by using mirrors whenever possible. Reflective optics scatter much less than refractive optics. Where refractive optics are required, it is better to use an uncoated single lens than a coated lens or a doublet with their inherent multiple reflections.

The scatterometer also uses apertures to block stray light. An aperture is placed before the lens to block extraneous light from the laser. Another aperture is placed in the focal plane of the lens. It was found experimentally that replacing the typical shiny pinholes (at the focus of the lens and at the detector) with dull black pinholes significantly reduces instrument scattering. Another aperture is placed at the binary optic. Each aperture and the mirrors are at least four times larger than the beam diameter to reduce diffraction. This is much larger than the typical spatial

filter pinhole size. A narrow band transmission filter centered at 633 nm was placed between the detector pinhole and the detector. The purpose of this filter is to block out ambient light. Since it is placed after the detector pinhole, it can scatter light in the forward direction without ill effect.

7.2.1.3 Air Currents

Laboratory air currents can cause very noticeable beam deflections. The solution was to encase the entire apparatus in plexiglass. The laser was just outside of the enclosure so that the heat generated by the laser would not create air turbulence. A small porthole allows the beam to enter the plexiglass enclosure. Originally an air cooled photomultiplier was used for the detector, but the air disturbance of the cooling fan made steady, accurate measurements impossible. Another benefit of the enclosure is to greatly reduce the settling of dust on the sample and optical elements. Dust can significantly increase scattering.

7.2.1.4 Resolution

The resolution of the instrument is determined by the detector pinhole and the focal length of the second mirror. The size of the pinhole is a trade-off; a large pinhole will increase the signal strength and average over unimportant intensity fluctuations, but the resolution will be reduced. The detector pinhole is designed to move with the detector to reduce the effects of differences in sensitivity from point to point on the detector. The detector is also placed sufficiently far behind the pinhole to cause diffraction to spread the signal over a large portion of the detector,

thus averaging the nonuniformity of the detector face. Repeating the measurements with various size pinholes can validate the linearity of the detector.

7.2.1.5 Measurement Procedure

In making the scattering measurements, the 15 mW Spectra-Physics Stabilite model 124A HeNe laser is turned on and allowed to achieve a thermal equilibrium for an hour or two. It was found that it is not necessary to continually monitor the laser output after this period of stabilization. The output varies very little. The typical scatterometer uses a beam splitter or beam chopper to constantly measure the laser output and this introduces instrument scatter and lowers the signal to noise ratio. Like the laser, the silicon detector with pico-watt digital power meter model 835 by Newport Corporation is very stable and repeatable after warming up for an hour or two.

The measurement of the instrument scatter is made by moving the detector pinhole with detector through the focussed spot when no sample is present in the beam. A linear translation stage is used because it is relatively inexpensive and very accurate. Theoretically the stage should be a section of a circle centered on the second mirror. In practice the difference is insignificant. The stage is only moved about an inch, while the focal length of the mirror f = 71.6 inches. The difference between a circle and the tangent line is at most

$$y = f - (f^2 - x^2)^{1/2}$$
$$= 71.6 - (71.6^2 - 1^2)^{1/2}$$

= 0.007 inches.

Experimentally it is found that the detector can be moved closer or further away from the mirror by an amount many times 0.007 inches and not change the detector reading a noticeable amount.

A measurement of the scattering is just a repeat of the instrument scattering measurement except this time the sample is in place. Front and back surface reflections are readily seen and the sample is tilted just enough to ensure that these reflections miss the pinhole at the focus of the lens. Multiple reflections can give false scattering measurements.

7.2.2 Comparison of Measurements to Theory

7.2.2.1 Comparison for Grating G1

We are now ready to present the results of scattering measurements of the Dammann grating G1 designed for 0.6328 μm, and to compare these measurements to calculations of scattering based on surface measurements. Figure (7.9) compares measured data, the normalized instrument scatter, and the 2-D roughness scattering formula (6.1). The instrument signature is the amount of scattering present from the scatterometer alone. The relative instrument scattering is about a factor of ten smaller than the sample scattering. In this figure only the random roughness with σ = 0.4 nm and τ = 1 μm is calculated. The theoretical scattering level for a binary optic with no processing errors, and with only the measured surface roughness values, is far below the measured scattering level.

In figure (7.10) the calculated scattering is based on the measured Wyko data for random etch depth error (δ = 1 nm). This small amount of error gives fairly good agreement with the measured scattering levels. We can not calculate the scattering from random line edge errors because the error was smaller than the resolution of the measurement equipment. Using random line edge error as a free parameter we can obtain somewhat better agreement of theory to experiment. In figure (7.11) we calculate scattering using the measured random etch depth error and assuming a reasonable line edge error of δ = 25 nm.

The calculation of the scattering shown in figure (7.11) fits the measured data well, though there is some scattering less than a few hundredths of a degree from each diffraction order that is not predicted by the roughness formula. Also the formula somewhat under predicts near the zero order, but over predicts near the fourth diffraction order. Since "grass" increases as the square of the diffraction order [equation (3.28)], this scattering could not be "grass."

The random line edge error formula we have derived (6.4) does *not* depend on the diffraction order, though it does depend on the diffraction efficiency. Our formula is more like the result of the classic grating "accidental error of amplitude", which is also independent of grating order. Evidently random line edge errors and random etch depth errors cause scattering by random changes in the "form" of the binary optic pattern, rather than the period of the grating. The measured scattering seems to confirm this hypothesis.

The sample was etched to a depth of 662 nm instead of the design value 554 nm. We expect this systematic error to reduce the intensity in the zero order by a factor of about ten. The scattering measurement agrees with this calculation.

7.2.2.2 Comparison for Grating G2

The scattering of the second optic G2, designed for 10.6 μm radiation, was measured at TMA using a CO_2 laser. This data was provided to the author by the courtesy of Van Hodgkin, NAWCWD. The TMA instrument was designed by John Stover, who has written extensively on the subject of scattering.[1] Figure (7.12) shows the BTDF from this binary optic. Also shown is a plot of the 1-D roughness scattering formula with an etch depth error 1.8 nm. The roughness formula has been scaled to match the peak at the zero order. The etch depth error does not account for the scattering at large angles.

Using the line edge error again as a free parameter we obtain a much better fit to the data, especially at large angles. Figure (7.13) shows the calculated scattering using the measured random etch depth error of 1.8 nm, and assuming a reasonable line edge error of 30 nm. Again there is some scattering near the diffraction orders that is not predicted by the roughness formula. The scattering is roughly constant a few degrees from the diffraction order indicating once again a random change in "form", and not "grass" or "ghosts".

There are systematic etch depth errors for G2. The groove bottoms do not have a constant value. We expect from Chapter 5 to see a redistribution of energy among the diffraction orders, and this is what we find.

7.3 Summary

We have seen that systematic errors redistribute energy among the diffraction orders. Random errors contribute to the measured scattering in between the diffraction orders. Scattering from random surface roughness and random etch depth errors tends to fall off with angle and does not account for all of the scattering measured. The addition of assumed random line edge errors gives a good agreement with the measured scattering, except very near each diffraction order.

The average random line edge error necessary to give the measured results is very small, about 25 or 30 nm *rms*. We cannot measure such small perturbations of the edge positions over a sufficiently long trace to give statistically meaningful results. A new, long trace, high accuracy instrument is needed to make the final verification of the theory of Chapter 6. It appears at this time that the theory gives reasonable results, and the theory compares well with computer simulations.

