NASA-CR-191307

NAG5-1346

OPTICAL DATA STORAGE CENTER

University-Industry Cooperative Research

Annual Report: March 15, 1992

Research studies on advanced optical module/head designs for optical

(NASA-CR-191307) RESEARCH STUDIES ON ADVANCED OPTICAL MODULE/HEAD DESIGNS FOR OPTICAL DATA STORAGE Annual Report, period ending 15 Mar. 1992 (Arizona Univ.) 265 p

N93-15505
--THRU--
N93-15520
Unclas

G3/74 0137281

Optical Sciences Center
University of Arizona
Tucson, Arizona

FROM THE DIRECTOR

This report should reach you about one month before our semiannual meeting at the Viscount Suite Hotel on April 22, It begins with brief summaries of progress and plans by each of the eight faculty who continue to participate in ODSC: Armstrong, Falco, Kostuk, Leavitt, Li, Mansuripur, Milster and Zelinski. Fifteen appendices follow the summary reports. Many are preprints of papers presented at the recent ODS '92 meeting in San Jose, where approximately 10% of the papers were presented by ODSC faculty and students. Several of these students expect to complete their graduate studies by the end of the summer: Roger Hajjar, Andy Zhou, Bruce Bernacki, Ron Roncone and Mark Shi Wang. We and they are very grateful to our sponsors for continuing support during their four- to five-year research program. That is the kind of support that is needed to do the research necessary to earn the doctoral degree.

I had hoped to have been able, in this report, to provide you with ODSC's financial plans for the next fiscal year. Unfortunately, neither the data nor the decisions are available at this time. We have been informed by the President of the University that there will be a 6% budget cut for the fiscal year beginning July 1. An additional 3% cut will fall on some University units. We do not yet know how these axes will fall. President Pacheco also has announced a reorganization of University management effective July 1. How OSC, and therefore ODSC, will fit into the new structure, has apparently not been decided. I hope, but cannot promise, that these uncertainties will be resolved by the time of our meeting.

Although three organizations have given us some indication that they wish to become sponsors of ODSC, none has yet sent a letter of intent. We will do the best we can with available resources. Our current sponsors remain AMC, EK, IBM and NASA Goddard.

We look forward to seeing you next month in Tucson.

Jim Burke
March 15, 1992

CONTENTS

SECTION 1. SUMMARY REPORTS

MAGNETO-OPTIC SUPERLATTICE THIN-FILMS: FABRICATION, STRUCTURAL AND MAGNETIC CHARACTERIZATION

C.M. Falco, B.N. Engel, R.A. Van Leeuwen and J. Yu

PROGRESS

This past quarter we have greatly enhanced our ability to study the structural influences on interfacial magnetic phenomena. We have constructed a computer based image capture system for quantitative analysis of RHEED and LEED patterns. This system allows us to digitize the electron diffraction patterns observed on the phosphor screens of our MBE with a high sensitivity CCD camera. The evolution of diffraction features can be followed easily during growth of epitaxial films and changes of intensity maxima positions caused by changes in lattice constant can now be measured with high precision.

We have used this system to search for coherent epitaxial strain in the Co/Pd system. From x-ray diffraction measurements reported earlier, we found the (111) oriented superlattices to be essentially unstrained (< 0.5%) while the (001) oriented films displayed a significant (8%) in-plane expansion of the Co layers. Preliminary measurements of the RHEED patterns of Co growth on oriented Pd buffer layers confirm these results.

In Fig. 1, intensity profiles of the RHEED patterns for 2Å and 6Å Co films deposited onto a Pd (111) buffer layer are shown. A change in diffraction maxima spacing between the two films

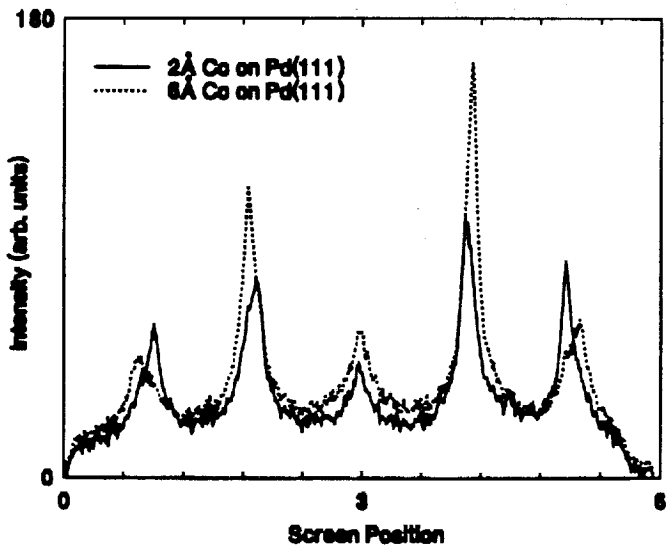


Figure 1. RHEED intensity line profiles of Co deposited on Pd (111).

is evident. From the larger distances between maxima for the 6Å Co film, it is clear that the Co is relaxing to its smaller bulk lattice spacing (larger distances in the reciprocal space of the diffraction pattern indicate a smaller lattice spacing in real space). From these measurements it appears that the 2Å of Co is strained to the bulk Pd lattice. However, we cannot conclude this from these data alone. The first atomic layer (~ 2Å) of Co is probably not continuous; hence, the RHEED beam will diffract from both Co islands as well as the underlying Pd substrate. Thus, the maxima will result from a convolution of diffraction from these surfaces. We are now working on separating these effects. We can conclude, however, that the Co relaxes to the bulk spacing by at least 6Å for crystals oriented along the (111) direction.

In contrast to the above case, Co deposited on Pd (001) displays coherently strained growth. Figure 2 is the RHEED intensity profile of a 2Å and 6Å Co film deposited on a Pd (001) surface. It is evident that there are no changes in lattice constant between these two films as indicated by the alignment of the diffraction maxima. This strained growth continues up to a thickness of at least 10 Å without signs of relaxation. This is consistent with our x-ray diffraction studies of our epitaxial (001) Co/Pd superlattices. We are now performing experiments to determine at what thickness the Co relieves the epitaxial strain.

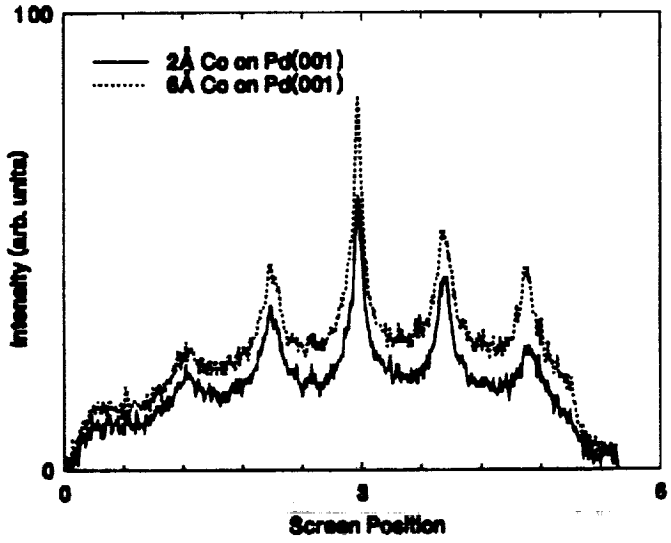


Figure 2. RHEED intensity line profiles of Co deposited on Pd (001).

The above measurements demonstrate the power of using an image capture system for analyzing our electron diffraction information. Using this capability with our *in situ* Kerr rotation

apparatus allows us to probe in detail the mechanisms responsible for perpendicular interface anisotropy.

PLANS

During the next quarter we will begin studying a new material system and will continue our effort on the Co/Pd interface. In our second MBE (Riber 1000) we now have the capability of depositing Bi. During the next quarter we will add an effusion cell with Mn to study MnBi alloys. MnBi alloys display an enhanced Kerr rotation of greater than one degree and, therefore, are of great interest for magneto-optic recording. These alloys are only ferromagnetic for a particular crystallographic phase. To date, workers on this system have concentrated on creating this phase by post-annealing polycrystalline Mn/Bi multilayer films under certain conditions. Our approach will be to use the MBE to "co-deposit" Mn and Bi onto single-crystal substrates to directly create epitaxial alloys with the proper crystallographic and stoichiometric phase for studying magneto-optic phenomena.

CHARACTERIZATION FACILITY FOR MAGNETO-OPTIC MEDIA AND SYSTEMS

M. Mansuripur, H. Fu, H. Sukeda, R. Hajjar, B. Bernacki, F.L. Zhou, T.H. Wu and J.K. Erwin

OBJECTIVES

- To measure the hysteresis loop, Kerr rotation angle, anisotropy energy profile, Hall voltage and magnetoresistance of thin-film magneto-optic media using our loop-tracer. This instrument has a maximum field capability of 21 kOe, and can measure the temperature-dependence of the media's magnetic properties in the range 77K to 475K.

- To measure the wavelength-dependence of the Kerr rotation angle, θ_k, and ellipticity, ε_k, for thin-film media using our magneto-optic Kerr spectrometer (MOKS). The spectrometer operates in the wavelength range 350 nm to 1050 nm.

- To measure the dielectric tensor of thin-film and multilayer samples using our variable-angle magneto-optic ellipsometer (VAMOE). This device measures (at the HeNe wavelength of 633 nm) the complex reflection coefficients of the samples at angles of incidence ranging from 26° to 88°. A computer program then estimates the dielectric tensor from the measured coefficients of reflection.

- To measure the hysteresis loop, coercivity, remanent magnetization, saturation magnetization, and anisotropy energy constant for thin-film magnetic media using vibrating sample magneto-metry. The temperature range of our VSM is 77K to 1000 K, and it has a maximum magnetic field capability of 12 kOe.

- To observe small magnetic domains, and to investigate their interaction with defects using magnetic force microscopy.

- To perform static read/write/erase experiments on thin-film magneto-optic media using our static test station. The test station is a polarized-light microscope, modified to include a laser diode, electromagnet, temperature-controlled stage, and a television camera, all under computer control.

- To perform dynamic read/write/erase experiments using a Nakamichi OMS-1000 test station.

- To integrate the existing models of magnetization, magneto-optic effects, coercivity, and anisotropy in an interactive and user-friendly environment. To analyze the characterization data obtained in the various experiments, using this modeling package.

- To measure focusing- and tracking-error signals on a static test-bed; to determine the "feedthrough" for various focusing schemes, investigate the effects of polarization, and compare the results with diffraction-based calculations. The test-bed has a HeCd laser (λ = 442 nm), a series of high-NA objectives (0.5-0.8), piezoelectric actuators (with better than 0.1 μm positioning accuracy) for driving the sample, CCD camera emulating multi-element detectors, several interchangeable focus-error modules, and in-built shearing and Twyman-Green interferometers for ascertaining beam quality and/or calibrating the focus position.

PROGRESS

Magnetic Force Microscopy

Our magnetic force microscope is now operational. We have been experimenting with various needle designs, trying to improve the resolution. We also are interested in finding ways to manipulate the recorded domains with the magnetic field of the needle. Figure 1(a) shows a picture of domains obtained with this MFM on a Co/Pt sample. Figure 1(b) shows the same area of the sample, scanned with the needle in contact; here, instead of magnetic domains, the physical structure of the surface is imaged (grooves are clearly visible).

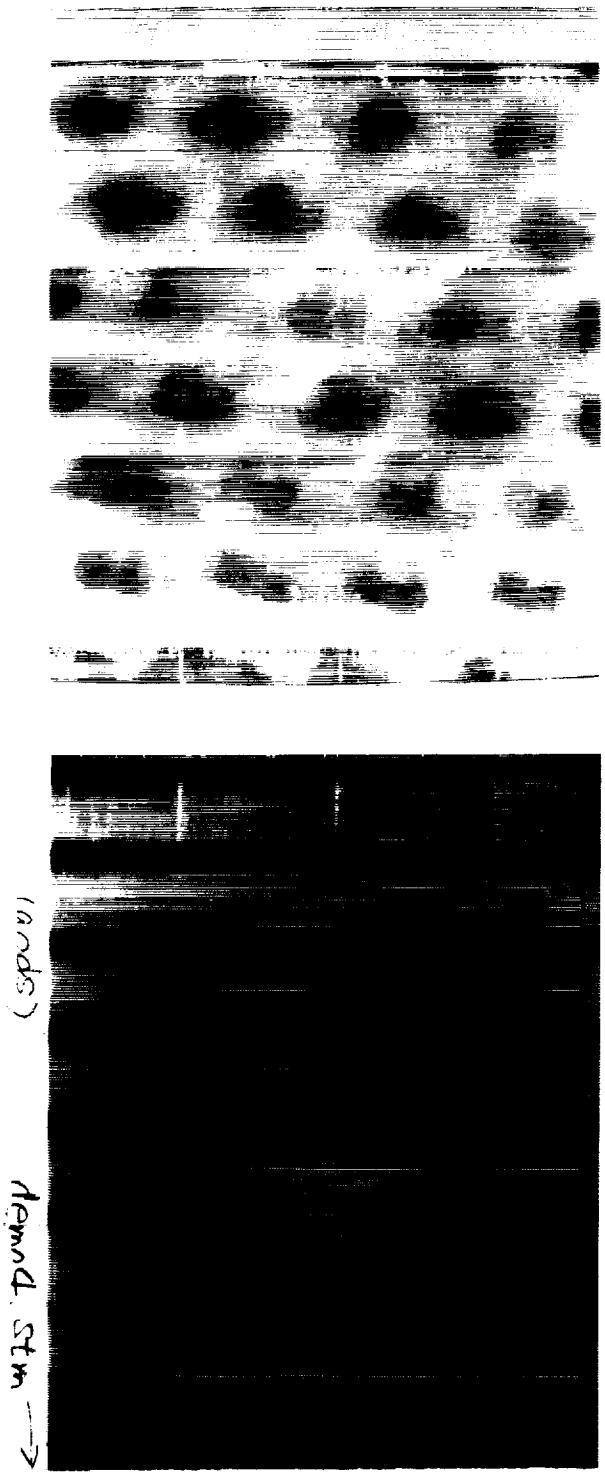


Figure 1. MFM images from a small region of a Co/Pt disk. The magnetic domains in (a) are about 1 μm in diameter. The image in (b), obtained with the needle in contact with the sample, shows the grooves.

Analysis of Focus//Track Servo Signals

We measured track-crossing signals on a static test-bed. The influence of track-crossing on the focus-error-signal was analyzed to understand "feedthrough." The data were obtained for various amounts of sample defocus, different focus-error-detection schemes, and different polarizations. Figure 2 shows a typical set of data obtained on a standard 1.6 μm track-pitch disk; the two curves in Fig. 2(a) correspond to linear polarization vector parallel and perpendicular to the grooves. Figure 2(b) shows the feedthrough signal for the astigmatic focus-error-detection scheme. This work is in progress and we shall report our latest results at the ODSC meeting in April.

With the experimental work, extensive computer modeling was carried out using DIFFRACT. The results were presented at the ODS meeting in San Jose, and are reproduced here as Appendix A. An interesting outcome of the work so far has been the discovery (by B. Bernacki) of a modified version of the astigmatic focus-error-detection scheme, which is fairly insensitive to cross-talk from the tracking channel. Appendix B is a document disclosing this invention.

Variable-Angle Magneto-Optic Ellipsometry

The ellipsometer has been used to determine the refractive indices and dielectric tensors of a large number of metal/dielectric/magnetic thin-film samples. The results were reported at the ODS meeting in San Jose, and are reproduced here as Appendix C.

Measurements of Anisotropy Constant K_u

Our latest measurement results on TbFeCo samples were presented at the ODS meeting in San Jose. Both in-plane and rotating fields were used to measure the temperature dependence of the anisotropy constant. Information concerning the deviation of magnetization from the easy axis was carried by the extraordinary Hall signal and the magneto-optic Kerr signal. Appendix D documents the results of this work.

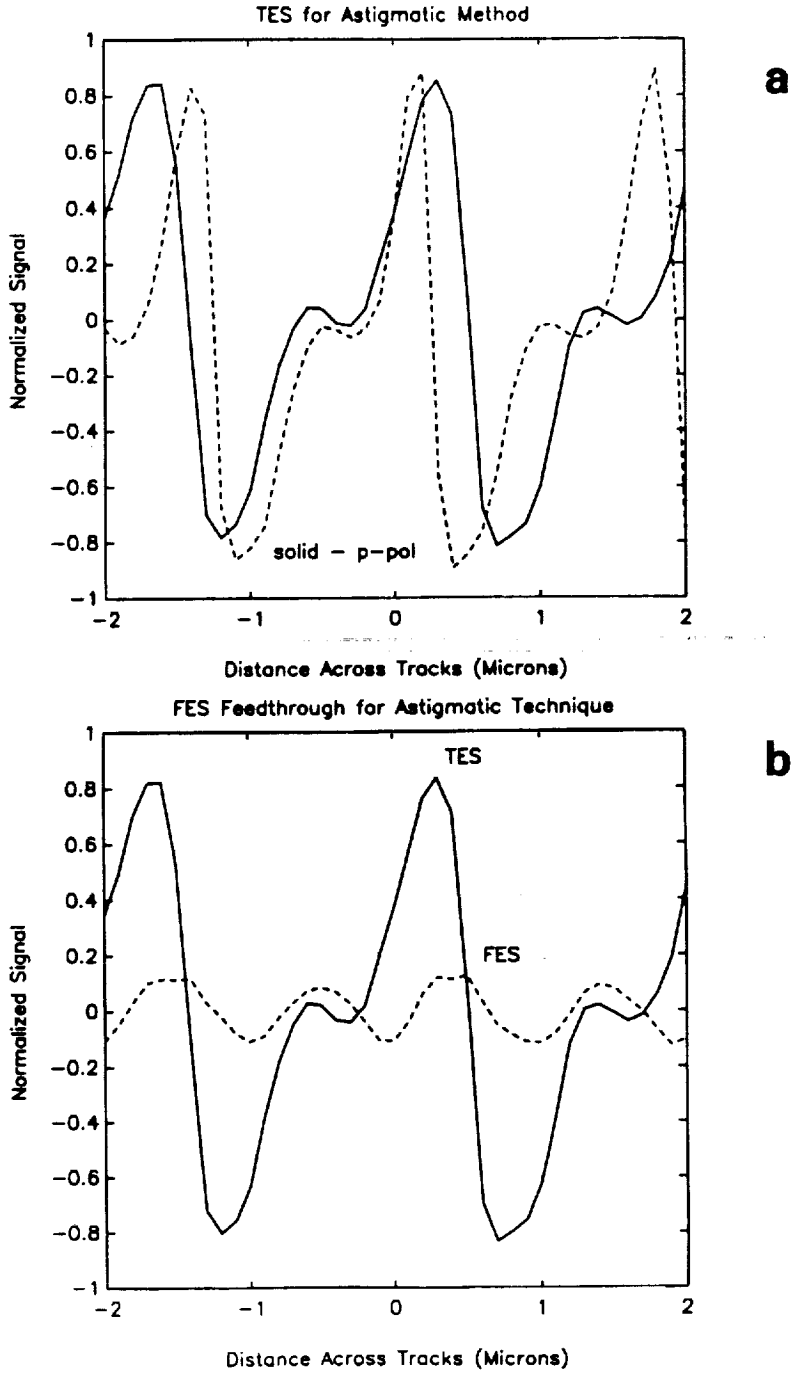


Figure 2. Measurement results on a section of a pre-grooved disk surface, showing the push-pull track-crossing signal, $(S_1-S_2)/(S_1+S_2)$. In (a) the solid curve is for polarization parallel to the grooves, while the dashed curve represents the perpendicular polarization. The relative shift of the curves along the horizontal axis is probably due to drift of the piezo-actuator between measurements. The FES curve in (b) shows the feedthrough signal. (λ = 442 nm, NA = 0.5, track-pitch = 1.6 μm, grooves V-shaped, astigmatic focus-error-detection scheme used.)

Other Activity

We explored the possibility of creating large data storage systems by integrating small drives in a parallel architecture. A computer simulation showed the tremendous potentials of this approach. Our preliminary results appear in Appendix E.

PLANS

In the months ahead we plan to use the magnetic force microscope to investigate the interaction of magnetic domains and domain walls with defects and structural features on the sample. These measurements are expected to provide guidance for our computer simulation effort aimed at understanding the origins of coercivity.

The static tester will provide information on focus/track error signals, their polarization sensitivity, and relative merits of the various focusing schemes.

We are looking into the initial magnetization curves of the magneto-optic films, using the loop tracer and the polarization microscope. This work is intended to provide insight into the nature of wall coercivity, and will be a major focus for our micromagnetic effort in the next period.

For three of our students this is the last semester at the University of Arizona . Roger Hajjar will graduate by the end of May; Bruce Bernacki and Fenglei Zhou are expected to complete their dissertations before the end of August. We will be looking for new students to replace them.

DYNAMIC TESTBED LABORATORY AND MICRO-OPTICS

T. Milster, C. Curtis, C. Campillo, K. Erwin, J. Kann and M. Wang

OBJECTIVE

Our objective is to advance understanding of optical systems in data storage devices for better data rate, capacity, and form factor.

PROGRESS

This has been a fruitful quarter in terms of our research results and papers/presentations. In this report, we concentrate on discussing noise in super-resolving systems and a new magneto-optic readout technique called magnetic circular dichroism (MCD) detection. Other activities include applying our ray-trace/diffraction code (Appendix F), transfer-function analysis of super-resolving systems (Appendix G), a computationally efficient method for calculating diffraction patterns from high-NA systems (Appendix H), modeling and measurement of the micro-optic beam deflector (Appendix I), and understanding the effects of residual higher-order aberrations in our optical data storage testbed (Appendix J).

Theory

Our super-resolution work continues with trying to understand how noise affects the system response. We start by making some simplifying assumptions about the form of the signal-to-noise ratio (SNR), as given by

$$\mathrm{SNR} = \frac{P_{sig}}{P_{elec} + c_{shot}\sqrt{\langle P_{sum}\rangle} + c_{las}\langle P_{sig}\rangle + c_{disk}\langle P_{sig}\rangle + c_{xy}\langle P_{sum}\rangle}, \quad (1)$$

where P_{elec} is the electronic noise power, P_{sum} is the sum electrical power, and P_{sig} is the signal power. Brackets denote average values over the frequency band. Multiplying constants c_{shot}, c_{las}, c_{disk}, and c_{xy} are factors that relate to shot noise, laser noise, disk noise, and depolarization noise, respectively. P_{elec}, c_{shot} and c_{las} are straightforward to calculate or measure experimentally. Disk noise, c_{disk}, and depolarization noise, c_{xy}, are not as easy to determine.

Disk noise is defined as the variation in reflected light amplitude due to defects in the disk reflectivity. Disk noise is illustrated in Fig. 1a, where the change in the length of the electric field vector is shown for an erased state of the track. Notice that the amplitude of the change in detector currents, δi_a and δi_b, are approximately equal, and they change in phase with respect to each other, that is, they both increase or they both decrease. The differential detection channel will reduce most of this noise contribution, since it is common mode. If the electric field vector is biased at exactly 45° and there is no Kerr effect, the disk noise is perfectly canceled. However, a small amount of disk noise will propagate through the channel because of the Kerr-signal imbalance in detector currents.

Depolarization noise is due to a fraction of the bias polarization (polarization direction emitted from the laser diode) that is rotated into the signal polarization. This could be due to diffraction off groove structures on the medium. Depolarization noise is illustrated in Fig. 1b, where the change in the angle of the electric field vector is shown for an erased state of the track. The amplitude of the changes in the detector currents, δi_a and δi_b, now change out of phase with respect to each other, that is, when $\delta i_a \uparrow \delta i_b \downarrow$ and visa versa. This type of noise is not canceled by the differential detection channel because it looks like the signal. Therefore, even a small amount of depolarization noise can be detrimental to the system.

The problem occurs when one tries to experimentally determine the relative contribution of disk noise and depolarization noise. They both appear as variations in the individual detector currents. One can block alternate detectors to estimate the amount of noise canceled by the differential channel, but the distinction between disk noise and depolarization noise can't be made without a more sophisticated experiment. The distinction is important because, as illustrated in Eq. (1), the disk noise is proportional to the signal current, $i_a - i_b$, while the depolarization noise is proportional to the sum current, $i_a + i_b$.

We propose an experiment to determine the relative contributions of disk noise and depolarization noise. We will measure the joint probability distribution function of the two noise sources. A positive correlation will be measured in the case of disk noise, and a negative correlation will be measured in the case of depolarization noise. Although the experiment sounds complicated, the implementation is very simple and straightforward. Outputs from the two data detectors are converted to digital signals via fast analog-to-digital converters (ADCs). The ADCs

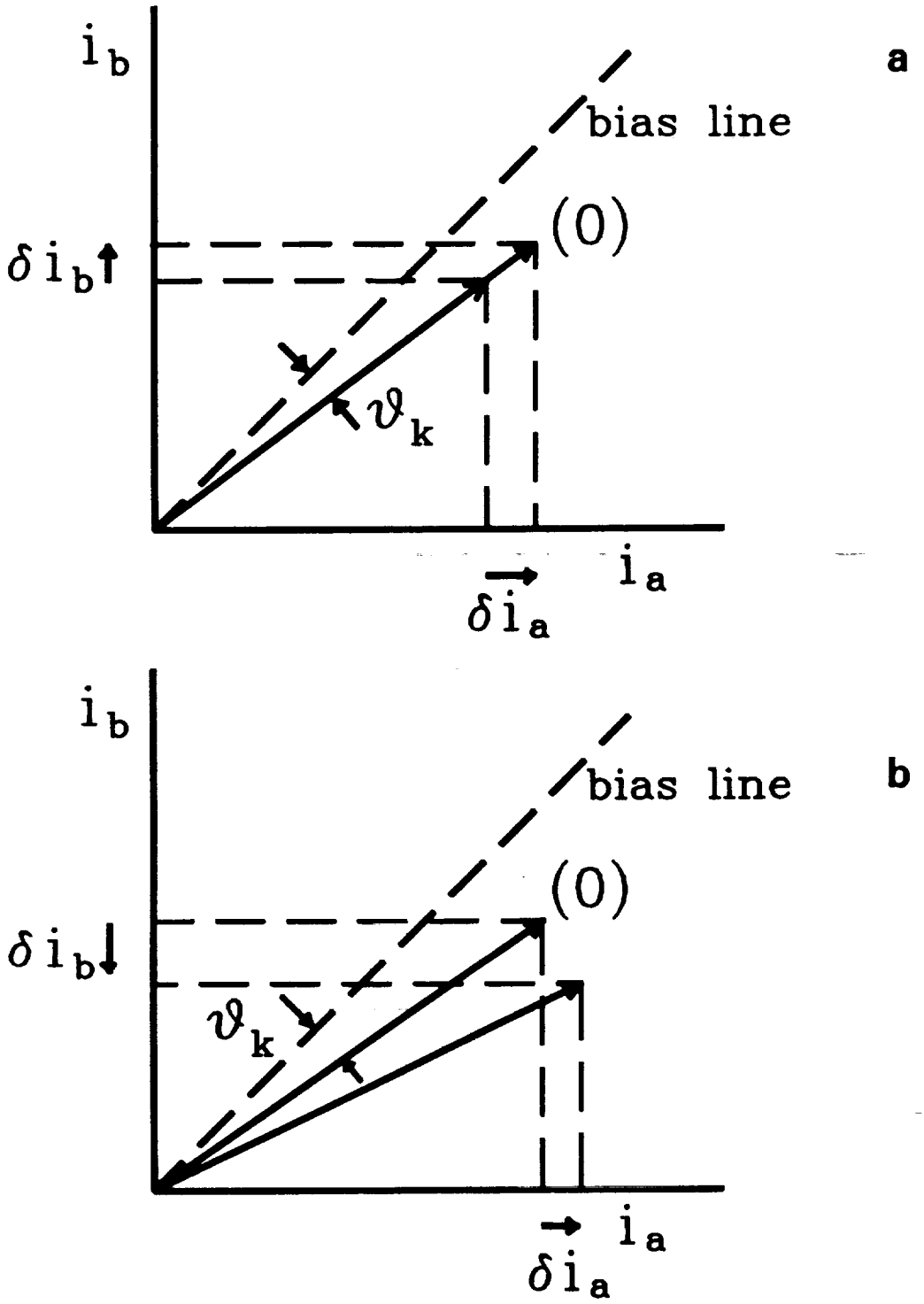


Figure 1. a) Disk noise is illustrated as a change in the length of the electric field vector for the erased (0) state of the medium. b) Depolarization noise is illustrated as a change in the Kerr angle of the electric field vector of the erased (0) state of the medium.

are triggered so that valid data are recorded at the same instant in time. Data points are accumulated in a histogram fashion by using the ADC outputs as addresses to a two-dimensional memory. Each time a particular memory location is addressed, its contents are incremented by one unit. Note that, if we assume that the signals are ergodic, there is no need to sample at extremely high frequency. Sufficient statistics can be compiled by using a moderate-speed ADC and associated electronics. The output from such a measurement might look as displayed in Fig. 2 for positively correlated, negatively correlated, and uncorrelated noise.

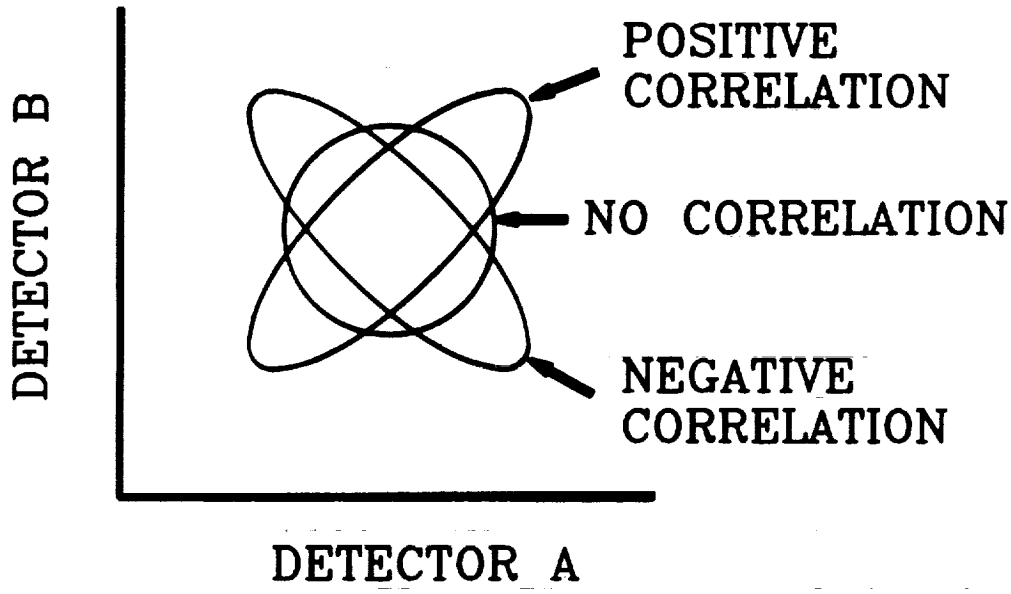


Figure 2. Correlation measurement output showing result for positive correlation, negative correlation, and uncorrelated noise.

Modeling

We have been pursuing several modeling activities this quarter to help us understand the optical system. Appendix F describes an application of our modeling code that uses a combination of ray-trace and diffraction analysis. Our hypothetical problem concerned the effects of tilting the objective lens in a data storage device. Appendix G describes transfer-function analysis of super-resolving systems. The full manuscript of our paper that was accepted by *Applied Optics* is available upon request. An interesting outcome of the modeling is shown in Fig.

3, where we display the two-point response for several optical systems. Curve A is simply the bit pattern on the medium. Curve B is the response from optical systems the way they are commonly configured. Curve C represents the response from the Sony self masking medium. Curve D represents the response from a simple shading band in the collection optics. Note that curves C and D are very similar, which suggests that super resolution may be obtained without the complicated self-masking medium. Appendix H describes a new technique for modeling extremely high numerical aperture systems with a greatly reduced computational overhead. The caveat is that we must observe around the axis in the image plane. (This approximation is okay for data storage.)

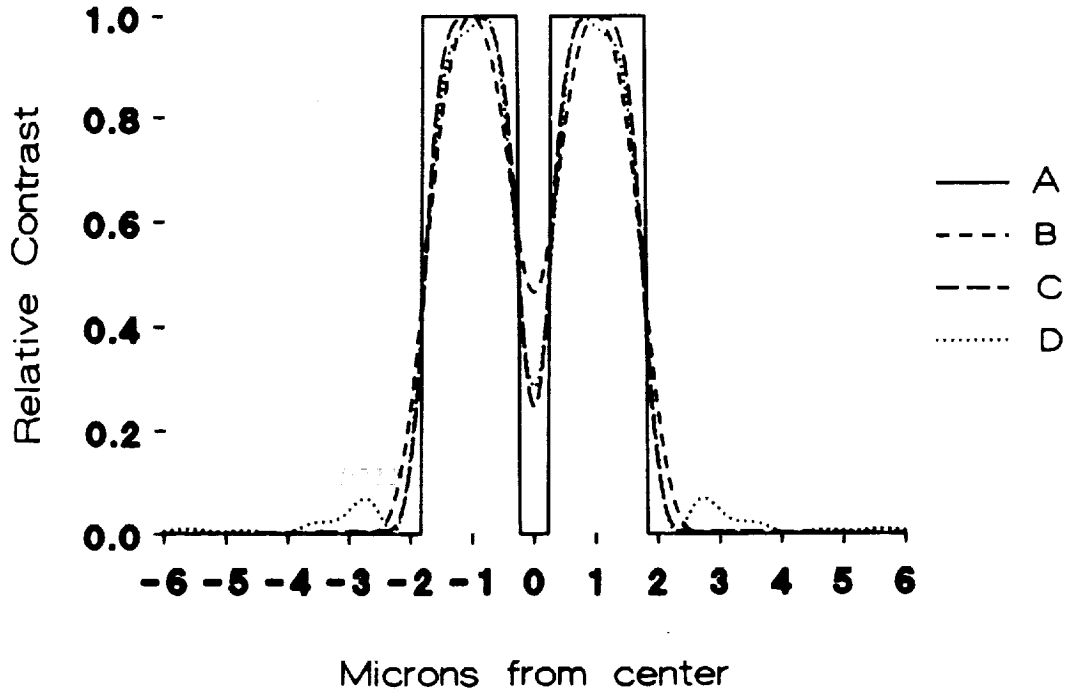


Figure 3. Two-point resolution of several systems: A) bit pattern on disk plane consists of two 1.56 m long marks separated by 0.52 m; B) reference system response; C) response from Sony self-masking medium; D) response from system with shading band in the collection optics.

Static Experiments

An interesting project that concerns magneto-optic readout has been investigated during the past six months. In our study, we investigated the detection of MCD. Major advantages of this technique include insensitivity to birefringence and depolarization noise. An additional benefit

is that the readout optics are extremely simple and that laser-diode feedback noise is suppressed. Figure 4 illustrates the optical system, which is exactly the same as those used in write-once devices. The advantages of simplicity and power throughput are obvious.

MCD detection works by sensing the change in ellipticity from the disk. A figure of merit (FOM) for MCD detection is given by

$$\mathrm{FOM} = r^2 k \sin(\kappa) \quad , \tag{2}$$

where r is the bulk reflectivity of the medium, k is the tangent of the Kerr angle, and κ is the s-p phase difference upon reflection from the medium. Note that the dependence of FOM on κ goes as $\sin(\kappa)$, where, in differential detection, the FOM goes as $\cos(\kappa)$. Therefore, to use MCD detection efficiently, the thin-film structure of the medium must be modified slightly. Given the potential benefits of MCD detection, we feel that this modification is justified.

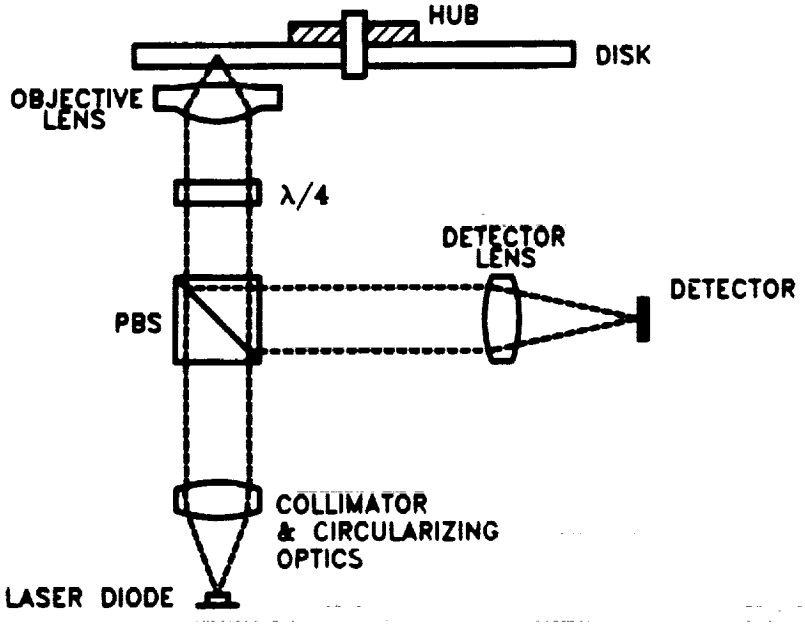


Figure 4. Optical system used with MCD detection.

C. Campillo has completed an initial study of this technique. He has constructed a static tester and has evaluated several samples. A detailed theoretical analysis of the technique also was performed. An example of the signal-to-noise ratio (SNR) performance of MCD detection

compared to differential detection is shown in Fig. 5. Note that, for increasing values of depolarization noise, MCD detection appears to have a significant advantage over differential detection. The complete analysis is documented in Campillo's Master's Report, which is available to sponsors upon request. (Incidentally, Campillo is seeking employment.)

A series of experiments have been conducted on the micro-optic beam deflector. Our progress to date is documented in Appendix I.

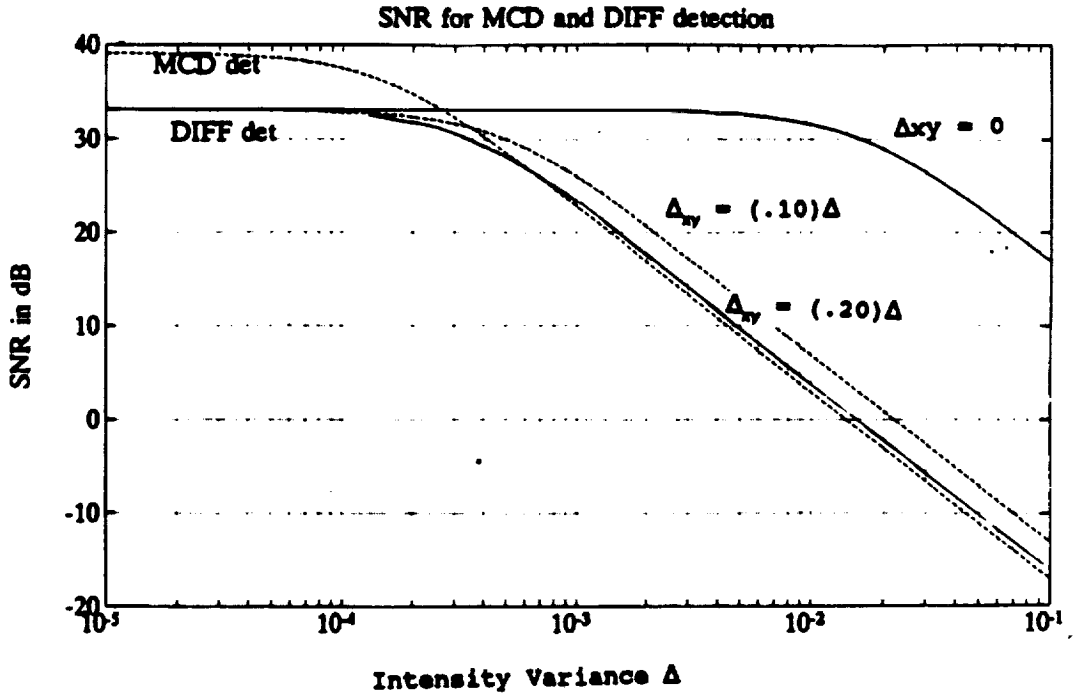


Figure 5. Plot of theoretical signal-to-noise ratios for MCD and differential detection as a function of the intensity variance, Δ. The depolarization noise component, Δ_{xy}, is shown at 0%, 10% and 20% of Δ.

Dynamic Testing

In the December 15, 1991, ODSC Quarterly Report, we described discrepancies between measured diffraction patterns and modeling results. Discrepancies took the form of energy redistribution on the focus-error detector. Through wavefront measurement and scalar diffraction modeling, we discovered that the energy redistribution is due to residual third-order and fifth-order spherical aberration of the objective lens and cover-plate assembly. The amount of residual

aberration is small, and the beam would be considered diffraction-limited by several criteria. Since the detector is not in a focal plane, even this small amount of aberration has a significant effect on the energy distribution. In Appendix J, M. Wong has completely analyzed this problem. We are preparing the manuscript for submission to *Optics Letters*.

MeV ION-BEAM ANALYSIS OF OPTICAL DATA STORAGE FILMS

J.A. Leavitt, L.C. McIntyre Jr. and Z. Lin

OBJECTIVES

Our objectives are threefold: 1) to accurately characterize optical data storage films by MeV ion-beam analysis (IBA) for ODSC collaborators; 2) to develop new and/or improved analysis techniques; and 3) to expand the capabilities of the IBA facility itself.

Using $^{1}H^{+}$, $^{4}He^{+}$ and $^{15}N^{++}$ ion beams in the 1.5 to 10 MeV energy range from a 5.5 MV Van de Graaff accelerator, we determine film thickness (in atoms/cm^2), stoichiometry, impurity concentration profiles and crystalline structure by Rutherford backscattering (RBS), high-energy backscattering, channeling, nuclear reaction analysis (NRA) and proton induced x-ray emission (PIXE). Most of these techniques are discussed in detail in the ODSC Annual Report (February 17, 1987), p. 74. The PIXE technique is briefly discussed in the ODSC Annual Report (March 15, 1991), p. 23.

PROGRESS

Film Characterization

From March 1, 1991, to March 1, 1992, we provided a total of 140 equivalent backscattering analyses of thin films provided by ODSC collaborators; Table 1 indicates the distribution.

Table 1. Numbers of analyses of thin films for ODSC

Period	Falco	Mansuripur	Zelinski	Armstrong	Totals
3/1/91-6/1/91	39	-	-	-	39
6/1/91-9/1/91	1	4	36	-	41
9/1/91-12/1/91	3	14	20	-	37
12/1/91-3/1/92	6	8	5	4	23
Totals	49	26	61	4	140

For Falco (with Engel and Van Leeuween), we performed extensive axial and planar channeling studies on two Pd/Ag/Co multilayers on crystalline GaAs, 17 RBS runs on these films as well as one NRA run to detect B.

For Mansuripur (with Wu and Shieh/IBM), we performed 26 RBS runs on SiN/TbFeCo/SiN/Al films to determine the stoichiometry and "thickness" of the layers.

For Zelinski (with Weisenbach), we performed 61 backscattering runs on SiO_2/TiO_2 waveguide films to determine stoichiometry, thickness and the percent of C present.

For Armstrong (with England), we performed four RBS runs to determine the stoichiometry and thickness of TbFeCo, TbFeCoTa and TbFeCoPt films.

Technique Development

We have continued our program of developing techniques for increasing sensitivity for quantifying light elements in/on heavy matrices. In this connection:

1) We report measured ^4He-^{10}B, ^4He-^{11}B backscattering cross sections for incident ^4He energies 1-3 MeV (see preprint: Appendix I, ODSC Quarterly Report, September 15, 1991).
2) We report development of a technique for quantifying ^{10}B and ^{11}B in thin films using (α,p) nuclear reactions for incident ^4He energies near 3 MeV (see preprint: Appendix H, ODSC Quarterly Report, September 15, 1991).
3) We describe development of a technique for quantifying C in near-surface materials using the strong non-Rutherford resonance at 4.26 MeV in the ^4He-^{12}C cross section (see the detailed discussion in ODSC Quarterly Report, September 15, 1991, p. 23).
4) We have attacked the problem of quantifying ^9Be in thin films using ion-beam analysis. We have made extensive measurements of both ^4He-^9Be and ^1H-^9Be backscattering cross sections. This work is discussed in ODSC Quarterly Report, December 15, 1991, p. 25.

We also have spent a considerable amount of time developing the PIXE technique for trace-element analysis during the past year. We obtained our first PIXE data on March 1, 1991, and have since concentrated on apparatus improvement and calibration to increase the accuracy and reliability of the results. Brief descriptions of this work are contained in ODSC Annual Report, March 15, 1991, p. 23; Quarterly Report, June 15, 1991, p. 23; and Quarterly Report, September 15, 1991, p. 26.

Facility Development

We have committed ourselves to installing a nuclear microprobe facility. The goal of this project is to produce an analysis beam (either ^{1}H or ^{4}He) of diameter about 1 μm. This microbeam is to be rastered over the sample to determine the lateral distribution of trace elements in features more than a few microns in size. The project is a collaborative effort involving Q. Fernando (University of Arizona, Department of Chemistry) and is partially supported by funds from an NIH Superfund grant. We have completed design of the overall system, and have ordered and received necessary components such as the object slits, the quadrupole doublet lens and the power supply. We are currently working on the design and construction of the beam-line and target chamber. We expect to be in rudimentary operation by June 1992. A figure showing the layout of the microprobe facility is contained in ODSC Quarterly Report, September 15, 1991, p. 27.

PLANS

During the next year, we plan to: 1) continue characterizing optical data storage films for ODSC collaborators; 2) acquire further experience with the PIXE trace element analysis technique; 3) continue measuring non-Rutherford cross sections for scattering of ^{4}He and ^{1}H on the light elements for analysis beams in the 1-5 MeV range; first priority is completion of the work on ^{9}Be; and 4) complete construction of the nuclear microprobe facility and learn how to use it.

PUBLICATIONS

1. J.A. Leavitt and L.C. McIntyre Jr., "Non-Rutherford ^{4}He Cross-Sections for Ion Beam Analysis," Nucl. Instr. Meth. **B56/57**, 734 (1991).

2. L.C. McIntyre, Jr., J.A. Leavitt, M.D. Ashbaugh, Z. Lin, and J.O. Stoner, Jr., "Cross Sections for 170.5° Backscattering of ^{4}He by the Isotopes of Boron for ^{4}He Energies between 1.0 and 3.3 MeV," Nucl. Inst. Meth., in press (1992).

3. L.C. McIntyre, Jr., J.A. Leavitt, M.D. Ashbaugh, Z. Lin, and J.O. Stoner Jr., "Determination of Boron using the B(α,ρ)c Nuclear Reaction at Incident Energies Near 3 MeV," Nucl. Instr. Meth., in press (1992).

CHARACTERIZATION FACILITY FOR MAGNETO-OPTIC MEDIA AND SYSTEMS

P.A. Lee, C.D. England and N.R. Armstrong

OBJECTIVES

The commercial availability of TbFeCo materials has not diminished the fact that they tend to oxidize under most atmospheric conditions, both wet and dry. Overcoating with both metallic and dielectric materials has been seen to slow down the process but not stop it. Fourth metal modifiers also have the same effect. The goal of this research is to study the reactivity of these materials to water and oxygen and determine what effect it is having on the composition of the near surface region that may lead to a more active or passive film.

In the last quarter we have studied two systems in which either Pt or Ta has been added as a fourth element modifier at concentrations of 1% to 5% which were supplied by 3M. Our approach to studying these material by XPS, AES and RBS has been documented elsewhere.[1] Some of this research was presented at the Optical Data Storage Symposium at the SPIE meeting in San Jose, California, Feb. 9-13. A preprint of the paper is attached as Appendix K.

CURRENT RESEARCH

Our current research focuses on the stability of FeTbCo materials that have been modified by adding Ta of Pt. We have come to expect metals like Ta, Ti, Zr and Nb to form protective oxides in the presence of oxygen. Oxides also can form by exposure to water but the reaction is somewhat slower because of the slow kinetics for the dissociation of H2O. Addition of these metals in low concentration to TbFeCo slow the oxidation and corrosion of the media in moist and dry environments.[2] Pt, on the other hand, is a noble metal and is not very reactive towards oxygen or water. The effect of adding Pt to the TbFeCo matrix has been expected to be primarily an electronic effect or surface work function change. At this time the role of Pt is not clear. It is important to note that oxidation and corrosion processes, while they are slowed by adding these types of modifiers, can and do still occur.

The importance of studying these materials is: 1) to assess the protective nature of the additive; 2) to determine whether suface modification of the film occurs and how it affects further oxidation; and 3) to try to predict how a modifier will affect the reaction pathway.

RESULTS AND CONCLUSIONS

Our results for Ta modified materials indicate that the Ta will migrate to the surface and oxidize to form a barrier oxide that will slow further oxidation of the media. Relative atomic ratios of Tb to Fe, calculated from XPS data, indicate that migration of all metals is still occurring but rediffusion or migration of Fe into the near surface region after oxidation of the Tb is diminished in both oxygen and water exposures. The Ta modified thin films also indicate that the Fe is more protected in the case of oxidation by O2. Initial results also indicate that lower amounts of Ta afford more protection. This last observation may be the result of oxide film disruption due to Ta migration to the surface that may open up defects where oxidants can enter the film to produce oxide at the metal/oxide interface.

Pt modified materials also show a protective effect. Even though Pt, as a metal, doesn't oxidize to any great extent in air, it may alter the oxidative pathway for oxygen and water. Pt exhibits a protective effect at concentrations of 5 to 10 atomic percent (at. %).[2,3] It is unclear now how the Pt affects the oxidation pathway. Detailed studies on Pt modified materials are now underway. Initial results indicate that at oxygen exposures below 500L Tb still appears to migrate to the near surface region without subsequent reaction of the Fe. Past this level the Fe appears to be rediffusing towards the surface. Pt appears to be increasing in concentration relative to Fe in the near surface. The nature of the interaction with oxygen is currently under study.

REFERENCES

1. P. Lee, K. Stork, B. Maschhoff and N. Armstrong, "Oxide formation on rare earth/transition metal thin films," *Materials for Magneto-Optic Storage*, T. Suzuki, C. Robinson and C. Falco, eds., vol. 150, 227-232, Materials Research Soc., New York, 1989.

2. F. Kirino, N. Ogihara and N. Ohta, "Corrosion resistance of TbFeCo-M magneto-optical Films," J. Electrochem. Soc., **138**(8), 2259-2262 (1991).

3. D. Majumbar and T.K. Hatwar, "Effects of Pt and Zr on the oxidation behavior of FeTbCo magneto-optic films: x-ray photoelectron spectroscopy," J. Vac. Sci. Technol. A **7**(4), 2673-2677 (1989).

APPLICATION OF HOLOGRAPHIC OPTICAL ELEMENTS TO MAGNETO-OPTIC READ/WRITE HEADS

R.K. Kostuk, E. Campbell and C. Haggans

OBJECTIVES

Our objectives are to determine the theoretical and practical performance limits of holographic optical elements (HOEs) formed in different recording materials, and to evaluate the application of these components to magneto-optic read/write heads.

PROGRESS

Rigorous Hologram Modeling and Verification

We completed a model based on rigorous coupled wave theory that allows both the average refractive index and the index modulation to vary with depth in the emulsion layer. This actually exists in many recording materials such as dichromated gelatin and photopolymers. It should significantly improve our ability to design holographic optical elements in these materials. We are currently using this model to investigate the theoretical limits of polarization-selective and non-selective holographic optical elements, and to experimentally verify the results. The findings from this analysis and set of experiments will be summarized in a research paper.

A different mathematical approach was taken in this model that reduces the size of the matrices used to describe the diffraction process. This also reduces compilation time. This approach also is being investigated as a means to improve the surface relief model's performance.

Grating Fabrication

We investigated the change in the average refractive index of holograms formed in dichromated gelatin. This was performed to understand some anomalous shifts in the Bragg angle that could not be explained by changes in emulsion thickness. This creates problems when trying to design polarization elements. Different methods were investigated for measuring the index including a Brewster angle and a Fabry-Perot comparison technique. These measurements show that the index at the emulsion surface and interface with glass are apparently higher than the index in the bulk of the film. Results by other researchers also have shown that the effective refractive index can be less than that of water due to the formation of air vacuoles in the emulsion

during processing. Understanding this phenomenon should significantly improve design capabilities with this material.

A special fixture for recording surface relief gratings in photoresist was designed and built with the help of L. Li. This mount has been used to fabricate a high spatial frequency photoresist mask pattern on a dielectric substrate. This mask then was coated with 0.2 μm of silver to form a metal grating. Illuminating this grating with a substrate mode reflected beam rotates the state of polarization of the TIR beam. This element is now being investigated for use in the magneto-optic sensing device described in the preprint of a paper presented at the Data Storage meeting in San Jose (Appendix L).

The first set of masks have been designed for fabricating lithographic binary phase gratings. We evaluated several mask writing formats that interface with standard mask writing foundaries. The masks have been used to form two-level gratings in photoresist and to produce on-axis focusing gratings that have efficiency and imaging performance near the predicted theoretical limits. We are currently working on making four-level gratings etched in glass using an ion beam etching machine in Li's lab. These elements should improve diffraction efficiency by a factor of 2.

Several beam splitting HOEs also have been fabricated for use in the tester built by Dr. Fujita. The design efficiency for the p-polarized beam in this element is 70% and 30% for the s-polarized beams; however, we have not yet formed a component with these specifications.

PLANS

During the next quarter we will concentrate on several problems including: 1) fabricating the components described in Appendix L on a substrate-mode magneto-optic head; 2) complete and summarize the magnitude and effects of index change in dichromated gelatin holograms; 3) fabricate the desired efficiency and phase properties of a holographic beam splitter for use in the tester; and 4) verify the index profile and use it to improve HOE designs.

SOL-GEL WAVEGUIDES AND GRATINGS: FABRICATION, CHARACTERIZATION AND MODELING

B.J.J. Zelinski, R. Roncone, L. Wiesenbach, J. O'Kelly, J. Morreale and J.J. Burke

OBJECTIVES

The overall goal of this research is to develop wet chemical techniques to synthesize low-loss dielectric planar waveguides and planar waveguides with surface relief structures. To achieve this goal three sub-objectives are being pursued: 1) development of coating solution chemistries and processing techniques that will reliably produce quality planar waveguides; 2) investigation of the theoretical and experimental aspects of loss in these waveguides; and 3) development of suitable solutions and techniques for embossing surface features into planar waveguides.

CURRENT STATUS

As a result of the reduction in industrial support of ODSC, the current activity for this program is almost entirely supported by the University of Arizona's Department of Materials Science and Engineering. Additional support is provided in the form of R. Roncone's fellowship administered by the Optical Science Center. Two proposals have been submitted to NSF that would provide support for this work. In addition, individual companies are being solicited for support.

Waveguide Synthesis

Research efforts this quarter have focused on the development of good quality optical films having a wide range of refractive indices. Within the SiO_2-TiO_2 system the range of investigated film compositions has been broadened and now includes 13, 35, 45, 50, and 90 mole% TiO_2 films. The refractive indices of these films range from 1.49 to 2.2. In addition, we have fabricated thin films of aluminosilicate alkaline earth-based glasses using wet chemical techniques. These films have low indices and so require greater thicknesses to confine the propagating mode. To achieve these thicknesses we have improved and standardized our processing procedures for fabricating multilayer sol-gel structures. For example, we have recently fabricated a crack-free, 15-layer film that is 1.8 μm thick and has an index of 1.464. This film was

made using the synthesis techniques developed for the aluminosilicate alkaline earth system. Initial coupling results with these new waveguides have been interesting and encouraging. Current efforts are aimed at documenting and understanding the structural development and loss characteristics of these new films as a function of processing using the techniques of FTIR, UV-VIS, TEM and SAXS.

Sol-Gel Derived Single Leakage Channel Grating Couplers

The investigation of additional compositions in the SiO_2-TiO_2 binary, and the exploration of films in the aluminosilicate alkaline earth system has been motivated by the need to develop high quality films for use in optical component fabrication, including fabrication of single leakage channel grating couplers (SLCGC). Details of the design and operation of this device are reported in Appendix M.

During the past quarter we have successfully fabricated SLCGCs using wet chemical techniques. For example, a SLCGC consisting of a five-layer reflective stack has been fabricated using the 90% TiO_2 (SiO_2) solution as the high index layer, the 13% TiO_2 solution for the low-index and buffer layers and the 35% TiO_2 solution for the waveguide layer. The index, thickness and loss characteristics of each of the layers used in the SLCGC are listed in Table 1.

Table 1. Measured layer parameters for sol-gel SLCGC

Layer	Index	Thickness (A)	Loss (dB/cm)
High Index	2.1	613	<12
Low Index	1.49	869	1-2
Buffer	1.49	6500	1-2
Waveguide	1.65	2300	<1

This SLCGC exhibited a branching ratio of 78% at 0.5145 μm, which is very close to the predicted value of 83%. We also have fabricated another SLCGC that possesses a branching ratio of 80.3% at 0.488 μm. Both of these devices used soda-lime slides as the substrate. We also have successfully fabricated an SLCGC based on a silicon substrate.

We are now making more efficient SLCGCs by increasing the number of layers in the reflective stack. This required us to focus on improving our ability to make thicker crack-free multilayer structures.

Waveguide Scatter

The next three months will see the culmination of the waveguide scatter modeling, and model validation. The computer models predict surface and volume-induced waveguide attenuation when given film index and thickness, wavelength, substrate index, rms roughness and correlation length of the film and substrate surfaces, and the (film volume) rms refractive index inhomogeneity and its correlation length. Work now is focused on experimental validation of these models.

Surface Scatter Model. By intentionally and "controllably" roughening both sides of polished pyrex substrates through either acid or base-etching, we aim to generate a variety of measurable roughnesses and correlation lengths upon which sol-gel films will be deposited. The optical characteristics (film index, thickness, and propagation loss) will be measured using guided-wave techniques. The surface statistics of the film and substrate will be measured at WYKO on an atomic force microscope. The experimental parameters then will be entered into the computer model and the predicted and actual propagation losses will be compared. Results of this work will be presented at the spring IAB meeting.

Volume Scatter Model. We are investigating two techniques, small angle x-ray scattering, and transmission electron microscopy, with an image processing technique, to experimentally detect and measure the (volume) rms refractive index inhomogeneity and its correlation length. These two parameters are critical to validation of the computer model.

Embossing

Recent efforts on embossing have focused on the development of an in-house capability for fabricating first-generation nickel replicas of masters using electroplating. Each master would then generate several replicas and each replica would be used to emboss gratings into waveguides. The ability to generate identical replicas will greatly enhance our efforts to establish suitable embossing conditions and solutions.

GRATINGS AND WAVEGUIDES

L. Li, K.A. Bates, N. Ramanujam, R.L. Roncone, and J.J. Burke

OBJECTIVE

Our immediate objective is to understand the limitations of guided-wave and grating coupler devices in their application to optical data storage. Our long-range goal is to develop and validate design codes for integrated optic devices.

PROGRESS

Our accomplishments in the past quarter have been documented in the three preprints appended to this report (Appendices M, N, O).

Single Leakage Channel Grating Couplers (SLCGC)

The work on the SLCGC was briefly reported at the last ODSC IAB in October and at the Optical Society of America annual meeting in November. Since then, further progress was made. Using sol-gel materials made by Weisenbach and Zelinski as buffer layers and the waveguiding layers on top of evaporated high-reflectivity dielectric stacks (ordered from an outside vendor), we achieved a branching ratio of 97.1%. The theoretical value is 96%. For details, see Appendix M.

Without any doubt; the design concept has been successfully demonstrated experimentally. To reduce the design concept to practical applications, the design principles and numerical models need to be carefully validated. We are currently collaborating with the integrated optics group of Kodak (John Brazas) on this project.

Modal Method of Lamellar Gratings in Conical Mountings

Li completed the writing of a theoretical paper (Appendix N) on the modal method for solving diffraction problems of rectangular-grooved gratings when the grating vector is not contained in the plane of incidence. Such a theoretical tool and its numerical implementation are indispensable in the study of diffractive optical elements where polarization conversion takes place. For example, see Appendix D of ODSC Report, December 15, 1991.

Grating Theory, Convergence Study

At last year's OSA meeting, Li (in collabortion with C. W. Haggans) briefly reported some numerical evidence that the popular coupled-wave method does not converge very well for metallic gratings in TM polarization. This report met with disbelief by the original authors of the coupled-wave method and some of their followers. After the meeting, Haggans and Li did further study and confirmed their earlier findings. Appendix O is a summary paper to be given at *the First Topical Meeting on Diffractive Optics* this April in New Orleans.

PLAN

Since the ODSC funding to our integrated optics group has stopped, our future work will shift toward the short wavelength laser project.

SECTION 2. APPENDICES

APPENDIX A

Diffraction analysis and evaluation of several focus- and track-error detection schemes for magneto-optical disk systems

Bruce E. Bernacki and M. Mansuripur

Optical Sciences Center, University of Arizona
Tucson, Arizona 85721

ABSTRACT

A commonly used tracking method on pre-grooved magneto-optical (MO) media is the push-pull technique, and the astigmatic method is a popular focus-error detection approach. These two methods are analyzed using DIFFRACT©, a general-purpose scalar diffraction modeling program, to observe the effects on the error signals due to focusing lens misalignment, Seidel aberrations, and optical crosstalk (feedthrough) between the focusing and tracking servos. Using the results of the astigmatic/push-pull system as a basis for comparison, a novel focus/track-error detection technique that utilizes a ring toric lens is evaluated as well as the obscuration method (focus error detection only).

1. INTRODUCTION

To achieve the highest recording densities promised by MO storage technology, it is essential that the axial position of the focusing lens that forms the optical stylus for thermo-magneto-optic recording and readout be held within tight tolerances. These tolerances result from the compromise between the high numerical aperture (NA) required to produce the smallest diffraction-limited spot size for high areal recording densities, and the low NA desired for a large depth of focus. Adding to this challenge is the requirement to move the optical head across the rotating disk to seek the desired track, and then, to maintain the head position on the correct track.

Although the tasks of focusing and tracking seem daunting, there is a plethora of focus error detection and tracking schemes from which to choose. In this paper, the astigmatic focus error detection method is studied in depth, along with the push-pull tracking method used on grooved MO media. The study is further broadened to include a novel focus error detection method using a ring toric lens, which may be used in conjunction with the push-pull tracking method. Additionally, the obscuration method for focus error detection is studied and comparison made with the other techniques.

2. METHODOLOGY/TOOLS

Each system is first modeled in its ideal form. The system is then perturbed to examine the effects of third-order aberrations, focusing lens misalignment and feedthrough. For this study, we define feedthrough as the unwanted focus error signal (FES) that is detected when the optical head is crossing the tracks, with the head held in nominal focus. The astigmatic/push-pull system is examined first since it is more familiar.

The computer tools used in this study have been described in detail elsewhere[1]. DIFFRACT© is a diffraction modeling program written in FORTRAN that was created as a general-purpose tool for optical data storage research. The user has a palette from which many types of lenses, polarization optics, media, and detectors encountered in optical data storage may be selected. DIFFRACT© permits the user to choose a uniform or truncated Gaussian amplitude distribution produced by a

collimated laser source, and then propagates this beam through the optical system by plane wave decomposition and Fourier methods. The polarization state and the intensity distribution of the propagating beam, as well as various detector outputs are available anywhere along the optical path during the simulation. The right-handed Cartesian coordinate system is used throughout this study, with the propagation of light taken to be in the +z direction. Positive angles are measured counter-clockwise with respect to the positive x-axis.

3. SYSTEM LAYOUT

In common with all the focusing and tracking techniques studied, the initial amplitude distribution is a collimated Gaussian laser beam with λ = 800 nm, 1/e radius of 3.2 mm and zero curvature. This beam is truncated by a 0.55 NA objective lens with focal length 3.76 mm: the truncated beam radius is 2.07 mm. A pre-grooved disk with rectangular cross section lies in the focal plane of the objective lens with groove geometry as follows: 1.6 μm pitch, 80 nm groove depth, 0.6 μm groove width, and 1.0 μm land width. The focused spot is reflected from the disk surface, and, 20 mm from the objective, the focus/track error detection components are positioned. For simulations of the non-ideal system, aberrations were introduced in the objective lens. Aberrations with angular pupil dependence were oriented to produce worst-case effects: coma was oriented 90° to the track direction, and astigmatism was oriented such that the sagittal and tangential foci were 45° to the track direction. Misalignments were treated by simulating the movement of the objective lens perpendicular to the groove direction.

4. SYSTEM SIMULATION RESULTS

4.1 Astigmatic focusing/push-pull tracking method

The theoretical background is well documented in the literature[2,3], hence, the technique, shown in Fig. 1, will only be described schematically. Light reflected from the disk is collected by an astigmat, i.e., a lens with two orthogonal line foci. A quadrant detector is placed between the two foci. For these simulations, f_1=19.75 mm and f_2=20.25 mm. The quadrant detector is treated as an ideal device with no dead zone between the quadrants. The orientation and numbering of the quadrants are as shown in Fig. 1.

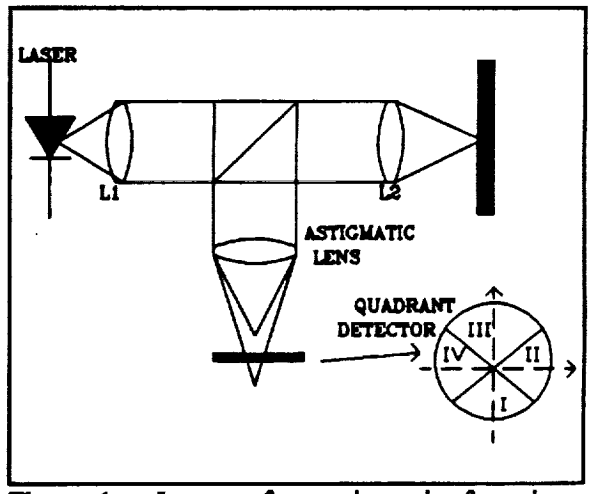


Figure 1. Layout for astigmatic focusing-push-pull tracking method showing geometry of quadrant detector.

The operation of the servo is readily understood using geometrical optics.[4] When the disk lies in the focal plane of the objective lens, the circle of least confusion coincides with the position of the quadrant detector, and light falls symmetrically on all four quadrants. However, when the light returning from the disk surface begins to diverge or converge due to disk movement in the ±z-direction, the location of the circle of least confusion changes, and the intensity pattern on the quadrant detector becomes elliptical, with the major axis of the ellipse corresponding to the nearer line focus of the astigmat. The error signal is obtained from the following combination of detector signals:

$$FES_{ast} = \frac{(I+III)-(II+IV)}{I+II+III+IV} \qquad (1)$$

Its ultimate sensitivity depends inversely[1] on the separation of the astigmatic foci. The result is a

bipolar error signal as shown in the solid curve of Fig. 2. Here, we define defocus as the amount of disk displacement from the focal plane of the objective lens along the ±z-direction. The focal plane coincides with the z=0 position. The gain G is the figure of merit for focus error detection systems and is defined here as the slope of the FES curve.

To understand the effect of Seidel aberrations on the FES, consider the nature of aberration present: Is it even or odd? Odd aberrations, e.g. coma, mimic the off-track condition and do little to degrade the FES since they produce a common mode signal on the quadrant detector. Even aberrations, such as spherical aberration, behave like defocus, causing the FES to be offset, as can be seen in Fig. 2. In practice, the offset caused by the even aberrations is nulled electronically or optically as part of a calibration procedure. The figure of merit, G, for this system is 0.16.

The operation of push-pull tracking can be understood by considering the disk surface as a reflective grating and observing its far-field diffraction pattern.[3] When the focused beam is exactly on track, diffraction from the edges of the land is symmetric, causing the far-field diffraction pattern to be symmetric. However, as the spot moves across the track, light from the 0 and ±1 orders is diffracted unequally from groove/land edges and destroys the symmetry of the far-field pattern, skewing the intensity distribution perpendicular to the groove direction. The orientation of the astigmat also causes the intensity distribution to be rotated 90°. The result is a bipolar track error signal (TES) as seen in Fig. 3 that is obtained from the quadrant detector using the relationship

$$TES_{push-pull} = \frac{(I+IV)-(II+III)}{I+II+III+IV}. \qquad (2)$$

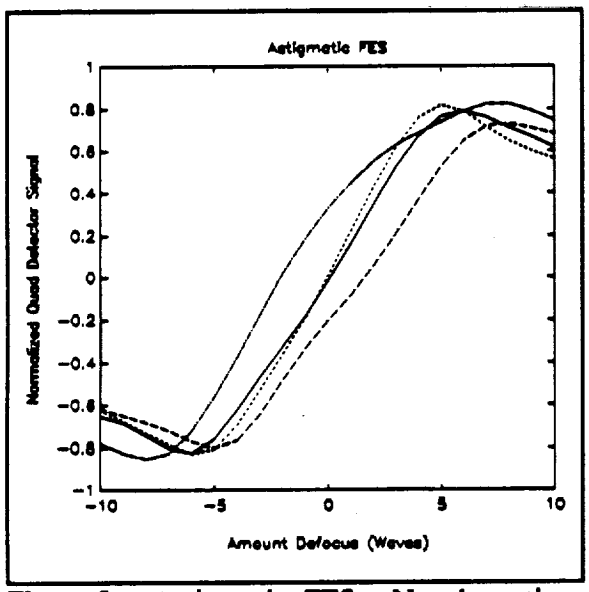


Figure 2. Astigmatic FES. No aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

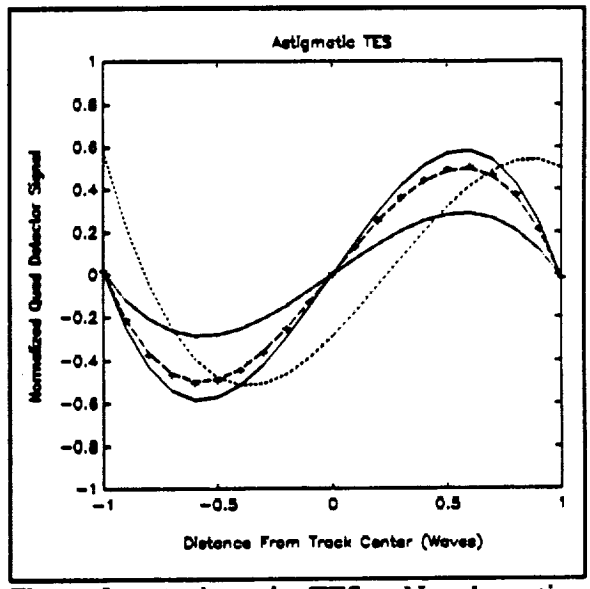


Figure 3. Astigmatic TES. No aberration (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), +0.25 λ coma, (dot), and +1 λ defocus (+).

For the push-pull TES, defocus reduces the slope of the signal, seen above in Fig. 3. Since the grating structure produces interference between the 0 and ±1 orders, it is a grating-type lateral shearing interferometer[4]. Therefore, an even aberration like defocus will produce a system of straight line fringes where the 0 and ±1 orders overlap, which reduces the contrast at the detector, and diminishes the slope of the TES. For the odd aberration studied, coma, the comatic tail was oriented in its worst-case orientation: perpendicular to the track direction. Its effect is to produce an offset in the TES, while maintaining the slope. As was the case with the FES, the reduced slope of the TES could be compensated by increasing the gain of the servo or adding some defocus to minimize

spherical aberration, while the offset of the zero crossing due to coma can be nulled by a corresponding electronic offset in the servo circuitry.

The fine adjustment of lens positioning used in tracking (usually accomplished using a voice coil arrangement) causes the lens to be displaced from the optical axis of the input beam by as much as ±50 μm. The effect of this operation on the TES was investigated by displacing the lens 0, 10, and 24 λ from the optical axis of the system along a line perpendicular to the groove direction. Then, the disk was scanned ±1 λ about the center of the original system, and the TES calculated. Moving an even number of waves puts the beam on a land center for this disk geometry. The results of the simulations, seen in Fig. 4, show that fairly good track error signals result for these amounts of lens decentering. No effect on the FES could be detected using these decentering values.

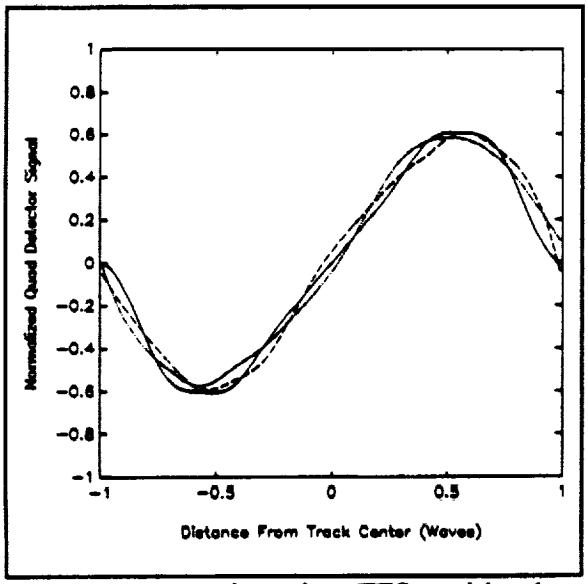


Figure 4. Astigmatic TES with lens decentering of 0 (solid), 10 λ (dash), and 24 λ (dash-dot) perpendicular to track direction.

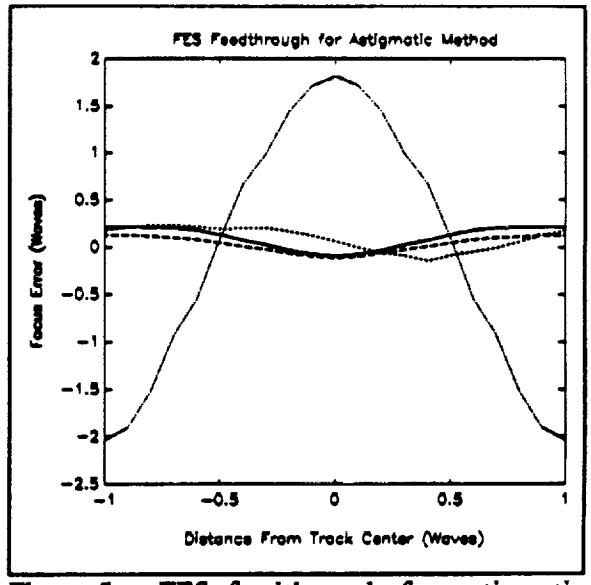


Figure 5. FES feedthrough for astigmatic method. No aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

Feedthrough is the unwanted FES that is detected when the optical head is seeking a new track. Ideally, the TES and FES should not be coupled. The reality of the situation is shown in Fig. 5. The focus error was calculated by determining the slope of the best fit line to the linear portion of the solid FES curve in Fig. 2, defined as G, and then its reciprocal was used to determine the error in the focus servo that results due to feedthrough. For feedthrough measurements, best focus was found to maximize the Strehl ratio for spherical aberration. In this case, the Strehl ratio improved from 0.81 uncompensated, to 0.98 with defocus added. The maximum 0.25 λ peak-to-peak error in the case of no aberration, coma, and spherical aberration could be tolerated by a well-designed control system, and any DC offsets in the signal can be nulled electronically. However, the 3.5 λ peak-to-peak focus error due to +0.25 λ of astigmatism oriented such that the astigmatic foci are 45° to the track would be intolerable, and therefore must be avoided in practice.

4.2 Ring toric lens/push-pull method

A focus- and track-error detection scheme has been described[5] that employs a ring toric lens and a four cell Φ detector, shown in Fig. 6. The Φ detector takes its name from the shape of the Greek letter suggested by the segmentation of the detector. The ring toric lens is the optical equivalent of a lens-axicon[6,7] combination. Collimated light is focused to a diffraction-limited ring

in the focal plane of the ring toric lens.

When the objective lens is in focus, the ring focus is centered on the dashed circle of the Φ detector. Diverging light returning from the disk surface shifts the ring focus outside the dashed circle shown in Fig. 6, while converging rays shift the ring focus inside the dashed circular area of the Φ detector. Unlike the astigmatic FES method, the ring toric approach is diffraction-limited, and its FES would approach a step function in the geometrical optics limit. Its slope is finite due to diffraction, however. The FES for the ring toric method, shown in Fig. 7, has G=0.42 and is obtained from the four-cell Φ detector by the following relation

$$FES_n = \frac{(I+II)-(III+IV)}{I+II+III+IV}. \tag{3}$$

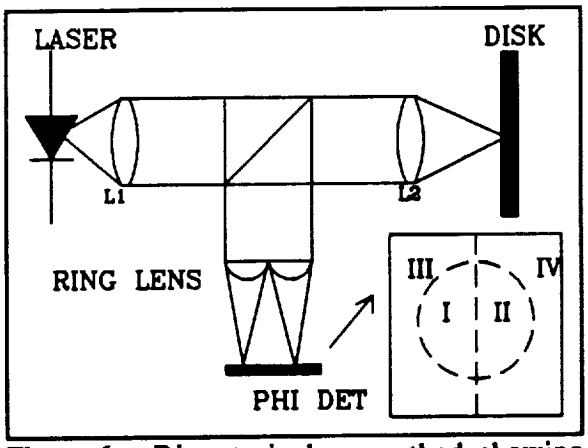


Figure 6. Ring toric lens method showing four-cell Φ detector. Focal length 25000 λ, inner ring radius 50 λ, outer ring radius 2750 λ, radius of ring in Φ detector 100 λ.

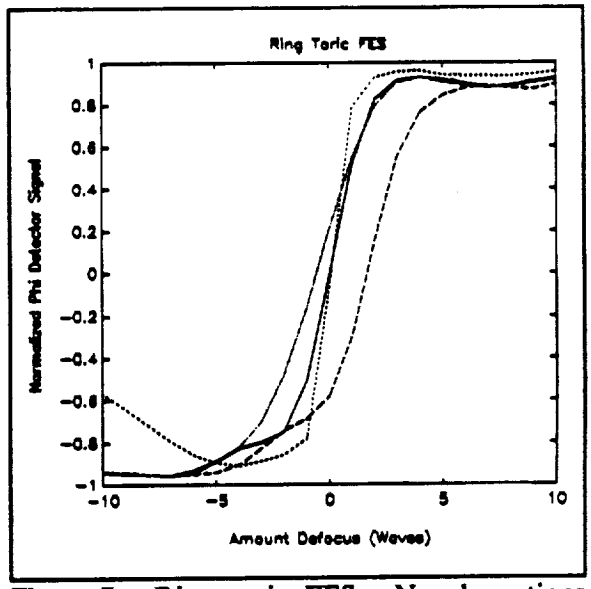


Figure 7. Ring toric FES. No aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

As was the case with the astigmatic method, the even aberrations have the greatest effect on the FES, since their behavior is similar to defocus, and offset the zero crossing. Also, since the gain G of the ring toric method is diffraction-limited, the presence of small amounts of even aberrations reduce its sensitivity as well as causing an offset.

Push-pull tracking is accomplished with the ring toric lens by summing the following regions of the Φ detector

$$TES_n = \frac{(I+III)-(II+IV)}{I+II+III+IV}. \tag{4}$$

The performance of the TES produced by the ring toric, including the effects of focusing lens displacement, is indistinguishable from that developed by the astigmatic lens approach.

With the exception of astigmatism, to which the astigmatic method is understandably more sensitive, the focus feedthrough signals are nearly equivalent in the two methods. However, since the focus error produced is proportional to 1/G, the greater gain of the ring toric method reduces the magnitude of the focus error due to feedthrough, as can be seen in Fig. 8. Best focus was also found to maximize the Strehl ratio, as was done for the astigmatic method.

