united nations educational, scientific and cultural organization

international atomic energy agency

the
abdus salam
international centre for theoretical physics

Winter College on Optics and Photonics
7 - 25 February 2000

1218-5

"Principles of Optical Disk Data Storage"

M. MANSURIPUR
Optical Sciences Center
University of Arizona
USA

Course description: Principles of magneto-optical (MO) recording and readout. The optical train from the laser to the detectors. Factors that affect the quality of the focused spot. The modulation transfer function (MTF). Sources of noise in readout. Super resolution techniques and the solid immersion lens (SIL). Polarization deterioration and the substrate birefringence.

Benefits/learning objectives: After completing this course the student should be able to:

1. Describe the requirements for a magneto-optical head
2. Identify the role of various components within a typical head
3. Differentiate between the various components of noise and jitter
4. Predict focused spot diameter, depth of focus, and relevant aberrations
5. Evaluate performance criteria for a given optical head design

Intended audience: This course is intended for the engineers, scientists, and technical managers working in the area of magneto-optical disk data storage. The required background is a BA or BS in physical sciences or engineering, and familiarity with the basic notions of classical optics.

Instructor: Masud Mansuripur (MS 1978, Phd 1981, Stanford University) is Professor of Optical Sciences at the University of Arizona in Tucson. He has worked in the field of optical disk data storage for the past 20 years, and has published two textbooks, five book chapters, and over 150 papers in this area. His latest textbook, *The Physical Principles of Magneto-optical Recording* was published by Cambridge University Press in 1995.

Bibliography

1. *Handbook of Magneto-optical Data Recording*, Edited by T. W. McDaniel and R. H. Victora, Noyes Publications, New Jersey, 1997.

2. *The Physical Principles of Magneto-optical Recording*, M. Mansuripur, Cambridge University Press, London, 1995.

3. Proceedings of SPIE, *Optical Data Storage '98*, Aspen, Colorado, Edited by S. Kubota, T. D. Milster and P. J. Wehrenberg, Volume **3401**, 1998.

4. Proceedings of SPIE, *Optical Data Storage '97*, Tucson, Arizona, Edited by H. Birecki and J. Z. Kwiecien, Volume **3109**, 1997.

5. See Proceedings of the *International Symposium on Optical Memory* (ISOM) and the *Magneto-optical Recording International Symposium* (MORIS) for the past few years.

Table of Contents

OPTICAL DISK TECHNOLOGY

1 - Read Only (CD-ROM)

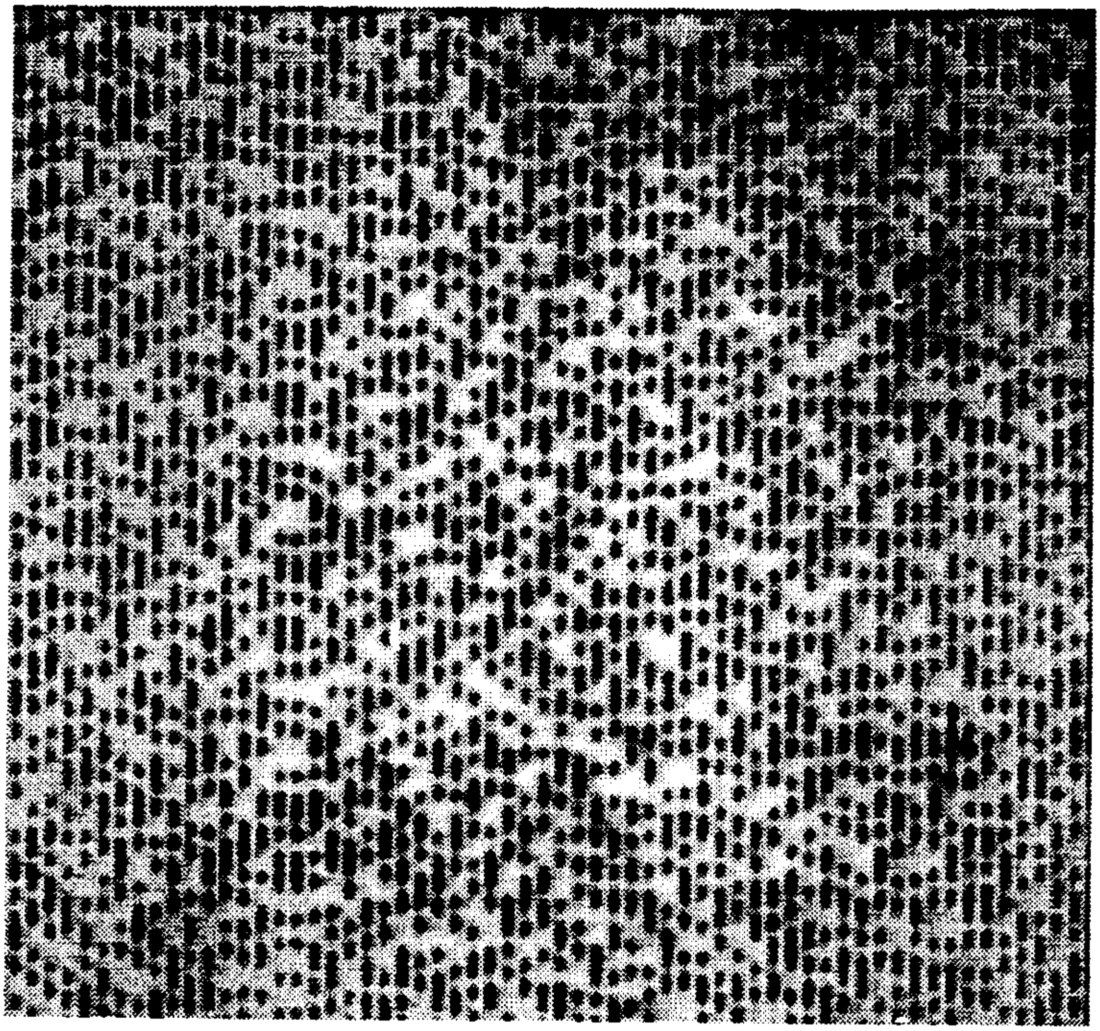
2 - Write-Once-Read-Many (WORM)

3 - Erasable
- Phase change
- Magneto-optic

- holographic
- spectral hole burning
- Electron trapping

IMAGE OF CD PITS TAKEN THROUGH MICROSCOPE ASSEMBLY

NA = 0.6, TRACK PITCH = 1.6 μm



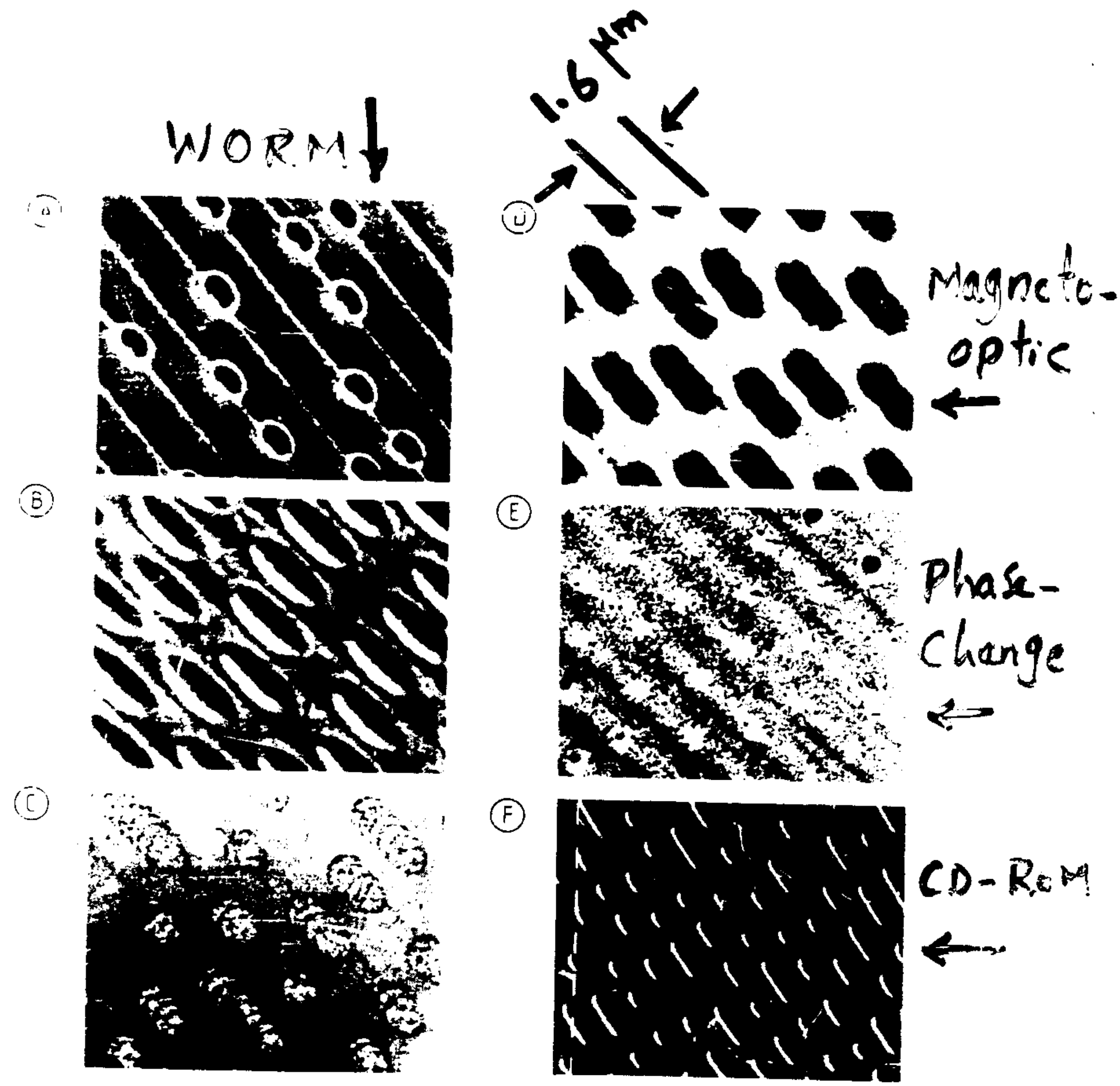
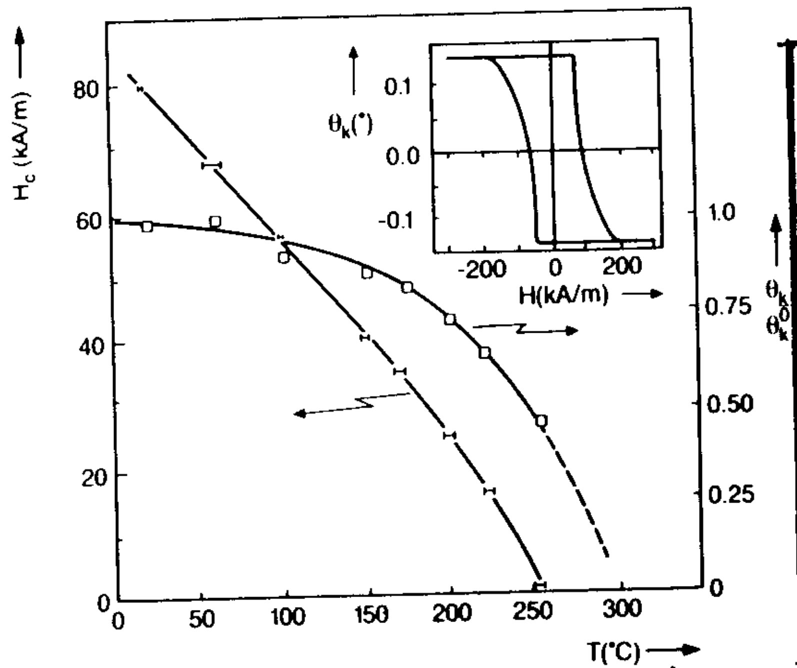


Figure 23. Micrographs of several types of optical storage media. The track pitch is 1.6 µm
A) Ablative, write-once tellurium alloy (SEM, see Section 2.4.3)
B) Ablative, write-once, organic squarylium dye SQS (SEM, see Section 2.4.4)
C) Amorphous to crystalline, write-once, phase change alloy GaSb (TEM, see Section [illegible])
D) Amorphous, reversible, magneto-optic alloy GdTb-Fe (Optical polarization micrograph, see Section 2.5.1)
E) Crystalline to amorphous, reversible, phase change tellurium alloy (TEM, see Section 2.5.2)
F) Molded, read-only CD-Audio or CD-ROM media with EFM code. Media injection molded from polycarbo
a nickel "son" stamper (SEM, see Sections 2.2 and 2.3)

Co/Pt superlattice for MO recording



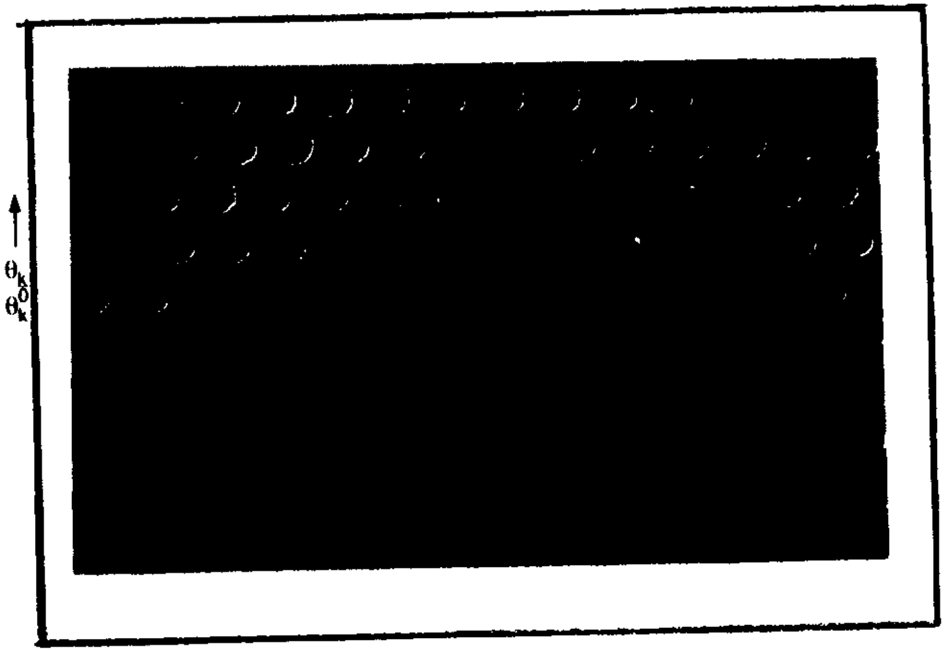


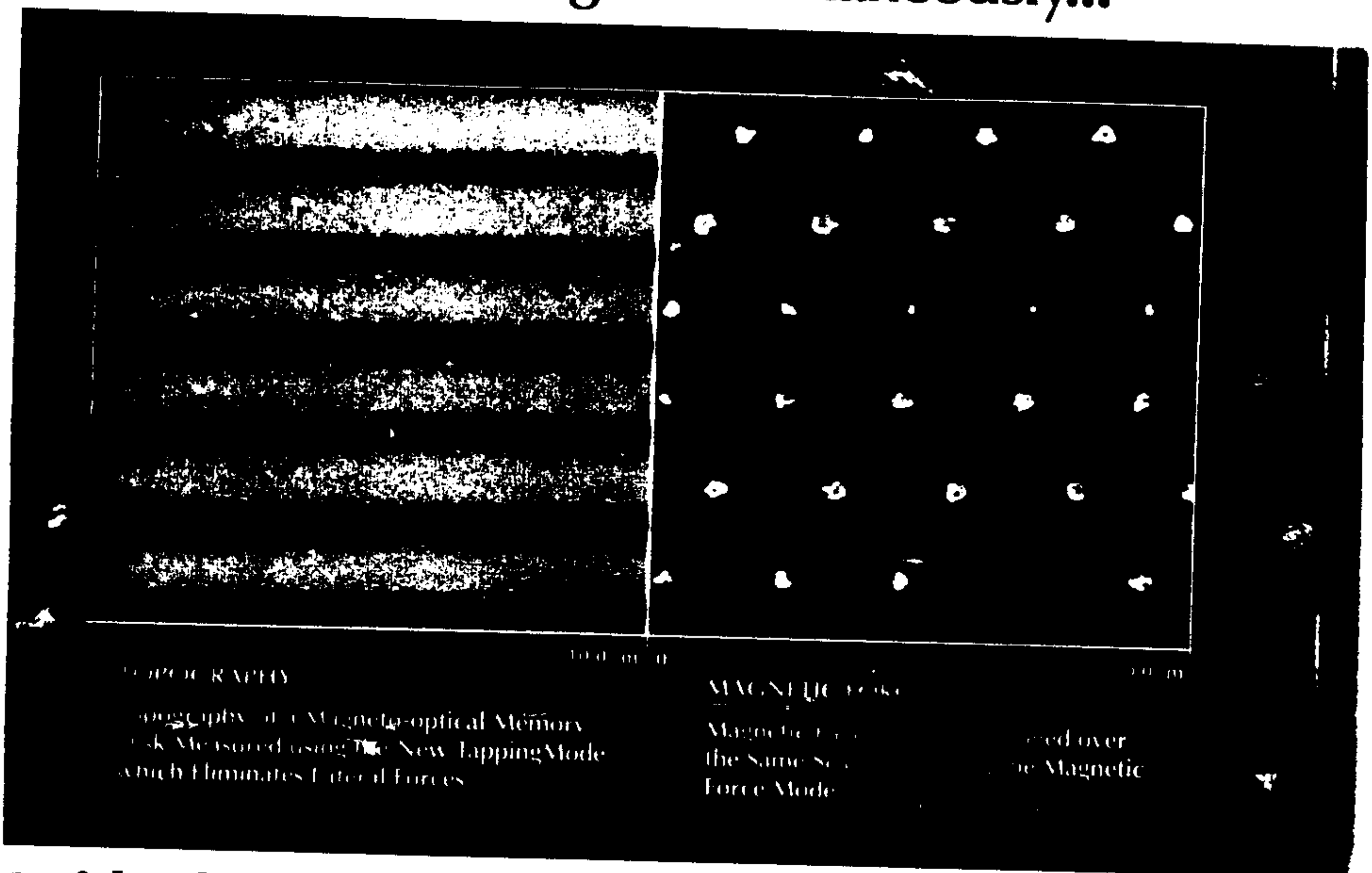
*Fig.1. Coercive field and Kerr rotation for a 25 * (4.1 Å Co + 19 Å Pt) multilayer. The inset shows the Kerr rotation as a function of applied field, measured at $\lambda = 530$ nm.*

*Fig.2. Lorentz micrograph of domains written thermomagnetically in a 9 * (3.5 Å Co + 18 Å Pt) multilayer at incident laser powers of 3.75, 4.0, 4.25, 4.50, 4.75, 5.0, 5.5, 6.0, 6.5, and 7.0 mW, respectively.*

Greidanus et al, Philips

NanoScope® MultiMode™ Scanning Probe Microscope

Now Get High Resolution Topographic And Magnetic Force Images Simultaneously...



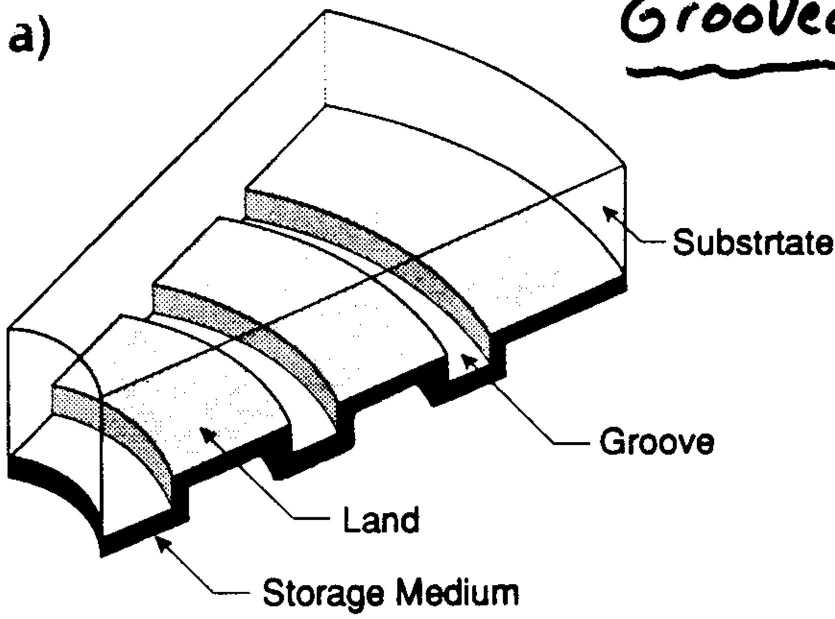
With The NanoScope MultiMode™ SPM Using NanoProbe® Magnetic Sensors.

These images of a magneto-optical disk were made using the NanoScope MultiMode AFM with Digital Instruments exclusive NanoProbe magnetic sensors–the only combination that can produce these images. The dual imaging capability of the NanoScope III allows two types of data to be taken during the same scan.

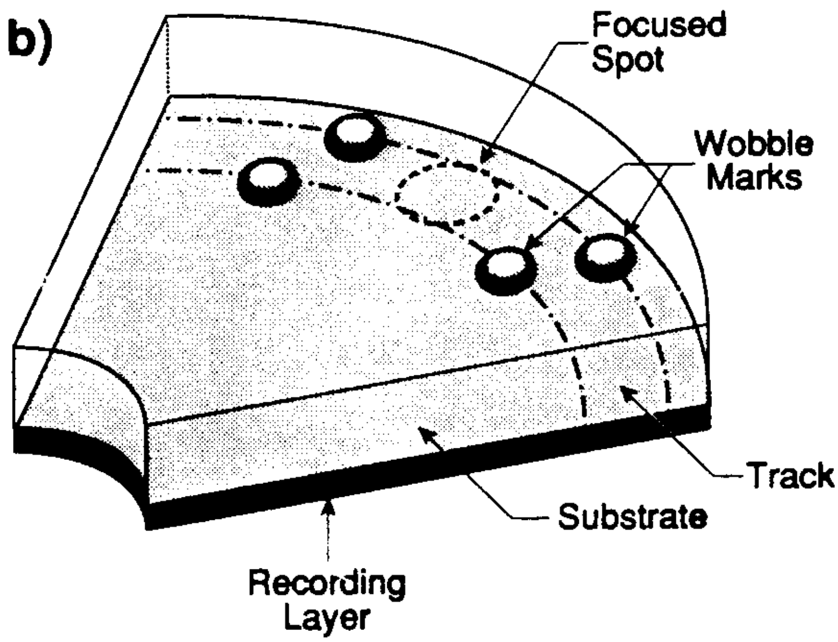
The power of the laser (blue light) was varied during the bit-writing process, resulting in magnetic bits of various sizes. The smallest bit shown here is less than 100nm across.

NanoScope delivers results, not promises.

Grooved media



Sampled-Servo



Eccentricity $\Delta r \approx \pm 50\,\mu m$

Tracking accuracy $\approx \pm 0.1\,\mu m$

Angular spacing of wobble pairs $= \frac{360^\circ}{500}$

Illustration by
Wally Varner / UA Graphics
Mansuripur fig 3(...)

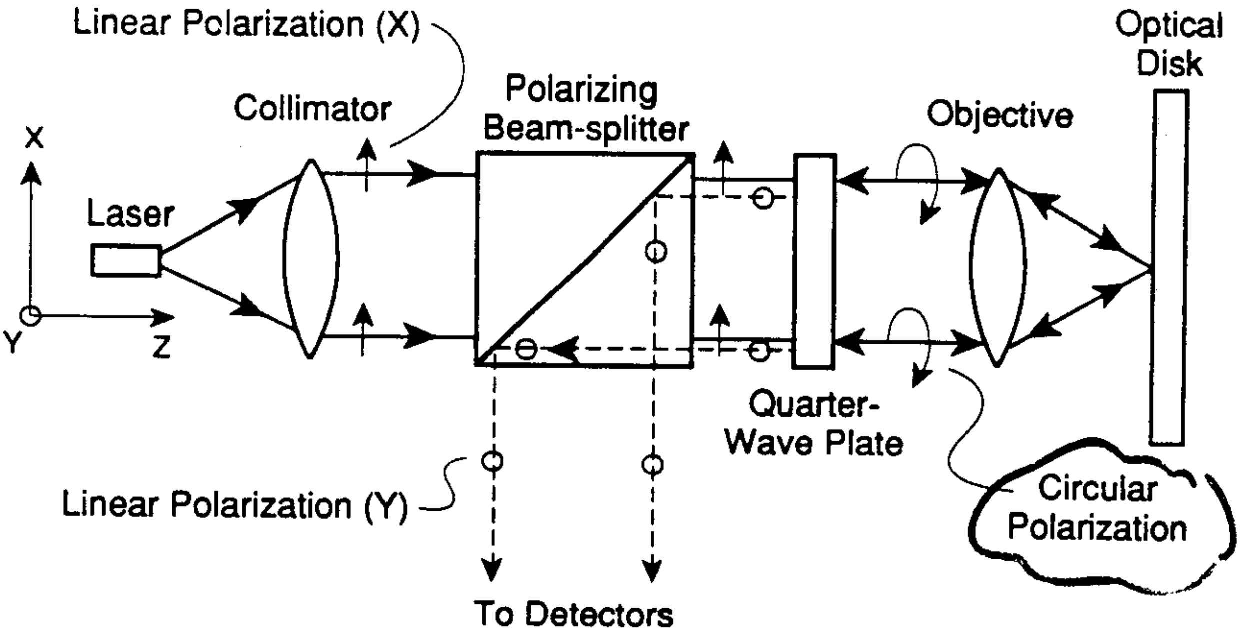
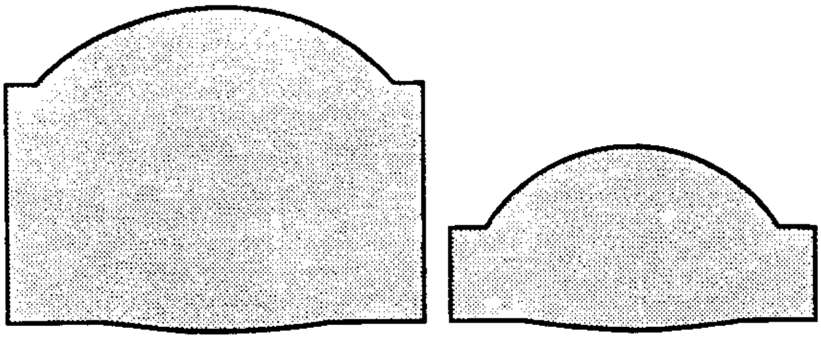


Illustration by
Wally Varner / UA Graphics
Mansuripur fig 10



Collimator Objective

Molded Glass Aspherics

Illustration by
Wally Varner / UA Graphics
Mansuripur fig 12

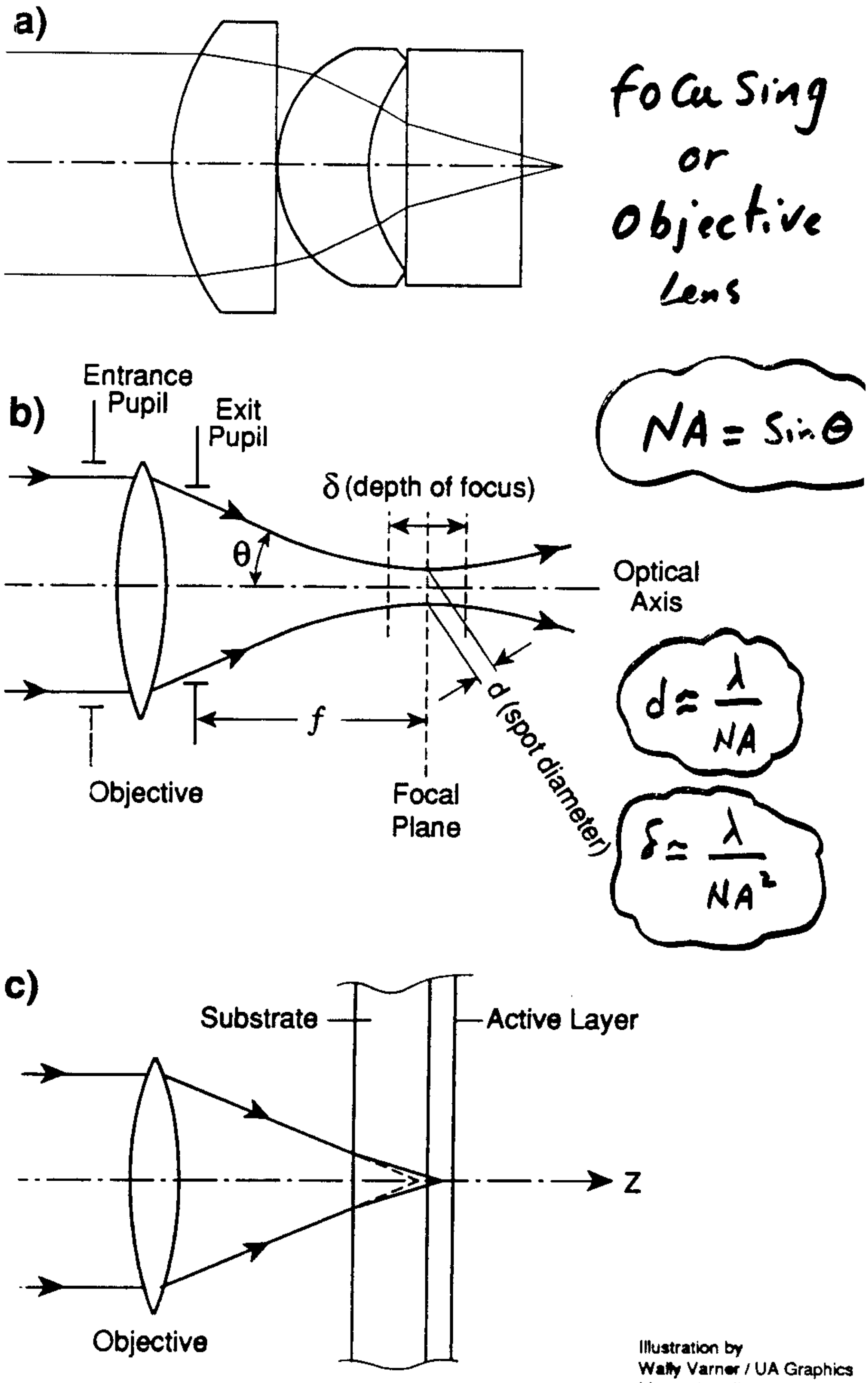
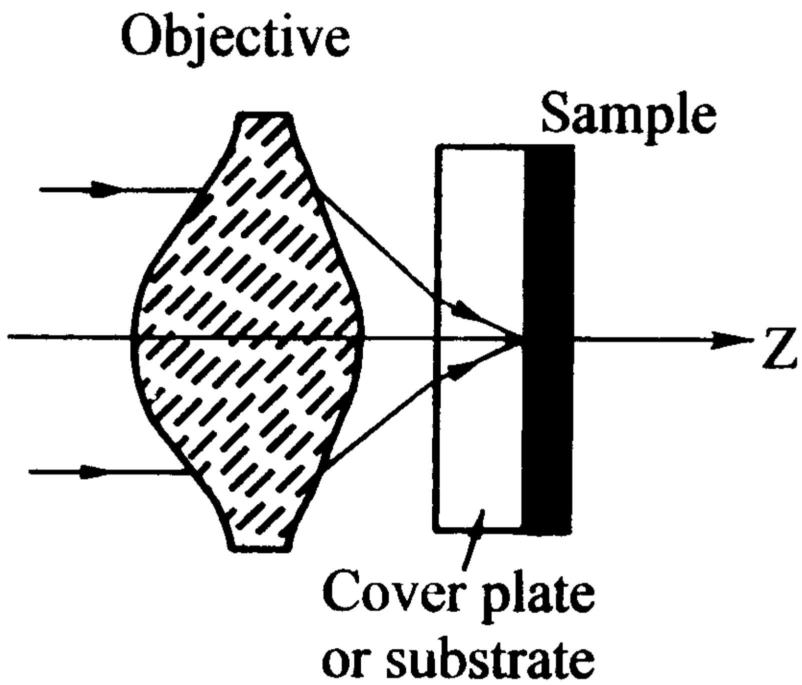


Illustration by
Wally Varner / UA Graphics
Mansuripur fig 11



Mansuripur, Figure 10

Focused Spot
(Airy Pattern)

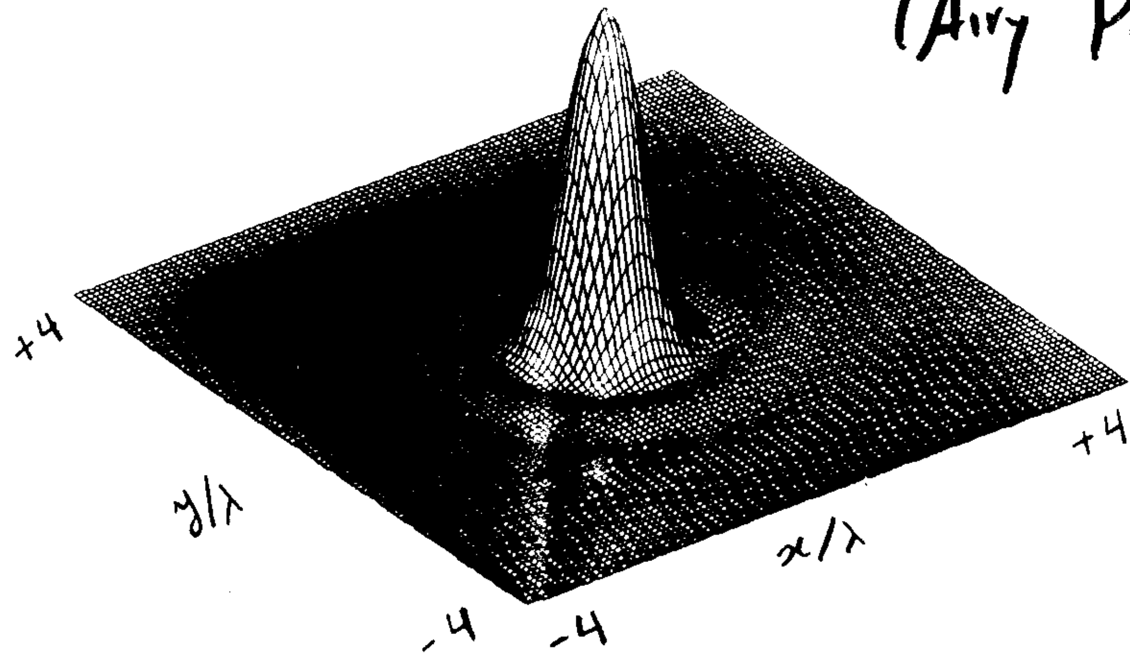


Fig. 6(a)

5λ defocus

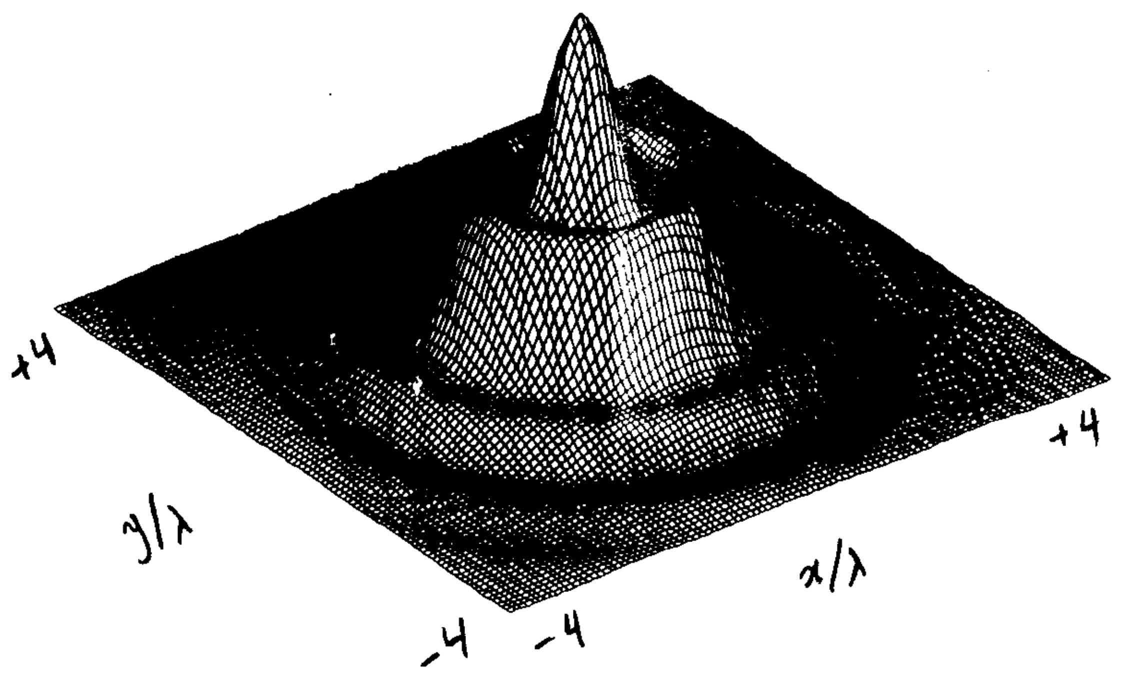


Fig.6(C)

+3λ defocus

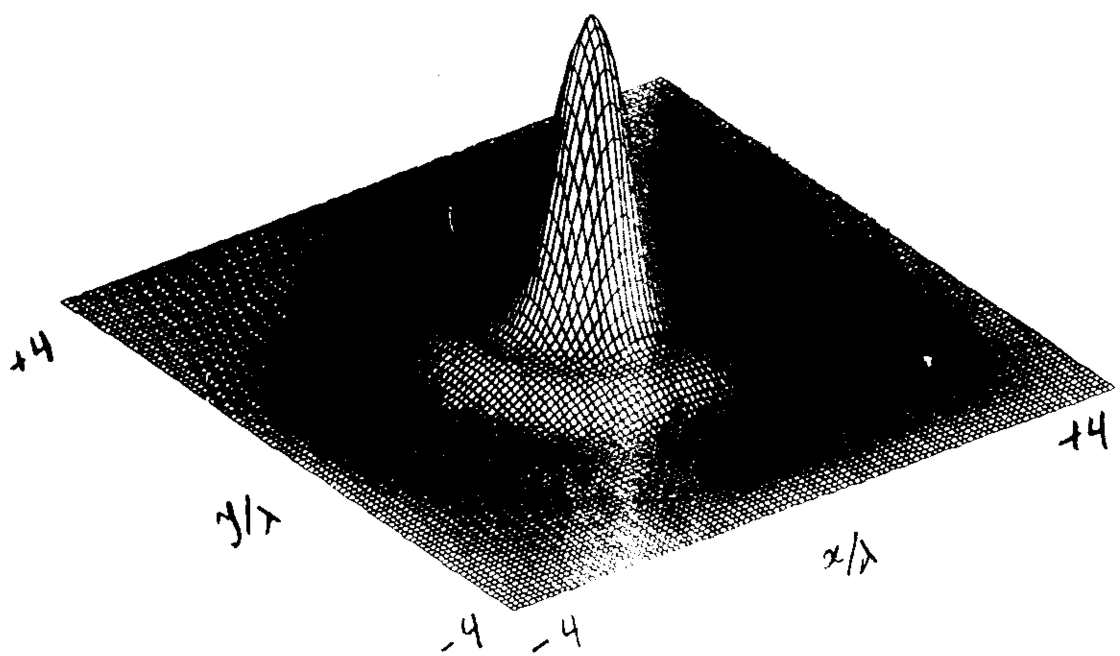


Fig. 6(b)

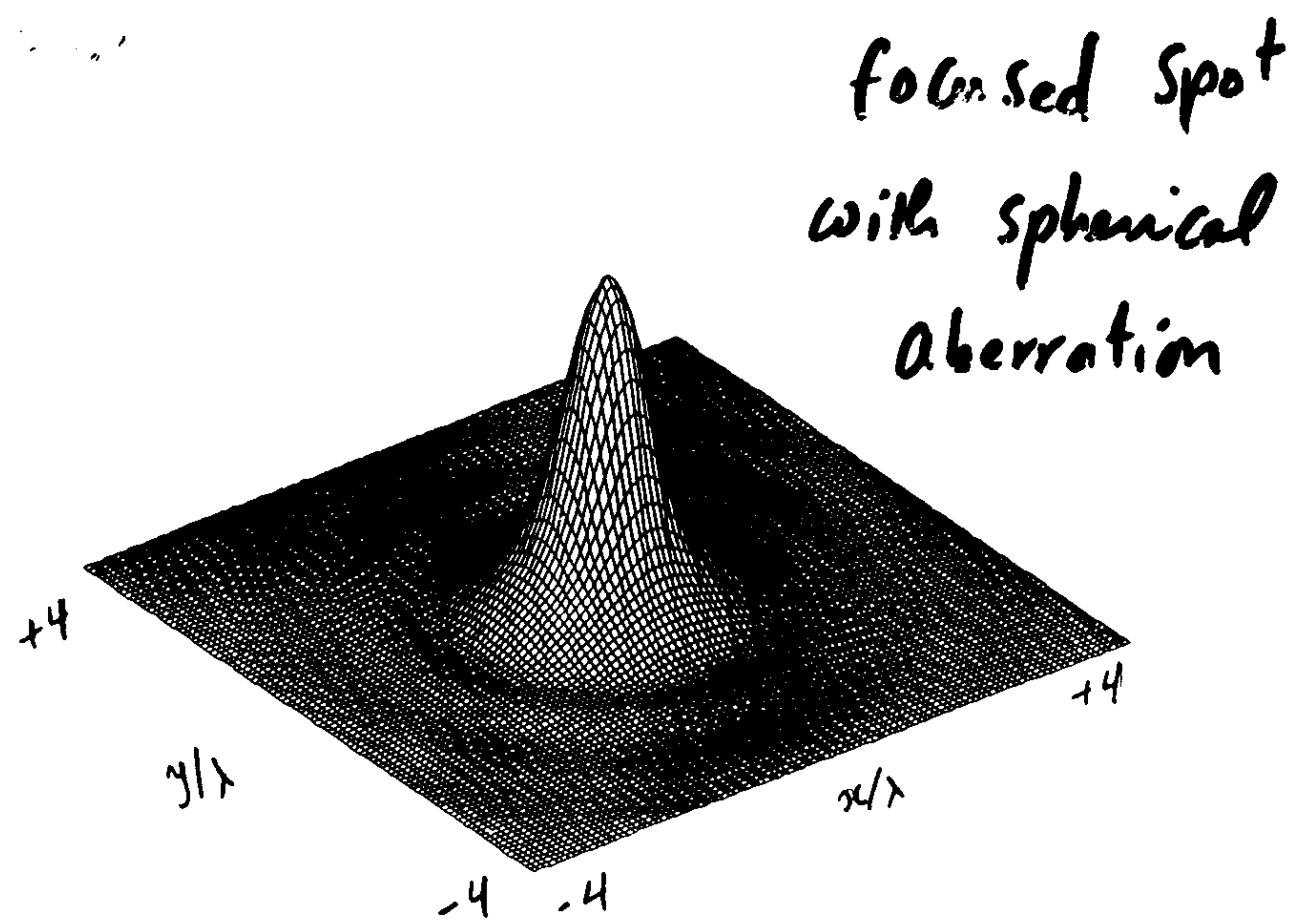


Fig. 7(a)