We should be able to determine the processing errors by a measurement of the scattering instead of the surface profile. Line edge errors may be distinguished from etch depth errors by the angle dependence of the scattering. Systematic errors will change the diffraction efficiencies, but not the interorder scattering.
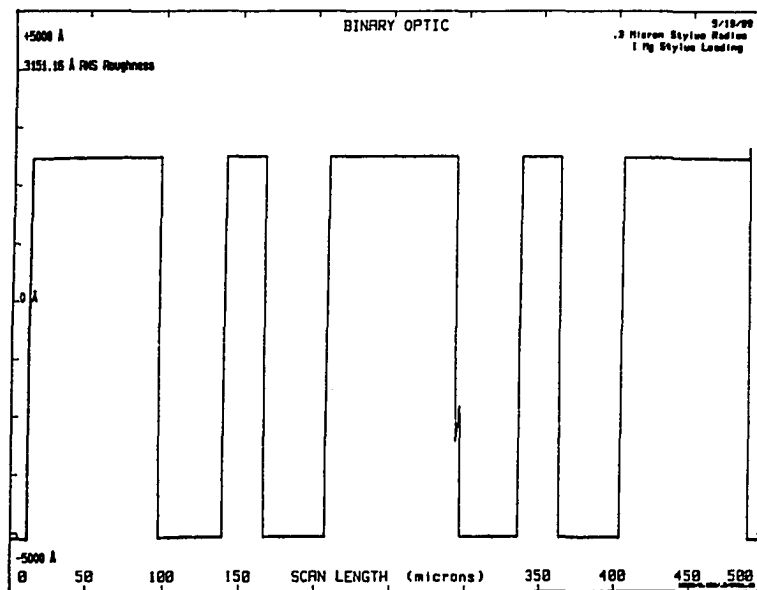
FIGURE 7.1, Profile of Dammann grating G1.
Talystep measurement.
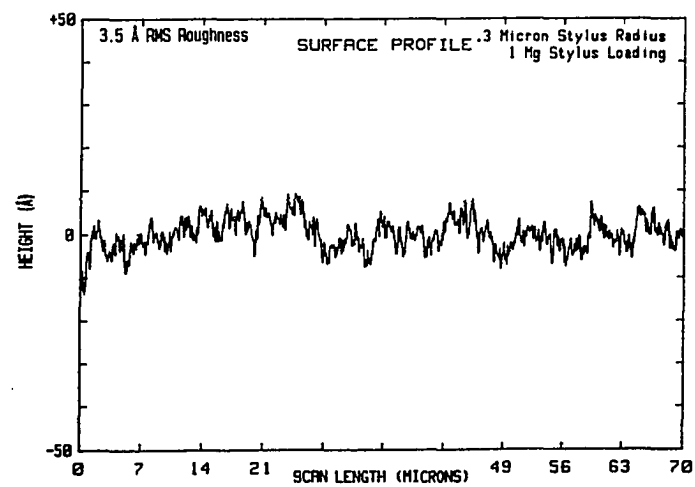
FIGURE 7.2, Surface roughness on the land (surface)
of a Dammann grating G1. Measurement by Talystep.
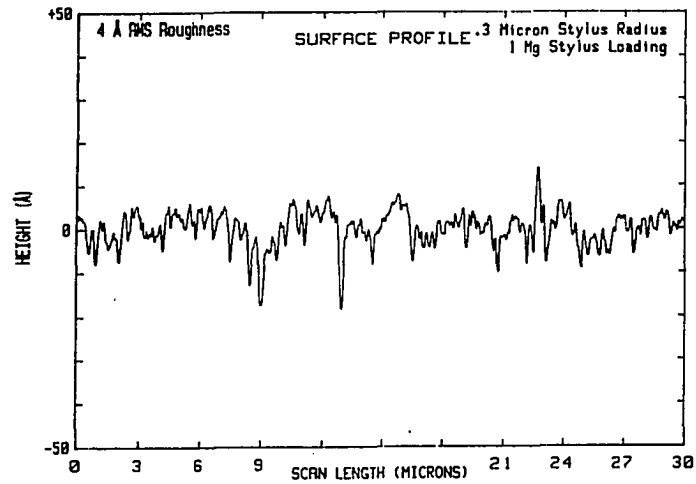
FIGURE 7.3, Surface roughness in the groove of a Dammann grating G1. Measurement by Talystep.
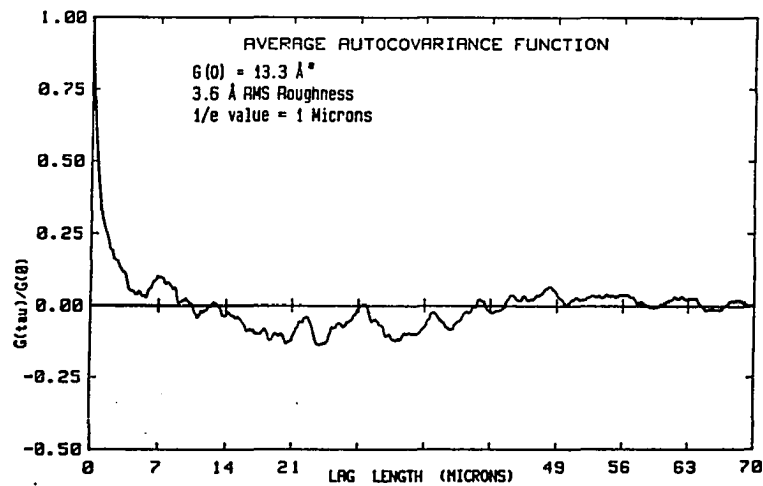
FIGURE 7.4, Autocovariance function for land of Dammann grating G1. Calculated from Talystep measurement.

```
1.00
                    AVERAGE AUTOCOVARIANCE FUNCTION
         G(0) = 16.1 Å*
0.75     3.9 Å RMS Roughness
         1/e value = .2 Microns
0.50

0.25
G(tau)/G(0)
0.00

-0.25

-0.50
    0    3    6    9   LAG LENGTH (MICRONS)  21   24   27   30
```
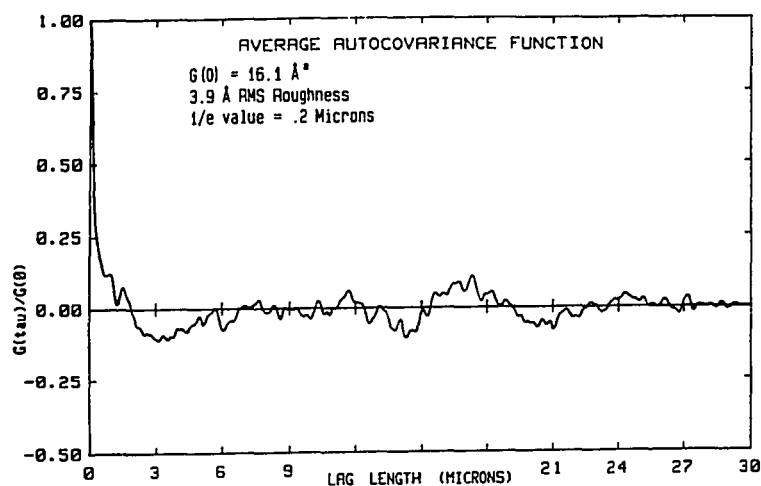
FIGURE 7.5, Autocovariance function for groove of Dammann grating G1. Calculated from Talystep measurement.


```
657/603          Time: 10:31     Date:5/20/91
RMS:  322nm                        PV:  701nm
PA:   315nm        PROFILE

        400    WL = 657.6nm



Nanometers  200



         0



       -200



       -400
         0.00   0.26   0.52   0.78   1.04   1.30

        Distance on Surface in Millimeters (10.1X)

           DOUG RICKS GRATING MODEL A-2-1 (WAVE CORRECTION)
                                              WYKO
```
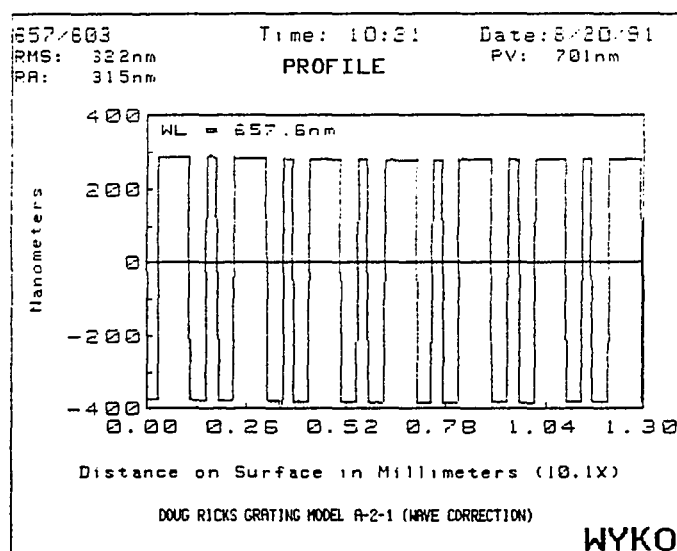