4.3 Obscuration method

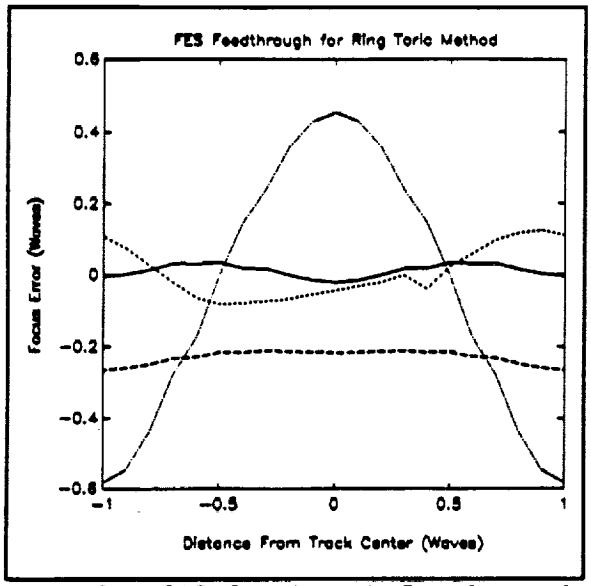


Figure 8. FES feedthrough for ring toric method. No aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), +0.25 λ coma (dot).

In this technique (a variation of the Foucault knife edge method) an obscuration is placed against the secondary lens which produces the focus error signal.[2,3] A split detector placed in the focal plane of the secondary lens produces a differential signal that is highly sensitive to movement of the centroid of the focused spot. Fig. 9 shows the schematic layout for this technique.

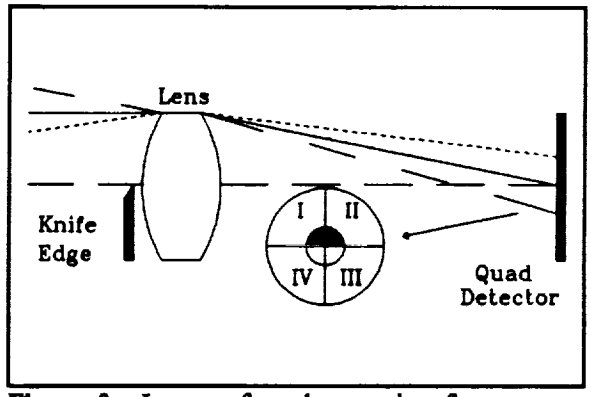


Figure 9. Layout for obscuration focus error detection method.

An out-of-focus condition causes the spot on the detector to move perpendicular to the orientation of the obscuration edge. In the limit of geometrical optics, the FES response approaches a step function, as was the case for the ring toric lens. The slope of the computed FES curve shown in Fig. 10 is due to diffraction. A drawback of this method is the loss of 50% of the light reflected from the disk due to the obscuration in the return path.

Since the obscuration method is diffraction-limited, its performance is similar to the ring toric method with regard to the Seidel aberrations, as can be seen in Fig. 10. Because of its high value of G = 0.32, it appears to be nearly as resistant to feedthrough as the ring toric lens method, as is seen in Fig. 11. Note, however, the greater peak-to-peak feedthrough signals in the cases of no aberration, coma, and spherical aberration. Push pull tracking may be done with this method if two Fresnel bi-prisms are employed to split the exit pupil of the secondary lens, or a double wedge-roof prism arrangement[8] is employed, but these variations were not investigated.

5. SUMMARY

Three focus/track error detection systems were examined in detail in this study: astigmatic/push-pull, ring toric/push-pull, and obscuration method for focus error detection only.

The astigmatic focus servo has a geometrical optics performance limit in the FES, which is degraded by the even Seidel aberrations, but insensitive to odd aberrations. The effect of the even aberrations is to offset the zero crossing of the FES response curve while maintaining nearly the same gain of G=0.16. With the exception of astigmatism, peak-to-peak error in the focus servo is less than 0.25 λ. However, +0.25 λ of astigmatism oriented such that the sagittal and tangential foci are at 45° to the track direction induces a focus error of nearly 4 λ peak-to-peak.

Push-pull tracking is degraded by both even and odd aberrations, including defocus. The even aberrations reduce the slope of the TES, but do not shift the zero crossing. The odd aberrations shift the location of the zero crossing, but the slope is preserved. Displacement of the objective lens from the optical axis, which occurs during tracking, was shown to have a small effect on the TES for the aberration-free, in-focus case.

The ring toric focus error detection method has diffraction-limited performance, which produces a steep FES curve and high value of G=0.42. It is much less prone to feedthrough than the

astigmatic/push-pull method when astigmatism with worst-case orientation is present. The performance of push-pull tracking with the ring toric lens is identical with that of the astigmatic approach.

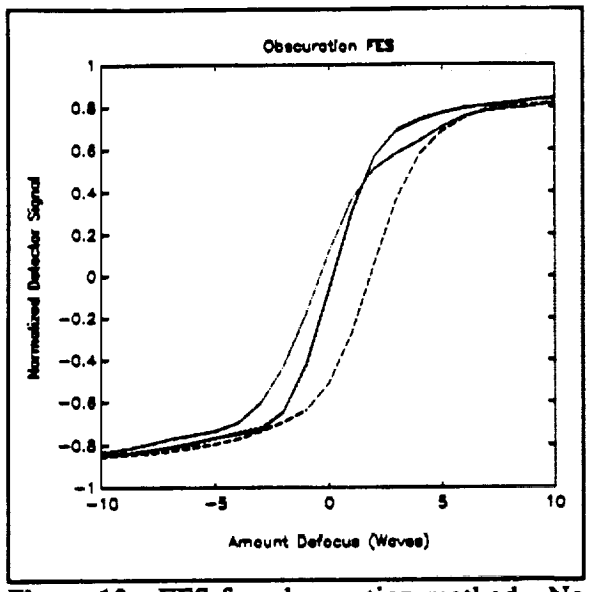


Figure 10. FES for obscuration method. No aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

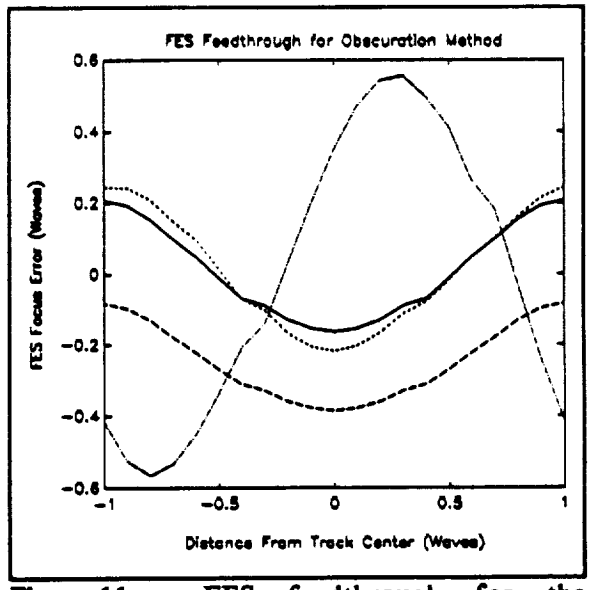


Figure 11. FES feedthrough for the obscuration method. No aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

The obscuration method is also diffraction limited, and with G=0.32, performs similarly to the ring toric method for focus error detection with regard to its behavior in the presence of aberrations, but performs somewhat worse than the ring toric method with respect to feedthrough. However, its use of the light returned from the disk is inefficient, since half of the lens must be blocked, unlike the astigmatic and ring toric methods, which use all of the available light. Push-pull tracking is possible only in the Fresnel bi-prism implementation of this method.

6. ACKNOWLEDGEMENTS

The authors wish to thank Dr. David Kay of Kodak Research Labs for helping us understand feedthrough in MO disk systems. Bruce Bernacki is supported by a Graduate Assistance in Areas of National Need (GANN) fellowship.

7. REFERENCES

1. M. Mansuripur, "Analysis of astigmatic focusing and push-pull tracking error signals in magnetooptical disk systems," *Appl. Opt.*, 26, No. 18, pp. 3981-3986, 1987.

2. G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K Schouhamer Immink, *Principles of Optical Disc Systems*, Chap. 2, Adam Hilger Ltd, Bristol, 1985.

3. A. Marchant, *Optical Recording-A Technical Overview*, Chap. 5,7, Addison-Wesley, Reading, 1990.

4. M. Mansuripur and C. Pons, "Diffraction Modeling of Optical Path for Magneto-Optical Disk Systems," Proc. SPIE, 899, pp. 592-597, 1988.

5. M.V.R.K. Murty, "Lateral Shearing Interferometers," *Optical Shop Testing*, D. Malacara,

Ed., pp. 105-148, John Wiley and Sons, New York, 1978.

6. J.H. McLeod, "The Axicon: A New Type of Optical Element,", *JOSA*, 44, No. 8, pp. 592-597, 1954.

7. J.H. McLeod, "Axicons and Their Uses," *JOSA*, 50, No. 2, pp. 166-169, 1960.

8. A. Smid, P.F. Ferdinand, and H. 't Lam, "Opto-electronic focussing-error detection system," *European Patent Application*, App. No. 85201588.2, Pub. No. 0 177 108, 1985.

APPENDIX B

IMPROVED ASTIGMATIC FOCUS ERROR DETECTION METHOD

Bruce E. Bernacki

Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

All easy-to-implement focus- and track-error detection methods presently used in magneto-optical (MO) disk drives using pre-grooved media suffer from a side effect known as feedthrough. Feedthrough is the unwanted focus error signal (FES) produced when the optical head is seeking a new track, and light refracted from the pre-grooved disk produces an erroneous FES. Some focus- and track-error detection methods are more resistant to feedthrough, but tend to be complicated and/or difficult to keep in alignment as a result of environmental insults. The astigmatic focus/push-pull tracking method is an elegant, easy-to-align focus- and track-error detection method. Unfortunately, it is also highly susceptible to feedthrough when astigmatism is present, with the worst effects caused by astigmatism oriented such that the tangential and sagittal foci are at 45° to the track direction.

This disclosure outlines a method to nearly completely eliminate the worst-case form of feedthrough due to astigmatism oriented 45° to the track direction. Feedthrough due to other primary aberrations is not improved, but performance is identical to the unimproved astigmatic method.

The new method works as follows: Light returning from the pre-grooved disk is split by a polarizing beam splitter into two equal halves when no MO signal is present. One path contains an astigmat oriented such that its axis is +45° to the track direction. The other astigmat is oriented with its axis -45°to the track direction and 90° with respect to the first astigmat. Quadrant detectors are positioned midway between the two astigmatic foci in the usual manner. Figure 1 depicts the scheme. The enhanced improvement over the common implementation is caused by the 90° rotation in the intensity pattern on each detector. As the head is moving across the tracks in the presence of astigmatism, the intensity pattern on the quadrant detector becomes elliptical in symmetry, with its major axis rotating 90° as the beam scans across the land and groove. One adds the quadrant signals in the following way

$$FES_{diff} = \frac{\dfrac{((I+III)-(II+IV))_{D1}}{(I+II+III+IV)_{D1}} - \dfrac{((I+III)-(II+IV))_{D2}}{(I+II+III+IV)_{D2}}}{2}$$

and this deleterious effect is canceled. Figure 2 shows the unimproved feedthrough plot for the astigmatic method. Defocus is added to improve the Strehl ratio in the case of spherical aberration. Figure 3 shows the increased performance with the newly invented differential approach. Figure 4 shows the increased performance with expanded scale.

This method is beneficial since it can be implemented easily in most existing designs, and also permits simultaneous differential detection of the MO data signal.

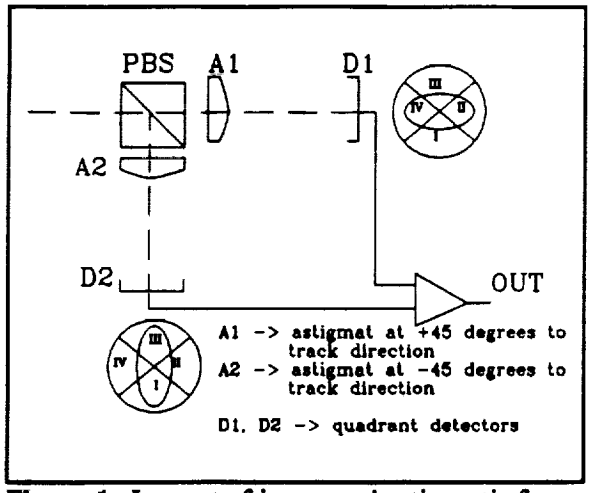


Figure 1. Layout of improved astigmatic focus error detection method.

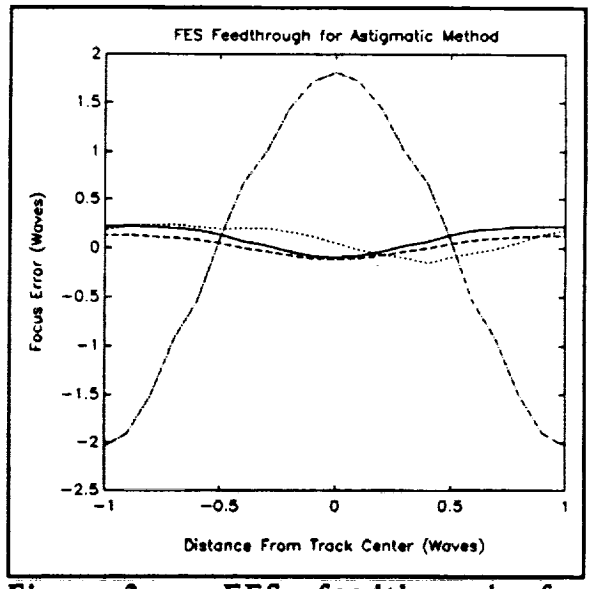


Figure 2. FES feedthrough for astigmatic/push-pull method for no aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

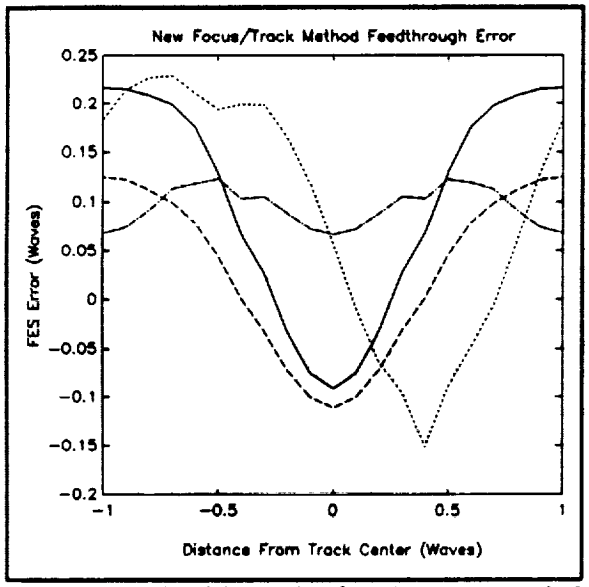


Figure 4. Feedthrough signal on expanded scale. No aberration (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

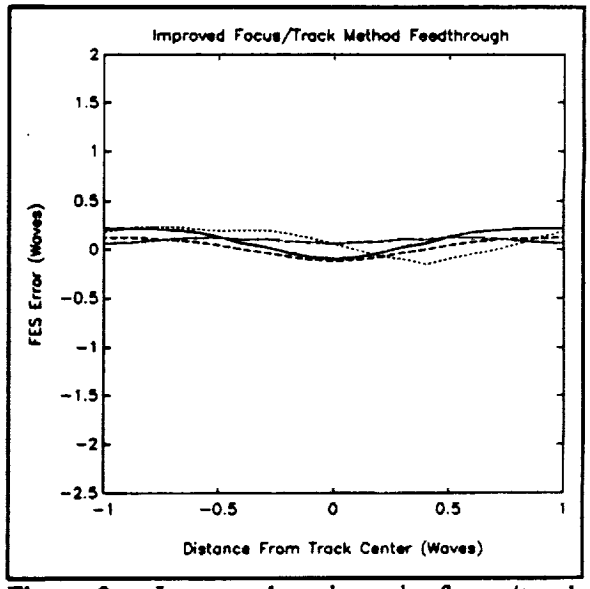


Figure 3. Improved astigmatic focus/track method with no aberrations (solid), +0.25 λ spherical aberration (dash), +0.25 λ astigmatism (dash-dot), and +0.25 λ coma (dot).

APPENDIX C

Instrumentation of the variable-angle magneto-optic ellipsometer and its application to m-o media and other non-magnetic films

Andy F. Zhou, J. Kevin Erwin and M. Mansuripur

Optical Sciences Center, University of Arizona, Tucson, AZ 85721

ABSTRACT

A new and comprehensive dielectric tensor characterization instrument is presented for characterization of magneto-optical recording media and non-magnetic thin films. Random and systematic errors of the system are studied. A series of TbFe, TbFeCo, and Co/Pt samples with different composition and thicknesses are characterized for their optical and magneto-optical properties. The optical properties of several non-magnetic films are also measured.

I. INTRODUCTION

Dielectric tensor ε characterization of M-O recording media is important for their application. It gives the important parameters like the reflectivity of the material, and, more importantly, the magneto-optic Kerr effect is determined from it. The traditional method of measuring ε consists of two steps. First using an ellipsometer[1] one obtains the refractive index n and absorption coefficient k, which are directly related to the diagonal element of the dielectric tensor. The second step consists of measuring the Kerr rotation angle θ_k and ellipticity ϵ_k using one of several available techniques[1-7] at normal incidence. From these measurements and the knowledge of n and k, the off-diagonal element of the dielectric tensor is calculated.[8] Connell used this method to obtain the dielectric tensor for a set of rare earth-transition metal alloy thin films.[9] Recently there have been some reported measurements for Co/Pt and Co/Pd superlattice samples.[10, 11] This traditional method usually requires that the magneto-optical film is thick enough to be opaque (thicker than 400 Å for TbFe and superlattice samples). This limits the application of this method and is not suitable for very thin samples which are important for application, especially the superlattice samples.

In this paper we use a new and comprehensive method[12, 13] to characterize the dielectric tensor of magneto-optical thin film samples. This method applies for all thicknesses and the measurements are done with a single apparatus. With a multilayer analysis program[14], the dielectric tensor as well as the film thickness for each layer of a multilayered film can be determined. The apparatus is also useful for determining the refractive index n, absorption coefficient k and film thickness of various layers of a multilayered non-magnetic film.

The paper is arranged as follows. In Sec. II, We discuss the principles of the technique. The experimental set up, calibration and performance analysis are discussed in detail. In Sec. III, we present some measurement results of TbFe and Co/Pt samples with various thicknesses and compositions. Various non-magnetic samples of glass, dielectric, sol-gel, organic polymer, and metals are studied in Sec. IV. Section V is the summary.

II. SYSTEM SETUP AND ANALYSIS

We first define the physical content of the dielectric tensor measured by this system. With the magnetization along the z-axis (perpendicular to the x-y plane of the film), the dielectric tensor ε is written as:[15]

$$\varepsilon = \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & 0 \\ -\varepsilon_{xy} & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{bmatrix} . \qquad (1)$$

Since the film is believed to be structurally isotropic, we shall assume $\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{xx}$. In general, the elements of the tensor are complex. Thus ε_{xx} (the diagonal element), describes the optical properties of the film in the absence of the magneto-optical effects according to the following relation

$$\sqrt{\varepsilon_{xx}} = n + ik \ . \qquad (2)$$

The off-diagonal element ε_{xy} is responsible for the magneto-optical Kerr and Faraday effects. Assuming that the Kerr rotation angle θ_k and ellipticity ϵ_k are small, and light is at normal incidence on thick film with no overcoating, the relation between θ_k, ϵ_k, ε_{xx}, and ε_{xy} can be written as

$$\theta_k - i\epsilon_k = \frac{\varepsilon_{xy}}{\sqrt{\varepsilon_{xx}}\,(\varepsilon_{xx} - 1)} \ . \qquad (3)$$

This relationship indicates that the magneto-optical Kerr effect is proportional to the off-diagonal element ε_{xy}.

Measurement of the dielectric tensor consists of measuring the complex reflection coefficients of the sample and fitting the parameters into the multilayer analysis program.

II.1. Experimental Setup

Figure 1 shows the experimental setup,[12] called variable-angle magneto-optic ellipsometer (VAMOE). The incident angle A varies from 30° to 88°. The HeNe laser beam (λ = 632.8 nm) passes through a polarizer and a quarter-wave plate whose fast axis is at 45° to the transmission axis of the polarizer. The polarizer helps to clean the laser beam to be linearly-polarized, and the quarter-wave plate makes the light to become circularly polarized at output. The polarizing beam splitter (PBS) divides the beam into two parts, sending one onto the monitor detector for laser power calibration. The other half passes through the PBS and is focused onto the sample. The PBS sits on a rotating mount and can polarize the incident beam at angle β with the plane of reference. Of those angles, the directions of p (parallel), s (perpendicular), and 45° are the ones used. The circularized light before the PBS keeps the light power incident on the sample constant for any β angle, which lets us use a constant gain for detecting circuitry. The electro-magnet behind the sample applies a magnetic field with the strength of ±7.5 kOe.

After reflecting from the sample, the beam passes through another quarter-wave plate, whose fast axis is at ζ = 45° to the reference plane. The beam then goes through a Wollaston prism and is finally detected by two detectors. The Wollaston prism and the photodetectors sit on a rotation stage whose axis is aligned with the beam. This rotation stage allows the axis of the Wollaston to be set at angle η relative to the reference plane; η = 0° and 45° are used in this experiment. The two photodetectors are identical, and their conversion factor, defined as the ratio of the output voltage to the input light intensity is α. We will denote the individual detector output by S_1 and S_2, their sum by σ, and their difference, which appears at the output of the differential amplifier, by ΔS.

To normalize the measurement results we must know the effective light amplitude that goes through the system. This can be measured by removing the sample and setting the incident angle A = 90°. If the effective light

amplitude is denoted C, the sum of the signals S_1 and S_2 will be given by

$$\sigma_0 = \frac{1}{2}\alpha C^2 \quad . \tag{4}$$

In the following discussion we will assume that the results are normalized by σ_0.

As for the discussion of the measurement technique, let us define the notation. The linearly polarized incident light with polarization direction in the plane of incidence (also the plane of reference) is the p light, while the light polarized perpendicular to this plane is the s light. For p incident light, the ordinary reflected amplitude and phase are defined as r_p and ϕ_p, respectively. Similarly for s incident light, the ordinary reflected amplitude and phase are r_s and ϕ_s. For magneto-optical samples the p (or s) incident light not only causes ordinary reflection in p (or s) direction but also induces a magneto-optical component in the s (or p) direction. We denote the amplitude and phase of this magneto-optical reflection coefficient by $r_\perp$ and $\phi_\perp$, respectively.

When the sample is placed in the system, the signals S_1 and S_2 will in general depend on r_p, ϕ_p, r_s, ϕ_s, $r_\perp$ and $\phi_\perp$, as well as on β, ζ and η. Then, the sum and difference signals between S_1 and S_2 are derived[12]

$$\begin{aligned} \sigma/\sigma_0 = {} & r_p^2\cos^2\beta + r_s^2\sin^2\beta + r_\perp^2 \\ & + [r_p r_\perp \cos(\phi_p-\phi_\perp) - r_s r_\perp \cos(\phi_s-\phi_\perp)]\sin 2\beta \end{aligned} \tag{5}$$

and

$$\begin{aligned} \Delta S/\sigma_0 = {} & [(r_p^2\cos^2\beta - r_s^2\sin^2\beta)\cos 2\zeta \\ & - r_\perp^2\cos(2\beta-2\zeta)]\cos(2\eta-2\zeta) \\ & + r_p r_s[\sin 2\zeta\cos(\phi_p-\phi_s)\cos(2\eta-2\zeta) \\ & + \sin(\phi_p-\phi_s)\sin(2\eta-2\zeta)]\sin 2\beta \\ & - 2r_p r_\perp[\cos\beta\sin(\phi_p-\phi_\perp)\sin(2\eta-2\zeta) \\ & + \sin(2\zeta-\beta)\cos(\phi_p-\phi_\perp)\cos(2\eta-2\zeta)]\cos\beta \\ & - 2r_s r_\perp[\sin\beta\sin(\phi_s-\phi_\perp)\sin(2\eta-2\zeta) \\ & - \cos(2\zeta-\beta)\cos(\phi_s-\phi_\perp)\cos(2\eta-2\zeta)]\sin\beta \end{aligned} \tag{6}$$

Seven different experimental set-ups are used to fully characterize the sample. They are combinations of different angles of β and η, while keeping $\zeta = 45°$. The seven curves with the setting specifications are listed at Table I.

For measuring the dielectric tensor, these seven different combinations of reflection coefficients versus the incident angle A are measured. Of these, three curves are optical reflectivities, which are independent of magnetic states of the samples, and are obtained with the magnet turned off, while the other four curves are measured with the sample saturated by the magnetic field and are called magneto-optic reflectivity curves. For non-magnetic materials, all the M-O reflection coefficients are zero and only the first three curves are measured.

After measuring these seven curves, a multilayer analysis program[14] is used to analyze the data and estimate the dielectric tensor of the film. This analysis program can deal with multilayer structures containing an arbitrary number of dielectric, metal and magnetic layers. An algorithm based on 2 × 2 matrices is used for the reflection calculation. In this calculation, the incident beam is assumed to be plane monochromatic with arbitrary angle of incidence. There are no approximations involved and the results are direct consequences of Maxwell's equations.

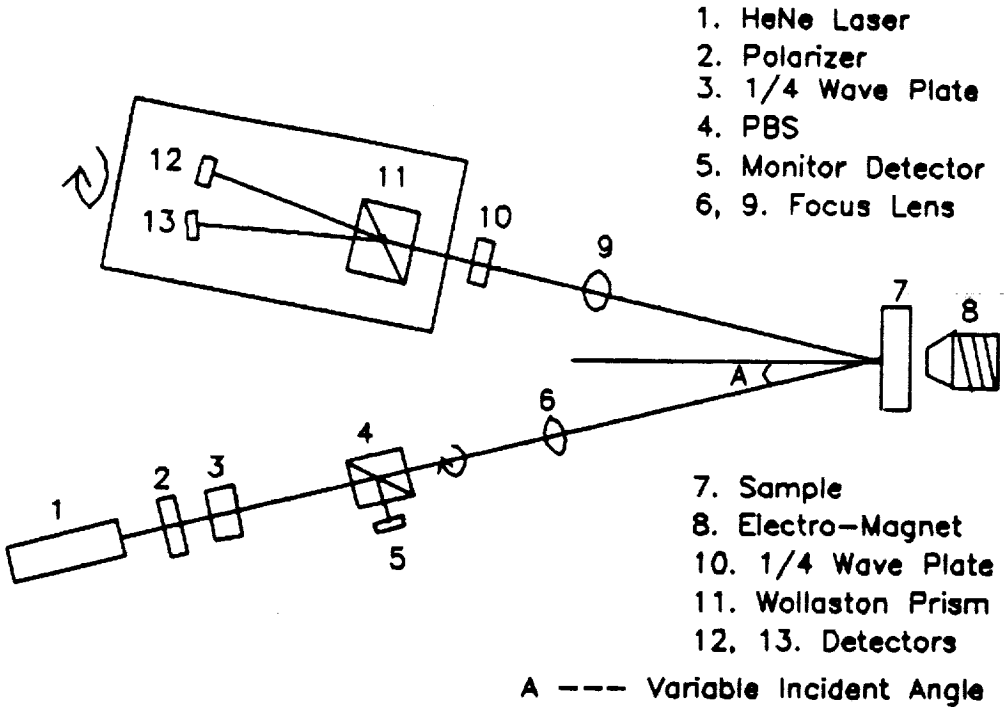


Figure 1. System setup for the variable-angle magneto-optic ellipsometer.

Table I. The measured reflectivity curves. ζ = 45° for all the measurements.

Curve #	Detector Signal Used	PBS Setting β(°)	Wollaston Setting η(°)	Physical description
1	σ/σ_0	0	45	r_p^2
2	σ/σ_0	90	45	r_s^2
3	$\Delta S/\sigma_0$	45	45	$r_p r_s \cos(\phi_p - \phi_s)$
4	$\Delta S/\sigma_0$	0	45	$r_p r_\perp \cos(\phi_p - \phi_\perp)$
5	$\Delta S/\sigma_0$	0	0	$r_p r_\perp \sin(\phi_p - \phi_\perp)$
6	$\Delta S/\sigma_0$	90	45	$r_s r_\perp \cos(\phi_s - \phi_\perp)$
7	$\Delta S/\sigma_0$	90	0	$r_s r_\perp \sin(\phi_s - \phi_\perp)$

II.2. Error Analysis

There are basically two types of error, random and systematic, in this dielectric tensor characterization system. Random errors come from the vibration of the system, fluctuations of the laser power, the random noise in the circuitry and the numerical errors of the multilayer analysis program. Since the individual analysis of them is complicated, we choose to experimentally measure their combined effect.

The random error in the VAMOE system is expressed in the random fluctuation of the measured data. The effect of this error could be different for different types of samples, due to their different shapes of reflectivity curves. To a good estimation, we measured three typical samples. A piece of glass (microscope-slide) which has only n value (k is so close to zero, that its measurement is below the resolution of this system) is typical for dielectric materials. A Co/Pd multilayered thin film with thickness of 165 Å is used as a typical sample for the superlattice materials. A TbFe sample (with overcoating) is typical for the RE-TM materials.

Figure 2 shows the measured reflectivity curves for four independent measurements of the glass sample. The back of the microscope-slide was ground and blackened by ink, in order to reduce reflection from the back surface. As can be seen in Fig. 2, the four independent measurements are coincide with each other. Table II lists the matched n values for each measurement. The data matching error (DME) in this and later tables is the average percent deviation for each data point between the measured data and the calculated data from the multilayer analysis program by using matched parameter(s). The matching error is less than 3% and the standard deviation of the four results is 0.5%.

Figure 3 is the corresponding three independent measurements for a Co(2Å)/Pd(9Å) sample. This sample is sputtered on glass at 7 mtorr of Kr gas, and the total thickness is 165 Å. There is no overcoating. The M-O reflectivity curves have larger difference among the three independent measurements than the curves 1 to 3, which is due to the much smaller signal for these M-O data. Table III is the corresponding results for the Co/Pd sample. The results in the parenthesis (also in the later tables) are measured by our variable wavelength Kerr rotation angle and ellipticity measuring system called magneto-optic Kerr spectrometer (MOKS).[5,6] MOKS measures the Kerr rotation angle and ellipticity using a different method from that of the VAMOE system.

The three measurements of the seven reflection curves for a $Tb_{22.1}Fe_{77.9}$ sample are shown in Fig. 4. The M-O film thickness is 1354 Å. The overcoating is $Si_{32.1}O_{67.9}$ and the estimated thickness is 1264 Å. The thickness measured by this system (the t_{eo} in the table) is 1326 Å. For the SiO overcoating we use n = 1.449 and k = 0 for the calculation. Table IV lists the computed parameters for the TbFe sample. Connell's measurement for a $Tb_{21}Fe_{79}$ sample[9] is cited here for comparison.

From the measurements for the three typical samples, we can see that the data matching error is less than 5% and the random error is within the range of tolerance. The slightly larger DME for the TbFe sample is due to its higher complexity with overcoating. The random errors for the Kerr rotation angle, ellipticity and reflectivity are much smaller, which indicates their less sensitivity to fluctuation. The smaller random error for the off-diagonal element of the TbFe sample is due to its much larger magneto-optical reflectivity.

There are at least five possible systematic error sources in the system. One is the incident angle error. We measured this angle by the top view pictures, taken by a camera, for angles from 26° to 90° (with one degree increment). After this, the error on the angle measurements, or the error on the mechanical movements on the rail will be random and they fall into the category of the random error analysis discussed above. The second error source is the gains in the two signal channels and the differential channel of the amplifier circuit. The circuit has been stabilized and calibration was done to wipe out any sizable systematic error on this part. The other three error sources are the accuracy of the polarizer angle for the incident beam, the accuracy of the quarterwave plate angle and the accuracy on the angle readings of the detector box.

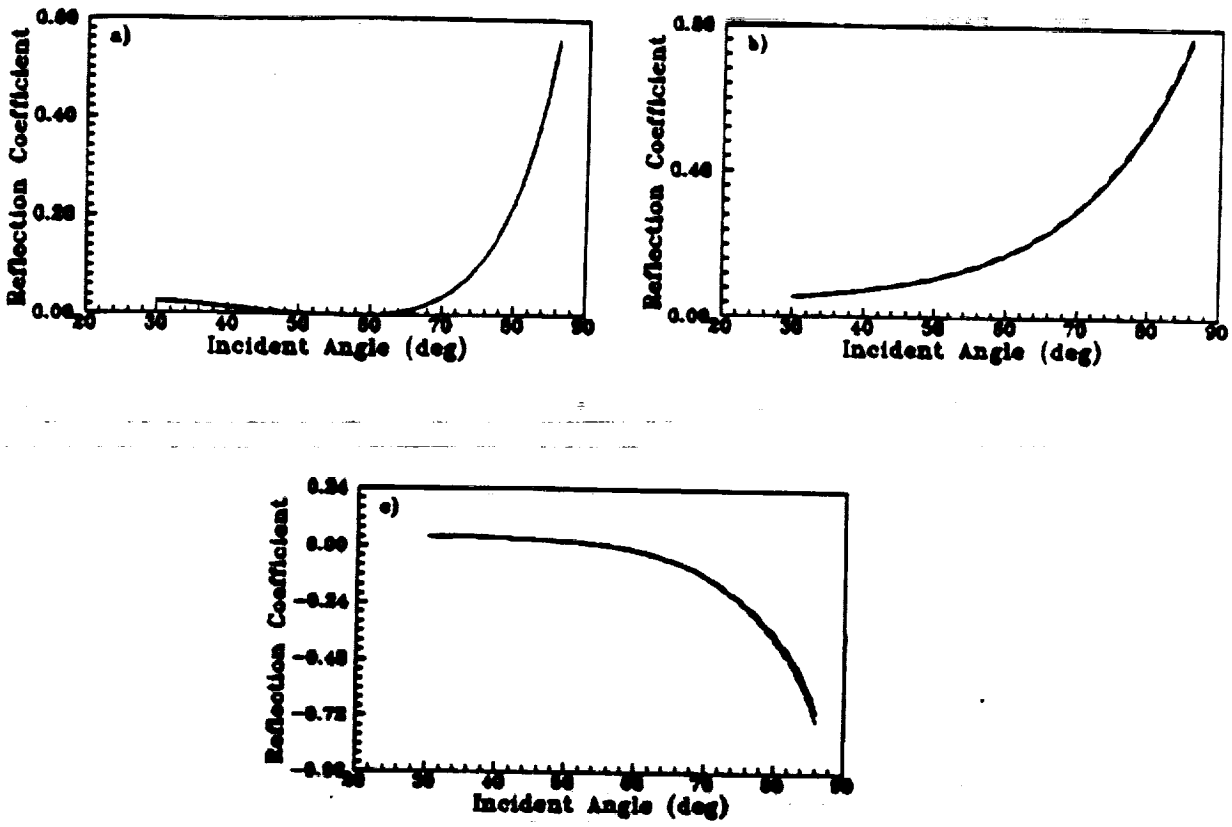


Figure 2. The four independently measured reflection curves (measurements 1 - 4 are "————", "- - - -", "— — —" and "—— ——" lines, respectively) of (a) curve 1, (b) curve 2 and (c) curve 3 for a glass sample.

Table II. Matched values of n for the four measurements of glass.

#	1	2	3	4	average	σ_{n-1}
n	1.496	1.499	1.498	1.512	1.501	0.007
DME (%)	1.9	2.8	2.9	1.1	—	—

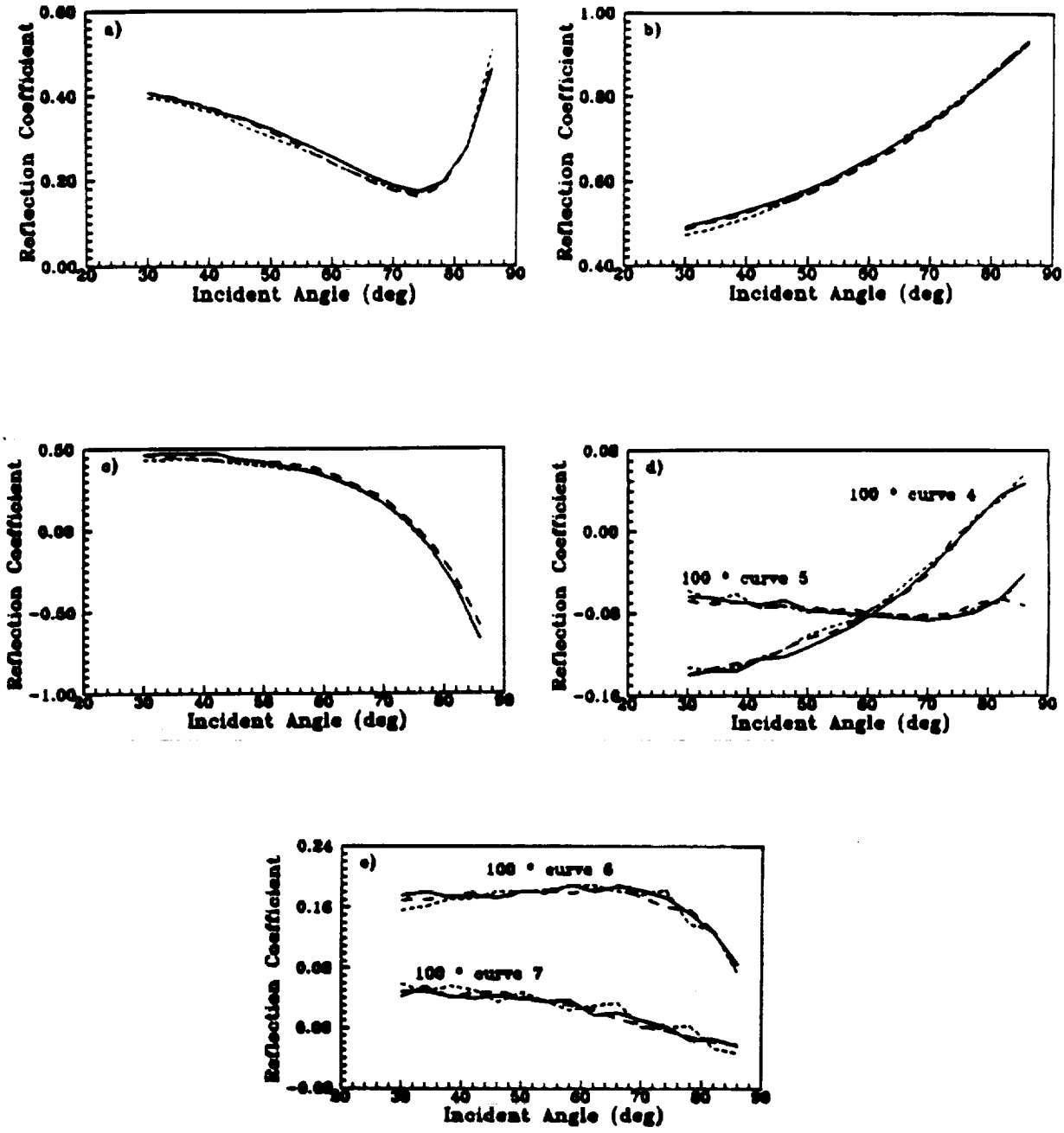


Figure 3. The three independently measured reflection curves (measurements 1 - 3 are "————", "- - - -" and "— — —" lines, respectively) of (a) curve 1, (b) curve 2, (c) curve 3, (d) curves 4 and 5, and (e) curves 6 and 7 for a Co/Pd sample.

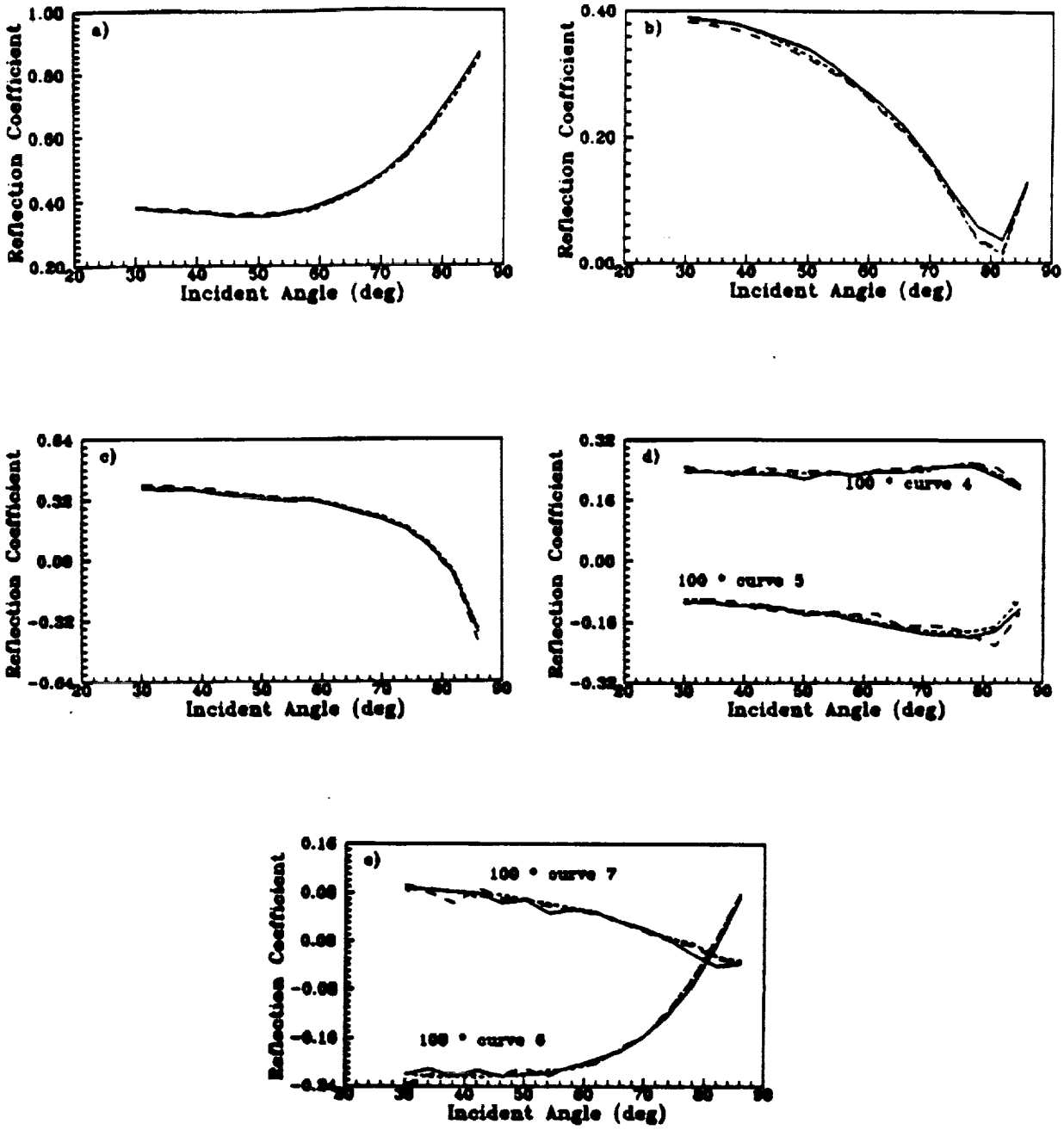


Figure 4. The three independently measured reflection curves (measurements 1 - 3 are "————", "- - - -" and "— — —" lines, respectively) of (a) curve 1, (b) curve 2, (c) curve 3, (d) curves 4 and 5, and (e) curves 6 and 7 for a TbFe sample.

Table III. Three measurements for the Co/Pd sample. The results in the parenthesis are measured from the MOKS system.

#	1	2	3	average	σ_{n-1}
n	1.46	1.42	1.75	1.54	0.2
k	3.81	3.71	3.84	3.79	0.07
ε_{xx}	-12.4 + i11.1	-11.8 + i10.5	-12.0 + i12.6	-12.1 + i11.4	0.3 + i1.0
ε_{xy}	0.148 — i0.100	0.138 — i0.100	0.148 — i0.122	0.145 — i0.107	0.006 — i0.013
DME (%)	2.8	3.0	3.0	—	—
θ_k	0.202	0.205	0.201	0.203 (0.19)	0.002
ϵ_k	0.074	0.080	0.073	0.075 (0.06)	0.004
R	0.450	0.436	0.451	0.446 (0.46)	0.008

Table IV. Three measurements for the TbFe sample. The results in the parenthesis are measured from the MOKS system.

#	1	2	3	average	σ_{n-1}	Connell's
n	2.32	2.38	2.40	2.37	0.04	—
k	3.08	3.10	3.14	3.11	0.03	—
t_{SiO}(Å)	1332	1323	1322	1326	6	—
ε_{xx}	-3.94 + i15.5	-3.82 + i15.9	-3.99 + i15.5	-3.92 + i15.6	0.09 + i0.2	-2.87 + i19.9
ε_{xy}	-0.250 — i0.069	-0.254 — i0.065	-0.247 — i0.065	-0.250 — i0.066	0.004 — i0.002	-0.338 — i0.028
DME (%)	3.0	3.0	4.9	—	—	—
θ_k	-0.310	-0.311	-0.312	-0.311 (-0.31)	0.001	-0.305
ϵ_k	0.131	0.131	0.125	0.129 (0.12)	0.003	0.127
R	0.422	0.417	0.418	0.419 (0.42)	0.003	0.420

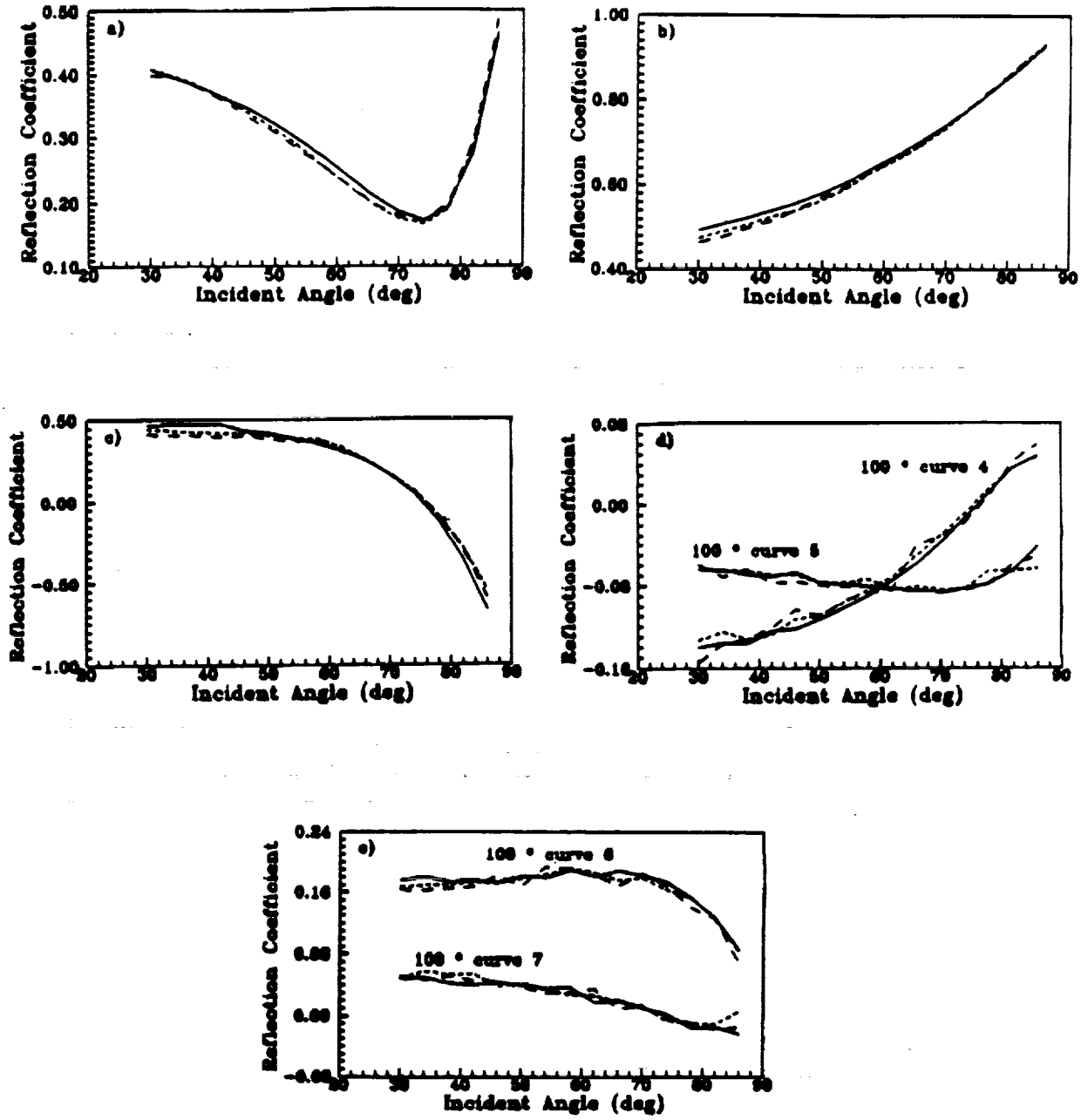


Figure 5. Reflection coefficient measurements for the case that the polarizer angles are at the calibrated angle ("————" line), 2° off ("- - - -" line) and -2° off ("— — —" line), respectively, for systematic error analysis.

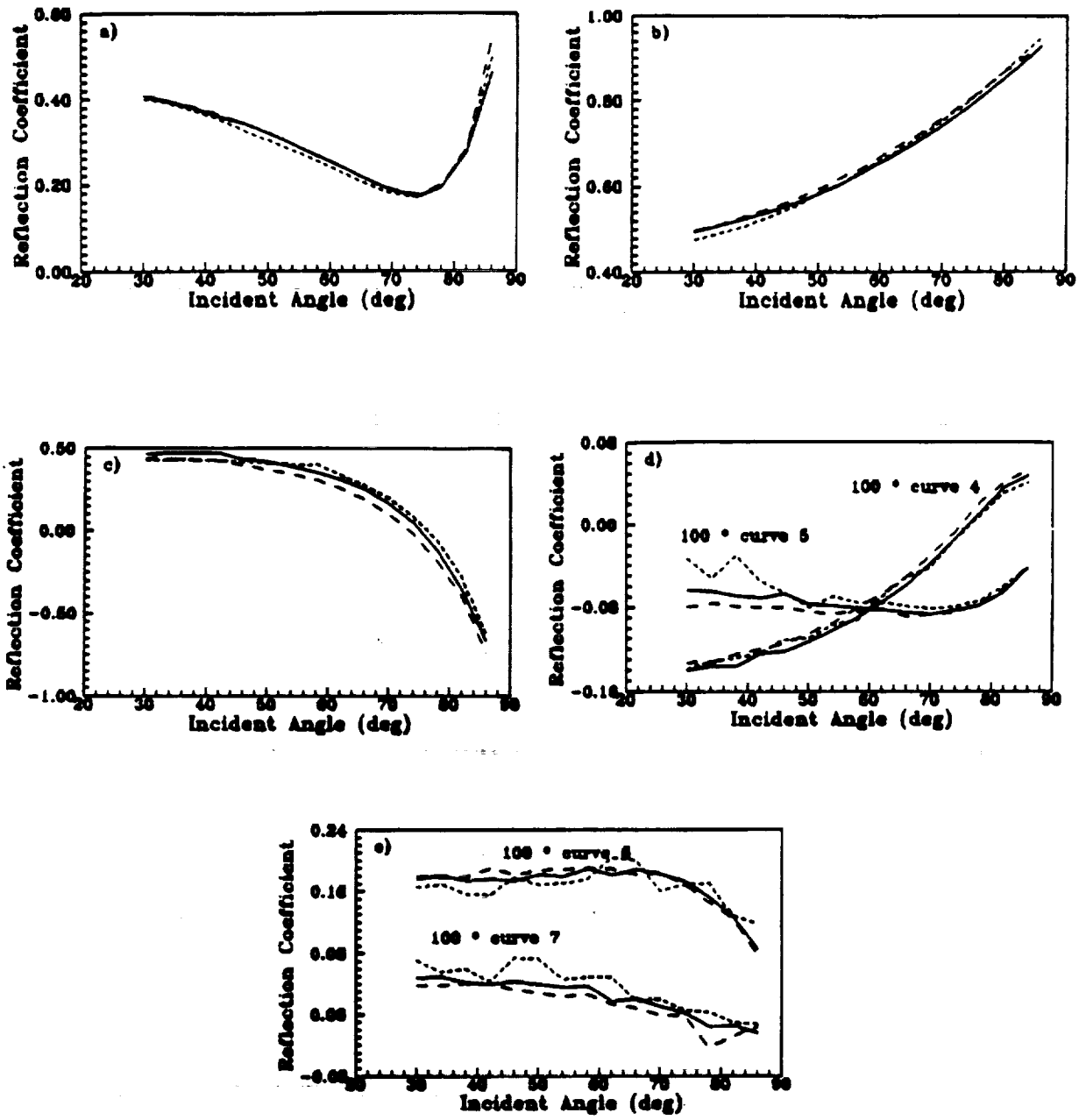


Figure 6. Reflection coefficient measurements for the case that the quarterwave plate are at the calibrated angle ("————" line), 2° off ("- - - -" line) and -2° off ("— — —" line), respectively, for systematic error analysis.

Table V. Measurements for the case that the polarizer angle is ±2° off the calibrated angle for the Co/Pd sample.

	average (no shift)	σ_{n-1} (no shift)	polarizer off 2°	Δ (off 2°)	polarizer off -2°	Δ (off -2°)
n	1.54	0.2	1.67	0.13	1.53	-0.01
k	3.79	0.07	3.78	-0.01	3.72	-0.07
ε_{xx}	-12.1 + i11.4	0.3 + i1.0	-11.4 + i12.4	-0.7 + i1.0	-11.7 + i10.6	0.4 + -i0.8
ε_{xy}	0.145 — i0.107	0.006 — i0.013	0.140 — i0.124	-0.005 — i0.017	0.137 — i0.101	-0.008 — i0.006
DME (%)	3.0	—	3.1	—	3.9	—
θ_k	0.203	0.002	0.203	0	0.205	0.002
ϵ_k	0.075	0.004	0.075	0	0.080	0.005
R	0.446	0.008	0.441	-0.005	0.436	-0.010

Table VI. Measurements for the case that the quarterwave plate angle is ±2° off the calibrated angle for the Co/Pd sample.

	average (no shift)	σ_{n-1} (no shift)	1/4 λ off 2°	Δ (off 2°)	1/4 λ off -2°	Δ (off -2°)
n	1.54	0.2	1.58	0.04	0.92	-0.62
k	3.79	0.07	3.78	-0.01	3.67	-0.12
ε_{xx}	-12.1 + i11.4	0.3 + i1.0	-11.5 + i12.7	0.06 + i1.3	-12.43 + i6.77	-0.3 + -i4.6
ε_{xy}	0.145 — i0.107	0.006 — i0.013	0.135 — i0.132	-0.01 — i0.025	0.130 — i0.047	-0.015 — -i0.060
DME (%)	3.0	—	5.5	—	4.0	—
θ_k	0.203	0.002	0.196	-0.007	0.208	0.005
ϵ_k	0.075	0.004	0.082	0.007	0.078	0.003
R	0.446	0.008	0.444	-0.002	0.441	-0.005

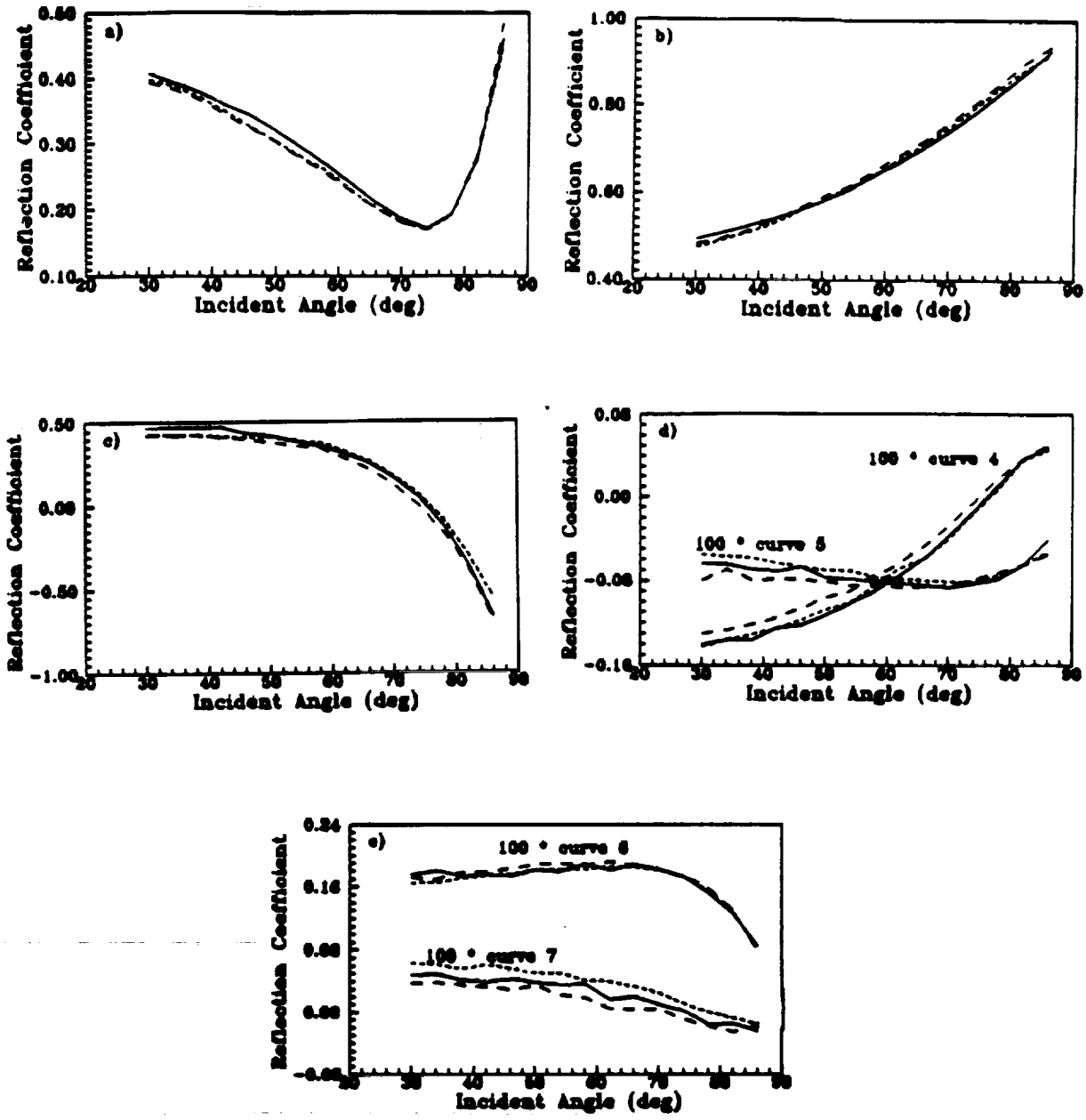


Figure 7. Reflection coefficient measurements for the case that the detector angles are at the calibrated angle ("————" line), 2° off ("- - - -" line) and -2° off ("— — —" line), respectively, for systematic error analysis.

Table VII. Measurements for the case that the detector angle is ±2° off the calibrated angle for the Co/Pd sample.

	average (no shift)	σ_{n-1} (no shift)	detector off 2°	Δ (off 2°)	detector off -2°	Δ (off -2°)
n	1.54	0.2	1.77	0.23	1.33	-0.21
k	3.79	0.07	3.78	-0.01	3.68	-0.11
ε_{xx}	-12.1 + i11.4	0.3 + i1.0	-11.1 + i13.5	1.0 + i2.1	-11.77 + i9.44	0.33 + -i2.0
ε_{xy}	0.145 — i0.107	0.006 — i0.013	0.135 — i0.149	-0.010 — i0.042	0.140 — i0.086	-0.005 — -i0.021
Q	1.3	—	1.2	—	1.6	—
DME (%)	3.0	—	2.9	—	4.0	—
θ_k	0.203	0.002	0.201	-0.002	0.215	0.012
ϵ_k	0.075	0.004	0.084	0.009	0.079	0.004
R	0.446	0.008	0.443	-0.003	0.432	-0.014

To study the possible effect of the three remaining systematic error sources, we measured the dielectric tensor with the polarizer, quarterwave plate and detector module deviated ±2° off the given angle, respectively, for the same Co/Pd sample which was used for the random error analysis. Figure 5 shows the measured reflection curves for the polarizer at normal position, off +2° and off -2° positions. The same measurements for the quarterwave plate and detector module are shown in Figs. 6 and 7, respectively. The calculated variation on the dielectric tensor, Kerr rotation, ellipticity and reflectivity for these systematic angle changes are listed in Table V, VI, and VII.

The variation on the measured results for the polarizer ±2° off is small and even within the random error range. Since the accuracy of the polarizer angle is better than that, it will cause no systematic error problem. The variation on the measured results for the quarterwave plate ±2° off is much larger and more attention needs to be paid for the accuracy of this angle. The resolution of this angle is about 0.25° and an accuracy of less than 1° can be achieved. This could make the possible systematic error much smaller. The variation on the results for the detector box angle ±2° off is between the above two cases. The accuracy of this angle is about 1° and the systematic error should therefore be under control.

Another way to check the systematic error is to compare the measured Kerr rotation angle, ellipticity and reflectivity with other established characterization system, like the MOKS system. The good agreement on the Kerr rotation, ellipticity and reflectivity measurements between these two systems (see Tables III and IV) also indicates that the systematic error in this VAMOE system is insignificant.

III. DIELECTRIC TENSOR FOR TbFe AND Co/Pt SAMPLES

Dielectric tensor characterization for the M-O samples is important for both understanding and application of these materials. Thickness dependence of the dielectric tensor for Co/Pd superlattice multilayered film has been studied in our previous paper,[16] and we found that both diagonal and off-diagonal elements of the tensor are more or less constant for film thicknesses greater than 150 Å. In the same paper, we also have studied the composition dependence of the dielectric tensor for both Co/Pt and Co/Pd multilayered films. The enhancement of the off-diagonal element for the multilayered samples over diluted pure cobalt explains the larger M-O Kerr effect for these multilayered samples. In this section, we first present the dielectric tensor measurement for several TbFe and TbFeCo samples with different compositions. Then, more measurements for Co/Pt samples are presented, which extends both the thickness and composition range of the samples for study.

For these TbFe and TbFeCo samples, they all have overcoating, and are with glass substrate. Their hysteresis loops are square. The detailed structural and magnetic data for all the TbFe and TbFeCo samples are listed in Table VIII. Figure 8 shows a typical measurement of the seven reflection curves along with the calculated curves from the matched dielectric tensor values, for sample TB6. The matching between the experimental and calculated reflectivity curves is excellent. All of the data matching errors are less than 6%, as listed in Table VIII. The DME decreases with larger reflectivity value which has better signal-to-noise ratio. The slightly larger difference on the Kerr rotation, ellipticity and reflectivity between the VAMOE and MOKS systems for TbFe and TbFeCo samples than that of the superlattice samples is caused by the higher complexity of the former types of sample. One notices that as TbFe sample changes from the Tb-rich (sample TF6) to Fe-rich (the rest), the off-diagonal element changes sign as well as the Kerr rotation.

Table VIIII lists all the structural and magnetic data for the Co/Pt samples studied here. There is no overcoating for any of them and their hysteresis loops are all square. (They are deposited on glass substrate.) Figure 9 shows a typical measurement for sample COPT3. The typical data for a Co/Pd sample has been shown in the previous error analysis section.

IV. OPTICAL MEASUREMENTS FOR NON-MAGNETIC SAMPLES

The VAMOE system developed for M-O media can also be used to measure the optical properties for various non-magnetic films. Since M-O data storage application also involves the characterizations for the substrate, overcoating and reflecting layer materials, it is useful that this system is able to perform this characterization. In this section, we present the optical refractive index n, absorption coefficient k and film thickness measurements for some non-magnetic materials.

Figures 10 to 13 show the measured reflection coefficient curves as a function of the incident angle for a glass substrate, a dielectric thin film, a sol-gel thin film and an organic polymer film. Table X lists the measurement results for these and other samples. The dielectric and organic polymer films are coated on glass substrate and the sol-gel films are coated on a GaAs substrate with $n = 3.85$ and $k = 0.077$.

Series of metallic films of Al, Al-Ti, Cu_3Al, Al_2Cr, Cu and Pt with different thicknesses are measured and their results are listed on Table XI. The typical reflection coefficient curves for each kind are shown in Figs. 14 to 19. The matching between the experimentally measured curves and those computer calculated from the estimated optical constants is quite good. All the DME are less than 3%. Also in this table, we cited the n, k values for Al, Cu and Pt from handbook.[17]

V. SUMMARY

In summary, we have constructed a new and comprehensive dielectric tensor characterization system, both for magneto-optical and non-magnetic thin films, at $\lambda = 632.8$ nm. The random and systematic error analysis of

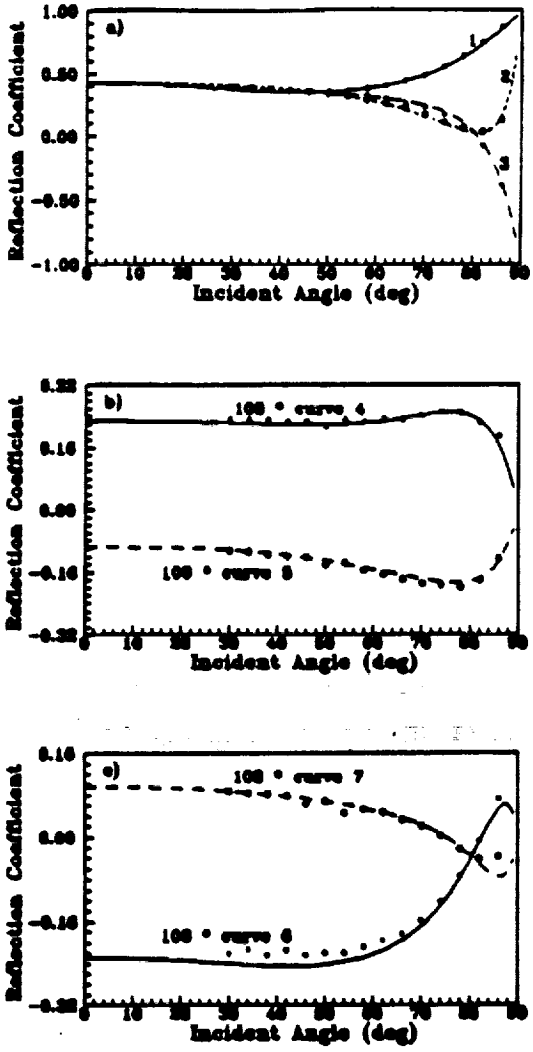


Figure 8. Seven reflection coefficients as function of the angle of incidence for TF6 sample. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated dielectric tensor.

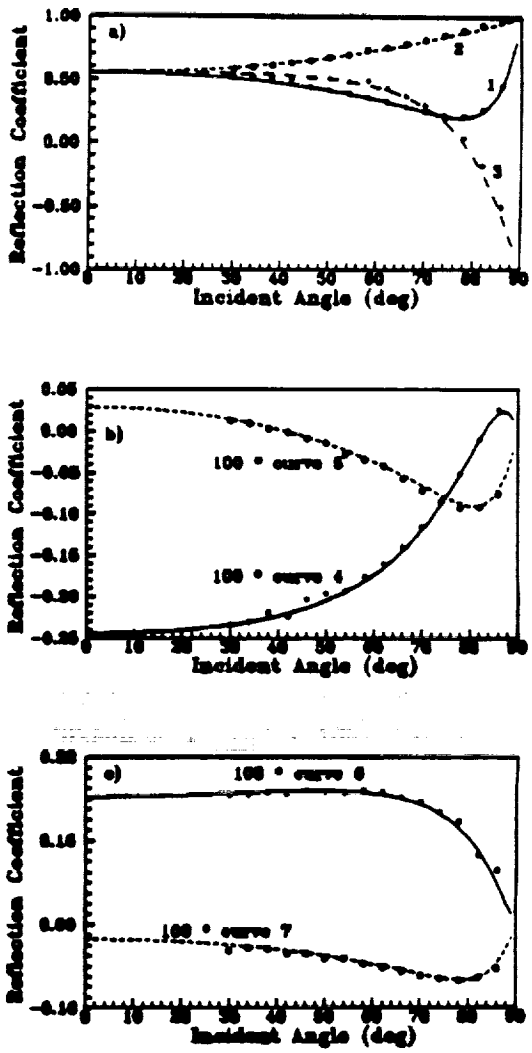


Figure 9. Seven reflection coefficients as function of the angle of incidence for COPT3 sample. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated dielectric tensor.

Table VIII. Measurement results of the TbFeCo and TbFe samples with various film compositions at λ=633nm. The values in the parentheses are measured by the magneto-optic Kerr spectrometer (MOKS) system.

Samples	TF1	TF2	TF3	TF4	TF5	TF6	TF7
MO composition (at. %)	Tb:16.1 Fe:77.0 Co: 6.9	Tb:20.1 Fe:71.3 Co: 8.6	Tb:20.3 Fe:70.9 Co: 8.9	Tb:16.7 Fe:83.3	Tb:14.5 Fe:85.6	Tb:28.1 Fe:71.9	Tb:18.3 Fe:74.5 Ar: 3.6
overcoating thickness (Å)	SiN 647	SiN 800	SiN 800	SiN 757	SiN 791	SiO 1326	SiO 1615
MO film thickness (Å)	431	1000	500	733	726	1354	554
underlayer thickness (Å)	-	-	-	SiN 716	SiN 723	-	-
coercivity (kOe)	2.42	3	2.22	2.24	0.92	1.63	1.26
n	2.32	2.43	2.91	2.30	2.42	2.37	2.72
k	3.78	3.24	3.31	3.23	3.19	3.11	2.85
ε_{xx}	-8.90 + i17.54	-4.60 + i15.80	-2.49 + i18.52	-5.14 + i14.83	-4.22 + i14.67	-3.92 + i15.60	-3.74 + i8.68
ε_{xy}	0.665 + i0.160	0.319 + i0.118	0.417 - i0.033	0.294 + i0.105	0.273 + i0.074	-0.250 - i0.066	0.121 + i0.041
DME (%)	5.4	3.2	3.1	3.0	3.5	3.0	1.2
θ_K(°)	0.675 (0.617)	0.631 (0.591)	0.679 (0.660)	0.738 (0.699)	0.682 (0.642)	-0.311 (-0.310)	0.189 (0.193)
ε_K(°)	-0.245 (-0.319)	0.033 (0.047)	0.077 (0.081)	0.044 (0.039)	0.034 (0.036)	0.129 (0.120)	-0.088 (-0.077)
R	0.180 (0.186)	0.301 (0.316)	0.221 (0.229)	0.246 (0.250)	0.240 (0.242)	0.419 (0.416)	0.501 (0.457)

Table VIIII. Measurement results of the Co/Pt samples with various film compositions at λ=633nm. The values in the parentheses are measured by the magneto-optic spectrometer (MOKS) system.

sample	COPT1	COPT2	COPT3	COPT4
composition of Co/Pt in Å	2.4/9.6	7.2/23.8	2.4/7.2	3/10
Co content	0.20	0.23	0.25	0.23
number of bilayers	14	12	17	20
MO film thickness (Å)	180	180	180	260
coercivity (kOe)	0.62	0.70	0.44	2.24
n	1.71	1.99	2.19	2.18
k	4.67	4.67	4.45	4.13
ε_{xx}	-18.9 + i16.0	-17.9 + i15.8	-15.3 + i17.6	-13.2 + i14.8
ε_{xy}	0.251 - i0.040	0.390 - i0.040	0.365 - i0.070	0.165 - i0.008
DME (%)	4.9	4.3	3.4	2.8
θ_K(°)	0.165 (0.16)	0.243 (0.23)	0.255 (0.24)	0.119 (0.114)
ε_K(°)	0.017 (0.01)	-0.018 (-0.02)	-0.030 (-0.03)	-0.013 (-0.012)
R	0.59 (0.60)	0.58 (0.57)	0.55 (0.53)	0.604 (0.612)

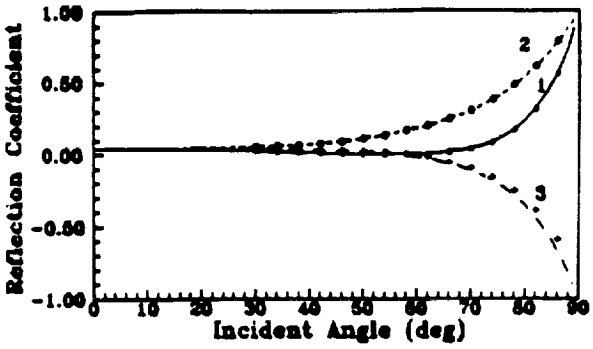


Figure 10. Three reflection coefficients as function of the angle of incidence for the glass sample on Table X. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 1.500, and k = 0 from the multilayer analysis program.

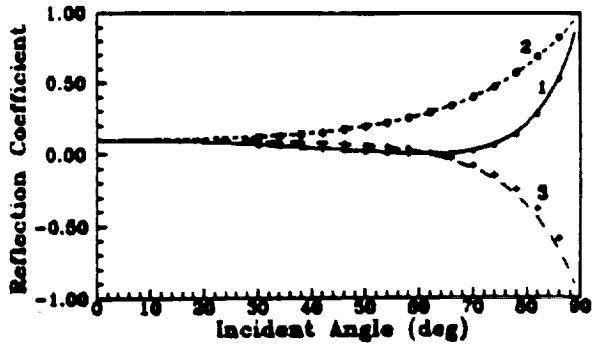


Figure 11. Three reflection coefficients as function of the angle of incidence for the dielectric film listed on Table X. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 2.12, k = 0 and film thickness = 235 Å from the multilayer analysis program.

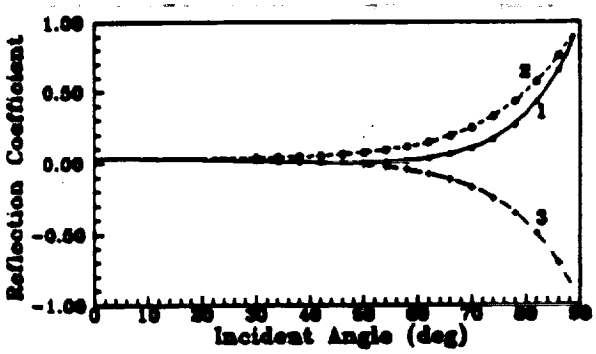


Figure 12. Three reflection coefficients as function of the angle of incidence for the sol-gel (high index) film listed on Table X. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 2.17, k = 0 and film thickness = 83 Å from the multilayer analysis program.

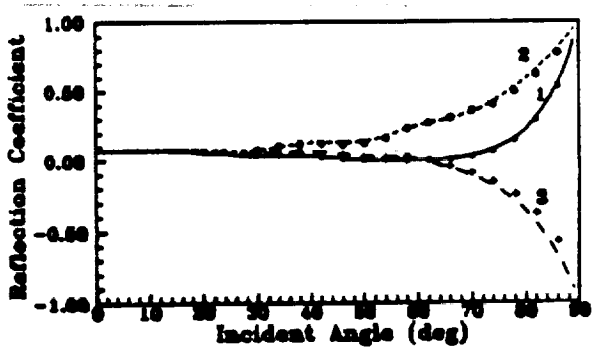


Figure 13. Three reflection coefficients as function of the angle of incidence for the organic polymer sample 9-D, listed on Table X. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 1.68, k = 0.030 and film thickness = 2700 Å from the multilayer analysis program.

Table X. Measurement results of several non-magnetic and non-metal materials at λ=633nm.

sample	n	k	film thickness (Å)	under-layer (Å)	DME (%)
Corning micro slide #2941	1.500	0	10^7	0	2.3
ZrO_2 dielectric	2.120	0	235	0	2.3
Sol-gel (high index)	2.17	0	83	0	0.8
Sol-gel (low index)	1.45	0	84	0	1.1
Organic polymer 7-D	1.71	0.010	5300	265 (ITO)	6.0
Organic polymer 9-B	1.61	0.044	1500	265 (ITO)	4.7
Organic polymer 9-D	1.68	0.030	2700	265 (ITO)	4.7

Figure 14. Three reflection coefficients as function of the angle of incidence for the aluminum sample #3, listed on Table XI. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 1.05, k = 5.88 and film thickness = 800 Å from the multilayer analysis program.

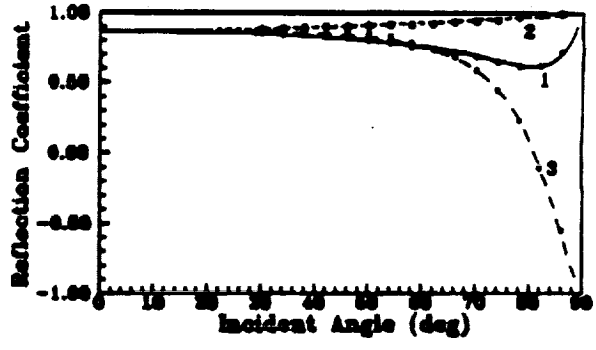


Figure 15. Three reflection coefficients as function of the angle of incidence for the Al-Ti sample #2, listed on Table XI. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 1.45, k = 6.11 and film thickness = 1000 Å from the multilayer analysis program.

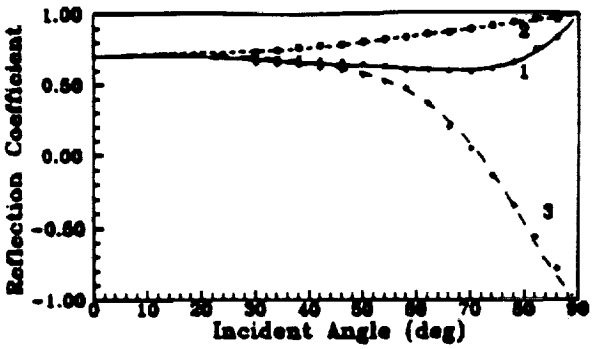


Figure 16. Three reflection coefficients as function of the angle of incidence for the Cu_5Al sample #1, listed on Table XI. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 1.41, k = 4.53 and film thickness = 500 Å from the multilayer analysis program.

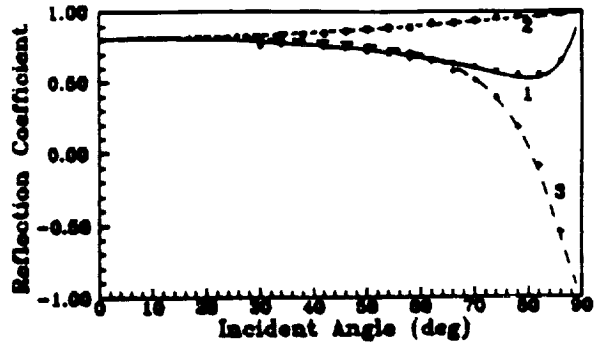


Figure 17. Three reflection coefficients as function of the angle of incidence for the Al_2Cr sample #2, listed on Table XI. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 1.76, k = 5.53 and film thickness = 800 Å from the multilayer analysis program.

