Subwavelength Structures, Optical Diffraction, and Optical Disc Memories

Thesis by

David S. Marx

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

1996
(Submitted January 18, 1996)

Acknowledgements

I am very grateful to my advisor, Professor Demetri Psaltis, for all the support and advice over the years. Clearly, Professor Psaltis was the keystone to all my work at Caltech, and I have never had a teacher who deserves more respect.

The samples used in my measurements and experiments came from several sources, and I am greatly indebted to each. Paul Maker and Rich Muller in the Microdevices Section at the Jet Propulsion Laboratory prepared several samples with electron beam lithography. They were also very helpful and generous with the use of their scanning electron microscope to inspect and photograph samples. I am very grateful for their excellent work and also for the atomic force microscope measurements made by Dan Wilson, also in the Microdevices Section. Another sample was generously provided by Dean Faklis, Rochester Photonics Corporation. Test samples made with silicon and galium arsenide were etched by Professor Scherer and his student Chuan-cheng Cheng, and I am very appreciative of their effort.

One highlight of my graduate studies was my two-week trip to Japan, sponsored by Seiji Kobayashi and Hiroshi Ooki of the Sony Corporate Research Laboratories. Not only did Seiji and the other engineers at Sony teach me the intricacies of optical disc design and manufacturing, but they, along with Seiji's wife Tomo, were very friendly, generous, and hospitable. I had a fantastic trip.

A second tremendous experience was my one-month visit to Gabriel Sirat and his company, C.D.O., in Paris. The trip was sponsored by both C.D.O. and Professor Psaltis. The purpose of the visit was to exchange ideas about conoscopic holography and to learn about their work designing conoscopic systems for surface metrology.

During my years in Professor Psaltis's research group, I have enjoyed all kinds of help, advice, and camaraderie from all the other members. They are: Sidney Li, Yong Qiao, Scott Hudson, Mark Neifeld, Alan Yamamura, Chuanyi Ji, Cheol-Hoon Park, Subrata Rakshit, Charlie Stirk, Steve Lin, Xin An, Robert Denkewalter, Annette

Grot, Jiafu Luo, Kevin Curtis, Geoffrey Burr, Jean-Jacques Drolet, Ernest Chuang, Michael Levene, Allen Pu, George Barbasthathis, Chuan Xie, Ali Adibi, Greg Billock, Steckman, Xu Wang, and George Ouyang.

Of course, no research can be conducted smoothly without the help and assistance of David Sieving, Yayun Liu, Su McKinley, Lucinda Acosta, and Helen Carrier.

Parts of the research presented in this thesis was supported by the Army Research Office.

Abstract

Conventional optical memory discs store information in the form of pits embossed on the disc. The minimum size of the pit marks is limited by the resolution of the optical system used to read the disc. Our investigations, presented in this thesis, are primarily concerned with the question, "Can an optical disc memory be designed so that an optical system can recover information from symbols (pit marks or otherwise) which are normally unresolved?" When an optical system can determine unresolved features of an object, then superresolution has been accomplished.

We describe an experiment to recover information about lines with a width one-fifth the minimum resolvable feature size. The result uncovers an important difference between an optical memory and a classical optical imaging system: in an optical memory, we can use *a priori* information about the finite number of possible stored states. The next investigation is for superresolution in depth, rather than for a lateral direction. We select the method of conoscopic holography and demonstrate the ability to measure the depth of a reflecting surface with an accuracy better than one-tenth the depth of focus of the optical system.

To allow the design and analysis of a memory format, we formulate an integral method to calculate diffraction for large numerical aperture focused beams on non-periodic two-dimensional structures. The numerical method is tested for numerical convergence and accuracy, and some comparisons of numerical results and experimental measurements are also shown. We then use the numerical method extensively to analyze a variety of formats and structures.

Contents

Acknowledgements iii

Abstract v

1 Introduction 1

2 Measurement of Unresolvable Binary Patterns 7

2.1 Introduction 7

2.2 Measuring the Presence of Unresolvable Patterns 9

2.3 The Inverse Problem: Decoding the Message from Far Field Measurements 14

2.3.1 The Test Patterns and Codewords 15

2.3.2 Forming the Training Set 16

2.3.3 Decoding the Test Set 17

2.3.4 Bit Error Rate and Information Capacity 19

2.4 Alternative Decoding Schemes Using Viterbi Decoding 20

2.4.1 Review of Viterbi Decoding 21

2.4.2 Applying Viterbi Decoding to Measurements of the Small Pits 23

2.4.3 Comparison of Majority Rule and Viterbi Decoding by Numerical Simulation 24

2.5 Conclusion 28

3 Surface Depth Measurement with Conoscopic Holography 29

3.1 Introduction 29

3.2 Conoscopic Holography 30

3.3 Design of a Pit Depth Memory 32

3.4 System Design for a Conoscopic Pit Depth Memory 34

3.5 Pit Depth Measurement Results 42
3.6 Conclusion 43

4 Rigorous Diffraction of Focused Spots on Small Structures 47

4.1 Introduction 47
4.2 Integral Method 49
4.2.1 Detailed Formulation 50
4.2.2 Numerical Solution of the Integral Equations 60
4.2.3 Testing the Program's Results 72
4.3 Comparison to Rigorous Coupled-Wave Analysis 80
4.3.1 Overview of RCWA 80
4.3.2 Comparison of Numerical Results 83
4.4 Experimental Verification 90
4.4.1 An Aluminum Grating with TM Incidence 91
4.4.2 A Single Groove in Gallium Arsenide 94
4.5 Conclusion: What's the Program Good For? 97
A Fourier Transform of the Free Space Green's Function 99
B Integral Equations for the TM case 99

5 Analysis and Design of Optical Disc Formats 102

5.1 Introduction 102
5.2 Current Formats 105
5.2.1 Audio CD 107
5.2.2 DVD (the new Digital Video Disc Format) 110
5.3 The Nature of the TM/TE Signal 113
5.3.1 DS-2 114
5.3.2 DS-3 115
5.4 TM/TE Signaling with Quadrature Detection 118
5.4.1 A TM/TE Quadrature Detector 119
5.4.2 DS-3.1 122
5.4.3 DS-6.0 8-PAM 124

5.4.4 DS-6.1 8-QAM 127

5.4.5 DS-6.2 8-QAM+8 130

5.5 TE/TM Quadrature Measurement for Surface Metrology 131

5.6 Conclusion 135

Bibliography **137**

Chapter 1 Introduction

The transmission of information via an optical imaging system is an incredibly important tool in life and science. From human vision to telescopes, microscopes, and optical discs, all optical imaging systems have a spatially distributed input (object), and they transmit information about the object to form an image at the output. The optical system's resolution is its ability to transmit size and spacing information about features of the object. For example, the resolving ability of a telescope determines how close two stars can be before the telescope transmits them as one smeared object rather than two distinguishable ones. Traditionally, resolution of an optical system is defined as the minimum distance between to point sources on the object so that two distinguishable reproductions appear in the image. Any technique to extract more accurate information about the spatial distribution in the image than what is allowed by resolved reconstructions is a technique for superresolution.

The literature on superresolution, the limitations of superresolution, and various methods to increase the capacity of optical data storage beyond the resolution limit is quite substantial. Most approaches involve estimating the amplitude and phase of spatial frequencies beyond the optical system's cutoff. The mathematical basis for these methods is the fact that if an analytic function is known exactly for a finite region, then the function can be determined uniquely for all space. Since the Fourier transform of a spatially bounded object is an analytic function, then, in principle, a limited number of measurements in the Fourier plane of an optical system should give enough information to reconstruct the object with any resolution desired. Methods of analytic continuation involve algorithms for estimating the entire Fourier transform, including evanescent waves, of an object based on measurements in the Fourier plane [16]. Schmidt-Weinmar expanded on this theme by representing the far field of an object in terms of nonuniform planewaves (planewaves with a complex wave vector) rather than simply uniform plus evanescent planewaves [43, 44, 45].

There was some work in the 1960's and 1970's to directly measure the evanescent fields, rather than attempt to estimate the evanescent spatial frequencies, of an object. Interference and diffraction involving evanescent waves generated at the surface of a medium by total internal reflection of light at the boundary with a denser medium was studied by Nassenstein [47, 48]. Lukosz and Wuthrich recorded holographic gratings where at least one of the incident fields was evanescent [46], and Bryngdahl recorded holograms constructed with a propagating object beam and an evanescent reference beam [49]. An example of how a hologram can be made using an evanescent field is to coat a planar waveguide with the recording film. The reference beam is then coupled into the waveguide, and the object beam is incident on the film from above. Because the field propagating in a dialectric waveguide has an evanescent field that travels outside the waveguide, the reference beam will have an evanescent field in the film. The resulting hologram recorded by the film will be the interference between the evanescent field and the object beam.

Any superresolution method is limited by the noise present in the measurement. The theoretical limit to superresolution in the presence of shot noise was treated separately by Gabor [1] and, specifically for astronomical observations, by Lucy [2]. Gabor also attempted to relate the field of information theory (only a few years old at the time) to optical imaging and communication.

An optical memory has an important advantage over other optical imaging systems: the input object can be designed using specifications that can serve as *a priori* information to aid the optical system when attempting to read. In contrast, an astronomer cannot design the composition of the stars, and so he has little *a priori* information about his object before he aims his telescope. However, a data storage system is designed with a specific format, and the set of states allowed by the format is given to the optical imaging system. Because of this difference, the possibilities for superresolution, or maximizing the information content, of an optical memory can be better than for other optical imaging systems.

Consider the optical disc memory as an information channel (figure 1.1). Source information is encoded onto the disc in the form of spatial structures according to

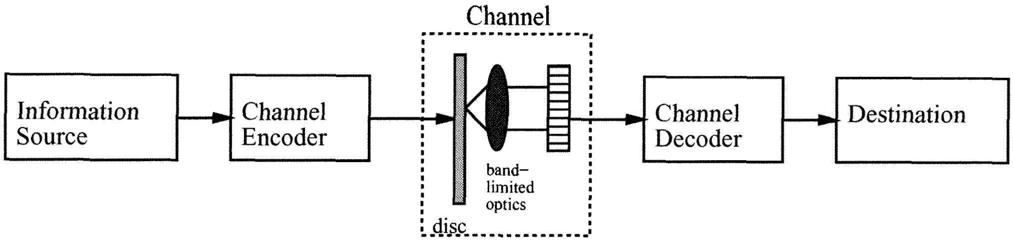


Figure 1.1: The optical disc memory as an information channel.

some convention. At a later time, an optical system attempts to read the information by measuring the optical response to the the spatial structures. Thus, the channel consists of the disc and the optical system. This channel has a different input alphabet and output alphabet. The input alphabet consists of the spatial structures, while the output alphabet is the set of corresponding measurements resulting from the optical imaging system. The channel encoder relates the source information (perhaps the output of a source encoder) to the physical structures on the disc, and the decoder works backward with the output of the optical system. Errors occur when noise in the channel causes an output to be either unrecognizable or resembling an output for the wrong input. A channel's capacity is defined as the maximum amount of information that can be reliably transmitted across the channel, and it depends on the nature of the channel's codewords and the amount of noise present.

The published work regarding the improvement of an optical disc's storage density can be divided into two broad categories. One general approach is to modify the optical system's transfer function through some nonlinearity in the optical medium or through optical techniques such as apodization. For example, Bouwhuis and Spruit [3] use a nonlinear substrate layer over the information pits and show that the optical transfer function widens with spot power. Yanagisawa and Ohsawa [4] coat a disc with a nonlinear material, whose transparency is dependent on the incident light intensity, modifies the illumination spot and effectively broadens the optical system's numerical aperture by up to a factor of three. Rather than use a nonlinearity to affect the illumination spot, a group at the Sony Corporate Research Laboratories [5] placed a nonlinear layer with the information layer in a magneto-optic memory that, when heated by the illumination spot, causes the information to temporarily disappear. Since, as the spot moves from one mark to the next, the first mark disappears, there will be no crosstalk between a mark and the previous mark. However, the crosstalk between a mark and the following mark will still be present. This system can be represented by an asymmetric optical transfer function.

The other general approach is to accept the linear optical system as given and modify the storage format and use signal processing to resolve the information. The work presented in this thesis is of this type. One example is pit edge position modulation, presented by Kobayashi [6]. In this format the center of each pit is constant and the minimum pit width is resolvable by the optical system, but the position of each edge of each pit is modulated and the minimum position difference is a small fraction of a resolution element. Another method [7] squeezes the track pitch to about half the spot size. The crosstalk between the tracks is reduced by reading with three staggered illumination spots. The spots are staggered at least one spot size in the track direction and half a spot size in the radial direction. The signal from each spot is given an appropriate time delay, and then a transversal filter and equalizer produce the desired output. Foreshadowing our approach in Chapter 5, Ooki [8] designed a format interleaving tracks of two different types of pits. One set of tracks use pits with the standard one-quarter wave depth, and the other set of tracks use pits with a depth of $\lambda/18$. The shallow pits are interleaved between the standard tracks, which are spaced at the standard pitch. Thus, the information density is doubled. To read from the memory, the detection includes a a waveguide incorporating the electro-optic effect in the detection arm. By adjusting the phase difference between two propagating modes in the waveguide, the phase signal from the shallow pits can be separated from the amplitude signal of the standard track. Although their intended application was magnetic media, the modified Viterbi algorithm developed by Zeng and Moon [9] improves the performance of Viterbi decoding in a jitter dominant channel, the type of channel that describes the optical disc. Perhaps, their Viterbi algorithm could be combined with Kobayashi's technique, or our technique of Chapter 2. Many more ideas for using superresolution in optical discs has been published in the literature, and we have only presented a sampling to introduce previous work related to the research presented in this thesis.

We start our investigation of the limits of information density of optical memories by considering the limits of the linear optical system. In Chapter 2 we describe an experiment to recover information about lines with a width one-fifth the minimum resolvable feature size. The result uncovers an important difference between an optical memory and a classical optical imaging system: in an optical memory, we can use *a priori* information about the finite number of possible stored states. Therefore, we only need to measure which state is present; we do not need to actually image the state, which would require far more spatial bandwidth. We then try a simple and direct approach to the recovery of stored states (pattern of subwavelength lines). By comparing far field diffraction measurements of the unknown state to previous measurements of known states and the use of a simple algorithm to make decisions, a sequence of states is decoded and the error rate is measured.

While Chapter 2 is concerned with the resolution of subwavelength patterns in the lateral dimensions, Chapter 3 investigates resolving information stored as depth, i.e. the direction of optical propagation. We select the method of conoscopic holography and demonstrate the ability to measure the depth of a reflecting surface with an accuracy better than one-tenth the depth of focus. However, since the stored information only increases with the logarithm of the number of available depths, but the inverse square of the useful lateral dimension, depth encoding alone cannot give much improvement in storage density.

After demonstrating the possibility of storing information with unresolvable features and the utility of depth in Chapters 2 and 3, the last two chapters are devoted toward the design a memory to best use these properties. We cannot be in a position to design a memory, or a set of states, that are in any way optimal without a satisfactory model of the optical diffraction for such structures. Chapter 4 is devoted to the development of a rigorous numerical method for calculating such diffraction. We give the detailed formulation of an integral method designed specifically for large numerical aperture focused beams with nonperiodic structures in two dimensions. The numerical method is tested for numerical convergence and accuracy, and some comparisons of numerical results and experimental measurements are also shown.

We start Chapter 5 by applying the numerical method developed in Chapter 4 to analyze the optical disc memory format now in use for the compact disc and the format proposed for the future digital video disc. With the use of the numerical method, we discover some subwavelength structures that exhibit remarkably different behavior for the two polarizations of the electric field with respect to the groove direction. To take advantage of these differences, we design a detector which measures the amplitude and phase relationship of one polarization with respect to the other. A series of futuristic formats which use subwavelength features and variable depth designed to make optimal use of the polarization differences. We dub this method "TM/TE Signaling."

Chapter 2 Measurement of Unresolvable Binary Patterns

2.1 Introduction

Resolution in an imaging system is limited by its spatial bandwidth. If two pits are close together in the object plane, then their images in an image plane of an optical system will be smeared together so that they will be difficult to distinguish. In the absence of noise, however, one can always determine that the object contains two pits, no matter how close they are or how smeared the image is, because an optical imaging system performs a linear and unique transformation from the object to the image. Since the Fourier plane contains the same information as the image plane, but in a different representation, the Fourier plane image is also unique. The method of analytic continuation[16] is an attempt to exceed the resolution limit by relying on the uniqueness of the Fourier image. This method predicts the Fourier plane image at spatial frequencies beyond those recovered by the imaging system based on the part of the Fourier plane that was measured. The reconstructed Fourier plane can then be used to calculate an image with greater resolution than otherwise permitted. In the presence of noise, however, the predicted part of the Fourier plane will be severely dominated by the noise, and this method will not provide any resolution improvement. Thus, noise is what makes two closely spaced objects indistinguishable in an optical imaging system, and analytic continuation can only be effective when the noise level is very low.

Recovering information in a noisy channel is a somewhat different problem from imaging a completely unknown object. In a digital memory the information is stored by recording codewords which are elements from a finite set. To read from the memory, the detection scheme merely needs to decide which of the possible codewords is present. This problem is very different from the classical imaging resolution problem. Imaging is essentially measurement without any *a priori* knowledge, whereas, in the memory detection problem, the set of possible codewords is known beforehand. Because of this difference, we would not expect the classical resolution limits for imaging to become fundamental limits for the density of an optical memory. When the codewords are constructed of features much smaller than the resolution limit of the optical system, the question becomes how do we determine which codeword is present without requiring a reconstructed image. As in analytic continuation, we can rely on the uniqueness of the Fourier plane representation. However, there is no need to predict the Fourier image for the higher spatial frequencies. We can simply compare the measured Fourier plane image to the finite set of Fourier plane images corresponding to the finite set of possible codewords and pick the "closest" one.

Our first task is to demonstrate that, in fact, codewords constructed of unresolvable elements can result in unique measurements in an optical imaging system. In our demonstration, we will use the Fourier plane of the optical imaging system. In order to limit the number of possible codewords, the illumination will be a tightly focused spot. As a result, the space bandwidth product of the optical system will be near one. Because the object has such a wide spatial bandwidth, while the space bandwidth product of the imaging system is near one, we expect the intensity in the Fourier plane to be slowly varying across the plane. We will measure changes in the Fourier plane intensity while scanning a test pattern, consisting of lines one-fifth to one-tenth the width of the minimum resolution element, across the focused illumination spot. The repeatability of these intensity changes will demonstrate that the presence of different codewords can be detected.

The same principle can be used with an image of the illuminated region rather than the Fourier plane representation. However, the image of the region illuminated by the small spot will be a bright spot, somehow modulated by the codeword. Because of the huge variation in brightmess in the image plane, in contrast to the slowly varying intensity across the Fourier plane, the dynamic range requirement of the recording device would be much larger if we were to use the image plane rather than the Fourier plan.

After the preliminary demonstration of uniqueness in the Fourier image plane for unresolvable codewords, we are still left with the problem of decoding the Fourier plane measurements and recovering the actual information. In section 2.3 we will discuss the design of the test patterns and codewords as well as a decoding algorithm. The decoding algorithm will use a majority rule to take advantage of repeated measurements while scanning the codewords. The decoding will be applied to Fourier plane measurements in an experiment to demonstrate the recovery of information and determine the bit error rate and the achievable channel information capacity. After the successful decoding demonstration using experimental data, we will determine the improvement that an optimal decoding scheme, such as Viterbi decoding, can offer. The Viterbi algorithm is a maximum likelihood decoder designed for decoding sequences of noisy channel symbols, and it is usually applied along with convolutional error correcting codes. We will adapt this algorithm to our specific problem and measure its performance improvement over majority rule with a computer simulation.

2.2 Measuring the Presence of Unresolvable Patterns

The first step in demonstrating the possibility of recovering information from optical features smaller than the classical resolution limit is to show that measurable differences exist in the optical measurement of different patterns of unresolvable features. The resolution of a microscope with coherent illumination is about λ/NA [12], where λ is the wavelength of the illumination and NA is the numerical aperture of the microscope objective. For $\lambda = 0.6328\mu\text{m}$ and $\mathit{NA} = 0.7$, the resolution limit is about $0.9\mu\text{m}$.

We had a sample fabricated that contained patterns of $0.1\mu\text{m}$ and $0.2\mu\text{m}$ lines. The sample was fabricated at the Jet Propulsion Laboratory (JPL) using electron beam lithography with a direct write, lift-off method. The patterns consisted of a

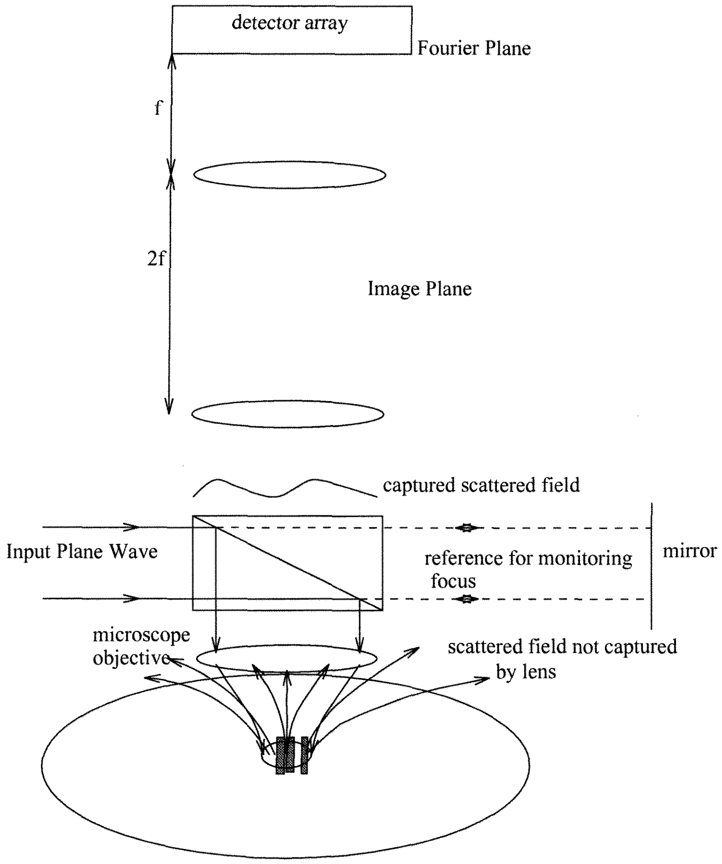


Figure 2.1: Experimental setup for measuring the intensity in the reflected Fourier plane

silicon substrate with lines of platinum on a titanium adhesion layer. The titanium layer was about 75Å high and the platinum added about another 300Å in height to the lines. All the lines were 10μm long so that we could restrict our superresolution measurement to one direction. In the other direction, the lines were arranged in repeating patterns with a period of 1μm. Having repeating patterns served two purposes: since the lines cannot be resolved we need to use the repetition of the patterns as an aid to independently determine the current location of the illumination spot, and by comparing measurements spaced 1μm apart we can estimate the amount of noise in the measurement process.

The optical measurement system is shown in figure 2.1. The laser illumination has a wavelength of $\lambda = 0.6328\mu$m, and $\mathcal{NA} = 0.7$ for the microscope objective. A spatial filter (not shown) expands the laser beam and ensures a nearly uniform incident beam to the objective. The objective focuses the incident beam to a spot on the test sample and then collects the reflected diffracted light. A pair of lenses, forming a 4-f system, image the back focal plane (the Fourier plane) of the microscope objective onto the CCD camera. The light separated by the beam splitter and not used for illuminating the sample can be used to adjust the focus of the illumination spot on the sample by inserting a mirror to form an interferometer. When the illumination spot is exactly on the surface of the sample, the reflected light becomes a collimated planewave (assuming a blank region on the sample is used). In this case, the interference pattern formed by the reflection of the focused spot and the reflected planewave from the mirror has either no fringes, or straight line fringes, depending on the angle of the mirror. If the illumination spot on the sample is out of focus, then the interference pattern has curved fringes. We found that this is a very sensitive method for guaranteeing the focus of the illumination spot.

Since the test pattern has variation in only one direction, we only need to use a cross section of the CCD image. For all of our measurements, we used a window on the CCD image that was the complete length of the image in the direction of variation and about five or ten pixels in the other direction. The intensity of the pixels within the window was then averaged across the narrow direction. The result was a one-dimensional graph of intensity versus position in one direction of the CCD. After each measurement, the test pattern was translated 0.1μm in the direction of variation and a new CCD image recorded. Since we are interested in determining the change in the CCD image resulting from changes in the test pattern, we calculated the difference between each successive intensity profile from the previous profile. These difference profiles are what we considered as the signal. Figure 2.2 shows some examples of the signal from different parts of the test pattern, and figure 2.3 shows two signals measured from locations on the test pattern 1μm apart. Because of the 1μm periodicity in the test patterns, the two signals in figure 2.3 should be identical, and any difference is due to noise in the optical system and inexact movement by the translation stage. We measured the noise level present by following the same differencing procedure for several steps when no pattern is present. The root mean

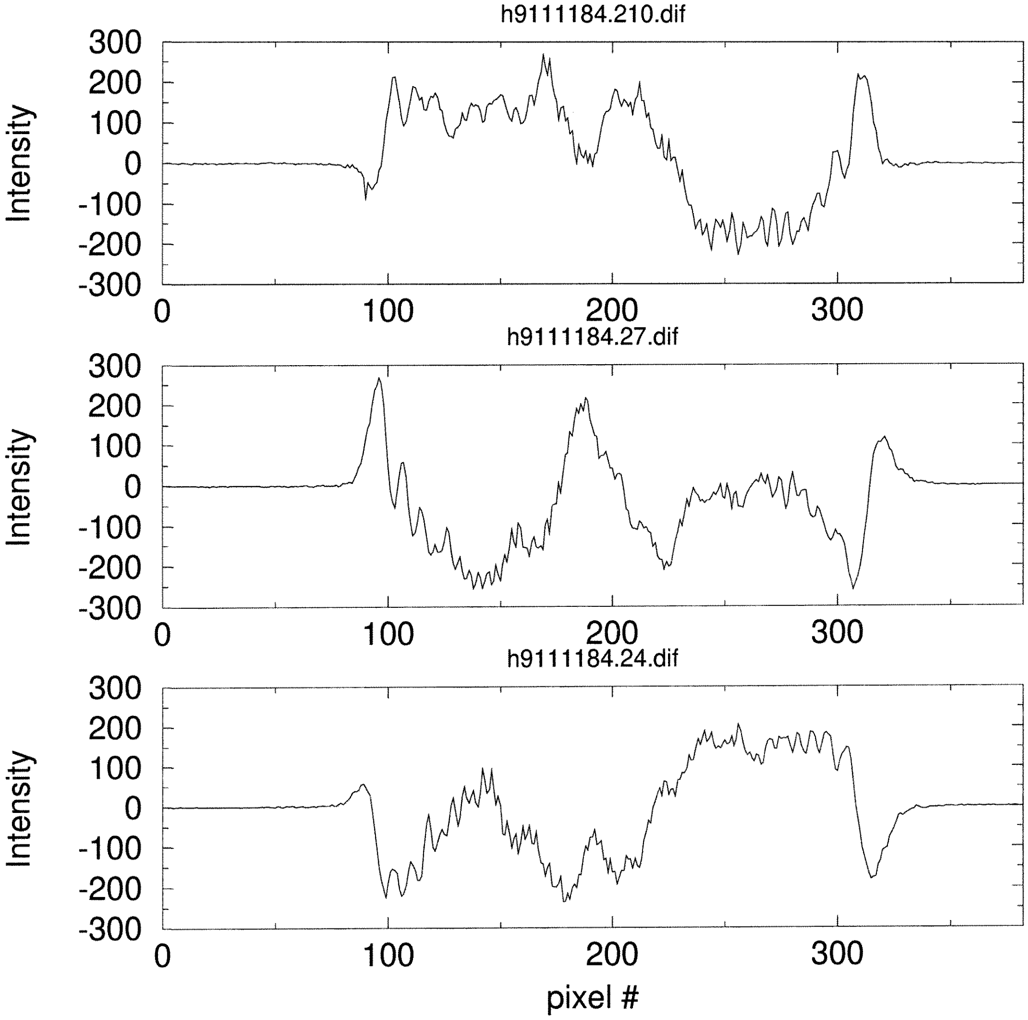


Figure 2.2: Intensity difference profiles measured for different test patterns

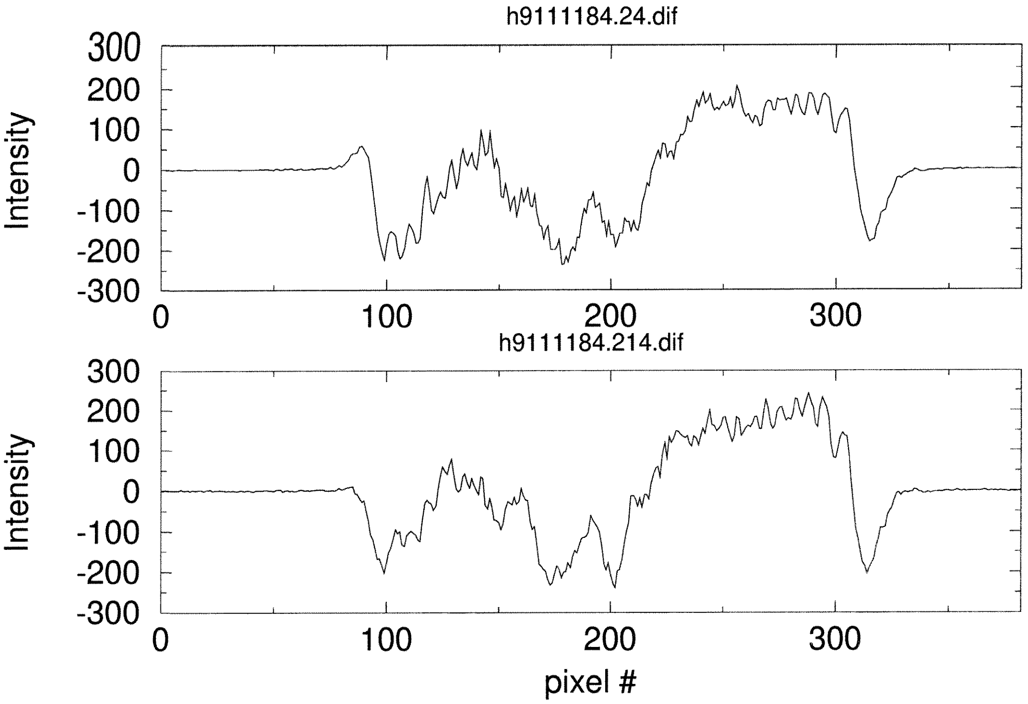


Figure 2.3: Intensity difference profiles measured 1μm apart show the measurement is repeatable

squared (rms) level was then calculated for each pixel. From this noise measurement and typical signal measurements, we surmise a signal to noise ratio of about five or six. We also found that a significant amount of the signal was simply due to the difference in reflection between the platinum and silicon. This difference in reflection gives the zero spatial frequency signal, and since our system misses the high spatial frequencies, this dominance of the reflectance is not surprising.

Figures 2.2 and 2.3 demonstrate our first goal. The first figure shows that significant, measurable features in the reflected field can be measured for features that are well beyond the classical resolution limit. Each profile in the figure, itself a measure of the change in the reflected intensity profile resulting from a 0.1μm shift of a pattern, shows a strong signal well above the noise level. In addition, the three graphs, taken from different regions of the test pattern, have significant differences in their shape, and these differences imply the possibility of assigning measured intensity patterns to different test patterns so that, given an intensity pattern, the corresponding test pattern can be determined. Figure 2.3 simply demonstrates that all these measured differences are not noise or somehow accidental. Of course, we can in no way claim that these patterns of 0.1μm lines are being resolved. Our only statement is that we can measure changes in the reflected intensity profiles for different patterns.

The results of these measurements can be better understood by considering that we made the intensity measurements close to the Fourier plane of the objective lens. Because of the linearity of the Fourier transform operation, the Fourier transform of each pattern is unique. However, because the pattern consists of lines that are smaller than the resolution limit, much of the power in the Fourier transform lies outside the aperture of the objective lens. Therefore, only a part of the Fourier transform of the pattern is available to be measured. An image of the pattern could not be formed with another lens because not enough of the Fourier field is available. However, our experiment shows that even though not enough of the Fourier field is available to resolve the pattern, enough of the field is present so that a difference can be measured between the Fourier fields for the different patterns. Naturally, because we are only measuring intensity and not phase, some degeneracy will be present between patterns that would only have a difference in phase in their Fourier patterns. Many of these degeneracies are eliminated, however, because of the shape of the illuminating spot.

2.3 The Inverse Problem: Decoding the Message from Far Field Measurements

Now that we have established that different intensity profiles can be reliably measured for different patterns of unresolvable lines, we would like to solve the inverse problem—determine which pattern is recorded by measuring its reflected intensity profile. To accomplish this data recovery, we take advantage of three principles. The first principle is to take advantage of the property established in the last section. Since intensity profiles are distinctive for different recorded patterns, we should be able to measure a profile and, somehow, relate it back to the recorded pattern. However, since the intensity profiles do not necessarily have a one to one correspondence with the set of codewords, this property alone is not enough. The second principle is that if the disc is shifted by one pit between each measurement, then each pit is illuminated N times, where N is the number of pits per spot size. This property implies that there is redundancy in the measurements. We will take advantage of this redundancy when decoding the measurements. The last principle is that we know *a priori* all the possible patterns; the pit pattern illuminated at any one time can only be one of 2^N patterns. We shall make use of each of these principles in an attempt to experimentally demonstrate reading a set of sub-wavelength lines.

2.3.1 The Test Patterns and Codewords

Our test patterns for this experiment were fabricated at JPL in a similar manner as the sample for the previous measurements. For this experiment, we wanted all the possible patterns that fall within a 1μm space. Using 0.2μm lines, there are 32 such patterns. As in the previous measurement, patterns were repeated every 1μm. Because many patterns are simply shifted versions of other patterns, only seven basic repeating patterns were required. For example, if the codeword 10000 (corresponding to a single 0.2μm line and a 0.8μm gap) is repeated, 1000010000, and then the illuminated spot is shifted 0.4μm, the resulting illuminated codeword is 00010.

Since the test patterns are made up of 0.2μm lines, and we can make translation steps of 0.1μm, we can add another level of redundancy. We consider the codewords to consist of five lines so that the illumination spot, which is about 1μm wide, covers one codeword. Since a measurement is made after each 0.1μm step, each line is measured, as part of some codeword, ten times. To ease the decoding algorithm, we will consider this scheme to consist of a precoder which codes an information codeword, $c \in C$, where C is the set of all codewords consisting of five information bits, $C = \{0,1\}^5$, to a channel symbol, $\hat{c} \in \hat{C}$, the set of codewords consisting of ten channel bits, $\hat{C} = \{0, 1\}^{10}$. The precoder performs the mapping,

$$\text{precoder}: \quad c = \{a_1, a_2, a_3, a_4, a_5\} \rightarrow \hat{c} = \{a_1, a_1, a_2, a_2, a_3, a_3, a_4, a_4, a_5, a_5\}. \tag{2.1}$$

With this precoder, each channel bit is now represented by a 0.1μm line. The precoder ensures, however, that a channel bit never appears alone, so that all the lines and spaces will be multiples of 0.2μm. The number of channel symbols is now 63, the 32 information codewords plus these codewords shifted by 0.1μm.

2.3.2 Forming the Training Set

We first need to record the Fourier plane patterns for the 63 channel symbols in order to form a training set. The training set can be considered a set of "perfect" examples matching a Fourier plane intensity pattern to each channel symbol. When attempting to decode measurements of unknown symbols, we will use the training set to match Fourier plane measurements. Creating the training set is a chicken-and-egg type problem, because we want to know which symbol we are measuring, but the only information we have is the measurement of the symbol. Our solution to this problem is to rely on the 1μm periodicity of the symbols. Because platinum has a greater reflectivity than silicon, the total reflected intensity oscillates with a period of 1μm as the patterns are translated under the illumination spot. For example, channel symbols with four neighboring platinum lines will have the brightest reflection when the symbol is $\{0, 0, 0, 1, 1, 1, 1, 0, 0, 0\}$, and the dimmest reflection when the symbol is $\{1, 1, 0, 0, 0, 0, 0, 0, 1, 1\}$. Then, from knowing which block is current and which specific intensity profiles correspond to the reflected intensity peaks, the specific assignments of channel symbols to measured intensity profiles can be made. Since each block contains seven repetitions of a channel symbol, we can record each symbol seven times and use as the training set profile for each symbol the average Fourier intensity profile for that symbol.

2.3.3 Decoding the Test Set

Our goal is to measure the far field intensity profiles of a set of unknown patterns and attempt to determine the arrangement of the measured lines. As explained above, each measurement is made after the sample is translated 0.1μm, resulting in a sequence of redundant measurements. Decoding a sequence of measured profiles resulting from a pattern of unkown channel symbols is a two step process. The first step is to compare each measured profile to the training set measurements to find the channel symbol that has an intensity profile closest (by some measure) to the measured profile. The result of this first step is a sequence of channel symbols which should overlap because of the redundancy generated by measuring with 0.1μm shifts. The second step is to recover the message symbols from the channel symbols, utilizing the redundancy to eliminate errors. Many different methods can be used to perform the decoding. The method that we used, and the one that we will describe here, uses a simple majority rule to decide each bit.

The unknown symbols are patterns similar to the ones used to generate the training set, but written in a different location on the sample. In the first decoding step, each measured profile is compared against the stored training set, and the codeword corresponding to the training set profile which has the smallest Euclidean distance from the measured profile is selected. The Euclidean distance between a measured and stored intensity profile is

$$D_k = \sum_{i=1}^{M} (p_i - p_i^k)^2, \tag{2.2}$$

where k is the index for the training set codeword, p_i is the intensity of pixel number i, and the profile has M pixels. The Euclidean distance metric is not the only choice for this type of decoding. Another choice might be a correlation type metric. The Euclidean distance actually contains the correlation of the profiles; this term becomes apparent when expanding the square in the Euclidean distance. The Euclidean distance metric, however, also contains terms relating to the total power in each profile. Since we already know that our measured profiles contain intensities only of the low spatial frequency components of the reflected field, we expect differences in total reflected power to be a very significant factor in distinguishing codewords. Our method described in the last section for measuring the training set also shows the importance in accounting for differences in total intensity.

Using the Euclidean distance measure, we want to choose the channel symbol from the training set that corresponds to the intensity profile with the least distance from the measured profile. For measurement number t, the decoded channel symbol is,

$$\hat{c}^t \in \hat{C} \quad \text{such that} \quad D^t = \min_k D^t_k. \tag{2.3}$$

A sequence of measurements, each after a 0.1μm step, produces a sequence of minimum distance channel symbols, $\{...\hat{c}^{t-1}, \hat{c}^t, \hat{c}^{t+1}...\}$. As explained in section 2.3.1, each channel symbol consists of ten bits, and as explained above, $\hat{c}^t$ and $\hat{c}^{t+1}$ overlap by nine bits. Conversely, each channel bit is incorporated into ten sequential channel symbols. By staggering the received channel symbols, each column should contain the identical channel bit. For example, consider the received channel symbol, $\hat{c}^t = \{a_0, a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9\}$. We can arrange it and the next nine codewords in the following manner:

$$\begin{array}{rccccccccc|c|l}
\hat{c}^t = & a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & \\
\hat{c}^{t+1} = & & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10} \\
\hat{c}^{t+2} = & & & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 & a_{10}\cdots \\
\vdots & & & & & & \vdots & & & & \vdots & \\
\hat{c}^{t+9} = & & & & & & & & & & a_9 & a_{10}\cdots
\end{array}$$

From the above diagram, we see that in the process of choosing the channel codeword with the minimum distance profile to the measured profile, we have taken ten guesses at the channel bit a_9. One simple solution for the final decision of this bit is a majority rule: count the number of ones in a column, and, if the number is greater than five, decode the channel bit as a one.