Focused spot
With Coma

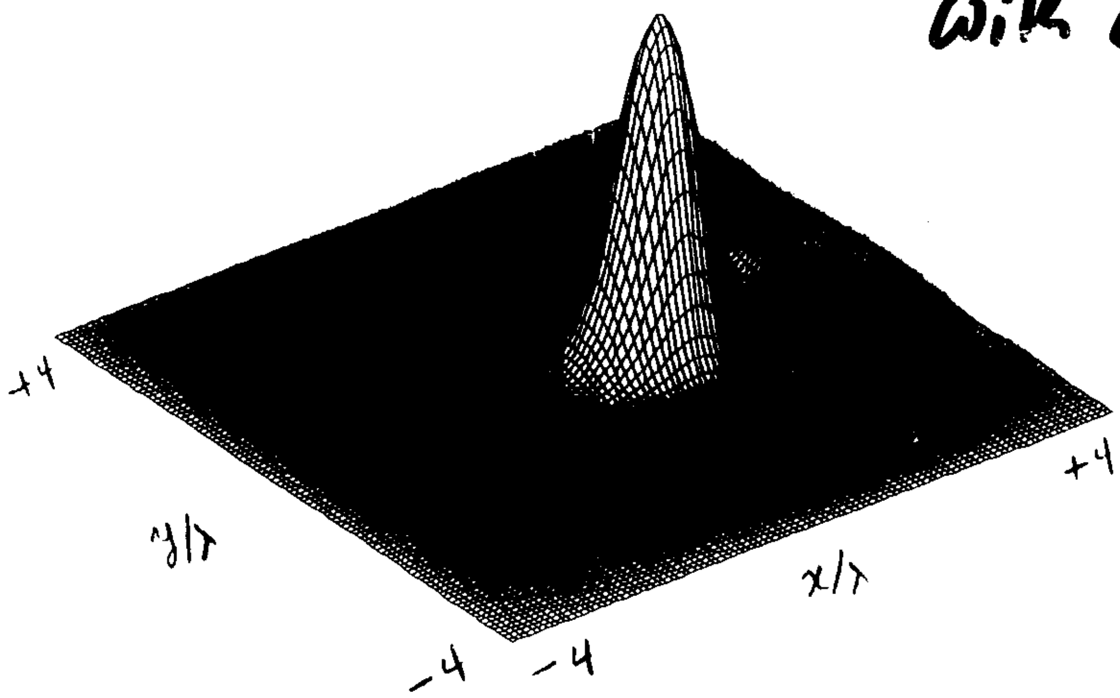
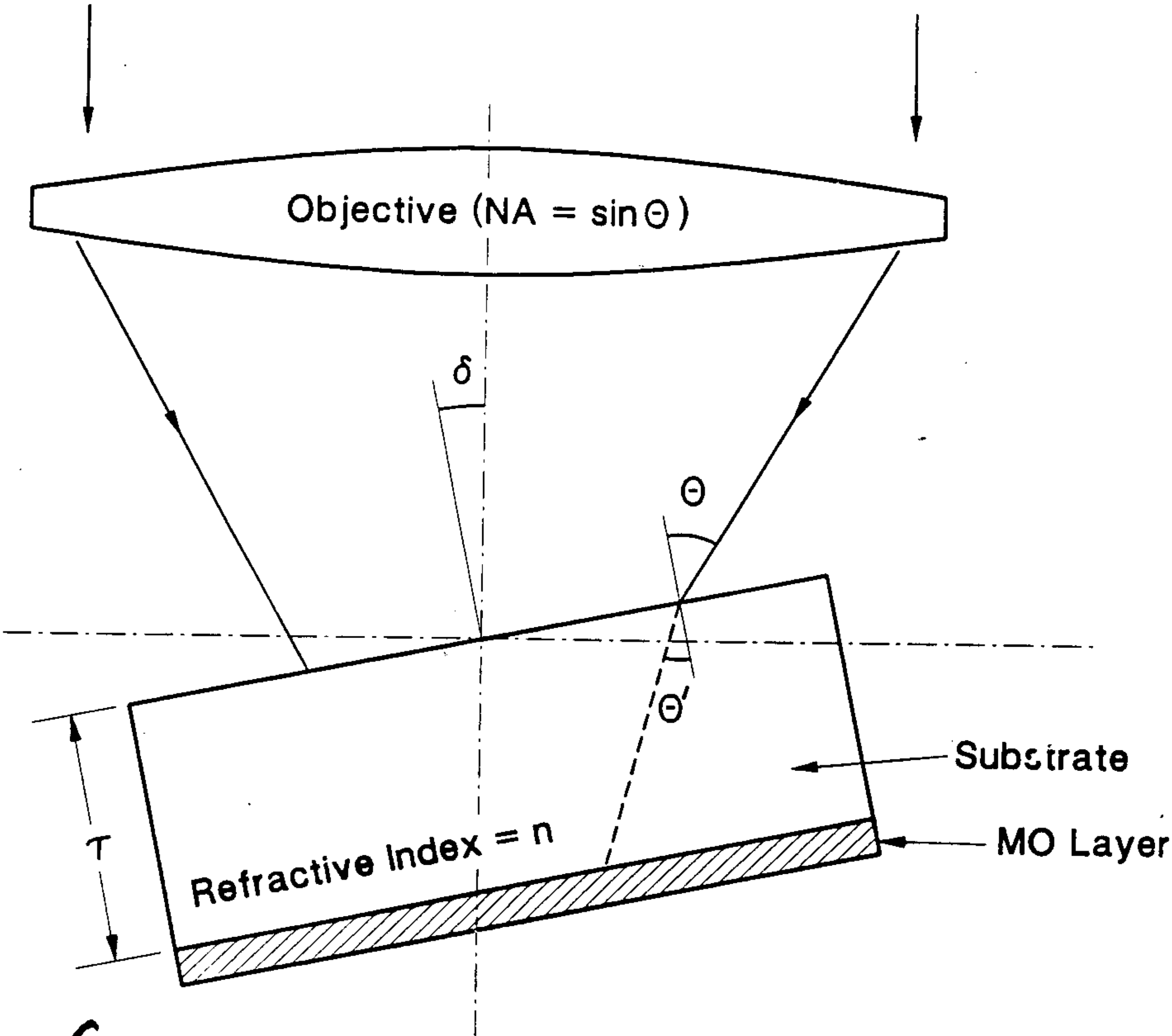


Fig. 7(b)

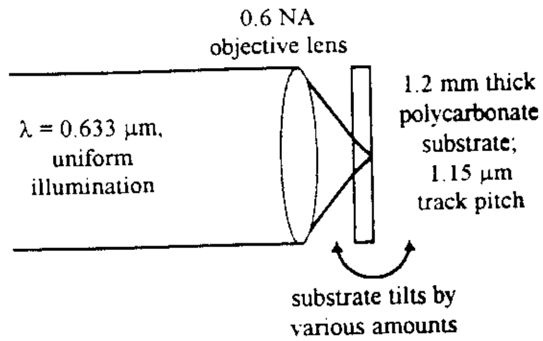


Coma:

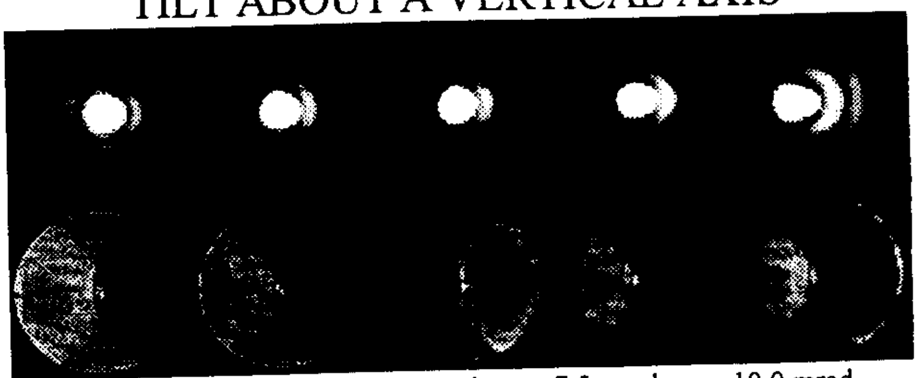
$$W_{31} = \left(\frac{1}{n} - \frac{1}{n^3}\right)\left(\frac{T}{\lambda_0}\right) NA^3 \delta$$

8-p6

MEASURED EFFECT OF DISK TILT



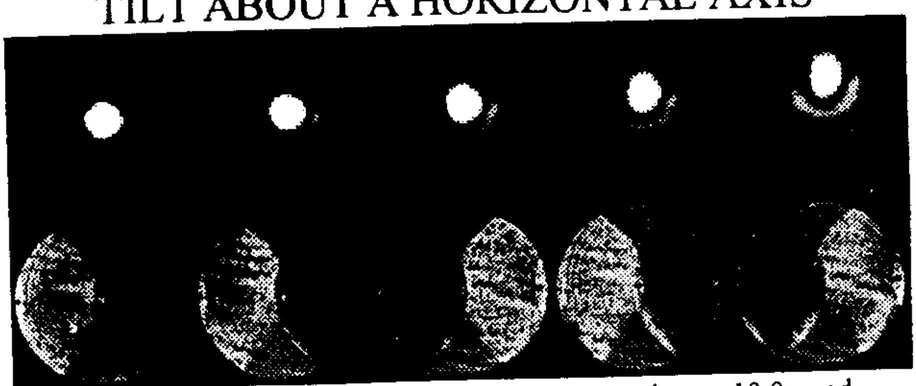
TILT ABOUT A VERTICAL AXIS



no tilt | 2.5 mrad | 5.0 mrad | 7.5 mrad | 10.0 mrad

$\delta = 0°$ | 0.15° | 0.3° | 0.45° | 0.6°

TILT ABOUT A HORIZONTAL AXIS



no tilt | 2.5 mrad | 5.0 mrad | 7.5 mrad | 10.0 mrad

focused spot
with astigmatism

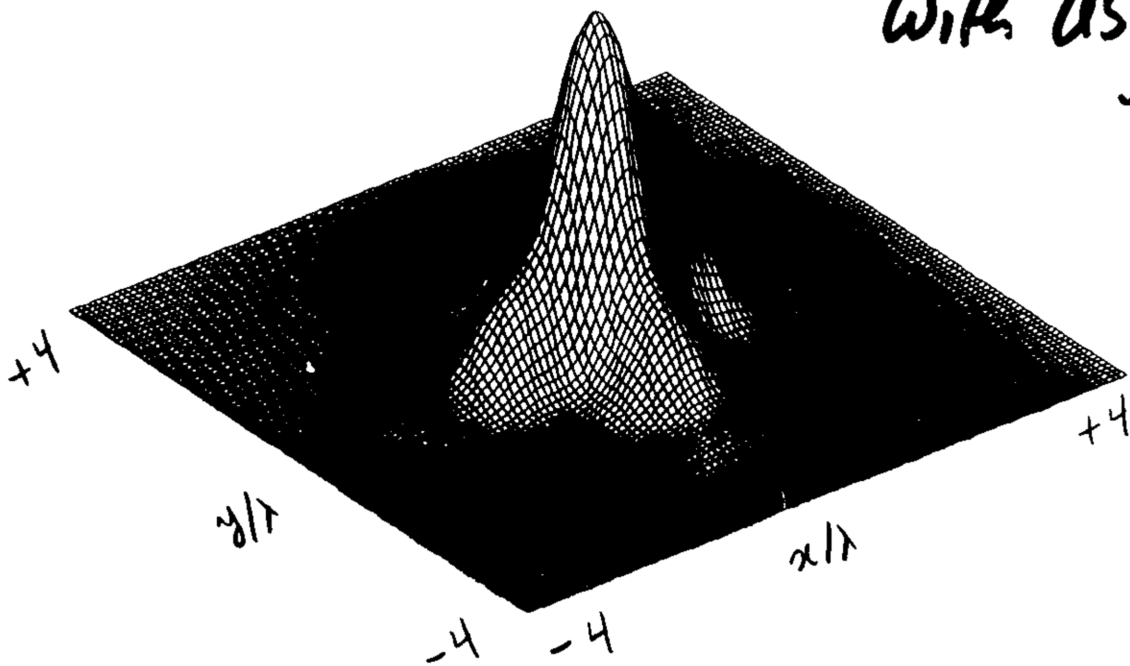
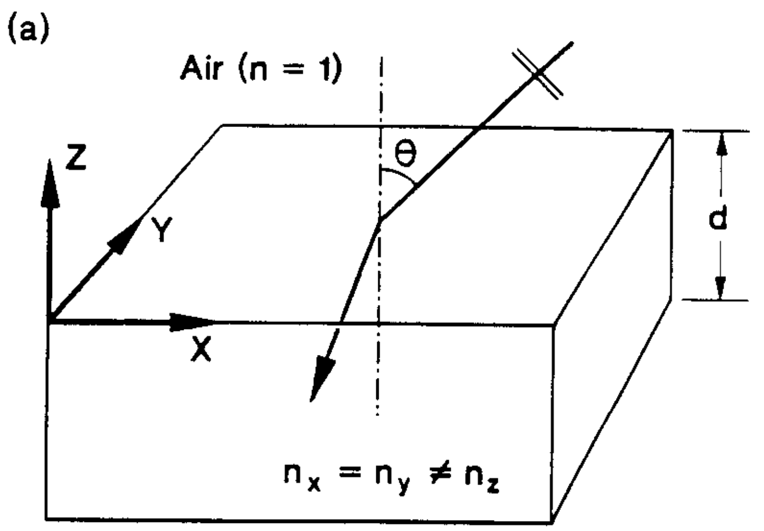
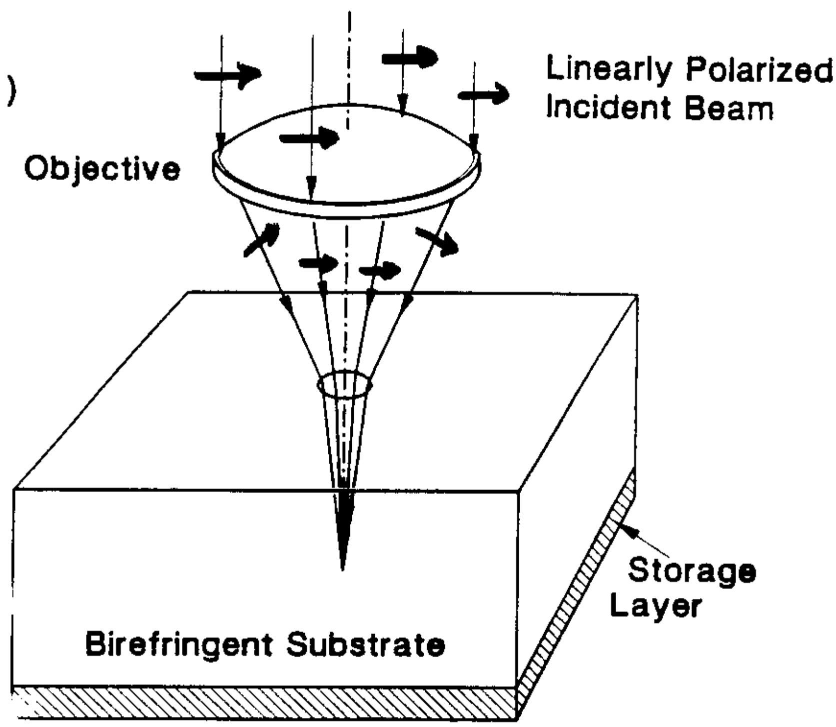


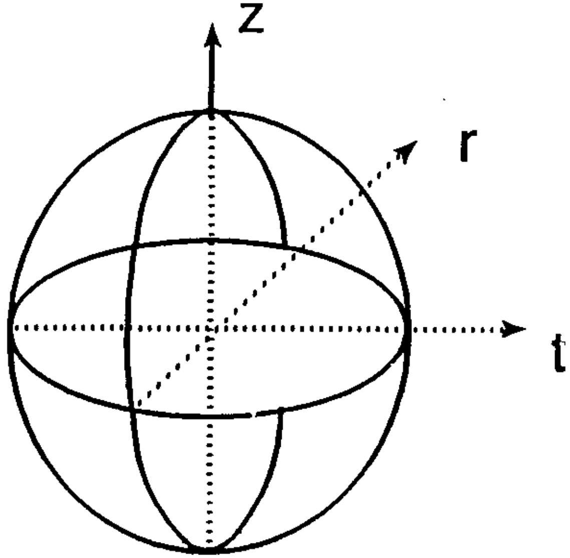
Fig. 7(C)



Vertical birefringence induces astigmatism



5-15



Index Ellipsoid

t : track direction

r : radial direction

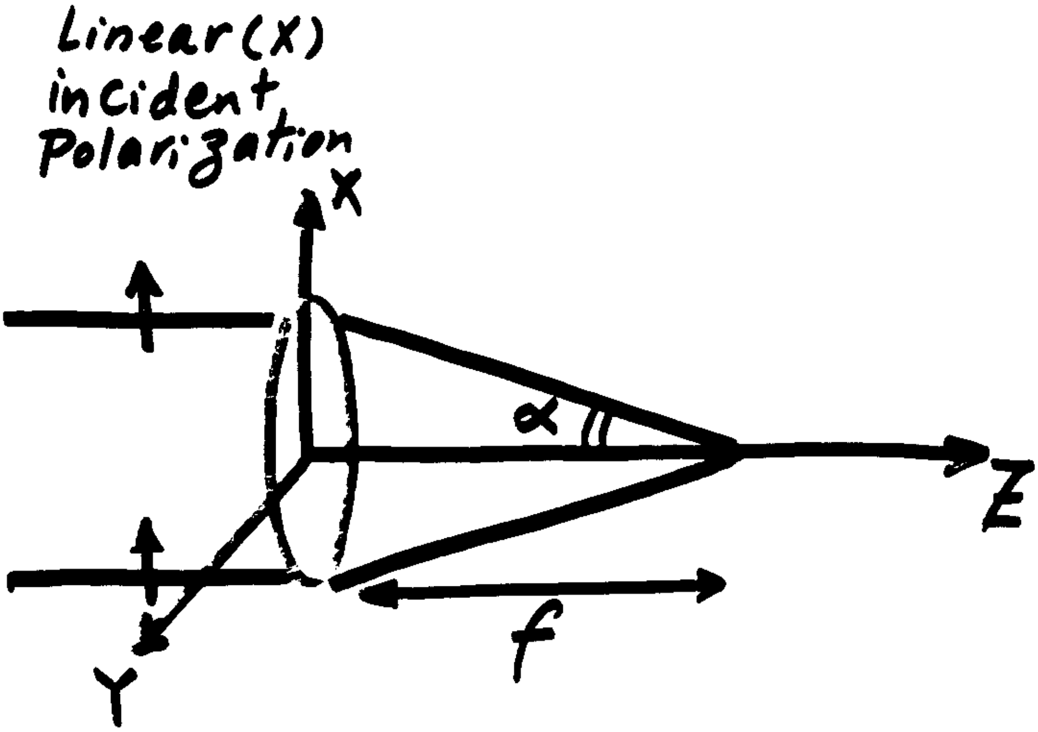
z : vertical direction

PC Substrate:

$$\Delta n_{\parallel} = n_r - n_t \approx 2 \times 10^{-5}$$

$$\Delta n_{\perp} = n_r - n_z \approx 5 \times 10^{-4}$$

High NA Spherical lens



$NA = \sin\alpha$

Incident Power = 1

X-polarized at focus

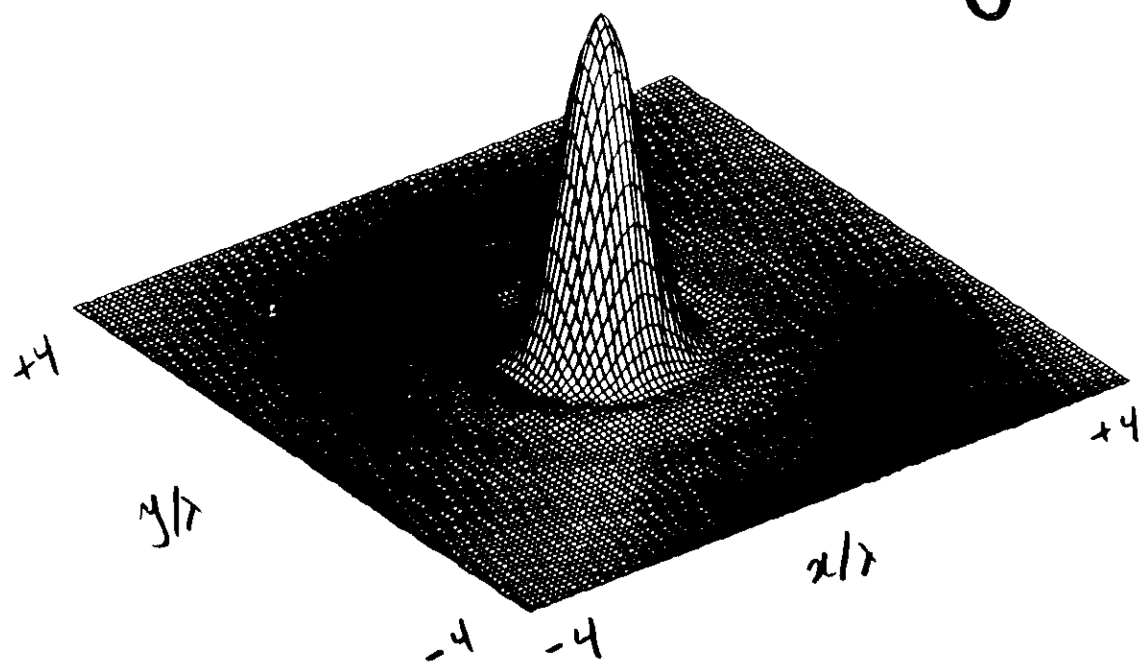


Fig. 9(a)

Y-polarized at focus

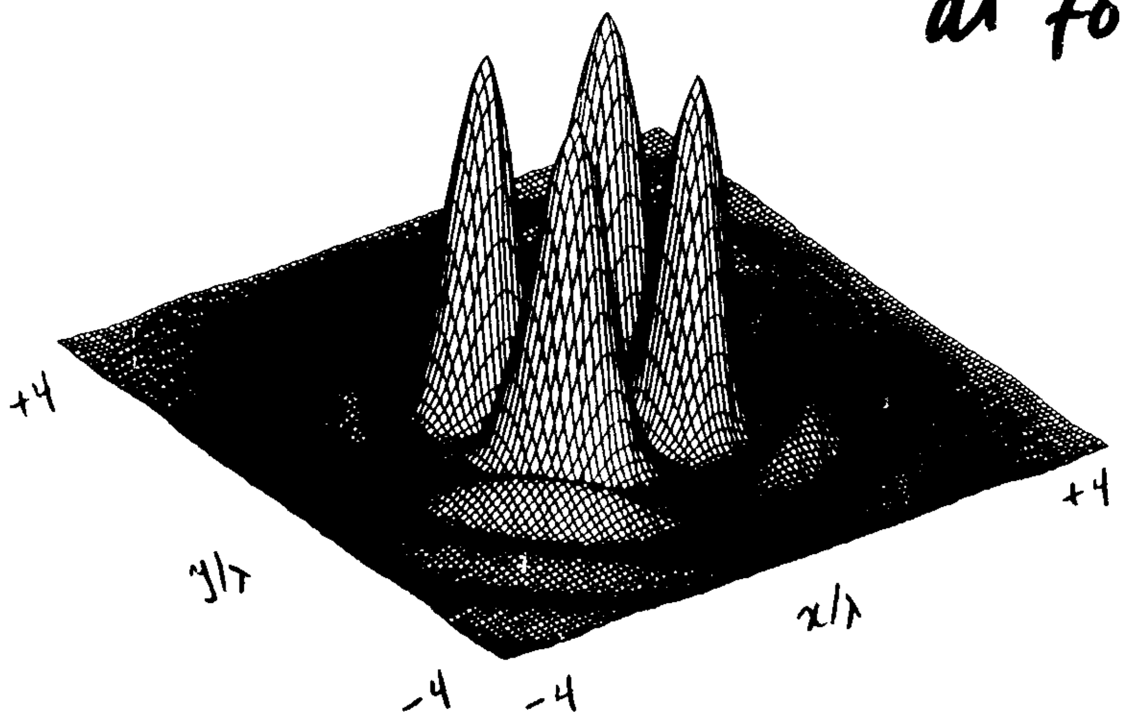


Fig. 9(b)

Z - polarized
at focus

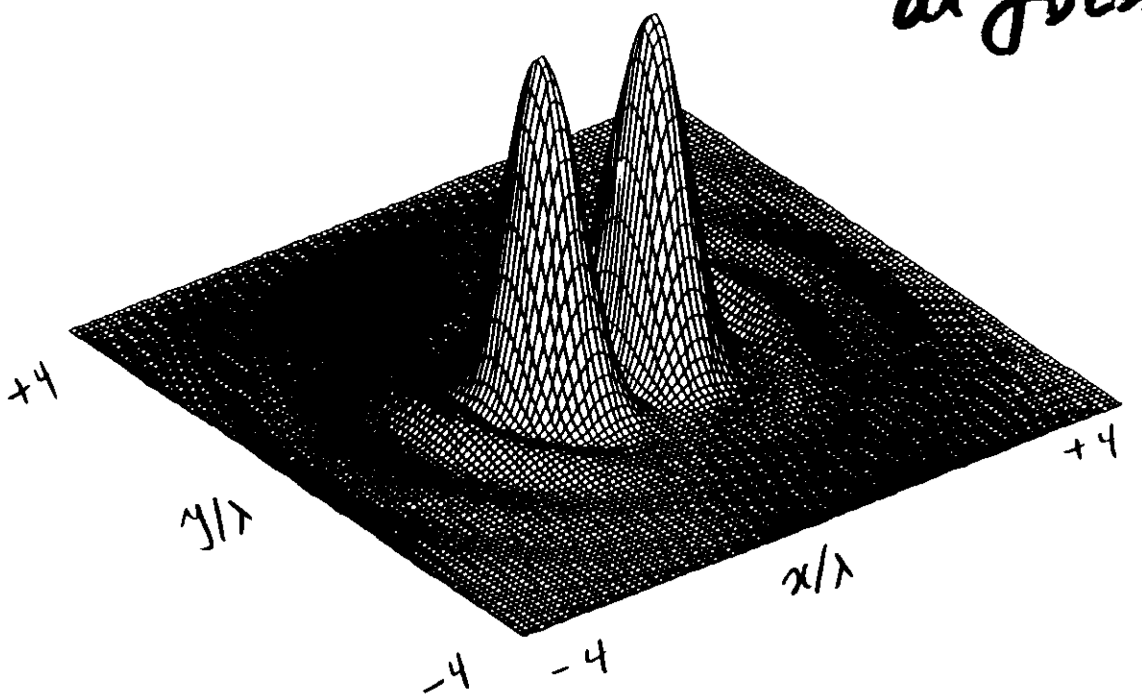
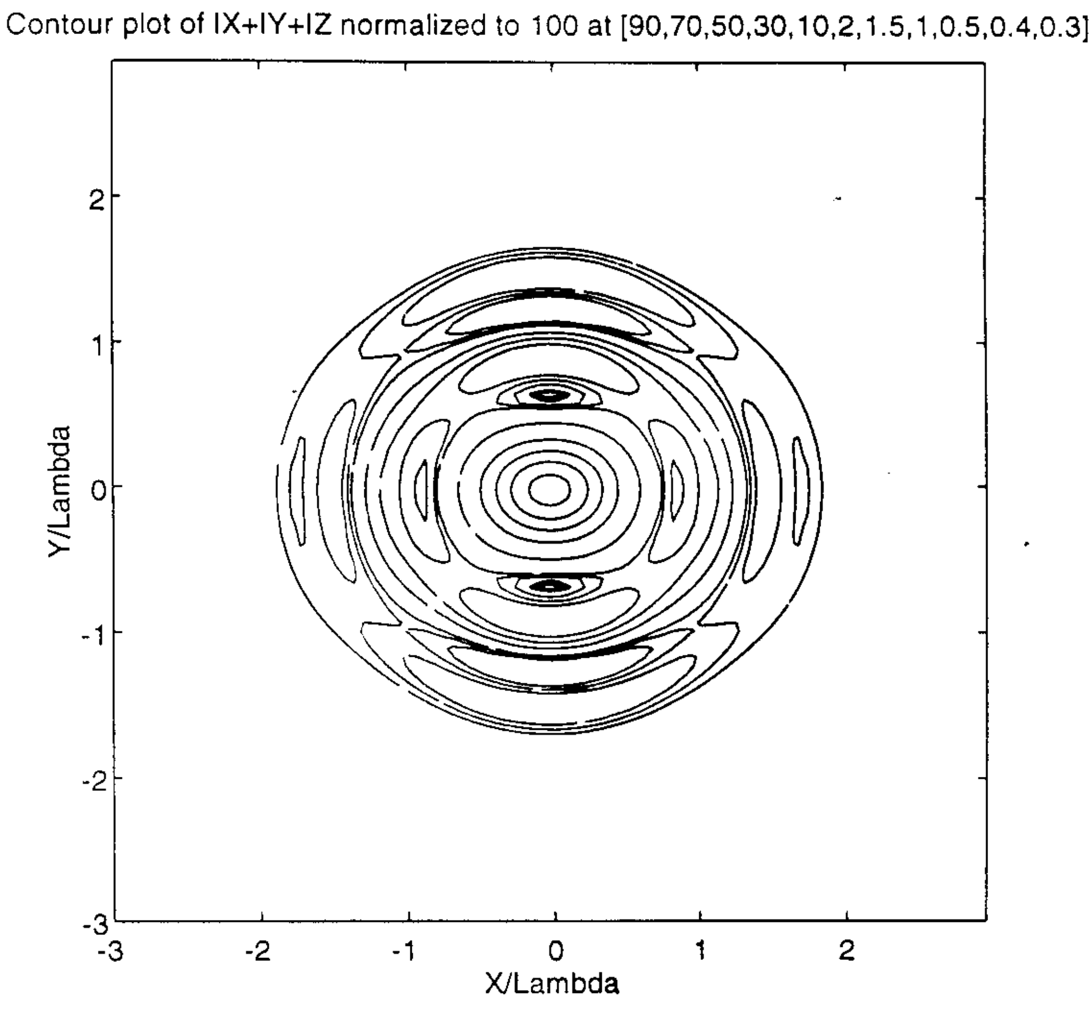
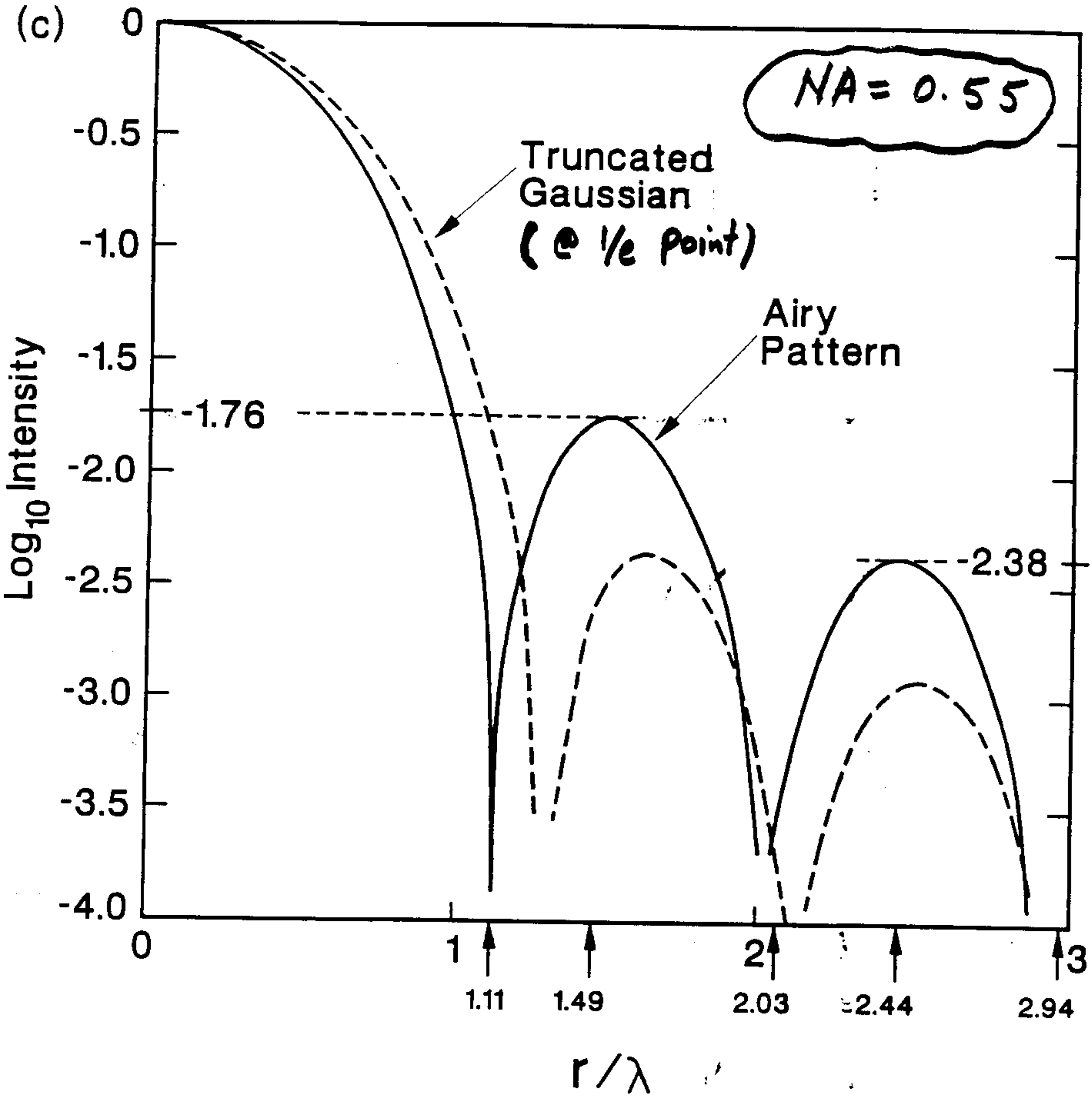


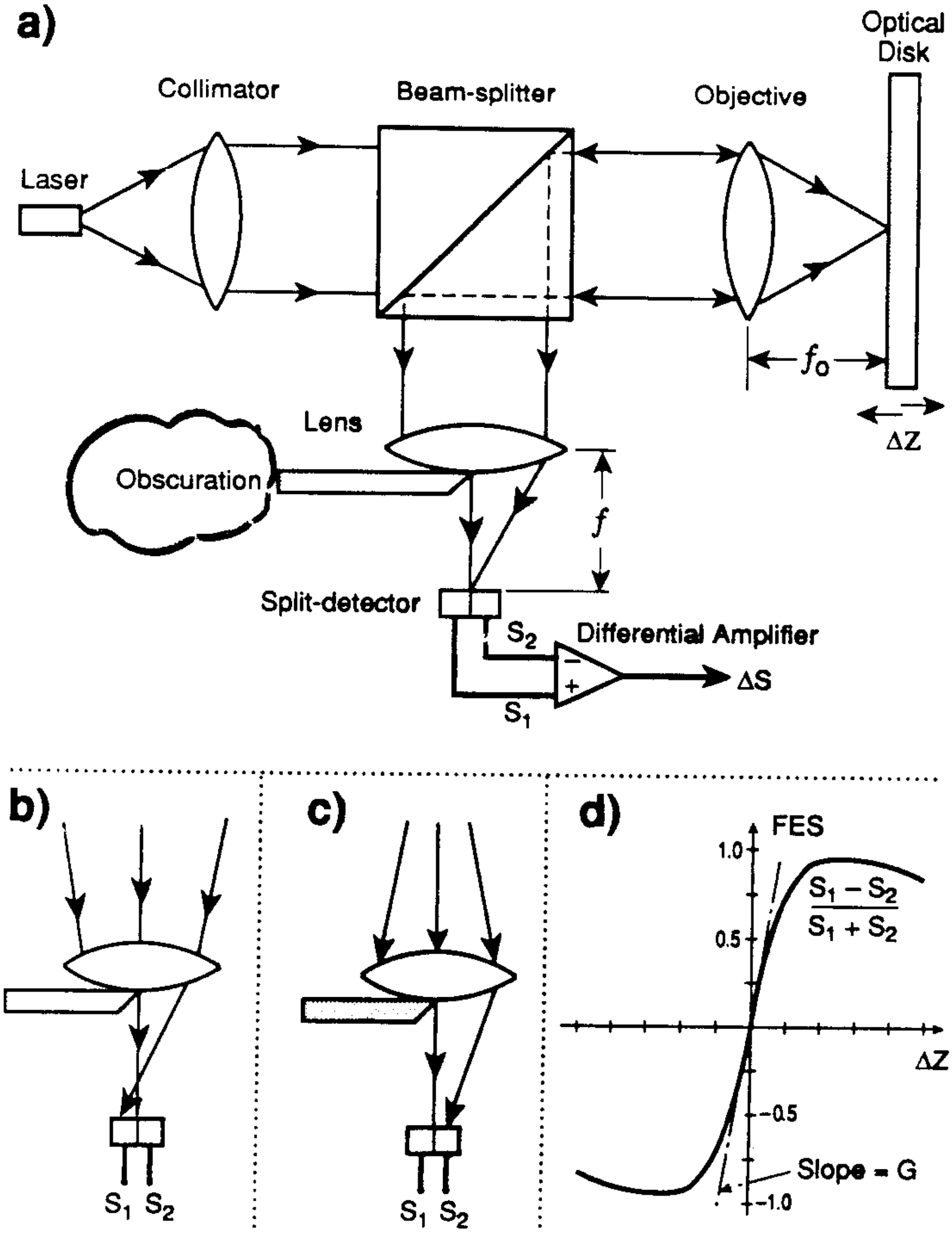
Fig. 9(c)


Contour plot of IX+IY+IZ normalized to 100 at [90,70,50,30,10,2,1.5,1,0.5,0.4,0.3]
2
1
0
-1
-2
-3
Y/Lambda
-3
-2
-1
0
1
2
X/Lambda

TK


(c)
NA = 0.55
Truncated
Gaussian
(@ 1/e point)
Airy
Pattern
-1.76
-2.38
0
-0.5
-1.0
-1.5
-2.0
-2.5
-3.0
-3.5
-4.0
Log_{10} Intensity
0
1
2
3
1.11
1.49
2.03
≈2.44
2.94
r/λ

8-2c



Focus Error Detection

Illustration by
Wally Varner / UA Graphics
Mansuripur fig 13

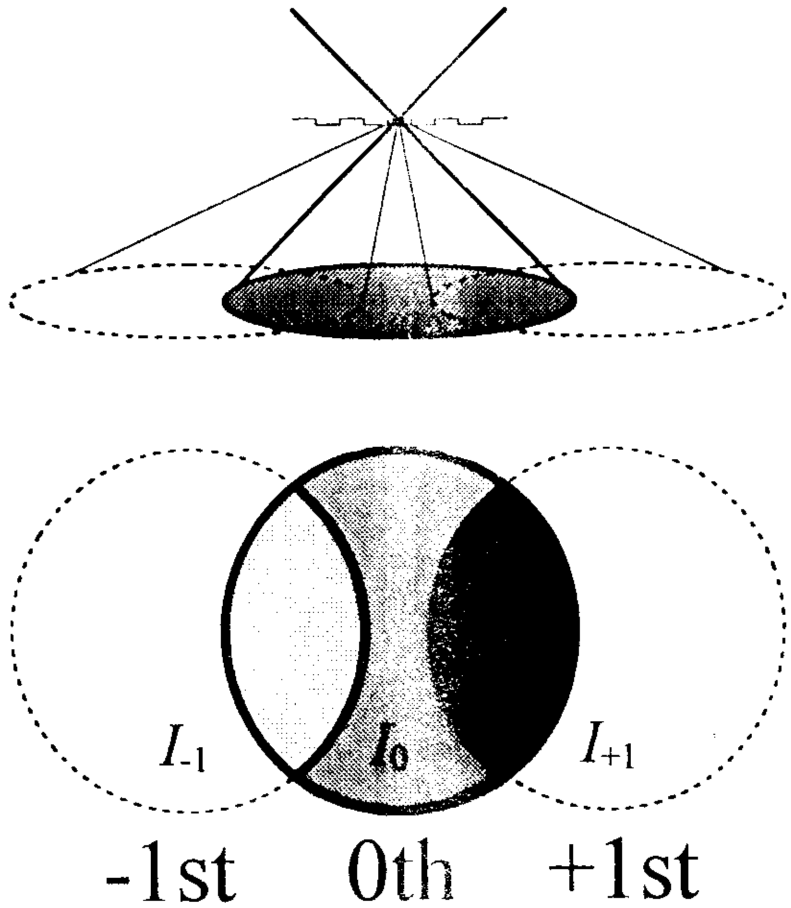
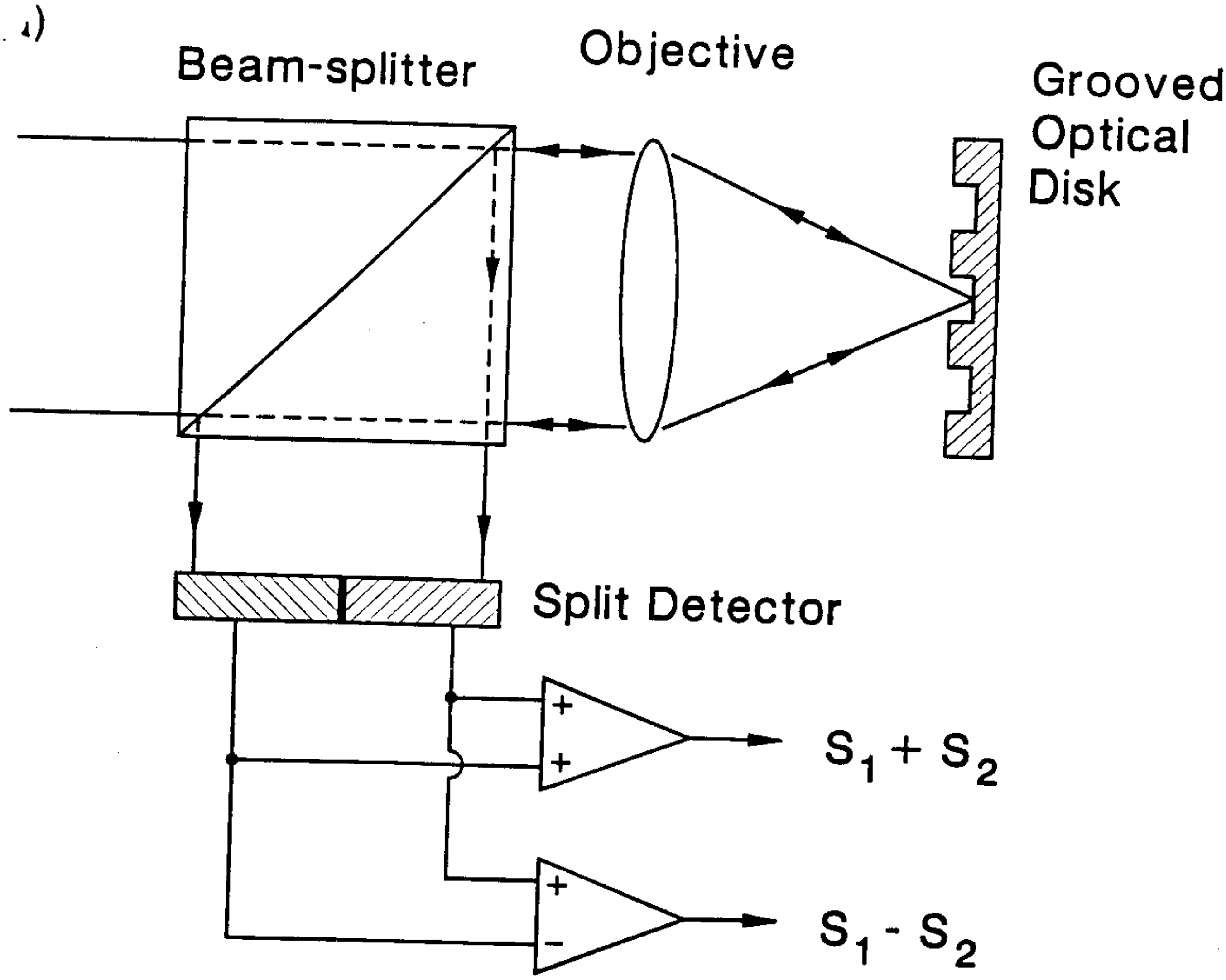


Fig. 2(a) Gerber

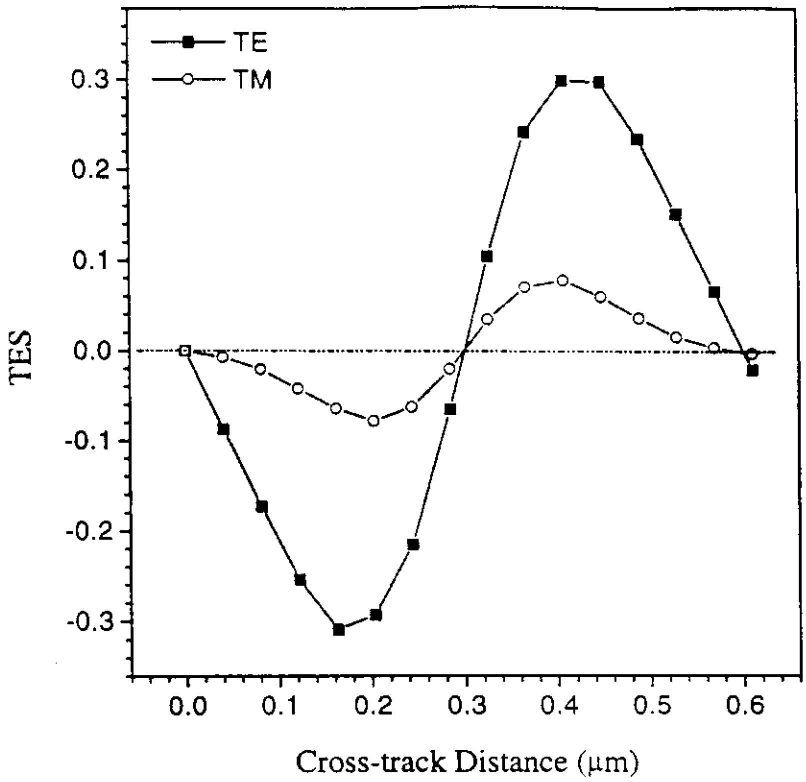
a)