FIGURE 7.6, Profile of Dammann grating G1, Wyko profilometer measurement.

```
1 LVL SI GRTNG     Time: 14:48      Date:12/6/91
PMS:  629nm                         PV:  3046nm
RA:   612nm          PROFILE        RC:  13.1 m
```
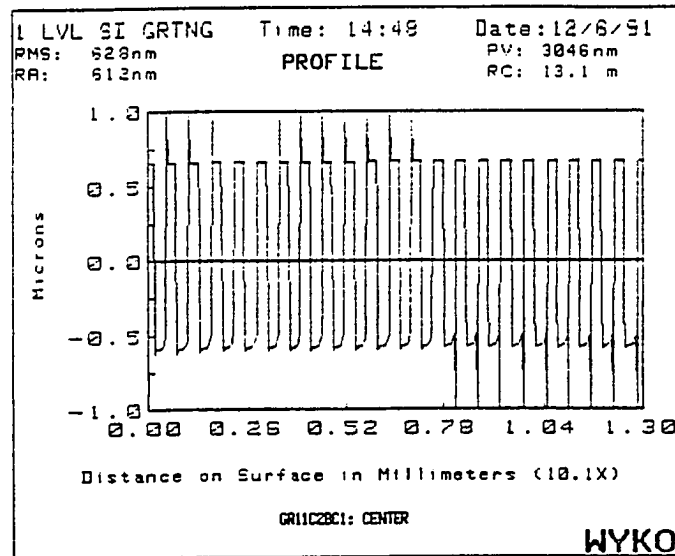
FIGURE 7.7, Profile of Dammann grating G2. Wyko profilometer measurement.
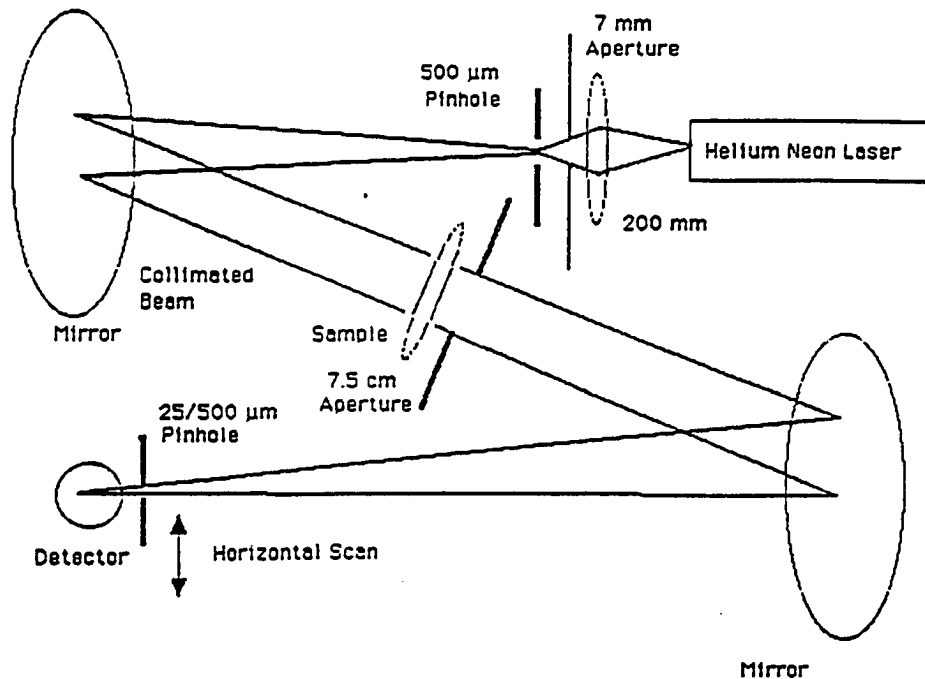
FIGURE 7.8, Diagram of scatterometer used to measure scatter at 0.6328 μm. Reference (6.1).
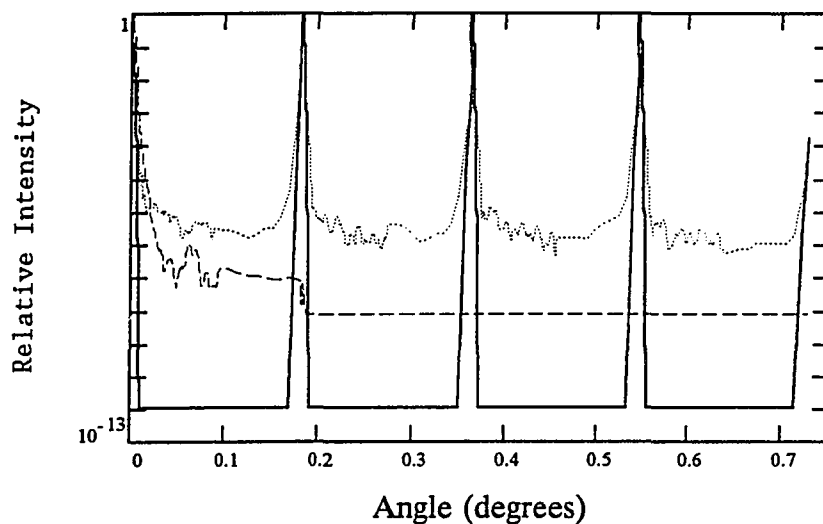
FIGURE 7.9, Relative intensity, Dammann grating G1 at 0.6328 μm. Measured scattering ($\cdots$) compared to normalized instrument signature (- - -), and 2-D scattering formula with measured roughness δ = 0.4 nm and τ = 1 μm (---).
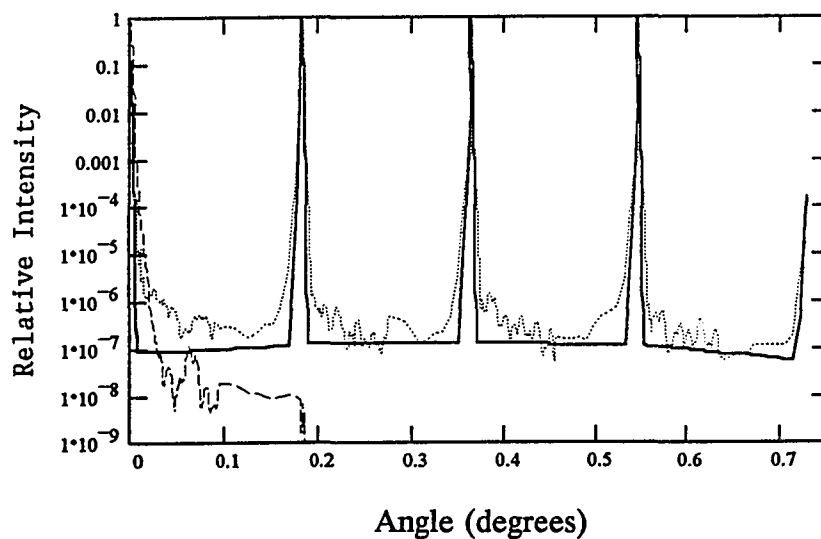
FIGURE 7.10, Relative intensity, Dammann grating G1 at 0.6328 μm. Measured scattering ($\cdots$) compared to normalized instrument signature (- - -), and 1-D scattering formula with measured etch depth error δ = 1 nm (---).
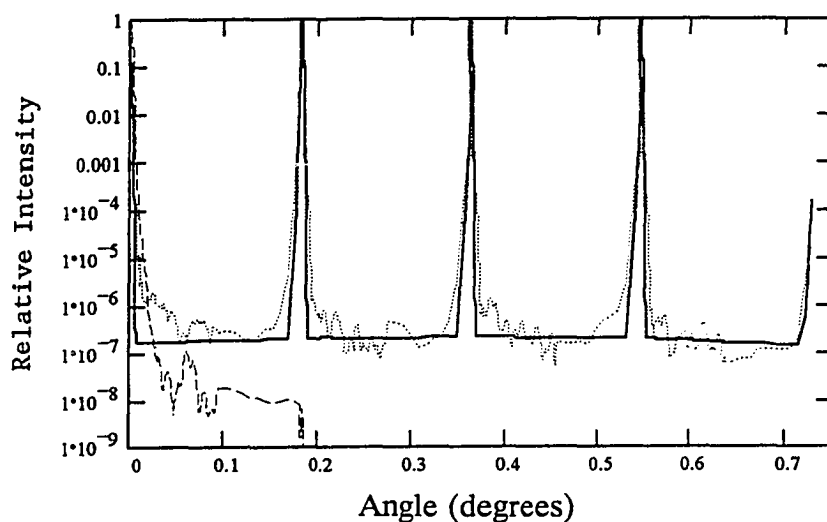
FIGURE 7.11, Relative intensity, Dammann grating G1 at 0.6328 μm. Measured scattering ($\cdots$) compared to normalized instrument signature (- - -), and 1-D scattering formula with measured etch depth error δ = 1 nm, and assumed line edge error δ = 25 nm (---).
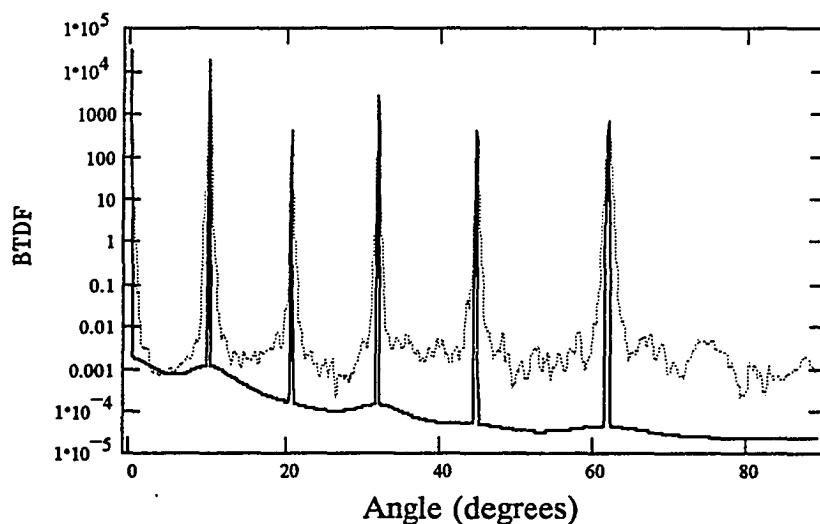
FIGURE 7.12, BTDF, Dammann grating G2 at 10.6 μm. Measured scattering ($\cdots$) compared to 1-D scattering formula with measured etch depth error δ = 1.8 nm (---).
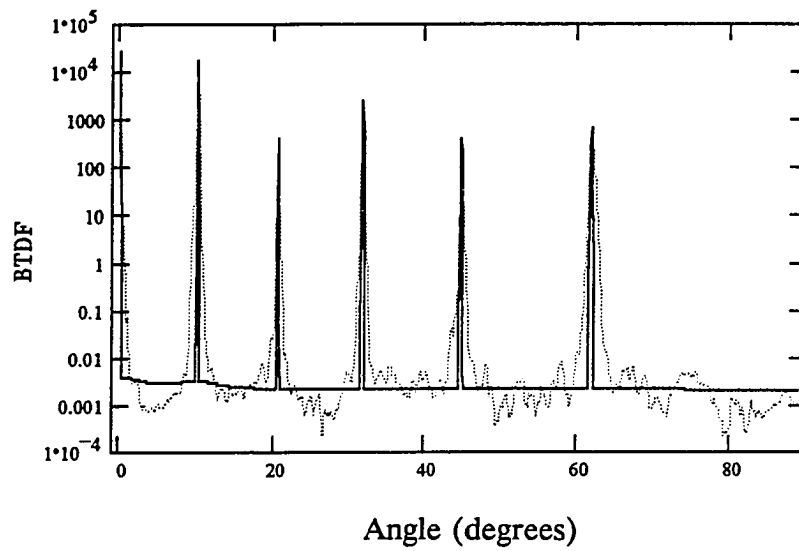
FIGURE 7.13, BTDF, Dammann grating G2 at 10.6 μm. Measured scattering ($\cdots$) compared to 1-D scattering formula with measured etch depth error δ = 1.8 nm and assumed line edge error 30 nm (---).