After the majority rule decision, we have estimates of the channel bits. However, we still need one more step in the decoding process—decoding the message bits from the channel bits. As described in section 2.3.1, the precoder forms two channel bits from each message bit. For this last task in recovering the message, errors in synchronicity will cause severe problems. For our purposes, we will assume perfect synchronicity, although we will see that in the measurement, synchronicity errors will be apparent. In a practical system the synchronicity problem is solved by using an additional source code. Assuming that we have perfect synchronicity and that we know which two channel bits belong to the same message bit, our majority rule decoding algorithm is modified to count the number of times two neighboring channel bits were decoded as a one.

2.3.4 Bit Error Rate and Information Capacity

In our measurement of an independent test set of 420 channel bits using the intensity profiles stored from a training set as described in the previous two sections, we obtained a channel symbol error rate of greater than 1/2. However, since there were 63 codewords, this error rate is significantly better than random guessing. After the decoding, only two message bit errors occurred. The occurrence of two bit errors out of 420 channel bits corresponds to a bit error rate on the order of 10^{-2}. In addition, 12 synchronicity errors occurred. Having measured the bit error rate, we can now make a statement about the information capacity of storing information and recovering it with this method. If we consider the whole process, starting with a random message, writing 0.2μm lines according to a precoded message, and measuring and decoding as described to produce the estimated message at the output, we can assume a binary symmetric channel as a representation. The information capacity of a binary symmetric channel is [14]

$$C = 1 - H(\epsilon) \quad \text{bits/channel symbol,} \tag{2.4}$$

where ϵ is the probability of a channel bit error, and $H(\epsilon)$ is the entropy function,

$$H(\epsilon) = -\epsilon \log_2(\epsilon) - (1-\epsilon)\log_2(1-\epsilon). \tag{2.5}$$

For our measurement of channel bit error rate of about 10^{-2}, the corresponding channel capacity is about 0.92 bits per channel bit, or 4.6 bits per μm since the channel has 5 message bits per μm.

2.4 Alternative Decoding Schemes Using Viterbi Decoding

The majority rule decoding, while simple to implement, is probably not the best method. Although our majority rule decoder utilizes the measurement redundancy for each pit, it does not account for the history of all the other pits sliding beneath the spot. Each measurement of a specific pit is affected by the other pits appearing within a small distance both before and after, and the measurement of each of these pits are affected by other pits, and so on. The majority rule decoder treats measurement as an independent trial, whereas if we were to account for previous measurements, certain possibilities in the current measurement would be eliminated. For example, assuming again perfect synchronicity, if a pit pattern is somehow known to be $\{a_0, a_1, a_2, a_3, a_4\}$, then the measurement after the disc is shifted by one pit must be $\{a_1, a_2, a_3, a_4, a_5\}$, and the only unknown for this pit pattern is a_5. The majority rule decoder would treat the shifted pattern as completely unknown and compare its measurement to all possibilities, even though only two possibilities exist.

A decoding method that fully accounts for the effect of history is Viterbi decoding [15]. The Viterbi decoder algorithm is designed to decode sequences of bit vectors generated by a convolutional code algorithm. Convolutional codes are characterized by an input information bit vector, a finite state machine, and an output bit vector. The output bit vector at any point in the sequence depends on the current state of the machine and the input bits. The input bits are the information bits, and the output bit vector is the channel symbol that is transmitted to the receiver. At the receiver, the Viterbi algorithm reconstructs the encoder's path through state space along with the information bit sequence. Convolutional codes purposely add redundancy to an information stream so that errors that occur after transmission over a noisy channel can be corrected by the receiver. The redundancy in our sub-wavelength pit memory was not designed in any optimal way as a convolutional code would be. However, a similar redundancy is present, and the Viterbi algorithm can be adapted to decode the measurements in order to take full advantage of the redundancy. The cost of using a Viterbi decoder is the added complexity of the decoder and increased sensitivity to synchronization errors. The purpose of this section is to explain how the Viterbi algorithm can be adapted to decode the redundant measurements of our pits, and then to compare it to the majority rule algorithm in both error rate and complexity.

2.4.1 Review of Viterbi Decoding

The best way to consider Viterbi decoding is to think of the information sequence as the description of the time evolution of a finite state machine. A finite state machine has a set of possible states, Q, and its state at time step i is $q_i \in Q$. For every information bit, a_i, input to the machine, the machine performs two operations: it produces an output vector, $\mathbf{y}_i(a_i; q_i)$, and it moves to a new state, q_{i+1}. The new state depends on the previous state and the input bit. In addition, the finite state machine is required to have a one-to-one correspondence between an input bit sequence and a path through state space, the sequence, $\bar{q}_i = \{q_0, q_1, \ldots, q_i\}$. Similarly, a one-to-one correspondence between a path through state space and the sequence of outputs must also exist.

When the outputs, $\mathbf{y}_i$, are transmitted across a noisy channel, a Viterbi decoder receiving the corrupted sequence, $\hat{\mathbf{y}}_i$, will compute the most likely path through state space that produced these outputs. The most likely information sequence is the sequence that corresponds to the most likely path through state space. For a decoding scheme to be optimal, it must consider all possible sequences. At each time step, the

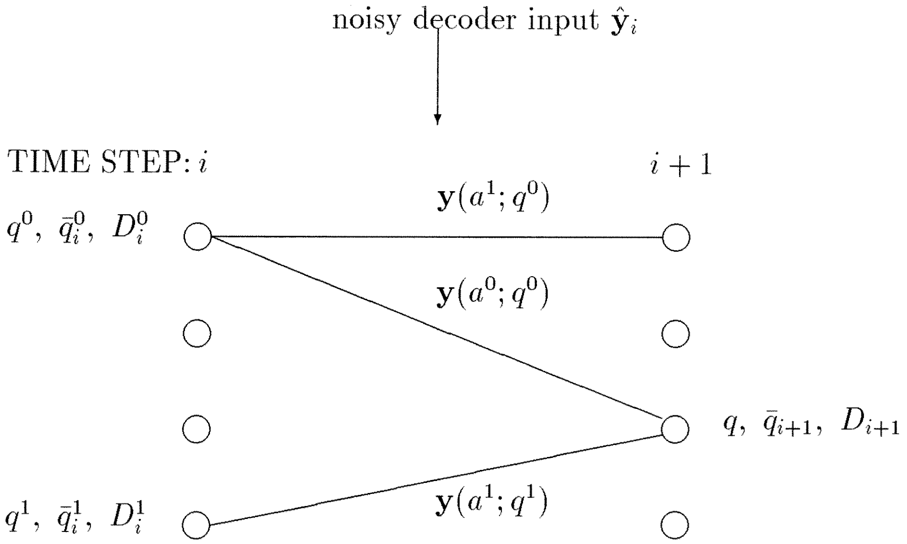


Figure 2.4: The path through state space with its associated metrics. At each time step, the Viterbi decoder considers all the possible states (aligned vertically), and their associated path history, $\bar{q}$, and its metric, D.

number of possible sequences doubles, since the next bit can be a one or a zero. To keep up with this exponential growth in possibilities, the decoder needs to eliminate half the possibilities at each time step. This principle of elimination at the same rate as the rate of growth is the basis for the Viterbi algorithm. Figure 2.4 outlines the process of one time sequence. Each vertical set of states represents all the possible machine states. Associated with each state at each time step is a path, $\bar{q}$, of minimum metric that ends at that state. Associated with that path is its metric, D. Each possible machine state has branches leading out to two states at the next time step, and each state at the next time step has two possible branches leading in from the previous time step. Each branch also has an associated output vector, $\mathbf{y}(a;q)$.

Consider the state, q, at time step $i+1$. It has two branches leading into it, and each branch comes from a state with an associated path history and metric. However, from time step $i+1$ forward, since the two branches merge at state q, they will follow identical paths. Therefore, at this time step, we can elliminate the less likely. For example, suppose at time step $i+1$ the two branches leading into state q are from the states q^0 and q^1. State q^0 has the associated path $\bar{q}_i^0$ and metric D_i^0, and q^1 has $\bar{q}_i^1$ and D_i^1. The output associated with the branch from q^0 to q is $\mathbf{y}(a^0; q^0)$, and the output for the other branch is $\mathbf{y}(a^1; q^1)$. We want to eliminate the least likely possibility. Then, the new metric for state q is

$$D_{i+1} = \min_{j=0,1}\Big[D_i^j + d\big(\hat{\mathbf{y}}_i, \mathbf{y}(a^j; q^j)\big)\Big], \tag{2.6}$$

where $d(\hat{\mathbf{y}}_i, \mathbf{y})$ is a distance measure between the received noisy channel symbol, $\hat{\mathbf{y}}_i$, and the branch label $\mathbf{y}$. The new path for state q is updated to $\bar{q}_{i+1} = \{\bar{q}_i^j, q^j\}$. In this manner, the path history and associated metric are updated for each state during each time step. At the end of the information sequence, the overall most likely path is the one with lowest associated metric. However, since most information sequences are quite long, waiting until the end is impractical. Typically, the path histories are truncated after a certain fixed length. The decoder output at each time step is then the oldest bit of the path with the lowest metric.

2.4.2 Applying Viterbi Decoding to Measurements of the Small Pits

To adapt the Viterbi algorithm to our problem of decoding the Fourier plane measurements of the small lines, we need to define the states, the branch labels, and the metric. There are many possibilities, and proving that one choice is better than another, except by simulation, is very difficult. Perhaps the most natural choice is to define the finite state machine as having thirty-two states, described by the five message bits that are illuminated by the focused spot. If the test pattern steps to the left, then the time evolution in state space describes the left most bit sliding out from under the spot while a new bit slides in from the right, and the other four bits shift one place to the left. The branch label, the finite state machine's output, is the intensity measurement in the Fourier plane for the current state. Since two steps and two measurements will be made before the next state is completely under the spot, we can add to the branch label the measurement after the first step. Thus, the branch label is the pair of vectors,

$$\mathbf{y}(a;q) = \Big[I(u;q), I(u;q_a)\Big]. \tag{2.7}$$

In the above equation, $I(u;q)$ represents the intensity in the Fourier plane along the u axis, when the state q is illuminated. The state q_a represents the state q after a single step of $0.1\mu\text{m}$ so that the first half of the input message bit a is now illuminated. Using channel symbols, if $q = \{a_0, a_0, a_1, a_1, a_2, a_2, a_3, a_3, a_4, a_4\}$, then $q_a = \{a_0, a_1, a_1, a_2, a_2, a_3, a_3, a_4, a_4, a_5\}$, where a_5 is the next input message bit. Looking at the state at the next time step, we clearly see that it has two branches leading in: one state is for the exiting bit equal to one, and the other for the exiting bit equal to zero. The last definition that we need to make is the metric. Assuming that we somehow know, possibly through measurement as we did for the majority rule decoding, "perfect" examples of the branch labels, we can define the metric using Euclidean distances between two successive noisy measurements and the two intensity profiles in the branch labels so that

$$d(\hat{\mathbf{y}}_i, \mathbf{y}(a;q)) = \sum_u (\hat{I}_{i0}(u) - I(u;q))^2 + \sum_u (\hat{I}_{i1}(u) - I(u;q_a))^2. \tag{2.8}$$

$\hat{I}_{i0}$ and $\hat{I}_{i1}$ are two noisy measurements made one step apart, and, as previously stated, we are assuming perfect synchronization.

2.4.3 Comparison of Majority Rule and Viterbi Decoding by Numerical Simulation

Now that we have described how Viterbi decoding can be used instead of the majority rule decoder employed in the experiment in section 2.3, we need to compare the two methods to determine if Viterbi decoding offers a significant improvement. In the experimental measurement of bit error rate, our sample contained only a few hundred independent message bits. If Viterbi decoding has any improvement on the majority rule decoding, we would not be able to experimentally measure the improved bit error rate simply because not enough bits are present to make a reasonable measurement. Instead, we use a numerical estimate.

Our numerical simulation uses scalar diffraction theory to calculate the Fourier plane intensity patterns. Normally, diffraction from objects with feature sizes less than the wavelength require vector diffraction theory for an accurate calculation. However, our main focus with this numerical simulation is the comparison of the two decoding methods. As long as the intensity patterns calculated with the simpler scalar diffraction maintain reasonable distances in relation to the message codewords, then any significant difference in the performance of the two decoding methods should generalize to the case of vector diffraction.

The simulation algorithm for measuring bit error rate as a function of signal to noise ratio (SNR) is shown in figure 2.5. The branch labels are first calculated with scalar diffraction and tabulated. A random number generator produces a message bit which is sent to the precoder. The precoder forms the appropriate channel symbol, and its channel output (Fourier plane measurement) is taken from the tabulated scalar diffraction calculation. White Gaussian noise is added to each of the eleven equally spaced points in the Fourier plane, thus forming the noisy channel data. Because the channel symbols do not have zero mean, the symbol energy used to determine the signal to noise ratio (SNR) must first have the mean subtracted. Our numerical simulation does not include jitter noise so that the assumption of perfect synchronization is satisfied. The decoder, whether the Viterbi algorithm or the majority rule algorithm, takes the noisy channel data and uses the stored branch labels to produce an output bit. Finally, the output bit is compared to the original message bit. Figure 2.6 plots the bit error rates as a function of SNR for both the majority rule algorithm used in the experiment and the adapted Viterbi algorithm as described above. We also included a plot for the raw symbol error rate. Specifically, we used the symbols for the Viterbi decoding scheme, and we counted a symbol error whenever the branch with the minimum distance to the noisy symbol was not the same as the symbol before noise was added.

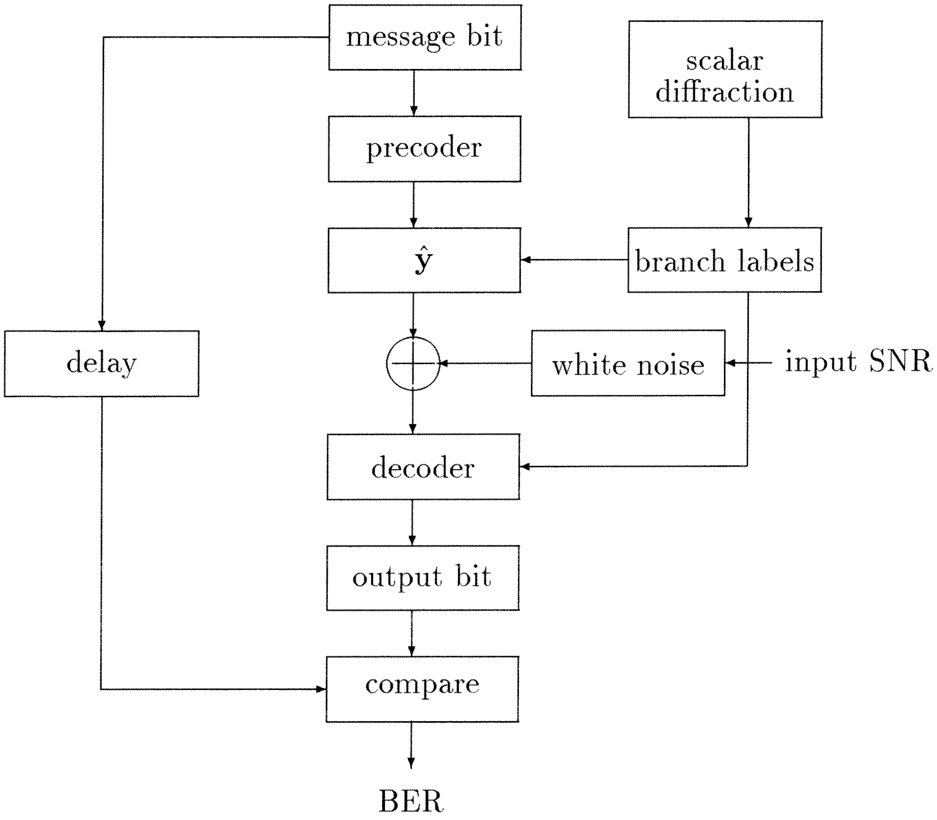


Figure 2.5: Simulating a noisy channel with Viterbi decoding or majority rule decoding

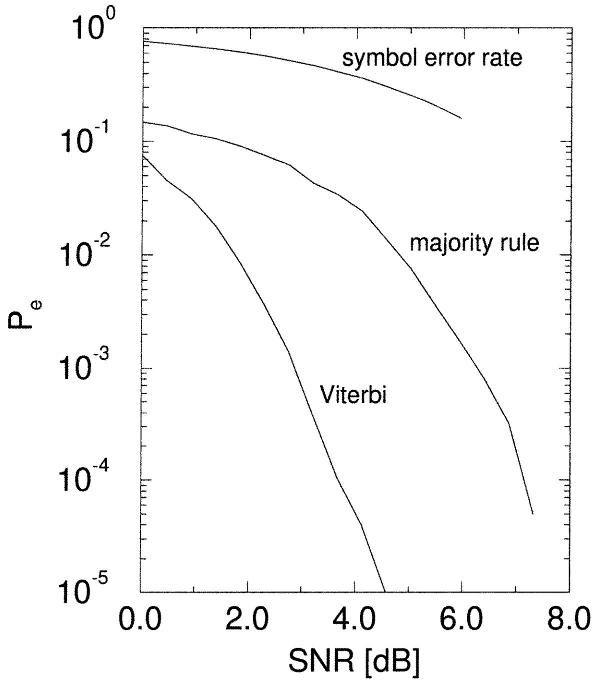


Figure 2.6: Comparison of the bit error rates for the Viterbi algorithm and the majority rule algorithm as well as the raw symbol error rate.

While our simulation demonstrates that Viterbi decoding can improve the decoding of unknown, unresolvable pit patterns, the simulation neglected several factors. The simulation assumes perfect synchronization between the channel bits and the measurements. If jitter noise were present, then there would be errors resulting from missing, repeated, or off center channel bits. In these cases, accounting for path history might actually become a liability, since a missing bit would have the effect of a state jump that has no corresponding branch. The noise that was added to the Fourier plane intensity profiles was independent from one point in the Fourier plane to another. This independence is mostly true for noise sources at the detectors, but for noise sources at the sample, for example, a dust spot covering a line, would have an effect across the Fourier plane. Thus, noise related to sample fabrication and the optical system would result in correlated noise in the Fourier plane measurement. One possible solution might be to add white noise to the channel symbol before using the scalar diffraction model, and then add more to the calculated Fourier plane intensity profile.

2.5 Conclusion

We started by stating the differences between imaging a completely unknown object, and deciding which pattern, from a finite set of possible patterns, is present. By making use of *a priori* information, patterns with features well beyond the resolution limit can be detected. First we demonstrated that simple measurements in the Fourier plane can distinguish between different patterns. We were then able to decode and make decisions about which patterns were present by making use of redundant measurements. Using a simple majority rule decoder, we were able to decode patterns of 0.2μm lines with a bit error rate of about 10^{-2}. Finally, we simulated the decision process to estimate the improvement that Viterbi decoding can offer over the majority rule algorithm.

Chapter 3 Surface Depth Measurement with Conoscopic Holography

3.1 Introduction

In the last chapter we considered the resolution of small features in the transverse direction across a sample. In this chapter we investigate the resolution of measurements of depth. Since we eventually want to measure the depth of small features, we will keep the same basic optical configuration of a beam focused through a high numerical aperture lens onto a sample. For this type of system, the depth of the illuminated region of the sample is best measured relative to the focal plane. Therefore, for the remainder of the chapter, we will refer to depth as the distance to the focal plane. The basic purpose of this investigation is to evaluate the possibility of using depth encoding in an optical memory.

There are several methods for measuring depth. One obvious method of measuring small changes in the depth of a reflecting surface is interferometry. In fact, we explained in the last chapter how we used an interferometer to accurately focus the incident spot. The basic problem with interferometry is that the fringes are ambiguous for depths greater than half a wavelength. Two ways to solve the ambiguity is to use two wavelengths or to use the information contained in the curvature of the fringes. The fringes will have curvature if the reflecting surface is not in the focal plane.

Another common system for measuring depth is the confocal microscope. In a confocal microscope, the reflected spot is imaged onto a pinhole. If the reflecting surface moves out of the focal plane, then the image of the focused spot at the pinhole will also move out of the pinhole plane, resulting in less light passing through the pinhole. Unless the pinhole is actively swept along the axis for each measurement, then simply measuring the intensity that passes through the pinhole is ambiguous, because the intensity alone does not determine whether the imaged spot is before or after the pinhole.

The method that we will consider in this chapter is conoscopic holography. Conoscopic holography has the advantage of being as accurate as confocal microscopy and much simpler to adapt to quick depth measurements. Also, conoscopic holography has no inherent ambiguity. In the next section, we describe conoscopic holography in detail. We then describe the optimal design of a pit depth memory with the restriction that the encoded pits are big enough to allow a simple depth measurement technique. In section 3.4 we merge conoscopic holography and the pit depth memory. In this section we also describe a significant feature of consocopic holography for depth measurement—the trade-off between unambiguous range and sensitivity is a simple design parameter. We tested our conoscopic measurement system with sample pits, and we describe the results at the end.

3.2 Conoscopic Holography

Conoscopic holography, invented by Sirat and Psaltis [10, 11], is similar to conventional holography and interferometry in the sense that it measures wavefront curvature. However, conoscopic holography works with the polarization of the light, and the light can be either coherent or incoherent. The basic configuration for conoscopic holography is shown in figure 3.1. Light emanating from a point source is polarized and passes through a birefringent crystal. Because the light from a point source has a spherical wavefront, the light passing through the crystal is comprised of a spectrum of planewaves. The electric field of each planewave is rotated by the crystal's birefringence, and the crystal's birefringence depends on the planewave's angle with respect to the crystal's optical axis. The analyzer after the crystal will pass only those planewaves whose polarization was rotated the right amount by the crystal. The resulting intensity on the screen consists of circular fringes, where the fringe spacing depends on the angular spectrum and distance to the point source, and the center of

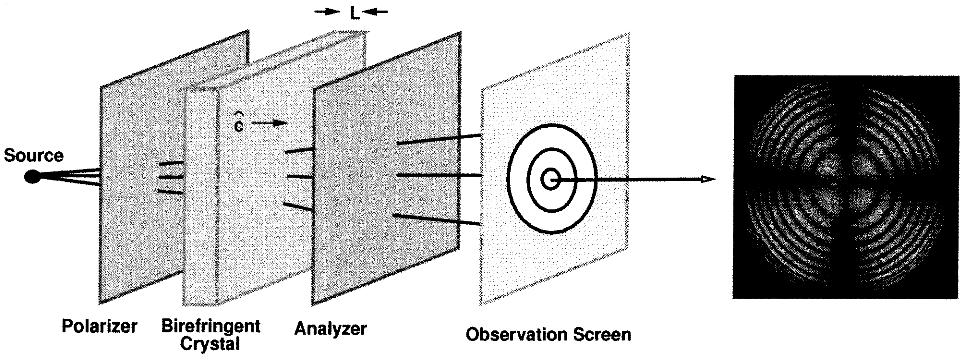


Figure 3.1: Basic conoscopic holography configuration.

the fringes corresponds to the lateral position of the point source.

Consider a single planewave emanating from the point source. The electric field of the planewave has both an E_x and an E_y component. Taking the optical axis of the crystal to be the z axis of the system, suppose that the direction of propagation of the planewave is $\cos(\theta)\hat{z} + \sin(\theta)\hat{y}$. Then, propagation through the crystal will cause a phase delay of approximately e^{jkn_oL} for E_x, and $e^{jkn_e(\theta)L}$ for E_y, where n_o and n_e are the ordinary and extraordinary index of refraction for the crystal. The extraordinary index of refraction is a function of the angle of propagation and is given approximately by [12]

$$n_e(\theta) \approx n_0 + \Delta n \sin^2(\theta). \tag{3.1}$$

If the polarizer is oriented to pass light with polarization 45° to the y axis, then "interference" between the E_x and E_y polarizations will result when the electric field is projected onto the analyzer,

$$\hat{p} \cdot \boldsymbol{E} = \frac{\hat{x} + \hat{y}}{\sqrt{2}} \cdot (E_x\hat{x} + E_y e^{jk\Delta nL\sin^2\theta}\hat{y}). \tag{3.2}$$

The intensity on the screen resulting from a point source located at the point, P, with intensity, $I(P)$, is then

$$I(\theta, P) = \frac{I(P)}{2}\Big(E_x^2 + E_y^2 + 2E_xE_y\cos(k\Delta nL\sin^2\theta)\Big). \tag{3.3}$$

The fringe spacing on the screen depends on the $\sin^2\theta$ term. If the angle θ is kept reasonably small, as is typically the case, then we can make the paraxial approximation

$$\sin^2\theta \approx [(x-x')^2 + (y-y')^2]/z^2. \tag{3.4}$$

We can now see that the fringe spacing on the screen has a strong dependence on z, the distance from the point source to the screen, and the center of the fringes corresponds to the lateral position of the point source. If more than one point source were present, then the intensity on the screen would be the sum of the fringes from each point source.

3.3 Design of a Pit Depth Memory

We wish to design the pit format to maximize the information density over the area of the memory. Obviously, we will achieve the maximum density when the pit area is as small as possible and the variation in pit depth is as large as possible. The amount of information that the depth of each pit can represent depends on the range of possible depths that the pit can have and the average uncertainty of the depth measurement. The uncertainty, or noise, that is inherently present when measuring the depth of a pit has several sources. Fabrication may have nonuniformities so that pits with the same nominal value will have different actual depths. Sources of noise for the measurement include all the traditional noises, such as laser noise and vibrations, as well as, for a spinning disc, jitter. If all these uncertanties are combined into one number, σ_z, then the number of discernible pit depths out of a maximum depth, L, is about L/σ_z. The amount of information carried by each pit, measured in bits, would then equal to $\log_2(L/\sigma_z)$.

The limitation to the maximum allowable pit depth comes from the depth of focus of the illuminating spot. When reading from an optical memory, laser light is focused by an objective lens onto each pit as the disc rotates. The width of the spot at the focal plane is approximately $\lambda/\mathcal{NA}$, where λ is the wavelength and $\mathcal{NA}$ is the objective's numerical aperture. The spot maintains this approximate width over the depth of focus, which is about $\lambda/2\mathcal{NA}^2$ [21]. However, the spot width becomes approximately proportional to z outside the depth of focus. If the focal plane corresponds to the middle pit depth level, then the spot width at the top— depth equal to zero—equals the spot size at the bottom—depth equal to maximum. The limitation on the maximum depth is then the depth at which the spot width becomes so wide that effects such as crosstalk with neighboring pits become a problem.

We have just explained two forces determining the maximum pit depth, L. The amount of information each pit can store is proportional to the logarithm of L and calls for a large L. In contrast, the limited depth of focus requires a limited L, and the spatial pit density, determined by the maximum spot width, is proportional to $1/L^2$. To calculate the optimum L, we plot the information density, $I = \log_2(L/\sigma_z)/w^2$ bits/μm^2, where w is the spot width for a defocus of $L/2$. (figure 3.2). The graph shows that for an objective with $\mathcal{NA} = 0.7$, the optimum L is about $L_{opt} \approx 0.6\lambda$, and that it does not strongly depend on σ_z.

In summary, if we are to improve the information density of an optical disc memory by employing pit depth encoding, then the range of available pit depths is limited. Using overly large pit depths would require increasing the lateral dimensions of the pit, causing a loss in pit density and overall information density. Given the limit on maximum pit depth, the measurement resolution, σ_z, becomes the crucial parameter. The uncertainty must be reduced, by good fabrication and a sensitive measurement technique, to the order of $L/4$ or $L/8$, without any increase in the pit width, for pit depth encoding to be a viable improvement.

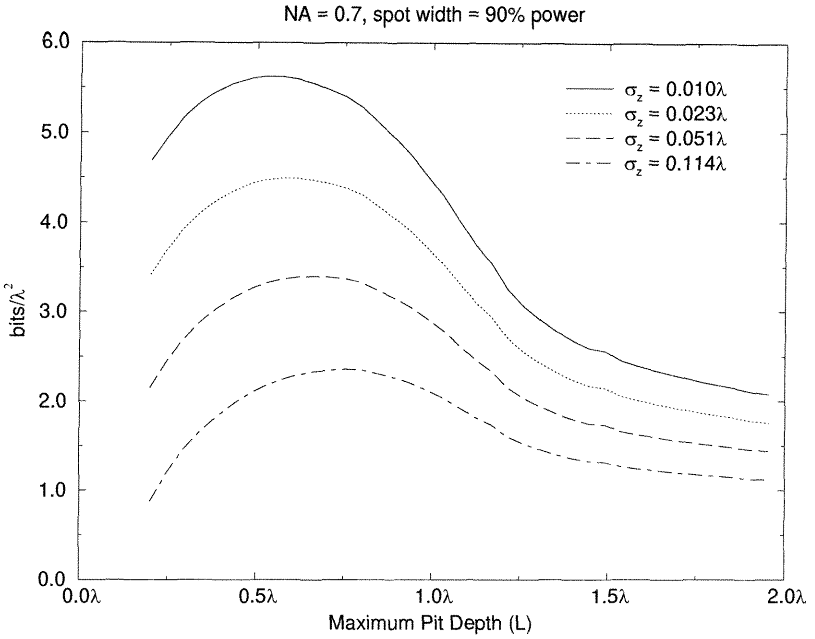


Figure 3.2: Information density as a function of the allowable pit depth, L.

3.4 System Design for a Conoscopic Pit Depth Memory

We now wish to adapt the conoscopic technique to measure the depth of pits in our optical memory. We will see that with conoscopic holography, we can design the system for specific sensitivity and unambiguous range of measurement. Another important advantage that this technique has over interferometry and confocal microscopy is that sensitivity to alignment and vibrations are reduced. For the pit depth memory, we are mainly interested in measuring the depth of a pit. Measurement of the lateral position of the pit is automatic from the rotation of the disc and the position of the objective lens along the track. We can assume that the current tracking techniques used on an optical disc can be used on our pit depth memory. If we consider the reflection of a focused spot from the bottom of a pit as the point source for a conoscopic system, then measuring the distance z gives the depth of the pit. If the objective is placed so that the reflecting surface of the pit is exactly in the focal plane of the lens, then the reflected light will be collimated by the lens. Variation in the pit depth will cause variation in the collimation of the return light, and this variation is what we will measure with the conoscopic system.

The placement of the birefringent crystal has three possibilities. It can be placed between the lens and the disc surface, in the collimated reflected light, or the lens itself can incorporate birefringent material. Placing the crystal in the path of the focused light has some severe disadvantages. First, propagation of the focusing light through the crystal will induce spherical aberration, increasing the size of the focused spot. The objective lens can be redesigned to account for the aberration, except that the aberration will be slightly different for each polarization. Further, the space between the objective lens and the disc is only a few millimeters. Since the length of the crystal directly affects the sensitivity of the measurement, we would like to use a crystal longer than a few millimeters. Finally, the sensitivity and unambiguous range of the measurement will be fixed by the numerical aperture of the objective lens. We would like to be able to adjust these parameters to fit our needs. We did not investigate the use of a lens made from a birefringent material because of the difficulty in obtaining such a lens. Perhaps in the future, with the development of birefringent diffractive optical elements, the use of a birefringent lens in conoscopic holography can be pursued. Figure 3.3 shows a tilted crystal placed in the path of the nearly collimated light reflected from the bottom of a pit. If the pit depth, z, is exactly equal to the focal plane of the objective lens, then the light propagating through the crystal is exactly collimated and no fringes will appear on the screen. By tilting the crystal, the angle θ in 3.3 becomes biased by the angle of the crystal. When the reflected light is collimated, the intensity on the screen oscillates from bright to dark as the crystal is tilted. Because the fringe spacing has a quadratic dependence on $\sin\theta$, the fringe spacing, or the rate of oscillation, becomes faster for greater tilts. Therefore, tilting the crystal increases the systems sensitivity to changes in the reflected light's degree of collimation. Since the degree of collimation depends on the pit depth in relation to the focal plane of the objective lens, the conoscopic system with the tilted crystal is an effective technique for measuring the pit depth.

Because speed is very important in an optical memory, we would like to keep the number of detectors small and the complexity of decoding limited. For example, if we

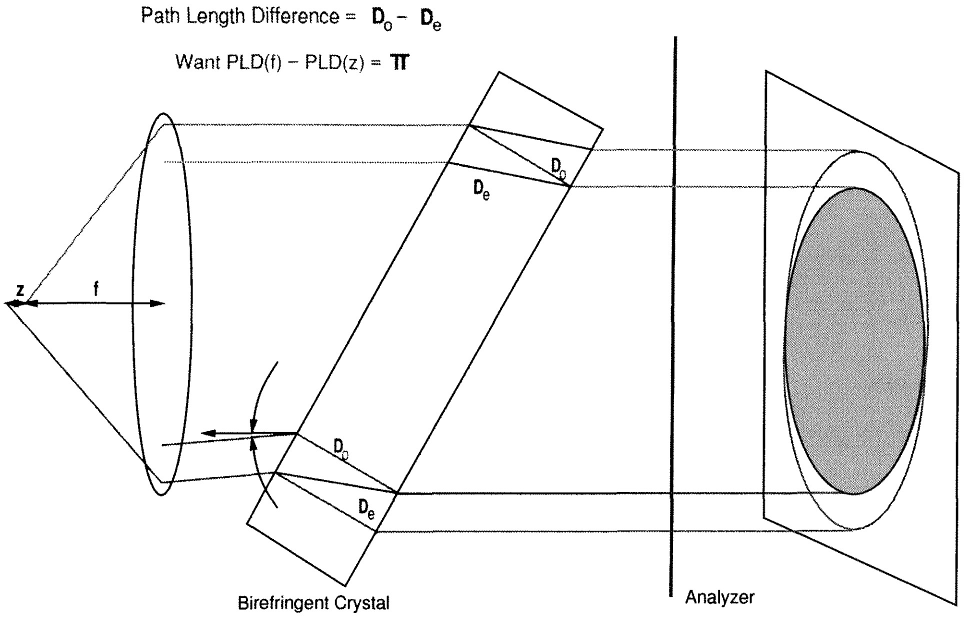


Figure 3.3: Conoscopic pit depth measurement with a tilted crystal.

were to design a system that counted fringes to determine the pit depth, then both the number of detectors would have to be large, and a time consuming algorithm would have to used. However, if the number of detectors is limited, then there will be ambiguous pit depths. The ambiguity results when the fringe spacing becomes smaller than twice the detector spacing. This aliasing will cause the detectors to measure the same intensity pattern as though the fringes were very large.

Our solution is to set the unambiguous range of measurement equal to the maximum range of pit depths. We can set the unambiguous range of measurement by adjusting the tilt of the crystal. As we have already mentioned, tilting the crystal increases the fringe sensitivity to changes in pit depth. Therefore, we need only to find the crystal tilt where a change from no fringes to one fringe corresponds to a change in the pit depth of half its maximum. The reason we specify a change in pit depth of half the maximum is because we are assuming that the focal plane of the objective lens is at half the maximum pit depth. In that case, the fringe caused by zero pit depth is 180° out of phase with the fringe caused by maximum pit depth. With this design where we use a single fringe for maximum pit depth, only two detectors are required in the output plane. The detectors can be connected to a differential amplifier so that the voltage is zero for a pit of half depth, positive for a pit of full depth, and negative for a pit of zero depth.

As is shown in figure 3.3, an optical path length difference, D, between the ordinary, D_o, and extraordinary, D_e, rays of half a wavelength will cause the formation of one fringe on the screen. From the figure, we know that the optical path length difference is

$$D = D_o - D_e = L\left(\frac{n_o}{\cos\phi_o} - \frac{n_e}{\cos\phi_e}\right), \tag{3.5}$$

where ϕ_o and ϕ_e are the angles the ordinary and extraordinary rays make with the crystal surface normal, and we have assumed that the optical axis of the crystal is perpendicular to the crystal surface. Snell's law relates the ray angles with the crystal tilt angle, ψ, so that D can be expressed as a function of the crystal tilt angle as

$$D = L\left(\frac{n_o^2}{\sqrt{n_o^2 - \sin^2\psi}} - \frac{n_e^2}{\sqrt{n_e^2 - \sin^2\psi}}\right). \tag{3.6}$$

By using Snell's law in equation 3.6, we have implicitly made the approximation that the direction of propagation inside the crystal is normal to the electric field, which is not correct for a birefringent crystal. However, as long as the angle ψ is kept small, this approximation yields useful results. When the point source (bottom surface of the pit) is exactly in the focal plane of the lens, then the angle ψ is constant over the whole crystal entrance. However, if the source is defocused, then the entrance angle depends on the lateral position along the crystal. We will call the angular deviation from collimation $\psi'(x, z)$, x being the lateral position and z the distance from the lens to the point source. With this notation the angle between a ray entering the crystal and the crystal normal is $\psi + \psi'(x, z)$. For small deviations from focus, the angular function is approximately,

$$\psi'(x,z) \approx x(\frac{1}{z} - \frac{1}{f}), \tag{3.7}$$

where f is the focal length of the lens. At the exit of the crystal, the ordinary and extraordinary rays that coincide did not enter the crystal at the same lateral position. The lateral shift, δx, can be calculated as a function of the crystal tilt from the geometry and Snell's law, and it is

$$\delta x \;=\; L \cos\psi \sin\psi \left(\frac{1}{\sqrt{n_0^2 - \sin^2\psi}} - \frac{1}{\sqrt{n_e^2 - \sin^2\psi}} \right). \tag{3.8}$$

The path length difference between the ordinary and extraordinary rays that coincide at the crystal exit can now be determined by substituting for ψ in equation 3.6

$$D(\psi; x, z) \;= \\ L\left(\frac{n_o^2}{\sqrt{n_o^2 - \sin^2(\psi + \psi'(x + \delta x, z))}} - \frac{n_e^2}{\sqrt{n_e^2 - \sin^2(\psi + \psi'(x, z))}} \right). \tag{3.9}$$

The ordinary and extraordinary rays that coincide at the crystal exit have slightly different propagation directions because they came from different lateral positions at the entrance, but we will ignore this difference so that the optical path length difference as a function of lateral position at the crystal exit is the same at the polarizer. As we have already discussed, the polarizer "interferes" the ordinary polarized light against the extraordinary polarized light by taking the projection of the electric field, preferably onto a direction 45° with respect to the ordinary or extraordinary directions. Since the ordinary light is phase delayed by $D(\psi; x)$ with respect to the extraordinary light, the intensity after the polarizer is proportional to

$$I \;\propto\; 1 + \cos(2\pi D(\psi; x, z)/\lambda). \tag{3.10}$$

Equations 3.7,3.9, and 3.10 completely determine the fringe spacing for various crystal

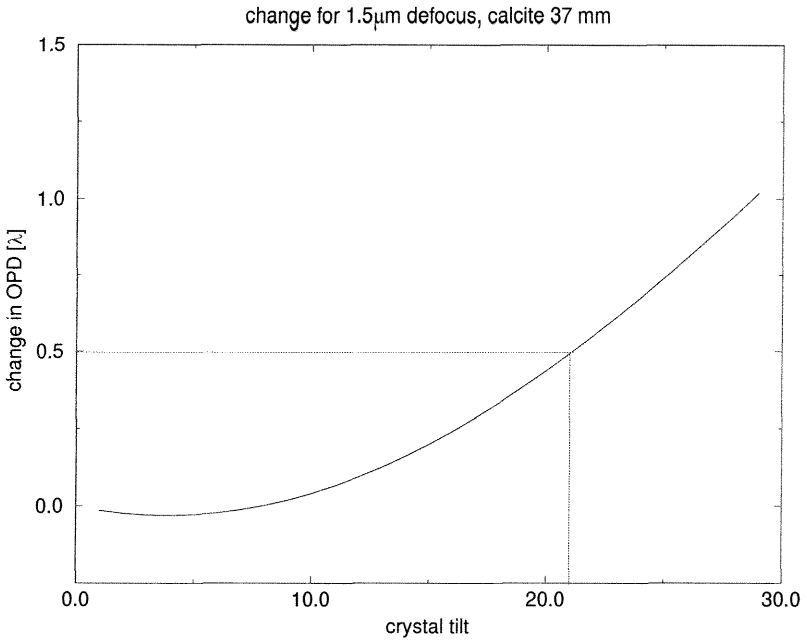


Figure 3.4: Optical path length difference as a function of tilt for a 37mm calcite crystal and a lens of $\mathcal{M} = 0.7$.