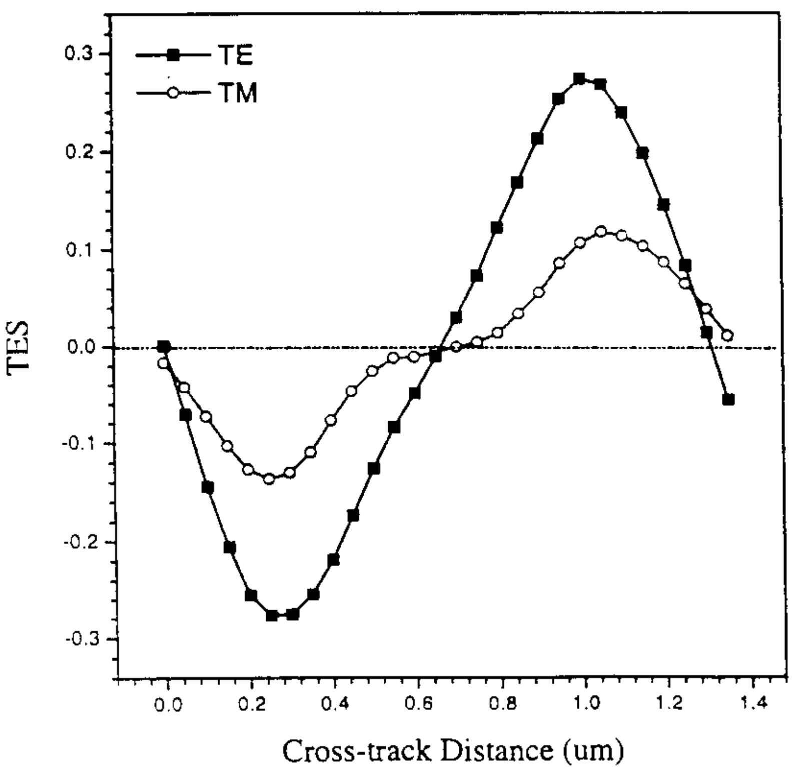


Track-crossing signal using push-pull method

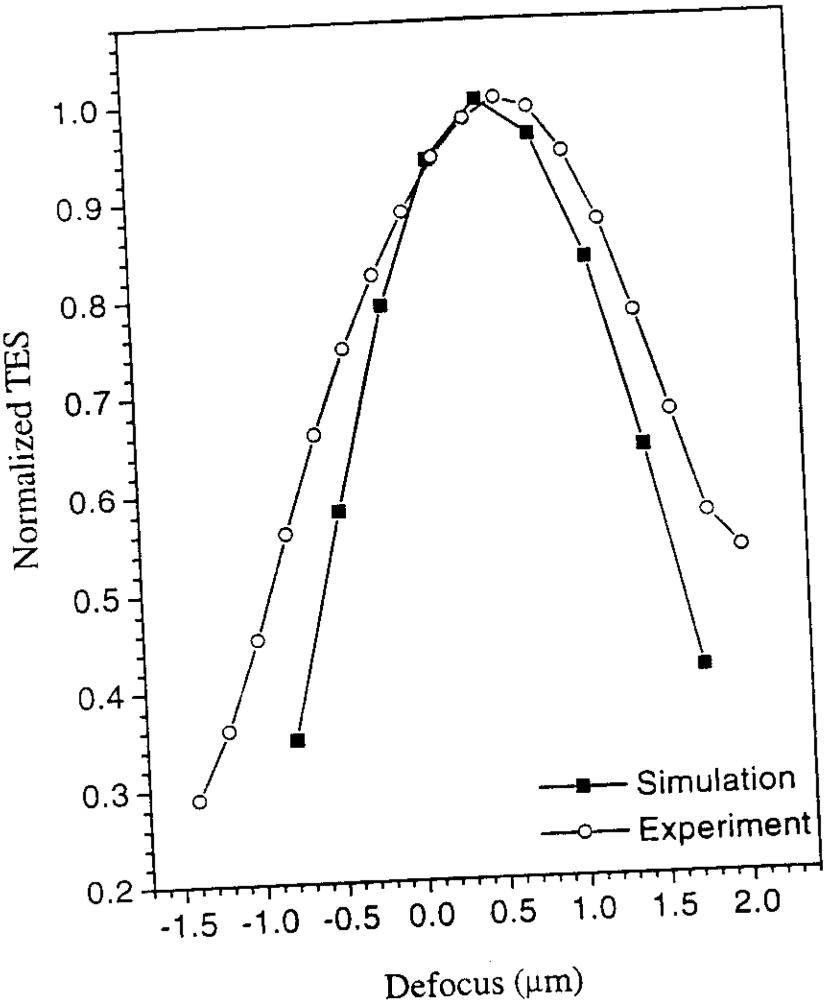
8-6a

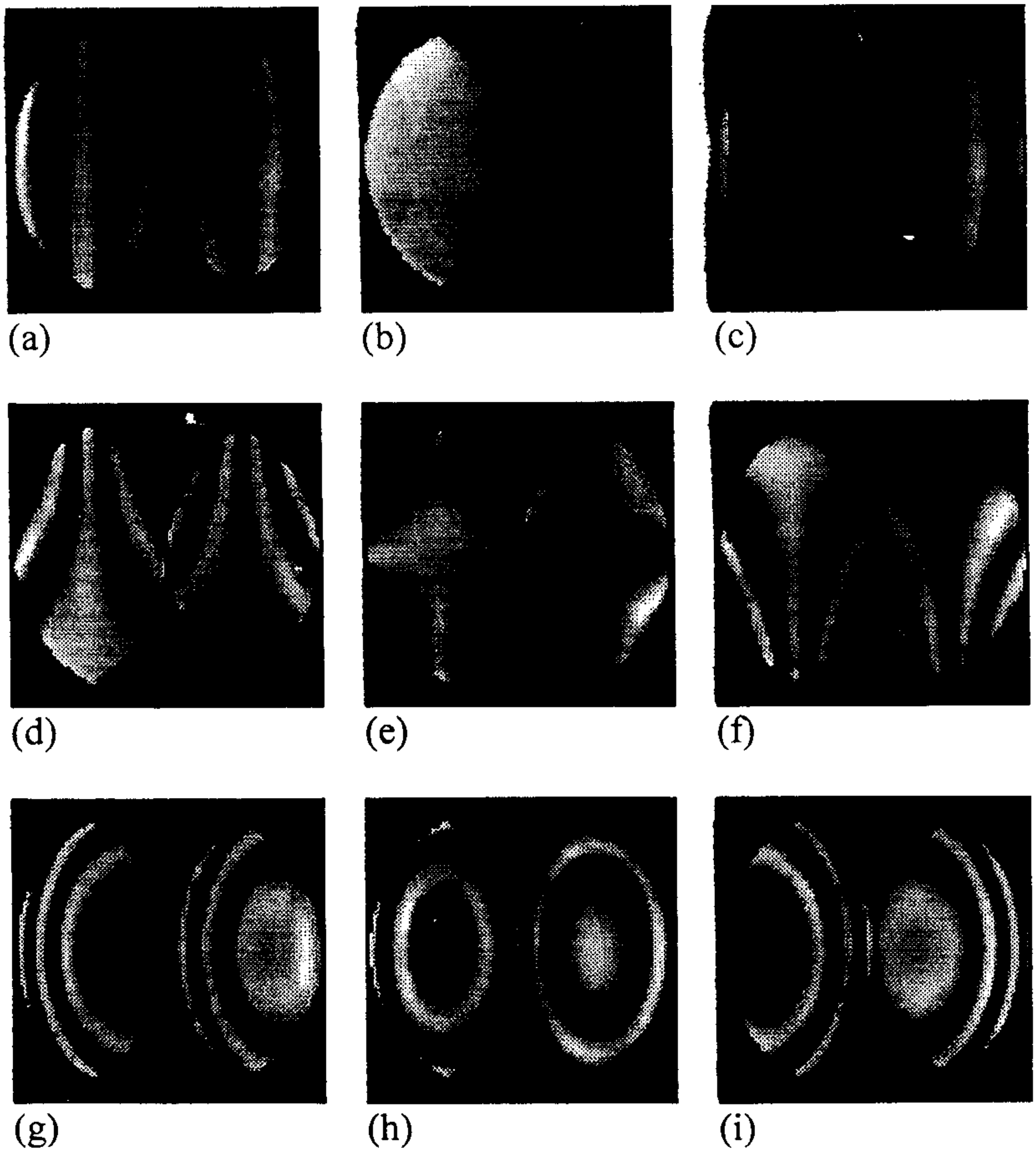
Simulation result for dielectric gratings with 0.6um.





The TES through focus for dielectric gratings with SIL.





Computed intensity Patterns at the exit pupil

Fig. 8 Gerber

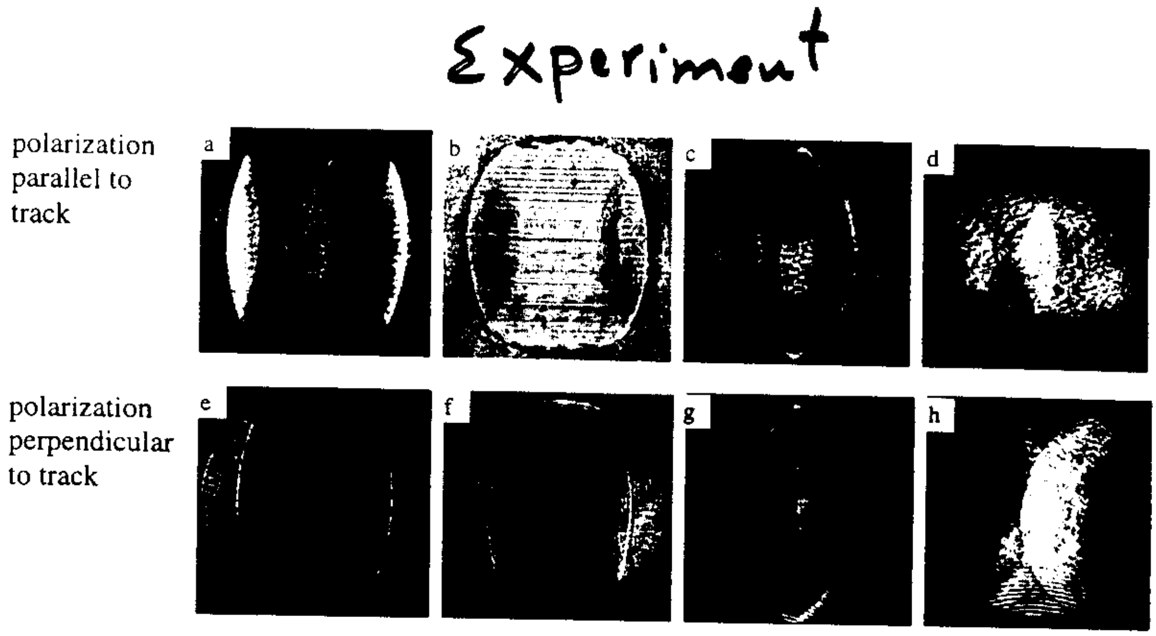


Figure 1. Images of the baseball pattern at the exit pupil of the objective lens.
(a) and (e): intensity distribution for the dielectric grating without SIL.
(b) and (f): phase distribution for the dielectric grating without SIL.
(c) and (g): intensity distribution for the dielectric grating with SIL.
(d) and (h): intensity distribution for the metal grating with SIL.

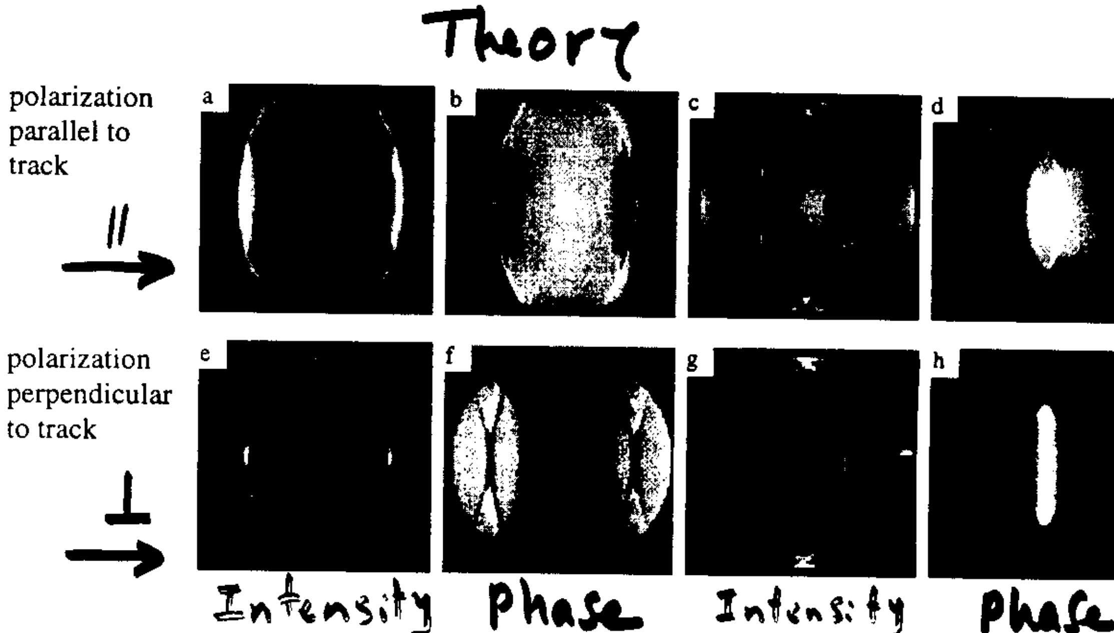
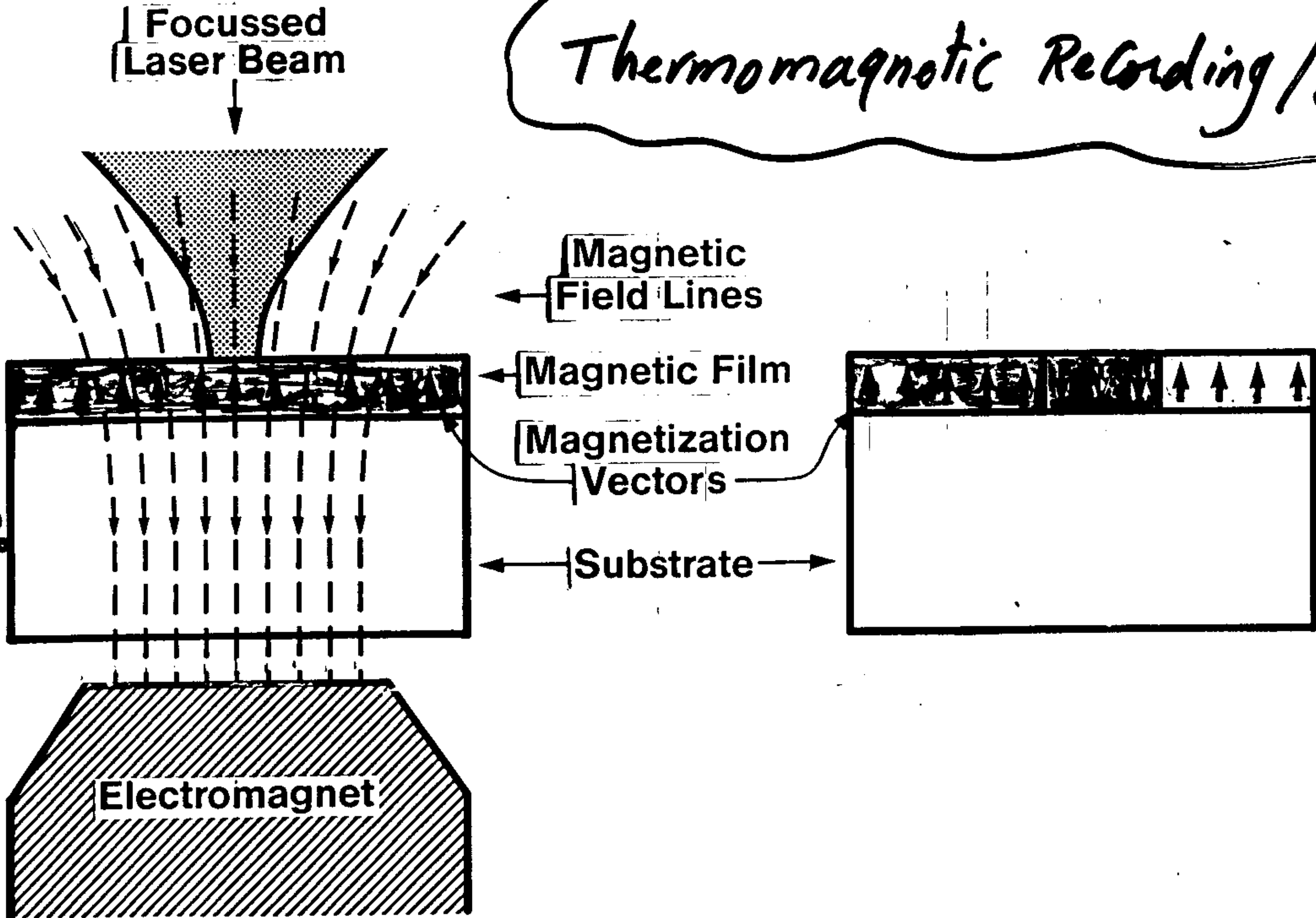


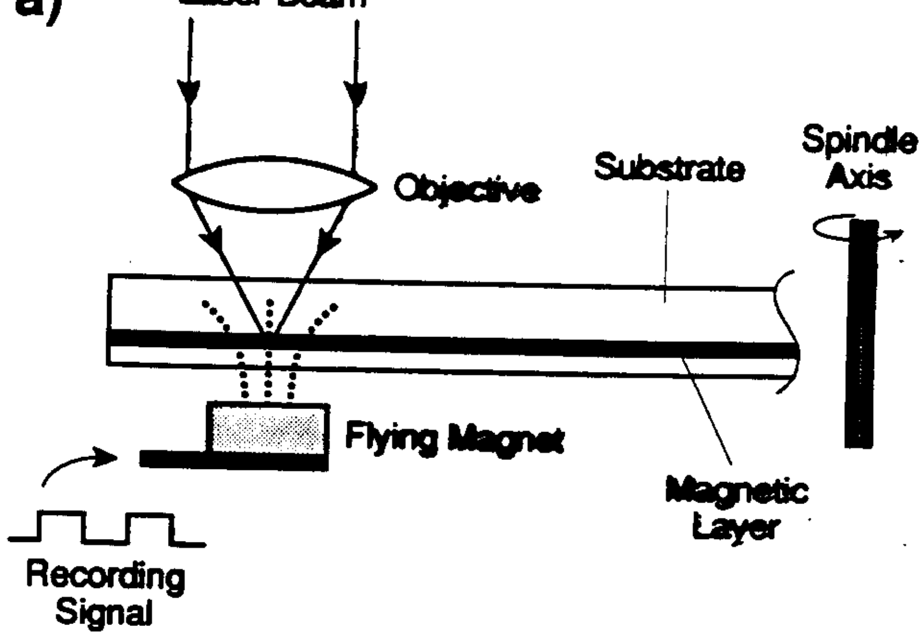
Figure 2. Simulation results corresponding to the images shown in Fig. 1.

Thermomagnotic Recording/Erasure



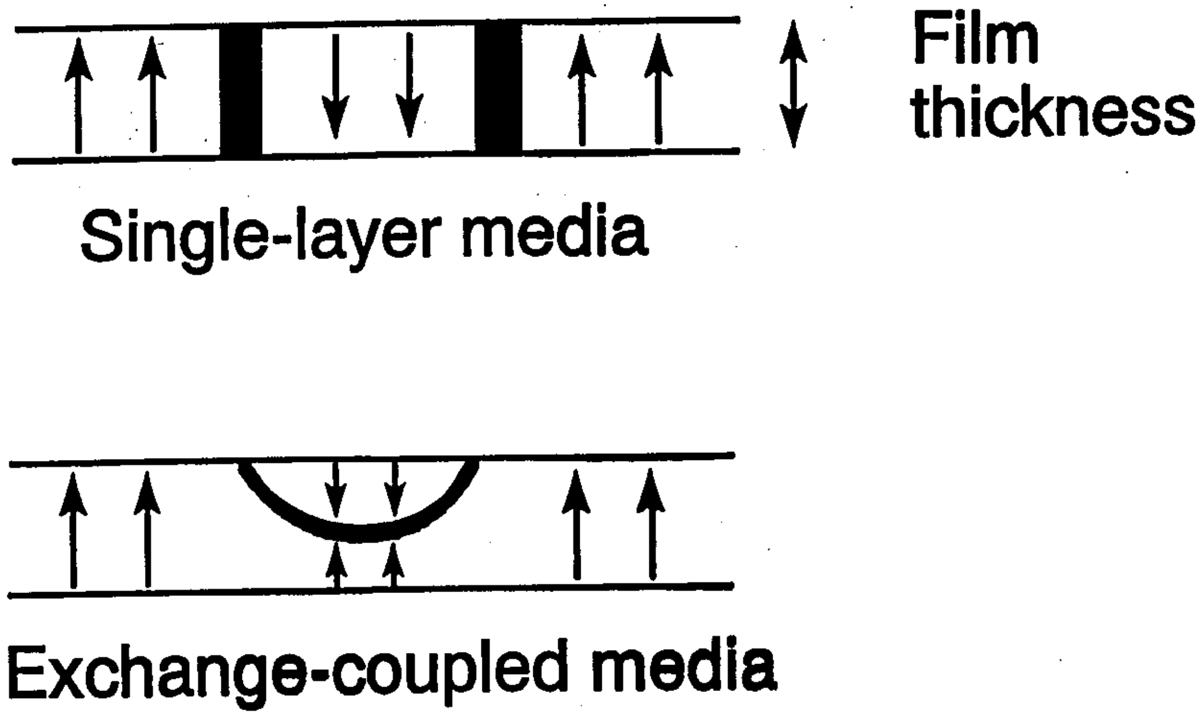
Recording by Magnetic Field Modulation

a)

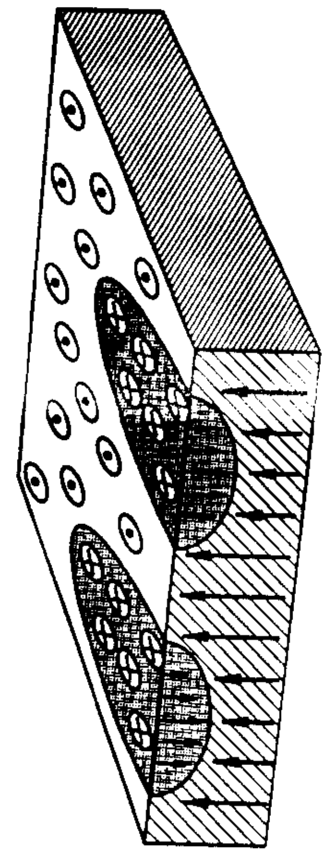


b)

Illustration by
Wally Varner / UA Graphics
Mansuripur fig 18 a b

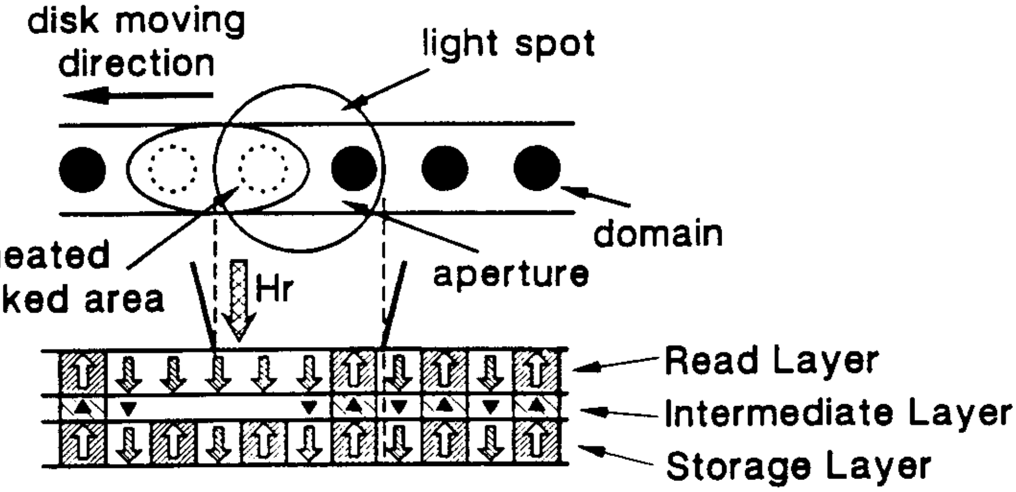

Film
thickness
Single-layer media
Exchange-coupled media

7.

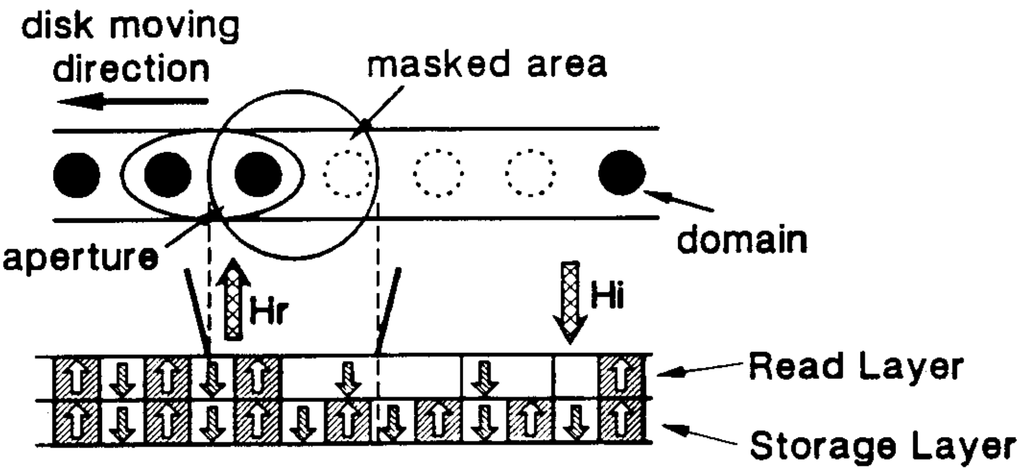


Magnetic Super Resolution (MSR)

(a)

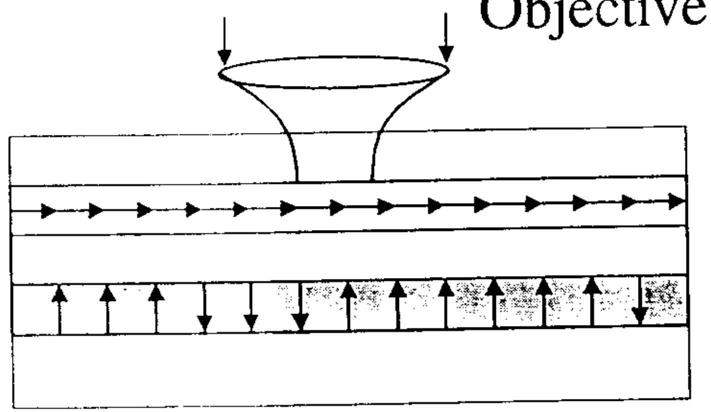


(b)



17-22

CAD-MSR
(Central Aperture Detection Magnetic Super Resolution)



- Laser spot heats up the read layer
- Magnetic domains are copied, one bit at a time, from the storage layer to the read layer

MAGNETO-OPTIC MATERIALS

__ GdCo

__ GdFe

__ TbFe

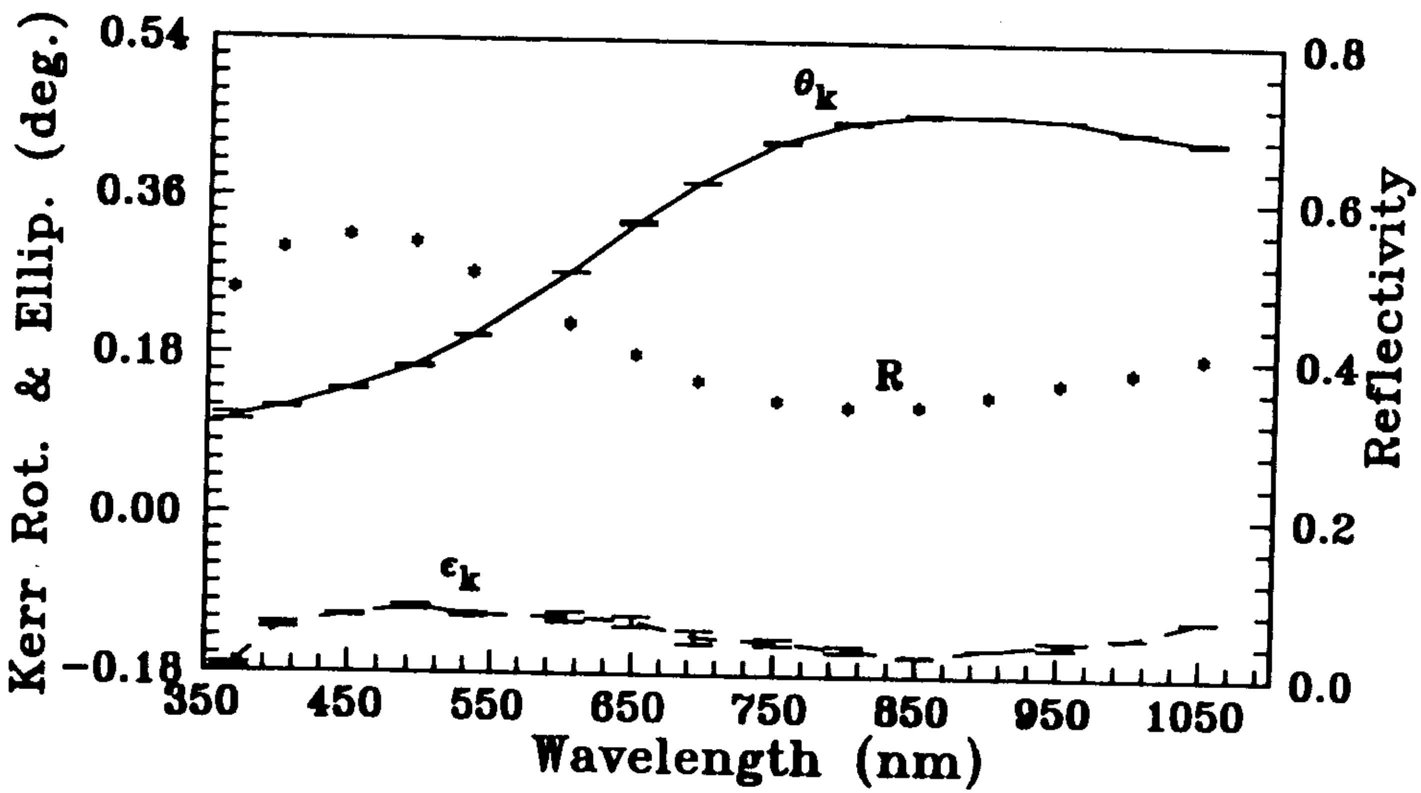
__ DyFe

__ GdTbFe

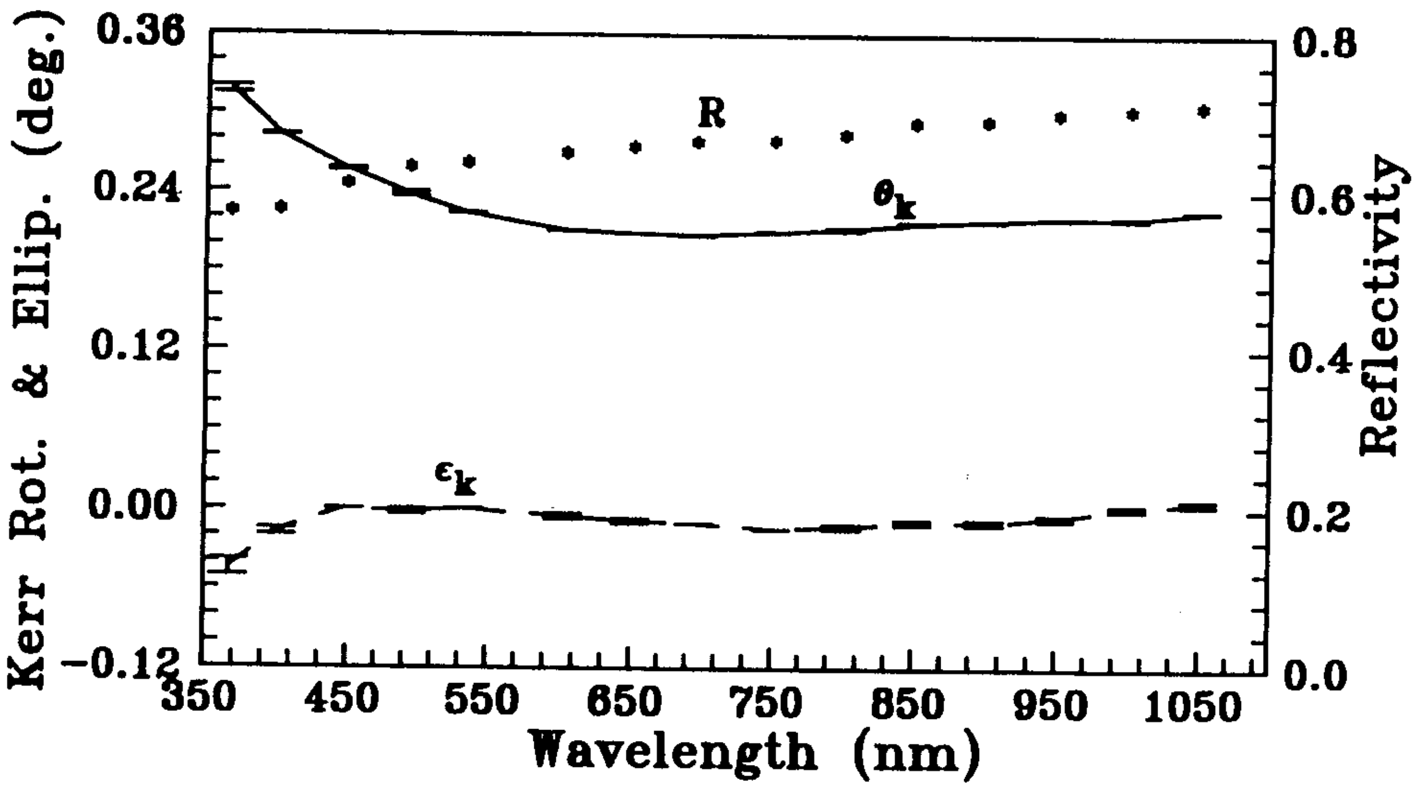
✓__ TbFeCo

✓__ GdTbFeCo

Sample: $Tb_{28.1}Fe_{71.9}$, MO film 1354 Å, SiO 1264 Å (overcoating).



Sample: Co(4Å)/Pt(10Å), 300 Å total, glass substrate.



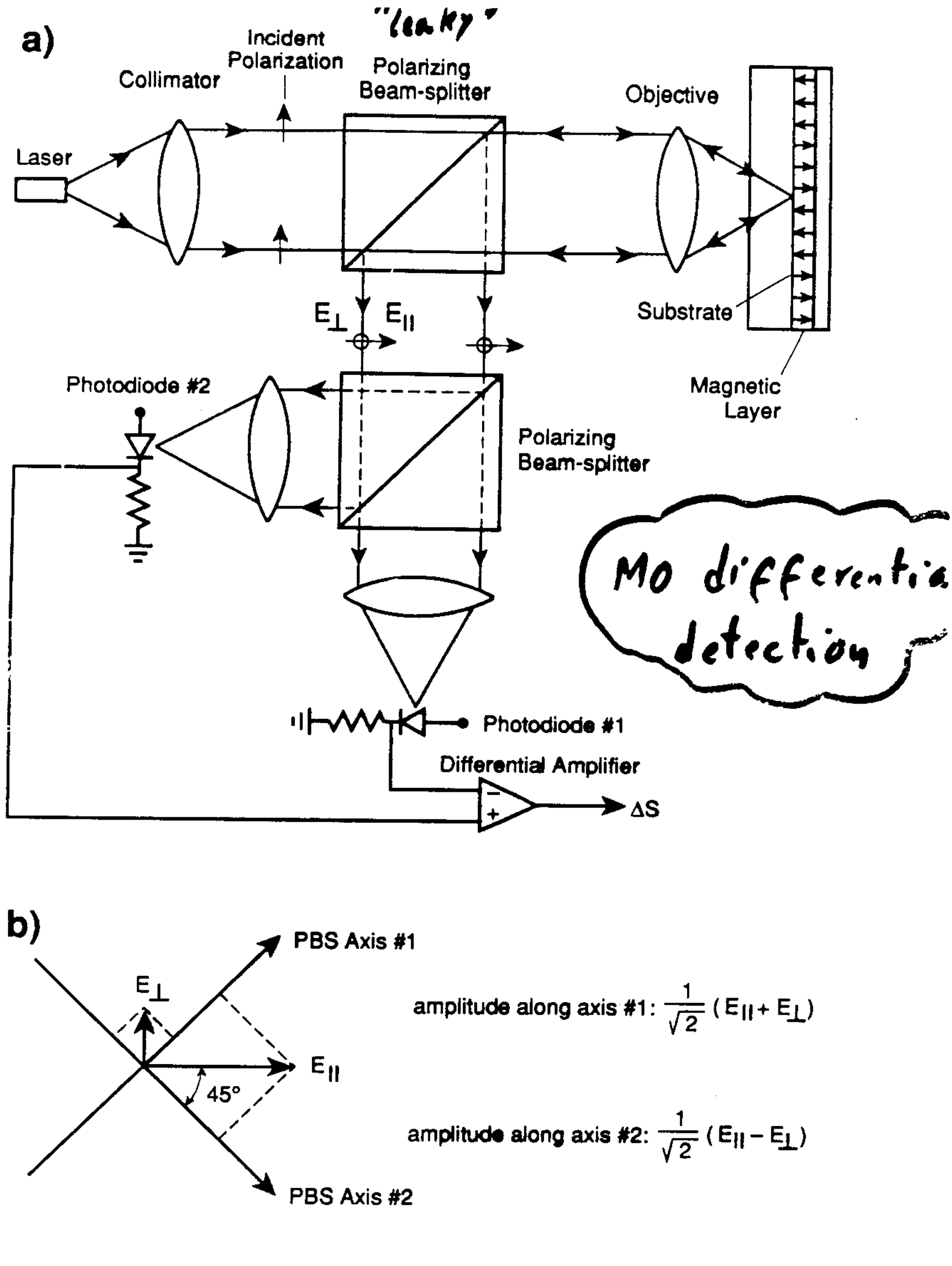
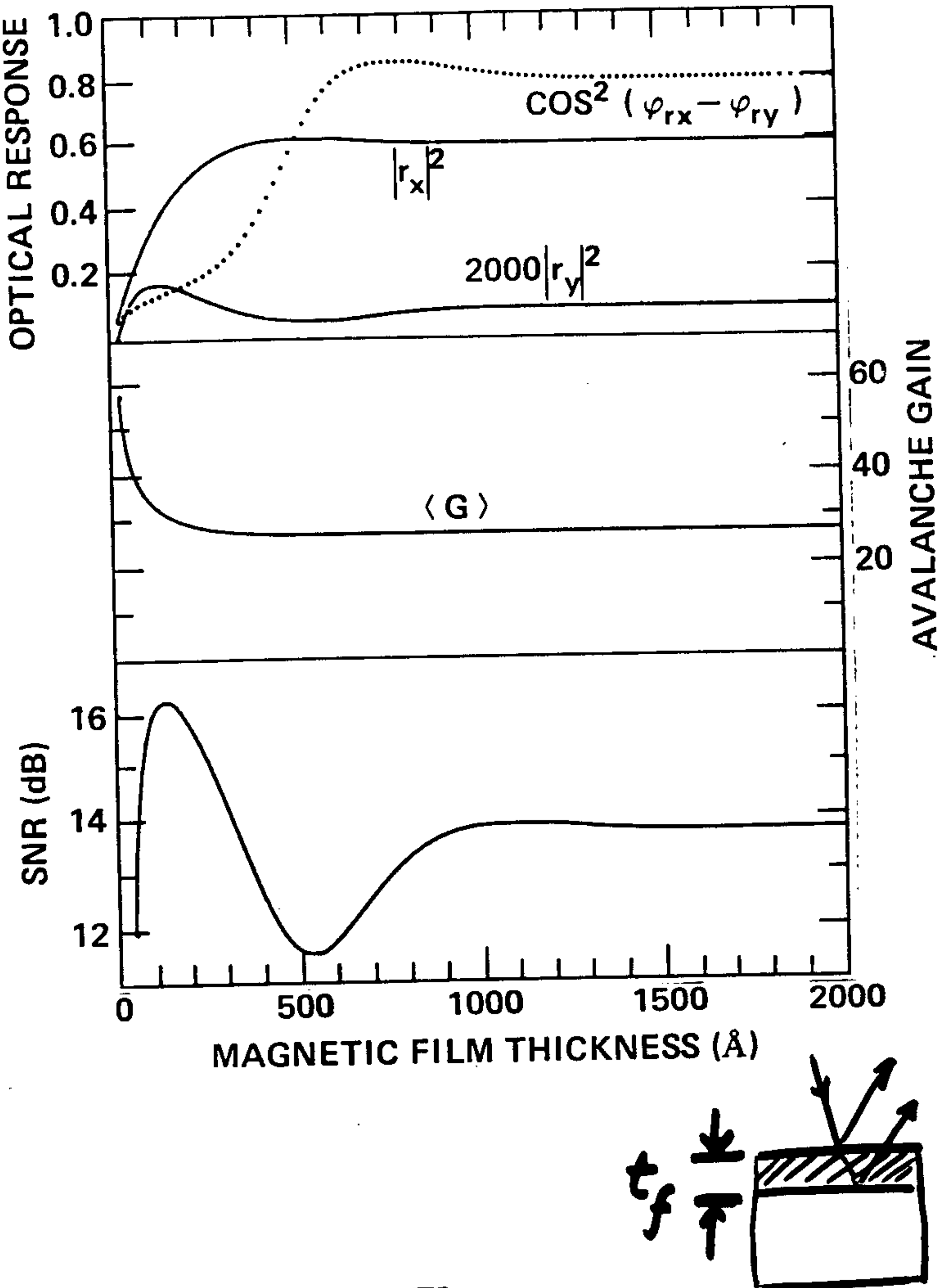
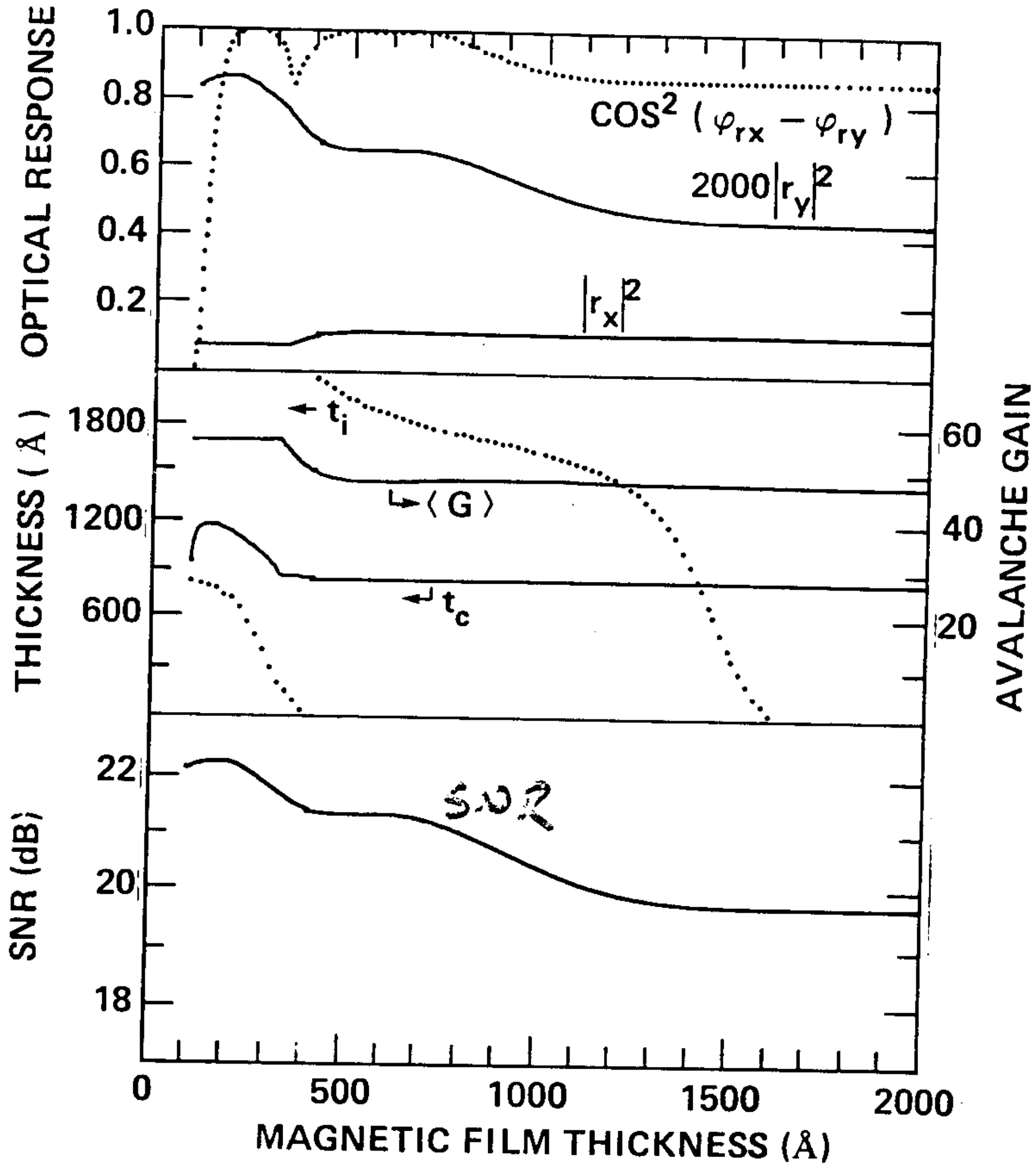
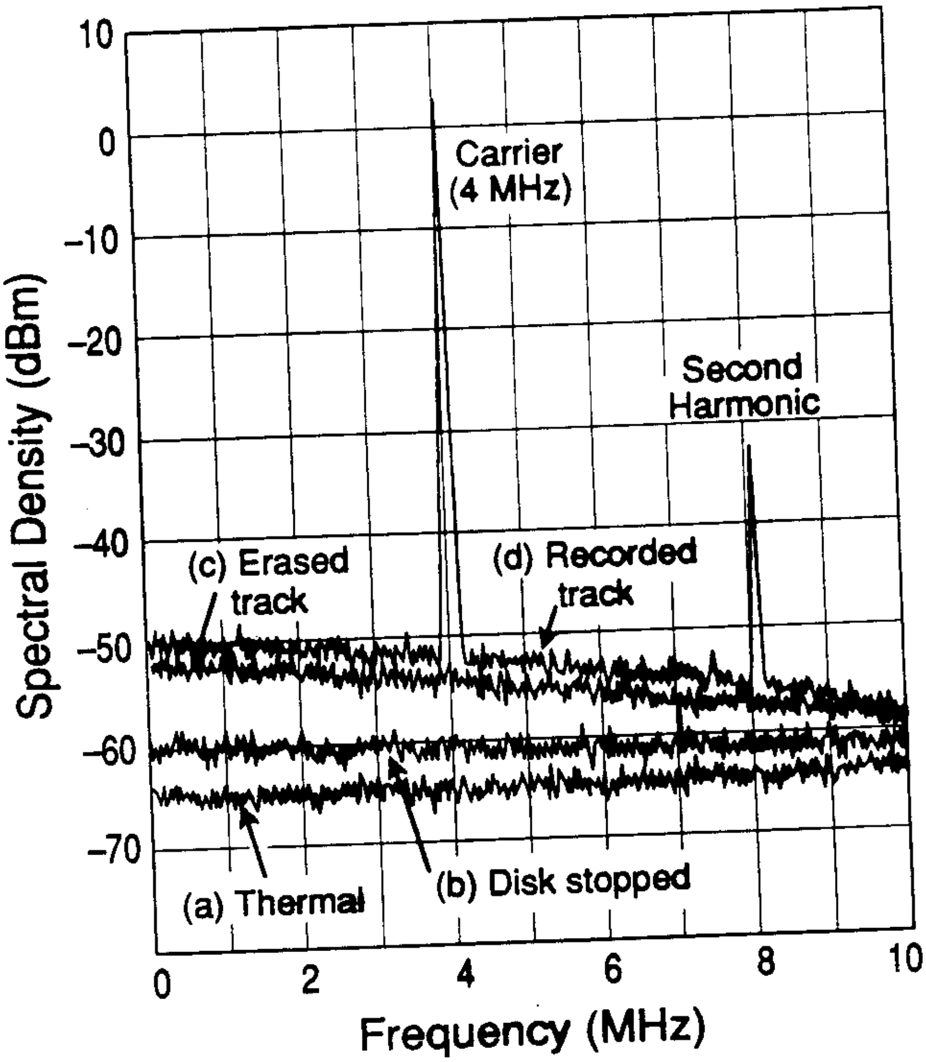