8. CONCLUSIONS

The diffraction grating was invented over a hundred and fifty years ago. The first zone plates were made more than one hundred years ago. Until recently these early "diffractive optics" were restricted by existing manufacturing technology to a few special applications. Recent advances in microlithographic technology have now made it possible to design and build a new class of optical elements, binary optics. This new class utilizes diffraction, as does the grating and the zone plate, but binary optics are much more efficient and versatile. In the next few years we can expect to see many diverse applications of binary optic technology.

How is light scattered from a binary optic? To what extent do the various manufacturing steps need to be controlled to achieve a specific level of scattering? Previous scattering formulas for slight roughness on flat surfaces are not adequate for binary optics. Formulas derived for diffraction gratings give a measure of intensity for "ghosts" and "grass", but the new design freedom and versatility of binary optics technology make possible many more sources of systematic and random error. What is needed is a fresh look at the calculation of scattering from binary optics.

In principle the electromagnetic vector theory of light could be used to calculate scattering from specific binary optics. There exist very few exact, analytical solutions to vector theory problems. In practice vector theory is very computer intensive, and is generally limited to strictly periodic (about one wavelength), one-dimensional surfaces of infinite extent. These computations give very little physical

insight into what is actually happening. The results also do not account for scattering between diffraction orders, which comes from random errors.

Scalar diffraction theory is a useful approximation to vector theory and can be applied to problems that would be impractical to solve using the vector theory. We have seen that in many cases the scalar theory is sufficiently accurate, and the limits of validity of the Rayleigh-Sommerfeld, Fresnel, and Fraunhofer scalar diffraction theories have been quantified. A new extended Fourier optics theory has been derived which retains the accuracy of the Rayleigh-Sommerfeld theory for off-axis observation points. This new theory is in better agreement with vector theory than the old Fourier optics theory.