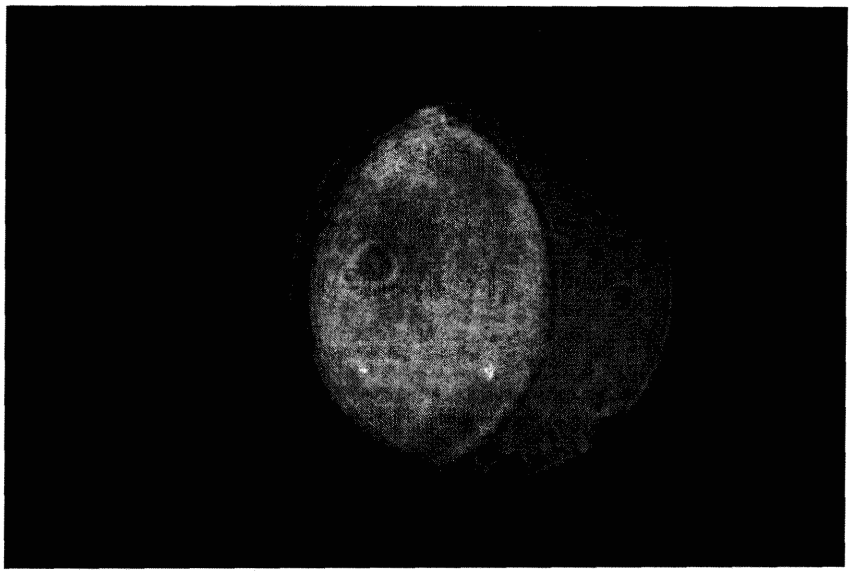
tilts and lens defocusing.

Our stated goal is to choose a crystal tilt that produces half a fringe when the defocus is half the maximum pit depth. The lateral extent (the maximum x value) is approximately the lens's focal length multiplied by its numerical aperture. To have half a fringe we require

$$D(\psi; x = f\mathcal{M}, z = f) - D(\psi; x = f\mathcal{M}, z = f + \ell) \;=\; \pm\lambda/2. \tag{3.11}$$

Figure 3.4 plots the above difference in optical path length as a function of crystal tilt, ψ, for L=37mm, $n_0 - n_e$=0.17 (calcite crystal), $\mathcal{M}$=0.7, and the maximum defocus, ℓ=1.5μm. From the graph we see that a crystal tilt of about 21° satisfies the condition 3.11. To verify this calculation we compared the image after the output polarizer for two cases (figure 3.5). The image, "21 degree tilt, focused," is the conoscopic image for a reflecting surface placed exactly in the focal plane of the objective lens, and the image, "21 degree tilt, defocused 1.5 micron," shows the resulting fringe when the reflecting surface is displaced from the focal plane by 1.5μm. Both figures

21 degree tilt, focused



21 degree tilt, defocused 1.5 micron

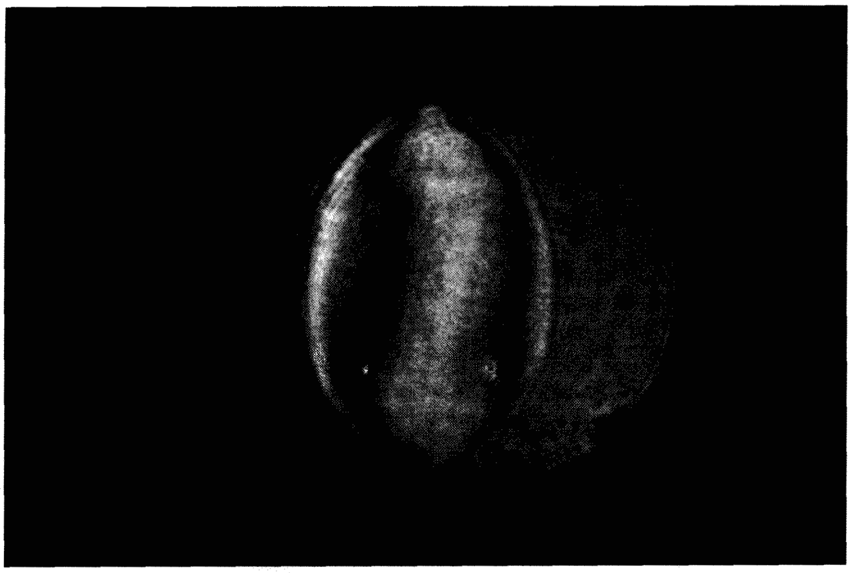


Figure 3.5: Measured conoscopic patterns

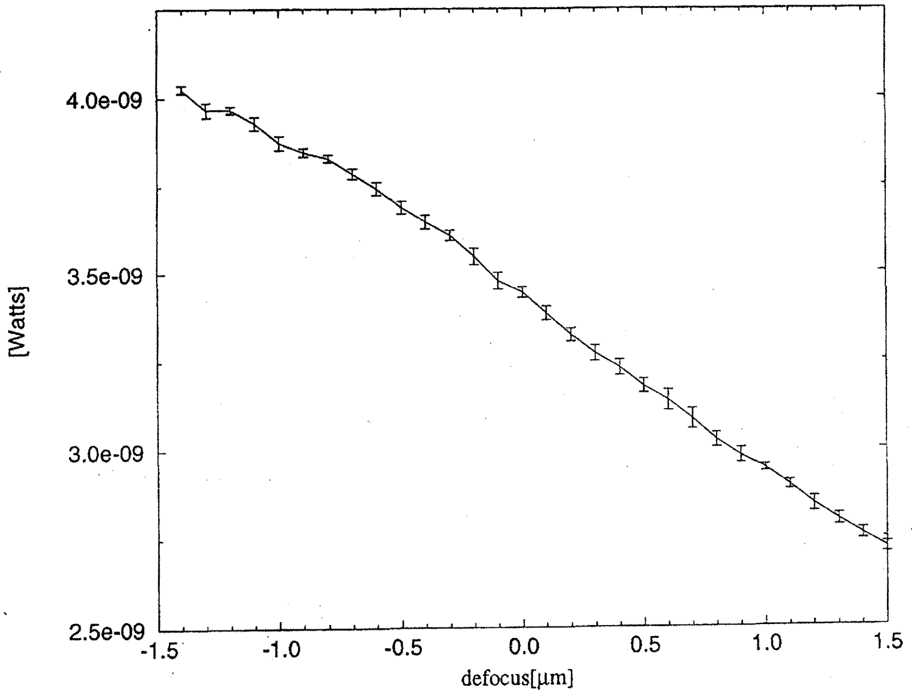


Figure 3.6: Linear Response

clearly show the lateral displacement between the ordinary and extraordinary beams, and the "interference" occurs only where they overlap. The image actually shows two fringes. Our calculation was for a point source movement of 1.5μm. However, because in our experiment we instead moved a reflective surface without changing the focus of the incident light, the corresponding point source movement is twice the movement of the reflective surface.

To measure the sensitivity to changes in z, we used one detector and an iris to block half the fringe. A reflecting planar surface was then stepped through the focal plane of an objective lens with $\mathcal{NA} = 0.7$. The steps were 0.1μm. Figure 3.6 shows the intensity at the detector as a function of defocus. The crystal tilt for this measurement was about 12°. The graph shows a relatively linear response over a 3μm range.

We now can design the conoscopic measurement to produce one fringe for a specified displacement (and the complimentary fringe for an equal displacement in the

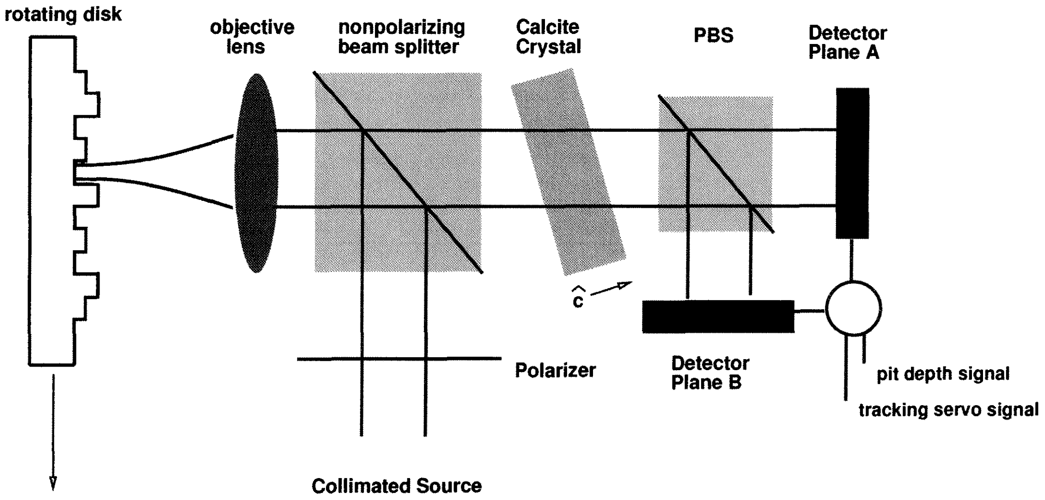


Figure 3.7: Conoscopic read head

opposite direction). The best way to measure the one fringe with a small number of detectors is to place two detectors symmetrically across the fringe and subtract their currents in a differential amplifier. Also, to utilize all the available light, a polarizing beam splitter should be used instead of the polarizer. Duplicate and complimentary detector planes are at the output of the beam splitter. Subtracting the signal from each detector pair gives a stronger conoscopic signal, and adding them gives a signal with the conoscopic part removed. The addition signal can be used for controls such as the tracking servo. The conoscopic read head for a pit depth encoded memory is shown in figure 3.7.

3.5 Pit Depth Measurement Results

To test the principle of the conoscopic read head with actual pits, we had a sample fabricated with patterns of pits with various widths and depths. The sample was fabricated at the Jet Propulsion Laboratory (JPL) using electron beam lithography.

Pits of varying depth were created in PMMA by modulating the electron beam exposure. The sample was then made reflective by evaporating about 200 Å of aluminum onto the surface. Figure 3.8 shows two scanning electron microscope (SEM) pictures taken from different regions of the sample.

Our measurements were made with a system similar to figure 3.7, except that we used only one detector and an iris in the detector plane so that only half a fringe would be visible. The wavelength was 0.6328μm, and the NA was 0.7. To measure the conoscopic signal as the staircase of pits passes under the spot, we first positioned a pit with half the full depth under the spot and set the objective lens focus and the crystal tilt. Figure 3.9a shows the conoscopic response as a staircase of 4μm pits passes underneath the spot. The response shows an overall linear dependence on pit depth. The fast oscillations are due to diffraction as the spot traverses the edges of the pits. In the SEM picture some roughness at the bottom of the deeper pits is evident, and diffraction from this roughness seems to disturb the conoscopic signal. We also performed the same measurement after removing the polarizer (figure 3.9b. Without the polarizer, we are simply measuring reflected intensity and no conoscopic signal is present. We know that the response measured in figure 3.9a is truly the conoscopic response because without the polarizer there is no signal present. The sharp drop that occurs at the same location in both responses is apparently diffraction due to a ridge that in the sample as a result of a small placement error during the lithography.

When we tried the same measurement with the 1μm wide pits, diffraction from the pits dominated over the conoscopic signal so that there was little difference between the conoscopic measurement and the measurement without the polarizer. Switching to an objective lens with a numerical aperture of 0.9 did not make any improvement.

3.6 Conclusion

Conoscopic holography is capable of measuring the depth of a surface with very high accuracy. In our first experiment, we translated a reflecting surface through a focused spot and measured the relative distance of the surface without the ambiguity that

JPL sample, 4 micron wide pits

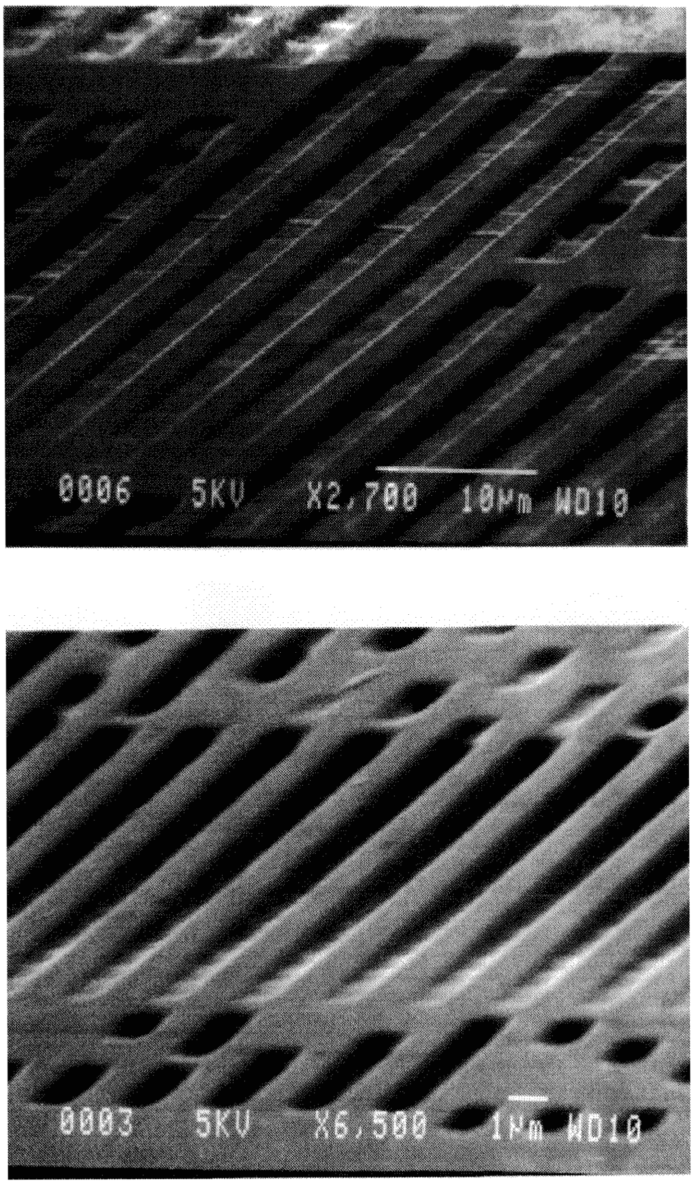


Figure 3.8: 4μm and 1μm wide staircases of pits in the JPL sample

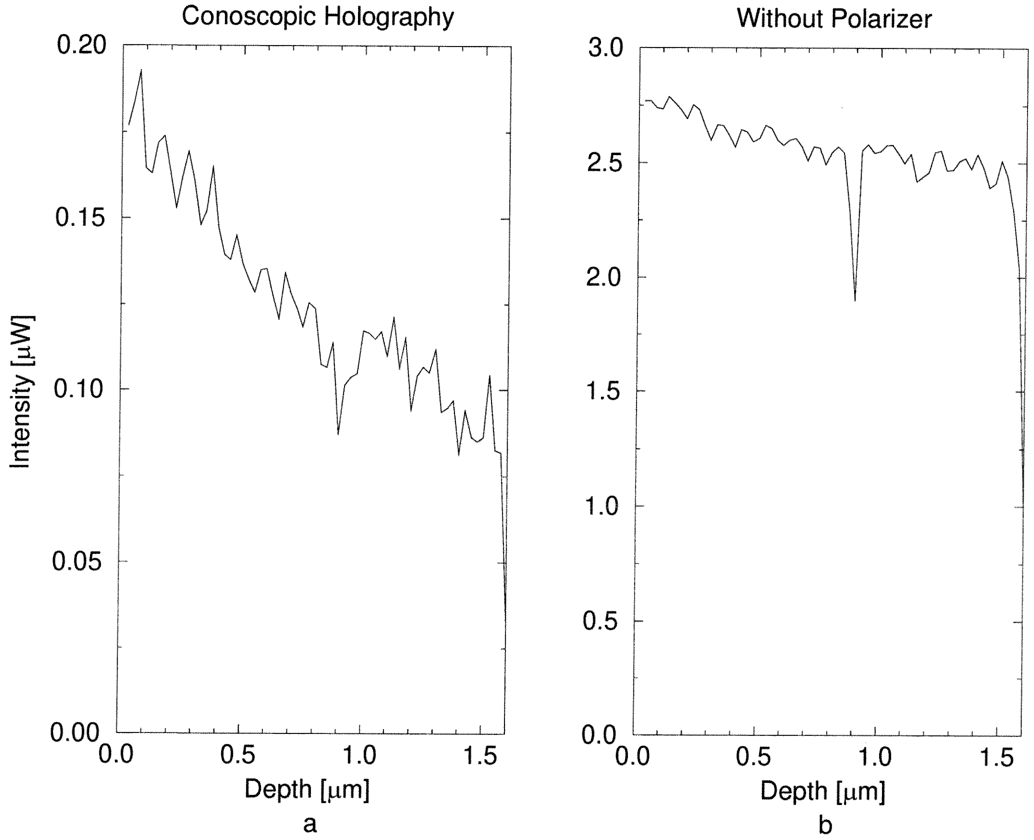


Figure 3.9: Detected light intensity versus pit depth a. conoscopic method b. polarizer removed from conoscopic head.

would be present in an interferometric system or a confocal system. The accuracy in depth of our basic conoscopic system was better than 0.1μm.

When we applied the conoscopic depth measuring system to an optical memory with pit depth encoding, we found that for the small pits of an ordinary optical memory, the conoscopic signal was lost to the effects of diffraction. We were successful in demonstrating the conoscopic system with pits that were 4μm wide. However, the information density with 4μm pits of 16 levels is only about 1 bit/μm^2. This density is not any real improvement over current state of the art. If we make the pit smaller, then diffraction effects ruin any direct measurement of the pit depth. Therefore, to make any improvement on the current optical disc storage density, even using pit depth encoding, we need a better model which includes accurate prediction of diffraction.

Chapter 4 Rigorous Diffraction of Focused Spots on Small Structures

4.1 Introduction

In Chapter 2 we learned that the far field diffraction of a subwavelength structure can be large enough to enable recovery of information about the structure, and the results from Chapter 3 illustrated the necessity of considering the diffraction if we are to improve data storage capacity of a planar medium. An accurate tool for modeling diffraction from subwavelength structures would be a useful tool to proceed further. With an accurate diffraction model, we would be able to design structures with desirable far field properties. For example, we could search for patterns that have far fields that are maximally separated in the manner of Chapter 2. Another use of a model is to explain the diffraction phenomena observed in Chapter 3. Other uses might be for the inspection of submicron features fabricated in the microelectronics industry and for the analysis of optical memory formats.

In this chapter, we will develop a method to numerically calculate diffraction for focused laser spots incident on subwavelength structures. Properties of a method that would be desirable are accuracy, simplicity, and computational efficiency. It must work for tightly focused spots with large numerical apertures, and nonperiodic, asymmetric surfaces. In the measurements described in Chapter 2, we found that significant differences existed for the TE and TM polarizations, especially for the deeper structures. Therefore, we cannot expect any scalar diffraction method to give satisfactory results. In fact, significant differences in the diffraction for TE and TM fields will be exploited in the next chapter.

There are several numerical vector diffraction methods described in various publications. We want to select the method that best suits our particular problem. Most numerical methods, however, are designed for periodic structures and planewave incidence. For example, the integral method invented by Petit and improved on by Wirgin [22, 24], the coupled-mode methods [41, 42], and the coupled-wave methods [34, 35, 36, 37] make these assumptions. Because these well-developed methods are not appropriate for our problem, we must take a step backwards and start with a more general approach. The two most general approaches to numerical solutions of Maxwell's equations are the integral approach and the differential approach. Examples of some differential approaches are finite differences, finite element analysis, and the more recent finite difference time domain (FDTD) method. In general, these methods are best when the problem is contained within finite boundaries, or when the fields on a boundary or initial conditions are known exactly. However, since we are working with incident fields with sources far from the structure, and the structure is not a perfect conductor—implying that our boundary conditions are continuity of the fields not the fields themselves—these types of differential methods would have to be adapted. The most natural approach seems to be the integral method as described in De Hoop [18].

In this chapter we will develop and test a numerical diffraction method based on coupled integral equations. Beginning with Maxwell's equations, the mathematical formulation derives expressions for the electric and magnetic fields inside a volume. Each expression involves a convolution of the tangential fields on the volume's surface with the electromagnetic Green's function. These same expressions are written a second time to use the tangential surface fields to calculate the fields outside the volume. After matching boundary conditions by equating the tangential components of the fields on either side of the surface (inside and outside the volume), coupled integral equations are formed for the tangential fields themselves. Writing a computer program to numerically solve the coupled integral equations is not simple, and we will make several choices in the implementation, which will be described. Any complicated numerical method needs to be fully tested for both mistakes in the formulation and overall numerical accuracy, and we will describe several methods of testing. In section 4.3 we will compare the convergence properties of our integral method to the

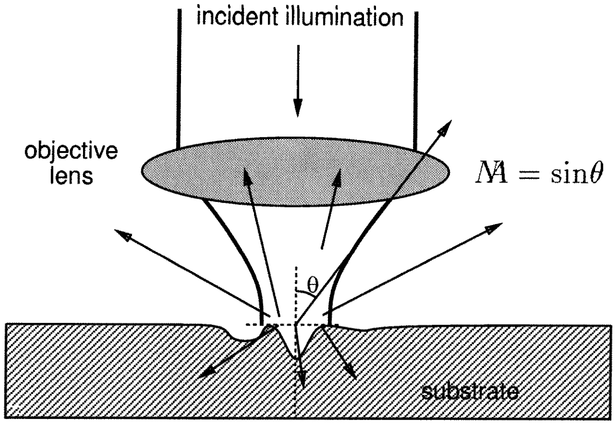


Figure 4.1: The basic diffraction model addressed in this chapter.

rigorous coupled-wave method[34]. We chose to compare against this method because of its popularity and simplicity. In the last section, we will compare results of our numerical method to experimental diffraction measurements for two cases.

4.2 Integral Method

Our diffraction model (figure 4.1) consists of a substrate surface with a known contour and an incident field resulting from a beam focused through an objective with a known numerical aperture. Most numerical diffraction approaches, such as the coupled-wave and coupled mode methods, require a periodic diffracting structure and a planewave incident field. For a focused incident spot on a single groove, these methods could be used by making the grating period large and separately accounting for each planewave that makes up the focused spot. However, the number of orders required in the diffraction calculation, because of the large period, will cause a burden on time and memory of the computation, and this computation time must be multiplied by the number of planewaves that are to be considered in composing the focused spot. Therefore, we decided to pursue a more direct integration technique which will allow us to calculate the diffraction directly from the focused incident field and a single groove.

Other alternatives to the integral method are finite differences, finite difference time domain, and finite element analysis. The advantages of these methods are that numerical integrations are not required, and, for finite differences, the formulation is somewhat simpler. The disadvantages of these methods are that grid points are placed throughout the whole volume, rather than just on the surface as in the integral method. With such a large number of grid points, the matrices become very large, although sparse. Other questions about these methods are the convergence properties and how to restrict the boundaries which are at infinity. In summary, however, one cannot decisively say that one method is better than all the rest for solving our class of problems, and our choice of the integral method is simply a choice.

We will formulate the numerical approach in two basic steps. First, we will derive the coupled integral equations for the fields in all space for the simple, general problem of an arbitrary incident field scattered from a single object. In the second part, we will make specific choices and shape the equations to numerically solve our specific type of diffraction problem. After describing our numerical algorithm, we will discuss various methods for testing the numerical results, and, in that last section, we will compare numerical results to experimental measurement.

4.2.1 Detailed Formulation

The formulation for the integral method of numerical diffraction is described in detail by de Hoop[18] and Pasman[19], and specific operator matrices for the two polarizations in two-dimensional diffraction problems are given in the appendix of Dil and Jacobs[17]. In the formulation presented here, we start with Maxwell's equations for the fields with an arbitrary geometry (figure 4.2), and solve for the fields in each region in terms of the tangential fields at the boundary between the two regions. The tangential fields can also be thought of as surface electric and magnetic currents, which act as sources for the scattered fields in each region. By integrating Maxwell's equations, we express the fields in each region with integral operators on the surface currents. The operators include convolution integrals of the surface currents with the

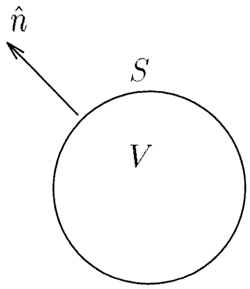


Figure 4.2: General Diffraction Geometry

free space Green's function. The boundary conditions are continuity of the tangential fields from one region to the other, and this condition is easily enforced with the surface currents. Finally, by setting the unknown fields expressed by the operators and the surface currents on the surface as well, we are left with two coupled integral equations.

Preliminaries

In the derivation, we need the following identity for the Fourier transform of the curl of a vector field. The Fourier transform of a vector field, $\boldsymbol{A}(\boldsymbol{r})$, is

$$\mathcal{F}\{\boldsymbol{A}(\boldsymbol{r})\} = \tilde{\boldsymbol{A}}(\boldsymbol{k}) = \iiint\limits_{V(\boldsymbol{r})} \boldsymbol{A}(\boldsymbol{r}) e^{-j\boldsymbol{k}\cdot\boldsymbol{r}}\, d\boldsymbol{r}. \tag{4.1}$$

Because we define the Fourier transform of the vector field only over a fixed volume, $V(\boldsymbol{r})$, we are implying that the inverse Fourier transform of $\tilde{\boldsymbol{A}}(\boldsymbol{k})$ returns a valid vector field only over $V(\boldsymbol{r})$. That is,

$$\mathcal{F}^{-1}\{\tilde{\boldsymbol{A}}(\boldsymbol{k})\} = (\frac{1}{2\pi})^3 \iiint\limits_{\text{all}\,\boldsymbol{k}} e^{j\boldsymbol{k}\cdot\boldsymbol{r}}\, \tilde{\boldsymbol{A}}(\boldsymbol{k})\, d\boldsymbol{k} = \begin{cases} \boldsymbol{A}(\boldsymbol{r}) & \boldsymbol{r} \in V \\ \frac{1}{2}\boldsymbol{A}(\boldsymbol{r}) & \boldsymbol{r} \in S \\ 0 & \text{elsewhere.} \end{cases} \tag{4.2}$$

We now form an identity for the Fourier transform of the curl of a vector by using the vector identity for the curl of a scalar times a vector along with Gauss's theorem.

First,

$$e^{-j\boldsymbol{k}\cdot\boldsymbol{r}}\,(\nabla\times\boldsymbol{A}) = \nabla\times(e^{-j\boldsymbol{k}\cdot\boldsymbol{r}}\,\boldsymbol{A}) - \nabla e^{-j\boldsymbol{k}\cdot\boldsymbol{r}}\times\boldsymbol{A}, \tag{4.3}$$

and then integrating and applying Gauss's theorem, we have the result,

$$\mathcal{F}(\nabla\times\boldsymbol{A}(\boldsymbol{r})) = \oiint_{S(\boldsymbol{r})} \hat{n}\times e^{-j\boldsymbol{k}\cdot\boldsymbol{r}}\,\boldsymbol{A}(\boldsymbol{r})\,d\boldsymbol{r} + j\boldsymbol{k}\times\tilde{\boldsymbol{A}}(\boldsymbol{k}), \tag{4.4}$$

where $\hat{n}$ is the unit normal vector on S pointing away from V.

Starting with Maxwell's Equations...

Starting with Maxwell's equations for time-harmonic ($e^{-j\omega t}$ dependence) fields in a linear, isotropic, homogeneous, and source free medium, we will move to the Fourier domain and solve for the fields in all space in terms of the tangential fields at the surface. At the end, we will take the inverse Fourier transform, using the convolution theorem. The result will be two coupled equations for the electric and magnetic fields in terms of integrals of the Green's function with the surface currents. In our formulation, the surface currents will act as sources so that once these currents are known, the fields at any point in space can be calculated by integrating over the sources with the appropriate Green's function. If the medium inside the volume were inhomogeneous, then diffraction would not only occur on the surface, but also inside the volume. In that case, we would require volume source currents as well as the surface currents. In the formulation in deHoop [18], inhomogeneous media are considered and the volume source currents are included. Our purpose is to consider the interface between two homogeneous media. For example, an optical disc memory consists of pits in an aluminum coated disc with a polycarbonate covering. Even though the aluminum covering is thin, aluminum is highly conductive so that the fields will not penetrate very deeply, and we can consider the volume to be filled with aluminum. The incident medium in this example is the polycarbonate. Because both the aluminum and polycarbonate are homogeneous, we do not need to consider the volume source currents, and our formulation will only use the surface currents.

We use, as a convention, a complex ϵ to account for induced current in the medium. If we let $\epsilon = \epsilon' + j\epsilon''$, then $-j\omega\epsilon'\boldsymbol{E} + \boldsymbol{J} = (-j\omega\epsilon' + \sigma)\boldsymbol{E} = -j\omega\epsilon\boldsymbol{E}$, with $\sigma = \omega\epsilon''$. Taking the Fourier transform of Maxwell's equations,

$$\nabla \times \boldsymbol{E} = j\omega\mu\boldsymbol{H} \tag{4.5a}$$

$$\nabla \times \boldsymbol{H} = -j\omega\epsilon\boldsymbol{E}, \tag{4.5b}$$

and using 4.4, we obtain

$$\mathcal{F}(\nabla \times \boldsymbol{E}) = \oiint_{S(\boldsymbol{r})} \hat{n} \times \boldsymbol{E}(\boldsymbol{r}_s)\, e^{-j\boldsymbol{k}\cdot\boldsymbol{r}_s}\, d\boldsymbol{r}_s + j\boldsymbol{k} \times \tilde{\boldsymbol{E}}(\boldsymbol{k}) = j\omega\mu\tilde{\boldsymbol{H}}(\boldsymbol{k}) \tag{4.6a}$$

$$\mathcal{F}(\nabla \times \boldsymbol{H}) = \oiint_{S(\boldsymbol{r})} \hat{n} \times \boldsymbol{H}(\boldsymbol{r}_s)\, e^{-j\boldsymbol{k}\cdot\boldsymbol{r}_s}\, d\boldsymbol{r}_s + j\boldsymbol{k} \times \tilde{\boldsymbol{H}}(\boldsymbol{k}) = -j\omega\epsilon\tilde{\boldsymbol{E}}(\boldsymbol{k}) \tag{4.6b}$$

We have used $\boldsymbol{r}_s$ to indicate the independent variable along the surface of the volume. The vector fields $\hat{n} \times \boldsymbol{E}(\boldsymbol{r}_s)$ and $\hat{n} \times \boldsymbol{H}(\boldsymbol{r}_s)$ are equal to the tangential components of $\boldsymbol{E}(\boldsymbol{r}_s)$ and $\boldsymbol{H}(\boldsymbol{r}_s)$, respectively, at the surface except that they point tangentially to the surface and perpendicularly to $\boldsymbol{E}(\boldsymbol{r}_s)$ and $\boldsymbol{H}(\boldsymbol{r}_s)$. These tangential components can be thought of as surface currents, and the boundary conditions require continuity of these surface currents. As a convenience, we define these surface currents as

$$\boldsymbol{J}_{s,m}(\boldsymbol{r}_s) = \hat{n} \times \boldsymbol{E}(\boldsymbol{r}_s) \tag{4.7a}$$

$$\boldsymbol{J}_{s,e}(\boldsymbol{r}_s) = -\hat{n} \times \boldsymbol{H}(\boldsymbol{r}_s). \tag{4.7b}$$

In equations 4.6a and 4.6b, the terms with surface integrals can now be written as Fourier transforms of the surface currents. Substituting the definitions into 4.6a and 4.6b, we obtain the coupled equations

$$\tilde{\boldsymbol{J}}_{s,m}(\boldsymbol{k}) = j\omega\mu\tilde{\boldsymbol{H}}(\boldsymbol{k}) - j\boldsymbol{k} \times \tilde{\boldsymbol{E}}(\boldsymbol{k}) \tag{4.8a}$$

$$-\tilde{\boldsymbol{J}}_{s,e}(\boldsymbol{k}) = -j\omega\epsilon\tilde{\boldsymbol{E}}(\boldsymbol{k}) - j\boldsymbol{k} \times \tilde{\boldsymbol{H}}(\boldsymbol{k}). \tag{4.8b}$$

Our next step is to solve equations 4.8a and 4.8b for $\tilde{\boldsymbol{E}}$ and $\tilde{\boldsymbol{H}}$ in terms $\tilde{\boldsymbol{J}}_{s,m}$ and $\tilde{\boldsymbol{J}}_{s,e}$. If we consider $\tilde{\boldsymbol{J}}_{s,m}$ and $\tilde{\boldsymbol{J}}_{s,e}$ as known, then 4.8a and 4.8b form two equations with two unknowns. We can write an equation for $\tilde{\boldsymbol{E}}$ ($\tilde{\boldsymbol{H}}$) by substituting 4.8a into 4.8b to eliminate $\tilde{\boldsymbol{H}}$ (4.8b into 4.8a to eliminate $\tilde{\boldsymbol{E}}$). Using the vector identity $j\boldsymbol{k} \times j\boldsymbol{k} \times \boldsymbol{A} = |\boldsymbol{k}|^2\boldsymbol{A} + j\boldsymbol{k}(j\boldsymbol{k} \cdot \boldsymbol{A})$ and the two additional relations,

$$j\boldsymbol{k} \cdot \tilde{\boldsymbol{H}} = (j\omega\mu)^{-1} j\boldsymbol{k} \cdot \tilde{\boldsymbol{J}}_{s,m} \tag{4.9a}$$

$$j\boldsymbol{k} \cdot \tilde{\boldsymbol{E}} = (j\omega\epsilon)^{-1} j\boldsymbol{k} \cdot \tilde{\boldsymbol{J}}_{s,e}, \tag{4.9b}$$

formed by taking $j\boldsymbol{k}\cdot$ equations 4.8a and 4.8b, we obtain expressions for $\tilde{\boldsymbol{E}}$ ($\tilde{\boldsymbol{H}}$) The resulting equations are

$$\tilde{\boldsymbol{E}}(\boldsymbol{k}) = \left[|\boldsymbol{k}|^2 - k^2\right]^{-1} \left[j\omega\mu(1 + \frac{j\boldsymbol{k}(j\boldsymbol{k}\cdot}{k^2})\tilde{\boldsymbol{J}}_{s,e}(\boldsymbol{k}) - j\boldsymbol{k} \times \tilde{\boldsymbol{J}}_{s,m}(\boldsymbol{k}) \right] \tag{4.10a}$$

$$\tilde{\boldsymbol{H}}(\boldsymbol{k}) = \left[|\boldsymbol{k}|^2 - k^2\right]^{-1} \left[j\omega\epsilon(1 + \frac{j\boldsymbol{k}(j\boldsymbol{k}\cdot}{k^2})\tilde{\boldsymbol{J}}_{s,m}(\boldsymbol{k}) + j\boldsymbol{k} \times \tilde{\boldsymbol{J}}_{s,e}(\boldsymbol{k}) \right]. \tag{4.10b}$$

In the above expressions, we have used the notation $k^2 = \omega^2\mu\epsilon$, which is a constant inside the volume and is not to be confused with $|\boldsymbol{k}|^2$, which is the squared magnitude of the independent variable in Fourier space.

The final step is to take the inverse Fourier transforms of equations 4.10, making use of the convolution theorem. As is shown in appendix A, the inverse Fourier transform of the first term in each of equations 4.10 is the Green's function for free space propagation:

$$\mathcal{F}^{-1}\left\{ \left[|\boldsymbol{k}|^2 - k^2\right]^{-1} \right\} = G(\boldsymbol{r}; \boldsymbol{r}') = \frac{e^{jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}, \tag{4.11}$$

where we have chosen as a convention that $\mathrm{Im}\{k\} > 0$. For convenience, we will notate the convolution operator as

$$G(\boldsymbol{r};\boldsymbol{r}_s) * \boldsymbol{A}(\boldsymbol{r}_s) = \oiint_{S(\boldsymbol{r}_s)} G(\boldsymbol{r};\boldsymbol{r}_s)\boldsymbol{A}(\boldsymbol{r}_s)\, d\boldsymbol{r}_s. \tag{4.12}$$

We are now ready to perform the inverse Fourier transforms on equations 4.10 to obtain our desired result for the fields in terms of the surface currents. Using the Fourier transform identities,

$$\mathcal{F}^{-1}\Big\{j\boldsymbol{k}(j\boldsymbol{k}\cdot\tilde{\boldsymbol{A}}(\boldsymbol{k}))\Big\} = \nabla\big(\nabla\cdot\boldsymbol{A}(\boldsymbol{r})\big), \tag{4.13}$$

and

$$\mathcal{F}^{-1}\Big\{j\boldsymbol{k}\times\tilde{\boldsymbol{A}}(\boldsymbol{k})\Big\} = \nabla\times\boldsymbol{A}(\boldsymbol{r}), \tag{4.14}$$

the inverse Fourier transform of equations 4.10 are

$$\begin{aligned}-\nabla\times\Big(G(\boldsymbol{r};\boldsymbol{r}_s) * \boldsymbol{J}_{s,m}(\boldsymbol{r}_s)\Big) + j\omega\mu(1+\frac{\nabla(\nabla\cdot}{k^2})\Big(G(\boldsymbol{r};\boldsymbol{r}_s) * \boldsymbol{J}_{s,e}(\boldsymbol{r}_s)\Big)&\\ = \begin{cases}\boldsymbol{E}(\boldsymbol{r}) & \boldsymbol{r}\in V\\ \frac{1}{2}\boldsymbol{E}(\boldsymbol{r}) & \boldsymbol{r}\in S\\ 0 & \text{elsewhere,}\end{cases}&\end{aligned} \tag{4.15a}$$

$$\begin{aligned}j\omega\epsilon(1+\frac{\nabla(\nabla\cdot}{k^2})\Big(G(\boldsymbol{r};\boldsymbol{r}_s) * \boldsymbol{J}_{s,m}(\boldsymbol{r}_s)\Big) + \nabla\times\Big(G(\boldsymbol{r};\boldsymbol{r}_s) * \boldsymbol{J}_{s,e}(\boldsymbol{r}_s)\Big)&\\ = \begin{cases}\boldsymbol{H}(\boldsymbol{r}) & \boldsymbol{r}\in V\\ \frac{1}{2}\boldsymbol{H}(\boldsymbol{r}) & \boldsymbol{r}\in S\\ 0 & \text{elsewhere.}\end{cases}&\end{aligned} \tag{4.15b}$$

In forming equations 4.15a and 4.15b, we cannot, in general, move the derivatives inside the convolution integrals because of the singularity of the Green's function. In fact, the singularity will only be of concern when $\boldsymbol{r}\in S$.