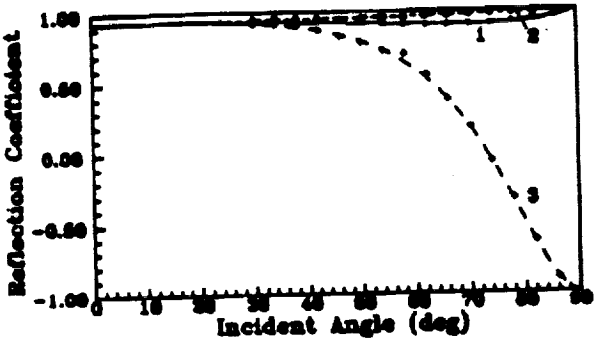


Figure 18. Three reflection coefficients as function of the angle of incidence for the copper sample #2, listed on Table XI. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 0.41, k = 5.83 and film thickness = 1000 Å from the multilayer analysis program.

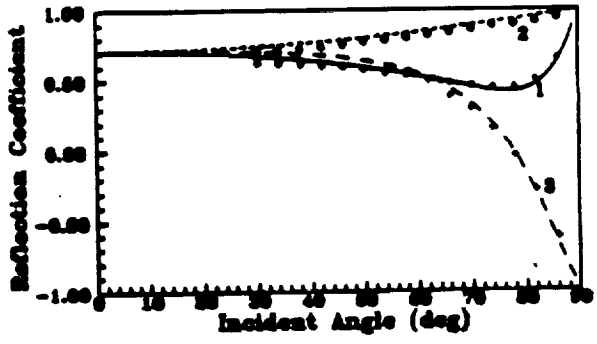


Figure 19. Three reflection coefficients as function of the angle of incidence for the platinum film, listed on Table XI. The symbols (star, circle and cross) are the measured ones and the continuous curves are calculated with the estimated n = 1.72, k = 4.05 and film thickness = 1510 Å from the multilayer analysis program.

Table XI. Measurement results of several non-magnetic metal samples with various film thicknesses at λ=633nm.

sample	over-coating (Å)	n	k	film thickness (Å)	DME (%)
Al #1	0	1.01	6.25	223	2.7
Al #2	0	1.05	5.83	500	2.3
Al #3	0	1.05	5.88	800	1.5
Al #4	0	1.03	5.97	1000	2.1
Al-Ti #1	0	1.67	6.08	438	1.9
Al-Ti #2	0	1.45	6.11	1000	1.1
Al-Ti #3	0	1.60	6.32	488	2.1
Al-Ti #4	0	1.35	5.51	650	1.5
Cu_9Al #1	166 (ZrO_2)	1.41	4.53	500	1.9
Cu_9Al #2	138 (ZrO_2)	1.31	4.59	1000	1.8
Al_2Cr #1	0	1.76	5.65	500	1.8
Al_2Cr #2	0	1.76	5.53	800	1.9
Cu #1	165 (ZrO_2)	0.48	5.42	500	2.3
Cu #2	173 (ZrO_2)	0.41	5.83	1000	1.3
Pt	0	1.72	4.05	1510	1.4
Al from handbook	—	1.396	7.650	infinite	—
Cu from handbook	—	0.24	3.450	infinite	—
Pt from handbook	—	2.330	4.150	infinite	—

the system was carried out to assure the accuracy and precision of the measurements. The dielectric tensor measurements for TbFe, TbFeCo, Co/Pt and Co/Pd thin films reveal valuable information about their optical and magneto-optical properties, and it helps their application performance improvement. This system also gives an accurate way to characterize the optical properties for non-magnetic multilayered thin films.

VI. REFERENCES

1. R. M. A. Azzam, and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, New York (1977).
2. D. Weller and W. Reim, H. Ebert, D. Jonson, and F. Pinski, "Correlation Between Bandstructure and Magneto-optical Properties of BCC Fe_xCo_{1-x}," Journal De Physique, Colloque C8, Supplement au n° 12, Tome 49, 41-42 (1988).
3. S.-C. Shin, and A. C. Palumbo, "Magneto-optical Properties of Co/Pd Superlattice Thin Films," J. Appl. Phys. 67(1), 317-320 (1990).
4. W. B. Zeper, F. J. A. M. Greidanus, P. F. Carcia, and C. R. Fincher, "Perpendicular Magnetic Anisotropy and Magneto-optical Kerr Effect of Vapor Deposited Co/Pt Layered Structures," J. Appl. Phys. 65(12), 4971-4975 (1989).
5. M. Mansuripur, F. L. Zhou, and J. K. Erwin, "Measuring the wavelength dependence of magnetooptical Kerr (or Faraday) rotation and ellipticity: a technique," Appl. Opt. 29, 1308-1311 (1990).
6. F. L. Zhou, J. K. Erwin and M. Mansuripur, "Spectral measurements of the magneto-optical Kerr rotation and ellipticity in the media of optical recording", J. Appl. Phys. 69(8), 5091-5093 (1991).
7. P. He, W. A. McGahan, J. A. Woollam, F. Sequeda, T. McDaniel, and H. Do, "Magneto-optical Kerr Effect and Perpendicular Magnetic Anisotropy of Evaporated and Sputtered Co/Pt Multilayer Structures," J. Appl. Phys. 69(7), 4021 (1991).
8. P. N. Argyres, "Theory of the Faraday and Kerr Effects in Ferromagnetics," Phys. Rev 97, 334 (1955).
9. G. A. N. Connell, "Measurement of the Magnetooptical Constants of Reactive Metals," Appl. Opt. 22, 3155-3159 (1983).
10. S. Hashimoto, Y. Ochiai, and K. Aso, "Film Thickness Dependence of Magneto-optical and Magnetic Properties in Co/Pt and Co/Pd Multilayers," J. Appl. Phys. 67(9), 4429-4431 (1990).
11. A. C. Palumbo and S.-C. Shin, "Thickness Dependence of Magneto-optical Effects in Ultrathin Multilayered Co/Pd Films," SPIE 1316, 363-366 (1990).
12. M. Ruane, M. Mansuripur, and R. Rosenvold, "Measurement of Reflectivities for Magnetooptical Media," Appl. Opt. 25, 1946-1951 (1986).
13. M. Mansuripur, M. Ruane, and R. Rosenvold, "Apparatus for Measuring Reflectivity", U.S. patent #4,838,695, June (1989).
14. M. Mansuripur, "Analysis of Multilayer Thin-film Structures Containing Magneto-optical and Anisotropic Media at Oblique Incidence Using 2×2 Matrices," J. Appl. Phys. 67, 6466-6475 (1990).
15. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 379-406, Pergamon, London (1958).
16. A. F. Zhou, J. K. Erwin, C. F. Brucker and M. Mansuripur, "Dielectric Tensor Characterization for Magneto-Optical Recording Media," accepted by Appl. Opt.
17. *Handbook of Chemistry and Physics*, CRC Press, Boca Raton, Florida (1989).

APPENDIX D

N93-15509

Measurement of the magnetic anisotropy energy constants for magneto-optical recording media

R.A. Hajjar, T.H. Wu, H. Fu and M. Mansuripur
Optical Sciences Center, University of Arizona, Tucson, AZ 85721

ABSTRACT

Measurement of the magneto-optical polar Kerr effect is performed on rare earth-transition metal (RE-TM) amorphous films using in-plane fields. From this measurement and the measurement of the saturation magnetization using a vibrating sample magnetometer (VSM), the magnetic anisotropy constants are determined. In this paper, we present the temperature dependence of the magnetic anisotropy in the range of $-175^{o}C$ to $175^{o}C$. The results show a dip in the anisotropy near magnetic compensation. This anomaly is explained based on the finite exchange coupling between the rare earth and transition metal subnetworks.

1. INTRODUCTION

Understanding the temperature dependence of magnetic anisotropy in the media of magneto-optical (MO) data storage is important for their applications. Measurements of the magneto-optical polar Kerr effect using in-plane external magnetic fields have been reported in several publications[1-3]. We have also reported recently, that the extraordinary Hall effect can be used to determine the magnetic anisotropy[4]. Due to the spin-orbit coupling, the perpendicular component of the magnetization of the transition metal subnetwork is proportional to the polar Kerr and Hall signals. Therefore, by proper normalization, the direction of magnetization θ with respect to the film normal can be obtained as function of the magnitude and direction α of the applied field H. Knowing θ, α, H and the saturation magnetization of the film M_s which is determined separately using a VSM, the uniaxial anisotropy energy $E_{anis}(\theta)$ can be calculated. We determine the anisotropy energy constants by matching $E_{anis}(\theta)$ with an expression of the form[5] $K_1 \sin^2(\theta) + K_2 \sin^4(\theta) + \ldots$. In this paper, we will present the temperature dependent anisotropy of TbFeCo films using the polar Kerr effect with perpendicular and in-plane applied magnetic fields. The films are of different terbium content ranging from 22.5% to 28%. The films were sputtered onto quartz substrates and consist of a quadrilayer structure of AlCr with MO layer sandwiched between two dielectric SiN layers. The magnetic anisotropy shows a dip near the magnetic compensation. We will discuss this anomaly based on the finite exchange coupling between rare earth (RE) and transition metal (TM) subnetworks.

2. TEMPERATURE DEPENDENT MAGNETIC PROPERTIES

Figure 1(a) displays the temperature dependence of M_s for the six films studied. Figure 1(b) displays the compensation temperature (T_{comp}) and the room temperature M_s versus terbium atomic percent ($T_{comp} \approx 25^{o}C$ when the terbium content is near 25%). Figure 2(a) displays the coercivity H_c as a function of temperature for these samples. H_c is obtained from the hysteresis loops measured using the polar Kerr effect. As expected, the coercivity diverges at T_{comp} which is a characteristic of ferrimagnetic materials. The room temperature H_c is shown in Fig. 2(b) as a function of the terbium content.

Figure 3(a) shows the magnetic anisotropy constant K_1 for these films as a function of temperature. The experimental set-up which uses the polar Kerr effect to determine K_1 and K_2 has been described in reference [2]. Since the maximum θ in the measurement was usually less than 15°, there was a large spread of about 50% in the fitted K_2 values. Figure 3(a) shows a dip in K_1 near the T_{comp} for the samples measured. This dip is clearly shown in the plot of Fig. 3(b) for K_1 and $K_1 + K_2$ at room temperature versus terbium atomic percent. The sample with 25% terbium could not be measured at room temperature since it requires infinite fields to tilt the magnetization. This point is represented by a question mark in Fig. 3(b). It is important to note that direct measurements of the anisotropy constants using a torque magnetometer show a similar rapid drop

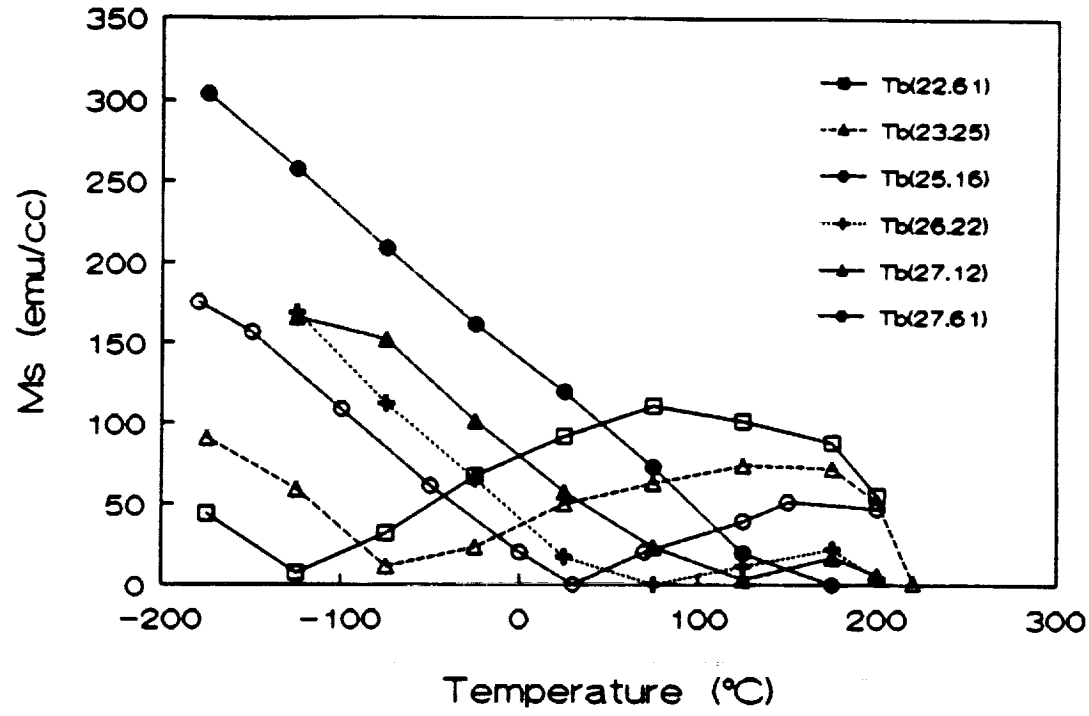


Fig. 1(a) Saturation magnetization M_s versus temperature for six TbFeCo samples.

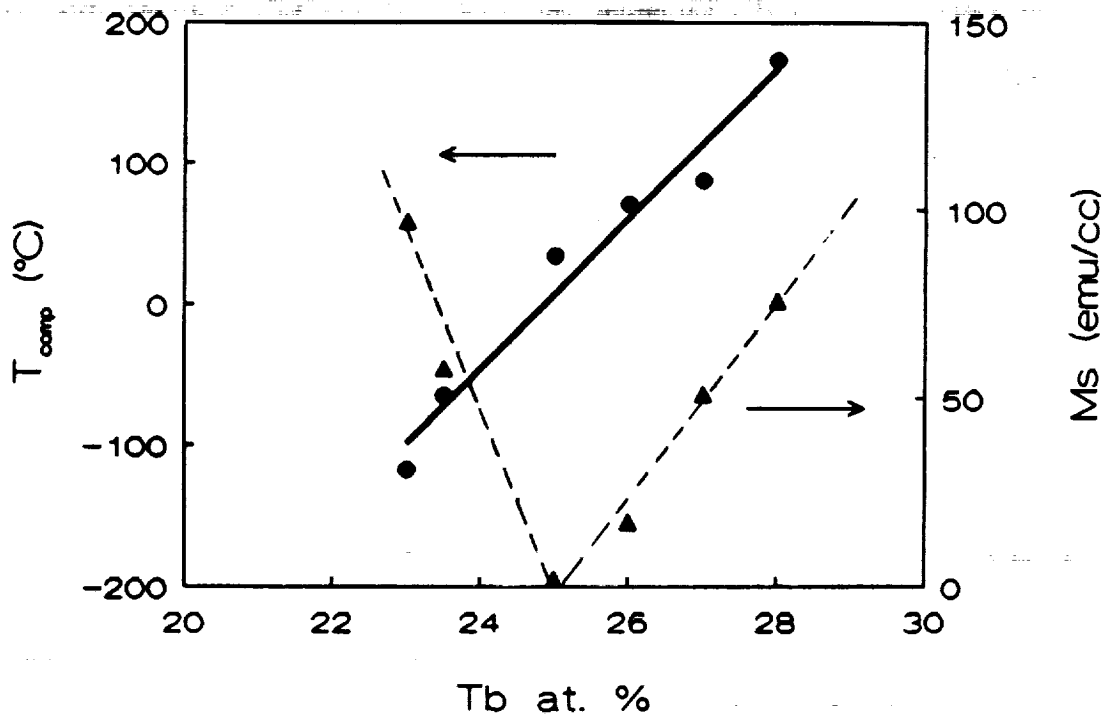


Fig. 1(b) Saturation magnetization M_s and compensation temperature T_{comp} at room temperature plotted as a function of terbium atomic percent.

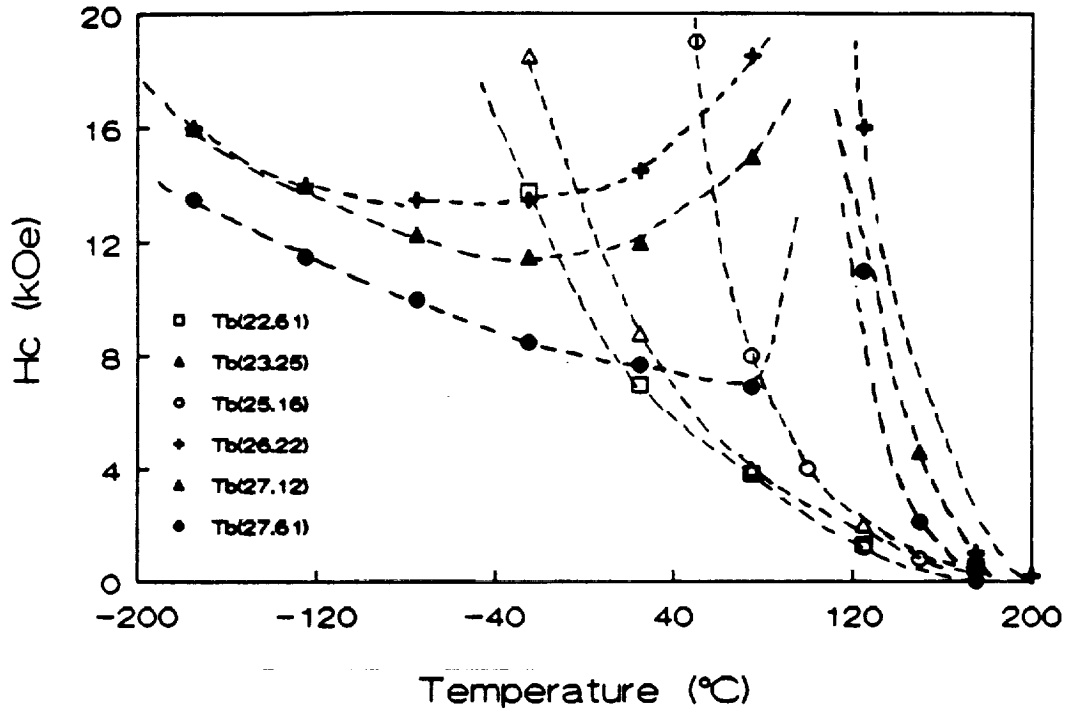


Fig. 2(a) Coercivity H_c versus temperature for six TbFeCo samples.

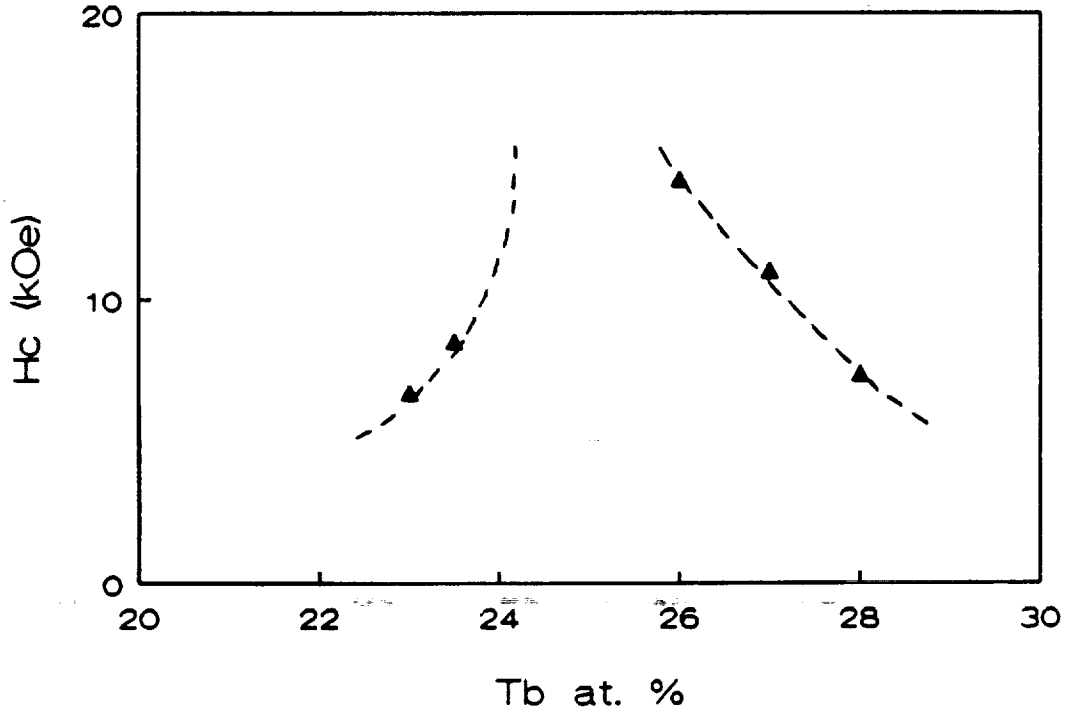


Fig. 2(b) Coercivity H_c at room temperature plotted as a function of terbium atomic percent.

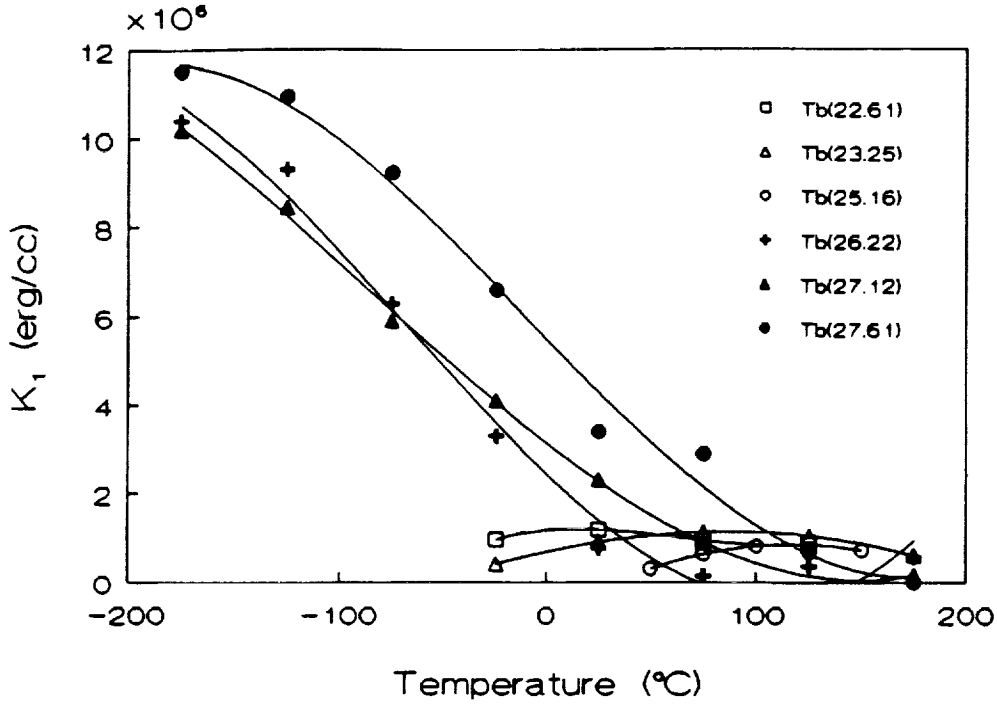


Fig. 3(a) Magnetic anisotropy constant K_1 versus temperature for six TbFeCo samples.

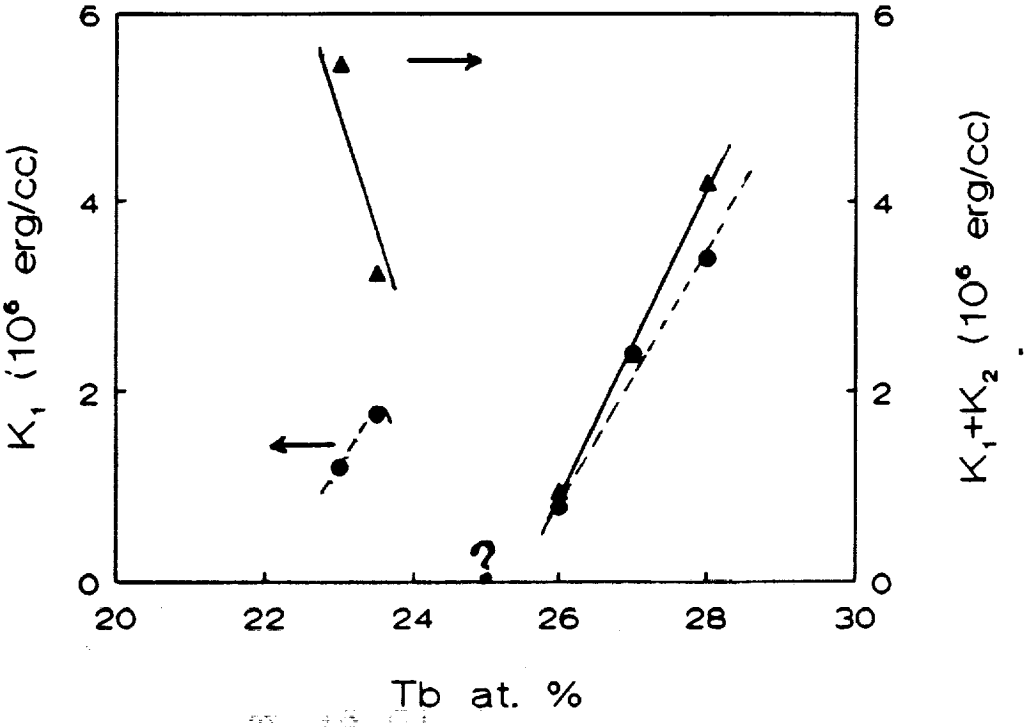


Fig. 3(b) Magnetic anisotropy constants K_1 and K_1+K_2 at room temperature plotted as a function of terbium atomic percent.

near T_{comp} as well. This anomaly is independent of the preparation or measurement technique and has been previously observed in similar samples[6-9]. The drop of anisotropy constant near compensation is quite unexpected, because compensation is simply a cancellation of the two subnetwork magnetizations and should have no effect on the anisotropy. Next, we discuss a mechanism that may explain this anomaly.

2.3 The Canting Model for the Anomaly

Sarkis and Callen[10] pointed out that the exchange coupling between RE and TM subnetworks in RE-TM ferrimagnets is not so large as to hold the RE and TM moments rigidly antiparallel. The external field used in measuring anisotropy can cause canting between the two sublattices. The amount of the canting is usually very small. For example, it is on the order of 1^o when a 20kOe field is applied to a typical TbFe thin film. However, neglecting this small canting by assuming infinite exchange between the two subnetworks in the total magnetic energy of the system and using this model to match the measured data, the K_1 found could be much smaller than the physical anistropy constant near compensation and would be zero at the exact compensation. Only in regimes far away from the compenstion, the fitted anisotropy approaches the physical anisotropy constant of the measured material. This problem has also been studied recently by Hellman[11].

In torque measurement, the reason why a small canting can cause a large discrepancy in the anisotropy constant can be easily understood in the following example. Consider the direction of the net magnetization M_s of the two sublattices of a RE-TM thin film under the influence of an in-plane applied field (Fig. 4). Usually, the tilted angle θ_N of the net magnetization from the normal direction is directly related to the anisotropy constant: small θ_N implies large K_1 and large θ_N indicates small K_1. However, when there is a canting between the sublattice magnetizations near compensation, θ_N can be very large even if the two sublattice magnetization are slightly tilted from the normal direction, see Fig. 4(a). Obviously, this large θ_N is not a result of small K_1 (and K_2), but a result of canting near the compensation. Far away from the compensation, the net magnetization will not tilt too much away from the major sublattice magnetization if the canting is small, see Fig. 4(*b*).

However, in the case of Kerr effect measurement, we only monitor the TM subnetwork magnetization direction θ_{TM}. The effect of the canting on the measured anisotropy constant becomes complicated. In fact, the fitted anisotropy based on infinite exchange can be smaller or larger than the physical anisotropy constant of the material, depending on the applied field direction α, which is 90^o in our case. Without going into a detailed theoretical treatment[12], it can be shown that if we assume a physical anisotropy constant of 5×10^6 erg/cc which is independent of M_s and take into account a finite exchange interaction between the RE and TM subnetworks[11], we can reproduce the measured θ_{TM}. Now, if we use the θ_{TM} found by the canting model and calculate K_1 with the assumption that the RE and TM moments are infinitely coupled to each other, we obtain the plot of Fig.5 showing a dip in K_1 (for the case of $M_{TM} > M_{RE}$). This finding suggests that the dip in the fitted anisotropy constants might be caused by an incomplete model that neglects the finite exchange between the sublattices. We will present in [12] a series of different measurements of K_1 and K_2 performed on a set of TbFeCo samples and discuss their differences based on the canting model.

3. ACKNOWLEDGMENTS

We wish to thank Dr. T. Suzuki of IBM Almaden Research Center for interesting discussions on this subject. We are also greatul to Dr. F. Sequeda and H. Notaris of IBM Almaden Reserch Center for providing the samples for this study. This work has been supported by the Optical Data Storage Center at the University of Arizona. One of us (R.A.H.) would like to acknowledge the support of a Komag Inc. fellowship.

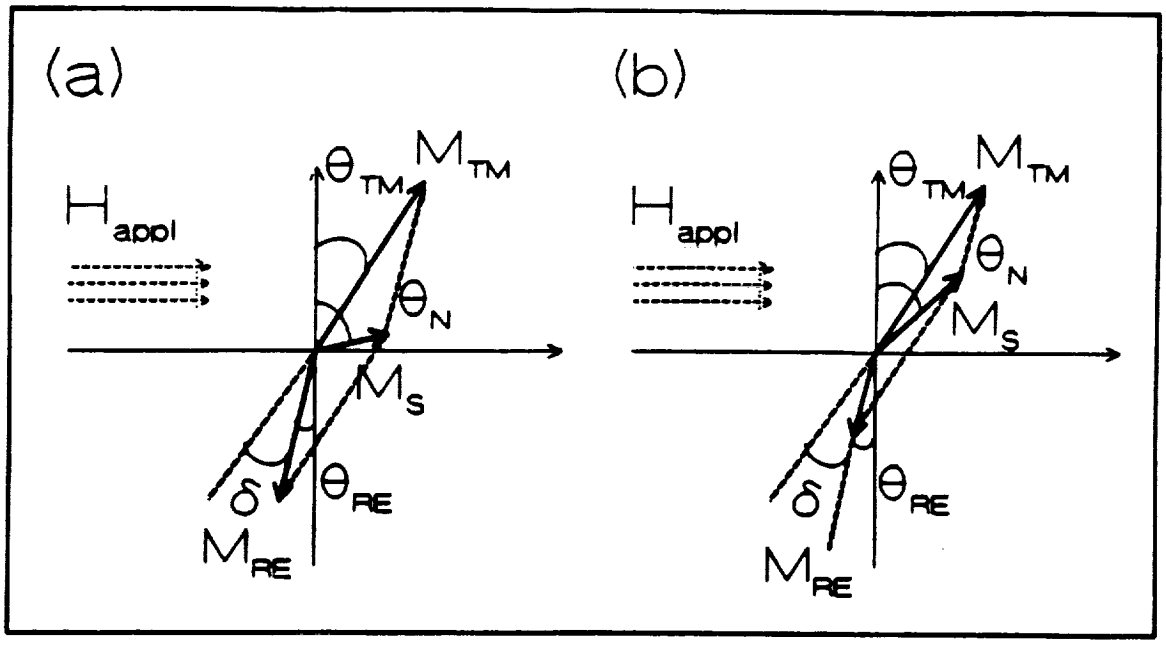


Fig. 4 The sublattice magnetization directions θ_{TM} and θ_{RE} and the net magnetization direction θ_N near (a) and far (b) from compensation. δ is the canting between the two sublattice magnetizations. In the case of (a) and when the canting is not zero, θ_N can be very large even when θ_{TM} and θ_{RE} are small. In the case of (b), the canting does not make θ_N and θ_{TM} too much different.

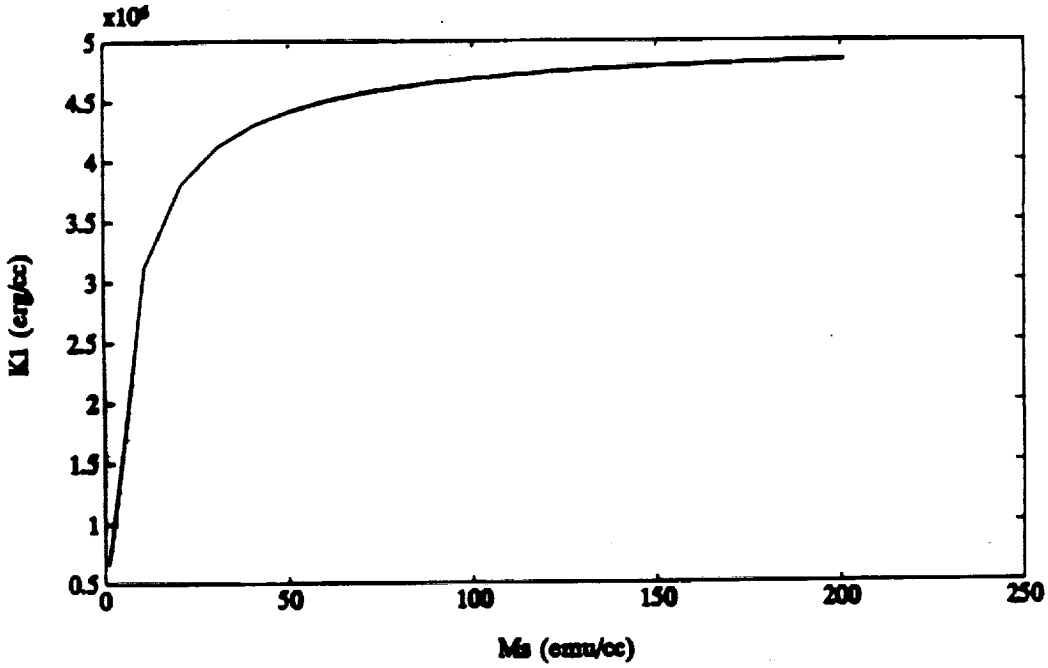


Fig. 5 Calculation of K_1 versus saturation magnetization M_s for a sample with a physical anisotropy constant of 5 x 10^6 erg/cc. The dip observed near magnetic compensation is a consequence of neglecting the finite exchange interaction between the two subnetworks.

4. REFERENCES

1.P.Wolniansky, S. Chase, R. Rosenvold, M.Ruane, and M. Mansuripur, "Magneto-Optical Measurements of Hysteresis Loop and Anisotropy Energy Constants on Amorphous Tb_x Fe_{1-x} Alloys", J. Appl. Phys. 60, 346 (1986).

2.R.A. Hajjar, F.L. Zhou and M. Mansuripur, " Magneto-Optical Measurement of Anisotropy Energy Constants on Amorphous Rare-Earth Transition-Metal Alloys", J. Appl. Phys.,67, 5328 (1990).

3.S.T. Purcell, M.T. Johnson, N.W.E. McGee, W.B. Zeper, and W. Hoving, "Spatially Resolved Magneto-Optical Investigation of the Perpendicular Anisotropy in a Wedge-Shaped Ultrathin Epitaxial *Co* Layer on *Pd*(111)", Submitted to J. Mag. Mag. Mat.

4.R.A. Hajjar, T. Wu, and M. Mansuripur, "Measurement of Anisotropy Energy for Magneto-Optical Media", Submitted to J. Appl. Phys.

5.C. Chikazumi,Physics of Magnetism edited by S. H. Charap (John Wiley & Sons, Inc. 1964).

6.N. Sato "Magnetic Properties of Amorphous *Tb-Fe* Thin Films with an Artificially Layered Structure", J. Appl. Phys. 59,2514 (1986).

7.H. Takagi, S. Tsunashima, S. Uchiyama, and T. Fujii, "Stress Induced Anisotropy in Amorphous *Gd-Fe* and *Tb-Fe* Sputtered Films", J. Appl. Phys. 50, 1642 (1979).

8.S. Yumoto, Y. Hidaka, and O. Okada, "Effects of Magnetic and Structural Properties on Magnetoresistance in Amorphous TbFeCo", J. Appl. Phys., 67, 5730 (1990).

9.F. Hellman, R.B. van Dover, S. Nakahara, and E.M. Gyrorgy, "Magnetic and Structural Investigation of the Composition Dependence of the Local Order in Amorphous *Tb-Fe*", Phys. Rev. *B*, 39, 591 (1989).

10.A. Sarkis and E. Callen, "Magnetic Anisotropy of Rare-Earth-Transition-Metal Compounds", Phys Rev. *B*, 26, 3870 (1982).

11.F. Hellman, "Measurement of Magnetic Anisotropy of Ferrimagnets Near Compensation", to appear in Appl. Phys. Let.

12.T. Wu, R.A. Hajjar, H. Fu, T. Suzuki and M. Mansuripur, "Measurements of the Magnetic Anisotropy Energy Constants of Magneto-Optical Recording Media Using Several Techniques", in preparation.

APPENDIX E

N93-15510

Proposal for Massively Parallel Data Storage System

M. Mansuripur, Optical Sciences Center, University of Arizona, Tucson, AZ 85721

ABSTRACT

An architecture for integrating large numbers of data storage units (drives) to form a distributed mass storage system is proposed. The network of interconnected units consists of nodes and links. At each node there resides a controller board, a data storage unit and, possibly, a local/remote user-terminal. The links (twisted-pair wires, coax cables, or fiber-optic channels) provide the communications backbone of the network. There is no central controller for the system as a whole; all decisions regarding allocation of resources, routing of messages and data-blocks, creation and distribution of redundant data-blocks throughout the system (for protection against possible failures), frequency of backup operations, etc., are made locally at individual nodes. The system can handle as many user-terminals as there are nodes in the network. Various users compete for resources by sending their requests to the local controller-board and receiving allocations of time and storage space. In principle, each user can have access to the entire system, and all drives can be running in parallel to service the requests of one or more users. The system is expandable up to a maximum number of nodes, determined by the number of routing-buffers built into the controller boards. Additional drives, controller-boards, user-terminals, and links can be simply plugged into an existing system in order to expand its capacity.

1. Background : The proliferation of computer networks in the near future is likely to create a tremendous demand for large data storage systems capable of handling massive amounts of information. In the foreseeable future, file-servers having capacities in the terabyte (10^{12}) range will serve a host of users whose demand for fast access and rapid data transfer rates can be satisfied only with sustained data rates of several hundred megabytes per second. At the present time, the use of high-capacity/high-data-rate storage devices may be confined to large banks and insurance companies, airline reservation systems, military and aerospace organizations, hospitals and certain medical facilities, major research and development centers, and so forth. Nonetheless, these applications represent a growing market for high-end data storage products, with performance requirements that in many instances outpace advances in the electronic storage technology.

Presently, three storage technologies dominate the mass-storage market. These are the magnetic tape, magnetic hard disk, and optical disk technologies. Tape storage is used mainly for backup and archival applications. Hard disks are predominant in high-density, fast access, high data-rate environments. Optical disks are now beginning to emerge as serious contenders for the same application area as has traditionally been assigned to magnetic disks; they also show great promise for large volume, archival applications. Roughly speaking, there is room for 1 GByte of storage on a 5.25" platter, and the achievable transfer rate to and from a disk drive (assuming reasonable disk rotation speeds), is in the range of 10 Mbit/sec (single channel). These numbers may be over-estimates for magnetic disks and under-estimates for optical disks; they are intended only as rough estimates here.

It is clear that in order to achieve high capacity and/or high data rate, small units of magnetic/optical storage must somehow be integrated. One approach to integration calls for the incorporation of multiple disks/heads within a single drive. This is certainly possible and, in fact, it is a path that has been vigorously pursued for many years. One disadvantage of this approach is the high cost and complexity of the resulting system. Also, since high performance

drives of this type concentrate large quantities of data in one location, their failure can be catastrophic. Another possible approach to integration involves networks of interconnected storage units. Here several drives are placed on a single bus (or on multiple buses), and controlled by an intelligent device driver. IBM Corporation's 3990 storage subsystem with a maximum capacity of 181 GByte (and a price tag of approximately $2,500,000) exemplifies the latter approach. Several other vendors offer integrated drives for the high-end storage market as well. Noteworthy among available products are Digital Equipment Corporation's DECarray SA900 with up to 40 GByte of storage and 2000 I/O sec, and the system DS 323 from Recognition Concepts, Inc. with 91 GByte of capacity at 72 MByte/sec data rate. These systems are typically configured around a powerful central controller, use only a few disk drives, have limited expandability, are very expensive, and are highly specialized and directed towards specific classes of applications.

We propose a massively parallel data storage system based on a network of distributed drives and controllers. The system that we envision will have the following distinct features and capabilities:

- o Constructed from small, inexpensive drives by large-scale system integration.
- o Storage capacity ranging from a few GByte to several TByte.
- o Data rate from a few MByte/sec up to several ten MBytes/sec.
- o Number of independent users from one to several hundred.
- o Expandable by adding plug-in modules.
- o Reconfigurable by changing the network of interconnections.
- o Immune to drive failures by adding redundancy to data and distributing blocks of data among several drives.
- o Designated drives for backup; automatic backup operations.
- o Automatic disk-space management and defragmentation throughout the system.
- o Open architecture; standard interface with the user.
- o Inexpensive.

In the following sections we shall describe our approach to drive integration, discuss certain important features of the distributed controller, and present preliminary results from a computer simulation program that has been used to evaluate the system's performance.

2. Introduction : The present document describes a system of interconnected data storage units (such as magnetic or optical disk/tape drives, or a mixture of them). The system allows massive amounts of data to be stored in a distributed network of storage units, with apparent data rates and access times which are substantial improvements over those available from the individual drives. Multiple users can access the system independently of each other, and will use the system by competing for resources. The architecture of the system has the geometry of a hypercube or that of an extended hypercube, depending on the required data rates which would be determined by the number of interconnections among the units. The backbone of the system is its communication network, with individual connections being made either by ordinary wires, or coaxial cables, or fiber-optic links. The control of the network is distributed, with one controller board at each node of the (extended) hypercube. Each controller manages the communication with the local user-terminal (if one is attached) and the local drive. The controller also handles the arriving message/data blocks from adjacent nodes, and determines the proper route for sending them towards their destination.

All controller boards within a network are identical in their basic design. A general-purpose board can be designed to handle a number of different protocols for communication with various types of storage devices and user terminals. Alternatively, separate boards may be built for handling specific protocols and device-interfaces. In any event, the differences among the boards for different nodes are confined to their interfaces with the local storage unit and/or the user-terminal; the routing mechanism remains the same for all boards. The boards are configured for a maximum number of adjacent-node connections. For instance, if 10 is the

maximum number of connections built into the boards, then hypercube networks designed around these boards can have as few as 2 and as many as $2^{10} = 1024$ nodes. The system is thus expandable, and can grow as the storage needs of the user increases. Details of the network's architecture and the controller board design are described in the following section.

Security against failures of individual units (head crashes, chip burnouts, etc.) may be provided by the addition of redundant blocks to the user data, and the distribution of these blocks among several drives. For instance, if five blocks of user-data are added (modulo-2) together to create a sixth block, and if these six blocks are stored on six differernt units, then failure of a single unit will not affect the integrity of the data, since the missing block can always be recovered by modulo-2 addition of the remaining five blocks. This is not a new idea; in fact the storage system architecture known as Redundant Array of Independent Drives (RAID) utilizes the same strategy for protection against drive failures. What is new here is that no specific drive is designated for storing the redundant block, and the allocation of resources involves a dynamic decision making process by the system itself. When a failure occurs, the faulty device can be replaced without halting the entire system. Subsequently the lost data is automatically reconstructed and stored on the new drive.

Another feature of the massively parallel data storage system is its ability to perform backups in the background, without requiring intervention from the users or from the system manager. Some of the nodes are designated as backup nodes, and the system is programmed to perform backups periodically, by storing the contents of all other drives at these designated nodes. Intelligence can be built into the system to perform backups with low priority, or to consider postponing such operations at times when the overall load/traffic within the system is heavy.

3. Definition of terms and description of the building blocks of the system : This section defines some of the technical terms used in the present document, and describes the building blocks of the massively parallel data storage system in detail.

Block of Data : A block of data is the minimum-size package of information that a storage system can handle. For example, a block may consist of 512 bytes of user data plus overhead (i.e., error control bits, synchronization bits, etc.). The users of the storage system submit their data in blocks, the system stores a given block contiguously on a sector somewhere within the system, and the block is returned in its entirety upon request. Error correction coding is used to ensure the integrity of the block in the presence of noise. Modulation coding may be used to tailor the block for storage on a specific device. Identification and synchronization bits may be added to the block for later identification and retrieval. All these additional bits constitute the overhead on the data. A block of user-data and its associated overhead are usually treated as a single unit of data, and recorded on a single sector. Within the storage system, each block is identified by a unique address. This address consists of two parts: The ID of the drive on which the block is stored, and the ID of the sector within that drive allocated to the block. For example, a system that stores up to 1 terabyte of data may have 1000 separate storage units (drives), each one of which handles 1 gigabyte of data. Assuming that each block is equal to one kilobyte, each drive can store up to 10^6 blocks; each sector, therefore, is identified by a 20 bit address ($2^{20} = 1,048,576$). Since each drive needs at least 10 bits for its unique identification ($2^{10} = 1024$), we see that a given block anywhere within this system is uniquely identified by a 30-bit address. A user submitting a block to the system for storage must receive this identifying address in return. All that the user needs to know about a block it submits is this address, until such time as the block is needed for processing. At that point the user will send the address to the system and request a retrieval. (Note: A data storage system dedicated to a single user can manage the task of address allocation at the user level, since the user always knows which sectors are available on various drives, where his stored data is located within the system, and so forth. On the other hand, a multi-user system that wants to keep the users independent of each other, must control the address allocation at a lower level. For example, the user in such a system must first send a request for allocation of an empty

sector. The system then sends a message, informing the user of an available address. Subsequently, the user sends its block of data for storage in that particular address, and keeps the address for future reference.)

Storage Unit (Drive) : This is a device capable of communicating via a well-defined protocol with its environment. The device must be able to accept blocks of data from a host and store that data internally. It must also be capable of delivering the stored blocks to the host upon request. The internal mechanism of storage is irrelevant as far as the outside world is concerned. The data might be stored on one or more magnetic disks, optical disks, magnetic/optical tapes, semiconductor memory chips, etc. The drive sends messages to its host, informing the host of its status. The busy message means that the drive is not available for new requests. This occurs, for instance, when the drive searches for a previously requested block of data in the read mode, or when it tries to record or erase a sector. The ready-to-transmit message means that the drive has found a previously requested block and is now ready to submit it to the host. There must be a table of contents for the drive that the host can access and modify. It is through this table that the host knows which sectors on the drive are available, what blocks are recorded on which sectors, and so forth.

Hypercube : This is a special geometry for connecting a number of devices. The simplest hypercube is a one-dimensional cube ($n = 1$), which has two nodes connected by a single edge (see Fig. 1(*a*)). The next hypercube has dimension $n = 2$, consists of $2^n = 4$ nodes, and each node is connected to two neighboring nodes via two edges, as shown in Fig. 1(*b*). The three dimensional cube is the ordinary cube with $n = 3$, has $2^n = 8$ nodes, and each node is connected to three other nodes via three edges. To construct an $n + 1$ dimensional cube, therefore, one must connect the corresponding nodes of two n-dimensional cubes via 2^n new edges. For example, Fig. 1(*d*) shows how a 4-dimensional cube may be constructed by connecting the eight nodes of two ordinary (i.e., 3*d*) cubes. In this way it is easy, for instance, to see that a 10-*d* cube consists of $2^{10} = 1024$ nodes, each node is connected directly to 10 neighboring nodes via 10 edges, and, in order to go from any node to any other node of the hypercube, one needs to go over at most 10 edges. In the architecture set forth in this document we shall connect data storage devices in the hypercube geometry (or an extension of the hypercube geometry to be described in the next paragraph). Thus individual drives are at the nodes of the hypercube, and the backbone of the communication network that connects these drives is formed by wires (i.e., twisted pairs, coax cables, fiber-optic links, etc.) that can be imagined as the edges of the hypercube. Of course we do not have access to a physical higher dimensional space than $n = 3$, so that the angles at the corners of our hypercubes are not 90^0 angles, but this is not a matter of concern, since all we are interested in is the number of connections between the nodes and the order in which these connections are made.

Extended Hypercube : In certain applications it might be desirable to increase the number of connections at each node, i.e., to have more adjacent nodes (as compared to the hypercube) for each node. In such cases we shall use simple extensions of the hypercube. In principle, an arbitrarily large number of connections per node can be achieved, provided that one is willing to use the required number of wires. In Fig. 2 we have depicted, by way of example, the geometry of an extended hypercube. Upon close examination of the strategy employed in Fig. 2 and comparison with Fig. 1, it should be trivial to create further extensions of the hypercube containing even more connections per node.

In Fig. 2(*a*) we have the smallest unit of an architecture with 4 nodes and 3 connections to each node. In Fig. 2(*b*) the basic unit is replicated 4 times, and each node is connected to 3 similar nodes. We now have a total of 16 nodes with each node connected directly to 6 other nodes. (Remember that the hypercube of dimension 4 which also has 16 nodes, has only 4 connections per node.) To go to the next step we can take the structure in Fig. 2(*b*), replicate it four times, and connect each one of the resulting nodes to their three counterparts. We will then have a total of 64 nodes, with each node directly connected to 9 other nodes (that is three connections more than afforded by a 6-*d* hypercube which has the same number of nodes). In

like manner, the procedure can be extended to structures with 4^n nodes and $3n$ connections at each node.

Link : Each edge of the hypercube is a "link" in the actual system. A link is *a* bi-directional communication channel between two adjacent nodes of the network. Links may be simple twisted-pair wires, coaxial cables, or fiber-optic channels. An *n*-dimensional hypercube will have a total of $n2^{n-1}$ links. A message/data block created at one node and addressed to another node, must travel through one or more links before it arrives at its destination. Typically, a message/data block arrives at a node, waits for a link to an appropriate adjacent node to become available, and then travels to that node; the process continues until the block arrives at its destination. Assuming that traffic jams do not force the controller to re-route, each message/data block traverses at most *n* links in an *n*-dimensional hypercube. For instance, in a 10-*d* hypercube architecture which has 1024 nodes, travelling blocks need at most 10 hops to reach their destination. Most blocks, however, travel a shorter distance than the above maximum, since they do not travel between extreme opposite corners of the cube. Since an extended hypercube will have more links than an ordinary hypercube with the same number of nodes, communication within the extended cube is faster and more efficient.

Buffer : There are several buffers on each controller board. A buffer is an electronic storage device, such as a shift register, used as temporary storage for commands and data blocks between origination and destination points. A buffer can store one block of user data plus additional information such as a command, one or more command qualifiers, flags, and command-related addresses. Figure 3 shows a typical structure of one such buffer. Within a board, buffers can exchange their contents with one another, transfer the contents to the local storage device, send/receive command and data from the local user-terminal, and transfer contents to another controller board (located at an adjacent node) via the links. The buffers do not make these decisions themselves; the control logic reads the command section of each buffer and initiates the necessary tranfer(s). The buffers may have multiple layers of depth, as shown in Fig. 4. When traffic congestions occur on the board, or when the buffer needs to wait for unloading its contents (and this may happen for a variety of reasons), the control logic pushes the contents one level down and clears the way for the next transaction. All previous commands in this case wait "below the surface" until congestion problems are resolved.

Controller Board : At each node of the (extended) hypercube there resides a controller board. A block diagram illustrating a possible configuration of this board is shown in Fig. 5. A storage unit (drive) communicates with the controller through the drive buffer (*d*-buffer for short). Similarly, the *I/O*-buffer handles traffic in and out of the user-terminal. The *d*-buffer is used to store command/data on the way to the storage device, or on the way out of the storage device and to a specific node. It may happen that a given node is not selected as a user-terminal node, in which case the flags within the *I/O*-buffer are properly set to indicate this fact to the control logic; the control logic then ignores the *I/O* buffer.

Routing buffers (*r*-buffers) receive command/data items from adjacent nodes via the links, or directly from other buffers within the same node. The number of *r*-buffers on each board is equal to the number of adjacent nodes for the particular architecture under consideration. For example, an *n*-cube architecture requires *n* routing buffers per board. When the contents of an *r*-buffer are addressed to the current node, the controller extracts them and places them either in the *d*-buffer or in the *I/O*-buffer. When the contents of an *r*-buffer are addressed to another node, the controller searches for the best route to direct the item, then transfers it to the proper adjacent node. The routing decisions are made locally at each node, based on the destination address, availability of adjacent-node buffers, and whatever local information may be available to each controller. The cross-bar switch allows any buffer to be connected to any link, connects *r*-buffers to *d*-buffer and *I/O*-buffer, and allows *d*-buffer and *I/O*-buffer to exchange their contents. The state of the cross-bar switch is determined by the control logic which is the CPU of the controller board. The control logic reads the command segment of each buffer, identifies the best route for the contents of that buffer, and programs the cross-

bar switch for the necessary connections.

The diagram in Fig. 5 also shows a connection between the control logic and a unit called the status table. This status table keeps useful information about the state of other nodes within the system. For instance, a drive upon becoming engaged/released by a user-terminal can send a message to all nodes, informing them of its new status. Each node then updates the status of that particular drive and keep the information in the status table, thus allowing the control logic to draw on this information for making allocation decisions.

4. Advantages : The massively parallel data storage system described in this document has several advantages over the existing mass storage devices. Some of the advantages of this system are briefly described in the following paragraphs.

i) The system is built from small units (drives) that are produced in large quantity for the personal computer market. These units are cheap, reliable, and are available from a number of different manufacturers. It is therefore possible to reduce the overall cost of the system and maintain the flexibility to respond to new technological developments.

ii) The system can be expanded by adding new storage units/links to the old ones. Thus when one's need for storage capacity/transfer rate increases, one can upgrade an existing system and pay only the cost of additional units. In the case of hypercube architecture, for example, if the controller boards were designed to handle 10 adjacent nodes, then the upper limit to the number of units would be 1024, if the boards were designed for 11 adjacent nodes the upper limit would be 2048, and so on. One might also choose to reconfigure a system without adding new drives, but by changing the architecture to an extended hypercube which has more links per node. For example, a 4-*d* hypercube with 16 drives and 4 links per node (total number of links = 32) can be reconfigured to the architecture shown in Fig. 2(*b*) which has the same number of drives but, with 6 links per node, requires 16 additional links. The new configuration will now have an increased data rate, since its nodes have more connections to the rest of the system.

iii) The massively parallel data storage system can handle as many independent user-terminals as there are nodes in the system. Each user has access to the entire system, and interacts with the system as though there were no other users. Of course the presence of many users on the network will affect the response-time and the traffic load, but it does not modify the logical modes of interaction between the user and the system.

iv) By adding redundancy to the data and distributing the blocks of data among various drives, the performance of the massively parallel data storage system becomes immune to failures. Thus when a storage unit fails, the system proceeds to reconstruct the lost data from the remaining blocks. In the meantime, the failed device may be replaced without having to shut down the entire system. Once the new drive is up and running, the system automatically replaces the lost data.

v) The system performs automatic backups during periods in which the network's load is light. One or more nodes may be assigned to backup drives (such as tape drives), and the system instructed to transfer the contents of other units to the backup units during appropriate time intervals. No action on the part of the user(s) is therefore required for safe-keeping of the data.

5. Computer Simulation : *A* highly simplified version of the hypercube-configured, massively parallel data storage system was simulated (in FORTRAN) on a VAX-Station. The purpose of the simulation was to demonstrate the feasibility of the concept, and also to investigate the performance of various routing and communication algorithms. Since the

simulated system incorporates some of the essential features of the proposed system, this section is devoted to a brief description of certain features of the simulation. The following terminology has been used in developing the simulation program.

ID: Number of dimensions of the hypercube ($ID_{max} = 16$)
NODES: Number of nodes in the network ($NODES = 2^{ID}$)
CNODE: Current node address (16-bit format)
DNODE: Destination node address (16-bit format)
GNODE: Origination node address (16-bit format)
SFLG: Status-flag, a logical array with as many elements as there are nodes. Each node keeps a copy of *SFLG* which is continually updated throughout the system to reflect the most recent state of the drives. $SFLG(n)$ = true/false depending on whether or not the drive at node n is being used by some other node.

Input/Output Buffer : Within each I/O-buffer two flags and six segments are utilized for command, DNODE/GNODE address, begin/end addresses for data-blocks, and the data-block itself.
IOFLG1: When this flag is false, buffer needs attention from the user-terminal.
IOFLG2: When this flag is false, buffer needs attention from the controller
IOFLG3: (not used)
IOBFR1: DNODE or GNODE address, depending on the type of command
IOBFR2: (not used)
IOBFR3: (not used)
IOBFR4: Command
IOBFR5: Command-parameter
IOBFR6: Begin address of the block
IOBFR7: End address of the block
IOBFR8: One data-block sent either from user-terminal for storage, or from a drive for pickup by the terminal

Drive Buffer : Within each d-buffer three flags and five segments are utilized for command, DNODE/GNODE address, begin/end addresses for data-blocks, and the data-block itself.
DFLG1: When this flag is false, buffer needs attention from the drive.
DFLG2: When this flag is false, buffer needs attention from the controller.
DFLG3: This flag becomes false upon reservation; it becomes true again when released by the node that made the reservation. When reserved, only commands arriving in *r*-buffers with the correct DNODE address are accepted. Whenever the status of DFLG3 changes, the corresponding SFLAG is reset throughout the system.
DBFR1: DNODE or GNODE address, depending on the type of command
DBFR2: (not used)
DBFR3: (not used)
DBFR4: Command
DBFR5: (not used)
DBFR6: Begin address of data-block
DBFR7: End address of data-block
DBFR8: Data-block

Routing Buffers : Each node has a number of r-buffers equal to ID, one for each link attached to the node. Each buffer can accept one command/data-item from its input link, or from a local *r*-buffer, *d*-buffer, *I/O*-buffer, or *x*-buffer. The contents of a buffer may be transferred to another node via the corresponding link and *r*-buffer, or they may be shipped to the local *d*-buffer or to the local *I/O*-buffer. The *r*-buffers utilize one flag and eight segments, as follows.
RFLG1: When this flag is true, *r*-buffer is empty in which case it can receive a command/data-item. When false, the buffer must be attended to by the control logic.

RFLG2: (not used)
RFLG3: (not used)
RBFR1: DNODE address for the command (16-bit binary format)
RBFR2: GNODE address for the command (16-bit binary format)
RBFR3: Counter. If 0, command does not propagate, otherwise, it is shipped to the adjacent *r*-buffer, with the corresponding bit of its DNODE address toggled and its counter decremented by 1.
RBFR4: Command
RBFR5: Command-parameter
RBFR6: Begin address of data-block
RBFR7: End-address of data-block
RBFR8: Data-block

Auxiliary Buffers : Each buffer in the controller-board has its own x-buffer corresponding to depth level #1 (see Fig. 4). When traffic congestions make it impossible for a buffer to ship its contents to the proper address, the contents are automatically pushed to the *x*-buffer immediately below for temporary storage. The *x*-buffer pops back up whenever its parent buffer becomes available again.
XFLG1: When this flag is false, *x*-buffer is occupied in which case it needs attention from the control logic.
XFLG2: (not used)
XFLG3: (not used)

Cache : The cache memory residing in each controller-board receives data-blocks from local user-terminal and stores them at the address indicated by BPNTR. Blocks move out sequentially from the address indicated by TPNTR. Cache is empty when BPNTR = TPNTR.
NMAX: Maximum number of blocks cache can store.
TPNTR: Pointer to the top of the occupied block within the cache.
BPNTR: Pointer to the bottom of the occupied block within the cache.
CFLG: When this flag is false, cache is full.

Internal system commands : The following commands were used in the computer simulation. This is a minimum command set required for basic routing and data transfer procedures.
F. Set at each and every node the status-flag of the drive at CNODE. SFLG(n) indicates whether or not the drive at node *n* accepts read/write/erase requests (when accepting, SFLG(n) = True).
H. Hold drive at DNODE on behalf of GNODE.
P. Reporting the status of drive in response to *a* H-request. Informs GNODE whether or not *d*-buffer at DNODE is available. If available, it also delivers an assigned address for the data-block.
L. Release drive at DNODE from an earlier hold made by GNODE.
S. Deliver data-block for storage at DNODE.
E. Erase data-blocks (address = begin : end) from drive at DNODE.
R. Retrieve data-blocks (address = begin : end) from drive at DNODE.
O. Transfer data-block to *I/O*-buffer at DNODE.

Results of simulations : Simulations were performed for hypercube networks of disk drives with varying dimensions. In each case, the number of users was varied in order to study the effect of overall load on the level of interaction between the system and individual users. Users were given equal priority, and were allowed to issue random requests for READ and WRITE operations. The WRITE commands were preceded by requests for drive allocation, without specifying the drive. The controller then tried to assign the local drive to the task, except when that drive was busy, in which case it selected another (free) drive. Subsequently the data-block was shipped to the reserved drive for storage. In case of a READ command, the user-terminal first made a request for reservation from the drive that contained the desired

block of data. If the request was granted, the drive would be asked to read a 5-sector block and forward it to the terminal that had issued the request. When a request for reservation was denied, the user-terminal simply abandoned the request. The simulation program was incapable of admitting concurrency of operation among the various nodes, leaving us with no option but to scan the nodes sequentially. The results obtained by simulation, therefore, are worst-case scenarios and do not reflect the true level of parallelism inherent in the system. Despite this and several other shortcomings of the simulation program, the results are impressive and indicate that significant gains can be expected from a massively parallel data storage system.

The unit of time T in these simulations is taken as the time needed to transfer the contents of one buffer over a single link (i.e., transfer between adjacent nodes). Each disk operation (sector-read or sector-write) is assumed to take 20T. Table I lists the total number of completed read/write operations in several simulations that were performed over a time interval of 500T. Notice that a single drive, working incessantly to perform sector read/write jobs, would complete only 25 such operations during the same period of time. There was no cache memory on the simulated boards, and the buffers had only one level of depth. When traffic jams occurred, the program was aborted and the simulation repeated with a reduced frequency of read/write requests on behalf of each user.

The simulation results in Table I indicate a significant level of drive activity within the system. For instance, the hypercube storage device with 16 drives and 4 users performs 371 read/write operations during the period under consideration, i.e., the equivalent of 15 full-time drives. Or consider the system with 64 nodes and 12 users, which delivers the performance of 30 full-time drives. Even ignoring the fact that the simulation under-estimates the power of the system, and also the fact that the users in these simulations make frequent requests for fairly short operations, the performance figures are still very impressive.

Table I. Simulated performance results for hypercube mass-storage systems. When more than one user is present, the users are given equal priority, and allowed to make random requests for READ and WRITE operations from the system. No attempt has been made to optimize the performance or to break traffic jams. The results shown are simply examples of sustained operation over a long time interval. (For comparison, note that a single drive running incessantly during the same period would perform 25 such operations.)

Nodes	Users	Requests issued/granted	Operations read/write
16	1	144/099	255/48
16	4	217/115	320/51
64	1	149/116	270/62
64	7	255/213	505/112
64	12	290/236	630/110
128	1	118/110	250/60
128	10	374/272	730/126

Figure Captions

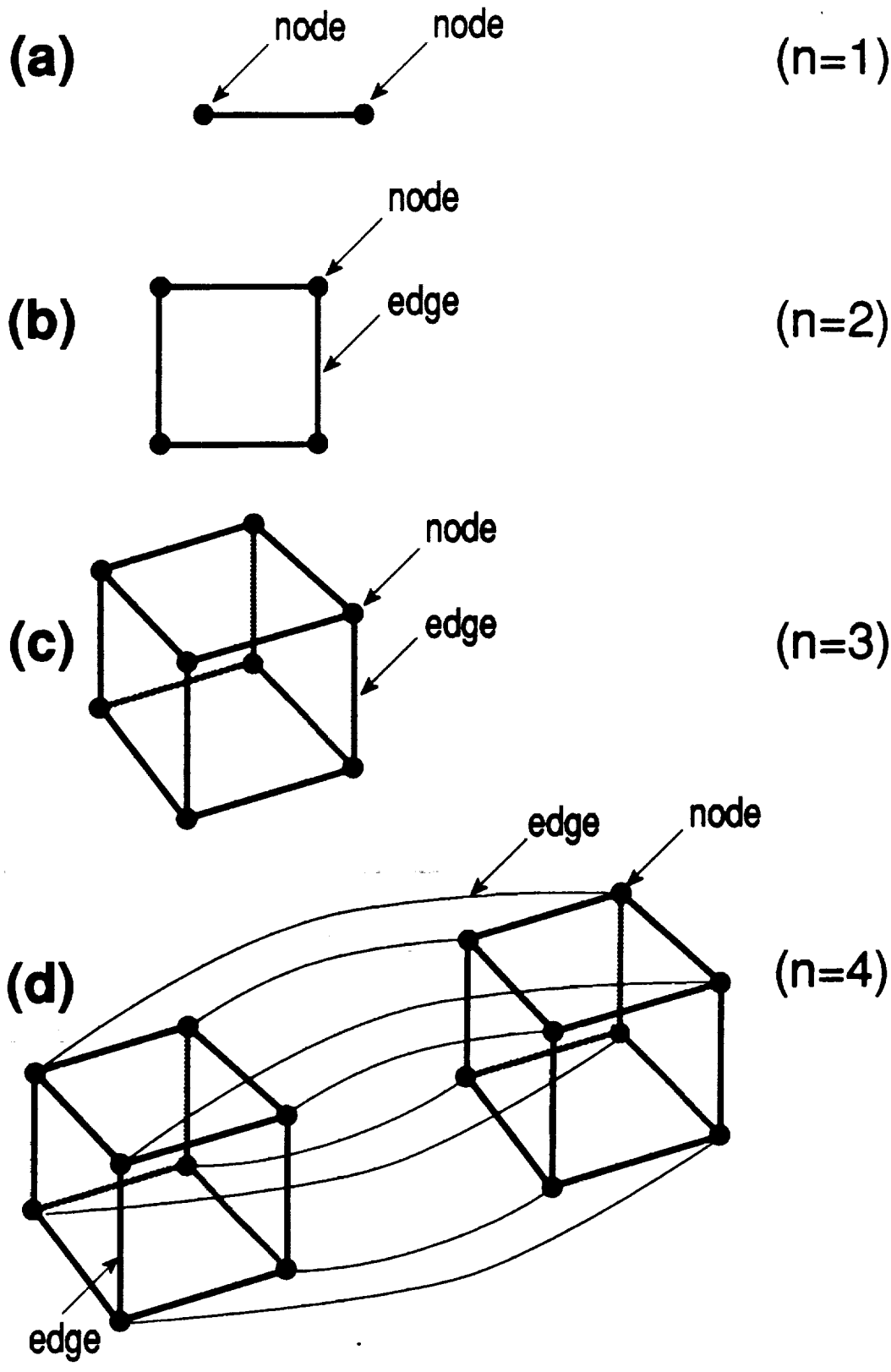
Fig. 1 shows n-dimensional cubes with n = 1, 2, 3 and 4. A cube of dimension $n > 3$ is often referred to as a hypercube. An n-dimensional cube will have 2^n nodes and $n2^{n-1}$ edges.

Fig. 2 shows the principle of construction of an extended hypercube network. The lowest-dimensional structure of this type shown in (a) has 4 nodes, where each node is directly connected to 3 other nodes. The next higher-dimensional structure shown in (b) consists of 16 nodes, where each node is directly connected to 6 other nodes.

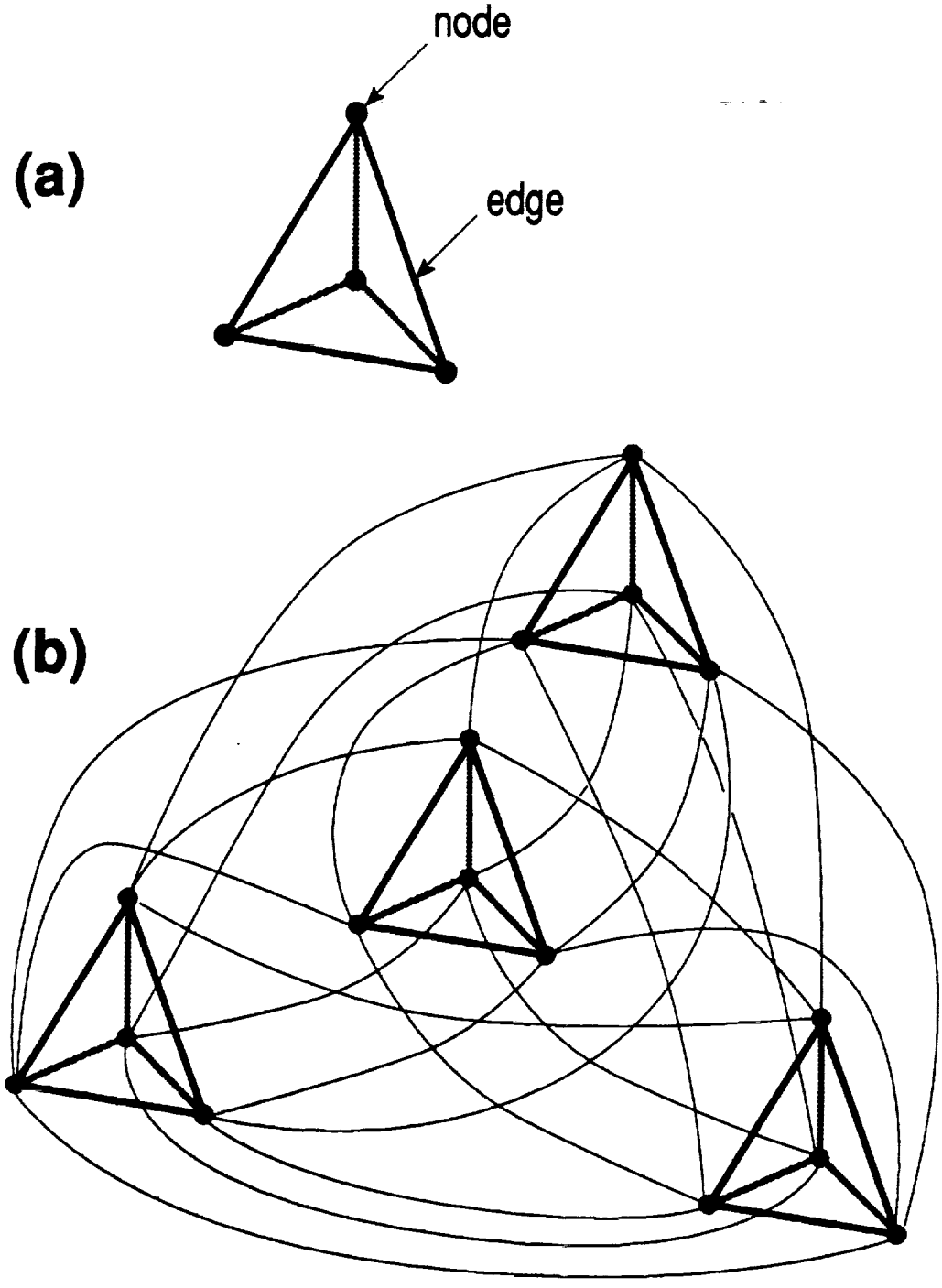
Fig. 3 is the block diagram of a typical buffer register. Several such registers are used as temporary storage units within each controller board. A buffer register is essentially a shift register which is loaded and unloaded electronically, either serially, or in parallel, or in a mixed mode. The control logic reads the "command & control" section of the register to decide what task(s) to perform and/or which route to send the block along. The "data" section of the register contains one block of user-data (typically 512 or 1024 bytes) which is on its way to be stored on a drive or to be retrieved by a user-terminal.

Fig. 4 shows a buffer register with multiple depth layers. At times of congestion, contents of the entire register are pushed one level down in order to free the main register (top) for continued operation. When the congestion condition disappears, contents of the register are pushed up (layer by layer) until all levels are cleared up.

Fig. 5 shows a block diagram of the controller board whose responsibility is the control of operations at each node of the system. To each node is attached a storage unit and, possibly, a user-terminal, both communicating with their respective buffers. The n routing buffers are associated with n adjacent nodes of the system. The control logic has access to the "command & control" section of each buffer, and makes all the decisions regarding transfer of data both within and without the board. The bidirectional cross-bar switch, which is capable of connecting any buffer to any other buffer or to any link, receives its commands from the control logic. The status table keeps track of the state of various nodes in the system; its contents are updated frequently by status messages that circulate through the network.

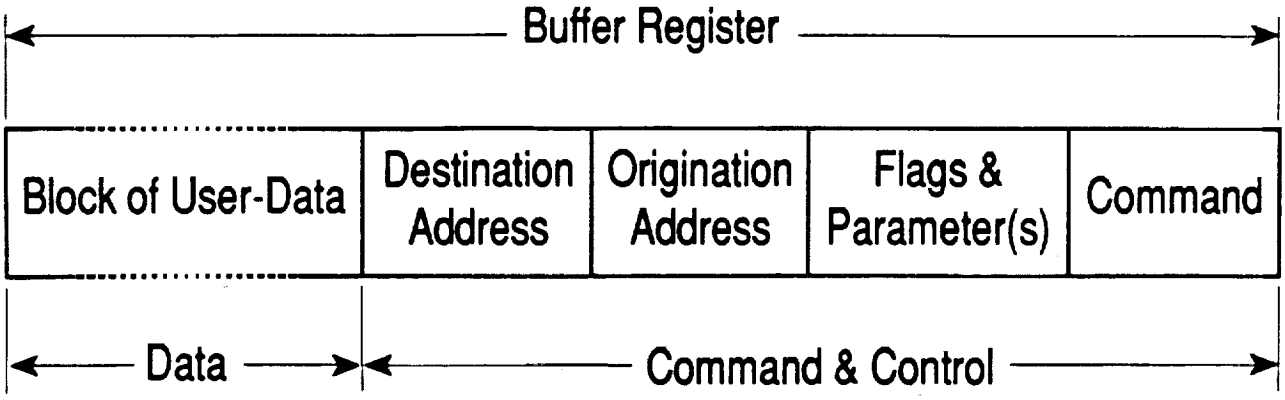


Illustraiton by Vamer
Mansuripur / Opt sci fig 1

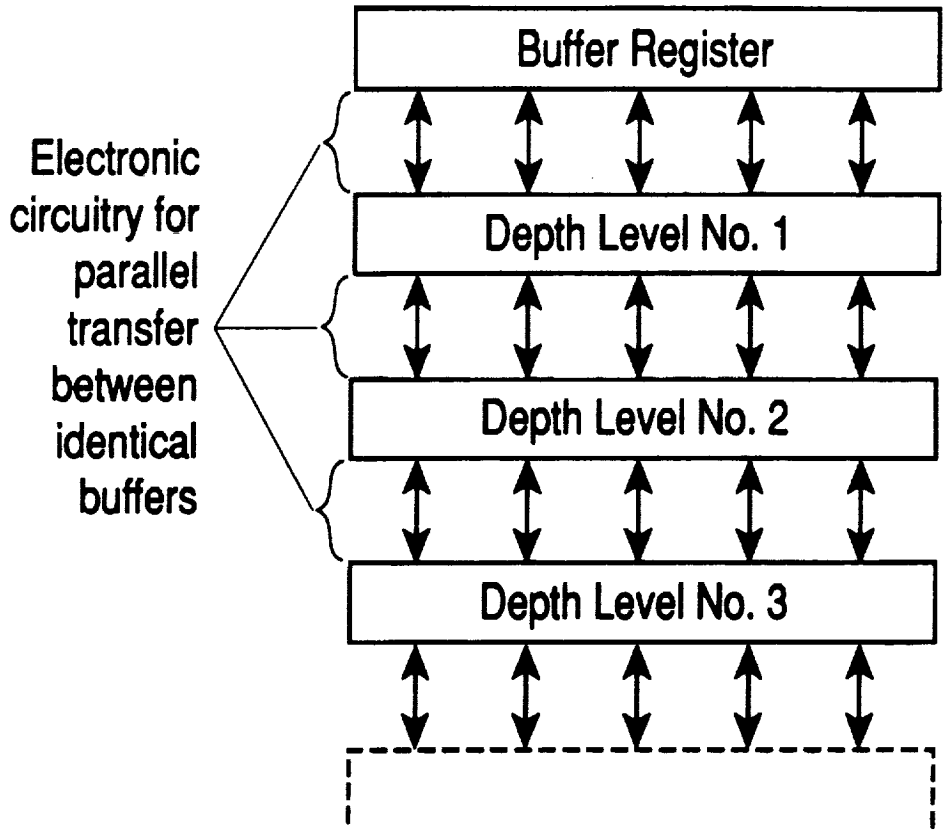


Illustraiton by Vamer
Mansuripur / Opt sci fig 2

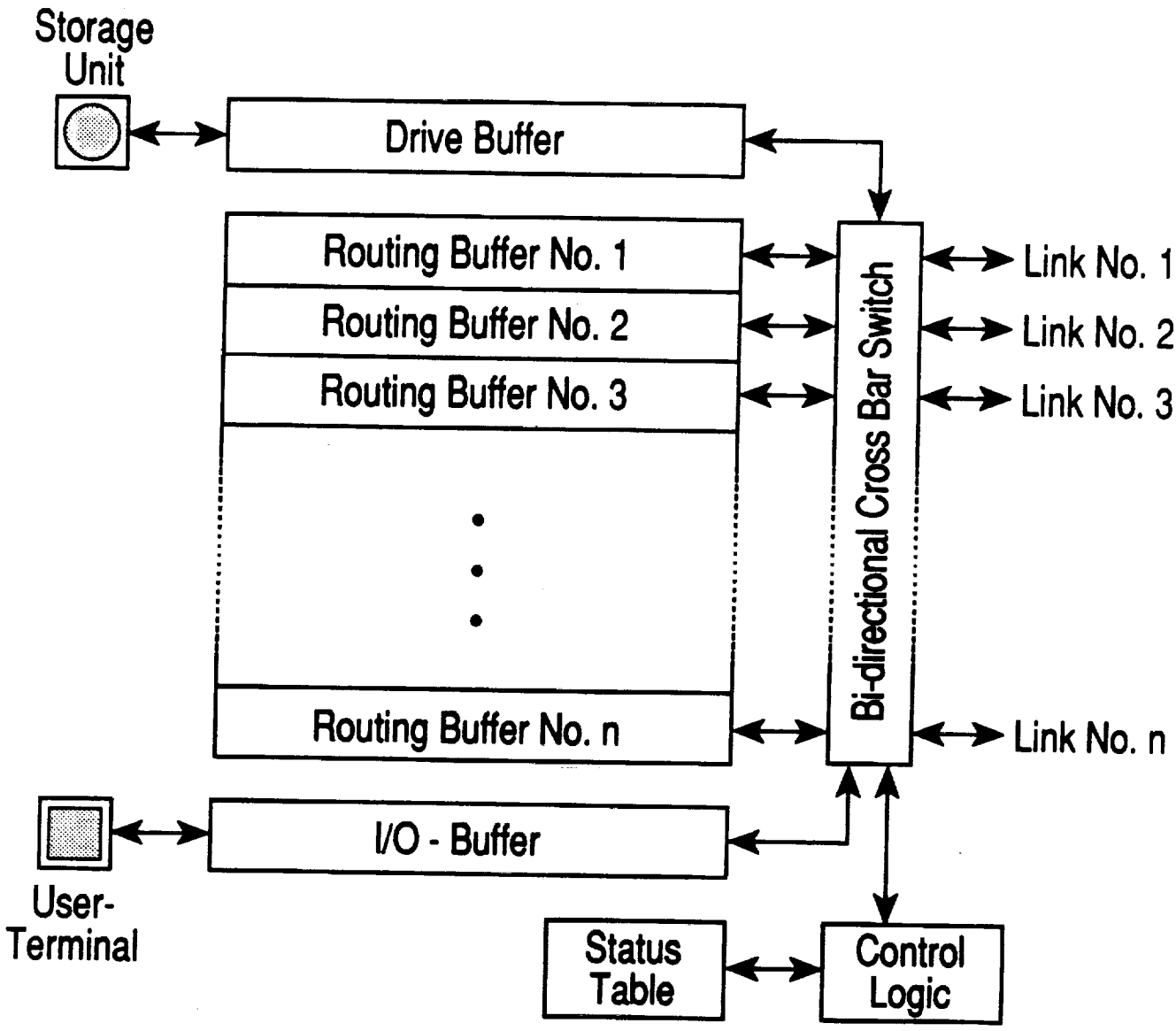
Q.2



Illustraiton by Vamer Mansuripur / Opt sci fig 3



Illustraiton by Varner Mansuripur / Opt sci fig 4



Illustraiton by Varner Mansuripur / Opt sci fig 5

APPENDIX F

COMBINING RAY-TRACE AND DIFFRACTION ANALYSIS: A DESIGN EXAMPLE

Tom D. Milster
Optical Sciences Center, University of Arizona, Tucson, Arizona 85721

Jeffrey P. Treptau
Cray Research, Chippewa Falls, WI 54729

ABSTRACT

We present an example of using a combined ray trace and diffraction modeling code to simulate effects of objective-lens tilt in an optical data storage device.