Illustration by
Wally Varner / UA Graphics
Mansuripur fig 27


1.0
0.8
0.6
0.4
0.2
OPTICAL RESPONSE
COS² (φrx − φry)
|rx|²
2000|ry|²
60
40
20
AVALANCHE GAIN
⟨ G ⟩
16
14
12
SNR (dB)
0
500
1000
1500
2000
MAGNETIC FILM THICKNESS (Å)
tf

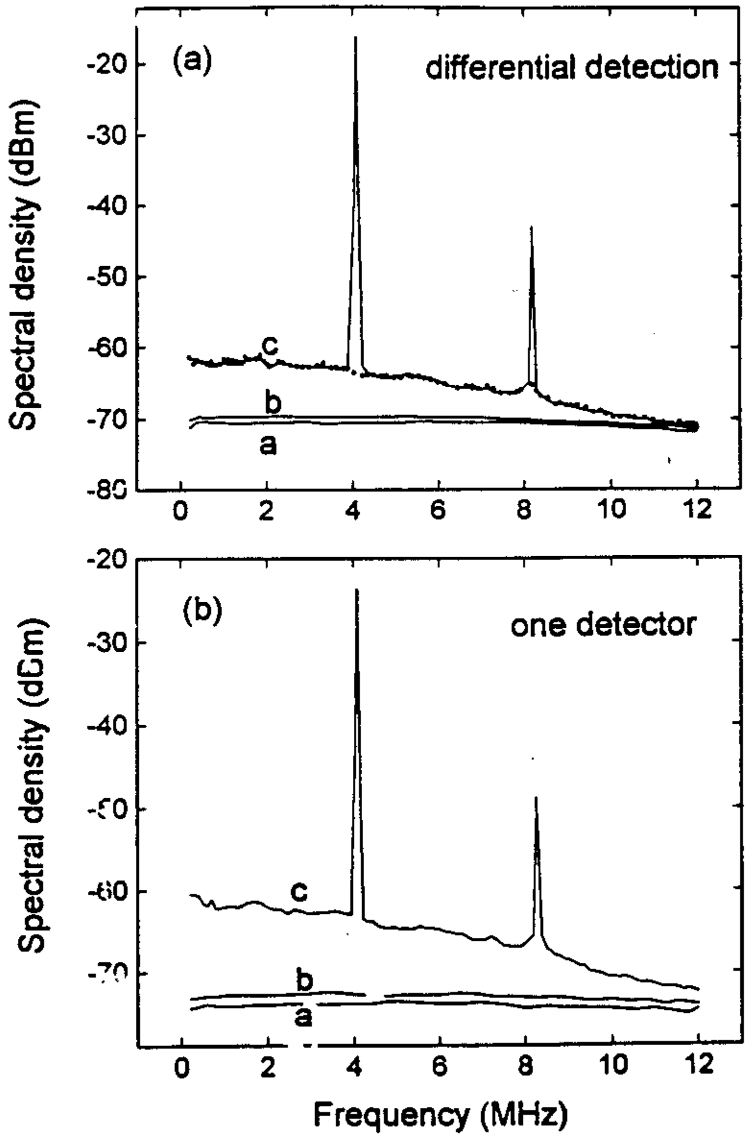

OPTICAL RESPONSE
1.0
0.8
0.6
0.4
0.2
COS^2 ($\varphi_{rx} - \varphi_{ry}$)
2000$|r_y|^2$
$|r_x|^2$
THICKNESS (Å)
1800
1200
600
t_i
⟨G⟩
t_c
AVALANCHE GAIN
60
40
20
SNR (dB)
22
20
18
SNR
0
500
1000
1500
2000
MAGNETIC FILM THICKNESS (Å)



Noise in MO Readout

Illustration by
Wally Varner / UA Graphi
Mansuripur fig 2?

4X MO disk (5.25"), $\lambda = 690$ nm



$\Delta r_{xy} \simeq 0.055\%$

depolarization ←

$\Delta r \simeq 0.23\%$

reflectivity fluctuation ←

Figure 3 Spectra of various noise components at the differential output of a magneto-optical readout system. The traces in (a) were obtained with the light blocked from both detectors (trace (a)), with the disk stationary and the light allowed to reach the detectors (trace (b)), and with the disk spinning and light reaching the detectors (trace (c)). The circles in the figure stand for the noise spectrum while reading on the erased track. The spectra in (b) were obtained reading one written track with only one detector. The light for these measurements had a wavelength of 690 nm.

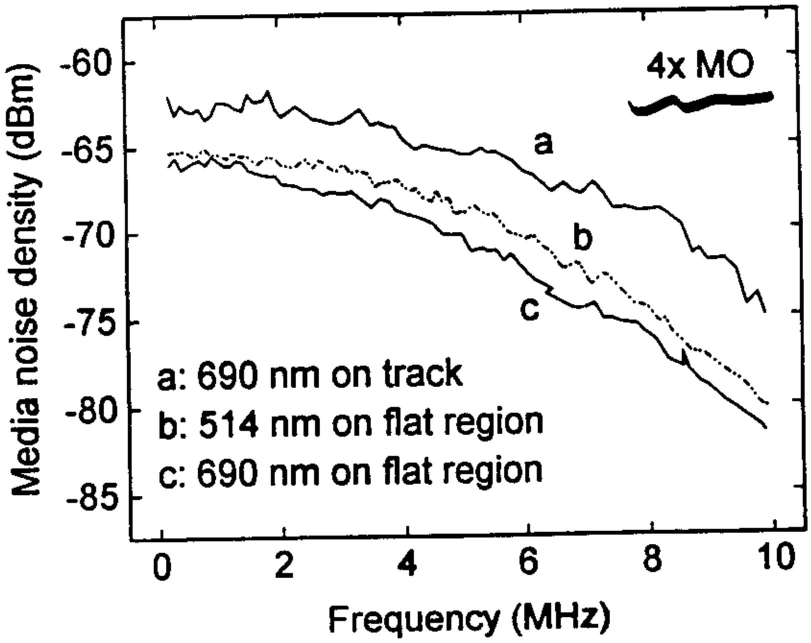
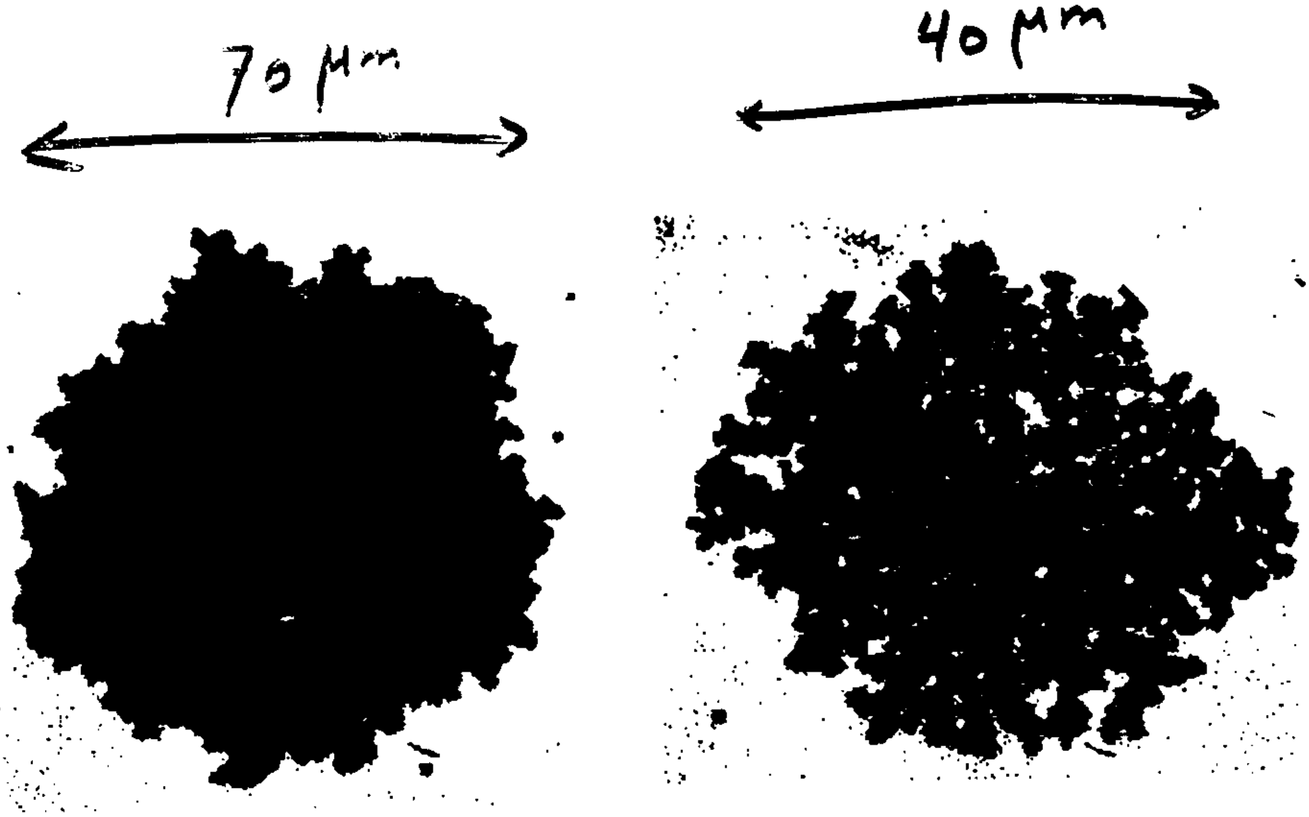
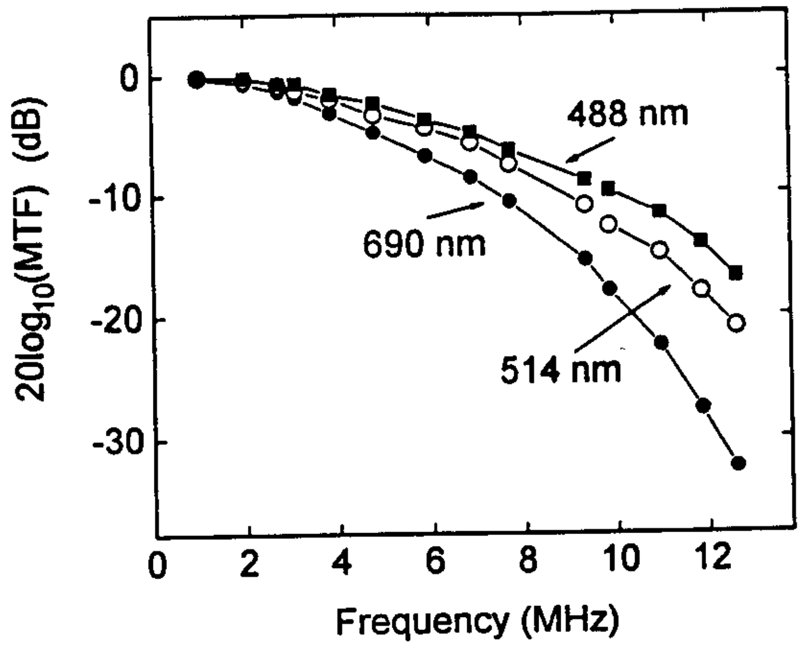


Figure 10 Media noise spectral density of reading on an erased track at red light and of reading on flat region at red and green lights for 4x MO disk. In the measurements, the disk was spinning at 1800rpm and the linear velocity was about 11.3m/s. The reflected power on each detector at red and green lights was about 20μW and 27μW, respectively.



Sample A   Sample B

Expanded magnetic domains observed with polarized light microscope

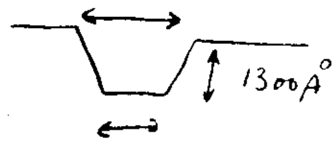


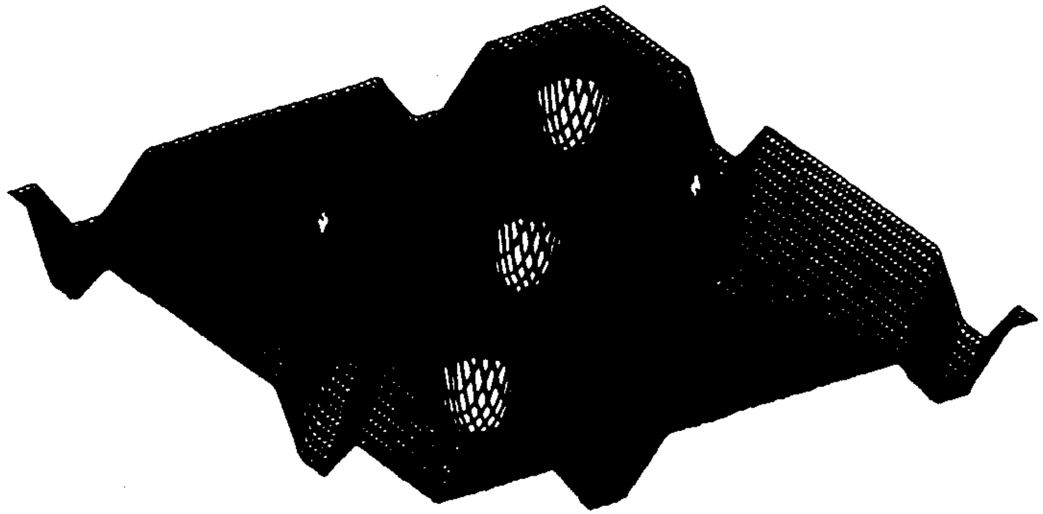
Minimum mark length = $\lambda / 4NA$

Preformat marks

0.545 X 0.3 Micron Pits, 1.1 Micron TP

Pit depth = 1500Å





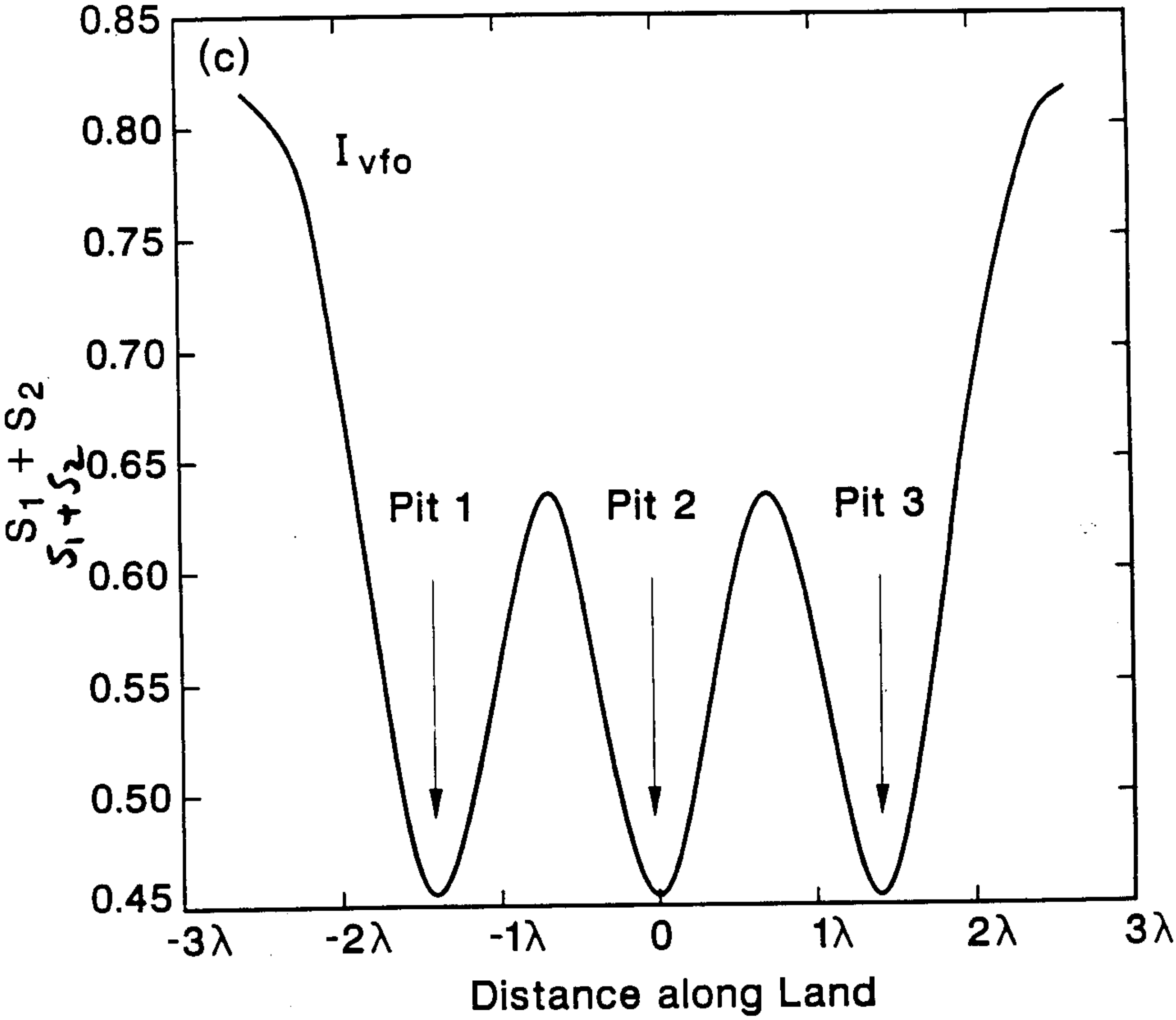
$N_A = 0.65$

$f_L = 1.23^{mm}$

May 8, '72

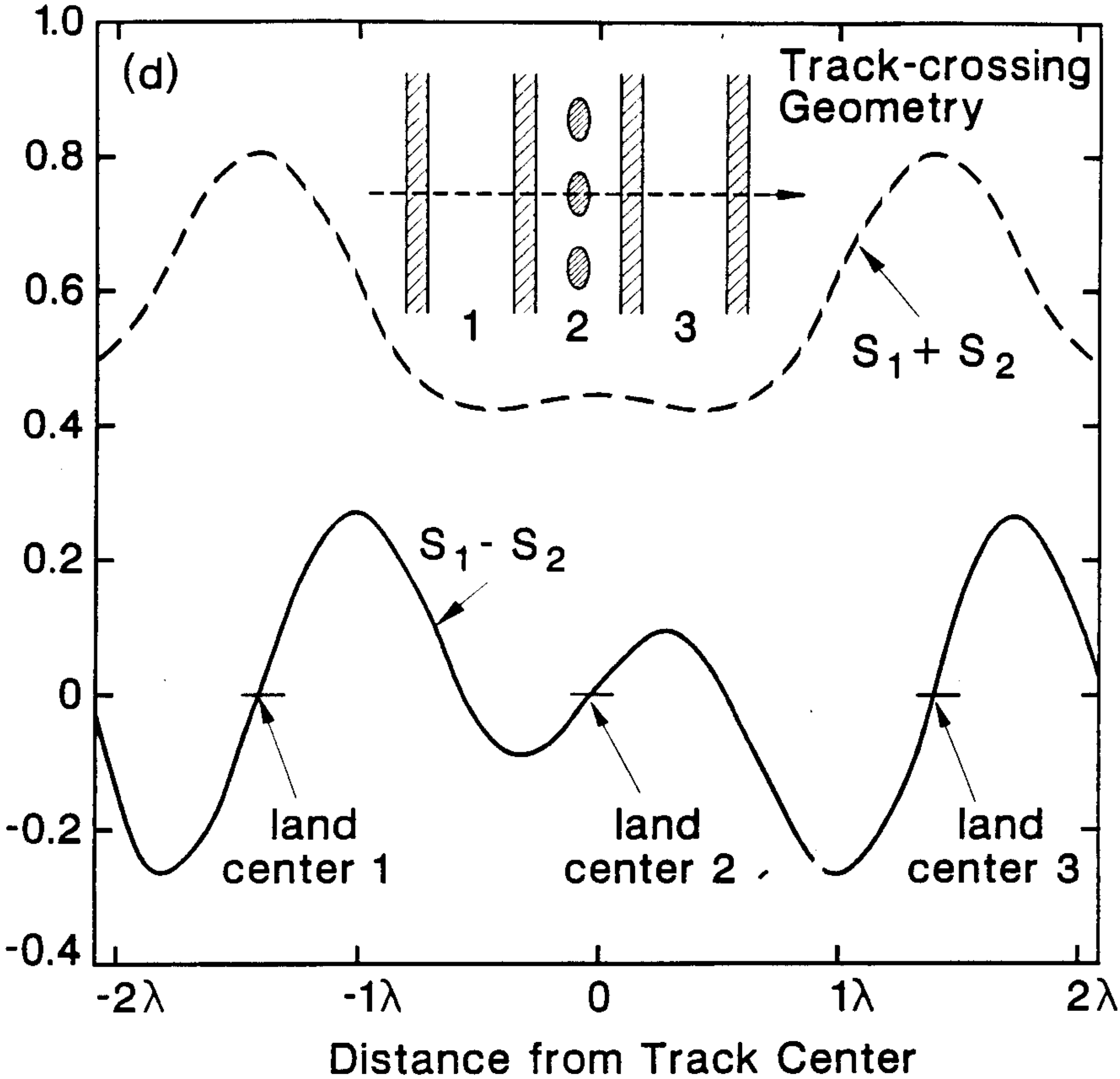
(1)

Read signal from preformat marks



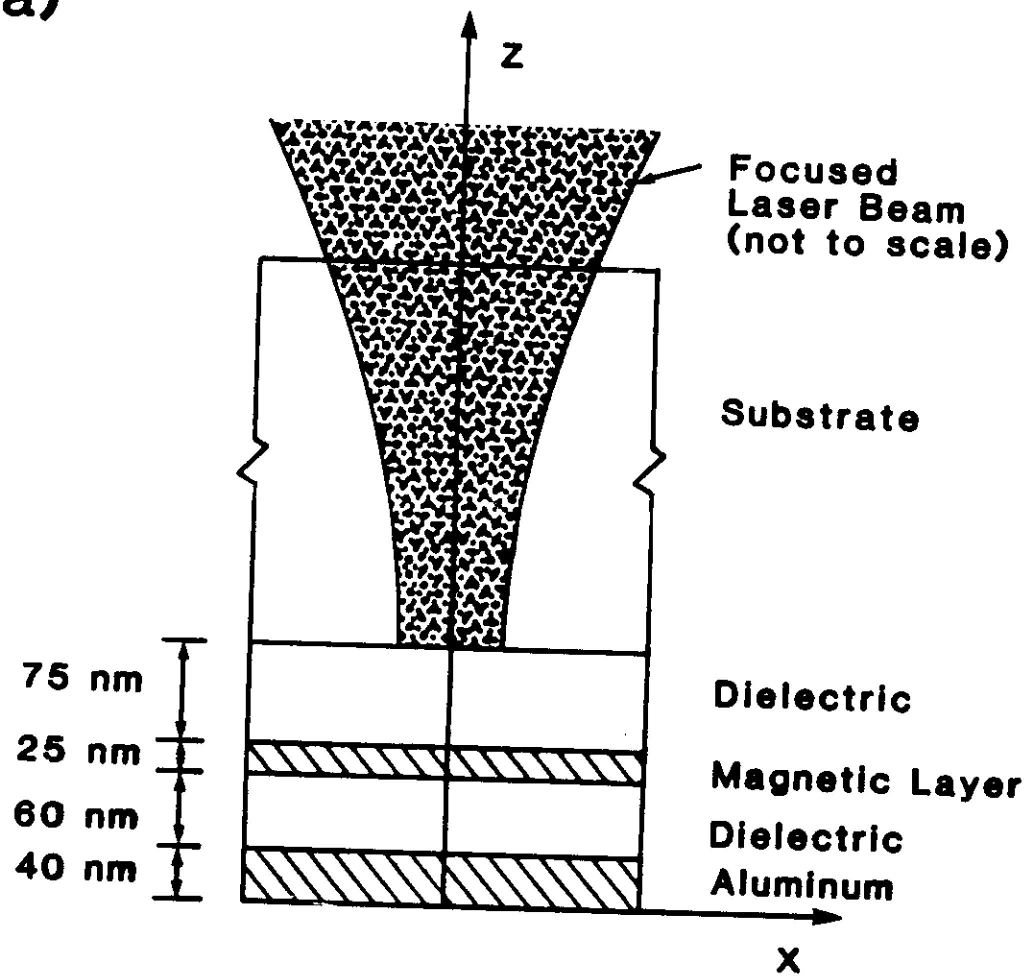
8-6c

Track-Crossing Signal in the preformat region



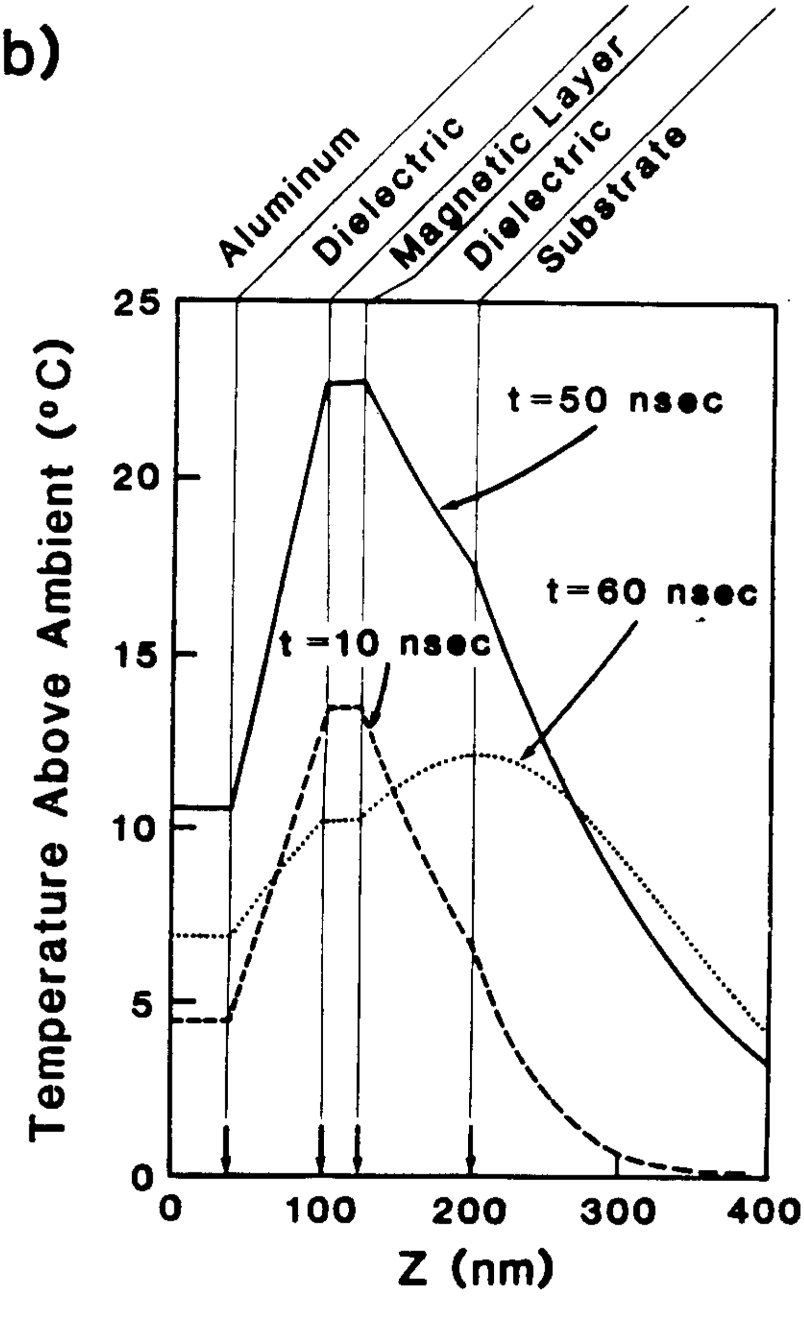
8-6d

a)

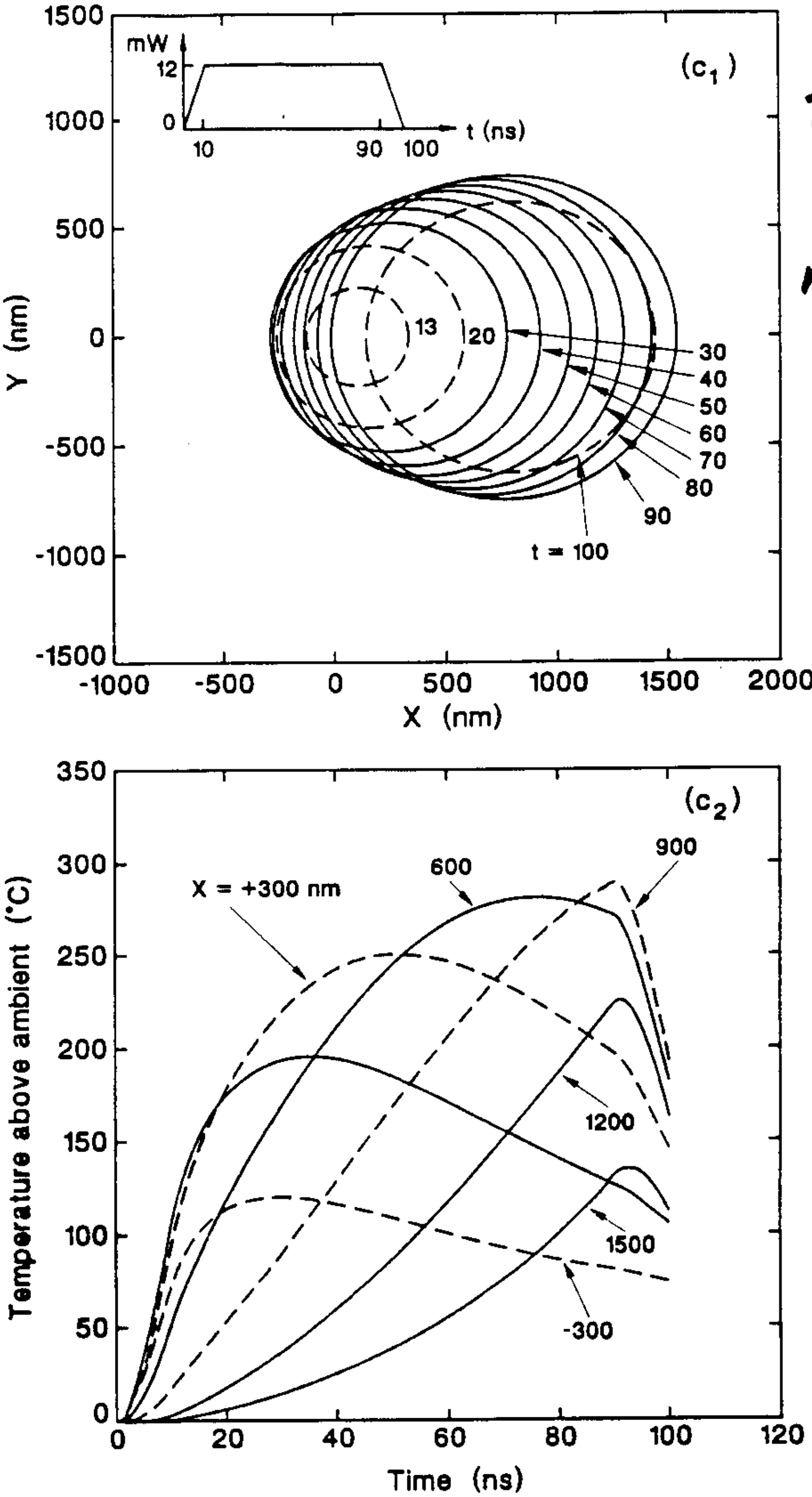


focusing (through the substrate)
on a quadrilayer MO disk

b)

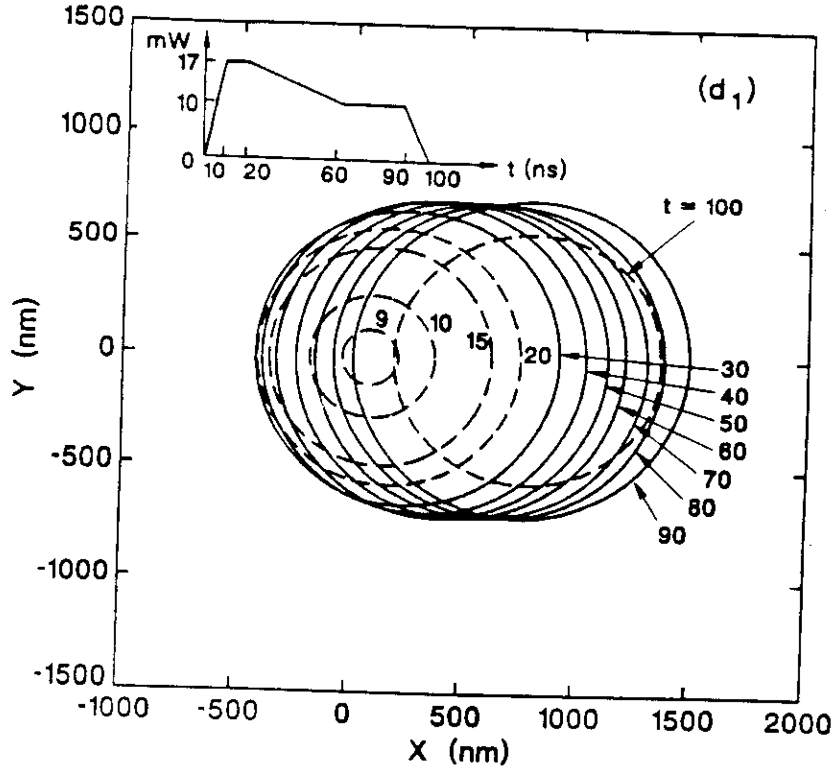


Depth profile of Temperature
(Quadrilayer)