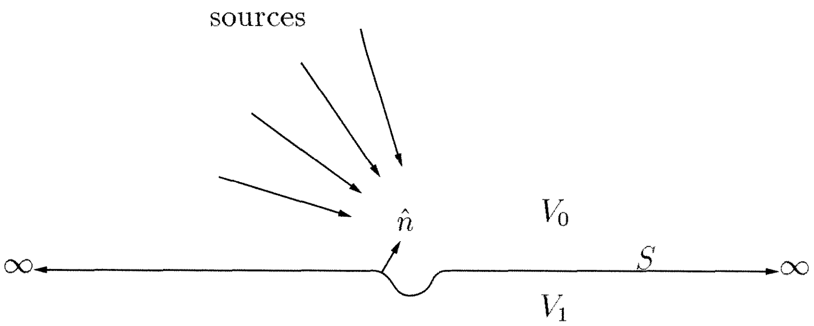


Figure 4.3: Diffracting surface with the incident field generated by distant sources

Field	description	source free in
$\mathbf{F}_1$	total field in V_1	V_1
$\mathbf{F}_0$	total field in V_0	
$\mathbf{F}_i$	incident field	V_1
$\mathbf{F}_s$	scattered field	V_0

Table 4.1: The fields and their source-free regions.

Matching Boundary Conditions

Our diffraction problem is for a boundary between two optically different, homogeneous media (figure 4.3). This geometry consists of two volumes, V_0 and V_1. Note that we place the sources of the incident field in V_0. In our initial statement of Maxwell's equations, we assumed a source-free region. Therefore, we can only write expressions for fields that are source-free in a particular region. Following de Hoop[18], we will work with the fields summarized in table 4.1. For each field, we will write an expression for it in the region where it is source-free. The boundary conditions require equality of the tangential components of the total fields in each region, $\mathbf{F}_0$ and $\mathbf{F}_1$, at the surface. However, note that $\mathbf{F}_0$ is not source free in any region. We can decompose $\mathbf{F}_0$ into its incident, $\mathbf{F}_i$, and scattered, $\mathbf{F}_s$, parts, and each of these components has a source free region. After equating the tangential components of $\mathbf{F}_0$ and $\mathbf{F}_1$ at the surface, the final result is a set of coupled integral equations for the unknown surface currents.

Equations 4.15 express the fields inside a volume in terms of operators on the tangential fields at a surface. Using operators, we can simplify the notation significantly. Let $\bar{\mathbf{\Gamma}}$ be the operator matrix, $\mathbf{J}$ be the vector of surface currents, and $\mathbf{F}$ be the vector of fields. Then,

$$\bar{\mathbf{\Gamma}} = \begin{bmatrix} -\nabla \times \Big(G(\boldsymbol{r};\boldsymbol{r}_s)* & j\omega\mu(1+\frac{\nabla^2}{k^2})\Big(G(\boldsymbol{r};\boldsymbol{r}_s)* \\ j\omega\epsilon(1+\frac{\nabla^2}{k^2})\Big(G(\boldsymbol{r};\boldsymbol{r}_s)* & \nabla \times \Big(G(\boldsymbol{r};\boldsymbol{r}_s)* \end{bmatrix} \tag{4.16}$$

$$\mathbf{J}(\boldsymbol{r}_s) = \begin{bmatrix} \boldsymbol{J}_{s,m}(\boldsymbol{r}_s) \\ \boldsymbol{J}_{s,e}(\boldsymbol{r}_s) \end{bmatrix}, \tag{4.17}$$

$$\mathbf{F}(\boldsymbol{r}) = \begin{bmatrix} \boldsymbol{E}(\boldsymbol{r}) \\ \boldsymbol{H}(\boldsymbol{r}) \end{bmatrix}, \tag{4.18}$$

and equations 4.15 become

$$\bar{\mathbf{\Gamma}}(\boldsymbol{r};\boldsymbol{r}_s)\mathbf{J}(\boldsymbol{r}_s) = \begin{cases} \mathbf{F}(\boldsymbol{r}) & \boldsymbol{r} \in V \\ \frac{1}{2}\mathbf{F}(\boldsymbol{r}) & \boldsymbol{r} \in S \\ 0 & \text{elsewhere.} \end{cases} \tag{4.19}$$

In region 0, $k_0 = \omega\sqrt{\mu_0\epsilon_0}$, $G_0(\boldsymbol{r};\boldsymbol{r}_s) = (1/4\pi|\boldsymbol{r}-\boldsymbol{r}_s|)e^{jk_0|\boldsymbol{r}-\boldsymbol{r}_s|}$, and $\bar{\mathbf{\Gamma}}^{(0)}$ is the corresponding operator. Similarly, $\bar{\mathbf{\Gamma}}^{(1)}$ is the operator for region 1.

We now have a convenient formula with which to express the fields in each region in terms of the tangential fields on the surface of our diffracting object. Solving first for the fields in V_1, we apply equation 4.19 and obtain the result,

$$\bar{\mathbf{\Gamma}}^{(1)}(\boldsymbol{r};\boldsymbol{r}_s)\mathbf{J}(\boldsymbol{r}_s) = \begin{cases} \mathbf{F}_1(\boldsymbol{r}) & \boldsymbol{r} \in V_1 \\ \frac{1}{2}\mathbf{F}_1(\boldsymbol{r}) & \boldsymbol{r} \in S \\ 0 & \boldsymbol{r} \in V_0. \end{cases} \tag{4.20}$$

The fields in region 0 can be separated into an incident field and a scattered field, $\mathbf{F}_0(\boldsymbol{r}) = \mathbf{F}_s(\boldsymbol{r}) + \mathbf{F}_i(\boldsymbol{r})$. By definition, the incident field has its sources somewhere in

V_0, and it is the field that would be present if the medium in region 1 were optically identical to the medium in region 0. The sources for the scattered field are the surface currents on S. In our formulation for the solution of the fields in a region (resulting in formula 4.19), we chose the form of Maxwell's equations for a source-free region. Therefore, we must be careful about how we apply the formula to the incident and scattered fields. Since the scattered field is source-free in V_0, we can solve for it there with

$$-\bar{\mathbf{\Gamma}}^{(0)}(\boldsymbol{r};\boldsymbol{r}_s)\mathbf{J}_s(\boldsymbol{r}_s) = \begin{cases} \mathbf{F}_s(\boldsymbol{r}) & \boldsymbol{r} \in V_0 \\ \frac{1}{2}\mathbf{F}_s(\boldsymbol{r}) & \boldsymbol{r} \in S \\ 0 & \boldsymbol{r} \in V_1. \end{cases} \tag{4.21}$$

The minus sign for the operator in the above formula indicates that we retain the convention that the surface normal points into V_0. The incident field, however, is not source–free in V_0, and so, we cannot directly write an expression for it there. Instead, we express the incident field in V_1,

$$\bar{\mathbf{\Gamma}}^{(0)}(\boldsymbol{r};\boldsymbol{r}_s)\mathbf{J}_i(\boldsymbol{r}_s) = \begin{cases} \mathbf{F}_i(\boldsymbol{r}) & \boldsymbol{r} \in V_1 \\ \frac{1}{2}\mathbf{F}_i(\boldsymbol{r}) & \boldsymbol{r} \in S \\ 0 & \boldsymbol{r} \in V_0. \end{cases} \tag{4.22}$$

We use the operator $\bar{\mathbf{\Gamma}}^{(0)}$ for the incident field in V_1 because we defined the incident field as the field that would be present in V_1 if this medium had no optical contrast with the medium in V_0. Subtracting 4.22 from 4.21, and substituting $\mathbf{F}_0 = \mathbf{F}_i + \mathbf{F}_s$, we obtain the relation,

$$-\bar{\mathbf{\Gamma}}^{(0)}(\boldsymbol{r};\boldsymbol{r}_s)\mathbf{J}(\boldsymbol{r}_s) = \begin{cases} \mathbf{F}_s(\boldsymbol{r}) & \boldsymbol{r} \in V_0 \\ \frac{1}{2}\mathbf{F}_0(\boldsymbol{r}) - \mathbf{F}_i(\boldsymbol{r}) & \boldsymbol{r} \in S \\ -\mathbf{F}_i(\boldsymbol{r}) & \boldsymbol{r} \in V_1. \end{cases} \tag{4.23}$$

We are now at a crossroads. Equations 4.20 and 4.23 make up a number of equations equal to four times the number of unknowns. Wirgin [24] and Pasman [19], when solving for diffraction from periodic gratings for planewave incidence, use an approach that avoids both explicit integral equations and integrals across singularities. In the third part of each of 4.20 and 4.23, the right-hand sides are known. Because the grating is periodic and only one planewave is incident, the fields and surface currents can be expanded into their Fourier components, and a simple matrix equation results for the solution of the coefficients. The last step is to use the coefficients in the forward matrix equation that results from the first parts of 4.20 and 4.23 to solve for the diffracted planewave coefficients. In addition, this whole process can be performed with either the electric fields or the magnetic fields (the second factor of two reduction in the number of required equations). To use this method to solve our problem, where the diffracting surface is not periodic and the incident field is not a planewave, three sampling grids would have to be generated: the surface S, a surface somewhere in V_1, and a surface somewhere in V_0. Then, basis functions for each grid would have to be chosen. Whether the cost of the added step is worth the benefit of avoiding the singularities is not clear. Perhaps this approach can be investigated in the future.

The approach followed by de Hoop[18] and Dil and Jacobs[17], and the one that we will follow here, is to concentrate on the fields at the surface for the right-hand sides of 4.20 and 4.23. Selecting the tangential components of the equations (multiplying the equations by $\hat{n}\times$), Fredholm integral equations of the first kind appear. The four integral equations, along with the boundary condition, need to be combined to form two equations with two unknowns, $\begin{bmatrix} \boldsymbol{J}_{s,m} \\ \boldsymbol{J}_{s,e} \end{bmatrix}$, and known right-hand sides, $\begin{bmatrix} \boldsymbol{J}^{(i)}_{s,m} \\ \boldsymbol{J}^{(i)}_{s,e} \end{bmatrix}$. The boundary condition, equality of the tangential components on the surface, are written as $\hat{n} \times \mathbf{F}_0 = \hat{n} \times \mathbf{F}_1 = \mathbf{J}$.

One solution is to select the tangential components of 4.20 and 4.23 and then add the two equations together (we will use $\mathbf{\Gamma} = \hat{n} \times \bar{\mathbf{\Gamma}}$):

$$(\mathbf{\Gamma}^{(1)} - \mathbf{\Gamma}^{(0)})\mathbf{J} = \frac{1}{2}\mathbf{J}_1 + \frac{1}{2}\mathbf{J}_0 - \mathbf{J}^{(i)}. \qquad (4.24)$$

Setting $\mathbf{J}_1 = \mathbf{J}_0 = \mathbf{J}$ (the boundary condition), we have our final coupled integral equations,

$$(\mathbf{\Gamma}^{(1)} - \mathbf{\Gamma}^{(0)} - \mathbf{1})\mathbf{J} = -\mathbf{J}^{(i)}, \tag{4.25}$$

for the unknown surface currents, **J**.

4.2.2 Numerical Solution of the Integral Equations

As is explained in [17, 18, 19, 26, 27], we numerically solve the integral equations in 4.25 by first approximating the surface currents with basis function expansions. As a result, each of the unknown surface currents in 4.25 becomes a vector of coefficients. The vector of basis functions is placed with the operators so that each of the four operators in 4.25 becomes a row vector of operators where each element includes a convolution of the Green's function and a basis function. If a basis function is chosen so that the expansion converges uniformly to the surface current, then the expansion itself introduces no error. However, the expansion must be truncated so that the numerical program can work with finite vectors, and this truncation introduces error.

The integrals in the operator matrix cannot be numerically evaluated until values for $\boldsymbol{r}$ are chosen. Choosing a value of $\boldsymbol{r}$ corresponds to choosing a point on the surface where we want to calculate the fields with the operator matrix. We choose values of $\boldsymbol{r}$ by representing the surface as a set of grid points. In the most general approach, rather than calculating the fields at specific grid points on the surface, weighting functions on the surface could be chosen and the fields calculated according to those weighting functions. Weighting functions made up of a set of evenly spaced delta functions is equivalent to choosing a set of grid points on the surface and calculating the fields at these grid points. Solving the fields only at a fixed set of grid points does not introduce error, except that the incident field on the right-hand side of the equation must also be sampled and truncated according to the weighting functions or grid points. Since each operator for each basis function must be numerically calculated for each grid point, the four operators in 4.25 now become four matrices. Each column of each matrix is for a particular basis function, and each row is for a specific grid point. De Hoop[18] and Poggio and Miller[27] give some discussion regarding various choices for basis functions and weighting functions.

We are now ready to tackle the computation of the diffracted fields. In the last section we formulated the integral equations in a manner independent of geometry and without any assumptions about the fields. From this point, we will make a series of choices which will tailor the numerical calculations to our specific needs. We are interested in tightly focused incident fields and single deep diffracting structures with a size on the order of the wavelength of the source. These structures are assumed to lie in an infinite half plane filled with a medium of complex refractive index n_1. After each simplification we will rewrite the operator matrices. Each operator matrix will be handled separately, with the labels $\mathbf{\Gamma}^{(1)} - \mathbf{\Gamma}^{(0)} = \begin{bmatrix} \mathbf{\Gamma}_{11} & \mathbf{\Gamma}_{12} \\ \mathbf{\Gamma}_{21} & \mathbf{\Gamma}_{22} \end{bmatrix}$.

Going to two dimensions: TE and TM

When the surface and all the fields are independent of a particular direction, say the $\hat{y}$ direction, then the integrals can be immediately performed along the y axis because only the Green's function has any dependence on y. The Green's function becomes

$$G(\boldsymbol{r};\boldsymbol{r}_s) = G(R) = \int_{-\infty}^{\infty} \frac{e^{jkR}}{4\pi R} dy = \frac{j}{4} H_0^{(1)}(kR), \tag{4.26}$$

which is the Hankel function of the first kind and order zero, and $R = |\boldsymbol{r} - \boldsymbol{r}_s|$. The remaining integration required to calculate each matrix element is a contour integral in the x-z plane. Since all the integrations are now contour integrations, the independent variable is better described as a scalar, s, which is the length along the contour starting from the point on the contour where $x = 0$ (see figure 4.4). The value of s at any point along the contour is calculated in the standard manner, $s(x) = \int_0^x (ds/dx)(x')\, dx'$. The unit vectors $\hat{\tau}$ and $\hat{n}$, which point tangentially and normally, respectively, to the contour, will ease the notation for the formulation in two dimensions. We have chosen the direction of $\hat{\tau}$ so that $\hat{\tau} \times \hat{n} = \hat{y}$.

Another consequence of the independence in the y direction is that the problem

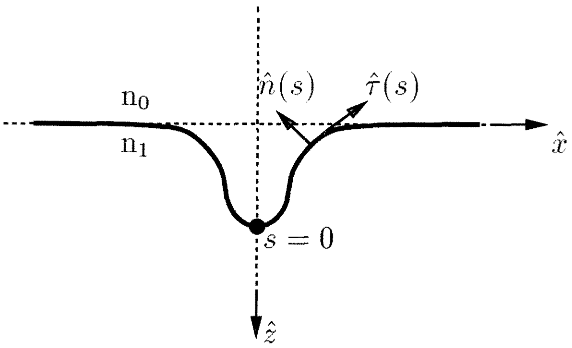


Figure 4.4: A diffracting structure and the coordinate systems for a two dimensional problem.

splits into two independent polarizations: $\boldsymbol{E} = E\hat{y}$, (TE) and $\boldsymbol{H} = H\hat{y}$ (TM). These polarizations are independent because an incident field with one of these polarizations will scatter only to fields of the same polarization. If we assume the TE polarization, for example, then the magnetic field has the direction $\boldsymbol{H} = H_\tau\hat{\tau} + H_n\hat{n}$, and the surface currents have the directions

$$\boldsymbol{J}_{s,m}(s) = \hat{n}(s) \times E(s)\hat{y} = E(s)\hat{\tau}(s) = J_{s,m}(s)\hat{\tau}(s) \tag{4.27a}$$

$$\boldsymbol{J}_{s,e}(s) = -\hat{n}(s) \times \Big(H_\tau(s)\hat{\tau}(s) + H_n(s)\hat{n}(s)\Big) = H_\tau(s)\hat{y} = J_{s,e}(s)\hat{y}. \tag{4.27b}$$

With the surface currents fixed to these directions and independent of y, the solution fields in equations 4.15 have directions consistent with the TE field. Therefore, the integral equations are completely self-consistent with only TE fields present. The principle is the same for the TM polarization.

For the reduction of the integral equations, we will assume the TE polarization. The results for the TM polarization are in appendix B. Knowing the specific directions of the surface currents, we can simplify the matrix of operators. Specifically, we see that in $\boldsymbol{\Gamma}_{12}$ there is a term,

$$\frac{1}{k^2}\nabla\Big(\nabla \cdot \Big(G(\boldsymbol{r}; s) * J_{s,e}(s)\hat{y}\Big)\Big). \tag{4.28}$$

Since we made the assumption that all functions are independent in y, this term must be zero. Substituting these surface currents into the matrix elements,

$$\mathbf{\Gamma}_{11}\boldsymbol{J}_{s,m}(\boldsymbol{r}_s) = -\hat{n}(s) \times \nabla \times \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) J_{s,m}(s')\hat{\tau}(s')\, ds' \tag{4.29a}$$

$$\mathbf{\Gamma}_{12}\boldsymbol{J}_{s,e}(\boldsymbol{r}_s) = j\omega\mu\hat{n}(s) \times \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) J_{s,e}(s')\hat{y}\, ds' \tag{4.29b}$$

$$\begin{aligned}\mathbf{\Gamma}_{21}\boldsymbol{J}_{s,m}(\boldsymbol{r}_s) = \frac{\hat{n}(s)\times}{j\omega\mu}\Big[&\int_{-\infty}^{\infty} \Big(k_1^2 G_1(R) - k_0^2 G_0(R)\Big) J_{s,m}(s')\hat{\tau}(s') ds' \\ &+ \nabla^2 \cdot \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) J_{s,m}(s')\hat{\tau}(s') ds'\Big]\end{aligned} \tag{4.29c}$$

$$\mathbf{\Gamma}_{22}\boldsymbol{J}_{s,e}(\boldsymbol{r}_s) = -\hat{n}(s) \times \nabla \times \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) J_{s,e}(s')\hat{y} ds', \tag{4.29d}$$

where we have assumed $\mu = \mu_1 = \mu_0$.

Choice of Basis Function

We now want to expand the surface current on a set of basis functions. In general, a function can be represented by the expansion,

$$f(x) = \sum_l a_l e_l(x), \tag{4.30}$$

where $e_l(x)$ are the basis functions, and a_l are coefficients. In [17, 19], the surface currents are assumed to be approximately constant over small segments, making the basis functions *rect* functions ($\mathrm{rect}(x) = 1$ when $|x| < \frac{1}{2}$, and 0 otherwise). This approach works well when the incident field has a small angular spectrum, such as a planewave. However, when the incident field is a tightly focused spot, then the number of grid points required to satisfactorily describe the incident field becomes very large, even when the diffracting surface is a flat plane. In our problem we know that the incident field has a large but finite angular spectrum. That the angular spectrum of the incident field is finite is guaranteed by the fact that the incident field comes from far away so that only propagating planewaves can reach the diffracting surface.

A good choice of basis functions might then be *sinc* functions ($\mathrm{sinc}(x) = \sin(x)/x$). The Nyquist sampling theorem states that a function of finite bandwidth (angular spectrum) is identically equal to a sum of evenly spaced samples weighted by *sinc* functions, if the sample rate is at least twice the highest frequency (planewave angle) of the function. That is, $f(x) = \sum_l f(l\pi/W)\,\mathrm{sinc}(\pi W x - \pi l)$, if W is greater than or equal to the bandwidth of $f(x)$. As a result, the incident field and surface currents can be perfectly represented by a set of *sinc* basis functions if the diffracting surface is a flat plane. For diffracting surfaces that are not a flat plane, however, the incident field is no longer a simple Fourier transform of its angular spectrum, and a sampling rate greater than the Nyquist frequency is normally required. This higher sampling rate corresponds to allowing for evanescent waves in other numerical approaches, such as coupled-wave. Intuitively, we would expect the sampled rate required for good results to increase as the surface becomes less like a flat surface, and our numerical results support this rule. The disadvantage of *sinc* basis functions is that the integral in each of the matrix elements now has infinite limits, whereas with *rect* basis functions, the integral is only over the region where the *rect* is nonzero. Compared to *rect* basis functions as used by [17, 19], *sinc* basis functions are superior for flat and nearly flat surfaces. However, for deep and narrow grooves in the diffracting surface, which basis function is better is not clear.

The expansion of the surface currents is of the form,

$$J_{s,m}(s) = \sum_{l=-N}^{N} a_l\,\mathrm{sinc}(k_0 W s - \pi l) \tag{4.31a}$$

$$J_{s,e}(s) = \sum_{l=-N}^{N} b_l\,\mathrm{sinc}(k_0 W s - \pi l). \tag{4.31b}$$

The unaliased bandwidth of the surface currents is determined by W. The coefficients, a_l and b_l, are samples of the surface currents at the grid points $s = \pi l/k_0 W$. We have truncated the expansion to $2N + 1$ terms. Substituting the basis functions into the matrix elements, and placing the coefficients in the vector of unknowns,

$$\mathbf{\Gamma}_{11_l} = -\hat{n}(s) \times \nabla \times \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) \operatorname{sinc}(k_0 W s' - \pi l) \hat{\tau}(s')\, ds' \tag{4.32a}$$

$$\mathbf{\Gamma}_{12_l} = j\omega\mu\, \hat{\tau}(s) \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) \operatorname{sinc}(k_0 W s' - \pi l)\, ds' \tag{4.32b}$$

$$\begin{aligned}\mathbf{\Gamma}_{21_l} = \frac{\hat{n}(s)\times}{j\omega\mu} \Big[&\int_{-\infty}^{\infty} \Big(k_1^2 G_1(R) - k_0^2 G_0(R)\Big) \operatorname{sinc}(k_0 W s' - \pi l)\, ds' \\ &+ \nabla^2 \cdot \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) \operatorname{sinc}(k_0 W s' - \pi l)\, ds' \Big] \hat{\tau}(s')\end{aligned} \tag{4.32c}$$

$$\mathbf{\Gamma}_{22_l} = \hat{y}\,(\hat{n}(s) \cdot \nabla) \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) \operatorname{sinc}(k_0 W s' - \pi l)\, ds', \tag{4.32d}$$

and the vector of unknowns is $\mathbf{J} = \begin{bmatrix} a_l \\ b_l \end{bmatrix}$.

Handling the Singularities

We now wish to resolve the derivatives in the matrix elements. The derivatives are with respect to the field point, $\boldsymbol{r}(s)$, not the independent variable of integration, s'. As we already mentioned, we cannot simply reverse the order of differentiation and integration because the integration includes the point, $s' = s$, where the Hankel function has a logarithmic singularity of the nature[29],

$$H_0^{(1)}(kR) \approx j\frac{2}{\pi}(\log(\frac{kR}{2}) + \gamma) \quad \text{as } R \to 0, \tag{4.33}$$

where γ is Euler's constant. Because each integrand contains the subtraction, $G_1(R) - G_0(R)$, the singularity cancels, and the integrand has the finite limit

$$\lim_{R \to 0} \Big[H_0^{(1)}(k_1 R) - H_0^{(1)}(k_0 R) \Big] = j\frac{2}{\pi}(\log(k_1) - \log(k_0)). \tag{4.34}$$

However, this cancelation does not occur for the first term of $\mathbf{\Gamma}_{21_l}$, where

$$\lim_{R\to 0}\left[k_1^2 H_0^{(1)}(k_1 R) - k_0^2 H_0^{(1)}(k_0 R)\right] \;=\; jk_1^2\log(\frac{k_1 R}{2}) - jk_0^2\log(\frac{k_0 R}{2}), \tag{4.35}$$

which has an integrable singularity at $R = 0$. Fortunately, this term does not involve a derivative. In the TM case, the term with the second derivative will have the integrable singularity. In that case, the form of the singularity must be subtracted from the integrand before differentiating. Subtracting singularities will be discussed in the next section.

To calculate the derivatives, we use the chain rule,

$$\nabla H_0^{(1)}(kR) = -kH_1^{(1)}(kR)(\nabla R), \tag{4.36}$$

$$\nabla R = \Big((x-x')\hat{x} + (z-z')\hat{z}\Big)/R. \tag{4.37}$$

Since the second term of $\mathbf{\Gamma}_{21_l}$ has two derivatives, we must now examine the nature of the singularity of $kH_1^{(1)}(kR)$. Fortunately, the subtraction in the integrand again causes the singularity to cancel:

$$\lim_{R\to 0}\left[k_1 H_1^{(1)}(k_1 R) - k_0 H_1^{(1)}(k_0 R)\right] \;=\; 0. \tag{4.38}$$

Taking the second derivative of this term will produce a singularity similar to the singularity in the first term of $\mathbf{\Gamma}_{21_l}$. We will discuss integrating across these singularities in the next section. The equations for the matrix elements are simplified by introducing the functions,

$$\mathcal{H}_0(R) = k_1^2 H_0^{(1)}(k_1 R) - k_0^2 H_0^{(1)}(k_0 R) \tag{4.39a}$$

$$\mathcal{H}_1(R) = k_1 H_1^{(1)}(k_1 R) - k_0 H_1^{(1)}(k_0 R) \tag{4.39b}$$

$$\mathcal{H}_2(R) = k_1^2 H_2^{(1)}(k_1 R) - k_0^2 H_2^{(1)}(k_0 R). \tag{4.39c}$$

Completing the derivatives in equations 4.32, they become

$$\mathbf{\Gamma}_{11_l} = \frac{j}{4}\hat{\tau}(s)\int_{-\infty}^{\infty} \mathcal{H}_1(R)\Big(\hat{n}(s')\cdot\nabla R\Big)\,\mathrm{sinc}(k_0 W s' - \pi l)\,ds' \tag{4.40a}$$

$$\mathbf{\Gamma}_{12_l} = \frac{-\omega\mu}{4}\,\hat{\tau}(s)\int_{-\infty}^{\infty} \Big(H_0^{(1)}(k_1 R) - H_0^{(1)}(k_0 R)\Big)\,\mathrm{sinc}(k_0 W s' - \pi l)\,ds' \tag{4.40b}$$

$$\begin{aligned}\mathbf{\Gamma}_{21_l} = \frac{-\hat{y}}{4j\omega\mu}\int_{-\infty}^{\infty}\Bigg[&\mathcal{H}_0(R)\Big(\hat{\tau}(s)\cdot\hat{\tau}(s')\Big)\\ &-\frac{1}{2}\Big(\hat{\tau}(s)\cdot\nabla R\Big)\Big(\hat{\tau}(s')\cdot\nabla R\Big)\Big(\mathcal{H}_0(R)-\mathcal{H}_2(R)\Big)\Big)\\ &-\hat{\tau}(s)\cdot\nabla\Big(\hat{\tau}(s')\cdot\nabla R\Big)\mathcal{H}_1(R)\Bigg]\,\mathrm{sinc}(k_0 W s' - \pi l)\,ds'\end{aligned} \tag{4.40c}$$

$$\mathbf{\Gamma}_{22_l} = \frac{-j}{4}\hat{y}\int_{-\infty}^{\infty} \mathcal{H}_1(R)\Big(\hat{n}(s)\cdot\nabla R\Big)\,\mathrm{sinc}(k_0 W s' - \pi l)\,ds'. \tag{4.40d}$$

With the aid of a few additional relations, all the integrands above can be numerically calculated. For the terms with $\hat{n}(s)\cdot$ and $\hat{\tau}(s)\cdot$, we need the relations,

$$\hat{n}(s) = \frac{\frac{dz}{dx}\hat{x} - \hat{z}}{ds/dx}, \tag{4.41}$$

and

$$\hat{t}(s) = \frac{\hat{x} + \frac{dz}{dx}\hat{z}}{ds/dx}. \tag{4.42}$$

Also,

$$\hat{\tau}(s)\cdot\nabla\Big(\hat{\tau}(s')\cdot\nabla R\Big) = \frac{1}{R}\Bigg[\hat{\tau}(s)\cdot\hat{\tau}(s') - \Big(\hat{\tau}(s')\cdot\nabla R\Big)\Big(\hat{\tau}(s)\cdot\nabla R\Big)\Bigg], \tag{4.43}$$

which is easily calculated. If the contour has a region where dz/dx becomes infinite (or large enough that numerical round off errors will cause problems), then in those regions, equations 4.41 and 4.42 must be replaced with forms using dx/dz and ds/dz. Finally, we will set the units of the electric and magnetic fields by choosing $\sqrt{\mu/\epsilon_0} = 1$.

This choice fixes $\omega\mu = k_0$. Equations 4.37, 4.39, 4.40, 4.41, 4.42, and 4.43 now give the completed recipe for calculating the integrands for any pair of s, s' on the contour.

Integrating along the Contour

Choosing as one of the grid points, $s_m = \pi m / k_0 W$, we calculate each matrix element for each m, l pair by numerically integrating the appropriate term from equation 4.40 along the contour. In principle any of the various numerical quadrature routines would work well. Since the grid points are evenly spaced in s, the simplest approach might be to use a standard Gaussian quadrature routine to integrate from grid point to grid point. With this idea, all the grid points, quadrature points, and their x, z, dz, ds, and Hankel function values can be tabulated at the start of the calculation. Obviously, the numerical integration cannot be performed to the infinite limits. However, since $H_0^{(1)}(kx)\,\mathrm{sinc}(kx) \propto (kx)^{-\frac{3}{2}} \sin(kx) e^{jkx}$ as $x \to 0$, the integration need not go too far before the error becomes small.

The last remaining hurdle, as we already mentioned, is integrating across the singularity in the matrix elements $\mathbf{\Gamma}_{21_{lm}}$. These singularities occur only when $m = l$. Otherwise, s_m falls on a null of the *sinc* function, and the *sinc* function goes to zero sufficiently fast to eliminate the singularity. One method to integrate those diagonal terms with the singularity is to subtract the form of the singularity and integrate it separately [25, 20]. The term becomes

$$\begin{aligned}\int_a^b \mathcal{H}_0(R)\,\mathrm{sinc}(k_0 W s' - \pi l)\,ds' = \int_a^b \Big[& \mathcal{H}_0(R)\,\mathrm{sinc}(k_0 W s' - \pi l) \\ & - j\frac{2}{\pi}\Big(k1^2 \log(\frac{k_1|s_l - s'|}{2}) - k_0^2 \log(\frac{k_0|s_l - s'|}{2})\Big)\Big]\,ds' \\ & + j\frac{2}{\pi}(b-a)\Big(k_1^2 \log(\frac{k_1(b-a)}{4}) - k_0^2 \log(\frac{k_0(b-a)}{4}) - (k_1^2 - k_0^2)\Big). \qquad (4.44)\end{aligned}$$

Outside the interval containing the singularity, $a < s' < b$, the subtraction is not necessary. In the formula, R_l indicates that R is calculated from the grid point s_l to s', and $s_l = s_m$ is the grid point for both the basis function and the source point.

The TM case is treated slightly differently because a singularity also appears in the term with the double derivative. The correct approach is to subtract the singularity before differentiating, and then differentiating the extra asymptotic terms along with the rest.

Incident Field

The right-hand side of the general matrix equation 4.25 is the tangential component of the incident field calculated at the grid points s_m. Many different models for the incident fields can be used. However, restrictions must be placed on the incident field if the numerical problem is to remain practical. As we mentioned in the formulation, one source of error is the truncation of the grid points. The truncation is only reasonable if the power flow across the surface at the missing grid points is negligible. Therefore, only incident fields with finite power concentrated over a relatively small region can be allowed. For example, planewaves cannot be handled with this method, but focused spots are permitted.

Figure 4.1 shows a lens focusing light onto the diffracting surface. A perfect lens transforms the light distribution at its front focal plane into a distribution of planewaves [16]. If $B(u)$ represents the amplitude and phase of the TE electric field at the front focal plane of the lens, where u is the transverse coordinate in this plane, then the incident field behind the lens is

$$\boldsymbol{E}^{(i)}(x,z) = \hat{y}\int B(u)e^{jk_0(ux+\sqrt{1-u^2}z)}\,du. \tag{4.45}$$

The incident magnetic field follows from Maxwell's equations,

$$\begin{aligned} H_x^{(i)}(x,z) &= \int B(u)\sqrt{1-u^2}e^{jk_0(ux+\sqrt{1-u^2}z)}\,du && \text{(4.46a)}\\ H_z^{(i)}(x,z) &= \int B(u)ue^{jk_0(ux+\sqrt{1-u^2}z)}\,du. && \text{(4.46b)} \end{aligned}$$

As long as x and z are sufficiently small, the integrals for the incident fields are easily calculated numerically. In most of our simulations, we used

$$B(u) = \begin{cases} 1 & |u| \le \mathcal{NA} \\ 0 & |u| > \mathcal{NA}, \end{cases} \tag{4.47}$$

where $\mathcal{NA}$ corresponds to the numerical aperture of the lens. The incident surface currents for the right-hand side of the matrix equation are then

$$\boldsymbol{J}_{s,m}^{(i)}(s) = \hat{n} \times \boldsymbol{E}^{(i)} = \hat{\tau}(s) E_y^{(i)}(x_s, z_s) \tag{4.48a}$$

$$\boldsymbol{J}_{s,e}^{(i)}(s) = -\hat{n} \times \boldsymbol{H}^{(i)} = \hat{y}\Big(n_x(s) H_z^{(i)}(x_s, z_s) - n_z(s) H_x^{(i)}(x_s, z_s)\Big). \tag{4.48b}$$

The incident fields for the TM case are completely analogous: $H_y(x, z)$ is calculated from the front focal plane distribution, and the electric field components are derived from the curl of $\boldsymbol{H}$.

Far Field

Once the matrix elements and the incident field vector are calculated, the matrix equation is solved using a standard numerical routine. The solution vector contains the coefficients a_l, b_l which can be substituted into 4.31 to find the actual surface currents anywhere on the contour. Unfortunately, the tangential fields at the contour cannot be directly measured. Useful results of this diffraction calculation would be the amplitude and phase in the far field, since these quantities can be experimentally measured. To calculate the far field, we use the first part of equation 4.23 to calculate the scattered electric field (magnetic field in the TM case). For this calculation, we use a far field approximation for the Hankel function, and we simplify the integrand in equation 4.23 by ignoring terms that have a dependence on $\boldsymbol{r}$ stronger than $1/\sqrt{|\boldsymbol{r}|}$. The electric far field can then be calculated by a simple numerical integration.

As a preliminary, we will first examine the Hankel function for large arguments:

$$H_n^{(1)}(kR) \approx \sqrt{\frac{2}{\pi k R}} e^{jkR} e^{j\frac{\pi}{4}(2n+1)} \quad \text{for } |kR| \gg |n|. \tag{4.49}$$

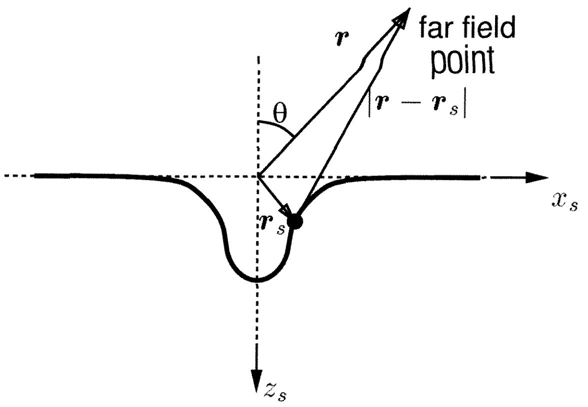


Figure 4.5: Coordinate system for the far field calculation.

In the far field (figure 4.5), $|\boldsymbol{r}| \gg |\boldsymbol{r}_s|$, so that using using cylindrical coordinates for $\boldsymbol{r}$,

$$R = \sqrt{(r\sin\theta - x_s)^2 + (-r\cos\theta - z_s)^2} \approx r - (x_s \sin\theta - z_s\cos\theta). \tag{4.50}$$

Substituting this value of R into the asymptotic form of the Hankel function, we obtain the Green's function for the far field,

$$G_0(\boldsymbol{r};\boldsymbol{r}_s) = H_0^{(1)}(k_0 R) \approx \sqrt{\frac{2}{\pi k_0 r}}\, e^{j\frac{\pi}{4}}\, e^{jkr}\, e^{-jk_0(x_s\sin\theta - z_s\cos\theta)}. \tag{4.51}$$

Equation 4.23 for the TE case is

$$\begin{aligned} -\boldsymbol{E}(\boldsymbol{r}) \;=\; & -\nabla\times\int_{-\infty}^{\infty} G_0(\boldsymbol{r};\boldsymbol{r}_{s'})J_{s,m}(s')\,\hat{\tau}(s')\,ds' \\ & + j\omega\mu_0\int_{-\infty}^{\infty} G_0(\boldsymbol{r};\boldsymbol{r}_{s'})J_{s,e}(s')\,\hat{y}\,ds'. \end{aligned} \tag{4.52}$$

Since the field point is now in the far field, the integration is not across a singularity, and the curl may be resolved directly:

$$-\nabla\times G_0(\boldsymbol{r};\boldsymbol{r}_{s'})\hat{\tau}(s') = \hat{y}\,\hat{n}(s')\cdot\nabla G_0(\boldsymbol{r};\boldsymbol{r}_{s'}) = (\hat{n}(s')\cdot\frac{\boldsymbol{r}}{|\boldsymbol{r}|})jk_0 G_0(\boldsymbol{r};\boldsymbol{r}_{s'}), \tag{4.53}$$

where we have ignored terms with $r^{-\frac{3}{2}}$ dependency. Making the appropriate substitutions and ignoring the $e^{jk_0 r}/\sqrt{r}$ factor, the electric far field is

$$-\boldsymbol{E}(\theta) = \hat{y}\, sqrt\frac{2k_0}{\pi}\, e^{j\frac{\pi}{4}} \int_{-\infty}^{\infty} \Big[\Big(n_x(s')\sin\theta - n_z(s')\cos\theta\Big) J_{s,m}(s') + J_{s,e}(s')\Big] e^{-jk_0(x(s')\sin\theta - z(s')\cos\theta)}\, ds'. \tag{4.54}$$

The integration in the above formula is easily performed numerically because all the grid points and surface currents have already been tabulated. Values for θ can be chosen arbitrarily, although values that lie within the objective lens are generally used. In some of our calculations we will use the far field intensity, which is proportional to the magnitude squared of the electric field, and in Chapter 5 we will also make use of the phase of the far field.