1.INTRODUCTION

In some cases, neither ray-trace analysis nor diffraction analysis can give an adequate description of an optical system. The designer that is faced with the problem of analyzing such a system is forced to use a ray-trace program to determine aberrations in the exit pupil and then introduce aberration coefficients into a diffraction model that simulate the propagation. We find this approach rather awkward, especially if complicated aberrations are present. Our approach is to integrate a diffraction analysis and a ray-trace description of an optical path into one program.

Our design is taken from a data storage application, where we must analyze the effects of objective-lens tilt. The optical path is shown in Figure 1. A diode laser beam is collimated, and it is circularized with an anamorphic prism pair. The beam then passes through a partially polarizing beam splitter, (PPBS), which passes 70% of the irradiance polarized in the *p* direction and reflects 100% of the *s* polarization. The transmitted beam then reflects off of the turning mirror into a high numerical aperture (NA) objective lens. The objective lens is usually constructed as a molded bi-aspheric element. The light is then focused down to the disk plane through a cover plate. At the recording layer, there is a complicated interaction between the laser spot and the medium. The disk usually has a groove structure, which means that the analysis must include interaction with a periodic phase object. If the medium is a magneto-optic film, the analysis must include effects of polarization rotation upon reflection. After reflection, the beam is recollimated and deflected into the servo and data detection optics.

This application requires many types of analysis. Since we are using a laser diode, we must consider the type of beam that a laser diode generates with a physical-optics description. Interaction with the medium also requires a physical-optics description. Magneto-optic media require polarization effects to be simulated. A ray-trace analysis is necessary to quantify the aberrations introduced by the tilted objective lens.

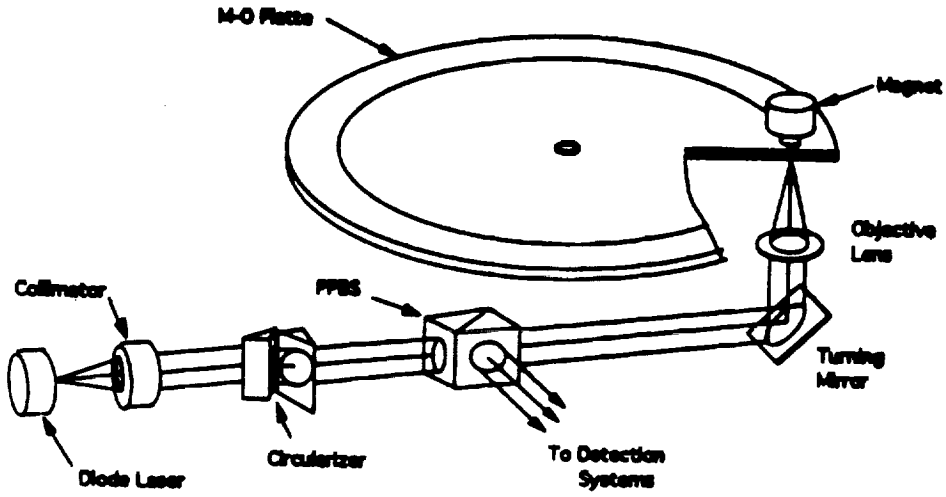


Figure 1. Layout of an optical data storage device.

Preprint from SPIE 1625

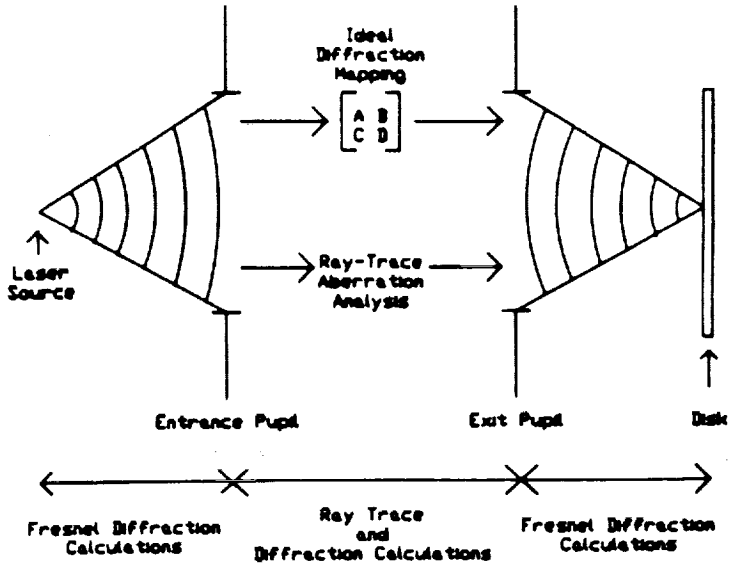


Figure 2. Diagram of entrance pupil to exit pupil mapping technique.

In a previous publication (1) we have described how to analyze these effects. Our model uses a ray trace from the object to the image plane and a scalar diffraction propagation to simulate physical optics phenomena. There are commercial modeling codes, like the GLAD analysis package (2), that come close to meeting all of our requirements. However, at the time of this writing, we are working with a code that has been developed in our laboratory. In this paper, we review basic principles of the model and study the effects of objective lens tilt at several planes in the optical system.

2. THEORY

Modeling of the optical system is accomplished in a two-pronged approach, as shown in Figure 2. To begin, the wavefront from the facet of the laser diode is propagated to the entrance pupil of the system with scalar diffraction. Next, we calculate a paraxial system quantifier, the ABCD matrix, which describes propagation from the entrance pupil to the exit pupil. Once the system ABCD matrix has been found, that information is placed into Huygens' integral for wave propagation (3). To this point, we have a physical-optics description of the field in the exit pupil.

The aberrations in the optical system must now be considered. We start with a point source at the laser diode and trace a matrix of rays to the disk plane. The optical path distance (OPD) traversed by each ray is accumulated during the trace. Each ray is then reverse traced to the exit pupil of the system. Ray sampling is chosen so that each ray is identified with a single pixel in a square matrix representing the exit pupil. The square matrix of sample points is the same as for the diffraction propagation discussed in the previous paragraph.

The ideal diffraction wavefront and the OPD in the exit pupil due to aberrations are combined to give the total exit-pupil wavefront. The resulting phase is just the addition of the two component phase fronts on a pixel-by-pixel basis. The amplitude of the beam at the exit pupil is given by the ideal diffraction propagation of the electric field amplitude incident on the entrance pupil.

The wavefront represented at the exit pupil is propagated to the disk using scalar diffraction calculations. After interaction with the medium, the reflected beam is propagated back through the optical system in a similar manner.

3. MODELING RESULTS

This model is implemented on a 486 computer that runs at 33MHZ. The modeling environment is the MATLAB programming language (4). MATLAB is an interactive mathematics programming environment that allows a user to

Table I

%	CURV	THICK	INDEX	DIAM	DECENX	DECENY	TILTX	TILTY	CONIC	D	E	F
laser2disk=[	Inf	22.590628e-3	AIR	0	0	0	0	0	0	0	0	0;
	42.5761e-3	2.542e-3	SF5	5.21e-3	0	0	0	0	0	0	0	0;
	10.4873e-3	3.558e-3	SK11	5.36e-3	0	0	0	0	0	0	0	0;
	-16.2137e-3	24.638042e-3	AIR	5.65e-3	0	0	0	0	0	0	0	0;
	Inf	10e-3	AIR	4.3e-3	0	0	0	0	0	0	0	0;
	Inf	20e-3	BK7	12.7e-3	0	0	0	0	0	0	0	0;
	Inf	10e-3	AIR	12.7e-3	0	0	0	0	0	0	0	0;
	Inf	1e-3	CRYSTAL	12.7e-3	0	0	0	0	0	0	0	0;
	Inf	19.27318e-3	AIR	12.7e-3	0	0	0	0	0	0	0	0;
	25.5864e-3	3.5e-3	SK11	15e-3	0	0	0	0	0	0	0	0;
	-17.1375e-3	2e-3	SF5	15e-3	0	0	0	0	0	0	0	0;
	-69.121e-3	37.264983e-3	AIR	15e-3	0	0	0	0	0	0	0	0;
	Inf	37.673408e-3	AIR	2.22e-3	0	0	0	0	0	0	0	0;
	69.121e-3	2e-3	SF5	15e-3	0	0	0	0	0	0	0	0;
	17.1375e-3	3.5e-3	SK11	15e-3	0	0	0	0	0	0	0	0;
	-25.5864e-3	25.202651e-3	AIR	15e-3	0	0	0	0	0	0	0	0;
	Inf	4.3e-3	AIR	4.32e-3	0	0	0	0	0	0	0	0;
	2.87398e-3	2.127607e-3	SK16	4.32e-3	0	0	0*pi/180	0	0	-.193727e-2	-.192818e-3	-.142685e-4;
	-32.43878e-3	2.396264e-3	AIR	4.2e-3	0	0	0*pi/180	0	0	.432639e-2	-.425675e-3	.113725e-3;
	Inf	1.2*e-3	PSK3	4e-3	0	0	0*pi/180	0	0	0	0	0;
	Inf	0	AIR	4e-3	0	0	0*pi/180	0	0	0	0	0;
	Inf	0	AIR	2e-3	0	0	0	0	0	0	0	0];

perform a variety of calculations, including those necessary for the scalar diffraction model of light propagating through the system.

The setup parameters are entered through a text file that is edited before beginning the master subroutine. Input parameters for laser diode beam widths, beam quality, wavelength, etc., are placed in the text file. Other parameters, like polarization characteristics, media format, data patterns, etc., are also placed in the text file. The optical system is entered in the text file as matrix, as shown in Table I for the beam path from laser to disk. The input format for the optical system closely resembles the format used in other ray-trace codes. The optical system consists of the collimator, a beam splitter, a relay lens pair, and a NA=0.55 objective lens. The return path to the servo detectors looks very similar, except that the path is reversed and a lower NA lens is used to focus down onto the detectors. The operating wavelength for this example is 780 nm. The stop is located at the objective lens and has a diameter of 4.0 mm.

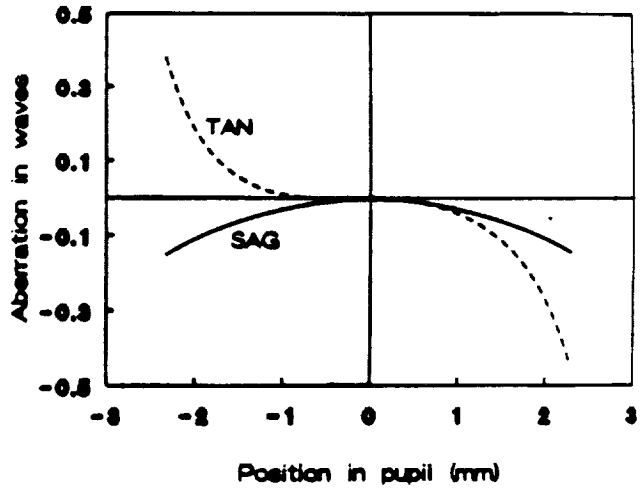


Figure 3. Sagittal (SAG) and tangential (TAN) wave fans for an objective lens tilted 0.50°.

Ideal diffraction propagation is accomplished first. Next, we calculate the OPD distribution due to the ray trace. The field in the exit pupil is calculated from the beam emitted by the laser diode and the system ABCD matrix. A profile of wavefront for a lens tilted 0.5° is shown in Figure 3, which indicates a combination of coma and astigmatism in the sagittal (SAG) and tangential (TAN) wave fans. Next, the diffraction to the disk plane is performed. Figure 4A-C displays spot field distributions for 0°, 0.25°, and 0.50°, objective lens tilt, respectively. The primary aberrations are coma and astigmatism. The field distribution is multiplied by the appropriate disk matrices. An example of a phase matrix representing a groove structure on an optical disk is shown in Figure 5. Propagation back to the detector plane results in the irradiance distributions shown in Figures 6A-C for objective-lens tilts of 0°, 0.25°, and 0.50°, respectively. The small bumps located on each side of the central lobe are due to diffraction from the groove structure. Coma, because it is an odd aberration, is effectively canceled when the reflected light propagates through the objective lens. Astigmatism, because it is an even aberration, is doubled. The irradiance distribution at the detector plane can now be summed over sections of the matrix to simulate effects of bi-cell, quad-cell, or other multi-element detector.

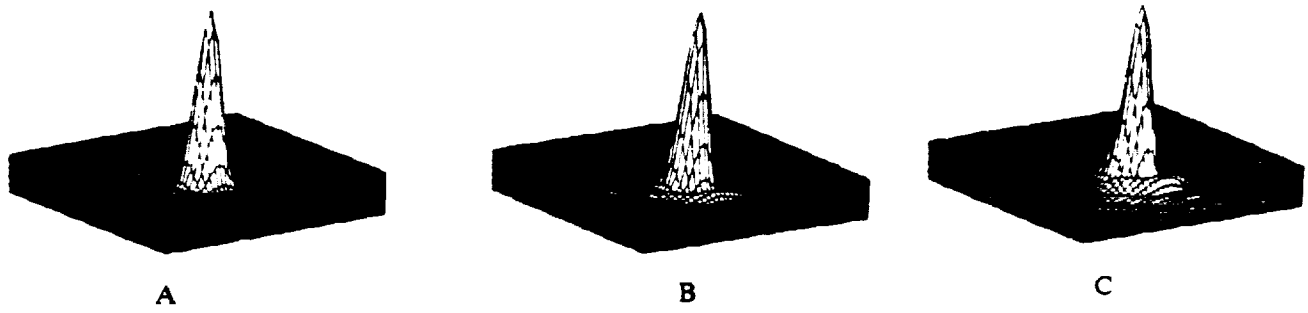


Figure 4. Spot irradiance distributions on the medium for objective-lens tilts of: A) 0.0°; B) 0.25°; and C) 0.50°.

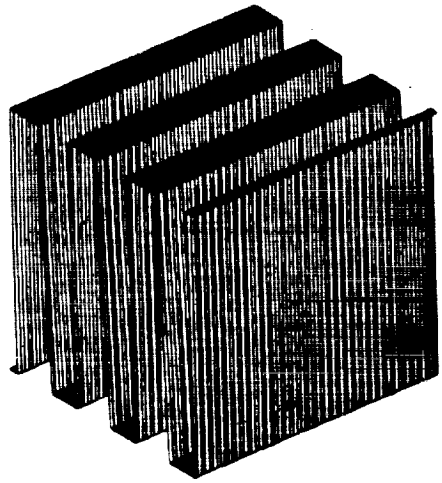


Figure 5. Phase matrix representing the groove structure on the disk.

4. SUMMARY

We have shown an implementation of a combined ray trace and diffraction optical modeling code. Our example application was the determination of effects due to objective-lens tilt in an optical data storage device. A physical-optics description of the beam emitted from the laser diode was used, along with a physical-optics description of the laser interaction with the storage medium. Scalar diffraction theory was used to propagate from plane to plane. The ray-trace analysis was used to determine aberrations due to the tilted lens. Other effects, such as polarization phenomena, can also be modeled.

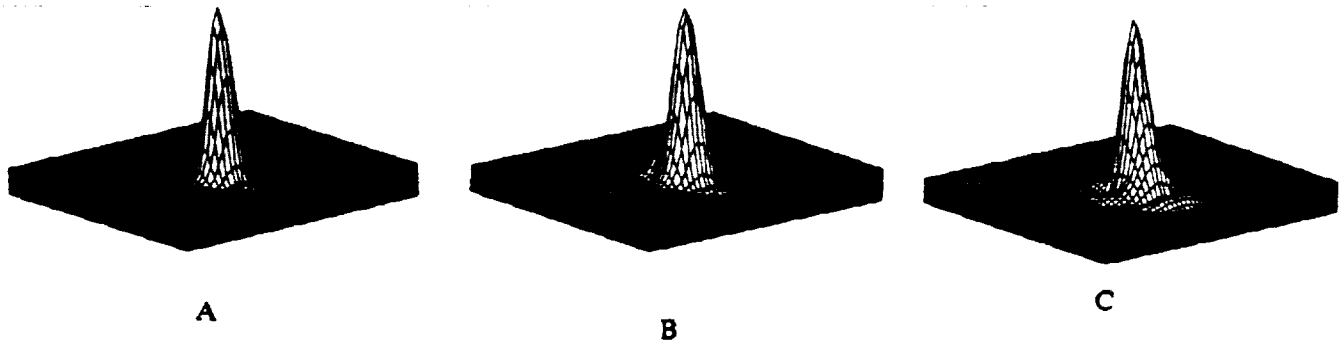


Figure 6. Irradiance distributions on the detector plane for objective lens tilts of: A) 0.0°; B) 0.25°; and C) 0.50°.

ACKNOWLEDGEMENTS

This work was supported by NASA grant no. NGT-50579 and the Optical Data Storage Center.

REFERENCES

1. J.P. Treptau, T.D. Milster and D.G. Flagello, "Laser Beam Modeling in Optical Storage Systems," *SPIE Modeling and Simulation of Laser Systems II*, 1415, pp. 317-321 (1991).

2. Applied Optics Research, commercial literature (1991).

3. A.E. Siegman, Lasers, University Science Books (1986).

4. The Math Works, Inc., (1989).

APPENDIX G

TRANSFER FUNCTION CHARACTERISTICS OF SUPER RESOLVING SYSTEMS

Tom D. Milster and Craig H. Curtis

Optical Sciences Center, University of Arizona, Tucson, Arizona 85721
(602) 621-8280

ABSTRACT

Signal quality in an optical storage device greatly depends on the optical system transfer function used to write and read data patterns. The problem is similar to analysis of scanning optical microscopes. Hopkins (1) and Braat (2) have analyzed write-once-read-many (WORM) optical data storage devices. In this paper, we discuss transfer-function analysis of magnetooptic (MO) data storage devices with respect to improving transfer-function characteristics. Several authors (3)(4)(5) have described improving the transfer function as *super resolution*. However, none have thoroughly analyzed the MO optical system and effects of the medium. In our development, we discuss both the optical system transfer function and effects of the medium.

1. INTRODUCTION

An unfolded magneto-optic (MO) readout optical system is illustrated in Figure 1. Only one detector path is shown. $P_1(\sigma)$ represents the objective-lens aperture. $D(r_d)$ is the disk plane. A collection lens is shown following reflection from the medium. A second aperture, $P_2(\sigma)$, is associated with the collection lens. For most MO systems, the collection lens is a mirror image of the objective lens. The beam then passes through the partially polarizing beam splitter (PPBS) and a half-wave plate. The detector lens focuses the beam onto a detector that exhibits a responsivity as a function of position given by $R(r_d)$. For simplicity, we model $R(r_d)$ as a binary 1/0 (on/off) function. A polarizing beam splitter is placed after the detector lens to provide a complimentary optical path for the second detector.

MO read-back optics can be considered as either nonconfocal or confocal scanning optical microscopes. Nonconfocal systems collect all of the energy on the detectors, that is, $R(r_d) = 1$. Confocal systems place pinholes in detector planes. Pinholes pass only the central portion of the spot on each detector. The effect of each pinhole can be modeled as $R(r_d) = 1$ inside the pinhole and $R(r_d) = 0$ outside the pinhole.

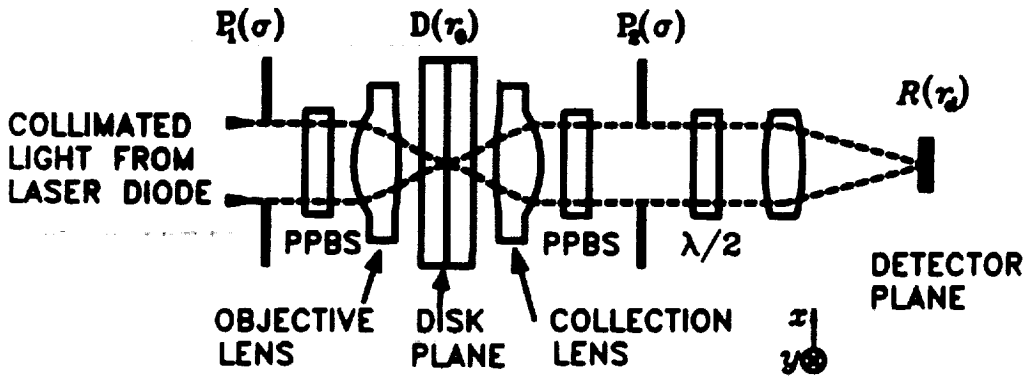


Figure 1. Unfolded MO detection system for one detector path.

Preprint from SPIE 1663

detector planes. Pinholes pass only the central portion of the spot on each detector. The effect of each pinhole can be modeled as $R(r_d) = 1$ inside the pinhole and $R(r_d) = 0$ outside the pinhole.

In a separate paper (6), we have derived an expression for the signal current, i_{sig}, as a function of scan position, r_s, that is given by

$$i_{sig}(r_s) = \int_{-}\int_{-} C(\nu;\nu')\,M(\nu;\nu')\exp\{-2\pi j(\nu - \nu')\cdot r_s\}\,d\nu d\nu' , \quad [1]$$

where $C(\nu;\nu')$ is the optical transfer function and $M(\nu;\nu')$ is the medium function described by

$$M(\nu;\nu') = Q(\nu)V^*(-\nu') + Q^*(-\nu')V(\nu) , \quad [2]$$

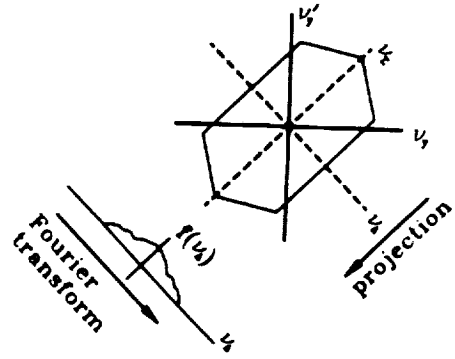


Figure 2. $i_{sig}(r_s)$ calculation. A typical $C(\nu_x;\nu_y')$ contour is shown.

$Q(\nu)$ is proportional to the Fourier transform of the multiplication of the Kerr signal and the reflectivity. $V(\nu)$ is the Fourier transform of the reflectivity. In most MO systems, the reflectivity does not vary across the medium except as a phase perturbation in the x direction due to the presence of grooves. With constant reflectivity, $V(\nu)$ becomes $V(\nu) = \delta(\nu)$. If we assume line objects with frequency components in the y direction and $\nu_x = 0$ and $\nu_x' = 0$, $M(\nu;\nu')$ is only defined along the $\nu_y = 0$ and $\nu_y' = 0$ axes.

We illustrate the calculation of $i_{sig}(y_s)$ in Figure 2. A typical $C(\nu_x;\nu_y')$ contour is sketched that exhibits symmetry about the ν_Δ and ν_Δ axes. Interaction with $C(\nu_x;\nu_y')$ is simplified due to the structure of $M(\nu_x;\nu_y')$. From Equation 2, the only overlap between $C(\nu_x;\nu_y')$ and $M(\nu_x;\nu_y')$ is along $\nu_y = 0$ and $\nu_y' = 0$ axes. After multiplication, a projection (integration) is performed along the ν_Δ axis. Values of $i_{sig}(y_s)$ are calculated by taking the Fourier transform of the result along the ν_Δ axis.

3. Super-resolving optical systems

We now study several super-resolving optical layouts. The numerical aperture (NA) is 0.50 and $\lambda = 830$ nm. This results in a cutoff frequency, ν_c of NA/λ ≈ 600 lines/mm. Gaussian beam width for the overfill on $P_1(\sigma)$ is such that the $1/e^2$ beam irradiance is located at 90% of the pupil radius. This results in an optimum peak irradiance at the medium and power throughput as defined by Haskal (7). Optical layouts we consider include variations of $P_1(\sigma)$, $P_2(\sigma)$, and $R(r_d)$. Due to the nature of $M(\nu_x;\nu_y')$, we only display $C(\nu_x;\nu_y')$ along the $\nu_y' = 0$ axis. Also, transfer functions are normalized to their peak values, which results in a comparison of *relative* contrast values.

An example of our analysis is shown in Figure 3. A central obscuration is used to block a width s of $P_2(\sigma)$. Both nonconfocal (curves B1, C1, and D1) and confocal (curves B2, C2, and D2) cases are shown. For reference, curve A is the nonconfocal configuration using an open aperture for $P_2(\sigma)$. Curves B, C, and D correspond to s/D = 0.25, 0.50, and 0.75, respectively. The frequency cutoff for all curves in Figure 3 is $2\nu_c$. For both the confocal and nonconfocal cases, a conservative obscuration width of 25% significantly improves the high-frequency response of the system without serious degradation at low frequencies. As s increases, high-frequency response improves at the cost of low-frequency response. Above 50% obscuration width, the confocal case does not significantly improve the transfer function.

Several super-resolving MO systems have been described in the literature. A system described by Yamanaka, Hirose, and Kubota (3) uses a prism arrangement to split and separate the Gaussian irradiance pattern on $P_1(\sigma)$. This distributes more power in the side lobes of the spot distribution on the medium. The spot is imaged onto a slit that is used to pass only the central lobe onto the detector. In Figure 4, curve A displays the nonconfocal configuration with $s/D = 0$ and the detector slit width, τ, completely open. This corresponds to curve A in Figure 3. Curve B

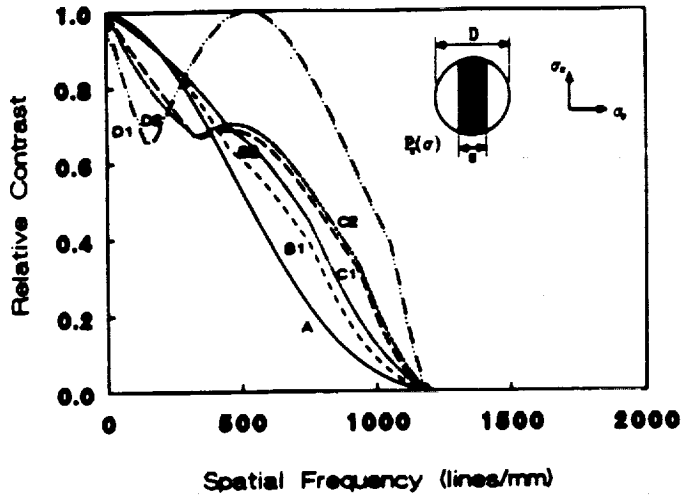


Figure 3. Transfer functions with rectangular obscurations in $P_2(\sigma)$.

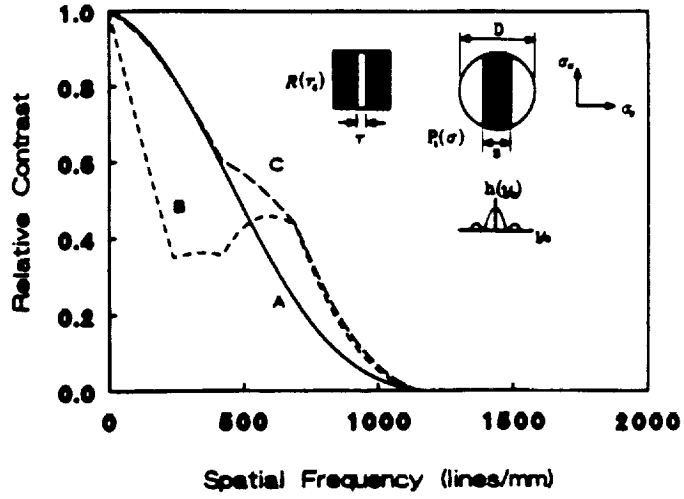


Figure 4. Transfer functions with rectangular obscurations in $P_1(\sigma)$.

corresponds to the case where s/D = 20% and τ is completely open. Even though the high-frequency response is moderately improved, the low-frequency response is severely degraded. Curve C corresponds to the case where s/D = 20% and τ is adjusted to pass only the central lobe to the detector. A significant improvement is obtained in high-frequency response while negligibly affecting low-frequency response.

It is worthwhile to compare curves from Figures 3 and 4. In Figure 3, the central obscuration is the collection optics, which decreases total power on the detectors. In Figure 4, the obscuration is in the illumination optics. This affects power throughput from laser to disk, which must be as high as possible in order to write data at high data rates. Transfer-function characteristics are similar for curve C of Figure 4 and curve B2 of Figure 3, so there appears to be no inherent advantage to placing the obscuration in the illumination optics.

Two other super resolving systems have been discussed. In the Suhara and Nishihara design of an integrated-optic CD disk pickup (4), super resolution improved readout performance. Another interesting system was proposed by Fukumoto et al. (5), in which the medium is self masking. A moderate improvement in the transfer function is observed in the mid-frequency range, and a significant improvement is observed in the high-frequency region. Other types of super resolving systems may be constructed from various apertures in $P_1(\sigma)$ and $P_2(\sigma)$. Aperture types include pinholes, annular rings, or slits. A more complete cataloog of our results is found in Reference 6.

It is desirable to have a sharp edge response in MO readout. The edge response of the system can be found by applying procedures outlined in Figure 2. We first find $M(v_y;v_y')$ by a Fourier transform of $a(y_o)$ = step(y_o). We then multiply by a transfer function and inverse Fourier transform. Edge responses of several systems are shown in Figure 5. Curve A corresponds to the reference system where $P_1(\sigma)$ and $P_2(\sigma)$ are unobstructed and $R(r_d) = 1$ (nonconfocal). Curves B and C correspond to systems with shading bands of 25% and 50%, respectively, in the collection optics and $R(r_d) = 1$. A significant improvement in edge slope is observed for curve C. However, a ringing is observed as the shading band width increases. Curve D is the simulated response from Suhara's and Nisihara's waveguide coupler with $R(r_d) = \delta(r_d)$. No significant advantage of this system is observed. Curve E is the Fukumoto et al. system with an open $P_2(\sigma)$ and $R(r_d) = 1$. A smooth edge response with a steep slope is observed. Not shown is the Fukumoto et al. system with shading bands in the collection aperture, in which we have observed minimal increase in edge slope and noticeable increase in ringing.

Figure 6 displays how well two closely spaced bits are resolved. Two bits 1.56 μm long are separated by 0.52 μm. Curve A represents the bits on the disk plane. Curve B is the reference system response with an open $P_2(\sigma)$ and $R(r_d) = 1$. A significant improvement is observed when compared to curve B. Curve D corresponds to a shading band of 50% in the collection optics and $R(r_d) = 1$. The dip in the detector current between the two bits is

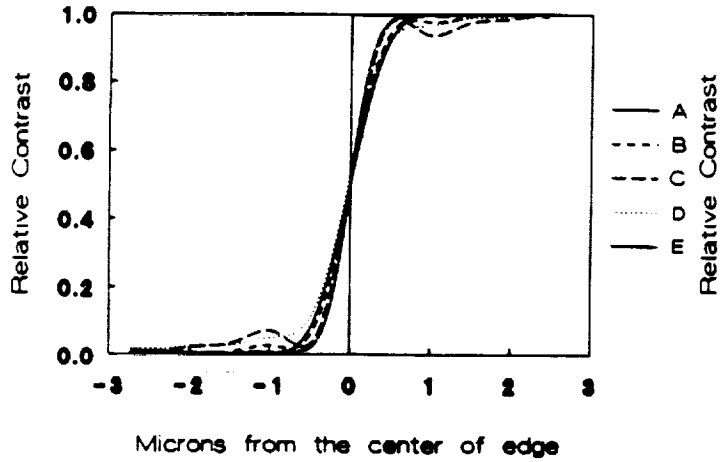


Figure 5. Edge responses for various super-resolving systems.

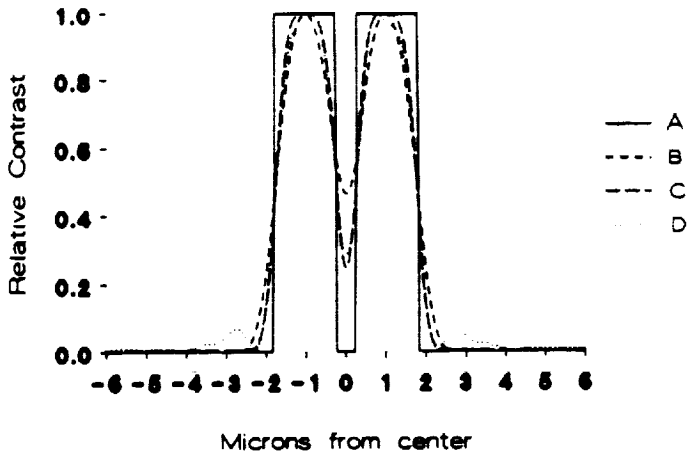


Figure 6. Two-point resolution of several systems.

comparable to the Fukumoto et al. system. However, the ringing observed in the edge response of Figure 5 exhibits itself as artifacts approximately ± 2.5 to 4.0 μm from the center of the pattern. Also, a slight disfiguring of the peak signal is observed.

4. Conclusions

We compared transfer functions of several super resolving MO data storage systems. We found that significant improvement can be obtained in several cases. Of the systems studied, the Fukumoto et al. system with an annular aperture for the collection optics gave the best edge response. In our analysis, we neglected effects of noise in the system. This is an active area in our research group, and we hope to augment our transfer-function analysis with noise studies in a later publication.

5. Acknowledgements

This work was supported by the Optical Data Storage Center. C. Curtis is supported by a fellowship under the Graduate Assistance in Areas of National Need (GAANN) Program.

6. References

1. H.H. Hopkins, "Diffraction theory of laser read-out systems for optical video disks," *JOSA A*, 22, p. 4 (1979).
2. J. Braat, "Optics of recording and read-out in optical disk systems," *Japanese Journal of Applied Physics*, 28, Supplement 28-3 pp. 103-108 (1989).
3. Y. Yamanaka, Y. Hirose and K. Kubota, " High density optical recording by superresolution," *Japanese Journal of Applied Physics*, 28, Supplement 28-3, pp. 197-200 (1989).
4. T. Suhara and H. Nishihara, "Possibility of super-resolution readout in integrated-optic disc pickup," Technical Digest of the International Symposium on Optical Memory, 1989, Kobe, Japan, September 26-28, pp. 97-98.
5. A. Fukumoto, et al., "Super resolution in a magneto-optical disk with an active mask," *SPIE Proceedings on Optical Data Storage*, 1499, pp. 216-225 (1991).
6. T.D. Milster and C.H. Curtis, "Analysis of super resolution in magneto-optic data storage devices," submitted for publication in *Applied Optics*.
7. H.M. Haskal, "Laser recording with truncated Gaussian beams," *Appl. Opt.*, vol 18, no. 13, pp. 2143-2146 (1979).

APPENDIX H

N93-15513

3D Modelling of High Numerical Aperture Imaging in Thin Films

D. G. Flagello

IBM T.J. Watson Research Center
Yorktown Heights, N.Y. 10598

Tom Milster

Optical Science Center
Univerisity of Arizona
Tucson, Arizona 85721

Abstract

This paper describes a modelling technique used to explore three dimensional (3D) image irradiance distributions formed by high numerical aperture (NA > 0.5) lenses in homogeneous, linear films. This work uses a 3D modelling approach that is based on a plane-wave decomposition in the exit pupil. Each plane wave component is weighted by factors due to polarization, aberration, and input amplitude and phase terms. This is combined with a modified thin-film matrix technique to derive the total field amplitude at each point in a film by a coherent vector sum over all plane waves. Then the total irradiance is calculated. The model is used to show how asymmetries present in the polarized image change with the influence of a thin film through varying degrees of focus.

1. Introduction

The demand for highly integrated electronic devices has motivated investigation into better lens resolution in micro-photolithography. Lens resolution is roughly proportional to λ/NA; therefore, we desire to maximize NA and minimize λ. Photoresist, source, and lens materials issues limit the operating wavelength much below 248nm. NAs are design and budget limited, with a possible practical maximum of 0.95. In this paper, we will show that sophisticated models are needed to understand fundamentals of imaging in the high NA regime.

Traditional models used in micro-photolithography are based on scalar image formation under the Fresnel approximations[1]. The central approximation is to treat spherical lens and pupil surfaces as pure quadratic paraboloids. This holds in the low NA regime but it breaks down when the exit pupil diameter is of the same order as the distance from pupil to image, i.e., for NA $\geq$ 0.5. Furthermore, the traditional scalar models cannot treat polarization. They, in effect, assume that each plane-wave component has the same polarization amplitude, with polarization vectors parallel to one another and perpendicular to the axis. In a high NA system, there is significant variation of these vectors across the exit pupil. Each x, y, and z cartesian component has a different polarization contribution. Furthermore, these scalar models treat irradiance within the photoresist film as if the aerial image could be propagated through the film stack as a normally incident plane wave. This approximation is reasonable for NAs below 0.5 or 0.6, but at higher NAs, the propagation angles of the electric field (E) become significant.

One of the early experiments of Wiener[2] shows that, at oblique angles, light polarized parallel to the plane of incidence (P or TM polarization) exhibits reduced contrast in photographic film compared to light polarized perpendicular to the plane of incidence (S or TE polarization). For example, a pair of P-polarized plane waves propagating within a film at $\pm 45^\circ$ produce no interference fringes because their E vectors are perpendicular to each other. Moreover, standing wave effects in the resist film stack are also different in the two polarizations. This is illustrated in Figure 1 with a plane wave incident

on a photoresist film over silicon at 33°, 44°, 58°, and 72° (or 0.55, 0.7, 0.85, and 0.95 NA respectively). The total power absorbed in the film, which is proportional to photoresist exposure, is plotted as a function of coated photoresist thickness in Figure 1. Although the differences between the polarizations are slight at 0.55 NA, they become appreciable high NAs. It becomes apparent that any imaging models for high NA must take into consideration the vector nature of light.

This work presents a 3D vector imaging model for homogeneous, linear, thin films. Previous work[3] concentrated on a 2D approach. Other authors[4-8] have looked at computationally intensive techniques for inhomogeneous films or required axially symmetric systems to obtain analytic solutions. In comparison, by limiting this work to homogeneous films, a technique can be presented that is easily visualized and computationally rapid. Since the final computation is numeric, no restriction is placed on the system symmetry. The model uses the Debye approach[9] to characterize the image field as a plane wave-decomposition. This restricts the propagation direction of the plane waves to the cone of rays subtended by the exit pupil of a high NA lens with its vertex at the geometrical focus. If a film stack is located at or near the focus, the amplitude and phase of each plane wave can be used as input into a thin-film matrix routine to calculate the local field in the top film. The total field is then the summation over all plane waves.

2. Model

Consider the Kohler projection system shown in Figure 2. It is assumed that the system is lossless, and has a Strehl ratio above 0.8. The film of interest is the 1st film, which is homogeneous and linear with thickness d_1. All optical elements from the source to the entrance pupil have low NA and can be modelled by traditional scalar methods. Additionally, the polarization remains constant throughout the optical system and is only altered at the exit pupil. These assumptions allow for a direct mapping of the amplitude and phase of the entrance pupil to the exit pupil of the imaging lens. In Figure 2, the source is initially polarized along the y axis. If a ray is considered to emerge from the exit pupil, it will now have polarization amplitudes for each cartesian component that depend on the exit pupil location.

The main assumption of this model is that each point in the exit pupil gives rise to a Huygens secondary spherical wavelet. If the image field size is much less than the pupil diameter and is located about the axis, the wavelets can be represented as plane waves propagating toward the the geometrical focus. Since each plane wave from an exit pupil point is normal to the localized surface about that point, the plane wave directions are limited to the cone (for circularly symmetric systems) formed by the vertex at the focus and the radius of the pupil. This is the Debye diffraction approach[9]. In the cases considered here, a local field size less than 10μm about the axis is sufficient to encompass all object points contributing significant amplitude to features of interest.

In Figure 3 we illustrate a general propagation vector, $\vec{k}$, for a plane-wave as

$$\begin{aligned}\vec{k} &= k_0\eta(\cos\phi\sin\theta\,\hat{x} + \sin\phi\sin\theta\,\hat{y} + \cos\theta\,\hat{z}) \\ &= k_0\eta(\alpha\,\hat{x} + \beta\,\hat{y} + \gamma\,\hat{z}) \quad ,\end{aligned} \qquad [1]$$

where k_0 is the free space propagation constant , and η is the defined as complex index of refraction given by

$$\eta = n - i\kappa \quad ,$$

where n is the real part of the refractive index and κ is the extinction coefficient.

Since θ is defined as the angle that $\vec{k}$ makes with the z axis and ϕ as the angle that the projection of $\vec{k}$ on the x-y plane makes with the x axis, the direction cosines have the standard definitions:

$$\alpha = \cos\phi\sin\theta$$

$$\beta = \sin\phi\sin\theta$$

$$\gamma = \cos\theta$$

$$\gamma = \sqrt{1 - \alpha^2 - \beta^2} \quad .$$

Propagation of a plane wave to the 1st film is illustrated by Figure 4. The pupil is given as a surface in direction cosine space over the variables (α, β). Since $z = 0$ at the center of curvature of the exit pupil, the direction of each plane-wave eminating from the exit pupil location (α_0, β_0) is completely defined by (α_0, β_0). A coordinate z' is defined as

$$z' = z - z_0 \quad . \qquad [2]$$

where z_0 is the distance from the geometrical focus of the system to the top surface of the film stack referenced to the incident medium. This places the top film surface at $z' = 0$ and the geometrical focus in the incident medium at $z' = -z_0$ or $z = 0$. If no aberrations are present in the optical system, all plane waves from the pupil are in phase at $z = 0$. Each plane wave will have a phase of $-k_0\eta_0\gamma_0 z_0$ at $z = z_0$, where subscripts on α_0, β_0, γ_0, and η_0 reference the incident medium. Also, the position vector, $\vec{r}$, has its origin at the top film surface, and the propagation vector in air, k_0, is coplanar with the refracted propagation vector in the film, k_1.

Each monochromatic plane wave propagating from the pupil to the top surface of the film has the form (with the periodic time factor $e^{i\omega t}$ suppressed)

$$\vec{E}(\alpha, \beta) = \vec{E}_{if}(\alpha, \beta)\, e^{-i\vec{k}_0 \cdot \vec{r}} = E_0(\alpha, \beta)\sqrt{\gamma_0}\; e^{ik_0 W(\alpha, \beta)}\; e^{-ik_0\eta_0\gamma_0 z_0}\; \vec{P}(\alpha, \beta)\; e^{-i\vec{k}_0 \cdot \vec{r}} \quad , \qquad [3]$$

and

$$\vec{k}_0 \cdot \vec{r} = k_0 \eta_0 (\alpha_0 x + \beta_0 y + \gamma_0 z') \quad .$$

$E_0(\alpha, \beta)$ describes the amplitude and phase on the entrance pupil due to the object diffraction. It is a scalar term due to the initial assumption of low NA on the object side. The $\sqrt{\gamma}$ term is arises from the requirement that the system has negligible energy losses and is aplanatic. It is derived[9] using the conservation of energy law, and it results in a complex amplitude at the exit pupil of $E_0(\alpha, \beta)\sqrt{\gamma_0}$. The next exponential term describes any residual aberration in terms of a wavefront phase error, $W(\alpha, \beta)$. Following this is the phase term due to the location of the top surface of the film. $\vec{P}(\alpha, \beta)$ is the polarization vector amplitude distribution across the exit pupil. It can be derived in terms of the propagation vector[10], k_0, since at all times it is assumed that the polarization vector is perpendicular to the propagation vector. Prior to the imaging lens, the initial polarization vector of the electric field is in some state of polarization that can be treated in terms of a orthogonal decomposition in x and y. Therefore, one only needs to examine the incoming component polarization vectors along the x or y axis. After passing through the lens, a given ray will have its polarization vector rotated accordingly. The cartesian components are found by applying the appropriate coordinate transformations. Upon interaction with the thin film surface, a plane of incidence about the z axis can be defined by the angle ϕ. This allows the components of the polarization vector to be further reduced in terms of contributions due to a component perpendicular to the plane of incidence, S or TE, and parallel to this plane, P or TM. Table I shows the values of the polarization amplitudes for each component. Note that $P_x^2 + P_y^2 + P_z^2 = 1$ and that the z component is only comprised of P polarization.

Figure 5 illustrates a plane wave from the pupil with amplitude and phase $\vec{E}_{if}(\alpha, \beta)$, described by Equation 3, arriving at the top surface of the film, interface I. The total field at any x,y,z point within the top film is found by summing the downward and upward fields at that location. The derivation presented here departs from the typical thin film-methods by presenting this in terms of the fields at bottom of the film, interface II, and then relating the result to the incident field at interface I through matrix formalism. The incident and reflected fields at interface II give for the field in the 1st film,

$$\text{Downward Field,} \quad \vec{E}^{+}(\alpha, \beta) = \vec{E}_{if}(\alpha, \beta)\; e^{ik_0 \eta_1 (\alpha_1 x + \beta_1 y + \gamma_1 z'')} \quad , \qquad [4]$$

$$\text{Upward Field,} \quad \vec{E}^{-}(\alpha, \beta) = \vec{E}_{rf}(\alpha, \beta)\; e^{ik_0 \eta_1 (\alpha_1 x + \beta_1 y - \gamma_1 z'')} \quad , \qquad [5]$$

where $z'' = d_1 - z'$.

The following relationships are found from Snell's law in direction cosine notation:

$$\begin{aligned} \eta_1 \alpha_1 &= \eta_0 \alpha_0 \\ \eta_1 \beta_1 &= \eta_0 \beta_0 \end{aligned} \quad , \tag{6}$$

and

$$\eta_1 \gamma_1 = \sqrt{\eta_1^2 - \eta_0^2 \sin^2(\cos^{-1}\gamma_0)} \quad . \tag{7}$$

The total field is given as

$$\begin{aligned} \vec{E}_T(\alpha, \beta) &= \vec{E}^+(\alpha, \beta) + \vec{E}^-(\alpha, \beta) \\ &= e^{ik_0 \eta_0 (\alpha_0 x + \beta_0 y)} \left(\vec{E}_{inf}(\alpha, \beta)\, e^{i\Phi(z'')} + \vec{E}_{ref}(\alpha, \beta)\, e^{-i\Phi(z'')} \right) , \end{aligned} \tag{8}$$

where

$$\Phi(z'') = k_0 z'' \sqrt{\eta_1^2 - \eta_0^2 \sin^2(\cos^{-1}\gamma_0)} \quad . \tag{9}$$

The amplitudes in Equation 8 are related to the incident field by using thin-film matrix techniques to derive reflection and transmission coefficients for the full film stack and a partial film stack. Since the basics of this technique[11,12] are well known, only the novel mathematics pertaining to the current discussion will be presented. It is assumed that the index of refraction above the first surface, η_0, is real, and that the film in question is the first film layer on an arbitrary film stack. Using the notation in Hecht[13], a matrix can be defined for each film, j, by

$$M_j = \begin{bmatrix} \cos\delta_j & \dfrac{i \sin\delta_j}{Y_j} \\ iY_j \sin\delta_j & \cos\delta_j \end{bmatrix} , \tag{10}$$

where Y_j is defined as the effective index for S and P polarization and δ_j is the phase for a film thickness d_j, given by

$$\begin{aligned} Y_{jS} &= \eta_j \gamma_j \\ Y_{jP} &= \frac{\eta_j}{\gamma_j} \\ \delta_j &= k_0 \eta_j d_j \gamma_j \quad . \end{aligned} \tag{11}$$

A total M matrix can be defined that is the product of all the individual film matrices, as well as a partial M′ matrix that represents the film stack without the contribution of the 1st film:

$$\text{Total Matrix,} \qquad M = \prod_{j=1}^{j=s} M_j = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} , \tag{12}$$

and

$$\text{Partial Matrix,} \qquad M' = \prod_{j=2}^{j=s} M_j = \begin{bmatrix} m'_{11} & m'_{12} \\ m'_{21} & m'_{22} \end{bmatrix} . \tag{13}$$

With respect to Figure 5, the reflection and transmission vector coefficients for the full film stack, $\vec{r}$ and $\vec{\tau}$ respectively, and at the 2nd interface, $\vec{r}_{II}$ and $\vec{\tau}_{II}$, can be defined with components given by

$$\tau = \tau_x = \tau_y = \frac{E_{\tau sx}}{E_{idx}} = \frac{E_{\tau sy}}{E_{idy}} \qquad \tau_z = \frac{E_{\tau sz}}{E_{idz}} , \tag{14}$$

$$\tau_{II} = \tau_{IIx} = \tau_{IIy} = \frac{E_{\tau sx}}{E_{idIIx}} = \frac{E_{\tau sy}}{E_{idIIy}} \qquad \tau_{IIz} = \frac{E_{\tau sz}}{E_{idIIz}} , \tag{15}$$

$$r = r_x = r_y = \frac{E_{rIx}}{E_{idx}} = \frac{E_{rIy}}{E_{idy}} \qquad r_z = \frac{E_{rIz}}{E_{idz}} , \tag{16}$$

and

$$r_{II} = r_{IIx} = r_{IIy} = \frac{E_{rIIx}}{E_{idIIx}} = \frac{E_{rIIy}}{E_{idIIy}} \qquad r_{IIz} = \frac{E_{rIIz}}{E_{idIIz}} , \tag{17}$$

where the original assumption of film homogeneity and linearity has been invoked to allow the equality of the x and y coefficients.

The coefficients can be expanded following a similar treatment outlined in Hecht in terms of the M and M′ matrix elements,

$$\tau = \frac{2Y_0}{Y_0 m_{11} + Y_0 Y_s m_{12} + m_{21} + Y_s m_{22}} \qquad \tau_{II} = \frac{2Y_1}{Y_1 m'_{11} + Y_1 Y_s m'_{12} + m'_{21} + Y_s m'_{22}} \tag{18}$$

$$r = \frac{Y_0 m_{11} + Y_0 Y_s m_{12} - m_{21} - Y_s m_{22}}{Y_0 m_{11} + Y_0 Y_s m_{12} + m_{21} + Y_s m_{22}} \qquad r_{II} = \frac{Y_1 m'_{11} + Y_1 Y_s m'_{12} - m'_{21} - Y_s m'_{22}}{Y_1 m'_{11} + Y_1 Y_s m'_{12} + m'_{21} + Y_s m'_{22}} , \tag{19}$$

where the dependence of the coefficients on α and β is given by Equation 11.

Propagating plane waves in the top film, which are represented by Equations 4 and 5, must also be solutions to the homogeneous wave equation and, therefore, Maxwell's equations. In particular, for a charge free media, it is required that the divergence of the field is zero in the incident medium and the film in question, hence

$$\vec{\nabla} \cdot \vec{E} = \vec{k} \cdot \vec{E} = 0 . \tag{20}$$

Equations 6 and 14 for the incident medium give

$$\vec{k}_0 \cdot \vec{E}_{id} = \eta_0 (\alpha_0 E_{idx} + \beta_0 E_{idy} + \gamma_0 E_{idz}) = \frac{\eta_0 (\alpha_0 E_{\tau sx} + \beta_0 E_{\tau sy})}{\tau} + \frac{\eta_0 \gamma_0 E_{\tau sz}}{\tau_z} = 0 , \tag{21}$$

which gives

$$\eta_0 (\alpha_0 E_{\tau sx} + \beta_0 E_{\tau sy}) = -\frac{\tau \eta_0 \gamma_0 E_{\tau sz}}{\tau_z} \quad . \tag{22}$$

Equation 20 for the downward field in the film and equation 22 give

$$\begin{aligned} \vec{k}_1 \cdot \vec{E}^+ &= \eta_1 (\alpha_1 E_{dIIx} + \beta_1 E_{dIIy} + \gamma_1 E_{dIIz}) \\ &= \frac{\eta_0 (\alpha_0 E_{\tau sx} + \beta_0 E_{\tau sy})}{\tau_{II}} - \frac{\eta_1 \gamma_1 E_{\tau sz}}{\tau_{IIz}} \\ &= \frac{-\tau \eta_0 \gamma_0 E_{\tau sz}}{\tau_z \tau_{II}} - \frac{\eta_1 \gamma_1 E_{\tau sz}}{\tau_{IIz}} = 0 \quad . \end{aligned} \tag{23}$$

Therefore,

$$\frac{\tau_z}{\tau_{IIz}} = \frac{\tau}{\tau_{II}} \frac{\eta_0 \gamma_0}{\eta_1 \gamma_1} \quad . \tag{24}$$

Similarly, for the upward field,

$$\vec{k}_1 \cdot \vec{E}^- = \eta_1 (\alpha_1 E_{rIIx} + \beta_1 E_{rIIy} - \gamma_1 E_{rIIz}) = r_{II} \eta_1 (\alpha_1 E_{dIIx} + \beta_1 E_{dIIy}) - r_{IIz} \eta_1 \gamma_1 E_{dIIz} = 0 \quad . \tag{25}$$

By Equation 23,

$$\eta_1 (\alpha_1 E_{dIIx} + \beta_1 E_{dIIy}) = -\eta_1 \gamma_1 E_{dIIz} \quad . \tag{26}$$

Then, by substituting Equation 26 into 25,

$$r_{II} \eta_1 \gamma_1 E_{dIIz} = -r_{IIz} \eta_1 \gamma_1 E_{dIIz} \quad , \tag{27}$$

and,

$$r_{IIz} = -r_{II} \quad . \tag{28}$$

Finally, using Equations 3, 14-17, 24, and 28, Equation 8 can be given for each cartesian component, linking the thin film field with each plane wave emanating from the pupil, that is,

$$E_{Tx}(\alpha, \beta) = e^{ik_0 \eta (\alpha x + \beta y)} e^{-ik_0 \eta \gamma z_0} E_0(\alpha, \beta) \sqrt{\gamma} e^{ik_0 W(\alpha, \beta)} P_x(\alpha, \beta) \frac{\tau}{\tau_{II}} (e^{i\Phi(z'')} + r_{II} e^{-i\Phi(z'')}) \quad , \tag{29}$$

$$E_{Ty}(\alpha, \beta) = e^{ik_0 \eta (\alpha x + \beta y)} e^{-ik_0 \eta \gamma z_0} E_0(\alpha, \beta) \sqrt{\gamma} e^{ik_0 W(\alpha, \beta)} P_y(\alpha, \beta) \frac{\tau}{\tau_{II}} (e^{i\Phi(z'')} + r_{II} e^{-i\Phi(z'')}) \quad , \tag{30}$$

and

$$E_{Tz}(\alpha, \beta) = e^{ik_0 \eta (\alpha x + \beta y)} e^{-ik_0 \eta \gamma z_0} E_0(\alpha, \beta) \sqrt{\gamma} e^{ik_0 W(\alpha, \beta)} P_z(\alpha, \beta) \frac{\tau}{\tau_{II}} \frac{\eta \gamma}{\eta_1 \gamma_1} (e^{i\Phi(z'')} - r_{II} e^{-i\Phi(z'')}) \quad , \tag{31}$$

where α, β, γ, and η are referenced to the incident medium.

Since reflection and transmission coefficients are only defined in terms of either the S or P orientation, Equations 29-31 are further reduced into S and P components. The total field at any point within the 1st film is then given by the component sum of $\vec{E}_T(\alpha, \beta)$ due to the S contributions and the P contributions over the solid angle subtended by the pupil. The normalized field is given by

$$\vec{E}(x,y,z) = K\int_{\Omega}\vec{E}_{TS}(\alpha,\beta)\,d\Omega + K\int_{\Omega}\vec{E}_{TP}(\alpha,\beta)\,d\Omega \quad , \tag{32}$$

where K is a the normalization constant.

Since the differential solid angle can be written as

$$d\Omega = \frac{d\alpha\,d\beta}{\gamma} \quad , \tag{33}$$

the final expanded form for the local electric field components in a film is:

$$\begin{aligned} E_x(x,y,z) = K\int_{\alpha}\int_{\beta}\frac{E_0(\alpha,\beta)\,P_{xS}(\alpha,\beta)\,e^{ik_0W(\alpha,\beta)}}{\sqrt{\gamma}}\,e^{ik_0\eta(\alpha x+\beta y-\gamma z_0)}\,\frac{\tau_S}{\tau_{IIS}}\,(e^{i\Phi(z'')} + r_{IIS}\,e^{-i\Phi(z'')})\,d\alpha\,d\beta \\ + K\int_{\alpha}\int_{\beta}\frac{E_0(\alpha,\beta)\,P_{xP}(\alpha,\beta)\,e^{ik_0W(\alpha,\beta)}}{\sqrt{\gamma}}\,e^{ik_0\eta(\alpha x+\beta y-\gamma z_0)}\,\frac{\tau_P}{\tau_{IIP}}\,(e^{i\Phi(z'')} + r_{IIP}\,e^{-i\Phi(z'')})\,d\alpha\,d\beta \end{aligned} \tag{34}$$

$$\begin{aligned} E_y(x,y,z) = K\int_{\alpha}\int_{\beta}\frac{E_0(\alpha,\beta)\,P_{yS}(\alpha,\beta)\,e^{ik_0W(\alpha,\beta)}}{\sqrt{\gamma}}\,e^{ik_0\eta(\alpha x+\beta y-\gamma z_0)}\,\frac{\tau_S}{\tau_{IIS}}\,(e^{i\Phi(z'')} + r_{IIS}\,e^{-i\Phi(z'')})\,d\alpha\,d\beta \\ + K\int_{\alpha}\int_{\beta}\frac{E_0(\alpha,\beta)\,P_{yP}(\alpha,\beta)\,e^{ik_0W(\alpha,\beta)}}{\sqrt{\gamma}}\,e^{ik_0\eta(\alpha x+\beta y-\gamma z_0)}\,\frac{\tau_P}{\tau_{IIP}}\,(e^{i\Phi(z'')} + r_{IIP}\,e^{-i\Phi(z'')})\,d\alpha\,d\beta \end{aligned} \tag{35}$$

$$E_z(x,y,z) = K\int_{\alpha}\int_{\beta}\frac{E_0(\alpha,\beta)\,P_{zP}(\alpha,\beta)\,e^{ik_0W(\alpha,\beta)}}{\sqrt{\gamma}}\,e^{ik_0\eta(\alpha x+\beta y-\gamma z)}\,\frac{\tau_P\eta\gamma}{\tau_{IIP}\eta_1\gamma_1}\,(r_{IIP}\,e^{-i\Phi(z'')} - e^{i\Phi(z'')})\,d\alpha\,d\beta\,. \tag{36}$$

The limits of integration for circularly symmetric systems extend from -NA to NA with the requirement that

$$\sqrt{\alpha^2+\beta^2} \le NA \quad . \tag{37}$$

The localized irradiance at any point within the film can be expressed as,

$$I(x,y,z) = [\vec{E}(x,y,z)\cdot\vec{E}^{*}(x,y,z)]\,Re(\eta_1) = (\,|E_x(x,y,z)|^2 + |E_y(x,y,z)|^2 + |E_z(x,y,z)|^2\,)\,Re(\eta_1) \quad . \tag{38}$$

This equation is easily evaluated numerically. The minimum sampling size can be determined either by trial and error, looking for convergence of the image, or by recognizing that Equations 34-36 represent a Fourier transform. If the image function is band limited within the image field, then by sampling theory[14], the image field sizes, X and Y, are inversely proportional to the sampling in the pupil. The minimum direction cosine interval can then be written as

$$\Delta\alpha = \frac{\lambda}{X}$$
$$\Delta\beta = \frac{\lambda}{Y} \quad [39]$$

In the cases of interest in this work the field size is limited to $5\mu m$ at $\lambda = 248nm$ giving $\Delta\alpha = 0.05$ For a 0.95 NA circularly symmetric system this results in 1100 samples in the pupil, i.e., only 1100 plane waves are needed to create the image within a film. This number is much less than other imaging models based on plane-to-plane propagation.

3. Results

Equation 38 was evaluated using a wavelength of $\lambda = 248nm$ and an incident medium of air with the wavefront aberration set to $W(\alpha, \beta) = 0$. Initially, the image of a point, or the point spread function (PSF), was computed for a non-absorbing photoresist film of $\eta_1 = 1.8$ on a matched substrate for an NA of 0.95. Figure 6 shows the normalized iso-irradiance contours for meridional slices along x and y for an initial source polarized along y and $z_0 = -.2\mu m$. Due to a significant z component contribution, the y slice is slightly wider than x and the interference effects are also less pronounced. This results in an asymmetry of the PSF in the x-y plane. There is also some asymmetry about the image in depth. This is explained by the spherical aberration caused by the top film surface. By Snell's law, the marginal ray at 0.95 NA will cross the optical axis at the bottom of the film or $z' = 1\mu m$. However, any paraxial ray will cross the axis at approximately $z' = .45\mu m$. The difference between these represents the longitudinal spherical aberration. The fact that most of the power is centered around a depth of $0.5\mu m$ is probably due to the smaller amplitude of the marginal plane waves from reflection losses at high angles and the greater number of paraxial plane waves .

Figure 7 shows a comparison of a the film PSF using a 0.6 and 0.95 NA system. The film is moderately absorbing with $\eta_1 = 1.8 - .04i$ over a silicon substrate with $\eta_s = 1.7 - 3.38i$. For each NA, the film surface is offset to result in the marginal ray crossing the axis at a depth of $0\mu m$ (no offset), $0.5\mu m$, and $1\mu m$. Since the substrate is no longer matched, standing wave interference patterns are observed. Decay of the irradiance due to absorption is also observed. The effects of polarization are more apparent at 0.95 NA, as evidenced by the difference in behavior between the x and y slices.

A tri-bar object with 0.17 lines and 0.17 spaces parallel to the x axis is shown in Figure 8a. The film, substrate, and z_0 offsets are identical to the previous 0.95 NA example. The resultant y meridional slice images using a 0.95 NA system are presented in Figures 8b-8d for x and y polarizations. Again, interference effects are less with the y polarization due to the influence of the z component. In Figure 8b the onset of spurious resolution is seen at the bottom of the film by the presence of four major irradiance lobes instead of three. The bending nature of the iso-irradiance contours are clearly the result of the vectorial nature of this model. Figure 9 shows a plot of the irradiance about the zero position at a depth of $z' = 0$ for the images in Figure 8b. Differences such as these become relevant in micro-lithography since the tolerances on the total processed structures maintained in photoresist must not exceed ± 10%.

4. Conclusion

A vector imaging model for thin films has been presented that is based on a plane wave decomposition of the radiation that propagates from the pupil, and a thin-film treatment of propagation into the film stack. The model is valid for NAs approaching unity and and tends to minimize the number of plane waves needed for an image. It is shown that, when the illumination is polarized, asymmetries are created due to high-NA polarization effects and spherical aberration caused by a film surface. The image iso-irradiance contours that are presented illustrate the vectorial nature of the model.

5. References

1 J. W. Goodman, Introduction to Fourier Optics (McGraw-Hill, 1968).

2 O. Wiener, Ann. d. Physik, 40 (1890), 203.

3 D. Flagello, A.E. Rosenbluth, C. Progler, and J. Armitage, Microcircuit Engineering 1991 (Elsevier Science)

4 M. Mansuripur, J. Opt. Soc. Am. A 6, 786-805 (1989).

5 M. Mansuripur, J. Appl. Phys. 67, 6466-6475 (1990).
6 B. Richards and E. Wolf, Proc. Royal Soc. A, 358 (1959).
7 M. Yeung, SPIE J., 922, 149-167 (1988).
8 J. J. Stamnes, Waves in Focal Regions (Adam Hilger, 1986).
9 M. Born and E. Wolf, Principles of Optics (Pergamon Press, 1980).
10 M. Mansuripur, J. Opt. Soc. Am A 3, 2086-2093 (1986).
11 P.H. Berning, Physics of Thin Films, G. Hass, Ed. (Academic Press 1963), 69-121.
12 H. A. MaCleod, Thin-Film Optical Filters (McGraw-Hill, 1989).
13 E. Hecht and A. Zajac Optics (Addeson-Wesley, 1979).
14 J. Gaskill, Linear Systems, Fourier Transforms, and Optics (John Wiley and Sons, 1978).

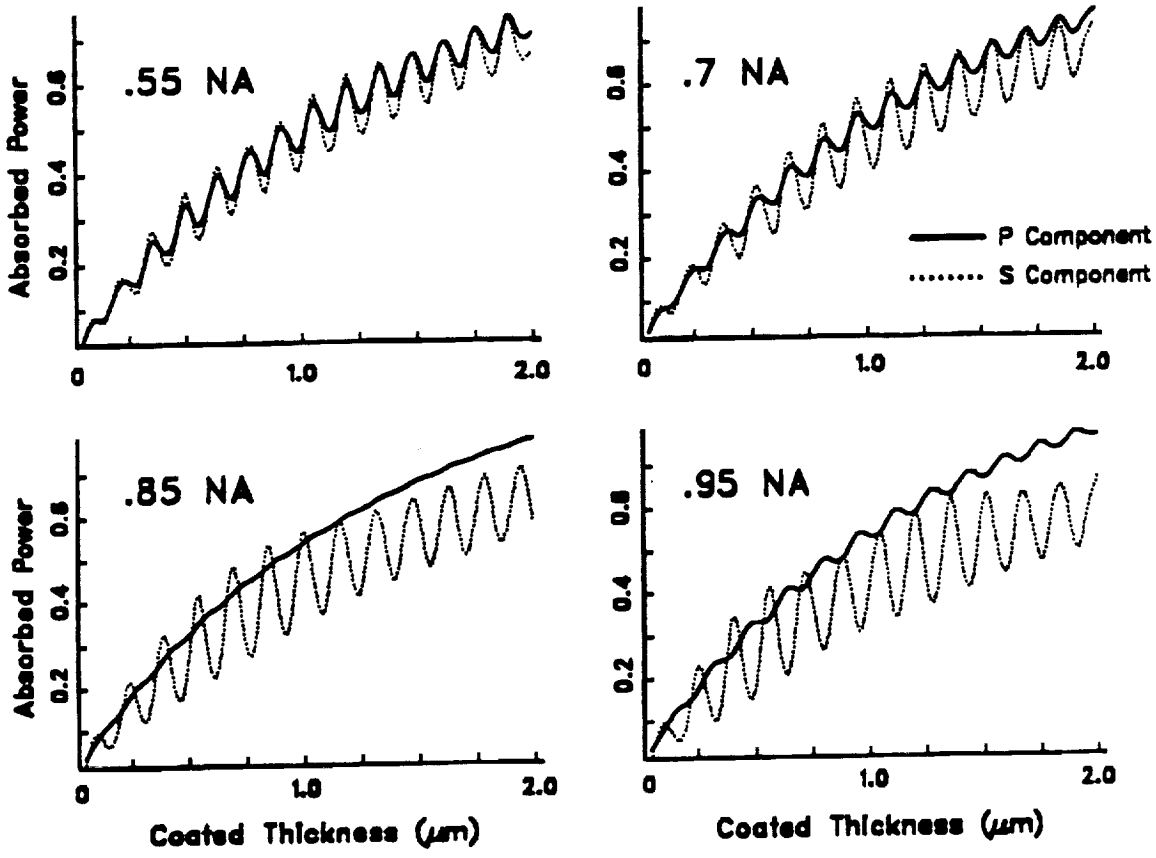


Fig. 1. Total absorbed power of an incident plane wave as a function of coated photoresist thickness for a moderately absorbing photoresist over a silicon substrate at $\lambda = 436nm$ The angle of incidence, θ, is given as the NA in air, i.e. $\sin\theta$

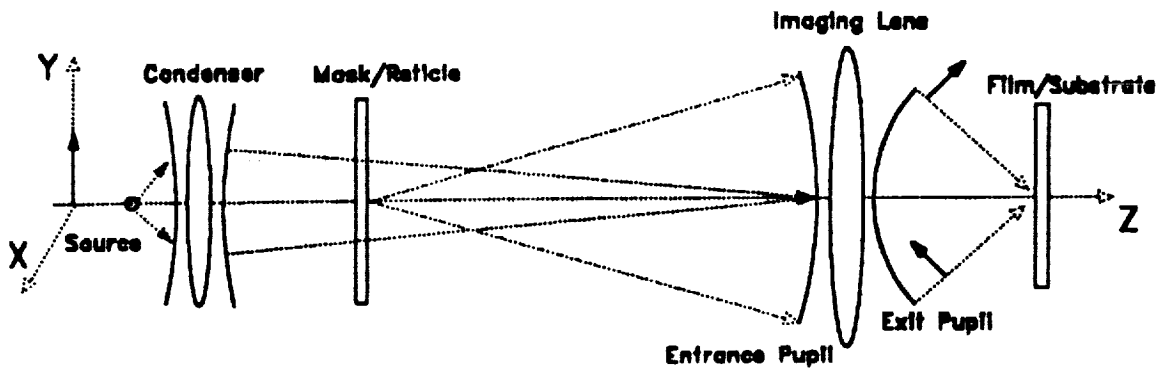


Fig. 2. General Kohler illuminated projection system.

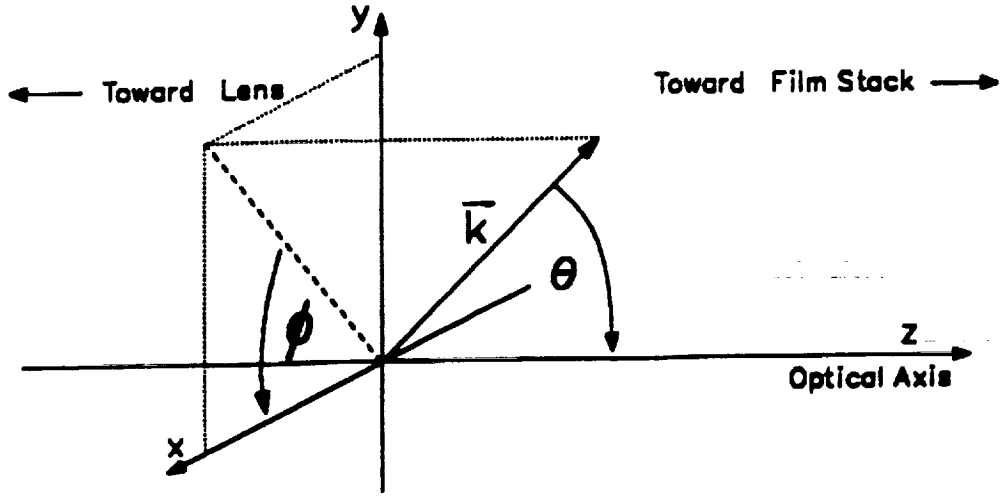


Fig. 3. Coordinate system used to define the projection vector, $\vec{k}$.

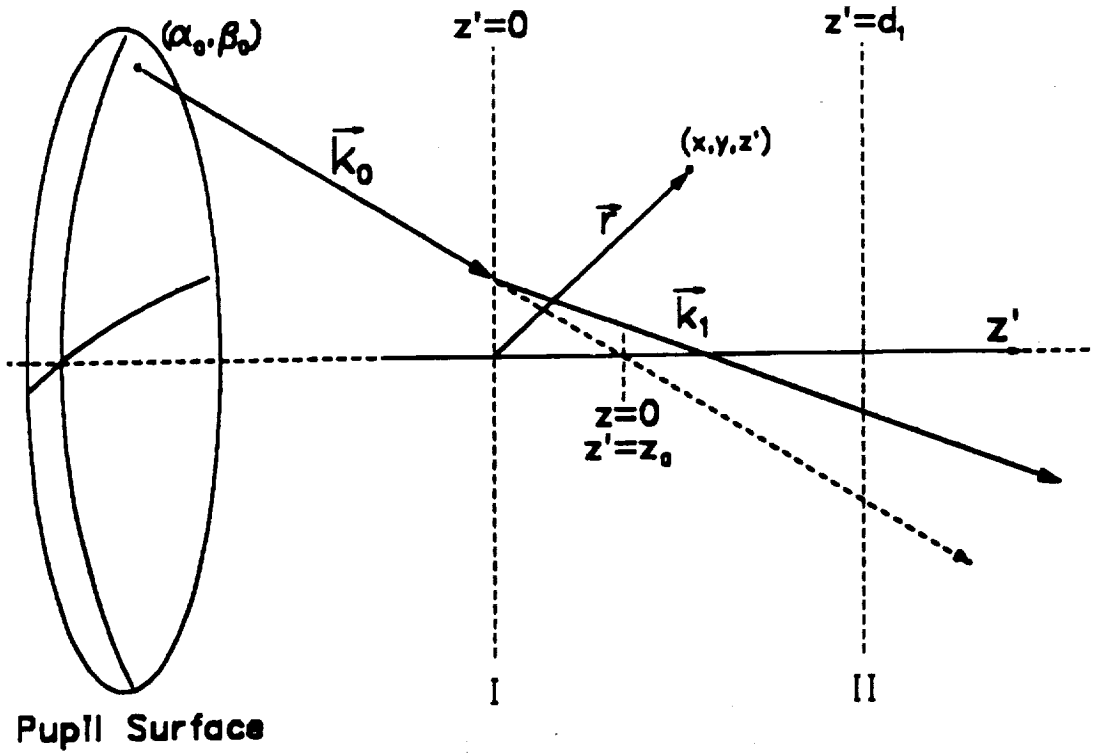


Fig. 4. Schematic of a plane wave propagation vector and position vectors in reference to the pupil and the first thin film.

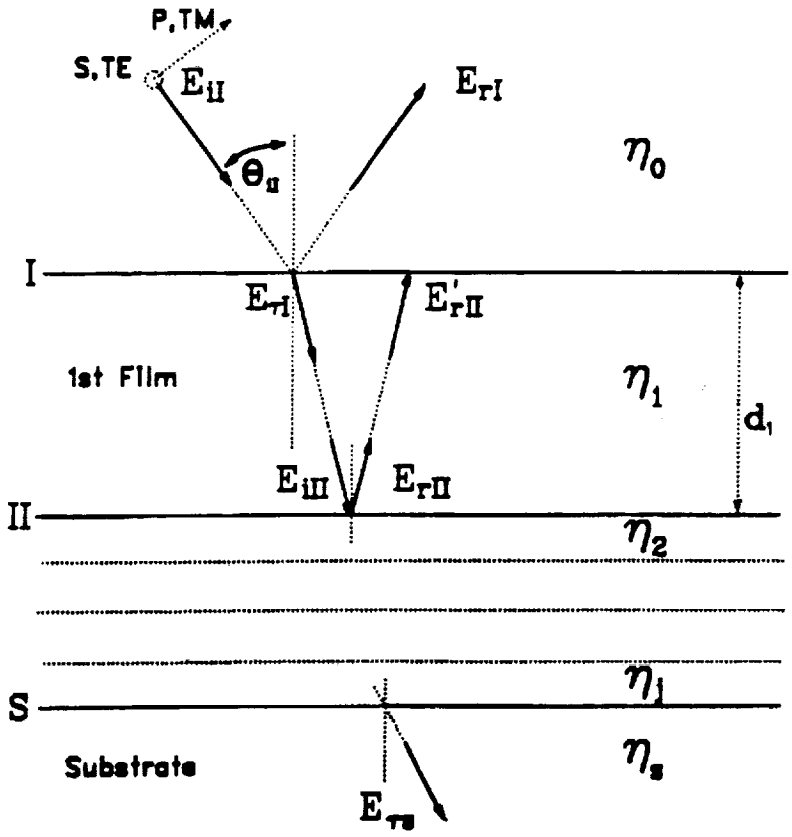


Fig. 5. Diagram of thin film stack with incident plane wave directions and electric field amplitudes.

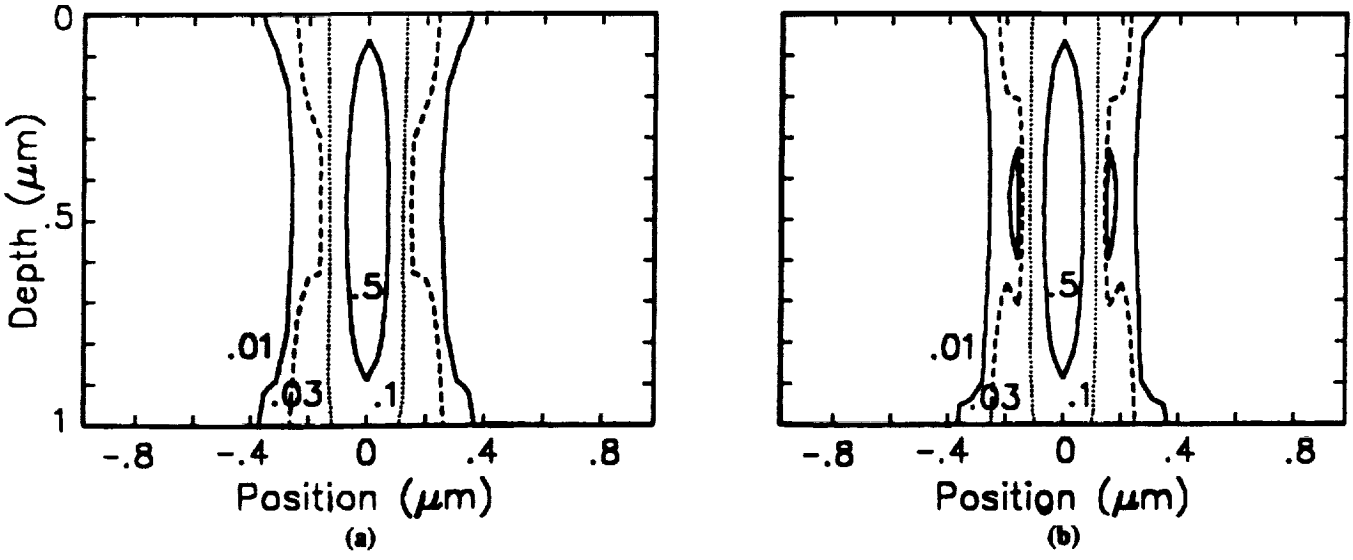


Fig. 6. Iso-irradiance contours of a point image in a $1\mu m$ photoresist film of $\eta_1 = 1.8$ on a matched substrate for a 0.95 NA system, $\lambda = 248nm$, and an initial polarization along the y axis. Meridional slices along (a) y axis, (b) x axis.