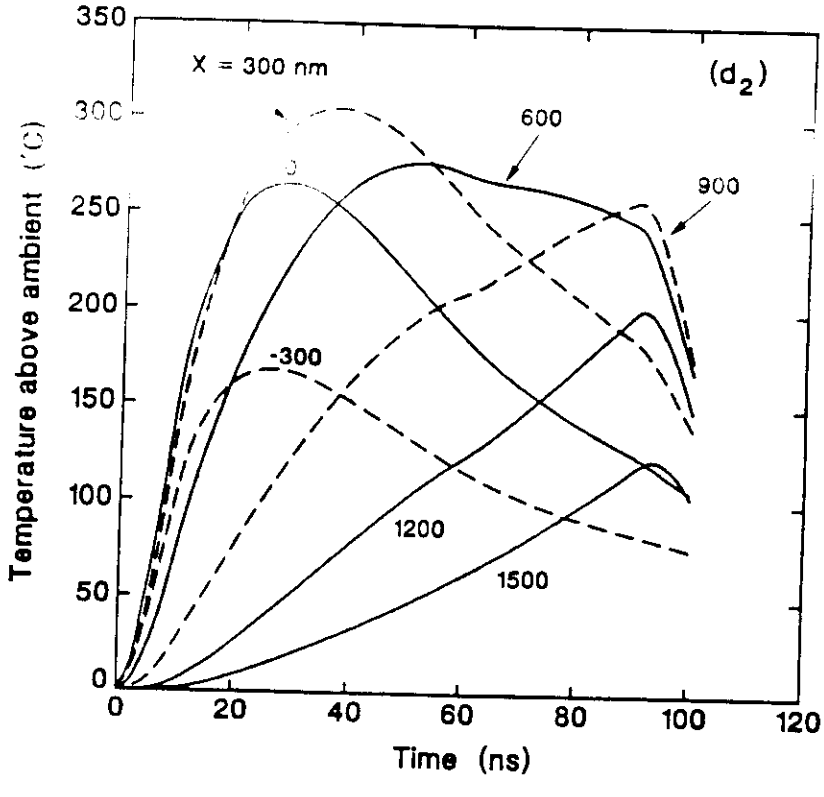


Tear-drop shaped mark

11-15c

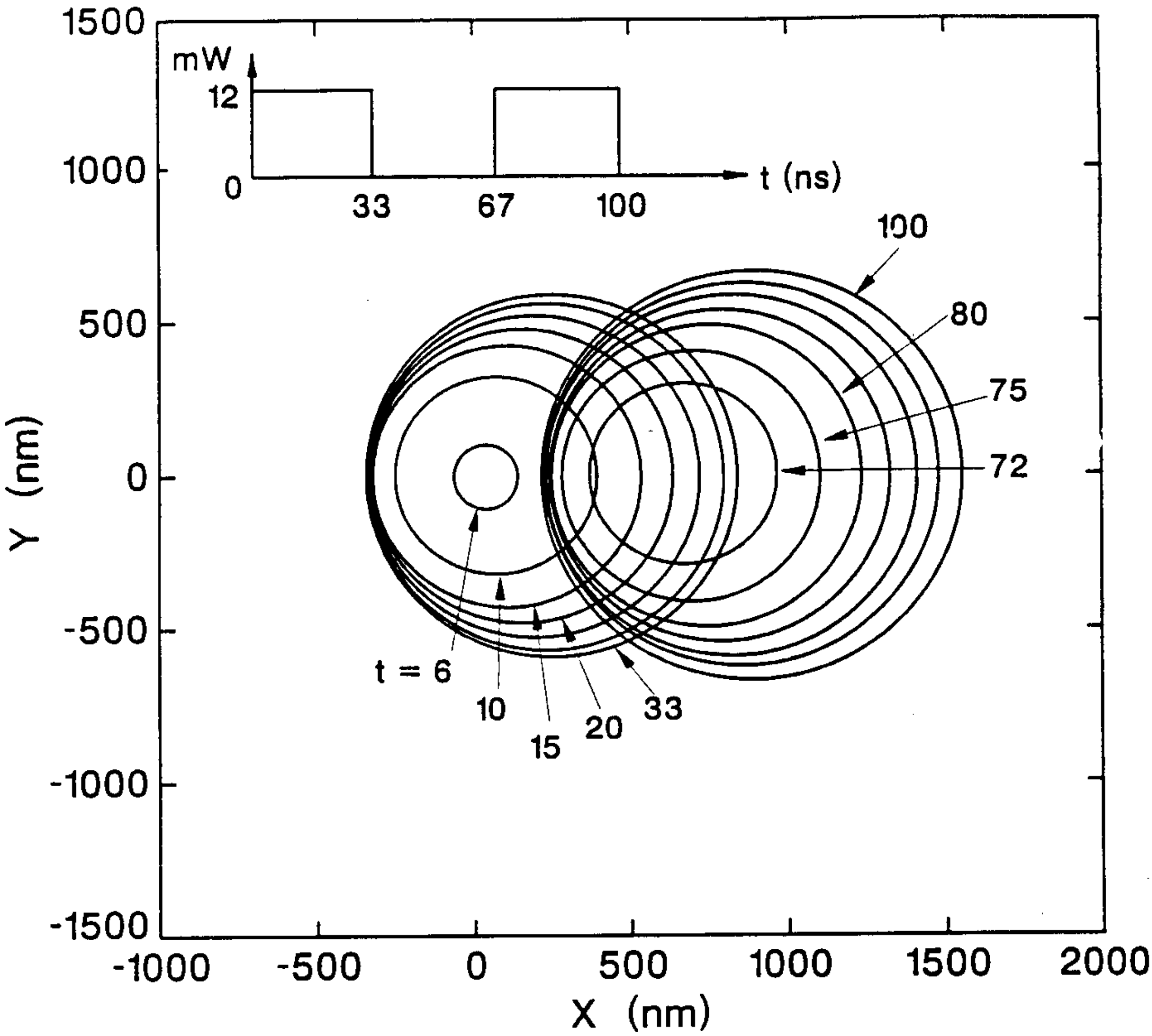


Symmetric mark

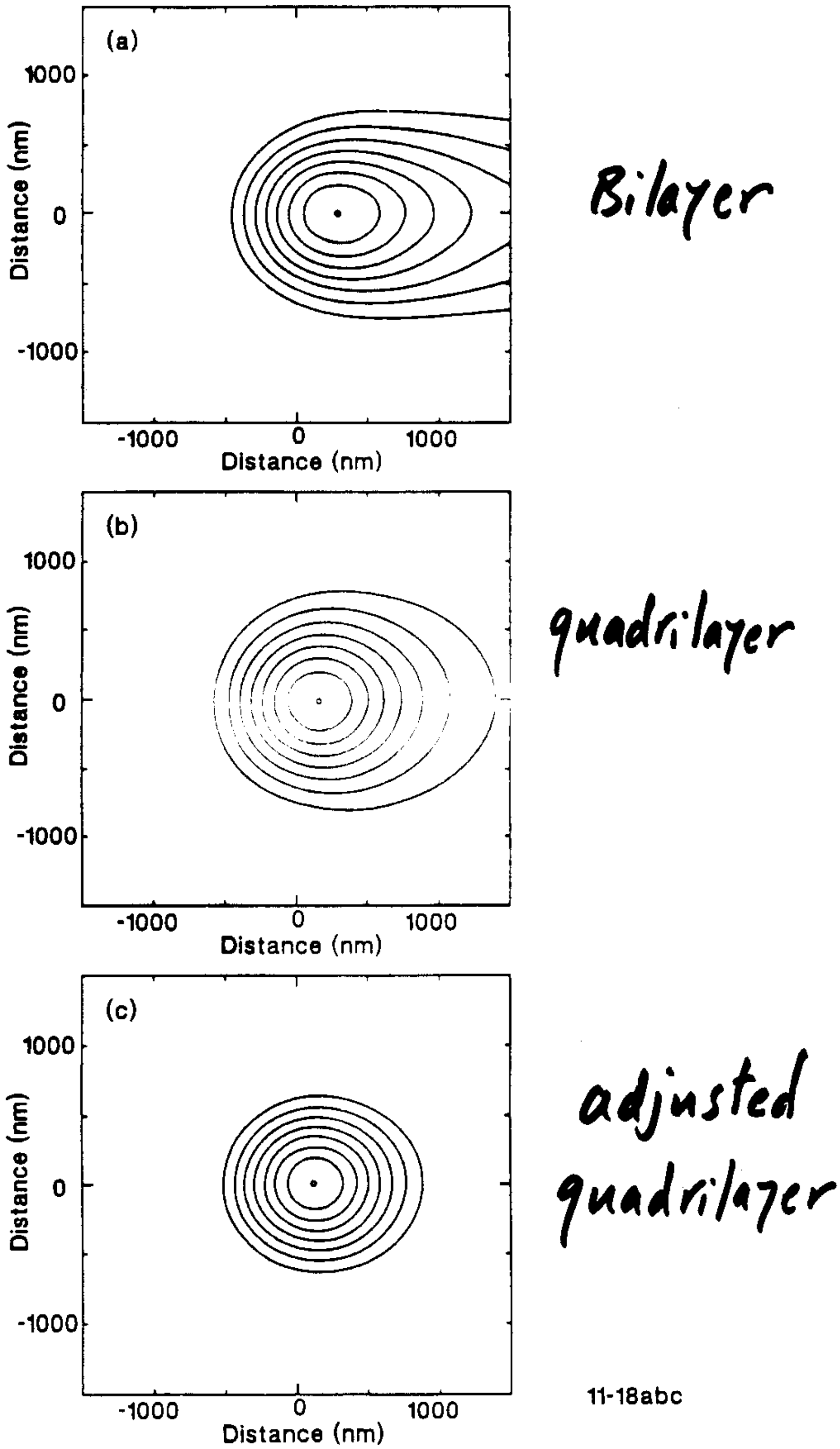


11-15d

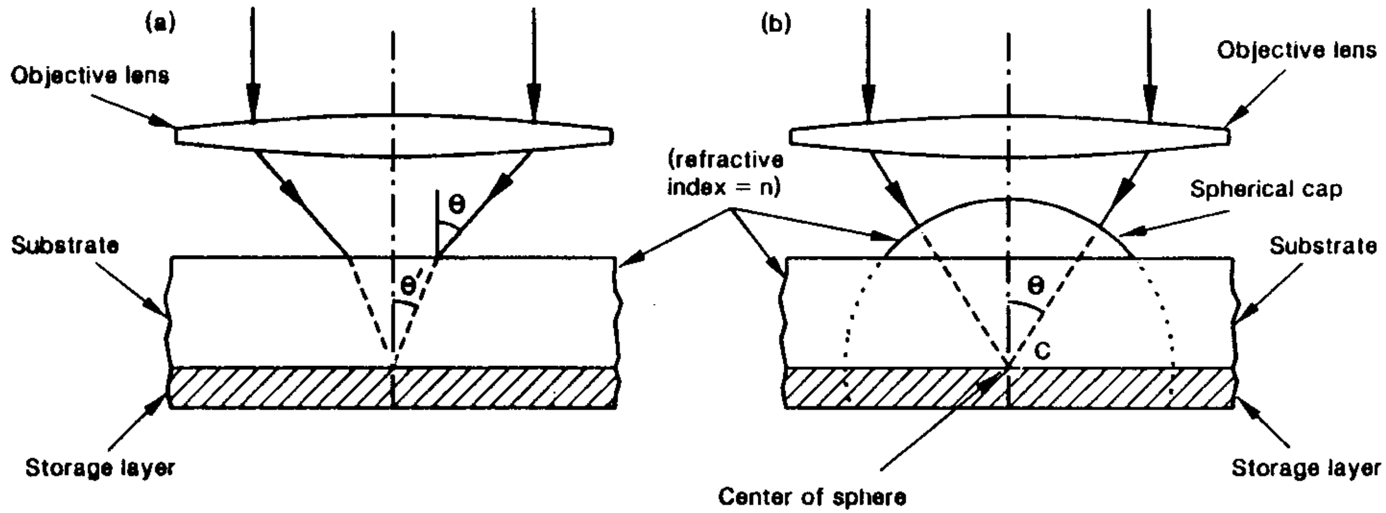
Thermal Cross-talk



11-16


(a)
1000
0
-1000
Distance (nm)
-1000
0
1000
Distance (nm)
Bilayer
(b)
quadrilayer
(c)
adjusted quadrilayer
11-18abc

Solid Immersion Lens



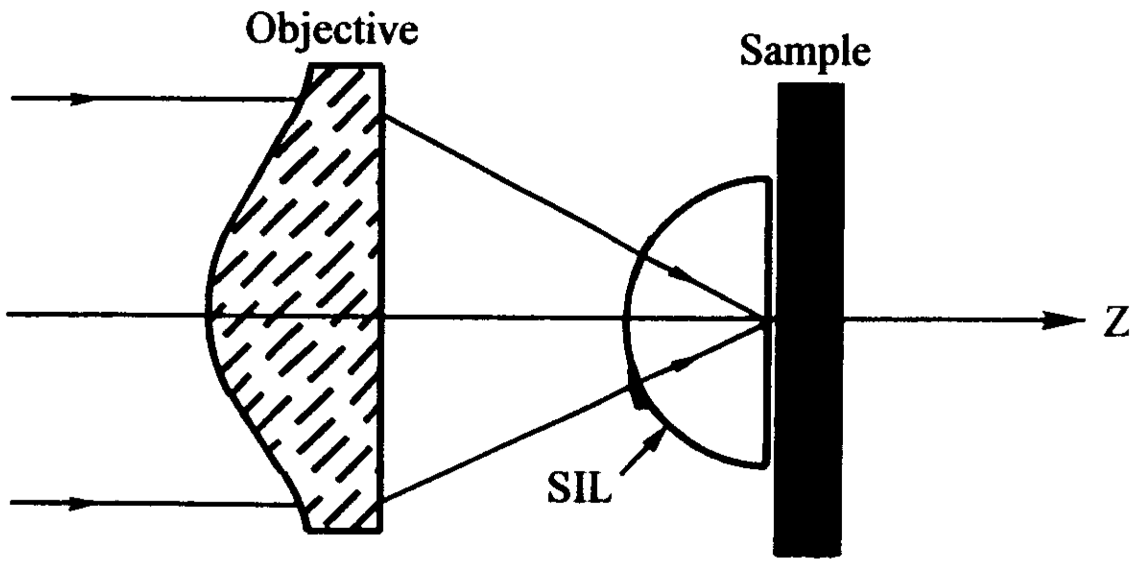
- Kino et al, Stanford Univ.
- Compatible with MSR, PRML, short-wavelength but <u>not</u> Land-groove

$$\sin\theta = n \sin\theta'$$

~~2-10~~ 2-9

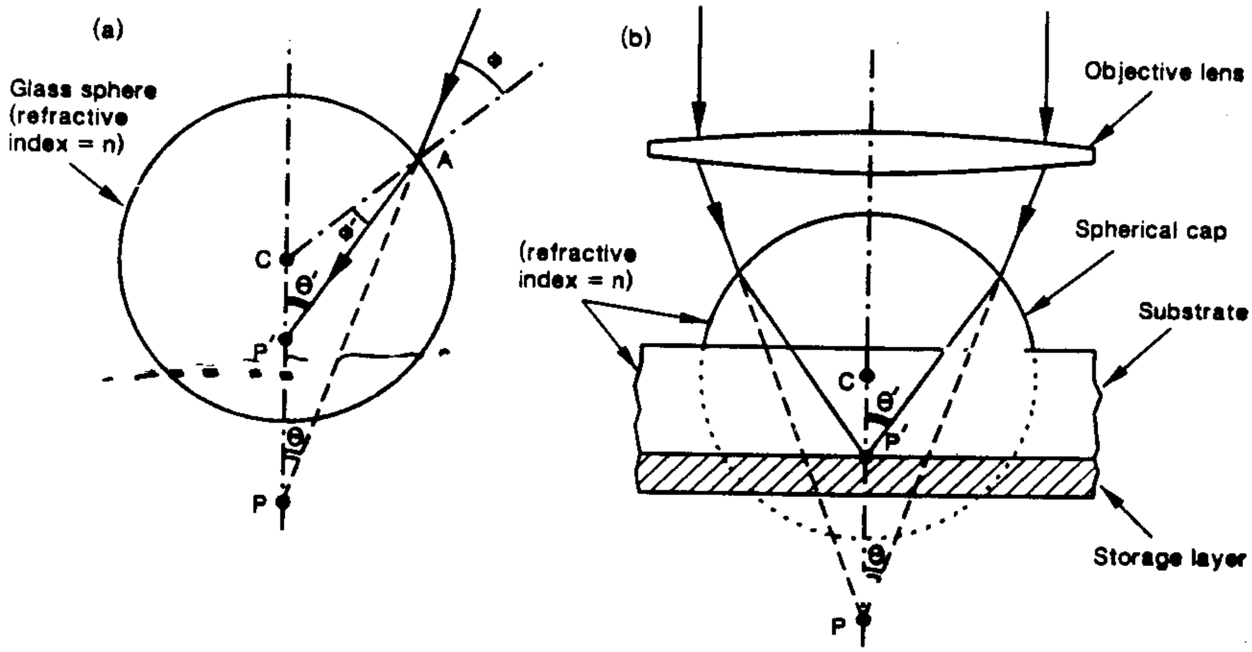
$$d \sim \frac{\lambda}{\sin\theta'} = \frac{\lambda/n}{\sin\theta/n} = \frac{\lambda}{\sin\theta}$$

$$d \approx \frac{\lambda/n}{N.A.}$$



Mansuripur, Figure 11

Solid Immersion Lens



2-10

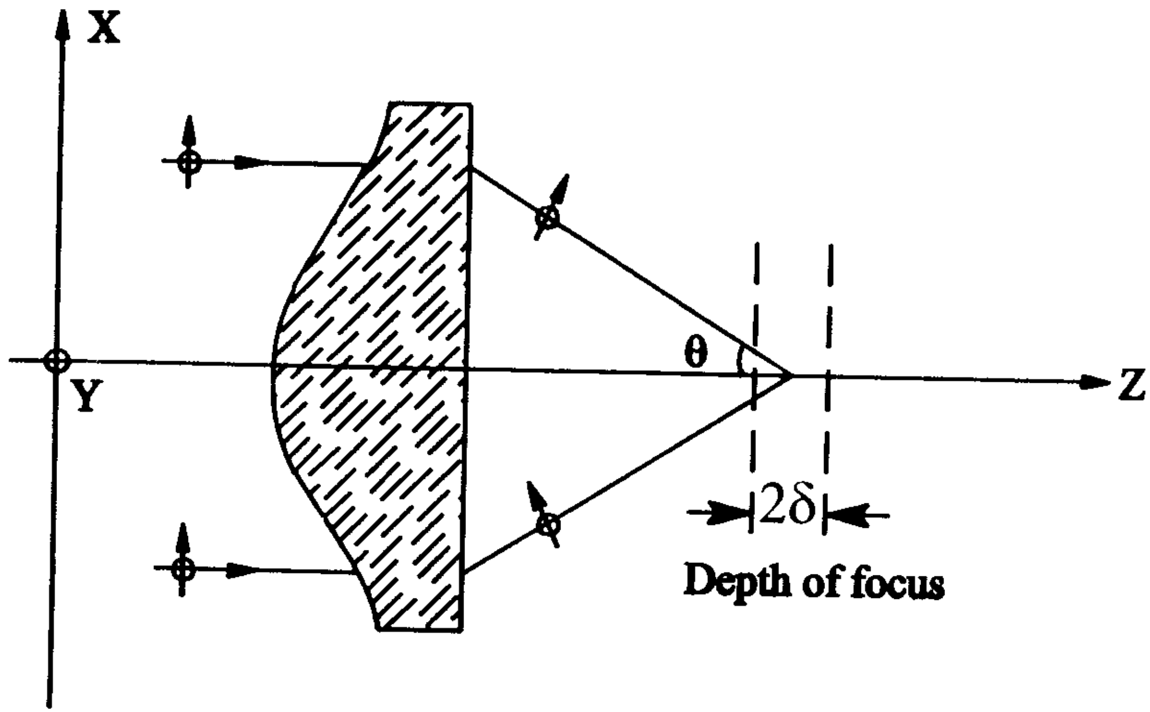
$$CP = nr$$

$$CP' = r/n$$

$$\sin\theta' = n \sin\theta$$

$$\lambda = \lambda_0/n$$

$$d^{-1} \sim \frac{\sin\theta'}{\lambda} = n^2 \frac{\sin\theta}{\lambda_0}$$

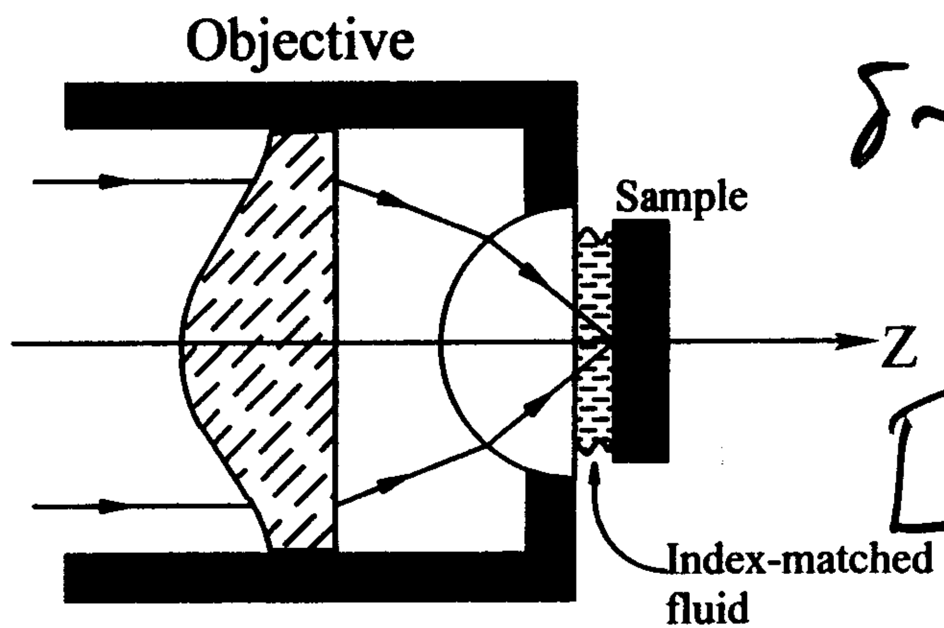


Mansuripur, Figure 2

$NA = n \sin\theta$

$FWHM \simeq \frac{0.6\lambda_0}{NA}$

$D.o.f \cong \frac{\lambda_0 n}{NA^2}$

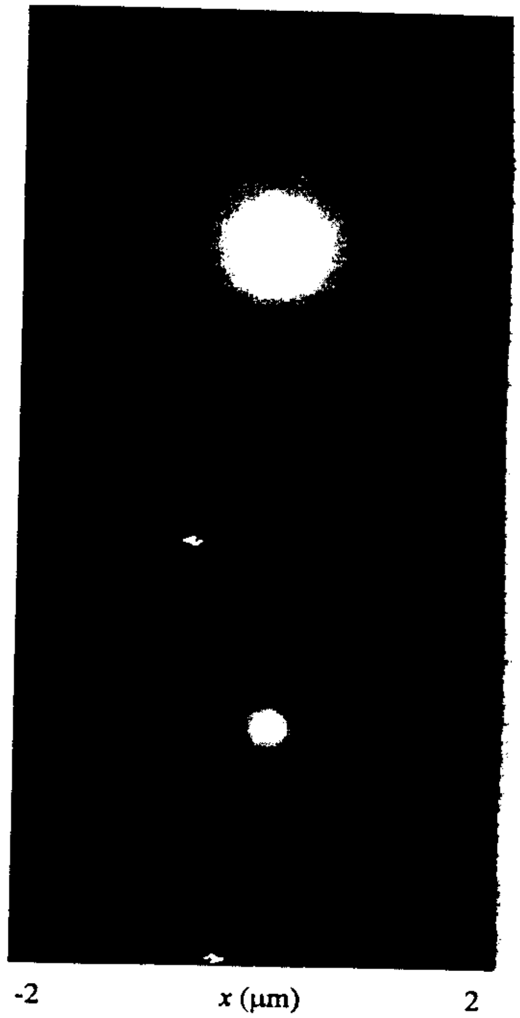


$\delta \sim \frac{\lambda_0}{NA^2}$

$NA = n \sin\theta$

7/A

Mansuripur, Figure 5



$\Delta z = 1\ \mu m$

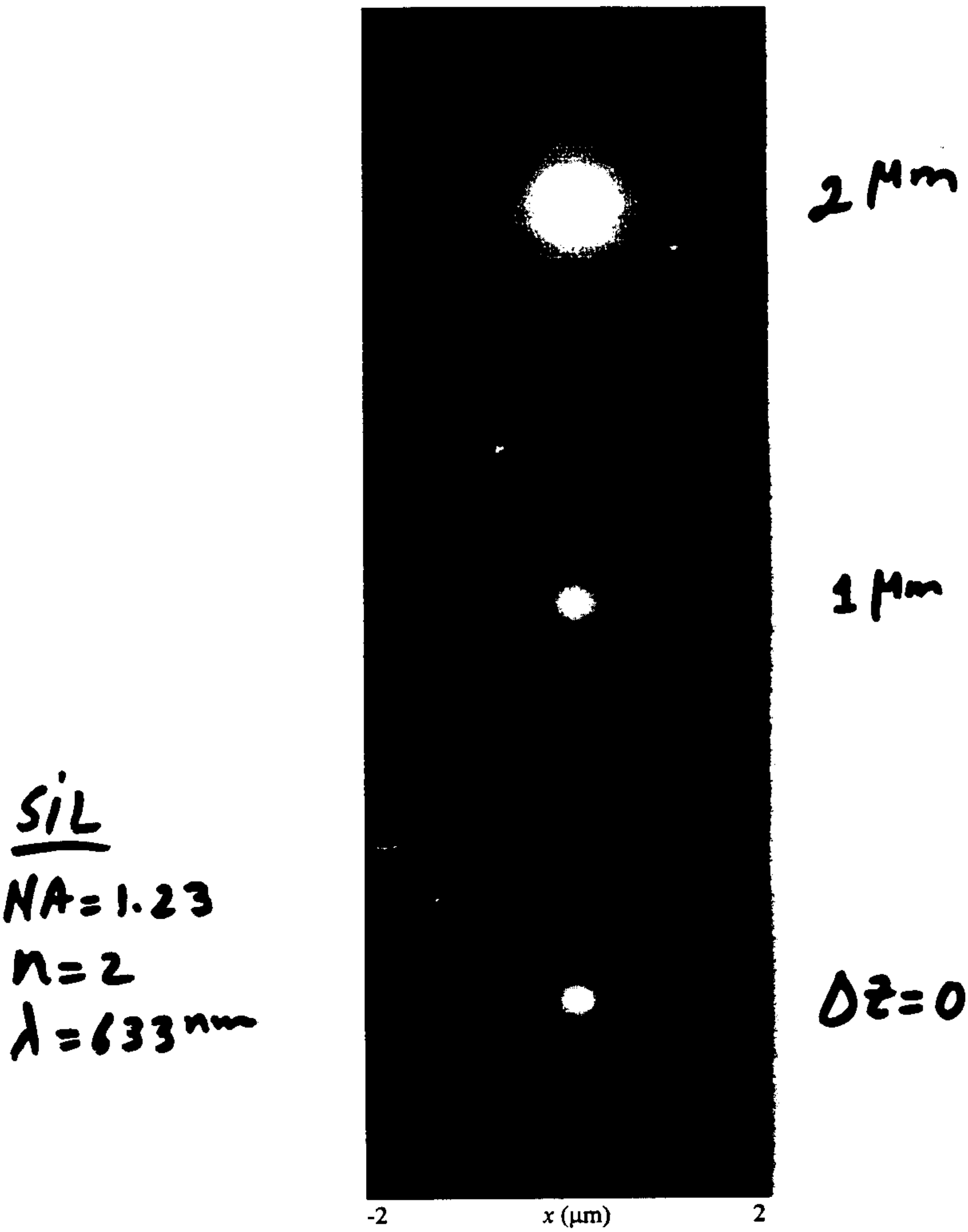
$\Delta z = 0$

Oil immersion

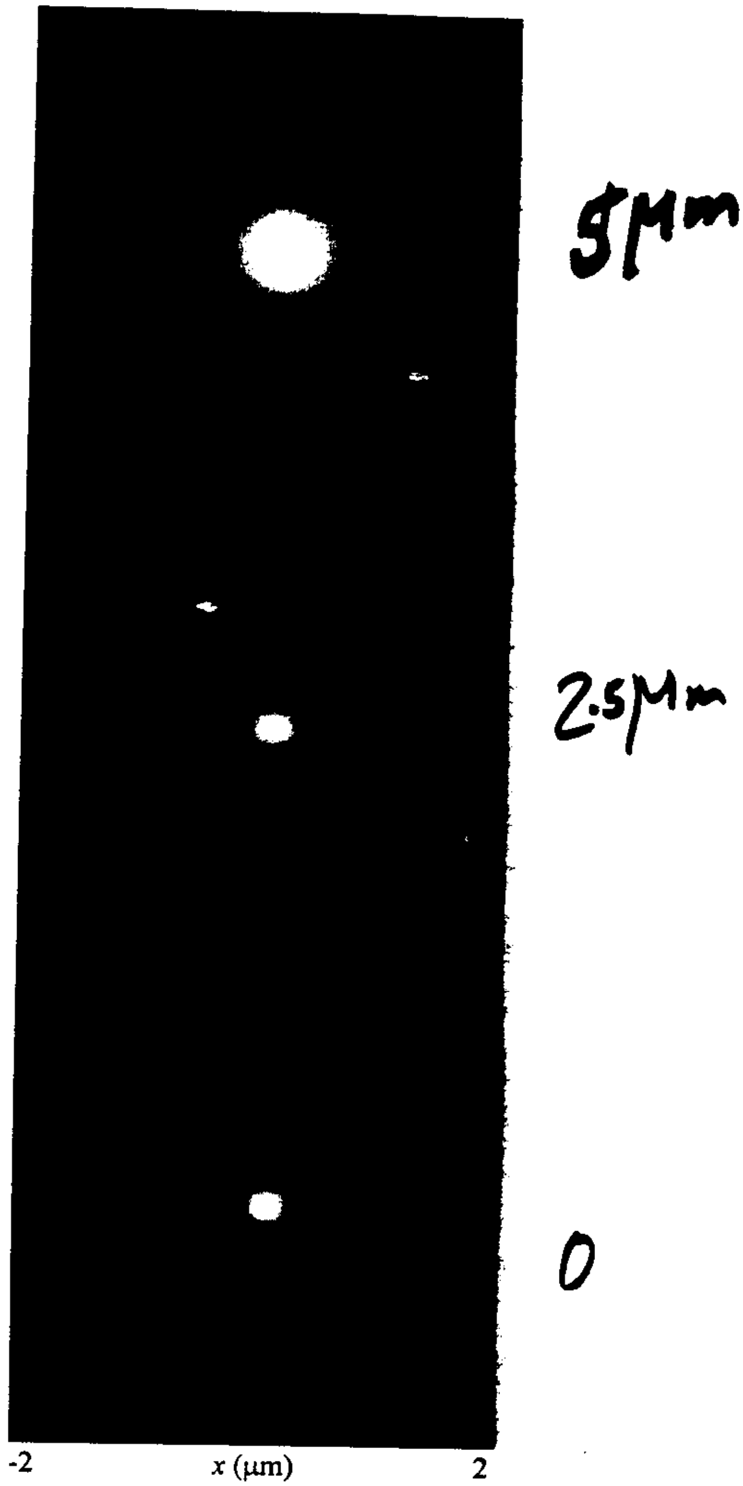
$\begin{cases} n = 2 \\ \sin\theta = 0.615 \end{cases} \Rightarrow NA = 1.23$

$\lambda = 633$ nm

Mansuripur, Figure 6



Mansuripur, Figure12

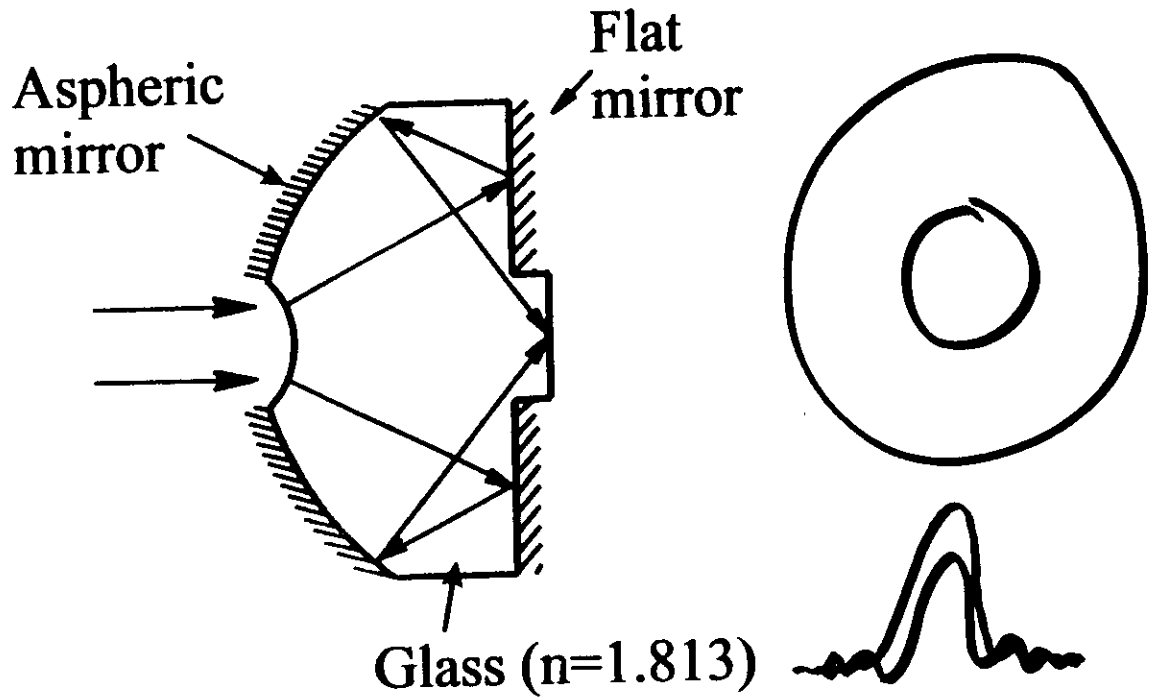


Super SIL

NA = 1.6
n = 2
λ = 633nm

Mansuripur, Figure 17

C.W. Lee et al
Samsung Electronics



Mansuripur, Figure 18

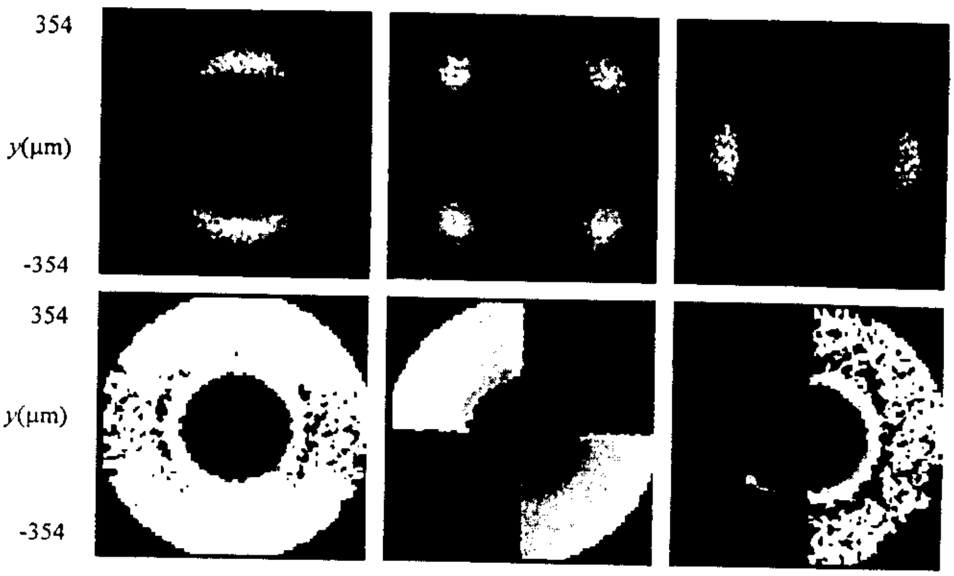


Figure 1. Distributions of intensity (top row) and phase (bottom row) at the Destination plane. From left to right: X-, Y-, and Z-components of polarization.

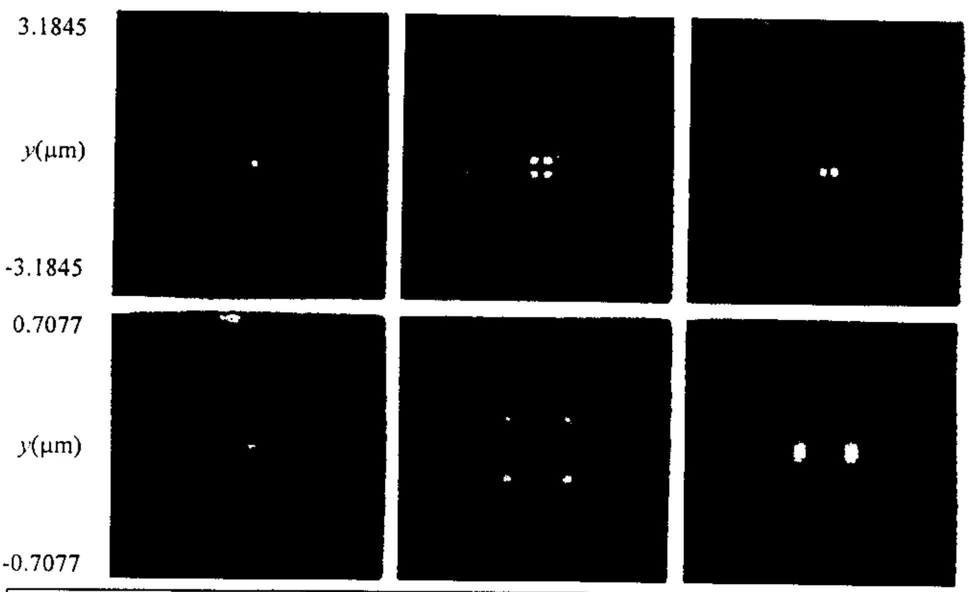
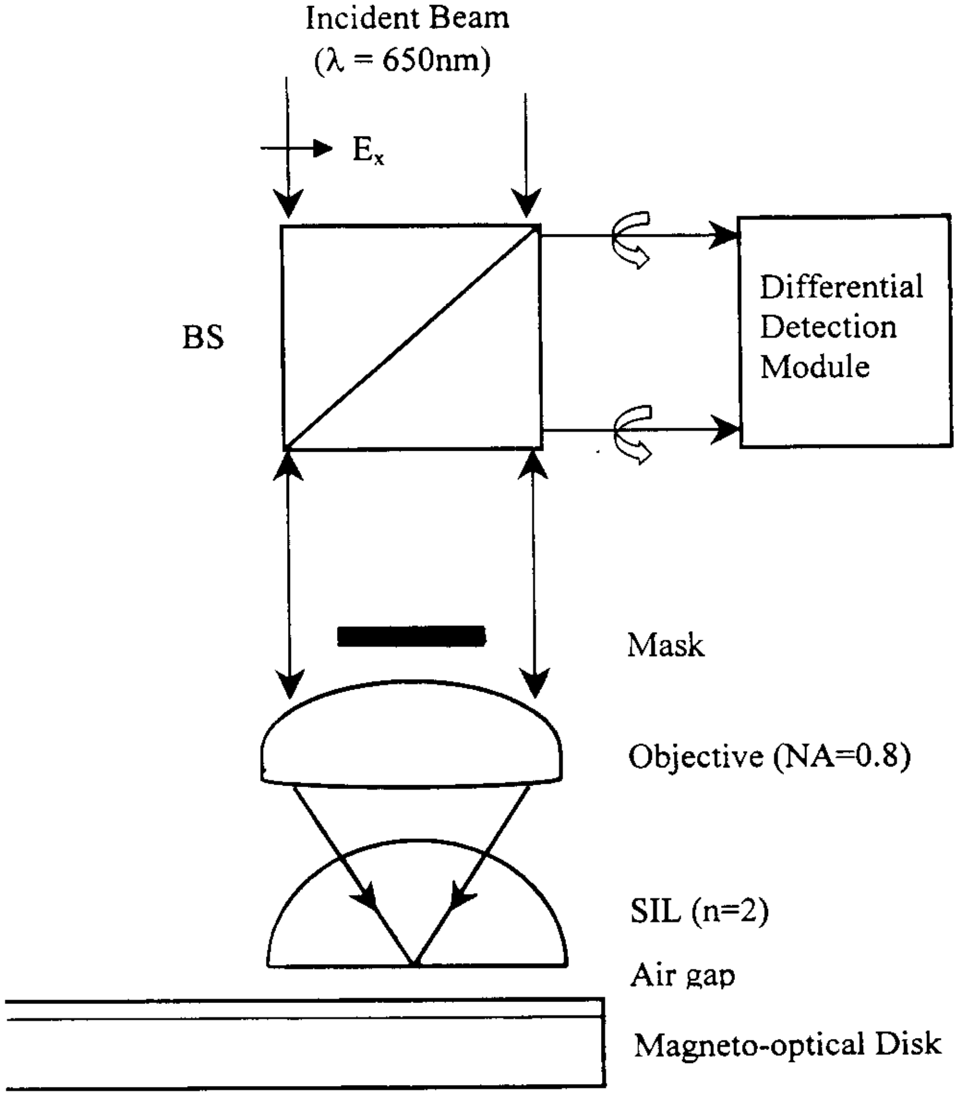
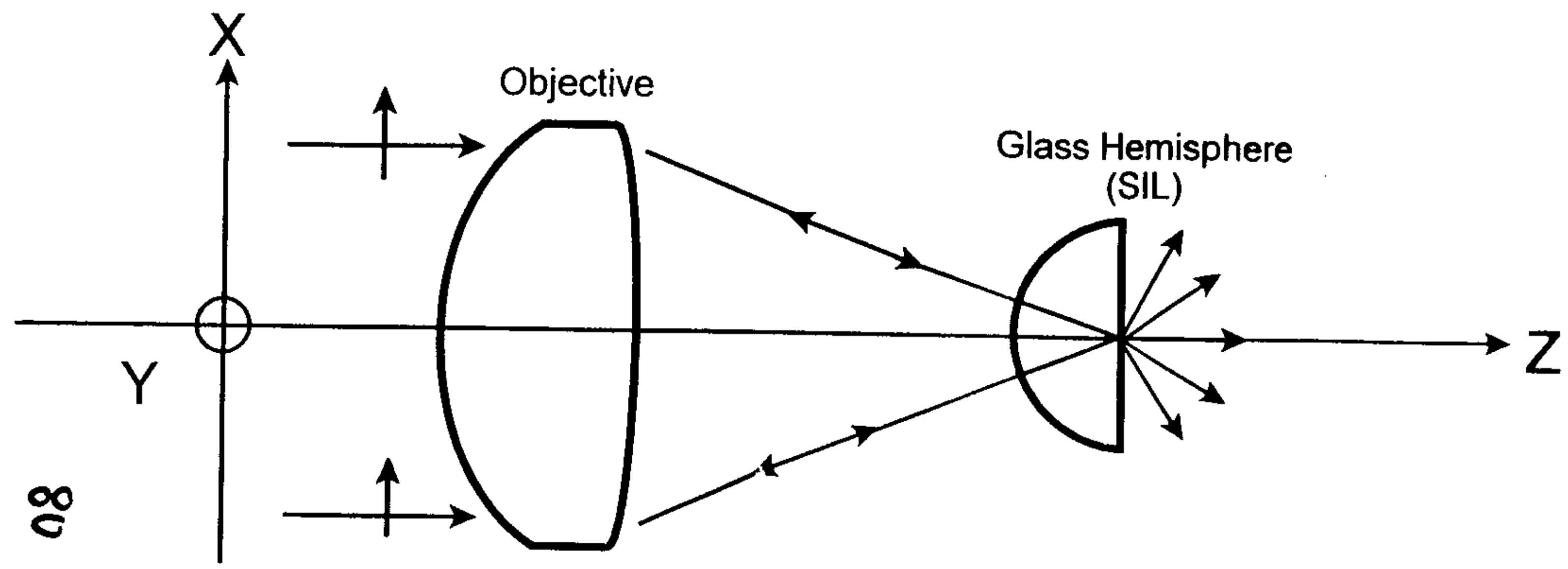
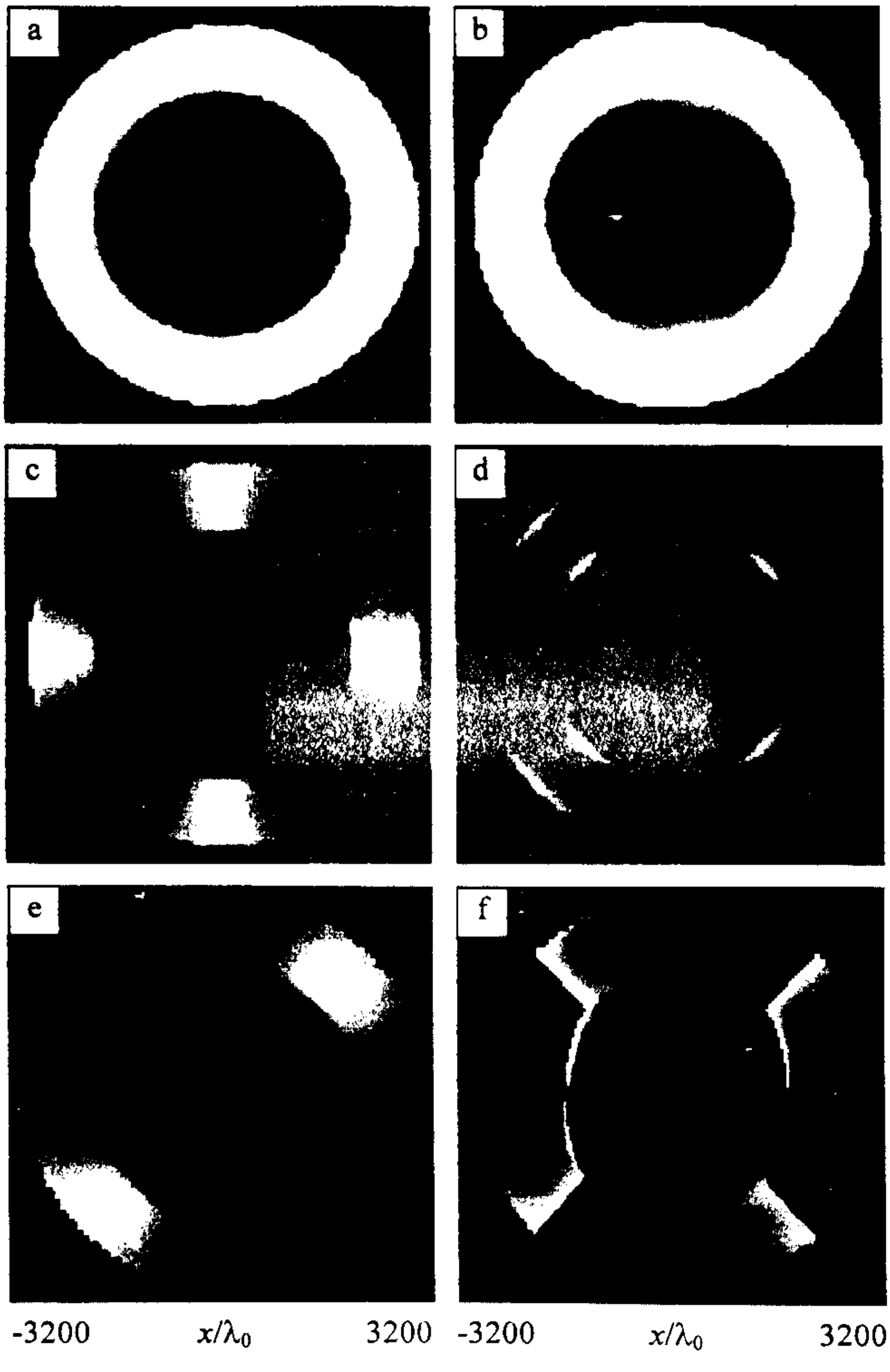


Figure 2. Plots of logarithmic (top row) and ordinary (bottom row) intensity distribution at the focal plane of the SIM, just inside the plateau. From left to right: X-, Y-, and Z-components of polarization.