4.2.3 Testing the Program's Results

Before we can start using a computer program to predict diffraction behavior, we must have some confidence that the program is both numerically stable and an accurate representation of the mathematical formulation. The literature suggests several methods for testing the numerical results[17, 18, 22]. These methods include comparing numerical results against analytic solutions for the small class of known solutions, checking that the numerical solution satisfies the appropriate power conservation laws, testing reciprocity relations with the numerical method, testing the numerical results with the extinction cross-section theorem, and checking the convergence of the numerical results for decreasing grid spacing and increasing matrix rank.

We found that different checks were useful at different stages of forming the numerical computer program. For the problem of a planar interface, a problem for which analytic solutions are known, only the matrix elements of $\boldsymbol{\Gamma}_{12}$ and $\boldsymbol{\Gamma}_{21}$ are nonzero. We took advantage of this fact by using the planar interface problem to scrutinize the calculation of the matrix elements $\boldsymbol{\Gamma}_{21}$, the most complicated calculations and, therefore, the most likely to contain errors. Power conservation relations, discussed in more detail in the next section, proved useful when putting together the various components of the overall calculation. If, for example, a sign error were present in one of the incident fields, then power conservation would not follow. These two tests ensure that the computer program is an error free representation of the formulated equations. However, they do not, in general, test the numerical stability and accuracy of the computer program. The best method to evaluate the numerical stability of the computer program is simply to run the program repeatedly for the same problem with an increasing number of grid points. If the numerical solution converges in a reasonable manner, then the numerical program is stable, at least for the particular problem that was computed. The numerical accuracy of the solution can be estimated from the convergence data. The last check that we will discuss is a new check specifically for the the parameter W, the scale of the *sinc* functions, and, thus, the inverse of the grid spacing. This method uses the Fast Fourier Transform (FFT) of the surface currents to check the validity of the assumption that they are bandlimited to W.

Power Flow

Our power conservation rule is simply that the time averaged power flow transmitted across the contour boundary plus the power reflected from the boundary is equal to the power in the incident field. Unlike formulations that use a planewave as the incident field, our treatment can only use incident fields with finite power. In this section, we will first show that the power flow across a boundary can be calculated with the surface currents. Applying this formulation and the orthogonality of the *sinc* functions, we will derive the formulas for calculating the transmitted, reflected, and incident power from the surface current coefficients.

The time averaged electromagnetic power flow across a surface is found by integrating the normal component of the real part of the Poynting vector [33]:

$$< P > = \int_{-\infty}^{\infty} \mathrm{Re}\{\boldsymbol{E} \times \boldsymbol{H}^{*}\} \cdot \hat{n}\, ds. \tag{4.55}$$

Having formulated the problem in two dimensions, we are ignoring the integration in the $\hat{y}$ direction. However, the normal component of the Poynting vector can be represented using surface currents. From the definition of the surface currents,

$$\boldsymbol{J}_{s,m} \times \boldsymbol{J}^*_{s,e} = (-\hat{n})\Big[\hat{n} \cdot (\boldsymbol{E} \times \boldsymbol{H}^*)\Big]. \tag{4.56}$$

In our two-dimensional treatment, we already know the directions of the surface current vectors. For the TE case,

$$J_{s,m}\,\hat{\tau} \times J^*_{s,e}\,\hat{y} = -J_{s,m}J^*_{s,e}\,\hat{n}, \tag{4.57}$$

and the time averaged power flow across the surface is

$$< P > = \int_{-\infty}^{\infty} \mathrm{Re}\{J_{s,m}(s)J^*_{s,e}(s)\}\, ds. \tag{4.58}$$

Note that we have changed the sign of the integrand so that the power flow is from V_0 into V_1, opposite the direction of $\hat{n}$. Substituting the basis function expansions of the surface currents, equations 4.31, into the expression for the average power 4.58, we obtain the power flow in terms of the surface current coefficients,

$$< P > = \frac{\pi}{2W}\sum_l \mathrm{Re}\{a_l b_l^*\}. \tag{4.59}$$

This last formula results from the orthogonality relation,

$$\int_{-\infty}^{\infty} \mathrm{sinc}(k_0 W s' - \pi l)\,\mathrm{sinc}(k_0 W s' - \pi l')\, ds' = \frac{\pi}{2W}\,\delta_{ll'}, \tag{4.60}$$

where l and l' are integers and $\delta_{ll'} = 1 \iff l = l'$.

The power calculated in equation 4.59 gives the total power flow across the boundary, which is the transmitted power. To test the power conservation relation,

$$< P > = < P^{(i)} > - < P^{(r)} >, \tag{4.61}$$

we also need to calculate the incident power, $< P^{(i)} >$, and reflected power, $< P^{(r)} >$. The incident power is simply the power flow across the boundary for the case where there is no optical contrast between the two regions. In that case, $a_l = a_l^{(i)}$ and $b_l = b_l^{(i)}$. Substitution of these coefficients, already calculated as the right-hand side of the matrix equation, into 4.59 gives $< P^{(i)} >$. To calculate the reflected power, we must first separate the tangential fields into incident and scattered fields. As stated in the formulation of the integral equations, the total fields in V_0 can be separated into the incident fields and the reflected scattered fields. The boundary condition ensures that the surface currents are representations at the boundary of the tangential fields of V_0 as well as V_1. Therefore, we can also separate the surface currents into incident and reflected components:

$$\boldsymbol{J}_{s,m}(s) = \boldsymbol{J}_{s,m}^{(i)}(s) + \boldsymbol{J}_{s,m}^{(r)}(s) \qquad \boldsymbol{J}_{s,e}(s) = \boldsymbol{J}_{s,e}^{(i)}(s) + \boldsymbol{J}_{s,e}^{(r)}(s). \tag{4.62}$$

Expanding the surface currents for the reflected fields on the same *sinc* basis set, the orthogonality of the basis functions gives

$$a_l^{(r)} = a_l - a_l^{(i)} \qquad b_l^{(r)} = b_l - b_l^{(i)}. \tag{4.63}$$

The reflected power can now be calculated using the coefficients $a_l^{(r)}$ and $b_l^{(r)}$ in 4.59.

Convergence

Convergence of the numerical results means that if a calculation is repeated several times with a smaller grid spacing each time, then the numerical results will become more accurate with each repetition. In our case, we must test convergence for both decreasing grid spacing (increasing W), and increasing number of grid points with a fixed grid spacing. The latter test is because we do not have a periodic structure, and we approximated the integral to infinity by just integrating as far as we made grid points. Convergence testing is very important and must always be performed before trustworthy results can be obtained. In every numerical calculation, there is a trade-off between accuracy and computation time. By considering the convergence properties, a grid spacing can be chosen which gives satisfactory accuracy for a reasonable computation time. Naturally, the convergence estimation itself will require long computation time since the test must necessarily extend to grid spacing smaller than required. Also, the test must be carried to grid spacings small enough to determine if the results are oscillating or converging uniformly. Finally, the numerical results for the surface currents and the far field will converge at different rates for different contour points or far field angles, depending on the particular contour.

Checking the Sample Rate with the FFT

The choice of *sinc* functions for the basis functions of the surface current allows for a new type of convergence check. As we stated earlier, expanding a function on a basis set of *sinc* functions makes the implicit assumption that the function is strictly bandlimited. Also, the coefficients, a_l, b_l, are samples of the surface currents at the points $s = \pi l/k_0 W$. If the surface currents are in fact bandlimited, then their Fourier transforms, which can be calculated since we know the functions exactly everywhere, should show a cutoff frequency. If we calculate the Fast Fourier Transforms (FFT) of the surface current coefficients, the highest frequency bin will correspond to the cutoff frequency of the bandlimited functions. Therefore, if the FFT's of the surface current coefficients have significant amplitudes in the high frequency bins, then we can reasonably assume that there is aliasing and that the grid spacing needs to be shortened (the bandwidth, W, needs to be increased).

Example: Single Groove in Silicon, TE Illumination

This example problem simulates measurement of the depth of a groove etched in silicon. In semiconductor fabrication, single grooves and pits with a lateral dimension as small as a quarter micron are etched one micron or more into a substrate, usually silicon. Conceivably, a useful quality control measurement would be to measure the diffraction of a spot of laser light focused onto the groove or pit. The measured diffraction could then be compared to a "perfect" example and a quality determination

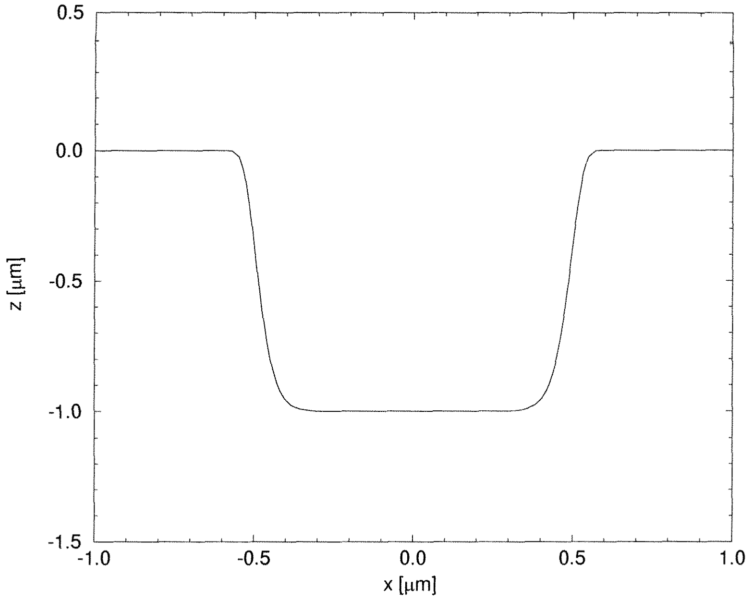


Figure 4.6: Groove contour for $a = 1$, $\sigma = 0.5$, and $\nu = 14$.

made. The sensitivity of this type of measurement could be numerically estimated by calculating the changes in diffraction patterns for grooves with varying parameters.

For the purpose of illustrating the convergence properties of the numerical method, we chose a single groove 1 μm deep and 1 μm wide. An objective lens with a numerical aperture of 0.6 focuses light onto the groove and captures the reflected far field with the same aperture. The wavelength of the light is 0.6328μm, and the index of refraction of silicon at this wavelength is approximately $3.88 + j0.02$ [30]. Because our integral method does have a requirement that the boundary have a continuous first derivative, we cannot use a rectangular groove with sharp corners. We used a super-gaussian profile which is of the form, $z(x) = a\exp(-(x/\sigma)^{\nu})$. Figure 4.6 shows the groove shape, and th contour was truncated at $s = \pm 4$ μm for these calculations.

Table 4.2 shows the convergence for the reflected, transmitted, and total power for an increasing number of grid points. Notice that for the calculation with 83 grid points, the reflected and transmitted power values are both nonphysical, but the sum is actually close to one. Also, when more grid points are used, R and T oscillate as they converge. As an absolute limitation to the value of using the numerical program to predict the far field intensity, we have plotted the successive change to the far field

# grid points	R	T	R+T
83	-3.171	4.1515	0.9802
99	0.2984	0.7019	1.0002
115	0.3006	0.6996	1.0002
131	0.2998	0.6993	0.9991
147	0.3000	0.6994	0.9994
165	0.3005	0.6991	0.9997

Table 4.2: Power conservation test for the integral method converges for a single groove in silicon.

calculation as the numerical solution converges (figure 4.7). The one graph that is very different from the rest is for the trial with 83 grid points. The other far field plots converge, but not uniformly for all the points. Notice the strong oscillation of the far field as $\sin\theta$, approaches the numerical aperture of 0.6. These oscillations are similar to the Gibbs phenomenon for truncated Fourier transforms. They appear here because of the similarity between the far field calculation (equation 4.54) and a Fourier transform. Also, the integration for the far field calculation is truncated as the contour is truncated. The Gibbs phenomenon oscillations can be reduced by weighting the contour integration for the far field calculation with an appropriate window function.

The final check is the FFT of the surface currents. Figure 4.8 shows the amplitude of the FFT of $\boldsymbol{J}_{s,m}$ for two different values of the grid spacing, W. A sample rate of $W = 5$ corresponds to 83 grid points, and $W = 10$ corresponds to 165 grid points. The highest contour spatial frequency for the first case is $5k_0$, and the FFT shows large amplitudes there. Therefore, the surface currents are highly aliased with this sample rate. However, for the greater sample rate of $10k_0$, the amplitude of the FFT goes to zero at the higher contour spatial frequencies. The result of aliasing brings energy that should be at higher frequencies into the lower contour spatial frequencies, distorting the calculated surface currents. This principle explains why the far field calculation for 83 grid points was significantly different in shape from the others.

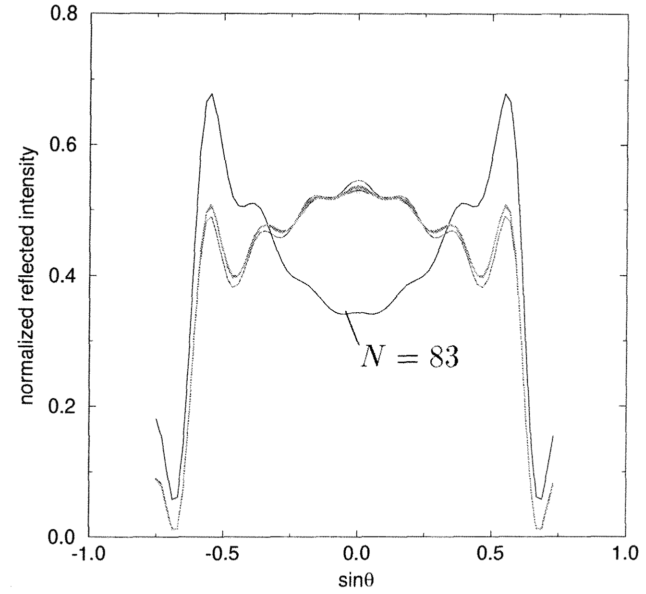


Figure 4.7: Convergence of the far field for the integral method. The plot marked $N = 83$ is for the calculation using 83 grid points.

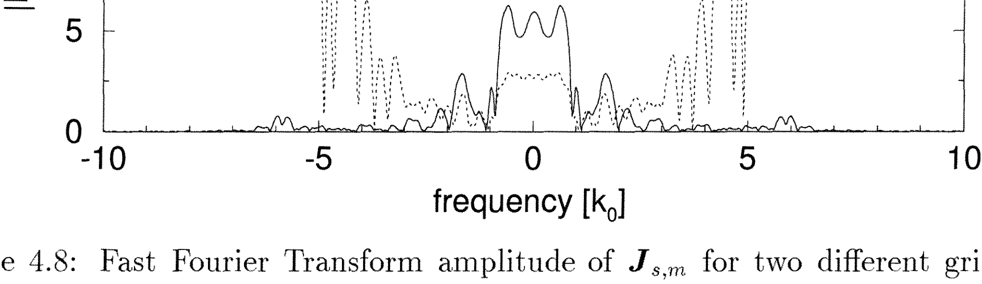


Figure 4.8: Fast Fourier Transform amplitude of $\boldsymbol{J}_{s,m}$ for two different grid spacings (sample rate). The sample rate of $5k_0$ corresponds to 83 grid points, and $10k_0$ corresponds to 165 grid points. The first is severely aliased, while the latter is not.

4.3 Comparison to Rigorous Coupled-Wave Analysis

To test the merits of our integral method approach, and to give further insight as to its application, we will compare it to the Rigorous Coupled Wave Analysis (RCWA). We chose RCWA as a basis for comparison because RCWA is very popular in the literature, and it has many advantages over other, older methods for calculating diffraction. RCWA's strengths, aside from being a completely rigorous method, are its versatility and simplicity. However, RCWA is not quite suited to the diffraction problems that we are concerned with. It requires a periodic structure, preferably with a period no larger than a few wavelengths, and a planewave incident field. Because we are concerned with diffraction from single structures, and we use a tightly focused spot as the incident field, we would have to make adaptations to both the statement of our diffraction problem and the RCWA method to obtain reasonable results. The nature of these adaptations will be discussed in section 4.3.2. In the next section, we will describe the method, although we will not give the complete mathematical formulation. We will also discuss the types of diffraction problems that RCWA is best suited for and other general features of RCWA that are instructive. Our goal is to compare numerical results from our integral method program to results from the RCWA for similar diffraction problems. Two specific examples will be discussed, and we will see that in our limited class of problems, our integral method will prove more reliable than the RCWA.

4.3.1 Overview of RCWA

The detailed formulation for RCWA was first published by Moharam and Gaylord [34, 35], and our favorite treatment was published by the binary optics group at the Massachusetts Institute of Technology Lincoln Laboratory [37]. They also supplied us with the FORTRAN source code for their coupled-wave program. In this section we will attempt to explain the method without a rigorous formulation. Our intention

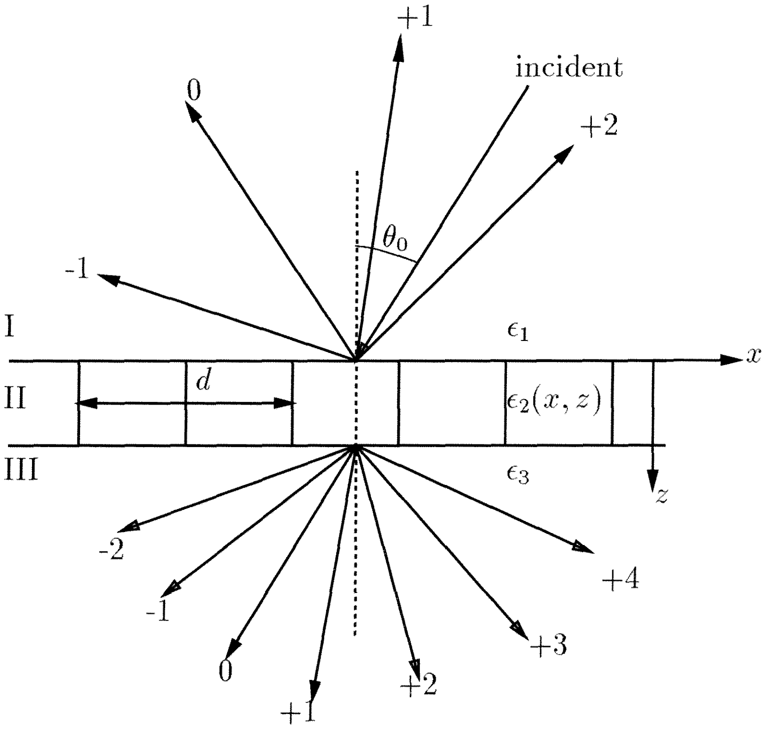


Figure 4.9: Diffraction geometry for the RCWA method.

is to give the reader who is unfamiliar with the method enough understanding so that the following discussions and comparisons can be understood. The diffraction geometry of RCWA is shown in figure 4.9. The figure shows three regions: the incident field is in region I; the grating is in region II; and the transmitted fields are in the substrate region III. In general, the grating region can be inhomogeneous in the z direction as well, so that the permittivity is a function of both x and z, $\epsilon_2(x,z)$. In that case, the region would be approximated by a stack of layers, where for each layer the permittivity is constant in the z direction. Naturally, the permittivity of each layer must be periodic in x with the same period. The only other requirement is that the boundary between each layer is planar and parallel to the x-y plane.

Because of the periodicity of the grating, an incident planewave interacts only with other planewaves which have angles related to the incident planewave by the Floquet condition, $k_1 \sin\theta_i = k_1 \sin\theta_0 - 2\pi i/d$, where θ_0 is the angle of the incident planewave, d is the grating period, and i is an integer. Consequently, an incident planewave can only diffract to planewaves with angles specified by the same relation.

The starting point for RCWA method is to write the fields in the three regions as sums of planewaves all related by the Floquet condition. For the TE case,

$$E_y(x,z) = \sum_i Q_i e^{-j(k_{x_i}x + k_{1z_i}z)} + \sum_i R_i e^{-j(k_{x_i}x - k_{1z_i}z)} \quad z \in \mathrm{I}, \tag{4.64a}$$

$$E_y(x,z) = \sum_i S_i(z) e^{-jk_{x_i}x} \quad z \in \mathrm{II}, \tag{4.64b}$$

$$H_x(x,z) = \sum_i U_i(z) e^{-jk_{x_i}x} \quad z \in \mathrm{II}, \tag{4.64c}$$

$$E_y(x,z) = \sum_i T_i e^{-j(k_{x_i}x + k_{3z_i}(z-a))} \quad z \in \mathrm{III}. \tag{4.64d}$$

The magnetic fields, $H_x(x,z)$, in regions I and III are easily found from the electric fields since these regions are homogeneous. The periodicity is manifested in the above equations because $k_{x_i} = k_1 \sin\theta_0 - 2\pi i/d$, $k_{1z_i} = \sqrt{k_1^2 - k_{x_i}^2}$, and $k_{3z_i} = \sqrt{k_3^2 - k_{x_i}^2}$. The Q_i are the known incident planewave coefficients, and θ_0 is the known angle of one of the incident planewaves. Note, however, that if the incident field consists of more than one planewave, they must be separated in angle by an integer multiple of $2\pi/d$. Otherwise, the different incident planewaves will have diffraction orders in different directions with no overlap.

After setting up the diffraction problem into these sums of planewaves, the RCWA method solves for the reflection and transmission coefficients, R_i and T_i, in two steps. The first step is to use Maxwell's equations in the grating region to express $S_i(z)$ and $U_i(z)$ as coupled first-order differential equations. These equations are numerically solved to obtain the eigenvalues and eigenvectors of the fields in terms of the planewaves. The $S_i(z)$ and $U_i(z)$ functions can then be written in terms of these eigenvalues and eigenvectors. The second step is to equate the tangential fields of the various regions at the boundaries to enforce the boundary conditions. A matrix equation for the coefficients is the result. Since this method naturally gives the coefficients of each diffracted planewave, the solution for the far field is automatic. In summary, the numerical method consists of first solving for the eigenvalues and eigenvectors in each grating layer, and then solving a matrix equation which equates the fields at the boundaries.

The power of this method results from its simplicity and versatility. The numerical methods are uncomplicated—high quality numerical routines for calculating the eigenvalues and eigenvectors are readily available. The versatility comes from the ability to construct almost any periodic structure by stacking layers. For example, stacking layers allows construction of sinusoidal boundaries, slanted phase and amplitude gratings, gratings with "hanging cliffs," and gratings with thin films and coatings. The cost of having several layers is that the eigenvalues and eigenvectors for each layer must be calculated separately. Also, the matrix equation for the boundary condition grows with the number of layers. Both of these calculations are computationally expensive.

The requirement of periodicity also imposes some limitations, and, as we shall see, this method loses its power when the period of the grating becomes too large. This requirement also becomes a burden if the incident field is composed of many planewaves, for example when the incident field is a focused spot. When the grating period is no larger than one or two wavelengths, very few of the incident planewaves will have the required relation mentioned above, and the whole calculation must be repeated for each incident planewave. Another disadvantage of allowing only planewave incidence is the possibility of missing diffraction effects where the reflection or transmission coefficient changes suddenly with incident angle. Examples of such phenomena are Wood's anomalies in metallic gratings and dielectric resonant gratings.

4.3.2 Comparison of Numerical Results

To further justify our use of the integral formulation, we will directly compare the numerical results of two specific cases where our integral method has superior qualities over the RCWA method. To make the comparison fair, we would like to use each method to calculate the diffraction for identical structures. However, each method makes inherently different assumptions about the structure.

The RCWA assumes that the diffracting structure is periodic, whereas our integral method assumes that all the relevant features and fields lie within a small region of the surface. Because of this difference, our integral method is better suited for calculating the diffraction from a single groove. To adapt the problem so that the RCWA can be used, we can define the structure as periodic with a large period. Since we are interested in the diffraction of a focused spot, making the period larger than the spot width should be sufficient.

Creating the focused spot for the incident field is also a problem with the RCWA. Because the RCWA allows only incident planewaves, the focused spot must be approximated by a finite sum of planewaves. Each calculation can include more than one incident planewave only if the planewave angles are separated by integer multiples of $2\pi/d$, as explained in the last section. In general, if other incident planewaves are desired, then the whole calculation must be repeated for each planewave. However, in the case where the grating period is made large to approximate a single groove, the spacing, $2\pi/d$, is small, and many planewaves can be included in a single calculation. Of course, the implication is that the number of diffraction orders in the calculation will also be very large resulting in a long calculation.

Another distinction between the two methods is the treatment of sharp corners in the diffracting structure. The RCWA treats the diffracting structure as a stack of layered gratings. At the interface between two layers, discontinuities between the gratings of each layer will have the effect of producing sharp corners, and stability of the numerical results should be affected by their presence. If rounded corners are desired— real physical structures never have infinitely sharp corners—then the structure will have to be represented by many layers, increasing the length of the calculation. Our implementation of the integral method requires that the contour have a finite first derivative at each grid point. We enforce this condition by requiring a smooth contour.

We will compare the convergence properties of each method for two separate problems. The first problem will be for a focused TE incident field on a single groove etched in silicon; we already discussed the convergence properties for the integral method for this diffraction problem in section 4.2.3. This problem will demonstrate the above principles regarding the treatment of a single, nonperiodic structure and a focused incident field. The second example will be for a TM incident field focused onto a sinusoidal aluminum grating. We made this choice in preparation for the analysis of compact disc formats which are usually constructed with an aluminum reflecting film. While the grating period for these structures is small enough for the practical application of RCWA, Li and Haggans[38] have reported that RCWA is slow to converge for gratings made in highly conductive media. Also, we still have the issue of replicating a focused spot, and, since the period is small, approximating the focused spot with many planewaves will require many repetitions of the calculation.

To measure the convergence properties of each method, we will study the behavior of power and far field quantities for increasing accuracy. Accuracy is determined by the number of diffraction orders included in the calculation for RCWA. We will consider the convergence of the reflected power, transmitted power, and their sum. As explained in section 4.2.3, the transmitted and reflected power flows are calculated across the same boundary in the integral method. Therefore, we would always expect their sum to equal one, even if there is absorption in the medium. For the RCWA method, however, the total reflected or transmitted power is calculated by summing the diffraction efficiencies for the reflected or transmitted diffracted orders. In effect, the summation of the reflected orders represents the power flow across the region I/region II boundary, and the summation of the transmitted orders represents the power flow across the region II/region III boundary. Since these boundaries are different, we would expect the sum of the reflected power and the transmitted power to be less than one if there is absorption in the grating region. However, we would still expect this sum to converge.

TE Example: Deep Groove with a Long Period

The first example problem is the same silicon groove that we used as an example for demonstrating the convergence tests in section 4.2.3. To make this problem suitable for the coupled-wave method, we set the grating period to $10\mu m$. Since the real part of the index of refraction for silicon is equal to 3.88 at this wavelength, the number of

# orders	R_0	T_0	R+T
120	0.2965	0.4239	0.9899
140	0.2777	0.4186	0.9928
160	0.9438	7.901	270
180	0.4170	14.48	2351

Table 4.3: Power conservation test for the coupled-wave method shows no convergence for a 10μm period grating in silicon.

transmitted propagating planewaves will be 122. Therefore, to test the convergence of the coupled-wave method, we will start with this number of orders and check if the results converge as the number of orders is increased. Table 4.3 shows the result of the reflected 0 order, transmitted 0 order, and the power conservation sum for all calculated orders for a single normally incident planewave. In fact, the result appears to be astable. We have traced the cause of this behavior to the presence of very large real eigenvalues. Because of the exponential decay of the planewaves with real eigenvalues, numerical accuracy is lost at the second boundary, and the matrix becomes nearly singular. The result is large round-off errors. If convergence cannot be attained for a single planewave incident, then, clearly, there is no need to pursue further the issue of filling the numerical aperture with planewaves.

For the case of a single deep groove, RCWA runs into a "catch-22" type problem. If the grating period is too small, the grating is a poor approximation of a single groove, and if the period is too large, the numerical results become unstable if all the propagating waves are to be used in the calculation. However, for approximately the same cost in computation time, the integral method converged to four decimal places for power conservation. The convergence was not uniform across the far field, however, and in the worst cases, only two decimal places were significant.

TM Example: Aluminum Grating with a Wood's Anomaly

This example is taken from the context of optical compact disc (CD) memories. Optical discs consist of an aluminum surface with small depressions written in circular tracks. The tracks form a grating in the radial direction. The information is represented by the presence (or lack of presence) of a depression, and it is read by measuring the diffraction from a focused spot of laser light. As the disc rotates, the laser spot moves along the track measuring the presence of a pit. The diffraction pattern is determined by the track period and the pit shape. If the track period is reduced, then the disc can hold more data. For example, the new SD CD format[39] reduces the track period to 0.74μm from 1.6μm for the older audio CD format.

Wood's anomalies[31] occur for gratings in highly conductive media. When a planewave is polarized with the electric field perpendicular to the grating (TM) and is incident on the grating at an angle so that a diffracted order propagates parallel to grating surface, then most of the incident power will be diffracted into this diffracted order, and very little light will be reflected. This phenomenon was first discovered by Wood[32] and explained by Lord Rayleigh[40]. That the RCWA method is slow to converge for this situation has been previously reported[38]. If our integral method is to be useful for analyzing diffraction from optical disc formats, then it must have reasonable convergence properties for highly conductive media, even in the difficult realm of a Wood's anomaly.

For our comparison, we will study the diffraction of TM polarized light from an aluminum grating with a grating period equal to the light's wavelength. Figure 4.10 shows the diffraction geometry. When the grating period is equal to the wavelength of the incident light, then the Wood's anomaly occurs for the angle of incidence equal to zero. At this incidence angle, only the plus and minus first diffraction orders are present, and they are at the grazing angle along the surface. Therefore, to use the RCWA method to calculate the diffracted far field for a focused spot would require repeated calculations to fill the the numerical aperture with incident planewaves. Also, the incident planewaves would have to be chosen densely enough to properly describe the variation around the anomaly.

Figure 4.11 shows the convergence results for both the RCWA and integral methods. The calculation using the integral method was performed for a $\mathcal{NA} = 0.6$, and the truncation length for the contour was $s = \pm 6.9\mu$m. The figure shows the convergence

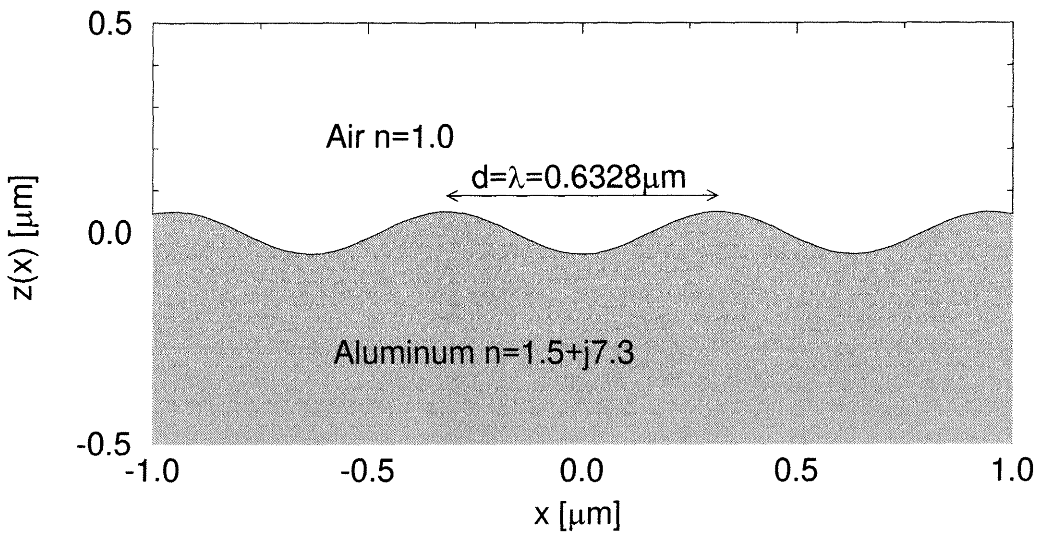


Figure 4.10: Diffraction grating used in comparison of RCWA and the integral method.

behavior of the amplitude of the far field at sinθ = 0, the total reflected power, the total transmitted power, and the power conservation test, $R + T$. Note that while the total reflected and transmitted power and the power conservation all seem to converge very quickly, the reflected far field at sinθ = 0 converges much more slowly. In fact, the value oscillates. However, no matter how finely we grid the contour, the solution remains numerically stable.

The RCWA calculation was for a single incident planewave at $\theta = 0$. The bottom graph in figure 4.11 plots the diffraction efficiency for the reflected zero order and the sum of the diffraction efficiencies of the reflected and transmitted orders. While the sum should be less than one since the aluminum is highly absorptive, it should nevertheless converge. These diffraction efficiencies show no convergence for an increasing number or orders. Since the numerical method does not converge for a single incident planewave, it would certainly not be useful to attempt the calculation for all

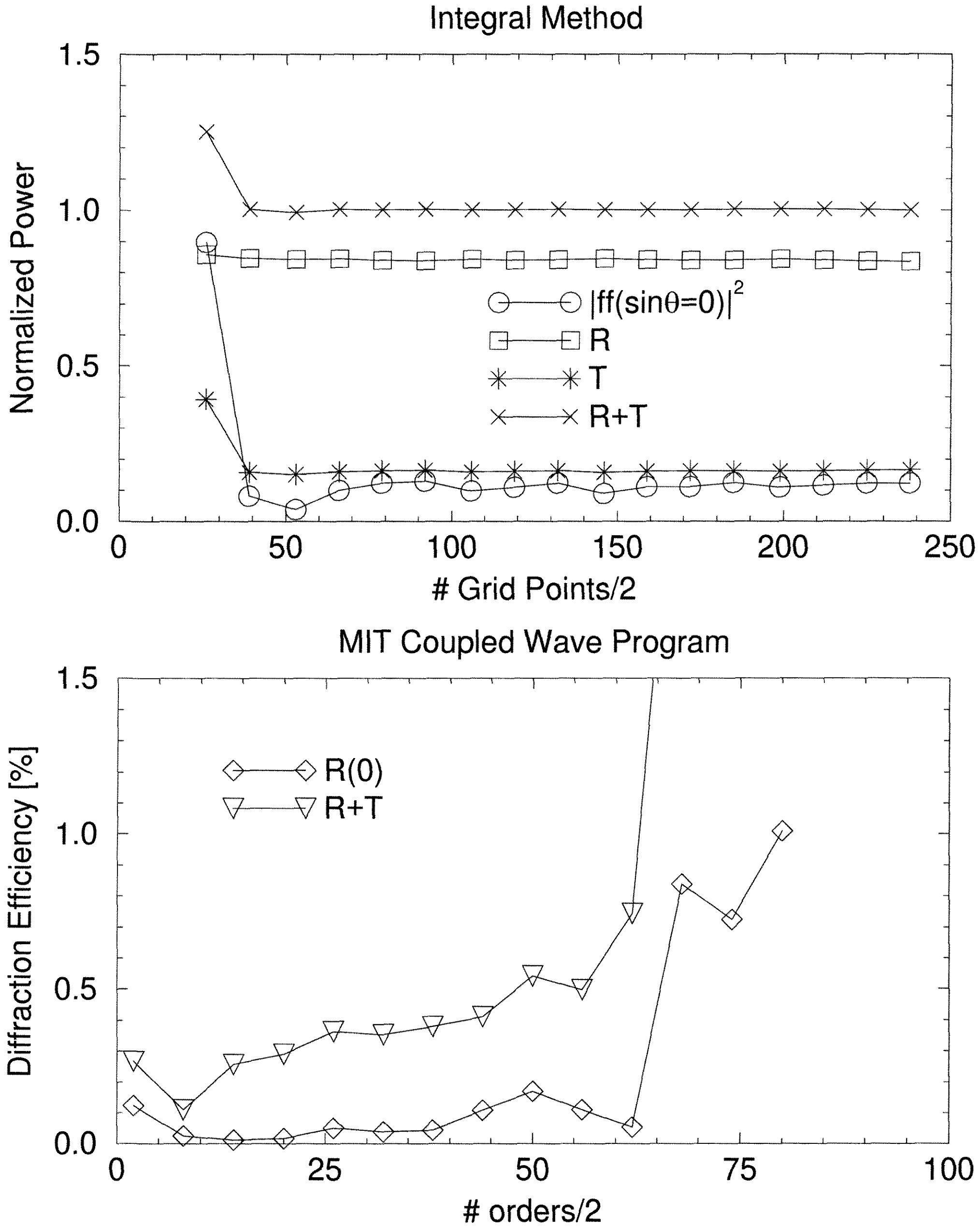


Figure 4.11: Comparison of RCWA and our integral method for a TM incident field onto a sinusoidal aluminum grating.

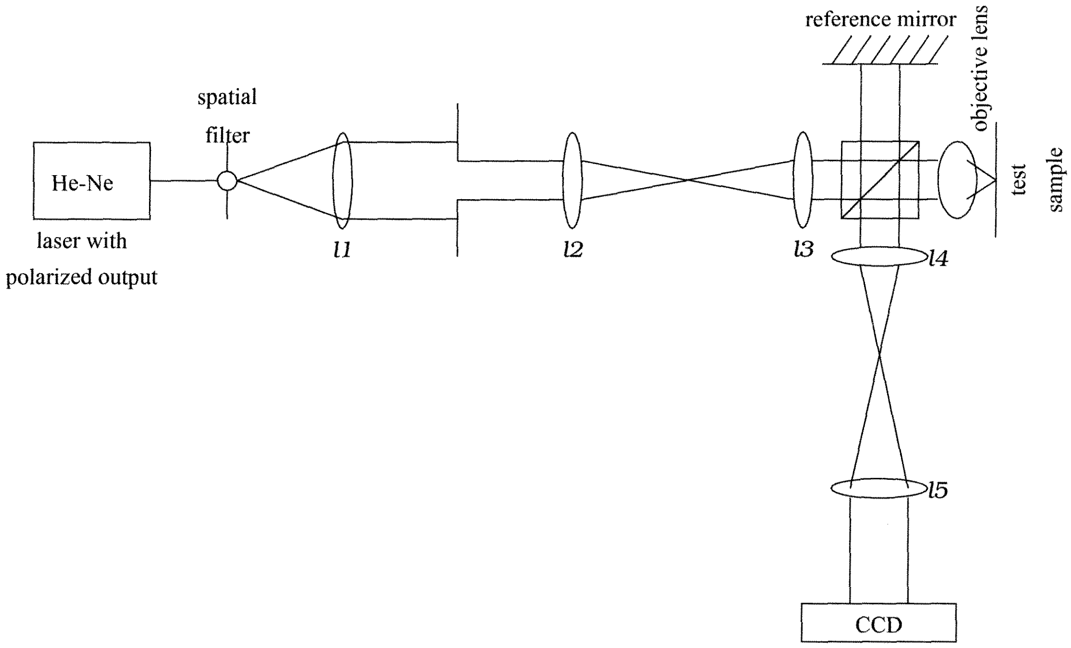


Figure 4.12: setup for measurement of diffracted far field for focused incident fields

the incident planewaves in the aperture. In conclusion, for this particular problem of a TM field with a shallow aluminum grating, our integral method is more useful then the RCWA method for calculating the diffracted field.