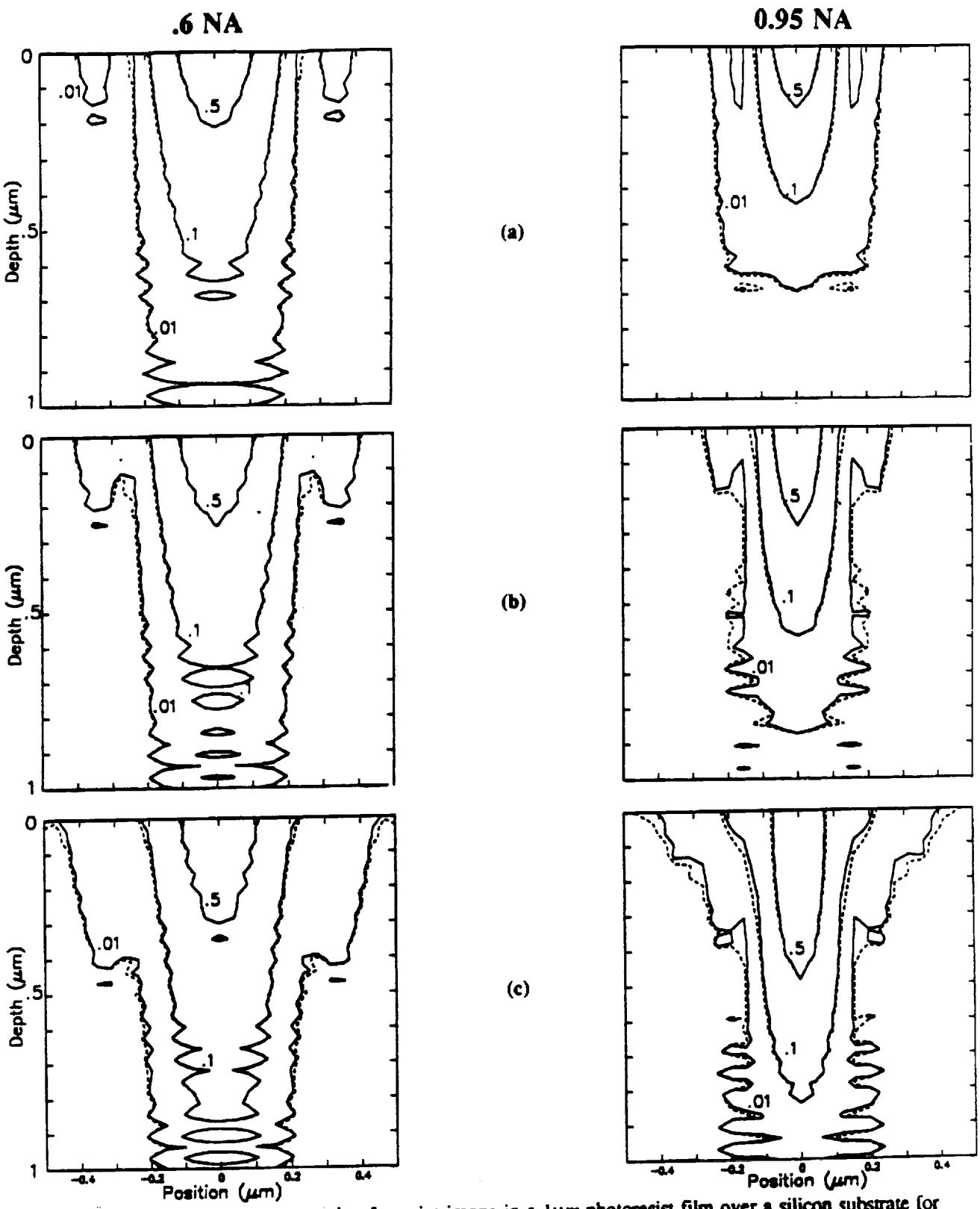


Fig. 7. Iso-irradiance contours in of a point image in a 1μm photoresist film over a silicon substrate for 0.6 and 0.95 NA. The source is initially y polarized and $\lambda = 248nm$ The solid line is a meridional slice along the y axis and the dotted line is a slice along the x axis. Marginal ray focus at (a) 0μm depth, (b) 0.5 μm depth, (c) 1.0μm depth.

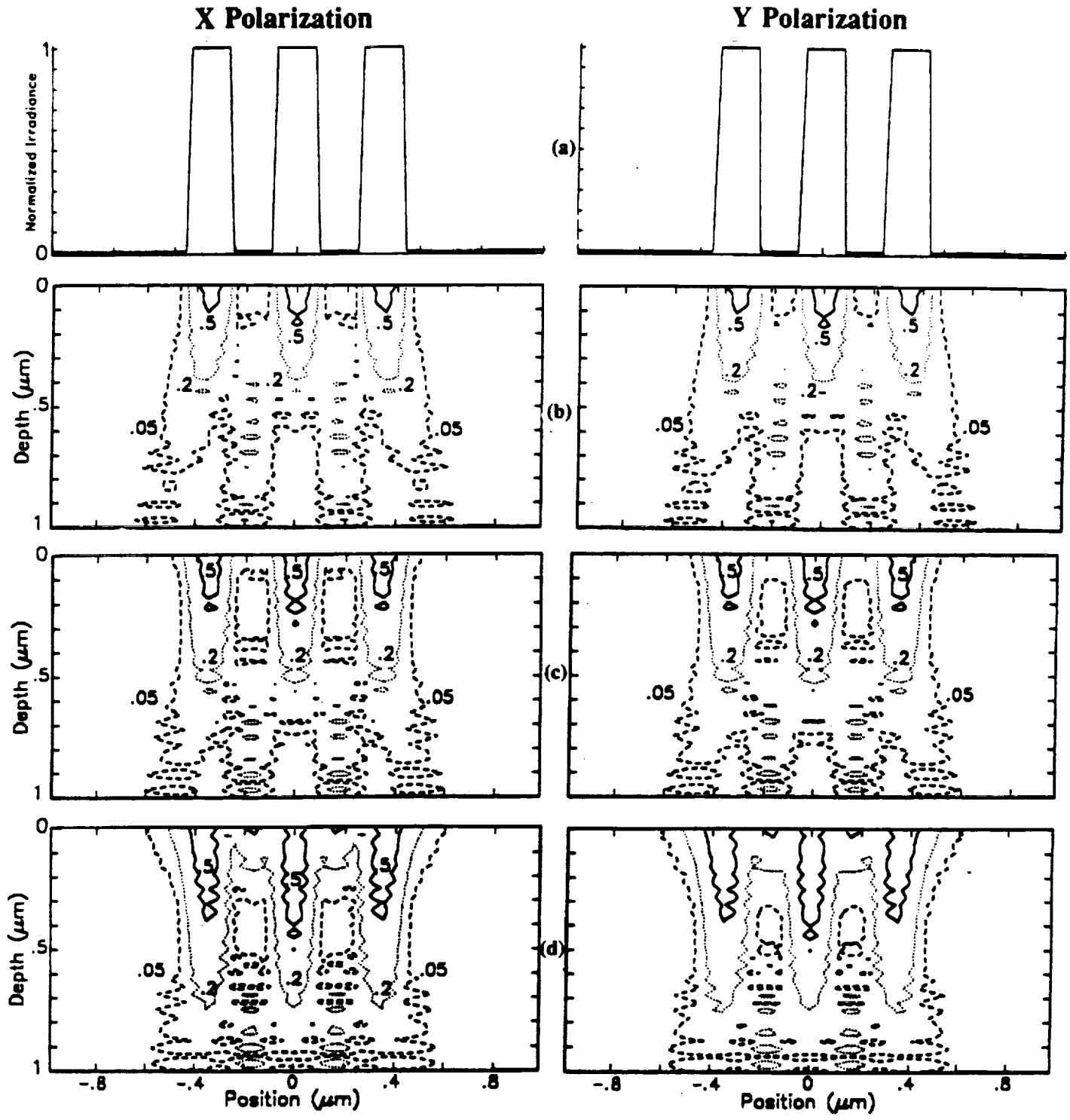


Fig. 8. (a) Transmission function of tri-bar object with lines and spaces parallel to x axis. y meridional slice images of 0.95 NA system with initial x and y polarizations for marginal ray focus at (b) 0μm depth, (c) 0.5μm depth, and (d) 1.0μm depth.

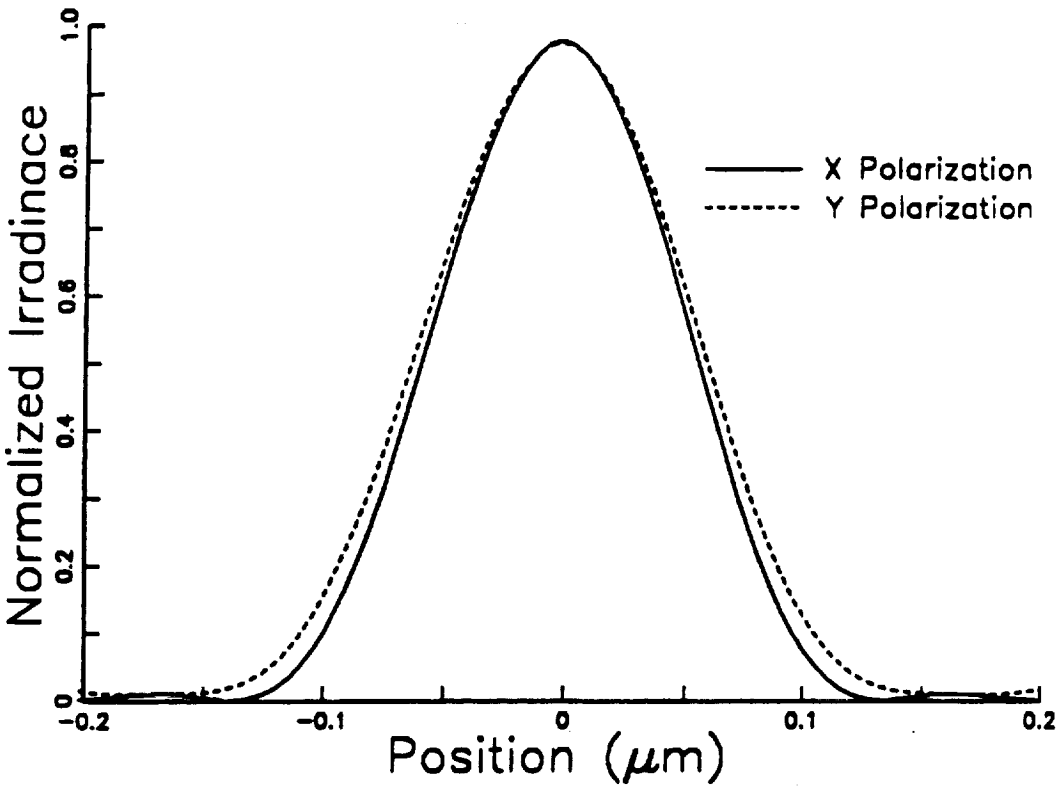


Fig. 9. Plots of irradiance for x and y polarizations at the top surface of the film for a zero offset using the parameters in Figure 8b.

Table I

$\vec{P}(\alpha, \beta)$ Components	Initial Polarization State X	Y
P_{xS}	$\frac{\beta^2}{1-\gamma^2}$	$\frac{-\alpha\beta}{1-\gamma^2}$
P_{xP}	$\frac{\gamma\alpha^2}{1-\gamma^2}$	$\frac{\alpha\beta\gamma}{1-\gamma^2}$
P_x	$1-\frac{\alpha^2}{1+\gamma}$	$\frac{-\alpha\beta}{1+\gamma}$
P_{yS}	$\frac{-\alpha\beta}{1-\gamma^2}$	$\frac{\alpha^2}{1-\gamma^2}$
P_{yP}	$\frac{\alpha\beta\gamma}{1-\gamma^2}$	$\frac{\gamma\beta^2}{1-\gamma^2}$
P_y	$\frac{-\alpha\beta}{1+\gamma}$	$1-\frac{\beta^2}{1+\gamma}$
P_{zP}	$-\alpha$	$-\beta$
P_z	$-\alpha$	$-\beta$

APPENDIX I

Modeling and measurement of a micro-optic beam deflector

Tom D. Milster

University of Arizona, Optical Sciences Center
Tucson, Arizona 85721

J. Nan Wong

Lawrence Livermore National Laboratory
Livermore, California 94550

ABSTRACT

We investigate the use of a unity-magnification micro-optic beam deflector. The deflector consists of two arrays of positively powered lenslets. The lenslets on each array are arranged in a square grid. Design criteria are based on usefulness in optical data storage devices. The deflector is designed to operate over a ± 1.6° range of deflection angles. We compare modeling results with interferometric analysis of the wavefront from a single lenslet pair. Our results indicate that the device is nearly diffraction limited, but there are substantial wavefront errors at the edges and corners of the lenslets.

1. INTRODUCTION

There are many ways to achieve fine position control in optical data storage systems. Most rely on the translational movement of an objective lens. Other methods rely on changing the angular position of the beam before the objective lens, as shown in Figure 1. Angular beam deflection for data storage has been discussed with respect to galvanometers[1,2] and acousto-optic cells[3]. There are other possibilities, which include prisms, electro-optic cells and holograms, but they are not well suited for data storage due to either performance limitations, massive components, or complicated electronics. Galvonometer mirrors provide the necessary function and achieve good bandwidth, but they have two disadvantages. First, they disturb the beam path beyond what is required for deflection at the objective lens. This is not a performance limitation, but it is a practical one. The beam path in an optical data storage device must be as compact as possible. Any unnecessary deviations in the beam path result in extra real estate that must be used to satisfy design requirements. Secondly, reflections off of a mirrored surface require that the surface be of extremely high quality. The surface quality is difficult to achieve in a lightweight mirror, which is a requirement for high bandwidth. In addition, as the lightweight mirror is rotated, vibrational modes of the mirror surface can affect beam quality. These considerations can be overcome with good engineering, but additional costs associated with the components are likely. There are also disadvantages to using acousto-optic cells for beam deflection. The primary disadvantages are cost and complexity of the electronics. What we seek is a low-cost alternative to these techniques that does not does not disturb the beam path beyond what is necessary for fine positioning.

An alternative technique for beam steering is to use two lenses arranged as an afocal telescope, as shown in Figure 2A. As one of the lenses is translated, the output beam is

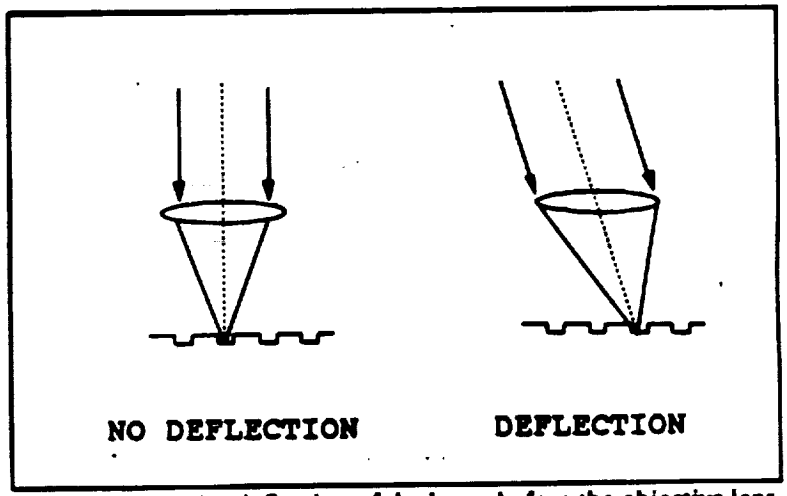


Figure 1. Angular deflection of the beam before the objective lens.

deflected. In order to avoid problems with aberrations, the lenses must operate at relatively high f/#, which results in a large form factor. The lenses are massive and require significant power from an actuator to achieve high bandwidth. The range of deflection is another disadvantage of this system. Since the f/# of the objective lens is typically f/1 and the telescope must be f/5 or greater, the translation of beam steering lens must be at least five times the distance required at the disk plane. As described, the afocal relay deflector is not useful as a beam steering device for optical data storage.

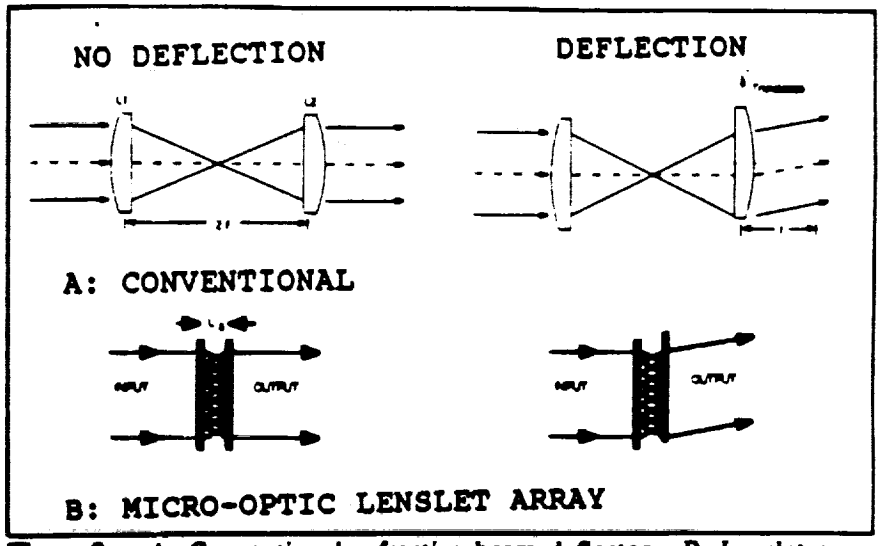


Figure 2. A: Conventional refractive beam deflector. B: Lenslet array beam deflector.

Instead of using two large lenses, we suggest using two arrays of micro-optic lenses, as shown in Figure 2B. Each lenslet pair performs as the afocal telescope in Figure 2A. The focal lengths of the lenslets are much shorter than the two-lens system, so the device is much more compact. The mass of a lenslet array is significantly less than the mass of a single lens, so high bandwidth can be achieved. The translation of the micro-lens array is much smaller than that required in the two-lens system. A similar device was discussed by Goltsos and Holz[4], as shown in Figure 3. Instead of two positively powered arrays, they use one positively powered array and one negatively powered array. As one array is translated relative to the other, the output beam is deflected. This device does not satisfy our requirement for unity magnification through the system. Also, the positive-negative system does not fill the output beam, which results in a complicated diffraction pattern at the disk plane. Conversely, the positive-positive system uniformly fills the output with unity magnification.

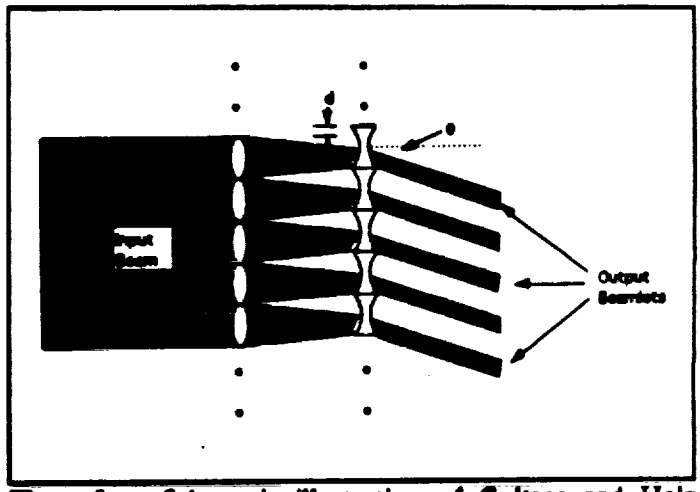


Figure 3. Schematic illustration of Goltsos and Holz device.

In the following paragraphs, we describe our device in detail. First, we describe system requirements for an optical data storage system. CODE V modeling is used to predict performance of the device. We then discuss boundary losses and the implementation of field lenses to limit these losses. Diffraction calculations are used to predict the pattern observed on the disk plane. Our experimental data include interferometric analysis of the lenslets. Finally, we draw conclusions from our work.

2. THEORY AND MODELING

Table I lists the operating requirements of the beam steering device. These numbers are based on typical values for the wavelength and aperture size used in optical data storage devices. The ± 1.6° range of deflection angles corresponds to about ± 120 μm of fine positioning at the disk plane with an f/1 objective lens. If the track pitch is 1.6 μm, ± 75 tracks can be accessed with the beam steering device. The field angle

requirement of ± 0.5° is to allow for some alignment tolerance.

We tested Monolithic Lenslet Modules[3] (MLMs), which are refractive microlens arrays in close-packed geometries. Our MLM arrays consisted of 400μm square lenslets on a grid spaced 400μm apart. Lenslets are formed in a thin layer of a single-part epoxy that has been molded on a substrate. In our case, the substrate material was a 1.27 mm thick piece of soda lime glass. The lenslets resemble simple plano-convex lenses. The advertised focal length for the arrays is 3.20 mm at λ = 633 nm. Our estimate of the focal length at λ = 830 nm is 3.27 mm. The f/# is 5.8 if one uses the diagonal of the lenslet in the calculation. Since the arrays are refractive and not diffractive, we expect high transmission efficiency and relatively little sensitivity to small wavelength variations. The depth of focus, which is given by

Table I. System design parameters.

INPUT BEAM SIZE (DIAMETER)	4.3 mm
OUTPUT BEAM SIZE (DIAMETER)	4.3 mm
WAVELENGTH	830 ±20 nm
RANGE OF DEFLECTION ANGLES	±1.6°
FIELD ANGLE	±0.5°

$$\Delta z = \pm 2\lambda(f/\#)^2 \quad ,$$

calculates to ± 56 μm at λ = 830 nm. The required translation for 1.6° is 91 μm.

CODE V was used to model the positive-positive array system. The lenslets were arranged as shown in Figure 2A, with the curved side toward the infinite conjugate. Wavefront profiles are shown in Figure 4 for the deflection system operating at λ = 633 nm. Figure 4A displays performance with no deflection. Figure 4B displays performance at 1.6°. In both cases, superb wavefronts of less than 0.05 waves peak to valley are predicted, which implies that diffraction-limited performance will be achieved.

The positive-positive system has one major loss mechanism. As the array is translated, the light that is incident on adjacent lenslets is lost, as shown in Figure 5. The boundary losses can be calculated by assuming square lenslets. The transmission efficiency of the system is given by

$$\eta = 1 - \sqrt{2}\, f/\# \tan\theta \quad , \tag{1}$$

where θ is the deflection angle. Therefore, we desire a low f/# in order to limit boundary losses.

Boundary losses can be virtually eliminated by using a field lens array, as shown in Figure 6. The field array is translated in unison with the last array. The field lenslets redirect the diverging beamlets so that they are centered and optimally fill the lenslets of the last array. The transmission efficiency of this system is given by

$$\eta' = 1 - f/\# \tan^2\theta \quad . \tag{2}$$

A comparison of the transmission factors resulting from Equations 1 and 2 is given in Table II. For all field

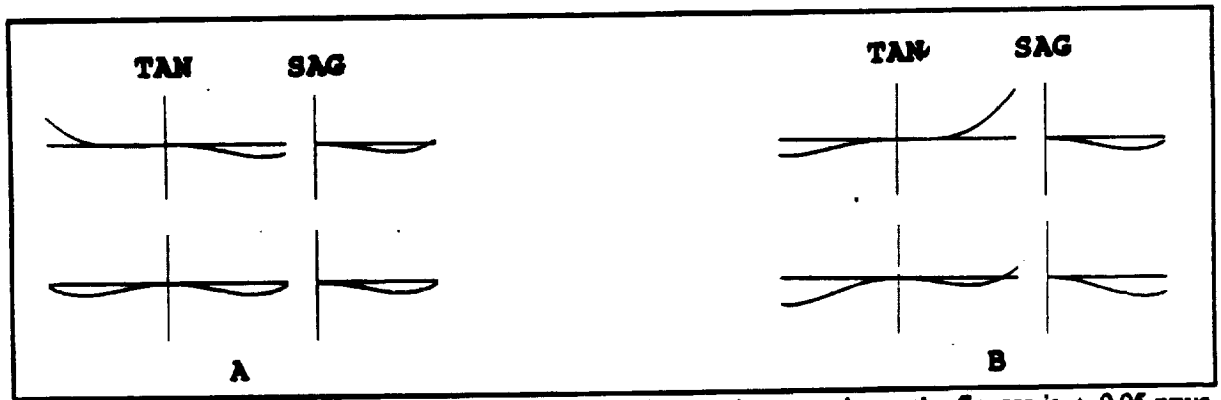


Figure 4. Wave fans resulting from a single lenslet. The maximum scale on the figures is ± 0.05 wave. Upper traces refer to performance at the maximum field angle. The arrays are set for: A) no deflection, and B) 1.6° deflection.

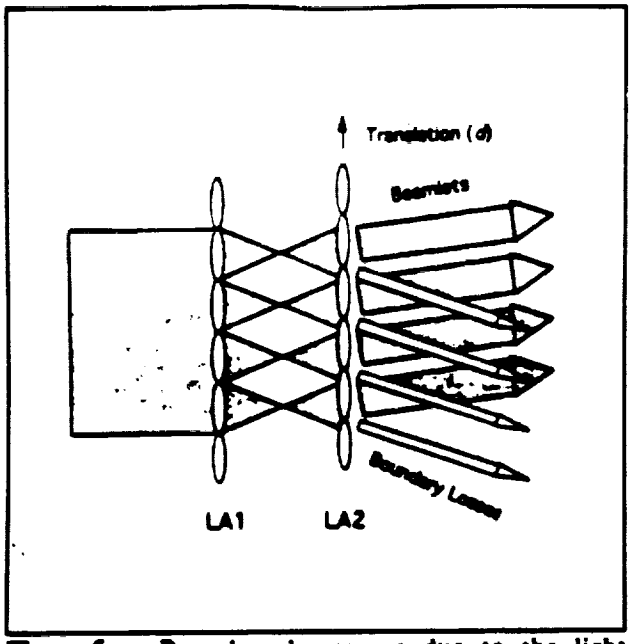


Figure 5. Boundary losses are due to the light incident on adjacent lenslets.

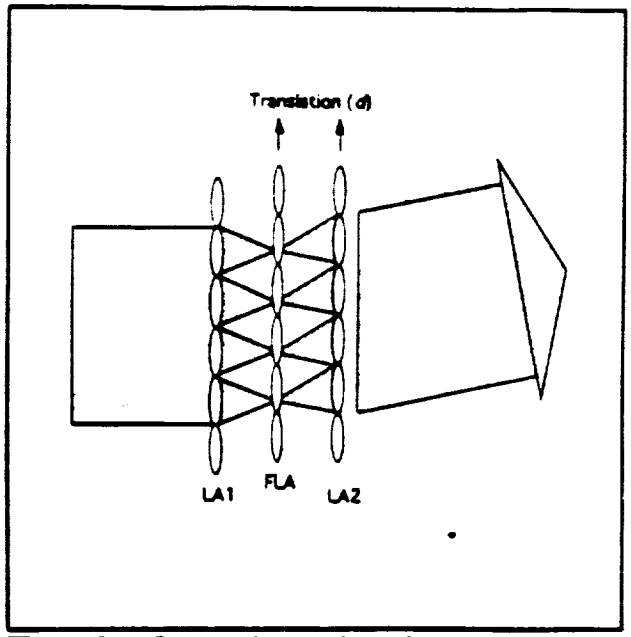


Figure 6. Losses due to boundary losses can be avoided by using a field-lens array. The field lens travels with LA2, the last array.

angles of interest, the addition of the field array significantly improves the transmission efficiency.

Another consideration for modeling is the diffraction pattern observed at the disk plane. The diffraction from each square lenslet results in a $(\sin\pi x)/\pi x$ distribution. Since the lenslets are spaced at a distance equal to their width, the zeros of the sinc function will coincide with higher orders of the diffraction pattern. Therefore, very little diffracted light should be observed. Figure 7 displays the result of a diffraction calculation for our system.

Table II. Efficiencies

Angle (°)	η	η′
1.0	0.86	0.99
2.0	0.71	0.99
4.0	0.43	0.97

3. EXPERIMENT

A layout of the interferometric experiment is shown in Figure 8. Collimated light from a 780 nm laser diode passed through the MLMs. A unity-magnification afocal telescope was used to image LA2, the last lenslet array, into the 8X beam expander. A Laser Diode Interferometric Tester (LADITE) from WYKO corporation was used for the interferometric analysis. The LADITE is a phase shifting and self referencing interferometer. The image of the array from the beam expander was placed at the entrance pupil of the LADITE.

The fringe pattern observed in one of the measurements is shown in Figure 9. The boundary of the lenslets are clearly visible, and straight fringes are observed over a large percentage of the lenslets. The corner areas display a diamond-like pattern that does not give useful lens action. A dust spec is observed in the upper center lenslet. In general, very careful procedures were necessary in order to limit

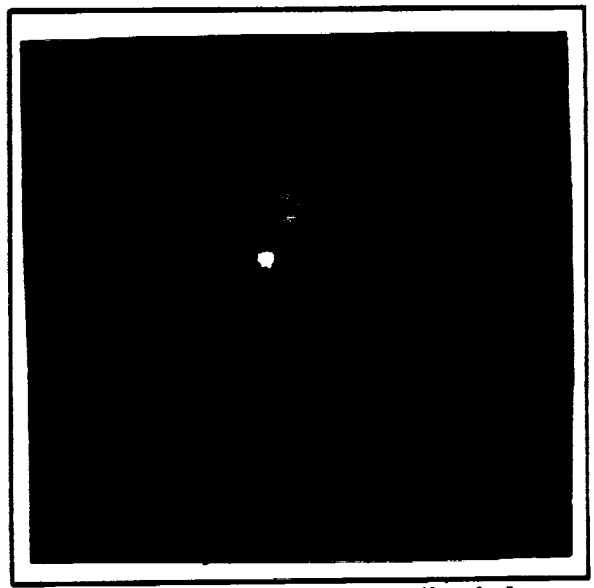


Figure 7. Far-field pattern predicted from a diffraction calculation of the lenslet arrays.

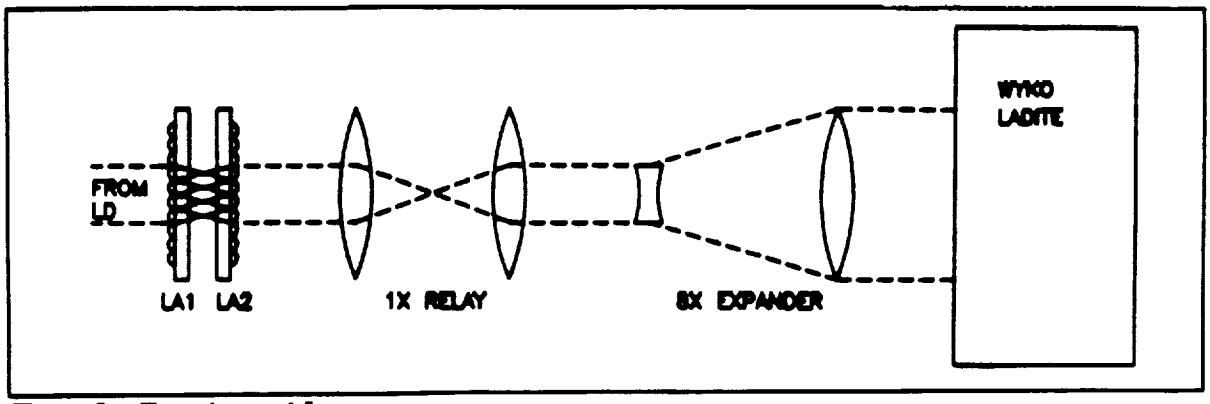


Figure 8. Experimental Layout.

contamination of the array. A small dust spec of 25 μm to 50 μm diameter could significantly affect the performance out of a lenslet. The epoxy material was also very easily scratched.

Before the MLM arrays were placed into the system, we obtained a reference map of the other optical components, which is subtracted from the data on the lenslets. When interferometric data were taken, one lenslets were masked individually. The best wavefront obtained for a lenslet was about 0.08 waves rms. Upon further investigation, we found that most of the wavefront error was located at the edge of the lenslets. Figure 10 displays an OPD map of the surface. The largest variations are observed in the corners and along the edges. Figure 11 displays an OPD profile of one edge of the lenslet. A large departure is observed near one corner of Figure 11. To illustrate this point further, Figure 12 displays the variation of wavefront quality observed on a typical lenslet when the mask size is reduced from the full aperture. Both a square and a round mask were used. It is clear that diffraction-limited performance can easily be achieved if performance in the corners and along the edges can be improved. We observed no significant change in wavefront characteristics as one array was translated up to 50 μm relative to the other.

Figure 13 displays the beam deflector illuminated over the full 4.3 mm. We are in the process of evaluating the far-field pattern from this device.

Figure 9. Fringe observed through the LADITE.

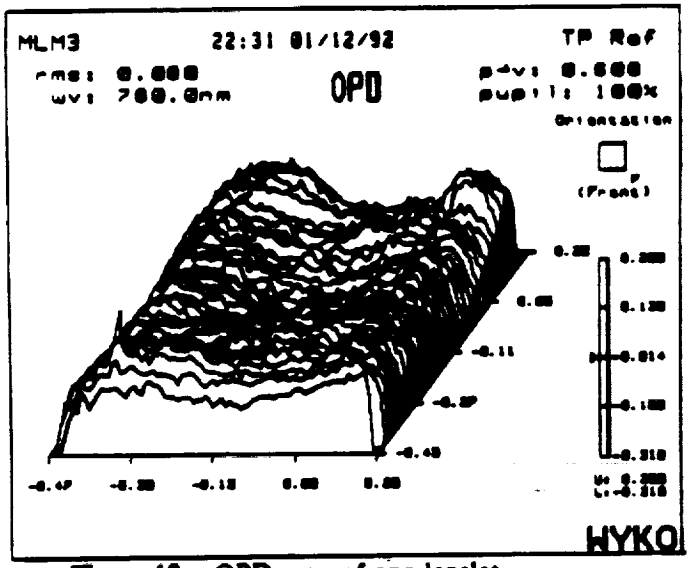


Figure 10. OPD map of one lenslet.

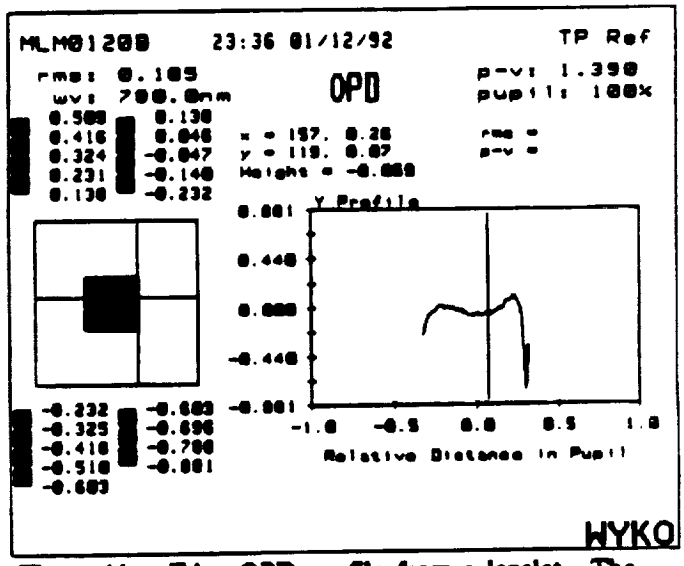


Figure 11. Edge OPD profile from a lenslet. The peak-to-peak variation at the edge is about 1.4 waves.

4. CONCLUSIONS

We have analyzed a beam deflector that consists of two arrays of refractive micro lenses. One array is translated in order to achieve deflection. Design analysis indicates that diffraction-limited performance is obtainable over a ± 1.6° range of deflection angles. Boundary losses can be eliminated with the use of a micro-optic field lens array. Diffraction analysis indicates that a bright central spot should be observed in the far-field pattern. Measurement of the wavefront quality from individual lenslets indicates that the lenslets are of good wavefront quality. However, the edges and corners of the lenslets have significant aberration. We are in the process of evaluating the far-field pattern from a 4.3 mm diameter beam passing through the beam deflector.

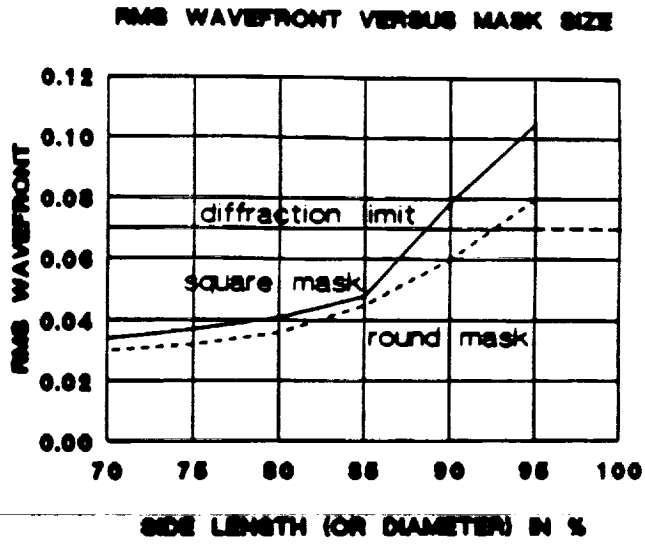


Figure 12. Wavefront variation as a function of mask diameter centered on the lenslet. The horizontal axis corresponds to the length of the mask: A) across one side of the square mask, B) across the diameter of the circular mask.

5. ACKNOWLEDGEMENTS

The authors wish to thank Kevin Erwin and Mark S. Wang, both of the Optical Sciences Center at the University of Arizona, for their help in acquiring data for this report. We give our sincere thanks to Steve Martinick at WYKO Corporation for loaning us the LADITE. This work was partially sponsored by a grant from the Joint Services Optical Program, contract #F49620-91-C-0009. Additional support was provided by the Optical Data Storage Center, and student support was provided for J.N. Wong from Lawrence Livermore National Laboratory.

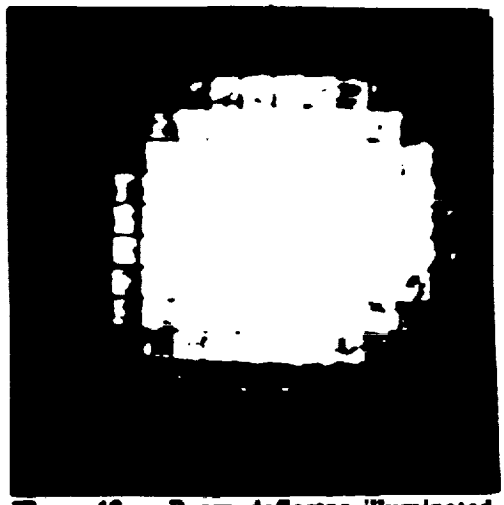


Figure 13. Beam deflector illuminated with a 4.3 mm laser beam.

6. REFERENCES

1. D.K. Towner, "Scanning techniques for optical data storage," *SPIE Proceedings: Optical Mass Data Storage II*, vol. 695, pp. 172-180, 1986.
2. A. Takahashi, H. Ohsawa, N. Seo, T. Takahara, H. Yaamasaki, and W. Shinohara, " High speed accessing of 90 mm optical disk drive," *SPIE Proceedings: Optical Data Storage*, vol. 1316, 1990.
3. K. Matsumoto and T. Maeda, "Acousto-optic accessing in optical disks," *Jap. J. Appl. Phys.*, vol. 28, supplement 28-3, pp. 335-340, 1989.
4. W. Goltsos and M. Holz, "Agile beam steering using binary optics microlens array," *Optical Engineering*, vol. 29, no. 11, pp 1392-1397, 1990.
5. Adaptive Optics Associates, commercial literature.

APPENDIX J

Effects of higher order aberrations on beam shape in an optical recording system

Mark S. Wang and Tom D. Milster
Optical Sciences Center, University of Arizona
Tucson, AZ 85721

Abstract

An unexpected irradiance pattern in the detector plane of an optical data storage system was observed. Through wavefront measurement and scalar diffraction modeling, we discovered that the energy redistribution is due to residual third-order and fifth-order spherical aberration of the objective lens and cover-plate assembly. The amount of residual aberration is small, and the beam focused on the disk would be considered diffraction limited by several criteria. Since the detector is not in the focal plane, even this small amount of aberration has a significant effect on the energy distribution. We show that the energy redistribution can adversely affect focus error signals, which are responsible for maintaining sub-micron spot diameters on the spinning disk.

In optical data storage systems, focusing and tracking servo controls are necessary to keep the micron-sized laser spot in focus and aligned with the proper track as the disk spins. Focusing and tracking error signals are commonly derived from the beam reflected from the optical disk[1]. The reflected beam exhibits a peak-to-peak wavefront deviation less than 0.25 wave and a rms deviation less than 0.07 wave, which is essentially diffraction limited. In our focus-error detection technique, the reflected beam is refocused by an auxiliary lens, and a detector is placed just outside of the auxiliary-lens focus[2]. The observed irradiance distribution on the detector plane is significantly different from that predicted by assuming that the beam has no aberration. The cause of this effect is mainly residual higher-order spherical aberration that redistributes the Gaussian beam energy into the caustic region. The energy redistribution can affect focus-error signals.

In this paper, we first present scalar diffraction modeling results of the reflected beam on the servo detector. Then we will show wavefront measurement results at various planes in the optical system. Finally, wavefront measurement and diffraction modeling are combined to explain our

experimental observation of the irradiance pattern at the focus-servo detector plane.

In Figure 1, the layout of the head optics is drawn in solid lines. The laser diode is a Sharp LT024 operating at 780 nm with beam divergences of 10° and 29° (FWHM) parallel and perpendicular to the junction, respectively. The laser diode output is collimated by a lens with a focal length of 6.25 mm and an aperture of 5.0 mm. The collimated beam is then circularized by an anamorphic prism pair with a magnification of 3x. The beam passes through a partially polarizing beam splitter (PPBS) and is focused onto the disk by an Olympus AV4350-3 objective lens mounted on TAOH-PB7 actuator. The objective lens is a glass triplet with designed rms aberration of $< 0.018\lambda$ on axis and $< 0.05\lambda$ at 100 μm field. The numerical aperture (NA) of the objective is 0.5 and the aperture diameter is 4.3 mm, which truncates the Gaussian beam profile at 0.76 of its $1/e^2$ width. The aperture of the objective lens is the stop for the optical system. The light reflected from the disk is re-collimated by the objective lens and directed to the detection optics by the PPBS. The equivalent air path length from the objective lens to the plano-convex detector lens is 150 mm. The

detector lens has a 124 mm focal length and focuses the light onto quad detector #2, which is 17.62 mm outside its focal plane.

A physical-optics analysis is accomplished through modeling the amplitude and phase of the laser beam's electric field[3]. Beam propagation is simulated through the use of Fresnel diffraction calculations[4]. In this model, a scalar diffraction approach is used. The beam is represented by a matrix of complex values, where each matrix element represents a different sample point in the wavefront. The complex values represent the amplitude and phase of the wavefront at the given point. Fresnel diffraction is computed with Fast Fourier Transforms (FFTs). Propagation of collimated beams between lenses is modeled using the angular spectrum method. Aberrations are represented by Zernike polynomials.

In our optical system there are several sources of aberration. The beam emitted from laser diode has astigmatism[5]. The collimating lens is carefully aligned to balance the laser diode astigmatism[6]. For light passing through, the prism pair and PPBS are just plane parallel plates that generate almost no aberration. The objective lens is designed to

compensate aberration caused by the disk cover glass, but it is not aberration free. Actually, the objective lens and cover plate are the main aberration sources in this system. Aberrations are induced when the beam is focused through the objective lens and as the beam is passed through the cover plate. When the beam is reflected back, odd aberrations are canceled and even aberrations are doubled. Another source of aberration is when the PPBS reflects light to detection optics. The reflecting surface generates astigmatism. The amount of aberration generated depends on the quality of the beam splitter. The detector lens and PBS have very little contribution to total aberration since they are used at very small NA.

In our model, aberrations are added in the entrance and exit pupil of the objective lens. Contributions of other components are traced back to either pupil. The amount of aberration refers to that of single pass.

We now discuss our modeling results. The beam is propagated from the collimating lens, through the objective lens, down to the disk, back through the objective lens, and to the detector plane. Line

profiles of the irradiance pattern at the detector plane are shown in Figure 2 for several situations. All the line profiles are normalized to the peak at the center. No aberration is present and the disk is in focus for Figure 2a. The edge of the intensity irradiance pattern decays like a Gaussian beam. The center shows diffraction rings caused by the stop. The beam profile is the image of Fresnel diffraction pattern in a plane 997 mm away from the detector lens, i.e., 847 mm from the stop. The stop size is 2.15 mm in radius, so the Fresnel number is:

$$\frac{r^2}{\lambda Z} = \frac{(2.15\times10^{-3})^2}{0.78\times10^{-6}\times0.847} = 7.0 \qquad (1)$$

The peak at the center is very clear. Figure 2b shows the line profile when -0.25 wave of balanced third-order spherical aberration is added at the objective pupil. Balanced third-order spherical aberration means that an appropriate amount of defocus has been added to minimize rms wavefront variance, as described in the eighth Zernike aberration polynomial. The pattern is wider as the spherical aberration spreads light to the edge of the irradiance pattern. In Figure 2c, -0.125 wave of balanced astigmatism, i.e., fourth Zernike polynomial, is introduced. The X-direction line

profile is shown as a solid line, while that in Y direction is shown as a dashed line. The line profile in Figure 2d has -1.0 waves of fifth-order balanced spherical aberration, i.e., a fifteenth Zernike coefficient of -0.1. The light is accumulated at the center by the fifth-order spherical and the balancing defocus while the balancing third-order spherical spreads light toward the edge. Consequently, a strong central peak and sidelobes are formed. Recall that the even aberration incident onto the detector lens is only half of that incident onto the disk. Thus, even if the focusing spot on the disk is well under the diffraction limit, the intensity pattern at the detector plane can be strongly modified.

We measured the wavefront at several planes with a LADITE laser-diode wavefront tester manufactured by WYKO Corporation. The objective lens was removed so that the entrance pupil of LADITE was located at the stop of the objective lens. The rms wavefront error was found to be 0.024λ and the Strehl Ratio was 0.985 as shown in Figure 3a. It was well collimated and diffraction limited[6].

We measured the PPBS separately. The objective lens was removed and a reference source was set in

the path of the reflected beam. The reference source was a laser diode source spatially filtered to be virtually aberration free. It was aligned and centered in the LADITE. This wavefront contained only aberration of the PPBS. We found that the PPBS had about 0.5 wave of astigmatism over the clear aperture.

The wavefront reflected from the disk was measured using the setup shown in Figure 1 with dashed lines. A high-quality first-surface mirror after the PPBS reflected the beam into a relay lens. The relay lens was used to relay the pupil of the objective lens to the entrance pupil of LADITE in order to avoid errors due to diffraction propagation. The relay lens consisted of two identical cemented doublets that form a 4f system, where the entrance and exit pupil of the relay lens was separated by four times the focal length of single lens. The relay lens was tested separately with the LADITE and exhibited only 0.012 wave rms aberration.

After data were acquired for the reflected wavefront after the PPBS, the wavefront of the PPBS was subtracted. This gives the wavefront before the PPBS that is due to a combination of laser optics and

objective/cover glass aberrations, as shown in Figure 3b. The aberration is mostly fifth-order spherical balanced with third-order spherical and defocus. The rms wavefront aberration is 0.054.

Recall that even aberrations are doubled in double pass. If the objective lens is well aligned, odd aberrations, like coma, are eliminated. The rms wavefront aberration focused down to the optical disk should be about half of the measured rms value of 0.054. This implies that the focused spot is definitely diffraction limited.

Reflected wavefronts after the PPBS are shown as 3D plot in Figure 3c and as X-profile in Figure 3d. The discrepancy between X and Y wave aberration fan is caused by the astigmatism of PPBS. The rms wavefront error is 0.079 wave, which is much worse than that before the PPBS. It is primarily a mixture of fifth-order spherical and astigmatism. Table I lists the first fifteen Zernike coefficients $Z_1(n)$ of the objective lens aberration and $Z_2(n)$ of the entire reflected wavefront. The first three polynomials, which represent tilt and defocus, are removed. As we discussed previously, the single-path aberration of the objective lens is half of the total, i.e., $Z_1/2$.

The aberration of the reflecting path includes contribution of objective lens $Z_1/2$ and beam splitter (Z_2-Z_1), the total is $(Z_2-Z_1/2)$.

A linear CCD array (CCD123 by Fairchild Weston Systems, 13μm by 10μm on 10μm pitch, dynamic range 5500:1) with 1728 elements was set in the detector plane to obtain line profiles of the irradiance pattern. Figure 4a shows the observed line profile. If no aberration were present, the edge should fall rapidly, as predicted in Figure 2a. However, a sidelobe shows up around a sharp central peak. Third-order aberration cannot produce this sidelobe and sharp central peak. The line profile was quite close to that predicted using fifth-order spherical aberration, as shown in Figure 2d. To model this line profile, we used the measured Zernike coefficients of Table I. Aberration of the objective before hitting the disk $(Z_1/2)$ is added in the entrance pupil of the objective, and aberration of the reflecting path $(Z_2-Z_1/2)$ is added in the exit pupil of the objective lens. The model was used to propagate the wavefront from objective lens down to the detector plane. The modeling result is shown in Figure 4b. A good agreement between modeled and measured beam profiles is achieved.

The Focusing Error Signal (FES) is generated from detector 2 using offset spot size method[2]. A typical FES vs defocus plot is shown in Figure 5. The solid line shows FES without aberration, while the dashed line shows FES with measured aberration. The aberrated FES has an offset at the center. The linearity of the aberrated FES is also changed significantly due to the presence of higher-order aberrations. The focus offset can produce a false focus. Nonlinearity decreases the performance of the servo system.

In conclusion, we reported an unexpected irradiance distribution in the detector plane of an optical data storage system. Through wavefront measurement and scalar diffraction modeling, we discovered that the energy redistribution was due to residual amounts of third-order and fifth-order spherical aberration of the objective lens and cover plate assembly. A significant amount of energy redistribution was observed, even though the beam was essentially diffraction limited.

The authors thank the Optical Data Storage Center at University of Arizona for financial support and WYKO Corporation for lending us LADITE

interferometer. We also wish to thank Steve Martinek of WYKO and Kevin J. Erwin of Optical Sciences Center for their technical support.

References

1. A.B. Marchant, Optical Recording, Addison-Wesley Publishing Co., New York, 1990, p. 8.

2. T.D. Milster, M.S. Wang, F.F. Froehlich, J.L. Kann, J.P. Treptau, and J.K. Erwin, "Differential Spot-Size Focus Servo", Proc. SPIE v. 1499 (1991).

3. J.P. Treptau, T.D. Milster, D.G. Flagello, "Laser Beam Modeling in Optical Storage Systems", Proc. SPIE v. 1415 (1991).

4. J.W. Goodman, Introduction To Fourier Optics, McGraw-Hill, Inc., 1968.

5. D.D. Cook, F.R. Nash, "Gain-induced guiding and astigmatic output of GaAs lasers", Journal of Applied Physics v. 46, no. 4 (1975).

6. F.F. Froehlich, M.S. Wang, and T.D. Milster, "Technique for simultaneous alignment and collimation of a laser diode in an optical data storage head", Appl. Opt. Vol. 30, No. 31 4481 (1991).

Tables

1. List of measured Zernike coefficients. $Z_1(n)$: aberration of the objective lens. $Z_2(n)$: Reflected wavefront.

n	$Z_1(n)$	$Z_2(n)$	n	$Z_1(n)$	$Z_2(n)$	n	$Z_1(n)$	$Z_2(n)$
1	0.000	0.000	6	0.008	-0.007	11	-0.019	-0.028
2	0.000	0.000	7	0.008	0.009	12	-0.014	-0.011
3	0.000	0.000	8	0.020	0.016	13	0.014	-0.001
4	0.032	0.143	9	-0.007	-0.014	14	0.033	0.031
5	0.021	0.036	10	-0.011	-0.007	15	-0.114	-0.111

Figure Captions

Figure 1. Optical system of a magneto-optical testbed. Testing path is drawn in dashed line.

Figure 2. Line profiles at detector plane when certain aberration is present: 2a. no aberration; 2b. -0.25 wave of spherical aberration; 2c. -0.125 wave of astigmatism in X (solid line) and Y (dashed line); 2d. -1.0 waves of fifth order spherical.

Figure 3. Measured wavefront. 3a. at objective lens; 3b. at objective lens, before PPBS; 3c. three dimensional wavefront, after PPBS; 3d. after PPBS in X direction.

Figure 4. 4a. Line profile at detector plane by a linear CCD array; 4b. prediction by using measured Zernike coefficients. Solid line: X direction. Dashed line: Y direction.

Figure 5. Calculated Focusing Error Signal (FES), solid line: without aberration; dashed line: with measured aberration.

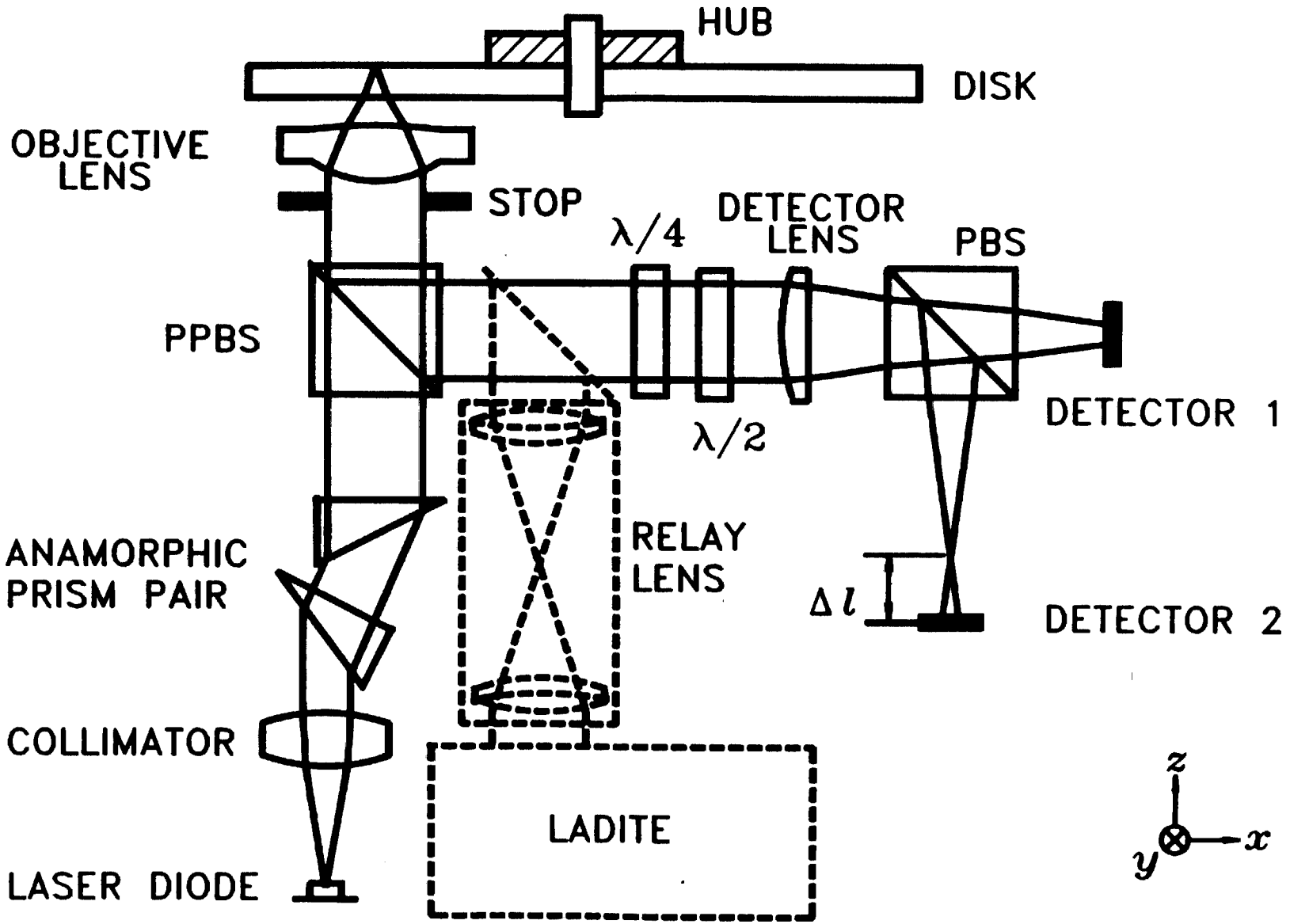

HUB
DISK
OBJECTIVE LENS
STOP
λ/4
DETECTOR LENS
PBS
PPBS
DETECTOR 1
λ/2
ANAMORPHIC PRISM PAIR
RELAY LENS
Δl
DETECTOR 2
COLLIMATOR
LADITE
LASER DIODE
z
y
x

Figure 1.

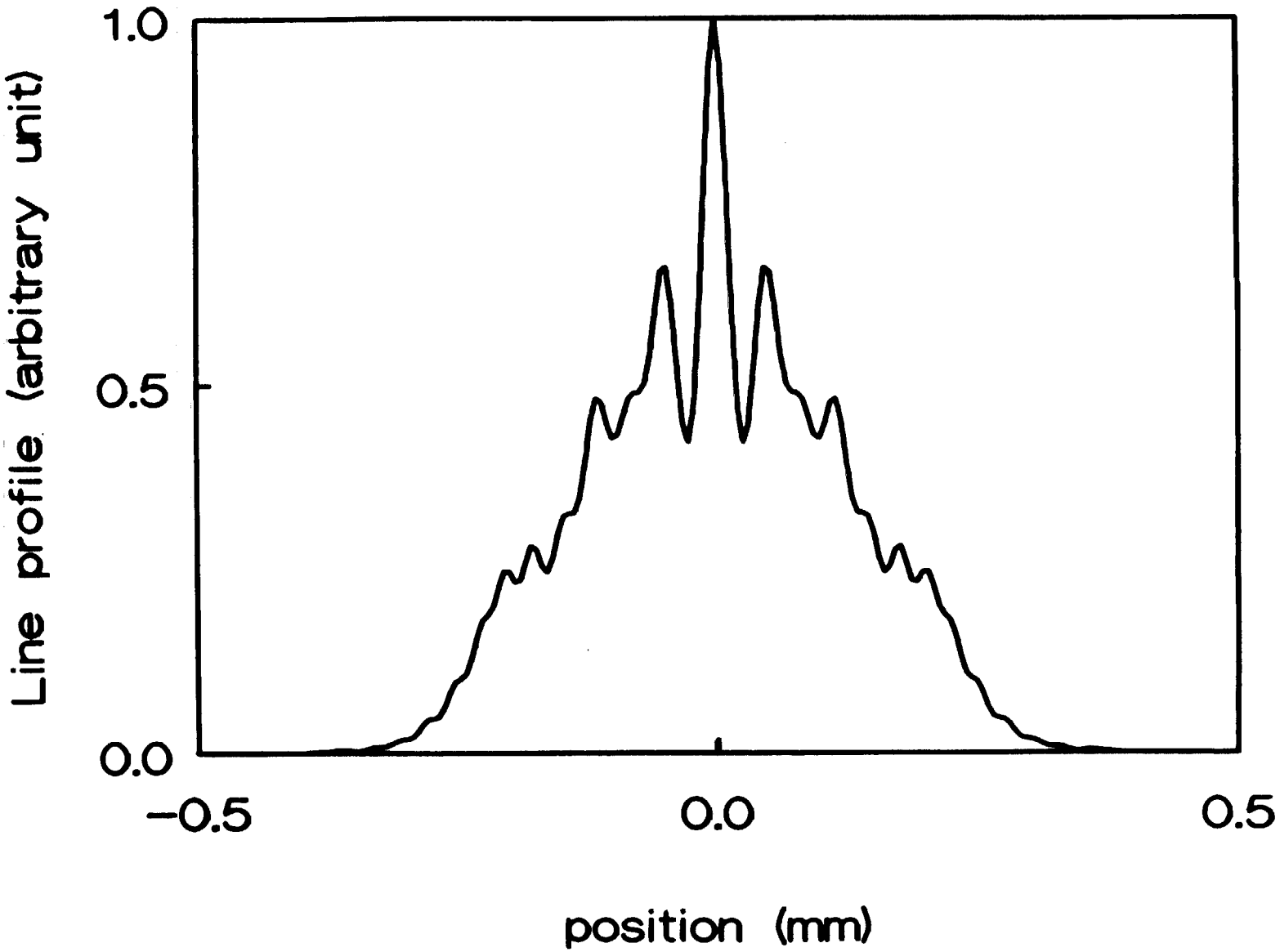


Figure 2a.

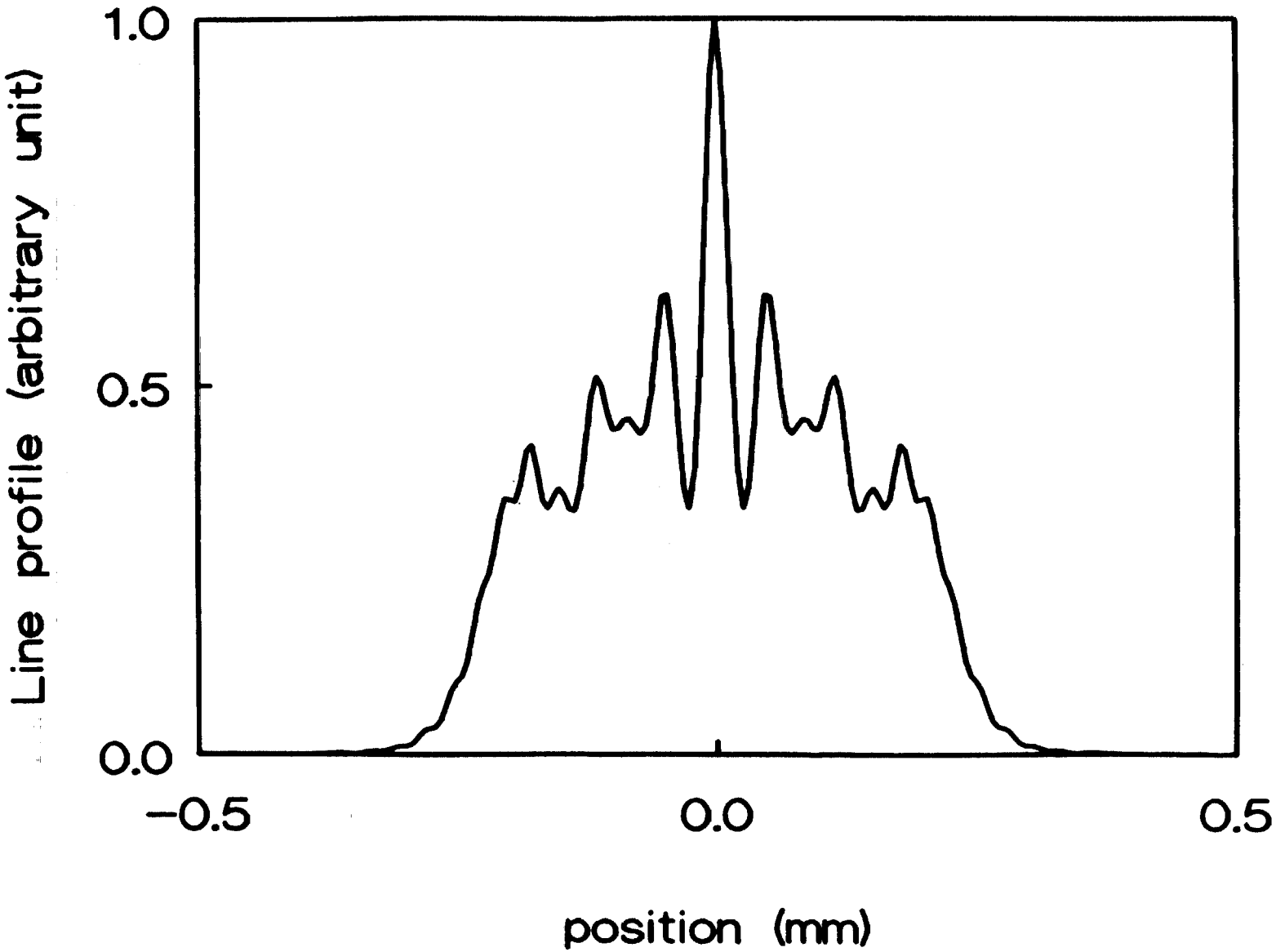


Figure 2b.

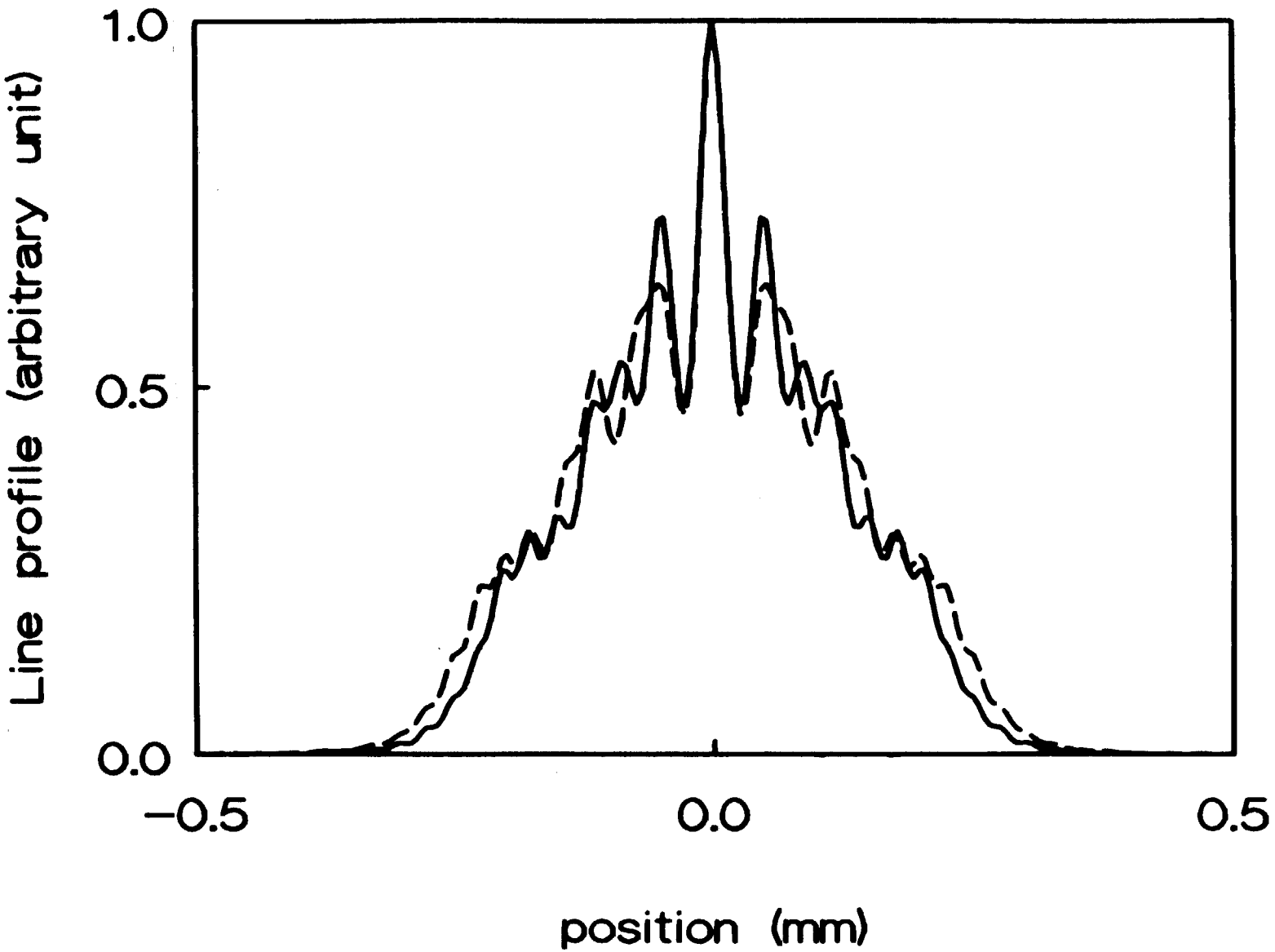


Figure 2c.

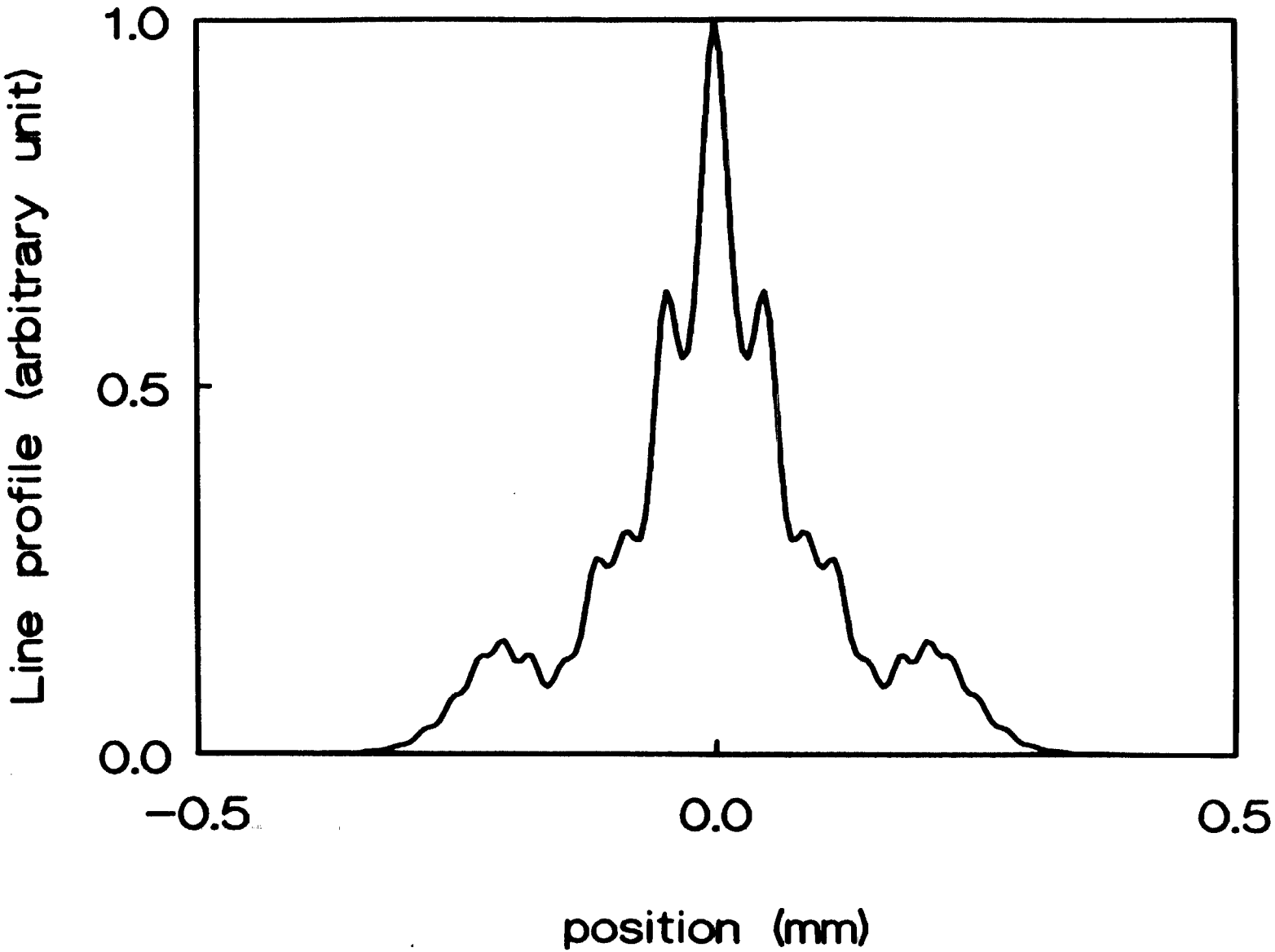


Figure 2d.

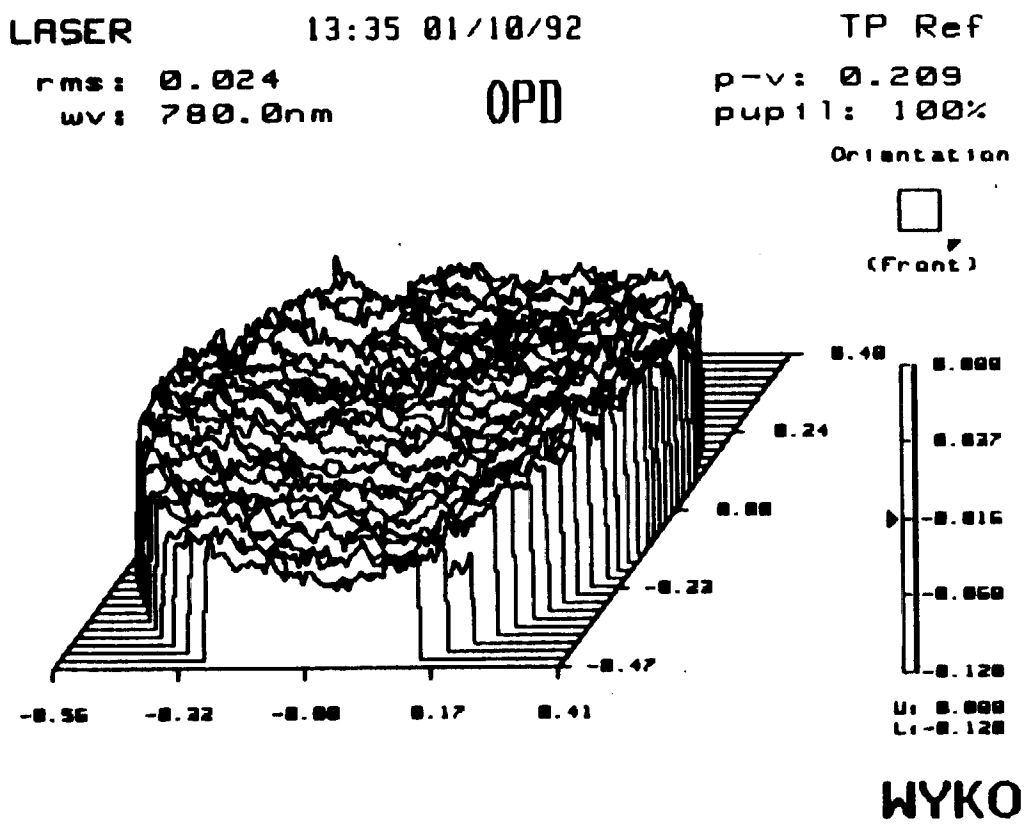

LASER
13:35 01/10/92
TP Ref
rms: 0.024
wv: 780.0nm
OPD
p-v: 0.209
pupil: 100%
Orientation
(Front)
0.48
0.24
0.00
-0.23
-0.47
-0.56
-0.22
-0.00
0.17
0.41
U: 0.000
L:-0.128
WYKO

Figure 3a.

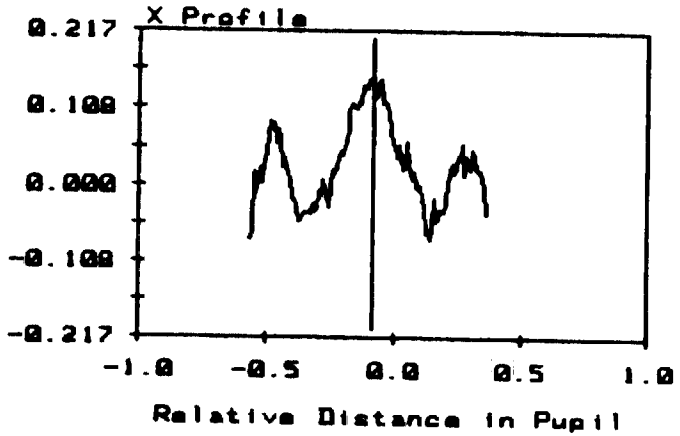


Figure 3b.

M.Wang
Optics Letters

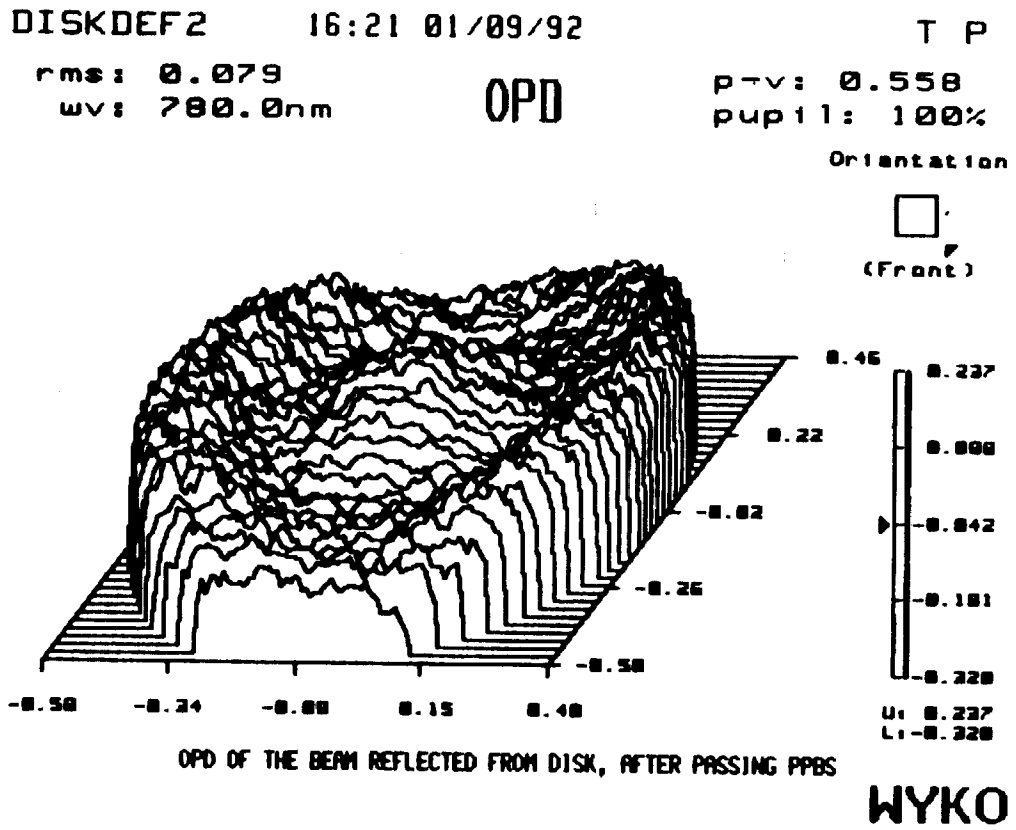


Figure 3c.

M. Wang
Optics Letters

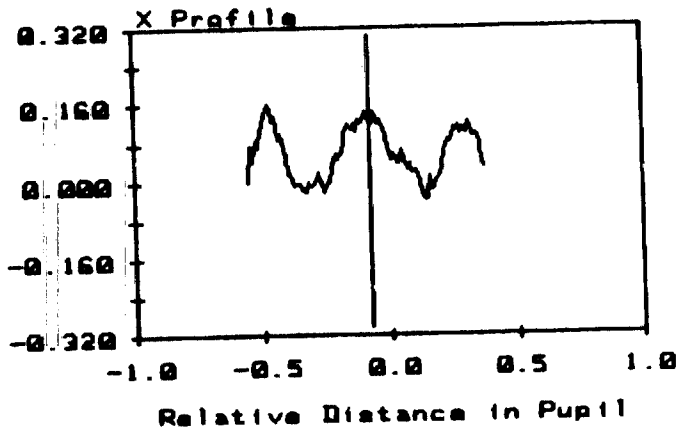


Figure 3d.