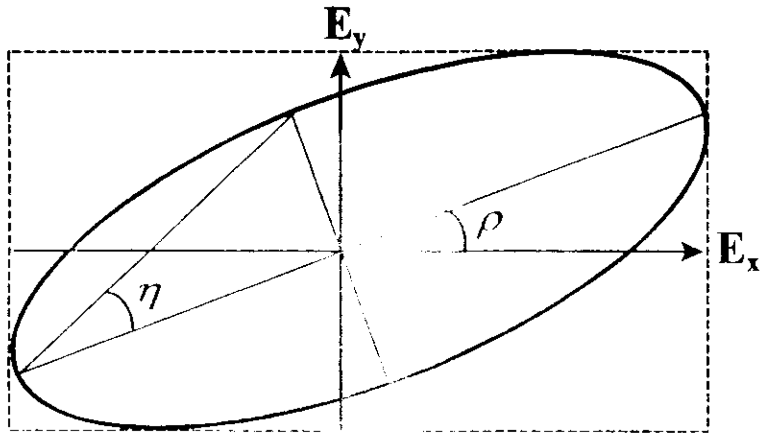

Incident Beam
(λ = 650nm)
E_x
BS
Differential
Detection
Module
Mask
Objective (NA=0.8)
SIL (n=2)
Air gap
Magneto-optical Disk



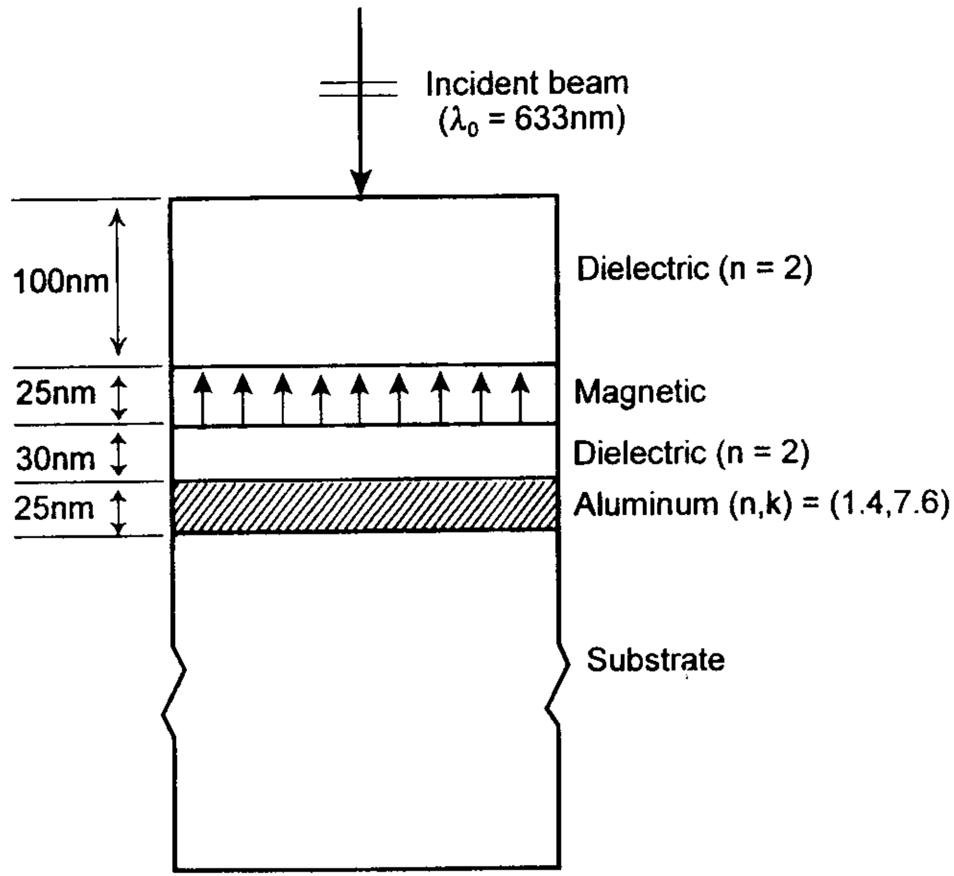
Mansuripur, Figure 3



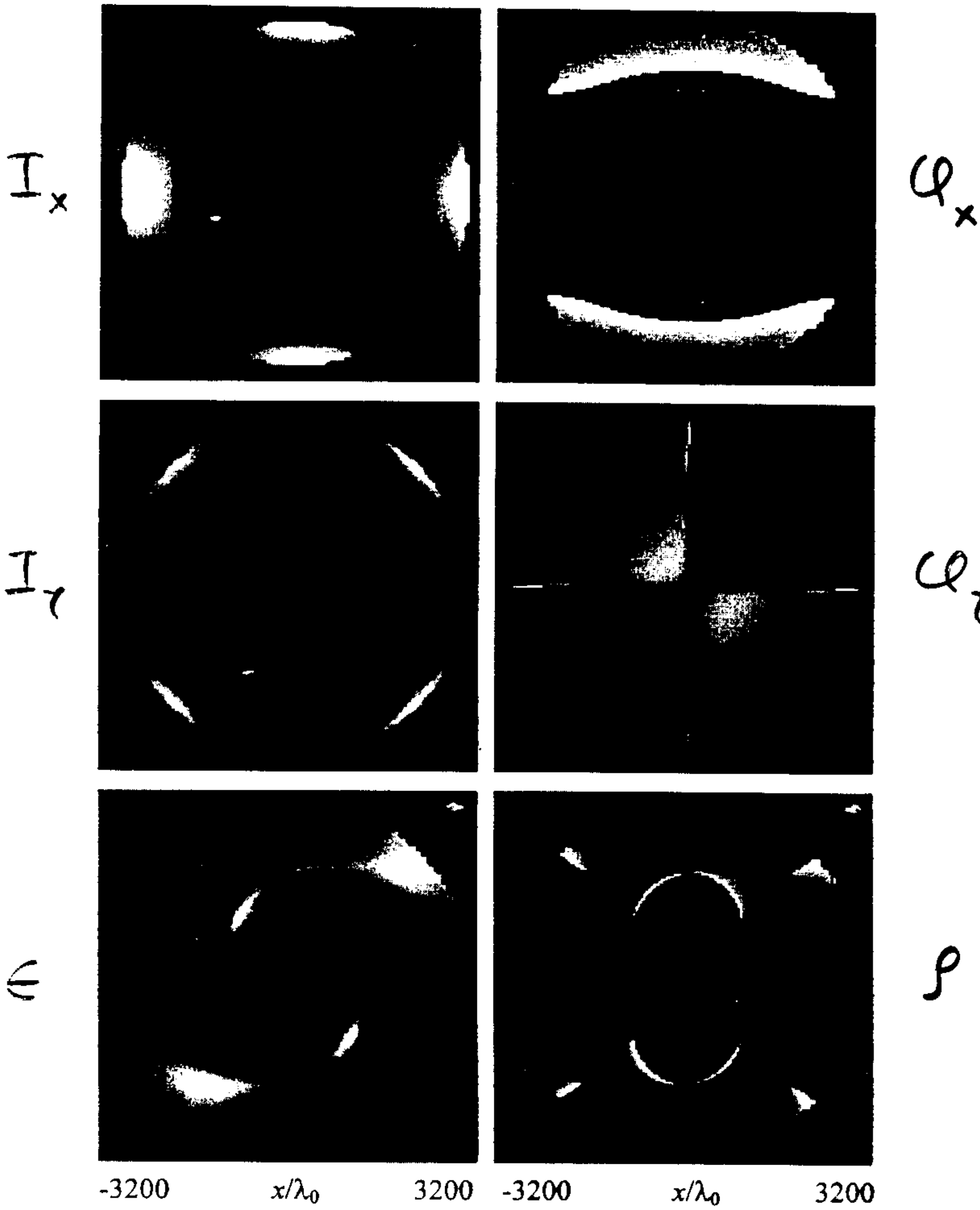
Mansuripur, Figure 4



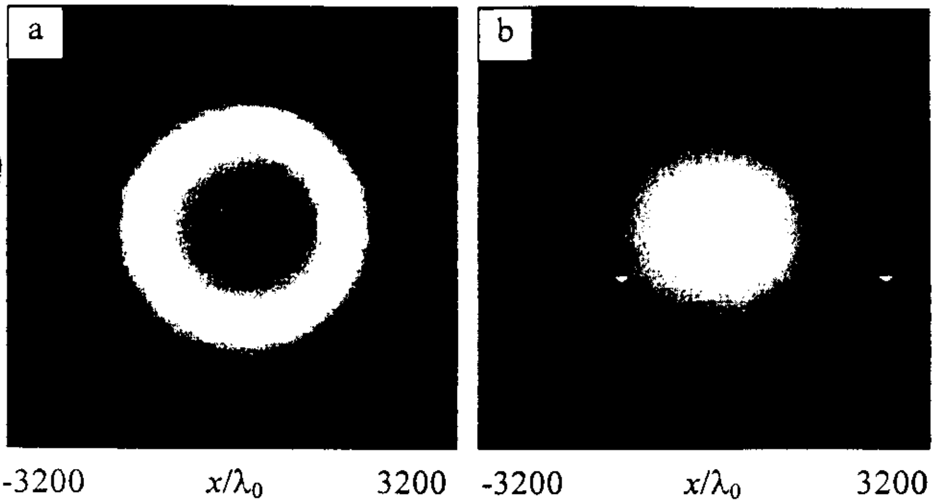
Mansuripur, Figure 2



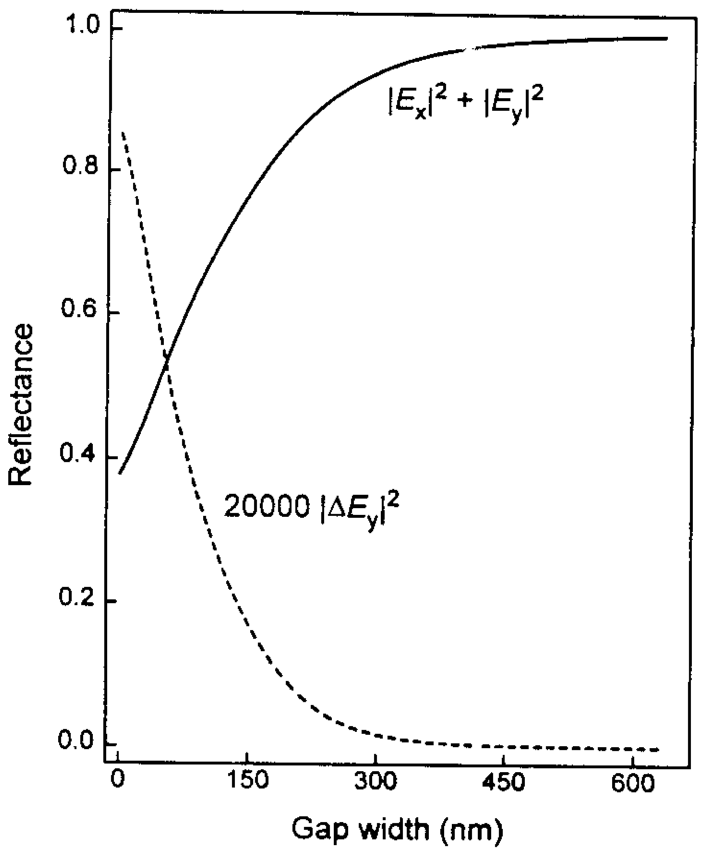
Mansuripur, Figure 10



Mansuripur, Figure 8

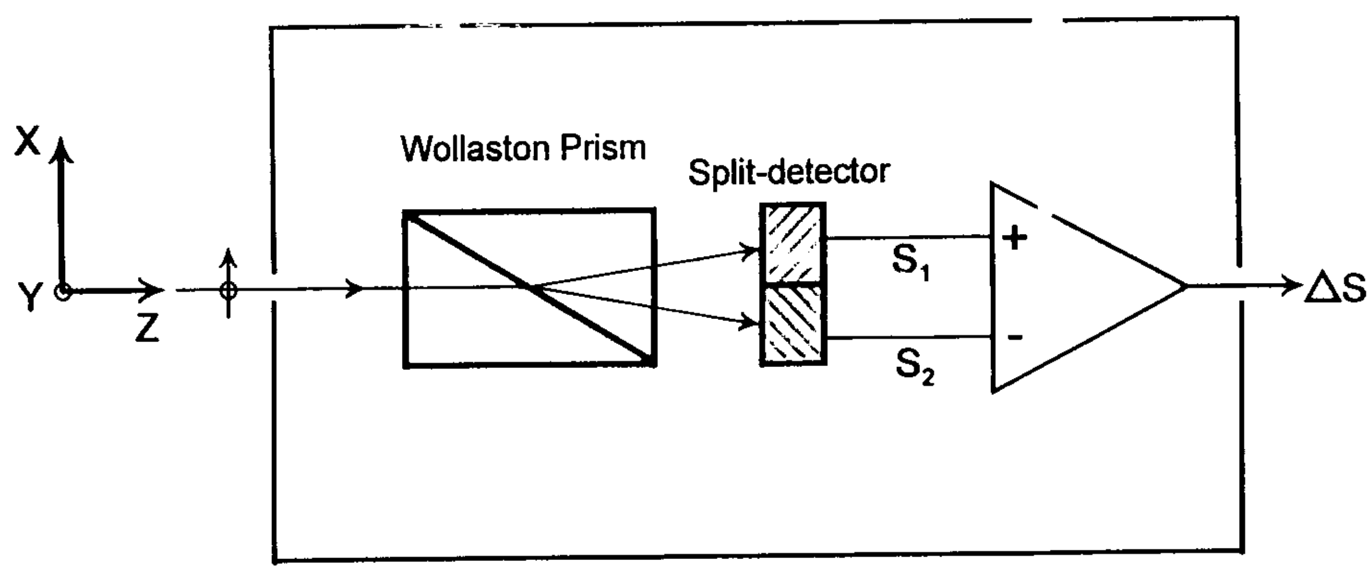


Mansuripur, Figure 12

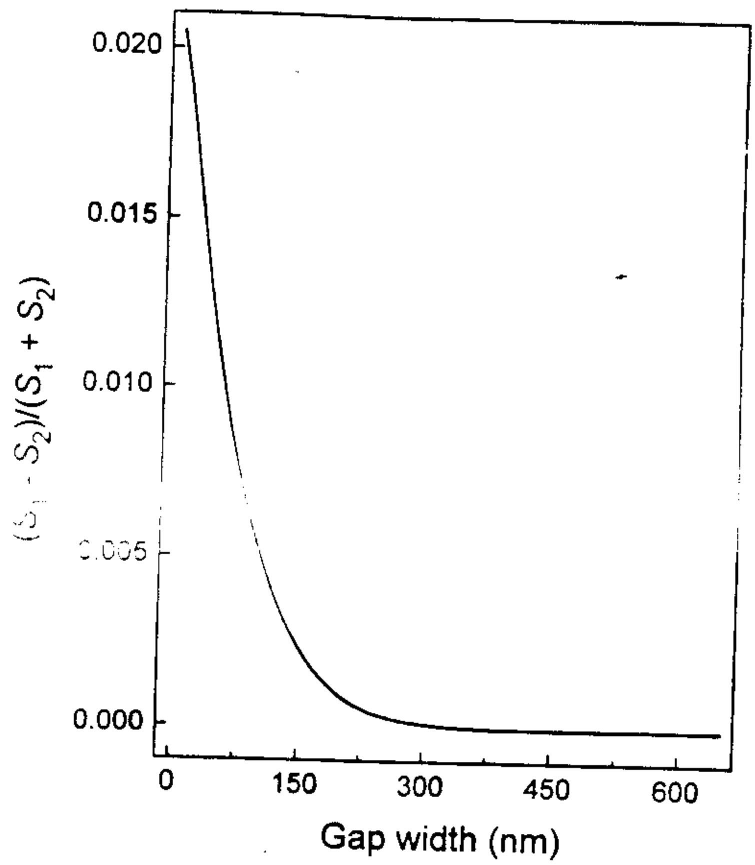


- $\lambda = 633$ nm
- Central region blocked

Mansuripur, Figure 13

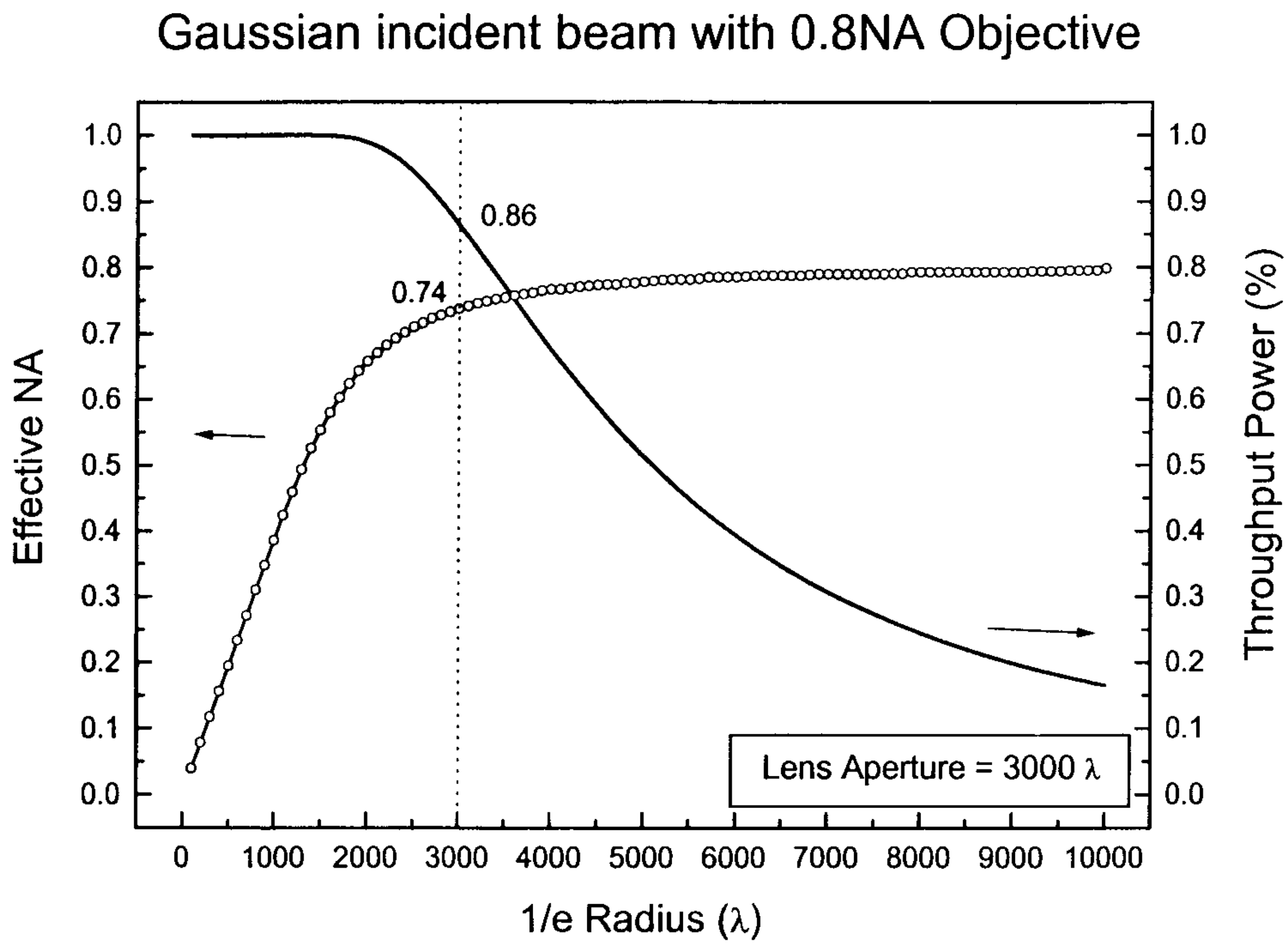


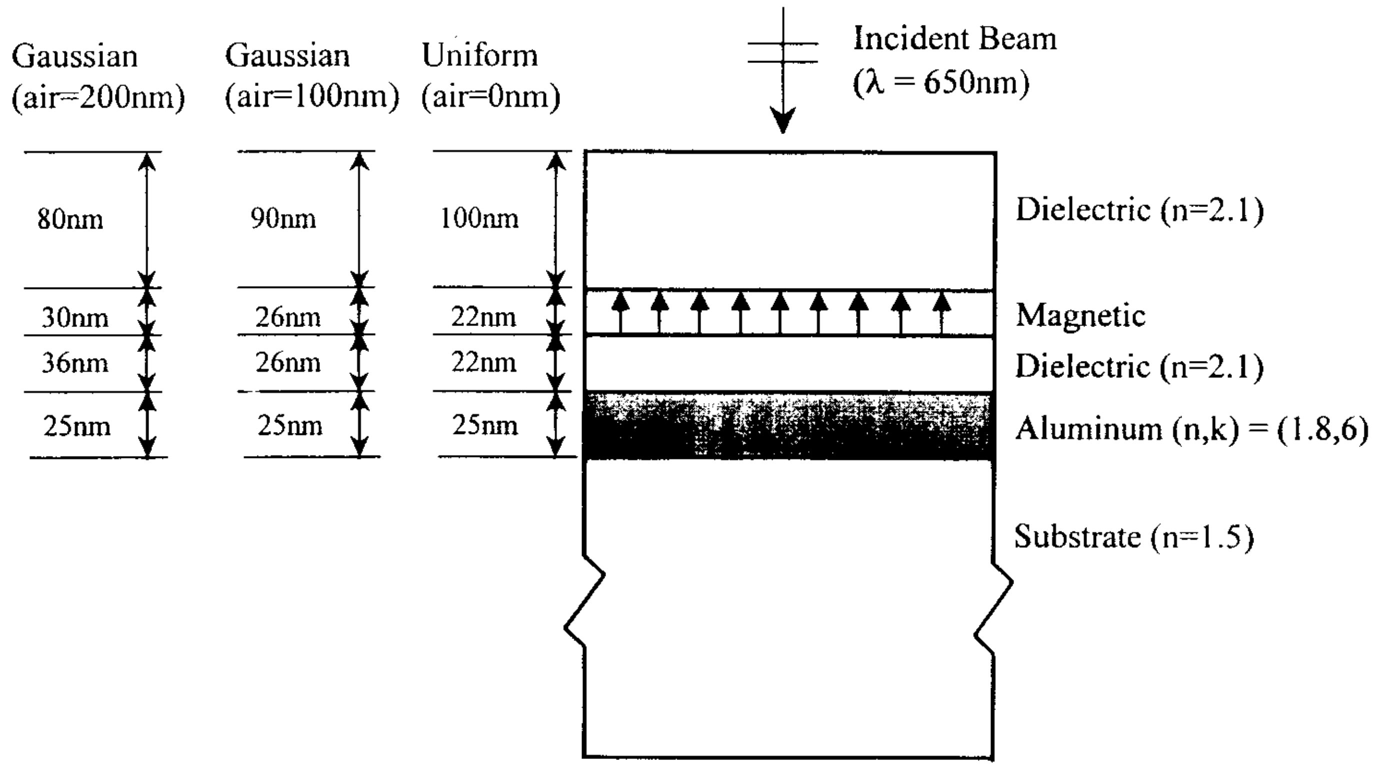
Mansur r, Figure 11



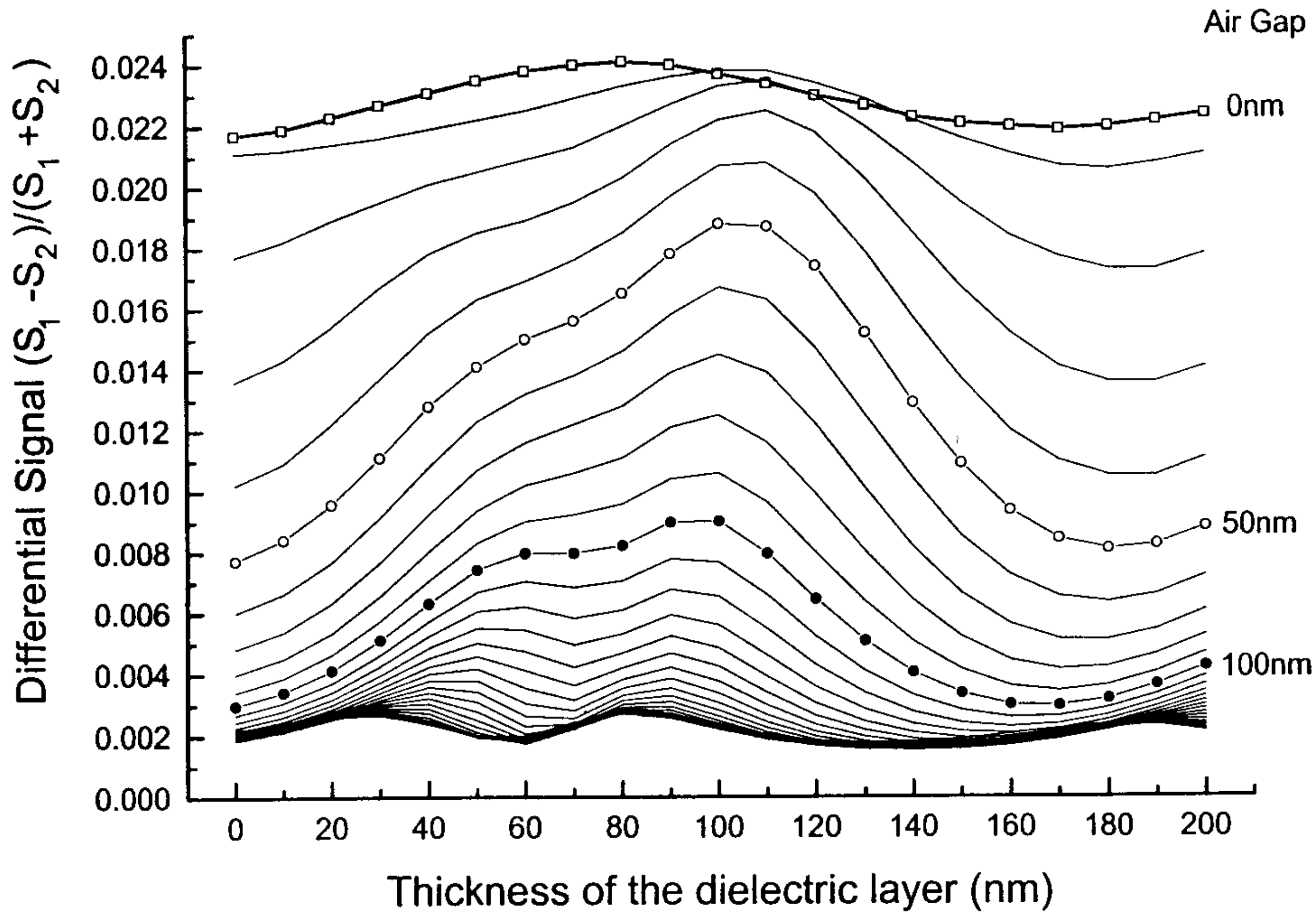
- $\lambda = 633$ nm
- Central region blocked

Mansuripur, Figure 15

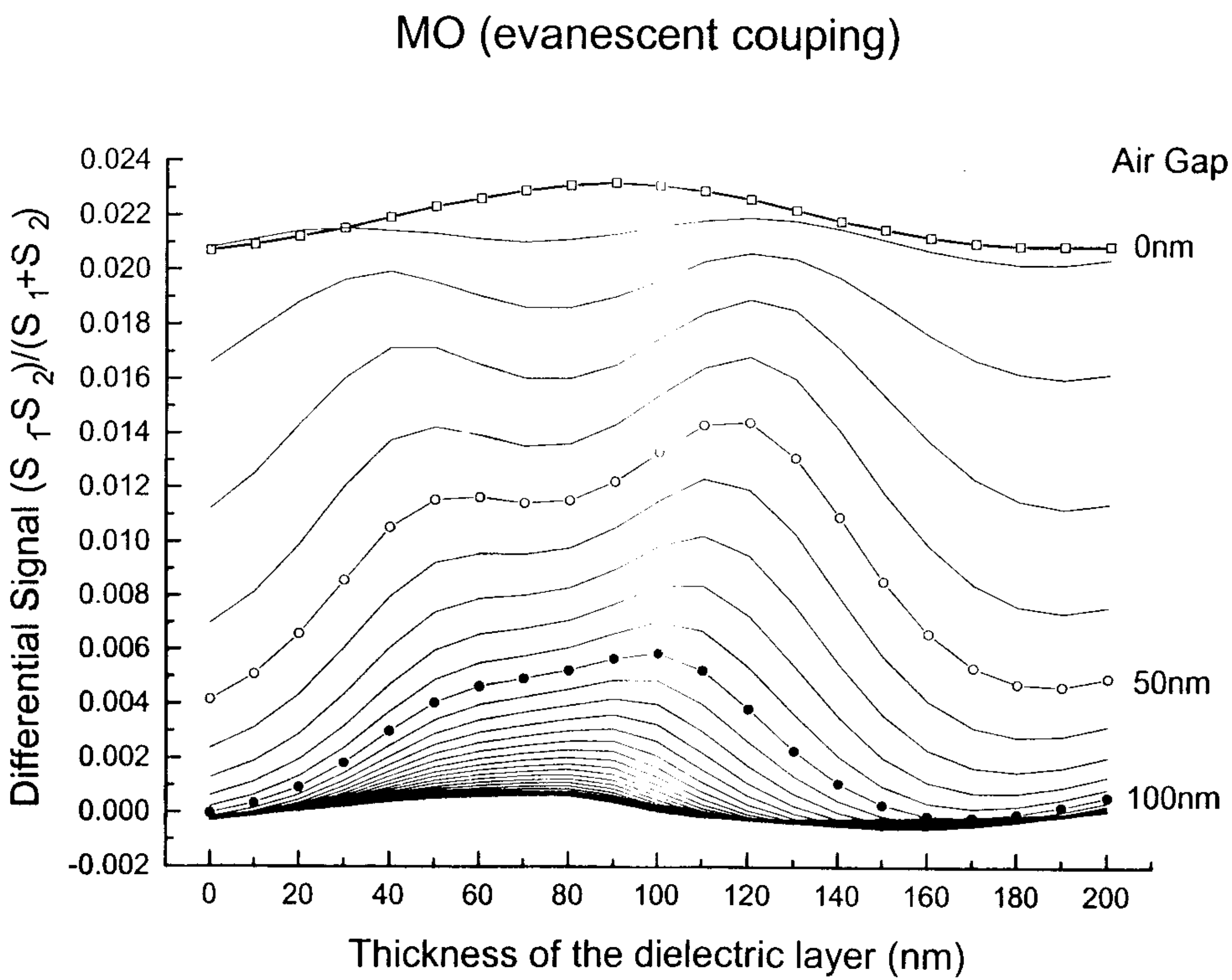

Gaussian incident beam with 0.8NA Objective
Effective NA
Throughput Power (%)
1/e Radius (λ)
0.86
0.74
Lens Aperture = 3000 λ
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000


Gaussian
(air=200nm)
Gaussian
(air=100nm)
Uniform
(air=0nm)
80nm
30nm
36nm
25nm
90nm
26nm
26nm
25nm
100nm
22nm
22nm
25nm
Incident Beam
(λ = 650nm)
Dielectric (n=2.1)
Magnetic
Dielectric (n=2.1)
Aluminum (n,k) = (1.8,6)
Substrate (n=1.5)

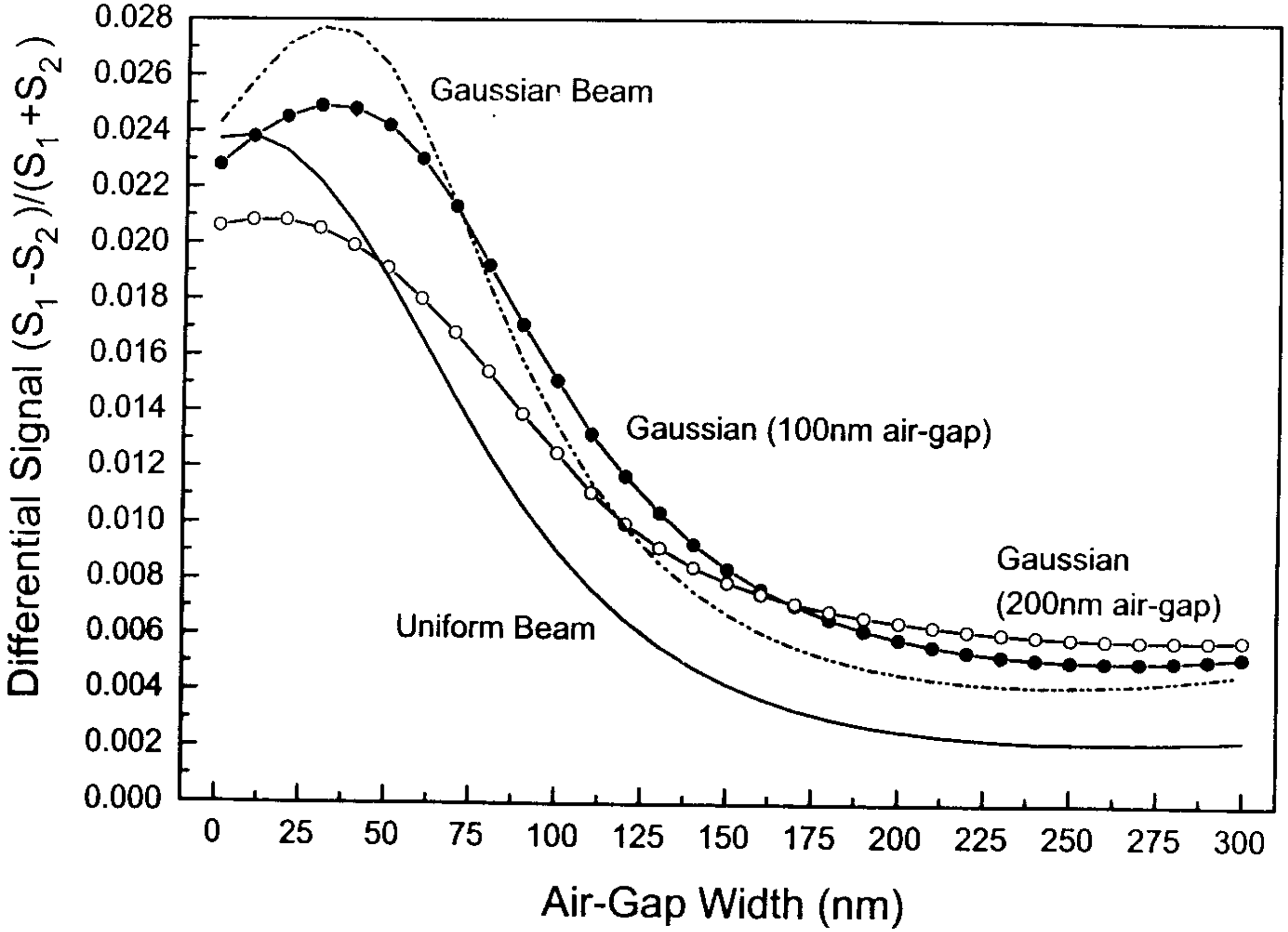
MO (whole aperture)

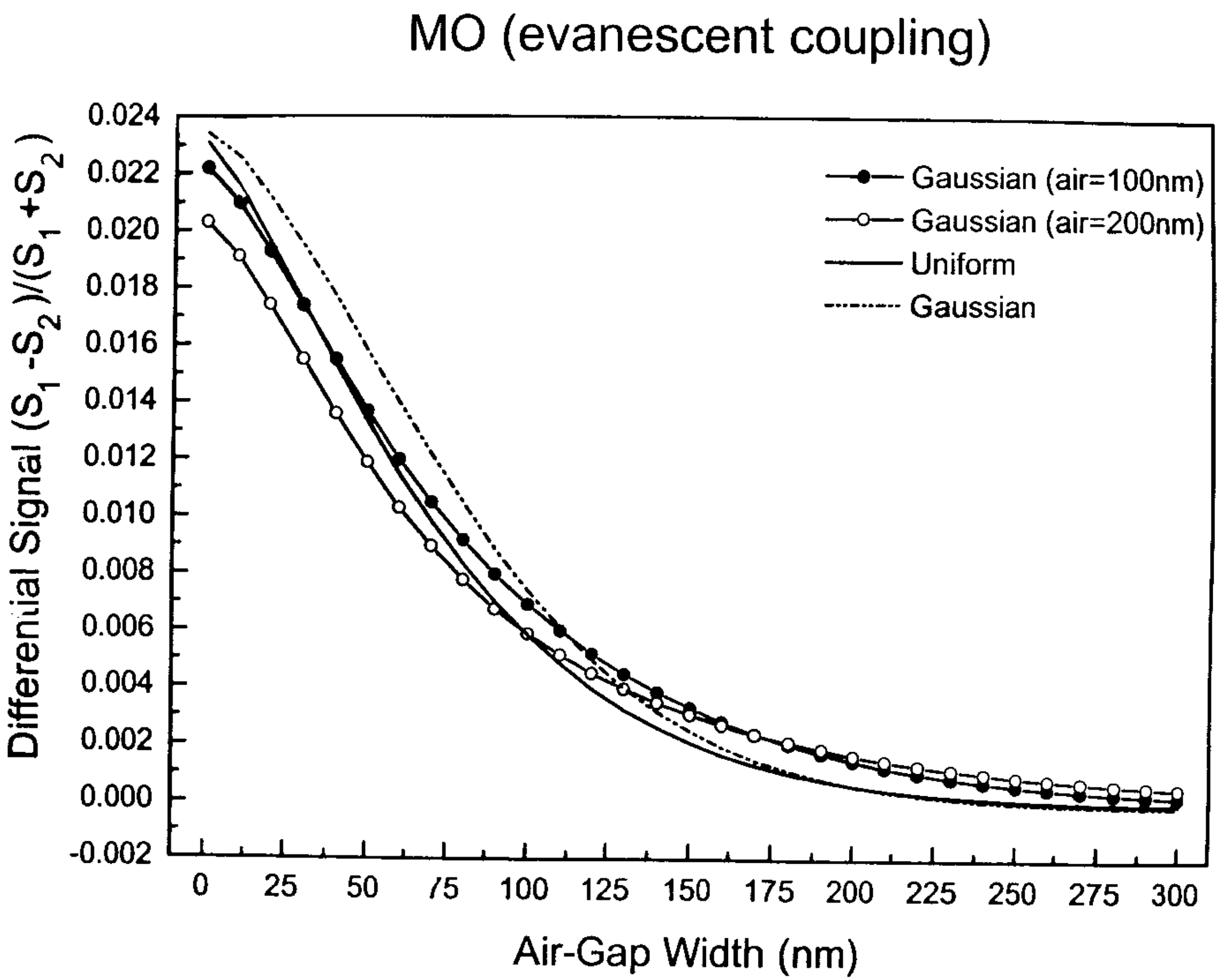


95


MO (evanescent couping)
Differential Signal (S 1-S 2)/(S 1+S 2)
Thickness of the dielectric layer (nm)
Air Gap
0nm
50nm
100nm
0.024
0.022
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
-0.002
0
20
40
60
80
100
120
140
160
180
200

MO (whole aperture)




MO (evanescent coupling)
Differential Signal (S1 -S2)/(S1 +S2)
0.024
0.022
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
-0.002
0
25
50
75
100
125
150
175
200
225
250
275
300
Air-Gap Width (nm)
Gaussian (air=100nm)
Gaussian (air=200nm)
Uniform
Gaussian

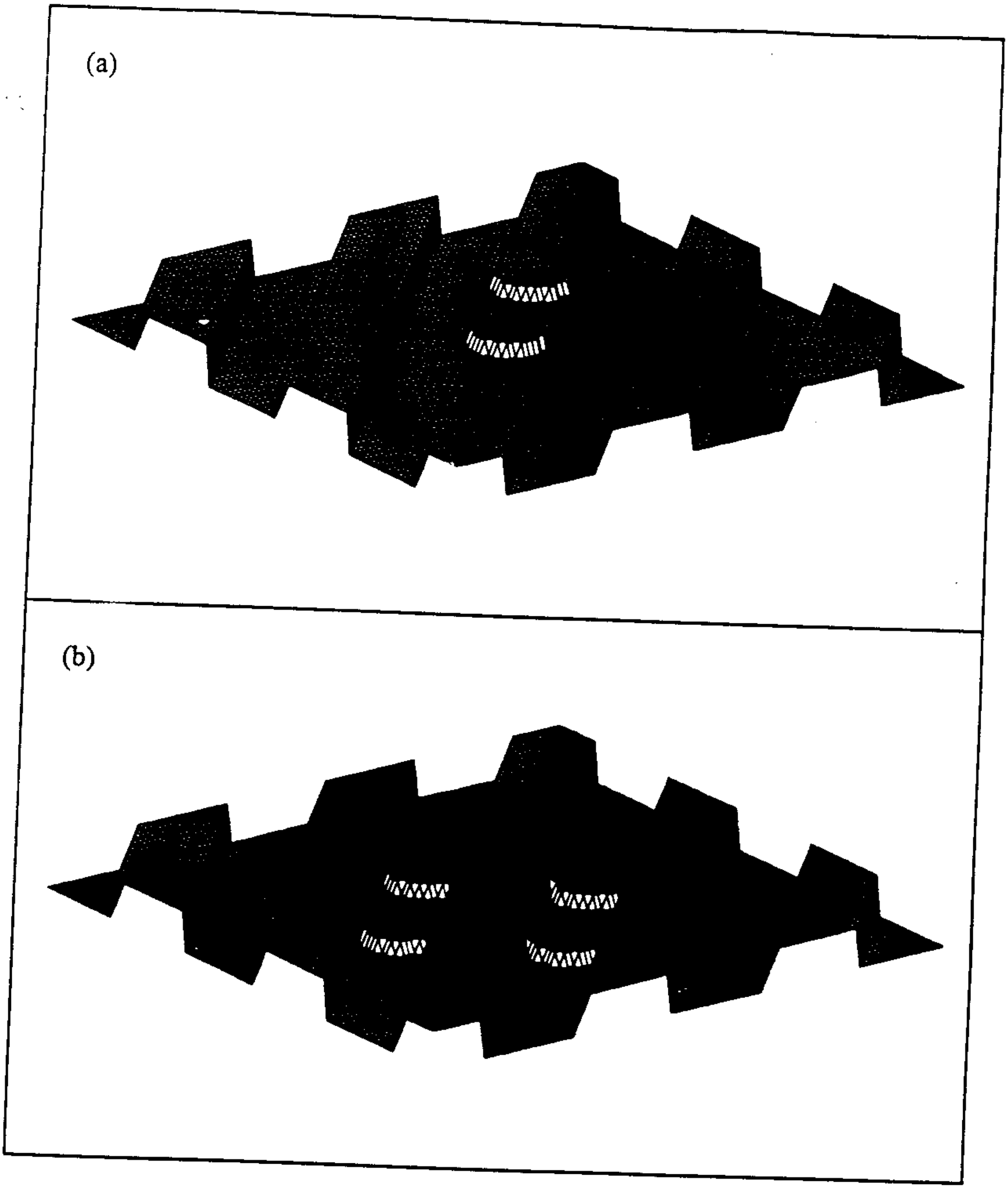


Figure 12

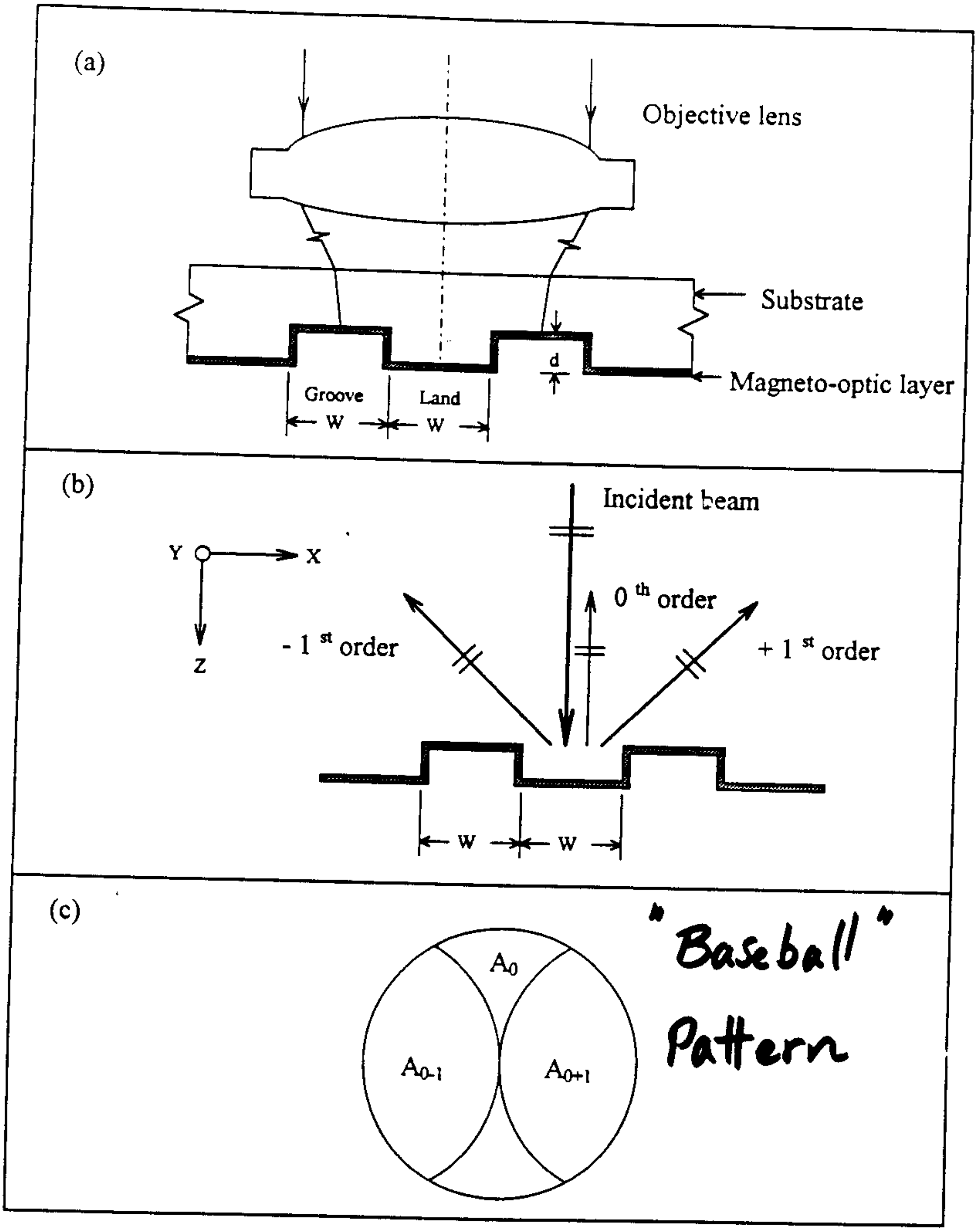
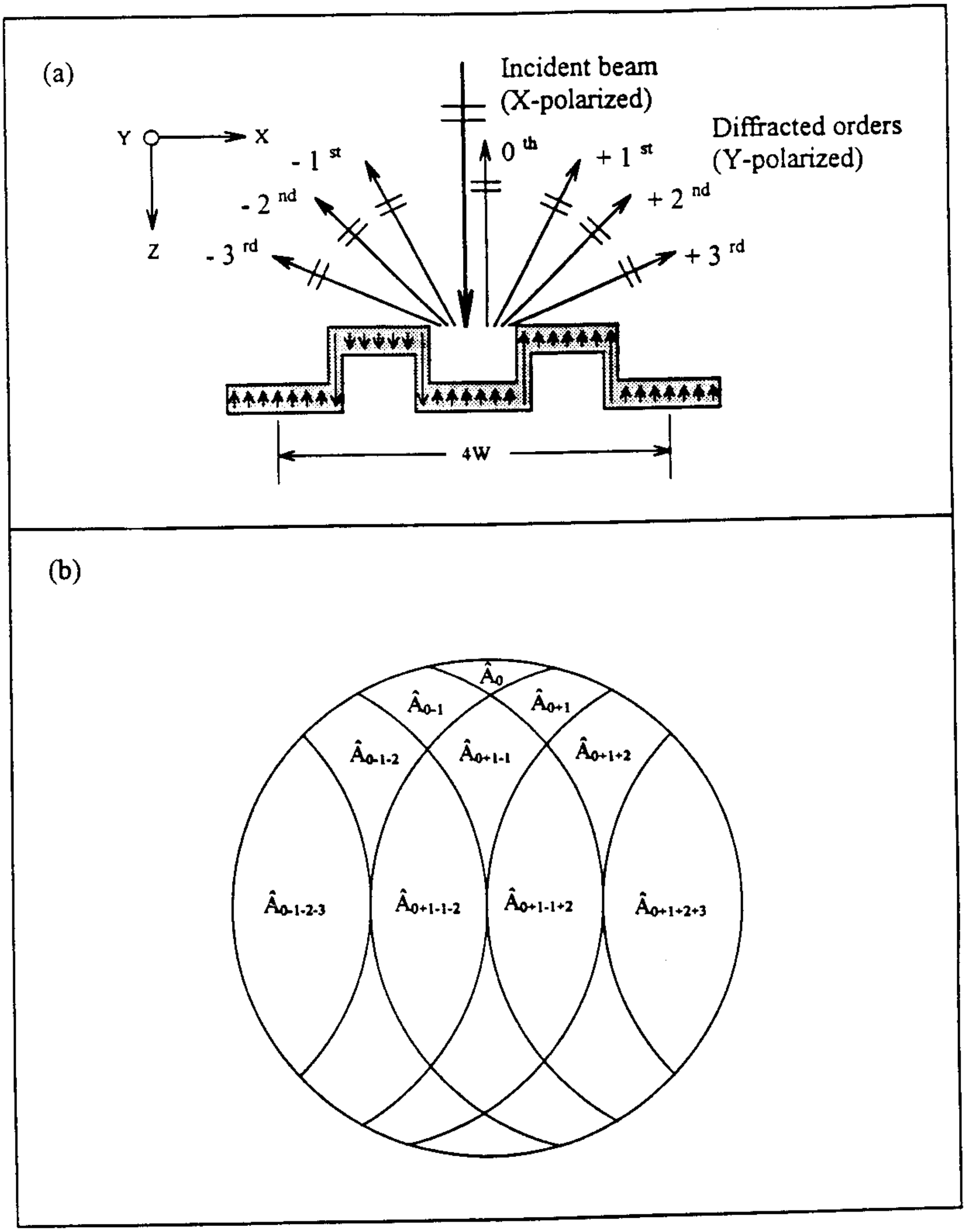