The calculation of scattering from binary optics using scalar diffraction integrals can be time consuming and fails to clarify the dependence of scattering on such factors as the wavelength. We have derived new analytical formulas for diffraction efficiency and scattering which can quickly and easily be evaluated using a personal computer. In addition, a study of the formulas shows the dependence of scattering on wavelength, standard deviation of error, number of levels, step widths and heights, refractive index, number of edges per period, diffraction efficiency, f-number, period length, probability density function, mask alignment errors, etch depth errors, and line edge errors.

To determine the accuracy of these formulas, we have performed numerous computer simulations. In addition, experimental results have been compared to the formulas. A new scatterometer was designed and built to measure scattering from

binary optics at very small angles. We find that improvements in the formulas could be made in the case of random line edge errors in binary optics with large step heights, for quantization errors, and for scattering at small angles. Better surface measuring equipment is needed to confirm experimentally the formulas for random line edge errors. In general, however, the derived formulas have been shown to be reasonably valid under a wide variety of conditions. It is possible in many cases to determine the cause of scattering by examining the angular dependence of the light scattering from binary optics.

APPENDIX A:

FFT CALCULATION OF CYLINDRICAL ZONE PLATES

The diffraction formulas for cylindrical Fresnel zone plates can be quickly solved using the Fast Fourier Transform (FFT) algorithm. We begin with the Fresnel formula, ignoring the y-axis and some phase terms that do not contribute to the intensity,

$$U(x_0, z) = 1/z\lambda \int_{-\infty}^{\infty} U(x_1) \exp(ikx_1^2/2z - x_1^2/b^2) \exp(-ikx_1x_0/z) \, dx_1 \qquad (1)$$

where

$$U(x_1) = \Sigma_j \, \phi_j \, \text{rect}\{[x_1 + \tfrac{1}{2}(a_{j+1} + a_j)]/a_{j+1} - a_j]\}$$

$$+ \Sigma_j \, \phi_j \, \text{rect}\{[x_1 - \tfrac{1}{2}(a_{j+1} + a_j)]/[a_{j+1} - a_j]\}, \qquad (2)$$

and where ϕ_j and a_j are the phase terms and transition points, respectively. We recognize that the integral of equation (1) is the Fourier transform of the product of two functions, $U(x_1)$ and the exponential function.

We could use Fast Fourier Transforms at this point to solve equation (1), but we would have to sample a great many points in the aperture plane to accurately model the small perturbations of the binary optic pattern which cause scattering. In addition, a discrete Fourier transform is limited in resolution in the transform plane; the smallest increment is $\lambda f^{\#}$. To avoid these problems we take a more roundabout method using the well known law that the Fourier transform of a product is given by the convolution of the individual transforms, and the Fourier transform of a

convolution is the product of the individual transforms.[1] We determine analytically the Fourier transforms of $U(x_1)$ and the exponential term. We then select the spacing we want in the transform plane and use a computer to take the FFT of each Fourier transform. The inverse FFT of the product of these two FFTs gives us the desired field amplitude in the transform plane.

The Fourier transform of $U(x_1)$ is given by

$$F\{U(x_1)\}|_{\xi = x0/\lambda z} = \Sigma_j \phi_j (a_{j+1} - a_j) \exp[i\pi(a_{j+1} + a_j)\xi] \text{sinc}[(a_{j+1} - a_j)\xi]$$

$$+ \Sigma_j \phi_j (a_{j+1} - a_j) \exp[-i\pi(a_{j+1} + a_j)\xi] \text{sinc}[(a_{j+1} - a_j)\xi]|_{\xi = x0/\lambda z}$$

$$= \Sigma_j \phi_j 2\cos[\pi(a_{j+1} + a_j)\xi] \sin[\pi(a_{j+1} - a_j)\xi]/\pi\xi|_{\xi = x0/\lambda z}$$

$$= \Sigma_j \phi_j [\sin(2\pi a_{j+1}\xi) - \sin(2\pi a_j\xi)]/\pi\xi|_{\xi = x0/\lambda z}$$

$$= \lambda z/\pi x_0 \Sigma_j (\phi_j - \phi_{j+1}) \sin(x_0 k a_{j+1}/z). \tag{3}$$

The Fourier transform of the exponential term in equation (1) is[1]

$$F\{\exp(ikx_1^2/2z - x_1^2/b^2)\}|_{\xi = x0/\lambda z}$$

$$= (1/\pi b^2 - i/\lambda z)^{-\frac{1}{2}} \exp[-\pi(x_0/\lambda z)^2/(1/\pi b^2 - i/\lambda z)] . \tag{4}$$

With these analytic expressions of the Fourier transforms we can convolute them using FFTs of each and then take the inverse FFT of the product to get our result.

APPENDIX B:

DERIVATION OF THE ROUGHNESS FORMULA

We consider the problem of a binary optic with slightly rough surfaces.[1] We treat the binary optic as a grating and use the Fraunhofer diffraction formula. The illumination of the grating will be by a collimated laser beam of wavelength λ, and a gaussian field amplitude $G(x_1, y_1)$ with beam radius b. The beam waist is at the grating. The field in the aperture is given by

$$U(x_1, y_1) = G(x_1, y_1) \, \Sigma_{m=-\infty} A_m \, \exp(2\pi i m x_1/a) \tag{1}$$

where a is the length of a single period. The roughness function, or the surface height variation, is $h(x_1, y_1)$. Using the Fraunhofer diffraction formula, equation (2.12), we can write the field at the point (x_0, y_0, z) by

$$U(x_0, y_0, z) = C \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Sigma_{m=-\infty} A_m \, \exp(2\pi i m x_1/a) \, \exp[k i h(x_1, y_1)(n-1)]$$

$$\exp[-ik(x_0 x_1 + y_0 y_1)/z] \, G(x_1, y_1) \, dx_1 \, dy_1 \tag{2}$$

where n is the index of refraction of the grating material, and C is equal to the terms in front of the integral of equation (2.12).

The intensity is given by multiplying by the complex conjugate, which we can write as

$$I(x_0, y_0, z) = CC^* \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Sigma_m \Sigma_j A_m A_j^* \, \exp[2\pi i(m x_1 - j x_2)/a]$$

$$\exp\{ki[h(x_1, y_1) - h(x_2, y_2)](n-1)\} \, \exp\{-ik[x_0(x_1 - x_2) + y_0(y_1 - y_2)]/z\}$$

$$G(x_1, y_1) \, G(x_2, y_2) \, dx_1 \, dy_1 \, dx_2 \, dy_2. \tag{3}$$

We now make a change of variables,

$$\delta_x = x_1 - x_2$$

$$\delta_y = y_1 - y_2$$

Then we can write equation (2) as

$$I(x_0, y_0, z) = CC^* \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Sigma_m \Sigma_j A_m A_j^* \exp[2\pi i(m - j)x_2/a]\, G(x_2, y_2)\, dx_2\, dy_2$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{ki[h(x_2 + \delta_x, y_2 + \delta_y) - h(x_2, y_2)](n - 1)\} \exp[2\pi im\delta_x/a]$$

$$\exp[-ik(x_0\delta_x + y_0\delta_y)/z]\, G(x_2 + \delta_x, y_2 + \delta_y)\, d\delta_x\, d\delta_y. \tag{4}$$