4.4 Experimental Verification

Before we can confidently use our numerical method to predict the diffraction behavior for various structures and incident fields, we would like to compare some calculations to actual diffraction measurements. If the comparison is reasonable, then we can proceed to use our method as a design tool without having to fabricate a sample for each intermediate design.

The diffracted far field for each sample was measured in a similar manner to the Fourier plane measurements in Chapter 2. Figure 4.12 diagrams the laboratory setup. The iris between *l1* and *l2* is imaged onto the entrance pupil of the objective lens, reducing the numerical aperture of the incident field. The full numerical aperture of the objective lens is used to capture the reflected field. The back focal plane, or Fourier plane, of the objective is imaged by *l4* and *l5* onto the CCD camera, where an intensity profile is measured. The reference mirror is used to focus the illumination spot, and the shutter is closed when diffraction measurements are recorded.

The measurement of the far field is for a three-dimensional spot, but our numerical calculation is only for two dimensions. A true comparison would be for the incident illumination being formed by a cylindrical lens rather than the spherical objective used. We use a cross section of the measured far field and ignore the effect of the spot profile in the other direction. If the test pattern is truly independent of one direction, then we can argue that no planewave with an angle in that dimension can be diffracted so that its angle in that dimension will change. Therefore, a cross section in the direction of interest should contain no diffraction of planewaves that would not be present if the objective were a cylinder.

4.4.1 An Aluminum Grating with TM Incidence

The first test sample consists of sinusoidal gratings in Aluminum. The gratings were fabricated by the Rochester Photonics Corporation using a rastered laser spot to expose a photoresist. After the photoresist was developed to form the grating, a coating of approximately 500Å of aluminum was evaporated onto the grating. Finally, the actual depth and profile of the grating was measured with an atomic force microscope at the Jet Propulsion Laboratory. The surface measurement showed that the grating was quite sinusoidal with a peak to valley height of 0.185μm. The grating period was 1.0μm.

For this grating, the Wood's anomaly should be quite evident. We can estimate the angular location where we expect to see strong absorption by calculating which planewave angle is coupled into the planewave traveling parallel to the surface by the grating. We will use the Floquet condition, $\sin\theta_r = \sin\theta_i + i\lambda/d$, where θ_r is the reflected angle, θ_i is the incident angle, and i is an integer. Then, setting the reflected angle to $\sin\theta_r = 1$, we expect the anomalous absorption at the incident angles,

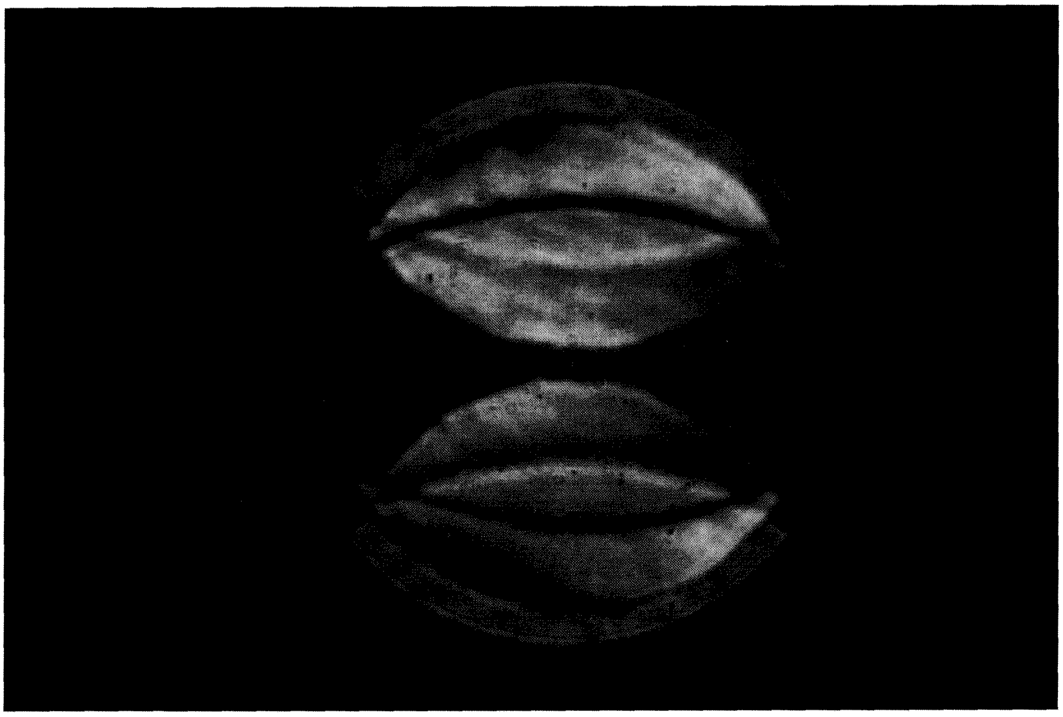


Figure 4.13: Reflected Fourier plane for aluminum grating with TM incidence clearly showing the Wood's anomalies.

i	$\sin\theta_i = 1 - i\lambda/d$
+1	0.367
+2	-0.266

for this aluminum grating with λ=0.6328μm. Consequently, we expect to measure notches in the reflection at these angles.

Figure 4.13 shows the CCD image of the Fourier plane for a focused TM spot on the aluminum grating. This image contains several interesting features. The numerical aperture for the incident spot was about 0.6, which is equal to the maximum $\sin\theta$ of the incident planewaves. The +1 and -1 diffracted orders are shifted $\sin\theta = \lambda/d = 0.6328/1.0 = 0.6328$, thus forming the two partial circles that do not quite meet in the center. The faint complete circle that fills in the background is the zero order reflected spot. Small bands at the top and bottom of the zero order spot are diffracted

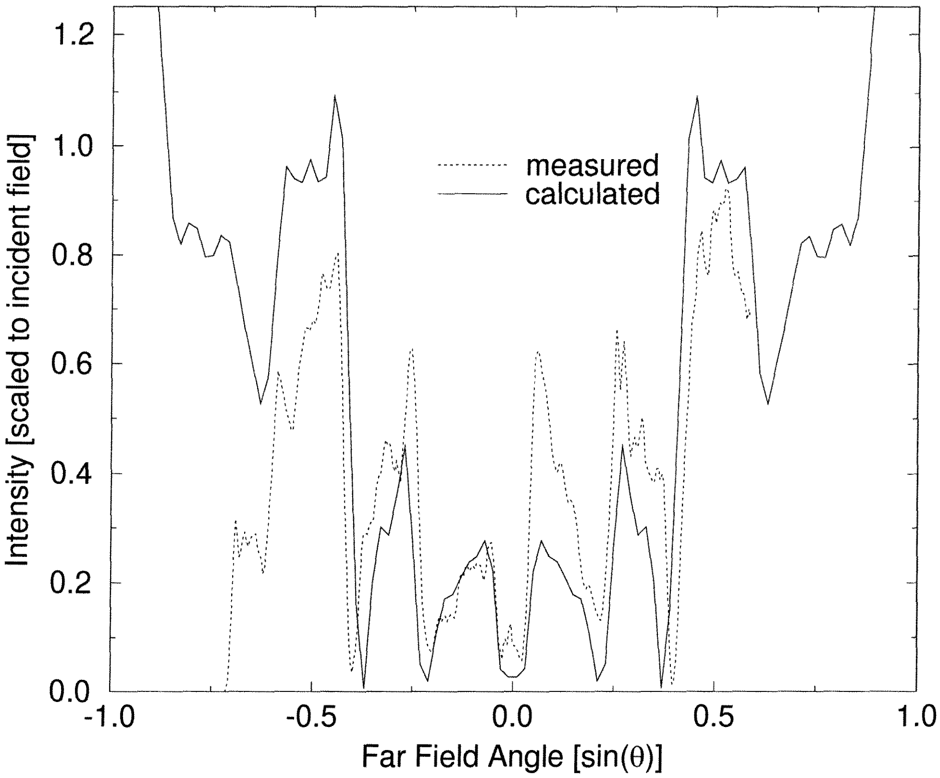


Figure 4.14: Measurement versus calculation for the sinusoidal aluminum grating with the Wood's anomalies clearly evident as null in the reflected far field.

light captured because the numerical aperture for the collection of the reflected light was about 0.7, which is larger than the numerical aperture for the incident light. Finally, the two dark arcs within each diffracted order result from the absorption of the Wood's anomalies. The darker band is for the +1 angle in the above table, and the other is for the +2 angle.

The numerical simulation for this problem was calculated for a sinusoidal surface in a medium with index of refraction 1.5+j7.3 [30]. With the imaginary part of the index of refraction so large, the optical field is only about 2.7% of the incident amplitude at a depth of 500Å. Therefore, approximating the thin film of aluminum by filling the whole volume with aluminum is reasonable. Figure 4.14 compares the numerical result to a vertical cross section taken from the image. Agreement is quite good at the smaller angles, and the nulls in the reflected power due to the Wood's anomalies are quite evident and occur very close to the predicted angles. The small difference in the null angle between the numerical result and the measured cross section could come from a difference in the actual numerical aperture and the one used in the numerical calculation, since the horizontal scale of the measured data was scaled to the calculated data assuming the numerical aperture used for each was identical. The vertical scale for each plot was derived by matching the measured and calculated reflection for a planar aluminum interface. We are pleased with the correspondence between calculated and measured data.

4.4.2 A Single Groove in Gallium Arsenide

Returning to diffraction from single structures, we used grooves etched into gallium arsenide for our second experiment. The fabrication was performed at Professor Scherer's laboratory by his student, Chuan-cheng Cheng, at Caltech. The photographs in figure 4.15 are of one of the etched grooves and were taken by a scanning electron microscope. The first photograph is a top view of the groove and shows the groove width, which is about 0.35μm wide. The bottom photograph shows a side view of the groove. To allow this view, a mask was formed across the grooves and then the substrate was etched away, opening the groove to the side. This photograph was taken at a 60° angle, and direct measurement of the groove's depth from this picture is difficult. The groove height was measured by comparing the groove to the large etch height, which could be measured with a surface profiler. The groove depth was measured to be about 0.65μm. The groove shape was estimated by the shape in figure 4.16, which was constructed from a raised cosine pulse.

The measurement of the diffracted far field was performed by sliding the groove underneath the illuminating spot. At each 0.1μm step, the intensity profile at the CCD was recorded. Figure 4.17 compares the measured and calculated far fields for the illuminating spot centered on the groove and for the illuminating spot away from the groove. While the overall change in intensity level between on and off the groove seems to agree between the calculated and measured data, there are a few disappointing features. The peak in the measured intensity in the middle of the far

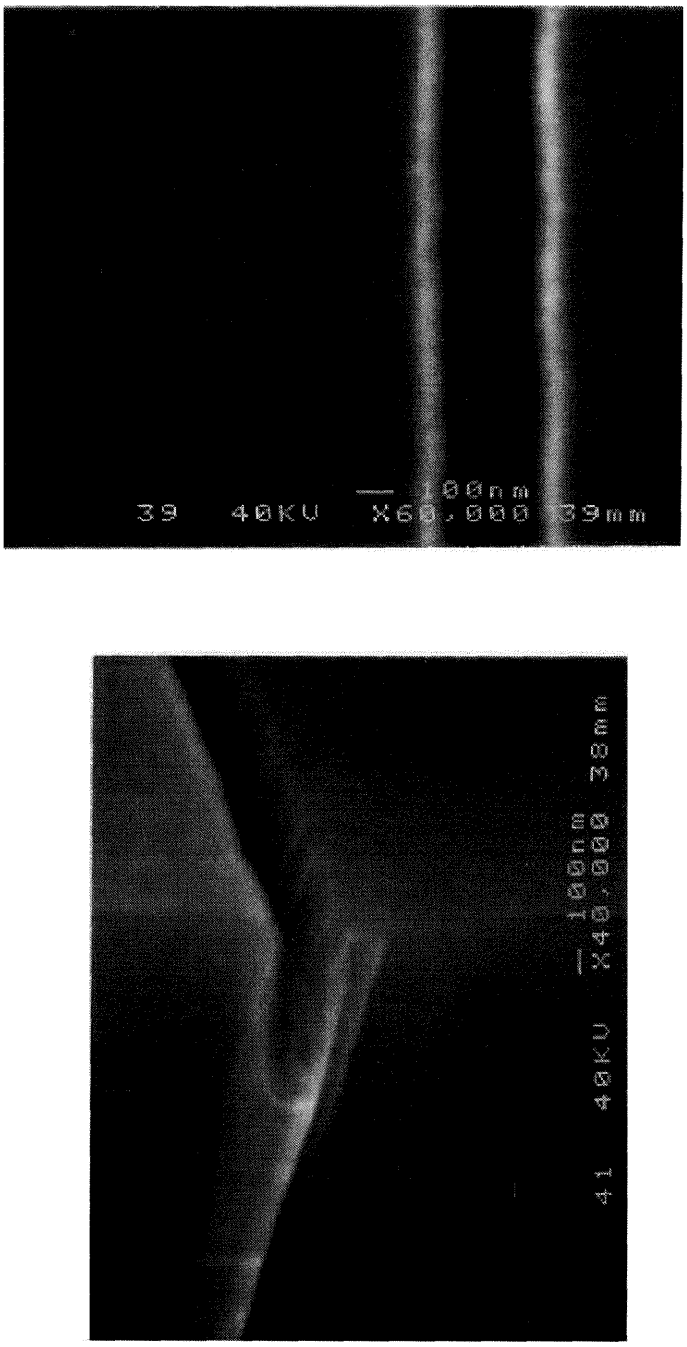


Figure 4.15: Scanning electron microscope photographs of the groove etched in GaAs and used in the diffraction measurement. The top photograph is a top view showing the width of the groove, and the bottom photograph is a side view showing the groove shape.

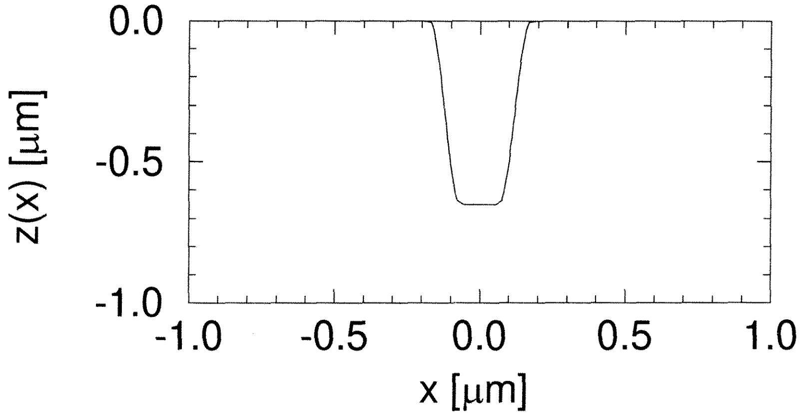


Figure 4.16: The contour used to represent the groove in gallium arsenide.

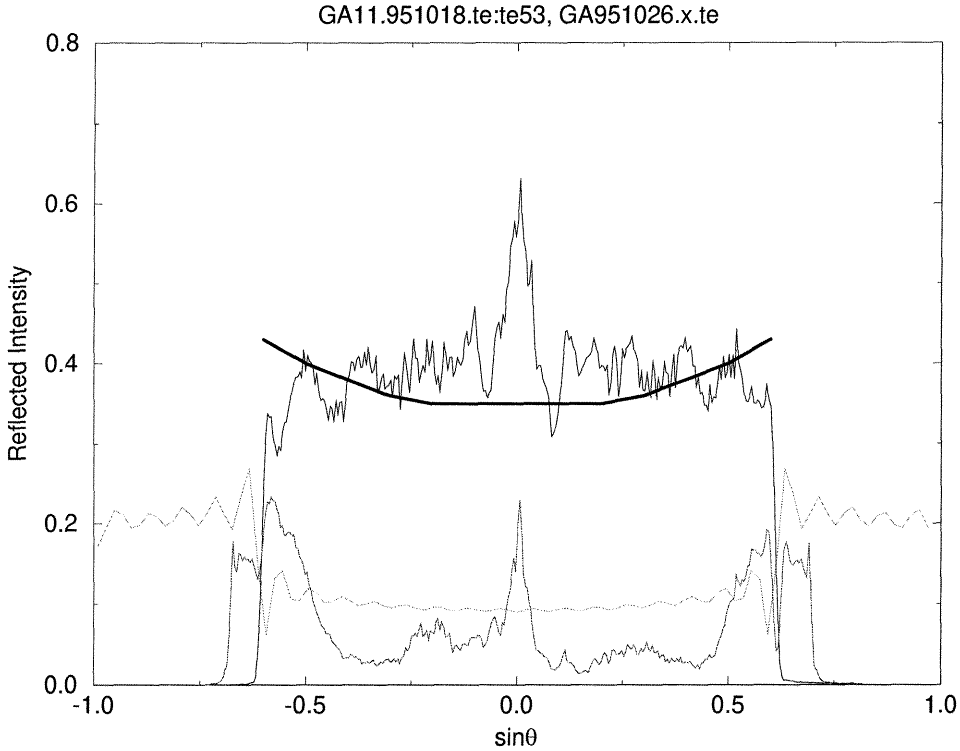


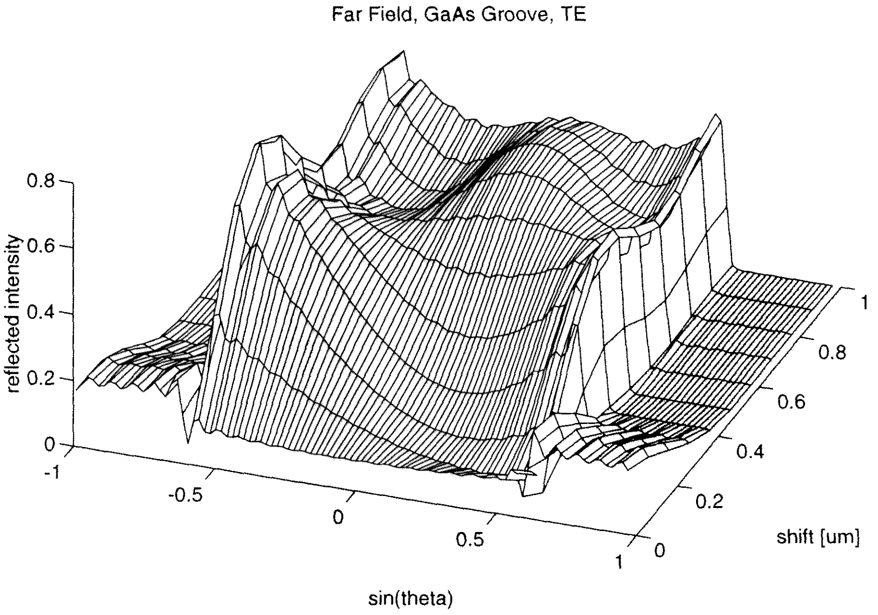
Figure 4.17: Measurement versus calculation for the groove etched in GaAs with TE incidence.

field is an artifact of light reflected from a lens surface that comes to a focus in the CCD plane; these peaks can be ignored. The reflected field calculated for the absence of a groove is simply the Fresnel reflection coefficients, and this reflection increases with far field angle for TE illumination. However, the measured intensity pattern does not seem to satisfy this. Figure 4.18 compares the complete calculation for the groove shifting underneath the spot with the measurement. Again, the peak running down the middle of the measured data can be ignored. Features in common with the calculated and measured data are the waves on the left side of the plots. One difference is that the slope on the front left part is steeper for the measured data than for calculated data. In conclusion, it appears that while the numerical calculation successfully predicts general shapes and trends for these single grooves, its value in quantitative measurement needs to further evaluated.

4.5 Conclusion: What's the Program Good For?

Need a method to calculate vector diffraction for focused spots. Most methods require planewaves and gratings. We are interested in two class of problems: deep structures in semiconductor materials and optical disc memory formats. We saw that for these problems, the integral method formulated in this chapter performs much better than the popular rigorous coupled-wave method.

When using our integral program, there are certain things that we must be aware of. 1. Must always check convergence and choose a sampling grid spacing that has good accuracy, but is not so fine that computation time is unreasonable. 2. Truncate the contour as wide as possible because of the Gibbs phenomenon. Of course, increasing the length of the contour used will also lengthen the computation time. Because of the Gibbs phenomenon oscillations, we cannot expect high frequency features in the far field to be accurate predictions of measurement. 3. The contour must be smooth. Vertical walls and ledges could be formulated by using the derivative in z for these parts. We have not tried any contours with these properties, and so we cannot make a statement about the convergence of the solution in these cases.



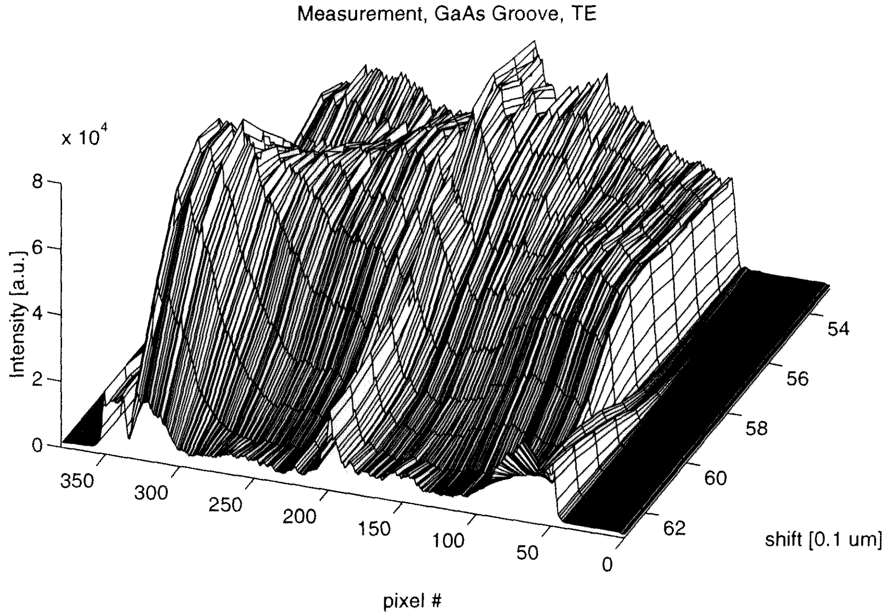


Figure 4.18: The far field intensity profile as the incident spot is shifted away from the groove. The top graph is calculated from the integral method and the bottom graph is the measured data.

Appendix A Fourier Transform of the Free Space Green's Function

To prove the inverse Fourier transform relation for the free space Green's function stated in equation 4.11, we start with the Green's function and compute the forward Fourier transform. Thus, the Fourier transform of the Green's function is

$$\mathcal{F}\left\{\frac{e^{jk_1R}}{4\pi R}\right\} = \iiint \frac{e^{jkR}}{4\pi R} e^{-j\boldsymbol{k}\cdot\boldsymbol{r}} R^2 \sin\theta \, dR \, d\theta \, d\phi. \tag{4.65}$$

This integral is easily computed in the θ and ϕ coordinates by expanding $|\boldsymbol{k}\cdot\boldsymbol{r}| = R|\boldsymbol{k}|\cos\theta$, and performing the change in variables $u = \cos\theta$. The final integral is

$$= \frac{1}{j2|\boldsymbol{k}|}\int_0^\infty \left[e^{j(k_1+|\boldsymbol{k}|)} - e^{j(k_1-|\boldsymbol{k}|)}\right] dR. \tag{4.66}$$

For this integral to be bounded, we must assume the $\mathrm{Im}\{k_1\} > 0$. The solution to this integral is our desired Fourier transform,

$$\mathcal{F}\left\{\frac{e^{jk_1R}}{4\pi R}\right\} = \frac{1}{k_1^2 - |\boldsymbol{k}|^2}, \tag{4.67}$$

and our proof is complete.

Appendix B Integral Equations for the TM case

For the TM case, the directions of $\boldsymbol{J}_{s,e}$ and $\boldsymbol{J}_{s,m}$ are reversed and mirrored with respect to their directions for the TE case, so that

$$\boldsymbol{J}_{s,e} = J_{s,e}\hat{\tau}(s), \quad \text{and} \quad \boldsymbol{J}_{s,m} = -J_{s,m}\hat{y}. \tag{4.68}$$

Because of this antisymmetry, the matrix elements $\mathbf{\Gamma}_{11}$ and $\mathbf{\Gamma}_{22}$ for the TM case are exactly mirrored to the TE case in the manner,

$$\mathbf{\Gamma}_{11_l}^{\mathrm{TM}} = \mathbf{\Gamma}_{22_l}^{\mathrm{TE}}, \quad \text{and} \quad \mathbf{\Gamma}_{22_l}^{\mathrm{TM}} = \mathbf{\Gamma}_{11_l}^{\mathrm{TE}}. \tag{4.69}$$

The other matrix elements are

$$\begin{aligned} \mathbf{\Gamma}_{12}^{\mathrm{TM}} \boldsymbol{J}_{s,e}(\boldsymbol{r}_s) = j\omega\mu\hat{n}(s) \times \Big[& \int_{-\infty}^{\infty} \Big(G_1(R) - G_0(R)\Big) J_{s,e}(s')\hat{\tau}(s')ds' \\ & + \nabla^2 \cdot \int_{-\infty}^{\infty} \Big(\frac{G_1(R)}{k_1^2} - \frac{G_0(R)}{k_0^2}\Big) J_{s,e}(s')\hat{\tau}(s')ds' \Big] \end{aligned} \tag{4.70a}$$

$$\mathbf{\Gamma}_{21}^{\mathrm{TM}} \boldsymbol{J}_{s,m}(\boldsymbol{r}_s) = \frac{-\hat{n}(s)}{j\omega\mu} \times \int_{-\infty}^{\infty} \Big(k_1^2 G_1(R) - k_0^2 G_0(R)\Big) J_{s,m}(s')\hat{y}\, ds'. \tag{4.70b}$$

The integrand in $\mathbf{\Gamma}_{21}^{\mathrm{TM}}$ is equal to $\mathcal{H}_0(R)$, which has an integrable singularity as $R \to 0$. The numerical integration of this term is explained in section 4.2.2, and we do not repeat it here. The second term of $\mathbf{\Gamma}_{12}^{\mathrm{TM}}$ is more troublesome, however, because it has an integrable singularity and a second derivative of the integral.

The solution is to subtract the form of the singularity and integrate it separately. Once the singularity is subtracted from the integrand, we execute the derivatives inside the integral. First, we ease the notation with a few definitions. We let

$$\mathcal{H}_3(R) = \frac{H_0^{(1)}(k_1 R)}{k_1^2} - \frac{H_0^{(1)}(k_0 R)}{k_0^2}, \tag{4.71}$$

and write the second term of $\mathbf{\Gamma}_{12}^{\mathrm{TM}}$ as

$$[\mathbf{\Gamma}_{12}^{\mathrm{TM}}]_2 = \nabla^2 \cdot \int_{-\infty}^{\infty} \mathcal{H}_3(R)\,\mathrm{sinc}(k_0 W s' - \pi l)ds'. \tag{4.72}$$

The singularity in the integrand has the limiting form as $R \to 0$,

$$\mathcal{H}_3(R) \sim \tilde{\mathcal{H}}_3(R) = j\frac{2}{\pi}\Big(\frac{1}{k_1^2}\log(\frac{k_1 R}{2}) - \frac{1}{k_0^2}\log(\frac{k_0 R}{2})\Big). \tag{4.73}$$

Within a fixed region around the singular point $s' = s$, we subtract $\tilde{\mathcal{H}}_3(R)$ from the integrand and integrate it separately,

$$\int_a^b \tilde{\mathcal{H}}_3(|s-s'|)ds' = \\ j\frac{2}{\pi}[-(b-a)+(s-a)\log(\frac{k_1}{2}(s-a))+(b-s)\log(\frac{k_0}{2}(b-s))]. \quad (4.74)$$

Applying the double derivative with respect to s, and assuming that a and b can be chosen such that $b-s=s-a$, the result is the constant

$$\alpha_{\mathcal{H}} = \frac{d^2}{ds^2}\int_a^b \tilde{\mathcal{H}}_3(|s-s'|)\,ds' \; = j\frac{4}{\pi(b-a)}(\frac{1}{k_1^2}-\frac{1}{k_0^2}). \quad (4.75)$$

After subtracting the singularity from the integrand, the integrand is no longer singular, and the derivatives may be carried out inside the integral. Similar to the TE case, we define the functions

$$\mathcal{H}_0^{\mathrm{TM}}(R) = H_0^{(1)}(k_1R) - H_0^{(1)}(k_0R) \quad (4.76\text{a})$$

$$\mathcal{H}_1^{\mathrm{TM}}(R) = \frac{H_1^{(1)}(k_1R)}{k_1} - \frac{H_1^{(1)}(k_0R)}{k_0} \quad (4.76\text{b})$$

$$\mathcal{H}_2^{\mathrm{TM}}(R) = H_2^{(1)}(k_1R) - H_2^{(1)}(k_0R) \quad (4.76\text{c})$$

and the resulting formula for the matrix element is

$$\begin{aligned}\mathbf{\Gamma}_{12_l}^{\mathrm{TM}} = \hat{y}\frac{k_0}{4}\int_{-\infty}^{\infty}\Bigg[&\mathcal{H}_0^{\mathrm{TM}}(R)\big(\hat{\tau}(s)\cdot\hat{\tau}(s')\big) \\ &-\frac{1}{2}\big(\hat{\tau}(s)\cdot\nabla R\big)\big(\hat{\tau}(s')\cdot\nabla R\big)\big(\mathcal{H}_0^{\mathrm{TM}}(R)-\mathcal{H}_2^{\mathrm{TM}}(R))\big) \\ &-\hat{\tau}(s)\cdot\nabla\big(\hat{\tau}(s')\cdot\nabla R\big)\mathcal{H}_1^{\mathrm{TM}}(R)\Bigg]\operatorname{sinc}(k_0Ws'-\pi l) \qquad (4.77) \\ &+j\frac{2}{\pi}\Big(\frac{1}{k_1^2}-\frac{1}{k_0^2}\Big)\Big(\frac{1}{|s-s'|^2}\Big)\,ds' \; + \alpha_{\mathcal{H}},\end{aligned}$$

completing the recipe for the TM matrix elements.

Chapter 5 Analysis and Design of Optical Disc Formats

5.1 Introduction

In the previous chapter, we developed a rigorous numerical diffraction method and demonstrated its utility by analyzing its convergence properties and its accuracy compared with real measured data. In Chapter 2 we discovered that we could measure differences in the far field diffraction pattern when the diffracting structure consisted of subwavelength structures, but we had no method of designing, or even predicting, the far field response. We stated that if we could design diffracting structures so that their differences in the far field were maximized, then we could improve the channel's SNR and bit error rate. The numerical method developed in Chapter 4 gives us that tool, and in this chapter we utilize this tool for the design of optical disc memory formats.

We start the investigation by analyzing currently used and proposed disc formats. Since our method is only designed for two-dimensional problems, we cannot solve the complete problem of diffraction from pits with finite width and length. We will use our numerical diffraction method to look exclusively at the diffraction from the track structures. We choose to analyze the track structure, as opposed to the pit length in the direction of disc movement, because a decrease in track spacing pays bigger dividends in storage density. The channel bit length (smallest incremental change in pit length that represents one bit of information) in the direction of movement is already a small fraction of the illumination spot in the current formats. Run length limiting type coding restricts the actual minimum pit length for a small cost in overall density, but the channel bit length is still the important parameter that determines information density in the direction of movement. No such coding can be used in the radial direction, and so squeezing the tracks together has an immediately significant effect on the storage density.

The recently proposed disc format for the Digital Video Disc (DVD) has a smaller track pitch in relation to the illumination spot size than the older audio Compact Disc (CD). However, the DVD format maintains the principle of avoiding cross-talk between tracks by specifying the track pitch just large enough so that neighboring tracks are outside the spot width when the spot is centered on one track. We would like to investigate the effect of diffraction on the far field signal if we make the tracks pitch even smaller so that the illumination spot covers more than one pit. We have two lessons that prompt us in this direction. In Chapter 2, we saw that when several pits lie within the illuminating spot, the information can still be recovered if the far field signals are distinct and the number of possible symbols is a finite set. With the development of our numerical method, we can now easily investigate the far field signal constellations for different formats and design one with good qualities. The second lesson is from the example in section 4.4.1 regarding TM diffraction from an aluminum structure. We saw some interesting effects when the incident illumination is polarized in the TM direction, and these effects are not present for the TE polarization. Therefore, we expect that we can design disc formats, which are usually fabricated by covering an embossed disc with an aluminum film, so that the diffraction for the two polarizations is very different. We will see when we design our experimental formats that we will be able to use the TE polarized reflected field as a reference for measuring amplitude and phase of the TM polarized reflected field.

Our perusal through the diffraction effects of different CD track structures will completely ignore some important issues. Because our numerical method is only for two-dimensional structures, we will completely ignore the effect of finite pit length. (We will call the pit dimension in the tracking direction the pit length and the pit dimension in the radial direction the pit width.) This effect ought not to be ignored, because just as the new DVD format reduces the track pitch, it also reduces the minimum pit length. In fact, the minimum pit length for this format is less than half the spot width. Clearly, a short pit will have an effect on the far field diffraction pattern. Another issue that we will ignore when we discuss our futuristic formats will be how to master and reproduce such formats. We assume that reproduction could be a stamping process similar to current methods, but the mastering process would be much more complicated because of the extremely close positioning of tracks and the possibility of multiple pit depths. The technique required for the optical disc head to maintain tracking will probably need to be significantly different for discs with a futuristic format than for a current format, and this issue will also be ignored. Although these issues are quite important, our results will still be informative because of the new ideas that we will present regarding the use of polarization in signaling constellations. Even if all our futuristic formats presented here prove impractical because of one of the ignored issues, we still believe that the general principle of polarization signaling can increase the information capacity of optical discs. In addition, the idea of polarization signaling and designing subwavelength structures to have desirable far field properties is an important extension of the superresolution ideas presented in Chapter 2.

We start our investigation by analyzing the current optical disc formats. The analysis of the DVD format will be the first to show large differences between the TE and TM reflected far fields. Proceeding with our own experimental formats (the DS formats), we will analyze the differences in the diffraction for the two polarizations, and we will develop a working model to explain the diffraction. After designing a quadrature detector to measure the amplitude and phase of the TM field with respect to the TE field, we use the working diffraction model in the last section to design pit structures with prescribed far field quadrature properties.

5.2 Current Formats

The following is a short table comparing the disc formats of the old CD and the new DVD[39].

	DVD	CD
wavelength	650nm or 635nm	780nm
wavelength in medium	433nm or 423nm	520nm
numerical aperture	0.6	0.45
spot size (λ/NA)	1.06μm	1.73μm
track pitch	0.740μm	1.6μm
smallest pit length	0.400μm	0.834μm
channel bit length	0.133μm	0.278μm
approximate density	5.8 bits/μm^2	1.2 bits/μm^2

The optical signal resulting from the interaction of the illuminating spot and the pit is usually described in the literature with scalar diffraction [19, 21]. We will see that a scalar description is satisfactory for the CD and only partly so for the DVD. According to a scalar model, a pit depth of a quarter wave would cause the center of the reflected far field to be dark because the light reflected from the bottom of the pit destructively interferes with the light reflected from the surface around the pit. To obtain the best contrast, the pit width is made about one-third the spot size, since the central one-third of the spot contains about half its power. Thus, the presence of a pit causes a dark reflection, and its absence causes a bright reflection.

The basic construction of these discs is by injection molding a polycarbonate substrate onto a nickel master, which is patterned with the information. The information side of the polycarbonate is then sputtered with a 100Å coating of aluminum, and then a protective coating is added on top of the aluminum [23]. To read the information, a focused spot of light enters the polycarbonate side of the disc and is reflected by the aluminum so that the reflected light is collected by the same lens and aperture that focused it. As a result the interface that serves as the diffraction contour is between polycarbonate and aluminum. To use our integral method program developed in the last chapter, we set n_0, the index of refraction of the incident medium, to 1.5, the approximate index of refraction of polycarbonate. Because of this incident medium, the wavelength of the incident light inside the medium is two-thirds the free space wavelength. However, the incident spot width is not different than if the incident medium were free space, because the angular spectrum of the incident field is also reduced by a factor of 1.5 due to refraction at the polycarbonate-air interface. The result is that inside the polycarbonate, the incident field wavelength and numerical aperture are reduced by the same factor. In the simulations, and in all the results that follow, we describe the pit's dimensions in terms of the wavelength inside the polycarbonate. The index of refraction of aluminum is somewhat wavelength dependent. For the CD format we use $\mathrm{n}_1 = 2.7 + j8.6$, and for the DVD simulations we use $\mathrm{n}_1 = 1.5 + j7.6$[30].

The contours used in the numerical simulation were generated with raised cosine functions which have the form

$$z(x) = \begin{cases} a & |x| < \sigma \\ \frac{a}{2}\left(1 + \cos(\pi \frac{|x|-\sigma}{\tau-\sigma})\right) & \sigma < |x| < \tau \\ 0 & |x| > \tau. \end{cases} \tag{5.1}$$

The raised cosine function has the convenient property that its first derivative is continuous, and the width and slope of the sides can be determined independently by the choice of σ and τ.

If a scalar model is to be a satisfactory predictor of actual measurement, then the diffraction for TE and TM waves must be the same. Dil and Jacobs[17] calculated that the phase depth (electric field phase at the bottom of the pit compared to the phase at the top surface) of aluminum pits is different for the TE and TM polarized fields. Our question for this section is: to what extent is the diffraction different for TE and TM fields for the CD format, and is there a significant change for the smaller pit width of the DVD format?

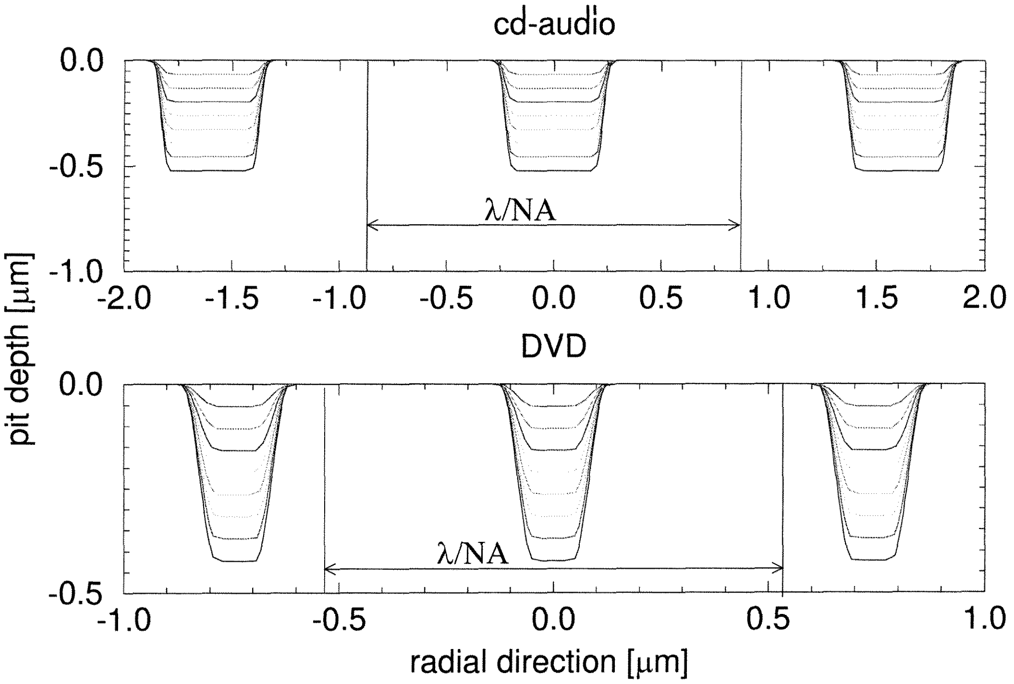


Figure 5.1: The contours used in the numerical simulations for CD and DVD formats. The pit depths were varied in steps of one-eighth wavelength.