(a)
Objective lens
Substrate
d
Magneto-optic layer
Groove
Land
W
W
(b)
Incident beam
Y
X
Z
0 th order
- 1 st order
+ 1 st order
W
W
(c)
A0
A0-1
A0+1
"Baseball" Pattern

Figure 4

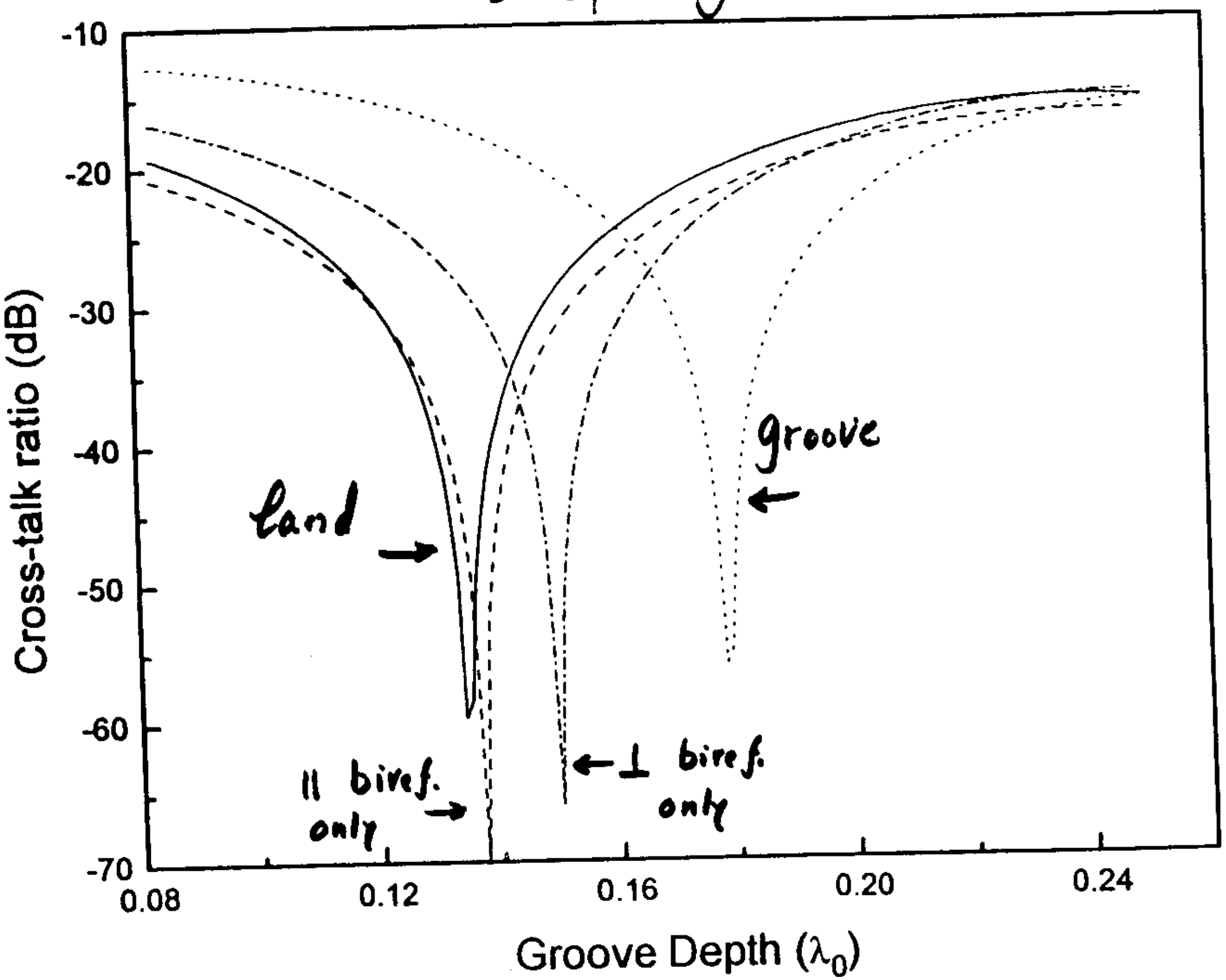


(a)
Incident beam
(X-polarized)
Diffracted orders
(Y-polarized)
Y
X
Z
0 th
- 1 st
- 2 nd
- 3 rd
+ 1 st
+ 2 nd
+ 3 rd
4W
(b)
Â0
Â0-1
Â0+1
Â0-1-2
Â0+1-1
Â0+1+2
Â0-1-2-3
Â0+1-1-2
Â0+1-1+2
Â0+1+2+3

Figure 5

Figure 18

Cross-talk Cancellation and the effect of Birefringence



If the sources in a laser diode array are separated by a large distance at the laser facet, the optical styli will also be widely spaced at the recording surface. Given a source spacing of d, the spot spacing is $\Delta = d/M$, where M is the net magnification of the optical system. For a simple collector/objective lens combination, the magnification is just the ratio of the numerical apertures, $M = NA_{obj}/NA_{coll}$. Anamorphic beam expansion acting along an axis parallel to the source junction can affect the magnification in this direction. If a typical laser beam (see the parallel profile in Figure 6.5) is expanded to fill a lens with $NA_{obj} \sim 0.5$, the net magnification is about 3 × to 4 ×. Therefore, the focused spots will be separated by more than 10 μm. Since the source locations are well defined on the laser facet, the relative spot positions at the media surface are very steady, insensitive to thermal changes and mechanical changes in the head.

Figure 8.16 illustrates an implementation of multi-channel recording with a three-source laser diode array. In order to record neighboring tracks, the laser array must be slanted nearly parallel to the

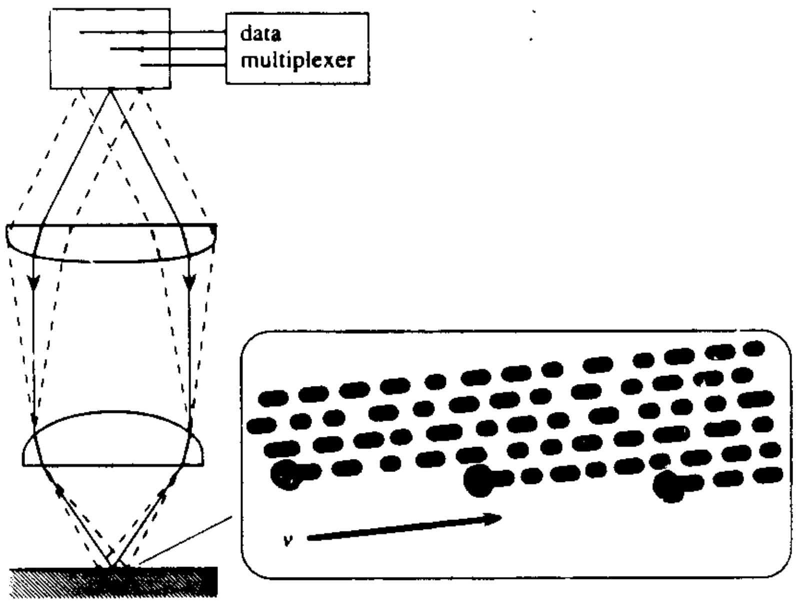


Figure 8.16 Multichannel recording with a laser diode array is illustrated here.

Multichannel Read/Write with a laser diode array.

Integrated optical head

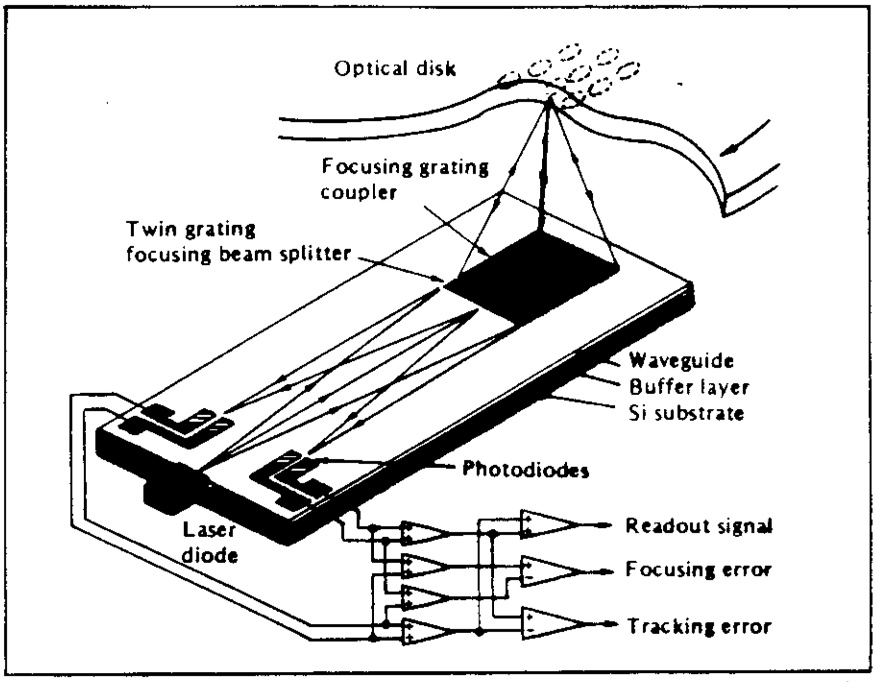


Fig. 8. Schematic view of the proposed integrated-optic pickup device

Suhara + Nishihara, OSAKA Univ.

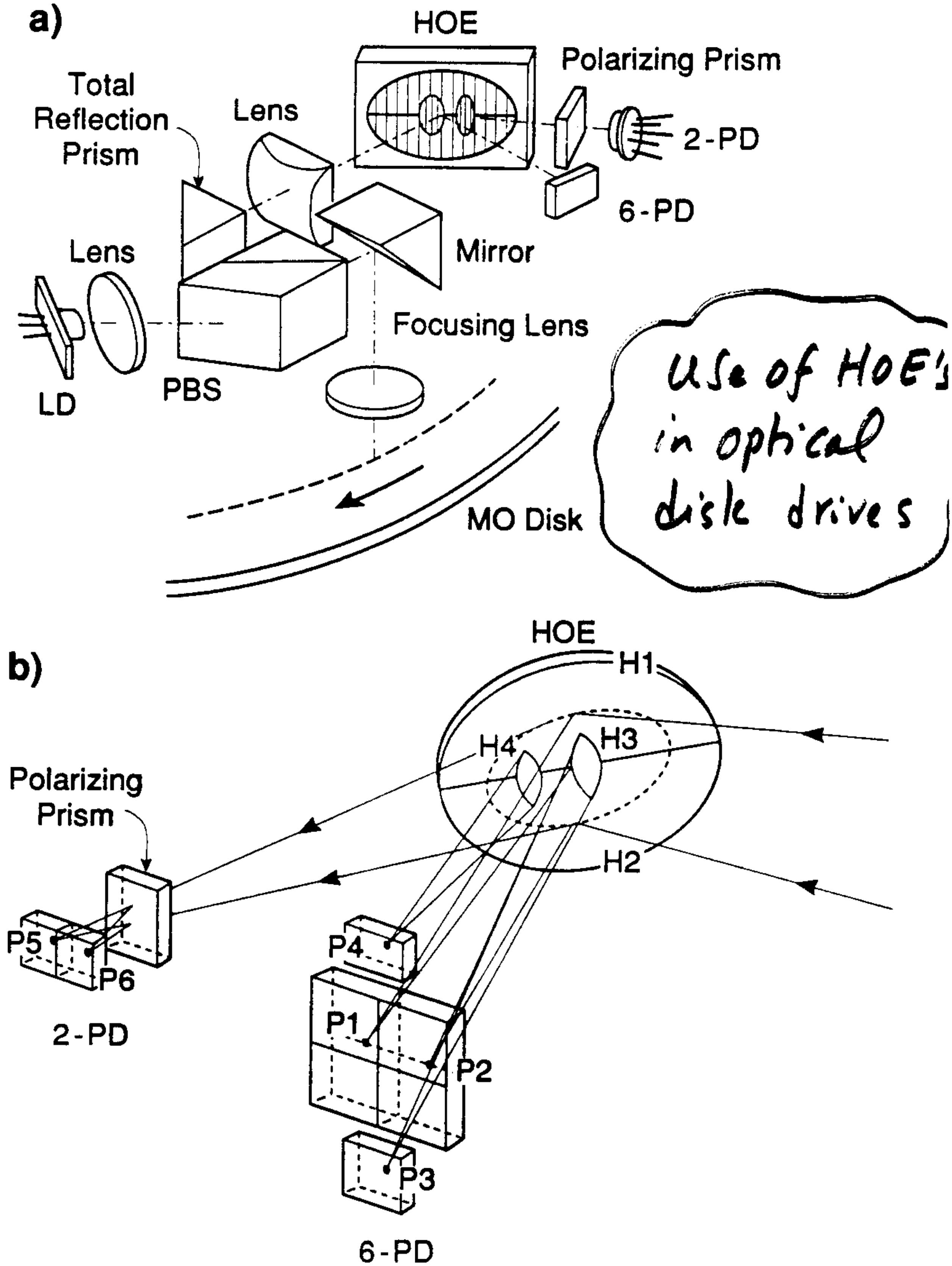


Illustration by
Wally Varner / UA Graphics
Mansuripur fig 27

MO Readout Using External Cavity Laser Diode

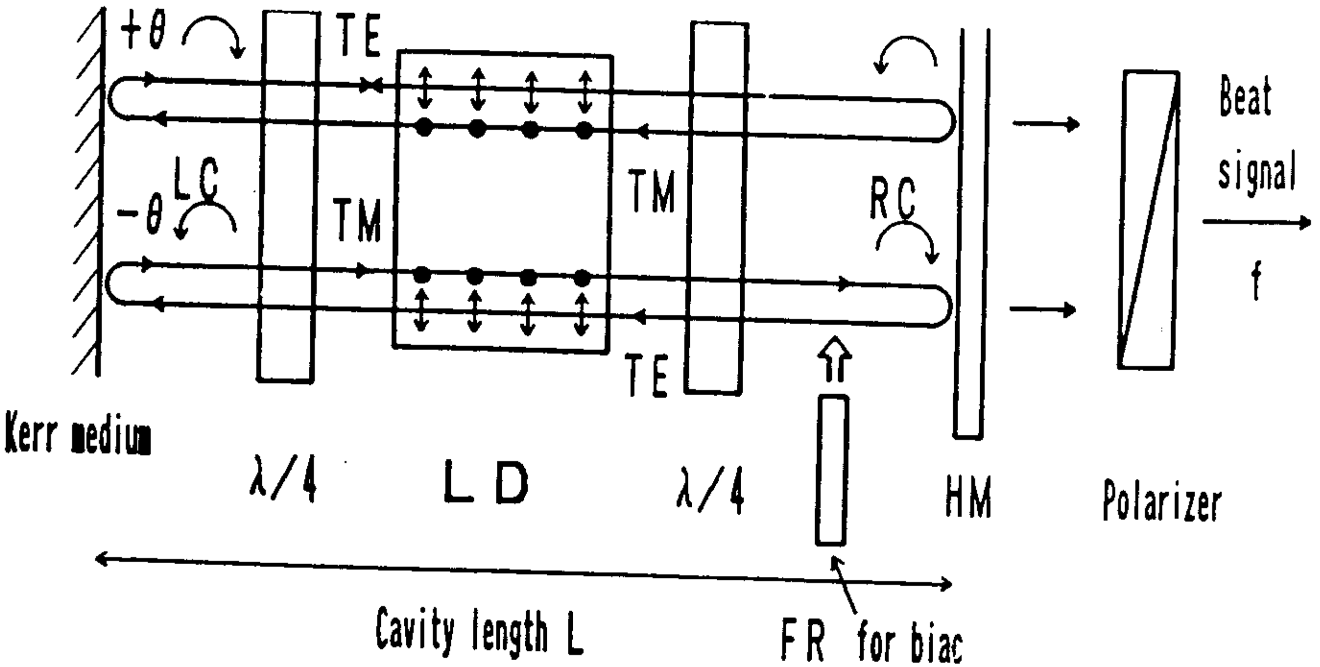


Fig. 1. Dual frequency generation by an external cavity laser diode.

N. Fukushima et al Fujitsu Labs

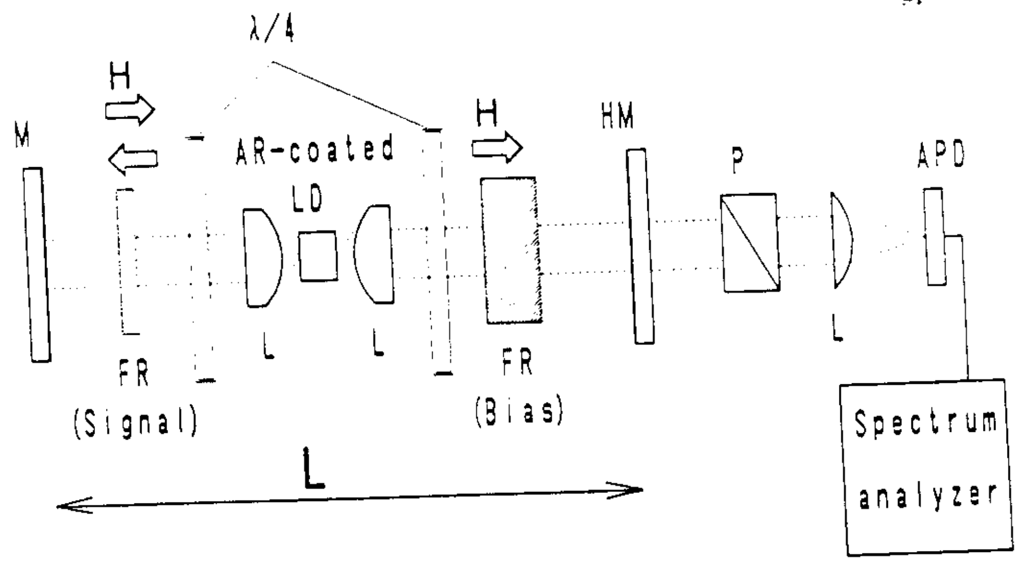


Fig. 2. Experiment setup for measuring beat frequency.

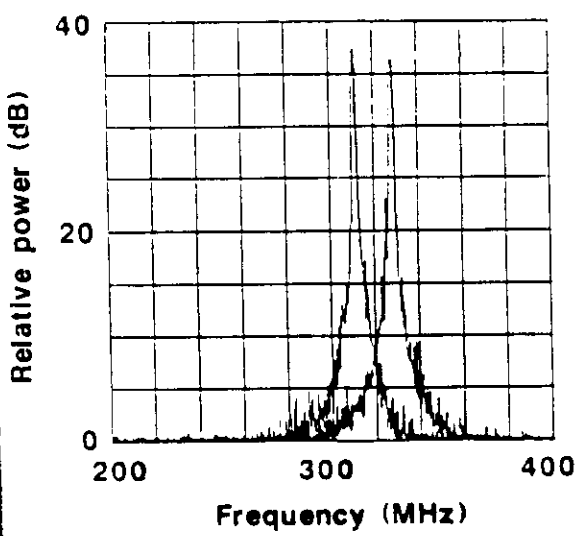


Fig. 3. Beat frequency shift induced by magnetization reversal of the signal Faraday rotator. Optical length of the cavity is 140 mm. The rotation angle of the Faraday rotators is 0.73° for the signal and 29° for the bias.

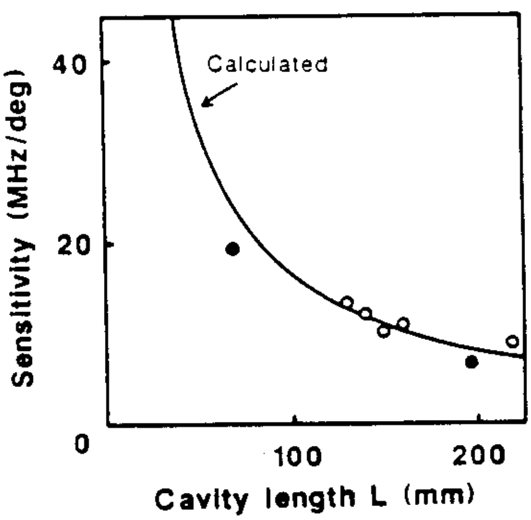


Fig. 4. Sensitivity of magneto-optic signal detection versus optical length of the cavity. The sensitivity is defined as the ratio of beat frequency shift to the Kerr rotation angle. Open circles are data for a laser of wavelength 830 nm and closed circles for 1300 nm.

N. Fukushima et al, Fujitsu Labs

Areas in which R&D will lead to better, high performance devices

1) Semiconductor diode lasers
2) Photodetectors
3) Plastic/Glass substrates
4) Lithography
5) Magneto-optical materials
6) Erasable phase-change materials
7) Miniaturization/Integrated optics
8) Signal Processing (optical and electronic)

Principles and Techniques of Optical Data Storage

MASUD MANSURIPUR AND GLENN SINCERBOX

Invited Paper

We review the field of optical data storage and describe the various technologies that either are in use today or are likely to play a role in the near future. Our emphasis will be on optical-disk and holographic optical storage.

Keywords— *Erasable optical data storage, holographic data storage, magnetooptical recording, optical disk, phase-change optical disk.*

I. Introduction

Our need for storage is explosive; fueled by multimedia requirements for text, images, video, and audio, storage requirements are growing at an exponential rate and are expected to exceed 10^{20} bits (12 exabytes) in the year 2000. With an increasing amount of information being generated or captured electronically, a large fraction, perhaps as much as 40%, will be stored digitally. To meet this need, the hierarchy of on-line, near-line and off-line storage systems will be composed of many diverse technologies: magnetic disk drives, magnetic tape drives and tape libraries, and optical disk drives and libraries. The mix of these subsystems will be very application specific so as to optimize the performance and cost of the overall system.

An optical storage system is a particularly attractive component of this hierarchy because it provides data-access times that are an intermediate solution between a hard disk drive and a tape drive. Access time is the time, including latency, required to start retrieving a random block of data, and typically ranges from less than 10 ms for a hard disk drive to 30–50 ms for an optical disk drive to several seconds for a tape drive. It becomes an important link in the chain as data are staged up and down between the CPU, memory, and storage.

Perhaps the most enabling feature of optical storage is the removability of the storage medium. With separations of a few millimeters between the recording surface and the optical "head," and with active servos for focusing and tracking, the medium can be removed and replaced with relatively loose tolerances. The infamous *head crashes* often experienced in hard disk drives do not occur in optical drives. Data reliability and removability is further enhanced by using the 1.2-mm transparent disk substrate as a protective cover to keep contamination away from the recording surface. (The recently announced digital versatile disk or DVD will have a substrate thickness of only 0.6 mm.)

Removability has created a whole new industry in compact disc (CD) audio. CD read only memory (CD-ROM) has enhanced the efficiency of distribution and use of software, games, and video. These read-only disks containing 680 Mbytes of information can be mass replicated by injection molding in a few seconds for less than ten cents each. CD's are virtually indestructible, and one-day express mail of 125 CD's is an effective data rate of 1.0 Mbytes/s.

At the other end of the spectrum, phase change and magnetooptical disks are used in write-once read-many (WORM) and read/write/erase systems, where a single disk can contain almost 5 Gbytes. Using robotics, storage libraries can be assembled with capacities of a terabyte with access to any disk in less than five seconds.

This review will discuss the fundamentals of optical storage, the components that compose a system, and the emerging technologies that will allow increased performance, higher storage capacities, and lower cost. We will then extend our discussion into the third dimension with some comments on multilayer recording, such as double-layer DVD, and conclude with some remarks on the latest advances in holographic storage—a 30-year-old optical storage technology that is generating renewed interest because of its potential for high data rates and rapid access times.

Several international conferences on optical data storage are held each year, and the proceedings of these conferences are a good source of information concerning the latest developments in this field. References [1]–[3] provide some information about these conference proceedings and/or di-

Manuscript received June 27, 1997; revised July 30, 1997.

The authors are with the Optical Sciences Center, University of Arizona, Tucson, AZ 85721 USA.

Publisher Item Identifier S 0018-9219(97)08230-3.

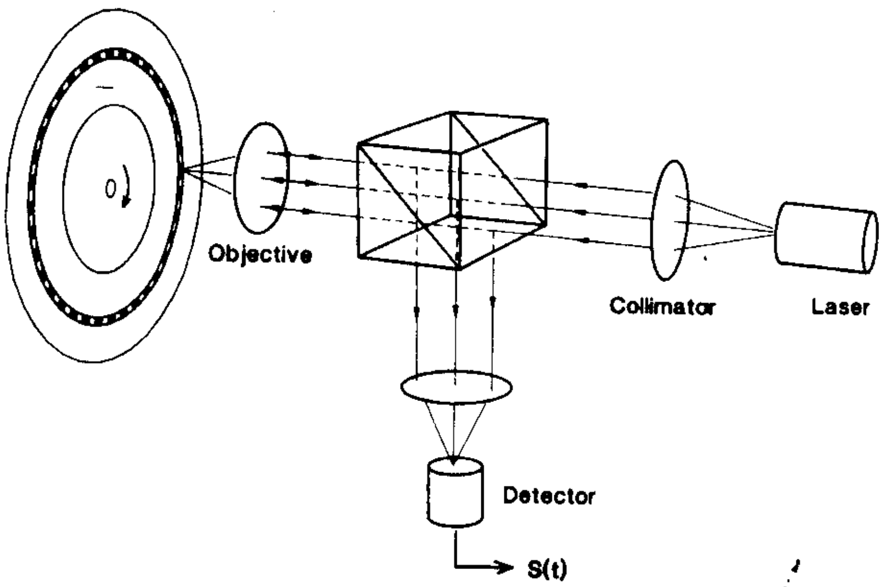


Fig. 1. Schematic diagram showing the basic configuration of an optical-disk system. The light from a semiconductor laser diode is collimated and directed toward a high-numerical-aperture objective lens. The objective brings the light to diffraction-limited focus on the surface of the disk, where information is written to or read from a given track. In the return path, a beam splitter directs the beam toward one or more detectors, where the recorded information as well as focusing- and tracking-error signals are extracted. In practical disk drives, the objective lens is mounted on a pair of actuators that carry the lens in response to movements of the disk, thus allowing the lens to stay locked in focus and on track.

gests. There also are several published books in this area, to which the readers are referred for an in-depth coverage of the various subjects [4]–[7].

II. Optical-Disk Data Storage

A. Recording and Readout of Information on Optical Disks

An optical disk is a glass or plastic substrate with one or more thin-film layers coated upon its surface(s). The information may be prerecorded on the surface of the substrate by the disk manufacturer, or it may be recorded on one or more of the thin-film layers by the user. Typical diameters of presently available optical disks are:

- 12 cm, as in CD audio, CD-ROM, CD recordable (CD-R), and (DVD);
- 2.5″, as in digital audio minidisk (MD);
- 3.5″, as in magnetooptical (MO) removable storage devices;
- 5.25″, as in both magnetooptical and phase-change (PC) rewritable disks;
- 12″ and 14″, such as the WORM media manufactured by Kodak and other companies for high-volume storage applications.

In the case of read-only media, the information is pressed onto the substrate by injection molding of plastics or by embossing of a layer of photopolymer coated on a glass substrate [4], [5]. The substrate is then coated with a thin metal layer (e.g., aluminum) to enhance its reflectivity. In other types of optical disks, some information, such as format marks and grooves, may be stamped onto the substrate itself, but then the substrate is coated with a storage layer that can be modified later by the user during recording of information. Typical storage layers are dye-polymer films for write-once applications, tellurium alloys for ablative recording (also write once), GeSbTe for phase-change rewritable media, and TbFeCo magnetic films for magnetooptical disks (also rewritable) [5]–[7].

In a typical disk, the storage layer is sandwiched between two dielectric layers, and the stack is capped with a reflector layer (e.g., aluminum or gold) and protected with a lacquer layer. The dielectric layers and the reflective layer perform several tasks, among them [6], [7]:

- protecting the storage layer;
- creating an optically tuned structure that has optimized reflectivity and/or absorptivity;
- allowing tailoring of the thermal properties of the disk for rapid cooling and reduction of thermal cross talk during writing.

Fig. 1 is a schematic diagram showing the basic elements of an optical disk drive. The laser is usually a semiconductor laser diode, whose beam is collimated by a well-corrected lens and directed toward an objective lens through a beam splitter. The objective lens focuses the beam onto the disk and collects the reflected light. This reflected light is directed (at the beam splitter) toward the detectors, which produce a data readout signal as well as servo signals for automatic focusing and tracking. The functions and properties of the various elements of this system will be described in some detail in the following sections.

B. The Light Source

All optical recording technologies rely on lasers as their source of light. The lasers used in optical-disk and tape data storage are semiconductor laser diodes of the shortest possible wavelength that can provide sufficient optical power for read/write/erase operations over a period of several thousand hours. Presently, the shortest wavelength available in moderate-power lasers (around 50 mW) is in the neighborhood of 680 nm; these lasers are being used in CD-R, MO, and PC products. For CD and DVD applications, where writing is not a concern, low-power lasers (e.g., ≈ 5 mW), which emit at 630 and 650 nm, are being considered. The emphasis on short-wavelength lasers for optical recording applications is due to the fact that shorter wavelengths can be focused to smaller spots at the diffraction limit. All things being equal, the diameter of a focused spot scales with its wavelength: a reduction of the wavelength by a factor of two, for example, will result in a reduction of the focused spot diameter by the same factor, and, consequently, a fourfold increase in the data storage density can be realized.

The wavelengths of optical data storage have continuously shrunk during the past 15 years; starting at 830 nm, they are now down to 630 nm, and there is every indication that they will continue to shrink in the foreseeable future. What is needed for optical data storage is compact and inexpensive laser diodes that can be incorporated into small, low-cost drives. The power requirement from such lasers is several milliwatts for read-only media and several tens of milliwatts for recordable media. The lasers should be capable of direct modulation (e.g., by modulating the electrical current input to the laser); otherwise, the cost and size of external modulators may become prohibitive. Spatial coherence and single transverse-mode operation is a requirement, because the beam must be focused to diffraction limit. (In this context, vertical-cavity surface emitting laser diodes, as well as arrays of such lasers, need to be improved since, at high powers, these lasers tend to operate in high-order modes.) Low-noise operation is very important, especially in applications such as magnetooptical readout, where the signal-to-noise ratio (SNR) is at a premium.

Although operation of the laser diode in several longitudinal modes is presently acceptable, for future devices, it may be important to add operation in a single, stable, longitudinal mode to all their other desirable characteristics. Mode hopping, wavelength shifts of several nanometers with temperature fluctuations and with operating current, manufacturing variability of the wavelength from batch to batch, etc. are so severe that at the present time, it is not possible to consider the use of diffractive lenses either for the collimator or for the objective lens. These high numerical aperture (NA), diffraction-limited lenses are still produced by molding of glass elements. In the future, when microminiaturization becomes a necessity and lenslet arrays begin to appear in commercial products, operation of the laser in a single, stable, longitudinal mode may be required.

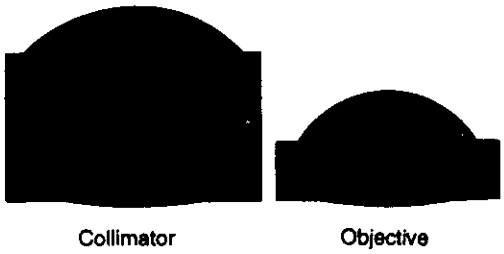


Fig. 2. The lenses used in optical disk drives are single-element aspherics made by glass molding. The collimator has a numerical aperture of ≈ 0.4 and is usually corrected for the astigmatism inherent in edge emitting laser diodes. The objective lens has an NA of 0.45–0.6 and is free from aberrations. Both lenses are designed for operation at a single wavelength. The objective's field of view is wide enough to allow beam steering, but it may also be used to focus multiple beams on several tracks in parallel.

C. The Objective Lens

Presently, disk and tape optical recording systems use molded glass lenses for focusing the laser beam to a diffraction-limited spot (see Fig. 2). These lenses consist of two aspheric surfaces on a single piece of glass, have fairly large NA's (in the range of 0.4–0.6), and are essentially free from aberrations. The NA of a lens is defined as $\sin\theta$, where θ is the half-angle subtended by the focused cone of light at its apex. A 0.5 NA lens, for example, will have a focused cone whose full angle is 60°. The diameter of the focused spot is of the order of λ_0/NA, where λ_0 is the vacuum wavelength of the laser beam. It is thus clear that higher NA's are desirable if smaller spots (and therefore higher recording densities) are to be attained. Unfortunately, the depth of focus of an objective lens is proportional to λ_0/NA^2, which means that the higher the NA, the smaller will be the depth of focus. It thus becomes difficult to work with high NA lenses and maintain focus with the desired accuracy in an optical disk drive.

But a small depth of focus is not the main reason why the present optical drives operate at moderate numerical apertures. The more important reason has to do with the fact that the laser beam is almost invariably focused onto the storage medium through the disk substrate. The disk substrate, being a slab of plastic or glass, has a thickness of 1.2 mm (DVD substrates are only half as thick, or 0.6 mm). When a beam of light is focused through such a substrate, it will develop an aberration, known as coma, as soon as the substrate becomes tilted relative to the optical axis of the objective lens. Even a 1° tilt produces unacceptably large values of coma in practice. The magnitude of coma is proportional to NA^3, and therefore, higher NA lenses exhibit more sensitivity to disk tilt. In the future, manufacturers will move toward higher NA's by adopting one or more of the following strategies.

1) Use thinner substrates or avoid focusing through the substrate altogether.
2) Make the substrates as flat as possible.
3) Develop a tilt servo mechanism whereby the tilt of the disk is automatically sensed and corrected for in the optical path.

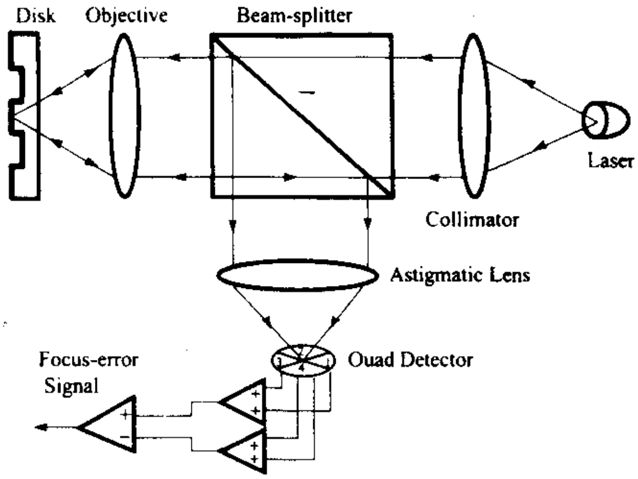


Fig. 3. Astigmatic focus-error detection system. The quad detector sits halfway between the two line-foci of the astigmat. When the disk is in focus, the beam arriving at the astigmat is collimated and the spot at the detector is more or less round and symmetric. When the disk moves out of focus, the beam arriving at the astigmat will be either convergent or divergent, forcing one or the other of the line-foci to move toward the detector. The distribution of light intensity at the detector thus becomes asymmetric, with one pair of opposite detectors receiving more light than the other pair.

As was mentioned earlier, the use of smaller lenses is always desirable in optical storage technology, particularly if lens arrays are being considered for parallel accessing of multiple tracks in a system. In this respect, gradient-index lenses, holographic optical elements, and binary diffractive optical lenses are all being considered for future generations of optical storage devices.

D. Automatic Focusing

A typical optical disk has a plastic substrate that is not perfectly flat but is slightly warped. Also, when mounted in a drive, small tilts of the axis could cause vertical motions of the disk surface during operation. It is not unusual to find vertical movements of as much as ± 100 μm during the operation of an optical disk. Now, a typical objective lens has a numerical aperture of 0.5 or higher, and therefore, the focused beam has a depth of focus of a fraction of λ_0/NA^2, which is less than a fraction of a micrometer. The focused spot must remain within the depth of focus, while the disk rotates at speeds of several thousand rpm and wobbles in and out of focus by as much as ± 100 μm in each revolution. Needless to say, without an autofocus mechanism to maintain the disk continually in focus, the operation of an optical disk drive is unthinkable.

In practice, the objective lens is mounted in a voice coil actuator (bandwidth = several kHz), and a feedback mechanism is used to drive the lens toward and away from the disk in such a way as to maintain focus at all times. The signal needed for this feedback mechanism is derived from the light that is reflected from the disk itself. Fig. 3 shows a diagram of the astigmatic focus-error detection system used in many practical devices these days. The light reflected from the disk and collected by the objective lens is either convergent or divergent, depending on whether the disk is farther away from best focus or closer to the lens than the plane of best focus. This returned beam goes through an astigmatic lens, which normally focuses the incident beam to a symmetric spot halfway between its focal planes. A quad detector placed at this plane (also called the plane of least confusion) then receives equal amounts of light on its four quadrants. When the disk is out of focus, however, the astigmat creates an elongated spot on the detector. Depending on the sign of defocus, this elongated spot may illuminate quadrants one and three or quadrants two and four of the detector. Therefore, the combination signal $(S_1+S_3)-(S_2+S_4)$ provides a bipolar focus-error signal, which is fed back to the voice coil for maintaining focus automatically.

E. Automatic Track Following

The information on an optical disk is recorded either around a series of concentric circular tracks or on a continuous spiral. Manufacturing errors and disk eccentricities caused by mounting errors, thermal expansion of the substrate, etc. will cause a given track to wobble in and out of position as the disk spins. Typically, a given track might be as much as ± 100 mm away from its intended position at any given time. The focused spot, of course, is only about 1 mm across, and cannot be at the right place at all times. An automatic tracking scheme is therefore desired. Given the mechanical rotation rates of the disks, the frequency response of the actuator needed for track following does not have to cover more then a few kilohertz, and a voice coil is usually sufficient for the purpose. The feedback signal for controlling the position of the objective lens within the tracking coil is again provided by the return beam itself. Several mechanisms have been proposed and have been put to use in commercial devices. We mention three of these schemes here.

The push-pull tracking mechanism relies on the presence of either grooves or a trackfull of data on the media. In the case of CD and CD-ROM, the data are prestamped along a spiral on the substrate, and the sequence of marks along the spiral comprises a sort of discontinuous groove structure. The discontinuity is irrelevant to the operation of the tracking servo, however, because it is at a much higher frequency than the tracking servo is designed to follow. Writable media such as CD-R, MO, and PC require a tracking mechanism distinct from the data pattern because prior to the recording of data, the WRITE head must be able to follow the track before it can record anything. Once the data are recorded, the system will have a choice as to whether to follow the original tracking mechanism or the recorded data pattern. Continuous grooves are a popular form of preexisting tracks on optical media [see Fig. 4(a)]. A typical groove is a fraction of a micrometer wide (say, 0.4 μm) and one-eight of a wavelength ($\lambda/8$) deep. As long as the focused beam is centered on a track, diffraction of light from the adjacent grooves will be symmetric. The symmetry of the reflected beam, as sensed by a split detector in the return path, would produce a zero error signal (see Fig. 5). When the focused spot moves away from the reflection. This is known as the polar magnetooptical Kerr effect. The sense of polarization rotation is dependent on the state of magnetization of the medium. Thus, when the magnetization is pointing up, for example, the polarization rotation is clockwise, whereas down-magnetized domains rotate the polarization counterclockwise. The polar Kerr effect provides the mechanism for readout in MO disk data storage. Typical materials used today impart about 0.5° of polarization rotation to the linearly polarized incident light. But, given the extremely low levels of noise in these media, the small Kerr signal nonetheless provides a sufficient SNR for reliable readout. Readout and writing in optical disk drives are typically carried out with the same laser. To avoid obliteration of the recorded data, the power of the read beam must be substantially below that of the write beam. Also, readout is carried out in CW, whereas writing is generally done in pulsed mode. To reduce the laser noise in readout, it is customary to apply a very-high-frequency modulation (several hundred MHz) to the laser driver. This high-frequency modulation is well outside the range of recording frequencies and does not affect the readout of data, but it forces the laser to operate in all its various longitudinal modes simultaneously, becoming temporally incoherent but, at the same time, less noisy.

5) Amorphous Nature of the Materials: The media of MO recording are amorphous. Lack of crystallinity in these media makes their reflectivity extremely uniform, thereby reducing the fluctuations of the read signal. This amounts to very low levels of noise in readout, which ultimately helps increase the data storage densities. (In general, the larger the available SNR from a given medium, the higher will be the achievable packing densities in that medium.) To be sure, noise from the MO medium is not the only noise in readout, but it is the dominant one. The other sources of readout noise are the thermal noise of the electronic circuitry, the shot noise of photodetection, and the laser noise. The method of MO readout is a differential method, whereby the signal is split between two photodetectors and the outputs of the two are then subtracted from each other to yield the final signal. The subtraction eliminates many of the common mode sources of noise, but of the noises that remain at the end, the media noise is still the dominant component.