M. Wang
Optics Letters

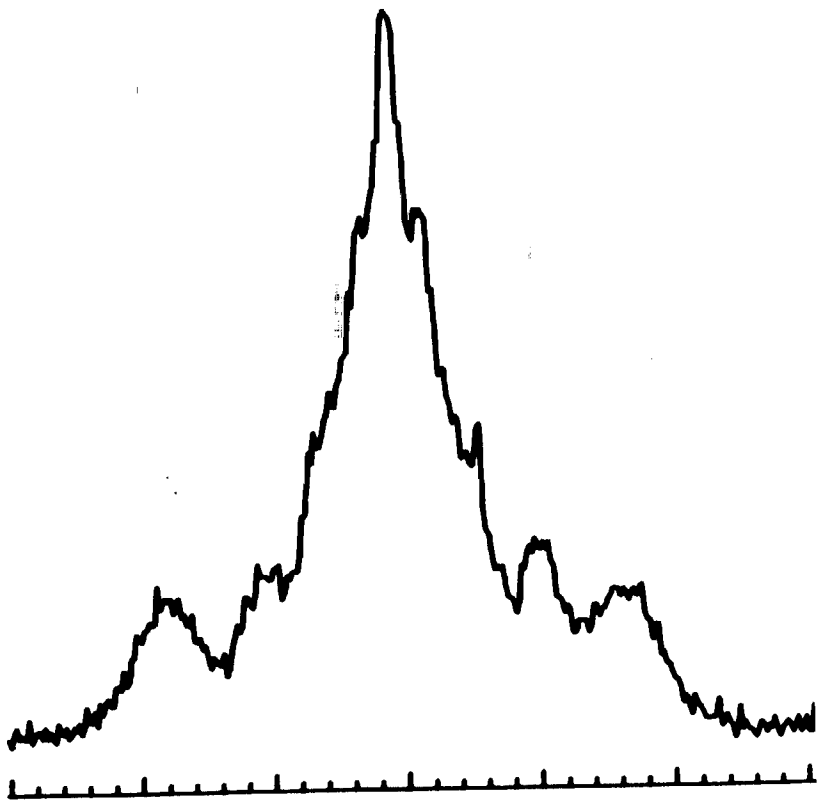


Figure 4a.

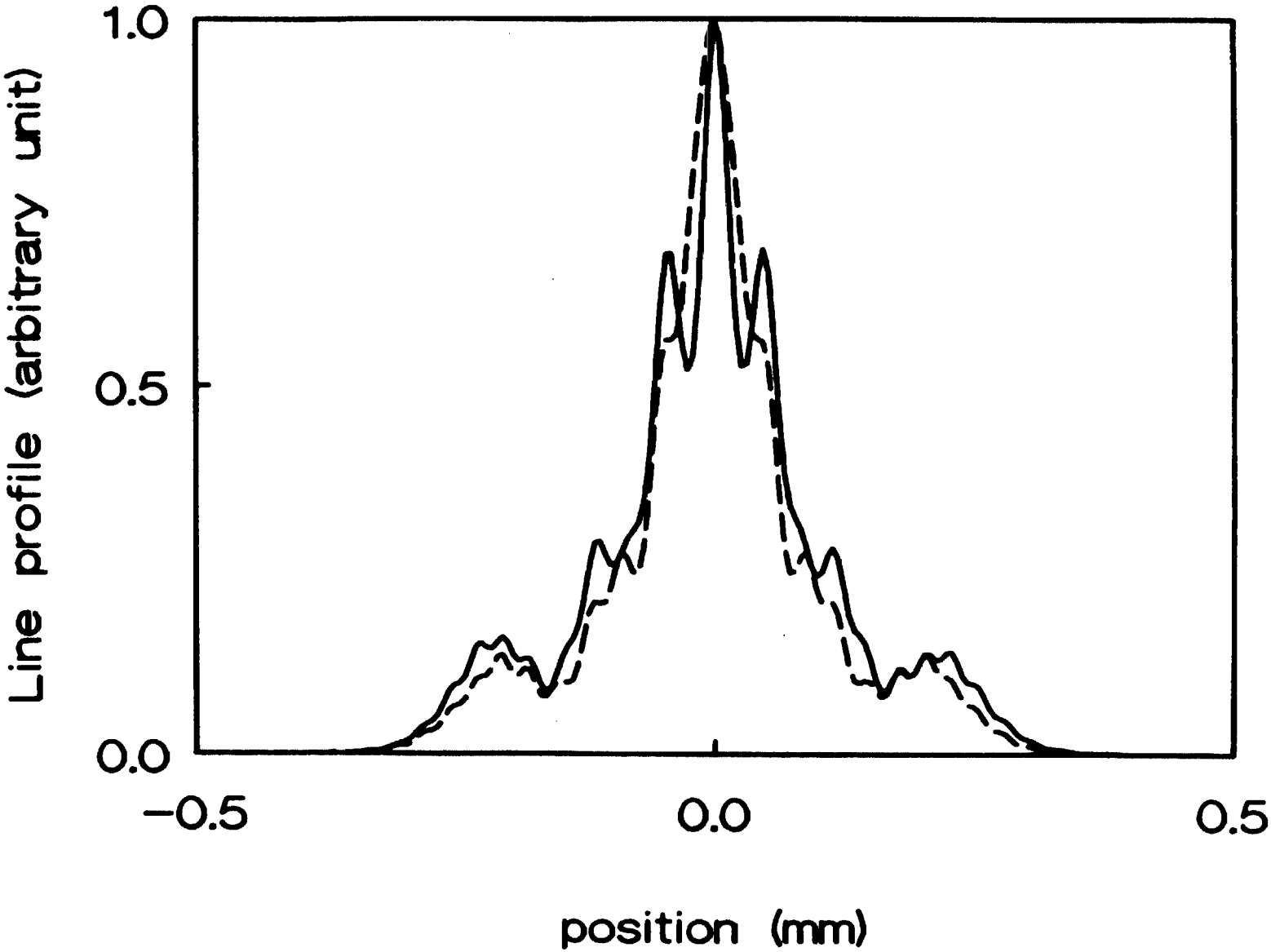


Figure 4b.

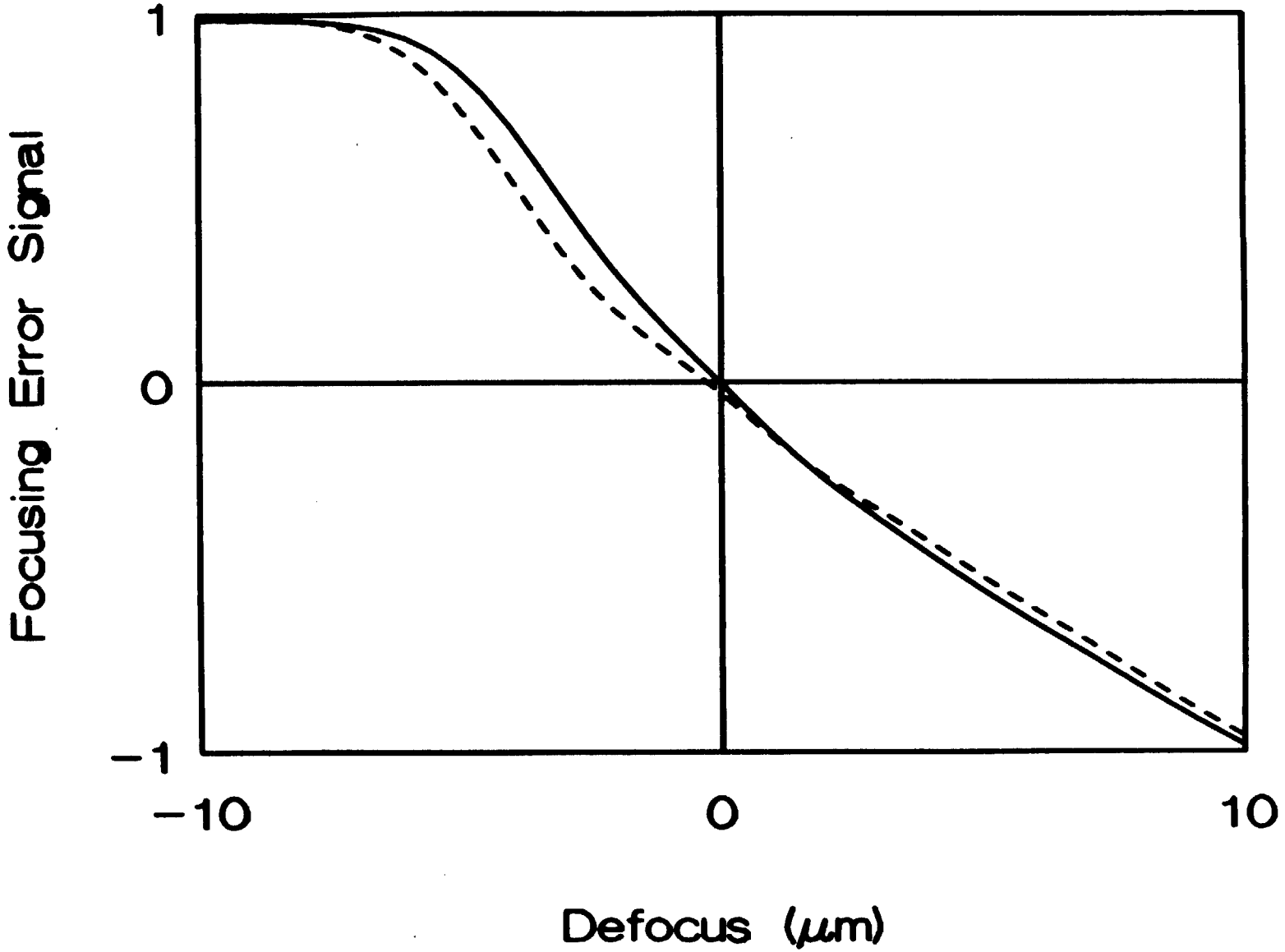


Figure 5.

APPENDIX K

SPIE Proceedings
submitted 2/12/92

Oxidation Processes in Magneto-Optic and Related Materials

Paul A. Lee#, Neal R. Armstrong, James L. Danziger and Craig D. England

Department of Chemistry
University of Arizona
Tucson, AZ 85721

ABSTRACT

The surface oxidation processes of thin films of magneto-optic materials, such as the rare-earth transition metal alloys have been studied, starting in ultrahigh vacuum environments, using surface analysis techniques, as a way modeling the oxidation processes which occur at the base of a defect in an overcoated material, at the instant of exposure to ambient environments. Materials examined have included FeTbCo alloys, as well as those same materials with low percentages of added elements, such as Ta, and their reactivities to both O_2 and H_2O compared with materials such as thin Fe films coated with ultrathin adlayers of Ti. The surface oxidation pathways for these materials is reviewed, and XPS data presented which indicates the type of oxides formed, and a critical region of Ta concentration which provides optimum protection.

1. INTRODUCTION

Oxidation and/or corrosion of magnetic and magneto-optic materials has become of considerable interest, since oxidation of these materials leads to loss of critical magnetic properties and thus, a loss of encoded data. Many of the materials in use or under consideration for MO data storage consist of binary, ternary and quaternary mixtures of transition metals and rare earth metal's (RE/TM), all of which react readily with ambient gases to form oxides with large heats of formation [1]. The RE/TM family is composed of TbFe, TbFeCo, GdFeCo, or TbFeCoM, where M is another metal such as Gd, Ti, Zr, Ta, Pt, etc., that has been added to enhance its stability towards oxidation and corrosion[1-9]. A systematic study of the various oxidation pathways in these materials has been carried out for certain systems[6-8,10,11].

The reaction of these materials can be classified as primarily surface oxidation processes under low and high humidity conditions, and "pitting corrosion". The former process appears to be primarily caused by the reaction with oxygen, and other strong oxidants in the atmosphere, and can occur uniformly along a metal surface, wherever contact with the atmosphere has been allowed (i.e. at the base of a defect in a protective overlayer used to generally protect the medium). Pitting corrosion represents a more destructive process, in that it generally involves the formation, dissolution and massive migration of oxides of the metals within the MO material. This process is facilitated by trace water in the environment, which can condense in defect sites, and by trace halides or other elements which accelerate the dissolution of oxides of many of the metals used in the RE-TM alloys[12].

Other recently developed MO materials consist of multilayers of two transition metals such as Pt/Co or Pd/Co which, due to the less reactive nature of Pt and Pd, and the lower reactivity of Co with respect to Tb, may promise to be more stable than RE-TM alloys[13-16] with the potential added cost of processing multilayer materials.

It has now been well documented that the addition of small levels of a fourth metal to the RE/TM alloys alters the oxidation pathways sufficiently to provide highly desirable protection of these materials, although the exact mechanism is still debated [2-9]. The addition of these metals, while satisfying the technological requirements for enhancing stability of the RE/TM alloy, without serious deterioration of the magneto-optical properties, poses more basic questions regarding the oxidation processes occuring at the surface of the material.

Many other methods have been used to protect the media from oxidation and subsequent corrosion, which include a variety of overprotection schemes[17-19]. With real materials, however, pinholes and other defects always develop, which expose the media to ambient conditions (albeit on a small distance scale initially) and there is no real way to prevent various forms of oxidation from occuring. Protection afforded by overcoating, and especially by an added fourth element, at low concentrations, can slow the process to bring it within technologically acceptable limits.

In our program, we have elected to model the initial steps of the corrosion process, which might occur at the base of a freshly produced defect site in a protective overlayer. We assume that, at the instant of formation of the defect, there is nearly clean and quite reactive metal alloy exposed, which undergoes rapid reaction to form oxides of various sorts, which ultimately control the stability of the RE-TM alloy in the vicinity of this defect. This initial reaction process can be modeled by producing a clean surface of the alloy, in ultrahigh vacuum conditions, and then exposing to low partial pressures of either O_2 or H_2O (or both), and systematically increasing their pressures until atmospheric conditions are achieved. X-ray photoelectron spectroscopy (XPS) is primarily used to characterize the extent of near surface reactions, as well as the molecular nature of the oxides formed, which confines the extent of our understanding to the first ca. 100Å of the RE-TM surface[20].

Most of our attention has been recently focused on the characterization of the differences in reactivity of RE-TM alloys (FeTbCo) to which have been added various amounts of a fourth element modifier[21,22]. Elements such as Zr, Ta, Nb, Ti, etc. are known to form stable, and stoichiometric oxides, which can act as a diffusion barrier to O_2 and H_2O. Migration of these metals has been seen in some cases, in reponse to exposure to O_2. Migration of both Fe and Tb is almost always observed, once their oxidation begins. It is of interest to determine if this simple protocol we have developed can lend some predictive capability to the determination of an optimum amount of the fourth element modifier, in addition to providing some mechanistic detail as to the corrosion protection mechanism. As will be discussed below, tantalum as a modifier can be quite efficient in its reaction with O_2, but it appears that there may be a critical Ta concentration which provides the best protection. Higher levels of Ta may actually form oxides which disrupt the microstructural environment sufficiently as to facilitate diffusion of O_2, and actually increase near surface reactivity.

2. EXPERIMENTAL

A number of methods have been used to study oxide film formation on a clean surface of the MO material in question, under highly controlled conditions (starting with UHV), and have been described elsewhere[21,22]. We have studied a number of metals directly and indirectly related to M-O systems using a combination of Auger Electron Spectroscopy (AES), X-ray Photoelectron Spectroscopy (XPS) and Rutherford Backscattering Spectrometry (RBS). These metals and alloys include Fe, FeTi, TbFeCo, TbFeCoM (where M = Zr, Ta, or Pt).

Exposure of the MO materials to oxygen and water was carried out by first removal of a thin Al protective coating (300Å) used to provide stability during transfer from the sputter deposition system. This sputtering step also removed small amounts of oxidized metal, to expose the alloy in as clean a state as it was produced (which means that there is always a few atomic percent oxide in the film to begin with). Dosing with O_2 and H_2O could then occur, between analysis by XPS and AES, at partial pressures of 1 x 10^{-7} to 1 x 10^{-5} torr, through a leakvalve, for Langmuir exposures of 1-10,000 L (where 1 langmuir = 1 x 10^{-6} torr-sec).

Clean Fe and FeTi thin films were made by evaporation from Fe and Ti wires. TbFeCo and TbFeCoTa were produced by both single and multiple target sputtering on doped Si wafers for RBS and XPS/AES analysis Analysis by RBS indicated that the films all had ca. 65-70% Fe, 19-21% Tb and 7-11% Co, and 1-5% Ta. Typical samples consisted of 1000Å of metal or alloy protected with 300Å of Al.

3. RESULTS AND DISCUSSION

3.1 Oxidation/Corrosion Processes, An Overview:

The reactions of O_2 and H_2O with clean, single element, metal surfaces have been extensively studied, and several excellent reviews exist detailing the present understanding of those processes[1,23-30]. Surface oxidation processes, through the reaction with O_2, occur either through the formation of $O_2^{=}$ or O_2^{-}, depending upon the partial pressure of O_2 at the interface with the ambient, which controls the concentration of anion states formed by the reduction of oxygen (Figure 1). Incorporation of the oxide next occurs by a place exchange mechanism, and subsequent growth of the oxide can occur by diffusion of the oxygen anion, and/or by diffusion of metal cations through the growing oxide. Both processes are field assisted, and therefore decrease in rate with time, since the concentration of anions at the surface of the material is fixed by the partial pressure of the oxidant, and the increasing thickness of the oxide decreases the field strength at the interface where oxide formation is occuring[24]. Secondary growth of the oxide can also continue through pores or other surface defects. The formation of this passivating oxide in some metal/alloy systems helps to prevent or inhibit the interdiffusion of oxidant to the metal/oxide interface.

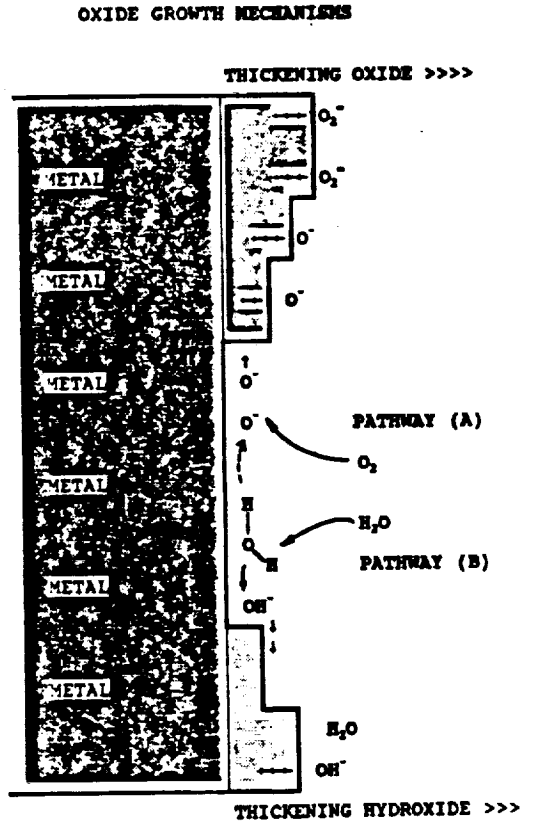


Fig. 1 Oxidation pathways for oxygen and water on a metal surface.

Given the tendency of metals such as iron and cobalt to form relatively stable oxides by this process, it might be expected that such near surface oxidation would not be a significant problem for the RE-TM alloy MO technologies[16,17,29,30]. The presence of the rare earth metal

however, coupled with the extremely thin films normally used, can add significance to this oxidation pathway. Surface oxides can often be formed, during exposure to atmospheric pressure O_2, which extend 100Å or more into the material, which can be an appreciable fraction of the total film thickness in a MO disk[6-8]. The oxides of the rare earth metals, which have a chemistry similar to calcium, can often support diffusion of both anions, cations, and even the zero valent metal itself, as has been shown previously by our studies of various alkali and alkaline earth metals[28].

The reaction of H_2O with these clean metals is believed to proceed by way of the chemisorption of molecular water (only seen at extremely low temperatures) and the subsequent formation of a surface hydride, and a hydroxide ion, followed by the further dissociation of the hydroxide to form another hydride and incorporated oxide anions (Figure 1)[1]. It has been our experience that many of the pure metals, and the metal alloys, show a significantly lower reaction tendency toward water than oxygen, in the gas phase, at low partial pressures, which we believe to be due to the presence of surface hydrides already present at near saturation levels under most vacuum and atmospheric characterization conditions. Only under conditions of high partial pressures of gaseous H_2O (i.e. the "wet oxidation" conditions described in other reports[9]), is appreciable reaction seen. This should not be construed, however, as inconsistent with the important role that H_2O plays in pitting corrosion, where the presence of condensed water acts to enhance diffusion of anions and cations in a microcorrosion cell, and supports the electrical double layer, which supports the electric field, which drives the pitting corrosion process[33]. When hydroxides are formed, as the result of the reaction with water, it is generally assumed that such hydroxide films are not chemically as stable as stoichiometric oxide films, and do not provide the type of diffusion barrier to O_2, etc. that stoichiometric oxide films provide. It is therefore essential to understand the nature of the reaction of both gaseous O_2 and H_2O, as the initial steps in the corrosion process.

The oxide layer, once formed, regardless of stability, can be an environment for further corrosion processes. In moist environments water condenses in microscopic regions on the surface of the metal or metal-oxide. Soluble species can then form (especially in the presence of halides) that leave pits upon dissolution, thus opening the surface of the alloy to further oxidation, a phenomenon known as pitting corrosion. Solution pH within the microscopic corrosion cell is often a controlling factor in the rate of pitting corrosion, as are mechanical stresses introduced into the metal film, and/or one of the protective overlayers[32].

3.2 Surface Oxidation Processes of Clean FeTbCo, FeTbCoTa:XPS Results

The reactions with O_2 and H_2O, of the clean metals, as would proceed at the base of a freshly opened defect site, have been largely studied by XPS, and compared with previous studies of pure Fe, and Fe films modified with adatom metal layers such as Ti[22].In the studies of the RE/TM alloys, the XPS core level spectra for the Fe(2p) and O(1s) transitions have been most revealing of the types of chemical processes ongoing during exposure to O_2 or H_2O. The photoelectron spectra for Tb or the other rare earths are complicated, show subtle shifts during the oxidation process, and cannot generally be used as an indication of the real growth of oxide films in the alloy. Fourth element modifiers such as Ta also have large enough XPS sensitivities as to provide for an additional monitor of oxidation processes.

A wide range of materials including Ti, Al, Ta, Zr, Au, Ag, Pt etc., have been used to try to inhibit surface oxidation and pitting corrosion[1-9]. Studies of Ti modified thin films showed a significant decrease in pitting corrosion[4,5]. In our initial studies we chose to

model the protection afforded by trace levels of Ti in the RE-TM alloy to both surface oxidation and the initial phases of pitting, by addition of discontinuous adlayers of Ti to a pure Fe surface, exposed to both O_2 and H_2O in the gas phase[33]. The initial oxidation of Fe is critical to the pitting corrosion process, and any protection which reduces the reactivity of Fe toward both O_2 or H_2O, and/or which may provide a diffusion barrier to those reactants. Thin Fe films were evaporated from resistively heated Fe filaments onto a Mo substrate and a 3-10 Å discontiuous layer of Ti was deposited over it. In Ti films greater than 10 Å the overlayer apparently became continuous and bulk oxidation properties were observed.

At Ti coverages of 3 Å some modification of the oxidation process for Fe did occur as shown in Figure 2. The three Fe(2p) XPS spectra show the lineshapes for the clean Fe (or FeTi) surface, and that same surface after exposure to low amounts (600L) of O_2, for both the bare surface, and one which had been modified with the 3Å Ti adlayer. For the unprotected surface there is extensive oxidation, as expected from several previous studies[25,26]. For the Ti-modified material, the extent of oxide formation has been reduced by ca. 50%, as deduced from the lower oxide film thickness on this sample (ca. 27Å versus 13Å on the unmodified Fe surface). In these studies the Ti overlayer is completely oxidized to TiO_2, and an exposure of 5L of O_2 to this Ti-modified Fe film is nearly enough to oxidize the entire Ti metal overlayer, which indicates how aggressive such a modifier can be in reacting with an incoming oxidant. Our previous work, along with that of several other investigators, has shown that the type of oxides formed on the Fe surface are combinations of FeO, Fe_3O_4, and at high enough oxygen exposures, Fe_2O_3[26]. This is essentially the type of oxidation pathway taken for Fe in the RE-TM alloy thin films, as seen below.

Low levels of water exposure to the Ti-modified Fe thin films does little to oxidize the Fe, but results in a range of oxidation states for the Ti, unlike the case for O_2 exposures (Figure 3). This wide range of oxidation states of Ti suggests that a very heterogeneous oxide will result from the reaction of H_2O with the modifier, which would be unlikely to provide a diffusion barrier (or an electronic barrier) toward further oxidation processes.

It has been widely reported that exposure of TbFeCo to O_2 typically results in initial oxidation of the Tb (in our studies to a depth less than 50Å), followed by oxidation and surface migration of Fe[10,11]. From these studies it appears that the surface of the oxidized TbFeCo thin film is composed of a terbium oxide layer over a mixed terbium/iron oxide at the interface with the unreacted

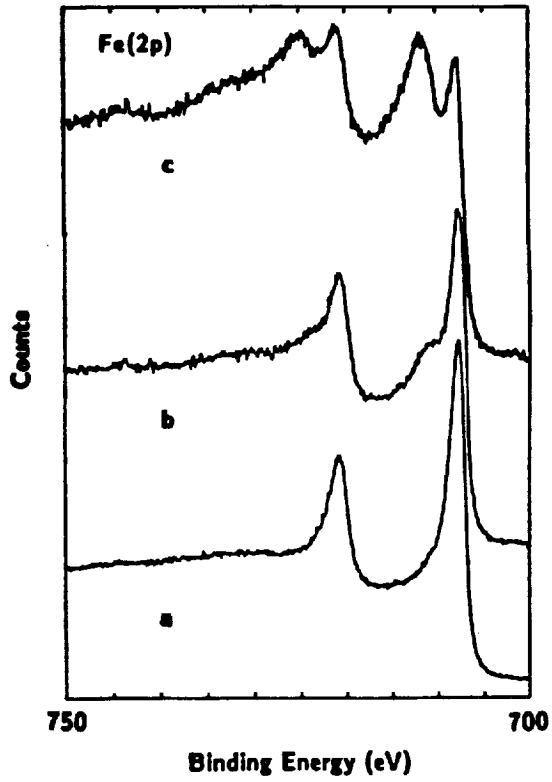


Fig. 2 Fe(2p) XPS lineshapes for 3Å of Ti on Fe. a) "Clean" Fe/Ti, b) after exposure to 600L O_2 and c) Clean Fe exposed to 600L O_2

metal alloy. Oxidation of the RE-TM thin film by H_2O results in the formation of a Tb oxide and hydroxide layer but does not oxidize the Fe appreciably, consistent with the results shown above for the Ti-modi e thin films. This oxidation by H_2O still causes the migration and enrichment of Tb in the near-surface region[7].

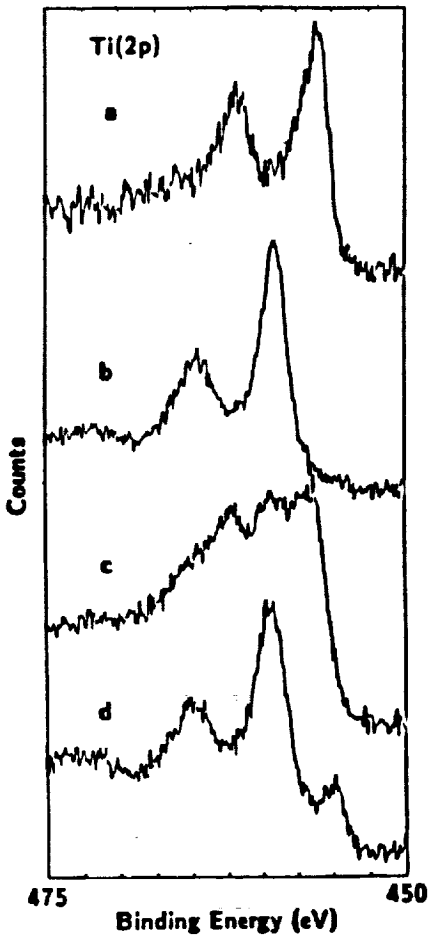


Fig. 3 Comparison of Ti(2p) Lineshapes for 3Å of Ti on Fe. a) Clean, b) 600L O_2, 600L H_2O, and 5L O_2 followed by 600L O_2.

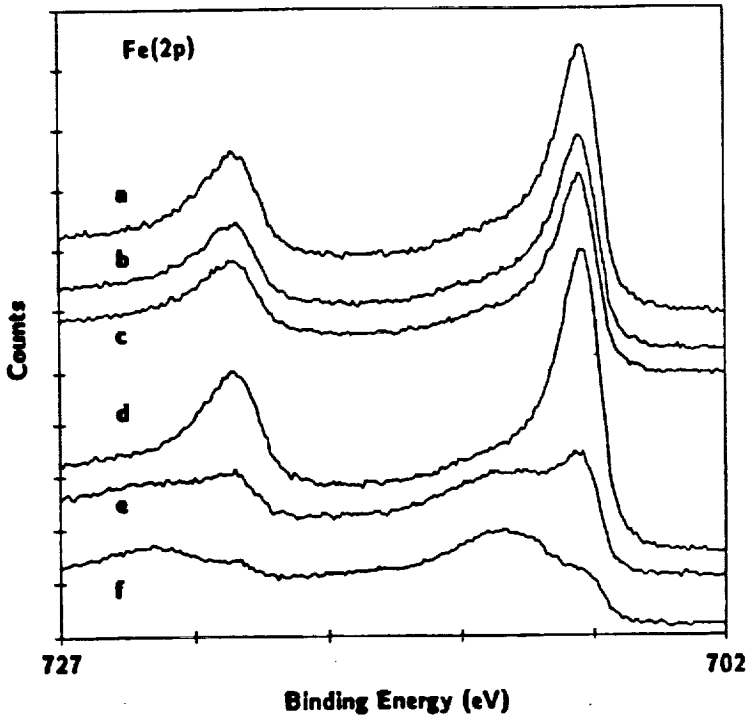


Fig. 4 Fe(2p) XPS lineshapes for TbFeCoTa 1% and 4% exposed to oxygen and water. a) "Clean", b) 1% Ta exposed to 1000L H_2O, c) 4% Ta exposed to 1000L H_2O, d) "Clean", e) 1% Ta exposed to 1000L O_2, f) 4% Ta exposed to 1000L O_2.

Oxidation studies of TbFeCo-Ta films, with Ta loadings of 1% to 4%, show that there are significant differences in the amount of protection associated with different Ta loadings in these thin films. From the Fe(2p) XPS spectra in Figure 4, it can be seen that there is considerable reaction of Fe with O_2 for both films, but that the extent of formation of the Fe^{+2} and Fe^{+3} oxides is lessened in the films with 1% Ta versus those with a 4% or higher loading. The Ta(4f) spectra for these samples (Figure 5) both show that extensive oxidation has occured to the Ta incorporated into these films, with the 1% Ta having been virtually completely oxidized by this low exposure to O_2 or by residual oxidants. Some migration of Ta occurs along with this sacrificial oxidation. The reasons for the better protection afforded by the lower Ta loading are not known, but since this is a consistent observation, it is tempting to suggest that the higher Ta loadings react with O_2 sufficiently to cause some microstructural changes in the near surface region, as a result of Ta

migration to form its oxide, which actually open up sections of the RE-TM alloy to O_2 which might otherwise have remained protected.

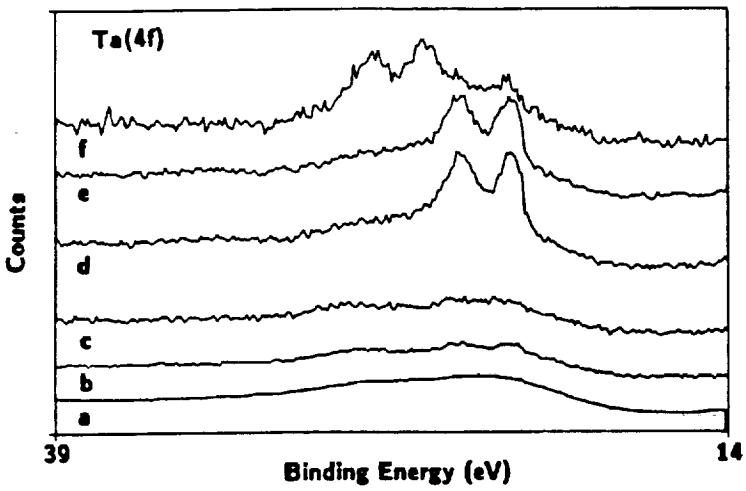


Fig. 5 Ta(4f) lineshapes for TbFeCoTa 1% and 4% Ta materials exposed to oxygen an water. a) "Clean" 1% Ta, b) 1% Ta exposed to 1000L H_2O, c) 1% Ta exposed to 1000L O_2, d) "Clean" 4% Ta, e) 4% Ta exposed to 1000L H_2O, f) 4% Ta exposed to 1000L O_2.

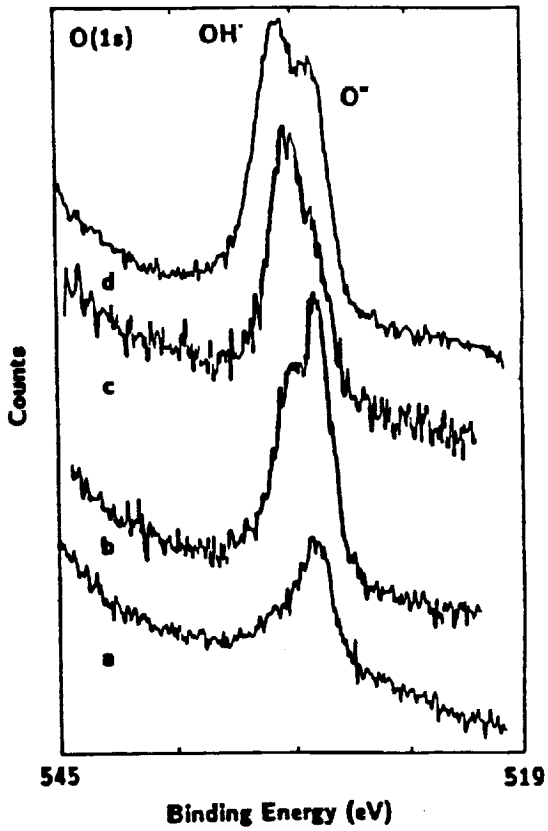


Fig. 6 O(1s) XPS lineshapes for water dosing on TbFeCo and TbFeCoTa 1% and 4% Ta films. a) Typical "Clean" film, b) 4% Ta exposed to 1000L H_2O, c) 1% Ta film exposed to 1000L H_2O and d) TbFeCo film exposed to 600L H_2O.

Oxidation of Ta-modified TbFeCo by water in this same near surface region results only in differences in the types of oxides seen in the 1% Ta film vs. the 4% Ta film. The difference is most exemplified in the O(1s) lineshapes of Figure 6. Unmodified TbFCo thin films exposed to low levels of H_2O show the existence of both oxide and hydroxide forms in the oxide layer, with the hydroxyl form being the most dominant. The assumption is that the oxide forms arise from reaction with both Tb and Fe (minimal reaction with Fe is seen in these cases) and that the hydroxides arise strictly from Tb. The 1% Ta material exhibited the same behavior as the unmodified FeTbCo films. Interestingly, in the 4% Ta materials the oxide formed was in greater proportion to the hydroxide than in the unmodified, or lower Ta content films. Since Ta will not readily form hydroxides, it is assumed that those hydroxides detected are strictly from the formation of Tb-hydroxides, and that these are favored in the lower Ta-content films.

4. CONCLUSIONS

The protection afforded a RE-TM alloy by a fourth element modifiers such as Ta may be viewed from a relatively simple perspective, even though the actual surface chemistries ongoing in the near surface region may be quite complex. There is a strong desire to add modifiers which will react preferentially with both O_2 and H_2O, migrate preferentially to the near surface region, and form protective barriers which prevent both diffusion of Tb outwards, and diffusion of O_2 and H_2O inwards, toward the unreacted metal. In the reactions with O_2 it is clear that it may be possible to select refractory oxide-forming metals such as Ta, Zr, and Nb which will fulfill most of those requirements. At high Ta loadings, however, the Ta oxidation chemistries may disrupt the microstructural environment sufficiently to actually increase the reactivity of the entire material in the vicinity of a newly formed defect. This question needs to be addressed in further experiments. The reaction with H_2O appears to be considerably more complicated. The fourth element modifiers that we have examined to date do not appear to affect the extent of oxide formation, as far as the XPS data for Fe reveal, but do play some role in the distribution of oxide versus hydroxide. The reasons for this change in distribution are still not clear, but may be important if it is assumed that the hydroxides formed are less passive than the stoichiometric oxides. Clearly one would like to have reactions involving H_2O proceed, as in Figure 1, so as to lose both hydrogens to hydride or H_2 formation, leaving the stoichiometric oxide. Except for the most reactive alkali metals, such as pure Li[28], this is rarely found to occur. The issue of oxide stability in the presence of water and halides has yet to be addressed in these studies, but is recognized as perhaps the central feature in determining the rate of pitting corrosion. Another alternative to protection is the addition of more noble metals such as Pt, which do not form stable oxides under the conditions cited here, but which may nevertheless change the reactivity of these RE-TM alloys, possibly through a purely electronic effect (e.g. change of surface work function). Detailed analyses of such processes are currently underway.

5. ACKNOWLEDGEMENTS

Partial support for this research is acknowledged from the Optical Data Storage Center at the University of Arizona, The National Science Foundation, and the Materials Characterization Program at the University of Arizona.

6. REFERENCES

1. Thiel, P. and Madey, T., "The Interaction of Water with Solid Surfaces: Fundamental Aspects", Surf. Sci. Rep., 7, 297-385, 1987.
2. Aratani, K., Kobayashi, T., Tsunashima, S. and Uchiyama, S., Magnetic and magnetic-optic properties of Tb-FeCo-Al Films", J. Appl. Phys., 57(1), 3903-3905, 1985.
3. Tanaka, S. and Inamura, N., "The Effect of Noble Metals on Oxidation Resistance of TbFe", Japan J. Appl. Phys., 24(5), L375-L376, 1985.
4. Koboyashi, M, Asano, M, Maeno, Y. Oishi, K. and Kawamura, K., High-Corrosion-Resistant Magneto-optical Recording Media Using TbFeCoTi Films", Appl. Phys. Lett., 50(23), 1694-1695, 1987.
5. Koboyashi, M, Maeno, Y., Oishi, K. and Kawamura, K., "Improvements in Corrosion Resistance of Tb-Fe thin Films", Appl. Phys. Lett., 52(6), 510-511, 1988.
6. Hatwar, T.K. and Debasis, M., "Oxidation Behavior of FeTbCoPt and FeTbCoZr Magneto-optic Films", Materials for Magneto-Optic Storage, T. Suzuki, C. Robinson and C. Falco, eds., vol. 150, 233-237, Materials Research Soc., New York, 1989.

7. Lee, P., Stork, K., Maschhoff, B. and Armstrong, N., "Oxide Formation on Rare Earth/Transition Metal Thin Films", Materials for Magneto-Optic Storage, T. Suzuki, C. Robinson and C. Falco, eds., vol. 150, 227-232, Materials Research Soc., New York, 1989.
8. Majumbar, D. and Hatwar, T.K., "Effects of Pt ans Zr on the Oxidation Behavior of FeTbCo Magneto-optic Films: X-Ray Photoelectron Spectroscopy", J. Vac. Sci. Technol. A, 7(4), 2673-2677, 1989.
9. Kirino, F., Ogihara, N and Ohta, N., "Corrosion Resistance of TbFeCo-M Magneto-optical Films" J. Electrochem. Soc., 138(8), 2259-2262, 1991.
10. Marinero, E., Miller, D., Bell, A., Gupta, A., Payne, R. and Notarys, H., "Oxidation Induced Magnetic and Structural Changes in Magneto-optical Alloys", IEEE Trans. on Magnetics, MAG23(5), 2629-2631, 1987.
11. Frankenthal, R., Siconolfi, D., van Dover, R. and Nakahara, S., "Oxidation of an Amorphous Iron-Terbium Alloy", J. Electrochem. Soc., 134(1), 235-239, 1987.
12. Farrow, M. and Marinero, E., "Corrosion Processes in Magneto-optic Films", J. Electrochem. Soc., 137(3), 808-814, 1990.
13. England, C.D., Engel, B.N., Falco, C.M., "Preparation and Structural Characterization of Epitaxial Co/Pt(111) Superlattices", J. Appl. Phys., 69(8), 5310-5312, 1991.
14. Chien, C.J., Farrow, R.F.C., Lee,C.H., Lin, C.J., Marinero, E.E., "High-resolution Transmission Electron Microscopy Studies of Seeded Epitaxial Co/Pt Superlattices", Magnetic Thin Films, Multilayers and Superlattices, European Materials Research Society Symposia Proceedings, A. Fert, G. Güntherodt, B. Heinrich, E.E. Marinero, M.Maurer, eds.,vol. 16, 47-52, Elsevier Science Publishers, Amsterdam, 1990.
15. Carcia, P.F. Shah, S.I., Zeper, W.B., Appl. Phys. Lett., 56, 2345 (1990)
16. Bardi, U., Beard B.C., and Ross, P.N., "Suface Oxidation of a Pt-20% Co Alloy: AnX-ray Photoelectron Spectroscopy and Low-energy Electron Diffraction Study on the [100] and [111] Oriented Single Crystal Surfaces", J. Vac. Sci. Technol.A, 6(3), 665-670, 1988.
17. Hartmann, M., Witter, K. Reck, J., Tolle, H.J.,"Improvement of Corrosion Resistance of GdTbFe by metal coatings", IEEE Trans. Mag., MAG-22(5), 943-945, 1986.
18. Hatwar, T.K., Shin, S. and Stinson, D.,"Corrosion Resistant Protective Overcoat for Magneto-optical Media", IEEE Trans. on Magn. MAG-22(5), 946-948, 1986.
19. Frankenthal, R., van Dover, R. and Siconolfi, D., "Duplex Coatings for the Protection of Magneto-optic Alloys Against Oxidation and Corrosion", Appl. Phys. Lett., 51(7), 542-544, 1987.
20. Orchard, A.F., in Handbook of X-ray and Ultraviolet Photoelectron Spectroscopy, D. Briggs, ed., p. 15, Heyden and Son, Philadelphia, 1977
21. Lee, P. and Armstrong, N., "Oxide Formation on Rare Earth/Transition Metal and Bimetallic Transition Metal Thin Films: Modeling the Effect of Fourth Element Modifiers on O_2 and H_2O Surface Chemistries", J. Magn. Magnetic Mat., 93, 159-163, 1991.
22. Lee, P.A., Armstrong, N.R., "Oxide Formation on Rare Earth/Transition Metal and Bimetallic Transition Metal Thin Films: Modeling the Effect of Fourth Element Modifiers on O_2 and H_2O Surface Chemistries", Magnetic Thin Films, Multilayers and Superlattices, European Materials Research Society Symposia Proceedings, A. Fert, G. Güntherodt, B. Heinrich, E.E. Marinero, M.Maurer, eds.,vol. 16, 159-163, Elsevier Science Publishers, Amsterdam, 1990.
23. Brundle, C.R., Broughton, J.Q., in Chemical Physics of Solid Surfaces and Heterogeneous Catalysis, Ch. 3, vol. 3A, D.A. King and D.P. Woodruff, eds., Elsevier, Amsterdam, 1990,

24. Felner, F.P. and Mott, N.F., "Low Temperature Oxidation", Oxid. of Met., 2(1), 59-99, 1970
25. Maschhoff, B.L., PhD Dissertation: Electron Spectroscopic Characterization of Corrosion Reactions of Active-metal Systems, University of Arizona, 1988.
26. Maschhoff, B.L., Armstrong, N.R., "Thin Oxide Layers on Clean Iron Surfaces under Vacuum and Characterization by Photoelectron Spectroscopy and Electrochemical Reactions of Probe Molecules at the Oxide/Electrolyte Interface" Langmuir, 7, 693-703, 1991.
27. Burrell, M.C., Armstrong,N.R., "Oxides Formed on Polycrystalline Titanium Thin Film Surfaces: Rates of Formation and Composition of Oxides formed at Low and High O_2 Partial Pressures" Langmuir, 2(30), 30-36, 1986.
28. Zavadil, K.R., Armstrong, N.R., "Surface Chemistries of Lithium: Detailed Characterization of the Reactions whth O_2 and H_2O Using XPS, EELS, and Microgravimetry", Surf. Sci., 230, 47-60, 1990.
39. Wang, N., Kaiser, U., Ganschow, O., Wiedmann, L., and Benninghoven, A., "Oxidation of Cobalt at Room Temperature, Studied by Combined Static SIMS, Static AES, XPS and Work Function Investigations' ,Surf. Sci., 124, 51-67, 1983.
30. Bogen, A., Küppers, J.,"Oxidation of Co(Poly) and Co(1010) Surfaces: Structure and Temperature Effects", Surf. Sci., 134, 223-236. 1983.
31. Padalia, B., Gimzewski, J., Affrossman, S., Lang, W., Watson, L. and Fabian, D., "The Reactions of Oxygen and Water with the Rare-Earth Metals Terbium to Lutetium Studied by X-Ray Photoelectron Spectroscopy", Surf. Sci., 61, 468-482, 1976.
32. Bockris, J. O'M. and Reddy, A.K.N. in "Modern Electrochemisty", v2, pp. 1267-1325, Plenum Publishing, New York, 1970
33. Lee, P., Stork, K., Maschhoff, B., Nebesny, K., Armstrong, N.," Oxide Formation on Fe an Ti Thin Films and on Fe thin Films Modified with Ultrathin Layers of Ti', Surf. Interface Anal., 17, 48-54, 1991.

APPENDIX L

Integrated device with diffractive polarization components for a magneto-optical disk head

Charles W. Haggans and Teruo Fujita*
Optical Sciences Center
University of Arizona
Tucson, AZ 85721
(602)-621-2031

Raymond K. Kostuk
Electrical and Computer Engineering Department
and Optical Sciences Center
University of Arizona
Tucson, AZ 85721

1. INTRODUCTION

The optical components in the detection train of a conventional magneto-optical (M-O) disk head include a half-wave plate and a polarization beamsplitter. These polarization components are bulky and require specialized mounting hardware. In order to realize a more compact head, we propose that these elements be replaced by an integrated device composed of cascaded volume and surface-relief gratings. In this paper, we describe the proposed system, detail designs for the individual elements, compare theoretical and prototype element performance, and discuss the operational tolerances of these elements.

Several authors have reported efforts to replace individual conventional elements in the M-O head detection train with diffractive elements.[1-5] The most comprehensive of these studies was that of Ono et. al.[1] in which focus and tracking error signals and the M-O signal were generated using a multifunctional holographic optical element (HOE). However, the angular separation of the transmitted and diffracted beams generated by this element (1°) limits the compactness of the assembled system (assuming a detector spacing of 500 μm, the leaky beamsplitter to detector plane spacing must be approximately 30 mm). Our proposed device is much more compact because all detection train elements are integrated onto two thin substrates. Thus, the leaky beamsplitter to detection plane distance is approximately 8 mm for a 4 mm diameter beam.

2. DEVICE LAYOUT

Figure 1 is a schematic diagram of a conventional M-O head. The detection train is indicated in this figure. We propose to replace the detection train components with a substrate-mode device, giving the system configuration shown in Figure 2. The leaky beamsplitter to detector distance with this device is reduced significantly from the conventional head design (~8 mm as opposed to ~30 mm). The optical path of this device is an integrated version of a detection train design presented by Milster[6] and Matsubayashi.[7] This differs from the conventional detection train in that circularly polarized light is incident on the polarization beamsplitter due to the presence of a quarter-wave retarder.

The proposed device is detailed in Figure 3. The beam incident on the device from the leaky beamsplitter is primarily p polarized (the electric field vector in the plane of the drawing). The beam also contains the small s polarized signal introduced by reflection from the M-O medium as well as any phase difference introduced between the two polarizations due to the M-O medium, the disk substrate, the folding mirror, and the leaky beamsplitter. This beam is coupled into the substrate mode by Element 1. This element is a slanted volume grating fabricated holographically in dichromated gelatin. This grating has a high diffraction efficiency and causes negligible polarization conversion (rotation) or phase shift between the s and p components of the incident beam. After propagating through the substrate, the beam is reflected from a high spatial frequency lamellar (rectangular) grating at the boundary between the glass substrate and a metallic layer. This grating acts as a quarter-wave retarder in that the reflected beam has an s-p phase difference of approximately -90°. Element 3 is a lamellar grating at the top of the substrate that is also formed between the glass and a metallic layer. This grating is oriented so that the polarization angle of the reflected 0th order beam is rotated by 45° with respect to that of the incident beam. Thus, the polarization angle of the beam reflected from this grating is oriented at $45° \pm |\Theta_K|$ with respect to the p polarization component direction. The phase difference between the s and p components of the reflected beam is 90°. This means that the beam reflected from Element 3 is nearly circularly polarized. Finally, a lamellar grating formed between the substrate and air (element 4) acts as a polarization beamsplitter, separating the s and p polarization components for differential signal detection with a high extinction ratio.

Focus and tracking error sensing is implemented on the s and p polarized beams exiting Element 4. The propagation directions of these beams vary with wavelength due to the diffractive nature of elements 1-4. However, a push-pull tracking error signal can be generated from the p polarized beam transmitted by Element 4 by aligning the split detector so that the beam moves along the separation between the detectors as the wavelength varies.

A focus error signal can be generated from the s polarized beam diffracted by Element 4 with an off-axis pupil obscuration method. Figure 3 shows a focus error sensing path consisting of a reflectively coated cylindrical lens, a knife-edge, and a split detector. The separation between the detector halves is parallel to the plane of the paper. In this path, the s polarized beam comes to a line focus that is inclined with respect to the detector plane. When the objective is properly focused on the M-O disk, the signal detected by the left and right detector halves cancels as in the standard pupil obscuration method. For an out of focus disk, the intensity pattern on the detector plane is asymmetric with respect to the split line of the detector. Thus, while the intensity pattern on the split detector is highly aberrated, its symmetry allows for focus error signal generation.

3. COMPONENT AND DEVICE MODELING

The polarization properties of the grating elements composing this device have been calculated using rigorous vector coupled-wave[8,9] and modal[10] grating diffraction models. The modal model has been used for metallic surface-relief grating cases where convergence was not achieved with the coupled-wave model.[11] In this paper, we use the notation of Born and Wolf[12] to describe the state of polarization of the substrate-mode beams. The electric field vector for an elliptically polarized beam can be written as

$$E=\$A_s\exp(i[k\cdot r-\omega t-\phi_s]) + \hat{p}A_p\exp(i[k\cdot r-\omega t-\phi_p]). \quad (1)$$

The polarization state of this beam is fully described by the angular ratio of the s and p polarization component amplitudes,

$$\alpha=\text{atan}(A_s/A_p), \quad (2)$$

and the phase difference between the s and p polarization components,

$$\delta=\phi_s-\phi_p. \tag{3}$$

An elliptically polarized beam can also be uniquely described by two other angular parameters: the polarization rotation angle Θ and the ellipticity ε. These parameters can be related to α and δ by

$$\tan 2\Theta = \tan 2\alpha \cos\delta \tag{4}$$

and

$$\sin 2\varepsilon = \sin 2\alpha \sin\delta. \tag{5}$$

The grating and incidence geometries used in this paper are defined in Figure 4. A subscript of i denotes parameters relating to the incident beam and subscripts of r and t denote reflected and transmitted beam parameters, respectively.

The polarization characteristics required for Elements 1-4 are summarized in Sections 3.1-3.4. In each section, grating dimensions and incidence parameters are given for an element designed for operation at a wavelength of 0.780 μm. Finally, in Section 3.5, the integrated performance of these four elements is detailed. For these computations, the propagation angle within the substrate was chosen to be 45° and the substrate refractive index was chosen to be 1.517. For Elements 2 and 3, the metallic external medium was chosen to be gold (n=0.175, κ=4.91).[13]

3.1 Unselective Incoupler

The first element of this device must couple the normally incident beam into the substrate with high efficiency and negligible polarization rotation (α_t≈45°). The period of this grating (Λ) must be 0.727 μm in order to obtain a propagation angle of 45° in the substrate. We investigated the potential for using a triangular profile (blazed) grating as the first element. However, rigorous modeling indicated that efficiencies greater than 0.5 were not achievable for a surface-relief element in this diffraction geometry. Thus, we propose that a slanted grating fabricated in dichromated gelatin (DCG) be used for this element.

The coupled-wave volume grating model indicates that a DCG grating can theoretically have an efficiency of 0.9 with negligible change to the incident values of α and δ. This is possible for a DCG grating with a period of 0.727 μm, an emulsion thickness of 8.75 μm, a bulk emulsion refractive index of 1.44, and an index modulation of 0.0435.

3.2 Quarter-Wave Retarder

The second element in this device is a metalized lamellar (rectangular) grating. This grating is oriented with the grating vector either parallel or perpendicular to the plane of incidence (Φ=0° or 90°). We have previously reported[14] that this type of grating can function as a quarter-wave retarder. In this device, this element introduces a -90° s-p phase difference to the incident beam (α_i=±0.5°, δ_i=0°) so that the reflected beam is nearly circularly polarized. All diffracted orders other than the specularly reflected order are suppressed upon reflection from this grating because the grating period satisfies the relation

$$\Lambda \le \lambda/[n(\sin\Theta+\sin 90°)], \tag{6}$$

where n is the substrate refractive index and Θ is the propagation angle of the substrate-mode beam with respect to the grating normal. For the parameters chosen for this device, the maximum allowable period for this grating is 0.301 μm.

A lamellar grating coated with gold can introduce the desired phase shift to the incident beam. A grating with d=0.205 μm, Λ=0.3 μm, and dc=0.5 that is oriented with Φ=90° gives δ_r=-90° and α_r=±0.5° for an incident beam with δ_i=0° and α_i=±0.5°. The efficiency for reflection from this grating is 0.92.

3.3 Polarization Rotator

The third element of this device is used for polarization rotation (modification of α_r) to nearly equalize the s and p polarization component amplitudes of the substrate-mode beam. This is done with a half-wave plate in conventional systems to maximize the differentially generated M-O signal. For conical incidence (Φ≠0°) on a lamellar grating at a glass-metal interface, the amount of polarization rotation imparted to the reflected beam can be chosen by varying d and Φ. A grating with Λ=0.3 μm, d=0.155 μm, and dc=0.5 oriented with Φ=44° changes the state of polarization from α_i=±0.5° and δ_i=-90° to α_r=45 ± 0.5° and δ_r=90°. The efficiency of reflection from this grating is 0.90.

3.4 Polarization Beamsplitter

Element 4 is a polarization beamsplitter that separates the components of the incident beam with high efficiency and a high extinction ratio. This is accomplished with a lamellar grating at the boundary between the first and second substrates of the device. These substrates must be assembled so that the grating grooves are filled with air (optical cement or index matching fluid must not fill the grating grooves).

A grating of this type with Λ=0.364 μm, d=0.98 μm, and dc=0.3 oriented with Φ=0° functions as a PBS for this device. The p polarization component is transmitted by this grating without angular deviation with an efficiency of 0.99 and an extinction ratio of 138 ($ER=A_{p,trans}/A_{s,trans}$). The s polarization component is diffracted into the +1st order which lies in a direction that is perpendicular to the p polarized beam. The diffraction efficiency for s polarized light is 0.92 and the extinction ratio is 74.

3.5 Performance of Cascaded Gratings

The previous sections indicate that it is theoretically possible for grating elements to function as the polarization components required in an M-O head detection train. Figure 5 summarizes the calculated properties of the grating elements. The throughput efficiency for elements 1-4 is 0.74 for p polarization and 0.7 for s polarization.

4. EXPERIMENTAL RESULTS

Efforts are underway in our laboratory to fabricate the four diffractive components used in this device. To this point, we have concentrated on fabricating Elements 2 and 3. These elements consist of silver-coated photoresist gratings on soda-lime substrates. These elements have been designed for use at a wavelength of 0.6328 μm to simplify the measurement of their polarization properties.

Figure 6 is a scanning electron micrograph of a photoresist grating fabricated in our laboratory. To fabricate this and other gratings, Shipley 1805 series positive photoresist was diluted to give nominal film thicknesses of 0.25 μm after spinning at 4000 rpm on soda-lime glass substrates. Gratings were then holographically generated in the photoresist layer using the Fresnel mirror method of Malag[15] at a wavelength of 0.4579 μm. The period of the gratings was chosen to be 0.24 μm to suppress all diffracted orders other than the 0th reflected. The gratings were monitored during development using a technique reported by Li et. al.[16] to ensure development to the substrate.

Following postdevelopment baking, 0.2 μm of silver was sputtered onto the photoresist gratings. Silver was used for the metallic coating because it is a better reflector than gold at λ=0.6328 μm. The refractive

indices of the substrate and silver were measured with a Gaertner L116C ellipsometer. The substrate refractive index was found to agree with the manufacturer's specification of 1.517. The real and imaginary parts of the silver refractive index were measured to be $n=0.12 \pm 0.02$, and $\kappa=2.48 \pm 0.01$. Measurement of the Fresnel reflection coefficient for incidence on the silver from the glass substrate confirmed the validity of these measurements. The period of this grating was measured optically to be 0.241 ± 0.001 μm, and the depth and duty cycle were estimated from the micrograph to be 0.28 ± 0.02 μm and 0.65 ± 0.05 respectively. The polarization state of a beam specularly reflected from the grating in Figure 6 was measured using the ellipsometer. A 45-45-90 prism was index matched to the back surface of the substrate to simulate substrate-mode incidence on the grating at 45°.

Figures 7a and 7b show experimental and calculated results for a polarization rotation grating (Element 3). Figure 7a is a plot of measured output polarization rotation angle (α_r) versus azimuthal angle Φ for p polarized light incident on the grating ($\alpha_i=0°$, $\delta_i=0°$). Note that at an azimuthal angle of incidence of ~55°, this grating gives the desired output polarization angle of 45°. Plotted along with the experimental data are calculated results for reflection from a lamellar photoresist grating coated with silver with dc=0.65, $\Lambda=0.241$, d=0.24, $n_{photoresist}=1.64$, and the substrate and silver indices as above. The discrepancies between the curves can be attributed to the poor approximation of the actual grating profile by a lamellar profile.

Figure 7b shows measured and calculated diffraction efficiencies for this grating. The efficiencies are low (~0.5) because of the small value of the imaginary part of the silver refractive index. This value is dependent upon the method of deposition, which was sputtering for this element. Typical values for evaporated silver are near 4.0, which would increase the element efficiency.

5. OPERATIONAL TOLERANCES FOR GRATING ELEMENTS

The theoretical analyses of Section 3 indicate that it is possible to design the gratings required for the proposed detection train device. However, the impact of wavelength and incident phase difference variations on the performance of the elements must be evaluated.

The wavelength of laser diodes produced for use in M-O heads varies due to manufacturing tolerances. A realistic estimate of this variation is ±0.01 μm about the design value, which is 0.780 μm for the device we have modeled. This wavelength variation directly impacts the performance of Elements 1-4. Additionally, the propagation angle within the substrate in our proposed device is wavelength dependent due to the incoupling volume grating. For a beam normally incident on the incoupling grating, the propagation angle of the substrate-mode beam is given by

$$\Theta=\text{asin}(\lambda/n\Lambda). \quad (7)$$

Substituting the period of this grating (0.732 μm), the substrate refractive index (n=1.517), and the range of the wavelength variation (0.77 μm to 0.79 μm) into this equation gives the range of propagation angles within the substrate for which the elements must perform.

We investigated the influence of wavelength variation (with associated angle of incident variation) on the throughput efficiency of this device, the polarization state incident on Element 4, and the extinction ratios for the beams transmitted by Element 4. Figure 8 is a plot of system throughput efficiency for $\alpha_i=\pm0.5°$ and $\delta_i=0°$ as a function of wavelength. The 1st order efficiency in this plot corresponds to the s polarized beam diffracted by the PBS for focus error sensing. The 0th order efficiency corresponds to the p

polarized beam used for tracking error sensing. An efficiency of greater than 0.69 for p polarization and 0.66 for s polarization is maintained over a wavelength range of 0.780 ± 0.01 μm.

Figure 9 is a plot of the rotation angle α_r of the beam incident on Element 4 as a function of wavelength for system input angles of $\alpha_i=\pm 0.5°$. Note that while the values of α_r for both incident rotation angles vary with wavelength, the separation between the two remains constant. Thus, minimal signal degradation occurs. The phase of this beam is not plotted because it is nearly invariant to wavelength variations.

Figure 10 is a plot of the extinction ratio of the s polarized 1st order and the p polarized 0th order diffracted by Element 4 (for the design of Section 3.4) as a function of wavelength. For this design, the extinction ratio of the s polarized 1st order is degraded with wavelength variation. However, the minimum extinction ratio of 31 still allows for differential signal generation.

An additional factor that influences the performance of the rotation grating (Element 3) is deviation of the incident state of polarization from left circular. As δ_i deviates from -90°, the magnitude of the Kerr rotation component is decreased after rotation (e.g, $\alpha_i=\pm 0.5°$ may be mapped to $\alpha_r=45.0 \pm 0.4°$ by reflection from the rotation grating). This is very undesirable, as this is a direct reduction of the signal. Figure 11 is a plot of rotation angle α_r versus δ_i for the design of section 3 with $\alpha_i=0.5°$. A typical range of δ_i for the incident beam is ±30° due to varying media ellipticities, M-O disk substrate birefringences, and phase differences introduced by the turning mirror and leaky beamsplitter. This figure shows that the maximum signal degradation is 20% over this range of δ_i.

6. CONCLUSION

In this paper, we have presented a novel device for providing M-O signal generation and error signal generation in a compact package. We also have presented gratings as polarization components in the M-O head detection train. We have experimentally demonstrated the polarization properties of one of the elements composing our proposed detection train device. This work suggests that significant miniaturization of the detection train can be accomplished if a minor decrease in detection train performance can be tolerated.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance of L. Li of the Optical Data Storage Center, University of Arizona in modeling the polarization properties of surface-relief gratings. The authors acknowledge numerous helpful discussions regarding grating fabrication with L. Li, K. Bates, and K. Erwin, all of the Optical Data Storage Center, University of Arizona. The authors wish to acknowledge support for this work from IBM SUR-0499, a Litton ITEK Optical Systems Graduate Research Fellowship, and the Optical Data Storage Center of the University of Arizona.

* Current address: Consumer Electronics Laboratory, Mitsubishi Electric Corporation, 1 Zusho-Baba, Nagaokakyo-City, Kyoto 617, Japan

8. REFERENCES

1. Y. Ono, "Polarizing holographic optical elements for optical data storage," SPIE 1545, in press.
2. Y. Ono, "Holographic optical elements for optical disk R/W heads," Digest from Optical Memory Symposium '90 (in Japanese), July 1990, Makuhari, Japan.
3. S. Aoyama and T. Yamashita, "Grating beam splitting polarizer using multi-layer resist method," SPIE 1545, in press.
4. H. Maeda et. al., "A high density dual type grating for magneto-optical disk head," Japanese Journal of Applied Physics 28, Supplement 28-3, 193-195 (1989).
5. M. C. Gupta and S. T. Peng, "Multifunction grating for signal detection of optical disk," in Optical Data Storage, 1991, Technical Digest Series, 5 (Optical Society of America, Washington, DC 1988), 164-168.
6. T. D. Milster, "Characteristics of phase compensation techniques in magneto-optical read-back systems," SPIE 1166, 355-364 (1989).
7. N. Matsubayashi et. al., "Magneto-optical readout system by detecting elliptically polarized light," Proceedings of the Japan Society of Applied Physics Autumn Conference, 30A-PB-10 (1989).
8. C. W. Haggans and R. K. Kostuk, "Use of rigorous vector coupled-wave theory for designing and tolerancing surface-relief diffractive components for magneto-optical heads," SPIE 1499, 293-302 (1991).
9. M. G. Moharam and T. K. Gaylord, "Three-dimensional vector coupled-wave analysis of planar-grating diffraction," J. Opt. Soc. Am. 73, 1105-1112 (1983).
10. L. Li, "A modal analysis of lamellar diffraction gratings in conical mountings," to be submitted to J. Mod. Optics.
11. L. Li and C. W. Haggans, "On the convergence of the coupled-wave approach for lamellar diffraction gratings," accepted for Topical Meeting on Diffractive Optics, New Orleans, April 1992.
12. M. Born and E. Wolf, Principles of Optics (Pergamon, 6th ed. 1980), 24-30.
13. Interpolation of values from D. W. Lynch and W. R. Hunter, "Optical Constants of Metals" in Handbook of Optical Constants of Solids, E. D. Palik, ed., (Academic Press, 1985), Table II, 294.
14. C. W. Haggans, L. Li, T. Fujita, and R. K. Kostuk, "Modeling of lamellar phase retardation and polarization rotation gratings for substrate-mode devices," submitted to J. Opt. Soc. Am. A.
15. A. Malag, "Simple interference method of diffraction grating generation for integrated optics by use of a fresnel mirror," Opt. Commun. 32, 54 (1980).
16. L. Li, M. Xu, G. I. Stegeman, and C. T. Seaton, "Fabrication of photoresist masks for submicrometer surface-relief gratings," SPIE 835, 72-81.

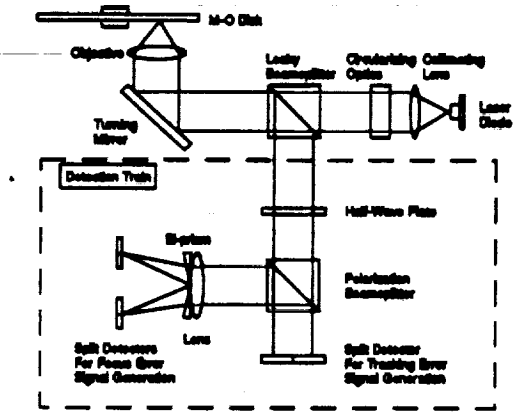


Fig. 1. Conventional M-O disk head layout.

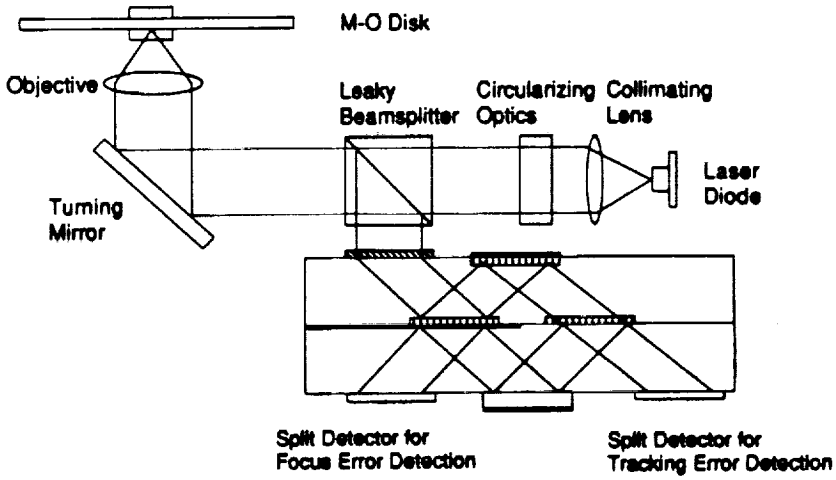


Fig. 2. Proposed M-O head with integrated detection train device.

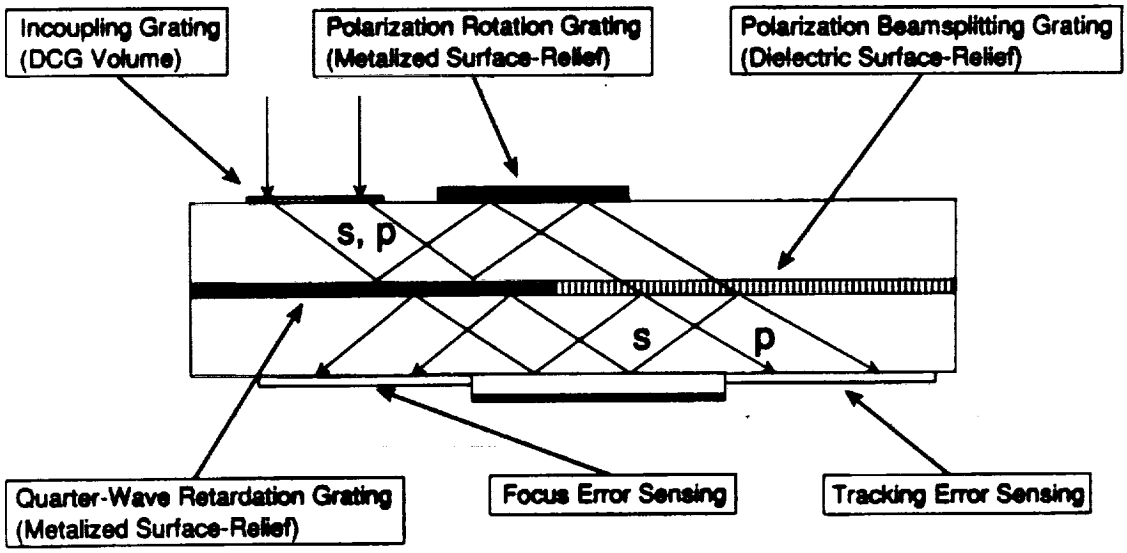


Fig. 3. Proposed detection train device.

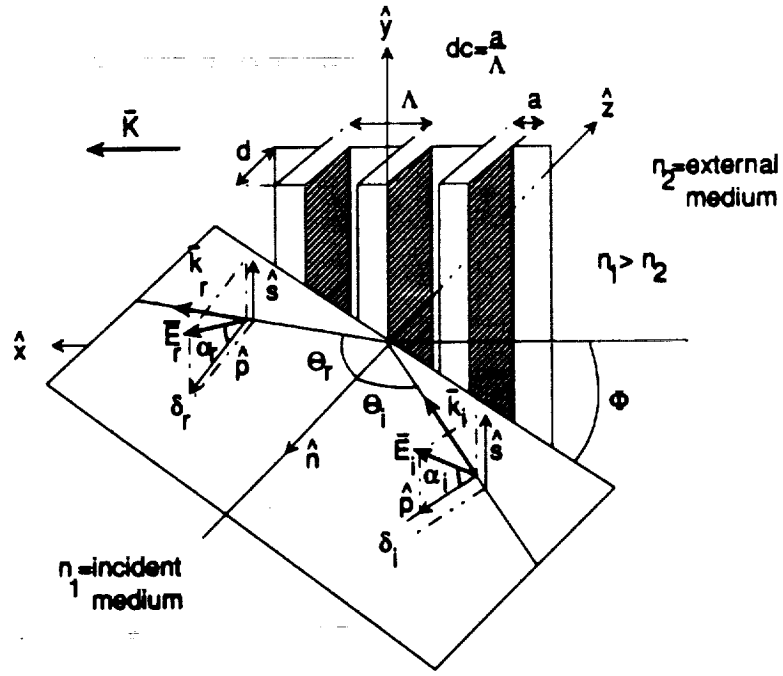


Fig. 4. Grating and incidence geometry.

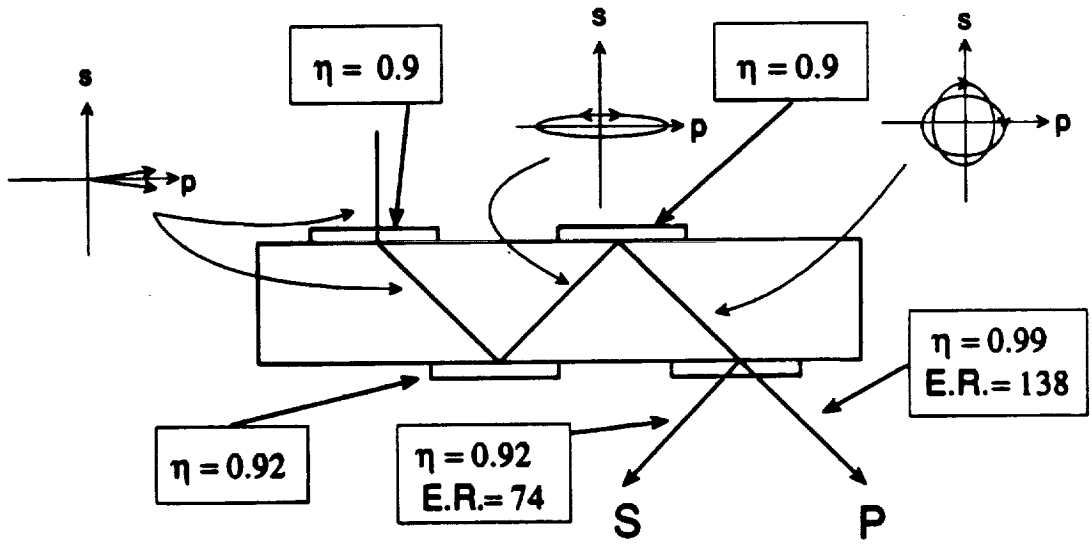


Fig. 5. Modeled device performance. The boxed quantities are individual element efficiencies. The polarization plots show the state of polarization at each point in the device.

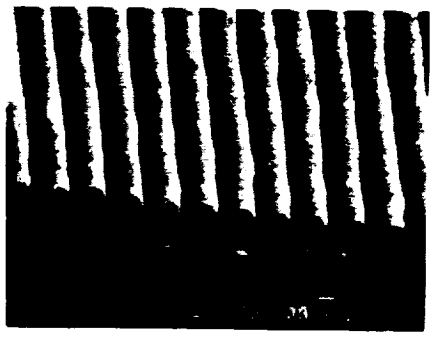


Fig. 6. Scanning Electron Micrograph of a photoresist grating fabricated in our laboratory.

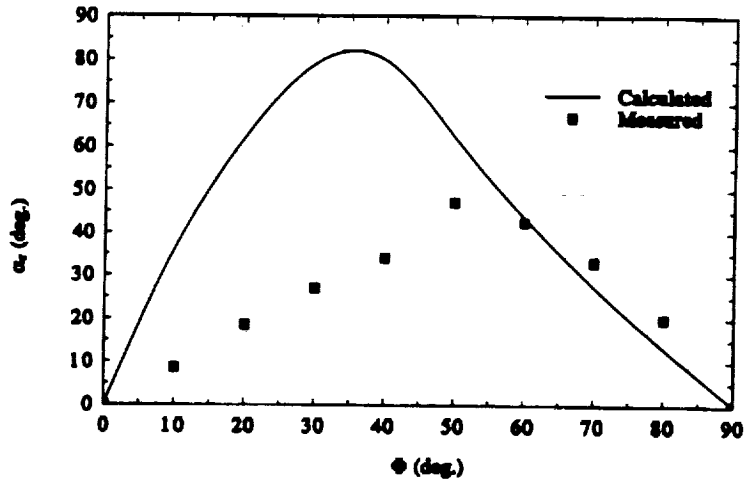


Fig. 7a. Measured and calculated values of the polarization rotation angle after reflection from the grating of Figure 6 (α_i=0°) as a function of the azimuthal angle of incidence.

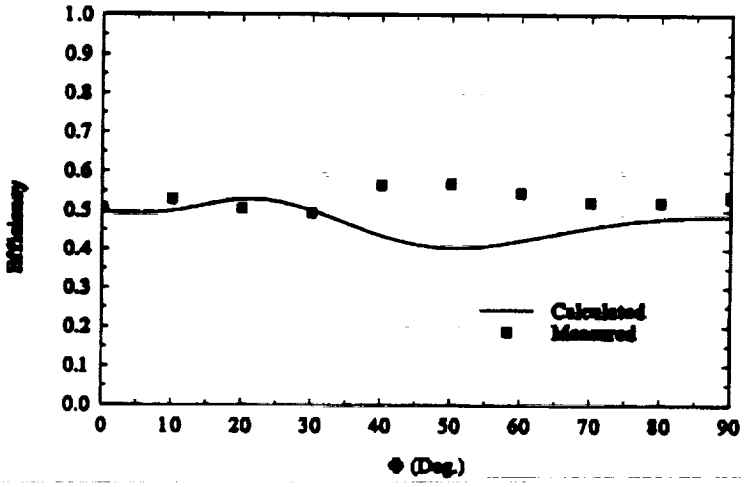


Fig. 7b. Measured and calculated values of the 0th reflected order diffraction efficiency as a function of azimuthal angle of incidence on the grating of Figure 6 .

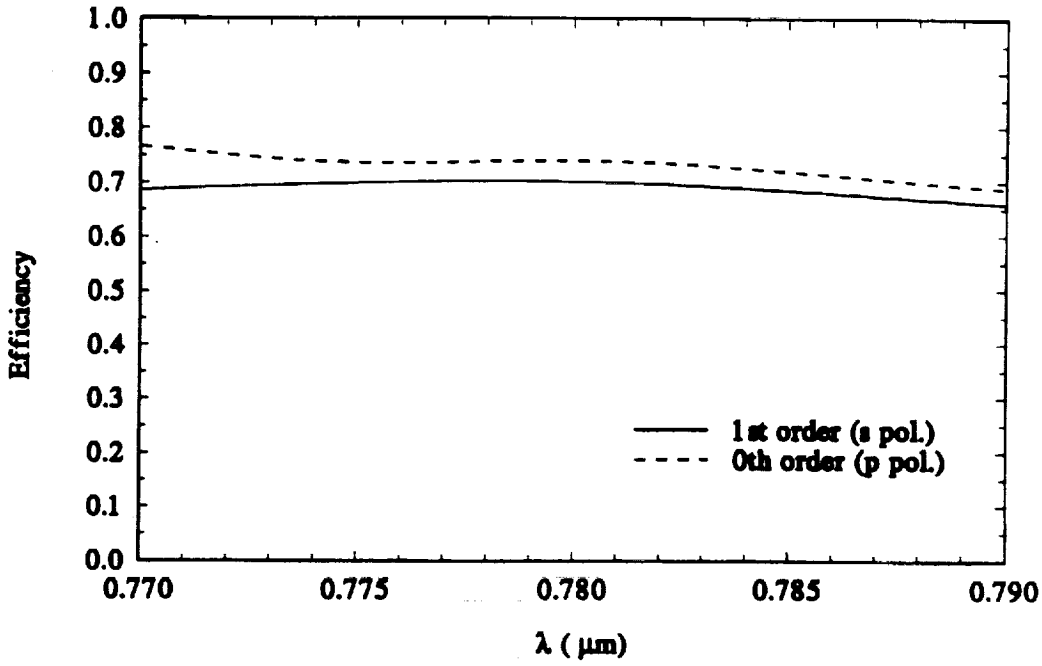


Fig. 8. Throughput efficiency for the device of Fig. 5 as a function of wavelength.

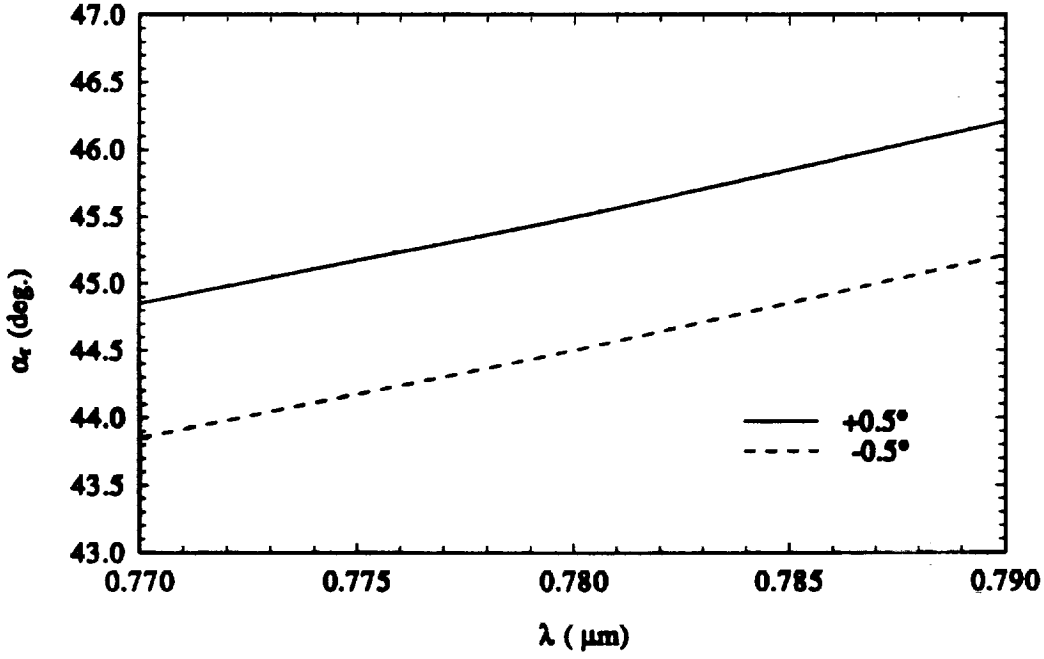


Fig. 9. Polarization rotation angle after Element 3 as a function of wavelength.