5.2.1 Audio CD

The track pitch and pit cross sections used in our calculations are shown in figure 5.1. The approximate spot width of the illumination ($\approx \lambda/NA$) is also indicated in the graphs. Note that for the CD format, the track pitch is about equal to the spot width, whereas the track pitch for the DVD format brings the neighboring pits up to the edge of the illumination spot. For each format the pit depth was varied in steps of $\lambda/8n_0$. As mentioned above, the real formats specify that the pit depth is $\lambda/4n_0$, a quarter wave.

In the scalar approximation of the reflected signal for the CD format, described by Braat[21], the tracks are treated as a diffraction grating. Even though the illuminating spot only covers one track, the system can still be considered as a diffraction grating with discrete diffraction orders if we consider the planewaves that compose the spot one at a time. Because the grating period is approximately equal to λ/NA, half of the first diffracted order overlaps with half of the zero diffracted order inside the objective

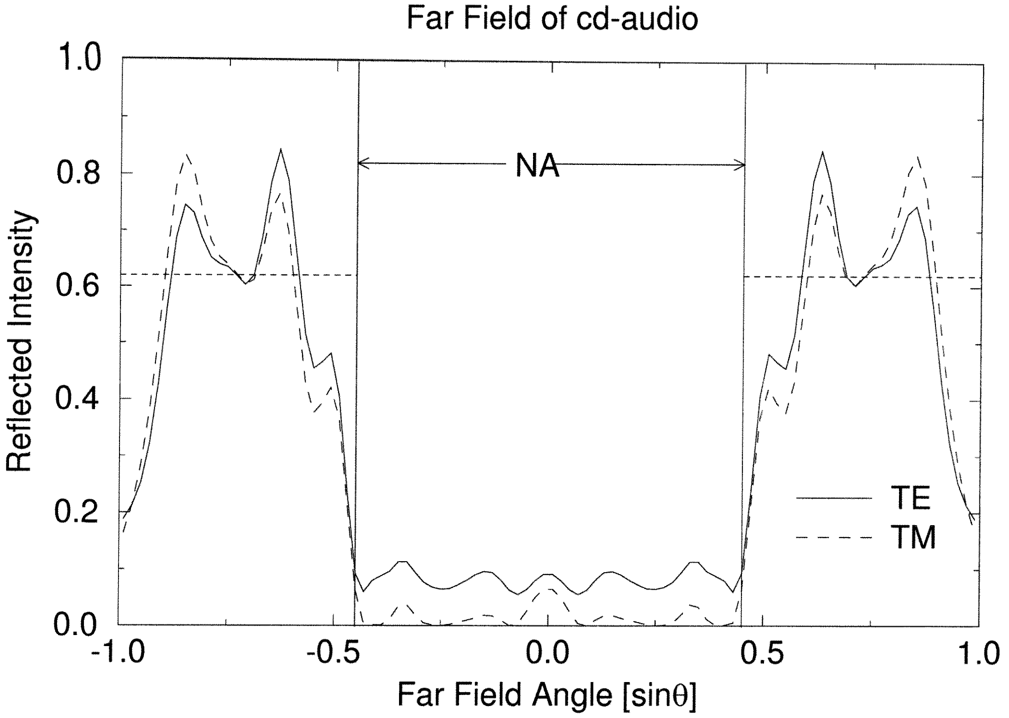


Figure 5.2: The far field intensity pattern for TE and TM illumination of the CD format with the pit depth equal to one quarter wave.

lens aperture. For a pit depth of one-quarter wave, the first diffracted order is π out of phase with the zero order so that there is no reflected field within the aperture. The far field intensity calculated by the integral method program is compared to the scalar prediction in figure 5.2, for a one-quarter wave pit depth. The results show that the TE and TM fields have very similar behavior for these pits, and the scalar approximation agrees well with both.

The computer simulations were calculated for increasing pit depth because we want to know if the scalar concept of a phase grating is appropriate, and if our results agree with the results of Dil and Jacobs regarding the difference in effective pit depth for the TE and TM polarizations. To measure the effective pit depth, we look at the phase of $\boldsymbol{J}_{s,e}(x = 0)$, the electric surface current at the center of the pit. Figure 5.3 plots the pit phase depth and corresponding central far field intensity (at $\sin\theta = 0$) as a function of actual pit depth. As also predicted by Dil and Jacobs, the phase of the electric field at the bottom of the pit is less for the TE field than for the TM field.

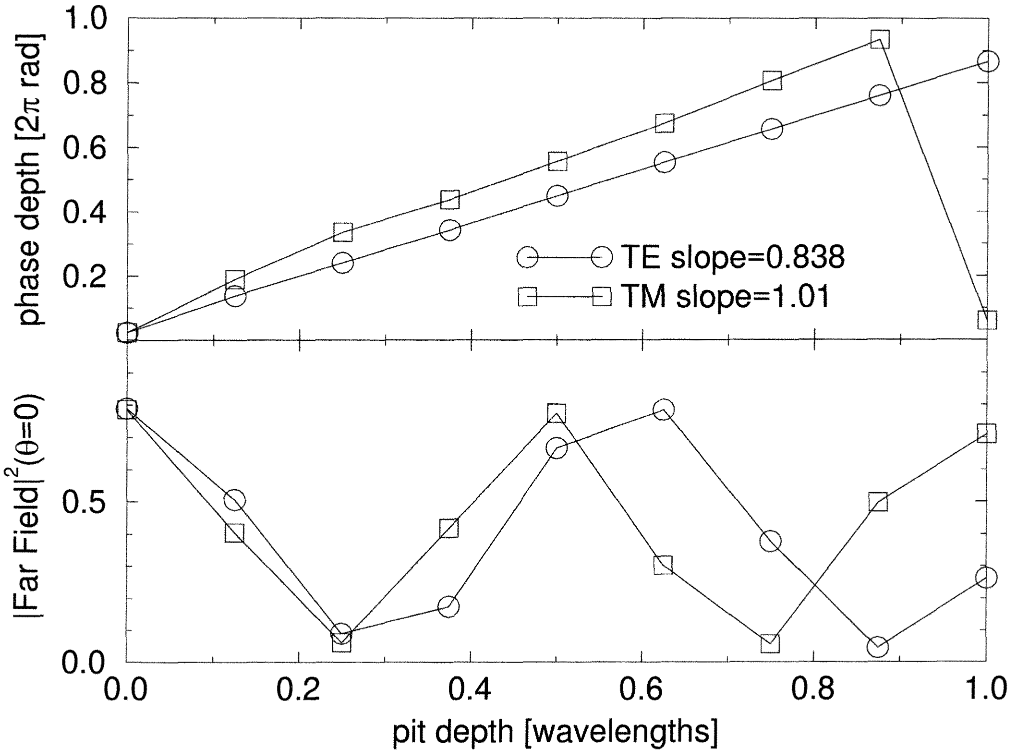


Figure 5.3: Calculated pit phase depth and far field intensity for the audio CD format. The pit phase depth is the the phase of the $J_{s,e}$ at the center of the pit. The bottom graph is the intensity in the center of the far field ($\sin\theta = 0$).

In both cases, the phase depth is quite linear with respect pit depth, implying that the phase grating principle is quite accurate. The far field intensity is also consistent since it varies approximately sinusoidally with twice the phase depth of the pit. Note that the slope of the phase depth change is slightly less for the TE than for the TM, and this difference has the expected effect on the far field intensity. The conclusion of our simulations for the CD format is that the scalar approximations used in the literature to design and explain the format is quite satisfactory.

5.2.2 DVD (the new Digital Video Disc Format)

The improvement in track density from CD to DVD is quite significant. Not only is the objective numerical aperture increased and the wavelength decreased to produce a smaller spot width, but the track pitch is significantly less than the spot width. As with the CD format, the pit width is about one-third the spot size, but since the spot size for the DVD format was decreased by increasing the numerical aperture, the pit width relative to the wavelength is smaller for the DVD format than for the CD format. This reduction in the ratio of the pit width to the wavelength has a significant effect on the phase grating view as described for the CD format. The phase depth of the pit is now quite different for the TE and TM fields, as is shown in figure 5.4. The phase of the TM field is still an accurate representation of the actual pit depth, but the TE field does not seem to be as strongly affected. The central far field intensity has the appropriate variation corresponding to the pit phase depth. This difference in the far field intensity variation for the different polarizations stimulates an interesting idea for the doubling of the information density. Rather than each pit having two possible depths, zero or one-quarter wave as in the current DVD format, each pit can have four possible depths, and the depth can be distinguished by the relative far field response of the TE and TM polarizations. The dashed vertical lines in far field plot of figure 5.4 indicate pit depths where four states are distinguishable. By placing a polarizing beamsplitter in the return path and duplicating the standard detector array in each leg, the reflection of each polarization can be measured independently,

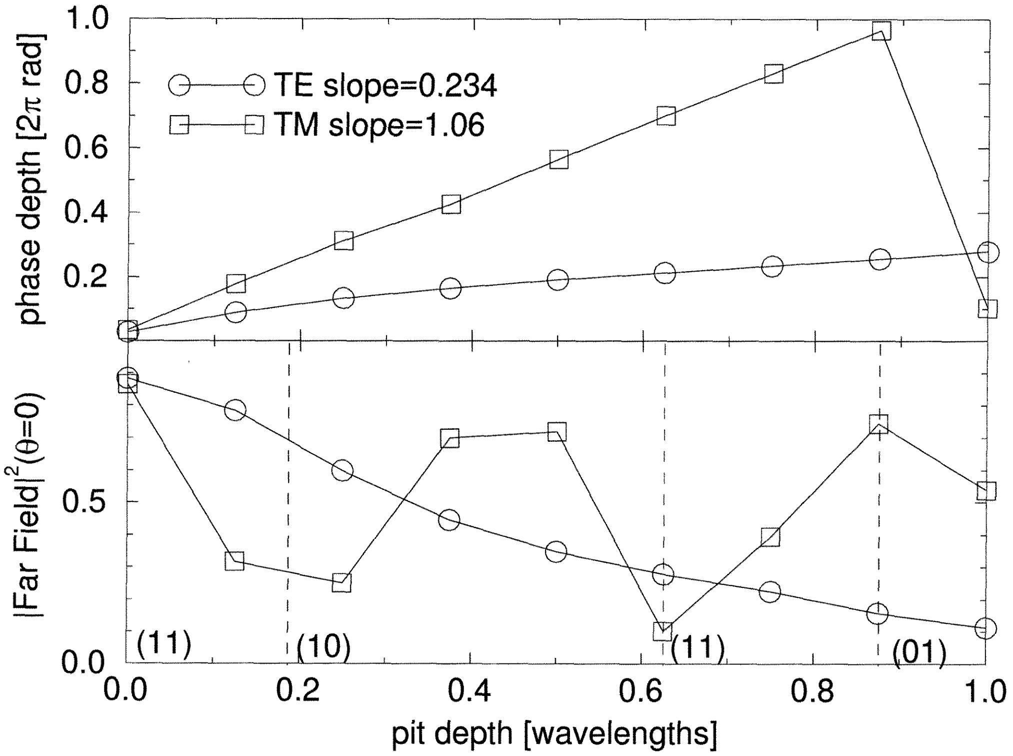


Figure 5.4: Calculated pit phase depth and far field intensity for the DVD format. The dashed vertical lines indicate which pit depths can be used to represent four logical states.

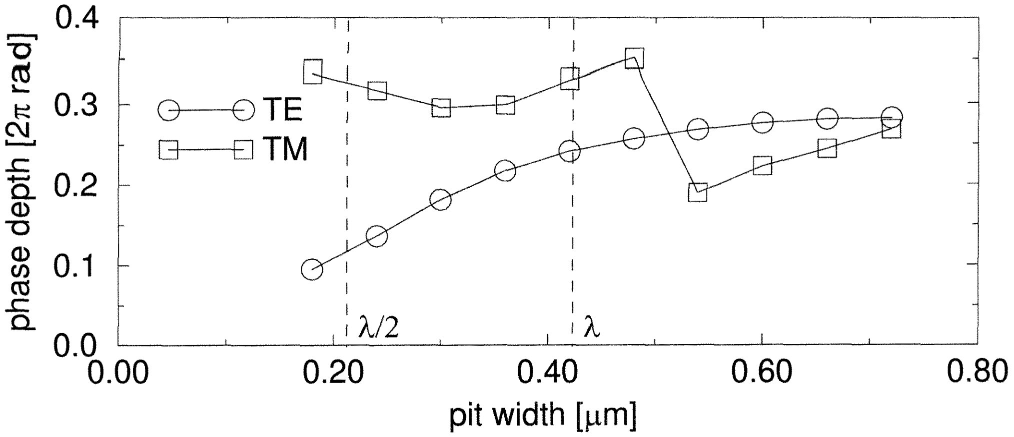


Figure 5.5: Pit phase depth for different pit widths with the pit depth equal to a quarter wave. In the region where the pit width is between $\lambda/2$ (λ is adjusted for the index of refraction of the incident medium) and λ, the phase depth for TE illumination varies almost a quarter wave, but it remains fairly constant for TM illumination.

doubling the information density of each pit.

Why should the TE polarization behave so differently for the CD and DVD formats? The biggest change between the two formats, from the point of view calculating diffraction, seems to be the pit width in relation to the wavelength. For both formats, the pit width is approximately one-third (λ/NA), but the numerical aperture increases by one-third for the DVD format, thereby reducing the pit width in relation to the wavelength. To study the effect of changing the pit width, we calculated the pit phase depth for the DVD format as a function of pit width (figure 5.5). In the region where the pit width varies between $\lambda/2$ and λ (here λ is the wavelength in the polycarbonate medium), the phase depth of the TM wave remains fairly constant while the phase depth of the TE wave increases close to the quarter wave depth of the pit. Therefore, the pit width appears to be crucial to the penetration of the TE wave into the pit.

We can explain the difference between the TE and TM dependence on the pit width by considering the case of a plane, parallel waveguide with perfectly conducting walls. The fields in this waveguide consists of discrete modes, as restricted by the boundary conditions of the tangential electric field at the walls. The TE polarization for the pit corresponds to the electric field parallel to the walls of the waveguide, and the solutions for the electric field inside the waveguide are,

$$E_y = \sin(p\pi x/d)e^{jk_z z}, \quad k_z = \sqrt{k^2 - (\frac{p\pi}{d})^2}, \tag{5.2}$$

where d is the width of the waveguide, and p is an integer. When the magnetic field is parallel to the walls, the solutions are,

$$H_y = \cos(p\pi x/d)e^{jk_z z}, \quad k_z = \sqrt{k^2 - (\frac{p\pi}{d})^2}. \tag{5.3}$$

Each mode, corresponding to an integer value of p, has a cutoff width, $d_c^p = p\pi/k$. For widths smaller than d_c^p, k_z is imaginary, and the field amplitude weakens exponentially in z. For the mode $p = 0$, the TE solution is $E_y = 0$; there can be no TE fields. However, there is a TM solution for $p = 0: \quad H_y = e^{jkz}$. The cutoff width for the $p = 1$ mode is $d_c^1 = \lambda/2$, and both TE and TM have propagating solutions for this mode when $d > d_c^1$. Therefore, when $d < \lambda/2$, only evanescent TE waves can exist while the TM zero mode can still propagate. Here lies the essential difference between the TE fields and the TM fields in our pit width calculations described above. When the pit width falls below half a wavelength, the TE fields can no longer penetrate the pit, while the TM fields can still propagate down the pit and back. Our basic principle, that we will exploit with our designs in the next section, is that by making the pit width small, we can have pits that the TM field sees, and the TE field does not.

5.3 The Nature of the TM/TE Signal

By making the pit width small enough, we have uncovered a powerful new design parameter with which we can increase the information capacity of each pit. The behavior of the TE and TM reflected fields can be affected in very different ways. In this section, we will further investigate the nature of the difference in reflected far field for the two polarizations. In particular, we would like to know just how closely the TM far field matches the scalar prediction. If the match is reasonable,

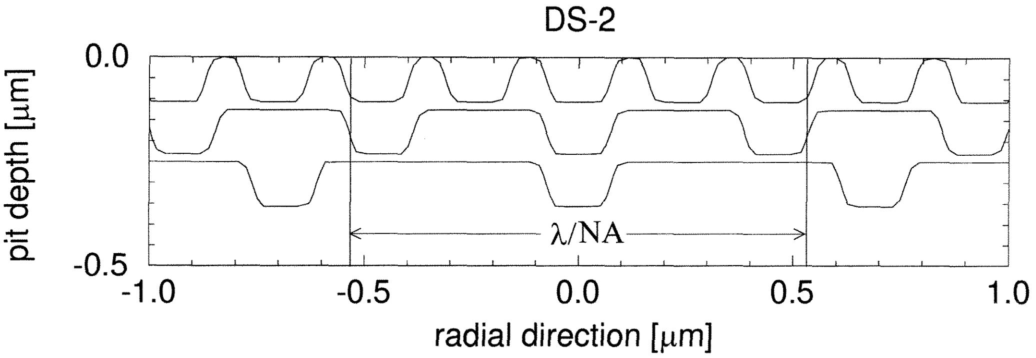


Figure 5.6: The DS-2 formats test the effect of variable track pitch using narrow pits.

then designing a format to create a specific far field signal would become very simple. In the following two simulations, we test the affect on the far field caused by different pit spacing and pit depth. Each simulation will use pit widths on the order of half wavelength, and the TM far field will be compared to a scalar model.

5.3.1 DS-2

Any format that uses only one pit depth has an obvious advantage in fabrication. If we also want to keep the pit width small, then the only parameter left is the distance between pits, and this is the parameter that is varied in the DS-2 format. Figure 5.6 shows the variation of the period used in the simulation, along with the spot size. The pit depth was set to a quarter wave to maximize the far field modulation. Figure 5.7 shows the central far field power and phase for the TE, TM and scalar cases. The scalar points were calculated with a simple model treating the pit structure as a phase mask. We used the formula

$$\mathrm{ff}_{\mathrm{sc}}(0) = \sum \mathrm{sinc}(\pi NA x/\lambda) e^{j4\pi n_0 z(x)/\lambda} \Delta x \tag{5.4}$$

for the scalar calculation. The far field phase for the scalar calculation agrees very well with the rigorous TM calculation. However, the far field power for each seems to have exactly the opposite trend. For both amplitude and phase, the TE field seems to be

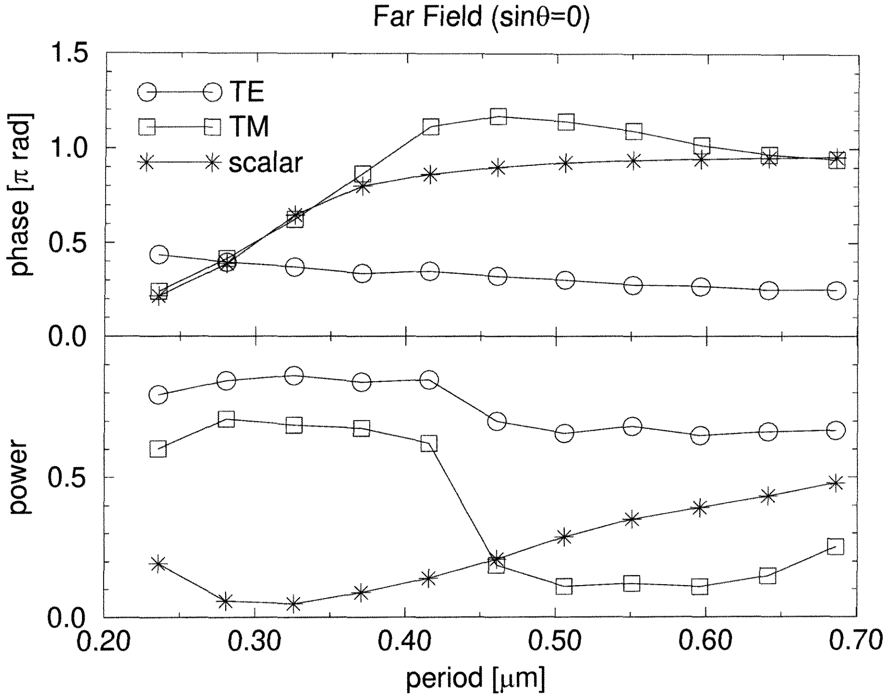


Figure 5.7: Power and phase in the center of the far field for the DS-2 format. The scalar calculation agrees well with the TM case in phase, but very poorly in power.

rather unaffected by the pits, as we expected. Our lessons from the DS-2 calculation are: a scalar model predicts the TM phase but not the amplitude; modulation of the distance between pits alone cannot modulate the far field phase the full range; and we can ignore the small region between the pits when using a scalar model to predict the TM far field phase.

5.3.2 DS-3

Using a small distance between pits, so that the illumination spot covers several pits, has at least the potential for significant gains in information density. We simply need to find a way to distinguish between various states. Therefore, in this format we keep the distance between pits small, and vary the depth of the pits. We know from the DS-2 format that with the pits close together and a quarter wave deep, the TM far field will have a large amplitude, even though the scalar model predicts a small amplitude. Without making an attempt to explain this phenomenon, we will simply

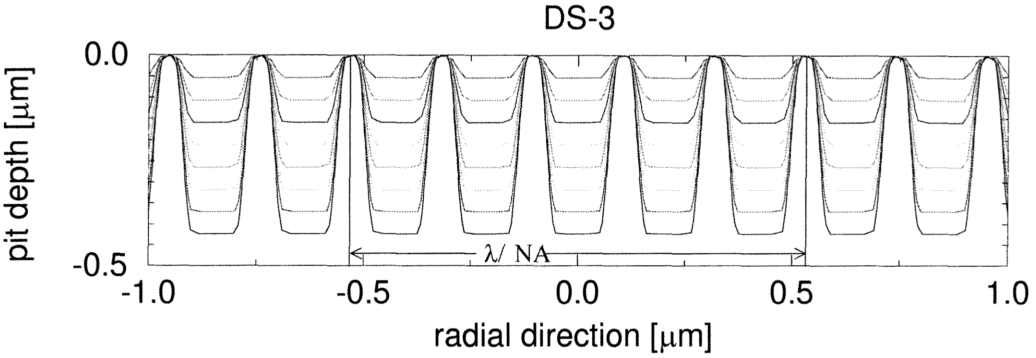


Figure 5.8: The DS-3 formats test the effect of variable pit depth using narrow pits with a high duty cycle.

modify our model by stating that the TM field is not affected by the small regions between the pits. With this adjusted model, we predict that with closely spaced pits of the same depth, the TM field will be phase modulated by the pits, and it will also have a large reflection in amplitude since there is no diffraction from an, effectively, planar surface. If the TM far field amplitude remains high, and the TM far field phase continues to be modulated by the pit depth, then we have discovered a method to modulate the TM phase with respect to the TE phase. We test this model with the DS-3 format. The pit structures for the DS-3 calculations (figure 5.8) consist of closely spaced pits, all with the same depth. The calculations are repeated for different depths in steps of one-eighth wavelength.

The first question is does the far field intensity remain approximately constant for all the pit depths. The graphs in figure 5.9 plot the far field intensity as a function of far field angle and pit depth. The far field intensity seems to fall gradually with increasing pit depth, probably as a result of diffraction with the pit walls. One important point is that the intensities for the TE and the TM fields behave similarly.

The far field phases for the TE and TM far fields is quite different, however (figure 5.10). Except for the step in pit depth, the TE far field phase remains almost constant, while the TM far field phase retains its linear dependence. This result is quite striking — while the TM far field is phase modulated by the pit structure, the

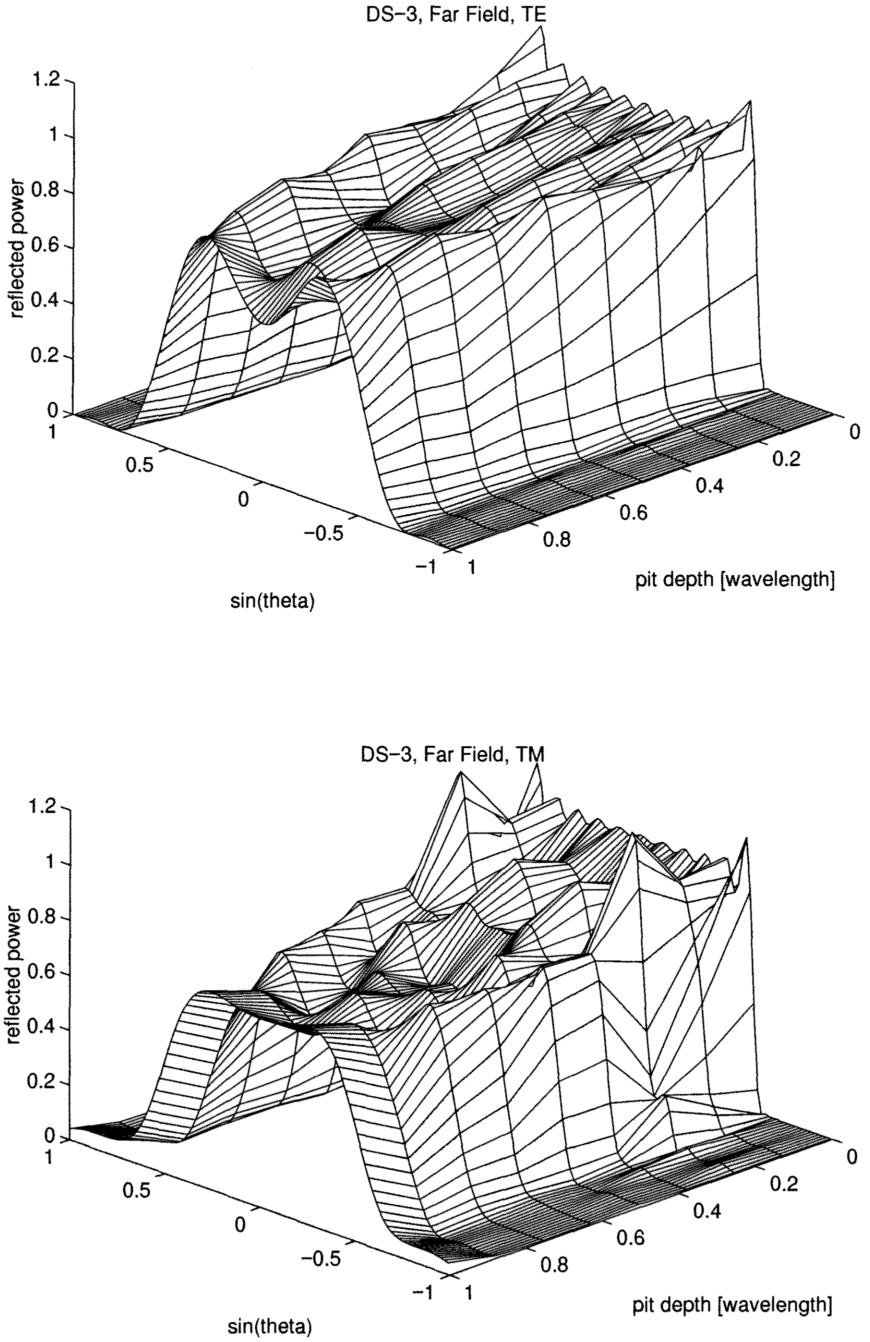


Figure 5.9: Far field intensity as a function of far field angle (sinθ) and pit depth for the DS-3 format.

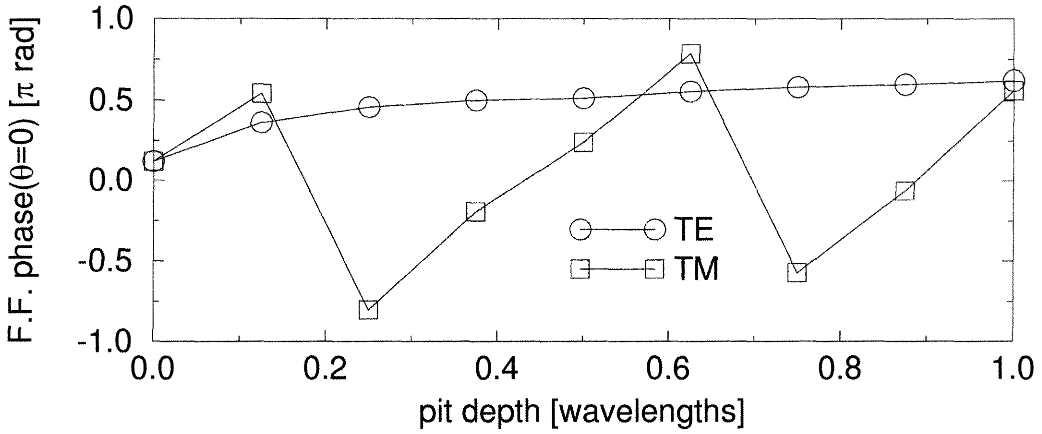


Figure 5.10: Electric field phase at the center of the reflected far field ($\sin\theta = 0$) for the DS-3 format with varying pit depth.

TE far field remains unaffected, and both retain large far field intensities for all the pit depths. The obvious suggestion is some sort of phase modulation of the TM field using the TE field as a reference, and this approach is pursued in the next section.

As a result of the DS-2 and DS-3 simulations, we have a working model of the diffraction of the TE and TM fields with these structures of small pits. As long as the pit width is on the order of half the wavelength, the TE field is unaffected by the pit structure (except for the first one-eighth wave in pit depth), and its far field amplitude and phase are approximately the same as if there were no pits. The TM field, on the other hand, seems to follow a scalar diffraction rule if we consider only the reflection from the pit bottoms and ignore the small regions between the pits. With this diffraction model, we can proceed to design some formats that have desirable features.

5.4 TM/TE Signaling with Quadrature Detection

Now that we know how to modulate the amplitude and phase of the TM far field with respect to the TE far field, we need to invent a method of detecting this modulation. The next section explains our proposal for detecting the TM field in quadrature using the TE field as a reference. This detection scheme has an advantage over interferometric detection because of the ease with which both the real and imaginary parts of the TM field can be detected. This type of detection allows full two-dimensional signaling in the manner of time domain communications systems. After the description of the quadrature detector, we design some signal constellations using our diffraction model developed in the last section, and we calculate the far field outputs. Analogous to time domain communications systems, we display the results in signal constellations.

5.4.1 A TM/TE Quadrature Detector

The fundamental idea behind TE/TM signaling is to modulate the phase and amplitude of the TM reflected far field with respect to the TE field. We have seen with the simulations that we can design the track format so that the TM field is significantly affected, while the TE field is essentially unaffected. Normally, phase of an optical field cannot be detected without interfering it with a reference optical field. In TE/TM signaling, we will illuminate the diffracting structure with both TE polarized and TM polarized fields. Because the TM field is much more strongly affected by the diffraction than the TE field, we can use the TE field as a reference to analyze the phase and amplitude change in the TM field. Measuring the relationship between the two orthogonally polarized fields is similar to the conoscopic principle discussed in chapter 3. The advantage of using the TE field as a reference, rather than constructing a traditional interferometer by splitting off part of the TM field before the diffraction and interfering the two parts, is that coherence length, stability, and vibrations are not an issue. Another advantage is the ease of measuring both the in-phase and quadrature components. To measure both in-phase and quadrature components of an electromagnetic field using traditional interferometry would require a separate reference leg with an optical path length one-quarter wave different from the first reference leg.

Consider the electric field at the center of the far field propagating in the z direction. We will assume that we have a single planewave, since, in reality, we can always place an aperture on the detector in the Fourier plane to restrict spatial components.

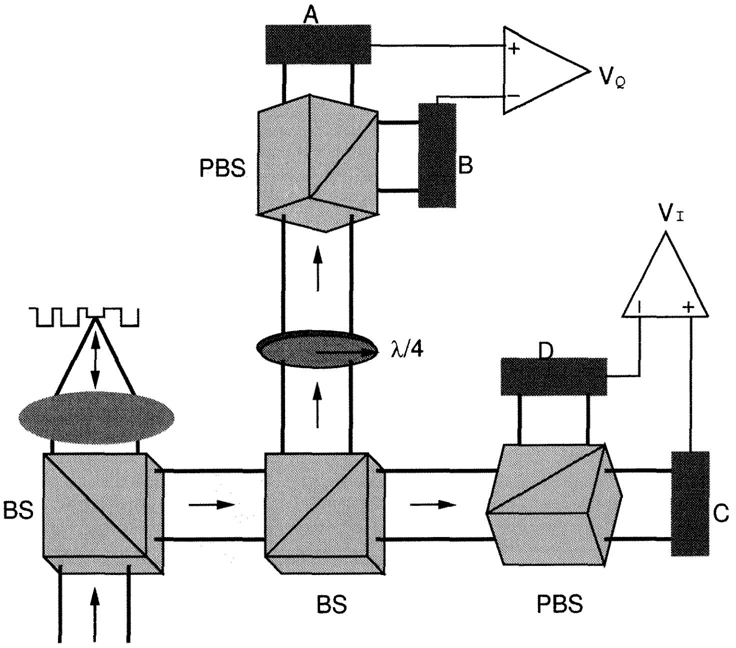


Figure 5.11: In-phase and quadrature detection of the TM far field using the TE field as a reference.

If desired, the electric fields and corresponding detector voltages used in the following formulation can be considered as functions of far field angle. In addition, we will use the convention that the TE reflected field is polarized in the x direction, so that the TM electric field points in the y direction. Then the total electric far field is

$$\boldsymbol{E}_1 = a_e e^{j\phi_e}\hat{x} + a_m e^{j\phi_m}\hat{y}, \tag{5.5}$$

where we have ignored the factor of $e^{j(kz-\omega t)}$. The quadrature detector illustrated in figure 5.11 is comprised of four detectors. The in-phase leg consists of a polarizing beam splitter rotated 45° with respect to the x axis. This rotated analyzer "interferes" the TE and TM polarized waves as in the conoscopic system. The quadrature leg is the same as the in-phase leg except that it also retards the TE component by one-quarter wave with respect to the TM component, so that the electric field in this leg is

$$\boldsymbol{E}_2 = a_e e^{j(\phi_e+\pi/2)}\hat{x} + a_m e^{j\phi_m}\hat{y}. \tag{5.6}$$

The $\pi/2$ phase shift in the reference results from the quarter-wave retardation and leads to the quadrature detection. In both legs, the signal bias due to measuring intensity is removed by subtracting the complementary outputs of the polarizing beamsplitters. The voltages at detectors A and C are

$$V_{A,C} \propto \left| \boldsymbol{E}_{1,2} \cdot (\hat{x} + \hat{y})/\sqrt{2} \right|^2, \tag{5.7}$$

and the voltages at detectors B and D are

$$V_{B,D} \propto \left| \boldsymbol{E}_{1,2} \cdot (-\hat{x} + \hat{y})/\sqrt{2} \right|^2. \tag{5.8}$$

The differential amplifiers in each leg subtract the bias terms inherent in measuring intensity. The resulting in-phase and quadrature voltages appearing at the outputs of the differential amplifiers are

$$V_I = V_A - V_B = a_e a_m \cos(\phi_e - \phi_m) \tag{5.9a}$$

$$V_Q = V_C - V_D = a_e a_m \sin(\phi_e - \phi_m). \tag{5.9b}$$

As long as a_e and ϕ_e remain stable, this quadrature detector will measure both the in-phase and quadrature components of the TM field. Because the TE field is also slightly modulated by the diffraction, however, we should keep in mind when designing the signal constellation that the TM field should be modulated relative to the TE field. If a_e is also somehow modulated, or is noisy, it can be normalized by adding to the detector a third leg with a polarizer positioned to pass the TE field and block the TM field. This leg will simply measure a_e, so that it may be divided out of V_I and V_Q. There is no simple way to also account for instability in the phase of the TE field.

5.4.2 DS-3.1

Since the quadrature detector allows the detection of the phase of the TM field with respect to the phase of the TE field, we can modify the DS-3 format into a phase shift keyed (PSK) channel. In an eight symbol PSK constellation (8-PSK), each symbol has unit amplitude but has a phase of $i\pi/8$, for integer i. Because the symbol phase represents the phase difference between the TE and TM reflected far fields, we can choose the pit depths for each symbol from figure 5.10. The following table lists the symbol phase and the corresponding pit depth used to create the DS-3.1 8-PSK format:

symbol #	$\phi_e - \phi_m$	depth	symbol #	$\phi_e - \phi_m$	depth
0	0	0	4	$-\pi$	0.32λ
1	$-\pi/4$	0.13λ	5	$-5\pi/4$	0.38λ
2	$-\pi/2$	0.20λ	6	$-3\pi/2$	0.44λ
3	$-3\pi/4$	0.26λ	7	$-7\pi/4$	0.50λ

Table 5.1: Quadrature phase and corresponding pit depth.

We calculated the diffracted far fields for the DS-3 pit structure (figure 5.8) using the pit depths in table 5.1. The in-phase and quadrature detector voltages were then calculated and plotted in the form of a signal constellation (figure 5.12). In this figure we plotted V_I along the real axis and V_Q along the imaginary axis for each calculated far field point where the far field angle $|\sin\theta| < 0.2$. Restricting the far field angle simply means that an aperture is placed in the Fourier plane. For the pit depth equal to zero, the constellation point lies on the real axis because there is no phase difference between the two polarizations. As the pit depth increases, the constellation points move in a clockwise direction because the phase of the TM far field is increasing faster than the TE far field phase. The result is a circular constellation characteristic of PSK modulation.