B.1 Some Assumptions of Surface Roughness

If we next assume that the distribution of surface heights **h** is random with a gaussian probability, then[2]

$$\langle \exp\{ki[h(x_2 + \delta_x, y_2 + \delta_y) - h(x_2, y_2)](n - 1)\}\rangle =$$

$$\exp\{-[k(n - 1)\sigma]^2 [1 - R(\delta_x, \delta_y)]\} \tag{5}$$

where $\langle \rangle$ denotes the average or expected value, σ is the *rms* surface roughness and R() is the autocorrelation function. If we next assume that the surface roughness is small, $k\sigma << 1$, then

$$\exp\{-[k(n - 1)\sigma]^2 [1 - R(\delta_x, \delta_y)]\} \approx 1 - [k(n - 1)\sigma]^2 [1 - R(\delta_x, \delta_y)]$$

$$= 1 - [k(n - 1)\sigma]^2 + [k(n - 1)\sigma]^2 R(\delta_x, \delta_y). \tag{6}$$

At this point we take the average or expected value of both sides of equation (4),

$$<I(x_0, y_0, z)> = CC^* \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Sigma_m \Sigma_j A_m A_j^* \exp[2\pi i(m - j)x_2/a]\, G(x_2, y_2)\, dx_2\, dy_2$$

$$[r(\sigma, x_2, y_2) + u(\sigma, x_2, y_2)] \tag{7}$$

where we have a roughness term $r(\sigma, x_2, y_2)$, and a smooth term $u(\sigma, x_2, y_2)$ given by

$$r(\sigma, x_2, y_2) = [k(n - 1)\sigma]^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(\delta_x, \delta_y) \exp[2\pi i m \delta_x/a]$$

$$\exp[-ik(x_0\delta_x + y_0\delta_y)/z] \, G(x_2 + \delta_x, y_2 + \delta_y) \, d\delta_x \, d\delta_y \qquad (8)$$

and

$$u(\sigma, x_2, y_2) = \{1 - [k(n - 1)\sigma]^2\} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[2\pi i m \delta_x/a]$$

$$\exp[-ik(x_0\delta_x + y_0\delta_y)/z] \, G(x_2 + \delta_x, y_2 + \delta_y) \, d\delta_x \, d\delta_y. \qquad (9)$$

B.2 Intensity of the Rough Component

We now assume that the autocorrelation function is exponential,

$$R(\delta_x, \delta_y) = \exp[-(\delta_x^2 + \delta_y^2)^{\frac{1}{2}}/\tau] \qquad (10)$$

where τ is the correlation length. Now if $G(x_2 + \delta_x, y_2 + \delta_y)$ is slowly varying compared to $R(\delta_x, \delta_y)$, or $\tau < b$, then $G(x_2 + \delta_x, y_2 + \delta_y)$ can be taken out of the integral in equation (8). Equation (8) can now be written as

$$r(\sigma, x_2, y_2) = [k(n - 1)\sigma]^2 \, G(x_2, y_2) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[-(\delta_x^2 + \delta_y^2)^{\frac{1}{2}}/\tau]$$

$$\exp[2\pi i m \delta_x/a] \exp[-ik(x_0\delta_x + y_0\delta_y)/z] \, d\delta_x \, d\delta_y. \qquad (11)$$

Recognizing equation (11) as a zero order Hankel Transform of the autocorrelation function we find

$$r(\sigma, x_2, y_2) = [k(n - 1)\sigma]^2 \, G(x_2, y_2) \, 2\pi \tau^2$$

$$\{1 + [(2\pi m/a - x_0 k/z)^2 + (y_0 k/z)^2]\tau^2\}^{-3/2} \qquad (12)$$

Substituting equation (12) into equation (7) gives the intensity distribution $<I_{rough}(x_0, y_0, z)>$ due to the roughness,

$$<I_{rough}(x_0, y_0, z)> = CC^* \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Sigma_m \, \Sigma_j \, A_m \, A_j^* \exp[2\pi i(m - j)x_2/a] \, G(x_2, y_2)$$

$$[k(n - 1)\sigma]^2 \; G(x_2, y_2) \; 2\pi\tau^2\{1 \; + \; [(2\pi m/a - x_0 k/z)^2 \; + \; (y_0 k/z)^2]\tau^2\}^{-3/2} \; dx_2 \, dy_2 \qquad (13)$$

which can be solved, giving us

$$<I_{rough}(x_0, y_0, z)> \; = \; CC^* \; \Sigma_m \; \Sigma_j \; A_m A_j^* \; \exp\{-[\pi b(m - j)/a]^2/2\} \; b^2\pi$$

$$[k(n - 1)\sigma]^2\pi\tau^2\{1 \; + \; [(2\pi m/a - x_0 k/z)^2 \; + \; (y_0 k/z)^2]\tau^2\}^{-3/2} \qquad (14)$$

The beam radius will be much greater than the grating period, $b >> a$, so

$$\exp\{-[\pi b(m - j)/a]^2/2\} \approx 0 \qquad \text{for } m \neq j.$$

Hence we can write equation (14) as

$$<I_{rough}(x_0, y_0, z)> \; = \; CC^* \; [\pi\sigma\tau b k(n - 1)]^2$$

$$\Sigma_m \; |A_m|^2 \; \{1 \; + \; [(2\pi m/a - x_0 k/z)^2 \; + \; (y_0 k/z)^2]\tau^2\}^{-3/2} . \qquad (15)$$

B.3 Intensity of the Smooth Component

Now for the smooth component, which gives the unscattered light. We can solve equation (9) to give

$$\mathbf{u}(\sigma, x_2, y_2) \; = \; b^2\pi \; \{1 \; - \; [k(n - 1)\sigma]^2\} \; \exp[2\pi i x_2(x_0/\lambda z - m/a) + y_0 y_2 i k/z]$$

$$\exp\{-b^2\pi^2[(x_0/\lambda z - m/a)^2 \; + \; (y_0/\lambda z)^2]\}. \qquad (16)$$

Substituting this into equation (7) and solving for the intensity, $I_{smooth}(x_0, y_0, z)$, of the beam not scattered by the roughness

$$<I_{smooth}(x_0, y_0, z)> \; = \; CC^* \; b^4\pi^2 \; \{1 \; - \; [k(n - 1)\sigma]^2\} \; \Sigma_m \; \Sigma_j \; A_m A_j^*$$

$$\exp\{-b^2\pi^2[(x_0/\lambda z - m/a)^2 \; + \; (x_0/\lambda z - j/a)^2 + 2(y_0/\lambda z)^2]\}. \qquad (17)$$

Equation (17) is approximately zero when $m \neq j$, so equation (17) becomes

$$<I_{smooth}(x_0, y_0, z)> \; = \; CC^* \; b^4\pi^2 \; \{1 \; - \; [k(n - 1)\sigma]^2\}$$

$$\Sigma_m \; |A_m|^2 \; \exp\{-2b^2\pi^2[(x_0/\lambda z - m/a)^2 \; + \; (y_0/\lambda z)^2]\} \qquad (18)$$

B.4 BTDF

We can compare our results for scattered light, equation (15), with the results of others by reducing our formula to that of the bidirectional transmission distribution function (BTDF). The incident power is given by

$$\mathbf{P}_{\text{incident}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (G(x_1, y_1))^2 \, dx_1 \, dy_1$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-2x_1^2/b^2 - 2y_1^2/b^2) \, dx_1 \, dy_1$$

$$= b^2\pi/2 \qquad\qquad (19)$$

The bidirectional transmission distribution function BTDF is therefore

$$\text{BTDF} = \mathbf{P}_{\text{scattered}}(\text{sr}^{-1})/\mathbf{P}_{\text{incident}}$$

$$= [\mathbf{I}_{\text{rough}}(x_0, y_0, z) \cos\theta_s \, dA \, /(dA/z^2)]/(b^2\pi/2)$$

$$= \mathbf{I}_{\text{rough}}(x_0, y_0, z) \cos\theta_s \, 2z^2/b^2\pi \qquad (20)$$

where θ_s is the scattering angle and dA is an increment of area.