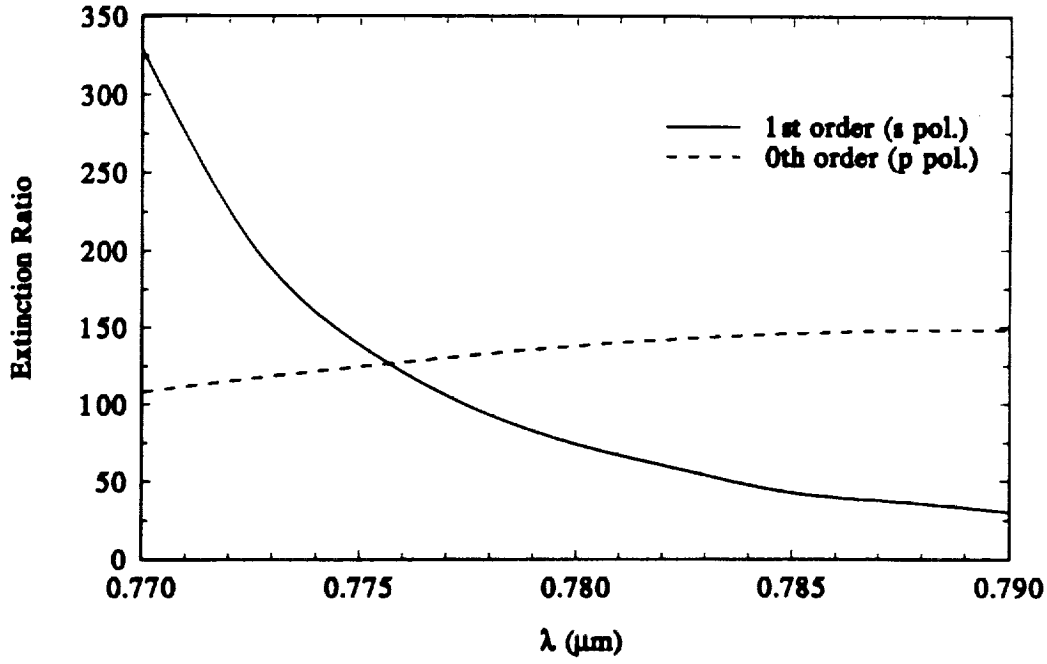


Fig. 10. Extinction ratio for Element 4 as a function of wavelength.

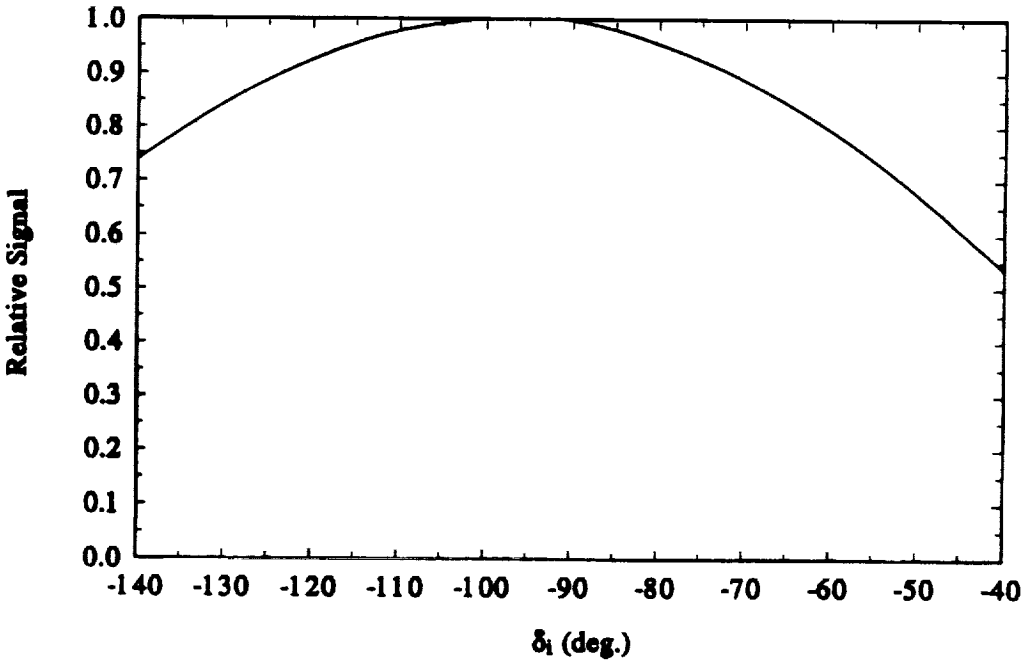


Fig. 11. Fraction of the ±0.5° input polarization rotation angle (input signal) remaining after Element 3 as a function of the phase difference incident on Element 1.

APPENDIX M

Submitted to *Applied Optics*
Feb. 17, 1992

The Design and Fabrication of a Single Leakage-Channel Grating Coupler

Ronald L. Roncone, Lifeng Li, Keith A. Bates*, and James J. Burke

Optical Sciences Center
University of Arizona
Tucson, AZ 85721

Lori Weisenbach, and Brian J. J. Zelinski

Department of Materials Science and Engineering
University of Arizona
Tucson, AZ 85721

ABSTRACT

We describe the modelling and fabrication of waveguide grating couplers with out-coupling efficiencies into a single diffracted order nearing 100%. Termed Single Leakage-Channel Grating Couplers (or SLCGCs), these devices utilize a high-reflectivity dielectric stack to reflect the out-coupled beam diffracted toward the substrate, back up into the air region where it constructively adds with the beam diffracted into the air region. Computer modelling shows that the branching ratio and the leakage rate can be independently controlled, and that the branching ratio is independent of grating depth and grating period. A SLCGC with a branching ratio of 97.1% was fabricated using a combination of vacuum evaporation and wet-chemical techniques.

Key words: grating coupler, thin film, integrated optics, multilayer waveguide

* Now with Storage Products Line of Business, IBM Corp., Tucson, AZ 85744

I. Introduction

The theoretical tools for understanding and analyzing waveguide grating couplers have been well-established for years[1,2]. In addition, numerous techniques exist for the fabrication of waveguide gratings[3,4]. However, several practical issues have limited the commercialization of devices which incorporate grating couplers. These include an undesirable exponentially decaying output beam profile, wavelength sensitivity, and low input and output coupling efficiencies. The first two issues have received considerable attention in the literature[5-9]. The last issue is the primary topic of this paper.

When used as output couplers, typical symmetric-groove gratings split (via diffraction) the guided mode into two beams, of approximately equal power, one going into the cladding and the other going into the substrate. Generally, during device operation only a single out-coupled beam is utilized. As a result about half of the guided mode energy is wasted. Although prism couplers provide 100% efficient out-coupling into a single beam, they are impractical for integrated-optic applications due to their bulkiness and non-planar nature.

As a potential solution to this problem, Peng and Tamir[10] first proposed the use of blazed (or asymmetric groove profile) grating couplers to selectively scatter the modal energy, and showed that these devices can have branching ratios (defined in eqn. 2 in the next section) close to 100%. Many authors[11-13] have reported on the experimental

demonstration of blazed gratings in both microwave and optical regimes, with branching ratios of 90-98.4% having been achieved. Without a doubt, blazing of the grating profile has proven to be an effective technique for fabricating very efficient couplers. However, technological problems persist in generating large groove slopes for near vertical incidence. Moreover (and more importantly), geometric constraints imposed by blazing can create a conflict between a chosen leakage rate, α, and the branching ratio. A typical example of this is illustrated with the help of figure 1.

Consider a typical triangular-profile blazed grating as shown in figure 1, which has period Λ, depth d_g, and slope angle θ. The leakage rate, (generally given in units of cm^{-1}) effectively is the rate at which the energy is coupled out of the guide and thus, it determines the out-coupled beam size. Although α depends on various waveguide parameters, it is most strongly dependent on d_g, and in fact, saturates beyond a certain d_g. The branching ratio typically is a single-maximum function of θ, with the maximum branching ratio generally occurring for θ=40°-60°. The problem arises when the device designer needs to choose an unsaturated value for α. In this case, once α is chosen, d_g becomes fixed. However, the d_g needed to give the desired α likely is not the same d_g required to give the optimal θ (and thus the optimal branching ratio). Thus, the interdependence of α, d_g, θ, and the branching ratio results in a loss of design flexibility. Moreover, note that (from fig. 1) for grating periods in the 0.20-0.50 μm range (typical for integrated optic applications), slope angles of 40°-60° require very deep gratings.

Waveguide grating couplers with branching ratios as high as 96% also have been demonstrated using holographically fabricated slanted index modulation volume phase gratings[14-16] on (or in) a waveguiding structure. However, the long-term stability and mechanical strength of the photosensitive and/or waveguiding materials, in addition to compatibility issues, make the broader application of this technique questionable.

Another recent (and somewhat different) approach[17,18] which yielded branching ratios of 90%, utilized two gratings of the same period, but different depths, separated by a thin coating, and phase shifted with respect to each other, to create an interference effect of the diffracted orders. While this scheme is quite attractive for certain applications, its implementation in a broad range of material systems and device configurations will likely prove to be difficult.

Recently, Agrawal et al.[19] proposed placing a dielectric stack on the substrate to reflect light diffracted toward the substrate back up and out of the device, and thus creating highly efficient coupling into a single diffraction order. However, such a device was never demonstrated. The remainder of this paper will examine such a structure in detail, and present design criteria, modelling and fabrication results, and alternate configurations of the basic device. We will show that this flexible design concept allows for easily implemented device configurations with independent control of the branching ratio and leakage rate, and that it also is applicable to variable depth gratings and chirped gratings. In addition, we will describe a device that has been fabricated using a

combination of vacuum-deposition and wet-chemical techniques. A branching ratio of 97.1% has been measured for this device.

II. Grating Coupler Background

Figure 2 shows a schematic illustration of a typical waveguide grating coupler. The indexes of the cladding, film, and substrate are given respectively by n_c, n_f, and n_s, while the film thickness is designated as d_f. The grating period and grating depth are represented by Λ and d_g. In the output coupling configuration shown, a guided wave incident on the grating region is diffracted (or "leaked" out) into two (leakage) channels, one into the substrate and one into the cladding. The diffraction angle, θ_c, is related to the waveguide and grating parameters by

$$n_c \sin\theta_c = N_{eff} + \frac{\upsilon\lambda}{\Lambda} \qquad (1)$$

where υ is a negative integer (usually -1), λ is the free-space wavelength, and N_{eff} is the effective refractive index of a particular guided mode. Generally, grating couplers with symmetric groove profiles are designed to have only two leakage channels, in an effort to eliminate energy loss into undesired diffraction orders. Typically, the branching ratios in these devices range between 40-60%. The branching ratio, R_i, is defined as the ratio of the power into a particular channel to the sum of the powers into all channels:

$$R_i = \frac{P_i}{\sum_i P_i} \times 100\% \quad . \tag{2}$$

For the remainder of this paper, when the term "branching ratio" is used, it is implied that we mean the branching ratio into the cladding (air) region.

To reduce effective losses during output coupling, it is highly desirable to have most or all of the incident energy diffracted into a single leakage-channel. The following section describes a device, which we will refer to as a Single Leakage-Channel Grating Coupler, or SLCGC, to accomplish such a task.

III. SLCGC Principle of Operation

The objective of the SLCGC structure shown in Figure 3 is to prevent the loss of energy coupled into the substrate by depositing a high-reflectivity quarterwave dielectric stack between the substrate and film to reflect the out-coupled light back toward the air region where it constructively adds with the light originally coupled into the air region. A simplistic ray-optic representation is shown to demonstrate such operation.

The reflectivity of the dielectric stack, with its alternating high (H) and low (L) refractive index quarterwave layers, ultimately determines the maximum branching ratio which can be achieved in a given device. As the number of stack layers increases, or as the refractive index difference between the H and L layers increases, the reflectivity

correspondingly increases.

The buffer layer serves a dual purpose. The first is to act as an auxiliary substrate to isolate the film layer from the potentially lossy stack. Since the stack layers generally have high propagation losses (as compared to the film) predominantly due to scattering at the interfaces (which is aggravated by the large index difference between adjacent layers), and/or due to inferior material qualities, an appropriate thickness is necessary to minimize the guided energy propagating within the stack layers, and to avoid perturbation of the desired propagation constant. The second purpose of the buffer layer is to serve as a phase compensation layer to ensure that constructive addition of the two beams occurs in the air region. As will be shown in the next section, the branching ratio can be strongly dependent on the buffer layer thickness.

IV. Design Considerations / Modelling Results

The theoretical data presented in figures 5-7 are generated by a computer program based on the modal method of Botten et al.[20] for analyzing plane wave diffraction by a rectangular-groove grating. The conversion, or extension of the diffraction grating theory to a waveguide grating theory is straight-forward[21]. The analysis is rigorous in the sense that the only approximation is an unavoidable matrix truncation during numerical implementation. Further details on the theoretical analysis and numerical modelling of the device (in conjunction with model validation) will appear in a future publication.

In the device configuration explicitly discussed in this paper, the SLCGC is a multimode structure, capable of supporting many TE and TM modes. Consider a device with a 9-layer stack (5H + 4L) of alternating H and L quarterwave layers (at λ=0.5145 μm), and having the configuration shown in figure 3, and the parameters shown in table 1. In the numerical modelling, the photoresist (grating) layer is treated as an equivalent uniform medium of thickness 0.100 μm and root-mean-square refractive index (between air, n_c=1.00, and photoresist, n_g=1.66) of 1.370.

Table 2 shows the effective indexes at λ=0.5145 μm for all the TE modes of the structure. The electric field intensity profiles for the TE_3 and TE_4 modes are shown in Figure 4. The high-loss TE_0-TE_2 modes have nearly all of their energy propagating within the stack layers, and thus are not of significant interest here. Note that the TE_3 and TE_4 modes have rather similar mode profiles. Based on the modelling of numerous different device configurations with varying parameters, the similarity in these mode profiles appears to be much more a special case than the general case (but happens to be the case for the device we have fabricated). The TE_4 mode is the desired mode in this structure because more of its energy is confined in the low-loss film layer.

In addition to mode profile considerations, the designer must be careful to ensure that the effective indexes (and thus the coupling angles) of adjacent modes (TE_3 and TE_4 in this case) are adequately separated to prevent inadvertent launching of the wrong mode. The easiest parameter to adjust for this purpose is film thickness.

The choice of the number of H-L layers to include in the stack is largely dependent upon the desired branching ratio, and the indexes of the materials to be used. As the number of layers in the stack increases,the stack reflectivity (and thus the branching ratio) correspondingly increases. The stack reflectivity is both polarization and angle dependent, and thus the grating period (which effectively controls the angle of the light incident on the stack) and the polarization state of the propagating mode become important considerations. The center wavelength of the reflectance/transmittance spectrum of the stack is chosen to coincide with the operating wavelength of the device.

The branching ratio as a function of buffer layer thickness is shown in Figure 5, for devices having either 3-layer or 9-layer stacks. Note the following:

1. The maximum achievable branching ratio depends on the number of layers in the stack, (99.2% for the 9-layer, and 87.4% for the 3-layer) and is periodic with buffer layer thickness.

2. The branching ratio becomes nearly independent of the buffer layer thickness over a broad range, as the number of stack layers increases. This considerably eases the fabrication tolerances on index and thickness control of the individual layers.

3. Although not explicitly shown in the plot, the minimum (practical) buffer layer thickness is dictated by material quality considerations. As d_B decreases, more

guided energy (from the desired mode) is propagated in the stack, and guided-mode losses increase. For example, with the materials we used, a minimum buffer layer of ~0.42 μm was necessary before losses stabilized to 2 dB/cm for the desired TE_4 mode.

Figure 6 and Figure 7 illustrate the flexibility and potential utility of the general design concept. Figure 6 shows the behavior of the branching ratio and the leakage rate, as a function of grating depth, while figure 7 shows the branching ratio as a function of grating period. A buffer thickness of 0.42 μm is assumed for both plots. Notice that in both cases, the branching ratio effectively is independent of the grating parameters. This has two important implications:

1. The design concept allows for independent control of both the leakage rate (which is strongly dependent on grating depth) and thus of out-coupled beam size, and the branching ratio.

2. Because the branching ratio is insensitive to the grating parameters, the design concept also should work with variable groove depth and chirped (variable groove spacing) gratings. As a result, high branching ratios can be engineered into grating devices which also possess beam-shaping and focussing capabilities.

V. SLCGC Fabrication / Results

A successful device exhibiting a 97.1% branching ratio into air has been fabricated using a combination of vacuum-evaporation and wet-chemical (sol-gel) techniques. A 9-layer stack, with alternating quarterwave layers of TiO_2 (n_H=2.38) and SiO_2 (n_L=1.46), and a center wavelength of λ=0.5145 μm, was deposited on a 2"x1"x1 mm fused silica substrate using vacuum evaporation[22]. After sonic cleaning the stack for 5 min in pure, dehydrated ethanol to prepare the surface for further processing, the stack was fired at 500° C for 30 min. The furnace bake was performed in an attempt to further densify the structure, and to minimize microstructural changes and stresses in the stack incurred during the subsequent heat-processing of the sol-gel buffer and film layers. The 500° C heat treatment shifted the stack spectrum ~10 nm toward the blue, probably due to a slight densification of the materials. Figure 8 shows the measured reflectance spectrum of the post-bake 9-layer stack, at a 5° angle of incidence.

The buffer layer and film layer were fabricated using wet-chemically derived solutions[23,24] spin-coated in a class 100 clean-room environment. The buffer layer was a 13:87 mol% TiO_2:SiO_2 solution, yielding amorphous glass films of refractive index 1.49. To achieve the desired buffer layer thickness, two layers were deposited. The film layer was fabricated by spin-coating a 35:65 mol% TiO_2:SiO_2 solution which yielded an amorphous glass film of refractive index 1.663 and thickness 0.20 μm (in a single layer). Each sol-gel layer was baked at 500° C for 30 min. Using these materials and buffer

layer thicknesses above ~0.42 μm, the loss of the TE_4 mode was 2.0 dB/cm at λ=0.5145 μm. As a comparison, the loss of the TE_3 mode was 6.8 dB/cm (recall the field profiles of figure 4). Entire 5 and 7-layer SLCGCs have been fabricated using the sol-gel materials, but these results are reported elsewhere[24]. It is interesting to note that the entire structure can be fabricated in only two materials, and by using any of the common thin-film deposition methods, or combinations thereof (as long as the materials are of sufficiently low loss).

The 0.30 μm period grating, designed to give a ~10° angle for the out-coupled beams, was fabricated using a holographic technique. Shipley 1805 positive photoresist was diluted 1:1 with Shipley type P thinner, to yield 0.1000 μm deep coatings (on top of the multilayer structure) when spun at 5800 rpm for 30 sec. The grating exposure was carried out using the 0.4579 μm line of an etalon-tuned Argon-ion laser. After this first exposure, a transparent mask (on a 2"x1"x1 mm glass slide) with two rectangular (15 mm x 4 mm) opaque aluminized regions separated by 1 cm, was placed against the exposed photoresist grating. The resulting mask/grating sandwich was held in front of a xenon-arc lamp for 60 sec to completely expose all regions of the photoresist, except for the two rectangular regions, and thus to delineate an input and output grating. At this point, an in-situ developing technique[25] was utilized to ensure that the resulting gratings were developed down to the film. The entire structure then was baked at 110° C for 30 min to harden the photoresist gratings.

To test the SLCGC, the 0.5145 μm line of an Argon-ion laser was run through a Glan-Thompson polarizer to give a TE polarized beam, and then gently focused (using a 200 mm focal length plano-convex lens) onto the input grating. The SLCGC was rotated and translated until the TE_4 mode was optimally coupled in. A Si PIN photodiode, connected to a picoammeter, was placed 5 mm from the output grating (first on the air side, and then on the substrate side) to capture the out-coupled light and convert it to a current. Care was taken to reduce background radiation. The resulting current on the picoammeter was used to calculate the branching ratio as follows:

$$R_{air} = \frac{I_{air}}{I_{air} + I_{sub}} \cdot 100\% \qquad (3)$$

where I_{air} and I_{sub} are the currents generated in the photodiode by the out-coupled light. Using this procedure, an average (over several locations on the device) branching ratio of 97.1% (±1%) was measured. The theoretical branching ratio from Fig. 5 (indicated by the "x") is 96.2%.

Considering the somewhat limited accuracy with which we know the buffer layer thickness (±0.02 μm), and that the theoretical design point shown by the "x" in figure 5 is not on a flat region of that plot, the agreement between the theoretical and experimentally measured branching ratio is surprisingly good. While this single sample is insufficient to validate the computer model, it does demonstrate the validity and usefulness of the design concept. Work continues in the fabrication of more devices in

order to obtain enough data points to validate the model.

VI. Future Applications

By extending the work reported in this paper, several interesting and potentially useful device configurations can be envisioned:

- As mentioned previously, the basic SLCGC design can be combined with variable depth gratings to form efficient couplers with beam-shaping capabilities, or with chirped gratings to form efficient focusing elements.

- Although only a simple broadband quarterwave stack design was utilized here, the reflecting stack easily could be designed to perform filtering and/or polarizing functions as well.

- Instead of placing the buffer layer and reflector under the guiding film and grating, they could be placed on top of the grating region. This would allow for the preferred single-mode device operation, while retaining a high branching ratio in structures where it is not possible to place the stack between the substrate and guiding region, such as in ion-exchanged or ion-diffused waveguides.

VII. Conclusions

We have modelled the characteristics of the SLCGC proposed by Agrawal et al.[19] and have fabricated the device using vacuum deposition and wet-chemical techniques. The branching ratio of the SLCGC has been measured to be 97.1%, much greater than the typical values of 40-60% exhibited by unmodified output couplers. Modelling indicates that the SLCGC branching ratio is independent of grating depth and period, but depends upon the buffer layer thickness when the number of layers in the reflective stack is small. As the number of stack layers increases, the branching ratio becomes independent of the buffer layer thickness over a broad range. These modelling results indicate that fabrication tolerances on the layer and grating components of the SLCGC are quite broad. They also indicate that the SLCGC design concept can be applied to more sophisticated grating applications including focussing and beam-shaping elements.

VIII. Acknowledgements

We express thanks to John O'Kelly, and Detrick Branston for their work with the sol-gel solution preparation and characterization. We also thank the Donnelly Corporation for performing the spectral characterization of the dielectric stacks. This work is sponsored in part by the Optical Data Storage Center at the University of Arizona. R.L. Roncone acknowledges the financial support of a United States Department of Education GAANN fellowship.

REFERENCES

1) T. Tamir (ed.): *Integrated Optics*, Topics in Applied Physics **7**, (Springer, New York, Heidelberg, Berlin 1975), Chap. 3.

2) T. Tamir and S.T. Peng, "Analysis and Design of Grating Couplers," Appl. Phys. **14**, 235-254 (1977).

3) M.L. Dakss, L. Kuhn, P.F. Heidrich, and B.A. Scott, "Grating coupler for efficient excitation of optical guided waves in thin films," Appl. Phys. Lett. **16**, 523-525 (1970).

4) M. Xu, R. Moshrefzadeh, U.J. Gibson, G.I. Stegeman, and C.T. Seaton, "Simple method for fabricating guided-wave gratings," Appl. Opt. **24**, 3155-3161 (1985).

5) S. Miyanaga, and T. Asakura, "Intensity Profile of Outgoing Beams From Uniform and Linearly Tapered Grating Couplers," Appl. Opt. **20**, 688-695 (1981).

6) A. Saad, H. Bertoni, and T. Tamir, "Beam Scattering by Nonuniform Leaky-Wave Structures," Proc. of IEEE **62**, 1552-1561 (1974).

7) L. Li, K.A. Bates, J.K. Erwin, R.L. Roncone, and J.J. Burke, "Planar Waveguide Grating Couplers With Variable Depth," in OSA Annual Meeting Technical Digest,

1991 (Optical Society of America, Washington, D.C., 1991), vol 17, p. 26.

8) K.E. Spaulding and G.M. Morris, "Achromatic Waveguide Input/Output Coupler Design," Appl. Opt. **30**, 1096-1112 (1991).

9) M.C. Gupta and L. Li, "Achromatic Compensation for Integrated Optic Grating Couplers With Focused Beams," Appl. Opt. **30**, 1461-1463 (1991).

10) S.T. Peng and T. Tamir, "Directional Blazing of Waves Guided by Asymmetrical Dielectric Gratings," Opt. Commun. **11**, 405-409 (1974).

11) T. Aoyagi, Y. Aoyagi, and S. Namba, "High-Efficiency Blazed Grating Couplers," Appl. Phys. Lett. **29**, 303-304 (1976).

12) A. Gruss, K.T. Tam, and T. Tamir, "Blazed Dielectric Gratings with High Beam-Coupling Efficiencies," Appl. Phys. Lett. **36**, 523-525 (1980).

13) Q. Gong, *Input and Output Waveguide Grating Couplers: Theory, Fabrication, and Testing*, Ph. D. dissertation, University of Arizona (1990).

14) H. Kogelnik, and T.P. Sosnowski, "Holographic Thin Film Couplers," Bell Syst. Technol. J. **49**, 1602-1608 (1970).

15) W. Driemeier, "Coupled-Wave Analysis of the Bragg Effect Waveguide Coupler," J. of Mod. Opt. **38**, 363-377 (1991).

16) W. Driemeier, "Bragg-Effect Grating Couplers Integrated in Multicomponent Polymeric Waveguides," Opt. Lett. **15**, 725-727 (1990).

17) I.A. Avrutsky, A.S. Svakhin, and V.A. Sychugov, "Interference Phenomena in Waveguides with Two Corrugated Boundaries," J. of Mod. Opt. **36**, 1303-1320 (1989).

18) I.A. Avrutsly, A.S. Svakhin, V.A. Sychugov, and O. Parriaux, "High-Efficiency Single-Order Waveguide Grating Coupler," Opt. Lett. **15**, 1446-1448 (1990).

19) N. Agrawal, T.C. Strand, and P. May, "High Directivity Waveguide Grating Couplers for Optical Storage," CLEO Proc., 1990, 40.

20) L.C. Botten, M.S. Craig, R.C. McPhedran, J.L. Adams, and J.R. Andrewartha, "The Dielectric Lamellar Diffraction Grating," Optica Acta **28**, 413-428 (1981).

21) M. Neviere, "The Homogeneous Problem," in *Electromagnetic Theory of Gratings*, ed. R. Petit (Springer-Verlag, Berlin, 1980).

22) Stacks provided by: Evaporated Coatings, Inc., 2365 Maryland Road, Willow Grove, PA 19090.

23) L. Weisenbach, B.J.J. Zelinski, J. O'Kelly, J. Morreale, R.L. Roncone, and J.J. Burke, "The Influence of Processing Variables on the Optical Properties of SiO_2-TiO_2 Planar Waveguides," to be published in SPIE Proc. of OE/Fibers 1991.

24) L. Weisenbach, B.J.J. Zelinski, R.L. Roncone, L. Li, and J.J. Burke, "Wet-Chemical fabrication of Single Leakage-Channel Waveguide Grating Couplers," submitted to Thin Solid Films, Feb., 1992.

25) L. Li, M. Xu, G.I. Stegeman and C.T. Seaton, "Fabrication of photoresist mask for submicrometer surface relief gratings," Proc. SPIE **835**, 72-82 (1987).

FIGURE CAPTIONS

Figure 1 - Configuration of a typical triangular-groove profile blazed grating.

Figure 2 - Schematic illustration of a symmetric-groove profile waveguide grating coupler, and it's basic operation.

Figure 3 - Ray-optic schematic illustration of a 9-layer stack SLCGC, and it's principle of operation for achieving high branching ratios.

Figure 4 - Transverse electric field intensity profiles for the device illustrated in figure 3, and with the parameters given in Table 1, for the TE_3 (dashed line) and TE_4 (solid line) guided modes. The vertical lines represent the boundaries between adjacent layers. The waveguiding layer is centered at 0.0. The horizontal axis is in units of μm.

Figure 5 - Plot of branching ratio (%) vs. buffer layer thickness (μm) for devices having 3 layers (dashed line) and 9-layers (solid line) in the dielectric stack. Devices having 5 and 7 layer stacks fall between these two plots. The 'x' indicates the point where the 9-layer stack device falls.

Figure 6 - Plot of branching ratio (%) and leakage rate (cm^{-1}) vs. grating depth (μm), for a 9-layer stack SLCGC. The solid line is the branching ratio. The dashed line is the leakage rate. The arrows indicate which vertical axis corresponds to each curve. Note that branching ratio effectively is independent of both grating depth, and leakage rate.

Figure 7 - Plot of branching ratio (%) vs. grating period (μm), for a 9-layer stack SLCGC, showing that branching ratio effectively is independent of the grating period.

Figure 8 - Plot of measured reflectance vs. λ for a 9-layer stack, after baking at 500° C for 30 min. Peak reflectivity is 97.5% in the range 500-507 nm.

Table 1 - Modelling and fabrication parameters for the SLCGC structure shown in figure 3.

Table 2 - Computed effective indexes for the TE modes (at λ=0.5145 μm) of a nine-layer stack SLCGC (shown in figure 3) having the parameters listed in table 1.

n_s=1.461

n_H=2.380 d_H=0.0525 μm

n_L=1.460 d_L=0.0856 μm

n_B=1.49

Table 1

n_f=1.663 d_f=0.200 μm

n_g=1.66 d_g=0.10 μm

Λ=0.30 μm λ=0.5145 μm

Mode	N_{eff}
TE_0	1.8870
TE_1	1.8204
TE_2	1.7056
TE_3	1.5460
TE_4	1.5328

Table 2

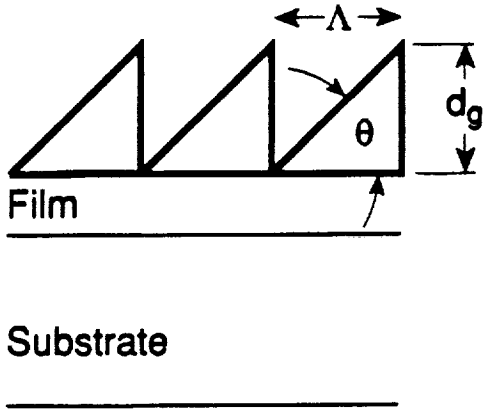

Λ
θ
d_g
Film
Substrate

Fig. 1

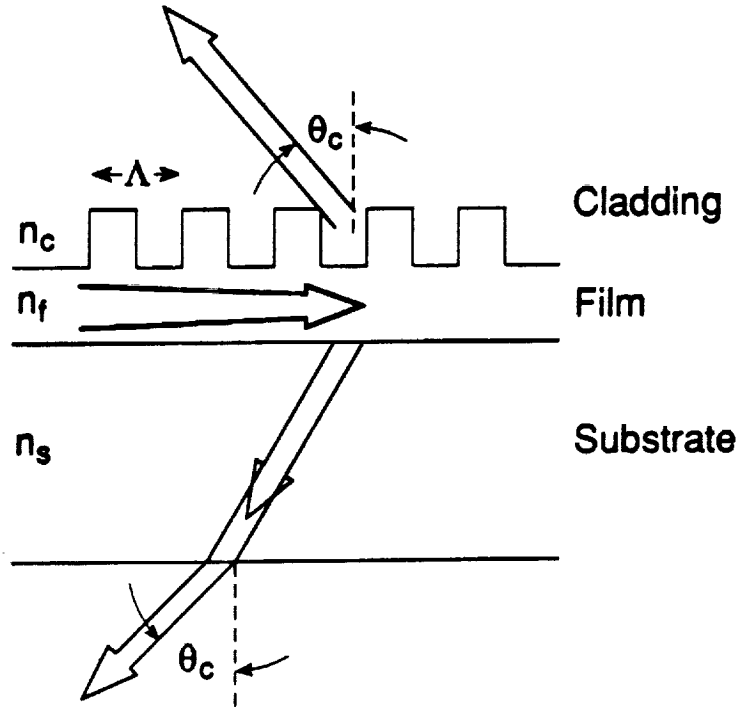

θ_c
Λ
Cladding
n_c
n_f
Film
n_s
Substrate
θ_c

Fig. 2

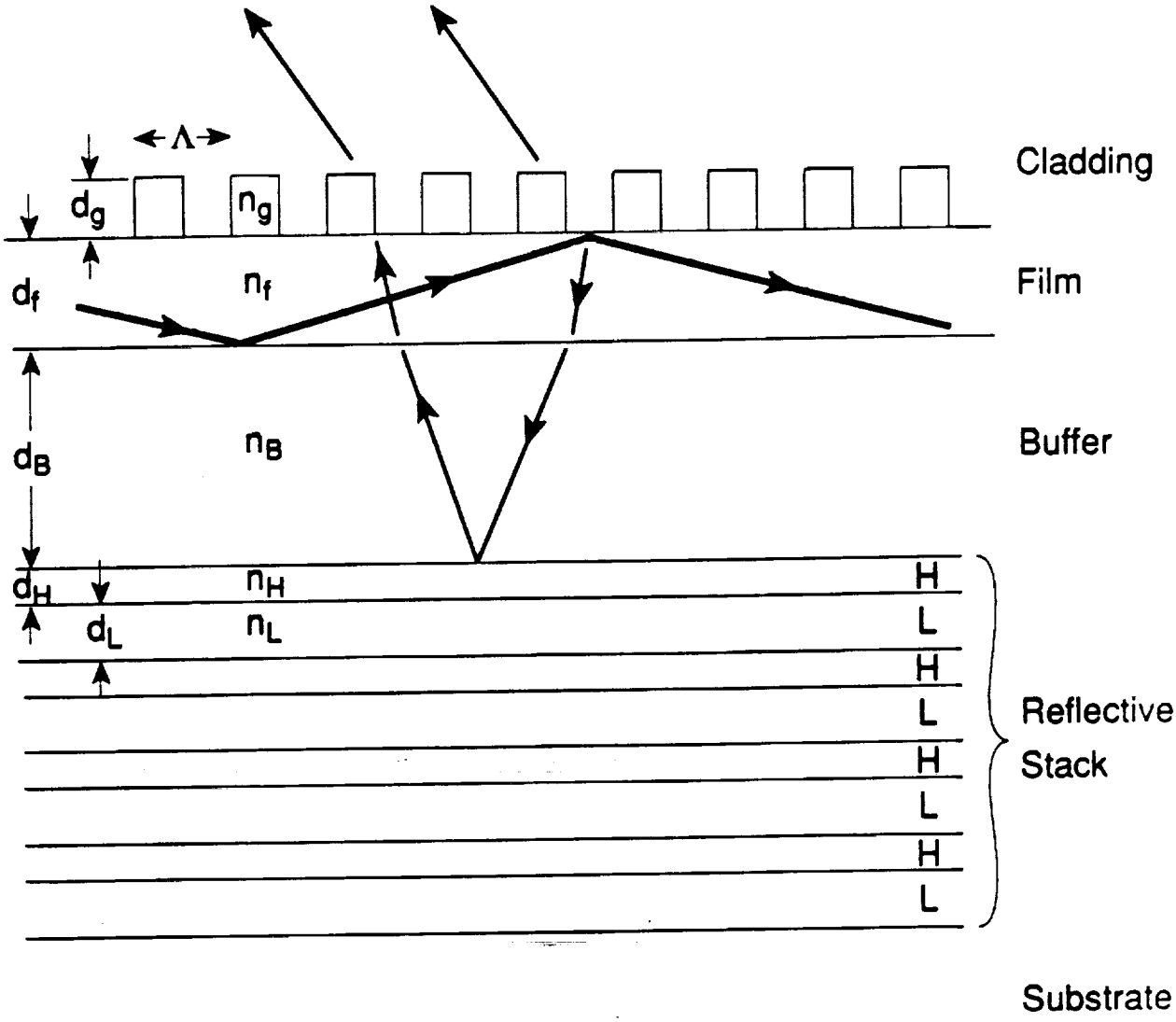

Λ
d_g
n_g
Cladding
d_f
n_f
Film
d_B
n_B
Buffer
d_H
n_H
H
d_L
n_L
L
H
L
Reflective
H
Stack
L
H
L
Substrate

Fig. 3

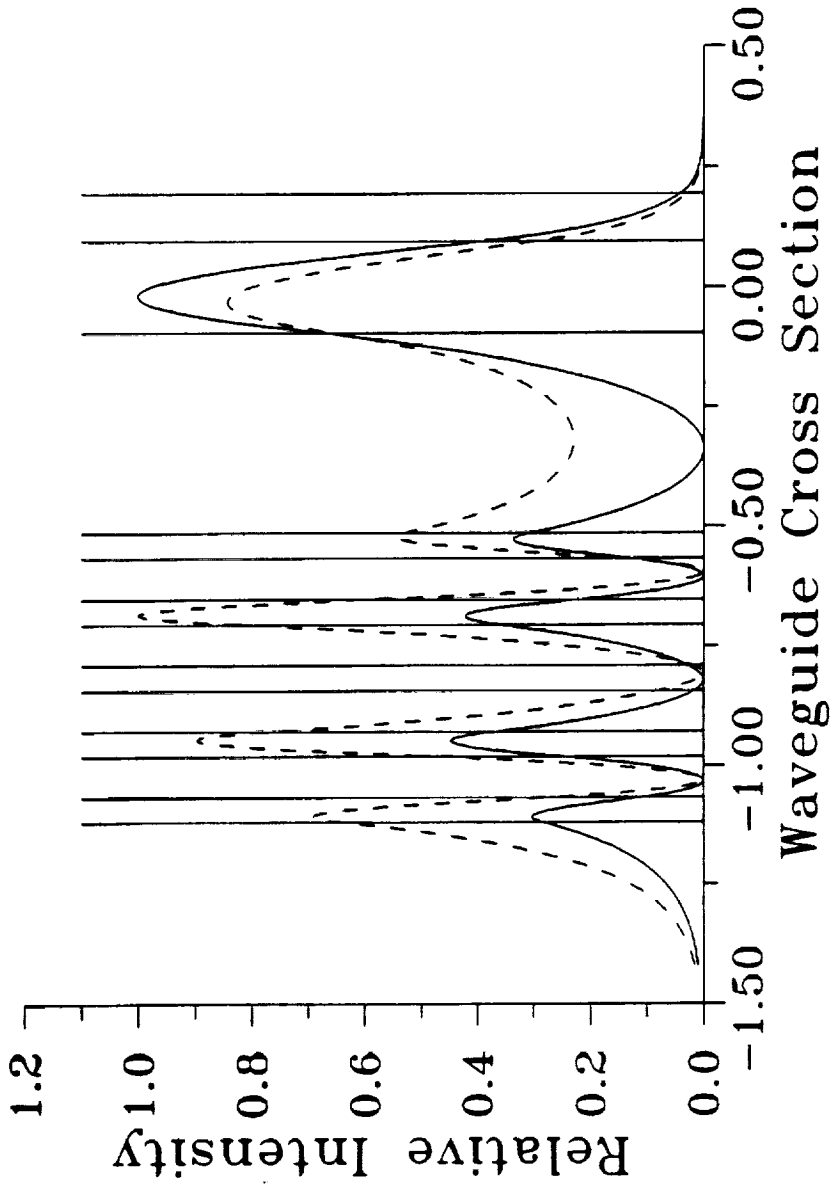

Relative Intensity
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−1.50
−1.00
−0.50
0.00
0.50
Waveguide Cross Section

Fig. 4

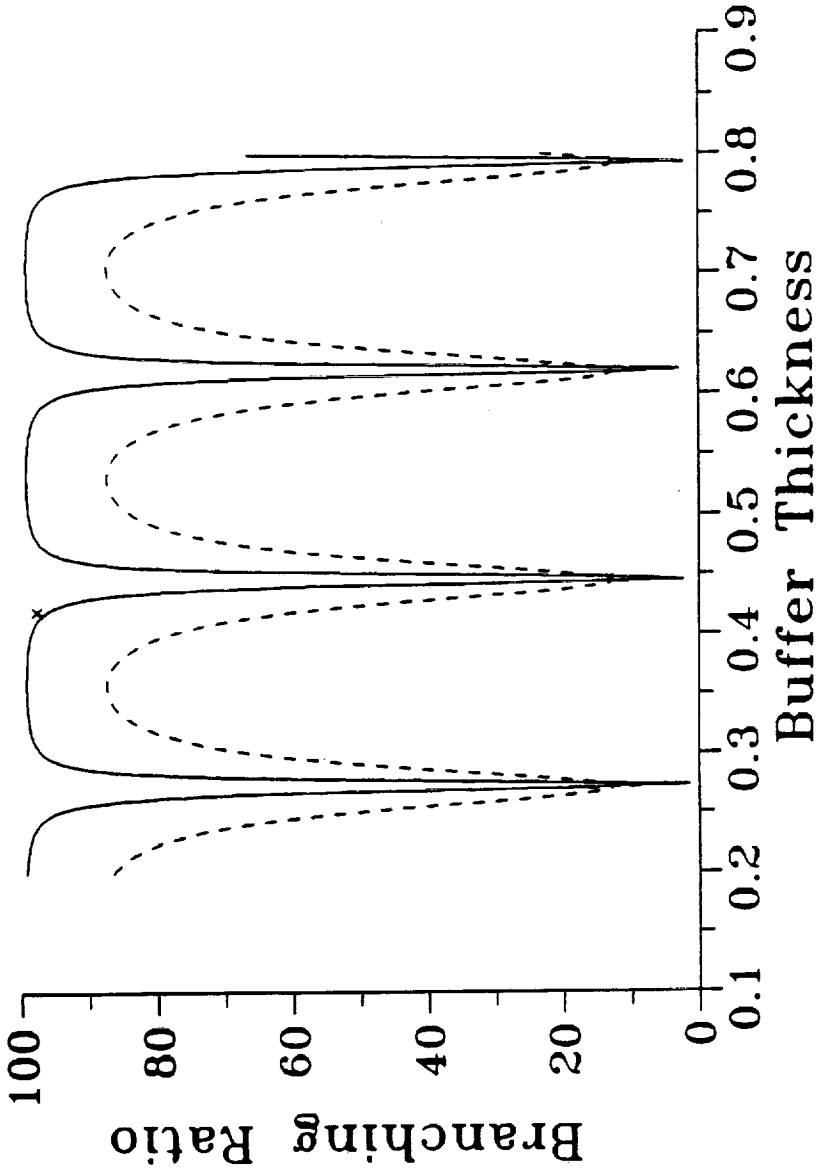

Branching Ratio
100
80
60
40
20
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Buffer Thickness

Fig. 5

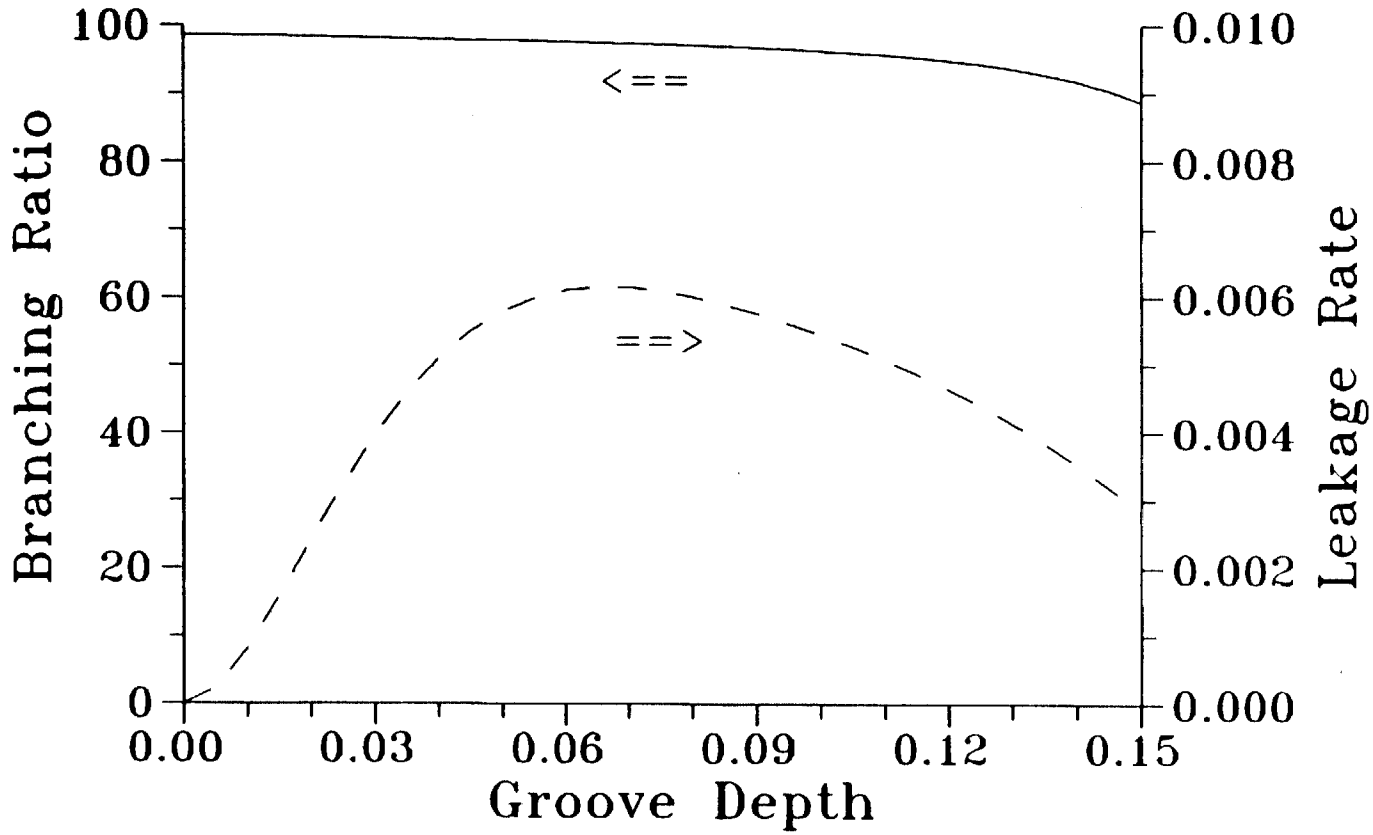

100
80
60
40
20
0
Branching Ratio
<==
==>
0.010
0.008
0.006
0.004
0.002
0.000
Leakage Rate
0.00
0.03
0.06
0.09
0.12
0.15
Groove Depth

Fig. 6

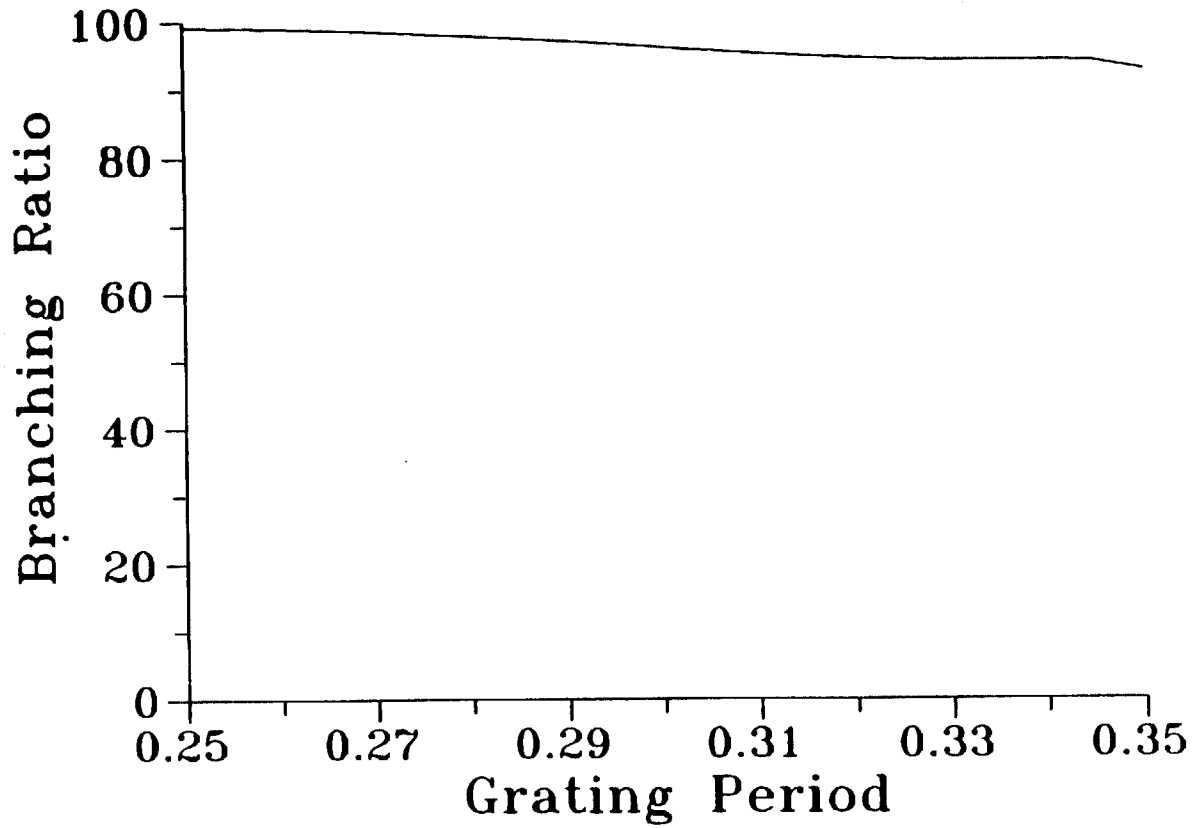

Branching Ratio
100
80
60
40
20
0
0.25
0.27
0.29
0.31
0.33
0.35
Grating Period

Fig. 7

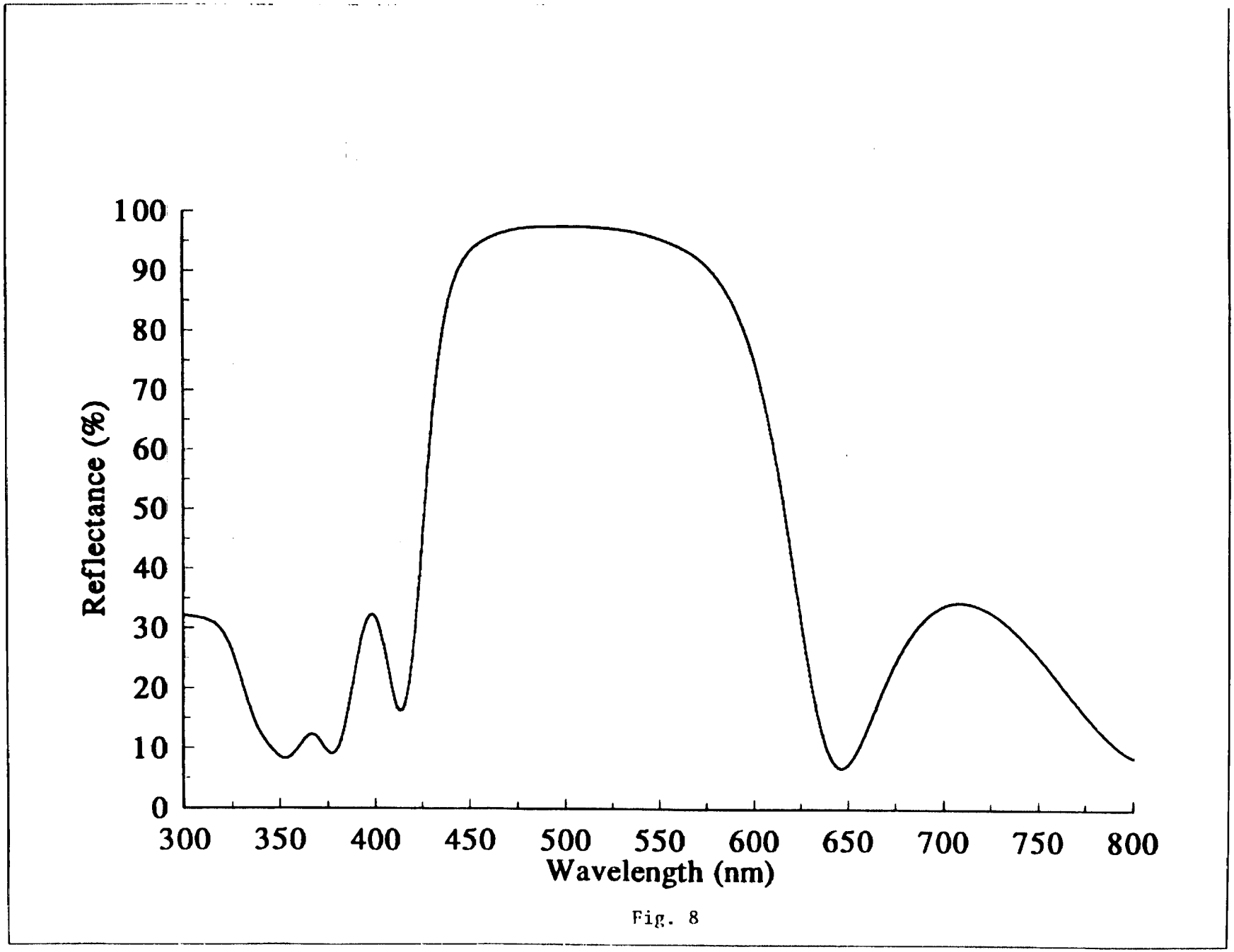

Reflectance (%)
100
90
80
70
60
50
40
30
20
10
0
300
350
400
450
500
550
600
650
700
750
800
Wavelength (nm)

Fig. 8

APPENDIX N

Submitted to the Special Issue on Diffractive Optics,
Journal of Modern Optics, Feb. 28, 1992

A MODAL ANALYSIS OF LAMELLAR DIFFRACTION GRATINGS IN CONICAL MOUNTINGS

Lifeng Li

Optical Sciences Center

University of Arizona

Tucson, AZ, 85721, U.S.A.

ABSTRACT. A rigorous modal analysis of lamellar gratings, i.e., gratings having rectangular grooves, in conical mountings is presented. It is an extension of the analysis of Botten *et al.* which considered non-conical mountings. A key step in the extension is a decomposition of the electromagnetic field in the grating region into two orthogonal components. A computer program implementing this extended modal analysis is capable of dealing with plane wave diffraction by dielectric and metallic gratings with deep grooves, at arbitrary angles of incidence, and having arbitrary incident polarizations. Some numerical examples are included.

1. INTRODUCTION

The modal approach has been applied by many authors to lamellar, non-perfectly conducting gratings in the past [1-10]. Most noticeably, Botten *et al.* [5-7] presented a series of three papers, in 1981, on the modal analysis of dielectric, finitely conducting, and highly

conducting lamellar gratings. Their work was later formulated in a more systematic way, and its certain numerical aspects were improved, by Suratteau *et al.* [8] and by Tayeb and Petit [9]. However, these analyses are limited to non-conical mountings. In many applications, lamellar gratings are used in conical mountings [11]. Recently, a modal analysis of lamellar gratings in conical mountings was presented by Peng [12].

The present work differs substantially from that of Peng [12] in mathematical formulation and numerical implementation. In Peng's work, the validity of the decomposition of the electromagnetic field in the corrugated region into two orthogonal components is assumed without proof. The eigen-functions (the modal fields in the corrugated region) for a grating in a conical mountings are constructed by geometrical means from the TE and TM eigen-functions for the grating in a equivalent non-conical (also called classical) mounting. The completeness and orthogonality of the eigen-functions assembled in this manner are not addressed. In this paper, the eigen-functions along with their completeness and orthogonality are derived rigorously and systematically from the boundary-value problems.

The present work can be considered as an extension of the works of Botten *et al.* and Suratteau *et al.* to conical mountings. A key step in its development is the proof of the field decomposition mentioned above. Once this is done, the task of finding eigenvalues and eigen-functions for a conical mounting reduces to that of a classical mounting, and the previous results of the above authors, including their powerful and sophisticated numerical

methods for finding the eigenvalues, can be used.

The mathematical formulation of the modal analysis is presented in section 2, and the numerical aspects of the analysis are addressed in section 3. Section 4 provides some numerical results. The validity of the field decomposition is proved in Appendix A. For the sake of the normal flow of the paper, some of the results available in Ref. 5-9 are re-stated in sections 2 and 3, but often are formulated differently. In other instances, the reader is referred to the original references.

2. MATHEMATICAL FORMULATION

2.1. *Notation*

A lamellar grating in a conical diffraction configuration is depicted in Fig. 1. The coordinate system is chosen such that the x-axis is perpendicular to, and the z-axis is parallel to, the grating grooves, and the y-axis is the normal of the overall structure. A monochromatic plane wave of vacuum wavelength λ_0 is incident on the lamellar grating at a polar angle θ and an azimuthal angle ϕ. The range of θ is $0 \le \theta < \pi/2$ and that of ϕ is $-\pi < \phi \le \pi$, with the clockwise direction being the positive direction for ϕ. The same conventions will be used for the diffracted waves that are not shown. The incident polarization is in general elliptical.

The geometry of the lamellar grating is shown in Fig. 2. The grating period is d, and the widths of medium 1 and medium 2 are d_1 and d_2. We shall call the regions of space

where $y > h/2$, $y < -h/2$, and $-h/2 < y < h/2$, regions 1, 2, and 0, respectively, where h is the grating groove depth. We shall use the superscript (j), where $j = 1, 2$, to denote quantities associated with regions j and the subscript j to denote quantities associated with the two media in region 0. Thus, the permittivity and permeability of the medium in region 1 are $\epsilon^{(1)}$ and $\mu^{(1)}$, and those in region 2 are $\epsilon^{(2)}$ and $\mu^{(2)}$. The permittivity and permeability of region 0 are periodic functions of x,

$$\left.\begin{aligned} \epsilon(x) &= \epsilon_1, \quad \mu(x) = \mu_1, \quad 0 \le |x| \le d_1/2, \\ \epsilon(x) &= \epsilon_2, \quad \mu(x) = \mu_2, \quad d_1/2 < |x| \le d/2. \end{aligned}\right\} \tag{1}$$

Although for most optical applications, the permeability is a constant and equals that of the vacuum, to reveal the symmetry of the electric and magnetic fields, μ_1 and μ_2 are formally assumed to be different. The Gaussian system of units is used in this paper.

Let k_0 be the magnitude of the vacuum wave-vector. The magnitudes of the wave-vector in regions $j = 1, 2$ and region 0 are denoted by $k^{(j)}$ and $k(x)$,

$$k^{(j)2} = \epsilon^{(j)} \mu^{(j)} k_0^2, \tag{2}$$

$$k^2(x) = \epsilon(x)\, \mu(x)\, k_0^2. \tag{3}$$

The wave-vector of the incident plane wave is

$$\vec{K} = k^{(1)} (\hat{x} \sin\theta \cos\phi - \hat{y} \cos\theta + \hat{z} \sin\theta \sin\phi). \tag{4}$$

We denote the z-component of the incident wave-vector by k_z,

$$k_z = k^{(1)} \sin\theta \sin\phi. \tag{5}$$

and we define the reduced magnitudes of the wave-vectors by $\bar{k}^{(j)}$ and $\bar{k}(x)$,

$$\tilde{k}^{(j)2} = k^{(j)2} - k_z^2, \tag{6}$$

$$\tilde{k}^2(x) = k^2(x) - k_z^2. \tag{7}$$

Since the physical problem is time and z-invariant, the electromagnetic field may be assumed to be of the form

$$E(x,y,z,t) = \vec{E}(x,y)\, e^{ik_z z - i\omega t}, \tag{8a}$$

$$H(x,y,z,t) = \vec{H}(x,y)\, e^{ik_z z - i\omega t}. \tag{8b}$$

Throughout this paper, $i = \sqrt{-1}$. Substituting (8a) and (8b) into Maxwell's equations, we may express the transverse components of the electromagnetic field in terms of the longitudinal (z) components,

$$\vec{E}_t(x,y) = \frac{i}{\tilde{k}^2(x)} \left[k_z \nabla_t E_z(x,y) - \mu k_0\, \hat{z} \times \nabla_t H_z(x,y) \right], \tag{9a}$$

$$\vec{H}_t(x,y) = \frac{i}{\tilde{k}^2(x)} \left[k_z \nabla_t H_z(x,y) + \epsilon k_0\, \hat{z} \times \nabla_t E_z(x,y) \right], \tag{9b}$$

where $\nabla_t = \hat{x}\partial_x + \hat{y}\partial_y$. So in solving the conical diffraction problem, it is only necessary to work with the z-components of the electric and magnetic fields.

As usual, the fields above and below the corrugated region may be written in Rayleigh expansions [13]:

$$E_z(x,y) = I_z^{(e)} e^{i\alpha_0 x - i\beta_0^{(1)} y} + \sum_{n=-\infty}^{+\infty} R_n^{(e)} e^{i\alpha_n x + i\beta_n^{(1)} y}, \tag{10a}$$

$$H_z(x,y) = I_z^{(h)} e^{i\alpha_0 x - i\beta_0^{(1)} y} + \sum_{n=-\infty}^{+\infty} R_n^{(h)} e^{i\alpha_n x + i\beta_n^{(1)} y} \tag{10b}$$

for $y > h/2$, and

$$E_z(x,y) = \sum_{n=-\infty}^{+\infty} T_n^{(e)} e^{i\alpha_n x - i\beta_n^{(2)} y}, \tag{11a}$$

$$H_z(x,y) = \sum_{n=-\infty}^{+\infty} T_n^{(h)} e^{i\alpha_n x - i\beta_n^{(2)} y} \tag{11b}$$

for $y < h/2$, where

$$\alpha_n = \alpha_0 + 2n\pi/d, \qquad \alpha_0 = k^{(1)} \sin\theta \cos\phi, \tag{12}$$

$$\beta_n^{(j)2} = k^{(j)2} - \alpha_n^2, \qquad Re[\beta_n^{(j)}] + Im[\beta_n^{(j)}] > 0. \tag{13}$$

In (10) and (11), $I_z^{(e)}$, $I_z^{(h)}$, $R_n^{(e)}$, $R_n^{(h)}$, $T_n^{(e)}$, and $T_n^{(h)}$ are, respectively, the complex amplitudes of the z-components of the incident and diffracted electric and magnetic fields in regions 1 and 2.

2.2. *Field Decomposition*

Suppose h tends to infinity in Fig. 2, yielding a medium that is periodic in x and infinite in y and z. If the general expression for the electromagnetic field in this infinite medium is known, by imposing the interface conditions at $y = \pm h/2$ between this field and those given in (10) and (11), we can determine all the unknown field amplitudes.

For a z- and y-invariant medium, the following system of equations for the fields E_z and H_z can be derived from the original Maxwell's equations

$$\begin{pmatrix} \frac{\bar{k}^2}{\epsilon}\frac{\partial}{\partial x}\left(\frac{\epsilon}{\bar{k}^2}\frac{\partial}{\partial x}\right) + \frac{\partial^2}{\partial y^2} + \bar{k}^2 & + \frac{k_z}{k_0}\frac{\bar{k}^2}{\epsilon}\left(\frac{d}{dx}\frac{1}{\bar{k}^2}\right)\frac{\partial}{\partial y} \\ -\frac{k_z}{k_0}\frac{\bar{k}^2}{\mu}\left(\frac{d}{dx}\frac{1}{\bar{k}^2}\right)\frac{\partial}{\partial y} & \frac{\bar{k}^2}{\mu}\frac{\partial}{\partial x}\left(\frac{\mu}{\bar{k}^2}\frac{\partial}{\partial x}\right) + \frac{\partial^2}{\partial y^2} + \bar{k}^2 \end{pmatrix} \begin{pmatrix} E_z \\ H_z \end{pmatrix} = 0. \qquad (14)$$

To determine the modal representation of the electromagnetic field we need to solve (14) subject to the pseudo-periodic boundary conditions [13]

$$\left.\begin{aligned} E_z(d/2,y) &= e^{i\alpha_0 d} E_z(-d/2,y), & H_z(d/2,y) &= e^{i\alpha_0 d} H_z(-d/2,y), \\ \frac{\partial E_z}{\partial x}(d/2,y) &= e^{i\alpha_0 d}\frac{\partial E_z}{\partial x}(-d/2,y), & \frac{\partial H_z}{\partial x}(d/2,y) &= e^{i\alpha_0 d}\frac{\partial H_z}{\partial x}(-d/2,y). \end{aligned}\right\} \qquad (15)$$

In addition, we also need the interface conditions for the fields and their derivatives at the medium discontinuities. However, the burden of mentioning the interface conditions can be relieved if we understand Maxwell's equations in the sense of distribution [13].

Equation (14) is a coupled system of equations for E_z and H_z, whose direct solution seems to be difficult. However, by the following field decomposition, its solution is simplified. In a z- and y-invariant medium, an electromagnetic field is said to be E⊥ (H⊥) if the x-component of its electric (magnetic) field vanishes. Let the superscript (e) denote the E⊥ field, and (h) the H⊥ field. Then from (9)

$$k_z \frac{\partial}{\partial x} E_z^{(e)} + \mu k_0 \frac{\partial}{\partial y} H_z^{(e)} = 0, \qquad (16a)$$

$$k_z \frac{\partial}{\partial x} H_z^{(h)} - \epsilon k_0 \frac{\partial}{\partial y} E_z^{(h)} = 0. \qquad (16b)$$

In Appendix A, we prove that any field $(E_z, H_z)^T$, where superscript T signifies matrix transpose, satisfying (14) and (15) can be uniquely expressed as a sum of an E⊥ field and

an $H_\perp$ field, and these two component fields satisfy (14) and (15) independently. Thus the task of solving the vector-valued boundary-value problem (14) plus (15) is reduced to two independent scalar ones to be derived below.

Using (16a) to eliminate H_z from the first equation of (14), and using (16b) to eliminate E_z in the second of (14), we obtain,

$$\mu \frac{\partial}{\partial x}\left(\frac{1}{\mu}\frac{\partial}{\partial x}E_z^{(e)}\right) + \frac{\partial^2}{\partial y^2}E_z^{(e)} + \tilde{k}^2 E_z^{(e)} = 0, \tag{17a}$$

$$\epsilon \frac{\partial}{\partial x}\left(\frac{1}{\epsilon}\frac{\partial}{\partial x}H_z^{(h)}\right) + \frac{\partial^2}{\partial y^2}H_z^{(h)} + \tilde{k}^2 H_z^{(h)} = 0. \tag{17b}$$

In these two equations, the electric field and the magnetic field are no longer coupled. Actually, these two equations are identical to the equations for the TE and TM polarized fields in classical mountings [5-8], provided that k^2 is replaced by $\tilde{k}^2$.

The symmetry exhibited by (17a) and (17b) with respect to $E_z^{(e)}$ and $H_z^{(h)}$ and with respect to ϵ and μ suggests that they can be rewritten as

$$\sigma^{(s)}\frac{\partial}{\partial x}\left(\frac{1}{\sigma^{(s)}}\frac{\partial}{\partial x}F^{(s)}\right) + \frac{\partial^2}{\partial y^2}F^{(s)} + \tilde{k}^2 F^{(s)} = 0, \tag{18}$$

where, and henceforth, $s = e, h$, and

$$F^{(e)} = E_z^{(e)}, \quad F^{(h)} = H_z^{(h)}, \quad \sigma^{(e)} = \mu(x), \quad \sigma^{(h)} = \epsilon(x). \tag{19}$$

Let $G^{(s)}$ denote quantities complementary to $F^{(s)}$ such that

$$G^{(e)} = H_z^{(e)}, \qquad G^{(h)} = E_z^{(h)}. \tag{20}$$

Then, (16a) and (16b) become

$$\frac{\partial}{\partial y} G^{(s)} = \frac{\delta^{(s)}}{\sigma^{(s)}} \frac{k_z}{k_0} \frac{\partial}{\partial x} F^{(s)}, \tag{21}$$

where

$$\delta^{(e)} = -1, \qquad \delta^{(h)} = 1. \tag{22}$$

The new notations introduced in (19), (20), and (22) allow the $E_\perp$ and $H_\perp$ fields to be treated identically in the rest of this paper.

Let a trial solution of (18) be

$$F^{(s)}(x,y) = u^{(s)}(x)\, \omega^{(s)}(y). \tag{23}$$

Then the standard procedure of separation of variables leads to

$$\sigma^{(s)} \frac{d}{dx}\left(\frac{1}{\sigma^{(s)}} \frac{d}{dx} u^{(s)} \right) + (k^2 - \lambda^{(s)2}) u^{(s)} = 0, \tag{24}$$

where $\lambda^{(s)}$ is a constant. The differential equation (24) and the boundary conditions (15) together pose a boundary-value problem which is considered in the next section.

2.3. *The Boundary-Value Problem*

In this section and the next section, for simplicity, we omit the superscript (s) in the relevant quantities. Let L be a differential operator defined by

$$L = \sigma \frac{d}{dx}\left(\frac{1}{\sigma} \frac{d}{dx} \right) + k^2. \tag{25}$$

Then the boundary-value problem for determining the eigenvalues and the eigen-functions of the modal fields is given by

$$\left.\begin{aligned} Lu &= \rho u, \\ u(d/2) &= e^{i\alpha_0 d} u(-d/2), \\ u'(d/2) &= e^{i\alpha_0 d} u'(-d/2), \end{aligned}\right\} \tag{26}$$

where a prime indicates the differentiation with respect to x, and

$$\rho = \lambda^2 . \tag{27}$$

Let us now define an inner product (,) for any two integrable, complex-valued functions $u(x)$ and $v(x)$ by

$$(u,v) = \int_{-d/2}^{d/2} \frac{1}{\sigma(x)} u(x)\, \bar{v}(x)\, dx , \tag{28}$$

where a bar indicates the complex conjugate. For a lossless dielectric grating ($\epsilon(x) > 0$, $\mu(x) > 0$), it is easily seen that L is self-adjoint, i.e.,

$$(Lu,v) = (u,Lv) . \tag{29}$$

From the theory of ordinary differential equations [14], we know that the eigenvalues determined by the boundary-value problem (26) are real and they form a denumerable sequence. Furthermore, the eigen-functions form a complete, orthonormal basis in the sense that any continuous and piecewise differentiable function $f(x)$ satisfying the pseudo-periodic boundary conditions in (26) can be expanded in the eigen-functions.

In order to embrace the most general cases, however, we assume that region 0 is composed of two media of complex permittivities and permeabilities, i.e., the functions $\epsilon(x)$ and $\mu(x)$ are in general complex valued. In addition, we assume α_0, and possibly k_z, to be complex. (This is a minor generalization of the works of Ref. 5-9.) This permits us to apply

the current model to the homogeneous problem of grating diffraction [15]. In either case, the operator L is no longer self-adjoint; therefore, the eigenvalues of (26) are no longer necessarily real and the eigen-functions are no longer orthogonal and complete. In order to be able to use the modal field expansions for the total field, it is necessary to consider the adjoint of (26), which is defined by

$$\left.\begin{aligned} L^{+}u^{+} &= \rho^{+}u^{+}, \\ u^{+}(d/2) &= e^{i\bar{\alpha}_0 d}u^{+}(-d/2), \\ u^{+\prime}(d/2) &= e^{i\bar{\alpha}_0 d}u^{+\prime}(-d/2), \end{aligned}\right\} \tag{30}$$

where the superscript + indicates the adjoint and L^{+} is the differential operator adjoint to L. It is easily seen that

$$(Lu, v^{+}) = (u, L^{+}v^{+}) \tag{31}$$

if

$$L^{+} = \bar{L} = \bar{\sigma}\frac{d}{dx}\left(\frac{1}{\bar{\sigma}}\frac{d}{dx}\right) + \bar{k}^2. \tag{32}$$

From the theory of non-self-adjoint boundary-value problems [14,16], we know that under certain conditions, which (26) and (30) satisfy, two mutually adjoint boundary-value problems have the following properties: (a) Both boundary-value problems possess an infinite number of eigenvalues and the eigenvalues can be ordered such that

$$\rho_m^{+} = \bar{\rho}_m, \quad m = 0, 1, 2, \ldots. \tag{33}$$

(b) The eigen-functions $\{u_m\}$ and $\{u_n^{+}\}$, are bi-orthonormal, i.e.,

$$(u_m, u_n^{+}) = \delta_{mn}. \tag{34}$$

(c) Any continuous and piecewise differentiable function $f(x)$ satisfying the boundary

conditions has a uniformly convergent formal expansion

$$f(x) = \sum_{m=0}^{\infty} (f, u_m^+)\, u_m(x). \tag{35}$$

Hence, even for lossy dielectric or metallic gratings, it is still mathematically justified to represent the total electromagnetic field in region 0 by a superposition of modal fields, as has been done by Botten *et al.*

Incidently, the Rayleigh expansions (10a,b) and (11a,b) can be viewed as expansions in basis functions

$$e_n(x) = e^{i\alpha_n x}. \tag{36}$$

It is easy to verify that $e_n(x)$ are eigen-functions of (26) with L replaced by d^2/dx^2. The adjoint of this new boundary-value problem, with respect to a new inner product $\langle\ ,\ \rangle$ defined by

$$\langle u, v\rangle = \int_{-d/2}^{d/2} u(x)\, \bar{v}(x)\, dx, \tag{37}$$

is (30) with L^+ again replaced by d^2/dx^2, and

$$e_n^+(x) = \frac{1}{d} e^{i\bar{\alpha}_n x} \tag{38}$$

are the adjoint eigen-functions.

2.4. *Eigenvalues and Eigen-Functions*

The explicit forms of the characteristic equation for determining the eigenvalues and

eigen-functions can be most conveniently derived by taking advantage of the simplicity and symmetry of $\epsilon(x)$ and $\mu(x)$ given in (1). It is easy to verify that the following two functions are two linearly independent solutions of (24)

$$\varphi_e = \begin{cases} \cos\gamma_1 x, & 0 \le |x| \le \dfrac{d_1}{2}, \\ \cos\dfrac{\gamma_1 d_1}{2}\cos\gamma_2\left(|x| - \dfrac{d_1}{2}\right) - \dfrac{\sigma_2\gamma_1}{\sigma_1\gamma_2}\sin\dfrac{\gamma_1 d_1}{2}\sin\gamma_2\left(|x| - \dfrac{d_1}{2}\right), & \dfrac{d_1}{2} \le |x| \le \dfrac{d}{2}, \end{cases} \tag{39}$$

$$\varphi_o = \begin{cases} \dfrac{1}{\gamma_1}\sin\gamma_1 x, & 0 \le |x| \le \dfrac{d_1}{2}, \\ \dfrac{1}{\gamma_1} sgn(x)\left[\sin\dfrac{\gamma_1 d_1}{2}\cos\gamma_2\left(|x| - \dfrac{d_1}{2}\right) + \dfrac{\sigma_2\gamma_1}{\sigma_1\gamma_2}\cos\dfrac{\gamma_1 d_1}{2}\sin\gamma_2\left(|x| - \dfrac{d_1}{2}\right)\right], & \dfrac{d_1}{2} \le |x| \le \dfrac{d}{2}, \end{cases} \tag{40}$$

where, for $j = 1, 2$,

$$\gamma_j^2 = k_j^2 - \rho\,. \tag{41}$$

Clearly, φ_e is an even function and φ_o is an odd function. The general solution of (24) is therefore given by

$$u(x) = A\varphi_e(x) + B\varphi_o(x), \tag{42}$$

where A and B are arbitrary constants. Imposition of the boundary conditions in (26) on (42) gives the characteristic equation

$$\Delta(\rho) = \begin{vmatrix} (1-\kappa)\varphi_e(d/2) & (1+\kappa)\varphi_o(d/2) \\ (1+\kappa)\varphi_e'(d/2) & (1-\kappa)\varphi_o'(d/2) \end{vmatrix} = 0, \tag{43}$$

where

$$\kappa = e^{i\alpha_0 d}. \tag{44}$$

It is obvious that when $\kappa = \pm 1$, i.e., in Littrow mountings for which the normal incidence is a special case, $\Delta(\rho)$ is a product of two factors. For general angles of incidence, factorization of $\Delta(\rho)$ is impossible, and the explicit form of the characteristic equation is

$$\cos\gamma_1 d_1 \cos\gamma_2 d_2 - \frac{1}{2}\left(\frac{\sigma_2\gamma_1}{\sigma_1\gamma_2} + \frac{\sigma_1\gamma_2}{\sigma_2\gamma_1}\right)\sin\gamma_1 d_1 \sin\gamma_2 d_2 - \cos\alpha_0 d = 0. \qquad (45)$$

This is a transcendental equation for ρ, whose solutions are, in general, complex numbers.

The eigen-functions of (26), expressed in terms of φ_e and φ_o, are given by

$$u(x) = \begin{cases} iC\,\varphi_o(x), & \text{if } \kappa = +1,\ \varphi_o(d/2) = 0, \\ iC\,\varphi_e(x), & \text{if } k = -1,\ \varphi_e(d/2) = 0, \\ C[(1+\kappa)\,\varphi_o(d/2)\,\varphi_e(x) - (1-\kappa)\,\varphi_e(d/2)\,\varphi_o(x)], & \text{else}, \end{cases} \qquad (46)$$

where C is the normalization constant. The eigen-functions of the adjoint problem (30) can be simply obtained by replacing the relevant quantities in (39-46) by their adjoint counterparts. It can be shown that for each eigenvalue, there is in general only one eigen-function and accidental degeneracy of an eigenvalue can only occur in Littrow mountings. Since the normalization constants C for $u(x)$ and C^+ for $u^+(x)$ are not individually fixed by (34), we can demand $\bar{C}^+ = C$. Then it can be shown that

$$\overline{u^+}(x) = \kappa^{-1} u(-x). \qquad (47)$$

This direct relationship between the two mutually adjoint eigen-functions is very useful in the numerical implementation of the theory.