H. Phase-Change Media and the Mechanism of Recording

Presently, the medium of choice for erasable phase-change recording is a $Ge_2Sb_2Te_5$ alloy, which is affectionately referred to as the GST material [8]. This alloy is sputter-deposited on a plastic substrate, with an undercoat and an overcoat of ZnS-SiO_2 dielectric layers. The stack is then capped with an aluminum alloy layer for making an antireflection structure. The quadrilayer stack is also effective as a rapid cooling structure, thanks to the heat-sinking properties of the aluminum layer. The as-deposited GST alloy is amorphous. Each disk is annealed at the factory, however, to transform the recording layer into its polycrystalline state. The recording process turns small regions of the GST medium into amorphous marks by raising the local temperature above the melting point and allowing a rapid cooldown quenching. The reflectivity of the amorphous mark is different from that of the polycrystalline background, and therefore, a signal is developed during readout. Erasure is achieved by using a laser pulse of an intermediate power level (i.e., between the read and write powers). If sufficient time is allowed for the laser beam to dwell on the amorphous mark, the mark will become crystalline once again (annealing). This process is compatible with direct overwrite and is therefore preferable to MO recording, where direct overwrite is harder to achieve.

Comparing PC and MO technologies, one can find several advantages and disadvantages for each. PC drives are simpler than MO drives because they do not need magnets to create external magnetic fields, and also because there is no need for sensitive polarization-detecting optics in PC readout. The read signal is very strong for PC media, so much so that despite the rather large component of media noise of PC, the SNR is still somewhat larger than that of MO media. On the other hand, repeated melting, crystallization, and amorphization of PC media result in material segregation, stress buildup, microcrack formation, etc. These factors tend to reduce the data reliability and cyclability of the PC media. MO disks are guaranteed to sustain more than 10^6 read/write/erase cycles and can probably do better than that in practice, but the corresponding figure for PC media typically is one to two orders of magnitude lower. The maximum temperature reached in MO media during recording and erasure typically is around 300°C, as opposed to 600°C in PC media. The lower temperatures and the fact that magnetization reversal does not produce material fatigue account for the longer life and better cyclability of the MO media. Writing and erasure in MO media can be very fast, fundamentally because spin flips occur on a subnanosecond time scale. In contrast, although amorphization in PC media can be very rapid, crystallization is a rather slow process: the atoms must be kept at elevated temperatures long enough to move around and find their place within the crystal lattice. As a result, in principle, high data rates are achievable with MO media, but there may be barriers to achieving them in PC media. On the other hand, PC media are directly overwritable, but MO media can be overwritten either with magnetic field modulation (which is a rather slow process) or by using exchange-coupled magnetic multilayers (which are more difficult to manufacture than the conventional single-magnetic-layer disks).

I. Solid Immersion Lens

A new approach to optical-disk data storage involves the use of near-field optics in general and the solid immersion lens (SIL) in particular [9]. The SIL approach requires that a part of the objective lens fly over the surface of the storage medium, as shown in Fig. 7. The hemispherical glass of refractive index n receives the rays of light at normal incidence to its surface. These rays come to focus at the center of the hemisphere and form a diffraction-limited

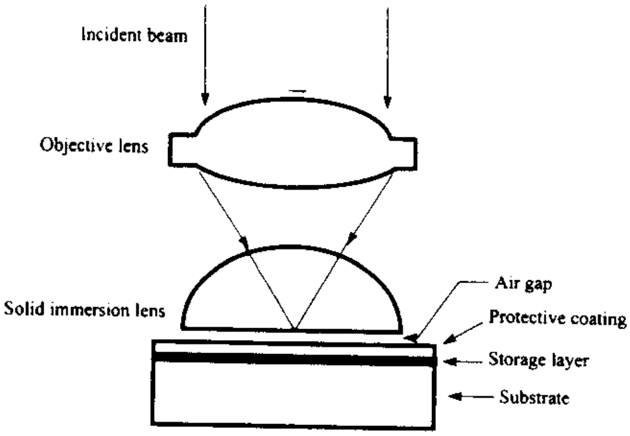


Fig. 7. An SIL is a hemispherical piece of glass of refractive index n suspended over the disk surface. The rays emerging from the objective lens enter the SIL at normal incidence and come to focus at the bottom of the hemisphere. Since the wavelength of the light inside the glass medium is shortened by a factor of n, the diameter of the focused spot will be reduced by the same factor. The SIL must either remain in contact with the active layer of the disk or fly extremely close to it (flying height $\approx \lambda/10$).

spot that is smaller by a factor of n compared to what would have been the case in the absence of the SIL. (This is a well-known fact in microscopy, where oil immersion objectives have been in use for many years.) A typical glass hemisphere having $n = 2$ will reduce the diameter of the focused spot by a factor of two, thus increasing the recording density fourfold. To ensure that the smaller spot size does indeed increase the resolution of the system, the bottom of the hemisphere must either be in contact with the active layer of the disk or fly extremely closely to it. For a disk spinning at several thousand rpm, it is possible to keep the SIL at a distance of less than 100 nm above the disk surface.

The rays of light that are incident at large angles at the bottom of the hemisphere would have been reflected by total internal reflection except for the fact that light can tunnel through and jump across gaps that are small compared to one wavelength. This tunneling mechanism is known as frustrated total internal reflection, and its presence qualifies the application of SIL in optical data storage as a near-field technique. The price one will have to pay for the increased storage density and data rate afforded by SIL is the necessity to enclose the disk permanently within the drive, thereby making it nonremovable. (The possibility of maintaining removability in a SIL system has been suggested, but it remains to be demonstrated in a practical setting.)

J. Partial Response Signaling and Maximum Likelihood Detection (PRML)

A technique that has been found very useful in communication systems and magnetic hard drives is beginning to make inroads in optical data storage as well. The PRML scheme allows intersymbol interference (ISI) to be removed from the read signal and the data to be optimally detected, provided a certain amount of SNR is available at the detectors [10]–[12]. The concept of PRML is based on the linearity of the read signal, in the sense that the signal as a whole must be the superposition of time-shifted and scaled versions of an individual signaling element. The time shifts must all be integer multiples of a basic unit interval T, which constitutes the clock period in the digital system. The scaling typically is as simple as a multiplication by one or by zero to indicate the presence or absence of the signaling element at a given point in time.

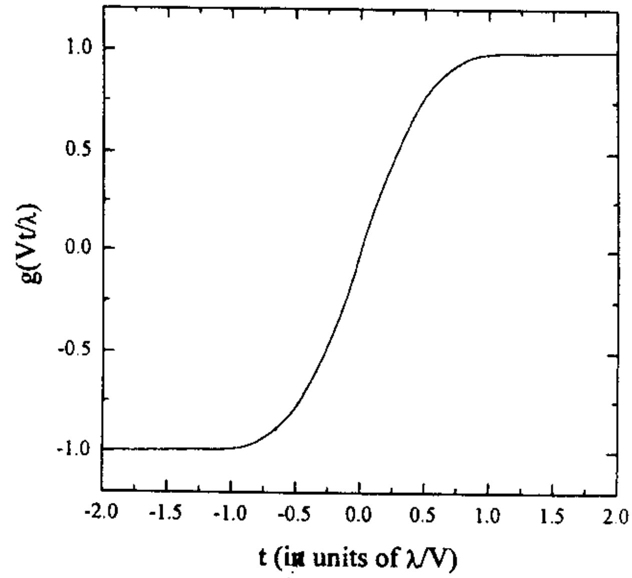
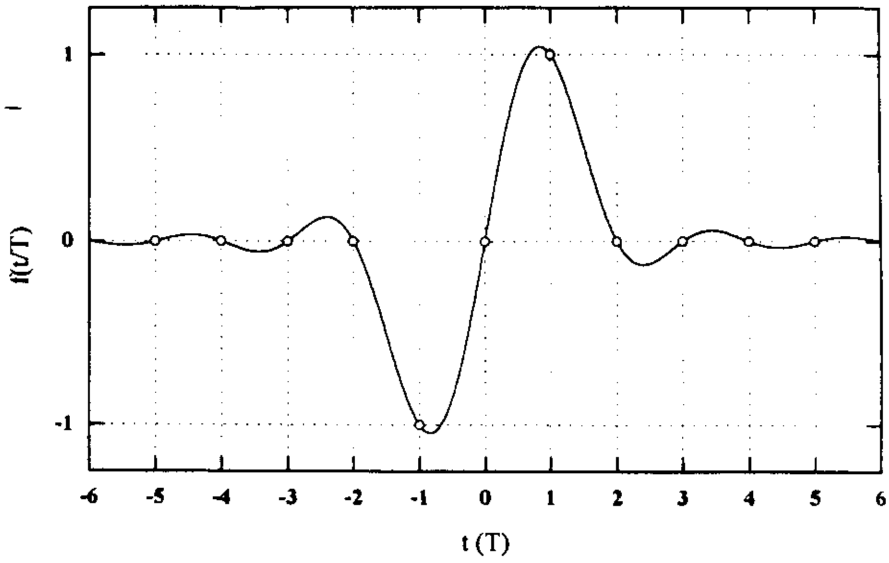
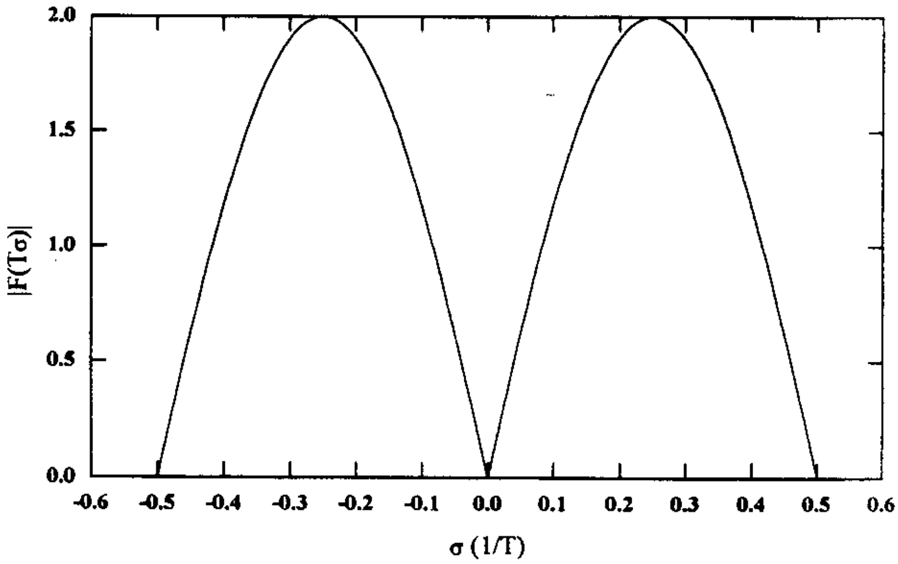


Fig. 8. Read signal obtained from a single mark-edge (i.e., a magnetic domain wall) recorded on an MO disk. The time scale is normalized by λ/V, where λ is the wavelength of the laser beam and V is the linear velocity of the disk moving under the focused spot. The lens used to read the data in this particular case had an NA of 0.55. As the focused spot moves along the track from a down-magnetized region to an up-magnetized region, the (normalized) readout signal goes from -1 to $+1$. The magnetic domain wall is assumed to be infinitely sharp, which is a good approximation to domain walls in TbFeCo media. The finite duration of the transition shown in the figure is due to the fact that the focused spot size is limited by diffraction effects to about one wavelength. When the track contains an arbitrary sequence of marks in the form of up- and down-magnetized domains, the complete readout signal is simply a superposition of shifted and scaled versions of the mark-edge signal shown in this figure. The shifting accounts for the fact that different edges arrive under the focused spot at different times, and the scaling in the form of multiplication by either $+1$ or -1 indicates whether a given edge is a leading edge or a trailing edge.

To recover the data from the read signal, one must use a linear electronic filter (often referred to as an equalizer) that converts the unit signaling element obtained from the disk into an electronic waveform that has limited ISI. For example, consider a magnetooptical disk system whose individual signaling element, the mark-edge signal, is depicted in Fig. 8. The linear filter might be one that converts this signal into the signal shown in Fig. 9(a). Because this output signal goes to zero at integer multiples of the clock period ($\pm 2T$, $\pm 3T$, ...), all ISI at times $2T$ and beyond is eliminated. The limited ISI corresponding to samples at $-T$ and T can then be optimally computed and canceled (using Viterbi's algorithm, for example). This is the essence of the PRML technique. One of the interesting features of



(a)



(b)

Fig. 9. One possible output signal from an electronic filter (i.e., equalizer) in response to the edge readout signal of Fig. 8. The waveform in (a) is the desired output signal in the time domain, while that in (b) is the Fourier transform of the output in the frequency domain. The shape of the output signal in the vicinity of the transition point (i.e., around $t = 0$) is similar to the input signal, but the output goes to zero at integer multiples of the clock period T for larger values of $|t|$. The fact that the edge-response signal in (a) is zero at sampling times $t = nT$ for $n = \pm 2, \pm 3$, etc. means that ISI is limited to only a few samples in the vicinity of any given transition. The waveform in (a) is constructed by superimposing a number of sinc functions ($\text{sinc}(t) = \sin \pi t / \pi t$). The Fourier transformed function in (b) shows that the equalizer is band limited and that it also blocks the DC content of the input signal.

the partial response equalizer is its frequency response. The Fourier transform of the output signal in Fig. 9(a) is shown in Fig. 9(b). Note that this signal is strictly band limited. The electronic filter, therefore, enhances certain midband frequencies but eliminates all the noise that is outside the range of useful signal frequencies.

The filtered edge-signal of Fig. 9 is but one possible example of signals that have limited ISI. Many filter designs are possible, and, in general, the output signal can be matched to the input signal in order to minimize the noise throughput of the filter. The output signal shown in Fig. 10, for instance, allows for more ISI than that in Fig. 9 because samples at $-2T, -T, T$, and $2T$ can now interfere. Consequently, this equalizer requires more computation to eliminate the ISI. As Fig. 11 indicates, however, the performance of this filter is superior to the previous one because it requires less SNR to achieve the same recording density. Fig. 11 is a plot of the required SNR versus the minimum mark length Δ that can be recorded on the disk. The performance of both filters I and II described above is shown and compared in this figure. Also indicated in the figure is a cutoff point for the curves; this is the minimum

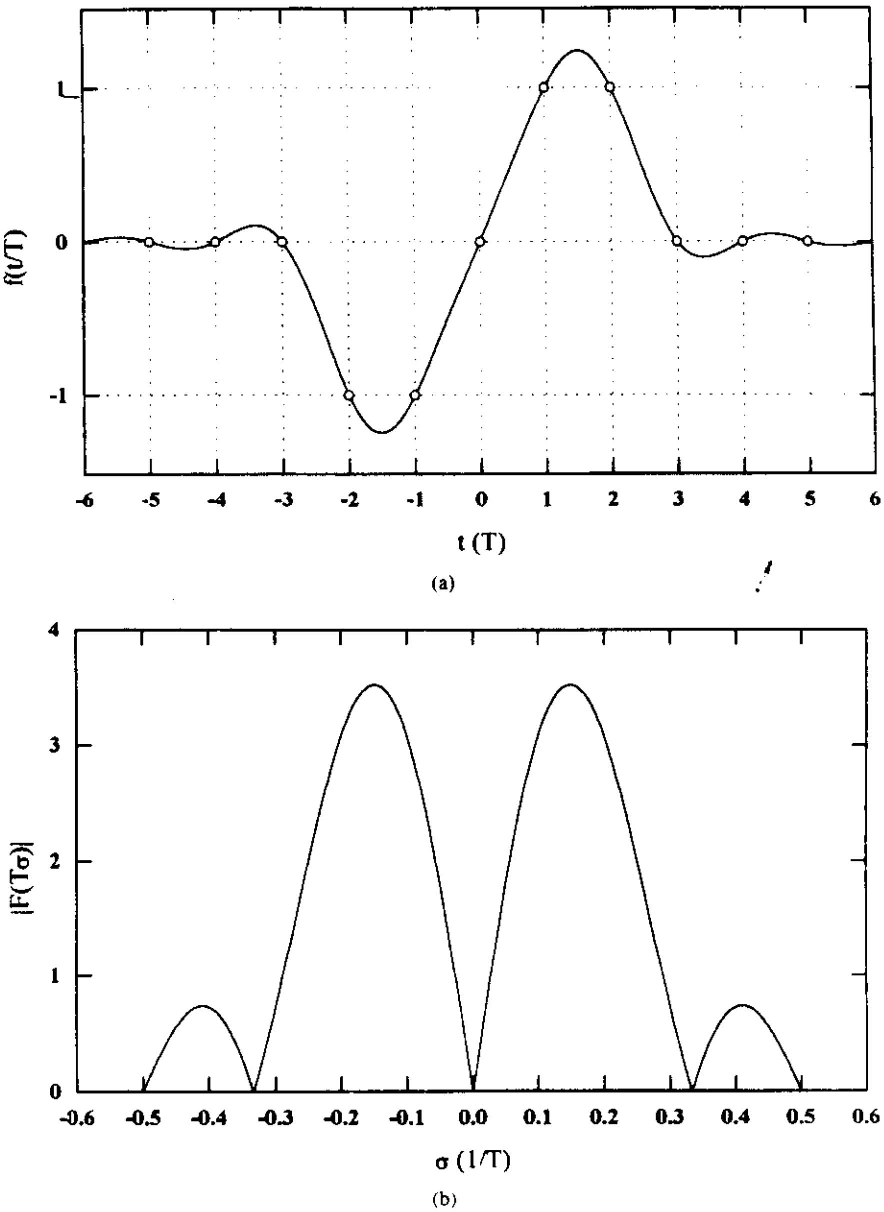


Fig. 10. Same as Fig. 9 except that the allowed range of ISI is now increased to cover samples at $\pm T$ and $\pm 2T$. This makes the Viterbi decoder for the present equalizer more complex than that for the equalizer of Fig. 9, but it improves the overall performance of the system by reducing its noise throughput.

mark length $\Delta = \lambda/(4\text{NA}) = 0.45\lambda$ that can possibly be detected with a linear system. The value of $\lambda/(4\text{NA})$ for the minimum mark length is the so-called optical cutoff of the readout system. If the recorded marks are any shorter than this cutoff value, no optical signal will return through the objective lens and, therefore, the electronic signal will be devoid of any information concerning the recorded data.

K. Land-Groove Recording

It has been found that by making the land and groove of equal width, and by recording the information on both lands and grooves, it is possible to eliminate (or at least substantially attenuate) the cross talk arising from the pattern of marks recorded on adjacent tracks [13]. Fig. 12(a) shows a typical pattern of recorded data on both lands and grooves. It turns out that for a particular groove depth (typically around $\lambda/6$), the cross talk from adjacent tracks reaches a minimum. Fig. 12(b) shows the computed cross-talk signal obtained from a theoretical model based on scalar diffraction theory. These results are in excellent agreement with the experimental data, indicating that cross-talk cancellation is a direct consequence of diffraction from the grooved surface and interference among the various diffracted orders.

Land-groove recording works well in phase-change media; in fact, this is where it was originally discovered. For MO systems, the birefringence of the substrate creates certain problems; specifically, the presence of in-plane birefringence will make the optimum groove depth for land size and low cost of magnetic heads allow them to use a separate head for each disk surface without compromising the overall size and price of the drive. In contrast, optical heads are very expensive and rather bulky. To achieve volumetric data storage with optical disks, designers have sought methods that rely on a single head for accessing multiple platters. In 1994, IBM researchers demonstrated a system that could read through a stack of six CD surfaces using a single optical head [15]. Their system was very similar to a conventional CD player except that they had taken special care to correct the spherical aberrations that result when the beam of light is focused through a substrate at different depths. The IBM stack consisted of three thin glass discs (thickness of ≈ 300 μm) on both surfaces of which the data pits had been embossed. Unlike standard CD's, these discs were not metallized because the focused laser beam had to pass through several such layers before reaching the desired surface. A bare glass surface reflects about 4% of the incident light, and this was apparently sufficient to enable the detection system to retrieve the data and to play back, with high fidelity, the recorded audio and video signals.

Technically, the method described above is straightforward and requires only that the objective lens be corrected for focusing through different thicknesses of the substrate. The separation between adjacent surfaces must be large enough to reduce cross talk from the data marks recorded on neighboring surfaces. With a 0.5-NA objective lens, a separation of 40 or 50 μm typically is enough to assure acceptable levels of cross talk from these surfaces.

In the case of writable media, recording and readout of information on multiple platters is more difficult, primarily because storage layers absorb a significant amount of the laser light. (In their 1994 demonstration, IBM researchers also described a four-layer WORM disk and a two-layer MO disk.) Recently, double-layer MO and PC disks have been proposed and demonstrated in several industrial laboratories around the world. Again, the separation between recording layers has been kept at about 40 μm to reduce cross talk, and the laser beam has been strong enough to write on the second layer even after half of its power has been absorbed by the first layer. (The power density at the first layer is reduced by a factor of almost 2000; hence, no writing occurs at that layer.) Since focusing through an additional 40 μm of plastic is well within the range of tolerance of typical objective lenses, correction for spherical aberration has not been necessary in these double-layer systems.

III. Holographic Storage

A. The Concept of Holographic Storage

Holographic storage is not a new technology [16]; its roots go back to the early 1960's, and over the years, several research organizations have explored its potential. Progress was hampered by a lack of key technologies such as compact lasers, spatial light modulators, detector arrays,

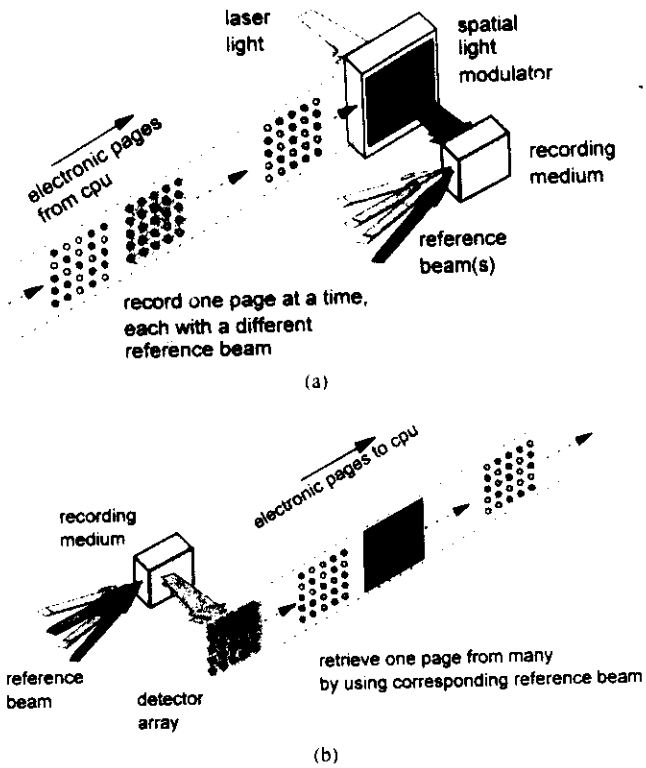


Fig. 15. Holographic storage recording and retrieval. (a) A digital pattern of ones and zeros, composed in the CPU, is used to address an SLM, converting the ones into bit locations that transmit light and the zeros into locations that do not transmit light. Light from the SLM is incident on a recording medium (optics not shown) and interferes with a reference beam derived from the same laser. Multiple pages are recorded in the same volume by altering the information on the SLM and using a different reference-beam angle for each new page. (b) Data are retrieved by illuminating the holograms with the reference-beam direction associated with the desired page. Reconstructed light is imaged onto a detector array (optics not shown), converting photons to electrons, and the signals are passed onto the thresholding, error correcting, and demodulation circuits.

and recording materials. As each of these subsystems became available, new efforts were mounted and eventually stopped for lack of another critical element. Today, we are seeing renewed interest in the technology and, to a large extent, most of the enabling technologies are in place, at least to the extent that serious product plans are being developed.

The underlying concept of holographic storage is much the same as in conventional holography. As shown in Fig. 15, an object is illuminated with a laser beam; in this case, the object is a spatial light modulator containing a pattern of ones and zeros. The light reflected or transmitted by this digital page composer falls upon a recording material, where it interferes with another beam of light (from the same laser) called the reference beam. The more intense regions of the interference pattern "expose" the material and locally alter its optical properties (e.g., absorption, refractive index, and/or thickness), forming a copy of the interference pattern. In conventional holography, processing may be used to create or augment this change in material properties; in holographic storage, the material usually self-

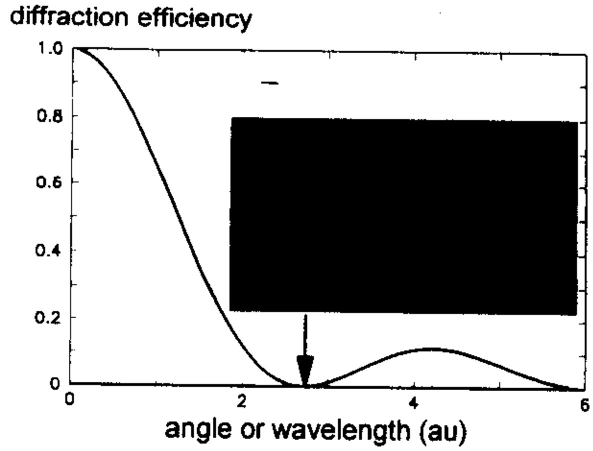


Fig. 16. The Bragg effect. Readout of a hologram at an angle (wavelength) different from the recording angle (wavelength) causes a decrease in the diffraction efficiency from the hologram. An angle (wavelength) exists where the diffraction efficiency is zero—a good location for recording another hologram with minimum cross talk. How closely the holograms can be placed in angle (wavelength) is directly proportional to the thickness of the recording medium, as shown in the inset.

processes *in situ*. The resulting structure, when illuminated with the reference beam alone, causes light to be diffracted so as to create a wavefront identical in all respects to the original wavefront containing the digital information that impinged upon the medium. This reconstructed wavefront is imaged onto a detector array such as a charge-coupled device (CCD), where it is sensed and the digital page retrieved.

If the recording material is sufficiently thick, multiple holograms can be recorded in the same volume by altering the angle between the two interfering beams—usually done by changing only the reference-beam direction. Diffraction from a thick hologram is governed by the Bragg effect and, as shown in Fig. 16, manifests itself as a rapid decrease in diffraction efficiency from a recorded hologram as the reference beam is brought into the medium at an angle different than that used during recording. For a given thickness, an angle can be found where the diffraction from one hologram is a minimum; at this new angle, another hologram can be recorded. Thousands of holograms can be recorded in a material 1 cm thick, thereby increasing the overall storage capacity and providing a means for fast access—in the above case, for example, the angle can be rapidly changed with an acoustooptic light deflector. Other means of storing multiple holograms, a process termed *multiplexing*, include changing the recording wavelength, changing the phase across the aperture of the reference beam, and translating or rotating the recording medium.

B. The Light Source

The light source for holography must have sufficient spatial and temporal coherence to allow the formation of an interference pattern over the desired volume of space (e.g., throughout the recording medium) and to keep this pattern stationary during the exposure time. With few exceptions, the source is a laser. In general, the laser must have sufficient power at the required wavelength (typically on the order of a few hundred milliwatts CW at the medium), be small enough to fit into a reasonable-sized system, and, of course, be cheap. An argon-ion laser satisfies the first two requirements but falls way short of the size and cost need. A reasonable choice is second harmonic generation of 530-nm light from a diode-pumped Nd:YAG laser. A better choice would be a near-infrared laser diode. The ultimate choice will depend on the wavelength sensitivity of the recording material.

C. The Spatial Light Modulator

The digital information must be organized into a page-like format of ones and zeros and subsequently modulated onto the object beam as a two-dimensional pattern of brightness and darkness. The analogy is an array of miniature shutters that are either open or closed. A more practical implementation is a device called a spatial light modulator (SLM). While these devices have been fabricated from many different technologies (electrooptic, magnetooptic, deformographic, etc.), the most common implementation is with liquid crystals. Having their genesis as watches, hand-held televisions, and computer displays, considerable engineering has gone into increasing the performance of these devices and driving the cost down through mass markets. An SLM for holographic storage need only be binary and does not require color; it does, however, require high contrast and rapid switching between the on and off states. Contrast levels of 5:1 (on a pixel by pixel basis) appear to be acceptable, and while frame rates as high as 1000 frames/s will ultimately be required, the recording speed of current holographic materials is such that frame rates of a few hundred frames per second are adequate. Since mainstream SLM applications only require video frame rates, the current state of the art and low cost cannot be fully leveraged. The pathways to higher speed are there, prototypes are available, additional engineering will be required, and it would be nice to have other applications such as optical computing to pull the cost down.

SLM's containing 640 $\times$ 480 pels are readily available, and prototype devices with 1024 $\times$ 1024 are starting to appear. Typical pel sizes are in the 30–50 μm range for the former and in the 15–20 μm range for the latter.

D. The Detector Array

The other end of the optical data channel is a detector array that receives the reconstructed image from the hologram and creates the signals that enter the thresholding, error-correcting, and demodulation circuits. Typically a CCD, there are several design parameters that directly affect the architecture of the overall system. These include quantum efficiency of detection at the design wavelength, number and size of the individual detector elements, and data rate. Most commercially available CCD's drop off in quantum efficiency rapidly at longer wavelengths—many have no sensitivity beyond 700 nm—affecting the choice

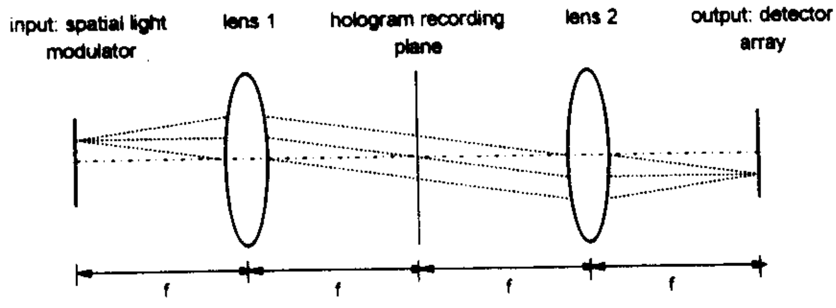


Fig. 17. Recording a Fourier transform hologram. Identical lenses, placed in a 4f configuration, convert the light distribution in an input plane (such as on an SLM) into its Fourier transform in the central (recording) plane, then invert the process, producing an image of the input plane on the output (detector) plane.

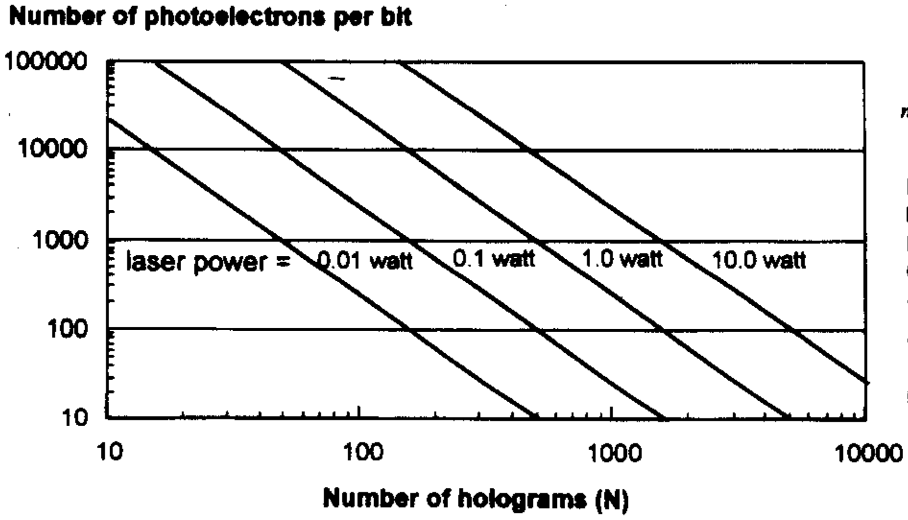
of recording material and laser. Data rates approaching 1.0 Gb/s are achieved by segmenting the CCD into subareas and reading them out in parallel. As many as 64 channels are in the prototype stage. One of the most difficult design parameters to resolve is the size of an individual detector element. For several reasons, the optimum holographic recording and reconstruction system is a 4f system of unit magnification, as shown in Fig. 17. The hologram, recorded in the central plane, is the Fourier transform of the SLM and enjoys some degree of translational invariance—important for accessing schemes that require mechanical motion and for removability. In addition, a system with unit magnification (all lens focal lengths are the same) is less sensitive to lens aberrations. These considerations essentially require that the SLM pixel and the detector pixel be the same size, and that 1:1 pixel matching occurs. That is, the SLM is imaged through the recording medium onto the detector such that each SLM pixel and corresponding detector pixel is a separate channel. As can be imagined, alignment is exceptionally critical. Unfortunately, the trend in CCD fabrication is to use smaller and smaller pixels; silicon real estate is expensive, and SLM pixels have a lower limit in size because the cell containing the liquid crystal material can only be made so thin. CCD's are at 9 μm and getting smaller, and SLM's are limited to about 15 μm. One solution is to demagnify the SLM image to preserve the pixel matching; the other is to oversample the reconstructed image and do some degree of image processing to recover the data. The former has cost- and form-factor implications, the latter will affect the data rate. Both methods are under active investigation.

E. Photon Budget and Data Reliability

The goal in any storage system is to cram the maximum amount of data into the smallest possible space. If the data can be retrieved with no errors (i.e., the raw error rate before applying error correction and demodulation decoding is zero), the general feeling is that the system design is not aggressive enough. A properly designed demodulation and error-correcting code can correct a 10^{-5} raw bit error rate (BER) to the 10^{-12} level. Error sources in data recovery will have some combination of Rician (optical) and Gaussian (electrical) statistics, and will require an SNR between seven and nine to achieve a 10^{-5} raw BER [17]. Since there are only so many photons to go around, the design must balance the required SNR with the number of bits that a reference beam will interact with at one time. In other words, the number of photons in a reconstructed bit is a function of the number of bits in a page (C) and the number of holograms (N) multiplexed in the same volume. Unfortunately, the diffraction efficiency of a single hologram contained within a set of N multiplexed holograms varies as $1/N^2$—hence, there are 10^6 times fewer photons per bit from a volume containing 1000 holograms compared to a single hologram. The functional dependency on these and other parameters is shown in Fig. 18. Multiplexing 500 hologram pages, each containing 1 Mb of information (each page coded to contain half ones and half zeros), and illuminating with a 1-W laser would provide 1000 photoelectrons for detection in 1 ms. This is probably the minimum needed for a 10^{-5} BER and an instantaneous data rate of 1 Gb/s. Variations in other parameters such as quantum efficiency, media figure of merit, fixing efficiency, etc. will alter this number slightly.

F. The Recording Material

If holographic storage has had an Achilles heel over the years, it has been the recording material. Certainly, many successful holographic materials have been developed, but the requirement for self-processing and thickness greatly reduces the number of choices. Most common are inorganic photorefractive materials such as lithium niobate ($LiNbO_3$), strontium barium niobate (SBN), and barium titanate ($BaTiO_3$). In the photorefractive effect, the bright regions of the exposing interference pattern excite electrons into the conduction band that drift into the dark regions and become trapped. This redistribution of charge sets up local space-charge fields that interact with the host material to alter its index of refraction through the electrooptic effect. The result is a phase hologram that efficiently diffracts light. Diffraction efficiency of 100% is not uncommon, as these materials can easily be fabricated in thicknesses of a centimeter—which also enhances their ability to multiplex many holograms in the same volume. This self-processing is, however, its own downfall. The light from a subsequent exposure of another hologram partially erases the first



$$n_{photoelectrons} = P_{inc}\left(\frac{M^2}{N^2}\right)\eta_{fixing}\left(\frac{2}{C}\right)\Delta t_i\ \phi\left(\frac{1}{h\nu}\right)$$

P_{inc}	= incident laser power
M	= media figure of merit (0.5)
N	= number of holograms
C	= capacity per page (10^6)
η_{fixing}	= fixing efficiency (0.5)
Δt	= detector integration time (1 msec)
ϕ	= detector quantum efficiency (0.3)
$h\nu$	= photon energy (3×10^{-19} @ 663 nm)

Fig. 18. Detecting a signal—the photon budget. The number of photoelectrons generated by a reconstructed "one" is a function of the laser power incident on the medium, the number of bits in a page (50% "ones," 50% "zeros" assumed), the number of holograms in a common volume, a media figure of merit that is a measure of the maximum diffraction efficiency for a single hologram, any loss in diffraction efficiency resulting from the fixing step, the integration time of the detector, the quantum efficiency of the detector, and the photon energy.

hologram, a third hologram exposure partially erases the first two, and so on. As a result, the efficiency of a single hologram falls off as $1/N^2$ rather than $1/N$. In addition, readout of a single hologram causes partial erasure of all the holograms in that same volume by the same photorefractive effect. The quest therefore has been to develop a fixing process to render the holograms insensitive to subsequent illumination. One such process is to use thermal fixing. The recorded holograms are elevated in temperature, and the electronic space charges interact with ions that are more mobile because of the higher temperature. The ions move to compensate for the space charge and become locked in place when the temperature is reduced. They are subsequently revealed and create their own index variations when the electronic space charge is erased with a second illumination. This process works quite well with only a small loss in diffraction efficiency.

A more elegant solution is to use a *gating* process during exposure of the hologram; i.e., two stimuli are required to be present for the creation of the hologram, and only one is required for readout. For example, a two-color process has been developed [18] in which the material is uniformly pumped to one level of traps with one wavelength (usually a long wavelength that need not be from a laser) and the hologram is formed with a longer wavelength, such as the near infrared that selectively excites some of these electrons into the conduction band. Without the pumping light, no hologram is formed; hence, the readout process is nondestructive. Erasure can be accomplished with uniform exposure of both beams. This process has been known for some time but generally required high-power pulsed lasers. The new news is that it can now be accomplished with modest CW power.

Organic photopolymeric materials also exist that permit hologram formation in real time without processing. These materials are attractive from an ease-of-fabrication viewpoint but currently suffer from two major drawbacks: as yet, they cannot be fabricated to a thickness greater than 100 μm (hence, the number of holograms that can be multiplexed is reduced) and they undergo about 5% shrinkage with exposure, which complicates retrieval of a multiplexed hologram and leads to cross-talk noise. Research into these and other polymeric materials continues, however, because of their inherent advantages over grown and polished inorganic crystals [19].

IV. Summary

Optical storage has made a lot of progress since its first introduction as laser video disc in the late 1970's. Granted, its growth has not been on the explosive 60% per year slope that magnetic storage has enjoyed over the last few years; but then, we should not expect it to be so. Removability, backward compatibility, and interchangeability carry with them a demanding burden, called *standards*, that must be developed and agreed to by the entire industry. Customers demand this, as they do not want a repeat of the VHS/Betamax situation of a few years ago. In contrast, the only standard imposed on the magnetic storage industry is the interface; the media and recording technology are captive within the drive, thus permitting tremendous freedom and competition.

Progress has been substantial, however, with CD technology jumping from 680 Mbytes to 4.7 Gbytes with the introduction of DVD this year. 3.5″ removable MO has steadily grown from 128 Mbytes to 640 Mbytes, and 5.25″ MO now contains 4.6 Gbytes on a single platter. These advances have come through a combination of laser wavelength reduction, objective lens NA increases, better ISI and cross-talk management, and coding improvements.

There is room for even greater advances in storage capacity as we make the transition to blue lasers, near-field optical recording, and multilayer systems; 50–100 times storage capacities are not unreasonable to expect.

Holographic storage has yet to get its foot in the door, but every indication is that it will make it this time. It will probably appear as a minilibrary containing several hundred gigabytes that is ideally suited for image storage, particularly if data rates approaching 1 Gb/s can be realized. The first material will be the old standard, $LiNbO_3$, operated in both read-only and write-once modes. Erasable materials will be down the road a bit, as will be their organic cousins.

REFERENCES

[1] *Tech. Dig. Int. Symp. Optical Memory (ISOM)*, Japan Society of Applied Physics, Aug. 30–Sept. 1, 1995.

[2] *Tech. Dig. Int. Symp. Optical Memory and Optical Data Storage*, Optical Society of America, Maui, HI, July 8–12, 1996.

[3] *Proc. Magneto-Optical Recording Int. Symp. (MORIS)*, vol. 20, supp. S1, 1996.

[4] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. Van Rosmalen, and K. S. Immink, *Principles of Optical Disk Systems*. Bristol, U.K.: Adam Hilger, 1985.

[5] A. B. Marchant, *Optical Recording*. Reading, MA: Addison-Wesley, 1990.

[6] M. Mansuripur, *The Physical Principles of Magneto-optical Recording*. London: Cambridge Univ. Press, 1995.

[7] T. W. McDaniel and R. H. Victora, Eds., *Handbook of Magneto-Optical Data Recording*. Westwood, NJ: Noyes Publications, 1997.

[8] S. Ohara, N. Akahira, and T. Ishida, "High density recording technology on phase change disk systems," in *Tech. Dig. Int. Symp. Optical Memory and Optical Data Storage*, Optical Society of America, Maui, HI, 1996, pp. 32–34.

[9] G. S. Kino, "Ultra high density recording using a solid immersion lens," in *Tech. Dig. Optical Data Storage Conference*, Optical Society of America, San Diego, CA, 1995, pp. 176–181.

[10] P. Kabal and S. Pasupathy, "Partial-response signaling," *IEEE Trans. Commun.*, vol. COM-23, pp. 921–934, 1975.

[11] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, 1972.

[12] L. Cheng, M. Mansuripur, and D. G. Howe, "Partial-response equalization in magneto-optical disk readout: A theoretical investigation," *Appl. Opt.*, vol. 34, pp. 5153–5166, 1995.

[13] A. Fukumoto, S. Masuhara, and K. Aratani, "Cross-talk analysis of land/groove magneto-optical recording," in *Proc. Soc. Photo-optical Instrumentation Engineers*, vol. 2429, pp. 41–42, 1994.

[14] M. Kaneko and A. Nakaoki, "Recent progress in magnetically induced super-resolution," in *Proc. Magneto-Optical Recording Int. Symp. (MORIS)*, vol. 20, supp. S1, 1996, pp. 7–12.

[15] K. Rubin, H. Rosen, T. Strand, W. Imaino, and W. Tang, "Multilayer volumetric storage," in *Tech. Dig. Optical Data Storage Conf.*, Optical Society of America, Dana Point, CA, 1994, pp. 247–253.

[16] G. Sincerbox, Ed., *Selected Papers in Holographic Storage*. Bellingham, WA: SPIE Press, 1994.

[17] C. Ju, F. Dai, and J. Hong, *Electron. Lett.*, vol. 32, no. 15, p. 1400, 1996.

[18] S. Orlov, A. Akella, L. Hesselink, and R. Neurgaonkar, "High sensitivity two-color non-volatile recording in lithium niobate," CLEO 1997 postdeadline paper.

[19] R. Wortmann, P. Lundquist, R. Tweig, C. Geletneky, C. Moylan, Y. Jia, R. Devoe, D. Burland, M.-P. Bernal, H. Coufal, R. Grygier, J. Hoffnagle, C. Jefferson, R. Macfarlane, R. Shelby, and G. Sincerbox, "A novel sensitized photochromic organic glass for holographic optical storage," *Appl. Phys. Lett.*, vol. 69, no. 12, pp. 1657–1659, 1996.

Masud Mansuripur received the Ph.D. degree in electrical engineering from Stanford University, CA, in 1981.

He is a Professor of Optical Sciences at the University of Arizona, Tucson. His areas of research have included magnetooptical disk data storage, information theory, micromagnetic simulations, optics of birefringent media, and the theory of diffraction. He is the author of *Introduction to Information Theory* (Englewood Cliffs, NJ: Prentice-Hall, 1987) and *The Physical Principles of Magneto-Optical Recording* (Cambridge: Cambridge Univ. Press, 1995). In addition, he has published more than 100 papers in scientific journals and given numerous technical presentations at international conferences and industrial laboratories.

Glenn Sincerbox received the B.S. degree from Rensselaer Polytechnic Institute, Troy, NY, and the M.S. degree from the University of Illinois, Urbana, both in physics.

He is a Professor of Optical Sciences at the University of Arizona, Tucson, and Director of the Optical Data Storage Center. During his 34-year tenure at IBM, he contributed to many areas of technology, including holographic bar-code scanners, printing and display devices, optical and holographic storage, and manufacturing inspection. He is the author of more than 40 publications, including three book chapters and two edited volumes and has given numerous presentations at international conferences. He has received 40 patents and holds 70 patent publications. In addition, he is very active in various professional societies, conference committees, and government advisory-group activities.