The value of CC^* comes from multiplying the factor in front of the integral in equation (4.13) with its complex conjugate. The result is

$$CC^* = 1/\lambda^2 z^2. \qquad\qquad (21)$$

The sine of the specular angle $\theta_{2,m}$ in the mth order is given by

$$\sin\theta_{2,m} = \lambda m/a. \qquad\qquad (22)$$

The sine of the scatter angle, to the approximations we are using, is

$$\sin\theta_s = x_0/z. \qquad\qquad (23)$$

We define the spatial frequency p_m by

$$p_m \equiv k(\sin\theta_s - \sin\theta_{2,m}). \qquad\qquad (24)$$

Along the axis $y_0 = 0$ we can write the BTDF using equation (15), and equations (20) to (24) by

$$\text{BTDF} = (2k^4/\pi) \, [\sigma\tau(n-1)/2]^2 \cos\theta_s \, \Sigma_m |A_m|^2 \, /[1 + p_m^2\tau^2]^{3/2}. \tag{25}$$

The typical two-dimensional bidirectional *reflection* distribution function (BRDF) with an exponential autocorrelation function, for normal incidence is given by[3]

$$\text{BRDF} = (2k^4/\pi) \, (\sigma\tau)^2 \cos\theta_s \, F(\theta_s) \, /[1 + p_m^2\tau^2]^{3/2} \tag{26}$$

where $F(\theta_s)$ is approximately one for small angles, and is exactly one in some theories. Comparing equation (26) with equation (25) we see that one difference is the factor $[(n-1)/2]^2$. This difference exists because the phase change on reflection from surface features is proportional to twice the height of the feature, but in transmission the phase change is $(n-1)$. The second difference is the term $\Sigma_m |A_m|^2$. Evidently the existence of the grating replicates the scattering pattern around each diffraction order. The amplitude of the scattering is scaled by the amplitude of the diffraction order.

As an illustration, consider a Dammann grating with relative intensities of 0.1, 0.1, 0.1, 0.1, 0.01, and 0.001, corresponding to diffraction orders 0, 1, 2, 3, 4, and 5, respectively. The scattering pattern around each diffraction order would be identical to the scattering pattern of a conventional optic with the same surface roughness. The magnitude of the scattering around the zero order would be one tenth the magnitude of the scattering from the conventional optic. Around the fourth order the scattering is reduced by a factor of a hundred, and a factor of a thousand for the

fifth order. However, one does not really reduce light scattering by using a binary optic. The total amount of scattered light from the binary optic is the same as the total for the conventional optic, and the ratio of unscattered to scattered intensity could decrease in cases of significant overlapping of scattering from different diffraction orders.

APPENDIX C:

FFT CALCULATION OF DAMMANN GRATINGS

For a Dammann grating we need only consider the one-dimensional problem in x_1, as we did for the cylindrical zone plate in Appendix A. Ignoring some unimportant phase terms in front of the Fraunhofer formula (2.12) the field at the distance z from the grating is given by

$$U(x_0, z) = 1/z\lambda \int_{-\infty}^{\infty} U(x_1) \, G(x_1) \, \exp(-ikx_0x_1/z) \, dx_1 \qquad (1)$$

where $G(x_1)$ is the incident gaussian field, and $U(x_1)$ is the phase profile imparted by the Dammann grating. Equation (1) is the Fourier transform of the product of $G(x_1)$ and $U(x_1)$, so we use FFTs to find the convolution of the Fourier transform of each term, as we did for the cylindrical Fresnel zone plate. For the symmetrical Dammann gratings with two transition points, the Fourier transform of $U(x_1)$ is given by

$$F\{U(x_1)\}\big|_{\xi=x0/\lambda z} = \Sigma_j \int_{ja}^{ja+a1} E(x_1) \, dx_1 + \int_{ja+a1}^{ja+a2} \phi 1_j \, E(x_1) \, dx_1$$

$$+ \int_{ja+a2}^{ja+a-a2} E(x_1) \, dx_1 + \int_{ja+a-a2}^{ja+a-a1} \phi 2_j \, E(x_1) \, dx_1 + \int_{ja+a-a1}^{ja+a} E(x_1) \, dx_1 \qquad (2)$$

where

$$E(x_1) = \exp(-ikx_0x_1/z) \qquad (3)$$

$$\phi 1_j = \exp[ik(n - 1)(h + e_j)]. \qquad (4)$$

The etch depth error term e_j in equation (4) is a different pseudo-random number generated by the computer for each $\phi 1_j$ and each $\phi 2_j$. For a line edge error

we add e_j to each a_j term instead. Mathcad generates uniformly distributed random numbers between 0 and y by the function $rnd(y)$. We can obtain a uniform, normal (gaussian), or an exponential distribution with zero mean and standard deviation δ as follows:[1]

$$e_j = (\delta\sqrt{3}) \; [rnd_j - 1] \qquad\qquad \text{uniform} \quad (5)$$

$$e_j = \delta \; \cos(2\pi \; rnd(1)) \; \{-2 \; \ln[rnd(1)]\}^{\frac{1}{2}} \qquad \text{normal} \quad (6)$$

$$e_j = (\delta/\sqrt{2}) \; (rnd_j - 1)/(|rnd_j - 1|) \; \ln(|rnd_j - 1|) \quad \text{exponential} \quad (7)$$

$$rnd_j = rnd(2). \qquad\qquad (8)$$

Solving equation (2) we find that

$$\mathbf{F}\{U(x_1)\}|_{\xi = x0/\lambda z} = \Sigma_j \; \mathbf{B}(x_0, j)\{1 - C_5(x_0) + (\phi 1_j - 1)[C_1(x_0) - C_2(x_0)]$$

$$+ \; (\phi 2_j - 1)[C_3(x_0) - C_4(x_0)]\} \qquad (9)$$

where

$$\mathbf{B}(x_0, j) = z/ikx_0 \; \exp(-ikajx_0/z)$$

$$C_1(x_0) = \exp(-ika_1x_0/z)$$

$$C_2(x_0) = \exp(-ika_2x_0/z)$$

$$C_3(x_0) = \exp(-ik(a - a_2)x_0/z)$$

$$C_4(x_0) = \exp(-ik(a - a_1)x_0/z)$$

$$C_5(x_0) = \exp(-ikax_0/z).$$

The other term we need to find is the Fourier transform of $G(x_1)$. This is readily given by

$$\mathbf{F}\{G(x_1)\}|_{\xi = x0/\lambda z} = \mathbf{F}\{\exp(-x_1^2/b^2)\}|_{\xi = x0/\lambda z}$$

$$= b\sqrt{\pi} \; \exp[-(\pi b x_0/\lambda z)^2]. \qquad (10)$$

REFERENCES

1.1 J. Hecht, "Binary optics emerging from labs," *Lasers & Optronics*, May 1990.

1.2. R.E. Fischer, "An interview with Wilfrid Veldkamp; binary optics and its applications," *OE Reports*, No. 67, July 1989.

2.1. A. Rubinowicz, "The Miyamoto-Wolf Diffraction Wave," *Progress in Optics*, Vol. IV, 1965.

2.2. *Aviation Week & Space Technology*, December 16/23, 1991.

2.3. D.C. Bertilone, "On the exact Kirchhoff and Rayleigh-Sommerfeld theories for the focusing of an infinite scalar spherical wave-field," *Optics Communications*, Vol. 85, No. 2/3, 1 Sept. 1991.

2.4. J.W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill Book Company, 1968.

2.5. M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, 6th edition, 1980.

2.6. F. Kottler, "Diffraction at a black screen, Part II; Electromagnetic Theory," *Progress in Optics*, Vol. VI, 1967.

2.7. W.B. Veldkamp, G.J. Swanson, S.A. Gaither, C-L. Chen, T.R. Osborne, "Binary Optics: a diffraction analysis," Massachusetts Institute of Technology, Lincoln Laboratory, Project Report ODT-20, 23 Aug. 1989.