2.5. *Modal Field Representation*

Since the basis functions $\{u_m^{(s)}\}$ are complete and bi-orthogonal in the sense stated in 2.3, the general solution of (18) for $F^{(s)}$ can be written as

$$F^{(s)} = \sum_{m=0}^{\infty} \omega_m^{(s)}(y)\, u_m^{(s)}(x), \tag{48}$$

where

$$\omega_m^{(s)}(y) = a_m^{(s)} \cos \lambda_m^{(s)} y + b_m^{(s)} \sin \lambda_m^{(s)} y, \tag{49}$$

and $a_m^{(s)}$ and $b_m^{(s)}$ are modal field amplitudes to be determined later. Substituting (48) into (21) and integrating with respect to y, we obtain an expression for $G^{(s)}$,

$$G^{(s)} = \sum_{m=0}^{\infty} \chi_m^{(s)}(y)\, w_m^{(s)}(x), \tag{50}$$

where

$$\chi_m^{(s)}(y) = \frac{k_z^2 + \lambda_m^{(s)2}}{\lambda_m^{(s)}} \left[- b_m^{(s)} \cos \lambda_m^{(s)} y + a_m^{(s)} \sin \lambda_m^{(s)} y \right] \tag{51}$$

and

$$w_m^{(s)}(x) = \frac{1}{k_z^2 + \lambda_m^{(s)2}} \frac{k_z}{k_0} \frac{\delta^{(s)}}{\sigma^{(s)}} \frac{d}{dx} u_m^{(s)}(x). \tag{52}$$

(The integration constant that would appear in (50) can be shown to be zero.) Thus, by (48), (50), and (A-1) we complete the derivation of the following modal field representation of the total electromagnetic field in region 0:

$$\begin{pmatrix} E_z(x,y) \\ H_z(x,y) \end{pmatrix} = \sum_{m=0}^{\infty} \begin{pmatrix} \omega_m^{(e)}(y)\, u_m^{(e)}(x) \\ \chi_m^{(e)}(y)\, w_m^{(e)}(x) \end{pmatrix} + \sum_{m=0}^{\infty} \begin{pmatrix} \chi_m^{(h)}(y)\, w_m^{(h)}(x) \\ \omega_m^{(h)}(y)\, u_m^{(h)}(x) \end{pmatrix}. \tag{53}$$

Before closing this section, we give the orthogonality relation between the vector-valued $E_\perp$ solution and $H_\perp$ solution. For this purpose, we define an inner product [,] of two vector-valued functions φ and ψ such that

$$[\varphi, \psi] = \int_{-d/2}^{d/2} \varphi^T \begin{pmatrix} -1/\mu & 0 \\ 0 & 1/\epsilon \end{pmatrix} \overline{\psi}\, dx. \tag{54}$$

Let $\varphi_m^{(e)} = (u_m^{(e)}, w_m^{(e)})^T$, and $\varphi_n^{+(h)} = (w_n^{+(h)}, u_n^{+(h)})^T$. Then it can be shown that

$$[\varphi_m^{(e)}, \varphi_n^{+(h)}] = 0. \tag{55}$$

2.6 *Matching Interface Conditions*

Having obtained the expressions of the electromagnetic fields in all regions of the space, we are now ready to form the final system of linear equations for determining the unknown field amplitudes by applying the interface conditions at $y = \pm h/2$. The interface conditions are the following:

$$E_{z+} = E_{z-}, \qquad H_{z+} = H_{z-}, \tag{56a}$$

$$E_{x+} = E_{x-}, \qquad H_{x+} = H_{x-}, \tag{56b}$$

where the subscripts $\pm$ indicate limits from above and below the interface respectively, and from (9a) and (9b),

$$E_x = \frac{i}{k^2}\left(k_z \frac{\partial}{\partial x} E_z + \mu k_0 \frac{\partial}{\partial y} H_z\right), \tag{57a}$$

$$H_x = \frac{i}{k^2}\left(k_z \frac{\partial}{\partial x} H_z - \epsilon k_0 \frac{\partial}{\partial y} E_z\right). \tag{57b}$$

Substituting (10a,b), (11a,b), and (53) into (56a) and (56b) and carrying out some tedious algebra, we have, for the continuity of E_z and H_z at $y = + h/2$,

$$\sum_{m=0}^{\infty}\left[\left(\tilde{a}_m^{(e)}+\tilde{b}_m^{(e)}\right)u_m^{(e)}(x)+\left(A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)}\right)w_m^{(h)}(x)\right]$$
$$=\tilde{I}_z^{(e)}e_0(x)+\sum_{n=-\infty}^{\infty}\tilde{R}_n^{(e)}e_n(x) \tag{58a}$$

and

$$\sum_{m=0}^{\infty}\left[\left(\tilde{a}_m^{(h)}+\tilde{b}_m^{(h)}\right)u_m^{(h)}(x)+\left(A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)}\tilde{b}_m^{(e)}\right)w_m^{(e)}(x)\right]$$
$$=\tilde{I}_z^{(h)}e_0(x)+\sum_{n=-\infty}^{\infty}\tilde{R}_n^{(h)}e_n(x). \tag{58b}$$

For the continuity of E_z and H_z at $y = -h/2$, we have

$$\sum_{m=0}^{\infty}\left[\left(\tilde{a}_m^{(e)}-\tilde{b}_m^{(e)}\right)u_m^{(e)}(x)+\left(-A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)}\right)w_m^{(h)}(x)\right]=\sum_{n=-\infty}^{\infty}\tilde{T}_n^{(e)}e_n(x) \tag{58c}$$

and

$$\sum_{m=0}^{\infty}\left[\left(\tilde{a}_m^{(h)}-\tilde{b}_m^{(h)}\right)u_m^{(h)}(x)+\left(-A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)}\tilde{b}_m^{(e)}\right)w_m^{(e)}(x)\right]=\sum_{n=-\infty}^{\infty}\tilde{T}_n^{(h)}e_n(x). \tag{58d}$$

For the continuity of E_x and H_x at $y = +h/2$, we have

$$\sum_{m=0}^{\infty}\left(A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)}\tilde{b}_m^{(e)}\right)\frac{1}{\mu}u_m^{(e)}(x)$$
$$=\left(\tau_1^{(1)}\beta_0^{(1)}\tilde{I}_z^{(e)}+\tau_3^{(1)}\alpha_0\tilde{I}_z^{(h)}\right)e_0(x)-\sum_{n=-\infty}^{\infty}\left(\tau_1^{(1)}\beta_n^{(1)}\tilde{R}_n^{(e)}-\tau_3^{(1)}\alpha_n\tilde{R}_n^{(h)}\right)e_n(x) \tag{59a}$$

and

$$\sum_{m=0}^{\infty}\left(A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)}\right)\frac{1}{\epsilon}u_m^{(h)}(x)$$
$$=\left(\tau_2^{(1)}\beta_0^{(1)}\tilde{I}_z^{(h)}-\tau_3^{(1)}\alpha_0\tilde{I}_z^{(e)}\right)e_0(x)-\sum_{n=-\infty}^{\infty}\left[\tau_2^{(1)}\beta_n^{(1)}\tilde{R}_n^{(h)}+\tau_3^{(1)}\alpha_n\tilde{R}_n^{(e)}\right]e_n(x). \tag{59b}$$

Finally, for the continuity of E_x and H_x at $y = -h/2$, we have

$$\sum_{m=0}^{\infty}\left(-A_m^{(e)}\tilde{a}_m^{(e)}+B_m^{(e)}\tilde{b}_m^{(e)}\right)\frac{1}{\mu}u_m^{(e)}(x)=\sum_{n=-\infty}^{\infty}\left(\tau_1^{(2)}\beta_n^{(2)}\tilde{T}_n^{(e)}+\tau_3^{(2)}\alpha_n\tilde{T}_n^{(h)}\right)e_n(x) \quad (59c)$$

and

$$\sum_{m=0}^{\infty}\left(-A_m^{(h)}\tilde{a}_m^{(h)}+B_m^{(h)}\tilde{b}_m^{(h)}\right)\frac{1}{\epsilon}u_m^{(h)}(x)=\sum_{n=-\infty}^{\infty}\left(\tau_2^{(2)}\beta_n^{(2)}\tilde{T}_n^{(h)}-\tau_3^{(2)}\alpha_n\tilde{T}_n^{(e)}\right)e_n(x). \quad (59d)$$

In the above equations the unknowns are

$$\tilde{a}_m^{(s)}=a_m^{(s)}\cos\left(\lambda_m^{(s)}h/2\right), \qquad \tilde{b}_m^{(s)}=b_m^{(s)}\sin\left(\lambda_m^{(s)}h/2\right), \quad (60)$$

$$\tilde{R}_n^{(s)}=R_n^{(s)}e^{i\beta_n^{(1)}h/2}, \qquad \tilde{T}_n^{(s)}=T_n^{(s)}e^{i\beta_n^{(2)}h/2}, \quad (61)$$

and the rest of the new notations are defined in Appendix B. The numerical solution of equations (58) and (59) will be considered in section 3.2.

2.7 *Diffraction Efficiency and Polarization*

Once $\tilde{R}_n^{(s)}$ and $\tilde{T}_n^{(s)}$ are solved from (58) and (59), $R_n^{(s)}$ and $T_n^{(s)}$ are given by (61). By virtue of (9a) and (9b), all quantities of practical interest can be readily expressed in terms of $R_n^{(s)}$ and $T_n^{(s)}$. Suppose the media in regions 1 and 2 are lossless, and the incident plane wave is normalized such that

$$\frac{\beta_0^{(1)}}{k^{(1)2}}\left(\epsilon^{(1)}|I_z^{(e)}|^2+\mu^{(1)}|I_z^{(h)}|^2\right)=1. \quad (62)$$

Then, the diffraction efficiencies for the reflected and the transmitted propagating waves of order n are given by

$$\eta_n^{(1)} = \frac{\beta_n^{(1)}}{k^{(1)2}} \left(\epsilon^{(1)} | R_n^{(e)} |^2 + \mu^{(1)} | R_n^{(h)} |^2 \right) \tag{63a}$$

and

$$\eta_n^{(2)} = \frac{\beta_n^{(2)}}{k^{(2)2}} \left(\epsilon^{(2)} | T_n^{(e)} |^2 + \mu^{(2)} | T_n^{(h)} |^2 \right). \tag{63b}$$

If the media in region 0 are also lossless, the energy balance theorem holds:

$$\sum_n \eta_n^{(1)} + \sum_{n'} \eta_{n'}^{(2)} = 1, \tag{64}$$

where n and n' run through all propagating orders in regions 1 and 2, respectively.

In many applications involving conical mountings, it is very important to be able to predict the states of polarization of the diffracted orders. Let us associate with a propagating order having a wave-vector

$$\vec{k}_n^{(j)} = \hat{x}\,\alpha_n \pm \hat{y}\,\beta_n^{(j)} + \hat{z}\,k_z, \tag{65}$$

two unit vectors $\hat{s}_n^{(j)}$ and $\hat{p}_n^{(j)}$ such that

$$\hat{s}_n^{(j)} = \frac{\vec{k}_n^{(j)} \times \hat{y}}{| \vec{k}_n^{(j)} \times \hat{y} |}, \qquad \hat{p}_n^{(j)} = \hat{s}_n^{(j)} \times \frac{\vec{k}_n^{(j)}}{k^{(j)}}. \tag{66}$$

In (65) the plus sign is for $j = 1$, and the minus sign is for $j = 2$. The polarization of this diffracted order can be described with the following two angular parameters [17]:

$$\left.\begin{aligned} \alpha_n^{(j)} &= \arctan\left(\frac{|E_{ns}^{(j)}|}{|E_{np}^{(j)}|}\right), \qquad 0 \le \alpha_n^{(j)} \le \frac{\pi}{2}, \\ \delta_n^{(j)} &= -\arg\left(\frac{E_{ns}^{(j)}}{E_{np}^{(j)}}\right), \qquad -\pi < \delta_n^{(j)} \le \pi, \end{aligned}\right\} \tag{67}$$

where $E_{ns}^{(j)}$ and $E_{np}^{(j)}$ are the *s*- and *p*-components of the electric field. (The notation $\alpha_n^{(j)}$ should not be confused with α_n defined in (12).) It is a simple exercise to show that for a diffracted wave in medium 1,

$$\frac{E_{ns}^{(1)}}{E_{np}^{(1)}} = \sqrt{\frac{\mu^{(1)}}{\epsilon^{(1)}}}\,\frac{\alpha_n \epsilon^{(1)} k_0 R_n^{(e)} + \beta_n^{(1)} k_z R_n^{(h)}}{\alpha_n \mu^{(1)} k_0 R_n^{(h)} - \beta_n^{(1)} k_z R_n^{(e)}}. \tag{68}$$

The expression for a diffracted wave in medium 2 may be obtained from (68) by replacing $R_n^{(s)}$ by $T_n^{(s)}$, superscript (1) by (2), and $\beta_n^{(1)}$ by $-\beta_n^{(2)}$.

3. NUMERICAL ASPECTS

3.1 *Solution of the Characteristic Equation*

An efficient, reliable, and accurate numerical method for solving the characteristic equation (45) is of vital importance to the performance of a computer program implementing the modal analysis. Two very different numerical methods have been employed. The method of Botten *et al.* [7,18] is a general one capable of finding all zeros of an analytic function in a prescribed region of the complex plane. That of Suratteau *et al.* [8] and Tayeb and Petit [9] is a problem-specific method that takes advantage of the fact that (45) can be factored in Littrow mountings. Both of these methods systematically find all eigenvalues of (45) in a prescribed region of the complex plane, and both of them

perform well even for highly conducting gratings. As reported by the authors of the second method, their method is as effective as the first method but requires significantly shorter computation time. For this reason, we have decided to use the second method in our numerical implementation of this work. For details of these numerical methods, the interested reader is referred to the references cited above.

3.2 *Solution of the Field Amplitudes*

Before embarking on the numerical solution of (58) and (59), we first analyze the composition of these equations. In (58) and (59) there are eight equations and eight sets of unknowns. The right-hand sides of these equations are expanded in basis functions $\{e_n\}$, which are bi-orthogonal to their adjoint $\{e_n^+\}$. The left-hand sides are expanded in four different sets of functions $\{u_m^{(e)}\}$, $\{u_m^{(h)}\}$, $\{w_m^{(e)}\}$, and $\{w_m^{(h)}\}$. Of these four, as scalar-valued functions, $\{u_m^{(e)}\}$ and $\{u_m^{(h)}\}$ are bi-orthogonal to their respective adjoints, but not to each other. The functions $\{w_m^{(e)}\}$ and $\{w_m^{(h)}\}$ are proportional to the derivatives of $\{u_m^{(e)}\}$ and $\{u_m^{(h)}\}$ and they are not orthogonal to any other functions.

Equations (58) and (59) constitute a system of equations in known function expansions with unknown expansion coefficients. Such a system can be solved by the method of moments [19], which consists of three steps. First, a projection basis, i.e., a set of linearly independent testing functions, is chosen. Then, both sides of the series expansion equations are projected onto this basis by forming appropriate inner products with the testing functions. This step eliminates the x-dependence of the equations and produces an

algebraic linear system of equations of infinite dimension. Finally, the linear system is truncated to a finite order and its solution is obtained by the standard numerical techniques.

Since equations (58) and (59) are already expanded in terms of bi-orthonormal basis functions, it is advantageous to choose the adjoints of these basis functions as the testing functions so that the subsequent numerical solution can be simplified. For each of the eight equations we have two convenient projection bases, $\{u_m^{+(s)}\}$ and $\{e_n^+\}$. Therefore, there can be many different combinations of choices of projection bases for the overall system. Following Suratteau *et al.*, a projection method in which the interface conditions for the z-components are projected onto one basis and those for the x-components are projected onto the other basis is called a hybrid method. A projection method in which all interface conditions are projected onto one basis is called a homogeneous method. For the case of non-conical mountings, Suratteau *et al.* have proved that the numerical solutions resulting from the hybrid methods satisfy the energy balance and reciprocity criteria automatically (independent of the truncation orders) while those resulting from the homogeneous methods do not. It can be shown that the above statement is also true in the case of conical mountings.

In this paper, we adopt the homogeneous method utilizing the projection basis $\{u_m^{+(s)}\}$, so that the energy balance and reciprocity criteria are not automatically satisfied. Multiplying equations (58) by $a_m^{+(s)}(x)/\sigma^{(s)}(x)$, and equations (59) by $a_m^{+(s)}(x)$, then integrating over a grating period, and making use of the bi-orthogonality relation (34), we have the

following matrix equations:

$$WY = UX + UI, \tag{69}$$

$$DY = QX + PI, \tag{70}$$

where

$$X = \begin{pmatrix} \tilde{R}^{(e)} \\ \tilde{T}^{(e)} \\ \tilde{R}^{(h)} \\ \tilde{T}^{(h)} \end{pmatrix}, \qquad Y = \begin{pmatrix} \tilde{a}^{(e)} \\ \tilde{b}^{(e)} \\ \tilde{a}^{(h)} \\ \tilde{b}^{(h)} \end{pmatrix}, \qquad I = \begin{pmatrix} \tilde{I}_z^{(e)} \delta_{0m} \\ 0 \\ \tilde{I}_z^{(h)} \delta_{0m} \\ 0 \end{pmatrix}, \tag{71}$$

δ_{0m} is the Kronecker delta, and the rest of the matrices are defined in Appendix B. In (71), each element of the column vectors is itself a column vector, and the elements of X and Y are related to the unknown field amplitudes listed in (60) and (61). The matrix D in (70) is diagonal, so vector Y can be expressed in terms of X without numerical matrix inversion. Substituting the expression of Y into (69), we have

$$(WD^{-1}Q - U)X = (U - WD^{-1}P)I. \tag{72}$$

This is the final linear system of equations from which we numerically determine the field amplitudes.

In order to solve the linear system (72) on a computer, we unavoidably have to truncate the matrices. We designate N as the total number of terms retained in Rayleigh expansions (we truncate the Rayleigh expansions symmetrically with respect to the zero diffraction order) and M as that retained in the modal expansions. The integers N and M are called the truncation orders. It is easily seen that for the solution of (72) to be well

(neither under- nor over-) specified, the two truncation orders must be the same. Thus, the matrices W, U, P, Q, and D in (72) are $4N$ by $4N$ square matrices.

4. NUMERICAL EXAMPLES

In this section we present some numerical results. The computer program is written in Fortran 77 and double precision is used for real and complex arithmetic. For the special case of non-conical mountings, the program has been checked using published data with good agreement. For the general case, it meets the energy balance and reciprocity criteria with reasonable accuracy. Table 1 tabulates TE and TM diffraction efficiencies of a metallic grating in a non-conical mounting. The data of Botten *et al.* are taken from Table 1 of Ref. 6. Note that Botten *et al.* adopted a hybrid projection method that allows unequal truncation orders N and M. Clearly, the agreement is very good, especially for the TE polarization.

To date, there is no numerical data, especially data of the polarization parameters, for non-perfectly conducting lamellar gratings in conical mountings available in the literature; therefore, we present some original data in the rest of this section. Listed in Table 2 are diffraction efficiencies (η), polarization angles (α,δ), and diffraction angles (θ,ϕ) of a dielectric grating in a conical mounting (the parameters are listed in the table caption). Note that the incident plane wave is right-hand, circularly polarized. Listed in Table 3 are diffraction efficiencies (η), polarization angles (α,δ), and diffraction angles (θ,ϕ) of a highly conducting grating in a conical mounting. The incident plane wave is linearly polarized with

equal s and p component amplitudes. In both cases, the truncation orders are chosen to ensure that the accuracy of the data is better than one percent (see discussion below).

In Figs. 3*a*, 3*b* and 3*c* we show the change of diffraction efficiencies η and polarization angles α and δ of a small period dielectric grating as the incident azimuthal angle ϕ sweeps through the first quadrant. The incident plane wave is always p polarized and it strikes the grating from the optically denser medium at a polar angle greater than the critical angle (for total internal reflection). This configuration is reminiscent of what occurs in a planar waveguide grating coupler for a TM polarized guided-wave [20]. The reflected and the transmitted negative first orders pass off at about $\phi = 37°$ and $\phi = 60°$, respectively. It is evident that as soon as ϕ is nonzero, the diffraction orders become elliptically polarized. The two first orders are nearly circularly polarized at $\phi \sim 20°$. Also, the senses of polarization of the two first orders remain right-handed throughout the angular range of their existence.

Next, we illustrate the excellent convergence rate of the modal method. For this purpose, we define a measure of error Δ_N as follows

$$\Delta_N = \log_{10}\left|\frac{f_N - f^*}{f^*}\right|, \tag{73}$$

where f_N stands for any one of the physical quantities, such as a diffraction efficiency or a polarization angle, computed with truncation order N, and $f^* = f_{N^*}$, where $N^* > N$ is an integer. If f stands for the sum of the diffraction efficiencies for a lossless grating, $f^* = 1.0$.

Roughly speaking, the negative of Δ_N gives the number of correct significant digits in the numerical results.

Fig. 4 shows the convergence of the sum of the diffraction efficiencies and the efficiencies of the negative first orders in reflection and in transmission for a dielectric grating. The physical parameters are the same as those of Table 2. The truncation order N varies from 11 to 61 in increment of 2, and $N^* = 63$. It is evident from the figure that better than one percent accuracy is achieved as soon as the truncation order is greater than 11. The convergence is not monotonic. The large oscillation in the convergence sequence of the reflected order is probably due to the smallness of the diffraction efficiency (see row 3, column 2 of Table 2). If we make a low order polynomial fit of each set of the data in the figure, the three resulting curves will have more or less similar shapes and close locations. This implies that the energy balance criterion can be used as a good accuracy indicator, thanks to our choice of the homogeneous projection method.

Fig. 5 shows the convergence of the diffraction efficiency and diffraction angles α and δ of the negative first order of a metallic grating. The physical parameters are the same as those of Table 3. The truncation order N varies from 11 to 69 in increment of 2, and $N^* = 71$. Since now the grating is metallic, the convergence in this case is, as expected, slower than that of Fig. 4. However, better than one percent accuracy can still be achieved with a truncation orders of 40 or greater.

As the authors of Ref. 5-8 have shown, one of the unique features of the modal method is that it works very well even for deep, highly conducting gratings. This feature is also true for the extension of the method to conical mountings as developed in this paper. This is understandable, since the fundamental analytical and numerical issues for conical and non-conical mountings are essentially the same. For the sake of saving space, however, we will not provide any numerical evidence here.

5. SUMMARY

In this paper, we have extended the rigorous modal method of Botten *et al.* to the case of conical mountings. A crucial step in accomplishing the extension is the field decomposition discussed in section 2.2 and Appendix A. The field decomposition reduces the vector-valued boundary-value problem given by (14) and (15) to a scalar one given by (26), thus tremendously simplifying the subsequent analysis and allowing the previous works of Botten *et al.* and Suratteau *et al.* be used here. The completeness and orthogonality of the modal fields in the corrugated region are carefully established. The computer program implementing the extended modal method can treat a plane wave of arbitrary angle of incidence and polarization. It converges very well for highly conducting grating materials and very deep grating grooves. We have included some original numerical data of both diffraction efficiencies and diffraction polarizations for conical diffraction configurations.

The mathematical formulation presented in this paper has been kept general. In fact, the explicit x-dependence of $\epsilon(x)$ and $\mu(x)$ given in (1) is not used, except for the derivation

of the characteristic equation and the eigen-functions in section 2.4. Therefore, the formulation is valid for any periodic $\epsilon(x)$ and $\mu(x)$. However, it is precisely the simple x-dependence of $\epsilon(x)$ and $\mu(x)$ in (1) that makes the easy derivations in section 2.4 possible. For any other permittivity and permeability variations, the solutions for the eigenvalues and eigen-functions become very complicated, and the modal analysis quickly loses its advantages over other grating methods.

The present analysis can be easily extended to treat gratings of arbitrary groove shapes and waveguide gratings with a number of uniform layers above and below the corrugated region. These extensions will be the subject of a future paper.

6. ACKNOWLEDGMENT

This work is supported by the Optical Data Storage Center at the University of Arizona.

7. APPENDICES

7.1 *Appendix A*

Theorem:

Any solution of (14) and (15) can be decomposed such that

$$\begin{pmatrix} E_z \\ H_z \end{pmatrix} = \begin{pmatrix} E_z^{(e)} \\ H_z^{(e)} \end{pmatrix} + \begin{pmatrix} E_z^{(h)} \\ H_z^{(h)} \end{pmatrix}, \tag{A-1}$$

where $(E_z^{(e)}, H_z^{(e)})^T$ and $(E_z^{(h)}, H_z^{(h)})^T$ are $E_\perp$ and $H_\perp$, respectively, and they satisfy (14)

and (15) independently. The decomposition (A-1) is, in general, unique.

Proof:

Since there is no explicit y-dependence in (14), any solution of (14) is necessarily of the following form

$$\begin{pmatrix} E_z \\ H_z \end{pmatrix} = \sum_\lambda \begin{pmatrix} E_{z\lambda}(x,y) \\ E_{z\lambda}(x,y) \end{pmatrix}, \tag{A-2}$$

where

$$\begin{pmatrix} E_{z\lambda}(x,y) \\ H_{z\lambda}(x,y) \end{pmatrix} = e^{i\lambda y} \begin{pmatrix} \xi_\lambda(x) \\ \eta_\lambda(x) \end{pmatrix}. \tag{A-3}$$

Substituting (A-3) into (14), and eliminating the y-dependence, we have

$$\begin{pmatrix} \frac{k^2}{\epsilon}\frac{d}{dx}\left(\frac{\epsilon}{k^2}\frac{d}{dx}\right) + k^2 - \lambda^2 & + i\lambda\frac{k_z}{k_0}\frac{k^2}{\epsilon}\left(\frac{d}{dx}\frac{1}{k^2}\right) \\ - i\lambda\frac{k_z}{k_0}\frac{k^2}{\mu}\left(\frac{d}{dx}\frac{1}{k^2}\right) & \frac{k^2}{\mu}\frac{d}{dx}\left(\frac{\mu}{k^2}\frac{d}{dx}\right) + k^2 - \lambda^2 \end{pmatrix} \begin{pmatrix} \xi_\lambda(x) \\ \\ \eta_\lambda(x) \end{pmatrix} = 0. \tag{A-4}$$

Equation (A-4) is a linear, two-dimensional vector-valued, second-order ordinary differential equation. It, therefore, has four linearly independent solutions. Suppose $(E_{z\lambda}, H_{z\lambda})^T$ is E. Then, from (16a)

$$\eta_\lambda(x) = -\frac{1}{i\lambda}\frac{1}{\mu}\frac{k_z}{k_0}\frac{d}{dx}\xi_\lambda(x). \tag{A-5}$$

Substitution of (A-5) into (A-4) results in two second-order equations for $\xi_\lambda(x)$. These two equations are not independent; one can be derived from the other. Thus, we have two linearly independent solutions of (A-4) that satisfy the E$\perp$ condition (16a). Similarly, if we demand $(E_{z\lambda}, H_{z\lambda})^T$ to be H$\perp$, then

$$\xi_\lambda(x) = \frac{1}{i\lambda}\frac{1}{\epsilon}\frac{k_z}{k_0}\frac{d}{dx}\eta_\lambda(x), \tag{A-6}$$

and we have two linearly independent solutions of (A-4) that satisfy the H⊥ condition (16b). Clearly, the solutions of E⊥ type and H⊥ type are linearly independent. Therefore, any solution of (A-4) has the following form

$$\begin{pmatrix} \xi_\lambda(x) \\ \eta_\lambda(x) \end{pmatrix} = c_{\lambda 1}^{(e)}\begin{pmatrix} u_{\lambda 1}^{(e)} \\ v_{\lambda 1}^{(e)} \end{pmatrix} + c_{\lambda 2}^{(e)}\begin{pmatrix} u_{\lambda 2}^{(e)} \\ v_{\lambda 2}^{(e)} \end{pmatrix} + c_{\lambda 1}^{(h)}\begin{pmatrix} v_{\lambda 1}^{(h)} \\ u_{\lambda 1}^{(h)} \end{pmatrix} + c_{\lambda 2}^{(h)}\begin{pmatrix} v_{\lambda 2}^{(h)} \\ u_{\lambda 2}^{(h)} \end{pmatrix}, \tag{A-7}$$

where $c_{\lambda l}^{(s)}$, $l = 1, 2$, $s = e, h$, are constants, $(u_{\lambda l}^{(e)}, v_{\lambda l}^{(e)})^T$ and $(v_{\lambda l}^{(h)}, u_{\lambda l}^{(h)})^T$ are E⊥ and H⊥ solutions of (A-4) respectively. So, the decomposition (A-1) is always possible.

Next, we impose the pseudo-periodic boundary conditions (15) on the general solution (A-7). This leads to the following characteristic equation

$$\begin{vmatrix} U_{\lambda 1}^{(e)} & U_{\lambda 2}^{(e)} & V_{\lambda 1}^{(h)} & V_{\lambda 2}^{(h)} \\ U_{\lambda 1}^{(e)'} & U_{\lambda 2}^{(e)'} & V_{\lambda 1}^{(h)'} & V_{\lambda 2}^{(h)'} \\ V_{\lambda 1}^{(e)} & V_{\lambda 2}^{(e)} & U_{\lambda 1}^{(h)} & U_{\lambda 2}^{(h)} \\ V_{\lambda 1}^{(e)'} & V_{\lambda 2}^{(e)'} & U_{\lambda 1}^{(h)'} & U_{\lambda 2}^{(h)'} \end{vmatrix} = 0, \tag{A-8}$$

where

$$\begin{aligned} U_{\lambda l}^{(s)} &= u_{\lambda l}^{(s)}(d/2) - e^{i\alpha_0 d}u_{\lambda l}^{(s)}(-d/2), \\ V_{\lambda l}^{(s)} &= v_{\lambda l}^{(s)}(d/2) - e^{i\alpha_0 d}v_{\lambda l}^{(s)}(-d/2), \\ U_{\lambda l}^{(s)'} &= u_{\lambda l}^{(s)'}(d/2) - e^{i\alpha_0 d}u_{\lambda l}^{(s)'}(-d/2), \\ V_{\lambda l}^{(s)'} &= v_{\lambda l}^{(s)'}(d/2) - e^{i\alpha_0 d}v_{\lambda l}^{(s)'}(-d/2). \end{aligned} \tag{A-9}$$

By elementary row manipulation and making use of (A-5) and (A-6) for the E⊥ and H⊥

solutions, it can be shown that the two off-diagonal two-by-two matrices in (A-8) can be made zero. Therefore, the E⊥ and H⊥ solutions of (A-4), and hence those of (14), satisfy the boundary conditions (15) independently.

Suppose the decomposition (A-1) is not unique. Then we may have a decomposition of the zero field into two non-zero orthogonal modal fields. Furthermore, each of these fields satisfies both the E⊥ and the H⊥ conditions simultaneously. However, this leads to $k_z^2 + \lambda^2 = 0$, a condition which is, in general, not true. This completes the proof of the theorem.

7.2 *Appendix B*

As in the main text of the paper, $j = 1, 2$, and $s = e, h$. In (58) and (59)

$$\tilde{I}_z^{(s)} = I_z^{(s)} e^{-i\beta_0^{(1)} h/2}. \tag{B-1}$$

$$A_m^{(s)} = \frac{k_z^2 + \lambda_m^{(s)2}}{\lambda_m^{(s)}} \tan\frac{\lambda_m^{(s)} h}{2}, \qquad B_m^{(s)} = \frac{k_z^2 + \lambda_m^{(s)2}}{\lambda_m^{(s)}} \operatorname{ctan}\frac{\lambda_m^{(s)} h}{2}, \tag{B-2}$$

$$\tau_1^{(j)} = i k_0^2 \frac{\epsilon^{(j)}}{\tilde{k}^{(j)2}}, \qquad \tau_2^{(j)} = i k_0^2 \frac{\mu^{(j)}}{\tilde{k}^{(j)2}}, \qquad \tau_3^{(j)} = i k_0 \frac{k_z}{\tilde{k}^{(j)2}}. \tag{B-3}$$

For the sake of clarity, the matrices in (69) and (70) are expressed in block forms, followed by the definitions of each sub-matrix.

$$W = \begin{pmatrix} 1 & 1 & W^{(e)} A^{(h)} & W^{(e)} B^{(h)} \\ 1 & -1 & -W^{(e)} A^{(h)} & W^{(e)} B^{(h)} \\ W^{(h)} A^{(e)} & W^{(h)} B^{(e)} & 1 & 1 \\ -W^{(h)} A^{(e)} & W^{(h)} B^{(e)} & 1 & -1 \end{pmatrix}. \tag{B-4}$$

$$U = \begin{pmatrix} U^{(e)} & 0 & 0 & 0 \\ 0 & U^{(e)} & 0 & 0 \\ 0 & 0 & U^{(h)} & 0 \\ 0 & 0 & 0 & U^{(h)} \end{pmatrix}. \tag{B-5}$$

$$D = 2\begin{pmatrix} A^{(e)} & 0 & 0 & 0 \\ 0 & B^{(e)} & 0 & 0 \\ 0 & 0 & A^{(h)} & 0 \\ 0 & 0 & 0 & B^{(h)} \end{pmatrix}. \tag{B-6}$$

$$Q = \begin{pmatrix} \tilde{U}^{(e)} & 0 & 0 & 0 \\ 0 & \tilde{U}^{(e)} & 0 & 0 \\ 0 & 0 & \tilde{U}^{(h)} & 0 \\ 0 & 0 & 0 & \tilde{U}^{(h)} \end{pmatrix} \begin{pmatrix} -\tau_1^{(1)}\beta^{(1)} & -\tau_1^{(2)}\beta^{(2)} & \tau_3^{(1)}\alpha & -\tau_3^{(2)}\alpha \\ -\tau_1^{(1)}\beta^{(1)} & \tau_1^{(2)}\beta^{(2)} & \tau_3^{(1)}\alpha & \tau_3^{(2)}\alpha \\ -\tau_3^{(1)}\alpha & \tau_3^{(2)}\alpha & -\tau_2^{(1)}\beta^{(1)} & -\tau_2^{(2)}\beta^{(2)} \\ -\tau_3^{(1)}\alpha & -\tau_3^{(2)}\alpha & -\tau_2^{(1)}\beta^{(1)} & \tau_2^{(2)}\beta^{(2)} \end{pmatrix}. \tag{B-7}$$

$$A^{(s)}_{mn} = A^{(s)}_{m}\,\delta_{mn}, \qquad B^{(s)}_{mn} = B^{(s)}_{m}\,\delta_{mn}. \tag{B-8}$$

$$\alpha_{mn} = \alpha_n\,\delta_{mn}, \qquad \beta^{(j)}_{mn} = \beta^{(j)}_{n}\,\delta_{mn}. \tag{B-9}$$

$$U^{(s)}_{mn} = (e_n, u_m^{+(s)})_s, \qquad \tilde{U}^{(s)}_{mn} = \langle e_n, u_m^{+(s)} \rangle. \tag{B-10}$$

$$W^{(e)}_{mn} = (w_n^{(h)}, u_m^{+(e)})_e, \qquad W^{(h)}_{mn} = (w_n^{(e)}, u_m^{+(h)})_h. \tag{B-11}$$

To get the expression for matrix P, we only need to reverse the sign of $\beta^{(j)}$ in (B-7). In (B-10) and (B-11) the inner products are those defined in (28) and (37), and the subscript s indicates that the weight function $\sigma^{(s)}$ should be used in the integral. By using (55) and (47), it can be shown that the two sub-matrices in (B-11) are related such that

$$W^{(e)}_{mn} + W^{(h)}_{mn} = 0. \tag{B-12}$$

Hence, there are five sub-matrices involving the eigen-functions to be calculated. It can be

shown that these sub-matrix elements can be expressed in terms of the left and right limits of $u_m^{(s)}$ and its derivative at $\pm d_1/2$, if the periodic medium is characterized by (1).

Since $w_m^{(s)}$ and $\tau_3^{(j)}$ are proportional to k_z, the matrices defined in (B-4) through (B-7) become block-diagonalized when $k_z = 0$. Of course, this means that in non-conical mountings, E_z *and* H_z are de-coupled.

8. REFERENCES

1. R. E. Collin, "Reflection and Transmission at a Slotted Dielectric Interface," *Canadian J. Phys.*, 34(4), 398-411 (1956).
2. C. B. Burckhardt, "Diffraction of a Plane Wave at a Sinusoidally Stratified Dielectric Grating," *J. Opt. Soc. Am.*, 56(11), 1502-1509 (1966).
3. K. Knop, "Rigorous Diffraction Theory for Transmission Phase Gratings with Deep Rectangular Grooves," *J. Opt. Soc. Am.*, 68(9), 1206-1210 (1978).
4. S. T. Peng, T. Tamir, and H. L. Bertoni, "Theory of Periodic Dielectric Waveguides," *IEEE Trans. Microw. Theory & Tech.*, 23(1), 123-133 (1975).
5. L. C. Botten, M. S. Craig, R. C. McPhedran, J. L. Adams, and J. R. Andrewartha, "The Dielectric Lamellar Diffraction Grating," *Optica Acta*, 28(3), 413-428 (1981).
6. L. C. Botten, M. S. Craig, R. C. McPhedran, J. L. Adams, and J. R. Andrewartha, "The Finitely Conducting Lamellar Diffraction Grating," *Optica Acta*, 28(8), 1087-1102 (1981).
7. L. C. Botten, M. S. Craig, and R. C. McPhedran, "Highly Conducting Lamellar

Diffraction Gratings," *Optica Acta*, 28(8), 1103-1106 (1981).

8. J. Y. Suratteau, M. Cadilhac, and R. Petit, "Sur la Détermination Numérique des Efficacités de Certains Réseaux Diélectriques Profonds," *J. Optics (Paris)*, 14(6), 273-288 (1983).

9. G. Tayeb and R. Petit, "On the Numerical Study of Deep Conducting Lamellar Diffraction Gratings," *Optica Acta*, 31(12), 1361-1365 (1984)

10. P. Sheng, R. S. Stepleman, and P. N. Sanda, "Exact Eigenfunctions for Square-Wave Gratings: Application to Diffraction and Surface-Plasmon Calculations," *Phys. Rev.*, B26(6), 2907-2916 (1982).

11. C. W. Haggans, L. Li, T. Fujita, and R. K. Kostuk, "Lamellar Gratings as Polarization Components for Specularly Reflected Beams," submitted to the special issue on diffractive optics, *J. Mod. Opt.*

12. S. T. Peng, "Rigorous Formulation of Scattering and Guidance by Dielectric Grating Waveguides: General Case of Oblique Incidence," *J. Opt. Soc. Am.*, A6(12), 1869-1883 (1989).

13. R. Petit, "A Tutorial Introduction," in *Electromagnetic Theory of Gratings*, ed. R. Petit (Springer-Verlag, Berlin, 1980).

14. R. H. Cole, *Theory of Ordinary Differential Equations* (Appleton-Century-Crofts, New York, 1968).

15. M. Neviere, "The Homogeneous Problem," in *Electromagnetic Theory of Gratings*, ed. R. Petit (Springer-Verlag, Berlin, 1980).

16. M. A. Naimark, *Linear Differential Operators, Part I* (Frederick Ungar, New York,

1967).

17. See, for example, M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Pergamon Press, Oxford, 1989).

18. L. C. Botten, M. S. Craig, and R. C. McPhedran, "Complex Zeros of Analytic Functions," *Computer Physics Communications*, **29**, 245 (1983).

19. R. F. Harrington, *Field Computation by Moment Methods* (Macmillan, New York, 1968).

20. L. Li, Q. Gong, G. N. Lawrence, and J. J. Burke, "Polarization Properties of Planar Dielectric Waveguide Gratings," *Appl. Opt.*, to be published.

TABLE CAPTIONS

1. Numerical comparison with the results of Botten *et al.* for a metallic grating in a non-conical mounting. Parameters: $d = 1.0\ \mu m$, $d_1 = 0.4001\ \mu m$, $h = 0.1\ \mu m$, $\epsilon^{(1)} = \epsilon^{(2)} = \epsilon_1 = 1.0$, $\epsilon_2 = (1.5+i1.0)^2$, $\lambda_0 = 0.8\ \mu m$, $\theta = 11.5°$, $\phi = 0°$. Truncation orders: Botten *et al.*, $M = 20$, $N = 51$; this paper, $M = N = 31$.

2. Diffraction efficiencies (η), polarization angles (α,δ), and diffraction angles (θ,ϕ) of a dielectric grating in a conical mounting. All angular values are in degrees. Parameters: $d = 1.0\ \mu m$, $d_1 = 0.5\ \mu m$, $h = 0.5\ \mu m$, $\epsilon^{(1)} = \epsilon_1 = 1.0$, $\epsilon^{(2)} = \epsilon_2 = 2.25$, $\lambda_0 = 0.5\ \mu m$. Incident polarization: $\alpha = 45°$, $\delta = 90°$. Incident angle: $\theta = \phi = 45°$. Truncation orders: $M = N = 31$.

3. Diffraction efficiencies (η), polarization angles (α,δ), and diffraction angles (θ,ϕ) of a metallic grating in a conical mounting. All angular values are in degrees. Parameters: $d = 1.0\ \mu m$, $d_1 = 0.5\ \mu m$, $h = 1.0\ \mu m$, $\epsilon^{(1)} = \epsilon_1 = 1.0$, $\epsilon^{(2)} = \epsilon_2 = (0.1+i5.0)^2$, $\lambda_0 = 0.5\ \mu m$. Incident polarization: $\alpha = 45°$, $\delta = 0°$. Incident angle: $\theta = 30°$, $\phi = 45°$. Truncation orders: $M = N = 51$.

FIGURE CAPTIONS

1. The coordinate system for a lamellar grating in a conical mounting.

2. The geometry of a lamellar grating.

3. Diffraction efficiencies η (Fig. *a*) and polarization angles α (Fig. *b*) and δ (Fig. *c*) of a dielectric grating versus the incident azimuthal angle ϕ. The polarization angles are measured in degrees. Parameters: $d = 0.3\ \mu m$, $d_1 = 0.15\ \mu m$, $h = 0.15\ \mu m$, $\epsilon^{(1)} = \epsilon_1 = 1.5$, $\epsilon^{(2)} = \epsilon_2 = 1.0$, $\lambda_0 = 0.5\ \mu m$. Incident polarization: $\alpha = 0°$. Incident polar angle: $\theta = 60°$. Truncation orders: $M = N = 45$.

4. Illustration of convergence for a dielectric grating. Round dots: sum of all diffraction efficiencies. Hollow squares: diffraction efficiency of - 1 order in reflection. Hollow triangles: diffraction efficiency of - 1 order in transmission. The parameters are the same as for Table 2.

5. Illustration of convergence for a metallic grating. Round dots: diffraction efficiency of -1 order. Hollow squares: polarization angle α. Hollow triangles: polarization angle δ. The parameters are the same as for Table 3.

Diffraction Order	Efficiency, TE		Efficiency, TM	
	Botten *et al.*	This Paper	Botten *et al.*	This Paper
R_{-1}	2.8529×10^{-2}	2.8526×10^{-2}	1.6772×10^{-2}	1.6756×10^{-2}
R_0	6.2128×10^{-2}	6.2124×10^{-2}	9.5755×10^{-2}	9.5799×10^{-2}
R_1	4.6011×10^{-3}	4.6011×10^{-3}	1.2779×10^{-3}	1.2757×10^{-3}
T_{-1}	3.8574×10^{-2}	3.8573×10^{-2}	2.1372×10^{-2}	2.1430×10^{-2}
T_0	4.6913×10^{-1}	4.6913×10^{-1}	3.9939×10^{-1}	3.9985×10^{-1}
T_1	5.4894×10^{-3}	5.4895×10^{-3}	3.6226×10^{-3}	3.6378×10^{-3}

Table 1

Order	η	α	δ	θ	ϕ
R_{-2}	1.6137×10^{-3}	64.3182	-30.2984	45.0000	135.0000
R_{-1}	3.8070×10^{-3}	65.9715	-157.1961	30.0000	90.0000
R_0	1.8548×10^{-2}	70.4908	-148.4611	45.0000	45.0000
T_{-3}	3.3631×10^{-2}	51.0569	32.2795	48.1897	153.4349
T_{-2}	1.0343×10^{-1}	56.2437	110.2136	28.1255	135.0000
T_{-1}	3.1868×10^{-1}	46.5484	99.0295	19.4712	90.0000
T_0	1.4186×10^{-1}	34.2601	68.3735	28.1255	45.0000
T_1	3.7827×10^{-1}	46.3291	86.8095	48.1897	26.5651

Table 2

Order	η	α	δ	θ	ϕ
R_{-2}	7.3099×10^{-2}	62.4788	52.7402	47.4606	151.3249
R_{-1}	1.3511×10^{-1}	15.3476	-12.0484	22.5000	112.5000
R_0	4.2986×10^{-1}	41.2528	171.2140	30.0000	45.0000
R_1	3.0238×10^{-1}	75.2325	168.7752	67.5000	22.5000

Table 3

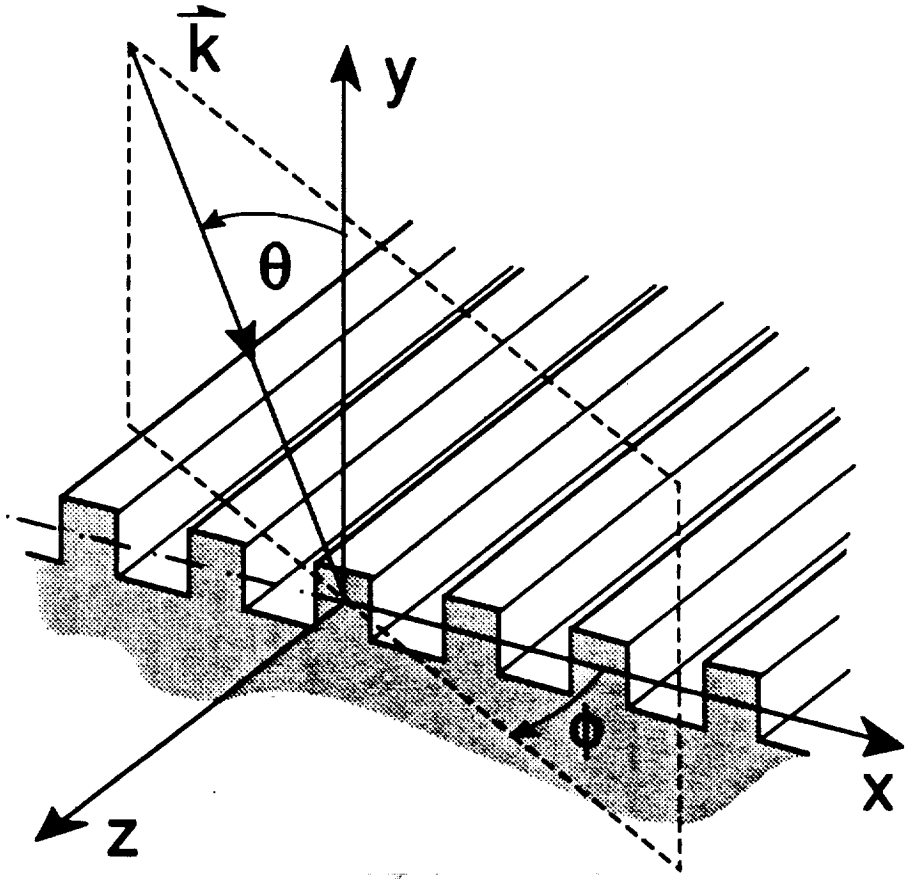


Fig. 1

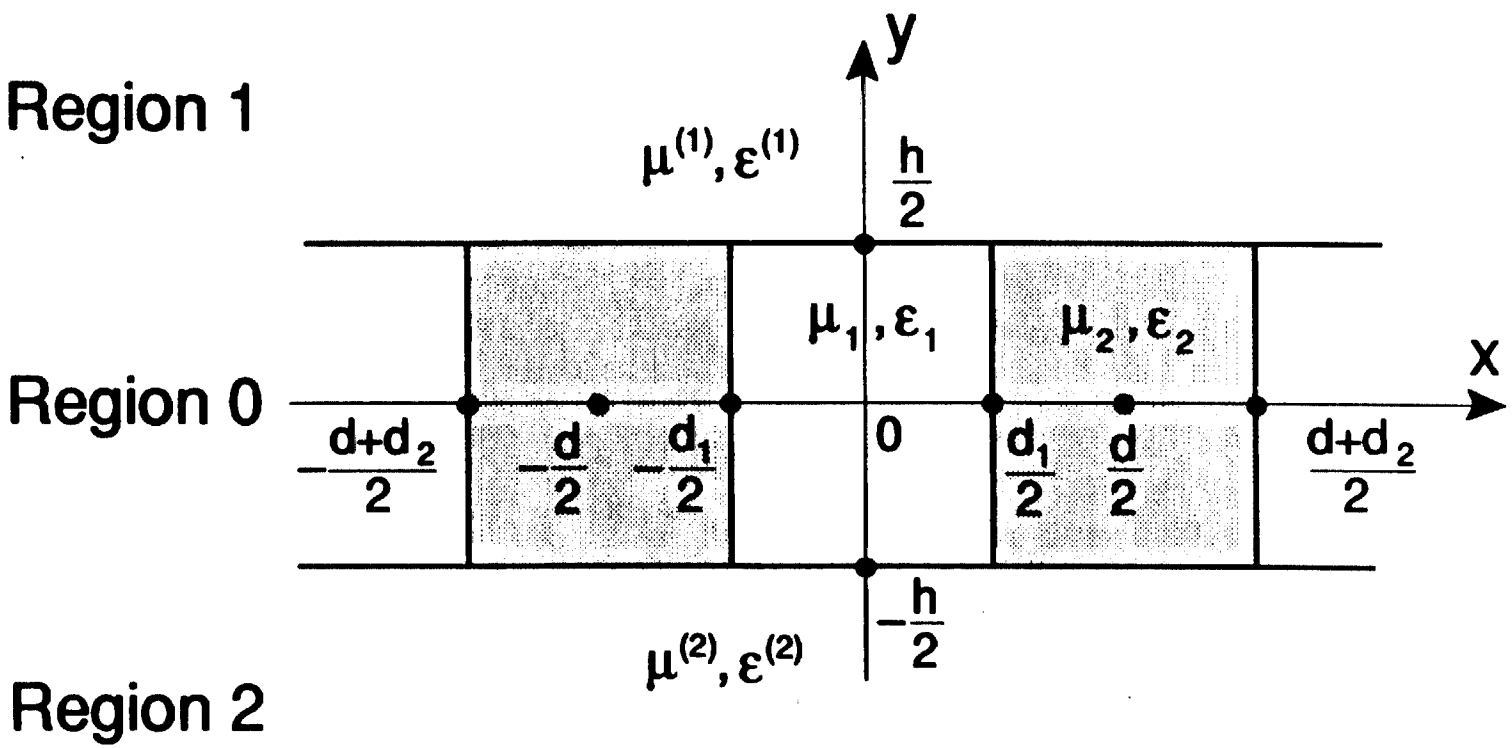

Region 1
Region 0
Region 2
y
x
$\mu^{(1)}, \varepsilon^{(1)}$
$\mu^{(2)}, \varepsilon^{(2)}$
μ_1, ε_1
μ_2, ε_2
$\frac{h}{2}$
$-\frac{h}{2}$
$-\frac{d+d_2}{2}$
$-\frac{d}{2}$
$-\frac{d_1}{2}$
0
$\frac{d_1}{2}$
$\frac{d}{2}$
$\frac{d+d_2}{2}$

Fig. 2

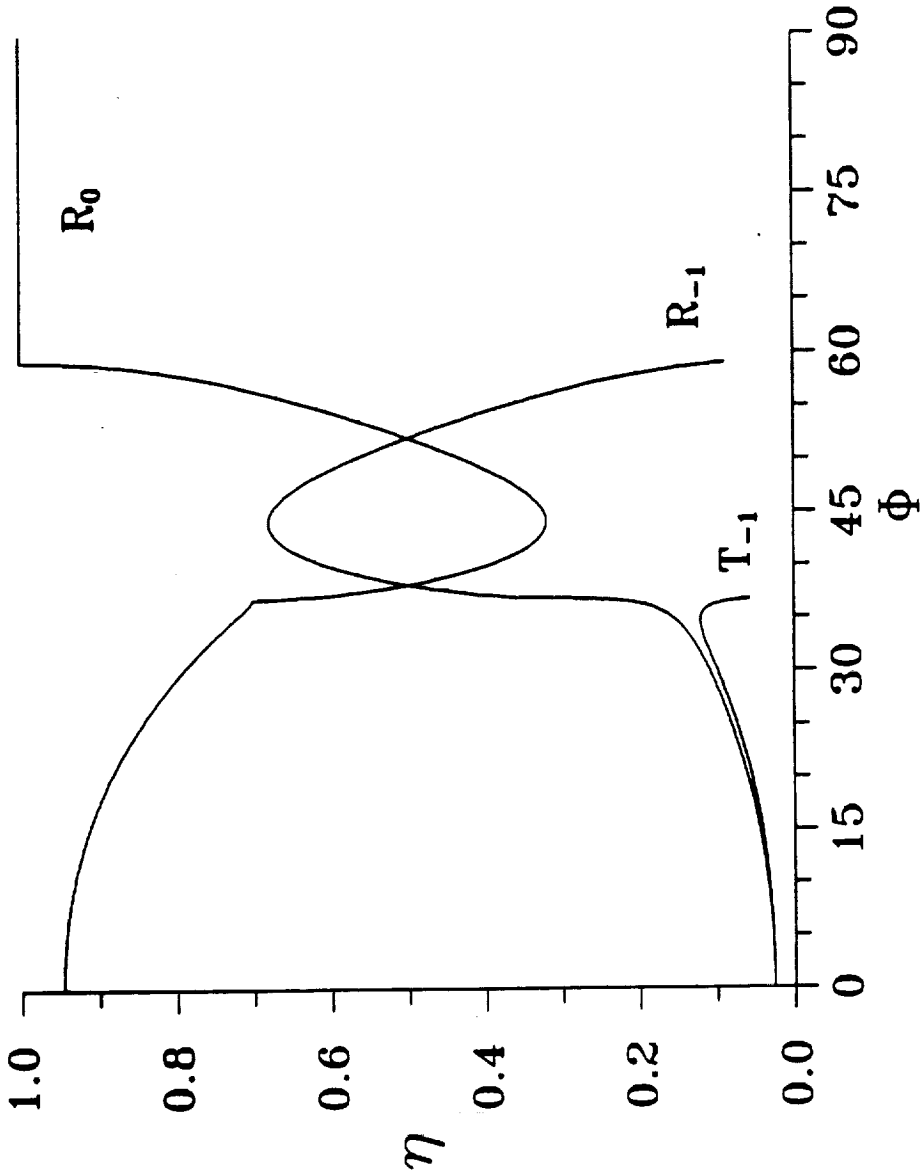


Fig. 3a

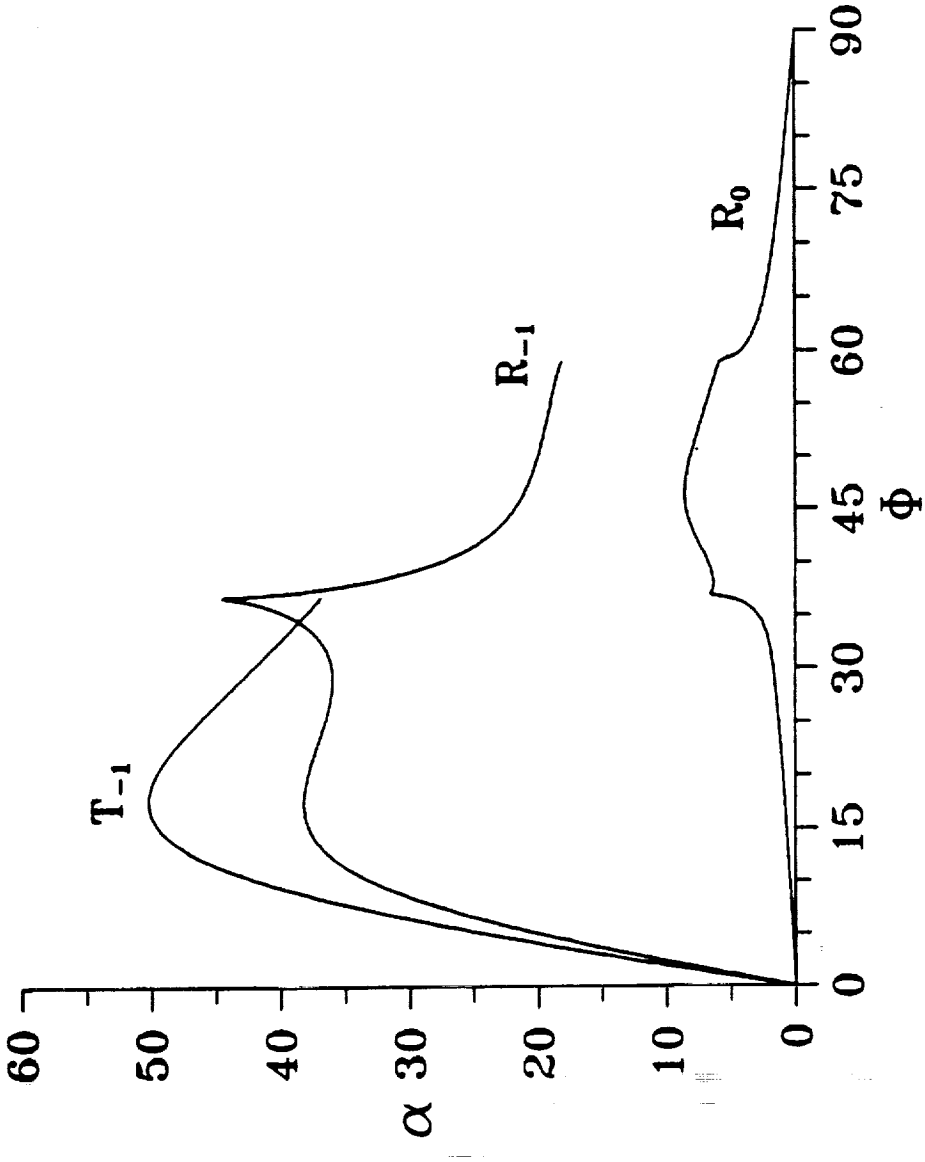


Fig. 3b

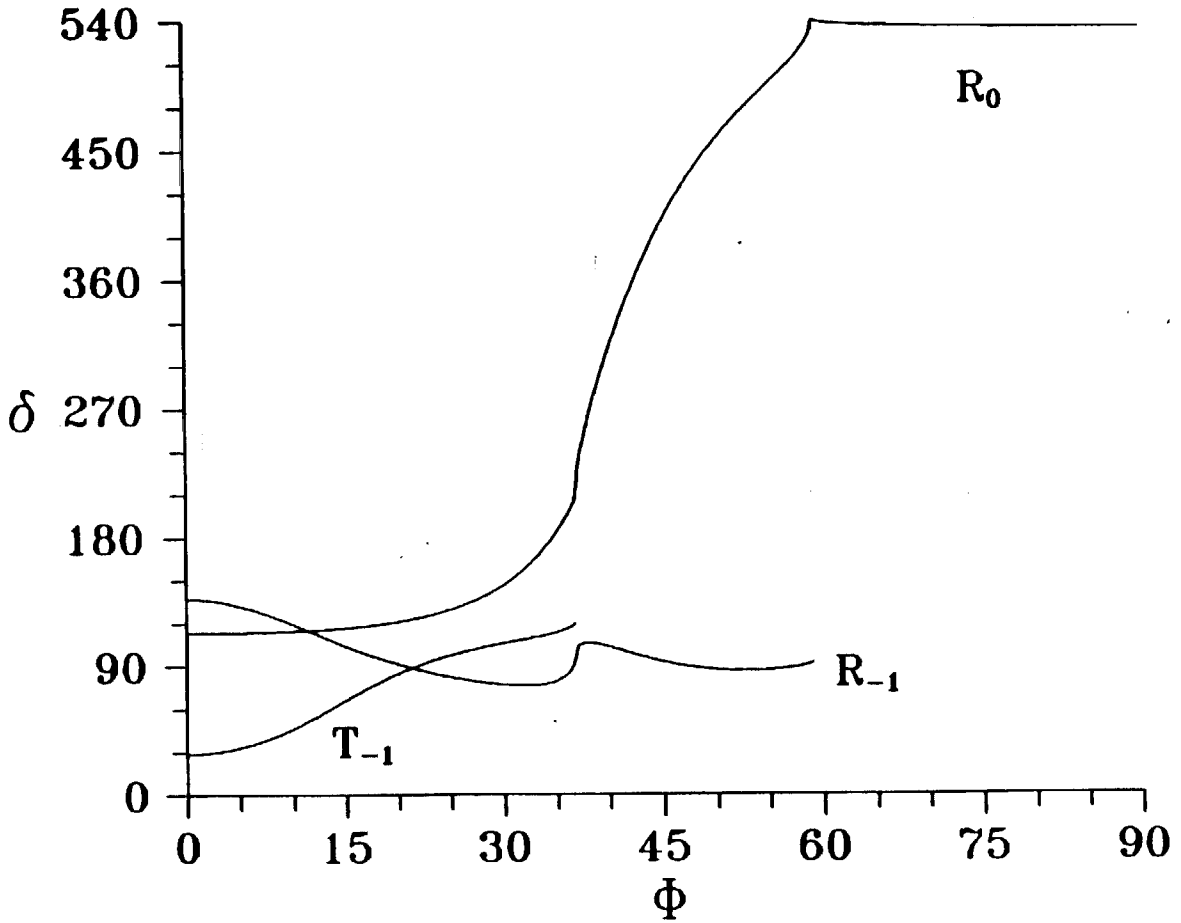


Fig. 3c

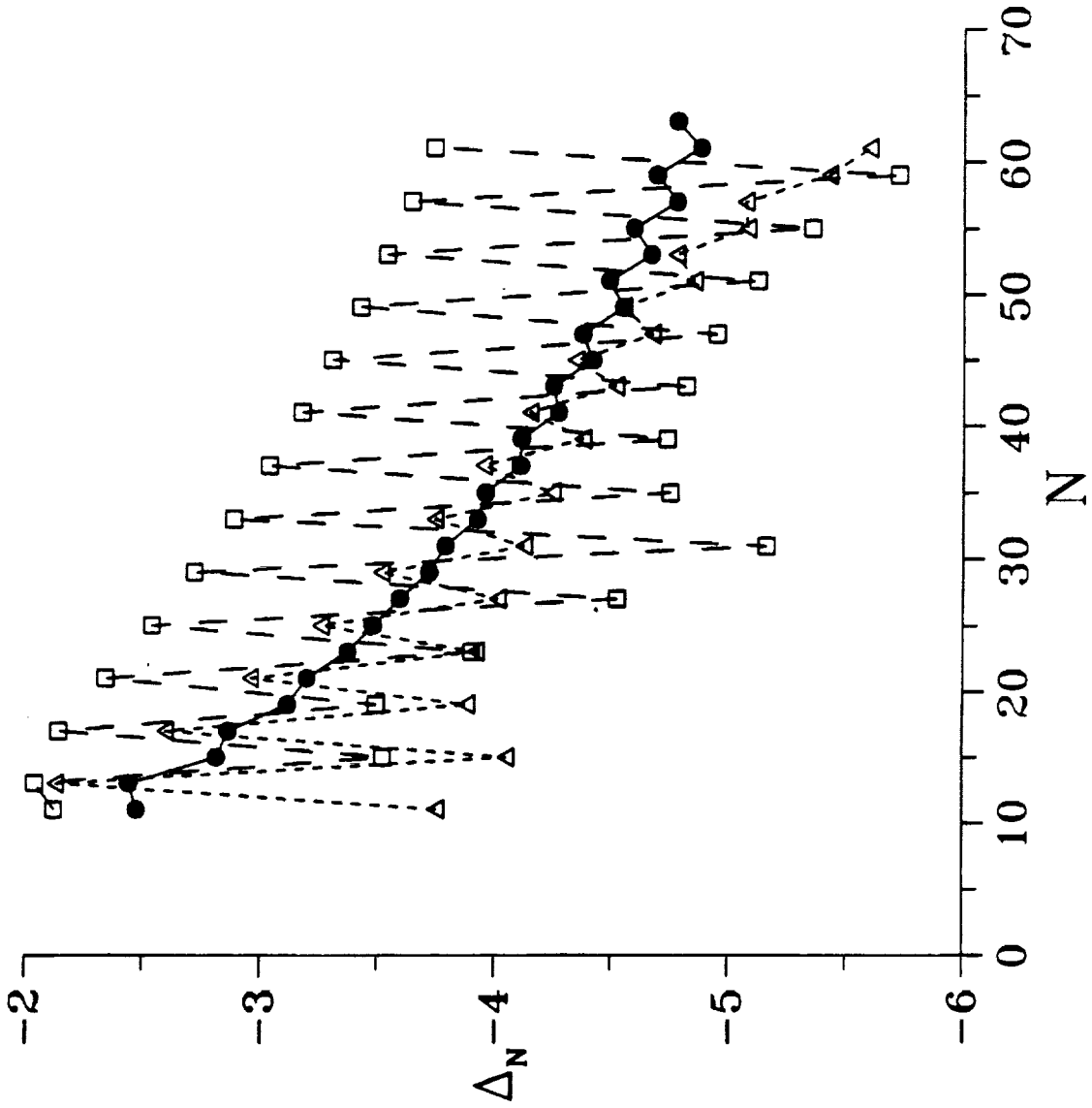


Fig. 4

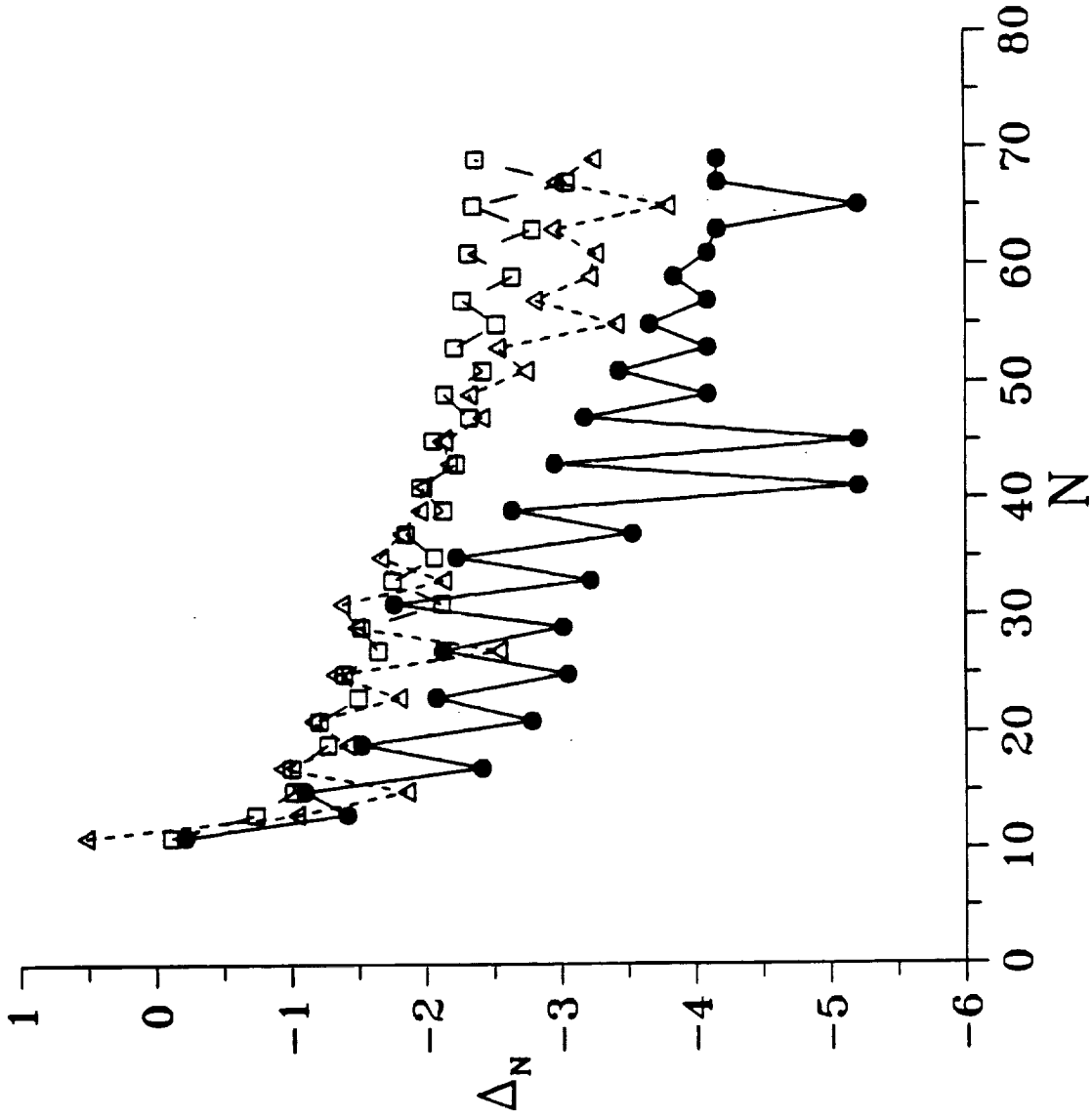


Fig. 5

APPENDIX O

Paper TuB4, The First Topical Meeting on Diffractive Optics,
April 13-15, 1992, New Orleans, LA

On the Convergence of the Coupled-Wave Approach for Lamellar Diffraction Gratings

Lifeng Li and Charles W. Haggans
Optical Sciences Center, University of Arizona, Tucson, AZ 85721. Phone: (602) 621-1789

I. INTRODUCTION

Among the many existing rigorous methods for analyzing diffraction of electromagnetic waves by diffraction gratings, the coupled-wave approach[1] stands out because of its versatility and simplicity. It can be applied to volume gratings and surface relief gratings, and its numerical implementation is much simpler than others. In addition, its predictions have been experimentally validated in several cases[2-4]. These facts explain the popularity of the coupled-wave approach among many optical engineers in the field of diffractive optics. However, a comprehensive analysis of the convergence of the model predictions has never been presented, although several authors[5,6] have recently reported convergence difficulties with the model when it is used for metallic gratings in TM polarization.

In this short paper, we will make three points: *(1) In the TM case, the coupled-wave approach converges much slower than the modal approach of Botten et al.[7,8]. (2) The slow convergence is caused by the use of Fourier expansions for the permittivity and the fields in the grating region, and (3) is manifested by the slow convergence of the eigenvalues and the associated modal fields.* The reader is assumed to be familiar with the mathematical formulations of the coupled-wave approach and the modal approach.

II. ANALYSIS

a. *Two Types of Modal Approaches*

There are two types of modal approaches: Modal Approach using a Scalar (characteristic) Equation[7,8] (MASE) and Modal Approach using a Matrix (characteristic) Equation[9] (MAME). *The coupled-wave approach (CWA) is equivalent to a MAME.* In the MASE, the eigenvalues and modal fields in the grating region are solved one at a time from a scalar characteristic equation. Each modal field thus found satisfies Maxwell's equations and the boundary conditions exactly. In the MAME, however, N eigenvalues and modal fields are solved simultaneously from a matrix characteristic equation. Unless $N = \infty$, none of these modes satisfies Maxwell's equation exactly.

b. *Two Kinds of Infinities*

When solving grating problems by any method, one always has to deal with two kinds of infinities: the infinity of the discrete set Z (the set of all integers), and the infinity of the continuum [0,d), where d is the grating period. The first infinity is due to the periodicity of the grating, which generates an infinite number of diffraction orders. The second infinity is related to the continuous nature of Maxwell's equations. In any method, one has to solve Maxwell's equations on [0,d), with the pseudo-periodic boundary condition. In the MASE, the infinity of [0,d) is handled analytically. The infinity of Z is not truncated until the matching of boundary conditions at the interfaces between uniform regions and corrugated region is completed. In the MAME, the infinity of Z is handled similarly. However, the infinity of [0,d) is transformed into that of the coefficients of the Fourier expansions, which happens to be degenerate with the infinity of the diffraction orders. *It is this tempering of the infinity of* [0,d) *that leads to the slow convergence of the CWA.*

c. *Two Convergence Processes*

In the CWA, the permittivity of the periodic medium in the grating region is represented by its Fourier expansion. When this infinite expansion is truncated, *the physical problem is changed* (the original discontinuous permittivity is changed to a smooth-varying one). Hence, both the electromagnetic field and the grating profile are approximated as a result of the truncation. This means that the convergence of the diffraction efficiency and phase values relies on the convergence

of both the modal fields and the permittivity representation.

d. The Convergence Rates of Related Fourier Series

The analysis above indicates that the CWA, or the MAME in general, should be expected to converge slower than the MASE. This point is reinforced by the following asymptotic analysis of the Fourier coefficients for the permittivity and the fields. Let the permittivity be $\epsilon(x+d) = \epsilon(x)$, $\epsilon(x) = \epsilon_1$, $|x| < a$, and $\epsilon(x) = \epsilon_2$, $a < |x| < d/2$. It is easy to check that $c_n = O(n^{-1})$ as $n \to \infty$, where c_n is a Fourier coefficient of $\epsilon(x)$. Evidently, the convergence of the permittivity expansion is very slow.

We now consider the Fourier expansion coefficient, f_n, of the fields. For TE polarization, the boundary conditions demand that both the electric field and its derivative be continuous at the permittivity discontinuity. Consequently, $f_n = O(n^{-3})$, as $n \to \infty$. For TM polarization, the magnetic field is continuous, but its derivative is not. Instead, we have $\epsilon_1^{-1}(dH/dx)|_{x=x_0=0} = \epsilon_2^{-1}(dH/dx)|_{x=x_0=0}$; therefore, $f_n = O(n^{-2})$, as $n \to \infty$. Hence the Fourier expansion for a TM field converges slower than that for a TE field. Next, suppose $\epsilon_1 = 1$, and the polarization is TM. For a lossless dielectric grating, ϵ_1 and ϵ_2 are both positive; while for a metallic grating, ϵ_1 and ϵ_2 have different signs, if the small imaginary part of ϵ_2 is neglected. Accordingly, the H-field at the permittivity discontinuity in a dielectric grating is "smoother" than it is in a metallic grating. In other words, the H-field has weaker high frequency components in a dielectric grating than it has in a metallic grating. The analysis above rationalizes why the CWA converges slower for TM polarization than for TE polarization, and why it converges slower for metallic gratings than for dielectric gratings.

e. Eigenvalues and Modal fields

The difficulty in the analysis of gratings lies in the accurate characterization of the fields in the grating region. For modal approaches of both types (including the CWA), this means that *it is imperative to determine the eigenvalues accurately* because it is trivial to determine the modal fields once the eigenvalues are known. In the CWA approach, the eigenvalues are solved from the characteristic matrix equation, which is obtained from the truncated and slowly converging Fourier expansions of the permittivity and the fields. Thus, the convergence of the eigenvalues is slow, as numerically demonstrated in the next section.

III. NUMERICAL EXAMPLES

Let us arbitrarily choose a fixed groove depth to wavelength ratio, say $d/\lambda = 1$, from Fig. 3 of Ref. 5 (a gold lamellar grating). We examine how the diffraction efficiencies and the eigenvalues converge with the CWA and with the MASE. Our numerical implementation of the CWA is based on the third paper in Ref. 1. The truncated matrix ϵ_{mn}^{-1} is obtained by numerically taking the inverse of ϵ_{mn}, as recommended by Moharam and Gaylord[10]. Our numerical implementation of the MASE is based on Refs. 7 and 8. Both computer programs have been thoroughly checked against the available results in the literature and against each other. In the following, N denotes the total number of space-harmonics retained in the computations. For the results of the MASE, the number of modal fields is set to equal N.

Figs. 1a and 1b show the convergence of diffraction efficiencies as N increases for TE and TM polarizations. For both polarizations, the MASE converges extremely fast. For TE polarization, the CWA converges reasonably fast toward the MASE. For TM polarization, however, the CWA converges very slowly toward the MASE; in fact, it does not begin converging until $N > 40$.

Figs. 2a and 2b show the convergence of the real and imaginary parts of the square of the tenth TE eigenvalue and the ninth TM eigenvalue, respectively. Similar to the convergence of diffraction efficiencies, the TE eigenvalue computed with CWA converges reasonably fast toward that computed with the MASE (dashed-lines), but the TM eigenvalue converges very slowly. We should mention that all the TE eigenvalues computed with CWA converge at a rate similar to that in Fig. 2a, and some of the TM eigenvalues do converge faster than the one shown in Fig. 2b.

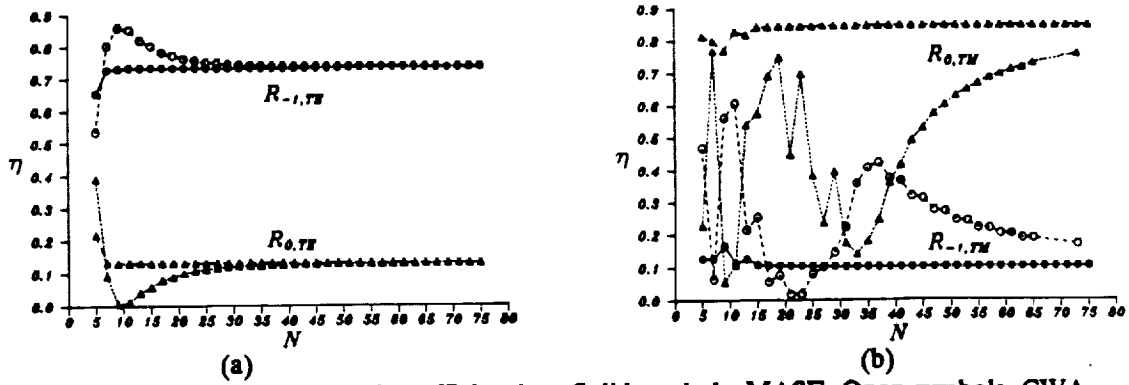


Fig. 1. Convergence of diffraction efficiencies. Solid symbols, MASE. Open symbols, CWA.

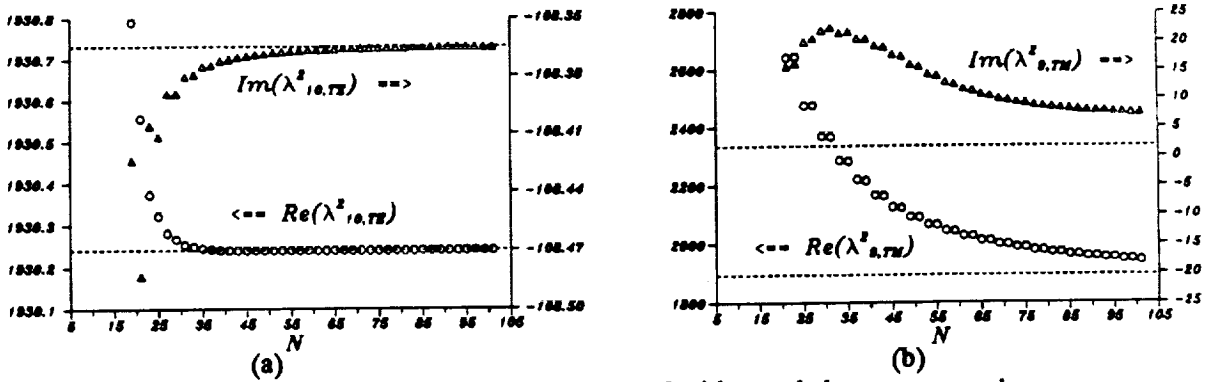


Fig. 2. Convergence of eigenvalues computed with coupled-wave approach.

IV. CONCLUSIONS

We have demonstrated with numerical examples that the predictions of the CWA converge slowly toward those of the MASE of Botten *et al.* for metallic gratings in TM polarization. We have shown that the slow convergence is caused by the use of slowly convergent Fourier expansions for the permittivity and the fields in the grating region. *Despite its popularity, versatility, and simplicity, the coupled-wave approach should be used with caution for metallic surface relief gratings in TM polarization.* Although we have only compared the CWA with the MASE, much of the analysis applies to the MAME and differential methods utilizing Fourier expansions in general.

Li's work is supported by the Optical Data Storage Center at the University of Arizona. Haggans' work is supported by IBM SUR-0499 and a graduate fellowship from Itek Optical Systems.

V. REFERENCES

1. M. G. Moharam and T. K. Gaylord, JOSA, 72, 1385(1982), 73, 451(1983), and 73, 1105(1983).
2. M. G. Moharam *et al.*, Appl. Opt., 23, 3214(1984).
3. T. Schimert and R. Magnusson, JOSA, A9, 1719(1990).
4. C. W. Haggans and R. K. Kostuk, SPIE 1499, 293(1991)
5. M. G. Moharam and T. K. Gaylord, JOSA, A3, 1780(1986).
6. C. W. Haggans *et al.*, paper submitted to JOSA.
7. L. C. Botten, *et al.*, Optica Acta, 28, 413(1981), 28, 1087(1981), and 28, 1103(1981).
8. J. Y. Suratteau, *et al.*, J. Optics (Paris), 14, 273(1983).
9. C. B. Burckhardt, JOSA, 56, 1502(1966).
10. Personal communication on November 8, 1991.