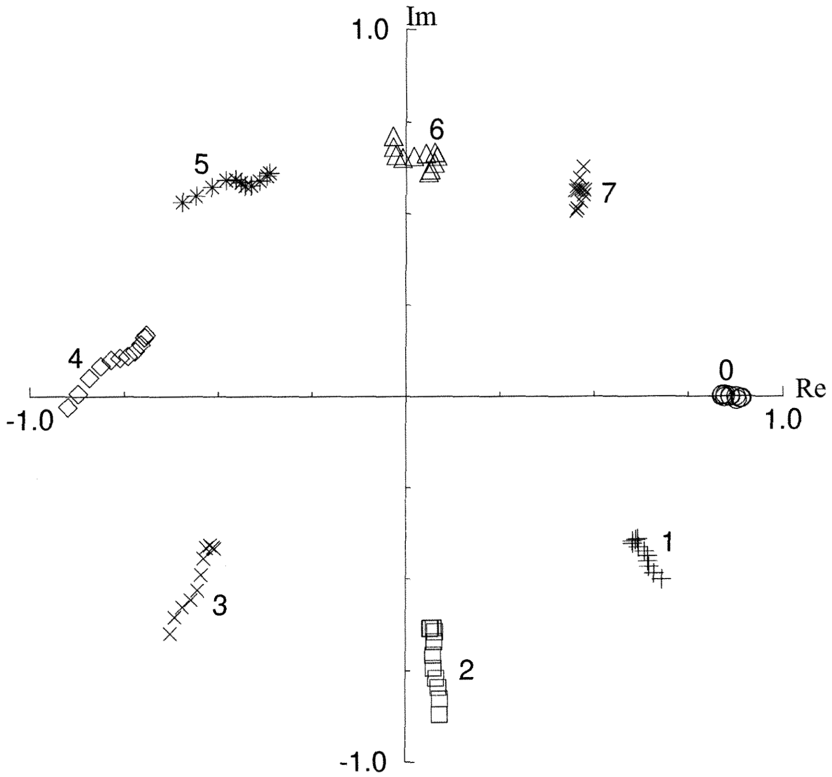


Figure 5.12: The in-phase (real) and quadrature (imaginary) constellation showing phase modulation for the DS-3 format. Each constellation point is constructed with the quadrature detector calculated points for the various far field angles, and each is labeled with the corresponding pit depth (in wavelengths). The constellation was generated by applying the quadrature detector to the calculated TE and TM far fields inside a numerical aperture of 0.2.

5.4.3 DS-6.0 8-PAM

The fabrication of multiple pit depths may prove costly. What sort of signal constellations are achievable if we are restricted to a single pit depth, but we use a DS-3 type of format with several closely spaced small pits? The limitation of only one pit depth implies that each pit can only contribute a phase of zero (zero pit depth) or a far field phase equal to twice the effective pit depth. However, since the amplitude of the incident field has a nonuniform shape across the track, some pits impress their phase shift more strongly than others onto the far field. The modulation on the center of the TM far field, using a scalar approximation, is of the form

$$E_{\mathrm{TM}}(u) = \int_{-\infty}^{\infty} \mathrm{sinc}(k\mathrm{NA}x) \sum_i \mathrm{rect}(\frac{x - i\nu}{\tau}) e^{ja_i 2\phi} e^{jkux}\, dx, \tag{5.10}$$

where a_i is 1 or 0, ν is the spacing between pits, τ is the width of each pit, and ϕ is the effective phase depth of a pit. Because the *rect* functions do not overlap (the pits are disjoint) we can rewrite the far field expression to consider the effect of each pit separately as

$$E_{\mathrm{TM}}(u) = \sum_i e^{ja_i 2\phi} \int_{i\nu-\tau/2}^{i\nu+\tau/2} \mathrm{sinc}(k\mathrm{NA}x) e^{jkux}\, dx, \tag{5.11}$$

ignoring the small space between pits. Each integral in equation 5.11 is essentially a weighting for each pit depending on the strength of the incident field on that pit. Since the phase contribution of each pit is restricted to either zero or 2ϕ, the possible signal constellations are straight lines. One end point of the constellation is for the symbol where each $a_i = 0$, and the other end point is for each $a_i = 1$. Therefore, the optimum phase depth is $2\phi = \pi$, producing a signal constellation along the real axis, and this type of constellation is known as Pulse Amplitude Modulation (PAM).

The next step for this design is to choose sets of a_i which form distinct constellation points. Since the central one-third of the illuminating spot contains about half of the illumination power, we will divide the track into six pits, as shown in figure 5.13, rather than five as in the DS-3 formats. To estimate the relative effect of each pit on

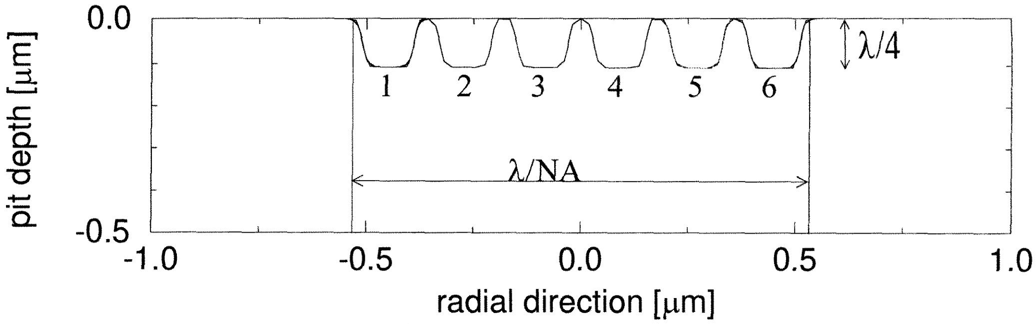


Figure 5.13: Track format for DS-6.0 PAM modulation.

the far field, we calculated the integral in equation 5.11 at $u = 0$, the center of the far field, for each pit, resulting in the simple approximation,

$$E_{\mathrm{TM}}(0) \; = \; 0.99[e^{ja_3\pi} + e^{ja_4\pi}] + 0.66[e^{ja_2\pi} + e^{ja_5\pi}] + 0.21[e^{ja_1\pi} + e^{ja_6\pi}]. \tag{5.12}$$

To make an eight point constellation, we want to choose eight sets of a_i so that $E_{\mathrm{TM}}(0)$ for each set is evenly spaced along the real axis. Fixing the end points at $E_{\mathrm{TM}}(0) = -1$, and $E_{\mathrm{TM}}(0) = 1$, the other six points will be at $E_{\mathrm{TM}}(0) = \{\pm 0.64, \pm 0.43, \pm 0.11\}$. The pit patterns that achieve this constellation are shown in table 5.2. Notice that the constellation points ($E_{\mathrm{TM}}(0)$) are not evenly spaced. Evenly spaced points would be separated by a distance of $2/7 \approx 0.29$, and from equation 5.12 we see that this separation distance is not possible. The tabulated choices are the best possible given

#	$E_{\mathrm{TM}}(0)$	$\{a_i\}$	#	$E_{\mathrm{TM}}(0)$	$\{a_i\}$
0	-1.0	111111	7	+1.0	000000
1	-0.64	111101	6	+0.64	000010
		101111			010000
2	-0.43	011100	5	+0.43	100011
		001110			110001
3	-0.11	111001	4	+0.11	000110
		110101			001010
		101011			010100
		100111			011000

Table 5.2: The choices of pit patterns for each signal constellation value.

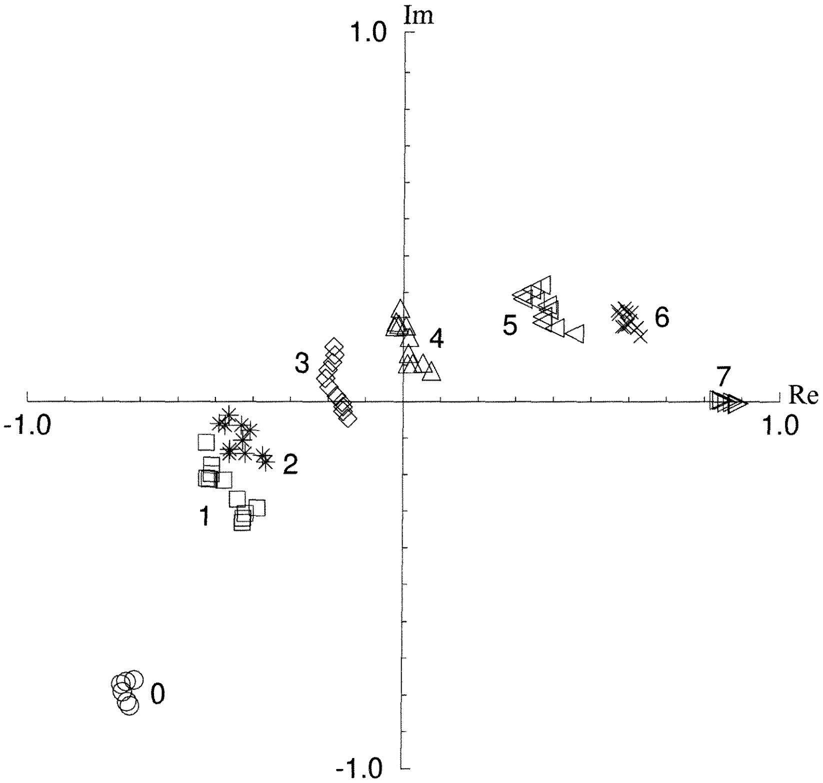


Figure 5.14: Signal constellation for the eight point DS-6.0 format with the detector $\mathcal{M} = 0.1$.

equation 5.12. Naturally, this design assumes that the TE field will be completely unaffected by the presence of the pits. We know from the DS-3 format that TE field will have some small phase change from the pits. Therefore, we should expect the calculated constellation for the above pit patterns to deviate somewhat from a straight line. Figure 5.14 shows the calculated signal constellation for the DS-6.0 PAM format. The constellation point corresponding to no pits present, symbol #7, is on the real axis, which is the correct position. However, because the TE field does become somewhat modulated by the presence of the pits, and the calculated TM field varies from the scalar design rule, the quadrature component does not remain zero. The distance between the constellation points reflects the unevenness described in the design. However, the symbols are unique and unambiguous, and we have successfully designed a signal constellation that requires only one pit depth.

5.4.4 DS-6.1 8-QAM

The formats for the 8-PSK (DS-3.1) and the 8-PAM (DS-6.0) point to the trade-off between the simplicity of having only one pit depth and a signal constellation which has greater distances between the signal points. A greater distance between constellation points implies better signal to noise ratio (SNR) which can be utilized to either improve the bit error rate or add more signal points and increase the information content. However, the 8-PSK constellation is not an optimum choice in this respect. We will now describe an 8 point Quadrature Amplitude Modulation (QAM) signal constellation that has better point separation properties and only requires half as many pit depths. Returning to the DS-6.0 8-PAM format, suppose we modulate the two central pits together with one depth and the four outside pits together with another. Then, the design equation equivalent to equation 5.12 has the form,

$$E_{\mathrm{TM}}(0) = \frac{1}{2}e^{j\phi_0}(1 + e^{j\phi_1}), \tag{5.13}$$

where the outside four pits have a phase depth equal to ϕ_0, and the central two pits have a phase depth equal to $\phi_0 + \phi_1$. Figure 5.15 plots the constellations for this design equation as a function of ϕ_1 for four different values of ϕ_0. Considering the signal points marked in the figure, we realize that only four phase depths are required. Each of the signal points corresponds to some combination of $\phi_0 = 0, \pi/2, \pi, 3\pi/2$, and $\phi_1 = 0, \pi/2, \pi, 3\pi/2$. For our simulation, we made the choices shown in table 5.3. To encode these phase depths onto a pit structure, we used six pits as in the DS-6.0 format (figure 5.13), except that pits 1,2,5,6 were given a phase depth equal to ϕ_0, and

#	ϕ_0	ϕ_1	#	ϕ_0	ϕ_1
0	0	0	4	π	0
1	$\pi/2$	$3\pi/2$	5	π	$\pi/2$
2	$\pi/2$	0	6	$3\pi/2$	0
3	$\pi/2$	$\pi/2$	7	$3\pi/2$	$\pi/2$

Table 5.3: The choices of pit phase depth patterns for the 8-QAM constellation.

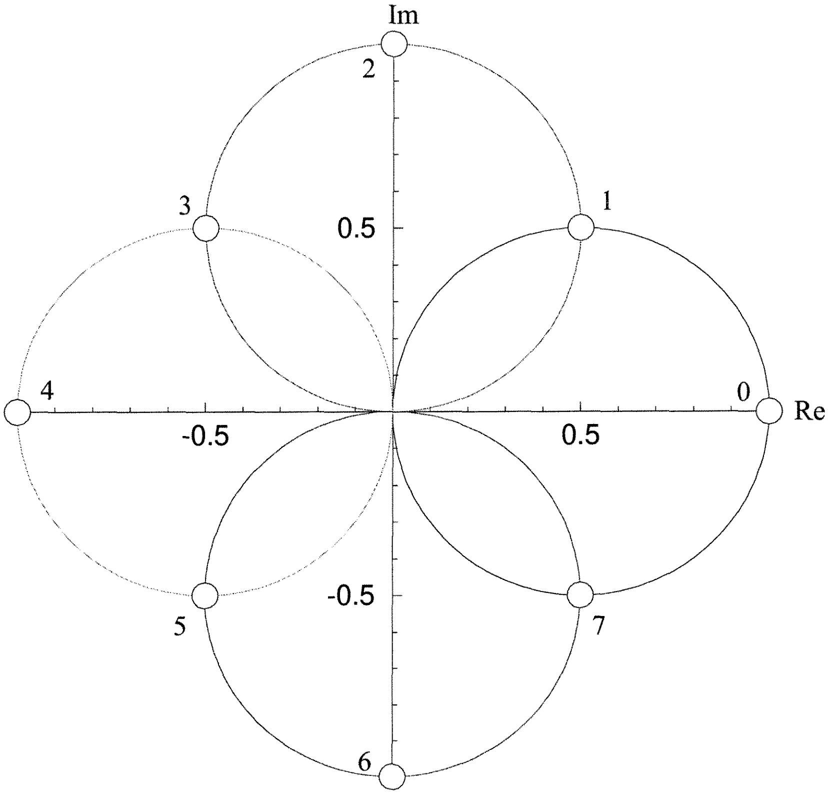


Figure 5.15: Plots of equation 5.13 for $\phi_0 = 0, \pi/2, \pi, 3\pi/2$. The signal constellation points for 8-QAM modulation are numbered.

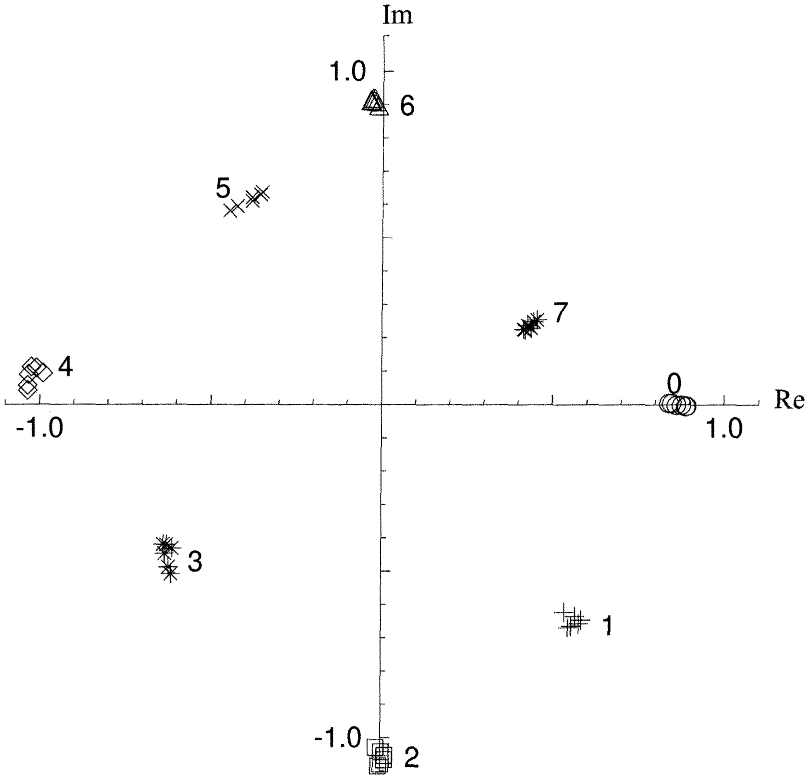


Figure 5.16: Signal constellation for the DS-6.1 format, using detector NA = 0.1.

pits 3 and 4 were set to a phase depth of $\phi_0 + \phi_1$. Notice that the phase depth choice for symbol #1 could have been $\phi_0 = 0$ and $\phi_1 = \pi/2$, but this region of shallow phase depth is where the TE polarization is most affected by changing pit depth. Therefore, better results of rotating the TM field without relative change to the TE field were attained with the phase choices given in the table. The pit depths used in the simulation were taken from table 5.1. The calculated constellation for the far field angle $|\sin\theta| < 0.1$ is shown in figure 5.16. The constellation points for symbols #0,2,4,6 are placed very close to their intended locations, but the other symbols, particularly #5,7, seem to be slightly offset. Perhaps different choices of ϕ_0 and ϕ_1 or the pit depths for these symbols would improve their locations. In general, however, the constellation demonstrates a signaling format having better symbol separation than the DS-6.0 8-PAM format with half the number of required pit depths as the DS-3.1 8-PSK format.

5.4.5 DS-6.2 8-QAM+8

When designing the 8-QAM system for the DS-6.1 format, we explained that some of the constellation points, namely symbols #1,3,5,7, could be encoded with several different choices of pit phase depth patterns. If we could find some other way to distinguish between two different pit patterns that have the same TE/TM quadrature signal, then we could add more symbols increasing the information density. In fact, there is one more degree of freedom yet to be exploited by these formats—variation in the far field pattern. In the DS-6.1 format, we chose pit patterns that were symmetric, and symmetric pit patterns result in symmetric far field patterns. If, instead, we set pits 1,2,3 to a phase depth equal to ϕ_0, and pits 4,5,6 to a phase depth of $\phi_0 + \phi_1$, then we can get the same quadrature constellation in the center of the far field, but the far field will be asymmetric for symbols #1,3,5,7. The far field would still be symmetric for symbols #0,2,4,6 because for these symbols all the pits have the same depth. Reversing the asymmetry would construct another four symbols, for a total of sixteen symbols.

However, the in-phase and quadrature voltages cannot distinguish between all the symbols. To determine the asymmetry in the far field, we measure the TM far field alone (which requires an additional leg in the quadrature detector). The TM field is used for the symmetry measure, and not the TE field, because the TM far field is more significantly affected by the pit structure than the TE, and, therefore, we would expect asymmetries to be more pronounced in the TM far field. The measurement is made by splitting the far field in half and detecting each half separately; one detector (V_-) measures the intensity for the far field angles $|\sin\theta| < 0$, and the other (V_+) for $|\sin\theta| > 0$. We define the measure of asymmetry, $M = V_+ - V_-$. The results of the sixteen point simulation is shown in a constellation plot (figure 5.17). In this representation, each of the constellation points #1,3,5,7 represents three symbols, each with a different value of M. Thus, each of the sixteen symbols can be distinguished by combination of V_I, V_Q, and M, and the format can represent four bits of information, whereas the DS-6.1 8-QAM format can only represent three bits. Naturally, the cost

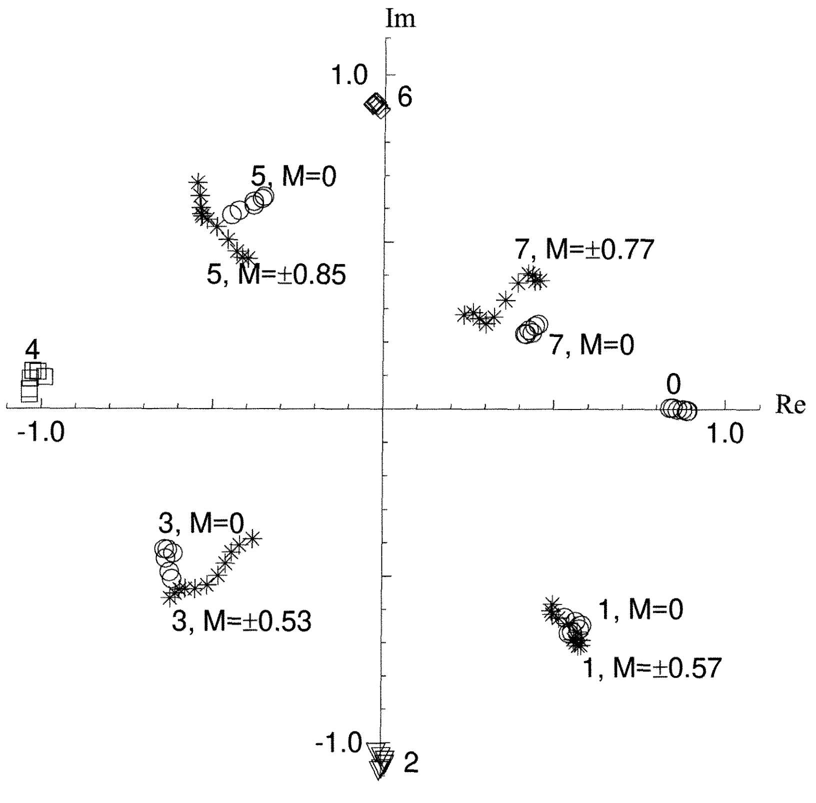


Figure 5.17: 16 point signal constellation for the DS-6.2 format. The quadrature detection is within a $\mathcal{M} = 0.1$, and the asymmetry detection is for a detector $\mathcal{M} = 0.5$.

is more complexity due to the requirements for the symmetry measure.

5.5 TE/TM Quadrature Measurement for Surface Metrology

As minimum feature sizes in microelectronic fabrication are reduced below 0.25μm, the quality control inspection required at each step of the fabrication process in a large scale microelectronics facility becomes a very difficult problem. Not only does the reduced feature size allow more devices per wafer, but with more devices, more layers of interconnections are required. Two specific features that are difficult to measure are contact holes etched through oxide layers for the purpose of contacting a circuit device with metallization, and capacitor trenches etched into the silicon substrate for

DRAM cell fabrication. Before each metallization step, contact holes are etched in the covering oxide down to the device layer where contact is desired. Because the device width can be as small as 0.25μm, and the contact hole can be micrometers in depth, these holes have a very large aspect ratio. A DRAM cell consists of a capacitor and a transistor. The information is stored by charging the capacitor, and the capacitor is accessed for reading and writing by the transistor. In general, the capacitance is proportional to the area of the capacitor, and a large capacitance is desired for a DRAM cell. However, to fit more capacitors onto a chip, the area of each capacitor must be made small, contrary to the desire for a large capacitance. To solve this problem of a capacitor that uses a small amount of chip surface area but also has the required capacitance, some DRAM processes form the capacitor by etching a deep and narrow trench, depositing a thin oxide along the walls of the trench, and then filling the trench with a conducting material. Thus, the capacitor is formed along the walls of the trench resulting in a large capacitor area. Yet, the chip surface area occupied by the trench capacitor is equal to the width of the trench, which can be as small as the lithography allows.

There are several quality control issues with both the contact hole and trench capacitor fabrication. For the contact hole, the depth of the hole is very important since the hole must make contact with the proper layer. For the trench capacitor, the depth and shape are very important, but the uniformity of the oxide layer is crucial. Currently, these structures can only be measured with an electron beam microscope. Not only in an electron beam microscope far to slow and expensive to inspect every wafer in a fabrication process, but many electron microscope techniques require cutting the wafer so that the structure's profile can be measured.

Electron microscopy is too slow and expensive to use on-line, and measuring trench depth requires destructively cutting the sample to expose its profile. Optical microscopes can barely resolve these structures, and they do not provide accurate depth information. Interference methods, such as the commercially available white light interference microscope, have excellent depth measuring ability, but they cannot measure the depth of structures smaller than the resolution limit or structures with large aspect ratios. As a result, when a wafer is etched defectively, the problem cannot be discovered until electrical tests are performed later in the fabrication process, wasting valuable time and resources on needless production steps on a defective wafer.

The previous examples in this chapter of diffraction from a groove were for aluminum structures with visible red light. At that wavelength, silicon is only about 30% reflective [4]. As a result, when a red spot is focused onto a silicon structure deeper than about a wavelength, most of the incident light that enters the groove is never reflected back out because of severe absorption by the groove walls. We saw this effect with the measurement of the gallium arsenide sample in Chapter 4. Without light reflected from the groove, there can be no polarization quadrature signal. Our solution is to illuminate the groove with DUV light. As an example, a KrF excimer laser emits at a wavelength of 248nm, and a high pressure mercury arc lamp can emit at 257nm. Both of these sources are commercially available and currently being used in microelectronic lithography. At these wavelengths, silicon is very conductive, and its reflectivity goes up to about 70% [4]. Another benefit of using the DUV light is that we can relax the numerical aperture requirements of the illuminating system, since the illuminating spot size is approximately the wavelength divided by the numerical aperture. We simulated the TE/TM quadrature detector response to a 0.25μm wide groove in silicon for increasing groove depth. We performed the simulation with a wavelength of 248nm, corresponding to a KrF excimer laser. Illuminating the groove with light at a wavelength where silicon is most reflective gives the best measurement signal since we are measuring the reflected light. At 248nm, the index of refraction of silicon is about 1.6+j3.6 [4]. Figure 5.18 shows the TE/TM quadrature response for the groove depth varying from zero to four wavelengths (0.992μm). The horizontal axis of this graph shows the in-phase voltage, V_I, from the polarization quadrature detector, and the vertical axis shows the quadrature voltage output. Each point plotted, represented by a triangle, is for a different trench depth in the calulation. Thus, the (V_I, V_Q) coordinate corresponds to the detector response for that trench depth. The graph shows how the phase difference between the TE and TM rotates as the trench deepens, causing a spiral graph. If the measurement of the trench could

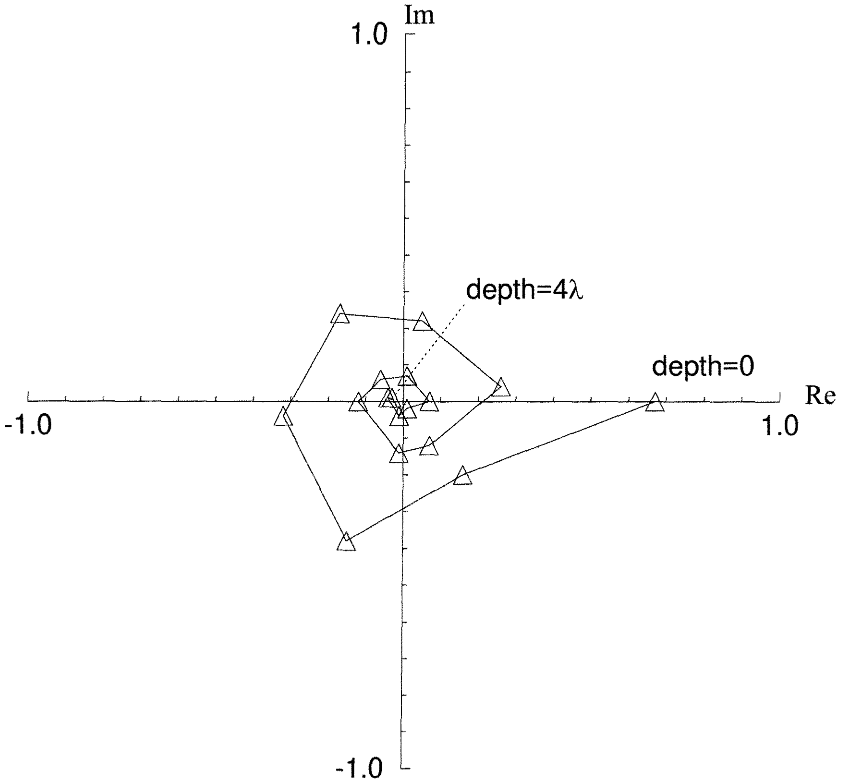


Figure 5.18: The TE/TM quadrature signal for a 0.25μm wide groove in silicon for increasing groove depth. The depth is incremented in steps of $\lambda/4$ from 0 to 4λ. The wavelength $\lambda = 0.248$nm, the numerical aperture for the incident light is 0.6, and the numerical aperture for collecting the reflected light is 0.1.

be made *in situ*, then the operator could precisely stop the etching process at the depth by watching the the detector output. If the measurement is not made *in situ*, then the an instrument would measure the voltage pair (V_I, V_Q) for each trench. If the measured voltage pair equals the expected voltage pair, then the trench passes. Otherwise, the etching process has been defective.

There is no requirement for a coherent source for the polarization quadrature measurement. The TE and TM fields travel exactly the same optical path. Therefore, coherence length is not an issue. In fact, there may be an advantage to using an incoherent source. With incoherent imaging, we can image the reflection of each trench in parallel. By placing lenses in the polarization quadrature system(figure 5.19, we can image the polarization quadrature response for each trench onto a detector

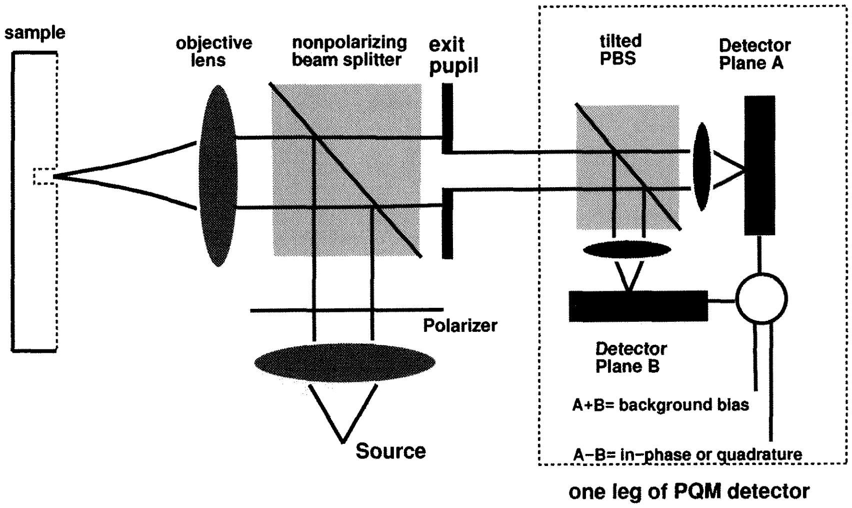


Figure 5.19: The polarization quadrature system for measuring the parameters of a trench. Only one leg of the polarization quadrature detector is shown. The added lenses before the detectors allows imaging as in a traditional microscope.

in an array of detectors. In this way, we can measure the polarization quadrature response for as many trenches in parallel as the optical imaging system and detector array can handle.

5.6 Conclusion

We started this chapter by simulating the diffraction for the current optical disc formats. While the generally accepted scalar model is reasonable for the older CD-audio format, it begins to fail for the smaller pit widths of the newer DVD format. In fact, the difference between the TE and TM diffraction for this format suggested a method of doubling the information density by using multiple pit depths and a polarizing beam splitter.

To gain a better understanding of the different diffraction effects for the TE and TM polarizations, we designed to experimental formats, DS-2 and DS-3, to test the applicability of normal scalar type diffraction models. We found that for pit widths smaller than about half a wavelength, the TE field is relatively undiffracted by the pit structure, and we explained this phenomenon by comparing the pit to a plane parallel waveguide. In this waveguide, no TE fields can propagate when the width is less than half the wavelength. For the TM diffraction, we found that describing the diffraction as phase modulation from the bottom surface of the pits, and ignoring altogether the small regions between closely spaced pits, was satisfactory in predicting the diffracted far field.

Making full use of our new-found ability to modulate the amplitude and phase of the TM far field independently of the TE field requires a special detector. Our quadrature detector measures the complete amplitude and phase of the TM far field using the TE far field as a reference. This detector can be useful in any application that uses some sort polarization modulation as a signal. Finally, we realized our ancient goal of Chapter 1; we designed a set of patterns consisting of subwavelength features that had some sort of optimal far field behavior. The far field behavior took the form of a signal constellation, and we were able to design the placement of the signal points, thereby maximizing the distance between symbols. Whereas in Chapter 2 we simply measured whatever distance happened to be present between symbols, the symbol constellations for the DS-3.1 and DS-6 formats were actually designed to have certain properties.

Bibliography

[1] D. Gabor, "Light and Information," in *Progress in Optics,* ed. by E. Wolf, Vol. 1, North-Holland Publishing Company, Amsterdam, 1961.

[2] L.B. Lucy, "Statistical Limits to Superresolution," Astronomy and Astrophysics, Vol. 261, p. 706, 1992.

[3] G. Bouwhuis and J.H.M. Spruit, "Optical Storage Read-Out of Nonlinear Disks," Applied Optics, Vol. 29, p. 3766, 1990.

[4] T. Yanagisawa and S. Ohsawa, "Superresolution in Optical Disk Systems with a Non-Linear Refraction Layer," Jpn. J. Appl. Phys., Vol. 32, p. 1971, 1993.

[5] K. Aratani, A. Fukumoto, M. Ohta, M. Kaneko, and K. Watanabe, "Magnetically Induced Super Resolution in Novel Magneto-Optical Disk," in *Optical Data Storage,* SPIE Vol. 1499, p. 209, 1991.

[6] S. Kobayashi, T. Horigome, J. P. deKock, H. Yamatsu, and H. Ooki, "Single Carrier Independent Pit Edge Recording," in *Optical Data Storage,* SPIE Vol. 2514, p. 73, 1995.

[7] T. Maeda, H. Sugiyama, A. Saitou, K. Wakabayashi, H. Miyamoto, and H. Awan, "High-Density Recording by Two-Dimensional Signal Processing," in *Optical Data Storage,* SPIE Vol. 2514, p. 70, 1995.

[8] H. Ooki, R. Arimoto, Y. Iwasaki, and J. Iwasaki, "A Novel Super Resolution Technique for High-Density Optical Data Storage using Mode Interference in Channel Waveguides," Jpn. J. Appl. Phys., Vol. 32, p. 1668, 1993.

[9] W. Zeng and J. Moon, "Modified Viterbi Algorithm for a Jitter-dominant $1-D^2$ Channel," IEEE Transactions on Magnetics, Vol. 28, p. 2895, 1992.

[10] Sirat and Psaltis, “Conoscopic Holography,” Optics Letters, Vol.10, p.4, 1985

[11] Sirat and Psaltis, “Conoscopic Holograms,” Optics Communications, Vol.65, p.243, 1988.

[12] M. Born and E. Wolf, *Principles of Optics,* Sixth Edition, Pergamon Press, 1989.

[13] M. Born and E. Wolf, *Principles of Optics,* Pergamon Press, New York, 1975, Ch. 14.

[14] R. Blahut, *Principles and Practice of Information Theory,* Addison-Wesley, 1987.

[15] R. McEliece, *The Theory of Information and Coding,* Cambridge University Press, 1984.

[16] J. Goodman, *Introduction to Fourier Optics,* McGraw-Hill Publishing Company, 1968.

[17] J. G. Dil and B. A. Jacobs, “Apparent Size of Reflecting Polygonal Obstacles of the Order of One Wavelength,” J. Opt. Soc. Am., Vol. 69, No. 7, p. 950, 1979.

[18] A. T. De Hoop, *Modern Topics in Electromagnetics and Antennas,* PPL Conference Publication 13, Peter Peregrinus Ltd., Stevenage, England, 1977, Ch. 6.

[19] J. Pasman, in *Principles of Optical Disc Systems,* Ch. 6, Adam Hilger Ltd., 1985.

[20] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes in C,* Cambridge University Press, Cambridge, 1988.

[21] J. Braat, *Principles of Optical Disc Systems,* Ch. 2, Adam Hilger Ltd., 1985.

[22] R. Petit, several papers on integral method for gratings, and paper on a single obstacle in a planar interface.

[23] A.B. Marchant, *Optical Recording,* Addison-Wesley Publishing Company, Inc., 1990.

[24] Wirgin, Optics Communications, Vol. 27, no. 2, p.189, 1978.

[25] L.M. Delves and J.L. Mohamed, *Computational Methods for Integral Equations,* Cambridge University Press, Cambridge, 1985.

[26] P.C. Waterman, "Numerical Solution of Electromagnetic Scattering Problems," in *Computer Techniques for Electromagnetics,* edited by R. Mittra, Pergamon Press, Ltd., 1973.

[27] A.J. Poggio and E.K. Miller, "Integral Equation Solutions of Three-Dimensional Scattering Problems," in *Computer Techniques for Electromagnetics,* edited by R. Mittra, Pergamon Press, Ltd., 1973.

[28] R.F. Harrington, *Field Computation by Moment Methods,* Macmillan, New York, 1968.

[29] F.W.J. Olver, "Bessel Functions of Integer Order," in *Handbook of Mathematical Functions,* ed. by M. Abramowitz and I. Stegun, Dover Publications, Inc., New York, 1965.

[30] *Handbook of Optical Constants of Solids,* ed./ by Edward D. Palik, Academic Press, Inc., 1985.

[31] *Electromagnetic Theory of Gratings,* ed. by R. Petit, Springer-Verlag, Berlin, New York, 1980.

[32] R. W. Wood, "On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum," Philos. Mag., Vol. 4, p.396, 1902.

[33] R.F. Harrington, *Time-Harmonic Electromagnetic Fields,* McGraw-Hill, 1961.

[34] M.G. Moharam and T.K. Gaylord, "Rigorous Coupled-Wave Analysis of Planar-Grating Diffraction," J. Opt. Soc. Am., Vol. 71, No. 7, p. 811-818, 1981.

[35] M.G. Moharam and T.K. Gaylord, "Diffraction Analysis of Dielectric Surface-Relief Gratings," J. Opt. Soc. Am., Vol. 72, No. 10, p.1385-1392, 1982.

[36] T.K. Gaylord and M.G. Moharam, "Planar Dielectric Grating Diffraction Theories," Appl. Phys. B, Vol. 28, p.1-14, 1982.

[37] W.B. Veldkamp, G.J. Swanson, S.A. Gaither, C-L. Chen, and T.R. Osborne, *Binary Optics: A Diffraction Analysis,* Project Report ODT-20, Massachusetts Institute of Technology Lincoln Laboratory, Lexington, Massachusetts, 1989.

[38] Lifeng Li and Charles W. Haggans, "Convergence of the Coupled-Wave Method for Metallic Lamellar Diffraction Gratings," J. Opt. Soc. Am. A, Vol. 10, No. 6, p.1184-1189, 1993.

[39] Toshiba SD Comparison Chart.

[40] J. W. S. Rayleigh, "Note on the Remarkable Case of Diffraction Spectra Discovered by Prof. Wood," Philos. Mag., Vol. 14, p.60, 1907.

[41] K. Knop, "Rigorous Diffraction Theory for Transmission Phase Gratings with Deep Rectangular Grooves," J. Opt. Soc. Am., Vol. 68, p.1206, 1978.

[42] F.G. Kaspar, "Diffraction by Thick, Periodically Modified Gratings with Complex Dielectric Constant," J. Opt. Soc. Am., Vol. 63, p.37, 1973.

[43] H.G. Schmidt-Weinmar, "Spatial Resolution of Subwavelength Sources from Optical Far-Zone Data," in *Inverse Source Problems,* ed. by H.P. Baltes, Springer-Verlag, Berlin, 1978.

[44] H.G. Schmidt-Weinmar, "Optical-Wave Near Field Specified from Far-Field Data," Journal of the Optical Society of America, Vol.65, p.1059, 1975.

[45] D.K. Lam, H.G. Schmidt-Weinmar, A. Wouk, "The 3-D Distribution of Sources of Optical Scattering Computed from Complex-Amplitude Far-Field data," Canadian Journal of Physics, Vol.54, p.1925, 1976.

[46] W. Lukosz and A. Wuthrich, "Holography with Evanescent Waves," Optik, Vol.41, p.191, 1974.

[47] H. Nassenstein, "Interference, Diffraction and Holography with Surface Waves ("Subwaves"). I.," Optik, Vol.29, p.597, 1969.

[48] H. Nassenstein, "Interference, Diffraction and Holography with Surface Waves ("Subwaves"). II.," Optik, Vol.30, p.44, 1969.

[49] O. Bryngdahl, "Holography with Evanescent Waves," Journal of the Optical Society of America," Vol.59, p.1645, 1969.