2.8. J. Chandezon, D. Maystre, G. Raoult, "A new theoretical method for diffraction gratings and its numerical application," *J. Optics (Paris)*, Vol. 11 No. 4, 1980.

2.9. G.M. Whitman, D.M. Leskiw, F. Schwering, "Rigorous theory of scattering by perfectly conducting periodic surfaces with trapezoidal height profile. TE and TM polarization," *J. Opt. Soc. Am.*, Vol. 70 No. 12, Dec. 1980.

2.10. W.B. Veldkamp and E.J. Van Allen, "Binary holographic LO beam multiplexer for IR imaging detector arrays," *Applied Optics*, Vol. 22 No. 10, 15 May 1983.

183

2.11.   V.L. Brudny and R.A. Depine, "Spectral impurity in gratings due to surface plasmon scattering," *Optics Communications*, Vol. 82 No. 5/6, 1 May 1991.

2.12.   SPIE Vol. 240, Periodic Structures, Gratings, Moiré Patterns and Diffraction Phenomena (1980).

2.13.   M.C. Hutley, *Diffraction Gratings*, Academic Press, 1982.

2.14.   A. Vasara, E. Noponen, J. Turunen, J.M. Miller, and M.R. Taghizadeh, "Rigorous diffraction analysis of Dammann gratings," *Optics Communications*, Vol. 81 No. 6, 15 March 1991.

2.15.   G.J. Swanson, "Binary optics technology: Theoretical limits on the diffraction efficiency of multilevel diffractive optical elements," MIT/LL, Technical Report 914, 1 March 1991.

2.16.   D.N. Qu, R.E. Burge and X. Yuan, "Analysis of surface electromagnetic fields of wavelength-sized surface relief gratings," *Optics Communications*, Vol. 84 No. 3/4, 15 July 1991.

3.1.   M. Born and E. Wolf, *Principles of Optics*, 6th Edition, Pergamon Press, 1980.

3.2.   H. Dammann and K. Görtler, "High-efficiency in-line multiple imaging by means of multiple phase holograms," *Optics Communications*, Vol. 3 No. 5, pp. 312-315, July 1971.

3.3.   U. Krackhardt and N. Streibl, "Design of Dammann-Gratings for Array Generation," *Optics Communications*, Vol. 74 No. 1&2, pp 31-36, 1 December 1989.

3.4.   J. Jahns, M.M. Downs, M.E. Prise, N. Streibl, S.J. Walker, "Dammann gratings for laser beam shaping," *Optical Engineering*, Vol. 28 No. 12, pp 1267-1275, December 1989.

3.5.   M. Heißmeier, U. Krackhardt and N. Streibl, "A Dammann Grating with Diffraction Orders of Arbitrary Intensity etched into Al_2O_3," *Optics Communications*, Vol. 76 No. 2, pp 103-106, 15 April 1990.

3.6.   Xuenong Lu, Ying Wang, Minxian Wu, and Guofan Jin, "The Fabrication of a 25x25 Multiple Beam Splitter," *Optics Communications*, Vol. 72 No. 3&4, pp 157-162, 15 July 1989.

3.7.   G.J. Swanson, "Binary Optics Technology: Theoretical Limits on the Diffraction Efficiency of Multilevel Diffractive Optical Elements," Massachusetts Institute of Technology Lincoln Laboratory, Technical Report 914, 1 March 1991.

3.8.   M.C. Hutley, *Diffraction Gratings*, Academic Press, 1982.

3.9.   G.J. Swanson and W.B. Veldkamp, "Binary lenses for use at $10.6\mu m$," *Applications of Holography*, Proc. SPIE 523, p. 244-250 (1985).

3.10.   C.M. Sparrow, "Theory of Imperfect Gratings," *Astrophysical Journal*, Vol. XLIX No. 2, March 1919.

3.11.   E.L. Church, H.A. Jenkinson, and J.M Zavada, "Measurement of the finish of diamond-turned metal surfaces by differential light scattering," *Optical Engineering*, Vol. 16 No. 4, July-August 1977.

4.1.   R.V. Shack and J.E. Harvey, "An Investigation of the Distribution of Radiation Scattered by Optical Surfaces," Optical Sciences Center (Univ. of Arizona) Final Report, August 1975, as outlined in J.D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, 385-390, John Wiley & Sons, 1978.

4.2.   G.J. Swanson, "Binary optics technology: Theoretical limits on the diffraction efficiency of multilevel diffractive optical elements," MIT Lincoln Laboratory, Technical Report 914, 1 March 1991.

4.3.   M.G. Moharam and T.K. Gaylord, "Diffraction analysis of dielectric surface-relief grating," *J. Opt. Soc. Am.* 72, 1383-1392, 1982.

4.4.   J.M. Elson, "Efficiency and wavelength scaling of diffractive optics," NAWCWPNS TP 8040, August 1992.

4.5.   E.G. Johnson and A.D. Kathman, "Rigorous electromagnetic modeling of diffractive optical elements," *International Conference on the Application and Theory of Periodic Structures*, J.M Lerner and W.R. McKinney, editors, Proc. SPIE 1545, 1991.

4.6.   D.A. Buralli and G.M. Morris, "Design of two- and three-element diffractive Keplerian telescopes," *Applied Optics*, Vol. 31, No. 1, 1 January 1992.

5.1.   J.A. Cox, T. Werner, J. Lee, S. Nelson, B. Fritz, and J. Bergstrom, "Diffraction efficiency of binary optical elements," *Computer and Optically Formed Holographic Optics*, SPIE Proceedings Vol. 1211, pp. 116-124 (1990).

5.2.   M.W. Farn, J.W. Goodman, "Effect of vlsi fabrication errors on kinoform efficiency," *Computer and Optically Formed Holographic Optics*, SPIE Proceedings Vol. 1211, pp. 125-136 (1990).

5.3.   J.A. Cox, B. Fritz, and T. Werner, "Process error limitations on binary optics performance," *Computer and Optically Generated Holographic Optics (Fourth in a Series)*, SPIE Proceedings Vol. 1555, pp. 80-88 (1991).

5.4.   C-L. Chen and T.R. Osborne, "Quantization effects on the fields of electron-beam generated cylindrical zone plates," *Binary Optics: A Diffraction Analysis* by W.B. Veldkamp, G.J. Swanson, S.A. Gaither, C-L. Chen, and T.R. Osborne, MIT Lincoln Laboratory, Project Report ODT-20, 23 August 1989. Also in *Applied Optics*, Vol. 26, No. 12, 15 June 1987.

5.5.   M. Kajanto, E. Byckling, J. Fagerholm, J. Heikonen, J. Turunen, A. Vasara, and A. Salin, "Photolithographic fabrication method of computer-generated holographic interferograms," *Applied Optics*, Vol. 28, No. 4, 15 February 1989.

6.1.   This derivation was worked out with the help of Dr. Merle Elson and Larry Chizek. It was published in Larry's thesis "Near-Angle Scattering and Binary Optics," Naval Postgraduate School, December 1989. The results were given in D.W. Ricks and L.V. Chizek, "Light scattering from binary optics," *Computer and Optically Formed Holographic Optics*, Ivan Cindrich and Sing H. Lee, Editors, Proc. SPIE 1211, 24-37 (1990).

7.1.   J.C. Stover, "Optical Scatter," *Lasers & Optronics*, Vol. 7 No. 7, July 1988.

A.1.   J.D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, 1978.

B.1.   See Reference 6.1.

B.2.   C.C. Sung and J.A. Holzer, "Scattering cf Electromagnetic Waves from a Rough Surface," *Applied Physics Letters*, Vol. 28, No. 8, pp. 429-431, 1976.

B.3.   See Reference 3.11.

C.1.   B.R. Frieden, *Probability, Statistical Optics, and Data Testing*, Springer-Verlag, 1983.